Collibra

Collibra Data Intelligence Cloud

Platform Configuration

Version 2022.06

Collibra

Collibra Data Intelligence Cloud - Platform Configuration

Release date: June 5th, 2022

Revision date: Thu Jun 02, 2022

You can find the most up-to-date technical documentation on our Documentation Center at

https://productresources.collibra.com/docs/collibra/latest/Content/Console/co_console.htm

Contents

Contents 1

DGC service configuration 1

- Configure Collibra email settings 4
- Edit hyperlink settings 8
- Configure the recommenders and matchers 11
- Customizing the search index 16
- Edit import configuration 49
- Edit Excel export configuration 50
- Edit CSV export configuration 51
- Configure the logging of Collibra DGC API calls 53
- Security configuration 56
- Configure Collibra Connect 96
- Edit global data source registration settings 98
- Add a Jobserver to the DGC service 101
- Configure data profiling behavior 104
- Configure Automatic Data Classification 107
- Enable or disable Catalog experience 109
- Enable registering a data source via Edge 110
- Enable profiling and classification via Edge 111
- Enable data quality synchronization via Edge 112
- Enable Tableau metadata API 114
- Anonymize data via Jobserver 115
- Enable or disable the Settings landing page 118

Configure the Search service ... 120
Search service configuration options ... 121
Jobserver service configuration ... 122
Jobserver authentication levels ... 123
Mutual authentication between Jobserver and DGC service ... 124
General specifications for certificates and private keys ... 126
Edit the Jobserver service settings ... 127
Connection from an on-premises Jobserver to a Collibra Data Intelligence Cloud ... 130
Connection from Collibra Data Intelligence Cloud to an on-premises Jobserver ... 147
Jobserver best practices ... 152
Service infrastructure configurations ... 156
Edit the DGC service infrastructure settings ... 157
Edit the Search service infrastructure settings ... 165
Managing environments ... 166
Create a Collibra environment ... 168
Environment statuses ... 169
Start an environment ... 170
Stop an environment ... 171
Delete an environment ... 171
Start a service ... 172
Stop a service ... 172
Reindexing Collibra Data Intelligence Cloud ... 173
Restore to factory defaults ... 173
Back up and restore ... 175
Create a backup of Collibra Data Intelligence Cloud 2022.06 ... 177
Backup options ... 179

The Backups page 181
Download a backup 182
Upload a backup 183
Delete a backup 184
Create a backup schedule 185
Backup schedule options 188
Backup schedules overview 190
Edit a backup schedule 192
Delete a backup schedule 192
Restoring a backup 193
Customize Collibra DGC with a backup restore 202
Back up and download with the REST API 202
REST API - List of backups 206
REST API - Delete a backup 207
Diagnostic files 209
Diagnostic files 210
Create a diagnostic file 212
Download a diagnostic file 214
Delete a diagnostic file 215
Edit the environment log settings 216
Logging 217
Content of a diagnostic file 219
Collibra Console settings 222
Collibra Console users 223
Edit the Collibra Console settings 227
Edit the Collibra Console server settings 253

Open a Collibra Console log file 259
Troubleshooting 261
DGC service fails to start due to invalid configuration 262
Finding resource IDs 262
DGC service configuration: options i
General settings i
Email configuration iii
Hyperlinking configuration vi
Recommender configuration vii
Search index configuration x
Upload configuration (*) xiii
Statistics configuration xiv
Import configuration xv
Excel export configuration xvi
CSV export configuration xvii
User interface configuration (*) xvii
API call logging xviii
System metrics xix
Security configuration xix
Workflow engine configuration (*) xli
Collibra Connect xlii
Register data source xlii
Jobserver (*) xliii
Data profiling xliv
Beta features xlvi
Throttling (*) xlvi

Hibernate cache configuration (*) xlviii
Graph query lii
Table liv
Purge configuration lv
Cloud Data Classification configuration lvi
Reporting lvi
Catalog Experience lviii
Diagrams lviii
Everywhere Desktop configuration lix
Everywhere Mobile configuration lxii
Everywhere Chrome Extension lxiii
Edge lxiii
Tableau Metadata API lxiv
Backup configuration management lxiv
Job Service (Activities) lxv
Appendix B - Spring Cron syntax lxvi
Special characters lxvii
Quartz Cron syntax lxxi
Special characters lxxiii

DGC service configuration

In Collibra Console, you can find all the Collibra Data Intelligence Cloud environment settings. After the installation of Collibra Data Intelligence Cloud, a default configuration is applied. If you have changed some settings, you can always go back to the original configuration by restoring the factory defaults. It is highly recommended to review the default settings and change them if necessary.

Disable or enable view permissions

Collibra Data Intelligence Cloud has a **View permissions** option that enables you to set permissions on objects. This option is enabled by default. When you disable the option, all objects are visible for every user.

> Note Depending on your environment, you have to follow this procedure either in the Services Configuration section of the Collibra settings or in Collibra Console. This topic shows the procedure in Collibra Console. If your Settings page has a **Configuration** tab, you can follow this procedure in that tab page. For more information, see the online documentation.

Prerequisites

- You have the ADMIN or SUPER role in Collibra Console.
- You have the SUPER role in Collibra Console.

Steps

1. Open the DGC service settings for editing:
    a. Open Collibra Console.
        » Collibra Console opens with the **Infrastructure** page.
    b. In the tab pane, expand an environment to show its services.

c. In the tab pane, click the Data Governance Center service of that environment.
d. Click **Configuration**.
e. Click **Edit configuration**.

2. In the **General settings** section, set the **Enable view rights** to **False** (disabled) or **True** (enabled).
3. Click the green **Save all** button.

Edit the Help menu

In Collibra Data Intelligence Cloud, you have a Help menu to access the product documentation and various Collibra sites.

> Note Depending on your environment, you have to follow this procedure either in the Services Configuration section of the Collibra settings or in Collibra Console. This topic shows the procedure in Collibra Console. If your Settings page has a **Configuration** tab, you can follow this procedure in that tab page. For more information, see the online documentation.

Prerequisites

- You have the ADMIN or SUPER role in Collibra Console.
- You have the SUPER role in Collibra Console.

Steps

1. Open the DGC service settings for editing:
   a. Open Collibra Console.
      » Collibra Console opens with the **Infrastructure** page.
   b. In the tab pane, expand an environment to show its services.
   c. In the tab pane, click the Data Governance Center service of that environment.
   d. Click **Configuration**.
   e. Click **Edit configuration**.

2. In the **Help menu** section, do one of the following:

<table>
<tr><th>Action</th><th>Description</th></tr>
<tr><td>Add a menu item</td><td>a. At the bottom of the section, click Add.
» A new link section appears.
b. Enter the required information:<table><tr><th>Setting</th><th>Description</th></tr><tr><td>Links</td><td>The list of links in the help menu.</td></tr><tr><td>Menu item name</td><td>The name of the menu item as it will appear in Collibra Data Intelligence Cloud's help menu.</td></tr><tr><td>Menu index</td><td>The position of the menu item in the help menu. The top position starts with the value 1.</td></tr><tr><td>Menu URL</td><td>The target URL of the menu item.</td></tr><tr><td>Show admin only</td><td>■ ✓ True: The menu item is only visible to users with the Sysadmin role.
■ ✕ False: The menu item is visible to every user.</td></tr></table></td></tr>
<tr><td>Edit an existing menu item</td><td>Make the necessary changes in the relevant link section.</td></tr>
<tr><td>Remove a menu item</td><td>In the upper-right corner of the menu section, click 🗑.</td></tr>
</table>

   - Add a new menu item:
   - Edit an existing menu item by making the necessary changes in the relevant link section.
   - Remove a menu item by clicking 🗑 in the upper-right corner of the menu section.
3. Click the green **Save all** button.

4. Refresh the Collibra page to see the changes.

Configure Collibra email settings

You can use notification emails to notify users of any changes made to assets. In the email settings, you can:

- Configure your email server settings.
- Specify which roles will receive emails on which days of the week.
- Enable monthly summary emails and specify which roles will receive them.
- Configure how emails are handled.

> Note Depending on your environment, you have to follow this procedure either in the Services Configuration section of the Collibra settings or in Collibra Console. This topic shows the procedure in Collibra Console. If your Settings page has a **Configuration** tab, you can follow this procedure in that tab page. For more information, see the online documentation.

> Note Collibra Console and Collibra Data Intelligence Cloud use different email services. This topic describes the email configuration of Collibra Data Intelligence Cloud. The configuration of the Collibra Console email settings is described in Configure Collibra Console email settings.

Prerequisites

- You have the ADMIN or SUPER role in Collibra Console.
- You have the SUPER role in Collibra Console.

Steps

1. Open the DGC service settings for editing:
    a. Open Collibra Console.
        » Collibra Console opens with the **Infrastructure** page.
    b. In the tab pane, expand an environment to show its services.
    c. In the tab pane, click the Data Governance Center service of that environment.

d. Click **Configuration**.
e. Click **Edit configuration**.

2. In the **Email configuration** section, make the necessary changes.

Setting	Description
Default schedule (Requires restart)	The Cron schedule to send emails only at specific times. With this, you can send emails in batches and avoid an overload of mails. Keep in mind that these emails are only workflow emails and have nothing to do with the notification schedule. If you create an invalid Cron pattern, Collibra Data Intelligence Cloud stops responding.
Template map	The location of template emails.
Password (*)	The password paired with your username to sign in to your SMTP server.
From address (*)	The email address used as the sender of all outgoing emails. Contact Collibra support to change the From address, see also Email configuration.
Port (*)	The port to connect to your SMTP server. The default value is **25**.
Host (*)	The hostname or URL of your SMTP server.
Start TLS (*)	◦ ✓ True: Use TLS (Transport Layer Security) to connect to your SMTP server. ◦ ✕ False (default): Do not use TLS to connect to your SMTP server.
Username (*)	The username to sign in to your SMTP server.

Setting	Description
Sending threads (*)	The number of threads that are used to send emails. The default value is **3**.
Max retries (*)	The maximum number of retries before the system aborts the sending of an email. The default value is **5**.

> Note
> To edit the DGC service configuration, you need the ADMIN role, except for settings or sections marked with (*), which require the SUPER role. As the SUPER role does not exist for cloud environments, settings that require the SUPER role are not available in Collibra Data Intelligence Cloud environments. If you want to edit one of these settings, please create a support ticket.

3. In the **Notifications** section, make the necessary changes.

Setting	Description
Notification days	The days of the week on which Collibra sends notifications. The days are represented by numbers from 1 to 7, where 1 represents Sunday. Per row you can add one day.
Daily roles	The roles that receive notifications on the days defined in **Notification days**.
Enable monthly notifications	◦ ✓ True: The users receive a monthly summary. ◦ ✗ False (default): The users do not receive a monthly summary.
Roles for monthly notifications	The roles that receive monthly notification emails. This is only relevant if **Enable monthly notifications** is ✓ True.

4. In the **Handlers** section, make the necessary changes.

Setting	Description
Host (*)	The hostname or URL of the incoming mail server.
Port (*)	The port to connect to your incoming mail server.
Protocol (*) 1.	The protocol to connect to your incoming mail server, with or without SSL (*POP3*, *POP3S*, *IMAP*, *IMAPS*). > Note The additional S at the end of the abbreviations stands for the secure version of the protocol using SSL. Using this requires the SSL certificates to be correctly configured.
Force domain (*)	◦ ✔ True: Only handle emails from the same domain as the handler's email address. ◦ ✖ False (default): Handle emails from any domain.
Handler list (*)	The configuration of email handlers, which can poll emails on an email server, process those emails and perform actions based on the contents.
Enabled (*)	◦ ✔ True: The handler is enabled. ◦ ✖ False (default): The handler is not enabled.
Name (*)	The name of the mail handler. We recommend to use a meaningful name to easily identify what this handler is used for.
Username (*)	The username to connect to the incoming mail server.
Password (*)	The password to connect to the incoming mail server.
Email address (*)	The email address to which workflow action mails are sent.

Setting	Description
Polling interval (*)	The time in milliseconds between two pollings of the mail server.
Delete (*)	◦ ✔ True: Delete messages from the mail server once the mail is processed. ◦ ✖ False (default): Keep messages on the mail server after the mail is processed. This option is only relevant if **Protocol** is *IMAP* or *IMAPS*.
Alias filter (*)	◦ ✔ True (default): Retrieve only the emails of which the **To** field contains the email address of the handler. ◦ ✖ False: Do not filter on the **To** field.

> Note
> To edit the DGC service configuration, you need the ADMIN role, except for settings or sections marked with (*), which require the SUPER role. As the SUPER role does not exist for cloud environments, settings that require the SUPER role are not available in Collibra Data Intelligence Cloud environments. If you want to edit one of these settings, please create a support ticket.

5. Click the green **Save all** button.

What's next?

Restart the environment to apply your changes. See Stop an environment and Start an environment.

Edit hyperlink settings

In Collibra Data Intelligence Cloud, each asset is a potential target of an automatically created hyperlink. If a text attribute contains the name of another asset, it is automatically converted to a link to that asset.

> Note The text editor on asset pages and domain pages includes an "Exclude from autohyperlinking" button, to disable autohyperlinking for text attributes.

> Note Depending on your environment, you have to follow this procedure either in the Services Configuration section of the Collibra settings or in Collibra Console. This topic shows the procedure in Collibra Console. If your Settings page has a **Configuration** tab, you can follow this procedure in that tab page. For more information, see the online documentation.

Prerequisites

- You have the ADMIN or SUPER role in Collibra Console.
- You have the SUPER role in Collibra Console.

Steps

1. Open the DGC service settings for editing:
    a. Open Collibra Console.
        » Collibra Console opens with the **Infrastructure** page.
    b. In the tab pane, expand an environment to show its services.
    c. In the tab pane, click the Data Governance Center service of that environment.
    d. Click **Configuration**.
    e. Click **Edit configuration**.

2. In the **Hyperlinking configuration** section, make the necessary changes.

Setting	Description
Enable hyper-linking	◦ ✔ True: Hyperlinks are created automatically. ◦ ✖ False (default): Hyperlinks are not created automatically. For more information about automatic hyperlinks, see Hyperlinking. Warning If you enable this setting, the performance of Collibra can decrease.
Enable case sensitivity	◦ ✔ True: Hyperlinks are case-sensitive. ◦ ✖ False (default): Hyperlinks are not case-sensitive. Note If you edit this setting, you have to reindex Collibra.
Batch size	The maximum number of text attributes that can be processed in a single transaction. The default value is *50,000*. The higher the value, the more memory used to create the hyperlinks, which happens each time you edit an asset or import data. The lower the value, the more time required for this process. Note To edit this setting, you need the SUPER role.

Setting	Description
Excluded asset type IDs	The list of asset types that are ignored by automatic hyperlinking. You can enter multiple asset type IDs, separated by commas. Excluding assets reduces the amount of hyperlinks, which improves performance. > Tip We recommend that you exclude technical asset types such as Column, Field, Table, Code Value and Code Set. > Note If you edit this setting, you have to reindex Collibra.

3. Click the green **Save all** button.
4. Reindex Collibra Data Intelligence Cloud.

Configure the recommenders and matchers

Collibra Data Intelligence Cloud contains recommenders and matchers that recommend data sets or business assets.

You can configure them to optimize the recommendations.

> Note Depending on your environment, you have to follow this procedure either in the Services Configuration section of the Collibra settings or in Collibra Console. This topic shows the procedure in Collibra Console. If your Settings page has a **Configuration** tab, you can follow this procedure in that tab page. For more information, see the online documentation.

Prerequisites

- You have the ADMIN or SUPER role in Collibra Console.
- You have the SUPER role in Collibra Console.

Steps

1. Open the DGC service settings for editing:
   a. Open Collibra Console.
      » Collibra Console opens with the **Infrastructure** page.
   b. In the tab pane, expand an environment to show its services.
   c. In the tab pane, click the Data Governance Center service of that environment.
   d. Click **Configuration**.
   e. Click **Edit configuration**.
2. In the **Recommender configuration** section, make the necessary changes.

Setting	impacts	Description
Catalog recommender enabled	All recommendations	◦ ✔ True (default): The "Data sets you might like" section is included on the Data Catalog Home page. This section shows data sets you might be interested in, as determined by the recommender, which takes into account your data sets and the data sets of similar users. ◦ ✖ False: The "Data sets you might like" section is not included on the Data Catalog Home page.
Data set recommender execution time	Recommendations of data sets to users	The schedule (CRON job) by which the data set recommender looks for recommended data sets for a user. By default the data set recommender does this every night.
Asset recommender execution time	Recommendations of business assets to data assets	The schedule (CRON job) by which the asset recommender looks for suggested relations between business assets and data sets.

Setting	impacts	Description
Data set matcher execution time	Data set matcher	The schedule (CRON job) by which the data set matcher looks for similar data sets.
Data set similarity threshold	Data set matcher	The amount of business assets that have to be related to two data sets before the data sets are considered to be similar. This percentage is expressed by a decimal where 1,00 equals 100%. Example If this value is 0.3 and at least 30% of the related business assets are related to both data sets, they are considered to be similar.
Duplicate schema threshold	Schema matcher	The amount of assets that have to be related to both schemas before the schemas are considered to be similar. This percentage is expressed by a decimal where 1,00 equals 100%.

Setting	impacts	Description
Fuzzy vs exact matching strategy for business assets	Recommendations of business assets to data sets and of business assets to column assets	The percentage that determines to what extent assets with a similar name become more important. The ranking in the search engine results always has an impact on the suggestion score. However, similarity between the asset names can also be taken into account. If you decrease this percentage, the ranking of the search results becomes more important for the suggestion score, while the similarity between the asset names becomes less important. If you increase the percentage, assets with similar names will receive a higher suggestion score. This percentage is expressed by a decimal where 1,00 equals 100%. You can enter a value greater than 1,00.

Setting	impacts	Description
Recommendation weights for data sets	Recommendations of data sets to users	An ordered comma-separated list of values that define the importance of properties for recommendations. The order of the values reflects the importance of the value. This setting is only used for data set recommendations if your Collibra does not yet have enough data for relevant results from the active recommendations algorithms. Possible values: ◦ *CERTIFIED*: Data sets that are certified are considered more relevant. ◦ *POPULARITY*: The number of visits to the data set page.
Active recommendation algorithms	Recommendations of data sets to users and of business assets to data sets	A comma-separated list of algorithms that calculate recommendations. By default, all available algorithms are listed. Possible values: ◦ *BASELINE* ◦ *USER_MEAN* ◦ *IICF* (Item-Item Collaborative Filtering) ◦ *SLOPE_ONE* ◦ *WEIGHTED_SLOPE_ONE*

Setting	impacts	Description
Data set elements threshold	Recommendations of data sets to users	The maximum number of elements per data set that the recommender will use to train the model. The data set elements are taken randomly. Lowering this number can prevent out-of-memory issues but also impacts the accuracy of recommendations for large data sets.

> Warning If you create an invalid Cron pattern, Collibra Data Intelligence Cloud stops responding.

3. Click the green **Save all** button.

> Note Depending on the configuration that you have applied, it is possible that you do not notice the recommendation updates immediately, but only the next day, for example when you update a schedule.

Customizing the search index

Before you customize the Collibra Data Intelligence Cloud search feature, it is important to learn how the search functionality works.

All text content in Collibra is stored in a search index to allow fast text search. To populate the search index, the text is split into separate words. The split is done by a component called the tokenizer and the words are often called tokens.

Every logical entity is stored in an index document. This document contains information about how many times a specific token occurs in the text. Separate index documents are stored for:

- Asset names
- Community names
- Domain names
- Text attributes
- Comments

When you search for text, the text is also tokenized in the same way. Then, the different words are searched for in the entire search index and a score is calculated for each of the matched documents. The calculation of this score is driven by different factors:

- The number of times the searched words occur in the document
- The size of the match relative to the size of the document

> Tip You can influence this score by changing the boost factor. For more information about the search functionality in Collibra, see Searching in Collibra DGC.

Configure the general search behavior

You can customize the search index configuration.

> Note Depending on your environment, you have to follow this procedure either in the Services Configuration section of the Collibra settings or in Collibra Console. This topic shows the procedure in Collibra Console. If your Settings page has a **Configuration** tab, you can follow this procedure in that tab page. For more information, see the online documentation.

Prerequisites

- You have the ADMIN or SUPER role in Collibra Console.
- You have the SUPER role in Collibra Console.

Steps

1. Open the DGC service settings for editing:
    a. Open Collibra Console.
        » Collibra Console opens with the **Infrastructure** page.

b. In the tab pane, expand an environment to show its services.
c. In the tab pane, click the Data Governance Center service of that environment.
d. Click **Configuration**.
e. Click **Edit configuration**.

2. In the **Search Index configuration** section, make the necessary changes.

Setting	Description
UI search appends wild-card	◦ ✔ True (default): A wildcard (asterisk) is automatically added to each search query. An asterisk is not added in the following exceptions: ▪ If the query contains a tilde (~). ▪ If the query ends with a quotation mark ("). > Note This applies only to queries via the user interface. A wildcard is not added automatically for REST API queries. ◦ ✖ False: No wildcard is added to the search query.
Maximum batch size	The amount of resources scanned in one go for the search query. The default value is 5,000. The maximum value is 30,000.
Maximum batch size for relations	Maximum batch size for relations reindex.

Setting	Description
Stop words (Requires restart)	A list of stop words that are ignored as tokens for the index. The default list of English stop words includes: a, an, and, are, as, at, be, but, by, for, if, in, into, is, it, no, not, of, on, or, such, that, the, their, then, there, these, they, this, to, was, will, with If you choose not to create your own list of stop words, the default list applies. If you create your own list of stop words, you have to: a. Reindexing Collibra Data Intelligence Cloud. b. Restart the environment to apply your changes. See Stop an environment and Start an environment.

3. Click the green **Save all** button.

Stop words

Stop words are specified words that the search engine ignores in a search text. The list typically contains articles, prepositions, pronouns, and auxiliary verbs.

The default list of English stop words includes: a, an, and, are, as, at, be, but, by, for, if, in, into, is, it, no, not, of, on, or, such, that, the, their, then, there, these, they, this, to, was, will and with.

> Note You can create your own list of stop words. If you choose not to, the default list applies.

Create a list of stop words

You can create a list of stop words, which tells the search engine which words in your search text to ignore.

> Note If you choose not to create a list of stop words, the default list applies.

> Note Depending on your environment, you have to follow this procedure either in the Services Configuration section of the Collibra settings or in Collibra Console. This topic shows the procedure in Collibra Console. If your Settings page has a **Configuration** tab, you can follow this procedure in that tab page. For more information, see the online documentation.

Prerequisites

- You have the ADMIN or SUPER role in Collibra Console.
- You have the SUPER role in Collibra Console.

Steps

1. Open the DGC service settings for editing:
   a. Open Collibra Console.
      » Collibra Console opens with the **Infrastructure** page.
   b. In the tab pane, expand an environment to show its services.
   c. In the tab pane, click the Data Governance Center service of that environment.
   d. Click **Configuration**.
   e. Click **Edit configuration**.
2. In the **Search index configuration** section, click in the **Stop words** field.
3. Type the first stop word, for example "a".
   » Another field appears below, in which you can type the next stop word.
4. Enter all of your stop words.

   > Warning To work correctly, you must ensure that each stop word field contains only one word.

5. Click the green **Save all** button.
6. Reindex Collibra Data Intelligence Cloud.
   For more information about reindexing, see Reindexing Collibra Data Intelligence Cloud.

What's next?

- Restart the environment to apply your changes. See Stop an environment and Start an environment.

Delete words from your list of stop words

You can delete some or all of the words in your list of stop words.

> Note If you delete all of the words in your list of stop words, the default list applies.

> Note Depending on your environment, you have to follow this procedure either in the Services Configuration section of the Collibra settings or in Collibra Console. This topic shows the procedure in Collibra Console. If your Settings page has a **Configuration** tab, you can follow this procedure in that tab page. For more information, see the online documentation.

Prerequisites

- You have the ADMIN or SUPER role in Collibra Console.
- You have the SUPER role in Collibra Console.
- You previously created a list of stop words.

Steps

1. Open the DGC service settings for editing:
   a. Open Collibra Console.
      » Collibra Console opens with the **Infrastructure** page.
   b. In the tab pane, expand an environment to show its services.
   c. In the tab pane, click the Data Governance Center service of that environment.
   d. Click **Configuration**.
   e. Click **Edit configuration**.
2. In the **Search index configuration** section, click in the **Stop words** field.
3. Click next to each of the stop words that you want to delete.
4. Click the green **Save all** button.

5. Reindex Collibra Data Intelligence Cloud.
   For more information about reindexing, see Reindexing Collibra Data Intelligence Cloud.

What's next?

- Restart the environment to apply your changes. See Stop an environment and Start an environment.

Edit the tokenizer settings

You can change the way text is tokenized (split into separate words).

> Note Depending on your environment, you have to follow this procedure either in the Services Configuration section of the Collibra settings or in Collibra Console. This topic shows the procedure in Collibra Console. If your Settings page has a **Configuration** tab, you can follow this procedure in that tab page. For more information, see the online documentation.

Prerequisites

- You have the ADMIN or SUPER role in Collibra Console.
- You have the SUPER role in Collibra Console.

Steps

1. Open the DGC service settings for editing:
   a. Open Collibra Console.
      » Collibra Console opens with the **Infrastructure** page.
   b. In the tab pane, expand an environment to show its services.
   c. In the tab pane, click the Data Governance Center service of that environment.
   d. Click **Configuration**.
   e. Click **Edit configuration**.
2. In the **Search index configuration** section, under **Tokenizer**, make the necessary

changes.

Setting	Description
Type	The tokenizer that is used. Currently two tokenizers are supported: ◦ **Standard** (default): This tokenizer uses the word break rules from the Unicode Text Segmentation algorithm, as specified in Unicode Standard Annex #29. ◦ **Character**: This tokenizer sees words as groups of all alpha-numeric characters together with a configurable list of extra characters. This can be used if you know for sure which characters should keep certain words together. For example, if you want to keep words with a dash (-) together, you have to add the dash in the **allowedCharacters** parameter.
Parameter map	The allowed characters if the **Type** is **Character**. ◦ **Field key**: This field has to contain *allowedCharacters*. ◦ **Field value**: The concatenated list of characters that does not split strings into separate tokens. For example, the concatenated list -'allows dashes and apostrophes in tokens.

3. Click the green **Save all** button.

What's next?

- Restart the environment to apply your changes. See Stop an environment and Start an environment.
- Reindex Collibra Data Intelligence Cloud. For more information, see Reindexing Collibra Data Intelligence Cloud.

Search boost settings

Search boost factors allow you to influence the order of the search results. You can edit search boost factors to increase or decrease the importance of a resource type, field or asset type, in the search ranking.

You can edit the boost settings.

Values

The default boost factor for all resource types and fields is 1, with the exception of Name, which has a default value of 3.

The higher the value, the greater the relevance score. Boost factors can have decimals but have to be bigger than 0.

Boosting object types

The search boost factor of a object type determines whether objects of this type appear higher or lower in the search results if the search string was found in any of its fields. Boosting applies to the following object types:

- Asset
- Community
- Domain
- User
- User group

Search boost factors only determine the relative importance of these five object types. They do not apply to the individual resources of a given object type. For example, you can boost the Community object type, but you cannot boost a specific community or any of the subcommunities, domains or assets in a specific community. Similarly, you can boost the User group object type, but you cannot boost a specific user group or the users in a specific user group.

You can use the facets on the Search page to navigate to specific resources.

> Example
> Let's assume that Asset has a search boost factor of 1, and Domain has a search boost factor of 3. If you search for a string that appears in any field of an asset and a domain, the domain will appear higher in the search ranking.

Boosting fields

The search boost factor of a field determines its importance for the search results, regardless of the resource type. The available fields are:

- Name
- Comment
- Tag
- <any attribute type>
- <any asset type>

> Example
> Let's assume that the Name field has a search boost factor of 2, and Tag has a search boost factor of 0,5. If you search for a string that appears in a Name field and a Tag field of different assets, the asset that has your search string in the Name field will be ranked higher than the asset that has the search string in the Tag field.

Boosting asset types

The search boost factor by asset type allows you to assign greater importance to specified asset types in the search results.

> Example
> Let's assume that the Business Process asset type has a search boost factor of 5, and the Risk asset type has a search boost factor of 1. If you search for a string that appears in any field of a Business Process asset and a Risk asset, the Business Process asset will be shown higher in the search ranking.

Edit search boost factors

You can edit search boost factors, to increase or decrease its importance of resource types, fields and asset types, in the search ranking.

By default, all resources and attribute types have a boost factor of 1. Values greater than 1 increase the importance, values less than 1 decrease the importance. All values must be greater than 0.

> Note Search performance can be affected by the number of edited boost factors and the range of the values. For example, assigning a boost factor of 2 will likely result in better Search performance than assigning a boost factor of 200. We recommend values no greater than 10. Keep in mind that you can use decimals, for example 0.5 and 6.3.

> Note Depending on your environment, you have to follow this procedure either in the Services Configuration section of the Collibra settings or in Collibra Console. This topic shows the procedure in Collibra Console. If your Settings page has a **Configuration** tab, you can follow this procedure in that tab page. For more information, see the online documentation.

Prerequisites

- You have the ADMIN or SUPER role in Collibra Console.
- You have the SUPER role in Collibra Console.

Steps

1. Open the DGC service settings for editing:
   a. Open Collibra Console.
      » Collibra Console opens with the **Infrastructure** page.
   b. In the tab pane, expand an environment to show its services.
   c. In the tab pane, click the Data Governance Center service of that environment.
   d. Click **Configuration**.
   e. Click **Edit configuration**.
2. In the **Search index configuration** section, edit the search boost factors of the following resource types and fields.

Setting	Description
Asset	The search boost factor of assets.

Setting	Description
Class Match	The search boost factor of data classes that were accepted during Automatic Data Classification. Note This is applicable only to users that have the Catalog global permission.
Community	The search boost factor of communities. Note This only affects the community asset itself, not the subcommunities, domains or assets in the community.
Domain	The search boost factor of domains. Note This only affects the domain asset itself, not the assets in the domain.
User	The search boost factor of users.
User group	The search boost factor of user groups.
Name	The search boost factor of the Name field.
Comment	The search boost factor of the Comment field.
Tag	The search boost factor of the Tag field.

3. Optionally, edit the boost factors of the existing additional attribute types in the list.
4. Optionally, add a boost factor of additional attribute types.
    a. Under **Attribute boost map**, click **Add**.
        » The **Add map option** dialog box appears.

b. Enter the required information.

Setting	Description
Field key	The resource ID of the attribute type. Tip You can find the resource ID of an attribute type in the Collibra settings.
Field value	The search boost factor of the attribute type.

c. Click **Add map option**.

5. Optionally, edit the boost factors of asset types.
   a. Under **Asset boost map**, click **Add**.
   b. Enter the required information.

Setting	Description
Field key	The resource ID of the asset type.
Field value	The search boost factor of the asset type.

   c. Click **Add map option**.

> Note Asset type hierarchy is not considered when boosting. This means that editing the boost factor of a parent asset type does not change the boost factor of any child asset types. You have to individually edit the boost factor of each relevant asset type.

6. Optionally, enable or disable exact matching:
   - ✔ True (default): If the name of an asset is exactly the same as the search text, put it at the top of the search results regardless of boost factors.
   - ✖ False: Use the regular search order, taking into account boost factors.
7. Optionally, enable or disable partial exact matching:
   - ✔ True (default): For multi-word search text, the search engine considers the exact match percentage with the resource name, when ordering the results.

> Example You enter search text "scheduled maintenance". Two example assets are ordered as follows:
>
> a. An asset named "daily **scheduled maintenance**", as two of the three words (66%) match exactly.
> b. An asset named "daily **scheduled maintenance** revised", as two of the four words (50%) match exactly.

- ✕ False: The exact match percentage is not taken into account in the score calculation.

8. Click the green **Save all** button.
9. Reindex Collibra Data Intelligence Cloud.

> Tip For examples of how boost factors affect search results, see Example search queries and analysis.

Find the resource ID of an attribute type

In some actions, you need the resource ID of an attribute type, for example, to edit search a boost factor.

Steps

1. In the main menu, click ⠿, then ⚙ **Settings**.
   » The Collibra settings page opens.
2. Click **Operating Model**.
   » The operating model settings appear on the **Asset types** tab page.
3. In the tab pane, click **Characteristics** → **Attributes**.
   » The table with attribute types appears.
4. In the content toolbar, click ⠿ → **Columns** → **Resource Id**.
   » The **Resource Id** column appears.

Delete an attribute boost map

You can delete an attribute boost map. As a result, the search boost factor of the attribute type is reset to 1.

> Note Depending on your environment, you have to follow this procedure either in the Services Configuration section of the Collibra settings or in Collibra Console. This topic shows the procedure in Collibra Console. If your Settings page has a **Configuration** tab, you can follow this procedure in that tab page. For more information, see the online documentation.

Prerequisites

- You have the ADMIN or SUPER role in Collibra Console.
- You have the SUPER role in Collibra Console.

Steps

1. Open the DGC service settings for editing:
   a. Open Collibra Console.
      » Collibra Console opens with the **Infrastructure** page.
   b. In the tab pane, expand an environment to show its services.
   c. In the tab pane, click the Data Governance Center service of that environment.
   d. Click **Configuration**.
   e. Click **Edit configuration**.
2. In the **Search index configuration** section, click **Boosting**.
3. In the **Attribute boost map** section, next to the boost map that you want to delete, click .
4. Click the green **Save all** button.
5. Reindex Collibra Data Intelligence Cloud.
   For more information about reindexing, see Reindexing Collibra Data Intelligence Cloud.

Exact match features

The exact match features help to ensure more intuitive ordering of search results.

You can enable and disable the exact match features in the boost settings. Both features are enabled, by default.

Exact Match Boost

This feature ensures that exact search text matches in the Name attribute of a resource always appear first in the search results, regardless of boost factors.

Example

Prerequisites

- The following assets exist in your Collibra environment:
  - A Schema asset with the name "Payment"
  - A Business Term asset with the name "Payment Type".
  - A Data Attribute asset with the name "Payment Type".
  - A Policy asset with the name "Payments".
  - Other assets, as shown in the following image.
- Edit asset type boost factors, in Collibra Console, as follows:
  - Business Term: 3
  - Data Attribute: 2
  - Policy: 1.5

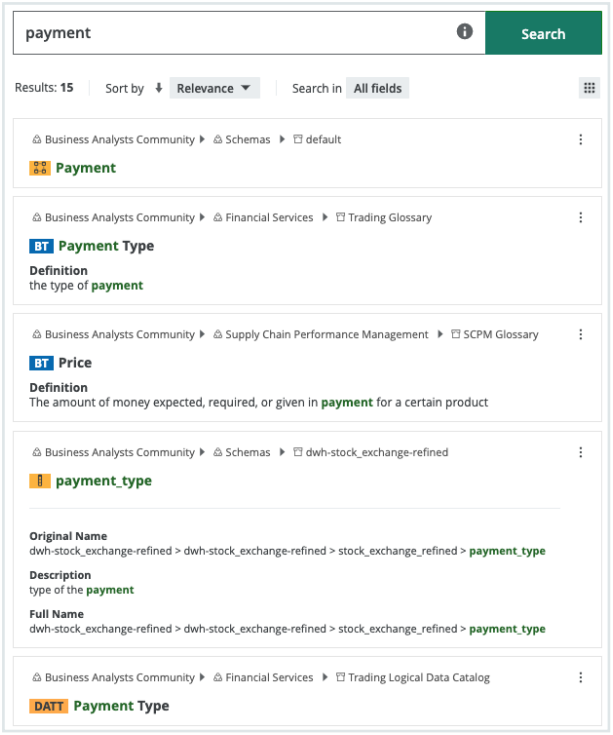
Search text

Enter search text "payment".

Results

The resources that exactly match the search text in the Name attribute of the resource appear first. In this example, there is only one exact match, the Schema asset with the name "Payment".

After the exact matches, the search results are sorted in order of descending relevance.



Partial Exact Match

This feature is only relevant for multi-word search text. In this case, the search engine considers the exact match percentage with the asset name, when ordering the results.

> Example You enter search text "scheduled maintenance". Two example assets are ordered as follows:
>
> 1. An asset named "daily **scheduled maintenance**", as two of the three words (66%) match exactly.
> 2. An asset named "daily **scheduled maintenance** revised", as two of the four words (50%) match exactly.

Relevance in search results

By default, search results are sorted in order of descending relevance.

What is relevance in the context of search results?

Relevance is a calculation of the similarity, measured across several lines of comparison, between your search text and the content of the resources in your Collibra environment.

In a set of search results, the relevance of each resource is represented by a positive number, or score. The higher the score, the more relevant the resource is to your search text.

How are relevance scores derived?

To derive relevance scores, the Collibra search engine uses a combination of query clauses and boost factors.

Query clauses

When you perform a search, the Collibra search engine queries the database, using various query clauses. Each query clause compares the similarity between your search text and your Collibra resources, along a different line of comparison.

The following are example objectives of different query clauses:

- Calculate the similarity between the spelling of your search term and the term found in a field in the database.
- Calculate how frequently your search term appears in a field. The more often it appears, the greater the relevance. A field containing five occurrences of a given term is more likely to be relevant than a field containing one occurrence of the term.
- Calculate the occurrence percentage of a term among all words in a particular field. For example, if your search term occurs twice in the 10-word description of an asset, that asset will have a higher relevance score than an asset for which your search term occurs twice in its 20-word description.

Boosting

Search boost factors allow you to influence the order of the search results. You can edit search boost factors to increase or decrease the importance of a resource type, field or asset type, in the search ranking.

For more information on boosting, see Search boost settings.

Example search queries and analysis

The following examples highlight common search text. We provide explanations for the results and explain how some configuration settings affect results.

> Note This information is intended to:
>
> - Help Collibra Data Intelligence Cloud administrators understand how the search configuration settings affect search results.
> - Help other Collibra users understand why some search queries might not provide the expected results.
>
> For more information, see the respective sections on customizing the search index, boosting, stop words and the exact match features.

Searching for assets that have stop words in their names

Prerequisites

- An asset named "On The Go" exists in your Collibra environment.

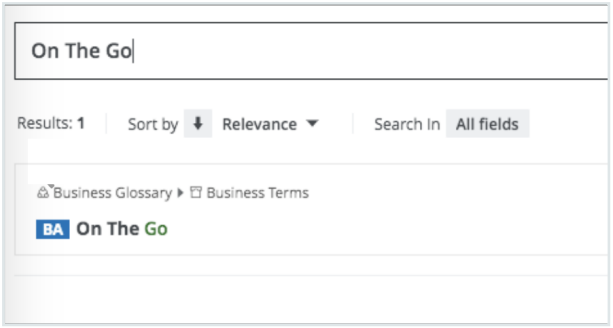
Search text

Enter search text "On The Go".

Results

The asset "On The Go" is found.

In the following image, the word "Go" is shown in green text. This indicates that "Go" was the match that produced the result. The words "On The", which are shown in black text, did not produce a match.



Furthermore, if you enter search text "On The", the asset "On The Go" will not be found. Nor will any other asset. This is because "on" and "the" are stop words, which are filtered from indexing and searches.

> Tip The best way to ensure thorough and intuitive search results is to name your assets, domains and communities as thoughtfully as possible.

Searching for assets that have more than one word in their names

Search text

Enter search text "marketing team summit".

The search engine handles the search text as: "marketing" OR "team" OR "summit*". Notice the wildcard (asterisk) at the end of the word "summit". This is determined by the default UI search appends wildcard setting, which adds the wildcard to the end of the search text.

Results

- An asset named "marketing_campaign_xyz" is not found.
- An asset named "team123' is not found.
- An asset named "summit_planning" is found.

To find the assets "marketing_campaign_xyz" and "team123", you have to add the wildcard after each word, for example: "marketing * team * summit". The search engine will handle this as: "marketing*" OR "team*" OR "summit*".

How boosting specified resource types affects search results

Prerequisites

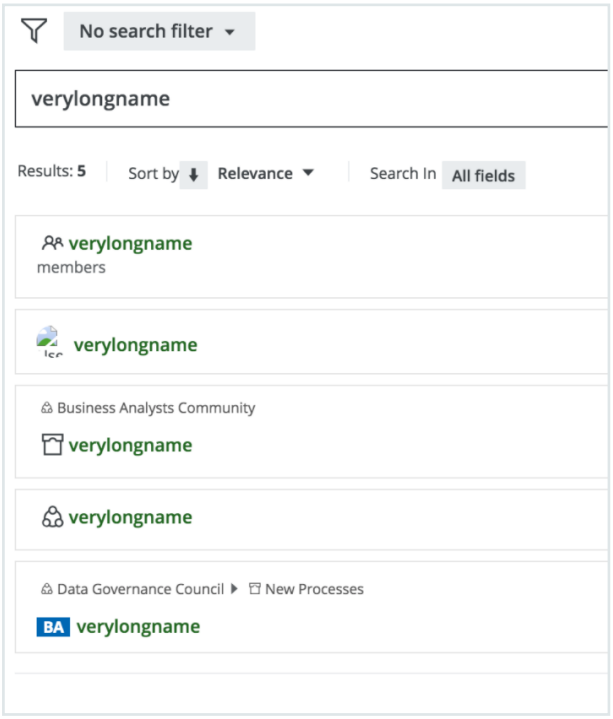
- A resource, user or user group with the name "verylongname" exists in your Collibra environment.
- Edit resource type boost factors, in Collibra Console, as follows:
  - Asset: 2
  - Community: 4
  - Domain: 6
  - User: 8
  - User group: 10

Search text

Enter search text "verylongname".

Results

The results are ordered in accordance with the boost factor values of the respective resource types. Resource type user group, with a boost factor of 10, is the most relevant of the results. Asset, with a boost factor of 2, is the least relevant resource type.



How boosting specified fields affects search results

Prerequisites

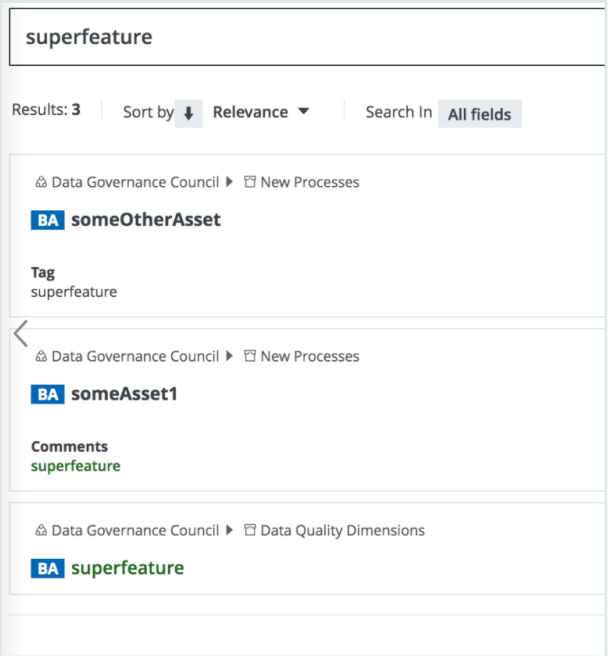
- The following assets exist in your Collibra environment:
  - An asset named "superfeature".
  - An asset named "asset1", with a tag "superfeature".
  - An asset named "asset2", with a comment "superfeature"
- Edit property boost factors, in Collibra Settings, as follows:
  - Name: 1
  - Comment: 5
  - Tag: 10

Search text

Enter search text "superfeature".

Results

The results are ordered in accordance with the boost factor values of the respective fields. The tag field, with a boost factor of 10, is the most relevant of the results. Name, with a boost factor of 1, is the least relevant field.



How boosting attributes affects search results

Prerequisites

- The following assets exist in your Collibra environment:
  - An asset with the description "terminator".
  - An asset with the definition "terminator".
  - An asset with the note "definition".
- Edit attribute boost factors, in Collibra Console, as follows:
  - Description: 1
  - Definition: 2
  - Note: 3

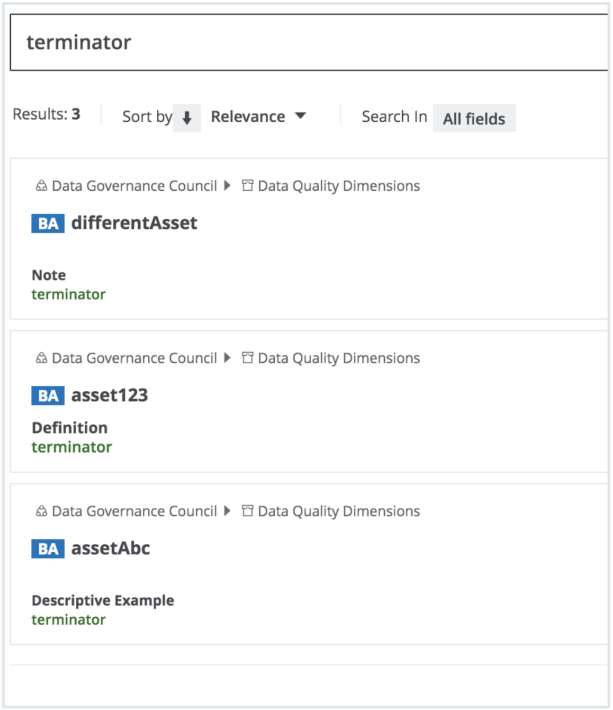
Search text

Enter search text "terminator".

Results

The results are ordered in accordance with the boost factor values of the respective attributes. The note attribute, with a boost factor of 3, is the most relevant of the results.

Description, with a boost factor of 1, is the least relevant attribute.



How boosting asset types affects search results

Prerequisites

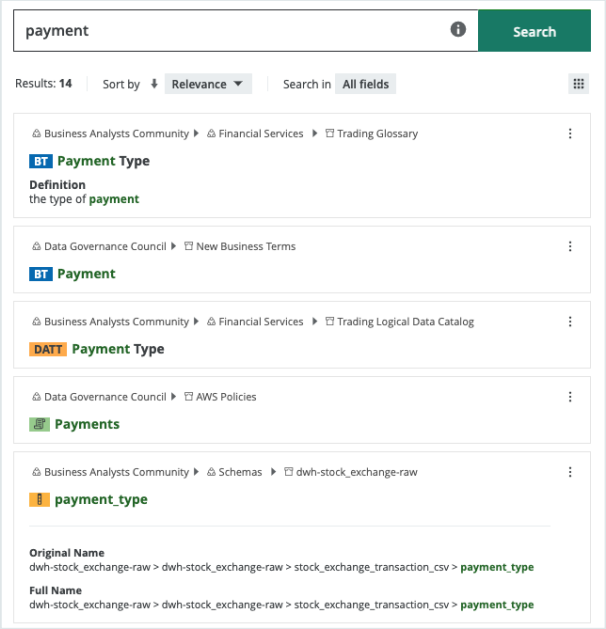
- The following assets exist in your Collibra environment:
  - A Business Term asset with the name "Payment".
  - A Data Attribute asset with the name "Payment Type".
  - A Policy asset with the name "Payments".
- Edit attribute boost factors, in Collibra Console, as follows:
  - Business Term: 3
  - Data Attribute: 2
  - Policy: 1.5

Search text

Enter search text "payment".

Results

The results are ordered in accordance with the boost factor values of the respective asset types. The Business Term asset type, with a boost factor of 3, is the most relevant of the results. The Policy asset type, with a boost factor of 1.5, is the least relevant asset type of the three boosted asset types.



How the exact match boost feature affects search results, regardless of boost factors

Prerequisites

- The following assets exist in your Collibra environment:
  - A Schema asset with the name "Payment"
  - A Business Term asset with the name "Payment Type".
  - A Data Attribute asset with the name "Payment Type".
  - A Policy asset with the name "Payments".
  - Other assets, as shown in the following image.
- Edit asset type boost factors, in Collibra Console, as follows:
  - Business Term: 3
  - Data Attribute: 2
  - Policy: 1.5

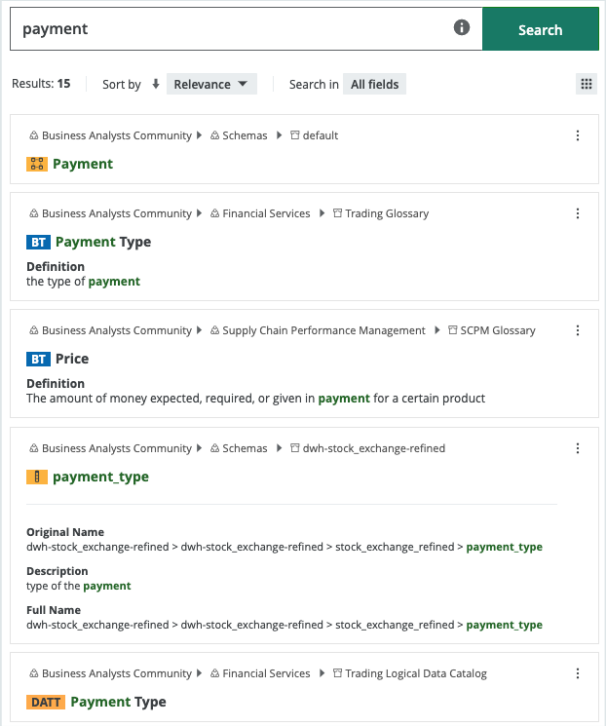
Search text

Enter search text "payment".

Results

The resources that exactly match the search text in the Name attribute of the resource appear first. In this example, there is only one exact match, the Schema asset with the name "Payment".

After the exact matches, the search results are sorted in order of descending relevance.



How term frequency contributes to relevance scores

Prerequisites

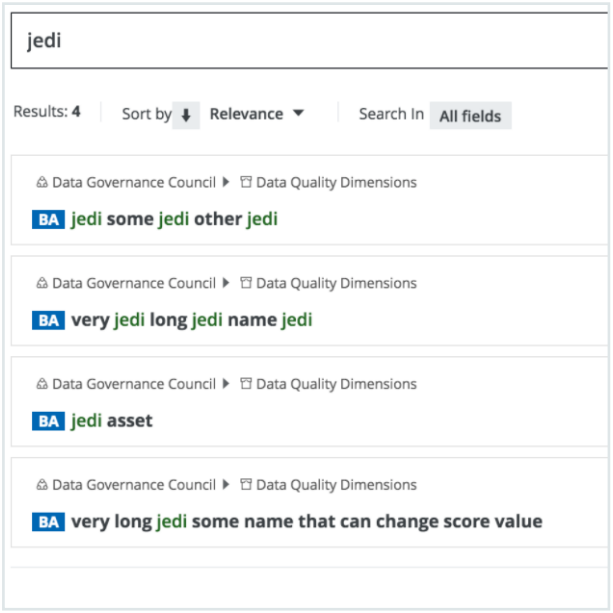
- Several assets, each with a variation of "jedi" in the name, exist in your Collibra environment.
- Default boost factors for all resource types, properties and attributes.

Search text

Enter search text "jedi".

Results

The results are ranked as shown in the following image. A combination of total number of matches in a name and the percentage of matches per total words in a name affect relevance scoring of the results.



Recommended file types for uploads

In general, you can allow any file type to be uploaded to Collibra Data Intelligence Cloud. However, you can limit the types of files that users can upload.

Below is a list with the minimum of allowed file types that you need for Collibra to work properly:

File type	MIME type
.xls	application/vnd.ms-excel

File type	MIME type
.xlsx	application/vnd.openxmlformats-office-document.spreadsheetml.sheet
.xml	application/xml
.zip	application/zip Note Only administrators can work with backups.
.java, .log, .txt, ...	text/plain Note For a list of all available file types, see http://www.sitepoint.com/web-foundations/mime-types-complete-list.
.csv	text/csv
.csv	text/plain
.bpmn	application/xml Note Only administrators can work with workflows.

Below is a list with an enumeration of additional recommended allowed file types:

File type	MIME type
.doc	application/msword
.docx	application/vnd.openxmlformats-office-document.wordprocessingml.document

File type	MIME type
.dotx	application/vnd.openxmlformats-office-document.wordprocessingml.template
.xltx	application/vnd.openxmlformats-office-document.spreadsheetml.template
.ppt	application/vnd.ms-powerpoint
.pptx	application/vnd.openxmlformats-office-document.presentationml.presentation
.potx	application/vnd.openxmlformats-office-document.presentationml.template
.ppsx	application/vnd.openxmlformats-office-document.presentationml.slideshow
.pdf	application/pdf
.html	text/html

Below is a list with an enumeration of recommended file types for avatars:

File type	MIME type
.jpeg, .jpg	image/jpeg
.png	image/png
.gif	image/gif

Edit the buffer settings of the Collibra DGC statistics

You can edit the buffer settings of the Collibra Data Intelligence Cloud

> Note Depending on your environment, you have to follow this procedure either in the Services Configuration section of the Collibra settings or in Collibra Console. This topic shows the procedure in Collibra Console. If your Settings page has a **Configuration** tab, you can follow this procedure in that tab page. For more information, see the online documentation.

Prerequisites

- You have the ADMIN or SUPER role in Collibra Console.
- You have the SUPER role in Collibra Console.

Steps

1. Open the DGC service settings for editing:
    a. Open Collibra Console.
        » Collibra Console opens with the **Infrastructure** page.
    b. In the tab pane, expand an environment to show its services.
    c. In the tab pane, click the Data Governance Center service of that environment.
    d. Click **Configuration**.
    e. Click **Edit configuration**.

2. In the **Statistics configuration** section, make the necessary changes.

Setting	Description
Buffer size	The maximum amount of statistics entries that the buffer can contain before saving them in the database. The default value is *10*.
Buffer flush time	The maximum amount of time in milliseconds to keep statistic entries in memory before saving them in the database. The default values is *10,000*.
Cron map	List of statistics, listed by their Cron name, and a Cron interval. These are the default values: **Field key** / **Field value** workflow-task: 0 59 23 * * ? active-users: 0 0/15 * * * ? term-count: 0 59 23 * * ? vocabulary-count: 0 59 23 * * ? page-hit: 0 0 * * * ? task-count: 0 0 * * * ? If you create an invalid Cron pattern, Collibra Data Intelligence Cloud stops responding.

3. Click the green **Save all** button.

Edit the schedule of Collibra DGC statistics

Collibra Data Intelligence Cloud collects statistics at regular intervals. The statistics are shown on various pages, such as the **Metrics** pages of the Collibra applications.

> Note Depending on your environment, you have to follow this procedure either in the Services Configuration section of the Collibra settings or in Collibra Console. This topic shows the procedure in Collibra Console. If your Settings page has a **Configuration** tab, you can follow this procedure in that tab page. For more information, see the online documentation.

Prerequisites

- You have the ADMIN or SUPER role in Collibra Console.
- You have the SUPER role in Collibra Console.

Steps

1. Open the DGC service settings for editing:
    a. Open Collibra Console.
        » Collibra Console opens with the **Infrastructure** page.
    b. In the tab pane, expand an environment to show its services.
    c. In the tab pane, click the Data Governance Center service of that environment.
    d. Click **Configuration**.
    e. Click **Edit configuration**.

2. In the **Statistics configuration** section, make the necessary changes.

Setting	Description
Buffer size	The maximum amount of statistics entries that the buffer can contain before saving them in the database. The default value is *10*.
Buffer flush time	The maximum amount of time in milliseconds to keep statistic entries in memory before saving them in the database. The default values is *10,000*.
Cron map	List of statistics, listed by their Cron name, and a Cron interval. These are the default values: **Field key** / **Field value** workflow-task: 0 59 23 * * ? active-users: 0 0/15 * * * ? term-count: 0 59 23 * * ? vocabulary-count: 0 59 23 * * ? page-hit: 0 0 * * * ? task-count: 0 0 * * * ? If you create an invalid Cron pattern, Collibra Data Intelligence Cloud stops responding.

3. Click the green **Save all** button.

What's next?

Restart the environment to apply your changes. See Stop an environment and Start an environment.

Edit import configuration

In Collibra Data Intelligence Cloud, you can import data from comma-separated value (CSV, Excel) files.

> Note Depending on your environment, you have to follow this procedure either in the Services Configuration section of the Collibra settings or in Collibra Console. This topic shows the procedure in Collibra Console. If your Settings page has a **Configuration** tab, you can follow this procedure in that tab page. For more information, see the online documentation.

Prerequisites

- You have the ADMIN or SUPER role in Collibra Console.
- You have the SUPER role in Collibra Console.

Steps

1. Open the DGC service settings for editing:
    a. Open Collibra Console.
        » Collibra Console opens with the **Infrastructure** page.
    b. In the tab pane, expand an environment to show its services.
    c. In the tab pane, click the Data Governance Center service of that environment.
    d. Click **Configuration**.
    e. Click **Edit configuration**.

2. In the **Imports configuration** section, make the necessary changes.

Setting	Description
Rebuild hyperlinks after import	◦ ✓ True (default): Automatically rebuild the hyperlinks after an import. ◦ ✕ False: Do not rebuild the hyperlinks after an import.
Asset responsibilities support	◦ ✓ True: Enable importing responsibilities at asset level. ◦ ✕ False (default): Disable importing responsibilities at asset level. Warning Setting specific responsibilities on a large number of resources will affect the performance and stability of the system.
Temp files storage	The location of the temporary files used by the import job. The default value is *FILE*.

3. Click the green **Save all** button.

Edit Excel export configuration

In Collibra Data Intelligence Cloud, you can export Collibra data to CSV or Excel files.

> Note Depending on your environment, you have to follow this procedure either in the Services Configuration section of the Collibra settings or in Collibra Console. This topic shows the procedure in Collibra Console. If your Settings page has a **Configuration** tab, you can follow this procedure in that tab page. For more information, see the online documentation.

Prerequisites

- You have the ADMIN or SUPER role in Collibra Console.
- You have the SUPER role in Collibra Console.

Steps

1. Open the DGC service settings for editing:
   a. Open Collibra Console.
      » Collibra Console opens with the **Infrastructure** page.
   b. In the tab pane, expand an environment to show its services.
   c. In the tab pane, click the Data Governance Center service of that environment.
   d. Click **Configuration**.
   e. Click **Edit configuration**.
2. In the **Excel export configuration** section, make the necessary changes.

Setting	Description
The default CSV separator character	The default separator character of the CSV fields for complex relations.
The default CSV quote character	The default quote character of the CSV fields for complex relations.
Number of rows per chunk of data	When exporting views, the database is called repeatedly, each time fetching a chunk of data to build the export file. This option defines how many rows each chunk of data can contain. Lower values reduce the burden on memory. Higher values require more memory, but may slightly increase the speed of the export. The default value is 5,000.

3. Click the green **Save all** button.

Edit CSV export configuration

In Collibra Data Intelligence Cloud, you can export Collibra data to CSV or Excel files.

> Note Depending on your environment, you have to follow this procedure either in the Services Configuration section of the Collibra settings or in Collibra Console. This topic shows the procedure in Collibra Console. If your Settings page has a **Configuration** tab, you can follow this procedure in that tab page. For more information, see the online documentation.

Prerequisites

- You have the ADMIN or SUPER role in Collibra Console.
- You have the SUPER role in Collibra Console.

Steps

1. Open the DGC service settings for editing:
   a. Open Collibra Console.
      » Collibra Console opens with the **Infrastructure** page.
   b. In the tab pane, expand an environment to show its services.
   c. In the tab pane, click the Data Governance Center service of that environment.
   d. Click **Configuration**.
   e. Click **Edit configuration**.

2. In the **CSV export configuration** section, make the necessary changes.

Setting	Description
Always use quotes	◦ ✔ True: Use quotes for every cell in the CSV. ◦ ✖ False (default): Only use quotes when necessary.
Number of rows per chunk of data	When exporting views, the database is called repeatedly, each time fetching a chunk of data to build the export file. This option defines how many rows each chunk of data can contain. Lower values reduce the burden on memory. Higher values require more memory, but may slightly increase the speed of the export. The default value is 5,000.

3. Click the green **Save all** button.

Configure the logging of Collibra DGC API calls

You can use Collibra Data Intelligence Cloud in a web interface, but also through REST API calls. Every API call is logged in a database. You can configure the methods that you want to log and the maximum number of entries to store.

Steps

1. Open the DGC service settings for editing:
2. In the **API call logging**, make the necessary changes.

Setting	Description
Enabled	◦ ✔ True: API call logging is enabled. ◦ ✖ False (default): API call logging is disabled.

Setting	Description
Maximum number of log entries (Requires restart)	The maximum number of API calls to store in the database. Once this number is reached, the oldest records are overwritten. The default value is 1,000,000.
Pattern duration list	The list of methods and a corresponding minimum duration time. The minimum duration time is the minimum time before the method is stored in the database.
Minimum duration	The time in milliseconds that an API call must last before it is logged.
Method pattern	The method that you want to log in the database. For each pattern that you want to log, you have to add a new pattern.

3. Click the green **Save all** button.

What's next?

Restart the environment to apply your changes. See Stop an environment and Start an environment.

Create a pattern duration for API call logs

You can create a pattern duration for API call logs.

> Note Depending on your environment, you have to follow this procedure either in the Services Configuration section of the Collibra settings or in Collibra Console. This topic shows the procedure in Collibra Console. If your Settings page has a **Configuration** tab, you can follow this procedure in that tab page. For more information, see the online documentation.

Prerequisites

- You have the ADMIN or SUPER role in Collibra Console.
- You have the SUPER role in Collibra Console.

Steps

1. Open the DGC service settings for editing:
    a. Open Collibra Console.
        » Collibra Console opens with the **Infrastructure** page.
    b. In the tab pane, expand an environment to show its services.
    c. In the tab pane, click the Data Governance Center service of that environment.
    d. Click **Configuration**.
    e. Click **Edit configuration**.
2. In the **API call logging** section under **Pattern duration list**, click **Add**.

3. In the new record, fill in the required fields.

Setting	Description
Enabled	◦ ✔ True: API call logging is enabled. ◦ ✖ False (default): API call logging is disabled.
Maximum number of log entries (Requires restart)	The maximum number of API calls to store in the database. Once this number is reached, the oldest records are overwritten. The default value is 1,000,000.
Pattern duration list	The list of methods and a corresponding minimum duration time. The minimum duration time is the minimum time before the method is stored in the database.
Minimum duration	The time in milliseconds that an API call must last before it is logged.
Method pattern	The method that you want to log in the database. For each pattern that you want to log, you have to add a new pattern.

4. Click the green **Save all** button.

What's next?

Restart the environment to apply your changes. See Stop an environment and Start an environment.

Security configuration

You can configure the security of Collibra Data Intelligence Cloud in various aspects, going from SSL settings to password settings.

Configuring LDAP

You can integrate Collibra Data Intelligence Cloud with your organization's LDAP servers. User data as well as group-related information can be synchronized and used by Collibra. When LDAP is configured, authentication (credential checking) is done directly on the LDAP server(s).

In order for Collibra to synchronize with the LDAP server, the LDAP fields have to be mapped to the corresponding Collibra fields and the LDAP server has to be configured correctly.

Configure one or more LDAP servers

To configure one or more LDAP servers, follow these steps:

1. Open the DGC service settings for editing:
2. Under **Servers** of the **LDAP** section, click **Add**.
3. Fill in the required fields. See Security configuration: options.
   For a complete tutorial, visit Collibra Support Portal.
4. Click the green **Save all** button.
5. Reindex the environment.

What's next?

Restart the environment to apply your changes. See Stop an environment and Start an environment.

How to enable logging for troubleshooting LDAP issues

Before you start to troubleshoot issues with LDAP, make sure that you have configured LDAP as documented here.

If you still encounter issues, enable logging for further investigation.

> Warning If you have investigated the LDAP issue, don't forget to revert all the changes from this section.

Steps

1. Open Collibra Console.
   » Collibra Console opens with the **Infrastructure** page.
2. In the tab pane, click the **Data Governance Center** service of the environment whose log settings you need.
3. Click **Logs**.
4. Above the table, to the right, click **Settings**.
5. Click **Add logger**.
6. In the **Add logger** dialog box, enter the necessary information.
   - **Logger name**: Enter *com.collibra.dgc.core.service.ldap.impl.LdapServiceImpl*
   - **Logger level**: Select *DEBUG*.
7. Click **Add logger**.

Define password strength

Every Collibra Data Intelligence Cloud user has a password to sign in to Collibra. You can define the password strength so that your system can resist guessing and brute-force attacks. You can set the password complexity, as well as limit the number of sign-in attempts or set an expiry date for the password.

Steps

1. Open the DGC service settings for editing:
2. In the **Security configuration** settings, click **Password**.
3. Set the necessary password policy parameters. See Security configuration: options, that also contains the default values for each of the password parameters.
4. Click the green **Save all** button.

What's next?

Restart the environment to apply your changes. See Stop an environment and Start an environment.

Configure REST security

To change the REST security settings, follow these steps:

1. Open the DGC service settings for editing:
2. In the **REST** section, make the necessary changes.

Setting	Description
CSRF token enabled	◦ ✔ True: The validity of a request is checked with a CSRF token. ◦ ✖ False (default): The validity of a request is not checked with a CSRF token.
Referrer enabled	◦ ✔ True: The HTTP referrer header is used to identify the origin of the request. ◦ ✖ False (default): The HTTP referrer header is not used to identify the origin of the request. It is recommended to leave this option disabled.
Referrer checking allow empty	◦ ✔ True (default): The HTTP referrer header can be empty. ◦ ✖ False: The HTTP referrer header cannot be empty.

3. Click the green **Save all** button.

What's next?

Restart the environment to apply your changes. See Stop an environment and Start an environment.

Configure the SSL settings

For secure communication from Collibra Data Intelligence Cloud to other services such as an LDAP server, you can activate SSL.

> Tip To configure SSL to securely access Collibra from your web browser, see Configure SSL to access Collibra DGC.

Steps

1. Open the DGC service settings for editing:
2. In the **SSL** section, make the necessary changes.

Setting	Description
Key store name	The name of the keystore file. The file is expected to be in the **<collibra_data>/dgc/security** folder.
Key store password	The password of the keystore.
Key store type	The type of the keystore file. For example, *JKS* or *PKCS12*.
Trust store name	The name of the truststore file. The file is expected to be in the **<collibra_data>/dgc/security** folder.
Trust store password	The password of the truststore.
Trust store type	The type of the truststore file. For example, *JKS* or *PKCS12*.

3. Click the green **Save all** button.

What's next?

Restart the environment to apply your changes. See Stop an environment and Start an environment.

Configure SSL to access Collibra DGC

If you want to connect to Collibra Data Governance Center in a secure way with your web browser, you have to use SSL. This procedure explains how you can activate SSL access to Collibra DGC.

> Tip For secure communication from Collibra DGC to other services, for example an LDAP server, see Configure the SSL settings.

Prerequisites

- You have knowledge of the JSON syntax.
- You have created a Java KeyStore according the procedure described by Oracle, for example **clientkeystore**.
- You have noted the following data while creating the Java KeyStore:
  - KeyStore file name: *clientkeystore* in the Oracle example.
  - KeyStore alias: *client* in the Oracle example.
  - KeyStore password: The password that you entered after executing the command of the first step in the Oracle example.
  - KeyStore alias password: The password that you entered as last step of step 2 in the Oracle example.
- You have stored the Java KeyStore on the DGC node in the **<collibra_data>/dgc/security** folder, for example **/opt/collibra_data/dgc/security**.

Steps

To configure Collibra Data Governance Center for access over SSL, follow these steps:

1. Open a terminal session on the DGC node.
2. Open the file **<collibra_data>/dgc/config/server.json** for editing.
3. Fill in the following parameters in the **httpsConnector** section:
   Add string values between double quotes.

Parameter	Description
port	The port on which the HTTPS connector must bind. The value must be higher than 1024 to avoid root permissions. > Note If you want to use the default SSL port 443, you have to use a reverse proxy.
keyAlias	The KeyStore alias.
keyPass	The KeyStore alias password.
keystorePass	The KeyStore password.
keystoreFile	The full path to the KeyStore file name, for example **/opt/collibra_data/dgc/security/clientkeystore**.

Example:

```
"httpsConnector" : {
    "port": 5404,
    "keyAlias": "your-alias",
    "keyPass": "your-password",
    "keystorePass": "your-password",
    "keystoreFile": "/opt/collibra_
data/dgc/security/collibradgc.jks" },
}
```

4. Save and close the file.
5. Open the DGC service settings for editing:
6. Click the **General settings** section.
7. Update the **Base URL** parameter with *https* and the new port.
8. Restart the environment.

Connect to your Collibra DGC environment via the Base URL.

Extra

To prevent regular HTTP traffic to Collibra DGC, update the **address** parameter with the value *127.0.0.1* in **<collibra_data>/dgc/config/server.json** and restart the environment.

This will not prevent the administration tools, for example Collibra Console, from connecting to Collibra DGC without SSL.

For more information, see the knowledge base on the Collibra Support Portal.

Single Sign-On (SSO)

Single Sign-On (SSO) enables users to access Collibra Data Intelligence Cloud using a web client, without having to explicitly type their login credentials (username and password).

Collibra provides support for two types of SSO. Each SSO type can be used with or without LDAP (Light-weight Directory Access Protocol), resulting in the following SSO modes:

- SAML 2.0, options:
  - SSO SAML with attributes sync
  - SSO SAML LDAP
- SSO through header information, options:
  - SSO Header with Collibra Data Intelligence Cloud users
  - SSO Header with LDAP sync

> Tip If you want to use a custom certificate in the SSO configuration for Collibra access, see this section.

This section explains how to:

- Choose the most appropriate mode of SSO.
- Configure the Collibra for the SSO mode you prefer.

Note On Collibra SAML support:

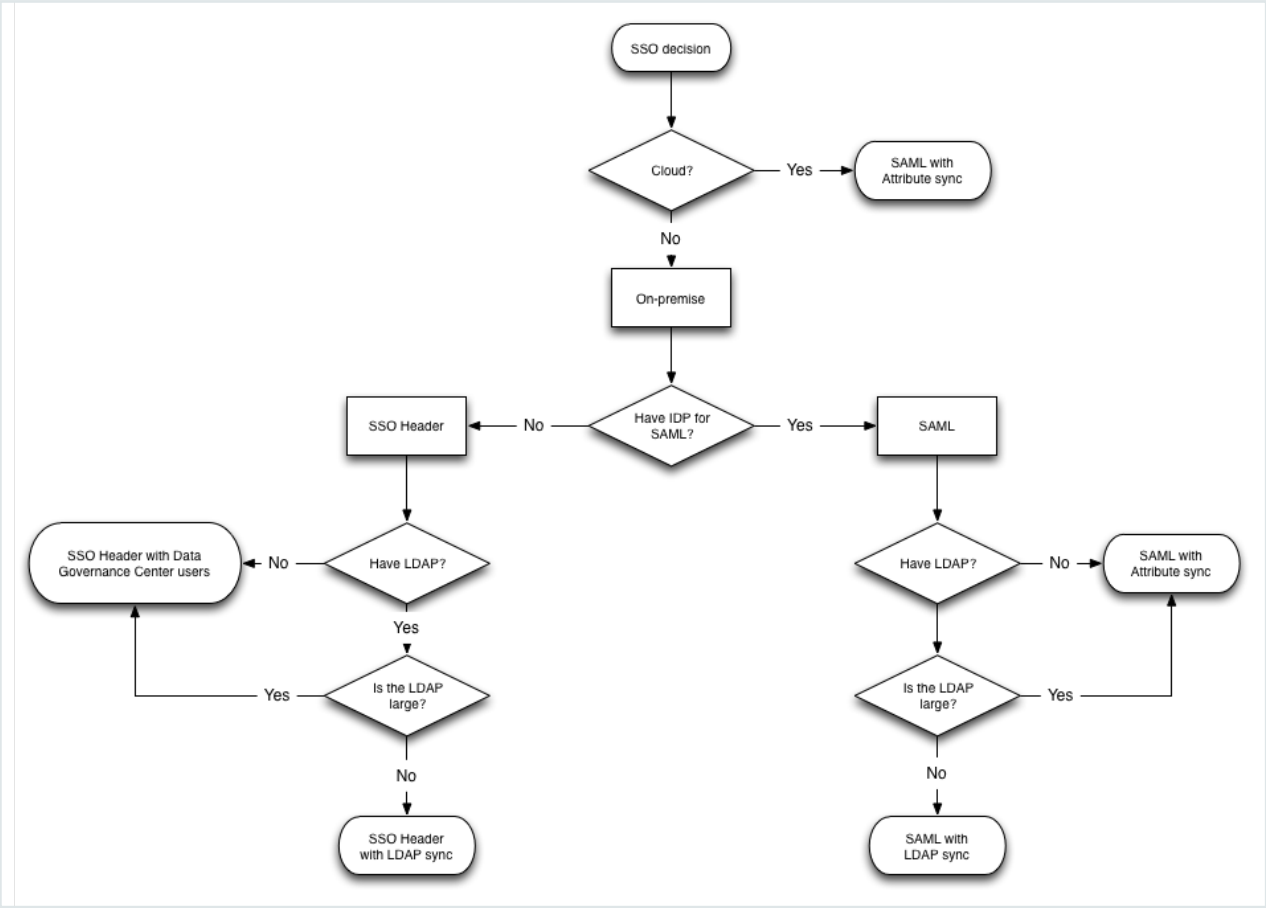
- The only supported SAML protocol version in Collibra is SAML 2.0 (urn:oasis:names:tc:SAML:2.0:protocol).
- When configuring SAML, in order to obtain the SP metadata from Collibra DGC, go to **https://<yourinstance>/rest/latest/saml/metadata**.
- If you want a full SP metadata, go to **https://<yourinstance>/rest/latest/saml/metadata?complete=true**.
- Collibra only supports assertions to come in through the HTTP-POST binding (as defined in the SP metadata file).
- For more information about this subject, see the knowledge base on the Collibra Support Portal.

Deciding which SSO mode to use

The following flow chart can help you decide which SSO mode to use. If you already know which mode is appropriate for your environment, just skip to the correct section in order to learn how to configure the relevant SSO mode.

Warning In case you are considering the SSO Header configuration, please make sure that you are aware of its limitations and security requirements. Please involve your security team while configuring SSO Header.

Note Collibra Data Intelligence Cloud Cloud only supports SAML with Attribute sync.



SSO header

SSO header with Collibra Data Intelligence Cloud users

Since there is no standard for the behavior of the SSO Header modes, Collibra makes certain assumptions.

Once configured with the appropriate header information (see configuration step), Collibra assumes that:

- You are running a proxy responsible for authenticating the users connecting to Collibra.
- This proxy fills in the value of the header to the username of the user that has been authenticated by the proxy. This is the case in every single request, including for resources like javascript, CSS and image files.

- Collibra uses this username to start a session in the application, assuming the above steps were performed properly.
- Collibra does not perform extra authentication.

To configure SSO header with Collibra users, consult the SSO header: configuration options section.

SSO Header with LDAP synchronization

Much like the SSO Header with Collibra users, there are no standards for proxy behavior.

Collibra assumes that:

- You are running an LDAP server and Collibra has been configured to sync with that LDAP. For more information about LDAP synchronization, see Configuring LDAP.
- You are running a proxy server that is responsible for authenticating the users who want to go to Collibra DGC.
- The proxy server fills in the value of the header to contain either the LDAP Distinguished Name (DN) or the identifying LDAP attribute of the user (as configured, see configuration step).
- Collibra:
  - Uses the DN or identifying attribute to fetch the username from the LDAP server.
  - Verifies that the user is allowed to use Collibra.
  - Uses this user to start a session.

To configure SSO header with LDAP synchronization, consult the SSO Header LDAP: configuration options section.

Single Sign-On FAQ

Search this FAQ section to see if you find an answer to your question. If you cannot find the right answer, please contact our support team at support@collibra.com.

Help! SSO is broken! I can’t get into Collibra!

Even though you seem to be locked out of Collibra Data Intelligence Cloud, you can always gain access by using a Collibra local account. You can change the address of the

page that you are visiting to end with /signin (So: https://<your instance>/signin)

A generic username/password form is displayed. You can only use accounts that have been created in Collibra itself.

Once you have regained access to your instance, you can try to figure out what exactly went wrong. To resolve any problems, you can start by retrieving the log files, and analyzing them. If something went wrong with the SAML authentication flow, some error will show up in these log files.

For more information about logging, see Logging.

I'm getting the "SAML authentication failed. Please contact your Administrator for more information." error, but nothing wrong shows up in the log

Most likely, this is because you are trying to sign in without global application permissions.

In most cases, you will want to use the **Everyone** group to assign application permissions to everybody. That is including new users and guest users. To do this, sign in manually, with the sign-in page, using an admin account. Then navigate to **Settings > Roles**, and assign the **Everyone** group to the appropriate global roles. For more details, see Collibra Console roles and Responsibilities.

In the cloud environment, when do we require a restart of the instance if we changed something?

A restart of the environment can be requested when:

- Your initial configuration has been done, and the SAML metadata file has been provisioned by uploading a SAML metadata file.

In order to help in the decision process, the following things are being reloaded automatically:

- The SAML metadata file is reloaded every 2 hours. The next reload point will be mentioned in the log file.

- All **configuration.xml** changes will be reloaded on every SAML Response (meaning: Collibra always uses the current setting in the **configuration.xml**).

You can request a restart of your cloud instance by contacting our support team at support@collibra.com.

How do I upload my SAML metadata file to a cloud instance?

If your environment is using a cloud instance of Collibra, the only supported mode of SSO is SAML with attribute sync.

To upload the SAML metadata file to your environment, see Upload a SAML metadata file.

How can I use SSO with SAML to assign groups/roles automatically to users?

In the SAML documentation, you can read how to set the attributes that Collibra expects. One of these attributes is to allow the IDP to send group information.

If you set the **Group** attribute to *user.group*, Collibra then parses the SAML Response to look for the *user.group* attribute. This attribute has to be a comma-separated list of groups to which this user has to be added.

Groups that do not exist, are created just in time. Users that are no longer part of a group per this definition are removed from that group.

Any groups that are empty after this operation are also removed.

Why is my Collibra user account removed from all groups when I log in with SSO?

In the context of group management, even if the Group attribute is not set, Collibra still assumes that you are trying to manage the groups from your IDP.

This question implies that you wish to use group management in Collibra itself. Set the option **Groups DGC managed** in **Security configuration > SSO** to *True*. Collibra then always assumes that group management is performed in Collibra itself.

Why is a new profile created each time I log in to Collibra?

Profiles are created when the username in the assertion is not yet found in the database.

You will notice that there is no <username> attribute linked. This is because the value Collibra uses to create and check usernames is actually the value of the **nameid** field in the SAML Response.

This field is governed by SAML and can ensure that the value is not Personally Identifiable Information (PII). The default nameid format used by Collibra is: `urn:oasis:names:tc:SAML:2.0:nameid-format:persistent`.

The problem with profiles being regenerated on every sign-on always has to do with this nameid. In some cases, the nameid is overridden by the IDP to be a random value, causing a user to be created many times. At other times, the IDP does not properly support our default nameid-format.

In order to fix this problem, first talk to your IDP team to figure out if the IDP is indeed supporting `urn:oasis:names:tc:SAML:2.0:nameid-format:persistent`. If it cannot support this in any case, Collibra can override the nameid-format used in the SAML Request.

> Warning Overriding the `nameid-format` is an advanced feature. Choosing to use it, makes the consequences your own responsibility.

If you still want to override the nameid-format, add the following tag in the <saml> section of the configuration.xml:

```
<force-nameid>Your fully qualified nameid-format</force-nameid>
```

Collibra expects you to provide the full string that describes the nameid that you want to use.

Configure Single Sign-On access

After you have decided which SSO mode you are going to use, you can configure Single Sign-On (SSO) for Collibra Data Intelligence Cloud.

> Note In your IDP configuration itself, make sure that the URL to your Collibra environment has a trailing slash, for example *https://dgc.yourcompany.com/*, without the trailing slash, SSO will not work in your environment.

Steps

1. Open the DGC service settings for editing:
2. In the **Security configuration** section, click **SSO**.
3. Enter the SSO parameters according to the selected SSO mode:
   - SAML_ATTRIBUTES, for a complete tutorial, see the knowledge base on Collibra Support Portal.
   - SAML_LDAP, for a complete tutorial, see the knowledge base on Collibra Support Portal.
   - SSO_HEADER
   - SSO_HEADER_LDAP
4. Click the green **Save all** button.

> Note
> On Collibra's Support Portal, you can find complete tutorials for the following SSO providers with SAML attributes:
>
> - Okta
> - OneLogin
> - PingOne Cloud
> - PingFederate
> - Azure
> - Active Directory Federation Service (ADFS)
> - ADFS with SAML LDAP mode

What's next?

Restart the environment to apply your changes. See Stop an environment and Start an environment.

SSO SAML with attributes: configuration options

To configure SSO as SAML with attributes sync, fill in the SSO fields as follows:

SSO configuration parameter	Value
Mode	*SAML_ATTRIBUTES*
Header	<leave empty>
DN	<leave empty>
Attribute	<leave empty>

SSO configuration parameter	Value
Metadata HTTP	URL to **saml.xml** file (http://url.to.your/saml.xml).

SSO configuration parameter	Value
Entity ID	The entity ID as defined in the metadata file. It defines which specific entity (IDP or SP) should be used in a metadata file. The SAML metadata file enables you to define multiple entities in one metadata file. This can also prove useful in combination with Collibra, in cases where planned upgrades are going to occur. You can then upload a new metadata file that contains both entities. When the time comes to switch, you only need to change the configuration option for the **Entity ID**.
Groups DC managed	• **False**: groups are managed by the SAML IDP. • **True**: groups are managed by DGC.
Service Provider Entity ID	Leave empty, unless the **Base URL** in **General settings** does not match the Service Provider Entity ID to be used.
Sign authentication requests	Set to **True** to use the SAML keypair to sign authentication requests. Note A SAML keypair in x509 is generated and stored in the SAML metadata file when Collibra is started for the first time.
Force authn	• **True** (default): Make the SAML request ask for authentication every single time, even when the user is already known. • **False**: Do not make the SAML request ask for authentication.
Force passive	• *False* (default): The IDP is allowed to take visible control of the user interface/authentication. • **True**: The IDP is forced to not take visible control. Only for specific setups, see the SAML 2.0 specifications for more information.

SSO configuration parameter	Value
Name ID	The **nameID** to be sent in the SAML Request. **nameID** has to have the following format: `urn:oasis:names:tc:SAML:2.0:nameid-format:persistent` For other options: See the SAML 2.0 specifications. IDP has to understand the nameID in the SAML Request. It is recommended to set this to what the IDP expects.
Name ID allow create	• **True** (default): Allow the IDP to create a new nameID to satisfy the SP SAML Request. • **False**: Do not allow the IDP to create a new nameID.
Disable	• **False** (default): Send the authentication context as configured in this section. • **True**: Disable the authentication context. Nothing in this section applies anymore. (see also Configuring requested authentication context
Comparison type	Defines the authentication strength that is to be used by the IDP compared to the SAML requested authentication context. This is advanced configuration, see the SAML 2.0 specifications for more information. Possible values: • *minimum* (default) • *maximum* • *better* • *exact* See also Configuring requested authentication context.

SSO configuration parameter	Value
Reference list	Contains a list of allowed references. Default: `urn:oasis:names:tc:SAML:2.0:ac:classes:PasswordProtectedTransport` This is advanced configuration, see the SAML 2.0 specifications for more information.
Declaration list	Similar to the Reference list, but is empty by default. This is advanced configuration, see the SAML 2.0 specifications for more information.
Ignore endpoint scheme differences	• **False**: The scheme is always taken into account for validating the SAML response. Any difference there invalidates the SAML response. • **True**: The scheme is not taken into account for validating the SAML response.

SSO SAML LDAP: configuration options

To configure SSO as SAML with LDAP user provisioning, fill in the SSO fields as follows:

SSO configuration parameter	Value
Mode	*SAML_LDAP*
Header	<leave empty>

SSO configuration parameter	Value
DN	If the **SSO mode** is **SSO_HEADER_LDAP** or **SAML_LDAP**, this field determines whether the distinguished name (DN) or attribute is used: • **True**: the header has to contain the distinguished name (DN) of the user in the LDAP. • **False**: the header has to contain the value of **Attribute**. If the **SSO mode** is **DISABLED**, **SSO_HEADER** or **SAML_ATTRIBUTES**, this field is ignored.
Attribute	Set to the unique identifier, usually uid of the LDAP directory when linking to LDAP through an LDAP attribute. The nameID contains the value of the attribute set here to look for in the LDAP service. For example, if the value equals *sAMAccountName*, the SAML response should contain the value for this attribute of the user being signed in.

SSO configuration parameter	Value
Metadata HTTP	Enter the URL to the SAML metadata file. For example: http://url.to.your/metadata.xml. Note This parameter is ignored if you uploaded the SAML metadata file.

SSO configuration parameter	Value
Entity ID	The entity ID as defined in the metadata file. It defines which specific entity (IDP or SP) has to be used in a metadata file. The SAML metadata file enables you to define multiple entities in one metadata file. This can also prove useful in combination with Collibra, in cases where planned upgrades are going to occur. You can then upload a new metadata file that contains both entities. When the time comes to switch, you only need to change the configuration option for the **Entity ID**.
Groups DC managed	• **False**: groups are managed by the SAML IDP • **True**: groups are managed by DGC
Service Provider Entity ID	Leave empty, unless the **Base URL** in **General settings** does not match the Service Provider Entity ID to be used.
Sign authentication requests	Set to **True** to use the SAML keypair to sign authentication requests. Note A SAML keypair in x509 is generated and stored in the SAML metadata file when Collibra is started for the first time.
Force authn	• *True* (default): Make the SAML request ask for authentication every single time, even when the user is already known. • **False**: Do not make the SAML request ask for authentication.
Force passive	• **False** (default): The IDP is allowed to take visible control of the user interface/authentication. • **True**: The IDP is forced to not take visible control. Only for specific setups, see the • SAML 2.0 specifications for more information.

SSO configuration parameter	Value
Name ID	The **nameID** to be sent in the SAML Request. **nameID** has to have the following format: `urn:oasis:names:tc:SAML:2.0:nameid-format:persistent` For other options: See the SAML 2.0 specifications. IDP has to understand the nameID in the SAML Request. It is recommended to set this to what the IDP expects.
Name ID allow create	• **True** (default): Allow the IDP to create a new **nameID** to satisfy the SP SAML Request. • **False**: Do not allow the IDP to create a new **nameID**.
Disable	• **False** (default): Send the authentication context as configured in this section. • **True**: Disable the authentication context. Nothing in this section applies anymore. See also Configuring requested authentication context.
Comparison type	Defines the authentication strength that is to be used by the IDP compared to the SAML requested authentication context. This is advanced configuration, see the SAML 2.0 specifications for more information. Possible values: • *minimum* (default) • *maximum* • *better* • *exact* See also Configuring requested authentication context.

SSO configuration parameter	Value
Reference list	Contains a list of allowed references. Default: `urn:oasis:names:tc:SAML:2.0:ac:classes:PasswordProtectedTransport` This is advanced configuration, see the SAML 2.0 specifications for more information.
Declaration list	Similar to the Reference list, but is empty by default. This is advanced configuration, see the SAML 2.0 specifications for more information.
Ignore endpoint scheme differences	• **False**: The scheme is always taken into account for validating the SAML response. Any difference there invalidates the SAML response. • **True**: The scheme is not taken into account for validating the SAML response.

When both **Attribute** and **DN** are defined, **DN** takes priority and the attribute-based configuration is ignored.

SSO header: configuration options

To configure SSO in SSO header mode, fill in the SSO fields as follows:

SSO configuration parameter	Value
Mode	*SSO_HEADER*
Header	The header to be monitored for the username to be signed in.

SSO configuration parameter	Value
DN	<does not apply>
Attribute	<does not apply>

SSO Header LDAP: configuration options

To configure SSO in SSO header mode with LDAP user provisioning, fill in the SSO fields as follows:

SSO configuration parameter	Value
Mode	*SSO_HEADER_LDAP*
Header	The header to be monitored for the username to be signed in.
DN	• **True**: The **nameID** in the SAML response refers to the DN (Distinguished Name) of the user to be signed in. • **False**: Fill in the attribute field. Set to true when linking to LDAP through the Distinguished Name.
Attribute	Set to the unique identifier, usually uid of the LDAP directory when linking to LDAP through an LDAP attribute. The nameID contains the value of the attribute set here to look for in the LDAP service. For example, if the value equals *sAMAccountName*, the SAML response should contain the value for this attribute of the user being signed in.

> Note When both **Attribute** and **DN** are defined, **DN** takes priority and the attribute-based configuration is ignored.

Configuring requested authentication context

You can configure in which context requested authentication happens on IDP side by changing the following settings:

- **Disable**
- **Comparison type**
- **Reference list**
- **Declaration list**

If you leave the above settings as they are and Collibra Data Intelligence Cloud is configured for SAML, Collibra sends a requested authentication context with:

- The **exact** comparison
- The **urn:oasis:names:tc:SAML:2.0:ac:classes:PasswordProtectedTransport** class-ref
- No class-decl

Enable SAML response encryption

To increase the security of the communication between Collibra Data Intelligence Cloud and an SSO provider, Collibra supports encrypted SAML responses.

Prerequisites

- You can access the Collibra REST API.

Steps

Enable response decryption

1. Open the DGC service settings for editing:
2. In the **Security configuration** section, click **SSO**.
3. Set the option **Response decryption mode** to *OPTIONAL* or *FORCED*.
   An encryption key pair is generated and added to the SAML keystore. A self-signed encryption certificate is generated and works in most situations.

4. Click the green **Save all** button.

Provide encryption certificate to Identity Provider

The encryption certificate of Collibra has to be imported in the Identity Provider. You can retrieve this encryption certificate via the Collibra REST API.

1. Retrieve the certificate via the endpoint at `/rest/2.0/security/saml` and copy the base64 representation of the encryption certificate.

2. Copy the content of the **ds:X509Certificate** element into a PEM file. A PEM file is a plain text file with the extension **pem**.

```
-----BEGIN CERTIFICATE-----
MIICrDCCAZSgAwIBAgI...
-----END CERTIFICATE-----
```

3. Provide this PEM file to an administrator of the Identity Provider who can load it into the IdP.

What's next?

Your Collibra Data Intelligence Cloud environment is configured to accept encrypted SAML responses.

If your Identity Provider does not accept self-signed certificates, contact Collibra Support.

Encrypted SAML response

Collibra Data Intelligence Cloud supports encrypted SAML responses. Collibra implements XMLEnc, the industry standard to encrypt SAML responses.

Encryption

SAML assertions are expected to be encrypted in XMLEnc mixed mode, which means that:

1. The Identity Provider (IdP) generates a random symmetric key and uses it to encrypt the assertion.
2. The symmetric key is encrypted using a public key provided by the Service Provider (SP), in this case Collibra.
3. The encrypted symmetric key is embedded into the SAML response alongside both the public key used in its encryption and encrypted assertion.
4. When Collibra receives the response, it decrypts the symmetric key using its own private key and then uses that symmetric key to decrypt the assertion.

Supported cryptographic algorithms

Different IdPs may use different cryptographic schemes and algorithms. As a Service Provider, Collibra has to decrypt the assertions of many IdPs, so Collibra supports the algorithms that are marked as "required" in the XMLEnc specifications.

The supported algorithms for symmetric cryptography are:

- 3DES
- AES-128-CBC
- AES-256-CBC
- AES-128-GCM
- AES-256-GCM

The supported algorithm for asymmetric cryptography is RSA-OAEP, including MGF1 with SHA1.

Collibra recommends to integrate with an IdP that uses AES-256-GCM for symmetric encryption.

Configure a custom certificate for SSO in Collibra

If you configure single sign-on for accessing Collibra Data Intelligence Cloud, a default certificate is used. You can use this certificate for signing SAML authn requests or for SAML assertion encryption/decryption.

Instead of the default certificate, you can use your own certificate. However, keep in mind that you can only configure SSO with your own certificate via a REST API call.

Prerequisites

- The certificate must meet the following requirements:
  - The certificate must be in PEM format.
  - The PEM file must be unencrypted (no password).
  - The PEM file must contain the server certificate the private key of that certificate.

> Tip To convert a key to a PEM key: `openssl rsa -in <pem-key>.key -out <rsa-key>.pem`

> Example PEM file:
>
> ```
> -----BEGIN PRIVATE KEY-----
> MIIEvgIBADA ... bml6YXRpb252YWxza
> ....
>  z3P668YfhUbKdRF6S42Cg6zn
> -----END PRIVATE KEY-----
>
> # Your certificate
>  -----BEGIN CERTIFICATE-----
> MIIFaDCCBFC ... bml6YXRpb252YWxza
> ...
>
> lffygD5IymCSuuDim4qB/9bh7oi37heJ4ObpBIzroPUOthbG4gv/5blW3Dc
> =
> -----END CERTIFICATE-----
> ```

- A base64 encoding hash of an API user.

  > Tip You can calculate the base64 hash of the user as follows: `echo '<username>:<password>' | base64`, for example `echo 'apiUser:apiUserpassword' | base64` results in `YXBpVXNlcjphcGlVc2VycGFzc3dvcmQK`

Use certificate for SAML assertion encryption

```
curl --location --request POST \
  'https://<your_dgc_environment_url>/rest/2.0/se-
curity/saml/certificate/ENCRYPTION' \
  --header 'Authorization: Basic <base-64 encoding hash>' \
  --form 'file=@"/path/to/pem-file"'
```

Use certificate for sigining SAML authn requests

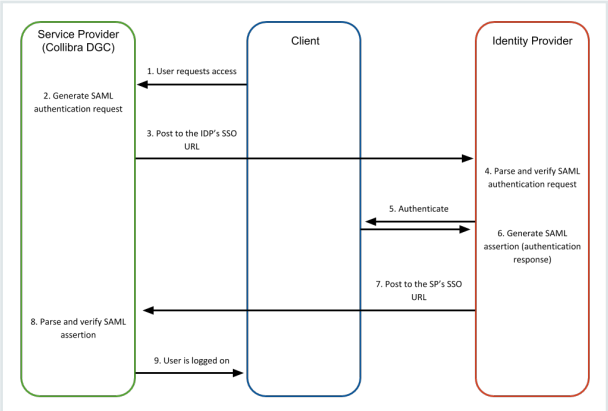
```
curl --location --request POST \
  'https://<your_dgc_environment_url>/rest/2.0/se-
curity/saml/certificate/SIGNING' \
  --header 'Authorization: Basic <base-64 encoding hash>' \
  --form 'file=@"/path/to/pem-file"'
```

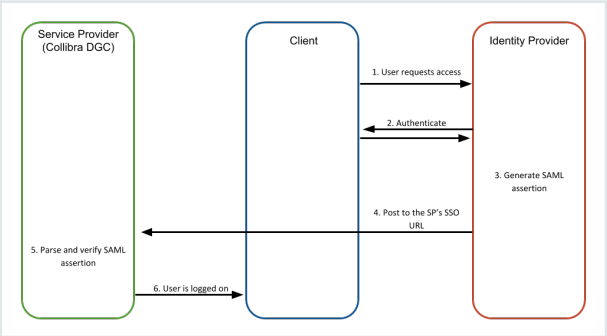
Working with SAML

SAML or Security Assertion Markup Language is an open standard that allows security credentials to be shared by multiple computers across a network. It describes a framework that allows one computer to perform some security functions on behalf of one or more other computers.

For more information, see the website of the OASIS Security Services (SAML) Technical Committee.

Terminology

Term	Explanation
SP	The Service Provider of the SAML integration. In this case, the SP is Collibra Data Intelligence Cloud.
IDP	The Identity Provider of the SAML integration. In this case , the IDP depends on your company. Examples of IDPs: Microsoft's Active Directory (if the federation services are installed), PingFederate, SiteMinder, Okta, OneSign...
SP-initiated sign-on	A type of sign-on that starts from the SP (Collibra). In this flow, the SP asks the IDP to authenticate the current user, using a SAML request. The IDP then returns a SAML response. This SAML response contains the assertion of the user currently visiting the SP. That assertion contains the properly configured attributes of the user that is requested. > Note If the assertion's **SubjectConfirmationData** element contains the **Address** attribute, ensure that the value is a URL. This value indicates the expected network address of the client. You can enable or disable the validation of the address attribute in Console. 

Term	Explanation
IDP-initiated sign-on	A type of sign-on that starts from the IDP portal towards the SP. This flow skips the SAML request sent in the SP-initiated flow, but still sends the SAML Response containing an assertion containing the properly configured attributes of the user that is requested. 

Supported SAML 2.0 profile

Although SAML 2.0 is a clear standard, it is too large to support in its entirety. The designers of the SAML standard also realized this and that's why they developed SAML Profiles. These profiles are defined in the SAML 2.0 specification's section.

The purpose of SAML profiles is to allow both IDP and SP to decide to only support a subset in the SAML protocol that is applicable to them, in their situation.

Collibra supports the Web Browser SSO Profile.

How does Collibra implement SAML?

Even though the web browser profile is properly defined in the standard's profile section, there are a lot of optional features, but not all of them are implemented. This section explains what is implemented and what not.

The profile defines that metadata may be used to build trust between SP and IDP. Collibra has chosen to enforce this, so the only way to integrate your IDP with Collibra is for the IDP to provide the metadata to Collibra. This can be done by either uploading a SAML metadata file, or by providing a public URL that contains the metadata.

The profile explains that signing and/or encryption of SAML requests and responses can be used and that Collibra expects the IDP to perform the signing and/or encryption. Since the communication between SP and IDP is always supposed to happen in the context of an HTTPS connection, encryption is not performed. The transport layer takes care of that. By default, SAML requests are not signed, but you can enable signed requests. SAML responses always have to be signed because the response has to come from the correct IDP.

The validUntil property in the SAML metadata

The SAML metadata may contain a `validUntil` attribute. This optional attribute indicates the expiration time of the metadata contained in the element, as well as its child elements.

The "valid until" field is taken into account by Collibra. SAML will stop working either when the validity period has expired or when the content in the metadata file is no longer valid.

Steps for integrating SAML in Collibra

Step	Task	Description
1	Bring in the IDP's SAML metadata in Collibra Console	Collibra needs the metadata of your IDP. You can bring in the metadata in two ways: • Upload the IDP's SAML metadata file in Collibra Console. This is the easiest approach, but the metadata will not be updated automatically if there are changes in the IDP. • Enter the URL to the IDP's metadata in the **Metadata HTTP** field of the SAML configuration of the SSO section of the environment settings. This requires the IDP to have a publicly accessible URL containing the SAML metadata. If the IDP's metadata changes, Collibra automatically updates the metadata as well.

Step	Task	Description
2	Configure DGC	Configure the environment settings for SSO SAML with attributes or SSO SAML with LDAP. We strongly recommend that you enable **Sign authentication requests**. If you do not enable **Sign authentication requests**, Collibra's outgoing requests are not signed. If your IDP demands that all requests are signed, the unsigned requests from Collibra will be ignored.
3	Restart Collibra	Stop and start the affected environment. This will generate a key pair, which is required to sign authentication requests.
4	Bring in Collibra's SAML metadata in your IDP	Your IDP needs the metadata of Collibra. Your IDP team should help with this. You can obtain the metadata in two ways: • Sign in and download the metadata file from Collibra, and import it in your IDP. You can download the metadata file here: `https://<your dgc instance>/rest/1.0/saml/metadata` • Enter the URL to the Collibra's metadata in the IDP: `https://<your dgc instance>/rest/1.0/saml/metadata`. Your IDP needs to be connected to the internet to access that URL.
5	Configure your IDP	You IDP may require further configuration. Your IDP team should help with or perform the complete configuration.

Upload a SAML metadata file

If you want to use SAML for user authentication, you need to provide your IDP's SAML metadata file.

If there is already a SAML metadata file, uploading another file will overwrite the existing one.

Steps

1. Open Collibra Console with a user profile that has at least the **ADMIN** role.
   » Collibra Console opens with the **Infrastructure** page.
2. Click the name of an environment to show its details.
3. Click the **SAML** tab.
4. Click **Upload**.
   » The **Upload SAML metadata** file dialog box appears.
5. Do one of the following:
   - Drag and drop a valid SAML metadata file in the **Upload a file...** field.
   - Click in the **Upload a file...** field, select the SAML metadata file and click **Open**.

What's next

Configure the environment settings for SSO SAML with attributes or SSO SAML with LDAP.

Download a SAML metadata file

You can download the IDP's SAML metadata file. For example, this can be used when you want to create a backup.

Prerequisites

- You have uploaded a SAML metadata file.

Steps

1. Open Collibra Console with a user profile that has the **SUPER** role.
   » Collibra Console opens with the **Infrastructure** page.
2. Click the name of an environment to show its details.
3. Click the **SAML** tab.
4. Click .
   » The SAML metadata file is downloaded.

Delete a SAML metadata file

You can delete a SAML metadata file from your configuration.

This is mandatory when you want to start using an online SAML metadata provider in the **Metadata HTTP** field of the SAML configuration. If you do not delete the SAML metadata file, the URL is ignored.

Prerequisites

- You have uploaded a SAML metadata file.

Steps

1. Open Collibra Console with a user profile that has the **SUPER** role.
   » Collibra Console opens with the **Infrastructure** page.
2. Click the name of an environment to show its details.
3. Click the **SAML** tab.
4. Click .
   » The **Delete SAML metadata** file dialog box appears.
5. Click **Delete**.

What's next

Enter the URL of the IDP's metadata in the **Metadata HTTP** field of the SAML configuration of the SSO section of the environment settings.

JSON Web Token authentication

You can use JSON Web Token (JWT) authentication for your applications that interact with the Collibra REST API. During this process, your application requests an access token from your Identity Provider (IdP). The IdP acts as the authentication server and returns a signed JWT access token. When your application makes API calls to Collibra, it provides the JWT access token as a **Bearer** token in the HTTP **Authorization** header.

Example

```
curl -H 'Authorization: Bearer <your_token>' ...
```

Benefits of using JWT token authentication include:

- Keeping the authentication details at the IdP and separate from Collibra.
- Simplifying your security management.
- Limiting the time a token is valid for.

Set up and use JWT

The following table summarizes acquiring and using JWT in Collibra REST API requests.

Process	Steps
Initial setup	1. Create a client credential account with a secret in your IdP. 2. Determine the JSON Web Key Set (JWKS) endpoint URL for your IdP. 3. Register the JWKS endpoint with Collibra. 4. Create a user in Collibra for your client application account. Tip Provide a meaningful first and last name to identify that this is a service account.

Process	Steps
When your application starts	1. Authenticate your client application with your IdP. 2. Save the returned access token for use in REST API calls.
When your application calls the Collibra REST APIs	1. Include the JWT token in the authorization HTTP header as a bearer token. 2. If the API call responds with **unauthorized**, the access token or JWKS credentials may have expired. Re-authenticate and retry the request.

JWT dependencies on your Identity Provider

The following details must be provided by your IdP in the JWT token for Collibra to accept the token.

Section	Field	Description
Header	alg	The encryption algorithm. Collibra supports all standard encryption algorithms.
	kid	The JWKS identifier of the public key used to sign the JWT token. An identity provider may have multiple public key certificates, for example multiple valid keys may exist while a key rotation is in progress.
	typ	Optional field that defines the JWT type. Tip Use the MIME type format, for example *application/jwt* or *application/secevent+jwt* with the *application/* prefix removed, for example *jwt* or *secevent+jwt*.

Section	Field	Description
Payload	iss	The issuer of the token. This field is used to check the token comes from the expected IdP.
	exp	The token expiry time.
	iat	The time the token was issued.
	sub	The subject or the principal ID.
Signature		The digital signature of the header and payload. The signature verifies the message did not change along the way. In the case of tokens signed with a private key, the signature also verifies the authenticity of the JWT sender.

Configure JWT settings

To enable or change the JSON Web Token configuration:

1. Open the DGC service settings for editing:
2. In the **JWT** section, make the necessary changes.

Setting	Description
JSON Web Key Set URL	The URL to retrieve public key information needed to verify the authenticity of JSON Web Tokens (JWTs), issued by an authorization server. This setting is required to enable JWT authentication.

Setting	Description
JWT Token Types	A case-insensitive comma-separated list of accepted JWT media types coming in the **typ** header parameter. Leave blank if the authorization server does not provide a media type parameter. The default values is **at+jwt,jwt**.
JWT Algorithms	A comma-separated list of accepted JWT algorithms coming in the **alg** header parameter. See https://tools.ietf.org/html/rfc7518#section-3.1 for details. Leave blank to accept all digital signature algorithms.
JWT Issuer	The accepted issuer coming in the **iss** JWT claim. Leave blank if the authorization server does not provide an issuer claim.
JWT Audience	A comma-separated list of accepted audience values for the **aud** claim. The value for this field is a configuration setting in your authorization server, which identifies your Collibra environment as the intended recipient of the JWT. Leave blank if the authorization server does not provide an audience claim.

Setting	Description
JWT Principal ID Claim Name.	The name of the JWT claim containing the principal's identity. See https://tools.ietf.org/html/rfc7519#section-4.1.2 for details. Defaults to the standard subject claim, **sub**. Change this setting only if your authorization server has other means of identifying the principal, for example, a **client_id** claim. This setting is required if JWT authentication is enabled.
JWT Maximum Clock Skew	The maximum acceptable difference in seconds between the clocks of the machines running the authorization server and Collibra. Differences smaller than the given amount are ignored when performing time comparisons for token validation. The default value is 60 seconds if left blank.

3. Click the green **Save all** button.
4. Restart the environment to apply your changes. See Stop an environment and Start an environment.

JWT troubleshooting

The following table contains JWT authentication error codes that may appear in the body of the **401 Unauthorized** HTTP response. You can use the error codes to determine the appropriate course of action for your application.

Error code	Description	Possible action
malformedToken	The JWT token is incorrectly encoded or has an incorrect syntax.	Check your IdP configuration to make sure it is returning a token in JWT format.

Error code	Description	Possible action
expiredToken	The JWT token has expired.	In your application, request a new token from the IdP and retry.
invalidToken	The JWT token is invalid. Common causes are: • Bad signature. • Wrong audience. • Wrong issuer. • Wrong media type. • Disallowed signing algorithm.	Check your IdP and Collibra JWT configuration to make sure the settings are consistent.
unableToProcessToken	The JWT token could not be processed.	Check your IdP and Collibra JWT configuration and if the problem persists contact Collibra Support.

Configure Collibra Connect

We have made the decision to transition away from Collibra Connect to provide customers a wider range of integration options.

Our native Collibra integrations (connectors) will be easier to implement and maintain, provide a better return on investment, and allow you to grow with and derive greater value from Collibra:

- Collibra integrations and Spring Boot based frameworks will replace Collibra Connect as options to build integrations going forward.
- You can choose any ESB or integration method for your use case.
- Our intention is to enable Collibra connectors to support ingestion as well as use cases for data profiling, data classification and other cloud functionalities.

- If you have an enterprise MuleSoft license, you can easily switch to it. For details on how to switch from Connect licenses to MuleSoft licenses see this Collibra Support article.

Rest assured Connect templates are and will remain compatible with our product, please contact us for any Connect-related question. Only support or any upgrades on these products will be discontinued.

> Note As of September 2022, you will need a MuleSoft Community Edition license or your own proprietary paid license to run Connect templates.

Resources:

- Spring Boot library.
- Spring Boot templates.
- Custom integrations.
- Learn more about different methods to build integrations.

With Collibra Connect, you can connect to Collibra Data Intelligence Cloud with a tool of your own choice. Collibra Connect acts as the gateway between your tool and Collibra. For more details on Collibra Connect, consult the Collibra Connect user guide.

In this section, you can learn how to set the credentials to access Collibra Connect with your own tool.

> Note Depending on your environment, you have to follow this procedure either in the Services Configuration section of the Collibra settings or in Collibra Console. This topic shows the procedure in Collibra Console. If your Settings page has a **Configuration** tab, you can follow this procedure in that tab page. For more information, see the online documentation.

Prerequisites

- You have the ADMIN or SUPER role in Collibra Console.
- You have the SUPER role in Collibra Console.

Steps

1. Open the DGC service settings for editing:
   a. Open Collibra Console.
      » Collibra Console opens with the **Infrastructure** page.
   b. In the tab pane, expand an environment to show its services.
   c. In the tab pane, click the Data Governance Center service of that environment.
   d. Click **Configuration**.
   e. Click **Edit configuration**.
2. Go to the **Data Classification** section.
3. Enter the required information:

Setting	Description
Machine Learning platform URL (*)	The address of the machine learning platform that will classify your data.
Requester Name (*)	The unique name to identify the client when using Machine Learning platform.
API key (*)	The API Key to authorize the requester when connecting to the Machine Learning platform.
Enable Data Classification	◦ ✔ True: Enable Collibra's data classification technology. ◦ ✖ False (default): Do not use Collibra's data classification technology are not accepted.

4. Click the green **Save all** button.

Edit global data source registration settings

In Collibra Data Intelligence Cloud, you can create a data profile when you register a data source.You can configure global settings of data source registration.

More information: Registering a data source in the User Guide.

> Note Depending on your environment, you have to follow this procedure either in the Services Configuration section of the Collibra settings or in Collibra Console. This topic shows the procedure in Collibra Console. If your Settings page has a **Configuration** tab, you can follow this procedure in that tab page. For more information, see the online documentation.

Prerequisites

- You have the ADMIN or SUPER role in Collibra Console.
- You have the SUPER role in Collibra Console.

Steps

1. Open the DGC service settings for editing:
   a. Open Collibra Console.
      » Collibra Console opens with the **Infrastructure** page.
   b. In the tab pane, expand an environment to show its services.
   c. In the tab pane, click the Data Governance Center service of that environment.
   d. Click **Configuration**.
   e. Click **Edit configuration**.

2. In the section **Data Ingestion**, make the necessary changes.

Setting	Description
Table types to ignore	A comma separated list of table types that are not ingested. For example, *INDEX* and *SEQUENCE*.
AWS regions restriction	A list of AWS regions Data Catalog is allowed to connect to. For example, *eu-west-3* and *us-east-2*. For a list of all AWS locations, see the AWS documentation. If you want to allow Collibra to make a connection to any AWS region, leave the field empty.
AWS API call rate	Allowed number of AWS API calls per second. Use this option to limit the number of API calls per second to prevent throttling errors from the AWS API.
Database registration via Edge	An option to enable database registration via Edge. ◦ ✔ True: Register a data source via Edge. ◦ ✖ False: Register a data source via Jobserver only. Note Enabling data source registration via Edge does not prevent you from registering a data source via Jobserver as well.
Data Quality Synchronization UI via DQ Connector on Edge	An option to enable the Data Quality extraction interface in Collibra ◦ ✔ True: The Quality extraction tab is available on the configuration page of a database asset ◦ ✖ False (default): The Quality extraction tab is not available and as such, it is not possible to extract and synchronize data quality information. You can only enable Data Quality synchronization if you also enabled Database registration via Edge.

3. Click the green **Save all** button.

Add a Jobserver to the DGC service

To register a data source and create a data profile in Collibra Data Intelligence Cloud, you need the Jobserver service.

If you don't have a Jobserver installed and configured in your environment, the **Register data source** action will be grayed out in the global create menu of Collibra Data Intelligence Cloud.

> Tip Execute this procedure on Collibra Console of your cloud environment. In this configuration, the DGC service will send jobs to the on-premises Jobserver, however we highly recommend to revert this communication path so that the Jobserver polls the DGC service for jobs.

Steps

1. Open the DGC service settings for editing:
2. In the **Jobserver** section, click **Add**.

3. Enter the necessary information:

Setting	Description
Jobserver list	The list of registered Jobserver instances.
Name	The name of the Jobserver as it will appear when you register a data source in Data Catalog. The name is a freely chosen name but it is recommended to only use alphanumerical characters and dashes, for example Jobserver-1. You will have to use this name as the ID of the gateway and in the address of this configuration.
Protocol	The protocol that is used for the communication between the Data Governance Center service and the Jobserver service. It is recommended to use HTTPS, especially if the services are hosted in different network segments.
Address	The address (IP address, URL, hostname) of the Jobserver.
Trusted server CA certificate	The certificate of the trusted CA needed to validate the server certificate. If blank, the default truststore will be used. The default truststore is defined in the SSL configuration section of the DGC service. The CA certificate of the server party (Jobserver).
Client certificate	The client certificate offered by the DGC service to the server. If blank, you cannot select mutual authentication as the Jobserver service authentication level.
Client private key	The private key of the DGC service's certificate.

Setting	Description
Table profiling data size	The approximate maximum disk size of the data in MB that will be used to profile a table. The value cannot exceed 10,000.
Test connection timeout	This timeout is a time limit (in seconds) after which the connection test is stopped and a timeout error is shown. The default value is 60 seconds.

4. Click the green **Save all** button.

> Tip You can add as many Jobserver services as you want.

Multiple Jobserver services

You can add more that one Jobserver service to one Data Governance Center service. To get the best performance out of the Jobserver service, is to install the service as close to the data source as possible, preferably in the same network to reduce the latency as much as possible.

If you use multiple data sources in different locations, you can install a Jobserver service in every installation. During the data source registration, you can then select the Jobserver that is closest to the selected data source.

> Note
> - A Jobserver can only run one ingestion job at a time. Multiple Jobservers can run multiple ingestion jobs in parallel.
> - In a Collibra Data Intelligence Cloud environment with an on-premises Tableau server, you can use an on-premises Jobserver to ingest Tableau data. Keep in mind that this Jobserver is dedicated for the Tableau server and that it cannot be used for ingesting JDBC or Amazon S3 data sources.

Configure data profiling behavior

In Collibra Data Intelligence Cloud, you can create a data profile when you register a data source. You can configure the behavior of data profiling.

> Note Depending on your environment, you have to follow this procedure either in the Services Configuration section of the Collibra settings or in Collibra Console. This topic shows the procedure in Collibra Console. If your Settings page has a **Configuration** tab, you can follow this procedure in that tab page. For more information, see the online documentation.

Prerequisites

- You have the ADMIN or SUPER role in Collibra Console.
- You have the SUPER role in Collibra Console.

Steps

1. Open the DGC service settings for editing:
   a. Open Collibra Console.
      » Collibra Console opens with the **Infrastructure** page.
   b. In the tab pane, expand an environment to show its services.
   c. In the tab pane, click the Data Governance Center service of that environment.
   d. Click **Configuration**.
   e. Click **Edit configuration**.
2. In the section **Data Profiling**, make the necessary changes.

Setting	Description
Maximum number of samples	The maximum number of rows taken as a sample during profiling.
Maximum value length	The maximum length of a value extracted during profiling or sampling. Additional characters are trimmed.

Setting	Description
Default date pattern	The default format used to decode dates. It is the default pattern used for detecting dates when the **Date Pattern** and/or **Time Pattern** attribute is not specified in Column assets.
Default time pattern	The default format used to decode times. It is the default pattern used for detecting times when the **Date Pattern** and/or **Time Pattern** attribute is not specified in Column assets.
Default combined date and time pattern	The default format used to decode combined dates and times. It is the default pattern used for detecting combined dates and times when the **Date Pattern** and/or **Time Pattern** attribute is not specified in Column assets.
Empty values	A comma separated list of strings enclosed in double quotes. A value that matches one of those expressions is considered an empty value. Please note that a database **null** value is always considered an empty value, for example *""*, *"na"* and *"none"*.
Data type detection threshold	The percentage of matching Column values to reach for an Advanced Data Type to be considered a possible Data Type for that Column. This is expressed as a value between 0.0 and 1.0).

Setting	Description
Anonymize data	An option to anonymize sensitive data. ◦ ✔ True: Content in columns with data type Text or Geo is removed or replaced by a random hash value before the profiling results are sent to the cloud. ◦ ✖ False (default): No content is removed or replaced by a random hash value. Tip If you profile and classify via Edge, the data in columns with data type Text or Geo is automatically anonymized before it is sent to Collibra Data Intelligence Cloud.
Database profiling via Edge	An option to enable profiling and classifying synchronized metadata via Edge instead of Jobserver. ◦ ✔ True: Profiling and classify via Edge. ◦ ✖ False: Profile via Jobserver and classify via the Data Classification Platform. Note You can only enable Database profiling via Edge if you also enabled Database registration via Edge.

3. Click the green **Save all** button.

Enable Tableau provisioning

You can enable Tableau provisioning to allow users to create Tableau provisioning files from Data Catalog data sets.

> Note Depending on your environment, you have to follow this procedure either in the Services Configuration section of the Collibra settings or in Collibra Console. This topic shows the procedure in Collibra Console. If your Settings page has a **Configuration** tab, you can follow this procedure in that tab page. For more information, see the online documentation.

Prerequisites

- You have the ADMIN or SUPER role in Collibra Console.
- You have the SUPER role in Collibra Console.

Steps

1. Open the DGC service settings for editing:
    a. Open Collibra Console.
        » Collibra Console opens with the **Infrastructure** page.
    b. In the tab pane, expand an environment to show its services.
    c. In the tab pane, click the Data Governance Center service of that environment.
    d. Click **Configuration**.
    e. Click **Edit configuration**.
2. In the **Beta Features** section, set the **Tableau provisioning enabled** field to **True**.
3. Click the green **Save all** button.

Configure Automatic Data Classification

When you want to use Automatic Data Classification in Data Catalog, you first have to configure the Automatic Data Classification.

> Note Depending on your environment, you have to follow this procedure either in the Services Configuration section of the Collibra settings or in Collibra Console. This topic shows the procedure in Collibra Console. If your Settings page has a **Configuration** tab, you can follow this procedure in that tab page. For more information, see the online documentation.

Prerequisites

- You have the ADMIN or SUPER role in Collibra Console.
- You have the SUPER role in Collibra Console.

Steps

1. Open the DGC service settings for editing:
   a. Open Collibra Console.
      » Collibra Console opens with the **Infrastructure** page.
   b. In the tab pane, expand an environment to show its services.
   c. In the tab pane, click the Data Governance Center service of that environment.
   d. Click **Configuration**.
   e. Click **Edit configuration**.
2. Go to the **Data Classification** section.
3. Enter the required information:

Setting	Description
Machine Learning platform URL (*)	The address of the machine learning platform that will classify your data.
Requester Name (*)	The unique name to identify the client when using Machine Learning platform.
API key (*)	The API Key to authorize the requester when connecting to the Machine Learning platform.
Enable Data Classification	◦ ✔ True: Enable Collibra's data classification technology. ◦ ✖ False (default): Do not use Collibra's data classification technology are not accepted.

4. Click the green **Save all** button.

Enable or disable Catalog experience

Catalog experience allows you to use the improved user experience in the Data Catalog asset pages.

> Note Depending on your environment, you have to follow this procedure either in the Services Configuration section of the Collibra settings or in Collibra Console. This topic shows the procedure in Collibra Console. If your Settings page has a **Configuration** tab, you can follow this procedure in that tab page. For more information, see the online documentation.

Prerequisites

- You have the ADMIN or SUPER role in Collibra Console.
- You have the SUPER role in Collibra Console.

Steps

1. Open the DGC service settings for editing:
    a. Open Collibra Console.
        » Collibra Console opens with the **Infrastructure** page.
    b. In the tab pane, expand an environment to show its services.
    c. In the tab pane, click the Data Governance Center service of that environment.
    d. Click **Configuration**.
    e. Click **Edit configuration**.

2. In the **Catalog Experience** section, enter the required information:

Setting	Description
Enable Catalog experience	◦ ✔ True: Catalog experience is enabled. This will improve the layout of Data Catalog's asset pages, such as those of Data Set, Schema, Table and Column assets. ◦ ✖ False: Catalog experience is disabled.
Catalog Experience Titlebar theme	The theme for the Catalog experience. You can choose between the LIGHT and DARK. This option is only applicable if the **Enable Catalog experience** option is enabled.

3. Click the green **Save all** button.

Enable registering a data source via Edge

You can enable Edge to register a data source. When you register a data source via Edge, an Edge site ingests data into Data Catalog instead of the Jobserver.

> Note Depending on your environment, you have to follow this procedure either in the Services Configuration section of the Collibra settings or in Collibra Console. This topic shows the procedure in Collibra Console. If your Settings page has a **Configuration** tab, you can follow this procedure in that tab page. For more information, see the online documentation.

Prerequisites

- You have the ADMIN or SUPER role in Collibra Console.
- You have the SUPER role in Collibra Console.

Steps

1. Open the DGC service settings for editing:
   a. Open Collibra Console.
      » Collibra Console opens with the **Infrastructure** page.
   b. In the tab pane, expand an environment to show its services.
   c. In the tab pane, click the Data Governance Center service of that environment.
   d. Click **Configuration**.
   e. Click **Edit configuration**.
2. In the **Register data source** section, enter the required information:

Setting	Description
Database registration via Edge	An option to enable database registration via Edge. ◦ ✔ True: Register a data source via Edge. ◦ ✖ False: Register a data source via Jobserver only. Note Enabling data source registration via Edge does not prevent you from registering a data source via Jobserver as well.

3. Click the green **Save all** button.

Enable profiling and classification via Edge

You can enable Edge to profile and classify synchronized metadata in Data Catalog.

Note Depending on your environment, you have to follow this procedure either in the Services Configuration section of the Collibra settings or in Collibra Console. This topic shows the procedure in Collibra Console. If your Settings page has a **Configuration** tab, you can follow this procedure in that tab page. For more information, see the online documentation.

Prerequisites

- You have the ADMIN or SUPER role in Collibra Console.
- You have the SUPER role in Collibra Console.

Steps

1. Open the DGC service settings for editing:
   a. Open Collibra Console.
      » Collibra Console opens with the **Infrastructure** page.
   b. In the tab pane, expand an environment to show its services.
   c. In the tab pane, click the Data Governance Center service of that environment.
   d. Click **Configuration**.
   e. Click **Edit configuration**.
2. In the **Data profiling** section, enter the required information:

Setting	Description
Database profiling via Edge	An option to enable profiling and classifying synchronized metadata via Edge instead of Jobserver. ◦ ✔ True: Profiling and classify via Edge. ◦ ✖ False: Profile via Jobserver and classify via the Data Classification Platform. Note You can only enable Database profiling via Edge if you also enabled Database registration via Edge.

3. Click the green **Save all** button.

Enable data quality synchronization via Edge

You can enable the Quality extraction tab on a Database asset page in Data Catalog.

> Note Depending on your environment, you have to follow this procedure either in the Services Configuration section of the Collibra settings or in Collibra Console. This topic shows the procedure in Collibra Console. If your Settings page has a **Configuration** tab, you can follow this procedure in that tab page. For more information, see the online documentation.

Prerequisites

- You have the ADMIN or SUPER role in Collibra Console.
- You have the SUPER role in Collibra Console.

Steps

1. Open the DGC service settings for editing:
   a. Open Collibra Console.
      » Collibra Console opens with the **Infrastructure** page.
   b. In the tab pane, expand an environment to show its services.
   c. In the tab pane, click the Data Governance Center service of that environment.
   d. Click **Configuration**.
   e. Click **Edit configuration**.
2. In the **Data profiling** section, enter the required information:

Setting	Description
Database profiling via Edge	An option to enable profiling and classifying synchronized metadata via Edge instead of Jobserver. ◦ ✔ True: Profiling and classify via Edge. ◦ ✖ False: Profile via Jobserver and classify via the Data Classification Platform. Note You can only enable Database profiling via Edge if you also enabled Database registration via Edge.

3. Click the green **Save all** button.

Enable Tableau metadata API

You can enable the Tableau metadata API if you want to ingest Tableau 2020.2 or newer in Data Catalog.

> Note Make sure that the Tableau metadata API is enabled in Tableau before you ingest or synchronize Tableau in Data Catalog.

> Warning If you upgrade to Tableau version 2020.2 or newer, but previously synchronized an older Tableau version via the REST API and XML mapping, you have to prepare the migration procedure to prevent losing manually added relations, attributes, tags, comments and stitching results.

> Note Depending on your environment, you have to follow this procedure either in the Services Configuration section of the Collibra settings or in Collibra Console. This topic shows the procedure in Collibra Console. If your Settings page has a **Configuration** tab, you can follow this procedure in that tab page. For more information, see the online documentation.

Prerequisites

- You have the ADMIN or SUPER role in Collibra Console.
- You have the SUPER role in Collibra Console.

Steps

1. Open the DGC service settings for editing:
   a. Open Collibra Console.
      » Collibra Console opens with the **Infrastructure** page.
   b. In the tab pane, expand an environment to show its services.
   c. In the tab pane, click the Data Governance Center service of that environment.
   d. Click **Configuration**.
   e. Click **Edit configuration**.

2. In the **Tableau metadata API** section, enter the required information:

Setting	Description
Enable Tableau metadata API	• ✓ True: Tableau metadata API is enabled. This enables you to ingest Tableau 2020.2 or newer into Data Catalog. • ✕ False: Tableau metadata API is disabled. If you ingest Tableau 2020.2 or newer, the ingestion will fail. This prevents data loss of manually added relations and attributes.

3. Click the green **Save all** button.

Anonymize data via Jobserver

You can enable or disable the option to anonymize the content of columns with data type TEXT and GEO after the profiling process via Jobserver.

Tip If you profile and classify via Edge, data in columns with data type Text or Geo is automatically anonymized before it is sent to Collibra Data Intelligence Cloud.

Warning Currently, if you enable the data anonymization process you can no longer use automatic data classification via the Data Classification platform. However, you can still classify and anonymize profiling results if you use Edge.

Note Depending on your environment, you have to follow this procedure either in the Services Configuration section of the Collibra settings or in Collibra Console. This topic shows the procedure in Collibra Console. If your Settings page has a **Configuration** tab, you can follow this procedure in that tab page. For more information, see the online documentation.

Prerequisites

- You have the ADMIN or SUPER role in Collibra Console.
- You have the SUPER role in Collibra Console.

Steps

1. Open the DGC service settings for editing:
    a. Open Collibra Console.
        » Collibra Console opens with the **Infrastructure** page.
    b. In the tab pane, expand an environment to show its services.
    c. In the tab pane, click the Data Governance Center service of that environment.
    d. Click **Configuration**.
    e. Click **Edit configuration**.

2. In the **Data Profiling** section, enter the required information:

Setting	Description
Maximum number of samples	The maximum number of rows taken as a sample during profiling.
Maximum value length	The maximum length of a value extracted during profiling or sampling. Additional characters are trimmed.
Default date pattern	The default format used to decode dates. It is the default pattern used for detecting dates when the **Date Pattern** and/or **Time Pattern** attribute is not specified in Column assets.
Default time pattern	The default format used to decode times. It is the default pattern used for detecting times when the **Date Pattern** and/or **Time Pattern** attribute is not specified in Column assets.
Default combined date and time pattern	The default format used to decode combined dates and times. It is the default pattern used for detecting combined dates and times when the **Date Pattern** and/or **Time Pattern** attribute is not specified in Column assets.
Empty values	A comma separated list of strings enclosed in double quotes. A value that matches one of those expressions is considered an empty value. Please note that a database **null** value is always considered an empty value, for example *""*, *"na"* and *"none"*.
Data type detection threshold	The percentage of matching Column values to reach for an Advanced Data Type to be considered a possible Data Type for that Column. This is expressed as a value between 0.0 and 1.0).

Setting	Description
Anonymize data	An option to anonymize sensitive data. ◦ ✔ True: Content in columns with data type Text or Geo is removed or replaced by a random hash value before the profiling results are sent to the cloud. ◦ ✘ False (default): No content is removed or replaced by a random hash value. Tip If you profile and classify via Edge, the data in columns with data type Text or Geo is automatically anonymized before it is sent to Collibra Data Intelligence Cloud.
Database profiling via Edge	An option to enable profiling and classifying synchronized metadata via Edge instead of Jobserver. ◦ ✔ True: Profiling and classify via Edge. ◦ ✘ False: Profile via Jobserver and classify via the Data Classification Platform. Note You can only enable Database profiling via Edge if you also enabled Database registration via Edge.

3. Click the green **Save all** button.

Enable or disable the Settings landing page

You can enable or disable the Settings landing page.

Note Depending on your environment, you have to follow this procedure either in the Services Configuration section of the Collibra settings or in Collibra Console. This topic shows the procedure in Collibra Console. If your Settings page has a **Configuration** tab, you can follow this procedure in that tab page. For more information, see the online documentation.

Prerequisites

- You have the ADMIN or SUPER role in Collibra Console.
- You have the SUPER role in Collibra Console.

Steps

1. Open the DGC service settings for editing:
   a. Open Collibra Console.
      » Collibra Console opens with the **Infrastructure** page.
   b. In the tab pane, expand an environment to show its services.
   c. In the tab pane, click the Data Governance Center service of that environment.
   d. Click **Configuration**.
   e. Click **Edit configuration**.
2. In the **Beta features** section, enter the required information:

Setting	Description
Settings landing enabled	◦ ✔ True (default): Use the new Settings landing page in the Collibra user interface. ◦ ✖ False : Do not use the Settings landing page in the Collibra user interface.

3. Click the green **Save all** button.

Configure the Search service

In the Search service configuration, you can edit the TCP and Transport port of the service.

1. Open Collibra Console with a user profile that has the **SUPER** role.
   » Collibra Console opens with the **Infrastructure** page.
2. In the tab pane, click the **Search** service of a Collibra DGC environment.
   » The details of the **Search** service are shown.
3. Click **Configuration**.
4. Click **Edit configuration**.
5. Edit the search service settings.

Setting	Description
HTTP Port (Requires restart)	The TCP port to access the Search service via REST API. The default port is *4421*.
Transport Port (Requires restart)	The TCP port for the communication between the DGC and the Search service. The default port is *4422*.

6. Click the green **Save all** button.

Search service configuration options

To edit the Search service configuration options, you need the SUPER role.

Setting	Description
HTTP Port (Requires restart)	The TCP port to access the Search service via REST API. The default port is *4421*.
Transport Port (Requires restart)	The TCP port for the communication between the DGC and the Search service. The default port is *4422*.

Jobserver service configuration

If you want to ingest data and profile that data, you need a Jobserver service, often referred to as Jobserver. If both services are in the same trusted network segment, you can use the unsecure HTTP protocol with no authentication. However, it is highly recommended to configure a secure communication.

The configuration of the Jobserver service is to enable the communication with the Data Governance Center service.

In this section, you will learn more about configuring the Jobserver security and adding the Jobserver service to the DGC service.

You can add more than one Jobserver service to the DGC service.

In this chapter

Jobserver authentication levels 123
Mutual authentication between Jobserver and DGC service 124
General specifications for certificates and private keys 126
Edit the Jobserver service settings 127
Connection from an on-premises Jobserver to a Collibra Data Intelligence Cloud 130
Connection from Collibra Data Intelligence Cloud to an on-premises Jobserver 147
Jobserver best practices 152

Jobserver authentication levels

The Jobserver service handles customer data to, for example, perform profiling on it or ingest data into Collibra Data Intelligence Cloud. As this is sensitive data, we have to make sure that the communication with the Jobserver is as secure as possible. This section provides more information about the authentication levels available at the Jobserver configuration.

The Jobserver service supports three authentication levels.

Authentication level	Description
MUTUAL	Mutual authentication or two-way authentication is the authentication system where the two parties authenticate each other during their handshake. We only support certificate based mutual authentication, not username/password-based.
SERVER	Server authentication is the authentication system where the client authenticates the server identity but not the other way around. Warning For security reasons, do not use this level when you are using an on-premises Jobserver and a Collibra Data Intelligence Cloud environment. We only support certificate based server authentication, not username/password-based.
NONE	No authentication, no encryption. This is plain HTTP communication. Warning This is an insecure connection, your data can be exposed.

No authentication level

If you don't set an authentication level, you communicate with the Jobserver via an unsecure connection, using the HTTP protocol. This authentication level should only be used when the Jobserver service runs in the same trusted network segment as the Data Governance Center service. But even when both services run in the same network segment, we recommend to avoid this authentication level as malicious software or network security breaches could still expose your data.

Server authentication

If you select server authentication, the client party (DGC service) authenticates the server party (Jobserver). If the authentication succeeds, the DGC service is certain that it communicates with the selected Jobserver service.

Mutual authentication

If you select mutual authentication, both parties (DGC and Jobserver services) authenticate each other at the same time. If the authentication succeeds, both parties are certain about each other's identity. This is the most secure authentication level and is therefore the recommended authentication level for the Jobserver service.

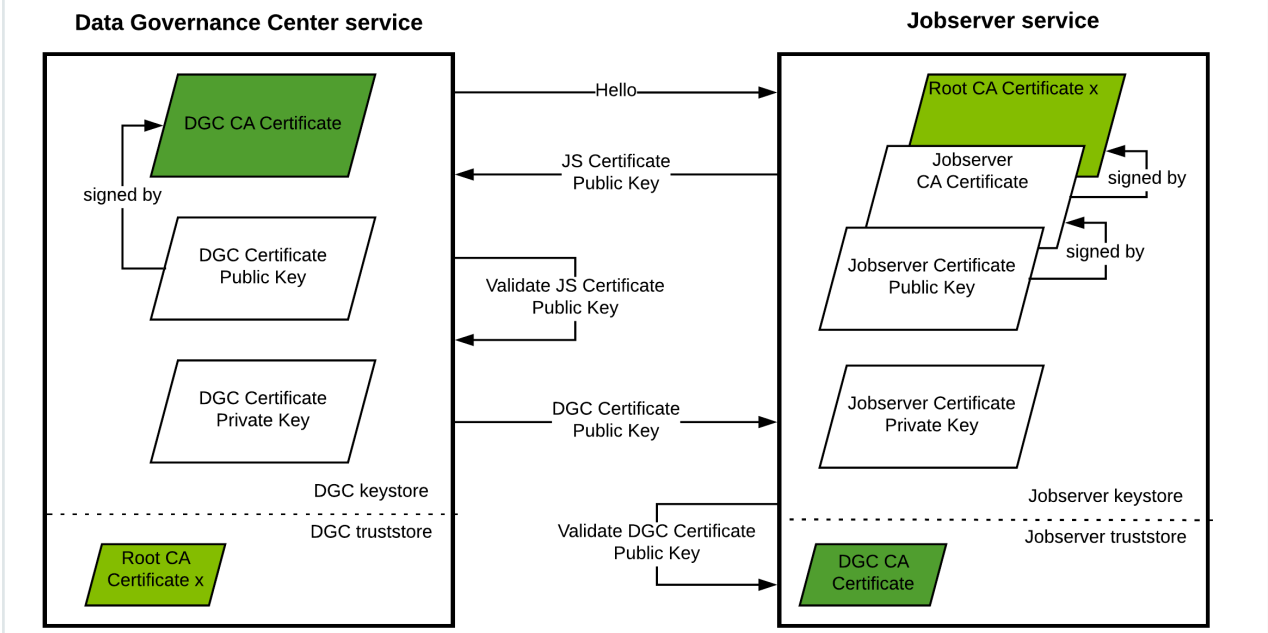
For more detailed information about the mutual authentication between the Jobserver and DGC service, see Mutual authentication between Jobserver and DGC service.

Mutual authentication between Jobserver and DGC service

If you want to ingest data and profile that data, you need a Jobserver service, often referred to as Jobserver. If both services are in the same network segment, you can use the insecure HTTP protocol with no authentication. However, it is highly recommended to configure a secure communication.

In the following schema, you can see how the communication is established for mutual authentication between the two services. This authentication level is the recommended level whenever both services are in a different network segment, for example when you have a cloud-based Collibra Data Intelligence Cloud environment and an on-premises Jobserver.

If there is a reverse proxy server between these services, see Mutual authentication with a reverse proxy server.



Step	Description
Hello	DGC service initiates the communication to the Jobserver service by sending a Hello message.
JS Certificate Public Key	The Jobserver service sends its public key to the DGC service. The DGC service can then authenticate the Jobserver service.

Step	Description
Validate JS Certificate Public Key	The DGC service validates the received public key with the Jobserver's CA certificate. This means that the DGC service has the Jobserver's CA certificate in its truststore.
DGC Certificate Public Key	The DGC service sends its public key to the Jobserver service. The Jobserver service can then authenticate the DGC service.
Validate DGC Certificate Public Key	The Jobserver service validates the received public key with the DGC CA certificate. This means that the Jobserver service has the DGC CA certificate in its truststore.

If all these steps are completed successfully, then both services are 100% sure that they are communicating with the counterparty that they expect.

General specifications for certificates and private keys

The certificates that you can use for the authentication between the Jobserver service and Data Governance Center service must meet the specifications as described in the following table.

Element	Specifications
Certificate	• The file name of the certificate can be freely chosen. • The certificate format must be in PEM format.

Element	Specifications
Private key	• The file name of the private key can be freely chosen. • The key format must be in PEM format. • The key must be packaged in PKCS#8 format. • The key is not protected by a passphrase.

If you already have a private key, but it does not meet the requirements, you can Convert a private key format the private key format.

Convert a private key format

If your company uses private keys, you can use them for server or mutual authentication between the services in a Collibra Data Intelligence Cloud environment, provided that they meet the specifications.

If they do not meet the specifications, you can convert the keys with the following command:

```
openssl pkcs8 -topk8 \
  -inform PEM \
  -outform PEM \
  -in <path_to_original_key> \
  -nocrypt \
  -out <path_to_new_key>
```

Edit the Jobserver service settings

The Jobserver service infrastructure settings allow you to edit the behavior of the connections to this service and the Java Virtual Machine (JVM) parameters.

1. Open Collibra Console with a user profile that has the **SUPER** role.
   » Collibra Console opens with the **Infrastructure** page.
2. In the tab pane, click **Jobserver** service of a Collibra DGC environment.
   » The details of the **Jobserver** service are shown.
3. Click **Infrastructure Configuration**.

4. Click **Edit configuration**.
5. Edit the Jobserver service infrastructure configuration.
6. Click the green **Save all** button.

Tip
- You can navigate to a specific section by clicking it in the tab pane.
- When you edit certain fields, the ⓘ icon is displayed next to it. When you click it, it displays the default value for that field and a **Reset** button.
- If you have to restart Collibra or execute extra actions to apply the new settings, it is indicated in the user interface.

Jobserver infrastructure configuration options

To edit the Jobserver infrastructure configuration options, you need the SUPER role.

Application server configuration

Setting	Description
Jobserver Monitoring Port (Requires restart)	The port that is used by the Monitoring service to monitor the Jobserver service. The default port is **4424**.
Spark Monitoring Port (Requires restart)	The port that is used by the Monitoring service to monitor the Spark service. The default port is **4434**.
Jobserver memory (Requires restart)	The memory that is assigned to the Jobserver service. The Jobserver controls the jobs that are executed for Data Catalog.

Setting	Description
Spark memory (Requires restart)	The memory that is assigned to the Spark application. This application is responsible for the actual data ingestion and profiling.

Security configuration

Setting	Description
Authentication level	The authentication level to communicate with the Jobserver. The client (DGC service) must be configured according the here selected authentication level.
Server certificate chain	The certificate or certificate chain with the public key that is offered by the Jobserver to the DGC service.
Server private key	The private key that is part of the Jobserver's certificate or certificate chain.
Trusted client CA certificate	The certificate of the trusted CA used to validate the client certificate (DGC service). To restrict authentication to this client, the CA should be exclusively used by this server.

JVM configuration

The Java Virtual Machine configuration parameters to run the Jobserver service.

> Warning
> Only apply changes to these settings if you are really experienced with JVM parameters. Changing parameters may cause serious performance issues.
>
> Restart the service after editing the JVM parameters.

Context JVM configuration

The context configuration parameters to run ingestion and profiling jobs.

> Warning
> Only apply changes to these settings if you are really experienced with JVM parameters. Changing parameters may cause serious performance issues.
>
> Restart the service after editing the JVM parameters.

Connection from an on-premises Jobserver to a Collibra Data Intelligence Cloud

In default installations, a Jobserver is installed on-premises and a Collibra Data Intelligence Cloud sends ingestion and profiling jobs to it. However, we highly recommend to reverse this communication, so that the on-premises Jobserver polls for jobs. This is often required for security reasons.

> Note If you want a Collibra Data Intelligence Cloud to send jobs to your on-premises Jobserver, contact your security officer and network administrator.

In this section, you get more information on how to set up the communication from an on-premises Jobserver to a Collibra Data Intelligence Cloud.

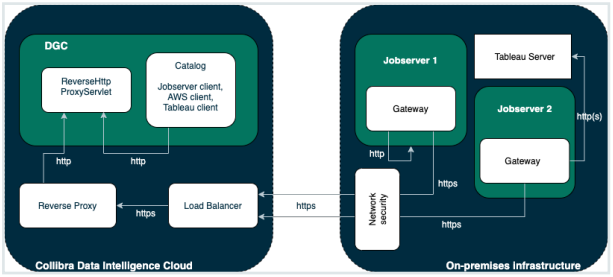
On-premises Jobserver to Collibra Data Intelligence Cloud communication

By default, the Data Governance Center service sends ingestion and profiling jobs to the Jobserver. This means that if you are using a Collibra Data Intelligence Cloud environment with an on-premises Jobserver, there is an inbound connection to the customer network, which is often not possible for security reasons. To allow such connections, you can use a reverse proxy server.

But, instead of using a reverse proxy, you also have to possibility to reverse this communication, where the on-premises Jobserver initiates the communication to Collibra Data Intelligence Cloud.

Communication overview

The following schema shows the communication paths from an on-premises Jobserver to Collibra Data Intelligence Cloud or to an on-premises Tableau server:



In case of an on-premises Tableau server, it is possible that inbound connections to it are not allowed. To establish a communication between your Collibra Data Intelligence Cloud and Tableau server, you will need a Jobserver that is dedicated to ingest from the Tableau server. The configuration of the communication from the Jobserver to the Tableau server is similar to the one from a Jobserver to a Collibra Data Intelligence Cloud environment.

Each Jobserver has to be a dedicated Jobserver, you cannot use a Jobserver to ingest from both Tableau server and S3 or JDBC data sources.

Components

To enable communication from an on-premises Jobserver to Collibra Data Intelligence Cloud, there are two new components:

New component	Description
Reverse HTTP proxy servlet	The reverse HTTP proxy is part of the DGC service. It acts as a server for all other Collibra services, whether they are installed on-premises or together with the DGC service in the cloud.

New component	Description
Gateway	The gateway is part of the Jobserver service. It polls the DGC service's reverse proxy to fetch tasks and send them to the Jobserver.

You only need the gateway to communicate with an on-premises Tableau server.

Encrypt passwords for basic authentication

In the Jobserver service configuration, you have to enter an encrypted password. To encrypt the password, use the **reversehttp-gateway-standalone** utility.

Prerequisites

- You have downloaded the reversehttp-gateway-standalone-6.3.7.jar file.
- You have the password of the Collibra user that you use to connect to your Collibra Data Intelligence Cloud environment.

Steps

> Note For security reasons, we have truncated the encrypted password in the example.

1. Open a terminal or command prompt session.
2. Go to the folder that contains the downloaded JAR file.
3. Execute the following command:

```
java -jar reversehttp-gateway-standalone-6.3.7.jar encrypt

Collibra Reverse HTTP Gateway
Enter value to encrypt: <password of Collibra user>
Re-enter value to encrypt: <password of Collibra user>
Encrypted value: encrypted:k7ScuJ3...
```

> Note If the entered values in this command don't match, it will ask the values again.

What's next?

If you use an encrypted password in a configuration, use the full string of the **Encrypted value** result. This includes the prefix "encrypted:".

Check connectivity between an on-premises Jobserver and Collibra Data Intelligence Cloud

Before you install your on-premises Jobserver, we highly recommend to do some basic connectivity tests between the node on which you are going to install the Jobserver service and your Collibra Data Intelligence Cloud environment. If the node cannot reach your Collibra Data Intelligence Cloud environment, then first fix the connectivity before you start the Jobserver installation.

1. Open the connection and port from your on-premises Jobserver to your Collibra Data Intelligence Cloud environment.
2. Whitelist the DNS name of your Collibra Data Intelligence Cloud environment.
3. Check if the Jobserver node can reach your Collibra Data Intelligence Cloud environment with the following command:

```
curl -x "" -i https://<your-environment>.collibra.com/reversehttp-poll/1
```

You should receive an HTTP 401 response. If you don't receive an HTTP 401 response, fix the connectivity before proceeding with the configuration.

Tip If there is a proxy server between your on-premises Jobserver and your Collibra Data Intelligence Cloud environment, then add the proxy address to the command:

```
curl -x http(s)://proxy:port -i https://<your-
environment>.collibra.com/reversehttp-poll/1
```

What's next?

When you receive an HTTP 401 response, you can start the installation of the Jobserver.

Configure the Jobserver to Collibra Data Intelligence Cloud communication

By default, Collibra Data Intelligence Cloud sends jobs to a Jobserver but you also have the possibility to have the Jobserver poll Collibra Data Intelligence Cloud for jobs.

If you want to have the Jobserver poll Collibra Data Intelligence Cloud for jobs, you have to configure both the Data Governance Center service and the Jobserver service.

Prerequisites

- You have Collibra Data Intelligence Cloud 2020.10 or newer.
- You have created a keystore in the PKCS#12 format on the node that hosts the Jobserver service.

Steps

Configure the Jobserver service

> Warning
> Only apply changes to these settings if you are really experienced with JVM parameters. Changing parameters may cause serious performance issues.
>
> Restart the service after editing the JVM parameters.

Execute the following steps in the Collibra Console instance that manages your Jobserver:

1. Open Collibra Console with a user profile that has the **SUPER** role.
   » Collibra Console opens with the **Infrastructure** page.

   > Tip The default address to access Collibra Console is *<server hostname>:4402*, but you may have set another port during the installation of Collibra Console. Keep in mind that a firewall of your operating system can block the access to Collibra Console.

2. In the tab pane, click the Jobserver service of a Collibra environment.
3. Click **Infrastructure Configuration**.
4. Click **JVM configuration**.
5. Click **Edit configuration**.
6. Add the following JVM settings:

Setting	Description
reversehttp.gateway	The setting to enable the Jobserver's gateway. To enable the communication from the on-premises Jobserver to the Collibra Data Intelligence Cloud environment., this value must be *true*. Example `-Dreversehttp.gateway=true`

Setting	Description
proxy.url	The URL of the Collibra environment, followed by *reversehttp-poll/<gateway-id>*. This "gateway-id" must be identical to the one used in the Name parameter when you add the Jobserver to the DGC service. The value of this setting is case-sensitive. Example `-Dproxy.url=https://<your-environment-url>/reversehttp-poll/Jobserver-1`
target.url	The URL of your the target system, either an on-premises Jobserver or a Tableau server. Example ◦ Jobserver: `-Dtarget.url=http://localhost:4404` ◦ Tableau server: `-Dtarget.url=https://tableau-sales.yourcompany.com`
http.proxy.host (optional)	The hostname of the HTTP proxy server for outbound connections to your Collibra Data Intelligence Cloud environment. This option is used to enable outbound traffic monitoring. Example `-Dhttp.proxy.host=proxy.yourcompany.com`

Setting	Description
http.proxy.port (optional)	The port of the HTTP proxy server for outbound connections to your Collibra Data Intelligence Cloud environment. This option is used to enable outbound traffic monitoring. Example `-Dhttp.proxy.port=8080`
username	The username of any Collibra user for basic authentication. Example `-Dusername=john.fisher`
password	The corresponding password of the Collibra user for basic authentication. Example `-Dpassword=ChangeMe` You can encrypt this password if necessary. Example `-Dpassword=enc_2:t2rklBY6699aWV0...`

Setting	Description
keystore.path	The full path to the PKCS12 keystore. This keystore should contain the private key to sign the basic authentication header. Example `-Dkeystore.path=/opt/collibra_data/spark-jobserver/security/jobserver-1-keystore.p12`

keystore.alias	The alias of the private key in the keystore. Each alias must be unique in your configuration.
	
	If you used the `name` argument during the creation of the keystore, then use the value of this `name` argument.
	
	If only 1 keystore is created, the default alias is "1".
	
	Example
	`-Dkeystore.alias=1`
	`-Dkeystore.alias=MyJobserver`

keystore.password (optional)	The password to access the keystore. If the keystore is not password-protected, don't add it to the JVM settings. You can encrypt this password if necessary. Example `-Dkeystore.password=ChangeMe`

Setting	Description
keystore.key.password (optional)	The password to use the private key, only applicable if you secured the private key with a password. If the key is not password-protected, don't add it to the JVM settings. You can encrypt this password if necessary. Example `-Dkeystore.key.password=ChangeMe`
polling.backoff	The time in milliseconds between a polling failure and a next polling attempt. We recommend to not define this parameter, it then uses the default value of 5,000 milliseconds. Example `-Dpolling.backoff=10000`
max.connections.route	The maximum number of HTTP connections per route. We recommend to not define this parameter, it then uses the default value of 20. Example `-Dmax.connections.route=30`
max.connections.total	The maximum number of all HTTP connections. We recommend to not define this parameter, it then uses the default value of 40. Example `-Dmax.connections.total=60`

Setting	Description
idle.connection.timeout	The time in milliseconds that an idle connection is kept in the connection pool. We recommend to not define this parameter, it then uses the default value of 5,000 milliseconds. Example `-Didle.connection.timeout=3000`
connection.timeout	The time in milliseconds that the reverse HTTP server waits for a response from your Collibra environment or from the value in target.url. If you don't set this parameter, the value is 60,000 milliseconds. Example `-Dconnection.timeout=30000`
connection.soTimeout	The time in milliseconds that the reverse HTTP server waits for a response from your Collibra environment or from the value in target.url on socket level. If you don't set this parameter, the value is 60,000 milliseconds. Example `-Dconnection.soTimeout=30000`

Setting	Description
polling.timeout	The time in milliseconds that the reverse HTTP server waits for a poll request from your Collibra environment to be submitted to the target.url. If you don't set this parameter, the value is 300,000 milliseconds. Example `-Dpolling.timeout=100000`
polling.period	The time in milliseconds that the reverse HTTP server waits in between poll request sessions. In other words, after having received a poll request or no request from your Collibra environment, the reverse HTTP server waits a certain amount of milliseconds before contacting the Collibra environment again. If you don't set this parameter, the value is 100 milliseconds. Example `-Dpolling.period=200`
health.check.period (optional)	The time in milliseconds that the reverse HTTP server waits between health checks of its connection with Collibra. If you don't set this parameter, the value is 5,000 milliseconds. Example `-Dhealth.check.period=10000`

Setting	Description
health.check.timeout (optional)	The time in milliseconds that the reverse HTTP server waits for a health check response from Collibra. If you don't set this parameter, the value is 5,000 milliseconds. Example `-Dhealth.check.timeout=10000`

> Note You have to use separate Jobservers for the ingestion of S3 or JDBC data sources and Tableau server data.

7. Click the green **Save all** button.
8. Click **Security configuration**.
9. Click **Edit configuration**.
10. Set the **Authentication level** to *NONE*.

> Note This means that there is a one-way outbound communication over TLS from the Jobserver to the Collibra environment, not that there is no authentication at all.

11. Click the green **Save all** button.

Add the Jobserver to the DGC service

Execute the following steps in Collibra Console of your Collibra Data Intelligence Cloud environment.

1. Open the DGC service settings for editing:
2. Go to the **Jobserver** section of the configuration.

3. Enter the required information.

Setting	Description
Name	The name of the Jobserver as it will appear when you register a data source. The name is a freely chosen name but it is recommended to only use alphanumerical characters and dashes, for example Jobserver-1. You will have to use this name as the ID of the gateway and in the address of this configuration.
Protocol	The protocol for this configuration has to be *HTTP* and not HTTPS.
Address	The loopback address of the DGC service, followed by */reversehttp/<gateway-id>*. The "gateway-id" must be identical to the one used in the Name parameter of this configuration. Do not use the scheme in the address. Example localhost:4400/reversehttp/Jobserver-1
Trusted server CA certificate	The certificate in PEM format that contains the public key of the Jobserver to validate the signature of the basic authentication header. In the example to create a keystore, this is the content of the file **cert.pem**. Example -----BEGIN CERTIFICATE----- MIICqDCCAZACCQCcy3Oq51c5YzANBgkqhkiG9w0BAQsF ADAWMRQwEgYDVQQDDAtq b2JzZXJ2ZXIt... -----END CERTIFICATE-----

Setting	Description
Client certificate	This field is not used in this configuration.
Client private key	This field is not used in this configuration. > Note This field always shows dots, even if it is empty.
Table profiling data size	The approximate maximum disk size of the data in MB that will be used to profile a table. The value cannot exceed 10 000.

4. Click the green **Save all** button.

If all settings and communication paths are correctly configured, you will see a notice on the Jobserver:

```
INFO [I/O dispatcher 1] reversehttp.gateway.PollingController -
proxy -> no requests polled (204)
```

What's next?

When you have set up this communication, you may want to monitor the outbound traffic. You can do so by enabling a man-in-the-middle proxy.

Monitor outbound traffic

If you set up the communication from your on-premises Jobserver to your Collibra Data Intelligence Cloud environment, you may want to monitor the outbound traffic. You can do so by setting up a man-in-the-middle proxy (MITM proxy).

1. Open Collibra Console with a user profile that has the **SUPER** role.
   » Collibra Console opens with the **Infrastructure** page.
2. In the tab pane, click the Jobserver service of a Collibra environment.

3. Click **Infrastructure Configuration**.
4. Click **JVM configuration**.
5. Click **Edit configuration**.
6. Add the following JVM settings:

Setting	Description
http.proxy.host	The hostname of the HTTP proxy server for outbound connections to your Collibra Data Intelligence Cloud environment. Example -Dhttp.proxy.host=proxy.yourcompany.com
http.proxy.port	The port of the HTTP proxy server for outbound connections to your Collibra Data Intelligence Cloud environment. Example -Dhttp.proxy.port=8080

7. Add the CA certificate of this MITM proxy in the Jobserver's truststore (`$(COLLIBRA_DIR)/jre/lib/security/cacerts`).

Generate keys, certificates and keystores

For a secure communication between the Jobserver and Collibra Data Intelligence Cloud, you can use certificates. In the current configuration, certificates are used as containers for public keys and the keystore is used to store private keys and certificates.

- On the node that hosts the Jobserver service, the keystore must be in PKCS#12 format.
- On the node that hosts the Data Governance Center service, you need a certificate, in PEM format, which includes the public key.

Steps

> Note The commands used in this procedure are only examples, ask your Security officer for more information.

1. On the node on which you want to install the keystore, certificate and private key, open a terminal or command prompt session.
2. Go to or create a directory in which you want to create the keystore.
3. Create the private key and certificate:

```
openssl req -x509 -newkey rsa -keyout key.pem -out cert.pem
-days 365

Generating a 2048 bit RSA private key
.....................................+++
.................+++
writing new private key to 'key.pem'
Enter PEM pass phrase: <optional password>
Verifying - Enter PEM pass phrase: <repeat password>
-----
You are about to be asked to enter information that will be
incorporated
into your certificate request.
What you are about to enter is what is called a Distin-
guished Name or a DN.
There are quite a few fields but you can leave some blank
For some fields there will be a default value,
If you enter '.', the field will be left blank.
-----
Country Name (2 letter code) [AU]:US
State or Province Name (full name) [Some-State]:New York
Locality Name (eg, city) []:New York
Organization Name (eg, company) [Internet Widgits Pty
Ltd]:Collibra
Organizational Unit Name (eg, section) []:
Common Name (e.g. server FQDN or YOUR name) []:Francois
Lemaire
Email Address []:francois.lemaire@collibra.com
```

4. Create a PKCS#12 keystore including a private key and certificate.

```
openssl pkcs12 -export -inkey key.pem -in cert.pem -out key-
store.p12 -name <meaningful name>
```

```
Enter pass phrase for key.pem:<if password added in pre-
vious step>
Enter Export Password:
Verifying - Enter Export Password:
```

> Important We recommend that you provide the `name` argument with a meaningful name. You then have to use this name as the keystore alias in the JVM configuration of the Jobserver service. If you don't use the `name` argument and there's only one keystore, then the keystore alias is 1.

5. Copy the **p12** file to **%collibra_data%/spark-jobserver/security/**.

Connection from Collibra Data Intelligence Cloud to an on-premises Jobserver

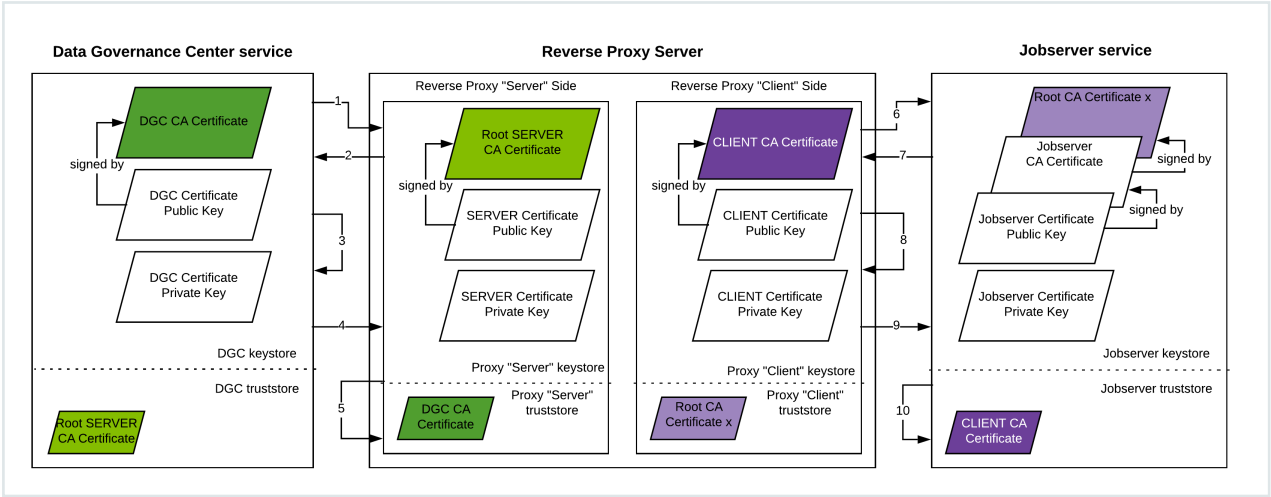
In a typical on-premises installation, you install all the services in the same network. To enable the communication between environment services that are installed in different networks, you can use a reverse proxy server between those networks, for example, when you use Collibra as a cloud service and an on-premises Jobserver.

> Note If you are using a Collibra Data Intelligence Cloud environment with an on-premises Jobserver, they must have the same installer version. You can find the installer version of your Collibra Data Intelligence Cloud environment at the bottom of the sign-in window of its Collibra Console, for example 5.7.12-32

Mutual authentication with a reverse proxy server

Mutual authentication is a means to ensure secure communications between the Data Governance Center service and the Jobserver service. If there is a reverse proxy server between these two services, you can still use mutual authentication.

In the following schema, you can see how the communication is established for mutual authentication between the two services with a reverse proxy server in between.



Step	Description
1	The DGC service initiates the communication to the proxy server by sending a Hello message.
2	The proxy server sends its public key to the DGC service. The DGC service can then authenticate the proxy server.
3	The DGC service validates the received public key with the proxy server's CA certificate. This means that the DGC service has the proxy server's CA certificate in its truststore.
4	The DGC service sends its public key to the proxy server. The proxy server can then authenticate the DGC service.
5	The proxy server validates the received public key with the DGC CA certificate. This means that the proxy server has the DGC CA certificate in its truststore.

Step	Description
6	The proxy server initiates the communication to the Jobserver service by sending a Hello message.
7	The Jobserver service sends its public key to the proxy server. The proxy server can then authenticate the Jobserver service.
8	The proxy server validates the received public key with the Jobserver service's CA certificate. This means that the proxy server has the Jobserver service's CA certificate in its truststore.
9	The proxy server sends its public key to the Jobserver service. The Jobserver service can then authenticate the client side of the reverse proxy server.
10	The Jobserver service validates the received public key with the proxy server's CA certificate. This means that the Jobserver service has the proxy server CA certificate in its truststore.

Set up mutual authentication with a proxy server

If you use mutual authentication with a proxy server between the Data Governance Center service and the Jobserver service, the configuration of both services is slightly different, especially with the certificates.

Prerequisites

- Your certificates meet the necessary specifications. If not, convert them to the right format.

Edit the Jobserver service settings

Execute the following steps in Collibra Console that manages the Jobserver service.

1. Open Collibra Console with a user profile that has the **SUPER** role.
   » Collibra Console opens with the **Infrastructure** page.
2. In the tab pane, click the **Jobserver** service of a Collibra environment.
   The details of the **Jobserver** service are shown.
3. Click **Infrastructure Configuration**.
4. Enter the security configuration:

Setting	Description
Authentication level	The authentication level to communicate with the Jobserver. The client (reverse proxy server) must be configured according the here selected authentication level.
Server certificate chain	The certificate or certificate chain with the public key that is offered by the Jobserver to the reverse proxy server.
Server private key	The private key that is part of the Jobserver's certificate or certificate chain.
Trusted client CA certificate	The certificate of the trusted CA used to validate the client certificate (reverse proxy server). To restrict authentication to this client, the CA should be exclusively used by this server.

5. Click the green **Save all** button.

Add a proxy server to the DGC service

Execute the following steps in Collibra Console that manages your DGC service.

1. Open the DGC service settings for editing:
2. In the **Jobserver** section, click **Add**.

3. Enter the necessary information:

Setting	Description
Jobserver list	The list of registered Jobserver instances.
Name	The name of the Jobserver as it will appear when you register a data source in Data Catalog.
Protocol	The protocol that is used for the communication between the Data Governance Center service and the reverse proxy server. It is recommended to use HTTPS, especially if the services are hosted in different network segments.
Address	The address (IP address, URL, hostname) of the reverse proxy server.
Trusted server CA certificate	The certificate of the trusted CA needed to validate the server certificate. If blank, the default truststore will be used. The default truststore is defined in the SSL configuration section of the DGC service. The CA certificate of the server party (reverse proxy server).
Client certificate	The client certificate offered by the DGC service to the server. If blank, you cannot select mutual authentication as the Jobserver service authentication level.
Client private key	The private key of the DGC service's certificate.
Table profiling data size	The approximate maximum disk size of the data in MB that will be used to profile a table. The value cannot exceed 10,000.

Setting	Description
Test connection timeout	This timeout is a time limit (in seconds) after which the connection test is stopped and a timeout error is shown. The default value is 60 seconds.

4. Click the green **Save all** button.

You can still add multiple Jobserver services but then you will need one reverse proxy server per Jobserver. In the unlikely event that there are multiple Jobservers behind one reverse proxy server, you have to configure the reverse proxy server in such a way that there is a unique port per Jobserver.

Configuring the proxy server

Consult the documentation of your reverse proxy server to configure the server side and client side.

Jobserver best practices

This article is specific for an on-premises Jobserver, and the corresponding Jobserver, that connect with Collibra Data Intelligence Cloud.

Installation

- Ensure that you use the latest available installer version that matches your Collibra environment. Collibra is deployed on a monthly basis while the on-premises installer versions are only available on a quarterly basis.

  > Example Collibra Data Intelligence Cloud is released on a monthly basis while the on-premises environments can only be upgraded on a quarterly basis. For example, Collibra 2020.11 has a corresponding on-premises version 5.7.7. This on-premises version will remain the latest available one for the next two monthly Collibra Data Intelligence Cloud releases.

- A user with the correct permissions to execute the installation:
  - Linux: Ideally, a root or sudo user should be used. The user must be able to execute the installation script on the mounted file system.
  - Windows: An administrator user must be used. This user must have full rights on the intended installation drive and directories.
- Ensure that all required ports are open/listening and not in use by other programs or processes on the new Jobserver.
  For the list of default TCP ports, see Overview default ports in Collibra.
- Ensure that you can connect to Collibra Console that is used to manage the on-premises Jobserver.
- Ensure that the server that will run the Jobserver meets the minimum system requirements.
- Ensure that the server that will run the Jobserver can communicate with Collibra Data Intelligence Cloud. For more information, see how you can check the communication.

Configuration

- Ensure that all JVM properties are entered correctly into Collibra Console.
  - Ensure that there are no spaces at the beginning or end of each JVM property added. These empty spaces will cause the Jobserver to enter a Failed state and prevent the creation of the **spark-jobserver.log** file.
  - Ensure the name specified in the **proxy.url** JVM parameter matches the name of the Jobserver defined in Collibra Console under section 18.a and 18.c.

    > Note This value is case-sensitive, so the name should match between both Collibra Console for Collibra Data Intelligence Cloud entries and Collibra Console for on-premises Jobserver entries.

- On Linux, the Jobserver may fail to start if the **keystore.p12** file is not owned by the "Collibra" user and group. You can verify this by looking at the last entry in the **spark-jobserver.log** once the Jobserver is in error state.
  If this is the issue, you can fix it as follows:

a. On the server, go to the directory where the p12 file is stored.
b. Run the following command:

```
chown collibra:collibra keystore.p12
```

c. Restart the Jobserver.

Jobserver usage

- Ensure that all drivers are configured properly in accordance to the related documentation. This includes applying all of the required properties, based on the Data Source and Driver type:
  - Manage Collibra-provided JDBC drivers
  - Manage your own JDBC drivers
- Ensure that your on-premises Jobserver can communicate with the servers that host your data sources. Please work with your Network/Server teams to resolve any network blockages or restrictions that would prevent the Jobserver from successfully connecting to the intended data sources.

Use the Jobserver behind a reverse proxy

See Connection from an on-premises Jobserver to a Collibra Data Intelligence Cloud.

If you have configured everything correctly, you will see a notice on the Jobserver:

```
INFO [I/O dispatcher 1] reversehttp.gateway.PollingController -
proxy -> no requests polled (204)
```

In the logs of the Data Governance Center service, you will see the following message stream:

```
[http-nio-0.0.0.0-4400-exec-12] INFO c.c.r.-
proxy.ReverseHttpProxyServlet - client -> eea3e5fc-3ed4-4095-
beb8-f365cd984973: GET /processes
[ajp-nio-0.0.0.0-8080-exec-7] INFO
```

```
c.c.r.proxy.ReverseHttpProxyServlet - gw <- gw polled eea3e5fc-
3ed4-4095-beb8-f365cd984973
[ajp-nio-0.0.0.0-8080-exec-2] INFO c.c.r.-
proxy.ReverseHttpProxyServlet - gw -> respond eea3e5fc-3ed4-
4095-beb8-f365cd984973
[http-nio-0.0.0.0-4400-exec-12] INFO c.c.r.-
proxy.ReverseHttpProxyServlet - client <- eea3e5fc-3ed4-4095-
beb8-f365cd984973: 200
```

Optional Jobserver service configurations

See Jobserver to Collibra communication.

- Dpolling.backoff
- Dmax.connections.route
- Dmax.connections.total
- Didle.connection.timeout
- Dhttp.proxy.host
- Dhttp.proxy.port

Chapter 4

Service infrastructure configurations

When you install the Collibra Data Intelligence Cloud environment, you can define some basic settings such as TCP ports and passwords. All other settings of the services will have default values.

Most of these default settings however, may not comply with your company's IT infrastructure. In this section, you will find more information about configuring the different service's infrastructure settings.

In this chapter

Edit the DGC service infrastructure settings 157

Edit the Search service infrastructure settings 165

Edit the DGC service infrastructure settings

The Data Governance Center service infrastructure settings allow you to edit the behavior of the connections to this service and the Java Virtual Machine (JVM) parameters.

1. Open Collibra Console with a user profile that has the **SUPER** role.
   » Collibra Console opens with the **Infrastructure** page.
2. In the tab pane, click **Data Governance Center** service of a Collibra DGC environment.
   » The details of the **Data Governance Center** service are shown.
3. Click **Infrastructure Configuration**.
4. Click **Edit configuration**.
5. Edit the DGC service infrastructure configuration.
6. Click the green **Save all** button.

Tip
- You can navigate to a specific section by clicking it in the tab pane.
- When you edit certain fields, the ℹ icon is displayed next to it. When you click it, it displays the default value for that field and a **Reset** button.
- If you have to restart Collibra or execute extra actions to apply the new settings, it is indicated in the user interface.

DGC infrastructure configuration options

To edit the DGC service infrastructure configuration options, you need the SUPER role.

Application server configuration

Setting	Description
Context path (Requires restart) *	The path that is added to the base URL to reach Collibra Data Intelligence Cloud. For example, if your base URL is `https://dgc.yourcompany.com:4400/` and your context path is `acceptance`, then your path to reach Collibra is `https://dgc.yourcompany.com:4400/acceptance`.

HTTP connector

A connector supporting the HTTP/1.1 protocol.

Setting	Description
URI encoding (Requires restart) *	The character encoding used to decode the URI bytes, URL-decoding has been performed.
Acceptor thread count (Requires restart) *	The number of threads to be used to accept connections. The actual request processing is done by separate threads, so you would never really need more than two.
Min spare threads (Requires restart) *	The minimum number of threads always kept running.

Setting	Description
Max threads (Requires restart) *	The maximum number of request processing threads to be created by this connector. This determines the maximum number of simultaneous requests that can be handled.
Accept count (Requires restart) *	The maximum queue length for incoming connection requests that can be assigned to a request processing thread. When the queue is full, requests will be refused.
Compression (Requires restart) *	The connector may use HTTP/1.1 GZIP compression. • **off**: No compression • **on**: Allow compression • **force**: Allow compress
Compression min size (Requires restart) *	The minimum amount of data before it is compressed. This setting only has effect if **Compression** is **on**.
Compressible MIME type (Requires restart)	A comma separated list of MIME types allowing HTTP compression.
Connection timeout (Requires restart) *	The number of milliseconds that this connector waits for the destination URI to be presented by the request, after accepting a connection.
Port (Requires restart) *	The TCP port to access your Collibra environment via your web browser.

HTTPS connector

A connector supporting the HTTP/1.1 protocol with SSL support enabled.

Setting	Description
URI encoding (Requires restart) *	The character encoding used to decode the URI bytes, URL-decoding has been performed.
Acceptor thread count (Requires restart) *	The number of threads to be used to accept connections. The actual request processing is done by separate threads, so you would never really need more than two.
Min spare threads (Requires restart) *	The minimum number of threads always kept running.
Max threads (Requires restart) *	The maximum number of request processing threads to be created by this connector. This determines the maximum number of simultaneous requests that can be handled.
Accept count (Requires restart) *	The maximum queue length for incoming connection requests that can be assigned to a request processing thread. When the queue is full, requests will be refused.
Compression (Requires restart) *	The connector may use HTTP/1.1 GZIP compression. • **off**: No compression • **on**: Allow compression • **force**: Allow compress

Setting	Description
Compression min size (Requires restart) *	The minimum amount of data before it is compressed. This setting only has effect if **Compression** is **on**.
Compressible MIME type (Requires restart)	A comma separated list of MIME types allowing HTTP compression.
Connection timeout (Requires restart) *	The number of milliseconds that this connector waits for the destination URI to be presented by the request, after accepting a connection.
Port (Requires restart) *	The TCP port to access your Collibra environment via your web browser.

AJP connector

A connector able to communicate with another web connector via the AJP protocol. Mainly for transparent integration with another HTTP server, for example Apache, nginx....

Setting	Description
URI encoding (Requires restart) *	The character encoding used to decode the URI bytes, URL-decoding has been performed.
Acceptor thread count (Requires restart) *	The number of threads to be used to accept connections. The actual request processing is done by separate threads, so you would never really need more than two.

Setting	Description
Min spare threads (Requires restart) *	The minimum number of threads always kept running.
Max threads (Requires restart) *	The maximum number of request processing threads to be created by this connector. This determines the maximum number of simultaneous requests that can be handled.
Accept count (Requires restart) *	The maximum queue length for incoming connection requests that can be assigned to a request processing thread. When the queue is full, requests will be refused.
Compression (Requires restart) *	The connector may use HTTP/1.1 GZIP compression. • **off**: No compression • **on**: Allow compression • **force**: Allow compress
Compression min size (Requires restart) *	The minimum amount of data before it is compressed. This setting only has effect if **Compression** is **on**.
Compressible MIME type (Requires restart)	A comma separated list of MIME types allowing HTTP compression.
Connection timeout (Requires restart) *	The number of milliseconds that this connector waits for the destination URI to be presented by the request, after accepting a connection.

Setting	Description
Port (Requires restart) *	The TCP port to access your Collibra environment via your web browser.

Static Resources

The static resources are the resources that are reserved for the service.

Setting	Description
Maximum cache size (Requires restart) *	The maximum size that can be assigned to the cache of the service.

JVM configuration

The Java Virtual Machine configuration parameters to run the Data Governance Center service.

> Warning
> Only apply changes to these settings if you are really experienced with JVM parameters. Changing parameters may cause serious performance issues.
>
> Restart the service after editing the JVM parameters.

Set the context path of the DGC service in Collibra Console

The context path allows you to serve the DGC service on a specific URL. You can set this in the installation wizard, but also after the installation in Collibra Console.

Steps

1. Open Collibra Console with a user profile that has the **SUPER** role.
   » Collibra Console opens with the **Infrastructure** page.
2. In the tab pane, expand an environment and then click the **Data Governance Center** service of that environment.
3. If the DGC service is running, click **Stop** to stop it.
4. In the details window, click **Infrastructure Configuration**.
5. Click **Application server configuration**.
6. Click **Edit configuration**.
7. In the **Context path** field, enter the name of your context path.

Setting	Description
Context path (Requires restart) *	The path that is added to the base URL to reach Collibra Data Intelligence Cloud. For example, if your base URL is `https://dgc.yourcompany.com:4400/` and your context path is `acceptance`, then your path to reach Collibra is `https://dgc.yourcompany.com:4400/acceptance`.

8. Click the green **Save all** button.

> Note If you didn't stop the DGC service, you will see an error message. Stop and start the DGC service to apply the change to the DGC service.

9. Click **Start** to start the DGC service.

What's next?

Connect to Collibra using the new URL, which is the base URL with the context path.

Edit the Search service infrastructure settings

The Search service infrastructure settings allow you to edit the Java Virtual Machine (JVM) parameters.

1. Open Collibra Console with a user profile that has the **SUPER** role.
   » Collibra Console opens with the **Infrastructure** page.
2. In the tab pane, click **Search** service of a Collibra DGC environment.
   » The details of the **Search** service are shown.
3. Click **Infrastructure Configuration**.
4. Click **Edit configuration**.
5. Edit the Search service infrastructure configuration.
6. Click the green **Save all** button.

Search infrastructure configuration options

To edit the Search infrastructure configuration options, you need the SUPER role.

JVM configuration

The Java Virtual Machine configuration parameters to run the Search service.

> Warning
> Only apply changes to these settings if you are really experienced with JVM parameters. Changing parameters may cause serious performance issues.
>
> Restart the service after editing the JVM parameters.

Managing environments

A Collibra Data Intelligence Cloud environment is a collection of services that allows you to govern your data. With Collibra Console, you can manage one or more environments. You can start and stop an environment, configure the environment behavior and so on.

Each environment must have Data Governance Center service, a Repository service and a Search service. You can install each service on a dedicated node or combine services on one node.

- Data Governance Center service: This is the service that contains the business logic of your environment.
- Repository service: This is the service that stores the data in the DGC database.
- Search service: This is the service that allows you to search for data in the Collibra user interface.

For more information about the services, take a look at the Product architecture section.

For more information about nodes, take a look at the Node management section.

In this chapter

Create a Collibra environment 168
Environment statuses 169
Start an environment 170
Stop an environment 171
Delete an environment 171
Start a service 172
Stop a service 172
Reindexing Collibra Data Intelligence Cloud 173

Restore to factory defaults 173

Create a Collibra environment

A Collibra Data Intelligence Cloud environment is a collection of services that are logically linked together.

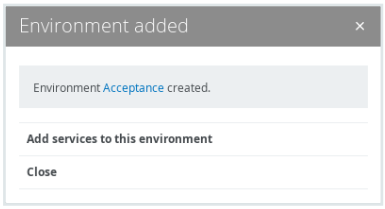
> Note With Collibra Console, you can manage many nodes, but these nodes must be on the same version as your Collibra Console.

Steps

1. Open Collibra Console with a user profile that has the **SUPER** role.
   » Collibra Console opens with the **Infrastructure** page.

   > Tip
   > - The default address to access Collibra Console is *<server hostname>:4402*, but you may have set another port during the installation of Collibra Console. Keep in mind that a firewall of your operating system can block the access to Collibra Console.
   > - The default credentials to sign in to Collibra Console are *Admin* / *admin*. We highly recommend that you edit the Collibra Console administrator's password after signing in for the first time.

2. In the tab pane, click **Add / Create**.
   » The **Add / Create** dialog box appears.
3. Click **Create environment**.
   » The **Create Environment** dialog box appears.
4. Enter a name.
5. Click **Create Environment**.



6. Do one of the following:
    - Click **Close** to end the wizard.
    - Click **Add services to this environment** to immediately add services.

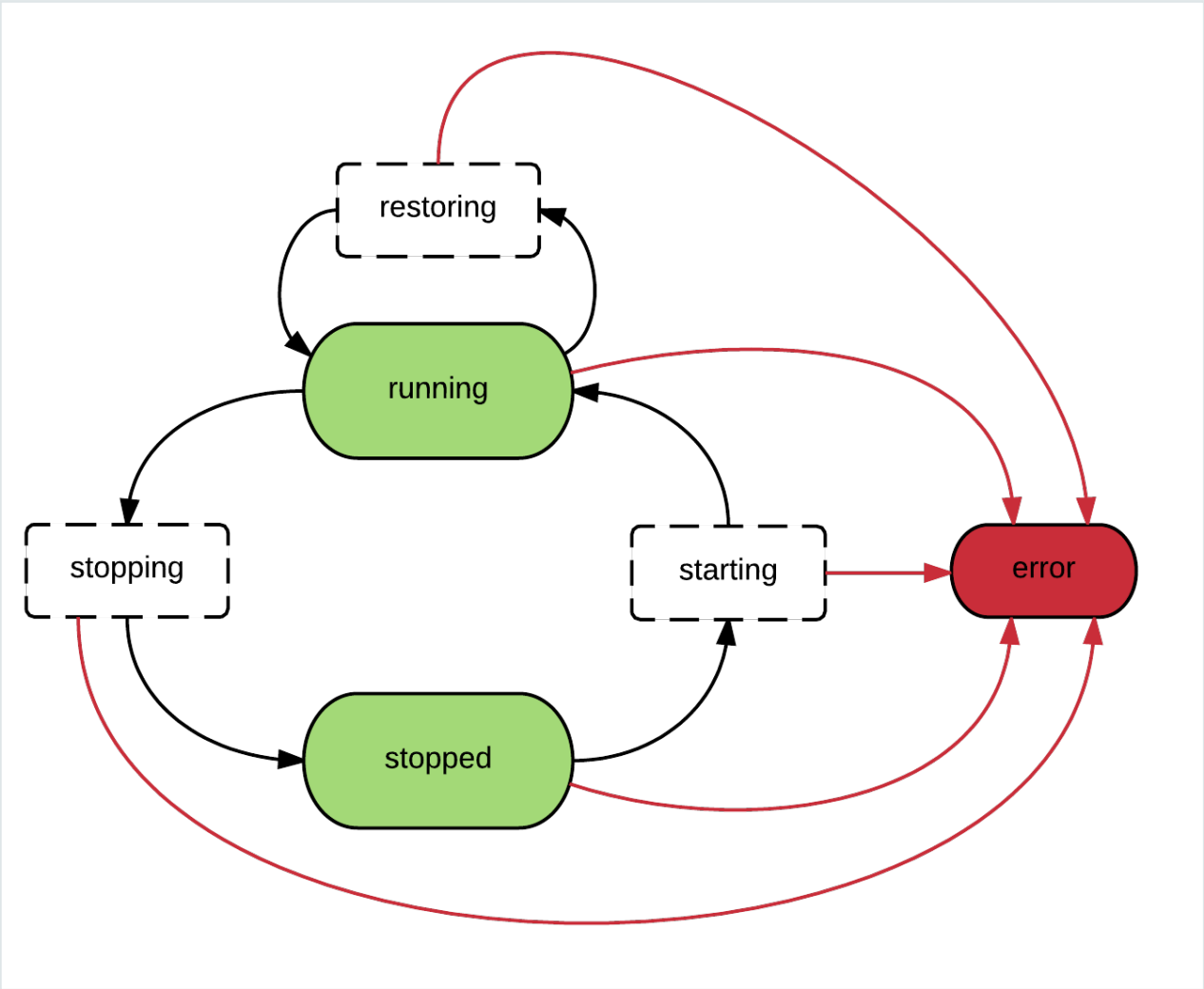
      > Note If the node that hosts the service you want to add is not yet available in Collibra Console, click **Add services from a new node** under the drop-down list and add the node details.

Environment statuses

An environment can have the following statuses:

- **starting** (transitional status)
- **running**
- **stopping** (transitional status)
- **stopped**
- **error**
- **restoring** (transitional status)

In the following diagram you can find the relation between the statuses:



Start an environment

Starting an environment will start all services of the environment. You can also start the services individually, for more information, see Start a service.

> Note When you use a network service account to start the agent and console services, the account must be available when the node starts. If the account is not available, the startup will fail.

Steps

To start an environment, follow these steps:

1. Open Collibra Console with a user profile that has at least the **ADMIN** role.
   » Collibra Console opens with the **Infrastructure** page.
2. Click the name of an environment to show its details.
3. Click ▶ **Start**.

What's next?

The environment first starts the Repository service and optionally the Jobserver service, then the Collibra service. When both services have the status **running**, the environment status becomes **running**.

Stop an environment

Stopping an environment will stop all services of the environment.

> Tip You can also stop individual services. More information: Stop a service.

Steps

1. Open Collibra Console with a user profile that has at least the **ADMIN** role.
   » Collibra Console opens with the **Infrastructure** page.
2. Click the environment that you want to stop.
3. Click ■ **Stop**.
   The **Stop environment** dialog box appears.
4. Click **Stop environment**.

What's next?

The environment first stops the Collibra service and then optionally the Jobserver service and finally the Repository service.

Delete an environment

To delete an environment, follow these steps:

1. Open Collibra Console with a user profile that has the **SUPER** role.
   » Collibra Console opens with the **Infrastructure** page.
2. In the list of environments, click 🗑 at the end of the line.
   The **Delete Environment** dialog box appears
3. Click **Delete**.

Start a service

By starting an environment, you start all services at once. Instead of starting the environment, you can also start the services manually.

Steps

1. Open Collibra Console with a user profile that has at least the **ADMIN** role.
   » Collibra Console opens with the **Infrastructure** page.
2. Click the environment whose service you want to start.
3. Click the service that you want to start.
4. Click ▶ **Start**.

Stop a service

Instead of stopping an environment, you can also stop a service manually. You can only stop services if no other services depend on it. For example, you cannot stop the Jobserver service if the Data Governance Center service is still running.

Steps

1. Open Collibra Console with a user profile that has at least the **ADMIN** role.
   » Collibra Console opens with the **Infrastructure** page.
2. In the infrastructure tree, click the service that you want to stop.
   » The service details apear.
3. Click ■ **Stop**.
   » The **Stop service** dialog box appears.
4. Click **Stop service**.

Reindexing Collibra Data Intelligence Cloud

When you edit the search or hyperlink settings, you have to reindex Collibra Data Intelligence Cloud to update the existing hyperlinks and search index.

Reindexing Collibra:

- Rebuilds automatic hyperlinks.
- Rebuilds the search index.

> Tip You can also rebuild the search index from the Collibra settings page.

Steps

1. Open Collibra Console with a user profile that has at least the **ADMIN** role.
   » Collibra Console opens with the **Infrastructure** page.
2. In the tab pane, click the **Data Governance Center** service of the environment whose search index and hyperlinks you want to reindex.
3. Click **Rebuild search index and automatic hyperlinks**.

Restore to factory defaults

If you no longer need the data from your environment, you can reset your environment to factory defaults. This action completely resets your repository and removes all the customizations of the Data Governance Center service, but it will keep the configuration of your environment.

> Warning This operation cannot be undone, it will delete permanently the content of the repository.

Steps

1. Open Collibra Console with a user profile that has at least the **ADMIN** role.
   » Collibra Console opens with the **Infrastructure** page.

2. Click the name of an environment to show its details.
3. Click **More** → **Restore to factory defaults**.
4. Click **Restore** to confirm that you want to restore the environment.
5. Optionally click **Continue in background** to continue working in Collibra Console.
   » The environment is reset. The DGC service will restart during this process.

Back up and restore

A backup corresponds to a picture of a Collibra Data Intelligence Cloud environment at a certain point in time. This picture consists of all application data, the operating model, assets, comments, and so on. You can restore a backup at a later point in time.

You can specify which aspects of the application data to include in the backup and again during the restore.

In this chapter

Create a backup of Collibra Data Intelligence Cloud 2022.06 177

Backup options 179

The Backups page 181

Download a backup 182

Upload a backup 183

Delete a backup 184

Create a backup schedule 185

Backup schedule options 188

Backup schedules overview 190

Edit a backup schedule 192

Delete a backup schedule 192

Restoring a backup 193

Customize Collibra DGC with a backup restore 202

Back up and download with the REST API 202

REST API - List of backups 206

REST API - Delete a backup 207

Create a backup of Collibra Data Intelligence Cloud 2022.06

If you want to take a full or partial backup of a Collibra Data Intelligence Cloud environment, follow these steps:

1. Open Collibra Console with a user profile that has at least the **ADMIN** role.
   » Collibra Console opens with the **Infrastructure** page.
2. In the main menu, click **Backups**.
   » The backups page appears.
3. In the upper right corner, click **Create backup**.
4. Enter the required information.

Option	Description
Environment	The Collibra Data Intelligence Cloud environment that you want to back up.
Include	You can create: ◦ a full backup: Includes all backup options. ◦ a partial backup: Includes only a couple of the backup options.
Data	The core (database) data contained in the Collibra environment. This is all the data related to communities, domains, assets, users, operating model and so on.

Option	Description
History	The history of the data (activity and snapshot-related). This can result in a very large file, depending on the activity in Collibra. If you create a backup of the data and exclude the data history, you are taking a backup of a subset of the database content. This results in a much smaller backup file, but it still contains a complete picture of the current content. This is useful when attaching a backup to a technical support ticket, if support has to reproduce a problem. Warning If you restore a backup that does not include history and was made from a version older than 5.7.x, the data quality metrics will not work correctly. To solve this issue, after you restore the backup in your latest environment, upgrade the data, create a new backup and restore the new backup in the same environment.
Configuration	The configuration of the Collibra environment.

Option	Description
Customizations	The customizations such as custom modules, styling, page-definitions and email template files. This will add files that are located in the Collibra user home directory and files that are used to customize the Collibra environment (**/opt/collibra_data/console/config**). This concerns all the files located in the following directories: ◦ **email-templates** ◦ **groovy-lib** ◦ **images** ◦ **page-definitions** ◦ **security** ◦ **styling** ◦ **translations** ◦ **modules** (if available)
Name	The name of the backup.
Description	The description to provide more information on the backup.

5. Click **Create backup**.

What's next?

A dialog box with a progress bar is displayed. You can let the backup continue in the background. If you do so, you can see the progress of the backup on the backups page.

Backup options

Option	Description
Environment	The Collibra Data Intelligence Cloud environment that you want to back up.

Option	Description
Include	You can create: • a full backup: Includes all backup options. • a partial backup: Includes only a couple of the backup options.
Data	The core (database) data contained in the Collibra environment. This is all the data related to communities, domains, assets, users, operating model and so on.
History	The history of the data (activity and snapshot-related). This can result in a very large file, depending on the activity in Collibra. If you create a backup of the data and exclude the data history, you are taking a backup of a subset of the database content. This results in a much smaller backup file, but it still contains a complete picture of the current content. This is useful when attaching a backup to a technical support ticket, if support has to reproduce a problem. Warning If you restore a backup that does not include history and was made from a version older than 5.7.x, the data quality metrics will not work correctly. To solve this issue, after you restore the backup in your latest environment, upgrade the data, create a new backup and restore the new backup in the same environment.
Configuration	The configuration of the Collibra environment.

Option	Description
Customizations	The customizations such as custom modules, styling, page-definitions and email template files. This will add files that are located in the Collibra user home directory and files that are used to customize the Collibra environment (**/opt/collibra_data/console/config**). This concerns all the files located in the following directories: • **email-templates** • **groovy-lib** • **images** • **page-definitions** • **security** • **styling** • **translations** • **modules** (if available)
Name	The name of the backup.
Description	The description to provide more information on the backup.

The Backups page

The **Backups** page contains a table of available backups. In the first column, you can find the name and date of the backup and who created it. In the second column, you can find the content that is included in the backup.

Asset View Column	Description
Name	The name of the backup. You can edit the name of the backup in this table.

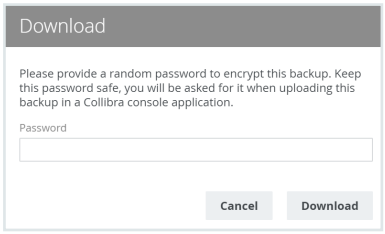
Asset View Column	Description
Description	The description that was entered when the backup was created. You can edit the description in this table.
Data	Indication whether the backup contains all current data (export of the underlying database schema).
History	Indication whether the data (if included) also holds the history (activity and snapshot data). Note Your backup can only include history data if it also includes the current data.
Configuration	Indication whether configuration files and license file are included in the backup.
Customizations	Indication whether customization files are included in the backup.
Created on	The date on which the backup was created.
Version	The version of Collibra Data Intelligence Cloud in the backup.
Size	The size of the backup
[empty]	Links to download (), restore () and delete () the backup.

Download a backup

If you want to restore a backup on another environment, you have to download it. When you download a backup, you can encrypt the backup for security reasons.

1. Open Collibra Console with a user profile that has at least the **ADMIN** role.
   » Collibra Console opens with the **Infrastructure** page.

2. In the main menu, click **Backups**.
   » The backups page appears.
3. In the row of the backup that you want to download, click **Download**.
   » The **Download** dialog box appears.
4. Type a password to encrypt the backup or leave the **Password** field empty for an unencrypted backup.



> Note The password is used to encrypt the backup file, not to password protect it. Encrypted or not, the backup file is a ZIP archive.
> You can only extract the unencrypted backup files with the usual archive managers.
> You can not extract an encrypted backup, it can only be used to upload it with Collibra Console.

5. Click **Download**.
   » The backup is downloaded to your default download folder as a ZIP archive file.

What's next?

To restore the backup on another environment, you have to upload it to that environment.

Upload a backup

If you are planning to restore a backup that you stored outside of Collibra Data Intelligence Cloud, or if you want to promote content from one Collibra environment to another, follow these steps:

1. Open Collibra Console with a user profile that has at least the **ADMIN** role.
   » Collibra Console opens with the **Infrastructure** page.
2. In the main menu, click **Backups**.
   » The backups page appears.

3. In the upper-right corner, click the **Upload backup** button.
   » The **Upload backup** dialog box appears.
4. Enter the required information:

Option	Description
Password protected	The option to choose when the backup was encrypted when downloading it.
Not password protected	The option to choose when the backup was not encrypted when downloading it.
Password	The password to unencrypt the backup when uploading it. This option is required when the **Password protected** option is selected.

5. Do one of the following:
   - Browse to the location of your backup file with a file explorer, grab the backup file with your mouse pointer and drag it to the dotted box in the **Upload backup** dialog box.
   - Click , browse to the location of your backup file in the window that opens, select the file and click **Open**.

   » Your backup upload starts automatically.

What's next?

You can now restore the backup on your environment.

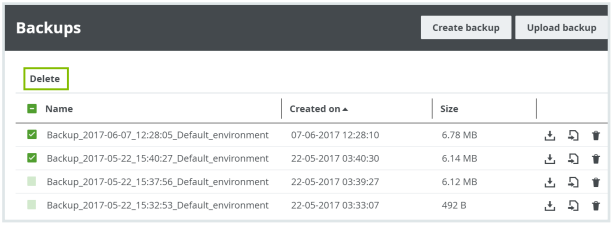
Delete a backup

If you no longer need a certain backup or backups, you can delete them. This will free up disk space.

Steps

1. Open Collibra Console with a user profile that has at least the **ADMIN** role.
   » Collibra Console opens with the **Infrastructure** page.

2. In the main menu, click **Backups**.
   » The backups page appears.
3. Do one of the following:

To delete only one backup at a time:	Click 🗑 at the end of the line of the backup that you want to delete.
To delete multiple backups at once:	Select the check boxes of the backups that you want to delete and click **Delete** at the top of the table. 

   » The **Delete Backups** dialog box appears.
4. Click **Yes** to confirm.

Create a backup schedule

A backup schedule allows you to automatically create backups in a defined schedule. This reduces the risk of forgetting to create backups and possible data loss.

> Note The backups that are taken with a backup schedule are full backups, not incremental backups.

Steps

1. Open Collibra Console with a user profile that has at least the **ADMIN** role.
   » Collibra Console opens with the **Infrastructure** page.
2. In the main menu, click **Backups**.
   » The backups page appears.
3. In the tab pane, click **Backup schedules**.

4. In the upper-right corner, click **Create backup schedule**.
5. Enter the required information:

Option	Description
Environment	The Collibra Data Intelligence Cloud environment that you want to back up.
Include	You can create: ◦ a full backup: Includes all backup options. ◦ a partial backup: Includes only a couple of the backup options.
Data	The core (database) data contained in the Collibra environment. This is all the data related to communities, domains, assets, users, operating model and so on.
History	The history of the data (activity and snapshot-related). This can result in a very large file, depending on the activity in Collibra. If you create a backup of the data and exclude the data history, you are taking a backup of a subset of the database content. This results in a much smaller backup file, but it still contains a complete picture of the current content. This is useful when attaching a backup to a technical support ticket, if support has to reproduce a problem. Warning If you restore a backup that does not include history and was made from a version older than 5.7.x, the data quality metrics will not work correctly. To solve this issue, after you restore the backup in your latest environment, upgrade the data, create a new backup and restore the new backup in the same environment.

Option	Description
Configuration	The configuration of the Collibra environment.
Customizations	The customizations such as custom modules, styling, page-definitions and email template files. This will add files that are located in the Collibra user home directory and files that are used to customize the Collibra environment (**/opt/collibra_data/console/config**). This concerns all the files located in the following directories: ◦ **email-templates** ◦ **groovy-lib** ◦ **images** ◦ **page-definitions** ◦ **security** ◦ **styling** ◦ **translations** ◦ **modules** (if available)
Name	The name of the backup.
Description	The description to provide more information on the backup.
Max retained	The number of backups that you want to retain. The default value is 30.
Cron expression	The schedule of the backups. For more information about the CRON syntax, see the CRON appendix. If you create an invalid Cron pattern, Collibra Data Intelligence Cloud stops responding.

6. Click **Create backup schedule**.
   » The schedule appears in the schedule table.

> Tip You can create multiple schedules for one environment. For example, one schedule for monthly backups that are kept for two years and another schedule for weekly schedules that are kept for only a couple of months.

Backup schedule options

Option	Description
Environment	The Collibra Data Intelligence Cloud environment that you want to back up.
Include	You can create: • a full backup: Includes all backup options. • a partial backup: Includes only a couple of the backup options.
Data	The core (database) data contained in the Collibra environment. This is all the data related to communities, domains, assets, users, operating model and so on.

Option	Description
History	The history of the data (activity and snapshot-related). This can result in a very large file, depending on the activity in Collibra. If you create a backup of the data and exclude the data history, you are taking a backup of a subset of the database content. This results in a much smaller backup file, but it still contains a complete picture of the current content. This is useful when attaching a backup to a technical support ticket, if support has to reproduce a problem. Warning If you restore a backup that does not include history and was made from a version older than 5.7.x, the data quality metrics will not work correctly. To solve this issue, after you restore the backup in your latest environment, upgrade the data, create a new backup and restore the new backup in the same environment.
Configuration	The configuration of the Collibra environment.
Customizations	The customizations such as custom modules, styling, page-definitions and email template files. This will add files that are located in the Collibra user home directory and files that are used to customize the Collibra environment (**/opt/collibra_data/console/config**). This concerns all the files located in the following directories: • **email-templates** • **groovy-lib** • **images** • **page-definitions** • **security** • **styling** • **translations** • **modules** (if available)

Option	Description
Name	The name of the backup.
Description	The description to provide more information on the backup.
Max retained	The number of backups that you want to retain. The default value is 30.
Cron expression	The schedule of the backups. For more information about the CRON syntax, see the CRON appendix. If you create an invalid Cron pattern, Collibra Data Intelligence Cloud stops responding.

Backup schedules overview

The **Backup schedules** page contains a table of the backup schedules.

> Note The backups that are taken with a backup schedule are full backups, not incremental backups.

Backup schedules column	Description
Name	The name of the backup schedule. You can edit the name of the schedule in this table.
Description	The description of the backup schedule. You can edit the description in this cell.
Data	Indication whether the backup includes the Collibra environment data. This is all the data related to communities, domains, assets, users, operating model and so on.

Backup schedules column	Description
History	Indication whether the data also holds the history. Note Your backup can only include the data history if **Data** is selected.
Configuration	Indication whether configuration files and license file are included in the backup.
Customizations	Indication whether customization files are included in the backup.
Max retained	The number of backups that are kept on disk. When reaching this number, a new backup will still be created, but the oldest backup will be deleted.
Environment	The name of the environment that will be backed up.
Cron expression	The schedule when the backups will be taken. For more information about the Cron syntax, see cron syntax. If your Cron expression could affect the performance of your environment, for example multiple times an hour, an exclamation mark will be shown. If you create an invalid Cron pattern, Collibra Data Intelligence Cloud stops responding.
Next execution	The date and time when the next backup will be created.
[empty]	Button to delete ( ) the backup.

Edit a backup schedule

When you have created a backup schedule, you can always change it afterwards, for example to change the name or schedule. For more information about the different columns, see Backup schedules overview.

Steps

1. Open Collibra Console with a user profile that has at least the **ADMIN** role.
   » Collibra Console opens with the **Infrastructure** page.
2. In the main menu, click **Backups**.
   » The backups page appears.
3. In the tab pane, click **Backup schedules**.
4. In the schedule overview table, double-click any value in the table and edit it.

   > Note You cannot change the **Environment** or **Next execution** columns.

5. Click ✔ to save the change.

> Tip If you edit the maximum number of retained backups to a number that is less than the number of currently retained backups, the first cleanup process will clean all necessary backups so that only the specified number of retained backups will remain.

Delete a backup schedule

When you no longer need one or more backup schedules, you can always delete them.

Steps

1. Open Collibra Console with a user profile that has at least the **ADMIN** role.
   » Collibra Console opens with the **Infrastructure** page.
2. In the main menu, click **Backups**.
   » The backups page appears.
3. In the tab pane, click **Backup schedules**.

4. In the schedule overview table, click next to the schedule that you want to delete.
   To delete multiple schedules at once, select them in the first column and click **Delete**.
   » The **Delete backups** dialog box appears.
5. Click **Delete**.

Restoring a backup

When you restore a backup, your Collibra Data Intelligence Cloud environment is reverted to the state it was in when the backup was created. You restore a backup via Collibra Console.

You have the following distinct back-up scenarios:

- You restore a backup that was created in the same Collibra major version as your current version.
- You restore a backup that was created in a previous version, via the environment page.
- You restore a backup that was created in a previous version, via the **Backups** page.

To restore a backup, you need:

- The ADMIN or SUPER role.
- Free disk space that is twice the backup size.

Restore a backup

In this section, we describe how to restore a backup that was created in the same Collibra Data Intelligence Cloud version as your current version.

> Warning If you restore a backup, you will lose all your current Collibra data.

Steps

1. Open Collibra Console with a user profile that has at least the **ADMIN** role.
   » Collibra Console opens with the **Infrastructure** page.

2. In the main menu, click **Backups**.
   » The backups page appears.
3. Click **Restore** in the row of the backup that you want to restore.
4. Fill in the **Restore from backup** form:

Option	Description
Environment	Select the environment in which you want to restore the backup.
Backup	Select the backup that you want to restore.
Include	You can restore a backup fully or partially. Note The backup itself should include the items you select below. For example, if the backup does not contain the configuration and you include configuration for restoring the backup, it will be restored successfully, though without the configuration.
Data	Select to include the core (database) data from the backup. This is all the data related to communities, domains, assets, users, operating model and so on.
History	Select to include all historic data (activity and snapshot-related) from the backup.
Configuration	Select to include the configuration and license from the backup. Warning Do not select this option if you restore a backup from another environment, as this will overwrite the configuration of your current environment. It can prevent your environment from starting up again.

Option	Description
Customizations	Select to include custom modules, styling, page-definitions and email template files. This will add files to the Collibra user home directory and files that are used to customize the Collibra DGC instance (**/opt/collibra_data/console/config**). This concerns all the files in the following directories: ◦ **email-templates** ◦ **groovy-lib** ◦ **images** ◦ **page-definitions** ◦ **security** ◦ **styling** ◦ **translations** ◦ **modules** (if available)

5. Click **Restore backup**.
6. Click **Close**.

What's next?

If the configuration file is restored, the database connection information in the backup is ignored. In other words, the restore process makes sure that the entire backed-up configuration is applied to the current configuration without modifying the current database connection. This means that the existing database connection is used to restore the data. Also, after the restore, the Collibra instance continues to use the same database connection.

Note
- You cannot restore a backup from a Collibra environment that is using a database with a case-sensitive collation sequence into a Collibra environment that is using a database with a collation sequence that is not case-sensitive. If you want to use backup and restore to migrate Collibra content, ask your database administrator to determine whether the collation sequences of the databases used by your source and target Collibra environments are compatible.
- If you restore a backup without the history, Collibra does not restore the activity stream as it was. Instead, it creates the entries in the activity stream as if they were done by the System User on the restore date.
- You can also use the restore functionality to promote content from one Collibra environment to another if the environments run the same version.

Restore a backup from a previous version

If you want to upgrade Collibra Data Intelligence Cloud to 2022.06, you can restore the backup from a previous version on a newly installed 2022.06 environment.

Warning
- If you restore a backup, you will lose all current Collibra data.
- Backups from versions older than 5.5.x do no include certificates. You have to reinstall any necessary certificates, for example for secure communication with the Jobserver.
- If you restore a backup that does not include history and was made from a version older than 5.7.x, the data quality metrics will not work correctly. To solve this issue, after you restore the backup in your latest environment, upgrade the data, create a new backup and restore the new backup in the same environment.

Steps

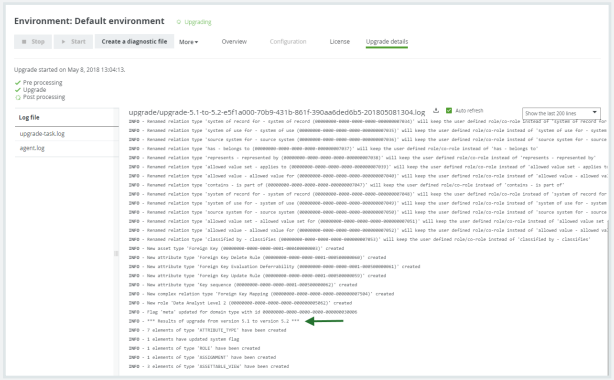
1. Open Collibra Console with a user profile that has at least the **ADMIN** role.
   » Collibra Console opens with the **Infrastructure** page.
2. Click the name of an environment to show its details.

3. Click **More** → **Restore from a previous version**.
4. Click the version of the backup:

Previous version	Substeps
4.5.x or 4.6.x	i. Accept the limitations for restoring this backup. See Limitations of the upgrade from 4.5.x / 4.6.x. ii. Click **Next**. iii. Upload or drop the backup in the **Upload backup** box. iv. Click **Restore backup and upgrade**. » The backup is restored and the upgrade of the repository is automatically started.
5.x	i. Select the **Upload backup** option: ■ **Upload backup**: Upload or drop the backup in the **Upload backup** box. ■ **Choose an existing backup**: Select the backup that you want to restore from the **Backup** drop-down menu. ii. Click **Restore backup and upgrade**. iii. Select the options that you want to restore. iv. Click **Restore backup**. » The backup is restored and the upgrade of the repository is automatically started.

Note
If you look in the upgrade logs (**upgrade-task.log**), you will see one of the following lines near the end:

- `INFO - COMPLETED Upgrade the database version to [5.2]`
- `INFO - *** Results of upgrade from version 5.1 to version 5.2 ***`



This means that the repository has upgraded to 5.2 at this point. The upgrade to a later version, for example 5.7.12, is performed during the start of the DGC service and is therefore not logged in this log file, but in **dgc.log**, which you can find in the diagnostic file.

Restore a backup from a previous version via Backups

If you restore a backup that was created in a previous major Collibra Data Intelligence Cloud version via the **Backups** page, you then have to upgrade the database, as described in this section.

However, we highly recommend to follow this procedure if you restore a backup from a previous major version.

Tip If your backup was created in the same version as your current environment, follow the restore a backup procedure.

Warning
- If you restore a backup, you will lose all current Collibra data.
- Backups from versions older than 5.5.x do no include certificates. You have to reinstall any necessary certificates, for example for secure communication with the Jobserver.
- If you restore a backup that does not include history and was made from a version older than 5.7.x, the data quality metrics will not work correctly. To solve this issue, after you restore the backup in your latest environment, upgrade the data, create a new backup and restore the new backup in the same environment.

Steps

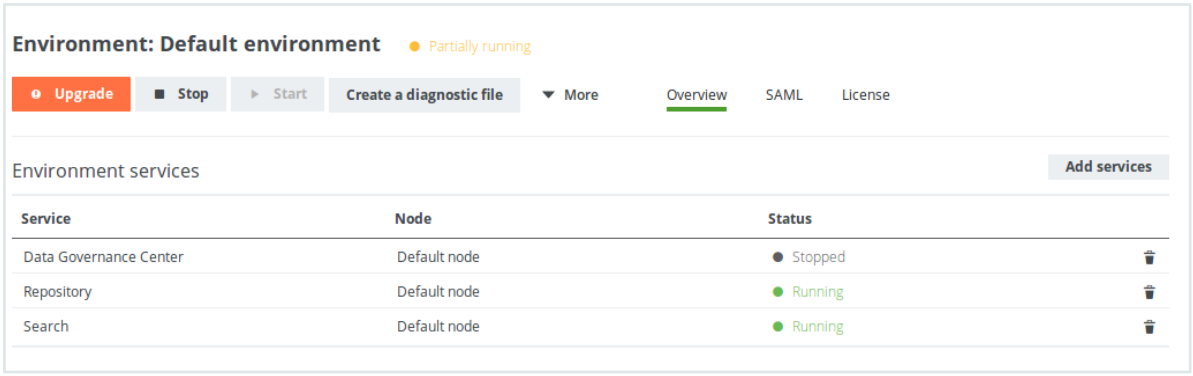
1. Open Collibra Console with a user profile that has at least the **ADMIN** role.
   » Collibra Console opens with the **Infrastructure** page.
2. In the main menu, click **Backups**.
   » The backups page appears.
3. Click **Restore** in the row of the backup that you want to restore.
4. Fill in the **Restore from backup** form:

Option	Description
Environment	Select the environment in which you want to restore the backup.
Backup	Select the backup that you want to restore.
Include	You can restore a backup fully or partially. Note The backup itself should include the items you select below. For example, if the backup does not contain the configuration and you include configuration for restoring the backup, it will be restored successfully, though without the configuration.

Option	Description
Data	Select to include the core (database) data from the backup. This is all the data related to communities, domains, assets, users, operating model and so on.
History	Select to include all historic data (activity and snapshot-related) from the backup.
Configuration	Select to include the configuration and license from the backup. Warning Do not select this option if you restore a backup from another environment, as this will overwrite the configuration of your current environment. It can prevent your environment from starting up again.
Customizations	Select to include custom modules, styling, page-definitions and email template files. This will add files to the Collibra user home directory and files that are used to customize the Collibra DGC instance (**/opt/collibra_data/console/config**). This concerns all the files in the following directories: ◦ **email-templates** ◦ **groovy-lib** ◦ **images** ◦ **page-definitions** ◦ **security** ◦ **styling** ◦ **translations** ◦ **modules** (if available)

5. Click **Restore backup**.
6. Click **Close**.
   » When the restore is complete, your environment will be partially running.

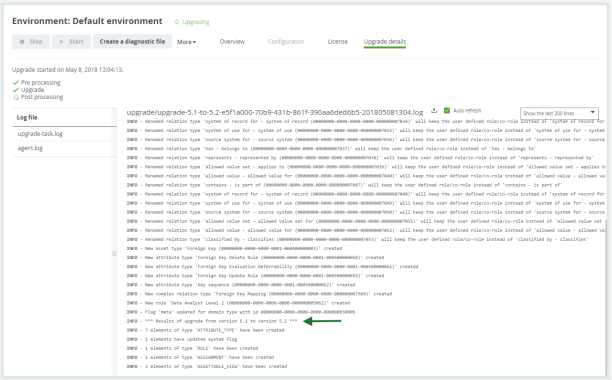
7. In the main menu, click **Infrastructure**.
8. Click the name of the environment that is partially running.
   » The details of the environment appear.



9. Click **Upgrade** to start the upgrade of your restored data.
10. In the **Upgrade** dialog box, click **Upgrade**.
    » After the upgrade, your environment is up and running again.

Note
If you look in the upgrade logs (**upgrade-task.log**), you will see one of the following lines near the end:

- `INFO - COMPLETED Upgrade the database version to [5.2]`
- `INFO - *** Results of upgrade from version 5.1 to version 5.2 ***`



This means that the repository has upgraded to 5.2 at this point. The upgrade to a later version, for example 5.7.12, is performed during the start of the DGC service and is therefore not logged in this log file, but in **dgc.log**, which you can find in the diagnostic file.

Customize Collibra DGC with a backup restore

If you have customized your Collibra Data Intelligence Cloudinstance and you want to keep those customizations after an upgrade, you can create a backup and restore the backup to replace the files in the Collibra user home directory.

1. Create a backup that includes customizations.
2. Download the backup.
3. Unzip the file in a new directory.
4. Open the new directory.
5. Modify the files.
6. Select all the unzipped folders and files, including **backup.properties** and zip them again.

   > Note Do not zip the containing directory but the individual files and directories, so that you get the same structure as the original ZIP file.

7. Upload this backup to your Collibra instance.
8. Restore the backup, including customizations.

What's next?

Test all customizations.

Back up and download with the REST API

It is highly recommended that you back up the Collibra Data Intelligence Cloud instances on a regular basis. Instead of doing this manually with Collibra Console, you can also use the Collibra REST API in combination with a scheduling tool like CRON. As such, you create backups at fixed moments and you can no longer forget to back up your Collibra DGC.

You can use the Chrome browser app POSTMAN to build and trigger REST calls. Below you can find examples of how to create and download a backup using the POSTMAN plugin.

1. Retrieve an environment ID in one of the following ways:
   - In Collibra Console, go to the **Environments** tab and click the relevant environment to open its details. The environment ID appears in the address bar of your browser, after the colon sign.
     Example: https://console.collibra.com/#/environments/Environment:**52df07dc-f69c-4995-af4a-547e5d3a83cf**
   - Retrieve a list of all environments with a GET operation:

```
GET http://<your_dgc_console_host>/rest/environment

[
  {
    "createdDate": 1474031943.171,
    "modifiedDate": 1474037344.178,
    "id": "3Acfb88804-c7da-4e02-b1a5-bf148691e91a",
    "name": "Default environment",
    "nodeList": [
      {
        "createdDate": 1474031943.278,
        "modifiedDate": 1474031943.278,
        "id": "b2bb7dc1-2fb3-4dc0-83af-23e80f28160e",
        "hostName": "localhost",
        "port": 4401,
        "name": "Data Governance Center",
        "managedServiceSet": [
          {
            "type": "DGC",
            "status": "RUNNING",
            "errorMessage": ""
          },
          {
            "type": "REPOSITORY",
            "status": "RUNNING",
            "errorMessage": ""
          }
        ]
      }
    ],
    "status": "RUNNING",
    "intendedStatus": "RUNNING"
  }
]
```

2. Use the following REST API data to create a backup:
    - **Path**: http(s)://<your_dgc_console_host>/rest/backup/<environment_id>
    - **HTTP method**: POST
    - **Authorization**:
        - **Type**: Select Basic Auth.
        - **Username** and **Password**: Provide a Collibra Console user with sufficient user rights.
    - **Headers**: Provide the following two key-value pairs:
        - Key-Value 1: The first key-value pair is already be filled in (Authorization / Basic QWRta...).
        - Key-Value 2: Content-Type - application/json
    - **Body**: Provide the backup information in JSON format, see the following example:

```
{
    "name": <name for the backup>,
    "description": <optional description>,
    "database": "dgc",
    "dgcBackupOptionSet":
["CUSTOMIZATIONS","CONFIGURATION"],
    "repoBackupOptionSet": ["DATA","HISTORY"]
}
```

> Note
> The **dgcBackupOptionSet** list can contain "CUSTOMIZATIONS" and "CONFIGURATION" or you can leave it empty.
>
> The **repoBackupOptionSet** list can contain "DATA" and "HISTORY". Keep in mind that when you add "HISTORY" to the list, you have to add "DATA" too.

3. Click **Send** to execute the API call.
4. Use the following data to download a backup, for example, by using cURL.
    - **Path**: http(s)://<your_dgc_console_host>/rest/backup/file/<backup_id>
    - **File name**: freely chosen file name
    - **HTTP method**: POST
    - **Username** and **Password**: valid credentials to download a backup from Collibra Console.

```
~$ curl -H "Content-Type:application/x-www-form-
urlencoded" -O \

   -X POST http://<your_dgc_console_
host>/rest/backup/file/<backup_id> \

   -d "key=<password_to_encrypt_backup>" \

   --user <username>:<password>

#for example

~$ curl -H "Content-Type:application/x-www-form-
urlencoded" -O \

   -X POST http://<your_dgc_console_
host>/rest/backup/file/6bb2bc51-b1d4-42a1-a72b-
668207eb5f11 \

   -d "key=aStrongPasswordToEncryptBackup" \

   --user Admin:<password>


* Server auth using Basic with user 'Admin'
> GET /rest/backup/6bb2bc51-b1d4-42a1-a72b-
668207eb5f11?my_downloaded_backup.zip HTTP/1.1
> Authorization: Basic QWRtaW46YWRtaW4=
> User-Agent: curl/7.19.7 (x86_64-redhat-linux-gnu)
libcurl/7.19.7 NSS/3.21 Basic ECC zlib/1.2.3
libidn/1.18 libssh2/1.4.2
> Host: localhost:4402
> Accept: */*
> Content-Type:application/zip
>
  % Total    % Received % Xferd  Average Speed   Time
 Time     Time  Current
                                 Dload  Upload   Total
 Spent    Left  Speed
  0     0    0     0    0     0      0      0 --:--:--
--:--:-- --:--:--     0< HTTP/1.1 200 OK
< Server: Apache-Coyote/1.1
< Set-Cookie: ConsoleSessionId=d4be0069-9770-45c0-8a65-
1ca240436b6a; Path=/; HttpOnly
< Set-Cookie: rememberMe=deleteMe; Path=/; Max-Age=0;
Expires=Mon, 19-Sep-2016 11:18:37 GMT
< Content-Disposition: attachment; filename = "DGC_
2016_09_19_13_51_28_Default_environment.zip"
< Cache-Control: private, must-revalidate
< Content-Type: application/zip
< Transfer-Encoding: chunked
```

```
< Date: Tue, 20 Sep 2016 11:18:37 GMT
<
{ [data not shown]
100 12.4M    0 12.4M    0     0  72.7M      0 --:--:--
--:--:-- --:--:-- 76.2M* Connection #0 to host
localhost left intact

* Closing connection #0

~$ ls
6bb2bc51-b1d4-42a1-a72b-668207eb5f11?my_downloaded_
backup.zip
```

This action downloads the backup as ZIP archive to your local workstation.

The argument `-d "key=<password...>"` is optional. It is used to encrypt the backup with a password. If you upload an encrypted backup, you have to provide this argument again.

REST API - List of backups

To retrieve the list of backups via the REST API:

```
GET https://<your_dgc_console_host>/rest/backup
```

```
[{"createdDate":1496985583.834000000,
  "modifiedDate":1496985583.846000000,
  "id":"c2a3e499-bb0f-4016-85aa-6efcbcd66df3",
  "backupInformation":{
    "appVersion":"5.1",
    "createdByEmail":"admin@example.com",
    "date":1496843737.891000000,
    "dgcBackupOptionSet":["CUSTOMIZATIONS"],
    "repoBackupOptionSet":["DATA","CONFIGURATION","HISTORY"]},
    "size":86869667,
    "stepStateMap":{},
    "inProgress":false},
 {"createdDate":1496906907.017000000,
  "modifiedDate":1496906910.141000000,
  "id":"3a749282-cbc7-4721-b598-67d785f416bc",
  "backupInformation":{
```

```
    "appVersion":"5.1",
    "createdByEmail":"john.doe@collibra.com",
    "date":1496906907.016000000,
    "dgcBackupOptionSet":["CUSTOMIZATIONS"],
    "repoBackupOptionSet":["DATA","CONFIGURATION","HISTORY"]},
    "size":7155216,
    "stepStateMap":{
      "REPOSITORY":{
        "status":"COMPLETED",
        "errorMessage":""},
      "DGC":{
        "status":"COMPLETED",
        "errorMessage":""},
      "POST_PROCESSING":{
        "status":"COMPLETED",
        "errorMessage":""},
      "PRE_PROCESSING":{
        "status":"COMPLETED",
        "errorMessage":""}},
  "inProgress":false}]
```


REST API - Delete a backup

To delete a backup via the REST API:

```
DELETE https://<your_dgc_console_host>/rest/-
backup/file/<backup-id>
```

If the operation succeeds, there is no response. If something went wrong, a response is shown in JSON format.

Example error:

```
{
    "statusCode": 500,
    "titleMessage": "Internal Server Error",
    "helpMessage": "Please report a problem as this situation
should not occur.",
    "userMessage": "The server encountered an unexpected con-
dition which prevented \
                   it from fulfilling the request. The message
```

```
was: null"
}
```

Diagnostic files

A diagnostic file is a ZIP file containing files that can be used by support to help solve a support ticket.

For example, it contains log files and some configuration files of the database. It does not contain any actual database content. For more information about logs, see Logging.

In this chapter

Diagnostic files 210
Create a diagnostic file 212
Download a diagnostic file 214
Delete a diagnostic file 215
Edit the environment log settings 216
Logging 217
Content of a diagnostic file 219

Diagnostic files

A diagnostic file is a ZIP file containing files that can be used by support to help solve a support ticket.

For example, it contains log files and some configuration files of the database. It does not contain any actual database content. For more information about logs, see Logging.

Content

> Tip You can also find the log files in the **collibra_data** directory, for example **/opt/collibra_data/agent/logs**.

Console

- config: The configuration files of all environments, JVM settings, log settings and server settings.
- installation: The installation log file and Collibra installation configuration.
- logs:
  - console_wrapper.log
  - console.log: Registration of all actions that are done in Collibra Console, for example sign in and out, start and stop an environment, create an environment, ...
  - database*.log: All transactions of the Collibra Console database.

Agent

- config: All configuration files of the agent.
- installation: The installation log file and Collibra Data Intelligence Cloud installation configuration.
- logs:
  - agent_commands.log
  - agent_wrapper.log

    - agent.log: Registration of all actions that are done by the Collibra agent, for example starting jobs, database operations, ...

DGC

- config: All configuration files of the DGC service such as the configuration, search properties, JVM configuration and so on.
- logs:
    - dgc_jobs.log: Logging of the actual execution of a job.
    - dgc_recommender.log: Logging of the recommendation tool to create data sets for Data Catalog.
    - dgc_validations.log
    - dgc.log: Logging of the DGC service.

Repo

- config: All configuration files of the PostgreSQL database.
- logs: Logging of the PostgreSQL operations.
- blockedLocksInfo.json
- componentCalls.dmp
- locksInfo.json
- customMetrics.txt: Statistics of the repository such as number of assets, assets per column, active users and number of jobs.

Search

- config: The physical configuration of the Search service such as port configuration, hostname, keystore and JVM settings.
- logs: All log files of the Search service.
- JSON files with the Search service's cluster configuration, status, statistics and so on.

Spark-jobserver

- config: All configuration files of the Jobserver service, including its database configuration.
- logs: All log files of the Jobserver service.

Monitoring

- config: All configuration files of the Monitoring service.
- logs: All log files of the Monitoring service.
- snapshot: The snapshot of the monitoring database that is created when you create the diagnostics file.

Other files

- benchmark-result.txt: This file contains the performance benchmark results if selected during the creation of the diagnostic file.
- infraReport.txt: This file contains the details of the infrastructure on which Collibra is installed.
- installation.log: This file contains the logging of the installation process.
- installationConfig.json: This file contains the settings that were selected during the installation process such as TCP ports and installation directory.

Create a diagnostic file

To create a diagnostic file, follow these steps:

1. Open Collibra Console.
   » Collibra Console opens with the **Infrastructure** page.

2. Do one of the following:

To create a diagnostic file...	Follow these steps
from an environment	a. On the **Infrastructure** page, click the environment for which you want to create a diagnostic file. b. Click **Create a diagnostic file**. » The **Create a diagnostic file** dialog box appears.
from the **Diagnostic files** view	a. In the main menu, click **Diagnostic files**. b. In the top right-corner, click **Create a diagnostic file**. » The **Create a diagnostic file** dialog box appears.

3. Enter the required information.

Field	Description
Environment	The environment from which you want to create a diagnostic file.
Name	The name for the diagnostic file. There are no specific limitations for the file name.
Select files to include	The files that you want to include in the diagnostic file.
Console files	The logging of the actions in Collibra Console.
DGC files	The logging of the DGC service and a Collibra thread dump.
Repository files	The logging of the Repository service and the Repository service configuration and metrics.

Field	Description
Jobserver files	The logging of the Jobserver service.
Node files	The logging of the nodes, these are the most important operating system log files.
Performance benchmarks	The performance benchmarks of your environment. These benchmarks will perform some predefined actions on your environment and create a report of it. > Note This option will slow down your system and will take a few minutes for the diagnostics to be created.
Monitoring data	The monitoring data of your environment. This will create a snapshot of the monitoring database. You can only select this option if you have installed the Monitoring service. > Note This option will significantly increase the diagnostic file size.

4. Click **Create a diagnostic file**.

What's next?

The diagnostic file appears in the **Diagnostic files** section of Collibra Console. For more information about the content of the diagnostic file, see Content of a diagnostic file.

Download a diagnostic file

To download a diagnostic file, follow these steps:

1. Open Collibra Console.
2. In the main menu, click **Diagnostic files**.
3. Do one of the following:

To download...	Follow these steps
a single diagnostic file	a. Click ⬇ on the far right of the diagnostic file that you want to download.
one or more diagnostic files	a. Select the checkbox in front of one or more diagnostic files. Tip You can select all checkboxes in one go by selecting the checkbox in the **Name** column. b. Click **Download**.

4. Depending on your browser and browser settings, the files are downloaded to a default location or a dialog box appears to specify the location for the downloads.

Delete a diagnostic file

To delete a diagnostic file, follow these steps:

1. Open Collibra Console.
2. In the main menu, click **Diagnostic files**.
3. Do one of the following:

To delete...	Follow these steps
a single diagnostic file	a. Click 🗑 on the far right of the diagnostic file that you want to download.

To delete...	Follow these steps
one or more diagnostic files	a. Select the checkbox in front of one or more diagnostic files. Tip You can select all checkboxes in one go by selecting the checkbox in the **Name** column. b. Click **Delete**.

Edit the environment log settings

Warning It is highly recommended not to edit the log levels. By doing so, you can influence the performance negatively.

To edit the environment log settings, follow these steps:

1. Open Collibra Console.
   » Collibra Console opens with the **Infrastructure** page.
2. In the tab pane, click the **Data Governance Center** service of the environment whose log settings you need.
3. Click **Logs**.
4. Above the table, to the right, click **Settings**.
5. Edit the settings to suit your company's needs by hovering over the **Level** column and click .

   See Environment log settings for DGC services and Environment log settings for Repository services.

Environment log settings for Repository services

Slow query logging: With this parameter, you can define the minimum amount of time a query has to last before it is logged in the Collibra Data Intelligence Cloud log files as a slow query. For example, if you set 10 milliseconds, then Collibra only logs queries that take longer than 10 milliseconds.

Logger level:

- **Repository messages**: The minimum severity level of the SQL messages that you want to log.
- **Repository statements**: The minimum severity level of the SQL statements that you want to log.

Repository log levels from low to high (the lower the log level, the more issues that are logged):

- DEBUG5
- DEBUG4
- DEBUG3
- DEBUG2
- DEBUG1
- INFO
- NOTICE
- WARNING
- ERROR
- LOG
- FATAL
- PANIC

> Warning It is highly recommended not to change the log levels. By doing so, you can influence the performance negatively.

Logging

When you encounter an issue in Collibra Data Intelligence Cloud, the log file can provide a lot of interesting information about what is happening/has happened.

Whenever there seems to be something wrong with the product, it is good practice to check the log file first.

When you report a problem to Collibra support, it is recommended to immediately attach the log file(s) to your support ticket.

Depending on the configuration, the log file shows you:

- Errors that have occurred on the server side
- Basic information about tasks that are performed
- Detailed information about every action that has been performed
- The date and time an event occurred
- The thread that is executing an action
- The log level of the message

You can see consult the log files in Collibra Console, or you can download the log files in Collibra Console.

If you download the log files, they are compressed (zipped). You have to decompress them before you can read them. You can use any text editor, like Notepad to read the log files.

> Tip If the log file is displayed as a single, long line in Notepad, download and use the (free) text editor Notepad++ from http://notepad-plus-plus.org

The log files can also be found on the file system of the server **/opt/collibra_data/<component>/logs**. For more information on the Collibra directory structure, see Product architecture.

Open a log file

Whenever there is something wrong with the product, it is good practice to check the log file first. You can open the log files of each service in Collibra Console.

To open the log files of Collibra Console itself, see Open a Collibra Console log file.

Steps

1. Open Collibra Console.
   » Collibra Console opens with the **Infrastructure** page.
2. In the tab pane, click the name of an environment.
3. In the **Environment Services** table, click the name of a service whose log files you want to consult.
4. On the details page of the service, click **Logs**.
5. Click the name of a log file to open it.

Content of a diagnostic file

These are the files contained in a diagnostic file.

> Tip You can also find the log files in the **collibra_data** directory, for example **/opt/collibra_data/agent/logs**.

Console

- config: The configuration files of all environments, JVM settings, log settings and server settings.
- installation: The installation log file and Collibra installation configuration.
- logs:
  - console_wrapper.log
  - console.log: Registration of all actions that are done in Collibra Console, for example sign in and out, start and stop an environment, create an environment, ...
  - database*.log: All transactions of the Collibra Console database.

Agent

- config: All configuration files of the agent.
- installation: The installation log file and Collibra Data Intelligence Cloud installation configuration.
- logs:
  - agent_commands.log
  - agent_wrapper.log
  - agent.log: Registration of all actions that are done by the Collibra agent, for example starting jobs, database operations, ...

DGC

- config: All configuration files of the DGC service such as the configuration, search properties, JVM configuration and so on.

- logs:
  - dgc_jobs.log: Logging of the actual execution of a job.
  - dgc_recommender.log: Logging of the recommendation tool to create data sets for Data Catalog.
  - dgc_validations.log
  - dgc.log: Logging of the DGC service.

Repo

- config: All configuration files of the PostgreSQL database.
- logs: Logging of the PostgreSQL operations.
- blockedLocksInfo.json
- componentCalls.dmp
- locksInfo.json
- customMetrics.txt: Statistics of the repository such as number of assets, assets per column, active users and number of jobs.

Search

- config: The physical configuration of the Search service such as port configuration, hostname, keystore and JVM settings.
- logs: All log files of the Search service.
- JSON files with the Search service's cluster configuration, status, statistics and so on.

Spark-jobserver

- config: All configuration files of the Jobserver service, including its database configuration.
- logs: All log files of the Jobserver service.

Monitoring

- config: All configuration files of the Monitoring service.
- logs: All log files of the Monitoring service.

- snapshot: The snapshot of the monitoring database that is created when you create the diagnostics file.

Other files

- benchmark-result.txt: This file contains the performance benchmark results if selected during the creation of the diagnostic file.
- infraReport.txt: This file contains the details of the infrastructure on which Collibra is installed.
- installation.log: This file contains the logging of the installation process.
- installationConfig.json: This file contains the settings that were selected during the installation process such as TCP ports and installation directory.

Chapter 8

Collibra Console settings

The Collibra Console settings allow you to define how you can access Collibra Console, to manage the Collibra Console users, to configure its mail settings and so on.

In this chapter

Collibra Console users 223

Edit the Collibra Console settings 227

Edit the Collibra Console server settings 253

Open a Collibra Console log file 259

Collibra Console users

The Collibra Console user is different from the Collibra Data Governance Center user. The Collibra Console user can only use Collibra Console and cannot access Collibra DGC, even if the credentials are identical.

Create a Collibra Console user

To create a user to Collibra Console, follow these steps:

1. In the main menu, click **Console settings**.
2. In the tab pane, click **Users**.
3. Above the table, to the right, click **Add User**.
   » The **Add User** dialog box appears.
4. Enter the required information.

Field	Description
Username	Provide a username for the new Collibra Console user.
Email	Provide the corresponding email address for this user.
Role	◦ **Role**: Select a role for the user. For more information about roles, see Collibra Console roles. ▪ READ ▪ ADMIN ▪ SUPER

5. Click **Add User**.

What's next?

The new Collibra Console user receives an email with the username and a link to set a password.

Change the role of a Collibra Console user

To change the role of a Collibra Console user, follow these steps:

1. In the main menu, click **Console settings**.
2. In the tab pane, click **Users**.
3. In the **Role** column, hover your pointer over the role next to the user whose role you want to change and click ✎.
4. Click ▼ and click on the new role for the user.

5. Click ✓ to save the change.

Reset the password of a Collibra Console user

If you have forgotten your Collibra Console password, a user with at least the ADMIN role can reset your password.

> Note
> - A user with the ADMIN role cannot reset the password of a user with the SUPER role.
> - You can only reset a password of an other user. To edit your own password, see Edit your Collibra Console user password.
> - At any time, users can edit their own password.

To reset the password of a Collibra Console user, follow these steps:

1. In the main menu, click **Console settings**.
2. In the tab pane, click **Users**.
3. Hover over 🔑 next to the user whose password you want to reset and click **Reset password**.
   » The **Reset password** dialog box appears.
4. In the **Reset password** dialog box, click **Send reset email**.

What's next?

The user receives an email with a link to reset password the password.

Edit your Collibra Console user password

At any time, Collibra Console users can change their own password.

To change your own password, follow these steps:

1. In the main menu, click **Console settings**.
2. In the tab pane, click **Users**.
3. Click next to your own username.
   » The **Change password** dialog box appears.
4. Enter the required information.

Field	Description
Old password	Type your current password.
New password	Type your new password.
Repeat password	Type your new password a second time.

5. Click **Change password**.

Delete a Collibra Console user

To delete a Collibra Console user, follow these steps:

1. In the main menu, click **Console settings**.
2. In the tab pane, click **Users**.
3. Click next to the user that you want to delete.

> Note Keep in mind that deleting a user is an irreversible action.

4. Click **Delete** to confirm your action.

What's next?

The user is deleted and can no longer access Collibra Console.

Collibra Console roles

A Collibra Console user can have three different roles:

- The READ role: This role gives you read-only access to Collibra Console.
- The ADMIN role: This role includes extra functionality on top of the READ role, for example restoring a backup.
- The SUPER role: This role includes all permissions, for example stopping an environment.

The READ role

Collibra Console users with the READ role have the following permissions:

- To change the email address at which they receive notifications.
- To change their own password.
- To view all the settings in Collibra Console.
- To view the configuration values that an ADMIN user can modify.

The ADMIN role

Next to the permissions of a user with the READ role, a Collibra Console user with the ADMIN role has the following permissions:

- To start and stop an environment and its services.
- To create, delete, download, upload and restore a backup.
- To restore Collibra to factory defaults.
- To create and download a diagnostics package.
- To manage general Collibra settings, but not infrastructure-specific settings, such as the configuration of a mail server.
- To upload a license.

The SUPER role

> Tip The SUPER role is not available in Collibra Console of Collibra Data Intelligence Cloud.The actions mentioned in this section can only be carried out by Collibra. For such actions, create a support ticket.

Next to the permissions of a user with the ADMIN role, a Collibra Console user with the SUPER role has the following permissions:

- Create and change an environment.
- Create, edit and delete Collibra nodes.
- Manage (edit) the infrastructure configuration, for example:
  - Configure the mail server connection.
  - Configure Collibra Connect.
  - Edit the base URL of Collibra.
  - Enable or disable guest access.
- Upgrade Collibra:
  - Migrate a repository.
  - Upgrade a repository.

Edit the Collibra Console settings

To edit the Collibra Console settings, follow these steps:

1. Open Collibra Console.
2. In the main menu, click **Console settings**.
3. In the left navigation pane, click **Configuration**.
4. Click **Edit configuration**.
5. Edit the settings.
6. Click the green **Save all** button.

Tip

- You can navigate to a specific section by clicking it in the tab pane.
- When you edit certain fields, the ⓘ icon is displayed next to it. When you click it, it displays the default value for that field and a **Reset** button.
- If you have to restart Collibra or execute extra actions to apply the new settings, it is indicated in the user interface.

Collibra Console configuration

Note

- Settings that are indicated with an asterisk (*) are mandatory settings.
- To edit the Console configuration, you need the ADMIN role, except for settings or sections marked with (**), which require the SUPER role.

Configuration

Setting	Description
Base URL* (**)	The base URL to access Collibra Console.
Startup timeout for Jobserver	The duration in milliseconds before the Jobserver service startup is considered as a failure. The default value is 300,000 milliseconds or 5 minutes.
Startup timeout for the Repository	The duration in milliseconds before the Repository service startup is considered as a failure. The default value is 300,000 milliseconds or 5 minutes.

Setting	Description
X-Frame options	The option to specify the content of the HTTP-header X-Frame-Options. It is set on all the rendered pages and should be used to avoid a click-jacking attack. By default, we specify that only pages from the same origin can use the rendered pages in a frame. > Note Changing this options requires a reboot of the server.

Email configuration (**)

The configuration of email notifications. More information: Configure Collibra Console email settings.

Setting	Description
Username	The username used to sign in to the SMTP server.
Password	The password used to sign in to the SMTP server.
From address	The email address used as the sender of all outgoing emails.
Host	The URL or IP address of the SMTP server.
Port	The port used to access the SMTP server. The default port is *25*.
Start TLS	• ✓ True: The insecure connections to the SMTP server will be upgraded to a secure connection using SSL or TLS. • ✕ False: The connection to the SMTP server does not use SSL or TLS.

LDAP configuration

The configuration of LDAP to handle the authentication. For a tutorial about the LDAP configuration of Collibra Console, see Collibra Support Portal.

Setting	Description
Enable LDAP integration *	Global setting to enable or disable LDAP integration.
Enable LDAP SUPER users (**)	Allow LDAP users to have the Collibra Console SUPER role. The default value is , LDAP users will have the ADMIN role.
Time limit *	Maximum time in milliseconds for LDAP search operations. The default value is 120,000 milliseconds or two minutes. Set it to *0* if you do not want a time limit.
Servers	This section configures one or more LDAP servers.
LDAP server URL	When using SSL, use the LDAP's protocol and use the correct port. The SSL section in the configuration should also be configured for this.
Bind DN	The main LDAP Domain Name (DN) to be able to connect to the LDAP server. This DN should be able to access all the users and groups that you want to sync.
Bind password	The password for the main LDAP DN to connect to the LDAP server.
Base DN	General base path to which all DNs are relative. As a general rule, this should stay empty for most configurations.

Setting	Description
User base	The base DN where the users are located. Subtree search is used, so all DNs located below the base are searched for matching users.
LDAP user authentication filter	User filter to which users authenticating in Collibra Data Intelligence Cloud should comply.
LDAP user synchronization filter	User filter to which users imported by the synchronization job should comply. This should be a subset of or equal to the Authentication user LDAP filter. When empty, the same filter will be used. as specified in the Authentication user LDAP filter.
Authentication type	LDAP authentication mechanism; must be one of the following: • *none*: No authentication is performed. • *simple*: Simple authentication is performed, using the Bind DN and Bind password as credentials. The credentials are sent as plain text. • *DIGEST-MD5*: Simple authentication is performed using the Bind DN and Bind password as credentials. The Bind password is hashed with the MD5 algorithm. • *TLS-SIMPLE*: A temporary secured TLS connection is set up before the credentials are sent as plain text. SSL must be configured. • *TLS-EXTERNAL*: A temporary secured TLS connection with external SASL authentication using a client certificate. SSL must be configured. > Note When using TLS, dont forget to configure the SSL security section.
Shutdown gracefully	In case of TLS usage, if set to true, Collibra Console tries to shutdown all TLS connections.

Setting	Description
Referral setting	Specifies how referrals should be handled; must be one of: • *throw* • *ignore* • *follow*
Group base DN	The base DN where all groups are located.
Membership attribute	The attribute to define that users are member of a group, for example uniqueMember. This field is mandatory when mapping LDAP groups to roles.
Group LDAP filter	The LDAP filter to which each group should comply to be synced.
Map the DN of an LDAP group to a role	The mapping of the LDAP users that are members of the group with a specified DN to a role.
User field mapping	Configuration mapping of all the user fields. Here you can configure to what LDAP field a given user field should be mapped. Leave empty if it should not be considered in the synchronization.
User Name *	Username mapping is mandatory for LDAP to be enabled.
Role	Role. Defaults to **READ** user if the role is not recognized.
Email *	Email mapping is mandatory for LDAP to be enabled.
Default role	The default role for LDAP users. The default role is READ. Only a Collibra Console SUPER user can grant the SUPER role to an LDAP user.

UI configuration (**)

The configuration of user interface features.

Setting	Description
Page definition overrides	Apply a new page style.
Optimize CSS	Optimize the CSS files to improve performance of the user interface.
Optimize Javascript	Optimize JavaScript code to improve performance of the user interface.
Modules JSON overrides	
Concatenate Javascript	
Modules properties overrides	

Backup configuration

Setting	Description
Backup process timeout*	The duration in milliseconds before the creation of a backup is considered as a failure. The default value is 43,200,000 milliseconds or 12 hours. During an upgrade, this value is also taken into account.

Setting	Description
Backup cleanup interval (Requires restart) (**)	Interval in minutes for the cleanup of backups. The default value is 720 minutes.

1.4.1 Backup schedule configuration

Setting	Description
Default max retained*	The default maximum number of retained backups related to a backup schedule. 0 signifies retain all.

CSRF configuration

Setting	Description
Enable CSRF protection	If enabled, Collibra Console checks the validity of the request using a CSRF token.

Session configuration

Setting	Description
Idle Session timeout (Requires restart)*	The duration in seconds of idle time before a Collibra Console session times out. The minimum value is 15 minutes, the maximum is 24 hours. The default value is 1,800 seconds or 30 minutes.

Security Configuration (**)

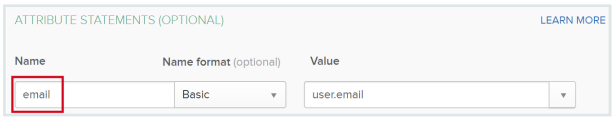
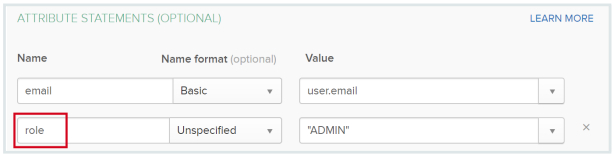
Setting	Description
User locked out interval (Requires restart)*	The number of seconds during which a user cannot sign in to Collibra Console after the specified number of incorrect **Login attempts**.
Login attempts (Requires restart)*	The number of times that a user can try to sign in to Collibra Console before being locked out for the time defined in **User locked out interval**.
Interval for consecutive login attempts (Requires restart)*	The number of seconds during which the user has the amount of login attempts specified in the **Login attempts** field.

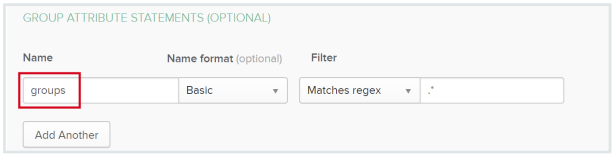
1.7.1 SSO

Setting	Description
Mode	The SSO mode of Collibra Console. Collibra Console only supports SAML_ATTRIBUTES.
Disable the Collibra Management console signin page	When SSO is enabled, a user can still navigate to the /signin page and try to log in via that page. However, you can disable that page. • ✓ True: Users cannot access the Collibra Console signin page. • ✕ False (default): Users can access the Collibra Console signin page

Setting	Description
Default role	The default role for all SSO users. The default role is READ. SSO users with a SUPER role will be downgraded to users with the ADMIN role. Only a Collibra Console SUPER user can enable the SUPER role for SSO users.
Enable SSO SUPER users(**)	• ✔ True: SSO users can have the SUPER role. • ✖ False (default): SSO users with the SUPER role will be downgraded to users with the ADMIN role. This option is only visible if you have the SUPER role.
SAML	The configuration of SAML.
Metadata HTTP	The URL of the SAML metadata file to be used. The URL always has to be reachable by the Collibra environment.
Entity ID	The entity ID inside the metadata to be referenced. Note A metadata file can describe multiple entity IDs, make sure to use in the entity ID from the correct metadata file.
Consumer service URL	By default, this URL is the same as the URL of your Collibra environment but if your IDP expects another value, you can fill it out here. Warning Make sure that the intended destination endpoint (The Destination attribute in the SAML response) matches the URL being used here. So this is only to be used in specific IDP circumstances When setting this, and getting the error "SAML message intended destination endpoint did not match recipient endpoint" check the Destination attribute in your SAML response and this parameter.

Setting	Description
Disable client address	• ✓ True: The validation of the client IP address in the assertion message is disabled. • ✕ False (default): The validation of the client IP address in the assertion message is enabled.
Sign authentication requests (Requires restart)	• ✓ True: Authentication requests have to be signed. Use a Collibra generated self-signed certificate to sign requests. The request that is generated by Collibra Console will be appended with a signature in the redirect URL of the response. • ✕ False (default): Authentication request don't have to be signed.
Force authn	• ✓ True (default): The SP authentication request forces re-authentication. • ✕ False: The SP authentication request does not force re-authentication.
Force passive	Configure whether the SP authentication should set the authentication to go passive. If True, the IDP or browser MUST NOT take visibly control of the user interface. See the SAML 2.0 specification for more details. This is only relevant if **Force authn** is *True*.
Name ID	Configure the nameID used in the SP authentication. If set, the full content will be sent as a nameID. Use a fully qualified nameID. Default nameID="urn:oasis:names:tc:SAML:2.0:nameid-format:persistent". See SAML 2.0 specification for accepted nameID values.
Group to Role mapping	Map groups from the group field to a role in the Collibra Console. The **Field key** must be a name of a group as defined in your IDP, the **Field value** is one of the Collibra Console roles (READ, ADMIN, SUPER).

Setting	Description
Attribute fields	The mappings of attributes in the SAML response. The values are used as keys to look for in the SAML response.
Username	The mapping for the user's username. The value of this attribute must be the name of the SAML attribute as defined in your IDP. In the following example, the value in Collibra Console must be *email*. This is the used name in your IDP software, see the following screenshot.  This field is mandatory.
Role	The mapping for the user roles. The value of this attribute must be the name of the SAML attribute as defined in your IDP. In the following example, the value in Collibra Console must be *role*. This is the used name in your IDP software, see the following screenshot. 

Setting	Description
Group	The mapping to define which attribute holds group information. The value of this attribute must be the name of the SAML attribute as defined in your IDP. In the following example, the value in Collibra Console must be *groups*. This is the used name in your IDP software, see the following screenshot.  If there is more than one group attribute statement, add them as comma-separated list.
SAML Requested authentication context	Settings for the SAML requested authentication context. The authentication context is the way in which the IDP authenticates the user. By default the authentication context will mandate user/password authentication over HTTPS.
Disable	Disable the SAML requested authentication context. Set to True if you wish to configure the IDP to use specific authentication contexts, without the need to send one in the request.

Setting	Description
Comparison type	Specifies the comparison method used to evaluate the requested authentication context. One of: "exact", "minimum", "maximum", "better". The industry default is "exact", other options are "minimum", "maximum" and "better". • **Exact**: The authn context in the assertion MUST exactly match the full expected context specified on the SP. • **Minimum**: The authn context in the assertion MUST be at least as strong as one of the contexts specified on the SP. • **Better**: The authn context in the assertion MUST be stronger than any of the contexts specified on the SP. • **Maximum**: The authn context in the assertion MUST be as strong as possible for all of the contexts specified on the SP, without exceeding the strength of at least one context. For more details, see the SAML 2.0 specification (Section 3.3.2.2.1).
Reference list	All SAML authentication classes to be sent in the SAML authentication request. Use to tune the authentication context on the IDP side. Default-="urn:oa-asis:names:tc:SAML:2.0:ac:classes:PasswordProtectedTransport" (Which means user/password over HTTPS). For more details, see the SAML 2.0 specification (Section 3.3.2.2.1 & Section 2.7.2.2).
Declaration list	All SAML authentication declarative classes to be sent in the SAML authentication request. For details, see the SAML 2.0 specification (Section 3.3.2.2.1 & Section 2.7.2.2).
Signout	The sign-out redirection settings.
Override signout URL	Enable this option to redirect to another page than the default sign-out page. The default page when you sign out is the Collibra Console sign-in page.

Setting	Description
Signout redirect URL	The URL to redirect to when you sign out.

1.7.2 HTTP headers

The configuration of the HTTP response headers of Collibra Console.

Field	Description
URL pattern	The pattern of the URLs to which the HTTP response header is applied. This field supports wildcards such as `**`, `*` and `?`. Tip This pattern matches all URLs: `/**`.
HTTP headers	The HTTP response headers in a key-value format. You can add new HTTP response headers by clicking **Add** at the bottom of the section, and entering the HTTP response header name as the field key and the HTTP response header value as the field value.

Set the context path of Collibra Console

The context path allows you to serve the Console service on a specific URL. You can set this in the installation wizard, but also after the installation in Collibra Console.

Steps

1. Open Collibra Console with a user profile that has the **SUPER** role.
   » Collibra Console opens with the **Infrastructure** page.

2. In the main menu, click **Console settings**.
3. In the tab pane, click **Application server configuration**.
4. Click **Edit configuration**.
5. In the **Context path** field, enter the name of your context path.

Setting	Description
Context path (Requires restart) *	The path that is added to the base URL to reach Collibra Console. For example, if your base URL is `https://dgc.yourcompany.com:4402/` and your context path is `console-acceptance`, then your path to reach Collibra Console is `https://dgc.yourcompany.com:4402/console-acceptance`.

6. Click the green **Save all** button.
7. Sign out from Collibra Console.
8. Restart the DGC service:
   - Collibra DGC services on Linux
   - Collibra DGC services on Linux as non-root user
   - Collibra DGC services on Windows

What's next?

Connect to Collibra Console using the new URL, which is the base URL with the context path.

Configure Collibra Console email settings

As for Collibra Data Intelligence Cloud, you can set email settings for Collibra Console to enable mails for Collibra Console user activation or password recovery links.

You have to configure the Collibra Console email settings via the Collibra Console user interface.

Note
- The Collibra email settings do not apply to the Collibra Console email settings.
- This section only applies to Collibra Console 5.2 or newer.

Prerequisites

You have the SUPER role in Collibra Console.

Steps

1. In the main menu, click **Console settings**.
2. In the tab pane, click **Configuration**.
3. In the middle column, click **Mail configuration**.
4. Above the table, to the right, click **Edit configuration**.
5. Enter the relevant data:

Setting	Description
Username	The username used to sign in to the SMTP server.
Password	The password used to sign in to the SMTP server.
From address	The email address used as the sender of all outgoing emails.
Host	The URL or IP address of the SMTP server.
Port	The port used to access the SMTP server. The default port is *25*.
Start TLS	◦ ✓ True: The insecure connections to the SMTP server will be upgraded to a secure connection using SSL or TLS. ◦ ✕ False: The connection to the SMTP server does not use SSL or TLS.

6. Click **Save all**.

Configure SSL to access Collibra Console

If you want to connect to Collibra Console in a secure way with your web browser, you have to use SSL. This procedure explains how you can activate SSL access to Collibra Console.

Prerequisites

- You have knowledge of the JSON syntax.
- You have created a Java KeyStore according the procedure described by Oracle, for example **clientkeystore**.
- You have noted the following data while creating the Java KeyStore:
  - KeyStore file name: *clientkeystore* in the Oracle example.
  - KeyStore alias: *client* in the Oracle example.
  - KeyStore password: The password that you entered after executing the command of the first step in the Oracle example.
  - KeyStore alias password: The password that you entered as last step of step 2 in the Oracle example.
- You have stored the Java KeyStore on the node hosting Collibra Console, in the **<collibra_data>/console/security** folder, for example **/opt/collibra_data/console/security**.

Steps

1. Open a terminal session on the node on which Collibra Console is installed.
2. Open the file **<collibra_data>/console/config/server.json** for editing.
3. Fill in the following parameters in the **httpsConnector** section, if the section does not exist, you have to create it manually:

   > Note Add string values between double quotes.

Parameter	Description
port	The port on which the HTTPS connector must bind. The value must be higher than 1024 to avoid root permissions. Note If you want to use the default SSL port 443, you have to use a reverse proxy.
keyAlias	The KeyStore alias.
keyPass	The KeyStore alias password.
keystorePass	The KeyStore password.
keystoreFile	The full path to the KeyStore file name, for example **/opt/collibra_data/console/security/clientkeystore**.

Example:

```
"httpsConnector" : {
    "port": 5404,
    "keyAlias": "your-alias",
    "keyPass": "your-password",
    "keystorePass": "your-password",
    "keystoreFile": "/opt/collibra_
data/console/security/console.jks" },
}
```

4. Save and close the file.
5. Open the DGC service settings for editing:
6. Click the **General settings** section.
7. Update the **Base URL** parameter with *https*and the new port.
8. Restart the environment.

What's next?

Connect to your Collibra Console instance via the Base URL.

> Tip To prevent regular HTTP traffic to Collibra Console, update the **address** parameter with the value *127.0.0.1* in **<collibra_data>/console/config/server.json** and restart the environment.
> For more information, see the knowledge base on the Collibra Support Portal.

Collibra Console SSO configuration

You can access Collibra Console with user accounts that are created within Collibra Console. Besides those user accounts, you can also configure single sign-on (SSO) access to enable integration with your SSO infrastructure.

> Tip If you want to use a custom certificate in the SSO configuration for Collibra Console access, see this section.

Prerequisites

- Your identity provider (IDP) supports SAML 2.0.

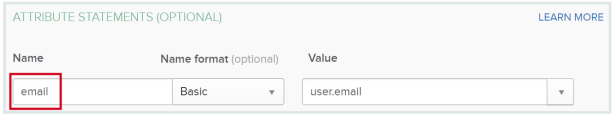
Steps

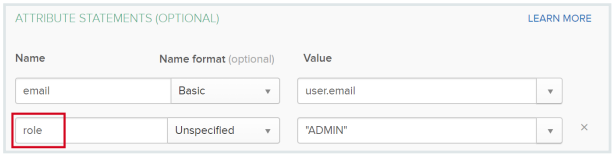
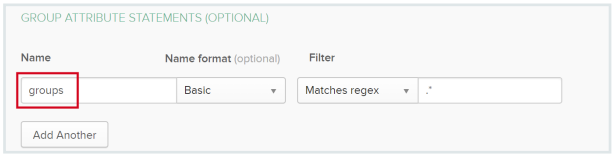
1. Open Collibra Console with a user profile that has at least the **ADMIN** role.
   » Collibra Console opens with the **Infrastructure** page.
2. In the main menu, click **Console settings**.
3. In the tab pane, click **Configuration**.
4. In the middle column, click **Security Configuration**.

5. Click **Edit configuration**.
6. Enter the required information:

Setting	Description
Mode	The SSO mode of Collibra Console. Collibra Console only supports SAML_ATTRIBUTES.
Disable the Collibra Management console signin page	When SSO is enabled, a user can still navigate to the /signin page and try to log in via that page. However, you can disable that page. ◦ ✔ True: Users cannot access the Collibra Console signin page. ◦ ✖ False (default): Users can access the Collibra Console signin page
Default role	The default role for all SSO users. The default role is READ. SSO users with a SUPER role will be downgraded to users with the ADMIN role. Only a Collibra Console SUPER user can enable the SUPER role for SSO users.
Enable SSO SUPER users(**)	◦ ✔ True: SSO users can have the SUPER role. ◦ ✖ False (default): SSO users with the SUPER role will be downgraded to users with the ADMIN role. This option is only visible if you have the SUPER role.
SAML	The configuration of SAML.
Metadata HTTP	The URL of the SAML metadata file to be used. The URL always has to be reachable by the Collibra environment.
Entity ID	The entity ID inside the metadata to be referenced. Note A metadata file can describe multiple entity IDs, make sure to use in the entity ID from the correct metadata file.

Setting	Description
Consumer service URL	By default, this URL is the same as the URL of your Collibra environment but if your IDP expects another value, you can fill it out here. Warning Make sure that the intended destination endpoint (The Destination attribute in the SAML response) matches the URL being used here. So this is only to be used in specific IDP circumstances When setting this, and getting the error "SAML message intended destination endpoint did not match recipient endpoint" check the Destination attribute in your SAML response and this parameter.
Disable client address	◦ ✔ True: The validation of the client IP address in the assertion message is disabled. ◦ ✖ False (default): The validation of the client IP address in the assertion message is enabled.
Sign authentication requests (Requires restart)	◦ ✔ True: Authentication requests have to be signed. Use a Collibra generated self-signed certificate to sign requests. The request that is generated by Collibra Console will be appended with a signature in the redirect URL of the response. ◦ ✖ False (default): Authentication request don't have to be signed.
Force authn	◦ ✔ True (default): The SP authentication request forces re-authentication. ◦ ✖ False: The SP authentication request does not force re-authentication.

Setting	Description
Force passive	Configure whether the SP authentication should set the authentication to go passive. If True, the IDP or browser MUST NOT take visibly control of the user interface. See the SAML 2.0 specification for more details. This is only relevant if **Force authn** is *True*.
Name ID	Configure the nameID used in the SP authentication. If set, the full content will be sent as a nameID. Use a fully qualified nameID. Default nameID="urn:oasis:names:tc:SAML:2.0:nameid-format:persistent". See SAML 2.0 specification for accepted nameID values.
Group to Role mapping	Map groups from the group field to a role in the Collibra Console. The **Field key** must be a name of a group as defined in your IDP, the **Field value** is one of the Collibra Console roles (READ, ADMIN, SUPER).
Attribute fields	The mappings of attributes in the SAML response. The values are used as keys to look for in the SAML response.
Username	The mapping for the user's username. The value of this attribute must be the name of the SAML attribute as defined in your IDP. In the following example, the value in Collibra Console must be *email*. This is the used name in your IDP software, see the following screenshot.  This field is mandatory.

Setting	Description
Role	The mapping for the user roles. The value of this attribute must be the name of the SAML attribute as defined in your IDP. In the following example, the value in Collibra Console must be *role*. This is the used name in your IDP software, see the following screenshot. 
Group	The mapping to define which attribute holds group information. The value of this attribute must be the name of the SAML attribute as defined in your IDP. In the following example, the value in Collibra Console must be *groups*. This is the used name in your IDP software, see the following screenshot.  If there is more than one group attribute statement, add them as comma-separated list.
SAML Requested authentication context	Settings for the SAML requested authentication context. The authentication context is the way in which the IDP authenticates the user. By default the authentication context will mandate user/password authentication over HTTPS.

Setting	Description
Disable	Disable the SAML requested authentication context. Set to True if you wish to configure the IDP to use specific authentication contexts, without the need to send one in the request.
Comparison type	Specifies the comparison method used to evaluate the requested authentication context. One of: "exact", "minimum", "maximum", "better". The industry default is "exact", other options are "minimum", "maximum" and "better". ◦ **Exact**: The authn context in the assertion MUST exactly match the full expected context specified on the SP. ◦ **Minimum**: The authn context in the assertion MUST be at least as strong as one of the contexts specified on the SP. ◦ **Better**: The authn context in the assertion MUST be stronger than any of the contexts specified on the SP. ◦ **Maximum**: The authn context in the assertion MUST be as strong as possible for all of the contexts specified on the SP, without exceeding the strength of at least one context. For more details, see the SAML 2.0 specification (Section 3.3.2.2.1).
Reference list	All SAML authentication classes to be sent in the SAML authentication request. Use to tune the authentication context on the IDP side. Default- ="urn:oasis:names:tc:SAML:2.0:ac:classes:PasswordProtectedTransport" (Which means user/password over HTTPS). For more details, see the SAML 2.0 specification (Section 3.3.2.2.1 & Section 2.7.2.2).
Declaration list	All SAML authentication declarative classes to be sent in the SAML authentication request. For details, see the SAML 2.0 specification (Section 3.3.2.2.1 & Section 2.7.2.2).

Setting	Description
Signout	The sign-out redirection settings.
Override signout URL	Enable this option to redirect to another page than the default sign-out page. The default page when you sign out is the Collibra Console sign-in page.
Signout redirect URL	The URL to redirect to when you sign out.

7. Click **Save all**.

Configure a custom certificate for SSO in Collibra Console

If you configure single sign-on for accessing Collibra Console, a default certificate is used. You can use this certificate for signing SAML authn requests.

Instead of the default certificate, you can use your own certificate. However, keep in mind that you can only configure SSO with your own certificate via a REST API call.

Prerequisites

- The certificate must meet the following requirements:
  - The certificate must be in PEM format.
  - The PEM file must be unencrypted (no password).
  - The PEM file must contain the server certificate the private key of that certificate.

> Tip To convert a key to a PEM key: `openssl rsa -in <pem-key>.key -out <rsa-key>.pem`

Example PEM file:

```
-----BEGIN PRIVATE KEY-----
MIIEvgIBADA ... bml6YXRpb252YWxza
....
 z3P668YfhUbKdRF6S42Cg6zn
-----END PRIVATE KEY-----

# Your certificate
 -----BEGIN CERTIFICATE-----
MIIFaDCCBFC ... bml6YXRpb252YWxza
...

lffygD5IymCSuuDim4qB/9bh7oi37heJ4ObpBIzroPUOthbG4gv/5blW3Dc
=
-----END CERTIFICATE-----
```

- A base64 encoding hash of an API user.

  Tip You can calculate the base64 hash of the user as follows: `echo '<username>:<password>' | base64`, for example `echo 'apiUser:apiUserpassword' | base64` results in `YXBpVXNlcjphcGlVc2VycGFzc3dvcmQK`

Steps

```
curl --location --request POST \
  https://<your-collibra-console-url>/rest/sam-
lconsole/certificate \
  --header 'Authorization: Basic <base-64 encoding hash>'
  --form 'file=@"/path/to/pem-file"'
```

Edit the Collibra Console server settings

To edit the Collibra Console server settings, follow these steps:

1. Open Collibra Console.
2. In the main menu, click **Console settings**.

3. In the left navigation pane, click **Application server configuration**.
4. Click **Edit configuration**.
5. Edit the Collibra Console - Application server configuration.
6. Click the green **Save all** button.

Tip
- You can navigate to a specific section by clicking it in the tab pane.
- When you edit certain fields, the ⓘ icon is displayed next to it. When you click it, it displays the default value for that field and a **Reset** button.
- If you have to restart Collibra or execute extra actions to apply the new settings, it is indicated in the user interface.

Collibra Console - Application server configuration

Application server configuration

Setting	Description
Context path (Requires restart) *	The path that is added to the base URL to reach Collibra Data Intelligence Cloud. For example, if your base URL is `https://dgc.yourcompany.com:4400/` and your context path is `acceptance`, then your path to reach Collibra is `https://dgc.yourcompany.com:4400/acceptance`.

HTTP connector

A connector supporting the HTTP/1.1 protocol.

Setting	Description
URI encoding (Requires restart) *	The character encoding used to decode the URI bytes, URL-decoding has been performed.
Acceptor thread count (Requires restart) *	The number of threads to be used to accept connections. The actual request processing is done by separate threads, so you would never really need more than two.
Min spare threads (Requires restart) *	The minimum number of threads always kept running.
Max threads (Requires restart) *	The maximum number of request processing threads to be created by this connector. This determines the maximum number of simultaneous requests that can be handled.
Accept count (Requires restart) *	The maximum queue length for incoming connection requests that can be assigned to a request processing thread. When the queue is full, requests will be refused.
Compression (Requires restart) *	The connector may use HTTP/1.1 GZIP compression. • **off**: No compression • **on**: Allow compression • **force**: Allow compress
Compression min size (Requires restart) *	The minimum amount of data before it is compressed. This setting only has effect if **Compression** is **on**.

Setting	Description
Compressible MIME type (Requires restart)	A comma separated list of MIME types allowing HTTP compression.
Connection timeout (Requires restart) *	The number of milliseconds that this connector waits for the destination URI to be presented by the request, after accepting a connection.
Port (Requires restart) *	The TCP port to access your Collibra environment via your web browser.

HTTPS connector

A connector supporting the HTTP/1.1 protocol with SSL support enabled.

Setting	Description
URI encoding (Requires restart) *	The character encoding used to decode the URI bytes, URL-decoding has been performed.
Acceptor thread count (Requires restart) *	The number of threads to be used to accept connections. The actual request processing is done by separate threads, so you would never really need more than two.
Min spare threads (Requires restart) *	The minimum number of threads always kept running.

Setting	Description
Max threads (Requires restart) *	The maximum number of request processing threads to be created by this connector. This determines the maximum number of simultaneous requests that can be handled.
Accept count (Requires restart) *	The maximum queue length for incoming connection requests that can be assigned to a request processing thread. When the queue is full, requests will be refused.
Compression (Requires restart) *	The connector may use HTTP/1.1 GZIP compression. • **off**: No compression • **on**: Allow compression • **force**: Allow compress
Compression min size (Requires restart) *	The minimum amount of data before it is compressed. This setting only has effect if **Compression** is **on**.
Compressible MIME type (Requires restart)	A comma separated list of MIME types allowing HTTP compression.
Connection timeout (Requires restart) *	The number of milliseconds that this connector waits for the destination URI to be presented by the request, after accepting a connection.
Port (Requires restart) *	The TCP port to access your Collibra environment via your web browser.

AJP connector

A connector able to communicate with another web connector via the AJP protocol. Mainly for transparent integration with another HTTP server, for example Apache, nginx....

Setting	Description
URI encoding (Requires restart) *	The character encoding used to decode the URI bytes, URL-decoding has been performed.
Acceptor thread count (Requires restart) *	The number of threads to be used to accept connections. The actual request processing is done by separate threads, so you would never really need more than two.
Min spare threads (Requires restart) *	The minimum number of threads always kept running.
Max threads (Requires restart) *	The maximum number of request processing threads to be created by this connector. This determines the maximum number of simultaneous requests that can be handled.
Accept count (Requires restart) *	The maximum queue length for incoming connection requests that can be assigned to a request processing thread. When the queue is full, requests will be refused.
Compression (Requires restart) *	The connector may use HTTP/1.1 GZIP compression. • **off**: No compression • **on**: Allow compression • **force**: Allow compress

Setting	Description
Compression min size (Requires restart) *	The minimum amount of data before it is compressed. This setting only has effect if **Compression** is **on**.
Compressible MIME type (Requires restart)	A comma separated list of MIME types allowing HTTP compression.
Connection timeout (Requires restart) *	The number of milliseconds that this connector waits for the destination URI to be presented by the request, after accepting a connection.
Port (Requires restart) *	The TCP port to access your Collibra environment via your web browser.

Static Resources

The static resources are the resources that are reserved for the service.

Setting	Description
Maximum cache size (Requires restart) *	The maximum size that can be assigned to the cache of the service, expressed in kilobytes.

Open a Collibra Console log file

To open the log files of Collibra Console, follow these steps:

1. In the main menu, click **Console settings**.
2. In the tab pane, click **Logs**.
3. Click the name of a log file to open it.

Chapter 9

Troubleshooting

DGC service fails to start due to invalid configuration 262

Finding resource IDs 262

DGC service fails to start due to invalid configuration

If Collibra detects an invalid configuration of the Data Governance Center service, it no longer automatically replaces the invalid configuration by the default configuration and the service does not start.

This situation can happen when you edit a configuration in a backup, introducing invalid data, and then you restore that backup.

In **dgc.log**, you can find a Collibra exception "configurationParsingFailed" and "DGCConfigurationServiceImpl.readDGCConfiguration (DGCConfigurationServiceImpl.java...)":

```
Caused by: com.collibra.common.exception.CollibraException: con-
figurationParsingFailed
Message: com.fasterxml.jackson.databind.JsonMappingException:
...


    at com.collibra.dgc.configuration.service.DGCCon-
figurationServiceImpl.readDGCConfiguration(DGCCon-
figurationServiceImpl.java:362)
```

Resolution

Revert the configuration changes in your backup.

Finding resource IDs

Every single resource in Collibra Data Intelligence Cloud has a unique ID. Sometimes it happens that you need this unique resource ID in procedures.

In this section, you can find some examples on how to do this.

Find the UUID of a metamodel element

It is possible that you need the UUID (Universally Unique Identifier) of an asset, domain or community, for example to use it in API calls.

You can find this UUID by clicking an asset, domain or community in Collibra Data Intelligence Cloud. The UUID of the selected resource is then shown in the address bar of your browser.

The URL looks like https://<yourdgcinstance>/<resource type>/bc40c085-352c-4a8c-8ee7-494fe821308e.

> Tip UUIDs that start with 00000000 are packaged resources.

Examples

- Community: https://<yourdgcinstance>/community/bc40c085-352c-4a8c-8ee7-494fe821308e
- Domain: https://<yourdgcinstance>/domain/00000000-0000-0000-0000-000000006019
- Asset: https://<yourdgcinstance>/asset/00000000-0000-0000-0000-000000008044

Find the resource ID of an asset type

You can find the resource ID of an asset type. This may be useful for certain operations, for example for API calls.

Prerequisites

You have a global role that has the System administration global permission.

Steps

1. In the main menu, click , then **Settings**.
   » The Collibra settings page opens.

2. Click **Operating Model**.
   » The operating model settings appear on the **Asset types** tab page.
3. In the content toolbar, click ⠿ → **Columns** → **Resource Id**.
   » The **Resource Id** column appears.

> Tip You can also find the resource ID by clicking the asset type. Then look in the URL of your browser to find the resource ID.
> The URL looks `like https://<yourdgcinstance>/assettype/00000000-0000-0000-0000-000000031101`.
> The resource ID of the selected asset type is `00000000-0000-0000-0000-000000031101`, in this example *Business Asset*.

Find the resource ID of a relation type

You can find the resource ID of a relation type. This may be useful for certain operations.

Prerequisites

You have a global role that has the System administration global permission.

Steps

1. In the main menu, click ⠿, then ⚙ **Settings**.
   » The Collibra settings page opens.
2. Click **Operating Model**.
   » The operating model settings appear on the **Asset types** tab page.
3. In the tab pane, click **Characteristics** → **Relations**.
4. If you don't see the **Resource Id** column, add it to the table.
   a. In the content toolbar, click ⠿ → **Columns** → **Resource Id**.
      » The **Resource Id** column appears.
   » The resource ID of the relation types appear in the **Resource Id** column of the table.

Find the resource ID of a complex relation type

You can find the resource ID of a complex relation type. This may be useful for certain operations.

Prerequisites

You have a global role that has the System administration global permission.

Steps

1. In the main menu, click , then **Settings**.
   » The Collibra settings page opens.
2. Click **Operating Model**.
   » The operating model settings appear on the **Asset types** tab page.
3. In the tab pane, click **Characteristics** → **Complex Relations**.
4. If you don't see the **Resource Id** column, add it to the table.
   a. In the content toolbar, click → **Columns** → **Resource Id**.
      » The **Resource Id** column appears.
   » The resource ID of the complex relation types appear in the **Resource Id** column of the table.

Find the resource ID of an attribute type

In some actions, you need the resource ID of an attribute type, for example, to edit search a boost factor.

Steps

1. In the main menu, click , then **Settings**.
   » The Collibra settings page opens.
2. Click **Operating Model**.
   » The operating model settings appear on the **Asset types** tab page.
3. In the tab pane, click **Characteristics** → **Attributes**.
   » The table with attribute types appears.

4. In the content toolbar, click ⠿ → **Columns** → **Resource Id**.
   » The **Resource Id** column appears.

DGC service configuration: options

> Tip
> To edit the DGC service configuration, you need the ADMIN role, except for settings or sections marked with (*), which require the SUPER role. As the SUPER role does not exist for cloud environments, settings that require the SUPER role are not available in Collibra Data Intelligence Cloud environments. If you want to edit one of these settings, please create a support ticket.

General settings

The general settings of Collibra Data Intelligence Cloud.

Setting	Description
Default locale (Requires restart)	The default locale for new users. It has to contain a language code and may contain a country code. • The language has to be an ISO language code. • The country has to be an ISO country code. Examples: *pl*, *en_US*, *nl_BE*.
Enable view rights	• ✔ True (default): The view permissions feature is enabled. • ✘ False: The view permissions feature is disabled.
Base URL (*)	The base URL for this Collibra instance, for example http://dgc.example.com. It is the consistent part of the URL to access Collibra Data Intelligence Cloud. This is used amongst others to: • construct hyperlinks to the system, for example in emails. • display your profile picture. • display Tableau report images. • ...
Enable auditing (Requires restart) (*)	• ✔ True (default): Audit and history information is stored. • ✘ False: Audit and history information is not stored.

Setting	Description
Google Analytics tracking ID (*)	The Google Analytics Tracking ID to which Collibra sends data about the page visits.
Show target asset type above relation table	• ✔ True (default): Show the asset type of the target asset in the title of relation tables on an asset page. The target asset can be either the head or the tail of the relation, depending on which asset page you have open. • ✘ False: Hide the asset type of the target asset. The default value is `true`.
Enable Usage analytics	The analytics are to understand how users interact with Collibra. The information will be used to provide historical reporting. When you enable the analytics in a Cloud environment, the data collection starts immediately. When you enable it in an on-premises environment, you have to whitelist the tracking script (pendo.io, app.pendo.io and cdn.pendo.io). • ✔ True: Gathering analytics is enabled and sent to Collibra. • ✘ False (default): Gathering analytics is disabled.
Usage analytics API key	The Pendo analytics API key that you need when you have enabled the usage analytics.
Refreshed Navigation	• ✔ True: Use the new, refreshed navigation of the Collibra user interface. • ✘ False (default): Use the classic navigation of the Collibra user interface.
Add Asset Grid link to menu	• ✔ True: Add the link to Asset Grid in the main menu of your environment. • ✘ False (default): The link to Asset Grid is not available in the main menu of your environment.

Help Menu

The configuration of the Help menu in Collibra Data Intelligence Cloud.

Setting	Description
Links	The list of links in the help menu.

Setting	Description
Menu item name	The name of the menu item as it will appear in Collibra Data Intelligence Cloud's help menu.
Menu index	The position of the menu item in the help menu. The top position starts with the value 1.
Menu URL	The target URL of the menu item.
Show admin only	• ✓ True: The menu item is only visible to users with the Sysadmin role. • ✕ False: The menu item is visible to every user.

Email configuration

The configuration of email notifications.

> Note In a Collibra Data Intelligence Cloud environment, you cannot update the email server settings, such as host and port. For more information, see Collibra Data Intelligence Cloud infrastructure.

Setting	Description
Default schedule (Requires restart)	The Cron schedule to send emails only at specific times. With this, you can send emails in batches and avoid an overload of mails. Keep in mind that these emails are only workflow emails and have nothing to do with the notification schedule. If you create an invalid Cron pattern, Collibra Data Intelligence Cloud stops responding.
Template map	The location of template emails.
Password (*)	The password paired with your username to sign in to your SMTP server.
From address (*)	The email address used as the sender of all outgoing emails. Contact Collibra support to change the From address, see also Email configuration.

Setting	Description
Port (*)	The port to connect to your SMTP server. The default value is **25**.
Host (*)	The hostname or URL of your SMTP server.
Start TLS (*)	• ✔ True: Use TLS (Transport Layer Security) to connect to your SMTP server. • ✖ False (default): Do not use TLS to connect to your SMTP server.
Username (*)	The username to sign in to your SMTP server.
Sending threads (*)	The number of threads that are used to send emails. The default value is **3**.
Max retries (*)	The maximum number of retries before the system aborts the sending of an email. The default value is **5**.

> Note
> To edit the DGC service configuration, you need the ADMIN role, except for settings or sections marked with (*), which require the SUPER role. As the SUPER role does not exist for cloud environments, settings that require the SUPER role are not available in Collibra Data Intelligence Cloud environments. If you want to edit one of these settings, please create a support ticket.

Notifications

The configuration of notification emails to users.

> Note These settings can be overridden for every user in the preferences.xml file.

Setting	Description
Notification days	The days of the week on which Collibra sends notifications. The days are represented by numbers from 1 to 7, where 1 represents Sunday. Per row you can add one day.
Daily roles	The roles that receive notifications on the days defined in **Notification days**.

Setting	Description
Enable monthly notifications	• ✔ True: The users receive a monthly summary. • ✘ False (default): The users do not receive a monthly summary.
Roles for monthly notifications	The roles that receive monthly notification emails. This is only relevant if **Enable monthly notifications** is ✔ True.

Handlers (*)

A mail handler can poll for emails on a mail server, process those emails and perform actions based on the contents.

Setting	Description
Host (*)	The hostname or URL of the incoming mail server.
Port (*)	The port to connect to your incoming mail server.
Protocol (*)	The protocol to connect to your incoming mail server, with or without SSL (*POP3*, *POP3S*, *IMAP*, *IMAPS*). > Note The additional S at the end of the abbreviations stands for the secure version of the protocol using SSL. Using this requires the SSL certificates to be correctly configured.
Force domain (*)	• ✔ True: Only handle emails from the same domain as the handler's email address. • ✘ False (default): Handle emails from any domain.
Handler list (*)	The configuration of email handlers, which can poll emails on an email server, process those emails and perform actions based on the contents.
Enabled (*)	• ✔ True: The handler is enabled. • ✘ False (default): The handler is not enabled.
Name (*)	The name of the mail handler. We recommend to use a meaningful name to easily identify what this handler is used for.
Username (*)	The username to connect to the incoming mail server.

Setting	Description
Password (*)	The password to connect to the incoming mail server.
Email address (*)	The email address to which workflow action mails are sent.
Polling interval (*)	The time in milliseconds between two pollings of the mail server.
Delete (*)	• ✔ True: Delete messages from the mail server once the mail is processed. • ✖ False (default): Keep messages on the mail server after the mail is processed. This option is only relevant if **Protocol** is *IMAP* or *IMAPS*.
Alias filter (*)	• ✔ True (default): Retrieve only the emails of which the **To** field contains the email address of the handler. • ✖ False: Do not filter on the **To** field.

Note
To edit the DGC service configuration, you need the ADMIN role, except for settings or sections marked with (*), which require the SUPER role. As the SUPER role does not exist for cloud environments, settings that require the SUPER role are not available in Collibra Data Intelligence Cloud environments. If you want to edit one of these settings, please create a support ticket.

Hyperlinking configuration

The configuration of automatic hyperlinks. When you change a setting, you have to rebuild the hyperlinks.

Setting	Description
Enable hyperlinking	• ✔ True: Hyperlinks are created automatically. • ✖ False (default): Hyperlinks are not created automatically. For more information about automatic hyperlinks, see Hyperlinking. Warning If you enable this setting, the performance of Collibra can decrease.

Setting	Description
Enable case sensitivity	• ✔ True: Hyperlinks are case-sensitive. • ✖ False (default): Hyperlinks are not case-sensitive. Note If you edit this setting, you have to reindex Collibra.
Batch size	The maximum number of text attributes that can be processed in a single transaction. The default value is *50,000*. The higher the value, the more memory used to create the hyperlinks, which happens each time you edit an asset or import data. The lower the value, the more time required for this process. Note To edit this setting, you need the SUPER role.
Excluded asset type IDs	The list of asset types that are ignored by automatic hyperlinking. You can enter multiple asset type IDs, separated by commas. Excluding assets reduces the amount of hyperlinks, which improves performance. Tip We recommend that you exclude technical asset types such as Column, Field, Table, Code Value and Code Set. Note If you edit this setting, you have to reindex Collibra.

Recommender configuration

The configuration of the recommender.

Setting	impacts	Description
Catalog recommender enabled	All recommendations	• ✓ True (default): The "Data sets you might like" section is included on the Data Catalog Home page. This section shows data sets you might be interested in, as determined by the recommender, which takes into account your data sets and the data sets of similar users. • ✕ False: The "Data sets you might like" section is not included on the Data Catalog Home page.
Data set recommender execution time	Recommendations of data sets to users	The schedule (CRON job) by which the data set recommender looks for recommended data sets for a user. By default the data set recommender does this every night.
Asset recommender execution time	Recommendations of business assets to data assets	The schedule (CRON job) by which the asset recommender looks for suggested relations between business assets and data sets.
Data set matcher execution time	Data set matcher	The schedule (CRON job) by which the data set matcher looks for similar data sets.
Data set similarity threshold	Data set matcher	The amount of business assets that have to be related to two data sets before the data sets are considered to be similar. This percentage is expressed by a decimal where 1,00 equals 100%. Example If this value is 0.3 and at least 30% of the related business assets are related to both data sets, they are considered to be similar.
Duplicate schema threshold	Schema matcher	The amount of assets that have to be related to both schemas before the schemas are considered to be similar. This percentage is expressed by a decimal where 1,00 equals 100%.

Setting	impacts	Description
Fuzzy vs exact matching strategy for business assets	Recommendations of business assets to data sets and of business assets to column assets	The percentage that determines to what extent assets with a similar name become more important. The ranking in the search engine results always has an impact on the suggestion score. However, similarity between the asset names can also be taken into account. If you decrease this percentage, the ranking of the search results becomes more important for the suggestion score, while the similarity between the asset names becomes less important. If you increase the percentage, assets with similar names will receive a higher suggestion score. This percentage is expressed by a decimal where 1,00 equals 100%. You can enter a value greater than 1,00.
Recommendation weights for data sets	Recommendations of data sets to users	An ordered comma-separated list of values that define the importance of properties for recommendations. The order of the values reflects the importance of the value. This setting is only used for data set recommendations if your Collibra does not yet have enough data for relevant results from the active recommendations algorithms. Possible values: • *CERTIFIED*: Data sets that are certified are considered more relevant. • *POPULARITY*: The number of visits to the data set page.
Active recommendation algorithms	Recommendations of data sets to users and of business assets to data sets	A comma-separated list of algorithms that calculate recommendations. By default, all available algorithms are listed. Possible values: • *BASELINE* • *USER_MEAN* • *IICF* (Item-Item Collaborative Filtering) • *SLOPE_ONE* • *WEIGHTED_SLOPE_ONE*

Setting	impacts	Description
Data set elements threshold	Recommendations of data sets to users	The maximum number of elements per data set that the recommender will use to train the model. The data set elements are taken randomly. Lowering this number can prevent out-of-memory issues but also impacts the accuracy of recommendations for large data sets.

> Warning If you create an invalid Cron pattern, Collibra Data Intelligence Cloud stops responding.

Search index configuration

The configuration of the search index.

Setting	Description
UI search appends wildcard	• ✔ True (default): A wildcard (asterisk) is automatically added to each search query. An asterisk is not added in the following exceptions: ◦ If the query contains a tilde (~). ◦ If the query ends with a quotation mark ("). > Note This applies only to queries via the user interface. A wildcard is not added automatically for REST API queries. • ✘ False: No wildcard is added to the search query.
Maximum batch size	The amount of resources scanned in one go for the search query. The default value is 5,000. The maximum value is 30,000.
Maximum batch size for relations	Maximum batch size for relations reindex.

Setting	Description
Stop words (Requires restart)	A list of stop words that are ignored as tokens for the index. The default list of English stop words includes: a, an, and, are, as, at, be, but, by, for, if, in, into, is, it, no, not, of, on, or, such, that, the, their, then, there, these, they, this, to, was, will, with If you choose not to create your own list of stop words, the default list applies. If you create your own list of stop words, you have to: 1. Reindexing Collibra Data Intelligence Cloud. 2. Restart the environment to apply your changes. See Stop an environment and Start an environment.

Tokenizer

The configuration of the tokenizer of the indexing mechanism. If you edit these settings, you need to restart and reindex your environment.

Setting	Description
Type	The tokenizer that is used. Currently two tokenizers are supported: • **Standard** (default): This tokenizer uses the word break rules from the Unicode Text Segmentation algorithm, as specified in Unicode Standard Annex #29. • **Character**: This tokenizer sees words as groups of all alphanumeric characters together with a configurable list of extra characters. This can be used if you know for sure which characters should keep certain words together. For example, if you want to keep words with a dash (-) together, you have to add the dash in the **allowedCharacters** parameter.
Parameter map	The allowed characters if the **Type** is **Character**. • **Field key**: This field has to contain *allowedCharacters*. • **Field value**: The concatenated list of characters that does not split strings into separate tokens. For example, the concatenated list -'allows dashes and apostrophes in tokens.

Boosting

The configuration of the boosting function.

Setting	Description
Asset	The boost factor of assets.
Class Match	The boost factor of data classes.
Community	The boost factor of communities.
Domain	The boost factor of domains.
User	The boost factor of users.
User group	The boost factor of user groups.
Name	The boost factor of names.
Comment	The boost factor of comments.
Tag	The boost factor of tags.
Attribute boost map	The boost factor of attribute types. • **Field key**: The attribute type ID. • **Field value**: The boost factor of the attribute type.
Display exact match of name as first	• ✓ True (default): If the name of an asset is exactly the same as the search text, put it at the top of the search results regardless of boost factors. • ✕ False: Use the regular search order, taking into account boost factors.
Asset boost map	The boost factor of asset types. • **Field key**: The asset type ID. • **Field value**: The boost factor of the asset type.

Setting	Description
Partial exact match enabled	Enables partial exact matching while searching for multi word phrases. • ✔ True (default): For multi-word search text, the search engine considers the exact match percentage with the resource name, when ordering the results. Example You enter search text "scheduled maintenance". Two example assets are ordered as follows: a. An asset named "daily **scheduled maintenance**", as two of the three words (66%) match exactly. b. An asset named "daily **scheduled maintenance** revised", as two of the four words (50%) match exactly. • ✘ False: The exact match percentage is not taken into account in the score calculation.

Slow logs configuration

The configuration of the slow logs function.

Setting	Description
Indexing threshold	The time limit, in milliseconds, after which an index query is logged in Elasticsearch. If the value is set to 0 (zero), all index queries are logged. Changes to this setting require a full reindex of your Collibra Data Intelligence Cloud environment.
Fetching threshold	The time limit, in milliseconds, after which a fetch query is logged in Elasticsearch. If the value is set to 0 (zero), all fetch queries are logged. Changes to this setting require a full reindex of your Collibra Data Intelligence Cloud environment.

Upload configuration (*)

The configuration of the file upload service.

The file upload restrictions apply to the following actions in Collibra:

- Importing and exporting assets and complex relations.
- Uploading attachments.
- Importing and exporting CMA files in the Migration feature.

Setting	Description
Max file size	The maximum file size in bytes for uploads. The default is 10 MB or 10 485 760 bytes.
Max per day	The maximum number of uploads per user per day. The default is 150 uploads.
Accepted content types	The MIME type names of the files you want to allow for uploads. For example, type *application/pdf* for PDF files.

Statistics configuration

The configuration of statistics.

Setting	Description
Buffer size	The maximum amount of statistics entries that the buffer can contain before saving them in the database. The default value is *10*.
Buffer flush time	The maximum amount of time in milliseconds to keep statistic entries in memory before saving them in the database. The default values is *10,000*.

<table>
<tr><th>Setting</th><th>Description</th></tr>
<tr><td>Cron map</td><td>List of statistics, listed by their Cron name, and a Cron interval.

These are the default values:

<table>
<tr><th>Field key</th><th>Field value</th></tr>
<tr><td>workflow-task</td><td>0 59 23 * * ?</td></tr>
<tr><td>active-users</td><td>0 0/15 * * * ?</td></tr>
<tr><td>term-count</td><td>0 59 23 * * ?</td></tr>
<tr><td>vocabulary-count</td><td>0 59 23 * * ?</td></tr>
<tr><td>page-hit</td><td>0 0 * * * ?</td></tr>
<tr><td>task-count</td><td>0 0 * * * ?</td></tr>
</table>

If you create an invalid Cron pattern, Collibra Data Intelligence Cloud stops responding.</td></tr>
</table>

Import configuration

The configuration for imports.

<table>
<tr><th>Setting</th><th>Description</th></tr>
<tr><td>Rebuild hyper-links after import</td><td>• ✓ True (default): Automatically rebuild the hyperlinks after an import.
• ✕ False: Do not rebuild the hyperlinks after an import.</td></tr>
<tr><td>Asset respons-ibilities support</td><td>• ✓ True: Enable importing responsibilities at asset level.
• ✕ False (default): Disable importing responsibilities at asset level.

Warning Setting specific responsibilities on a large number of resources will affect the performance and stability of the system.</td></tr>
<tr><td>Temp files stor-age</td><td>The location of the temporary files used by the import job.

The default value is FILE.</td></tr>
</table>

Excel import configuration

The configuration of Excel import.

Setting	Description
The default CSV separator character	The default separator character of the CSV fields for complex relations.
The default CSV quote character	The default quote character of the CSV fields for complex relations.
Number of rows per chunk of data	When importing views, the database is called repeatedly, each time importing a chunk of data from the import file. This option defines how many rows each chunk of data can contain. Lower values reduce the burden on memory. Higher values require more memory, but may slightly increase the speed of the export. The default value is 5,000.

Excel export configuration

The configuration of Excel export.

Setting	Description
The default CSV separator character	The default separator character of the CSV fields for complex relations.
The default CSV quote character	The default quote character of the CSV fields for complex relations.

Setting	Description
Number of rows per chunk of data	When exporting views, the database is called repeatedly, each time fetching a chunk of data to build the export file. This option defines how many rows each chunk of data can contain. Lower values reduce the burden on memory. Higher values require more memory, but may slightly increase the speed of the export. The default value is 5,000.

CSV export configuration

The configuration of CSV export.

Setting	Description
Always use quotes	• ✔ True: Use quotes for every cell in the CSV. • ✘ False (default): Only use quotes when necessary.
Number of rows per chunk of data	When exporting views, the database is called repeatedly, each time fetching a chunk of data to build the export file. This option defines how many rows each chunk of data can contain. Lower values reduce the burden on memory. Higher values require more memory, but may slightly increase the speed of the export. The default value is 5,000.

User interface configuration (*)

The configuration of user interface features.

Setting	Description
Optimize CSS	• ✔ True (default): The CSS files are optimized to improve performance of the user interface. • ✘ False: The CSS files are not optimized.

Setting	Description
Optimize JavaScript	• ✓ True (default): The JavaScript code is optimized to improve the performance of the user interface. • ✕ False: The JavaScript code is not optimized
Concatenate JavaScript	• ✓ True (default) • ✕ False
Velocity cache	• ✓ True: The velocity cache is enabled. • ✕ False: The velocity cache is disabled. This allows you to reload velocity templates without restarting Collibra.
Modules JSON overrides	
Modules properties overrides	
Page definition overrides	

API call logging

The configuration of the API call logging.

Setting	Description
Enabled	• ✓ True: API call logging is enabled. • ✕ False (default): API call logging is disabled.
Maximum number of log entries (Requires restart)	The maximum number of API calls to store in the database. Once this number is reached, the oldest records are overwritten. The default value is 1,000,000.
Pattern duration list	The list of methods and a corresponding minimum duration time. The minimum duration time is the minimum time before the method is stored in the database.

Setting	Description
Minimum duration	The time in milliseconds that an API call must last before it is logged.
Method pattern	The method that you want to log in the database. For each pattern that you want to log, you have to add a new pattern.

System metrics

The configuration of metric collection.

Setting	Description
Enable (Requires restart)	• ✓ True (default): Metric collection is enabled. • ✕ False: Metric collection is disabled.
Enable JVM metrics (Requires restart)	• ✓ True (default): JVM metric collection is enabled. • ✕ False: JVM metric collection is disabled.
Enable advanced metrics (Requires restart)	• ✓ True: Advanced metrics collection is enabled. Enabling this option has a negative impact on the performance of your environment. • ✕ False (default): Advanced metrics collection is disabled.
Enable minimal monitoring (Requires restart)	• ✓ True: Monitoring of the metrics is enabled. • ✕ False (default): Monitoring of the metrics is disabled.

Security configuration

The configuration of security.

Setting	Description
X-Frame options (Requires restart)	The content of the HTTP-header `X-Frame-Options`. This is set on all rendered pages and is used to avoid clickjacking attacks. By default, only pages with the same origin can use the rendered pages in a frame.

Setting	Description
Limit user sessions	• ✔ True: A user can only open one session. • ✘ False (default): A user can open multiple sessions.
Office research guest access	• ✔ True: The Office research integration is always allowed guest access via REST, regardless of the general Guest access setting. • ✘ False (default): The general Guest access setting is kept. Note Currently, The Office research integration is only available when Collibra Data Intelligence Cloud is publicly available, which is why this override setting is necessary.
Prevent advanced html features in text dashboard	Text widgets can contain full HTML. However, this means an attacker could potentially execute an XSS attack by injecting malicious HTML. For more information, see the Troubleshooting section. • ✔ True: Potentially dangerous HTML elements are removed from text attributes when you save the text field. • ✘ False (default): No HTML elements are removed from text attributes when you save the text field. Note If you enable this setting, the following HTML elements are deleted when you save: • script (including JavaScript) • svg • frame • frameset • iframe • any event handlers
Guest access	• ✔ True: Anyone that can access the URL, has viewing rights to the system. • ✘ False (default): The user is asked to sign in before having access to any data.
Enable schema introspection	• ✔ True: Schema fields are shown during an introspection. • ✘ False (default): Schema fields are hidden during an introspection.

Setting	Description
Enabel customer validation functions	• ✔ True (default): Groovy scripts with custom validation functions can be loaded. • ✖ False: Groovy scripts with custom validation functions cannot be loaded.

LDAP

The configuration of an LDAP server to handle the authentication.

Setting	Description
Enable LDAP integration (Requires restart)	• ✔ True: The LDAP integration is enabled. • ✖ False (default): The LDAP integration is disabled.
Sync after restore	• ✔ True (default): LDAP data is synchronized with Collibra when an initial data set is bootstrapped. • ✖ False: LDAP data is synchronized with Collibra only when the LDAP synchronization job is triggered.
User page size	The page size that is used when retrieving users during synchronization. The default value is 500. You can set it to *0* to disable paging. Note This is a global setting. If you are working with multiple LDAP servers, only the value for the main server is taken into account.
Group page size	The page size that is used when retrieving groups. You can set it to *0* to disable paging. Note This is a global setting. If you are working with multiple LDAP servers, only the value for the main server is taken into account.

Setting	Description
Time limit	Specifies the time limit in milliseconds for all LDAP searches. The default value is 120,000. You can set it to *0* to disable the time limit. Tip • If you get **Time limit Exceeded** error messages, increase the default value or check why the LDAP search takes too long. • We recommend that you modify the **User page size** and **Group page size** settings before you modify this setting.
Sync job enabled	• ✔ True (default): The synchronization job is enabled. • ✖ False: The synchronization job is disabled.
Sync job cron	The schedule to perform an LDAP synchronization (CRON). The default value for this setting is daily at midnight. If you create an invalid Cron pattern, Collibra Data Intelligence Cloud stops responding.
User field mapping	The configuration mapping of all the user fields. This determines which LDAP field ismapped to which user field. Empty fields are ignored during the synchronization.
Username	The unique user ID in the LDAP, typically UID. This is a mandatory field.
Email	The corresponding email field in the LDAP directory. This is a mandatory field.
First name	The first name field in the LDAP directory.
Last name	The last name field in the LDAP directory.
Middle name	The middle name field of the LDAP directory, this is usually givenName.
Enabled	Indication whether a user is active or inactive in LDAP.

Setting	Description
Language	The language and locale of the user. It has to contain a language code and may contain a country code. • The language has to be an ISO language code. • The country has to be an ISO country code. Examples: *pl*, *en_US*, *nl_BE*.
Group	The LDAP property that defines to which groups the user belongs. If there is a group entry in the LDAP directory, use the **Group field mapping** settings.
Additional email list	An additional email list.
Instant messaging fields	The mapping for the user's IM locations.
AIM	The mapping for the user's AOL IM account.
Google Talk	The mapping for the user's Google Talk IM account.
Icq	The mapping for the user's ICQ IM account.
Jabber	The mapping for the user's Jabber IM account.
Messenger	The mapping for the user's Live Messenger IM account.
Skype	The mapping for the user's Skype IM account.
Yahoo Messenger	The mapping for the user's Yahoo Messenger IM account.
Website map	Enter the field value and field key to map a social media website.
Phone	The mapping for the user's phone.
Fax	The mapping for the user's fax number.
Mobile	The mapping for the user's mobile number.
Pager	The mapping for the user's pager number.

Setting	Description
Private	The mapping for the user's private number.
Work	The mapping for the user's work number.
Other	The mapping for any other phone number for this user.
Home address	The mapping for the user's home address.
Street	The mapping for the user's street.
Number	The mapping for the user's number.
City	The mapping for the user's city.
Post code	The mapping for the user's postal code.
State	The mapping for the user's state.
Country	The mapping for the user's country.
Work address	The mapping for the user's work address.
Street	The mapping for the user's street.
Number	The mapping for the user's number.
City	The mapping for the user's city.
Post code	The mapping for the user's postal code.
State	The mapping for the user's state.
Country	The mapping for the user's country.
Gender	The mapping information for the user's gender.
Mapping	The attribute key for the gender value. If the content equals one of the male or female mappings, the user will be saved as male or female. Otherwise a default of *UNKNOWN* will be used.

Setting	Description
Male value	The value for male users.
Female value	The value for female users.
Group field mapping	Groups can be defined as a separate structure or as a userField. The following section allows you to sync with a group structure that is unrelated to the user structure.
Group name field	The name of the group to use in the application.
Users field	The user DNs that are member of the group.
Servers	The Collibra parameters to map with your LDAP server parameters.
LDAP server URL	The URL or IP address to the LDAP server, for example *ldap://ldap.yourcompany.com:389* or *ldaps://ldap.yourcompany.com:636*.
Bind DN	The DN of the administrator user that is used for authentication, for example *admin*.
Bind password	The password of the administrator user.
Base DN	The base DN for when you are working with relative DNs. This base DN is used for all LDAP look-ups.
User base	The base DN of where the LDAP users for Collibra are located. If a base has been specified, it is used as a prefix for this user base. Subtree search is used, so all DNs located below are searched for matching users.
Authentication user LDAP filter	The filter that specifies which users can authenticate in the application. By default, all the objects found in the user base are selected, including the root.

<table>
<tr><th>Setting</th><th>Description</th></tr>
<tr><td>Synchronization user LDAP filter</td><td>The filter that specifies which users are imported by the synchronization job. The users have to be the same as, or a subset of, the Authentication user LDAP filter.

If you provide no value for this setting, the same filter as specified for the Authentication user LDAP filter setting is used. That allows you to synchronize only the users that have to have access to the application, even if they have not logged in yet. Users in the Authentication user LDAP filter are synchronized each time they authenticate and are only available after the first sign-in to the application. This is the default setting.</td></tr>
<tr><td>Authentication type</td><td>The authentication mechanism for authenticating users on the LDAP servers.

<table>
<tr><th>Authentication type</th><th>Explanation</th></tr>
<tr><td>none</td><td>No authentication is performed.</td></tr>
<tr><td>simple</td><td>Simple authentication is performed, using the Bind DN and Bind password as credentials. The credentials are sent as plain text.</td></tr>
<tr><td>DIGEST-MD5</td><td>Simple authentication is performed, using the Bind DN and Bind Password as credentials. The Bind password is hashed with the MD5 algorithm.</td></tr>
<tr><td>TLS-SIMPLE</td><td>A temporary secured TLS connection is set up before the credentials are sent as plain text. SSL must be configured.</td></tr>
<tr><td>TLS-EXTERNAL</td><td>A temporary secured TLS connection with external SASL authentication using a client certificate. SSL must be configured.</td></tr>
</table></td></tr>
<tr><td>Shutdown gracefully</td><td>• ✔ True: The LDAP context is destroyed immediately. When using TLS, some servers require the connection to be shut down by the client before the LDAP context is destroyed.
• ✖ False (default): The LDAP context is not destroyed immediately.</td></tr>
</table>

<table>
<tr><th>Setting</th><th>Description</th></tr>
<tr><td>Referral Setting</td><td>Specifies what to do with referrals. Possible values:
<table>
<tr><th>Referral setting</th><th>Explanation</th></tr>
<tr><td>throw</td><td>Throws an exception if a referral is encountered.</td></tr>
<tr><td>ignore (default)</td><td>All referrals are ignored.</td></tr>
<tr><td>follow</td><td>Follows the referral to the actual location of the entry on another server.
This is recommended when using Microsoft Active Directory.</td></tr>
</table>

Note If you are experiencing slow searches on Microsoft Active Directory with the follow value for the Referral setting, try using the Global Catalog as Active Directory domain controller. The Global Catalog enables searching for Active Directory objects in any domain in the forest without the need for subordinate referrals. This can dramatically speed up searching. However, the Global Catalog only contains a subset of the attributes of an object. This solution is only viable if the attributes requested for the search results are stored in the global catalog. Note that the Global Catalog is accessible on port 3268/3269, not the standard 389/636 LDAP ports.</td></tr>
<tr><td>Group base DN</td><td>The base DN of where all the groups are located. If a base has been specified, that base is used as the prefix for this group base.</td></tr>
<tr><td>Group LDAP filter</td><td>The LDAP filter to which each group has to comply to be synchronized.</td></tr>
<tr><td>Batch synchronization</td><td>The synchronization of the users with the LDAP server happens in batches.</td></tr>
<tr><td>Batch size</td><td>The number of users in each batch. If a batch fails, none of the users in that batch is updated and the user names are listed in the DGC service log. Other batches are processed as normal. After processing all batches, Collibra disables users that are no longer in LDAP, unless one ore more batches failed.
Set the value to 0 to disable batch processing.</td></tr>
</table>

Password

The configuration of passwords.

Setting	Description
Minimum length (Requires restart)	The minimum length of passwords. The default minimum length is 12.
Maximum length (Requires restart)	The maximum length of passwords. The default maximum length is 1,024.
Digits required (Requires restart)	• ✓ True (default): Passwords have to contain one or more digits. • ✕ False: Passwords do not have to contain digits.
Non alpha-numeric required (Requires restart)	• ✓ True (default): Passwords have to contain one or more non-alpha-numeric (special) characters. • ✕ False: Passwords do not have to contain non-alphanumeric characters.
Uppercase required (Requires restart)	• ✓ True (default): Passwords have to contain one or more upper-case characters. • ✕ False: Passwords do not have to contain upper-case characters.
Lowercase required (Requires restart)	• ✓ True (default): Passwords have to contain one or more lower-case characters. • ✕ False: Passwords do not have to contain lower-case characters.
Username disallowed (Requires restart)	• ✓ True (default): Passwords cannot be the username. • ✕ False: Passwords can be the username.
Expiration interval (months)	The number of months before users have to change their passwords. Set it to *0* if users never have to change their passwords. The default interval is 6 months.
Allowed login failures	The number of consecutive failed login attempts that are allowed before the user account is disabled. Set it to *0* for unlimited attempts. The default is 3 login failures.
No reuse count	The number of previous passwords users cannot reuse. The default is *1*: the user cannot change his password to what it currently is. Set this to *0* to allow using the same password.

Setting	Description
Password reset link validity period	The number of minutes that a link to reset a password remains valid. Beyond this time, the user has to request a new password reset link. The default value is 60 minutes. The minimum value is 15 minutes, the maximum value is 1,440 minutes (24 hours).
Account lock-out duration	The number of minutes that a user cannot sign in after too many failed sign-in attempts. If the number of minutes is set to 0, a Collibra administrator must reset the password to unlock the account. This setting is only applicable if the "Allowed sign-in failures" setting is defined. A locked-out account does not mean that your account is disabled.

REST

The security configuration of the REST interface.

Setting	Description
CSRF token enabled	• ✓ True: The validity of a request is checked with a CSRF token. • ✗ False (default): The validity of a request is not checked with a CSRF token.
Referrer enabled	• ✓ True: The HTTP referrer header is used to identify the origin of the request. • ✗ False (default): The HTTP referrer header is not used to identify the origin of the request. It is recommended to leave this option disabled.
Referrer checking allow empty	• ✓ True (default): The HTTP referrer header can be empty. • ✗ False: The HTTP referrer header cannot be empty.

SSL

The configuration of SSL.

Setting	Description
Key store name	The name of the keystore file. The file is expected to be in the **<collibra_data>/dgc/security** folder.
Key store password	The password of the keystore.
Key store type	The type of the keystore file. For example, *JKS* or *PKCS12*.
Trust store name	The name of the truststore file. The file is expected to be in the **<collibra_data>/dgc/security** folder.
Trust store password	The password of the truststore.
Trust store type	The type of the truststore file. For example, *JKS* or *PKCS12*.

SSO

The configuration of Single Sign-On (SSO) authentication.

Setting	Description
Mode	The SSO mode of Collibra. The possible values are: • SAML_ATTRIBUTES • SAML_LDAP • SSO_HEADER • SSO_HEADER_LDAP • DISABLED
Header	The name of the header to be checked. The contents of this header is used for the search query, which is *SSO_HEADER = username*. The value of the actual query depends on **DN** and possibly **Attribute**.

Setting	Description
DN	If the **SSO mode** is **SSO_HEADER_LDAP** or **SAML_LDAP**, this field determines whether the distinguished name (DN) or attribute is used: • ✔ True: The header has to contain the distinguished name (DN) of the user in the LDAP. • ✘ False (default): The header has to contain the value of **Attribute**. If the **SSO mode** is **DISABLED**, **SSO_HEADER** or **SAML_ATTRIBUTES**, this field is ignored.
Attribute	This field is only used if the **SSO mode** is **SSO_HEADER_LDAP** or **SAML_LDAP**, and if **DN** is **False**. If the above criteria are met, the LDAP has to contain this value. Example If **Attribute** is *FirstName*, then the header should contain the **FirstName** of the user that was signed in.
Disable automatic user creation when signing in via SSO	If users try to sign in via SSO, they still need a user account in Collibra. You can either create the user accounts automatically when they sign in, or create the user accounts manually or via LDAP synchronization • ✔ True: User accounts are not created automatically. • ✘ False (default): User accounts are created automatically.
Disable the Collibra signin page	When SSO is enabled, a user can still navigate to the /signin page and try to log in via that page. However, you can disable that page. • ✔ True: Users cannot access the Collibra signin page. • ✘ False (default): Users can access the Collibra signin page
SAML	The configuration of SAML.
Metadata HTTP	The URL of the SAML metadata file to be used. The URL always has to be reachable by the Collibra environment.
Entity Provider Entity ID	The entity ID inside the metadata to be referenced. Note A metadata file can describe multiple entity IDs, make sure to use in the entity ID from the correct metadata file.

Setting	Description
Attribute fields	The mappings of attributes in the SAML response. The values are used as keys to look for in the SAML response. Examples of attribute fields are first name, last name, address information, phone numbers and so on.
First name	The mapping for the user's first name. This attribute is optional. The value can be empty.
Last name	The mapping for the user's last name. This attribute is optional. The value can be empty.
Email	The mapping for the user's email address. This attribute is optional for existing users, but mandatory for new users. Warning If the email address is invalid when you synchronize, the user is deactivated and the user information is not updated.
Enabled	The mapping that indicates whether the account of the incoming user is enabled.
Group	The mapping (attribute) which indicates to which Collibra groups the user should be added. If the groups don't exist yet, they will be created. This attribute can have multiple values (groups) or the groups can be sent as a comma-separated list of groups. If passing groups in this attribute, you must set **Groups DGC Managed** to *False*.
Phone	The mapping for the user's phone.
Fax	The mapping for the user's fax number.
Mobile	The mapping for the user's mobile number.
Pager	The mapping for the user's pager number.
Private	The mapping for the user's private number.
Work	The mapping for the user's work number.

Setting	Description
Other	The mapping for any other phone number for this user.
Home address	The mapping for the user's home address.
Street	The mapping for the user's street.
Number	The mapping for the user's number.
City	The mapping for the user's city.
Post code	The mapping for the user's postal code.
State	The mapping for the user's state.
Country	The mapping for the user's country.
Work address	The mapping for the user's work address.
Street	The mapping for the user's street.
Number	The mapping for the user's number.
City	The mapping for the user's city.
Post code	The mapping for the user's postal code.
State	The mapping for the user's state.
Country	The mapping for the user's country.
Instant messaging	The mapping for the user's IM locations.
AIM	The mapping for the user's AOL IM account.
Google Talk	The mapping for the user's Google Talk IM account.
Icq	The mapping for the user's ICQ IM account.

Setting	Description
Jabber	The mapping for the user's Jabber IM account.
Messenger	The mapping for the user's Live Messenger IM account.
Skype	The mapping for the user's Skype IM account.
Yahoo Messenger	The mapping for the user's Yahoo Messenger IM account.
Gender	The mapping information for the user's gender.
Mapping	The attribute key for the gender value. If the content equals one of the male or female mappings, the user will be saved as male or female. Otherwise a default of *UNKNOWN* will be used.
Male value	The value for male users.
Female value	The value for female users.
Groups DGC managed	Option to configure that groups should be managed by Collibra, or that groups should be set by the SAML assertion (SAML+Attributes mode). This option is only relevant if **Mode** is *SAML_ATTRIBUTES*. • ✓ True: The groups are fully managed by Collibra. In the UI the admin has the option to assign groups to users, without it being overwritten by SAML. • ✗ False (default): The groups are managed by the SAML assertions. In this case the groups are managed by the SAML IDP. Be sure to configure the **Group** attribute in the **Attribute Fields** section.

Setting	Description
Service Provider Entity ID	Field that determines the value of the `Entity ID` parameter in the service provider metadata returned by Collibra. The default value is empty, in which case Collibra uses the value of the `Base URL` field. Enter a custom value if the base URL does not match the `audience` configured in your SAML identity provider. Warning The value of the `audience` restriction in the SAML response has to be exactly the same as the value of this field. Note SSO does not work if the `Service Provider Entity ID` field contains the base URL with trailing forward slash (for example `www.collibra.com/`), and the `audience` of your IDP contains the base URL without a trailing forward slash (for example `www.collibra.com`). Both values need to be exactly the same. In this case, you can resolve the issue by changing the value in the configuration of your IDP, or the value of this field. It does not matter whether both have a trailing forward slash or not, as long as they contain the same value.
Sign authentication requests (Requires restart)	• ✔ True: Authentication requests have to be signed. • ✘ False (default): Authentication request don't have to be signed.
Force authn	• ✔ True (default): The SP authentication request forces re-authentication. • ✘ False: The SP authentication request does not force re-authentication.
Force passive	• ✔ True: The reauthentication has to happen in the background. • ✘ False (default): The reauthentication does not have to happen in the background. This is only relevant if **Force authn** is *True*.
Name ID	Name ID that is used in the SP authentication. The default value is *urn:oasis:names:tc:SAML:2.0:nameid-format:persistent*. The Name ID value is mandatory.

Setting	Description
Name ID allow create	• ✔ True (default): The IDP can create a name ID to fulfill the SP authentication request. • ✘ False: The IDP cannot create a name ID to fulfill the SP authentication request.
Disable client address	• ✔ True: The validation of the client IP address in the assertion message is disabled. • ✘ False (default): The validation of the client IP address in the assertion message is enabled.
SAML Requested authentication context	Settings for the SAML requested authentication context. The IDP uses the authentication context to authenticate the user. By default, the authentication context mandates user/password authentication over HTTPS.
Disable	• ✔ True: The requested authentication context section is not sent in the SAML request. • ✘ False (default): The requested authentication context section is sent in the SAML request.
Comparison type	The comparison type that is transmitted in the requested authentication context. Possible values: • *minimum* • *maximum* • *better* • *exact* (default value) For more information about the comparison type values, refer to the SAML specifications.
Reference list	The list of class references in the requested authentication context. You can separate list items with the pipe character (|). For more information about this list, refer to the SAML specifications.
Declaration list	The list of class declarations in the requested authentication context. You can separate list items with the pipe character (|). For more information about this list, refer to the SAML specifications.

Setting	Description
Response decryption mode	Enable the support for encrypted SAML responses. • DISABLED: Collibra only accepts plain-text SAML responses. • OPTIONAL: Collibra can handle both encrypted and plain-text SAML responses. • FORCED: Collibra only accepts encrypted SAML responses. Once OPTIONAL or FORCED is selected, the encryption key pair is generated and added to the Collibra SAML keystore. A self-signed certificate is generated and works in most situations. If your IdP rejects self-signed certificates, you will have to add a certificate that is signed by a trusted 3rd party.
Validity period of the SAML certificate	The SAML certificate expiry date in years. By default, the SAML certificate expires after 20 years.

Signout

The configuration of redirecting after signing out of Collibra.

Setting	Description
Override signout URL (Requires restart)	• ✓ True: Redirect the user to a specific website after signing out. • ✗ False (default): Redirect the user to the sign-in page after signing out.
Signout redirect URL (Requires restart)	The URL to be redirected to when signing out.

Session (*)

The configuration related to sessions.

Setting	Description
Idle Session timeout	The time after which you are logged out if you are inactive. The minimum value is 5 minutes, the maximum is 24 hours. The default is 1,800 seconds.

Import/Export

The configuration to avoid the Formula Injection vulnerability in Excel.

Setting	Description
Escape Excel formulas	The option to disable Formula Injection into Excel. When enabling this option, an escape character is added at the beginning of Excel formulas during the export and is removed when importing formulas. The escape character will be added to fields that start with one of the following characters: • equation: = • plus: + • minus: - • at-sign: @ This option is enabled by default.
Excel formulas escape character	The escape character for Excel formulas when exporting or importing data.

JWT

The JSON Web Token configuration.

Setting	Description
JSON Web Key Set URL	The URL to retrieve public key information needed to verify the authenticity of JSON Web Tokens (JWTs), issued by an authorization server. This setting is required to enable JWT authentication.

Setting	Description
JWT Token Types	A case-insensitive comma-separated list of accepted JWT media types coming in the **typ** header parameter. Leave blank if the authorization server does not provide a media type parameter. The default values is **at+jwt,jwt**.
JWT Algorithms	A comma-separated list of accepted JWT algorithms coming in the **alg** header parameter. See https://tools.ietf.org/html/rfc7518#section-3.1 for details. Leave blank to accept all digital signature algorithms.
JWT Issuer	The accepted issuer coming in the **iss** JWT claim. Leave blank if the authorization server does not provide an issuer claim.
JWT Audience	A comma-separated list of accepted audience values for the **aud** claim. The value for this field is a configuration setting in your authorization server, which identifies your Collibra environment as the intended recipient of the JWT. Leave blank if the authorization server does not provide an audience claim.
JWT Principal ID Claim Name.	The name of the JWT claim containing the principal's identity. See https://tools.ietf.org/html/rfc7519#section-4.1.2 for details. Defaults to the standard subject claim, **sub**. Change this setting only if your authorization server has other means of identifying the principal, for example, a **client_id** claim. This setting is required if JWT authentication is enabled.
JWT Maximum Clock Skew	The maximum acceptable difference in seconds between the clocks of the machines running the authorization server and Collibra. Differences smaller than the given amount are ignored when performing time comparisons for token validation. The default value is 60 seconds if left blank.

HTTP headers

The configuration of the HTTP headers

Field	Description
URL pattern	The pattern of the URLs to which the HTTP response header is applied. This field supports wildcards such as `**`, `*` and `?`. > Tip This pattern matches all URLs: `/**`.
HTTP headers	The HTTP response headers in a key-value format. You can add new HTTP response headers by clicking **Add** at the bottom of the section, and entering the HTTP response header name as the field key and the HTTP response header value as the field value.

Whitelists

The configuration for whitelist placeholders that can be used in security headers.

Option	Description
connect-src whitelist	The 'connect-src' whitelist. To use this whitelist in a security header, use the '{connectSrcWl}' placeholder.
font-src whitelist	The 'font-src' whitelist. To use this whitelist in a security header, use the '{fontSrcWl}' placeholder.
frame-src whitelist	The 'frame-src' whitelist. To use this whitelist in a security header, use the '{frameSrcWl}' placeholder.
img-src whitelist	The 'img-src' whitelist. To use this whitelist in a security header, use the '{imgSrcWl}' placeholder.
script-src whitelist	The 'script-src' whitelist. To use this whitelist in a security header, use the '{scriptSrcWl}' placeholder.
style-src whitelist	The 'style-src' whitelist. To use this whitelist in a security header, use the '{styleSrcWl}' placeholder.
frame-ancestors whitelist	The 'frame-ancestors' whitelist. To use this whitelist in a security header, use the '{frameAncestorsWl}' placeholder.

Option	Description
Tableau frame-ancestors whitelist	The tableau 'frame-ancestors' whitelist. To use this whitelist in a security header, use the '{tableauFrameAncestorsWl}' placeholder.

Disclaimer

The configuration of a disclaimer upon signing in to Collibra.

Setting	Description
Disclaimer	• ✓ True: Upon signing in, show a disclaimer that you have to agree with before you can continue. • ✕ False (default): Don't show a disclaimer.
Disclaimer message	The disclaimer message that is shown after signing in. If you leave this field empty, there is a default message. You can use basic html tags, such as headers, paragraphs, images and hyperlinks.

Workflow engine configuration (*)

The configuration of the workflow engine.

Setting	Description
Activate default escalation	• ✓ True (default): Automatically add an escalation function to user tasks after a configurable period of time. • ✕ False: Don't escalate task automatically. This option only works for tasks that do not yet have a configured escalation path configured in Collibra Data Intelligence Cloud.
Default escalation timer duration	The duration before a task is escalated. For more information on the format, please refer to ISO 8601.

Setting	Description
Activate default task email notification	• ✔ True (default): Send email notifications to any candidate user of a user task in the workflow engine. • ✖ False: Do not send email notifications to candidate users.

Collibra Connect

The configuration to communicate with Collibra Connect.

Setting	Description
Base URL	The URL to Collibra Connect.
Username	The username to connect to Collibra Connect.
Password	The password to connect to Collibra Connect.

Register data source

Global parameters that apply to Data Source Registration.

Setting	Description
Table types to ignore	A comma separated list of table types that are not ingested. For example, *INDEX* and *SEQUENCE*.
AWS regions restriction	A list of AWS regions Data Catalog is allowed to connect to. For example, *eu-west-3* and *us-east-2*. For a list of all AWS locations, see the AWS documentation. If you want to allow Collibra to make a connection to any AWS region, leave the field empty.
AWS API call rate	Allowed number of AWS API calls per second. Use this option to limit the number of API calls per second to prevent throttling errors from the AWS API.

Setting	Description
Database registration via Edge	An option to enable database registration via Edge. • ✔ True: Register a data source via Edge. • ✘ False: Register a data source via Jobserver only. Note Enabling data source registration via Edge does not prevent you from registering a data source via Jobserver as well.
Data Quality Synchronization UI via DQ Connector on Edge	An option to enable the Data Quality extraction interface in Collibra • ✔ True: The Quality extraction tab is available on the configuration page of a database asset • ✘ False (default): The Quality extraction tab is not available and as such, it is not possible to extract and synchronize data quality information. You can only enable Data Quality synchronization if you also enabled Database registration via Edge.

Jobserver (*)

The configuration of the Jobserver service.

Setting	Description
Jobserver list	The list of registered Jobserver instances.
Name	The name of the Jobserver as it will appear when you register a data source in Data Catalog. The name is a freely chosen name but it is recommended to only use alphanumerical characters and dashes, for example Jobserver-1. You will have to use this name as the ID of the gateway and in the address of this configuration.
Protocol	The protocol that is used for the communication between the Data Governance Center service and the Jobserver service. It is recommended to use HTTPS, especially if the services are hosted in different network segments.

Setting	Description
Address	The address (IP address, URL, hostname) of the Jobserver.
Trusted server CA certificate	The certificate of the trusted CA needed to validate the server certificate. If blank, the default truststore will be used. The default truststore is defined in the SSL configuration section of the DGC service. The CA certificate of the server party (Jobserver).
Client certificate	The client certificate offered by the DGC service to the server. If blank, you cannot select mutual authentication as the Jobserver service authentication level.
Client private key	The private key of the DGC service's certificate.
Table profiling data size	The approximate maximum disk size of the data in MB that will be used to profile a table. The value cannot exceed 10,000.
Test connection timeout	This timeout is a time limit (in seconds) after which the connection test is stopped and a timeout error is shown. The default value is 60 seconds.

Data profiling

The global configuration of Data Profiling. Profiling must be executed again after a change in this section.

Setting	Description
Maximum number of samples	The maximum number of rows taken as a sample during profiling.
Maximum value length	The maximum length of a value extracted during profiling or sampling. Additional characters are trimmed.
Default date pattern	The default format used to decode dates. It is the default pattern used for detecting dates when the **Date Pattern** and/or **Time Pattern** attribute is not specified in Column assets.

Setting	Description
Default time pattern	The default format used to decode times. It is the default pattern used for detecting times when the **Date Pattern** and/or **Time Pattern** attribute is not specified in Column assets.
Default combined date and time pattern	The default format used to decode combined dates and times. It is the default pattern used for detecting combined dates and times when the **Date Pattern** and/or **Time Pattern** attribute is not specified in Column assets.
Empty values	A comma separated list of strings enclosed in double quotes. A value that matches one of those expressions is considered an empty value. Please note that a database **null** value is always considered an empty value, for example *""*, *"na"* and *"none"*.
Data type detection threshold	The percentage of matching Column values to reach for an Advanced Data Type to be considered a possible Data Type for that Column. This is expressed as a value between 0.0 and 1.0).
Anonymize data	An option to anonymize sensitive data. • ✔ True: Content in columns with data type Text or Geo is removed or replaced by a random hash value before the profiling results are sent to the cloud. • ✘ False (default): No content is removed or replaced by a random hash value. Tip If you profile and classify via Edge, the data in columns with data type Text or Geo is automatically anonymized before it is sent to Collibra Data Intelligence Cloud.
Database profiling via Edge	An option to enable profiling and classifying synchronized metadata via Edge instead of Jobserver. • ✔ True: Profiling and classify via Edge. • ✘ False: Profile via Jobserver and classify via the Data Classification Platform. Note You can only enable Database profiling via Edge if you also enabled Database registration via Edge.

Beta features

The configuration of features in beta state.

Setting	Description
Tableau provisioning enabled	• ✓ True: Provisioning to Tableau is enabled. • ✕ False (default): Provisioning to Tableau is disabled.
Max number of concurrent import jobs	The maximum number of import jobs that can be executed at the same time via the API. This is to avoid memory issues. Default value is 4, set to *0* if there is no limit.
Task sidebar	• ✓ True: Workflow tasks appear in the sidebar on both resource pages an the task management page. Task forms appear in the sidebar instead of dialog boxes. Users can seamlessly complete their tasks from the task management page and have a side-by-side view of the tasks and resource details on resource pages. • ✕ False (default): Workflow tasks appear in the task bar on resource pages and in a sidebar on the task management page. Task forms appear in dialog boxes. The behavior is the same as with older versions of Collibra.
Settings landing enabled	• ✓ True: Show the new Settings landing page in your Collibra environment. • ✕ False (default): Use the classic Settings page in your Collibra environment.
Search reindex using Output Module	• ✓ True: Enable the use of the Output Module when executing a search reindex. • ✕ False (default): Disable the use of the Output Module when executing a search reindex.

Throttling (*)

Throttling is a security mechanism where you can limit the number of requests per seconds to ensure security and performance of your environment.

Setting	Description
Collect metrics without throttling	• ✓ True (default): Apply throttle logic without actual throttling to collect metrics. • ✕ False: No throttle logic is applied to collect metrics.

The throttling metrics are only used for evaluation by Collibra support engineers if you report significant performance loss of your environment. We only track the number of times the throttling limit is exceeded.

By default, the throttling limit is 100 API requests per second.

REST API version 1.0 throttling (*)

Throttle configuration for REST API version 1.0.

Setting	Description
Throttling enabled	• ✓ True: REST API v1 throttling is enabled. • ✕ False (default): REST API v1 throttling is disabled.
Number of requests	The number of allowed request for the configured number of seconds.
Number of seconds	The number of seconds during which the configured number of requests can be performed.

REST API version 2.0 throttling (*)

Throttle configuration for REST API version 2.0.

Setting	Description
Throttling enabled	• ✓ True: REST API v2 throttling is enabled. • ✕ False (default): REST API v2 throttling is disabled.

Setting	Description
Number of requests	The number of allowed request for the configured number of seconds.
Number of seconds	The number of seconds during which the configured number of requests can be performed.

GraphQL throttling (*)

Throttle configuration for GraphQL.

Setting	Description
Throttling enabled	• ✓ True: GraphQL throttling is enabled. • ✕ False (default): GraphQL throttling is disabled.
Number of requests	The number of allowed request for the configured number of seconds.
Number of seconds	The number of seconds during which the configured number of requests can be performed.

Hibernate cache configuration (*)

The configuration of the hibernate second level caching. Hibernate caching uses a buffer memory that lies between Collibra and your repository database. It stores recently used data in this buffer memory to reduce the number of requests to your database, thereby improving the performance of your environment.

> Tip We recommend that you use the default values. If you choose to edit the values, contact the Collibra support department before doing so.

Setting	Description
Enabled	• ✔ True (default): Hibernate caching is enabled. • ✘ False: Hibernate caching is disabled.
Configuration list	The list of associated cache configurations.
<cache configuration>	The cache configurations. Click **Add** to create a new cache configuration.

Setting	Description
Enabled	• ✔ True: This cache configuration is enabled. • ✘ False: This cache configuration is disabled.
Name	The code of the assets that you want to cache. The codes are displayed in the following table. Do not use other codes.
Max elements in memory	The maximum amount of elements in the memory.
Eternal	• ✔ True: The cache configuration is eternal. • ✘ False: The cache configuration is not eternal. This is the default value. We recommend that you do not change the default value.
Overflow to disk	• ✔ True: Write data to disk if the cache if full. • ✘ False: Do not write data to disk if the cache is full. This is the default value. We recommend that you do not change the default value.

Name	Resource type	Recommendation for Max elements in memory
GR	Groups	Set a number that is a percentage of the total amount of groups.

Name	Resource type	Recommendation for Max elements in memory
CO	Communities	Set a number that is at least the total amount of communities if possible.
UR	Users	Set a number that is a percentage of the total amount of users.
VC	Domains	Use as many domains as possible in cache.
RP	Assets	Set a number that is a percentage of the total amount of assets.
CT	Asset Types	Set a number that is at least the total amount of asset types if possible.
VT	Domain Types	Set a number that is at least the total amount of domain types if possible.
TY	Attribute Types	Set a number that is at least the total amount of attribute types if possible.
BF	Relation Types	Set a number that is at least the total amount of relation types if possible.
ST	Status	Set a number that is at least the total amount of statuses if possible.

Default configuration (*)

The default configuration for this cache.

Do not change any of the default values.

Setting	Description
Max elements in memory	Type the maximum amount of elements in the memory.
Eternal	• ✔ True: The cache configuration is eternal. • ✘ False: The cache configuration is not eternal.
Time to idle (seconds)	Type the amount of seconds until the cache is idle.
Time to live (seconds)	Type the amount of seconds until the cache starts.
Overflow to disk	• ✔ True: The cache configuration overflows to the disk. When the cache has reached its maximum number of elements, the next elements will be stored on disk. • ✘ False: The cache configuration does not overflow to the disk. This is the default value.
Disk persistent	• ✔ True: The disk store is persistent between CacheManager instances. • ✘ False: The disk store is not persistent between CacheManager instances. This is the default value. This option is only relevant if **Overflow to disk** is set to `True`.
Disk expiry thread interval (seconds)	Enter the amount of seconds between runs of the disk expiry thread. This value is how often we check for expiry. This option is only relevant if **Overflow to disk** is set to `True`.
Memorystore eviction policy	Enter the policy how items are deleted from disk: LRU: Least recently used items are removed as first from disk. This is the default value. LFU: Least frequently used items are removed as first from disk. FIFO: First in first out principle to remove items from disk. This option is only relevant if **Overflow to disk** is set to `True`.
Statistics	This option is not used by a Collibra environment.

Graph query

The configuration of the Graph query engine which is used to retrieve data from the repository.

For the general Graph query settings, you need the SUPER role. Contact Collibra support if you want to edit these settings.

Setting	Description
work_mem setting for output module queries	A custom amount of memory that is reserved for the output module SQL queries. This setting should only be used for diagnosing potential lack of memory in case of performance issues. Performance issues may arise when large sorting or large joins are needed. By default, this option is disabled.
Enable simple joins	• ✓ True: Enable an alternative logic in the query engine to execute join operations. The alternative logic improves the query engine performance if queries don't have "to-many" relations. • ✕ False (default): Keep the default logic of the query engine.
Enable joins for view permission filtering	• ✓ True: Enable an alternative logic to calculate view permissions in output module queries to improve the performance. • ✕ False (default): Keep the default logic of the output module.

Graph query limits

Setting	Description
Enables limiting of the number of root nodes in result	• ✓ True: Enable limiting the number of root elements as result of a Graph query. • ✕ False (default): Disable limiting the number of root elements as result of a Graph query.
Maximum number of root nodes that can be requested with graph query API	The maximum number of root nodes that you can request in the view configuration of an API call (REST or workflow). If you exceed this value in the view configuration, an exception is shown. If no value is defined in the view configuration, then the default value is taken. The default value is 100,000. > Note If the number of asset types or domain types exceeds the set number, the hierarchy will be incomplete. Make sure that the limit is always higher than the actual number of asset and domain types.

Graph query timeouts

Setting	Description
Maximum number of minutes a graph query can run	The maximum number of minutes that the graph query runs before it will time out. The maximum is 1,440 minutes (1 day). The default value is 480.

Table

The configuration of tables.

Setting	Description
Time limit for loading data in tables in seconds	The time limit after which a table stops loading on a page. Example A value of *600* means that if a table hasn't loaded within 600 seconds, the task is canceled and a timeout error is shown. The default value is 60, the maximum value is *720* seconds.

Multi-column sort

The configuration of multi-column sorting.

Setting	Description
Multi-column sorting on tables	• ✔ True: Tables can be sorted on multiple columns. • ✘ False (default): Tables can be sorted on one column.
Number of columns available for multi-sort	Type the maximum number of columns that can be used to simultaneously sort tables. The default value is 3, the minimum is 1, the maximum is 9. This setting is only relevant is **Multi-column sorting on tables** is *True*.

Inherited responsibilities

Setting	Description
Enable Inherited Responsibilities	• ✔ True: Show inherited responsibilities on asset views. • ✘ False (default): Do not show inherited responsibilities on asset views. Note This setting only affects asset views and tile sets. It does not affect the Responsibilities tab page of asset pages.

Purge configuration

The configuration of the automatic purging of data from the repository database. Purging means to delete data of a specified age. This helps to keep your data relevant and keep the database from growing infinitely.

Setting	Description
Purge schedule (Requires restart)	A Cron expression specifying the timing and frequency of purge cycles. The default scheduled time is 002 * *, which equates to 02:00 every day. If you create an invalid Cron pattern, Collibra Data Intelligence Cloud stops responding.
Maximum time for each purge cycle (Requires restart)	Maximum amount of time (in seconds) allowed for each purge cycle. The default value is 7,200, which is two hours. Any qualifying data that is not purged in the allowed time will be addressed in a subsequent purge cycle, which picks up where the previous cycle left off.

List of data elements and age at which each will be purged (Requires restart)	The data elements that will be purged at the specified age, in a key-value format:
	• The data elements to be purged (**Field key**). Possible data elements:
	  ◦ **Statistics**: Data that has been used to calculate data quality.
	  ◦ **Authentication events**: Data about authentication in your environment.
	  ◦ **Validation results**: Information about data validation.
	  ◦ **Jobs**: Data about all jobs that are created in your environment.
	• The age (in months) at which each individual data element will be purged (**Field value**).
	> Note If you don't select any data elements, the purge job will still run according to schedule, but nothing will be deleted.

Enable removal of orphaned tags (Requires restart)	Option to enable the automatic deletion of tags that are not assigned to any assets. • ✓ True (default): Orphan tags are deleted according to the timing and frequency that you specify. • ✕ False: Orphan tags are not deleted.

Setting	Description
Orphaned tags removal schedule (Requires restart)	A Cron expression specifying the timing and frequency of the deletion of orphan tags. The default scheduled time is 001* *, which equates to 01:00 every day. If you create an invalid Cron pattern, Collibra Data Intelligence Cloud stops responding.

Cloud Data Classification configuration

With data classification you can automatically assign data classes to ingested data.

> Note In a Collibra Data Intelligence Cloud environment, you have to create a support ticket to configure this feature.

Setting	Description
Machine Learning platform URL (*)	The address of the machine learning platform that will classify your data.
Requester Name (*)	The unique name to identify the client when using Machine Learning platform.
API key (*)	The API Key to authorize the requester when connecting to the Machine Learning platform.
Enable Data Classification	• ✓ True: Enable Collibra's data classification technology. • ✕ False (default): Do not use Collibra's data classification technology are not accepted.

Reporting

For more information about these settings, see Introduction to the Reporting Data Layer.

Setting	Description
Cloud Provider	Cloud provider - AWS or GCP
Customer GUID	The GUID of your Collibra environment. Note This field is configured by Collibra Cloud Ops.
Collibra Insights download bucket name	The name of the AWS S3 bucket in which your reporting data is stored. Note This field is configured by Collibra Cloud Ops.
Collibra Insights AWS S3 Region	The AWS S3 region in which your data is processed. Note This field is configured by Collibra Cloud Ops.
Collibra Insights zip location pattern	A pattern with the format "/zip/insights_%s.zip", where "%s" is replaced by the Collibra Insights snapshot date. Note This field is configured by Collibra Cloud Ops.
Tableau report URL pattern	The Tableau URL pattern, which should contain {reportName}. Tip You can paste the URL from the Link field in Tableau, as described in Generate the dashboard reports you configured in Collibra Data Intelligence Cloud Settings.
Reports definitions	
Report view name	The report name, as you want it to appear on the report button in the Insights widget, for example "Data Maturity Dashboard".
Report name	The report name, as it appears in the URL of the Tableau report, for example "DataMaturityDashboard".

Catalog Experience

Data Catalog Experience improves the layout of Data Catalog's asset pages.

Setting	Description
Enable Catalog experience	• ✓ True: Catalog experience is enabled. This will improve the layout of Data Catalog's asset pages, such as those of Data Set, Schema, Table and Column assets. • ✕ False: Catalog experience is disabled.
Catalog Experience Titlebar theme	The theme for the Catalog experience. You can choose between the LIGHT and DARK. This option is only applicable if the **Enable Catalog experience** option is enabled.

Diagrams

These settings determine dialog loading time and size limits.

Setting	Description
Maximum loading time for the back end	The time limit, in seconds, after which a diagram stops fetching data. The value must be a positive integer and cannot be greater than 3,600 (one hour). The default value is 300. > Example A value of *300* means that if a diagram hasn't fetched all data within 300 seconds, the diagram stops fetching data and an empty diagram with a notification is shown.
Size limit for the backend	The maximum number of nodes plus edges that will be fetched by the backend, to build a diagram. The value must be a positive integer and cannot be greater than 100,000. The default value is 10,000. > Example A value of *10,000* means that if the total number of nodes plus edges is greater than 10,000, the diagram does not load and a notification is shown.

Setting	Description
Size limit for the frontend	The maximum number of visible nodes plus edges that can be shown on the page. The value must be a positive integer and cannot be greater than 10,000. The default value is 2,000. Example A value of *2,000* means that if the total number of visible nodes and edges is greater than 2,000, the diagram does not load and a notification is shown.
Maximum flow depth	The system-wide maximum number of flow relations between the start node and any other diagram node. The value must be an integer between 1 and 100. The default value is 50. Note • If the maximum flow depth is specified in the selected diagram view, that value supersedes the maximum you specify here. • You can also manually adjust the flow depth in the diagram.
Diagrams Business Qualifier Filter (*)	• ✔ True: Users can filter diagrams by a specified Business Qualifier asset. • ✘ False (default): Users are unable to filter diagrams by Business Qualifier.

Everywhere Desktop configuration

These settings determine some of the ways in which Collibra for Desktopinteracts with Collibra Data Intelligence Cloud.

Note These settings are only applied in Collibra for Desktop if you have Collibra 2021.01 or newer in combination with Collibra for Desktop 1.2.1 and newer.

Setting	Description
Default search filter	The filter that is applied, by default, to search results. The value must be the UUID of the filter. To find the UUID, open the Collibra environment and click in the Search box. Click the name of a search filter. In the address bar you will see the UUID of the filter. > Note Specifying a default search filter in the application will override the default filter that you specify here.
Custom Search box placeholder	Placeholder text that appears in the Search field before a user enters search text. The default text is "Search in Collibra".
Shortcut Search	Enable or disable the use of a keyboard shortcut to search for selected text in Collibra Data Intelligence Cloud from within your browser or another application. • ✔ True (default): Users can use the keyboard shortcut you specify in the following setting to search in Collibra Everywhere. • ✖ False: Keyboard shortcut is disabled.
Custom Shortcut Search	The keyboard shortcut to search for selected text in Collibra Data Intelligence Cloud from within your browser or another application. > Tip The keyboard shortcut has to be a combination of Control, Alt or Shift with one letter or number. On macOS you can also use the Command key. > Note > • To make available the keyboard shortcut, you have to enable the feature in the previous setting. > • Specifying a keyboard shortcut in the application will override the shortcut that you specify here.

Setting	Description
Enable Auto Hyperlinking	Option to enable automatic hyperlinking within Collibra for Desktop. With this option enabled, the name of an asset automatically becomes a hyperlink when you fill out a text attribute. This option only works if the **Enable hyperlinking** option in Collibra Console is also enabled.
Enable Workflows	Option to enable workflows in Collibra Everywhere. This allows you to complete tasks or start a workflow in the app. The available workflows depend on the ones that you add to the Global workflows and Asset workflow configuration.
Recommender	The Recommender helps users by suggesting relevant business assets and data sets, based on certain relation types and the past actions of similar users. • ✔ True: Recommender is enabled. • ✖ False (default): Recommender is disabled. This feature only works if Analytics is enabled. You can enable Analytics in section 1 General of the DGC service configuration.
Auto-updater	Option to automatically upgrade Collibra Everywhere when a new version is available. • ✔ True (default): Collibra Everywhere is automatically upgraded when a new version is available. • ✖ False: You need to manually upgrade Collibra Everywhere when a new version is available. > Note If you enable automatic updates, you have to whitelist the S3 bucket collibra-otg-desktop-installers in the region eu-west-1.
Allow User Configuration	Option to allow users to edit personal settings in Collibra Everywhere. • ✔ True (default): Users can edit personal settings. • ✖ False: Users cannot edit personal settings.

Setting	Description
Global workflows	The list of workflows that is available in the app's main menu. Enter the UUIDs of the workflows. An example workflow could be "Create issue".
Asset workflow	The list of workflows that is available on an asset page in the app. Enter the UUIDs of the workflows. An example workflow could be "Ask the expert".
No search result workflows	The workflows that are available if there are no search results found. Enter the UUIDs of the workflows. An example workflow could be "Propose new business term".
Enable autostart	Option to automatically start Collibra for Desktop when signing in to your operating system. • ✔ True: The app starts automatically when signing in to your operating system. • ✖ False (default): The app does not start automatically. If you have set this option in the Collibra for Desktop settings, this option is neglected.

Everywhere Mobile configuration

These settings determine some of the ways in which Collibra for Mobileinteracts with Collibra Data Intelligence Cloud.

Setting	Description
Default search filter	The filter that is applied, by default, to your search results. The value must be the UUID of the filter. This filter overrules any search filter that is set in the app. To find the UUID, open the Collibra environment and click in the Search box. Click the name of a search filter. In the address bar you will see the UUID of the filter.

Setting	Description
Custom Search box placeholder	The text that is shown in the search box of Collibra for Mobile before a user enters search text.
Enable Workflows	Option to enable workflows in Collibra for Mobile. This allows you to complete tasks or start a workflow in the app. The available workflows depend on the ones that you add to the Global workflows and Asset workflow configuration.
Global workflows	The list of workflows that is available in the app's main menu. Enter the UUIDs of the workflows. An example workflow could be "Create issue".
Asset workflow	The list of workflows that is available on an asset page in the app. Enter the UUIDs of the workflows. An example workflow could be "Ask the expert".
No search result workflows	The workflows that are available if there are no search results found. Enter the UUIDs of the workflows. An example workflow could be "Propose new business term".

Everywhere Chrome Extension

The settings determine how and where you can use the Everywhere Chrome Extension.

Setting	Description
Domains	Add a web domain, for example of a web application such as Power BI or Tableau, on which the Browser Extension automatically appears as overlay.

Edge

Edge configuration options, when changing an option, you only have to refresh the page that runs your Collibra environment.

Setting	Description
Enable Edge jobs feature (beta)	• ✓ True: Enable the Edge jobs page, this page gives you a overview of all jobs and their status. • ✕ False (default): Disable the Edge jobs page.

Tableau Metadata API

You need the Tableau metadata API to ingest Tableau 2020.2 and newer.

> Warning If you upgrade to Tableau version 2020.2 or newer, but previously synchronized an older Tableau version via the REST API and XML mapping, you have to prepare the migration procedure to prevent losing manually added relations, attributes, tags, comments and stitching results.

Setting	Description
Enable Tableau metadata API	• ✓ True: Tableau metadata API is enabled. This enables you to ingest Tableau 2020.2 or newer into Data Catalog. • ✕ False: Tableau metadata API is disabled. If you ingest Tableau 2020.2 or newer, the ingestion will fail. This prevents data loss of manually added relations and attributes.

Backup configuration management

Setting	Description
Backup service URL	The URL of the backup service.

Job Service (Activities)

Setting	Description
Number of executor threads for the Job Service	The maximum number of threads, or jobs, that the Job Service can run in parallel. Generally speaking, increasingly the number of jobs running in parallel reduces overall processing time. Conversely, it requires more system resources, which can negatively impact performance. It also increases the risk of job conflicts.

Appendix B - Spring Cron syntax

Cron is a software utility that specifies commands to run on a given schedule. This schedule is defined by a Cron pattern, which has a specific syntax that will be described in this section.

> Warning If you create an invalid Cron pattern, Collibra Data Intelligence Cloud stops responding.

> Note By default, Collibra Console uses Spring Cron expressions to schedule back-ups, while you use Quartz Cron expressions, for example, to schedule your mail, LDAP synchronizations, Purge cycles, Tableau and S3 synchronizations or to create a statistics cron map.

The Cron pattern consists of six space-separated fields:

```
<second> <minute> <hour> <day of month> <month> <day of week>
```

Position	Field	Allowed values	Allowed special characters	Examples
1	second	0-59	, - * /	• *10*: at the 10th second. • **/10*: every 10 seconds.
2	minute	0-59	, - * /	• *30*: at the 30th minute. • **/15*: every 15 minutes. • *5/10*: every 10 minutes starting at the 5th minute after the hour
3	hour	0-23	, - * /	• *10*: at 10 o'clock. • *8-10*: at 8,9 and 10 AM. • *6,18*: at 6 AM and at 6 PM.
4	day of the month	1-31	, - * ? / L W	• *3*: on the 3rd day of the month. • *1-4*: every first four days of the month. • *1,15*: the first day of the month and the 15th day of the month.

Position	Field	Allowed values	Allowed special characters	Examples
5	month	1-12 or JAN-DEC	, - * /	• *12*: in December. • *1-3*: every first three months of the year. • *JUL,AUG*: every July and August. Tip The names of the months are not case-sensitive.
6	day of the week	0-7 or MON-SUN where 0 and 7 is Sunday.	, - * ? / L #	• *TUE*: every Tuesday. • *1-5*: every weekday, Monday to Friday. • *MON,WED,FRI*: every Monday, Wednesday and Friday. • *L*: on Sunday, the 7th day of the week. • *1L*: at the last Monday of the month. • *5#3*: on the 3rd Friday of the month. Tip The names of the days are not case-sensitive.

For more information, see the Spring Cron documentation.

Special characters

Character	Description
*	Used to select all values within a field. Example * in the minute field corresponds with every minute.

Character	Description
?	Used to specify something in one of the two fields in which the character is allowed, but not the other, mainly used for days of the week. Example If you want your trigger to fire on a particular day of the month, for example the 10th, but don't care what day of the week that happens to be, you could put "10" in the day-of-month field, and "?" in the day of the week field.
-	Used to specify ranges. Example `10-12` in the hour field means “the hours 10, 11 and 12”.
,	Used to specify additional values. Example `MON,WED,FRI` in the day-of-week field means “the days Monday, Wednesday, and Friday”.
/	Used to specify increments. Example `0/15` in the seconds field means “the seconds 0, 15, 30, and 45”. And `5/15` in the seconds field means “the seconds 5, 20, 35, and 50". You can also leave out the number before /, which is equivalent to having 0 before / . `1/3` in the day-of-month field means “fire every 3 days starting on the first day of the month”.

Character	Description
L	Has different meaning in each of the two fields in which it is allowed. Example The value `L` in the **day-of-month field** means "the last day of the month" - day 31 for January, day 28 for February on non-leap years. You can also specify an offset from the last day of the month, such as "L-3" which would mean the third-to-last day of the calendar month. If you use `L` in the **day-of-week field** by itself, it means "7" or "SUN". But if used in the day-of-week field after another value, it means "the last xxx day of the month" - for example "6L" means "the last Saturday of the month". When using the `L` option, it is important not to specify lists, or ranges of values, because you may get unexpected results.
W	Used to specify the weekday (Monday-Friday) nearest the given day. Example `15W` in the value for the day-of-month field, means the nearest weekday to the 15th of the month: • If the 15th is a Saturday, the trigger will fire on Friday the 14th. • If the 15th is a Sunday, the trigger will fire on Monday the 16th. • If the 15th is a Tuesday, then it will fire on Tuesday the 15th. However if you specify `1W` as the value for day-of-month, and the 1st is a Saturday, the trigger will fire on Monday the 3rd, as it will not 'jump' over the boundary of a month's days. The 'W' character can only be specified when the value in the day-of-month field specifies a single day, not a range or list of days. Tip The 'L' and 'W' characters can also be combined in the day-of-month field to yield 'LW', which translates to *"last weekday of the month"*.
#	Used to specify "the nth" XXX day of the month. Example `6#3` in the day-of-week field means "the third Saturday of the month" (day 6 = Friday and "#3" = the 3rd one in the month). Other examples: `2#1` is the first Tuesday of the month and `4#5` is the fifth Thursday of the month. Note that if you specify `#5` and there is not 5 of the given day-of-week in the month, then no firing will occur that month.

Example

- `0 0 * * * *` = the top of every hour of every day.
- `*/10 * * * * *` = every ten seconds.
- `0 0 8-10 * * *` = 8, 9 and 10 o'clock of every day.
- `0 0 6,19 * * *` = 6:00 AM and 7:00 PM every day.
- `0 0/30 8-10 * * *` = 8:00, 8:30, 9:00, 9:30, 10:00 and 10:30 every day.
- `0 0 9-17 * * MON-FRI`= on the hour nine-to-five weekdays.
- `0 0 0 25 12 ?`= every Christmas Day at midnight, no matter what weekday it is.

Quartz Cron syntax

Cron is a software utility that specifies commands to run on a given schedule. This schedule is defined by a Cron pattern, which has a specific syntax that will be described in this section.

For example, you can create a schedule for LDAP synchronizations, Purge cycles or to automatically send emails using cron patterns. You can also use it to create a Cron map for your statistics.

> Note By default, you use Spring Cron expressions to schedule Collibra Console back-ups.

> Warning If you create an invalid Cron pattern, Collibra Data Intelligence Cloud stops responding.

The Cron pattern consists of six or seven space-separated fields:

`<second> <minute> <hour> <day of the month> <month> <day of the week> <year>`

Position	Field	Mandatory	Allowed values	Allowed special characters	Examples
1	second	Yes	0-59	, - * /	• *10*: at the 10th second. • **/10*: every 10 seconds.
2	minute	Yes	0-59	, - * /	• *30*: at the 30th minute. • **/15*: every 15 minutes. • *5/10*: every 10 minutes starting at the 5th minute after the hour
3	hour	Yes	0-23	, - * /	• *10*: at 10 o'clock. • *8-10*: at 8,9 and 10 AM. • *6,18*: at 6 AM and at 6 PM.

Position	Field	Mandatory	Allowed values	Allowed special characters	Examples
4	day of the month	Yes	1-31	, - * ? / L W	• *3*: on the 3rd day of the month. • *1-4*: every first four days of the month. • *1,15*: the first day of the month and the 15th day of the month. • *L*: on the last day of the month. • *L-3*: on the third-to-last day of the month. • *15W*: on the nearest weekday to the 15th of the month. If the 15th is a Saturday, then the trigger will be on the 14th, if the 15th is a Sunday, then the trigger will be on the 16th. Note If the 1st day of the month is a Saturday, then *1W* corresponds to the 3rd day of the month, since the month is specified in the 5th value of the Cron expression. *LW*: on the last weekday of the month.
5	month	Yes	1-12 or JAN-DEC	, - * /	• *12*: in December. • *1-3*: every first three months of the year. • *JUL,AUG*: every July and August. Tip The names of the months are not case-sensitive.

Position	Field	Mandatory	Allowed values	Allowed special characters	Examples
6	day of the week	Yes	1-7 or SUN-SAT	, - * ? / L #	• *TUE*: every Tuesday. • *2-6*: every weekday, Monday to Friday. • *MON,WED,FRI*: every Monday, Wednesday and Friday. • *L*: on Saturday, the 7th day of the week. • *2L*: at the last Monday of the month. • *6#3*: on the 3rd Friday of the month. Tip The names of the days are not case-sensitive.
7	year	No	empty, 1970-2099	, - * /	• *<empty>*: if your schedule doesn't require a year, you can leave this value empty. • *2021*: in 2021. • *2021-2025*: in the years 2021, 2022, 2023, 2024 and 2025. • *2021,2022,2025*: in the years 2021, 2022 and 2025.

Special characters

Character	Description
*	Used to select all values within a field. Example * in the minute field corresponds with every minute.

Character	Description
?	Used to specify something in one of the two fields in which the character is allowed, but not the other, mainly used for days of the week. Example If you want your trigger to fire on a particular day of the month, for example the 10th, but don't care what day of the week that happens to be, you could put "10" in the day-of-month field, and "?" in the day of the week field.
-	Used to specify ranges. Example `10-12` in the hour field means “the hours 10, 11 and 12”.
,	Used to specify additional values. Example `MON,WED,FRI` in the day-of-week field means “the days Monday, Wednesday, and Friday”.
/	Used to specify increments. Example `0/15` in the seconds field means “the seconds 0, 15, 30, and 45”. And `5/15` in the seconds field means “the seconds 5, 20, 35, and 50". You can also leave out the number before /, which is equivalent to having 0 before / . `1/3` in the day-of-month field means “fire every 3 days starting on the first day of the month”.

Character	Description
L	Has different meaning in each of the two fields in which it is allowed. Example The value `L` in the **day-of-month field** means "the last day of the month" - day 31 for January, day 28 for February on non-leap years. You can also specify an offset from the last day of the month, such as "L-3" which would mean the third-to-last day of the calendar month. If you use `L` in the **day-of-week field** by itself, it means "7" or "SAT". But if used in the day-of-week field after another value, it means "the last xxx day of the month" - for example "6L" means "the last Friday of the month". When using the `L` option, it is important not to specify lists, or ranges of values, because you may get unexpected results.
W	Used to specify the weekday (Monday-Friday) nearest the given day. Example `15W` in the value for the day-of-month field, means the nearest weekday to the 15th of the month: • If the 15th is a Saturday, the trigger will fire on Friday the 14th. • If the 15th is a Sunday, the trigger will fire on Monday the 16th. • If the 15th is a Tuesday, then it will fire on Tuesday the 15th. However if you specify `1W` as the value for day-of-month, and the 1st is a Saturday, the trigger will fire on Monday the 3rd, as it will not 'jump' over the boundary of a month's days. The 'W' character can only be specified when the value in the day-of-month field specifies a single day, not a range or list of days. Tip The 'L' and 'W' characters can also be combined in the day-of-month field to yield 'LW', which translates to *"last weekday of the month"*.
#	Used to specify "the nth" XXX day of the month. Example `6#3` in the day-of-week field means "the third Friday of the month" (day 6 = Friday and "#3" = the 3rd one in the month). Other examples: `2#1` is the first Monday of the month and `4#5` is the fifth Wednesday of the month. Note that if you specify `#5` and there is not 5 of the given day-of-week in the month, then no firing will occur that month.

Example

- `0 0 * ? * * *` = the top of every hour of every day.
- `*/10 * * * * ?` = every ten seconds.
- `0 0 8-10 * * ? 2020` = 8, 9 and 10 o'clock of every day during the year 2020.
- `0 0 6,19 ? * *` = 6:00 AM and 7:00 PM every day.
- `0 0/30 8-10 ? * *` = 8:00, 8:30, 9:00, 9:30, 10:00 and 10:30 every day.
- `0 0 9-17 * * MON-FRI` = on the hour nine-to-five weekdays.
- `0 0 0 25 12 ?` = every Christmas Day at midnight, no matter what day of the week it is.
- `0 15 10 ? * 6L 2022-2025` = 10:15 AM on every Friday of every month during the years 2022, 2023, 2024 and 2025.
- `0 30 11 ? * 6#2` = 11:30 AM on the second Friday of every month.

Warning Quartz Cron only supports a value in either the 4th or the 6th position, but not in both. At the same time, both positions cannot be empty.