Collibra

Collibra Data Intelligence Cloud

Cloud Infrastructure

Version 2024.02

Collibra

Collibra Data Intelligence Cloud - Cloud Infrastructure

Release date: February 4, 2024

Revision date: February 01, 2024

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

https://productresources.collibra.com/docs/collibra/latest/#cshid=cloud-infrastructure

Contents

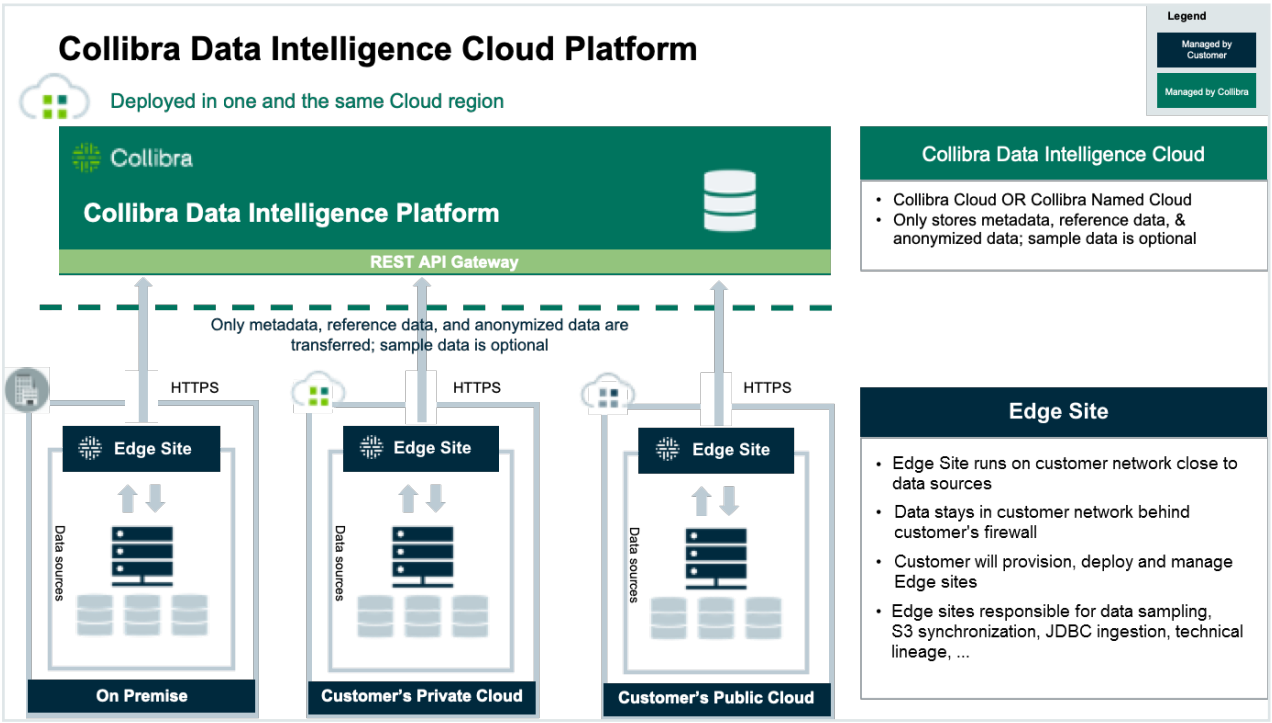
Collibra Data Intelligence Cloud

The Collibra Data Intelligence Cloud is delivered as a Software-as-a-Service offering. This section describes the Collibra Cloud Infrastructure processes including network security, backup and recovery processes, monitoring and security.

You can keep your data where it is today or where it makes sense tomorrow - while benefiting from the enhanced scalability, availability and performance that cloud offers.

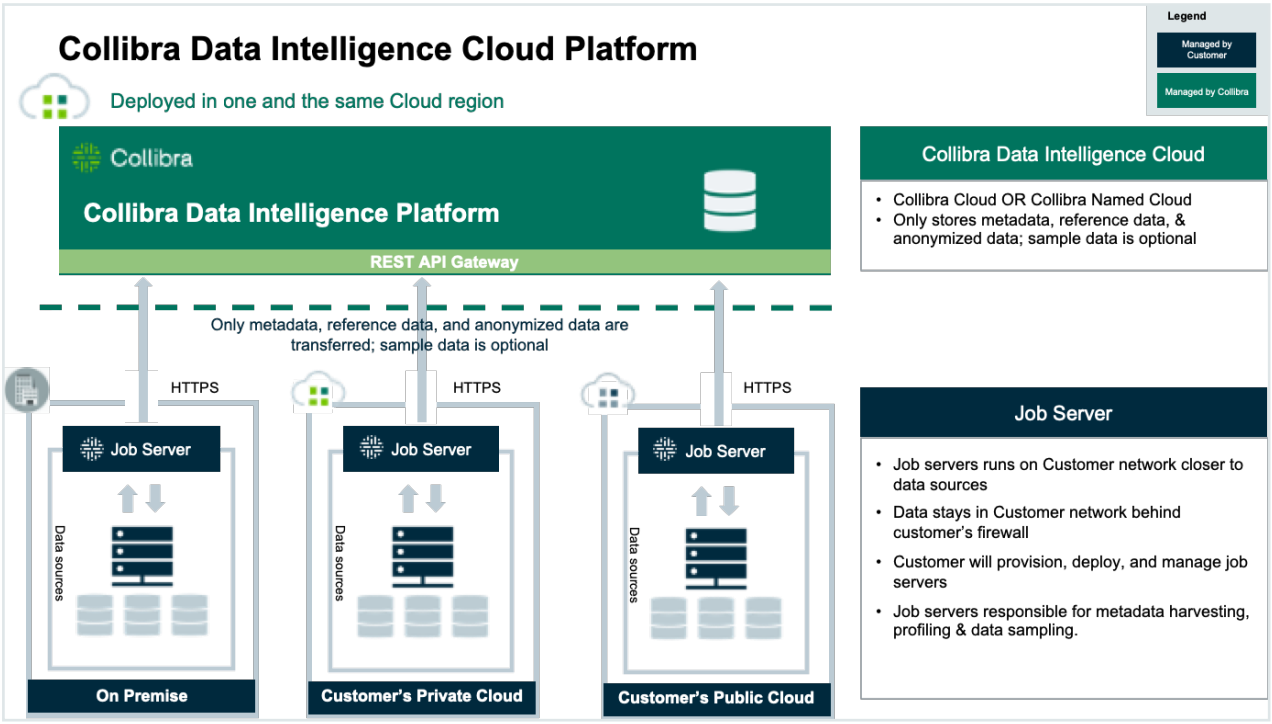
Collibra and Edge

The following diagram shows the role of Edge and how it connects to Collibra. Edge connects to your data sources to perform actions like data profiling and then provides that information back to Collibra.



Collibra and Jobserver

The following diagram shows the role of a Jobserver and how it connects to Collibra. The Jobserver connects to your data sources to perform actions like data profiling and then provides that information back to Collibra.



Collibra Cloud Infrastructure overview

This section provides an overview of the infrastructure and architecture of a Collibra Cloud.

Cloud regions and availability zones

Collibra has a cloud presence in many places around the world. Cloud centers are located so that each customer's data can be kept in the appropriate regulatory region. The on-boarding process includes selection of an appropriate regulatory region based on each unique customer's requirements. Under no circumstances does a customer's data ever leave the regulatory region agreed upon. Data may be moved around for resilience or performance reasons within a regulatory region.

Customer separation

Collibra uses a variety of technical means to keep each customer's data separate from every other customer's data. Separation is ensured during storage, internal transit, and processing at all times.

Data privacy and security

Data privacy and security are an essential aspect of Collibra Data Intelligence Cloud.

Collibra is constantly improving its security procedures in order to guarantee the best possible security standards.

Cloud infrastructure

Collibra works closely with various Infrastructure-as-a-Service cloud providers to provide a flexible and secure environment. For Collibra's cloud offering, these services include:

- Data center
- Server hardware
- Network infrastructure
- Cloud provider specific services

Physical access to the servers is subject to the privacy statements of cloud providers.

Encryption

All data is encrypted in transit between the Jobserver (installed at client) and the Collibra Data Intelligence Cloud. It uses mutual authentication via certificates over TLS 1.2. When data is stored in Collibra, it is always encrypted at rest using AES 256.

Key management

Keys are managed by Collibra using the key management systems native to the cloud provider. The KMS used is FIPS 140-2 compliant 140-2, uses AES 256-bit encryption and the keys are rotated at standard intervals.

Collibra employees

Collibra is aware that, when it comes down to security, people are often the weakest link.

- Every Collibra employee signs a non-disclosure agreement (NDA) contract concerning all company and customer data.
- Before new employees are hired, background checks are done, to the best of our abilities.
- All desktops and laptops used by Collibra employees are encrypted using a unique (randomly generated) key.

- We have several roles at Collibra, and only qualified people will have access to the customer virtual machine.
- An employee termination process is in place in order to remove all access to Collibra internal services and data.

Password management

All customer data is protected by highly secure (random) passwords. Different passwords are used for:

- The virtual machine running the application.
- The database
- The administrator access to the application
- The backup encryption

Passwords are kept in a highly secure digital vault, AES-256 algorithms with encryption keys of at least 80 bit. This vault allows us to make specific passwords available to specific people.

Database, virtual machine and backup passwords are only accessible by senior management and infrastructure support employees.

Application administrator passwords are only accessible by senior management and application support employees.

Additionally, temporary access to the necessary passwords is provided to development or pre-sales employees to solve specific problems in case of an emergency.

Employee training

Collibra employee trainings are organized internally during the last quarter of every year. This training covers the following topics:

- Password vault usage
- Data Privacy and Data Security awareness and best practices
- Securing your desktop/laptop
- Cloud infrastructure

New employees are given this training within the first three months of their employment.

Chapter 4

Access and authentication

This section provides more information on access and authentication to Collibra Data Intelligence Cloud environments by customers and the Collibra Cloud Infrastructure team.

Customer access to Collibra cloud

You can only access Collibra Data Intelligence Cloud by going to https://<customer-name>.collibra.com.

> Note It's impossible to access the actual servers that contain the data. This also means that you cannot install other services and applications.

On request, it is possible to limit inbound access to a specific IP-range (IP Whitelisting). To enforce a limit, you have to create a ticket on the Collibra support portal.

Whitelisting IP addresses

By default, all source IP addresses on the Internet can reach your environment. Whitelisting is a cybersecurity enhancement that prevents unexpected sources from reaching your Collibra environment.

When whitelisting is enabled for your environments, all source IP addresses are blocked, except for the source IP addresses that you provided in your support ticket.

To avoid interruptions and access issues to your environments, ensure that you provide a comprehensive and validated list of your internal egress IP addresses. These egress IP addresses could include your corporate offices, VPN, integrations and partners; generally any client that should be able to access your environment.

We strongly recommend that you work with your internal departments to identify who needs access, and then apply the whitelisting on your non-production environments. This allows you to validate the list of source IP addresses while mitigating access interruptions to mission-critical or production environments.

IP address format

In your support ticket, you must provide a comma-separated list of all IP addresses that you want whitelisted, in Classless Inter-Domain Routing (CIDR) notation. The IP addresses should be routable.

Addresses in private spaces, such as 10.0.0.0 or 192.168.0.0, are not valid and will be ignored.

Collibra Cloud Infrastructure team access

The Collibra Cloud Infrastructure team can sign in to the cloud servers to perform necessary management and maintenance tasks. Collibra Security does not allow direct management access to the actual servers containing the data, and must be done through a dedicated management server and then through a bastion host.

Collibra has an Identity and Access Management (IAM) system in place to provide separation of rights. Only dedicated and trained personnel have access to the entire system. Each user has its own unique user ID to track responsibility of changes to the system. Access is handled through private/public key authentication secured by a FIPS-140 second factor.

Chapter 5

Monitoring a Collibra cloud environment

Collibra uses multiple monitoring services to provide a near-real-time visibility into the internal cloud infrastructure.

Monitoring alerts

Every customer cloud environment is monitored on a minute-by-minute basis from multiple external locations around the world. This monitoring checks the availability of the service and sends alerts to the Collibra Cloud Infrastructure team in case of an outage. An incident management process is then initiated to resolve the problem as soon as possible.

The Cloud Infrastructure team also uses a continuous monitoring solution internally to manage performance and capacity for every customer.

Uptime monitoring

A system run by Collibra fully monitors and reports on uptime. This system is able to make a difference between maintenance mode and other downtime to measure the uptime with higher accuracy.

Detailed reporting on uptime is available on a monthly basis by request to your Customer Success Manager or Collibra support.

Performance monitoring

Performance of the cloud resources is monitored using a specialized service.

This includes:

- Memory usage
- CPU utilization
- Disk usage
- Network usage
- (Average) Throughput (requests per minute)
- (Average) Response times

When performance problems occur for any of these resources, alerts are sent to the Collibra Cloud Infrastructure team so that prompt action can be taken.

Security monitoring

In order to track changes to our cloud servers, the following actions are logged:

- SSH authentications
- User commands
- Changing user privileges
- Sign-in attempts
- Other indicators specific to our internal platform

Host Intrusion Detection System

A Host Intrusion Detection System (HIDS) monitors network traffic for suspicious activity and alerts the system or network administrator. In some cases, the HIDS may also respond to anomalous or malicious traffic by taking action, such as blocking the user or source IP address from accessing the network.

The tools, as implemented by Collibra, detects the following intrusion possibilities:

- File integrity (system and application files)
- Sign-in attempts
- Portscanning
- Brute force attacks
- Rootkit detection

All alerts are collected in a central place, and alerts with high priority are reviewed by both Production Engineering and Security Operations.

Host Intrusion Prevention System

A Host Intrusion Prevention System (IPS) is a network security/threat prevention technology that examines network traffic flows to detect and prevent vulnerability exploits.

Collibra has an active response system in place that does not only send out alerts (HIDS), but also takes action (HIPS). The HIPS can take the following actions when needed:

- Deny IP addresses to access the system (firewall, hosts.deny).
- Disable the user accounts under attack.

- Drop packets.

Antivirus and malware

Collibra is always checking for the latest virus and malware (including rootkits) vulnerabilities. Our automated process will download the newest threats and search the system for possible virus and malware presence.

If a threat is detected, the customer is notified.

Bring Your Own Key (BYOK)

For security reasons, we encrypt the virtual hard disks that contains your Collibra Data Intelligence Cloud repository, with an encryption key. By default, this key is provided by Collibra but you can also use your own encryption key. By using your own key, you have full control over your data and can revoke access to that data at any time.

Collibra supports AWS Key Management Service (KMS) and GCP Key Manager.

Note that encryption is done at the virtual hard disk level, not the database level, and that we encrypt only the data drives that contain the Collibra repository.

Prerequisites

- You can use Bring Your Own Key (BYOK) to encrypt your virtual hard disks (VHD) only in Amazon Web Services (AWS) and Google Cloud Platform (GCP).
- You can use only your own encryption key using the open-source software AWS CloudHSM (Hardware Security Modules) or Google Cloud Key Management.
- You must provide us with the Amazon Resource Name (ARN) or Google Resource Name of your encryption key via your Customer Success Manager.

Good to know

- We encrypt only the virtual hard disks that contain the Collibra Data Intelligence Cloud repository with your encryption key. We do not encrypt theVHDs that contain databases of Metadata Lake, Usage Analytics, Collibra Console or Jobserver, provided that these are on different VHDs.
- Encrypting the VHDs with your own key does not have any impact on future Collibra upgrades.

Key management

By using your own encryption key to encrypt the virtual hard disks (VHD) that contain the Collibra Data Intelligence Cloud repository, you have full control over the access to the VHD.

Disabling keys

If you suspect that Collibra has been compromised, you can disable the key so that no one can decrypt or even access the VHDs anymore.

We scan AWS or GCP to check if you have disabled your key. If we detect that a key is disabled, our security systems shut down all services in the affected Collibra environment. This ensures that the data is no longer accessible. By re-enabling the key, we can restart your environment.

To disable or enable keys in AWS Key Management Service, go to the AWS documentation. To disable or enable keys in Google Cloud Key Management, go to the GCP documentation.

We do not store your encryption key on our servers. It is your responsibility to manage and protect it. We can only access the VHDs if you give us access to this encryption key. If you delete or lose the key, your data is permanently lost.

Rotating keys with AWS Key Management Service

AWS KMS automatically rotates AWS managed keys every year. You cannot enable or disable key rotation for AWS managed keys.

If a key rotates via the automatic key rotation mechanism, it does not affect your environment, meaning that your environment remains operational. For more information about automatic key rotation, go to the AWS documentation. Keep in mind that an automatic key rotation does not re-encrypt your VHDs.

If you want to re-encrypt VHDs, we have to migrate your data. For this purpose, you have to create a second encryption key. With this second key, we create new encrypted VHDs and migrate your data to these new VHDs. This data migration causes downtime of your Collibra environment. As such, you must provide a maintenance window via your Customer Success Manager.

> Important During the re-encryption process, the previous encryption key must remain enabled until the migration is completed.

Note that backups will not be re-encrypted, therefore, we recommend that you disable the previous encryption key rather than deleting it.

Rotating keys with GCP Key Management

With GCP key management, you can configure automatic key rotations, which is a recommended security practice. If a key rotates via the automatic key rotation mechanism, it does not affect your environment, meaning that your environment remains operational. For more information about automatic key rotation, go to the GCP documentation. Keep in mind that an automatic key rotation does not re-encrypt your VHDs.

If you want to re-encrypt VHDs in GCP, we have to decrypt your VHDs with the current encryption key and re-encrypt them with the new key. For this purpose, you have to create a second encryption key, so that we can complete this action. Keep in mind that the re-encryption causes downtime of your Collibra environment. As such, you must provide a maintenance window via your Customer Success Manager.

> Important During the re-encryption process, the previous encryption key must remain enabled until the re-encryption is completed.

Note that backups will not be re-encrypted, therefore, we recommend that you disable the previous encryption key rather than deleting it.

Create an encryption key and enable BYOK

This section explains the steps to create the encryption key in AWS or GCP, to encrypt the virtual hard disks that will contain the Collibra Data Intelligence Cloud repository.

AWS

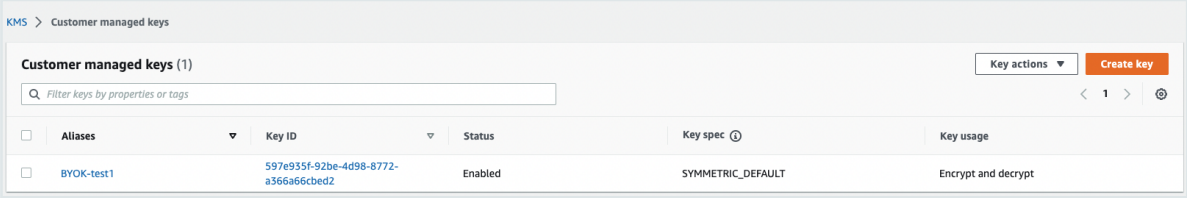
1. Sign in to your AWS Console and go to the Key Management Service.

   > Note You must be in the same region as your Collibra Data Intelligence Cloud environment.

2. Generate an encryption key with the following specifications:

Encryption key specification	Value
Key type	Symmetric
Key usage	Encrypt and decrypt

   For more details on creating encryption keys, go to the AWS documentation.
   These keys are protected by FIPS 140-2 validated cryptographic modules (hardware security modules).

3. Contact your Customer Success Manager (CSM) to enable BYOK on your Collibra environment and provide the following information:
   - The list of environments on which you want to enable BYOK.
   - The Amazon Resource Names (ARN) address of the created encryption key. You can find the ARN in the **General Configuration** section of your key.
4. Your CSM then provides you with our Collibra AWS account ID.
5. In the KMS key settings, add the Collibra AWS account ID.
6. Provide a maintenance window in which we can enable BYOK on the selected environments.

During this maintenance window, we enable BYOK by using the ARN of your encryption key.

GCP

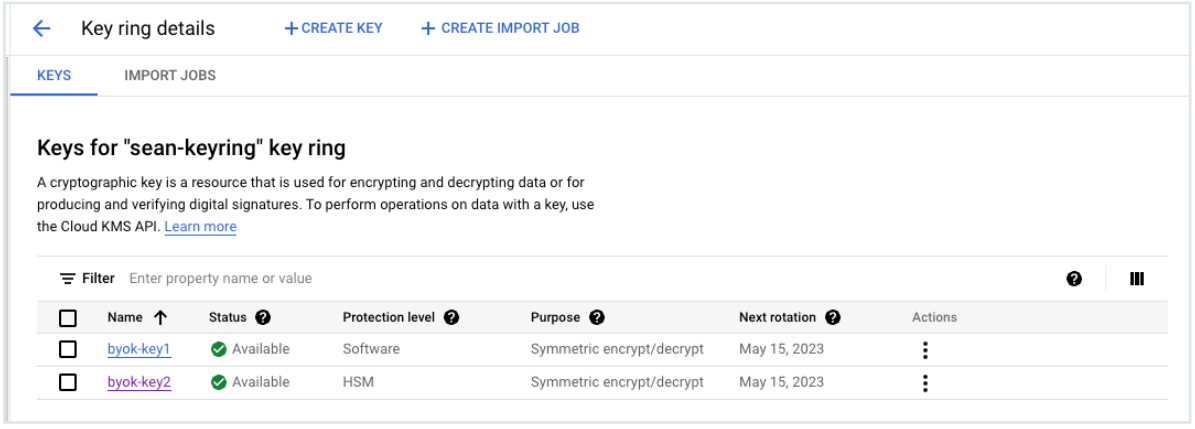
1. Sign in to your Google Cloud console and go to the Key Management section.
2. Create a new key ring

> Note You must be in the same region as your Collibra Data Intelligence Cloud environment. If you choose to create a multi-region key, make sure that Collibra's region is included.

3. In this key ring, create a new key with the following specifications:

Encryption key specification	Value
Purpose	Symmetric encrypt/decrypt
Duration of 'scheduled for destruction' state	We recommend at least 90 days. We recommend that you choose to disable the key rather than deleting it. Deleting the key is permanent and cannot be undone. For more information about the destruction state of keys, go to the GCP key destroy and restore documentation.

For more details on creating encryption key rings, go to the GCP documentation.



4. Contact your Customer Success Manager (CSM) to enable BYOK on your Collibra environment and provide the following information:

- The list of environments on which you want to enable BYOK.
- The ID of the symmetric encryption key. To find the ID of the key, go to the GCP Key Management documentation.
  The ID must be in the following format:

```
projects/${my-gcp-project}/locations/${my-location}/keyRings/${my-key-ring-name}/cryptoKeys/${my-key-name}
```

5. Your CSM will deliver you two GCP principals and their required permissions.
6. In the permission settings of the encryption key, add the principals with the provided permissions.
7. Provide a maintenance window in which we can enable BYOK on the selected environments.
   During this maintenance window, Collibra enables BYOK by using the ID of your encryption key.

Backup and recovery

As a SaaS Data Intelligence company, Collibra is responsible for backing up your data and then restoring that data, when required, in response to various types of incidents. The following describes our backup and recovery practices and timelines for recovery from different incident types.

Repository backups

We perform backups of databases and file systems according to the practices below. With these, we can restore customer data for the last 30 days with an accuracy window of 15 minutes.

Category	Backup practice
Database	We support point in time recovery with an accuracy window of 15 minutes for the past 30 days using a combination of full and incremental backups. This system is fully separated from the Collibra Console backup and restore feature.
File systems	We take a nightly snapshot of all disk volumes, except for the root partition. The snapshot is taken according to the time zone of the server.

For disaster recovery purposes, we ensure that backups are replicated across multiple availability zones located in the same cloud region. These backups are encrypted at rest using AES-256 encryption.

We can restore your system or data as required throughout the subscription term. At the end of the subscription term, the final backup is retained and available for 30 days after subscription termination.

> Note Currently, we don't support multi-region backups.

Recovery

The procedure to recover an encrypted backup depends on the event that triggers the need for recovery.

Incident type	Recovery practice
Data loss	We restore data to the last known good state, within 15 minutes of accuracy.
Database corruption	We analyze the database dump to find the time stamp of corruption. After this investigation, we restore the database to the last known point in time prior to the corruption.
Server problems	We analyze server problems. If the problem is not found in a reasonable time, then we schedule a full restore, but only with your approval. The application and the data can be restored within up to 8 business hours.
Availability zone crash and data loss	We restore an off-site backup in a different zone.

> Note We only offer hot/cold recovery across multiple availability zones within the region.

About RPO and RTO

RPO or Recovery Point Objective, is the time from the last data backup until an incident occurred that may have caused data loss.

RTO or Recovery Time Objective, is the time that you set to recover the lost data.

Given the backup and recovery practices, we support the following RPO and RTO:

RPO	30 days with a granularity of 15 minutes.
RTO	Up to 8 hours under normal circumstances, depending on the incident type and the volume of the data to be restored.

Chapter 5

Adding Edge or a Jobserver to Collibra Data Intelligence Cloud

When you want to ingest data, you need Edge or the Jobserver service in your environment, whether it's an on-premises Collibra environment or a cloud environment. From a network point of view, we recommend in both environment types to install Edge or the Jobserver service as close to the data source as possible for optimal ingestion performance. In all cases, we recommend to use Edge, as the Jobserver's end-of-life is announced for September 2024.

Whether you use Edge or Jobserver, ensure that they always meet the system requirements. To add Edge to your environment, go to the Edge configuration guide.

This section focuses on installing and adding the Jobserver service to your Collibra Data Intelligence Cloud environment. Note that Jobserver is nearing its end of life and that it is highly recommended to use Edge.

> 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 Jobserver and Collibra Console on the same server.
>
> Additionally, you can install the Monitoring service to monitor the Jobserver service.
>
> You can find the version of your Collibra Data Intelligence Cloud environment at the bottom of the sign-in window, for example 2024.02.0. Always use a Jobserver version that is compatible with your environment.

Supported operating systems

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

Linux operating systems

- Red Hat Enterprise Linux/CentOS 7.x
- Red Hat Enterprise Linux/Rocky Linux 8
- Red Hat Enterprise Linux/Rocky Linux 9

> 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 2016
- Windows Server 2019
- Windows Server 2022

Jobserver requirements

The following requirements are for an on-premises Jobserver that will be used with a Collibra Data Intelligence Cloud environment.

On Linux, you have installed PostgreSQL 14.9 before installing the Jobserver. On Windows, PostgreSQL 14.9 is included in the Jobserver-only installer.

Data ingestion

- 64 GB RAM
- 500 GB free disk space
- Hard disk type: SSD
- Number of CPUs: 16

Tableau ingestion

- 6 GB RAM
- 35-50 GB free disk space
- Hard disk type: SSD
- Number of CPUs: 4

Collibra Console requirements

The following requirements are for an on-premises Collibra Console that will be used to manage an on-premises Jobserver.

On Linux, you have installed PostgreSQL 14.9 before installing Collibra Console. On Windows, PostgreSQL 14.9 is included in the Jobserver-only installer.

- 1 GB RAM
- 1 GB free disk space for the installation.
- 1 GB free disk space for the Collibra Console database.

As this Collibra Console is only used to manage the on-premises Jobserver, you don't need extra disk space for backups.

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

Before you install your on-premises Jobserver, we highly recommend to do some basic connectivity tests between the node on which you are going to install the Jobserver service

and your Collibra Data Intelligence Cloud environment. If the node cannot reach your Collibra Data Intelligence Cloud environment, then first fix the connectivity before you start the Jobserver installation.

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

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

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

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

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

What's next?

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

Install Jobserver and Collibra Console on Linux

This section describes how to install the Jobserver and Collibra Console on Linux. There is no need to install these components on different servers.

> Important If you do the installation on RHEL/Rocky Linux or Suse, see the services section of the installation guide. This is also valid if you install without root permissions.

Prerequisites

- You have downloaded the latest Jobserver-only installer for your operating system. See also the compatibility list to know which installer you have to download.
- You have installed PostgreSQL 14.9. If you install Jobserver and Collibra Console on separate servers, then both servers require PostgreSQL 14.9.
- We recommend to use a static IP address for the node.
- You have a connection between the on-premises Jobserver and Collibra Data Intelligence Cloud.

Steps

1. Run the installer:
   - Linux as user with sudo rights: `sudo ./dgc-linux-jobserver-only-2024.02.0-x.sh`
     Linux as root user: `./dgc-linux-jobserver-only-2024.02.0-x.sh`
   - Linux as standard user: `./dgc-linux-jobserver-only-2024.02.0-x.sh`
2. Follow the command-line wizard. If you don't enter a value, the value between brackets or the value in capital is used.

   > Note The Jobserver encryption key in the wizard is a passphrase that is used to generate the actual encryption key.

```
Verifying archive integrity...  100%   All good.
Uncompressing DGC Installer  100%
Specify the installation directory [/opt/collibra]:

Please specify the data directory [/opt/collibra_data]:

Do you want to install the Collibra Jobserver component?
[Y/n]
```

```
Do you want to install the Collibra Management Console com-
ponent? [Y/n]

Are you sure these are the components you want to install?
[Management Console, Jobserver] [Y/n]

Specify the postgresql 14 path [/usr/pgsql-14]:

Specify the Jobserver port [4404]:

Specify the Jobserver database port [4414]:

Specify the Jobserver monitoring port [4424]:

Specify the Jobserver Spark monitoring port [4434]:

Specify the Management Console port [4402]:

Specify the Management Console database port [4420]:

Specify the Agent port [4401]:

Specify the Agent address [localhost]:
Note: with a loopback address (localhost, 127.0.0.1, et
al.) you will not be able to use a multi node setup

2022-09-26 15:23:19.123 - SUCCESS - Create user and group
2022-09-26 15:23:19.133 - SUCCESS - Check umask settings

...

2022-09-26 15:23:40.167 - SUCCESS - Start Console
2022-09-26 15:23:44.411 - SUCCESS - Start Agent
2022-09-26 15:23:44.411 - COMPLETED - Installation finished
in 25739ms.
```

What's next?

> Tip If you don't install these services on the same node, then you have to add the Jobserver node to your on-premises Collibra Console.

Configure the connection between the Jobserver and your Collibra Data Intelligence Cloud environment by creating a connection from the Jobserver to the DGC service. In this configuration, the on-premises Jobserver will poll the DGC service for ingestion tasks.

> Note You can also add the Jobserver service to the DGC service. In this configuration, the DGC service from your cloud environment will send ingestion tasks to the on-premises Jobserver. For security reasons, this configuration may not be allowed by your organization's network infrastructure department.

Install Jobserver and Collibra Console on Windows

This section describes how to install the Jobserver and Collibra Console on Windows. There is no need to install these components on different servers.

Prerequisites

- You have downloaded the latest Jobserver-only installer for your operating system. See also the compatibility list to know which installer you have to download.
- We recommend to use a static IP address for the node.
- You have a connection between the on-premises Jobserver and Collibra Data Intelligence Cloud.

Steps

> 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: Windows Server: double-click **setup.bat**The path of the installer file cannot contain spaces.
   If you run the installation without Administrator rights, an error is shown.
2. In the wizard introduction, click **Next**.
3. Enter the **Installation directory** and click **Next**.
4. Enter the **Data directory** and click **Next**.
5. Select **Management Console** and **Jobserver** and click **Next**.

6. Enter the Jobserver settings and click **Next**.

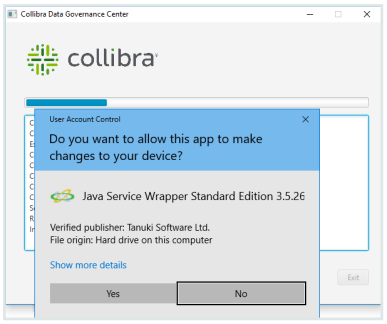
Setting	Description
Jobserver port	The TCP port to access the Jobserver service. The default port is *4404*.
Jobserver database port	The TCP port to access the Jobserver database. The default port is *4414*.
Jobserver monitoring port	The TCP port that is used by the monitoring service to monitor the Jobserver service. 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. The default port is *4434*.

7. Enter the Console settings.

Setting	Description
Console port	The TCP port to access your Collibra Console via your web browser. 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. The default port is *4420*.

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

8. Click **Install**.
   » The installation of the components starts.
9. 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 fails.

10. Click **Exit**.
    » Collibra is installed on your system.

What's next?

> Tip If you don't install these services on the same node, then you have to add the Jobserver node to your on-premises Collibra Console.

Configure the connection between the Jobserver and your Collibra Data Intelligence Cloud environment by creating a connection from the Jobserver to the DGC service. In this configuration, the on-premises Jobserver will poll the DGC service for ingestion tasks.

> Note You can also add the Jobserver service to the DGC service. In this configuration, the DGC service from your cloud environment will send ingestion tasks to the on-premises Jobserver. For security reasons, this configuration may not be allowed by your organization's network infrastructure department

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

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

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

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

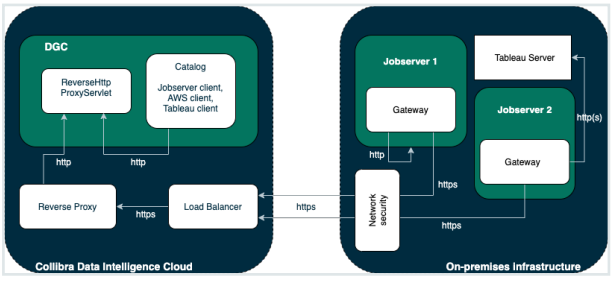
On-premises Jobserver to Collibra Data Intelligence Cloud communication

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

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

Communication overview

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



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

server. The configuration of the communication from the Jobserver to the Tableau server is similar to the one from a Jobserver to a Collibra Data Intelligence Cloud environment.

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

Components

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

New component	Description
Reverse HTTP proxy servlet	The reverse HTTP proxy is part of the DGC service. It acts as a server for all other Collibra services, whether they are installed on-premises or together with the DGC service in the cloud.
Gateway	The gateway is part of the Jobserver service. It polls the DGC service's reverse proxy to fetch tasks and send them to the Jobserver.

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

Configure the Jobserver to Collibra Data Intelligence Cloud communication

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

In this section, we describe how to configure the Jobserver to poll Collibra for jobs. You will have to configure both the Data Governance Center service and the Jobserver service.

Prerequisites

- You have Collibra Data Intelligence Cloud2020.10 or newer.
- You have created a keystore in the PKCS#12 format on the node that hosts the Jobserver service.

Steps

Configure the Jobserver service

> Warning
> Only apply changes to these settings if you are really experienced with JVM parameters. Changing parameters may cause serious performance issues.
>
> To remove an individual JVM property you must use the delete icon (🗑) alongside the property, otherwise, the service will interpret it as a blank line and fail to start correctly.
>
> Restart the service after editing the JVM parameters.

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

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

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

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

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

Setting	Description
proxy.url	The URL of the Collibra environment, followed by *reversehttp-poll/<gateway-id>*. This "gateway-id" must be identical to the one used in the Name parameter when you add the Jobserver to the DGC service. The value of this setting is case-sensitive. Example `-Dproxy.url=https://<your-environment-url>/reversehttp-poll/Jobserver-1`
target.url	The URL of your the target system, either an on-premises Jobserver or a Tableau server. Example ◦ Jobserver: `-Dtarget.url=http://localhost:4404` ◦ Tableau server: `-Dtarget.url=https://tableau-sales.yourcompany.com`
http.proxy.host (optional)	The hostname of the HTTP proxy server for outbound connections to your Collibra Data Intelligence Cloud environment. This option is used to enable outbound traffic monitoring. Example `-Dhttp.proxy.host=proxy.yourcompany.com`
http.proxy.port (optional)	The port of the HTTP proxy server for outbound connections to your Collibra Data Intelligence Cloud environment. This option is used to enable outbound traffic monitoring. Example `-Dhttp.proxy.port=8080`

Setting	Description
username	The username of any Collibra user for basic authentication. Example `-Dusername=john.fisher`
password	The corresponding password of the Collibra user for basic authentication. Example `-Dpassword=ChangeMe` You can encrypt this password if necessary. Example `-Dpassword=enc_2:t2rklBY6699aWV0...`
keystore.path	The full path to the PKCS12 keystore. This keystore should contain the private key to sign the basic authentication header. Example `-Dkeystore.path=/opt/collibra_data/spark-jobserver/security/jobserver-1-keystore.p12`

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

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

Setting	Description
idle.connection.timeout	The time in milliseconds that an idle connection is kept in the connection pool. We recommend to not define this parameter, it then uses the default value of 5,000 milliseconds. Example `-Didle.connection.timeout=3000`
connection.timeout	The time in milliseconds that the reverse HTTP server waits for a response from your Collibra environment or from the value in target.url. If you don't set this parameter, the value is 60,000 milliseconds. Example `-Dconnection.timeout=30000`
connection.soTimeout	The time in milliseconds that the reverse HTTP server waits for a response from your Collibra environment or from the value in target.url on socket level. If you don't set this parameter, the value is 60,000 milliseconds. Example `-Dconnection.soTimeout=30000`
polling.timeout	The time in milliseconds that the reverse HTTP server waits for a poll request from your Collibra environment to be submitted to the target.url. If you don't set this parameter, the value is 300,000 milliseconds. Example `-Dpolling.timeout=100000`

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

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

7. Click the green **Save all** button.
8. Click **Security configuration**.

9. Click **Edit configuration**.
10. Set the **Authentication level** to *NONE*.

> Note This means that there is a one-way outbound communication over TLS from the Jobserver to the Collibra environment, note that there is no authentication at all.

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

Add the Jobserver to the DGC service

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

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

Setting	Description
Name	The name of the Jobserver as it will appear when you register a data source. The name is a freely chosen name but it is recommended to only use alphanumerical characters and dashes, for example Jobserver-1. You will have to use this name as the ID of the gateway and in the address of this configuration.
Protocol	The protocol for this configuration has to be *HTTP* and not the recommended HTTPS, this is because of the Collibra internal architecture.
Address	The loopback address of the DGC service, followed by */reversehttp/<gateway-id>*. The "gateway-id" must be identical to the one used in the Name parameter of this configuration. Do not use the scheme in the address. Example localhost:4400/reversehttp/Jobserver-1

Setting	Description
Trusted server CA certificate	The certificate in PEM format that contains the public key of the Jobserver to validate the signature of the basic authentication header. In the example to create a keystore, this is the content of the file **cert.pem**. Example -----BEGIN CERTIFICATE----- MIICqDCCAZACCQCcy3Oq51c5YzANBgkqhkiG9w0BAQsF ADAWMRQwEgYDVQQDDAtq b2JzZXJ2ZXIt... -----END CERTIFICATE-----
Client certificate	This field is not used in this configuration.
Client private key	This field is not used in this configuration. Note This field always shows dots, even if it is empty.
Table profiling data size	The approximate maximum disk size of the data in MB that will be used to profile a table. The value cannot exceed 10 000.

If you use a truststore, go to Generate keys, certificates and keystores.

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

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

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

What's next?

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

Encrypt passwords for basic authentication

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

Prerequisites

- You have downloaded the reversehttp-gateway-standalone-7.0.0.jar file.
- You have the password of the Collibra user that you use to connect to your Collibra Data Intelligence Cloud environment.

Steps

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

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

```
java -jar reversehttp-gateway-standalone-7.0.0.jar encrypt

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

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

What's next?

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

Monitor outbound traffic

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

1. Open Collibra Console with a user profile that has the **SUPER** role.
   » Collibra Console opens with the **Infrastructure** page.
2. Click the Jobserver service of a Collibra environment.
3. Click **Infrastructure Configuration**.
4. Click **JVM configuration**.
5. Click **Edit configuration**.
6. Add the following JVM settings:

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

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

Generate keys, certificates and keystores

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

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

Steps

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

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

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

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

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

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

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

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

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

Add a Jobserver to the DGC service

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

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

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

Steps

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

3. Enter the necessary information:

Setting	Description
Jobserver list	The list of registered Jobserver instances.
Name	The name of the Jobserver as it will appear when you register a data source in Data Catalog. The name is a freely chosen name but it is recommended to only use alphanumerical characters and dashes, for example Jobserver-1. You will have to use this name as the ID of the gateway and in the address of this configuration.
Protocol	The protocol that is used for the communication between the Data Governance Center service and the Jobserver service. It is recommended to use HTTPS, especially if the services are hosted in different network segments.
Address	The address (IP address, URL, hostname) of the Jobserver.
Trusted server CA certificate	The certificate of the trusted CA needed to validate the server certificate. If blank, the default truststore will be used. The default truststore is defined in the SSL configuration section of the DGC service. The CA certificate of the server party (Jobserver).
Client certificate	The client certificate offered by the DGC service to the server. If blank, you cannot select mutual authentication as the Jobserver service authentication level.
Client private key	The private key of the DGC service's certificate.
Table profiling data size	The approximate maximum disk size of the data in MB that will be used to profile a table. The value cannot exceed 10,000.
Test connection timeout	This timeout is a time limit (in seconds) after which the connection test is stopped and a timeout error is shown. The default value is 60 seconds.

4. Click **Save all**.

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

Upgrade the Jobserver and Collibra Console on Linux

When your Collibra Data Intelligence Cloud is upgraded, you have to update all your on-premises Jobservers and Collibra Console instances to the latest available on-premises version. The installers are released on a quarterly basis, check the compatibility list to know which installer you have to download.

This section describes how you can upgrade your on-premises Jobservers and Collibra Console on Linux.

> Tip If you already installed the on-premises Jobserver and Collibra Console with the latest available installer, there's no need to upgrade these.

Prerequisites

- You have downloaded the latest Jobserver-only installer from the Collibra Community Downloads page.
- You have installed PostgreSQL 14.9. If you install Jobserver and Collibra Console on separate servers, then both servers require PostgreSQL 14.9

> Note
> - 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 in the Troubleshooting section.

Steps

1. Stop the environment:
   a. Open Collibra Console with a user profile that has at least the **ADMIN** role.
      » Collibra Console opens with the **Infrastructure** page.
   b. Click the environment that you want to stop.
   c. Click ■ **Stop**.
      » The **Stop environment** dialog box appears.
   d. Click **Stop environment**.
   b. Wait until all the nodes of the environment show **Stopped**.
   c. Stop Collibra Console and the Collibra DGC Agent. In the terminal of the node that runs Collibra Console:
      a. As root:

```
service collibra-agent stop
service collibra-console stop
```

      b. Manual stop of the services:

```
/opt/collibra/console/bin/console stop
/opt/collibra/agent/bin/agent stop
```

2. Install PostgreSQL 14.9:

   a. Install PostgreSQL 14.9 with the following commands.

> Important Run the commands as root.

```
#Clean the YUM cache and update existing packages for
your current Linux repository. Note that this makes
system changes.

yum clean all && yum update -y

#Prepare the PostgreSQL repository and packages:
```

```
yum -y install
https://download.postgresql.org/pub/repos/yum/reporpms
/EL-$(rpm -E %{rhel})-x86_64/pgdg-redhat-repo-
latest.noarch.rpm

#Update the packages in the repository:
yum -y update

#Install the PostgreSQL 14.9 packages:
yum -y install postgresql14 postgresql14-server
postgresql14-contrib
```

b. Update the file **/usr/lib/tmpfiles.d/postgresql-14.conf** to set the correct permissions for some PostgreSQL folders. Open the file for editing, for example with vim or nano and update the line `d /run/postgresql 0755 postgres postgres -` to:

```
d /run/postgresql 2777 postgres postgres - -
```

> Important Do not use the `chmod` command on any directories or files, edit this configuration file instead.

c. If you are upgrading from version 5.8.x to 5.9, you also have to update the configuration file of PostgreSQL 11 (**/usr/lib/tmpfiles.d/postgresql-11.conf**) as described in the previous step.

d. Reboot the server.

> Tip The default PostgreSQL 14 path on RHEL/Rocky/CentOS is **/usr/pgsql-14**.

a. Install PostgreSQL 14.9 with the following commands:

> Important Run the commands as root.

```
sh -c 'echo "deb
http://apt.postgresql.org/pub/repos/apt $(lsb_release -
cs)-pgdg main" > /etc/apt/sources.list.d/pgdg.list'

wget --quiet -O -
https://www.postgresql.org/media/keys/ACCC4CF8.asc |
sudo apt-key add -

apt-get update && apt-get upgrade -y

apt-get install -y postgresql-14
```

b. Update the file **/usr/lib/tmpfiles.d/postgresql-common.conf** to set the correct permissions for some PostgreSQL folders. Open the file for editing, for example with vim or nano and update the line `d /run/postgresql 1775 postgres postgres` - to:

```
d /run/postgresql 2777 postgres postgres - -
```

c. Reboot the server.

> Tip The default PostgreSQL 14 path on Debian/Ubuntu is **/usr/lib/postgresql/14**.

a. Install PostgreSQL 14.9 with the following commands:

> Important Run the commands as root. In the first of the following commands, the packages are for Suse 15, replace "15" by your used major version, for example 12. You can check your Suse version with the command: `cat /etc/os-release`.

```
zypper addrepo
https://download.postgresql.org/pub/repos/zypp/repo/pg
dg-sles-15-pg14.repo
```

```
zypper refresh

zypper install postgresql14 postgresql14-server
postgresql14-contrib
```

b. Update the file **/usr/lib/tmpfiles.d/postgresql-14.conf** to set the correct permissions for some PostgreSQL folders. Open the file for editing, for example with vim or nano and update the line `d /run/postgresql 0755 postgres postgres -` to:

```
d /run/postgresql 2777 postgres postgres - -
```

c. If you are upgrading from version 5.8.x to 5.9, you also have to update the configuration file of PostgreSQL 11 (**/usr/lib/tmpfiles.d/postgresql-11.conf**) as described in the previous step.
d. Reboot the server.

> Tip The default PostgreSQL 14 path on Suse is **/usr/lib/pgsql-14**.

If you don't update the PostgreSQL configuration file, you get an error that a test file could not be written.

3. Run the installer:
   - Linux as user with sudo rights: `sudo ./dgc-linux-jobserver-only-2024.02.0-x.sh`
     Linux as root user: `./dgc-linux-jobserver-only-2024.02.0-x.sh`
   - Linux as standard user: `./dgc-linux-jobserver-only-2024.02.0-x.sh`
4. Follow the command-line wizard. If you don't enter a value, the value between brackets or the value in capital is used.

> Note
> - The path to your PostgreSQL 14.9 installation differs per Linux operating system. In the following example, the path is of a default installation on CentOS 7.
> - The amount of time it takes to upgrade your environment depends on the size of your repository. The larger the database, the more time it takes to upgrade.

```
Verifying archive integrity...  100%   All good.
Uncompressing DGC Installer  100%
Specify the installation directory [/opt/collibra]:

/opt/collibra contains a previous installation. Do you want
to perform an update? [y/N]
y
Before you can schedule an upgrade, you need to:
- Create a backup of the entire environment.
- In Collibra Console, stop all running services on this
node.
Have you completed these steps? [yes,NO]

yes
Specify the postgresql 14 path [/usr/pgsql-14]:

2022-09-26 15:15:00.542 - SUCCESS - Check umask settings
2022-09-26 15:15:00.552 - SUCCESS - Create installation and
data directories

...
2022-09-26 15:15:34.063 - SUCCESS - Start Console
2022-09-26 15:15:34.064 - COMPLETED - Installation finished
in 33881ms.
```

5. Start the Jobserver.

Upgrade the Jobserver and Collibra Console on Windows

When your Collibra Data Intelligence Cloud is upgraded, you have to update all your on-premises Jobservers and Collibra Console instances to the latest available on-premises version. The installers are released on a quarterly basis, check the compatibility list to know which installer you have to download.

This section describes how you can upgrade your on-premises Jobservers and Collibra Console on Windows.

> Tip If you already installed the on-premises Jobserver and Collibra Console with the latest available installer, there's no need to upgrade these.

Prerequisites

- You have downloaded the latest Jobserver-only installer from the Collibra Community Downloads page.

> Note
> - 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 in the Troubleshooting section.

Steps

1. Stop the environment.
2. Stop the Collibra Agent and Collibra Console.
3. Run the installer: Windows Server: double-click **setup.bat**The path of the installer file cannot contain spaces.
   If you run the installation without Administrator rights, an error is shown.
4. Click **Next**.
5. Select the installation directory of the old version and click **Update**.
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 upgraded.
8. Click **Exit**.
9. Start Collibra Console.

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

Email configuration for Collibra Data Intelligence Cloud and Collibra Console

Collibra Data Intelligence Cloud and Collibra Console use Mailgun to reliably and securely send emails to the customer inboxes. For more information about the email security, see the Mailgun security pages.

> Important This is only applicable for commercial cloud customers.
>
> - GovCloud customers use AWS SES as mail provider.
> - On-premises customers have to set up their own email provider.

You can find the IP addresses from which you can expect emails via its SPF record. You find the necessary information online or with the following command.

```
dig @8.8.8.8 +short -t TXT alerts.collibra.com
```

By default, the From-address is no-reply@alerts.collibra.com. On request, we can update the From-address to no-reply+<environment name>@alerts.collibra.com. We do not support any other changes to the From-address.

> Example We can change the From-address to no-reply+acme-dev@collibra.com.

To request this change, contact Collibra support.

Upgrading Collibra cloud environments

As a cloud customer, your Collibra customer success representative will contact you to plan the upgrade of your Collibra Data Intelligence Cloud environment to the latest version. Alternatively, you can request an upgrade yourself through a support ticket. In the ticket, specify the version to which you want to upgrade and your preferred date and time. We will then schedule the upgrade at the earliest.

Every upgrade causes a planned downtime. To minimize the impact, we plan your upgrade outside your regular business hours (09:00-17:00, in your time zone) whenever possible.

> Note
> - If we upgrade your Collibra Data Intelligence Cloud environment and you have an on-premises Jobserver installation, you need to upgrade your on-premises installation if possible. If your Collibra environment does not have a corresponding on-premises installer, ensure that your on-premises Jobserver and Collibra Console are installed with the latest available installer. To know which installer you need to use, go to the compatibility list.
> - After your Collibra environment is upgraded, the search index is automatically rebuilt. The rebuild process can take some time depending on the number of assets. During this period, the search function is unavailable.

Environment upgrades

Every environment upgrade can contain new features and bug fixes. The Collibra Data Intelligence Cloud data structure can change during an upgrade, or that existing content

will be migrated as well. This is an irreversible action, unless you restore a backup from your environment prior to the upgrade.

The proposed date and time of an upgrade is communicated to you through email, at least two weeks prior to the planned upgrade. These upgrades are done over the weekend to avoid any impact on the users.

Within these two weeks, you can request to delay this upgrade by replying to the email. You can propose a later date and time, up to two months after the initially proposed upgrade date.

> Note If you don't reply to our upgrade proposal, we assume that you approve, and we will upgrade your Collibra Data Intelligence Cloud environment at the proposed date and time.

The downtime for an upgrade depends on the size and complexity of your Collibra Data Intelligence Cloud content.

After the upgrade to 2020.11, you have to upgrade the activity history. This is a one-time only action, subsequent upgrades don't require this extra step.

> Important After an upgrade of your environment, verify if you have to upgrade your Jobserver. Check the release notes to know which Jobserver version you need. See the installation section for the upgrade steps.

PostgreSQL 14.9 FAQ for cloud upgrades to 2023.04 or newer

The FAQ in this section mainly focuses on the upgrade of PostgreSQL to version 14.9 in your cloud environments. You can find the FAQ of the PostgreSQL upgrade in your on-premises environments on this page.

Why are we upgrading PostgreSQL?

PostgreSQL releases new major versions every year and provides updates for each major version for five years, after which support for fixes ends. Our cloud solution Collibra Data

Intelligence Cloud currently uses PostgreSQL version 11.17, for which support will end on November 9, 2023.

What is the impact on my cloud environment after November 9, 2023?

None. The PostgreSQL version in your 2023.03 or older cloud environments will still work beyond November 9, 2023 but after this date, there will be no further updates for PostgreSQL, not even for critical security issues.

What is the timeline for the upgrade to PostgreSQL 14.9?

The PostgreSQL 14.9 upgrade will be generally available on April 23, 2023 as part of the 2023.04 release and deployed to all customers who upgrade.

What are the implications of waiting beyond PostgreSQL 11 end-of-life (EOL)?

In the event of a functional bug or security vulnerability, the PostgreSQL community is not going to backport a fix/patch to version 11.

Which Collibra Data Intelligence Cloud environments are impacted?

Collibra Data Intelligence Cloud 2023.03 is the last version that uses PostgreSQL 11. Collibra 2023.04 or newer makes use of PostgreSQL 14.9. We do not support PostgreSQL 14.9 on Collibra 2023.03 or older.

Is it mandatory to upgrade to Collibra Data Intelligence Cloud 2023.04?

No. You can create a support ticket to opt out from the 2023.04 release as in the regular release schedule.

Which on-premises Jobserver do I use?

To avoid any future PostgreSQL upgrade scenarios, we strongly recommend that you migrate from Jobserver to Edge.

If you decide to keep using an on-premises Jobserver, the compatibility between your cloud environment and that Jobserver remains as is. An on-premises Jobserver is always compatible until the next quarterly cloud release. This means that for 2023.04, you can still use the on-premises Jobserver 2023.02. However, we will release a 2023.04 Jobserver-only installer. This version is identical to the 2023.02 version but uses PostgreSQL 14.9 instead of PostgreSQL 11.

For the on-premises Jobserver compatibility, go to the Jobserver compatibility page.

What is the feature compatibility between cloud and on-premises environments?

The on-premises Collibra Data Governance Center 5.7.13 is feature compatible with the cloud Collibra Data Intelligence Cloud 2022.05. The on-premises version 5.9.0 is identical to 5.8.2 but makes use of PostgreSQL 14.9 and includes extra bug fixes. Version 5.8.2 on its turn is identical to 5.7.13 but makes use of PostgreSQL 11 and includes extra bug fixes.

For the feature compatibility of older on-premises versions, go to the compatibility page.

Can I restore backups from on-premises environments in my cloud environment?

Yes, but keep the version compatibility in mind.

- You can restore a 5.7.13 backup only on 2022.05 cloud environments or newer.
- You can restore a 5.8.x backup only on 2022.12 (on-demand version) cloud environments or newer.
- You can restore a 5.9.0 backup only on 2023.04 cloud environments or newer.

For the backup compatibility of older versions, go to the compatibility page.

Following this upgrade, will customers using Jobserver be responsible for subsequent PostgreSQL upgrades?

Future PostgreSQL upgrades will follow the same process. Collibra drives the cadence to upgrade PostgreSQL as it is tied to our releases.

Chapter 8

Troubleshooting

DGC service fails to start due to invalid configuration

If Collibra detects an invalid configuration of the Data Governance Center service, it no longer automatically replaces the invalid configuration by the default configuration and the service does not start.

This situation can happen when you edit a configuration in a backup, introducing invalid data, and then you restore that backup.

In **dgc.log**, you can find a Collibra exception "configurationParsingFailed" and "DGCConfigurationServiceImpl.readDGCConfiguration (DGCConfigurationServiceImpl.java...)":

```
Caused by: com.collibra.common.exception.CollibraException: con-
figurationParsingFailed
Message: com.fasterxml.jackson.databind.JsonMappingException:
...


    at com.collibra.dgc.configuration.service.DGCCon-
figurationServiceImpl.readDGCConfiguration(DGCCon-
figurationServiceImpl.java:362)
```

Resolution

Revert the configuration changes in your backup.

Limitations

Running Collibra Data Intelligence Cloud has a couple of limitations:

- Integrations: Connecting to your company's LDAP server is only possible if the LDAP server is accessible from the outside world (from our servers).
- Sending emails is always done through the Collibra SMTP server. For this we use mailgun.
  When receiving emails from your cloud environment, you see the following:

  > Example Received: from cloud-mail.collibra.com (cloud-mail.collibra.com. [198.2.180.134])

  This can be used to filter on your mail server (either by DNS name or IP).
- Workflow action emails requires Collibra to connect to an IMAP or POP server.
  When requested, Collibra can setup a private email address for you.
- Collibra does not support a custom domain or SSL certificate to be used for the URL of the cloud environment.
  To solve this, a custom proxy server can be used on the customer side.
- Collibra automatically applies product updates when necessary.
- Because the cloud environment are publicly accessible, guest access is not possible in the cloud environment.
- On a Collibra Data Intelligence Cloud environment, you cannot use the HTTP method PATCH.
  Use the HTTP method POST with X-HTTP-Method-Override header instead.
- Microsoft Azure Application Gateway has a limitation of 4 GB on uploads. Therefore, we can only upload backups with a maximum size of 4 GB. We are looking for solutions to work around this limitation.