Collibra

Collibra Data Governance Center

# Platform Installation

Version 5.7

Collibra

Collibra Data Governance Center - Platform Installation

Release date: February 28, 2022

Revision date: Fri Mar 04, 2022

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

https://productresources.collibra.com/docs/collibra/latest/Content/Installation/to_installation.htm

# Contents

Chapter 1

# Product architecture

## In this chapter

# Collibra DGC overview

In this section, you will learn more about the Collibra DGC services, their purpose and how they communicate with each other.

## Global overview

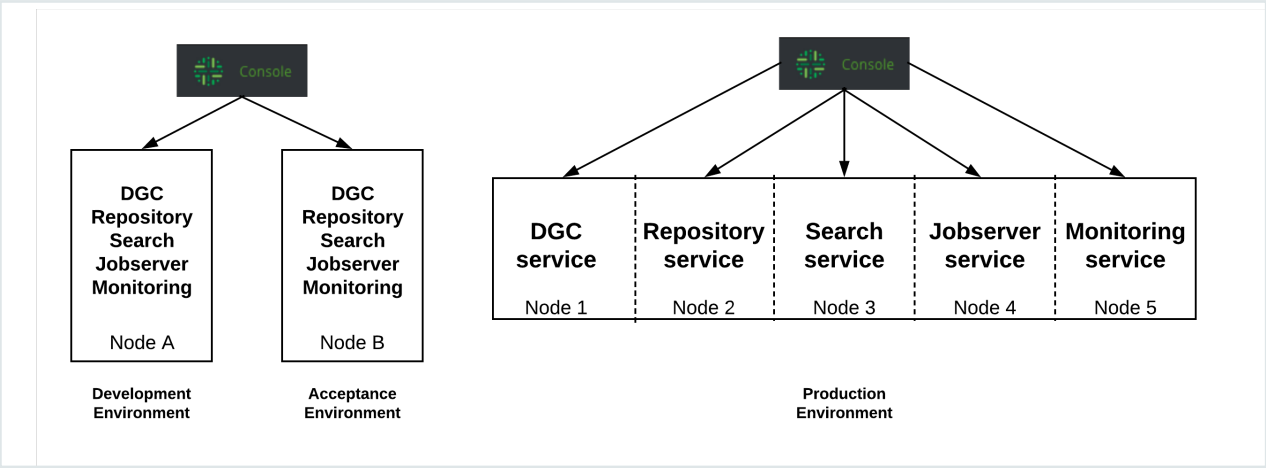
Collibra Data Governance Center consists of the following components:

- Services:
  - DGC service
  - Repository service
  - Search service
  - Jobserver service
  - Monitoring service
- Collibra Console

With Collibra Console, you can monitor and maintain multiple Collibra DGC environments. You can also install multiple Collibra Console instances, each managing a different set of environments.

> **Note** The version of Collibra Console and your environments must be identical.

In the following schema, you see a typical enterprise setup, with two Collibra Console instances. One manages a development and an acceptance environment, the other manages a production environment:




In the schema, you see that in the development and acceptance environment, all services run on the same node. In the production environment, the services run on separate nodes. You can also run Collibra Console on a separate node, which is recommended, but you can also choose to run it on a node in combination with one of the services

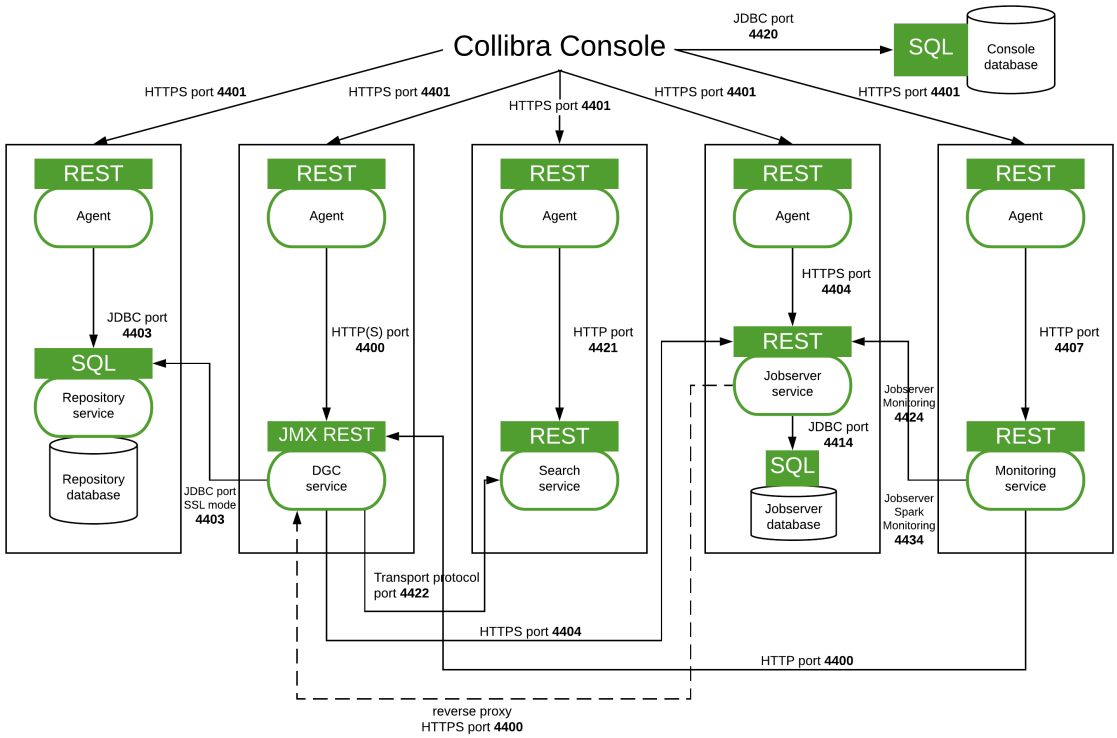
Every node of an environment has a running Agent. The Agent is the communication bridge between Collibra Console and the Collibra services.

> **Tip** It is not necessary to run each service on a single node. We recommend that you install the Repository service and Jobserver service on dedicated nodes. You can install all other services on another node, provided that it has enough memory. Make sure that you have a fast network between the nodes.

# Internal communication

In the following schema, you can see the communication paths between the different Collibra Data Governance Center components.

> **Note** HTTPS is always used for Cloud environments. For on-premises environments, HTTPS is only used if you enable it.
> The Agent and all services are all separate processes.



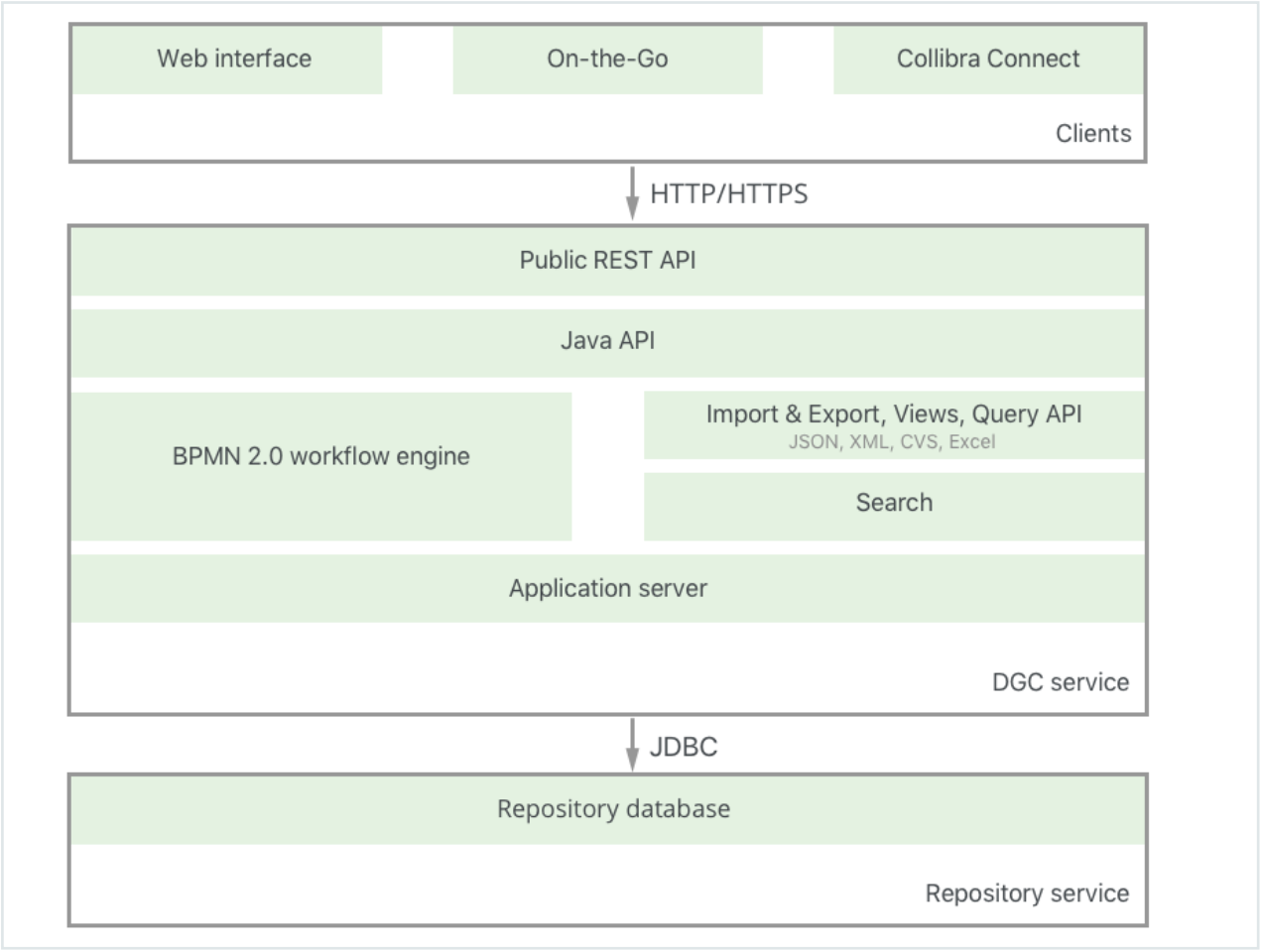

| Communication path | Description |
|---|---|
| • Agent → Repository service<br>• DGC service → Repository service | Send SQL statements over JDBC, via port 4403. |
| Agent → DGC service | Send management-specific commands with a private REST interface (JMX REST) over HTTP, via port 4400. |
| DGC service → Search service | Send search requests using the Transport protocol via port 4422. |
| Agent → Search service | Send management specific commands with a private REST interface over HTTP(S), via port 4421 |

| Communication path | Description |
| --- | --- |
| Collibra clients → Collibra Data Governance Center | Access to Collibra DGC with the public REST interface (REST) over HTTPS. |
| • Agent → Jobserver service<br>• DGC service → Jobserver service | Send job commands using a REST interface over HTTPS, via port 4404. |
| Jobserver service → DGC service | Poll the DGC service if there are any jobs to be executed. This uses the built-in reverse proxy servlet of the DGC service.<br>If this type of communication is set up, the DGC service does never initiate a communication to the Jobserver. |
| Jobserver service → Jobserver database | Send SQL statements over JDBC, via port 4414. |
| • Agent → Monitoring service<br>• Monitoring service → DGC service | Send job commands using a REST interface over HTTPS, via port 4407.<br>The monitoring service connects to the DGC service via port 4400. |
| Console → Console database | Send SQL statements over JDBC, via port 4420 |

# Collibra DGC services

In the following schema, you can see the three-tiered architecture of Collibra Data Governance Center. It is a web application, fully implemented in Java and it is platform independent.

> **Note** HTTPS is always used for Cloud environments. For on-premises environments, HTTPS is only used if you enable it. Note that for some clients HTTPS is mandatory, such as On-the-Go for iOS and macOS.




All the services and their components, such as the application server, repository database and APIs, are prepackaged in the Collibra DGC installer. During the installation, these components are all automatically configured to make the installation experience as smooth as possible. Collibra also provides support for all installed components.

> **Note** Collibra Everywhere and Collibra Connect are not included with the Collibra DGC installer because they are separate Collibra products.

## Collibra DGC environment data storage

All the data of a Collibra Data Governance Center environment is stored in two locations:

- The actual data (communities, assets, domains, users, comments, ...) of Collibra DGC is stored in the repository.
- The Collibra DGC software metadata (temporary files, log files, license file) is stored in the Collibra **data** directory on the local file system.
    - Default location on Linux as root or user with sudo privileges: **/opt/collibra_data**
    - Default location on Linux as standard user: **~/collibra_data**
    - Default location on Windows Server: **C:\collibra_data**

      > **Important** On Windows, that target data directory cannot contain spaces.

The Collibra **collibra_data** directory contains a subdirectory for every service that is installed on the server.

- agent: Agent service
- console: Collibra Console
- dgc: Data Governance Center service
- monitoring: Monitoring service
- repo: Repository service
- search: Search service
- spark-jobserver: Jobserver service

# Collibra DGC clients

To support multiple client types, the client side is separated from the back-end side. The back-end web application layer provides a public API for Java and REST.

This means that the packaged clients communicate with the backend through the same API as custom-written applications would. This way, the full power of Collibra DGC is available to any external application.

The following clients are delivered out of the box:

- **Collibra DGC Web Interface**: This is the standard way to interact with Collibra DGC. It connects with the REST API over HTTP(S). Almost all functions that are supported by the server application can be accessed from the web interface. The web interface is supported by the usual web browsers.
- **Collibra Everywhere**: Purpose-built applications that allow you to access your organization's trusted, governed data wherever they happen to be working.

# DGC service

## Architecture

The Data Governance Center service is a web application containing the business logic of Collibra Data Governance Center. It is fully developed in Java.

The DGC service has the following major components:

| Component | Description |
|---|---|
| Java API | The Java API contains the business logic like the CRUD (Create, Read, Update, Delete) operations on assets. It is accessible through an open and fully-documented Java API. |
| REST API | All the methods that are available in the Java API are also available in the REST API. The REST API makes it easy to connect from external applications (browser clients, desktop clients, other servers) to the DGC service. The REST API is implemented with the Jersey REST server and is fully documented. All the Collibra Client applications use the same public REST API that is available to customers. |
| Import/Export, Views, Query APIs | These APIs make it very easy to effectively access and manipulate application data in Collibra DGC. Collibra DGC supports various file formats such as JSON, XML, CSV and Excel. Because all of these methods are available through the Java and REST API, you can easily integrate them with other applications by using external tools, such as ETL (Extract, Transform, Load) or ESB (Enterprise Service Bus) middleware applications. |
| Search | With the search API, you can search for specific application data in Collibra DGC. For example, Collibra Everywhere uses the search API to let the user find assets in Collibra DGC. |

| Component | Description |
|---|---|
| BPMN 2.0 Workflow Engine | The workflow engine, Flowable, supports the execution of BPMN 2.0 (Business Process Model and Notation) processes. The prepackaged workflows are completely configurable and it is very easy to add, modify, and deploy the workflows to support the governance processes that best suit your organization. Worfklow service tasks can use the available Java API, which enables you to automate various application tasks, like email notification, creating comments, adding assets, and so on. |

# Data storage

The data of the DGC service is located in:

- Linux with root permission: **/opt/collibra_data/dgc**
- Linux without root permission: **~/collibra_data/dgc**
- Windows: **C:\collibra_data\dgc**

In the data directory of the DGC service, you can find several files and directories serving different purposes:

| Directory or file | Purpose |
|---|---|
| cache | Contains the cache files of the DGC service. |
| collibra.license | Contains the authorized applications and allowed number of users for Collibra DGC. |
| config | Contains the configuration files used by Collibra DGC. |
| email-templates | Overrides the built-in email templates to customize the emails that are sent to the users. |
| groovy-lib | Contains additional Groovy library functions to be used in validation rules. |
| images | Contains images that can be referenced directly as a URL, for example to set another logo. |
| indexes | Contains the search index files for quickly searching Collibra DGC content. If the directory is not present when Collibra DGC is started, it is automatically created. |

| Directory or file | Purpose |
|---|---|
| logs | Contains the log files of the DGC service. The log files are important to troubleshoot possible problems in the product. |
| modules | Custom UI modules to extend or override the existing UI. |
| page-definitions | Overrides page definitions to customize pages in Collibra DGC. |
| security | Contains security related files such as SSL and SAML metadata files. You should not change anything in this directory. |
| styling | Overrides the CSS styling of the web interface. |
| translations | Used to override the built-in user interface labels or to add new languages.<br>**Tip** You can also do this in the **Settings** section of Collibra DGC. |

# Repository service

## Architecture

The Repository service is a PostgreSQL database that is managed and maintained by Collibra Console and its agents.

The repository settings are configured during the initialization phase and Collibra Console takes care of doing periodic maintenance operations on the database. You can use the back up and restore feature to restore the repository to a given moment in time.

Monitoring, logs and diagnostics are available to help troubleshoot possible issues.

## Data storage

The data of the Repository service is located in:

- Linux with root permission: **/opt/collibra_data/repo**
- Linux without root permission: **~/collibra_data/repo**
- Windows: **C:\collibra_data\repo**

It contains the following subfolders:

| Directory name | Content |
|---|---|
| config | The configuration of the repository database. It overrules matching configurations in the **postgresql.conf** file. |
| logs | All the log files created by the Repository service. |
| data | The actual repository data. Therefore it is the most important directory of the repository. Make sure that this folder is secured against file system failures and other possible defects.<br>This folder also contains the PostgreSQL configuration file, **postgresql.conf**. This is the main configuration file of PostgreSQL. However, the actual settings of the database are saved in the **config** directory (**collibra_data/repo/config/configuration.json**).<br>Settings that are not defined in **configuration.json** are taken from the **postgresql.conf** file. |

# Search service

## Architecture

The Search service is built on Elasticsearch and is configured and maintained by Collibra Console and an agent.

## Data storage

The data of the Search service is located in:

- Linux with root permission: **/opt/collibra_data/search**
- Linux without root permission: **~/collibra_data/search**
- Windows: **C:\collibra_data\search**

It contains the following subfolders:

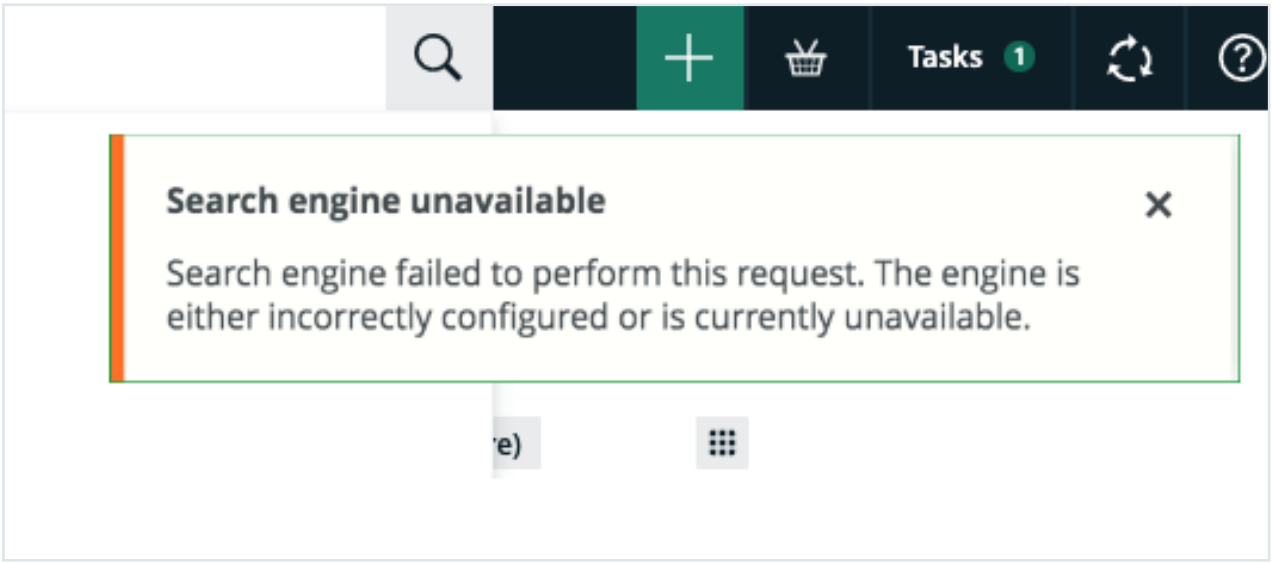
| Directory name | Content |
|---|---|
| config | The configuration of the Search service. |
| logs | All the log files created by the Search service. |

| Directory name | Content |
|---|---|
| data | The actual search database. |
| tmp | The directory where the service saves its temporary files. |

# Synchronization with the database

The Search service is mandatory for every Collibra DGC environment, and generally speaking, the Search service has to be synchronized with the database in order for you to search for and work with Collibra DGC resources. In several situations, for example after editing search or hyperlink settings or upgrading your Collibra DGC environment, you have to synchronize the Search service with the database. Synchronization is achieved by reindexing Collibra DGC.

If, by chance, the Search service is incorrectly configured or becomes unavailable, a Search engine unavailable warning is shown, but Collibra DGC will still continue to function.




Users can still create, edit and delete resources; however, such activity will result in an asynchronous state between the Search service and the database. When the Search service becomes available again, if no resources were created, edited or deleted, there is no need to reindex Collibra DGC. If, however, such changes were made in the environment, a full reindex is necessary. Your Collibra DGC administrator is notified via a warning, as shown in the following image.


# Monitoring service

## Architecture

The Monitoring service is configured and maintained by Collibra Console and an agent.

It will monitor the Data Governance Center service, such as JVM and memory usage, and store the data in its database. The monitoring interval can be configured in Collibra Console.

The monitoring data can be saved by creating a diagnostics file.

## Data storage

The data of the Monitoring service is located in:

- Linux with root permission: **/opt/collibra_data/monitoring**
- Linux without root permission: **~/collibra_data/monitoring**
- Windows: **C:\collibra_data\monitoring**

It contains the following subfolders:

| Directory name | Content |
|---|---|
| config | The configuration of the Monitoring service. |
| logs | All the log files created by the Monitoring service. |
| data | The data that is stored by the Monitoring service. |

# Jobserver service

## Architecture

The Jobserver is used to ingest data and to execute data profiling or to create sample data on the ingested data. You can ingest data when you register a data source.

It is an application that relies on Apache Spark to perform CPU and memory intensive computations quickly and efficiently. More specifically, the Jobserver acts as an interface between the Data Governance Center service and Spark, sending Spark job execution requests through a REST interface. The Jobserver also provides control over the single Spark jobs and the data used by Spark.

When running a profiling operation, the Jobserver starts a new Java Virtual Machine (JVM), running a Spark Context. The profiling operations are executed within this JVM and returns to the DGC service through the main Jobserver application.

Only one profiling operation can run at a time. If there are several profiling operations, they are queued for execution.

## Data storage

The Jobserver must be installed on a dedicated server and is managed by Collibra Console through an agent.

The data of the Jobserver service is located in:

- Linux with root permission: **/opt/collibra_data/spark-jobserver**
- Linux without root permission: **~/collibra_data/spark-jobserver**
- Windows: **C:\collibra_data\spark-jobserver**

It contains the following subfolders:

| Directory name | Content |
|---|---|
| logs | All the log files created by the Jobserver service. |
| data | The data used by the Jobserver service during runtime, it does not contain any critical state for the application to maintain. |

| Directory name | Content |
|---|---|
| config | The data of the Jobserver memory and CPU usage. |
| security | The public and private keys needed to use SSL encryption when communicating with the Jobserver REST API. |
| pgsql-data | The data that is stored by the Jobserver service, such as job information and the JAR files to register data sources. |
| spark-warehouse | The directory where the Spark tables are persisted. |
| temp-files | The directory to store temporary files during ingestion and profiling jobs. |

# Jobserver memory and CPU usage

The most demanding jobs in terms of computing resources are the ingestion and profiling processes. You should make sure to provide the minimum requirements to perform ingestion and profiling successfully.

The Jobserver and the Spark Context run in two separate Java Virtual Machines, which means that the memory is shared between them.

We highly recommend you to install the Jobserver on a dedicated server. However, if you install the Jobserver on the same server as other Collibra nodes, the minimum hardware requirements of the Jobserver must be added to those of the other Collibra nodes on the same server.

## Ingestion

During the ingestion of a schema, the schema is analyzed by a process of the Jobserver Spark context, split in pages and then sent to the Jobserver page per page. Each page it is stored in memory until Collibra DGC fetches it.

## Profiling

The amount of data processed per table is limited to a certain threshold. You can customize this threshold in the Data Governance Center service configuration with a maximum of 10 GB of disk space. The profiling restarts on a subset of the data when that threshold is reached. It extracts a random subset of an approximate size defined by the threshold. This gives an upper estimation of the largest data set the Spark context may have to process. Taking into account that data size in memory is larger than on disk, we consider a heap size of 40 GB.

> **Warning** If you are using an installation of JobServer older than the 5.7.2 version, you may experience memory errors. See the Troubleshooting section for more information.

### CPU usage

In the `spark` section of the **jobserver.conf** file, located in **/opt/collibra/spark-jobserver/conf/**, the `local[N]` parameter determines how many CPUs can be used by Jobserver Spark for profiling. The original setting (`local[*]`) enables the usage of all the CPUs available to the machine. We recommend keeping the original settings for best performance.

# Collibra Console

## Architecture

The management console is a Java web application. It runs independently and does not rely on any other components. It uses a local file-based database to store some of its data, like the environment configuration and the user information.

## Data storage

The data of the Collibra Console is located in:

- Linux with root permission: **/opt/collibra_data/console**
- Linux without root permission: **~/collibra_data/console**
- Windows: **C:\collibra_data\console**

Collibra Console contains the following data directories:

| Directory or file | Purpose |
|---|---|
| cache | Contains the cache files of Collibra Console. |
| config | Contains the configuration files used by Collibra Console. |
| email-templates | Overrides the built-in email templates to customize the emails that are sent to the users. |
| images | Contains images that can be referenced directly as a URL, for example to set another logo. |
| logs | Contains the log files of Collibra Console. The log files are important to troubleshoot possible problems in Collibra Console. |
| modules | Custom UI modules to extend or override the existing UI. |
| page-definitions | Overrides page definitions to customize pages in Collibra Console. |
| pgsql-data | The data that is stored by Collibra Console. It contains the information of all resources that are managed by Collibra Console, such as environments and nodes. |
| security | Contains security related files such as SSL and SAML metadata files. You should not change anything in this directory. |
| styling | Overrides the CSS styling of the web interface. |
| translations | Used to override the built-in user interface labels or to add new languages. |
| backups | All the backups that are taken via Collibra Console. The folder size depends on the number of backups and the database size. |

# Security architecture

Collibra Data Governance Center is by default a very secure application, but it offers some extra configuration options to make Collibra DGC meet your organization's security level.

For more information about setting the security options, see Security configuration.

# Passwords

User passwords are never stored in plain text, nor in encrypted form. Only a salted hash is stored in the database, so that users can be authenticated:

- The salt is constructed with a fixed salt, concatenated with a private salt. The private salt changes constantly and is stored in the database next to the hash.
- The SHA-512 algorithm is used for hashing.
- 10 000 iterations are executed to get an extremely secure hash.

This means that passwords cannot be recovered in any way.

The password specifications are set in Collibra Console. The passwords are encrypted the first time they are read, which is during server start-up.

This way, you can type passwords in plain text but you prevent malicious use.

There are no other locations in the product where passwords or other credentials are stored.

# Cookies

Collibra DGC uses cookies for some of its functions:

- **JSESSIONID**: This cookie is used to store the ID of the session between the client and the server. This session expires as soon as the browser session ends and is not accessible with scripts.
- **rememberMe**: This cookie is used to remember a user on a certain browser. This cookie is only set when a user selects the **Remember me** option when signing in. It has an expiration time of one year and is not accessible with scripts. In the cloud environment, this cookie is also secure.

When supported and enabled in the browser, the Collibra DGC web interface uses local storage to improve performance.

# Sign-in attempts

All the sign-in attempts to Collibra DGC or Collibra Console, whether they are successful or not, are tracked and stored in dgc.log and console.log respectively. Each entry contains the

following information:

- ID of the user
- Status of the sign-in attempt
- Source IP address of the sign-in attempt
- Time of the sign-in attempt

**Example**

```
2020-07-28 07:27:36.467 [http-nio-0.0.0.0-4400-exec-10]
INFO c.c.d.c.a.s.DgcAuthenticationStrategy -
Authentication for username [edafaefeafaef] & realm
[DGCRealm] failed.
```

# Audit logging

For all the resources in the system (communities, domains, assets, attributes, comments, user information, meta model, and so on), Collibra Data Governance Center stores the full history in the database.

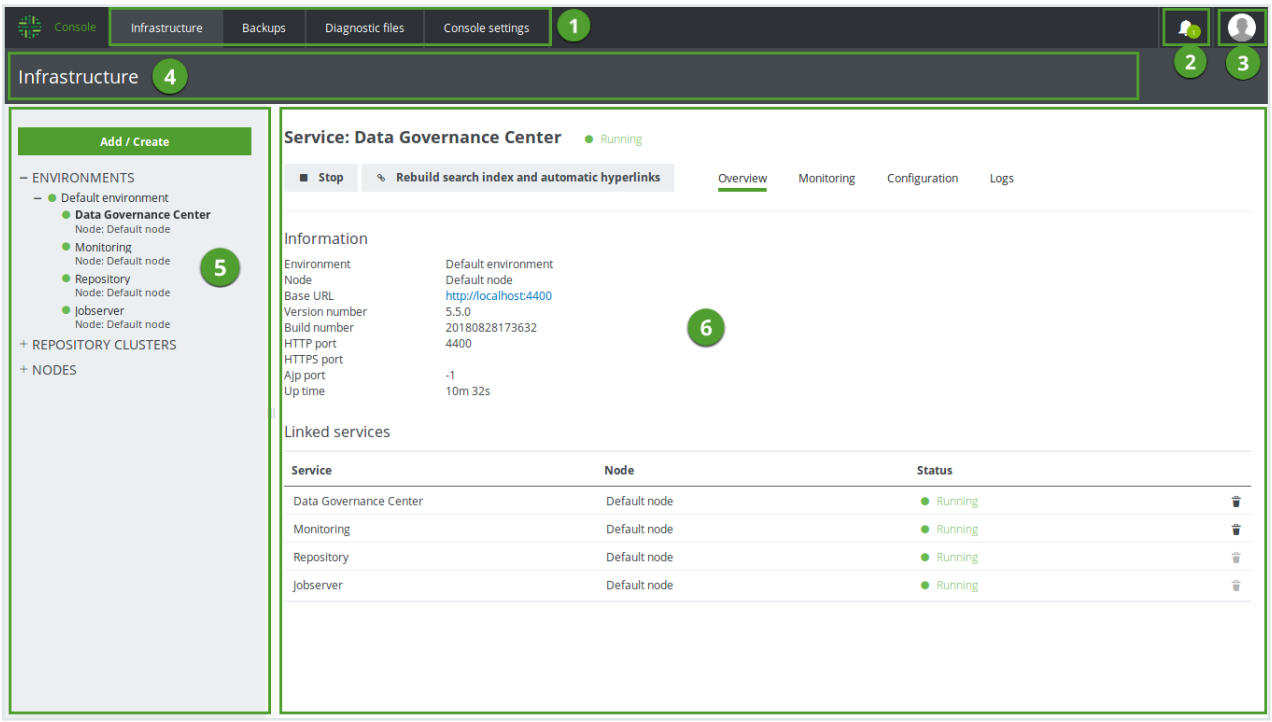
This means that for every change made in Collibra DGC, you can consult the following information:

- Who made the change (which user).
- When was the change made.
- What was done as part of the change.

Next to the above, it is possible to enable detailed logging for each action taken in the product (both updates and data retrievals). For more information, see Logging.

# Navigating in the Console user interface

The Collibra Console user interface contains a number of common elements that appear throughout the product.




| No. | Element name | Description |
|---|---|---|
| 1 | Main menu | Provides access to the following:<br>• Infrastructure: Environments and nodes<br>• Backups<br>• Diagnostic files<br>• Collibra Console settings |
| 2 | Notification center | Stores the notifications you receive while working with Collibra Console. |

| No. | Element name | Description |
|---|---|---|
| 3 | Console user profile | Enables you to view your username and to sign out. |
| 4 | Title bar | Shows the name of the selected **Main menu** item. |
| 5 | Tab pane | Shows a list of items, depending on the selected **Main menu** item. |
| 6 | Table | Shows relevant resources in a structured way in table format. Depending on the resource, you can find menus, action buttons or resource details above the table. |

# Notification center

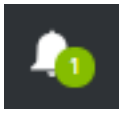
The Notification center stores the notifications you receive while working with Collibra Console.

The bell icon accompanied by a number indicates you have that number of unread notifications.

- A number in a red circle indicates an unread error notification:

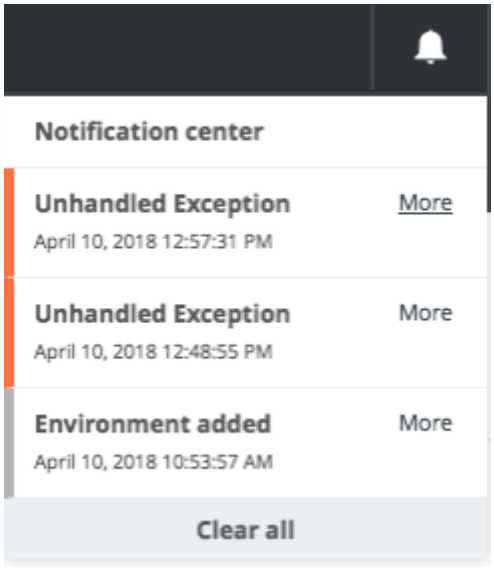
- All other notifications are indicated by a number in a green circle:

> **Note** If you have a number of unread notifications and at least one of them is an error notification, the number is shown in a red circle.

## Actions

- Click the bell icon to view all notifications.
  » All notifications are considered read and the notification count is removed from the bell icon.

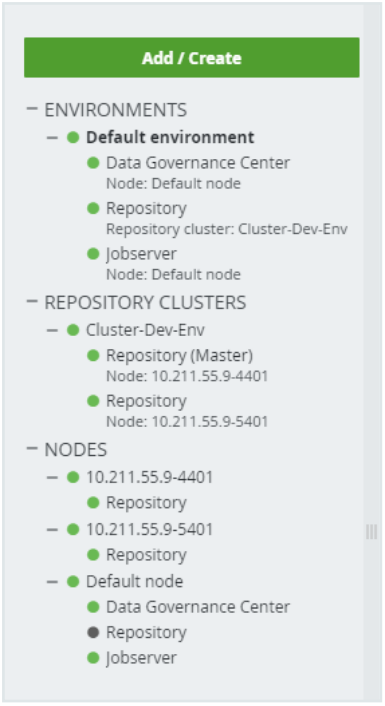



- Click **More** for more information on a specific notification.
- Click **Clear all**, to permanently remove all notifications.

# Infrastructure tree

The infrastructure tree on the **Infrastructure** page of Collibra Console contains the environments, repository clusters, nodes and services in a visual, logic representation.

Example of an infrastructure tree:




# Tree structure

> **Note** The colored dot in front of each item (environment, cluster, node, service) in the tree indicates whether the item is running (green) or not (gray). The red dot mean that there is an issue with that specific item.

| | |
|---|---|
| – ENVIRONMENTS | This section contains all environments managed by this instance of Collibra Console.<br><br>Click **ENVIRONMENTS** to open the list of environments.<br><br>You can collapse and expand this section by clicking – or +. |

| | | |
|---|---|---|
| − <Environment> | | This section contains the services and nodes of a specific environment.<br><br>Click it to show the environment details.<br><br>You can collapse and expand this section by clicking − or + in front of the environment name. |
| | **Data Governance Center**<br>**Node:** <node of the service> | The DGC service, followed by the node on which the service runs. |
| | **Repository**<br>**Node:** <node of the service> | The Repository service, followed by the node on which the service runs or by the repository cluster. |
| | **Jobserver**<br>**Node:** <node of the service> | The Jobserver service, followed by the node on which the service runs.<br><br>This service is not always present. |
| − **REPOSITORY CLUSTERS** | | This section contains all repository clusters managed by this instance of Collibra Console.<br><br>Click **REPOSITORY CLUSTERS** to show the list of clusters.<br><br>You can collapse and expand this section by clicking − or +. |
| − <cluster> | | This section contains the repository services of a specific cluster.<br><br>Click the cluster name to show the its details.<br><br>You can collapse and expand this section by clicking the − or + in front of the cluster name. |

| | |
|---|---|
| Repository (Master) | A Repository service of the cluster. The master repository service has the postfix **(Master)**, all others are slaves.<br><br>Click a repository service to show its details. |
| − NODES | This section contains all nodes managed by this instance of Collibra Console.<br><br>Click **NODES** to show the list of nodes.<br><br>You can collapse and expand this section by clicking − or +. |
| − <Node> | This section contains the services on a specific node.<br><br>Click it to show the node details.<br><br>You can collapse and expand this section by clicking the − or + in front of the node name. |
| Data Governance Center | The DGC service.<br><br>Click it to show the service details.<br><br>This service is not always present, but there could also be more than one. |
| Repository | The Repository service.<br><br>Click it to show the service details.<br><br>This service is not always present, but there could also be more than one. |
| Jobserver | The Jobserver service.<br><br>Click it to show the service details.<br><br>This service is not always present, but there could also be more than one. |

# Getting Collibra DGC up and running

This section describes how to install and configure Collibra Data Governance Center. If you are running an old version of Collibra DGC, see the upgrade section.

## In this chapter

# System requirements

Before installing an on-premises Collibra Data Governance Center environment, check if your nodes meet the hardware and software requirements.

If you use Collibra Data Lineage, check the system requirements that you need to install the lineage harvester.

**Tip**

- We recommend that you install the Repository service and Jobserver service on dedicated nodes. You can install all other services on another node, provided that it has enough memory. Make sure that you have a fast network between the nodes.
  For test environments, you can install all the services on one node.
- The system requirements in this setting are only meant for on-premises environments. The system requirements for a cloud environment are described in Collibra Data Intelligence Cloud infrastructure.

## Supported operating systems

**Note**

- Only 64-bit operating systems are supported.
- Linux operating systems are recommended over Windows operating systems.
- Windows Administrator rights with full rights on the intended installation drive/directories are mandatory.

### Linux operating systems

- Red Hat Enterprise Linux/CentOS 6.x
- Red Hat Enterprise Linux/CentOS 7.x
- Debian 9
- Ubuntu 16.x

- Ubuntu 18.x
- Suse 12

> **Note**
> - You have to set the locale to *en_US.UTF-8* on all Linux systems.
> - Root permissions are not mandatory but preferred.
>
> If you install the Jobserver without root permissions, see the services section.

## Microsoft Windows operating systems

- Windows Server 2012 R2
- Windows Server 2016
- Windows Server 2019

# Minimum system requirements

| Service | Minimum system requirements |
|---|---|
| DGC | • 2 GB RAM<br>• 50 GB free disk space |
| Repository | • 1 GB RAM<br>• 100 MB free disk space for the installation.<br>• 125 GB free disk space for the data. |
| Jobserver | • Data ingestion<br>  ◦ 64 GB RAM<br>  ◦ 500 GB free disk space<br>  ◦ Hard disk type: SSD<br>  ◦ Number of CPUs: 16<br>• Tableau ingestion<br>  ◦ 6 GB RAM<br>  ◦ 35-50 GB free disk space<br>  ◦ Hard disk type: SSD<br>  ◦ Number of CPUs: 4<br><br>We highly recommend you to install the Jobserver on a dedicated server. However, if you install the Jobserver on the same server as other Collibra nodes, the minimum hardware requirements of the Jobserver must be added to those of the other Collibra nodes on the same server. |

| Service | Minimum system requirements |
|---|---|
| Monitoring | • 3 GB RAM<br>• 1 GB free disk space for the installation of the service.<br>• At least 10 GB free disk space for the monitoring database. The disk space to store the data depends on the configuration of the data retention time, so the disk space needs to be monitored. |
| Search | • 1 GB RAM<br>• Solid state disks or high-performance server disks (15k RPM drives) to store the search index.<br>• For more hardware requirements, see the Elasticsearch website.<br>• Before the Search service will be installed, there are bootstrap checks. If one or more checks fail, the service will not be installed and will cancel the complete installation process. |
| Agent | • 512 MB RAM<br>• 1 GB free disk space |
| Console | • 1 GB RAM<br>• 1 GB free disk space for the installation.<br>• 1 GB free disk space for the Collibra Console database.<br>• Extra free disk space to store the backups. |
| Lineage harvester | • Java Runtime Environment version 8 or newer.<br>• 2 GB RAM<br>• 1 GB free disk space<br><br>**Note** To install and use the lineage harvester, you first have to purchase Collibra Data Lineage. This feature is only available for customers that use Collibra Data Intelligence Cloud version 5.7.3 or newer. |

| Service | Minimum system requirements |
| --- | --- |
| Power BI harvester | • Microsoft .NET Framework 4.7.2.<br>• One of the following:<br>  ◦ Client operating system: Windows 7 SP1, 8.1 or 10, version 1607.<br>  ◦ Server operating system: Windows Server 2008 R2 SP1.<br>• 2 GB RAM<br>• 1 GB free disk space<br><br>**Note** To install and use the Power BI harvester, you first have to purchase the Power BI integration feature. This feature is only available for customers that use Collibra Data Intelligence Cloudversion 2020.11 or newer. |

## Other requirements

- Use the same installer version to install each service.
- If you install each service on a dedicated server, ensure that the servers can communicate with each other over SSL.
- For the installation, you need 5 GB of free disk space in the temporary folder.

# Recommended system requirements

The minimum system requirements are most likely insufficient for production environments. In the next sections, you can find some guidelines for the system requirements per Data Governance Center service.

> **Note** These guidelines are only recommendations. Ultimately, the performance of a Collibra DGC environment depends on many additional factors, for example, network performance, load balance and data volume.

## DGC service

| Number of concurrent users | Recommended CPU cores | (*) Recommended memory (GB) |
| --- | --- | --- |
| 1 -15 | 4 | 8 |

| Number of concurrent users | Recommended CPU cores | (*) Recommended memory (GB) |
|---|---|---|
| 16 - 50 | 4 | 12 |
| 51 - 100 | 12 | 12 |
| 101 - 200 | 16 | 16 |

(*) The amount of memory indicates the memory dedicated to the DGC service.

CPU power influences the performance of transactions, while memory has a significant influence on data imports.

The recommended CPU cores and memory can be different according to your usage:

- If you use Collibra DGC for reference data management, you can use more CPU cores than indicated in the table.
- If you regularly import large amounts of data, we recommend that you increase the memory.

The DGC service does not store any real Collibra DGC data, therefore the disk space usage of this component is more stable. We recommend at least 50 GB.

## Repository service

| Number of concurrent users | Recommended CPU cores | (*) Recommended memory (GB) |
|---|---|---|
| 1 -15 | 4 | 8 |
| 16 - 50 | 4 | 12 |
| 51 - 100 | 6 | 12 |
| 101 - 200 | 10 | 16 |

(*) The amount of memory indicates the memory dedicated to the repository service.

The recommended CPU cores and memory can be different according to your usage:

- If you have a large amount of assets in your database, the repository requires more RAM memory. The more RAM available to the operating system, the greater the role the file

system cache plays in storing the data.
- If the size of your content is large, you will need more disk space. For example, the history of performed actions in the system is stored in the database.

We recommend that you start with 125 GB of disk storage and monitor the usage.

## Search service

| Number of assets | Search service memory |
|---|---|
| < 500k | 1 GB |
| < 1M | 2 GB |
| > 1M | 4 GB |

Rule of thumb for assigning memory to the search service:

```
#million assets x 2 = GB of memory
```

For example, for 3 million assets in the repository, assign 6 GB of memory.

## Jobserver service

The recommended requirements are identical to the minimum requirements:

- Total memory: 64 GB
- Total free disk space: 500 GB
- Hard disk type: SSD
- Total #CPUs: 16

We highly recommend you to install the Jobserver on a dedicated server. However, if you install the Jobserver on the same server as other Collibra nodes, the minimum hardware requirements of the Jobserver must be added to those of the other Collibra nodes on the same server.

## Collibra Console

Collibra Console requires free disk space as all backups are stored on the node with the Collibra Console component.

## Collibra Agent

The agent works perfectly when you take only the minimum system requirements into account. There's no memory scaling required as its memory consumption remains stable.

## Lineage harvester

The recommended software requirements are identical to the minimum software requirements. However, the minimum hardware requirements are most likely insufficient for production environments. We recommend to provide the following hardware requirements:

- 4 GB RAM
- 20 GB free disk space

> **Note** To install and use the lineage harvester, you first have to purchase Collibra Data Lineage. This feature is only available for customers that use Collibra Data Intelligence Cloud version 5.7.3 or newer.

## Power BI harvester

We recommend to provide the following system requirements:

- 4 GB RAM
- 20 GB free disk space
- Microsoft .NET Framework 4.7.2 or higher.
- Client operating system: Windows 10 April 2018 update, version 1803 or newer.
- Server operating system: Windows Server 2016 version 1803 or newer.

> **Tip** To ingest Power BI metadata in Data Catalog, you need to run two different harvesters: the Power BI harvester and the lineage harvester.

> **Note** To install and use the Power BI harvester, you first have to purchase the Power BI integration feature. This feature is only available for customers that use Collibra Data Intelligence Cloudversion 2020.11 or newer.

## Supported web browsers

> **Important** Beginning September 2021, Collibra will no longer support Internet Explorer 11 to ensure security and optimize platform performance. Although you will still be able to use Collibra DGC on IE11, we will deprecate our support for IE11 and recommend you move to another supported browser.

| Browser | Version |
|---|---|
| Mozilla Firefox | 52.4.1 or newer |
| Google Chrome | 31 or newer |
| Microsoft Edge | All versions |
| Microsoft Internet Explorer | 11 |

# Before you install Collibra DGC

Before you start the installation of Collibra Data Governance Center, follow these steps:

1. Download the installer from the Collibra Downloads page.
   - Linux: **dgc-linux-5.7.12-0.sh**
   - Windows: **dgc-windows-5.7.12-0.zip**
2. Save the installer on each server that is used to run Collibra DGC.

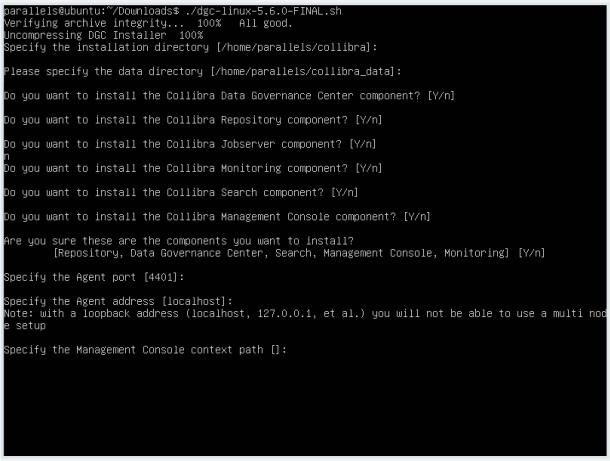
3. On Linux servers:
    a. Make the installer executable: `chmod a+x dgc-linux-5.7.12-0.sh`
    b. Set the locale to *en_US.UTF8*.
    c. Ensure that the user account that will perform the installation or upgrade has execute access on **/tmp**.
4. On Windows Server 2012 R2, install all the latest updates and patches.
5. On Windows, extract the installer ZIP archive.

> **Important** The target directory to extract the ZIP archive cannot contain spaces.

**Note**
On Linux systems without a graphical user interface, the installation wizard remains identical but you need your keyboard to make selections.

- Default values are displayed between square brackets, for example for the **Installation directory**. Press `Enter` to accept the default value or enter a new value.
- If there is a yes or no question, the upper-case character is the default selection. Press `Enter` to select the default value.

In the next image, you can see a snippet of the wizard:




# Install Collibra DGC on multiple nodes

This section describes how to install the Collibra Data Governance Center services on separate nodes.

Keep in mind that you can install all services on one node or two services on one node and the others on a second node.

> **Note** We recommend that you install the Repository service and Jobserver service on dedicated nodes. You can install all other services on another node, provided that it has enough memory. Make sure that you have a fast network between the nodes.
> Also make sure that you use the same installer version on all nodes. You can find the installer version of your environment at the bottom of the sign-in window of Collibra Console, for example 5.7.12-37

# Install Collibra Console

This section describes how to install the Collibra Console software. The Collibra Console is required to configure and manage one or more Collibra DGC environments.

> **Tip** For the installation on Linux without root permissions, also read the services section.

## Steps

> **Note** If you want to configure the init daemon on Linux systems, you have to execute an unattended installation. For more information, see also the unattended installation configuration parameters.

> **Note** Anti-virus and/or security software may block the installation on Windows. Make sure that these allow the installation of software and services.
> For more information, see also the Collibra University course.

1. Run the installer:
   - Linux as user with sudo rights: `sudo ./dgc-linux-5.7.12-0.sh`
     Linux as root user: `./dgc-linux-5.7.12-0.sh`
   - Linux as standard user: `./dgc-linux-5.7.12-0.sh`

- Windows Server: double-click **setup.bat**

> **Important** The path of the installer file cannot contain spaces.

If you run the installation without Administrator rights, an error is shown.

> **Tip** If you don't want to use the user interface even if it's available, add the following to the command:
> `-- --nox11`

2. In the wizard introduction, click **Next**.
3. Enter the **Installation directory** of the Collibra Console component.
   - Default location on Linux as root or user with sudo privileges: **/opt/collibra**
   - Default location on Linux as standard user: **~/collibra**
   - Default location on Windows Server: **C:\collibra**

   > **Important** On Windows, the target installation directory cannot contain spaces.

4. Click **Next**.
5. Enter the location of the **Collibra Data Directory**.
   - Default location on Linux as root or user with sudo privileges: **/opt/collibra_data**
   - Default location on Linux as standard user: **~/collibra_data**
   - Default location on Windows Server: **C:\collibra_data**

   > **Important** On Windows, that target data directory cannot contain spaces.

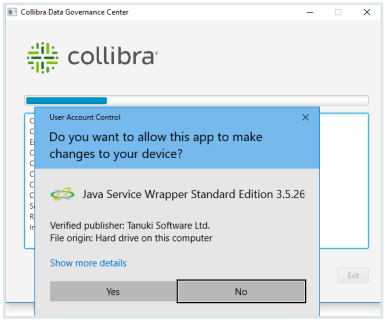
6. Click **Next**.
7. Clear all components except **Management Console**.
8. Click **Next**.
9. Enter the required information.

| Setting | Description |
|---|---|
| Console port | The TCP port to access your Collibra Console via your web browser.<br>The default port is *4402*. |

| Setting | Description |
| --- | --- |
| Console database port | The TCP port to access the Collibra Console database. This is the database where the data and configuration of Collibra Console is stored.<br>The default port is *4420*. |
| Console context path | The path that is added to the base URL to reach Collibra Console.<br>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`.<br>See also Set the context path of Collibra Console. |

If you run multiple Collibra Console instances on one node, this port must be unique for each instance.

10. Click **Install**.
    » The installation of Collibra Console starts.
11. On Windows, you may see User Account Control warnings requesting to make changes to your device.




Click **Yes** for each of the requests, if you click **No**, the installation will fail.

12. Click **Exit**.
    » Collibra Console is installed on your node.

## What's next?

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.

Access to Collibra Console does not require a license.

# Install the DGC service

The DGC service contains the business logic of Collibra Data Governance Center and contains amongst others a workflow engine and various APIs (Java, REST, Search, ...).

This is a mandatory service in every Collibra DGC environment.

> **Tip** For the installation on Linux without root permissions, also read the services section.

## Steps

> **Note** If you want to configure the init daemon on Linux systems, you have to execute an unattended installation. For more information, see also the unattended installation configuration parameters.

> **Note** Anti-virus and/or security software may block the installation on Windows. Make sure that these allow the installation of software and services.
> For more information, see also the Collibra University course.

1. Run the installer:
   - Linux as user with sudo rights: `sudo ./dgc-linux-5.7.12-0.sh`
     Linux as root user: `./dgc-linux-5.7.12-0.sh`
   - Linux as standard user: `./dgc-linux-5.7.12-0.sh`
   - Windows Server: double-click **setup.bat**

     > **Important** The path of the installer file cannot contain spaces.

     If you run the installation without Administrator rights, an error is shown.

   > **Tip** If you don't want to use the user interface even if it's available, add the following to the command:
   > `-- --nox11`

2. In the wizard introduction, click **Next**.

3. Enter the **Installation directory** of the Data Governance Center service.
    - Default location on Linux as root or user with sudo privileges: **/opt/collibra**
    - Default location on Linux as standard user: **~/collibra**
    - Default location on Windows Server: **C:\collibra**

    > **Important** On Windows, the target installation directory cannot contain spaces.

4. Click **Next**.
5. Enter the location of the **Collibra Data Directory**.
    - Default location on Linux as root or user with sudo privileges: **/opt/collibra_data**
    - Default location on Linux as standard user: **~/collibra_data**
    - Default location on Windows Server: **C:\collibra_data**

    > **Important** On Windows, that target data directory cannot contain spaces.

6. Click **Next**.
7. Clear all components except **Data Governance Center**.
8. Click **Next**.
9. Enter the required information.

| Setting | Description |
|---|---|
| DGC port | The TCP port to access your Collibra DGC environment via your web browser.<br>The default port is *4400*. |
| DGC Shutdown port | The TCP port to stop the DGC service.<br>The default port is *4430*. |
| DGC minimum memory | The minimum amount of memory in megabytes for the DGC service.<br>This must be at least 1024 MB and no greater than 32 768 MB (32 GB). |
| DGC maximum memory | The maximum amount of memory in megabytes that can be assigned to the DGC service.<br>This must be at least 2048 MB and no greater than 32 768 MB (32 GB). |

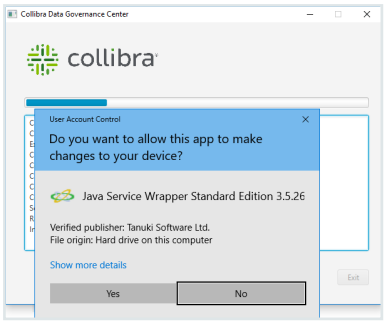
| Setting | Description |
| --- | --- |
| DGC context path | The path that is added to the base URL to reach Collibra Data Governance Center.<br>For example, if your base URL is `https://dgc.yourcompany.com:4400/` and your context path is `acceptance`, then your path to reach Collibra DGC is `https://dgc.yourcompany.com:4400/acceptance`.<br>See also Set the context path of the DGC service in Collibra Console. |

If you run multiple environments on one node, all ports must be unique for each environment.

10. Click **Next**.
11. Enter the Agent service settings and click **Next**.

| Setting | Description |
| --- | --- |
| Agent port | The TCP port that is used by Collibra Console to manage the services of an environment.<br>The default port is *4401*.<br>If you run multiple agents on one node, this port must be unique for each agent. |
| Node address | The hostname of the node on which the Agent service is running.<br>You cannot use a loopback address if you want to use the node in a multinode environment. |

12. Click **Install**.
    » The installation of the DGC service starts.
13. On Windows, you may see User Account Control warnings requesting to make changes to your device.




Click **Yes** for each of the requests, if you click **No**, the installation will fail.

14. Click **Exit**.
    » The DGC service is installed on your node.

## What's next?

Add the DGC service to your environment, this service is mandatory for a functional Collibra DGC environment.

# Install the Repository service

The repository service is used to run the database of Collibra Data Governance Center.

This is a mandatory service in every Collibra DGC environment.

> **Tip** For the installation on Linux without root permissions, also read the services section.

## Steps

> **Note** If you want to configure the init daemon on Linux systems, you have to execute an unattended installation. For more information, see also the unattended installation configuration parameters.

> **Note** Anti-virus and/or security software may block the installation on Windows. Make sure that these allow the installation of software and services.
> For more information, see also the Collibra University course.

1. Run the installer:
   - Linux as user with sudo rights: `sudo ./dgc-linux-5.7.12-0.sh`
     Linux as root user: `./dgc-linux-5.7.12-0.sh`
   - Linux as standard user: `./dgc-linux-5.7.12-0.sh`
   - Windows Server: double-click **setup.bat**

     > **Important** The path of the installer file cannot contain spaces.

     If you run the installation without Administrator rights, an error is shown.

   > **Tip** If you don't want to use the user interface even if it's available, add the following to the command:
   > `-- --nox11`

2. In the wizard introduction, click **Next**.
3. Enter the **Installation directory** of the Repository service.
   - Default location on Linux as root or user with sudo privileges: **/opt/collibra**
   - Default location on Linux as standard user: **~/collibra**
   - Default location on Windows Server: **C:\collibra**

     > **Important** On Windows, the target installation directory cannot contain spaces.

4. Click **Next**.
5. Enter the location of the **Collibra Data Directory**.
   - Default location on Linux as root or user with sudo privileges: **/opt/collibra_data**
   - Default location on Linux as standard user: **~/collibra_data**
   - Default location on Windows Server: **C:\collibra_data**

     > **Important** On Windows, that target data directory cannot contain spaces.

6. Click **Next**.
7. Clear all components except **Repository**.
8. Click **Next**.

9. Enter the required information.

| Setting | Description |
|---|---|
| Repository port | The TCP port to access the repository service. It is only used by the DGC service and the Collibra agent.<br>The default port is *4403*.<br>If you run multiple environments on one node, all ports must be unique for each environment. |
| Repository memory | The amount of memory for the Repository service in megabytes.<br>This must be at least 512 MB and no greater than 16 384 MB (16 GB). |
| Repository admin password (*) | The password that is used by the agent to access the Repository service. |
| Confirm repository admin password | The password as entered in the Repository admin password field. |
| Repository dgc password (*) | The password that is used by the DGC service to access the repository. |
| Confirm repository dgc password | The password as entered in the Repository dgc password field. |

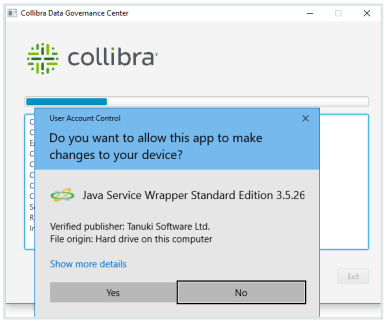
> **Note** (*) These passwords can contain the following characters:
> - lowercase letters
> - uppercase letters
> - numbers
> - the following special characters: #?!@$%&*-

10. Click **Next**.

11. Enter the Agent service settings and click **Next**.

| Setting | Description |
|---|---|
| Agent port | The TCP port that is used by Collibra Console to manage the services of an environment.<br>The default port is *4401*.<br>If you run multiple agents on one node, this port must be unique for each agent. |
| Node address | The hostname of the node on which the Agent service is running.<br>You cannot use a loopback address if you want to use the node in a multinode environment. |

12. Click **Install**.
    » The installation of the Repository service starts.
13. On Windows, you may see User Account Control warnings requesting to make changes to your device.




    Click **Yes** for each of the requests, if you click **No**, the installation will fail.
14. Click **Exit**.
    » The Repository service is installed on your node.

## What's next?

Add the Repository service to your environment, this service is mandatory for a functional Collibra DGC environment.

# Install the Search service

The Search service allows you to search for any asset in Collibra Data Governance Center.

This is a mandatory service in every Collibra DGC environment.

> **Tip** For the installation on Linux without root permissions, also read the services section.

## Prerequisites

If you install the Search service on a Linux system, the node that will run the Search service must pass the following bootstrap checks:

- File descriptor
- Maximum number of threads check
- Maximum file size
- Maximum size virtual memory check
- Maximum map count check

| Type | Check description | Minimum value | Applies for installation type | Setting name |
|---|---|---|---|---|
| User limit | Maximum number of open file descriptors | 65536 | ▪ Without root permissions<br>▪ With root permissions, using System V init daemon | nofile |
| | Maximum number of open threads/processes | 4096 | | nproc |
| | Maximum file size | unlimited | | fsize |
| Kernel parameter | Maximum virtual memory areas | 262144 | ▪ All | vm.max_map_count |

For more information on these settings, see the Troubleshooting section.

## Steps

> **Note** If you want to configure the init daemon on Linux systems, you have to execute an unattended installation. For more information, see also the unattended installation configuration parameters.

> **Note** Anti-virus and/or security software may block the installation on Windows. Make sure that these allow the installation of software and services.
> For more information, see also the Collibra University course.

1. Run the installer:
   - Linux as user with sudo rights: `sudo ./dgc-linux-5.7.12-0.sh`
     Linux as root user: `./dgc-linux-5.7.12-0.sh`
   - Linux as standard user: `./dgc-linux-5.7.12-0.sh`
   - Windows Server: double-click **setup.bat**

     > **Important** The path of the installer file cannot contain spaces.

     If you run the installation without Administrator rights, an error is shown.

   > **Tip** If you don't want to use the user interface even if it's available, add the following to the command:
   > `-- --nox11`

2. In the wizard introduction, click **Next**.
3. Enter the **Installation directory** of the Search service.
   - Default location on Linux as root or user with sudo privileges: **/opt/collibra**
   - Default location on Linux as standard user: **~/collibra**
   - Default location on Windows Server: **C:\collibra**

     > **Important** On Windows, the target installation directory cannot contain spaces.

4. Click **Next**.
5. Enter the location of the **Collibra Data Directory**.
   - Default location on Linux as root or user with sudo privileges: **/opt/collibra_data**
   - Default location on Linux as standard user: **~/collibra_data**
   - Default location on Windows Server: **C:\collibra_data**

     > **Important** On Windows, that target data directory cannot contain spaces.

6. Click **Next**.
7. Clear all components except **Search**.

8. Click **Next**.
9. Enter the required information.

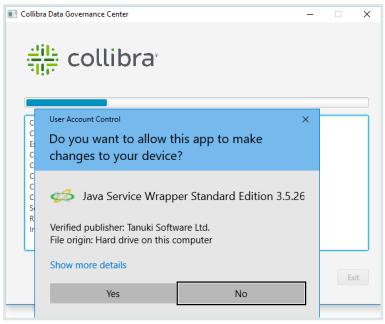
| Setting | Description |
|---|---|
| Search http port | The TCP port to access the Search service via REST API.<br>The default port is *4421*. |
| Search transport port | The TCP port for the communication between the DGC and the Search service.<br>The default port is *4422*. |
| Search memory | The amount of memory in megabytes that is assigned to the Search service.<br>The default value is *1024*. |

If you run multiple environments on one node, all ports must be unique for each environment.

10. Click **Next**.
11. Enter the Agent service settings and click **Next**.

| Setting | Description |
|---|---|
| Agent port | The TCP port that is used by Collibra Console to manage the services of an environment.<br>The default port is *4401*.<br>If you run multiple agents on one node, this port must be unique for each agent. |
| Node address | The hostname of the node on which the Agent service is running.<br>You cannot use a loopback address if you want to use the node in a multinode environment. |

12. Click **Install**.
    » The installation of the Search service starts.
13. On Windows, you may see User Account Control warnings requesting to make changes to your device.




Click **Yes** for each of the requests, if you click **No**, the installation will fail.

14. Click **Exit**.
    » The Search service is installed on your node.

## What's next?

Add the Search service to your environment. Th Search service is mandatory for a functional Collibra DGC environment.

# Install the Monitoring service

The Monitoring service allows you to gather metrics from Collibra Data Governance Center. The service also provides extensive monitoring and diagnostics capabilities.

This is a mandatory service in every Collibra DGC environment.

If you don't install the Monitoring service on a dedicated node, we recommend to install the service on the node that hosts the Repository service.

> **Tip** For the installation on Linux without root permissions, also read the services section.

## Steps

> **Note** If you want to configure the init daemon on Linux systems, you have to execute an unattended installation. For more information, see also the unattended installation configuration parameters.

> **Note** Anti-virus and/or security software may block the installation on Windows. Make sure that these allow the installation of software and services.
> For more information, see also the Collibra University course.

1. Run the installer:
   - Linux as user with sudo rights: `sudo ./dgc-linux-5.7.12-0.sh`
     Linux as root user: `./dgc-linux-5.7.12-0.sh`
   - Linux as standard user: `./dgc-linux-5.7.12-0.sh`
   - Windows Server: double-click **setup.bat**

     > **Important** The path of the installer file cannot contain spaces.

     If you run the installation without Administrator rights, an error is shown.

   > **Tip** If you don't want to use the user interface even if it's available, add the following to the command:
   > `-- --nox11`

2. In the wizard introduction, click **Next**.
3. Enter the **Installation directory** of the Monitoring service.
   - Default location on Linux as root or user with sudo privileges: **/opt/collibra**
   - Default location on Linux as standard user: **~/collibra**
   - Default location on Windows Server: **C:\collibra**

     > **Important** On Windows, the target installation directory cannot contain spaces.

4. Click **Next**.
5. Enter the location of the **Collibra Data Directory**.
   - Default location on Linux as root or user with sudo privileges: **/opt/collibra_data**
   - Default location on Linux as standard user: **~/collibra_data**
   - Default location on Windows Server: **C:\collibra_data**

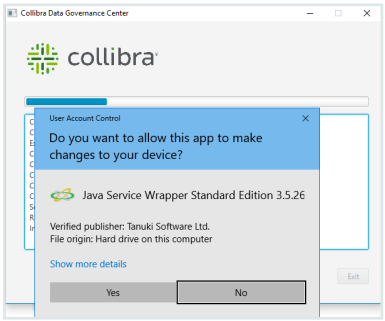
     > **Important** On Windows, that target data directory cannot contain spaces.

6. Click **Next**.
7. Clear all components except **Monitoring**.

8. Click **Next**.
9. Enter the port of the service. The default port is *4407*.
   If you run multiple environments on one node, this port must be unique for each environment.
10. Click **Next**.
11. Enter the Agent service settings and click **Next**.

| Setting | Description |
|---|---|
| Agent port | The TCP port that is used by Collibra Console to manage the services of an environment.<br>The default port is *4401*.<br><br>If you run multiple agents on one node, this port must be unique for each agent. |
| Node address | The hostname of the node on which the Agent service is running.<br><br>You cannot use a loopback address if you want to use the node in a multinode environment. |

12. Click **Install**.
    » The installation of the Monitoring service starts.
13. On Windows, you may see User Account Control warnings requesting to make changes to your device.




    Click **Yes** for each of the requests, if you click **No**, the installation will fail.
14. Click **Exit**.
    » The Monitoring service is installed on your node.

## What's next?

Add the Monitoring service to your environment in Collibra Console.

# Install the Jobserver

The Jobserver is used to ingest data and to execute data profiling or to create sample data on the ingested data. You can ingest data when you register a data source.

The Jobserver must be installed on a dedicated server.

> **Tip**
> When you install an on-premises Jobserver for use in a Collibra Data Intelligence Cloud environment, you also have to install Collibra Console, to manage and configure this Jobserver. You can install both services on the same server.
>
> You can find the version of your Collibra Data Intelligence Cloud environment at the bottom of the sign-in window, for example 5.7.0. Always use the latest available on-premises installer to install the Jobserver.

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 Governance Center.

## Steps

> **Tip** For the installation on Linux without root permissions, also read the services section.

> **Note** If you want to configure the init daemon on Linux systems, you have to execute an unattended installation. For more information, see also the unattended installation configuration parameters.

> **Note** Anti-virus and/or security software may block the installation on Windows. Make sure that these allow the installation of software and services.
> For more information, see also the Collibra University course.

1. Run the installer on a dedicated server:
   - Linux as user with sudo rights: `sudo ./dgc-linux-5.7.12-0.sh`
     Linux as root user: `./dgc-linux-5.7.12-0.sh`
   - Linux as standard user: `./dgc-linux-5.7.12-0.sh`

- Windows Server: double-click **setup.bat**

> **Important** The path of the installer file cannot contain spaces.

If you run the installation without Administrator rights, an error is shown.

> **Tip** If you don't want to use the user interface even if it's available, add the following to the command:
> `-- --nox11`

2. In the wizard introduction, click **Next**.
3. Enter the **Installation directory** of the Jobserver service.
   - Default location on Linux as root or user with sudo privileges: **/opt/collibra**
   - Default location on Linux as standard user: **~/collibra**
   - Default location on Windows Server: **C:\collibra**

   > **Important** On Windows, the target installation directory cannot contain spaces.

4. Click **Next**.
5. Enter the location of the **Collibra Data Directory**.
   - Default location on Linux as root or user with sudo privileges: **/opt/collibra_data**
   - Default location on Linux as standard user: **~/collibra_data**
   - Default location on Windows Server: **C:\collibra_data**

   > **Important** On Windows, that target data directory cannot contain spaces.

6. Click **Next**.
7. Clear all components except **Jobserver**.
8. Click **Next**.
9. Enter the required information.

| Setting | Description |
|---|---|
| Jobserver port | The TCP port to access the Jobserver service.<br>The default port is *4404*. |

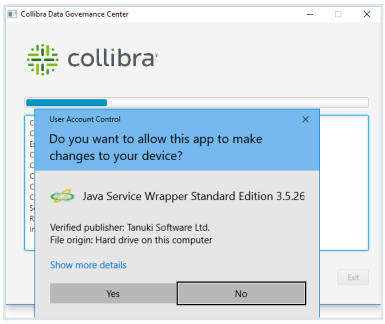
| Setting | Description |
| --- | --- |
| Jobserver database port | The TCP port to access the Jobserver database.<br>The default port is *4414*. |
| Jobserver monitoring port | The TCP port that is used by the monitoring service to monitor the Jobserver service.<br>The default port is *4424*. |
| Jobserver Spark monitoring port | The TCP port that is used by the monitoring service to monitor the Jobserver Spark service.<br>The default port is *4434*. |

If you run multiple environments on one node, all ports must be unique for each environment.

10. Click **Next**.
11. Enter the Agent service settings and click **Next**.

| Setting | Description |
| --- | --- |
| Agent port | The TCP port that is used by Collibra Console to manage the services of an environment.<br>The default port is *4401*.<br>If you run multiple agents on one node, this port must be unique for each agent. |
| Node address | The hostname of the node on which the Agent service is running.<br>You cannot use a loopback address if you want to use the node in a multinode environment. |

12. Click **Install**.
    » The installation of the Jobserver starts.
13. On Windows, you may see User Account Control warnings requesting to make changes to your device.




Click **Yes** for each of the requests, if you click **No**, the installation will fail.

14. Click **Exit**.
    » The Jobserver service is installed on your node.

## What's next?

Add the Jobserver service to your environment and then add the Jobserver to the Data Governance Center service in Collibra Console.

For more information about the Jobserver configuration, consult the Jobserver memory and CPU usage section.

# Complete the setup

In the previous sections, you have installed all possible Collibra DGC services on two or more nodes. To complete the setup, you have to create an environment and add the services to this environment. You need Collibra Console to do this.

## Prerequisites

You have installed at least:

- Collibra Console.
- the DGC service.
- the Repository service.
- the Search service.
- the Monitoring service.

The Jobserver service is optional.

# Steps

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

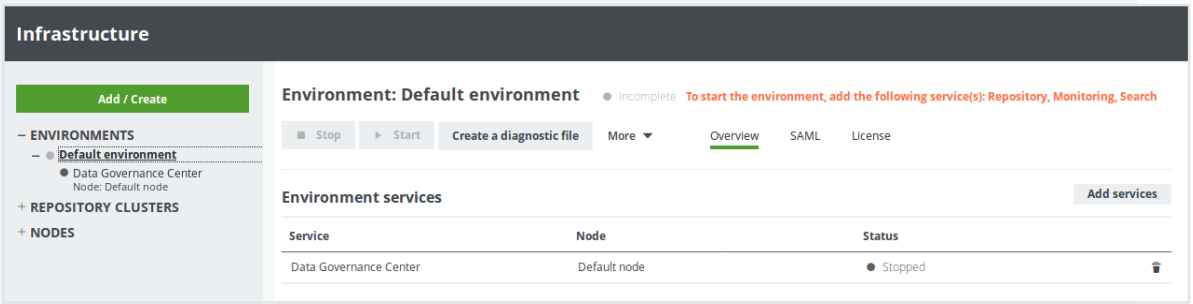
   **Tip**
   - The default address to access Collibra Console is *<server hostname>:4402*, but you may have set another port during the installation of Collibra Console. Keep in mind that a firewall of your operating system can block the access to Collibra Console.
   - The default credentials to sign 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.
   - Access to Collibra Console does not require a license.

2. Add the necessary nodes to the infrastructure. Each node hosts one or more services of your Collibra DGC environment.
3. Create a new environment.
4. Add services to the environment.
   A fully functional environment requires the Data Governance Center, Repository, Search and Monitoring services.

   **Note** The Jobserver service is only required if you are ingesting data with Data Catalog. See Add a Jobserver to the Data Governance Center service for more information.

5. In the tab pane, click the name of the created environment.

   **Tip** If your environment is missing a mandatory service, it will be indicated and you cannot start the environment.



6. Click ▶ **Start**.
   » If all services and the environment have the **running** status, you have successfully started Collibra DGC.

## What's next?

Upload your license to start using the environment.

> **Tip** If you want to register data sources in Data Catalog, you have to configure the Jobserver.

## Create a Collibra DGC environment

A Collibra Data Governance Center 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.

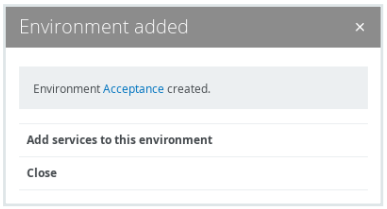
### Steps

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.

## Add a node to your infrastructure

A node is a physical server that runs one or more services of a Collibra Data Governance Center environment.

### Prerequisites

- The node that you want to add to your infrastructure must be up and running and reachable from the Console that you are using.
- The version of the node must match the version of Collibra Console.

> **Tip** To add a node that was previously registered to another Collibra Console, see the knowledge base on the Collibra Support Portal.

### Steps

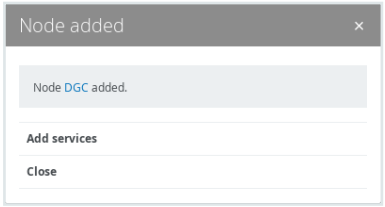
1. Open Collibra Console with a user profile that has the **SUPER** role.
   » Collibra Console opens with the **Infrastructure** page.

> **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 **Add node**.
   » The **Add node** dialog box appears.
4. Enter the necessary information.

| Field | Description |
|---|---|
| Node name | Enter a meaningful name for the node. |
| Hostname | Enter the hostname or IP address of the node, for example *192.168.1.100* or *repository-node-A*.<br>If you use a hostname, make sure that the Collibra Console can resolve the hostname.<br>**Note** Do not reuse hostnames, every hostname must be unique. If you reuse a hostname for a node that will be used in a repository cluster, the cluster won't synchronize. |
| Port | Enter the agent port. This is the port through which Collibra Console connects to the node. The default value is 4401. |

5. Click **Add node**
6. Click **Close**.



# Add a service to an environment

A Collibra Data Governance Center environment consists of a collection of services, such as the DGC service and the Repository service. A service is hosted on a node. To add a service to an environment, the node must be added to the infrastructure that is managed by Collibra Console.

You can add a service to an environment in the following ways:

- Add a service via the global **Add / Create** button.
- Add a service via the environment details.
- Add services while creating an environment.

> **Tip** Make sure that the environment is stopped before adding services.

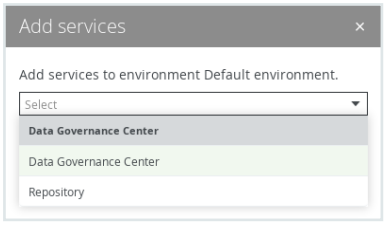
## Via global Add / Create button

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.

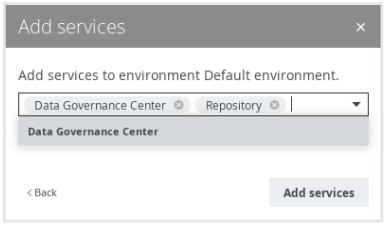
2. In the tab pane, click **Add / Create**.
   » The **Add / Create** dialog box appears.
3. Click **Add services to environment / cluster**.
   » The **Select environment** dialog box appears.
4. Select the **Environment** option and select an environment from the drop-down list.
5. Click **Next**.
   » The **Add services** dialog box appears.

6. Click the relevant services in the drop-down list.

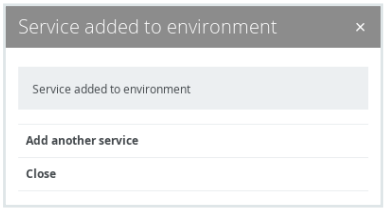



> **Note** If the node that hosts the service you want to add is not yet available in Collibra Console, click **Add services from a new node** under the drop-down list and add the node details.

7. Click **Add services**.



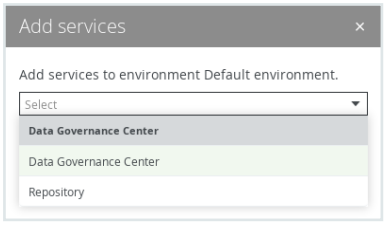

8. When you have added all services, click **Close**.




## Via environment details

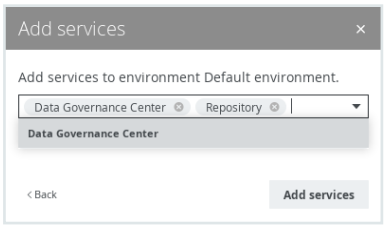
1. Open Collibra Console with a user profile that has the **SUPER** role.
   » Collibra Console opens with the **Infrastructure** page.
2. In the tab pane, click the name of your environment.
   » The environment details appear.
3. Click **Add services**.
   » The **Add services** dialog box appears.

4. Click the relevant services in the drop-down list.

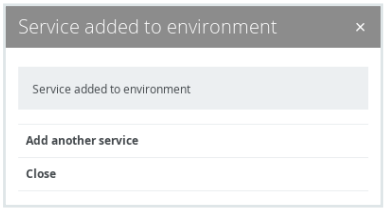



> **Note** If the node that hosts the service you want to add is not yet available in Collibra Console, click **Add services from a new node** under the drop-down list and add the node details.

5. Click **Add services**.




6. When you have added all services, click **Close**.




# Install Collibra DGC on a single node

This section describes how to install Collibra Data Governance Center on a single node.

> **Tip** We recommend to use single-node installations only for testing environments.

## Install Collibra DGC on a single node

This section describes how to install the Collibra Data Governance Center's services on a single node. Environments that run on a single node should only be used for testing purposes.

For production environments, install the services on multiple nodes.

> **Tip** For the installation on Linux without root permissions, also read the services section.

## Prerequisites

- You have downloaded the installer for your operating system.
- If you install the Search service on a Linux system, the node that will run the Search service must pass the following bootstrap checks:
  - File descriptor
  - Maximum number of threads check
  - Maximum file size
  - Maximum size virtual memory check
  - Maximum map count check

| Type | Check description | Minimum value | Applies for installation type | Setting name |
|---|---|---|---|---|
| User limit | Maximum number of open file descriptors | 65536 | ▪ Without root permissions<br>▪ With root permissions, using System V init daemon | nofile |
| | Maximum number of open threads/processes | 4096 | | nproc |
| | Maximum file size | unlimited | | fsize |
| Kernel parameter | Maximum virtual memory areas | 262144 | ▪ All | vm.max_map_count |

For more information on these settings, see the Troubleshooting section.

## Steps

> **Note** If you want to configure the init daemon on Linux systems, you have to execute an unattended installation. For more information, see also the unattended installation configuration parameters.

> **Note** Anti-virus and/or security software may block the installation on Windows. Make sure that these allow the installation of software and services.
> For more information, see also the Collibra University course.

1. Run the installer:
   - Linux as user with sudo rights: `sudo ./dgc-linux-5.7.12-0.sh`
     Linux as root user: `./dgc-linux-5.7.12-0.sh`
   - Linux as standard user: `./dgc-linux-5.7.12-0.sh`
   - Windows Server: double-click **setup.bat**

     > **Important** The path of the installer file cannot contain spaces.

     If you run the installation without Administrator rights, an error is shown.

   > **Tip** If you don't want to use the user interface even if it's available, add the following to the command:
   > `-- --nox11`

2. In the wizard introduction, click **Next**.
3. Enter the **Installation directory** and click **Next**.
   - Default location on Linux as root or user with sudo privileges: **/opt/collibra**
   - Default location on Linux as standard user: **~/collibra**
   - Default location on Windows Server: **C:\collibra**

     > **Important** On Windows, the target installation directory cannot contain spaces.

4. Enter the **Data directory** and click **Next**.
   - Default location on Linux as root or user with sudo privileges: **/opt/collibra_data**
   - Default location on Linux as standard user: **~/collibra_data**
   - Default location on Windows Server: **C:\collibra_data**

     > **Important** On Windows, that target data directory cannot contain spaces.

5. Select the components and click **Next**.
   You must select at least Data Governance Center, Repository, Management Console, Search and Monitoring.

> **Note** The Jobserver service is only required if you are ingesting data with Data Catalog. See Add a Jobserver to the Data Governance Center service for more information.

6. Enter the DGC service settings and click **Next**.

| Setting | Description |
|---|---|
| DGC port | The TCP port to access your Collibra DGC environment via your web browser.<br>The default port is *4400*. |
| DGC Shutdown port | The TCP port to stop the DGC service.<br>The default port is *4430*. |
| DGC minimum memory | The minimum amount of memory in megabytes for the DGC service.<br>This must be at least 1024 MB and no greater than 32 768 MB (32 GB). |
| DGC maximum memory | The maximum amount of memory in megabytes that can be assigned to the DGC service.<br>This must be at least 2048 MB and no greater than 32 768 MB (32 GB). |
| DGC context path | The path that is added to the base URL to reach Collibra Data Governance Center.<br>For example, if your base URL is `https://dgc.yourcompany.com:4400/` and your context path is `acceptance`, then your path to reach Collibra DGC is `https://dgc.yourcompany.com:4400/acceptance`.<br>See also Set the context path of the DGC service in Collibra Console. |

If you run multiple environments on one node, all ports must be unique for each environment.

7. Enter the Repository service settings and click **Next**.

| Setting | Description |
|---|---|
| Repository port | The TCP port to access the repository service. It is only used by the DGC service and the Collibra agent.<br>The default port is *4403*.<br>If you run multiple environments on one node, all ports must be unique for each environment. |
| Repository memory | The amount of memory for the Repository service in megabytes.<br>This must be at least 512 MB and no greater than 16 384 MB (16 GB). |
| Repository admin password (*) | The password that is used by the agent to access the Repository service. |
| Confirm repository admin password | The password as entered in the Repository admin password field. |
| Repository dgc password (*) | The password that is used by the DGC service to access the repository. |
| Confirm repository dgc password | The password as entered in the Repository dgc password field. |

> **Note** (*) These passwords can contain the following characters:
> - lowercase letters
> - uppercase letters
> - numbers
> - the following special characters: #?!@$%&*-

8. If you selected the Jobserver in step 5, enter the Jobserver service settings and click **Next**.

| Setting | Description |
|---|---|
| Jobserver port | The TCP port to access the Jobserver service.<br>The default port is *4404*. |
| Jobserver database port | The TCP port to access the Jobserver database.<br>The default port is *4414*. |
| Jobserver monitoring port | The TCP port that is used by the monitoring service to monitor the Jobserver service.<br>The default port is *4424*. |
| Jobserver Spark monitoring port | The TCP port that is used by the monitoring service to monitor the Jobserver Spark service.<br>The default port is *4434*. |

If you run multiple environments on one node, all ports must be unique for each environment.

9. Enter the port of the Monitoring service and click **Next**.The default port is *4407*.
10. Enter the Search service settings and click **Next**.

| Setting | Description |
|---|---|
| Search http port | The TCP port to access the Search service via REST API.<br>The default port is *4421*. |
| Search transport port | The TCP port for the communication between the DGC and the Search service.<br>The default port is *4422*. |
| Search memory | The amount of memory in megabytes that is assigned to the Search service.<br>The default value is *1024*. |

If you run multiple environments on one node, all ports must be unique for each environment.

11. Enter the Agent service settings and click **Next**.

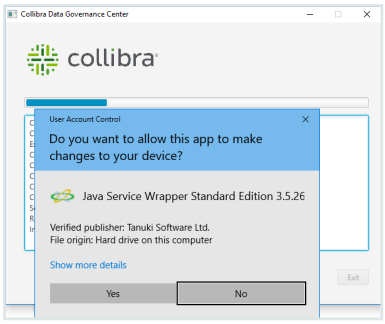
| Setting | Description |
|---|---|
| Agent port | The TCP port that is used by Collibra Console to manage the services of an environment.<br>The default port is *4401*.<br>If you run multiple agents on one node, this port must be unique for each agent. |
| Node address | The hostname of the node on which the Agent service is running.<br>You cannot use a loopback address if you want to use the node in a multinode environment. |

12. Enter the Console service settings.

| Setting | Description |
|---|---|
| Console port | The TCP port to access your Collibra Console via your web browser.<br>The default port is *4402*. |
| Console database port | The TCP port to access the Collibra Console database. This is the database where the data and configuration of Collibra Console is stored.<br>The default port is *4420*. |
| Console context path | The path that is added to the base URL to reach Collibra Console.<br>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`.<br>See also Set the context path of Collibra Console. |

If you run multiple Collibra Console instances on one node, this port must be unique for each instance.

13. Click **Install**.
    » The installation of the components starts.
14. On Windows, you may see User Account Control warnings requesting to make changes to your device.




Click **Yes** for each of the requests, if you click **No**, the installation will fail.

15. Click **Exit**.
    » Collibra DGC is installed on your system.

## What's next?

Start your environment for the first time.

# Start an environment for the first time

After the installation of the software on a single node, you need to configure an environment to start using Collibra Data Governance Center.

## Steps

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

**Tip**

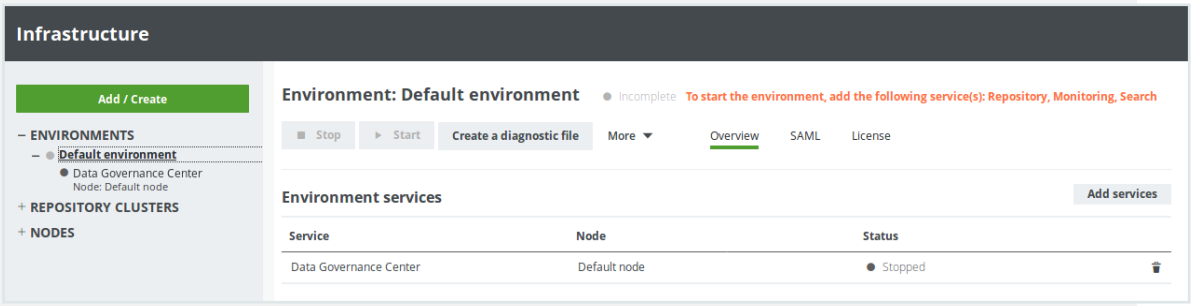
- The default address to access Collibra Console is *<server hostname>:4402*, but you may have set another port during the installation of Collibra Console. Keep in mind that a firewall of your operating system can block the access to Collibra Console.
- The default credentials to sign in to Collibra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

  **Note** The credentials are case-senstive.

- Access to Collibra Console does not require a license.

2. In the tab pane, click **Default environment**.
   All the services that you selected during the installation are added to this environment.

   **Tip** A valid environment needs the Data Governance Center, Repository, Search and Monitoring services. If one of those services is missing, your environment will have the status incomplete and you cannot start the environment.




3. Click ▶ **Start**.
   » If all services and the environment have the **running** status, you have successfully started Collibra DGC.

## What's next?

Upload your license to start using the environment.

**Tip** If you want to register data sources in Data Catalog, you must have selected Jobserver during the installation and you have to configure this service.

# Run multiple Collibra DGC environments on one server

This procedure describes how you can install multiple Collibra Data Governance Center environments on one server. Even though you should only do this for testing purposes, we don't support this setup.

The basic idea of running multiple Collibra DGC environments on one server is to install each Collibra DGC in a separate directory and to assign a different set of ports per environment.

> **Note** You can only run multiple Collibra DGC installations on one server on Linux systems without root permission.

For example:

| Parameters | Installation 1 | Installation 2 |
|---|---|---|
| **Installation directory** | /home/johndoe/collibra1 | /home/johndoe/collibra2 |
| **Data directory** | /home/johndoe/collibra_data1 | /home/johndoe/collibra_data2 |
| **Context path** (optional) | dgc_dev | dgc_test |
| **DGC port** | 4400 | 4500 |
| **DGC Shutdown port** | 4430 | 4530 |
| **Agent port** | 4401 | 4501 |
| **Repository port** | 4403 | 4503 |
| **Jobserver port** | 4404 | 4504 |
| **Jobserver database port** | 4414 | 4514 |
| **Jobserver monitoring port** | 4424 | 4524 |
| **Jobserver Spark monitoring port** | 4434 | 4534 |
| **Monitoring port** | 4407 | 4507 |

| Parameters | Installation 1 | Installation 2 |
|---|---|---|
| **Console port** | 4402 | 4502 |
| **Console database** | 4420 | 4520 |
| **Console context path** (optional) | console_dev | console_test |
| **Search HTTP port** | 4421 | 4521 |
| **Search Transport port** | 4422 | 4522 |

To run multiple Collibra DGC installations on one server:

1. Install Collibra DGC on a single node, using one set of the parameters.
2. Execute another single node installation using another set of parameters.
3. Start both installations for further configuration.

# Upload a Collibra license

When you have installed and started an environment or you have upgraded your Collibra Data Governance Center 4.x environment to 5.x, you must upload a valid 5.x Collibra license file to start using Collibra Data Governance Center.

> **Note** If you upload an invalid license, for example it has been tampered with, there are no changes in the **Usage information** and **Collibra products** sections.
>
> Access to Collibra Console does not require a license.

## 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 **License** tab.
4. Click **Upload new license**.

5. In the **Upload new license** dialog box, do one of the following:
    - Drag and drop a valid license file in the **Upload a file or drop ...** field.
    - Click in the **Upload a file or drop ...** field, select the Collibra license file and click **Open**.
6. In the upper-right corner of the license information section, click **Refresh**.
   The **Usage information** and **Collibra products** sections are updated.

# Collibra Data Governance Center license file

To be able to use Collibra Data Governance Center, you need a valid license file, named **collibra.license**, that you have to upload in Collibra Console.

Your organization's license file defines:

- The Collibra products and applications that your organization can use.
- The expiration date of the user licenses.
- The maximum number of users.

This license file also contains a signature string, to avoid tampering.

## Example license

The following is an example of a Collibra DGC license file:

```
customer = Collibra
writerCount = 10
contributorCount = 0
readerCount = 50
apiUserCount = 0
maxAssets = 2147483647
maxWorkflows = 2147483647
product-connect = true
product-catalog = true
product-glossary = true
product-reference-data = true
product-helpdesk = true
product-policy = true
product-stewardship = true
product-data-dictionary = false
product-onthego-mobile = true
product-onthego-windows = true
```

```
expirationDate = 2100-12-31
guestAccess = true
Vqr27XTn0Swuax...
```

# Maximum number of users per license type

The license file contains the maximum number of user licenses of each license type.

For each license type, Collibra Data Governance Center counts the number of enabled users.

> **Note** The apiUserCount parameter in the license file is no longer taken into account in Collibra DGC.

# Applications

Collibra DGC consists of the following applications and external products.

Applications:

- Catalog (product-catalog)
- Business Glossary (product-glossary)
- Reference Data (product-reference-data)
- Data Helpdesk (product-helpdesk)
- Policy Manager (product-policy)
- Stewardship (product-stewardship)
- Data Dictionary (product-data-dictionary)

External products:

- Collibra Connect (product-connect)
- Collibra Everywhere for iOS (product-onthego-mobile)
- Collibra Everywhere for Windows (product-onthego-windows)

# License violations

You can encounter the following license violations:

- Invalid or no license file: You have no license file or have tampered with it.
- Expired: Your license has passed the expiration date that was defined in the license file. When you are close to the expiration date of your license, every user with the Sysadmin role gets a notification message at the top of every page.

  > **Note** In Collibra Console, every user will see the license expiry message.

- Author user limit exceeded: Contact Collibra Support, at support@collibra.com, for further assistance.

  > **Note** This does not prevent you from creating new users, or restrict Collibra DGC in any other way. It is, however, a license violation, and may lead to legal action.

In case of license violations, all Collibra DGC users get an error message at the top of every page. You can only remove the error by taking one or more of the following actions:

- Upload a new, suited license file.
- Disable users.
- Delete users.

If there is no license file, an invalid license, or an expired license, Collibra DGC blocks all upgrades.

# Change the Collibra Console administrator password

The initial password of the Collibra Console administrator is "admin". It is highly recommended to change this password after your first sign in.

> **Note**
> - Keep in mind that Collibra Console's Admin user has the SUPER role.
> - Passwords are case-senstive.

# Steps

1. Open Collibra Console with the user **Admin**.
2. In the main menu, click **Console settings**.
3. In tab pane, click **Users**.
4. Click in the row of the user **Admin**.
   » The **Change password** dialog box appears.
5. Under **Old password**, enter the current password.
6. Under **New password**, enter the new password.
7. Under **Repeat**, enter the new password again.
8. Click **Change password**.

# What's next?

The password is changed. Next time you log in with the **Admin** user, use the new password.

Chapter 4

# Unattended installation

Instead of following the installation wizard, whether you use the installation user interface or the command line, you can also install the software in an unattended way by executing the installation command in combination with a configuration file.

This allows you to automate the installation process on various servers.

## In this chapter

# Install the software unattendedly

In this section, you will learn how you can install Collibra DGC without manual interaction.

## Prerequisites

- The prerequisites of a normal installation, see System requirements.
- A valid configuration file in JSON format, see Unattended installation configuration parameters.
  None of the configuration parameters is required. For every parameter that is not provided, the system will use a default value.
- You have write access to the temporary folder in which the installer extracts the required files.

  > **Tip** The default temporary folder of your operation system is used, but you can choose another temporary folder as well.

- If you install the Search service on a Linux system, the node that will run the Search service must pass the following bootstrap checks:
  - File descriptor
  - Maximum number of threads check
  - Maximum file size
  - Maximum size virtual memory check
  - Maximum map count check

| Type | Check description | Minimum value | Applies for installation type | Setting name |
|---|---|---|---|---|
| User limit | Maximum number of open file descriptors | 65536 | ▪ Without root permissions<br>▪ With root permissions, using System V init daemon | nofile |
| | Maximum number of open threads/processes | 4096 | | nproc |
| | Maximum file size | unlimited | | fsize |

| Type | Check description | Minimum value | Applies for installation type | Setting name |
|---|---|---|---|---|
| Kernel parameter | Maximum virtual memory areas | 262144 | ▪ All | vm.max_map_count |

For more information on these settings, see the Troubleshooting section.

# Linux

1. Open a terminal session and go to the directory with the installer.
2. Run the following command:
   - As root: `sudo ./dgc-linux-5.7.12-0.sh -- --config /full-path/to/config`
   - As non-root: `./dgc-linux-5.7.12-0.sh -- --config /full-path/to/-config`

> **Tip**
> - You can replace `--config` by `-c`.
> - Use the full path to the configuration file, even if it is in the same directory as the installer.

Example output:

```
~$ ./dgc-linux-5.7.12-0.sh -- --config
/home/johndoe/Downloads/config.json
Verifying archive integrity... 100% All good.
Uncompressing DGC Installer 100%
10:49:02.235 - Using configuration file :
/home/johndoe/Downloads/config.json
10:49:02.324 - SUCCESS - Check umask settings
10:49:02.326 - SUCCESS - Create installation and data directories
10:49:02.353 - SUCCESS - Create installation configuration file
10:49:02.454 - SUCCESS - Create uninstall script.
10:49:04.559 - SUCCESS - Extract JRE

...
```

```
10:49:20.826 - SUCCESS - Start Agent
10:49:24.464 - SUCCESS - Start Console
10:49:24.464 - Installation finished in 22184ms.
```

## Windows

1. Open a command-line session (Command Prompt or Windows PowerShell) as Administrator and go to the directory with the installer.
2. Run the following command: `setup.bat --config <full-path/to/config>`

> **Tip**
> - You can replace `--config` by `-c`.
> - Use the full path to the configuration file, even if it is in the same directory as the installer.

## What's next?

After you have installed all the services, follow either one of the following instructions to create an environment:

- Complete the setup if you installed all the services on separate nodes.
- Start an environment for the first time if you installed all the services on a single node.

# Unattended installation configuration parameters

The following table contains the parameters that you can use in the JSON installation file for an unattended installation of Collibra Data Governance Center. If the parameter is not provided, a default value is used.

| Parameter | Description | Type | Linux example | Windows example |
|---|---|---|---|---|
| installationDirectory | Name of the directory where Collibra DGC will be installed.<br>On Windows, the directory must have a URL format (file:///path). | string | • Default (Linux with root permission): */opt/collibra*<br>• Default (Linux without root permission): **/home/<user>/collibra** | Default: *file:///c:/collibra* |
| dataDirectory | Name of the directory where the Collibra data will be stored.<br>On Windows, the directory must have a URL format (file:///path). | string | • Default (Linux with root permission): */opt/collibra_data*<br>• Default (Linux without root permission): **/home/<user>/collibra_data** | Default: *file:///c:/collibra_data* |
| repositoryMemory | Reserved random access memory in MB for the repository service. | int | Default value: *1024* | Default value: *1024* |
| dgcMinMemory | Minimum amount of memory in MB for the DGC service. | int | Default value: *1024* | Default value: *1024* |
| dgcMaxMemory | Maximum amount of memory in MB for the DGC service. | int | Default value: *2048* | Default value: *2048* |
| dgcPort | TCP port to access the DGC service. | long int | Default value: *4400* | Default value: *4400* |

| Parameter | Description | Type | Linux example | Windows example |
|---|---|---|---|---|
| dgcShutdownPort | TCP port to shut down a Collibra DGC environment. | long int | Default value: *4430* | Default value: *4430* |
| repositoryPort | TCP port to access the repository database. | long int | Default value: *4403* | Default value: *4403* |
| consolePort | TCP port to access Collibra Console. | long int | Default value: *4402* | Default value: *4402* |
| consoleDatabasePort | TCP port to access the Collibra Console database. | long int | Default value: *4420* | Default value: *4420* |
| consoleDatabasePassword | Password used by Collibra Console to store data in its database. | string | There is no default value but you have to fill in a password if you add this parameter. Empty strings are not allowed.<br>If you don't add this parameter, the password will be automatically generated. | There is no default value but you have to fill in a password if you add this parameter. Empty strings are not allowed.<br>If you don't add this parameter, the password will be automatically generated. |

| Parameter | Description | Type | Linux example | Windows example |
|---|---|---|---|---|
| consoleDatabaseAdminPassword | Password to directly access the Collibra Console database. | string | There is no default value but you have to fill in a password if you add this parameter. Empty strings are not allowed.<br>If you don't add this parameter, the password will be automatically generated. | There is no default value but you have to fill in a password if you add this parameter. Empty strings are not allowed.<br>If you don't add this parameter, the password will be automatically generated. |
| agentPort | TCP port that is used by Collibra Console to connect to the Collibra agent for management purposes. | long int | Default value: *4401* | Default value: *4401* |
| jobserverPort | TCP port to access the Jobserver. | long int | Default value: *4404* | Default value: *4404* |
| jobserverDatabasePort | TCP port to access the Jobserver database. | long int | Default value: *4414* | Default value: *4414* |
| monitoringPort | TCP port to access the Monitoring service | long int | Default value: *4407* | Default value: *4407* |
| searchHttpPort | TCP port to access the Search service via REST API | long int | Default value: *4421* | Default value: *4421* |

| Parameter | Description | Type | Linux example | Windows example |
|---|---|---|---|---|
| searchTransportPort | TCP port for the communication between the DGC and Search service. | long int | Default value: *4422* | Default value: *4422* |
| searchMemory | The memory in MB assigned to the Search service. | int | Default value: *1024* | Default value: *1024* |
| dgcContextPath | Context path for the DGC service. | string | Default value: empty | Default value: empty |
| consoleContextPath | Context path for the Collibra Console service. | string | Default value: empty | Default value: empty |
| nodeHostName | The hostname of the node on which you are installing services.<br>If you are installing a multinode environment, you have to use this parameter with another name than *localhost*. | string | Default value: *localhost*<br>If you use this default value, the node cannot be used in multinode environments. | Default value: *localhost*<br>If you use this default value, the node cannot be used in multinode environments. |

| Parameter | Description | Type | Linux example | Windows example |
|---|---|---|---|---|
| repoAdminPassword | Admin password to access the repository database directly.<br>This should only be done by experienced database administrators. | string | There is no default value but you have to fill in a password if you add this parameter. Empty strings are not allowed.<br>If you don't add this parameter, the password will be automatically generated. | There is no default value but you have to fill in a password if you add this parameter. Empty strings are not allowed.<br>If you don't add this parameter, the password will be automatically generated. |
| repoDgcPassword | Password for the DGC service to obtain access to the repository database. | string | There is no default value but you have to fill in a password if you add this parameter. Empty strings are not allowed.<br>If you don't add this parameter, the password will be automatically generated. | There is no default value but you have to fill in a password if you add this parameter. Empty strings are not allowed.<br>If you don't add this parameter, the password will be automatically generated. |

| Parameter | Description | Type | Linux example | Windows example |
|---|---|---|---|---|
| componentSet | List of services to install:<br>• DGC<br>• REPOSITORY<br>• JOBSERVER<br>• AGENT<br>• CONSOLE<br>• MONITORING<br>• SEARCH<br>**Tip** If you install DGC, REPOSITORY, MONITORING, SEARCH and/or JOBSERVER, AGENT is automatically included. | string | Example: *DGC,CONSOLE* | Example: *DGC,CONSOLE* |
| initDaemon | Select a custom init daemon:<br>• systemv<br>• upstart<br>• systemd<br>This is a Linux only parameter.<br>Be careful when you specify an init daemon, it may result in an unstable operating system. | int | The default value is the one that is the most appropriate for your Linux system. | Not applicable |

| Parameter | Description | Type | Linux example | Windows example |
|---|---|---|---|---|
| userName | The name of the user who will install the software.<br>This is only required if the userGroup is not the same as the userName.<br>This is a Linux only parameter. | string | The default value is the one that is used to execute the installation command. | Not applicable |
| userGroup | The group to which the user belongs.<br>This is only required if the userGroup is different from the userName.<br>This is a Linux only parameter. | string | The default value is the same as the userName. | Not applicable |

**Note**

- Make sure that you add the escape character ( \ ) in the Windows paths in front of a backslash.
  Example: **C:\\collibra_data**
- Only use double quotes in the configuration file.

.

# Example input file

```
{
   "installationDirectory" : "/home/johndoe/collibra/",
   "dataDirectory" : "/home/johndoe/collibra_data/",
   "repositoryMemory" : 1024,
   "dgcMinMemory" : 1024,
   "dgcMaxMemory" : 2048,
   "dgcPort" : 4400,
```

```
    "dgcShutdownPort" : 4430,
    "repositoryPort" : 4403,
    "consolePort" : 4402,
    "consoleDatabasePort" : 4420,
    "agentPort" : 4401,
    "jobserverPort" : 4404,
    "jobserverDatabasePort" : 4414,
    "monitoringPort" : 4407,
    "searchHttpPort": 4421,
    "searchTransportPort": 4422,
    "searchMemory": 1024,
    "dgcContextPath" : "",
    "consoleContextPath" : "",
    "repoAdminPassword" : "aV3r4Str0ngP@sww0rd",
    "repoDGCPassword": "aV3r4Str0ngP@ssw0rd",
    "userName" : "johndoe",
    "userGroup" : "johndoe",
    "initDaemon" : null,
    "componentSet" : [ "CONSOLE", "JOBSERVER", "AGENT",
"REPOSITORY", "DGC", "SEARCH" ]
}
```

# Upgrade the software unattendedly

Similar to an unattended installation, you can also upgrade the software in an unattended way. To upgrade to version 5.7 or newer, you have to add the Search and Monitoring services.

## Prerequisites

- The prerequisites of a normal installation.
- You must upgrade with the same user account that was used for the installation, both on Linux and Windows. If the user account is no longer active, see Upgrade an environment with another user account.
- If you install the Search service on a Linux system, the node that will run the Search service must pass the following bootstrap checks:
    - File descriptor
    - Maximum number of threads check
    - Maximum file size
    - Maximum size virtual memory check
    - Maximum map count check

| Type | Check description | Minimum value | Applies for installation type | Setting name |
|---|---|---|---|---|
| User limit | Maximum number of open file descriptors | 65536 | ▪ Without root permissions<br>▪ With root permissions, using System V init daemon | nofile |
| | Maximum number of open threads/processes | 4096 | | nproc |
| | Maximum file size | unlimited | | fsize |
| Kernel parameter | Maximum virtual memory areas | 262144 | ▪ All | vm.max_map_count |

For more information on these settings, see the Troubleshooting section.

**Note**

- If you use a configuration file, you can edit the necessary parameters of your existing services, for example to edit a TCP port.
- In version 5.5 or newer, there is a database port, with default value 4414 for the Jobserver. If you want to use a different port after the upgrade, add the Jobserver database port in the configuration file.
- Upgrading to 5.7.0 or newer, requires you to add the Search and Monitoring services on one of the nodes.

# Linux

1. Open a terminal session.
2. Go to the directory with the installer.
3. Run the following command:

**Tip**

- If you use a configuration file, you can replace `--upgrade-config` by `-uc`.
- Use the full path to the configuration file, even if it is in the same directory as the installer.

| OS | Command |
|---|---|
| Linux (root) | ◦ Upgrade with the default options:<br>`sudo ./dgc-linux-5.7.12-0.sh -- \`<br>`--upgrade /path/to/installation`<br>◦ Upgrade with a configuration file:<br>`sudo ./dgc-linux-5.7.12-0.sh -- \`<br>`--upgrade /path/to/installation \`<br>`--upgrade-config /path/to/config`<br>◦ Upgrade and add a service:<br>`sudo ./dgc-linux-5.7.12-0.sh -- \`<br>`--upgrade /path/to/installation \`<br>`--upgrade-config /path/to/extra-service` |
| Linux (non-root) | ◦ Upgrade with the default options:<br>`./dgc-linux-5.7.12-0.sh -- \`<br>`--upgrade /path/to/installation`<br>◦ Upgrade with a configuration file:<br>`./dgc-linux-5.7.12-0.sh -- \`<br>`--upgrade /path/to/installation \`<br>`--upgrade-config /path/to/config`<br>◦ Upgrade and add a service:<br>`./dgc-linux-5.7.12-0.sh -- \`<br>`--upgrade /path/to/installation \`<br>`--upgrade-config /path/to/extra-service.cfg` |

# Windows

1. Open the command prompt.
2. Go to the directory with the installer.

3. Run the following command:

> **Tip**
> - If you use a configuration file, you can replace `--upgrade-config` by `-uc`.
> - Use the full path to the configuration file, even if it is in the same directory as the installer.

| OS | Command |
|---|---|
| Windows | ◦ Upgrade with the default options:<br>`setup.bat --upgrade \path\to\installation`<br>◦ Upgrade with a configuration file:<br>`setup.bat --upgrade /path/to/installation \`<br>`--upgrade-config /path/to/config`<br>◦ Upgrade and add a service:<br>`setup.bat --upgrade /path/to/installation \`<br>`--upgrade-config /path/to/extra-service.cfg` |

## What's next?

All Collibra DGC services will be upgraded and be readily available upon the upgrade completion.

# About the unattended upgrade

When you upgrade a node in an unattended way, you can now add extra services on that node. To do so, you have to create a new configuration file that includes at least the key `componentSet` and as value the list of services that you want to add, for example:

```
{"componentSet":["SEARCH","MONITORING"]}
```

If you don't want to use the default parameters for the added service(s), you also have to add the configuration key-value pairs for each service.

For an upgrade to 5.7 or newer, you have to add the Search and Monitoring services on one of the nodes. Adding a service to a node requires you to add an extra parameter (`--upgrade-config` or `-uc`) to the upgrade command.

Example Search service configuration:

```
{
    "componentSet": ["SEARCH"],
    "searchHttpPort": 4421,
    "searchTransportPort": 4422,
    "searchMemory": 1024
}
```

Chapter 5

# Reinstall Collibra DGC

By reinstalling the Collibra Data Governance Center's services, you can fix potential corrupted files. The reinstallation also allows you to add extra services.

## In this chapter

# Reinstall Collibra DGC

In this section, we describe a reinstallation of Collibra Data Governance Center. This will only reinstall the services that were previously installed on the node with the same installer as the original installation.

## Prerequisites

- You must reinstall with the same user account that was used for the installation, both on Linux and Windows. If the user account is no longer active, see Upgrade an environment with another user account.
- You have enough free disk space in the volume that hosts the data folder, **collibra_data**. The free disk space must be at least the size of the current data.
  For example, if your data in the data folder takes 5 GB, you need at least 5 GB of free disk space on that volume to upgrade.
- You have the same installer as the original installation.

## Steps

1. Stop the environment.
2. Stop the Collibra Agent and Collibra Console.
3. Start the installation wizard:
   - Linux as user with sudo rights: `sudo ./dgc-linux-5.7.12-0.sh`
     Linux as root user: `./dgc-linux-5.7.12-0.sh`
   - Linux as standard user: `./dgc-linux-5.7.12-0.sh`
   - Windows Server: double-click **setup.bat**

     > **Important** The path of the installer file cannot contain spaces.

     If you run the installation without Administrator rights, an error is shown.

> **Tip** If you don't want to use the user interface even if it's available, add the following to the command:
> `-- --nox11`

4. Click **Next**.
5. Select the original installation directory and click **Update**.

   > **Note** If you select a different installation directory, you will add a new Collibra DGC installation next to the existing installation.

6. Click **Yes** to confirm that you have created a backup and that all the services are stopped on the node.
   » The **Component selection** dialog box appears, indicating which services are installed on the node.
7. Click **Update**.
   » The installed services on the node are reinstalled.
8. Click **Exit**.

   > **Note** If you have a multinode installation, repeat the previous steps for every node of the environment until you have reinstalled all nodes.

9. Start Collibra Console.
10. Start the environment.

# Reinstall Collibra DGC and add service

When you reinstall the services on a node, you can also add one or more extra services.

> **Tip** You can also add services when you upgrade the services.

## Prerequisites

- You must reinstall with the same user account that was used for the installation, both on Linux and Windows. If the user account is no longer active, see Upgrade an environment with another user account.
- You have enough free disk space in the volume that hosts the data folder, **collibra_data**. The free disk space must be at least the size of the current data.
  For example, if your data in the data folder takes 5 GB, you need at least 5 GB of free disk space on that volume to upgrade.

- You have to use the same installer as the original installation.
- If you install the Search service on a Linux system, the node that will run the Search service must pass the following bootstrap checks:
  - File descriptor
  - Maximum number of threads check
  - Maximum file size
  - Maximum size virtual memory check
  - Maximum map count check

| Type | Check description | Minimum value | Applies for installation type | Setting name |
|---|---|---|---|---|
| User limit | Maximum number of open file descriptors | 65536 | ▪ Without root permissions<br>▪ With root permissions, using System V init daemon | nofile |
| | Maximum number of open threads/processes | 4096 | | nproc |
| | Maximum file size | unlimited | | fsize |
| Kernel parameter | Maximum virtual memory areas | 262144 | ▪ All | vm.max_map_count |

For more information on these settings, see the Troubleshooting section.

# Steps

1. Stop the environment.
2. Stop the Collibra Agent and Collibra Console.
3. Start the installation wizard:
   - Linux as user with sudo rights: `sudo ./dgc-linux-5.7.12-0.sh`
     Linux as root user: `./dgc-linux-5.7.12-0.sh`
   - Linux as standard user: `./dgc-linux-5.7.12-0.sh`
   - Windows Server: double-click **setup.bat**

     > **Important** The path of the installer file cannot contain spaces.

     If you run the installation without Administrator rights, an error is shown.

> **Tip** If you don't want to use the user interface even if it's available, add the following to the command:
> `-- --nox11`

4. Click **Next**.
5. Select the installation directory of the old version and click **Update**.

> **Note** If you select a different installation directory, you will add a new Collibra DGC installation next to the existing installation.

6. Click **Yes** to confirm that you have created a backup and that all the services are stopped on the node.
   » The **Component selection** dialog box appears, indicating which services are installed on the node.
7. Select the services that you want to add to the node and click **Next**.
8. In the next dialog boxes, configure the newly selected services. After the last configuration dialog box, click **Update**.
   » The installed services on the node are upgraded and newly selected services are installed.
9. Click **Exit**.

> **Note** If you have a multinode installation, repeat the previous steps on every node of the environment until you have reinstalled all nodes.

10. Start Collibra Console.
11. Add the new services to your environment.

You have reinstalled the services on a node and added one or more extra services.

Chapter 6

# Uninstall Collibra DGC

In this section, you learn how to remove the Collibra Data Governance Center software from your servers.

## In this chapter

# Uninstall Collibra DGC on Linux

When you no longer need the Collibra Data Governance Center software on your server(s), you can remove it.

This section describes how to do this safely on Linux systems.

## Prerequisites

- If you have installed the software with root permissions, you must have root privileges to uninstall the software.
- You have stopped the environment.

## Steps

1. Open a terminal session on your system.
2. Go to the installation directory
   - Default location on Linux as root or user with sudo privileges: **/opt/collibra**
   - Default location on Linux as standard user: **~/collibra**
3. Start the uninstall script:
   - Uninstall with root permissions: `sudo ./uninstall.sh`
   - Uninstall without root permissions: `./uninstall.sh`

   » The **Uninstall** wizard starts.
4. Enter **yes** to continue.
5. Enter **y** if you want to delete the data directory or press **Enter** to keep the data.
6. Press **Enter** to end the wizard.

```
user@linux:~/collibra$ ./uninstall.sh
00:27:03.327 - SUCCESS - Validate user
Installer will proceed to the removal of Collibra services.
Please make sure you have backed up all important data and shut
down all services.
Have these steps been done ?
Type "yes" to continue.
yes
00:27:10.396 - SUCCESS - Confirm uninstallation
```

```
00:27:13.553 - SUCCESS - Stop Agent service
00:27:16.678 - SUCCESS - Stop Console service
Delete data directory? [y/N]
y
00:27:23.267 - SUCCESS - Delete data directory
00:27:23.648 - SUCCESS - Delete installation directory
Uninstallation finished.
Press enter to exit.
```

If Collibra DGC has been installed on multiple nodes, repeat this procedure on every node.

> **Note** If you did not stop the environment before uninstalling the software:
>
> - Restart the server and remove the services manually.
> - Check whether the Collibra installation and data directory still exist. If they do, remove them manually.

# Uninstall Collibra DGC on Windows

When you no longer need the Collibra Data Governance Center software on your server(s), you can remove it. In this section, you learn how to do this safely on Windows systems.

## Prerequisites

- You have administrative rights on your Windows system.
- You have stopped the environment. See Stop an environment.

## Steps

1. In Windows Explorer, go to your Collibra installation directory, the default is **C:\collibra**.
2. Double-click **uninstall.bat** to start the **Uninstall** wizard.
3. In the command prompt, type **yes** to continue.
   If you see user account control warnings, click **Yes** for each of the requests, if you click **No**, the removal of the software will fail.
4. Enter **y** if you want to delete the data directory or press **Enter** to keep the data.

5. Press **Enter** to complete the uninstall.
6. Delete the Collibra installation directory.

If Collibra DGC has been installed on multiple nodes, repeat this procedure on every node.

> **Note** If you did not stop the environment before uninstalling the software:
>
> - Restart the server and remove the services manually.
> - Check whether the Collibra installation and data directory still exist. If they do, remove them manually.

# Uninstall Collibra Console

If you have a multinode installation and you want to uninstall the Collibra Console node, you have to follow these steps before uninstalling the Collibra Console software.

These are the steps to execute:

1. In Collibra Console, stop every environment, see Stop an environment.
2. Remove all services of all environments:
   a. In the main menu, click **Infrastructure**.
   b. Click an environment to open its details.
   c. Click 🗑 next to a service.
   d. Click **Remove** to confirm the removal of the environment's service.
   e. Repeat this for all the services of an environment.
3. Remove all nodes.
4. Sign out from Collibra Console

You can now uninstall the software from the Collibra Console node, see Uninstall Collibra DGC on Linux or Uninstall Collibra DGC on Windows.

Chapter 7

# Collibra DGC service management

When you install Collibra Data Governance Center, services are installed on your system (Linux as well as Windows). In this section, you learn more about these services and how you can manage them.

## In this chapter

# Collibra services on Linux with root permissions

If you install Collibra Data Governance Center as a root user on a Linux system, the following items are also installed:

- The **Collibra Agent** service
- The **Collibra Management Console**
- An extra user: **collibra**. The extra user is necessary to run the Collibra DGC software. You cannot sign in with this user.

> **Note** If you installed Collibra DGC as a normal user, see Collibra DGC services on Linux as non-root user.

To manage the Collibra DGC services, use the default service management tool of your operating system:

```
service collibra-agent <command>
service collibra-console <command>
```

You can use the following commands to control the services:

- start
- stop
- restart
- status

> **Note** The services are automatically started after a restart of a node.

# Collibra services on Linux as non-root user

If you have installed Collibra Data Governance Center as a non-root user on Linux, there are no services added to the Linux services list.

> **Note** If you installed Collibra DGC as root, see Collibra DGC services on Linux.

You have to manage the Collibra services manually or install them afterwards.

Similar to the installation with the root user, you can only manage the agent and Collibra Console services. The other services in an environment are then managed via the Collibra Console user interface.

In the installation directory, by default **~/collibra**, you find each Collibra service as a separate directory:

- agent
- console

Each of these directories has a **bin** directory, which contains the script to manage these services.

By default, you have to start these services manually after the restart of the node.

# Run the script

- agent: `./agent <command>`
- console: `./console <command>`

> **Tip** If you use the install command, the service is added as daemon, which can then be configured to start with the start of your operating system.

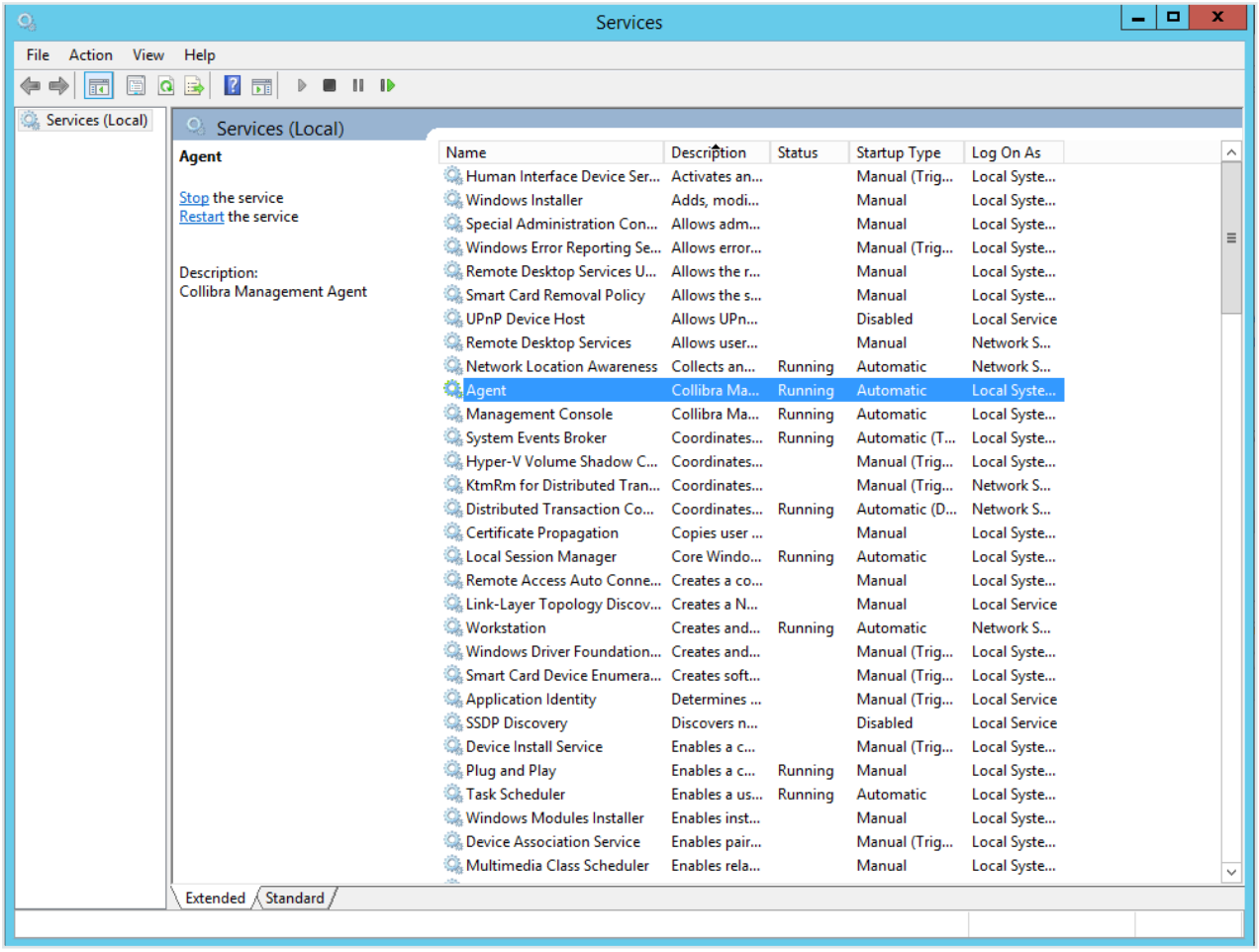
# Commands

| Command | Description |
|---|---|
| console | Start in the current console. |
| start | Start in the background as a daemon process. |
| stop | Stop if running as a daemon or in another console. |

| Command | Description |
|---|---|
| restart | Stop if running and then start. |
| condrestart | Restart only if already running. |
| status | Display the current status. |
| install | Install the service as daemon and manage further through the operating system's service management. |
| remove | Remove the service from the node. |
| dump | Request a Java thread dump if running. |

# Collibra services on Windows

When you install Collibra Data Governance Center on a Windows system, two services are installed as well, **Agent** and **Management Console**.




The **Startup Type** of both services must be *automatic* or Collibra DGC is not available when the server is restarted.

To manage these services, right-click on the service and select the operation that you want to perform.

Chapter 8

# Troubleshooting

# DGC service fails to start due to invalid configuration

If Collibra DGC 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.

# Error "unexpected EOF" in monitoring logs

The Collibra Data Governance Center monitoring service encounters connectivity issues if you have configured your network to forward all traffic via a proxy.

In the monitoring log file (**~/collibra_data/monitoring/logs/prometheus.log**), you find the following message:

```
yyyy/mm/dd hh:mm:ss transport: http2Client.notifyError got noti-
fied that the client transport was broken unexpected EOF.
```

## Resolution

Add a `no_proxy` entry for traffic to the monitoring service.

> **Tip** The monitoring service is configured to listen on 0.0.0.0:4407 by default.

# Installation log files

If you encounter an installation failure, you can consult the installation log files for a first investigation.

The installation log file is stored in the Collibra installation directory:

- Default location on Linux as root or user with sudo privileges: **/opt/collibra**
- Default location on Linux as standard user: **~/collibra**
- Default location on Windows Server: **C:\collibra**On Windows, the target installation directory cannot contain spaces.

The log file is **installation.log**.

Only in the event that the installer was not able to create the installation directory, there will be no installation log file.

# Upgrade fails on backup

One of the steps during an upgrade is the creation of a backup. If creating the backup takes too much time, the upgrade will fail.

## Resolution

Increase the timeout settings of the backup, the default value is 12 hours or 43 200 000 milliseconds.

# Environment in error state during the upgrade

If you upgrade an environment, it may occur that starting the Data Governance Center service times out, 30 minutes is the default. If this occurs, your environment will be in an Error state.

> **Warning** The upgrade process is not completed yet, do not start any other action.

## Resolution

Most likely, the upgrade is still running. You can follow the upgrade process in the DGC log file.

To do so, follow these steps:

1. In the tab pane, click the environment that is in error.
2. Click the Data Governance Center service and go to the **Logs** tab.
3. Click **dgc.log** to open the log file
4. Click ↻ or select the **Auto refresh** option.

The upgrade is completed when the DGC service and the whole environment is back up and running.

# Error "Version mismatch"

If you upgrade a Collibra Data Governance Center 5.0 to 5.1 or newer by overwriting an existing setup, it may occur that you get an error if you start the Data Governance Center service before you have upgraded the repository.

## Error

Version mismatch
There is a version mismatch between Data Governance Center and linked repository. Please upgrade the environment to be able to start the service.

## Resolution

Upgrade the repository as described in Upgrade - scenario 2.

# Upgrade an environment with another user account

If you upgrade an environment by installing a new version over an old version, you have to use the same user account that was used for the old installation.

However, it is possible that the specific user account is no longer available, for example because the user has left the company. If you upgrade an environment with another user account, you will receive an error message during the procedure:

```
connection to database failed: FATAL: role "<other user>" does
not exist.
could not connect to source postmaster with the command: ...
```

> **Tip** This is only applicable for Linux operating systems, on Windows you can install with any user who has administrator rights.

## Prerequisites

- The Repository service is online.
- You know the password to access the Repository service.

# Resolution

1. On the node that hosts the Repository service, open a shell session.
2. Retrieve the username that you will use to upgrade the environment: `whoami`
   If there is a domain name/username displayed, for example `mydomain\john.smith`, you can ignore the domain name. The username is `john.smith`.
3. Go to the repository directory: `cd <install location>/collibra/repo/bin`, for example `cd /opt/collibra/repo/bin`
4. Connect to the database using the repository's administrator password.
   » A PostgreSQL session starts.

```
./psql -p 4403 -U collibra postgres
Password for user collibra:<repo admin password>
psql.bin (10.3)
Type "help" for help.

Cannot read termcap database;
using dumb terminal settings.
postgres=#
```

5. Update the bootstrap super user with the new username, where you have to replace francois.lemaire by your own username, found in step 2.

```
update pg_authid set rolname='francois.lemaire' where oid=10;

select rolname from pg_authid where oid=10;
```

   For example:

```
postgres=# update pg_authid set rolname='francois.lemaire'
\
    where oid=10;
UPDATE 1
postgres=# select rolname from pg_authid where oid=10;
 rolname
---------
 francois.lemaire
(1 row)
```

Make sure that the second command returns the username that you will use to do the upgrade.

6. Leave the PostgreSQL session: `\q`
7. Close the shell session.
8. In Collibra Console, stop the complete environment and upgrade the environment.

# DGC does not start after an upgrade

If you upgraded a multi-node environment, it is possible that the environment doesn't start anymore.

In Collibra Console you see the following error message:

```
Multi nodes environment cannot mix loopback and public
addresses.
```

To work around this issue, you can choose one of the following procedures:

- via the Console configuration file.
- via the Collibra Console user interface.

## Collibra Console configuration file

1. Open an SSH session to the node on which Collibra Console runs.
2. Back up **console.db**:
   ```
   cp /collibra_data/console/console.db \
    /collibra_data/console/console.db.bck
   ```
   > **Note** The path to the file can be different for your environment.
3. Stop Collibra Console.
4. Open the file **/collibra_data/console/console.db** for editing.
5. In the **nodeSet** section, look up the nodes that have a key **hostName** with value `localhost`.

```
"nodeSet" : [ {
     ...
    "hostName" : "localhost",
    "port" : 8081,
    "name" : "NODE NAME",
    "registered" : true,
    ...
  } ],
```

6. Replace `localhost` by the private IP address of that node and repeat this for all nodes in this **nodeSet** section.
7. Save and close the file.
8. Start Collibra Console.
9. Start the environment.

# Collibra Console user interface

1. Open Collibra Console with a user profile that has the **SUPER** role.
   » Collibra Console opens with the **Infrastructure** page.
2. Remove the relevant services from the environment and optionally repository clusters. The relevant services are the services that are installed on the nodes which are configured as localhost.
3. Remove the node from Collibra Console.
4. Add the node again in Collibra Console using the private IP address of the node.

> **Note** If you receive the following error, follow the procedure as described on Collibra Community.
>
> ```
> Error while trying to set up a new node. Is the node
> up and address correct?
> certificateRegistrationFailed Message:
> javax.ws.rs.ProcessingException:
> java.net.SocketException:
> Unexpected end of file from server.
> ```

5. Add the services on that node to the environment and or repository cluster.
6. Start the environment.

# SAML no longer works after an upgrade

If you upgrade a Collibra Data Governance Center 5.5.x or older with a SAML configuration to 5.6 or newer, everything should be upgraded in a seamless way.

In the event that the upgrade did break the SAML configuration, upload your SAML configuration via Collibra Console.

# Select a custom temporary folder

When you install Collibra Data Governance Center, you need about 5 GB of free disk space in the temporary folder. This is the location where the installer will first extract all files before it can install the software in the selected location.

If your operating system does not have enough free disk space in its default temporary folder, you can select a temporary folder of your choice with the `--target` option.

- Linux as root user: `sudo ./dgc-linux-5.7.12-0.sh --target /path/to/-custom/tmp/folder`
- Linux as non-root user: `./dgc-linux-5.7.12-0.sh --target /path/to/-custom/tmp/folder`
- Windows: `setup.bat --target /path/to/custom/tmp/folder --config <path/to/config>`

# No write access to /tmp

On Linux, you need write permissions to the **/tmp** folder to install Collibra Data Governance Center, even if you select a custom temporary folder.

If for any reason you don't have this permission, you have to use the following command to install our software:

```
JAVA_TOOL_OPTIONS="-Djava.io.tmpdir=$HOME" sh dgc-linux-5.7.1-23.sh \
--target $HOME
```

You can choose the value of $HOME freely, but make sure that the location has enough space and that you use the same path in both occurrences of the command.

# Solve JobServer memory errors for 5.7.1 or older

If you are using an installation of JobServer version 5.7.1 or older, you may experience memory errors. To resolve these errors, do the following:

1. Open Collibra Console.
   » Collibra Console opens with the **Infrastructure** page.
2. In the tab pane, click the Jobserver whose configuration you want to edit.
3. Click **Infrastructure configuration**.
4. Click **Edit configuration**.
5. Edit the option **Spark memory** to *40G*.
6. Click **JVM configuration**.
7. Edit **-XX:+UseG1GC** to *-XX:+UseParallelGC*.
8. Click **Context JVM configuration** if available. If not, skip the next step.
9. Edit **-XX:+UseG1GC** to *-XX:+UseParallelGC*.
10. Click **Save all**.
11. Restart the Jobserver.

If the above is not available in your Collibra Console, then proceed as follows:

1. On the server that runs the Jobserver service, go to **%<collibra installation directory>%/spark-jobserver/conf/**.

   > **Tip** The default installation directory on Linux is **/opt/collibra**, on Windows **C:\collibra**

2. Open the file **jobserver.conf** for editing.
3. In the **spark.context-settings** section, edit the driver memory (heap memory) to 40 GB:

   ```
   driver-memory="40G"
   ```

4. Save and close the file.
5. Open the file **jobserver.default.conf** for editing.

6. In the **spark.context-settings** section, edit the driver memory (heap memory) to 40 GB:

```
driver-memory="40G"
```

7. Save and close the file.
8. Open the file **jvm.conf** for editing.
9. Replace the **-XX:+UseG1GC** option by *-XX:+UseParallelGC*.
10. Save and close the file.
11. Open the file **jvm.default.conf** for editing.
12. Replace the **-XX:+UseG1GC** option by *-XX:+UseParallelGC*.
13. Save and close the file.
14. Open the file **context_jvm.conf** for editing, if available. If the file is not available, skip the next two steps.
15. Replace the **-XX:+UseG1GC** option by *-XX:+UseParallelGC*.
16. Save and close the file.
17. Restart the Jobserver service via Collibra Console.

# Settings for the Search service requirements

## Kernel parameter

```
# Get kernel parameter value
$ sysctl vm.max_map_count
vm.max_map_count = 65530

# Update kernel parameter
$ sysctl -w vm.max_map_count=262144
vm.max_map_count = 262144
```

# User limits

## Session

```
# Get user limit values
$ ulimit -n -f -u
open files                      (-n) 4096
file size               (blocks, -f) unlimited
max user processes              (-u) 1024

# Update user limits
$ ulimit -n 65536 -f unlimited -u 4096
```

## Process

```
$ prlimit -u -n -f --pid <pid>
RESOURCE DESCRIPTION                    SOFT      HARD UNITS
NPROC    max number of processes        4096      4096
NOFILE   max number of open files      65536     65536
FSIZE    max file size             unlimited unlimited blocks

$ cat /proc/<pid>/limits
Limit                     Soft Limit           Hard Limit
    Units
Max cpu time              unlimited            unlimited
    seconds
Max file size             unlimited            unlimited
    bytes
Max data size             unlimited            unlimited
    bytes
Max stack size            8388608              unlimited
    bytes
Max core file size        0                    unlimited
    bytes
Max resident set          unlimited            unlimited
    bytes
Max processes             4096                 4096
    processes
Max open files            65536                65536
    files
Max locked memory         65536                65536
    bytes
Max address space         unlimited            unlimited
```

```
    bytes
Max file locks            unlimited            unlimited
    locks
Max pending signals       15078                15078
    signals
Max msgqueue size         819200               819200
    bytes
Max nice priority         0                    0
Max realtime priority     0                    0
Max realtime timeout      unlimited            unlimited
    us
```

## System

To change the user limits you can edit the **/etc/security/limits.conf** file or **/etc/security/limits.d/*.conf**, as an example:

```
#<domain>      <type>  <item>         <value>
collibra        soft    nproc           4096
collibra        hard    nproc           4096
collibra        soft    nofile          65536
collibra        hard    nofile          65536
collibra        soft    fsize           unlimited
collibra        hard    fsize           unlimited
```

**Tip**
It is possible that an upgrade fails due to issues with these settings. In that case, add the same configuration lines as in the above example, but replace `collibra` by the user account that is executing the upgrade.

Error message excerpt:

```
Maximum file descriptors [4096] for Search is too low,
increase to at least [65536]. ...
```

## Daemon/services

| Daemon | Specification |
| --- | --- |
| systemd | systemd configuration file:<br>`...`<br>`[Service]`<br>`...`<br>`LimitNOFILE=65536`<br>`LimitNPROC=4096`<br>`LimitFSIZE=unlimited`<br>`...` |
| upstart | upstart configuration file:<br>`...`<br>`limit nofile 65536 65536`<br>`limit nproc 4096 4096`<br>`limit fsize unlimited unlimited`<br>`exec ...` |
| System V | System V services inherit the user limits. |

# No improvements in Escalation Process after upgrade to Collibra DGC 5.7.2-13 or newer

The Collibra Data Governance Center 5.7.2-13 release fixed performance issues with the Escalation Process workflow.

To take advantage of the improvements, you must deploy the new version of the Escalation Process workflow in your Collibra DGC 5.7.2-13 or newer.

> **Note** If you are using a modified Escalation Process workflow, you must port your changes to the new workflow.

**Warning** The new Escalation Process workflow is only valid for Collibra DGC 5.7.2-13 or newer.

# Workflow is broken after upgrade from Collibra DGC 5.5.2 or older to 5.6.0 or newer

Valid BPMN workflow files must have the same value for :

- the **`processRef`** attribute of the `participant` tag.
- the **`id`** attribute of the `process` tag.

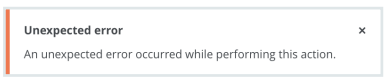
The workflow engine in Collibra DGC 5.5.2 or older ignores this requirement and accepts the file.

**Example of an invalid BPMN file**

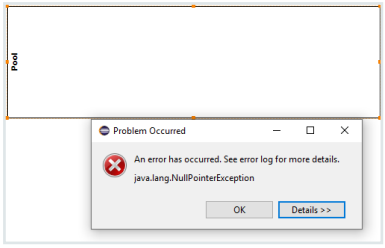
```
...
<collaboration id="Collaboration">
    <participant id="pool1" name="Pool" processRef="process_
pool1"></participant>
</collaboration>
<process id="process_pool2" name="Process Pool" isEx-
ecutable="true">
    <laneSet id="laneSet_process_pool2">
        <lane id="lane1"></lane>
    </laneSet>
</process>
...
```

Invalid BPMN workflow files may lead to the following:

- After an upgrade to Collibra DGC 5.6.0 or newer, the workflows with invalid BPMN files do not work and result in a `java.lang.NullPointerException` error, visible in the Collibra Console logs.
- Uploading the invalid BPMN file to Collibra DGC 5.6.0 or newer results in an **Unexpected error**.



- Editing the invalid BPMN file in Eclipse IDE with the Flowable Diagram Editor results in an empty canvas or an empty pool, and a `java.lang.NullPointerException` error.




To solve this issuse, edit the invalid BPMN file with a text editor and use the same value for the **`processRef`** and **`id`** attributes.

# Overview build numbers

For 5.7 versions older than 5.7.7, there is a difference between the build number that is shown in Collibra Data Governance Center and in Collibra Console. For those versions, the installation files have the build number that is shown in Collibra Console. The installer contains among other files, the Collibra DGC package, which has a different build number. This is the build number that is shown in Collibra DGC.

These are the build numbers of all 5.7 releases prior to 5.7.7:

| Installer build number | Collibra DGC build number |
|---|---|
| 5.7.0-60 | 5.7.0-59 |
| 5.7.1-19 | 5.7.1-20 |
| 5.7.1-23 | 5.7.1-22 |
| 5.7.2-4 | 5.7.2-3 |
| 5.7.2-9 | 5.7.2-8 |
| 5.7.2-13 | 5.7.2-12 |
| 5.7.2-14 | 5.7.2-12 |
| 5.7.2-16 | 5.7.2-14 |
| 5.7.3-25 | 5.7.3-19 |
| 5.7.4-14 | 5.7.4-10 |
| 5.7.4-20 | 5.7.4-18 |
| 5.7.4-22 | 5.7.4-20 |
| 5.7.4-26 | 5.7.4-22 |
| 5.7.4-29 | 5.7.4-26 |
| 5.7.4-39 | 5.7.4-40 |

| Installer build number | Collibra DGC build number |
|---|---|
| 5.7.4-42 | 5.7.4-43 |
| 5.7.5-44 | 5.7.5-46 |
| 5.7.5-49 | 5.7.5-54 |
| 5.7.5-50 | 5.7.5-55 |
| 5.7.5-53 | 5.7.5-58 |
| 5.7.5-65 | 5.7.5-71 |
| 5.7.5-77 | 5.7.5-90 |
| 5.7.6-103 | 5.7.6-125 |

# Appendix A - Component versions

| Component | Product | Version |
|---|---|---|
| Virtual Machine (jreVersion) | Azul Zulu JRE | Azul Zulu 8.0.312 |
| Repository (post-gresVersion) | PostgreSQL | 10.17 |
| Jobserver (sparkVersion) | Spark | 2.4.8-collibra-11 |
| Search | Elasticsearch | 7.16.2 |

# Appendix B - Overview default ports in Collibra DGC

The following table contains an overview of the default ports that are used in Collibra Data Governance Center.

| Port | Default value | Description |
|---|---|---|
| DGC service | 4400 | The TCP port to access your Collibra DGC environment via your web browser. |
| Repository service | 4403 | The TCP port to access the repository service. It is only used by the DGC service and the Collibra agent. |
| Console application | 4402 | The TCP port to access your Collibra Console via your web browser. |
| Console database | 4420 | The TCP port to access the database of Collibra Console. |
| Agent application | 4401 | The TCP port that is used by Collibra Console to manage the services in a Collibra DGC environment. |
| Jobserver service | 4404 | The TCP port to access the Jobserver service. |
| Jobserver database | 4414 | The TCP port to access the Jobserver database. |
| DGC shutdown port | 4430 | The TCP port through which you can stop the DGC service. |
| Console AJP port | n/a | The Apache Jserv Protocol to access Collibra Console. |
| Agent AJP port | n/a | The Apache Jserv Protocol to connect to the agent. |
| DGC AJP port | n/a | The Apache Jserv Protocol to access the DGC service. |
| Monitoring port | 4407 | The TCP port to access the monitoring service. |

| Port | Default value | Description |
|---|---|---|
| Search HTTP port | 4421 | The TCP port to access the Search service. |
| Search Transport port | 4422 | The TCP port used by the DGC service to communicate with the Search service. |
| Jobserver monitoring port | 4424 | The port that is used by the Monitoring service to monitor the Jobserver service. |
| Jobserver Spark monitoring port | 4434 | The port that is used by the Monitoring service to monitor the Spark service. |

# Appendix C - Plain-text attributes

When you upgrade to 5.7 or newer, characteristics of some Catalog-related assets are converted from rich-text format to plain-text format.

The following table contains the list of characteristics with their corresponding unique identifiers:

| Name | Unique identifier |
|---|---|
| Original name | 00000000-0000-0000-0001-000500000032 |
| Location | 00000000-0000-0000-0000-000000000203 |
| Technical Data Type | 00000000-0000-0000-0000-000000000219 |
| Schema name | 00000000-0000-0000-0000-000000000226 |
| File location | 00000000-0000-0000-0001-000500000004 |
| Table Type | 00000000-0000-0000-0001-000500000008 |
| Primary Key Name | 00000000-0000-0000-0001-000500000016 |
| Minimum Value | 00000000-0000-0000-0001-000500000040 |
| Maximum Value | 00000000-0000-0000-0001-000500000041 |
| Date and/or Time Pattern | 00000000-0000-0000-0001-000500000044 |
| Mode | 00000000-0000-0000-0001-000500000048 |
| 1st Percentile | 00000000-0000-0000-0001-000500000049 |
| 5th Percentile | 00000000-0000-0000-0001-000500000050 |
| 1st Decile | 00000000-0000-0000-0001-000500000051 |
| 1st Quartile | 00000000-0000-0000-0001-000500000052 |
| Median | 00000000-0000-0000-0001-000500000053 |

| Name | Unique identifier |
| --- | --- |
| 3rd Quartile | 00000000-0000-0000-0001-000500000054 |
| Category | 00000000-0000-0000-0001-000500000046 |
| 9th Decile | 00000000-0000-0000-0001-000500000055 |
| 95th Percentile | 00000000-0000-0000-0001-000500000056 |
| 99th Percentile | 00000000-0000-0000-0001-000500000057 |
| Empty values definition override | 00000000-0000-0000-0001-000500000063 |
| File Type | 00000000-0000-0000-0001-002500000012 |
| Glue database name | 00000000-0000-0000-0001-000500000066 |
| Glue table name | 00000000-0000-0000-0001-000500000067 |

# Appendix D - 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 Governance Center 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.<br>• **/10*: every 10 seconds. |
| 2 | minute | 0-59 | , - * / | • *30*: at the 30th minute.<br>• **/15*: every 15 minutes.<br>• *5/10*: every 10 minutes starting at the 5th minute after the hour |
| 3 | hour | 0-23 | , - * / | • *10*: at 10 o'clock.<br>• *8-10*: at 8,9 and 10 AM.<br>• *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.<br>• *1-4*: every first four days of the month.<br>• *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.<br>• *1-3*: every first three months of the year.<br>• *JUL,AUG*: every July and August.<br><br>**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.<br>• *1-5*: every weekday, Monday to Friday.<br>• *MON,WED,FRI*: every Monday, Wednesday and Friday.<br>• *L*: on Sunday, the 7th day of the week.<br>• *1L*: at the last Monday of the month.<br>• *5#3*: on the 3rd Friday of the month.<br><br>**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.<br><br>**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.<br><br>**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.<br><br>**Example** `10-12` in the hour field means “the hours 10, 11 and 12”. |
| , | Used to specify additional values.<br><br>**Example** `MON,WED,FRI` in the day-of-week field means “the days Monday, Wednesday, and Friday”. |
| / | Used to specify increments.<br><br>**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".<br>You can also leave out the number before /, which is equivalent to having 0 before /.<br>`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.<br><br>**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.<br><br>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".<br><br>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.<br><br>**Example** `15W` in the value for the day-of-month field, means the nearest weekday to the 15th of the month:<br><br>• If the 15th is a Saturday, the trigger will fire on Friday the 14th.<br>• If the 15th is a Sunday, the trigger will fire on Monday the 16th.<br>• If the 15th is a Tuesday, then it will fire on Tuesday the 15th.<br><br>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.<br><br>**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.<br><br>**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).<br>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 Governance Center 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.<br>• **/10*: every 10 seconds. |
| 2 | minute | Yes | 0-59 | , - * / | • *30*: at the 30th minute.<br>• **/15*: every 15 minutes.<br>• *5/10*: every 10 minutes starting at the 5th minute after the hour |
| 3 | hour | Yes | 0-23 | , - * / | • *10*: at 10 o'clock.<br>• *8-10*: at 8,9 and 10 AM.<br>• *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.<br>• *1-4*: every first four days of the month.<br>• *1,15*: the first day of the month and the 15th day of the month.<br>• *L*: on the last day of the month.<br>• *L-3*: on the third-to-last day of the month.<br>• *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.<br>> **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.<br>*LW*: on the last weekday of the month. |
| 5 | month | Yes | 1-12 or JAN-DEC | , - * / | • *12*: in December.<br>• *1-3*: every first three months of the year.<br>• *JUL,AUG*: every July and August.<br>> **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.<br>• *2-6*: every weekday, Monday to Friday.<br>• *MON,WED,FRI*: every Monday, Wednesday and Friday.<br>• *L*: on Saturday, the 7th day of the week.<br>• *2L*: at the last Monday of the month.<br>• *6#3*: on the 3rd Friday of the month.<br><br>**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.<br>• *2021*: in 2021.<br>• *2021-2025*: in the years 2021, 2022, 2023, 2024 and 2025.<br>• *2021,2022,2025*: in the years 2021, 2022 and 2025. |

# Special characters

| Character | Description |
|---|---|
| * | Used to select all values within a field.<br><br>**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.<br><br>**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.<br><br>**Example** `10-12` in the hour field means “the hours 10, 11 and 12”. |
| , | Used to specify additional values.<br><br>**Example** `MON,WED,FRI` in the day-of-week field means “the days Monday, Wednesday, and Friday”. |
| / | Used to specify increments.<br><br>**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".<br>You can also leave out the number before /, which is equivalent to having 0 before /.<br>`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.<br><br>**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.<br><br>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".<br><br>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.<br><br>**Example** `15W` in the value for the day-of-month field, means the nearest weekday to the 15th of the month:<br><br>• If the 15th is a Saturday, the trigger will fire on Friday the 14th.<br>• If the 15th is a Sunday, the trigger will fire on Monday the 16th.<br>• If the 15th is a Tuesday, then it will fire on Tuesday the 15th.<br><br>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.<br><br>**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.<br><br>**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).<br>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.