Collibra

Collibra Data Intelligence Cloud

Edge Infrastructure

Version 2022.03

Collibra

Collibra Data Intelligence Cloud - Edge Infrastructure

Release date: March 6, 2022

Revision date: Thu Mar 03, 2022

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

https://productresources.collibra.com/docs/collibra/latest/Content/Edge/to_edge.htm

Contents

Introducing Edge

Edge is a cluster of Linux servers for accessing and processing data close to where it resides. It helps to connect to data sources and process information within your data landscape.

Edge enables Collibra Data Intelligence Cloud to safely connect to your data sources hosted in an on-premise or cloud environment. It processes the data source information on the Edge site and sends the process results to Collibra Data Intelligence Cloud.

Edge components

Edge consists of three main components:

- An Edge configuration page in Collibra Data Intelligence Cloud to create and install Edge sites.
- An Edge integration capability repository that resides on the Collibra Platform and contains all capabilities that can run on an Edge site.
- An Edge site that is installed close to a data source in the customer's environment, whether it's in the cloud or on the customer's premises.

Integration steps

The following table shows which steps you have to take to set up Edge.

Step	Description	Required permissions
1	Create an Edge site via Collibra Data Intelligence Cloud Settings.	You have a global role with the Manage Edge sites global permission in Collibra Data Intelligence Cloud.

Step	Description	Required permissions
2	Install the Edge site close to the data source you want to access. You can only install an Edge site on a Linux system that meets the necessary system requirements.	You have a global role with the Install Edge sites global permission in Collibra Data Intelligence Cloud.
3	Update the credentials of the Edge site user.	You have a global role with the Connect Edge sites to Collibra global permission in Collibra Data Intelligence Cloud.

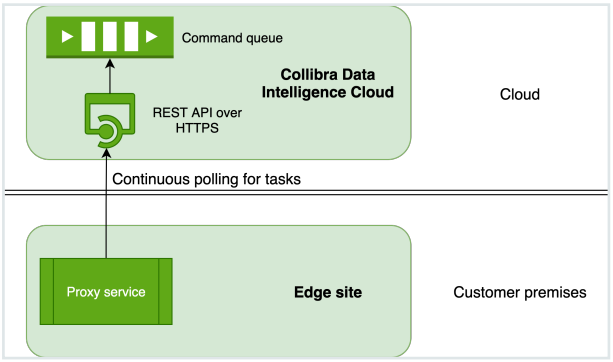
Chapter 3

Edge security

Edge is built with security first approach. All communication channels are secured by TLS and all endpoints outside Edge are accessible only via authentication. Edge does not send or store any customer data, its purpose is to host capabilities that process the data in its own environment and to send only processing results to Collibra Data Intelligence Cloud.

Communication between Edge and Collibra

Edge does not need incoming access - it executes tasks as commands stored in the Edge management queue in Collibra. Communication to Collibra uses basic authentication over TLS 1.2. User account for communicating to Collibra is generated on each Edge site installer download, and it is unique for each Edge site. It is possible to change the password of this user account through a documented procedure.



- Edge sites always use REST API endpoints to establish connections.
- Edge does not store data after the data is processed on the Edge site, not even sample data.
- Edge manages Collibra Data Intelligence Cloud and data source credentials. This has the following consequences:
  - Credentials are not accessible outside of Edge.
  - Credentials used on an Edge site are encrypted with a key that is secured in Collibra.
  - Credentials of data sources and Collibra can be updated if necessary.
- All configuration parameters, files or strings marked as secret are stored on the Edge site encrypted with a public key that resides in Collibra. The private part of that key is encrypted with a public key from the Edge site. As a result, secrets can only be decrypted with both key pairs, one residing on the Edge site and the other on Collibra.
- An Edge site communicates in a secure way with your Collibra environment using certificates, issued by a Collibra-chosen Certificate Authority (CA). However, if there is a forward proxy server between the Edge site and Collibra, you have to use the proxy server's CA.

Authentication to data sources

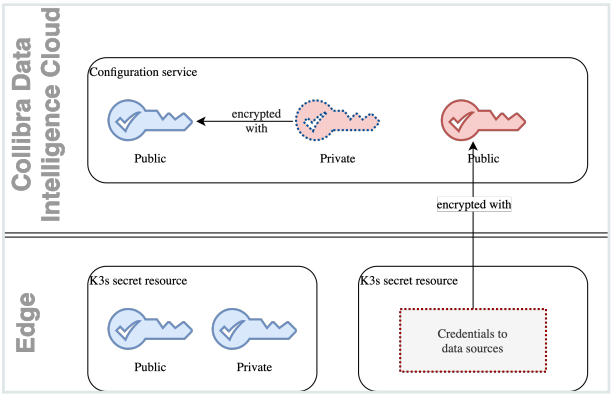
Edge connections and capabilities use different ways to connect to data sources The required level of privileges or security greatly depends on the data source type and supported Catalog Connectors.

Collibra regularly adds and certifies Catalog connectors. To understand the authentication methods and the level of security, consult the Catalog connector documentation.

Storing secrets

Secrets for connections and capabilities are stored solely on the Edge site. While at rest, secrets are using envelope encryption where the secret is encrypted by a key, which on its turn is encrypted by another key.

An Edge site owns the Blue key pair, with the Blue private key stored on Edge. Similar to that, Collibra Data Intelligence Cloud owns the Red key. Every secret on Edge is encrypted with the private Red key, which is sent to the Edge site for each capability execution, encrypted with the Blue public key. Once on the Edge site, Red private key is decrypted, and secrets needed to execute a connection or a capability are decrypted and injected into the capability container.



Data samples in Edge

Edge by design doesn't send any samples. Edge capabilities such as Profiling and Classification use data in memory to execute, after which data is discarded.

It is not possible to view data samples in Collibra Data Intelligence Cloud. We are currently developing a new capability for this. You will have the choice to install it after the Edge deployment.

Edge service repository

To keep Edge synchronized with your Collibra Data Intelligence Cloud version, we deploy core Collibra services and business capabilities in the Collibra repository of your environment. An Edge site uses token-based authentication with read privileges to download services for each release. The authentication and endpoint to access the Collibra repository are stored in the **registries.yaml** file as part of the Edge site installer.

Collibra uses XRay scanning for image scanning on every build, as well as Contrast scanning for dependency in runtime scanning. Every critical vulnerability automatically fails the build.

For 2-day vulnerability, you can edit **registries.yaml** and access the registry independently, and download images for Edge to scan them. Currently there is no SLA for vulnerabilities that you may find. The standard support SLAs are applied.

Monitoring and logging

We monitor and log all interaction between an Edge site and Collibra Data Intelligence Cloud, as well as the Edge site infrastructure health. All logs are kept in the Collibra Datadog account.

> **Note** We don't send Catalog connector logs to your environment. These Catalog connector logs are by default turned off. If they are enabled, they are kept on the Edge site itself. If you have Catalog connector issues, you have to extract these logs and send them to Collibra Support via a support ticket.

Installing an Edge site

An Edge site is a component installed in a customer's environment. Each Edge site has a unique identifier and hosts an Edge capability that can access a data source.

This section contains the information that you need to know to install an Edge site.

About an Edge site installation

After creating the Edge sites in Collibra Data Intelligence Cloud, you have to install the Edge software on either K3S or EKS. You typically install Edge sites within the same secure environment as the relevant data source. A customer usually has several Edge sites, depending on his requirements. The required amount of Edge sites is can be based for example on the number of networks and secure environments but it can also on the technical and legal spread of data sources.

An Edge site can have:

- Zero or more predefined connections to data sources via a JDBC driver.
- One or more integration capabilities to process data on site and send the results to Collibra.

An Edge site is a compute runtime on K3S or EKS, that executes capabilities close to your data but that is configurable from the Collibra Data Intelligence Cloud settings. It has a dedicated unique identifier and handles data sources that it can reach within its network. You can have more than one Edge site, depending on the number of networks, security domains, regions or VPCs that you have.

Properties

Property	Description
Name	The name of the Edge site. Use a meaningful name, for example NetherlandsDataCentre1. Do not use spaces or special characters. This field is mandatory and the name must be globally unique.
Status	The status of the Edge site. The status is automatically shown when you create an Edge site.

Property	Description
ID	The unique ID of the Edge site, which is created automatically when you created the Edge site.
Description	The description of the Edge site. We recommend to put at least basic location information of the Edge site. This field is mandatory.
Installer and property files	A section where you can download the installer and property files to install an Edge site on a server. This section is only visible when the Edge site has status **To be installed**.

Statuses

The status of an Edge site indicates if the Edge site can be used or not. The status is shown on the **Edge** settings page of the Collibra settings. An Edge site can have one of the following statuses:

Status	Description
To be installed	The Edge site is created, but not installed yet.
Offline	Collibra cannot reach the Edge site. This can be caused by an unsuccessful installation or a lost connection. See the installation logs for more information.
Unhealthy	Collibra can connect to the Edge site, but some functions don't work correctly. This is typically caused by problems during the installation. See the installation logs for more information.
Healthy	The Edge site installation was successful.

Installation directories

The Edge site installer installs files in the following directories on your host server:

- /var/lib/rancher/
- /var/log/
- /etc/
- /usr/local/bin/

System requirements of an Edge site

If you want to use Edge, you must make sure that the following system requirements are met.

Software requirements

- You can install the Edge software on CentOS/RedHat Enterprise Linux 8.x
- The **sudo** package is installed on the Linux host.
- The user who installs Edge has full sudo access (`ALL=(ALL) ALL`)

> **Tip** If you are an early adopter or you use Edge for beta testing purposes, we highly recommend to disable SELinux.

Hardware requirements

You need the following minimum hardware requirements:

- 64 GB memory
- 16 core CPU with x86_64 architecture
- You have mounted at least 50 GB of dedicated storage for the core installation of Edge sites on the mountpoint **/var/lib/rancher/k3s**.

  > **Warning** Any data in this location is fully managed by the Edge site, do not save any other data in here as it can be removed by Edge without notification.

- You have mounted at least 500 GB of dedicated storage for the Edge site data on a freely chosen mountpoint, for example **/var/edge/storage**.

```
mkdir -p /var/edge/storage
mkfs.xfs /dev/<block-device-name>
mount /dev/<block-device-name> /var/edge/storage
echo '/dev/<block-device-name> /var/edge/storage xfs
defaults 0 0' >> /etc/fstab
```

> **Note** Change `<block-device-name>` to the name of the device that contains the storage.

> **Warning** This dedicated storage should not be shared with other services because Edge can delete and overwrite files on this location without notice, so don't use **/home/<username>** or **/var**.

> **Warning** When new capabilities are added in the future, the hardware requirements may change.

Network requirements

- An Edge site needs outbound connections to all of the following:
  - The URL of your Collibra Data Intelligence Cloud environment
  - https://ingest.eu0.signalfx.com
  - https://*.datadoghq.com

    > **Note** If wildcard whitelisting is not accepted:
    > - https://http-intake.logs.datadoghq.com
    > - https://7-30-1-app.agent.datadoghq.com
    > - https://agent-http-intake.logs.datadoghq.com
    > - https://api.datadoghq.com

  - https://*.repository.collibra.io

    > **Note** If wildcard whitelisting is not accepted:
    > - https://repository.collibra.io
    > - https://edge-docker-delivery.repository.collibra.io

  - https://otlp-http.observability.collibra.dev/
- Your Edge site has to be able to connect to port 443.
- The resolve configuration file of your Linux host has maximum three search domains and two name servers.

Note If a firewall is enabled, run the following commands to add the cni0 and loopback interfaces to a trusted zone, so that Kubernetes can use it between its services:

```
firewall-cmd --zone=trusted --change-interface=cni0 --
permanent
firewall-cmd --zone=trusted --change-interface=lo --
permanent
firewall-cmd --reload
```

EKS requirements

You can install the Edge software on managed Kubernetes clusters:

- AWS EKS 1.21 (only with `--container-runtime containerd`)
- AWS EKS worker nodes use the EKS optimized Amazon Linux 2 AMI
- EKS cluster has IRSA enabled
- AWS EKS worker nodes need to be in the same (one) Availability Zone!
  This can be implemented by creating just one node group for the EKS cluster, which limits the subnets to just one subnet, that is one of the subnets of the EKS cluster.

```
module "eks" {
                                source             = "terraform-aws-modul
                                version            = "17.24.0"

                                cluster_name       = "${var.vpc_name}-${v
name}-eks"
                                cluster_version    = "1.21"
                                vpc_id             = var.vpc_id
                                subnets            = data.aws_subnet_ids.
subnet_ids.ids              # Subnets specified must be in
at least two different AZs
                                worker_additional_security_group_ids = [
security_group.worker_sg.id]
                                enable_irsa        = true
                        # enable iam role for service
account, for later use

                                worker_groups = [
```

```
                                {
                                name                            = "${var.v
${var.cluster_name}-eks-workers"
                                instance_type                   = var.work
                                asg_desired_capacity            = var.inst
workers
                                key_name                        = aws_key_
pair.cluster-ssh-keypair.key_name
                                bootstrap_extra_args            = "--conta
runtime containerd"      # mandatory to run with containerd
if on 1.21
                                subnets                         = [subnet1
                    # restriction for now to use only 1
subnet due to EBS tied to AZ
                                },
                                ]

                                map_accounts = [
                                data.aws_caller_identity.current.account
                                ]

                                tags = {
                                Name            = "${var.vpc_name}-${var
name}-eks"
                                }
                            }
```

Software requirements

- A Linux server with bash available. This is the server from which you install the Edge software on EKS.

  > **Tip** This server will also contain the Edge tools.

- Plain cluster_admin kubectl access to the EKS cluster using its kubeconfig. With this kubeconfig, you must be able to use the kubectl command to communicate with the Kubernetes API server with full cluster access.
- Kubectl client version 1.21.6, supports EKS 1.20 and 1.21.

Hardware requirements

You need an operational EKS cluster with at least 1 worker node. The cluster must meet the following requirements:

- Cluster capacity of at least 16 core CPU and 64 GB memory, for example 1 m5.4xlarge node or 4 m5.xlarge nodes..
- Each worker node needs at least 100 GB free disk space to store Docker images.
- Ability to create EBS-based persistent volumes as a default storage class, at least 500 GB in total.

Network requirements

- An Edge site needs outbound connections to all of the following:
  - The URL of your Collibra Data Intelligence Cloud environment
  - https://ingest.eu0.signalfx.com
  - https://*.datadoghq.com

    > **Note** If wildcard whitelisting is not accepted:
    > - https://http-intake.logs.datadoghq.com
    > - https://7-30-1-app.agent.datadoghq.com
    > - https://agent-http-intake.logs.datadoghq.com
    > - https://api.datadoghq.com

  - https://*.repository.collibra.io

    > **Note** If wildcard whitelisting is not accepted:
    > - https://repository.collibra.io
    > - https://edge-docker-delivery.repository.collibra.io

  - https://otlp-http.observability.collibra.dev/
- Your Edge site has to be able to connect to port 443.
- The resolve configuration file of your Linux host has maximum three search domains and two name servers.

Note If a firewall is enabled, run the following commands to add the cni0 and loopback interfaces to a trusted zone, so that Kubernetes can use it between its services:

```
firewall-cmd --zone=trusted --change-interface=cni0 --
permanent
firewall-cmd --zone=trusted --change-interface=lo --
permanent
firewall-cmd --reload
```

Create an Edge site

You create an Edge site to have a processing runtime at your premises.

Prerequisites

- You have a global role that has the System administration global permission.
- You have a global role that has the Manage Edge sites global permission.
- You have enabled database registration via Edge in Collibra Console.

> **Note** You must restart the Data Governance Center service when you have enabled this option.

Steps

1. In the main menu, click ⋮⋮⋮, then ⚙ **Settings**.
   » The Collibra settings page opens.
2. Click **Edge**
   » The Edge sites overview appears.
3. Above the table, to the right, click **Create Edge site**.
   » The **Create Edge site** wizard starts.
4. Enter the required information.

Field	Description
Name	The name of the Edge site. Use a meaningful name, for example NetherlandsDataCentre1. Do not use spaces or special characters. This field is mandatory and the name must be globally unique.
Description	The description of the Edge site. We recommend to put at least basic location information of the Edge site. This field is mandatory.

5. Click **Create**.
   » The Edge sites overview appears, including the new Edge site with the status **To be installed**.

What's next?

You can now install the Edge site.

Configure a forward proxy

For security reasons, it is possible that an Edge site has to connect via a forward HTTP proxy. In that case, you have to update **proxy.properties** before installing the Edge site.

If the forward proxy server is responsible to decrypt TLS traffic, you also have to use the proxy server's CA certificate during the installation.

Steps

1. Download the Edge site installer:
    a. Open an Edge site.
        a. In the main menu, click , then **Settings**.
            » The Collibra settings page opens.
        b. Click **Edge**
            » The Edge sites overview appears.
        c. In the Edge site overview, click the name of an Edge site.
            » The Edge site page appears.
    b. In the **Installer and properties files** section, click **Download**.
    c. Depending on your operating system and browser, follow the regular steps for downloading files.
    » The installer file is a TGZ archive that contains the files **proxy.properties**, **properties.yaml** and **registries.yaml**.

    > **Note** If you download an installer, all previously downloaded installers become invalid.

2. Open the **proxy.properties** file.
3. Uncomment and update the outbound-proxy properties by removing "#" at the beginning of the following lines:

```
#noProxy=<host IP addresses>,<host DNS names>,<k8s-svc-ip-addresses>,<k8s-pod-ip-addresses,<others>
#proxyHost=<proxy domain name or IP address>
#proxyPort=<proxy-port>
```

```
#proxyUsername=<proxy username>
#proxyPassword=<proxy password>
```

Setting	Value
noProxy	A comma-separated list of IP or DNS addresses that can bypass the proxy server. This list must include at least the Kubernetes cluster's internal IP addresses and the Kubernetes nodes' IP and DNS addresses. The list may not contain spaces. **Tip** To get the values for this setting, you can use the **edge-get-noproxy.sh** script. However, make sure that your network administrator reviews these values. where: ◦ <host-ip-addresses>: for example `172.20.0.0/16`. ◦ <host-dns-names>: for example `*.compute.internal`. ◦ <k8s-svc-ip-addresses>: is by default `10.43.0.0/16`, but this can differ for other k8s flavors or configurations. ◦ <k8s-pod-ip-addresses>: is by default `10.42.0.0/16`, but this can differ for other k8s flavors or configurations. ◦ <others>: other IP addresses that don't need to be proxied. Add at least `169.254.169.254.` for AWS. **Example** `noProxy=172.20.0.0/16,*.compute.internal,10.43.0.0/16,10.42.0.0/16,169.254.169.254`

Setting	Value
proxyHost	The IP or DNS address of the proxy server. **Example** `proxyHost=site4-proxy.shared.edge.collibra.dev`
proxyPort	The TCP port of the proxy server. This value must be a quoted string and not an integer value. **Example** `proxyPort="3128"`
proxyUsername	The username to authenticate at the proxy server. **Example** `proxyUsername=edge`
proxyPassword	The password to authenticate at the proxy server. **Example** `proxyPassword=la;fs90jpo4j3rR%`

```
#noProxy=<host IP addresses>,<host DNS names>,<k8s-svc-ip-addresses>,<k8s-pod-ip-addresses,<others>
#proxyHost=<proxy domain name or IP address>
#proxyPort=<proxy-port>
#proxyUsername=<proxy username>
#proxyPassword=<proxy password>
```

Setting	Value
noProxy	A comma-separated list of IP or DNS addresses that can bypass the proxy server. This list must include at least the Kubernetes cluster's internal IP addresses and the Kubernetes nodes' IP and DNS addresses. The list may not contain spaces. **Tip** To get the values for this setting, you can use the **edge-get-noproxy.sh** script. However, make sure that your network administrator reviews these values. where: ◦ <host-ip-addresses>: for example `172.20.0.0/16`. ◦ <host-dns-names>: for example `*.compute.internal`. ◦ <k8s-svc-ip-addresses>: depends on your EKS installation. Typically this is `10.100.0.0/16` or `172.20.0.0/16`. ◦ <k8s-pod-ip-addresses>: depends on your EKS installation. Typically they are the same subnets as in the VPC, for example `172.20.0.0/16`. ◦ <others>: other IP addresses that don't need to be proxied, for EKS, always add `169.254.169.254`.. **Example** `noProxy=172.20.0.0/16,*.compute.internal,10.100.0.0/16,169.254.169.254`
proxyHost	The IP or DNS address of the proxy server. **Example** `proxyHost=site4-proxy.shared.edge.collibra.dev`

Setting	Value
proxyPort	The TCP port of the proxy server. This value must be a quoted string and not an integer value. **Example** `proxyPort="3128"`
proxyUsername	The username to authenticate at the proxy server. **Example** `proxyUsername=edge`
proxyPassword	The password to authenticate at the proxy server. **Example** `proxyPassword=la;fs90jpo4j3rR%`

Important When you add a new node to a cluster, review and update, if necessary, the noProxy and implicitly forward proxy settings, unless the subnet used for nodes and their DNS suffix are added to noProxy.

4. If you use a forward proxy that decrypts TLS traffic, a so-called man-in-the-middle proxy or MITM proxy, then on-the-fly TLS certificates that are generated by the MITM proxy, must use the subjectAltName (SAN) extension. To enable Edge via a MITM proxy, perform also the following steps.
   If the proxy server does not decrypt the TLS traffic, you can skip the following steps.
   a. Export your proxy server's CA certificate in PEM format.
   b. Save this certificate as **ca.pem** in the same directory as the Edge site installer.

      Note If you save the certificate in another directory, use the `--ca` argument in the Edge site installation command.

What's next?

Install the Edge site

If you want to update the forward proxy afterwards, you can use the update script.

Install an Edge site

After you have created the Edge site in Collibra Data Intelligence Cloud, you have to install the Edge software on a server.

> **Tip**
> Every time you download an Edge site installer, the previously downloaded Edge site installer becomes outdated. If you use this outdated installer, the Edge site cannot communicate with Collibra.

Prerequisites

- You have a global role with the Install Edge sites and the User Administration global permission, for example Edge site administrator
- You have a global role that has the System administration global permission.
- You have created an Edge site.
- You have configured the forward proxy, if a forward proxy is required for Edge to connect to Collibra, Datadog, SignalFX and jFrog. Contact your network administrator if this is applicable.
- Your server meets all system requirements.

> **Tip** If you are an early adopter or you use Edge for beta testing purposes, we highly recommend to disable SELinux.

Steps

1. Download the installer:
    a. Open an Edge site.
        a. In the main menu, click , then **Settings**.
            » The Collibra settings page opens.
        b. Click **Edge**
            » The Edge sites overview appears.
        c. In the Edge site overview, click the name of an Edge site.
            » The Edge site page appears.

b. In the **Installer and properties files** section, click **Download**.

> **Tip** When you download the installer, an Edge user is automatically created in Collibra.

c. Depending on your operating system and browser, follow the regular steps for downloading files.
» The installer file is a TGZ archive that contains the files **proxy.properties**, **properties.yaml** and **registries.yaml**.

> **Warning** If you download an installer, the previously downloaded Edge site installer becomes invalid.

2. Extract the TGZ archive on the server on which you are going to install the Edge site software.

```
tar -xf <edge-site-id>-installer.tgz
```

> **Tip** Keep the installer or the content of the extracted installer somewhere safe on your server. It contains various tools that you may need later, for example to troubleshoot issues.

3. Execute the installation. Use the correct path to the mounted storage as described in the prerequisites.

> **Important** If the Edge site has to connect via a forward HTTP proxy, then first configure the forward proxy before executing the installation.

```
sudo sh install-master.sh --storage-path
/path/mounted/storage properties.yaml -r registries.yaml

for example:

sudo sh install-master.sh --storage-path /var/edge/storage
```

```
properties.yaml -r registries.yaml
```

» In the Edge sites overview, you can see the status of the deployment.

Tip If you want to enable classification on the Edge site, add the following extra argument to the command:

```
--set collibra_
edge.collibra.classification.enabled=true
```

4. Run the following commands to verify the status of the installation.
   - To ensure that Kubernetes is running and that there is an existing node:

     ```
     sudo /usr/local/bin/kubectl get nodes
     ```

   - To ensure the state of all pods are installed and running:

     ```
     sudo /usr/local/bin/kubectl get pods --all-namespaces
     ```

Tip If you already installed Edge site and you want to enable classification afterwards, execute the following command:

```
sudo /usr/local/bin/kubectl  -n collibra-edge \
 exec -it deploy/argo-cd-argocd-application-controller  \
 --  bash -c 'argocd admin cluster kubeconfig
https://kubernetes.default.svc  /tmp/config --namespace
collibra-edge ; env KUBECONFIG=/tmp/config  argocd app
set collibra-edge   --core  -p
collibra.classification.enabled=true'
```

1. Download the installer:
    a. Open an Edge site.
        a. In the main menu, click ⠿, then ⚙ **Settings**.
            » The Collibra settings page opens.
        b. Click **Edge**
            » The Edge sites overview appears.
        c. In the Edge site overview, click the name of an Edge site.
            » The Edge site page appears.
    b. In the **Installer and properties files** section, click **Download**.

> **Tip** When you download the installer, an Edge user is automatically created in Collibra.

    c. Depending on your operating system and browser, follow the regular steps for downloading files.
    » The installer file is a TGZ archive that contains the files **proxy.properties**, **properties.yaml** and **registries.yaml**.

> **Warning** If you download an installer, the previously downloaded Edge site installer becomes invalid.

2. Extract the TGZ archive on the server on which you are going to install the Edge site.

```
tar -xf <edge-site-id>-installer.tgz
```

> **Tip** Keep the installer or the content of the extracted installer somewhere safe on your server. It contains various tools that you may need later, for example to troubleshoot issues.

3. Execute the installation.

```
./run-installer-job.sh properties.yaml --repositories
repositories.json --set collibra_
edge.collibra.minio.persistence.size=120Gi
```

» In the Edge sites overview, you can see the status of the installation.

Tip If you want to enable classification on the Edge site, add the following extra argument to the command:

```
--set collibra_
edge.collibra.classification.enabled=true
```

4. Run the following commands to verify the status of the installation.
   - To ensure that Kubernetes is running and that there is an existing node:

     ```
     kubectl get nodes
     ```

   - To ensure the state of the installation is either running or finished:

     ```
     kubectl get pods --all-namespaces
     ```

Tip If you already installed Edge site and you want to enable classification afterwards, execute the following command:

```
kubectl  -n collibra-edge exec -it deploy/argo-cd-argocd-
application-controller --  bash -c 'argocd admin cluster
kubeconfig https://kubernetes.default.svc  /tmp/config --
namespace collibra-edge ; env KUBECONFIG=/tmp/config
argocd app  set collibra-edge   --core  -p
collibra.classification.enabled=true'
```

JDBC connections

JDBC connections, or simply connections, define how an Edge capability accesses a data source.

To create a connection to your data source, you need to select a connection provider, which determines the available properties of the connection, such as the authentication method and connection string and driver.

> **Example** If you want to ingest data from an Amazon Redshift data source, you need a specific JDBC driver for Amazon Redshift. You use that driver to create a connection between your Edge site and your Amazon Redshift data source.

> **Tip** Collibra provides a selection of certified JDBC drivers on Collibra Marketplace. We highly recommend to only use JDBC drivers that are certified for Edge.

Data sources supported by Edge

You can register, profile and classify several data sources via Edge. Depending on your data source, you can use a Collibra-provided Catalog connector, or your own JDBC driver when you create a JDBC connection.

The following data sources have been tested for registering, profiling and classifying via Edge.

Create a JDBC connection

You can create a JDBC connection from an Edge site to a data source. You can then register the data source via Edge.

> **Warning** Currently, Data Catalog does not support registering Oracle and data sources without schemas via Edge.

Available Catalog connectors

Edit a JDBC connection

You can edit a JDBC connection, for example if you want to change one of its connection properties. You can then register the data source via Edge.

> **Warning** Currently, Data Catalog does not support registering Oracle and data sources without schemas via Edge.

Available Catalog connectors

Delete a JDBC connection

You can delete an JDBC connection from an Edge site to a data source if you no longer need it.

Prerequisites

- You have a global role that has the System administration global permission.
- You have a global role that has the Manage connections and capabilities global permission.
- You have created and installed an Edge site.
- You have created a JDBC connection.

Steps

1. Open an Edge site.
   a. In the main menu, click , then **Settings**.
      » The Collibra settings page opens.
   b. Click **Edge**
      » The Edge sites overview appears.
   c. In the Edge site overview, click the name of an Edge site.
      » The Edge site page appears.
2. In the JDBC Connections section, click the name of a JDBC connection.
   » The **Connection settings** page appears.
3. At the bottom of the page, click **Delete**.
   » The **Delete confirmation** dialog box appears.
4. Click **Delete Connection**.

Chapter 6

Edge capabilities

An Edge capability is an application that runs on an Edge site to extract and process data. It delivers the results to Collibra Data Intelligence Cloud.

About Edge capabilities

An Edge capability is an application that runs on an Edge site. The Edge capability can access a data source to extract and process data. It delivers the output to Collibra Data Intelligence Cloud in a secure and reliable way.

An Edge capability has a capability template that defines a specific use case, for example data source ingestion.

Capability templates

A capability template is developed for a specific task on a specific data source type. The capability template also determines which properties are available to configure the Edge capability.

Currently, the following capability templates are available:

- Catalog JDBC ingestion: A capability template you use to register a data source and synchronize schemas from a data source via a JDBC connection.
- JDBC Profiling: A capability template you use to profile and classify data from a registered data source.
- DQ Connector: A capability template you use to ingest Data Quality user-defined rules, metrics, and dimensions into Collibra Data Catalog.
- S3 synchronization: A capability template you use to connect to Amazon S3.

> **Important** While these capability templates are available for all customers, the features that you use them for might still be in beta.

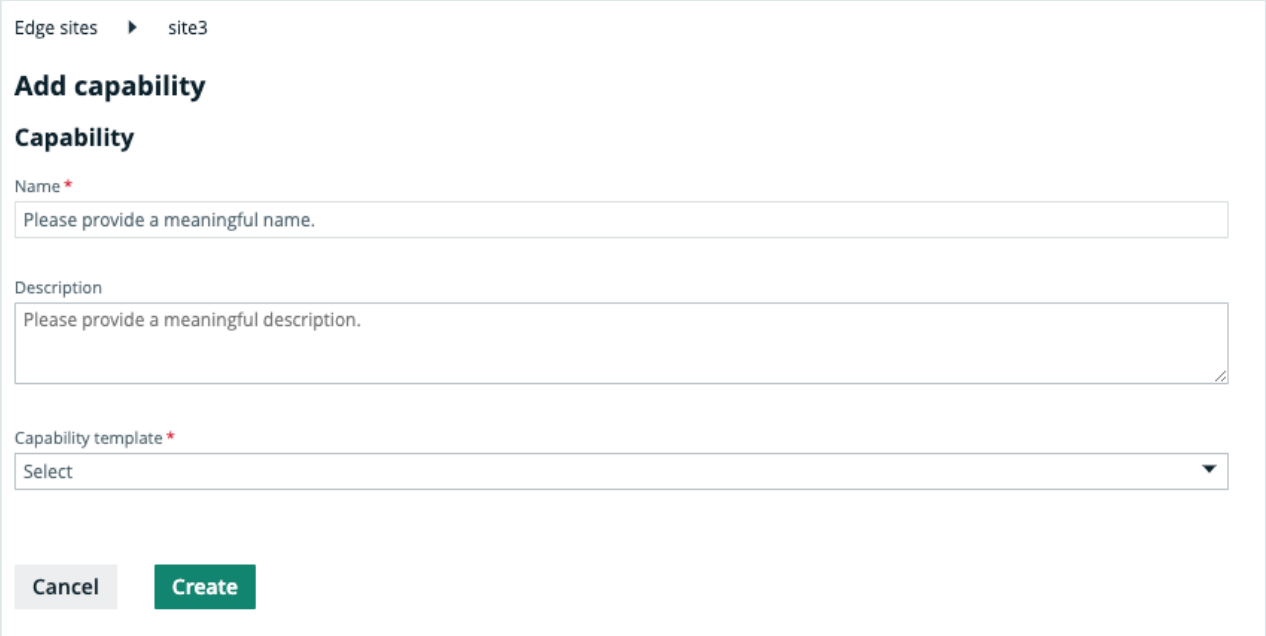
Capability template structure

Each Edge capability template contains the following:

File	Description
A manifest file (YAML)	This file contains the capability metadata and input parameter requirements.
A workflow file (YAML)	This file defines the workflow and binds the parameters to capability containers.
Docker images	One or more Docker images that implement the business logic.

Page layout

The following image shows the page for adding an edge capability.



Field	Description	Required
Capability	This section contains the general information about the capability.	
Name	The name of the Edge capability.	✓ Yes

Field	Description	Required
Description	The description of the Edge capability.	✕ No
Capability template	The capability template, which determines the next available sections.	✓ Yes
Custom properties	Custom properties as required by the capability template. You see the custom properties after you select a capability template.	✓ Yes

Add an Edge capability to an Edge site

After you created and installed an Edge site, you can add an Edge capability to perform specific tasks on a data source. For example, you can register a data source using a JDBC connection that belongs to an Edge capability.

Prerequisites

- You have a global role that has the System administration global permission.
- You have a global role with the Manage connections and capabilities global permission, for example Edge integration engineer.
- You have a global role with the Register profiling information global permission. (optional)
- You have created and installed an Edge site.
- You have created a JDBC connection.

Steps

> **Tip** For more information about all fields in the capability, go to the online version of the documentation.

1. Open an Edge site.
   a. In the main menu, click , then **Settings**.
      » The Collibra settings page opens.
   b. In the tab pane, click **Edge**.
      » The Edge sites overview appears.
   c. In the Edge site overview, click the name of an Edge site with the status **Healthy**.
      » The Edge site page appears.
2. In the **Capabilities** section, click **Add capability**.

» The **Add capability** page appears.

3. Enter the required information.

Field	Description	Required
Capability	This section contains the general information about the capability.	
Name	The name of the Edge capability.	✓ Yes
Description	The description of the Edge capability.	✕ No
Capability template	The capability template, which determines the next available sections. Select the Catalog JDBC ingestion capability template to register a data source. Select the JDBC Profiling capability template to profile and classify your data. Select the DQ Connector capability template to ingest Data Quality user-defined rules, metrics, and dimensions into Collibra Data Catalog. **Important** Data Quality is only available in beta. Please create a support ticket to get access. Select the S3 synchronization capability template to connect to Amazon S3.	✓ Yes
Custom properties	Custom properties as required by the capability template. You see the custom properties after you select a capability template.	✓ Yes

Field	Description	Required
Connection	This section contains information to connect to the data source.	
JDBC connection	The connection to the data source.	✓ Yes
JDBC data source type	The data source type of the data source that you want to ingest.	✓ Yes
Supports schemas	A text field where you have to enter *True* to enable database registration of data sources that have no schema. If the data source has schemas, you can ignore this field. **Tip** If the data source does not have a schema, Data Catalog creates a Schema asset with the same name as the full name of the database.	✕ No
DQ	This section contains information about the Collibra Data Quality connection.	
Base URL	Your Collibra Data Quality URL	✓ Yes
Username	The Collibra Data Quality username for this connection.	✓ Yes
Password	The Collibra Data Quality password for this connection.	✓ Yes
Encryption options	Select the type of encryption to use. Default: *To be encrypted by Edge management server*.	

Field	Description	Required
Issuer of the JWT	If you have selected *Encrypted with public key*, enter your JWT issuer.	✕ No
Collibra metadata model	This section contains information about where to ingest Data Quality assets.	
DQ Rules domain id	The UUID of the **Rulebook Domain** for the ingested Data Quality rules.	✓ Yes
DQ Metrics domain id	The UUID of the **Business Asset Domain** for the ingested Data Quality metrics.	✓ Yes
DQ Dimensions domain id	The UUID of the **Business Asset Domain** for the ingested Data Quality dimensions.	✓ Yes
Default DQ Dimension name	The default **Data Quality Dimension**, for example *Accuracy*, *Completeness*, *Consistency* and so on. Default: *Completeness*.	✓ Yes
DQ Metric classified by DQ Dimension relation type id	The UUID of the **Data Quality Metric classified by / classifies Data Quality Dimension** relation. If left unspecified, this relation will not be added.	✕ No
Assets are imported in batches of this size	The batch size of the ingestion. Default: *5000*.	✓ Yes
S3 service account	This section contains the information on how to connect to Amazon S3.	

Field	Description	Required
AWS Connection	The AWS connection to be used.	✓ Yes
IAM role	The IAM role used by AWS Glue crawlers.	✓ Yes
Encryption options	Select the type of encryption used to store the IAM role. Default: *To be encrypted by Edge management server*.	✓ Yes
General	This section contains general information about logging. **Note** This section only applies to JDBC capabilities.	
Debug	An option to automatically send Edge infrastructure log files to Collibra Data Intelligence Cloud. By default, this option is set to *false*. **Note** We highly recommend to only send Edge infrastructure log files to Collibra Data Intelligence Cloud when you have issues with Edge. If you set it to *true*, it will automatically revert to *false* after 24h.	✗ No
Log level	An option to determine the verbosity level of Catalog connector log files. By default, this option is set to *No logging*.	✗ No

4. Click **Save**.
   » The capability is added to the Edge site.
   » The fields become read-only.
5. Click **Run**.
   » The Edge site connects to your data source.

What's next?

You can now register a data source via Edge.

Edit an Edge capability of an Edge site

You can edit an Edge capability of an Edge site, for example to change the custom properties.

Prerequisites

- You have a global role that has the System administration global permission.
- You have a global role with the Manage connections and capabilities global permission, for example Edge integration engineer.
- You have a global role with the Register profiling information global permission. (optional)
- You have created and installed an Edge site.
- You have added an Edge capability to the Edge site.

Steps

> **Tip** For more information about all fields in the capability, go to the online version of the documentation.

1. Open an Edge site.
   a. In the main menu, click , then **Settings**.
      » The Collibra settings page opens.
   b. In the tab pane, click **Edge**.
      » The Edge sites overview appears.
   c. In the Edge site overview, click the name of an Edge site with the status **Healthy**.
      » The Edge site page appears.
2. In the **Capabilities** section, click the name of an Edge capability.
   » The **Capability** page appears and shows a read-only overview of the capability.
3. Click **Edit**.
4. Enter the required information.

Field	Description	Required
Capability	This section contains the general information about the capability.	

Field	Description	Required
Name	The name of the Edge capability.	✓ Yes
Description	The description of the Edge capability.	✕ No
Capability template	The capability template, which determines the next available sections. Select the Catalog JDBC ingestion capability template to register a data source. Select the JDBC Profiling capability template to profile and classify your data. Select the DQ Connector capability template to ingest Data Quality user-defined rules, metrics, and dimensions into Collibra Data Catalog. **Important** Data Quality is only available in beta. Please create a support ticket to get access. Select the S3 synchronization capability template to connect to Amazon S3.	✓ Yes
Custom properties	Custom properties as required by the capability template. You see the custom properties after you select a capability template.	✓ Yes
Connection	This section contains information to connect to the data source.	
JDBC connection	The connection to the data source.	✓ Yes

Field	Description	Required
JDBC data source type	The data source type of the data source that you want to ingest.	✓ Yes
Supports schemas	A text field where you have to enter *True* to enable database registration of data sources that have no schema. If the data source has schemas, you can ignore this field. **Tip** If the data source does not have a schema, Data Catalog creates a Schema asset with the same name as the full name of the database.	✕ No
DQ	This section contains information about the Collibra Data Quality connection.	
Base URL	Your Collibra Data Quality URL	✓ Yes
Username	The Collibra Data Quality username for this connection.	✓ Yes
Password	The Collibra Data Quality password for this connection.	✓ Yes
Encryption options	Select the type of encryption to use. Default: *To be encrypted by Edge management server*.	
Issuer of the JWT	If you have selected *Encrypted with public key*, enter your JWT issuer.	✕ No
Collibra metadata model	This section contains information about where to ingest Data Quality assets.	

Field	Description	Required
DQ Rules domain id	The UUID of the **Rulebook Domain** for the ingested Data Quality rules.	✓ Yes
DQ Metrics domain id	The UUID of the **Business Asset Domain** for the ingested Data Quality metrics.	✓ Yes
DQ Dimensions domain id	The UUID of the **Business Asset Domain** for the ingested Data Quality dimensions.	✓ Yes
Default DQ Dimension name	The default **Data Quality Dimension**, for example *Accuracy*, *Completeness*, *Consistency* and so on. Default: *Completeness*.	✓ Yes
DQ Metric classified by DQ Dimension relation type id	The UUID of the **Data Quality Metric classified by / classifies Data Quality Dimension** relation. If left unspecified, this relation will not be added.	✗ No
Assets are imported in batches of this size	The batch size of the ingestion. Default: *5000*.	✓ Yes
S3 service account	This section contains the information on how to connect to Amazon S3.	
AWS Connection	The AWS connection to be used.	✓ Yes
IAM role	The IAM role used by AWS Glue crawlers.	✓ Yes

Field	Description	Required
Encryption options	Select the type of encryption used to store the IAM role. Default: *To be encrypted by Edge management server*.	✓ Yes
General	This section contains general information about logging. **Note** This section only applies to JDBC capabilities.	
Debug	An option to automatically send Edge infrastructure log files to Collibra Data Intelligence Cloud. By default, this option is set to *false*. **Note** We highly recommend to only send Edge infrastructure log files to Collibra Data Intelligence Cloud when you have issues with Edge. If you set it to *true*, it will automatically revert to *false* after 24h.	✕ No
Log level	An option to determine the verbosity level of Catalog connector log files. By default, this option is set to *No logging*.	✕ No

5. Click **Save**.
   » The fields become read-only.
6. Click **Run** to rerun the Edge capability.
   » The Edge site reconnects to your data source with the updated properties.

Delete an Edge capability from an Edge site

You can remove an Edge capability from an Edge site if you no longer need it.

> **Warning** If you delete a JDBC Profiling capability and synchronize previously profiled and classified schemas again, the profiling and classification results are removed.

Prerequisites

- You have a global role that has the System administration global permission.
- You have a global role with the Manage connections and capabilities global permission, for example Edge integration engineer.
- You have created and installed an Edge site.
- You have added an Edge capability to the Edge site.

Steps

1. Open an Edge site.
   a. In the main menu, click , then **Settings**.
      » The Collibra settings page opens.
   b. Click **Edge**
      » The Edge sites overview appears.
   c. In the Edge site overview, click the name of an Edge site.
      » The Edge site page appears.
2. In the **Capabilities** section, click the name of a capability.
   » The **Capability** page appears and shows a read-only overview of the capability.
3. Click **Delete**.
4. Click **Delete Capability**.
   » The **Capability** is deleted from the Edge site.

Chapter 7

Maintaining Edge sites

In this section, you get to know how you can maintain your Edge site installations, such as taking backups or updating credentials.

Running Edge tools

This section contains an overview on how to use the Edge tools, for example to create a backup of your Edge site.

Prepare the Edge tools on K3S

On k3s, the Edge tool is downloaded at the end of a successful installation.

Alternatively, you can download it from the cluster:

```
TOOLS_POD=$(sudo /usr/local/bin/kubectl -n collibra-edge get
pod -l edge.collibra.com/contains=tools -o jsonpath='{.items
[0].metadata.name}')

sudo /usr/local/bin/kubectl cp collibra-edge/$TOOLS_POD:edge
/usr/local/bin/edge

sudo chmod +x /usr/local/bin/edge
```

The Edge command is in **/usr/local/bin** on the host, that is your first worker node, so you run the Edge command on the actual host where k3s runs.

Overview Edge commands on K3S

Edge tool	Command for K3S
Uninstall Edge	`/usr/local/bin/uninstall-edge.sh`
Create Edge diagnostics file	• Edge site is not yet installed: `<extracted installer directory>/resources/tools/edge-diagnostics.sh -d <file name>.tgz` • Edge site is up and running: `edge diagnostics -d <file name>.tgz`

Edge tool	Command for K3S
Create an Edge site backup	`edge backup -o /<path to folder>/<backup-name>.yaml`
Set Edge storage cache ttl	`edge cachettl --ttl <value in days>`
Retrieve logs from a catalog connector	`edge catalog-connector --jobid <Edge job ID> \ --dst <path to destination>`
Update Collibra credentials	• Interactive way: `edge update-dgc-creds -i` • Explicit update: `edge update-dgc-creds <username> <password> <url collibra environment>`
Update forward proxy settings	`edge update-outbound-proxy --update-outbound-proxy /path/to/proxy.properties`
Get help to set up no_proxy configuration	• Edge site is not yet installed: `<extracted installer directory>/resources/tools/edge-get-noproxy.sh k3s` • Edge site is up and running: `edge get-noproxy k3s`

Prepare Edge tools on EKS

Edge is installed from a Linux machine that has access to the actual k8s cluster.

There is no automatic download of Edge tool after installation, because we don't want to enforce it in some location. Therefore, you have to download the Edge tool to your Linux machine, for example in your current folder:

```
TOOLS_POD=$(kubectl -n collibra-edge get pod -l
edge.collibra.com/contains=tools -o jsonpath='{.items
[0].metadata.name}')

kubectl cp collibra-edge/$TOOLS_POD:edge edge

chmod +x edge
```

You can now run Edge commands from your current folder.

> **Note** As you are not on the worker node itself, you cannot collect worker node diagnostics. If you need these diagnostics, create a support ticket.

Overview Edge commands on EKS

Edge tool	Command for EKS
Uninstall Edge	`<extracted installer directory>/resources/tools/installer-job/uninstall-edge-on-managed-k8s.sh`
Create Edge diagnostics file	• Edge site is not yet installed: `<extracted installer directory>/resources/tools/edge-diagnostics.sh -d <file name>.tgz` • Edge site is up and running: `edge diagnostics -d <file name>.tgz`
Create an Edge site backup	`edge backup -o /<path to folder>/<backup-name>.yaml`
Set Edge storage cache ttl	`edge cachettl <value in seconds>`

Edge tool	Command for EKS
Retrieve logs from a catalog connector	`edge catalog-connector --jobid <Edge job ID> \ --dst <path to destination>/<file name>.txt`
Update Collibra credentials	• Interactive way: `edge update-dgc-creds -i` • Explicit update: `edge update-dgc-creds <username> <password> <url collibra environment>`
Update forward proxy settings	`edge update-outbound-proxy --update-outbound-proxy /path/to/proxy.properties`
Get help to set up no_proxy configuration	• Edge site is not yet installed: `<extracted installer directory>/resources/tools/edge-get-noproxy.sh eks <clustername>` • Edge site is up and running: `edge get-noproxy eks <clustername>`

Edit an Edge site

You can edit a Edge site to give it another name or description.

Prerequisites

- You have created an Edge site.
- You have a global role that has the System administration global permission.
- You have a global role that has the Manage Edge sites global permission.

Steps

1. Open an Edge site.
   a. In the main menu, click , then **Settings**.
      » The Collibra settings page opens.
   b. Click **Edge**
      » The Edge sites overview appears.
   c. In the Edge site overview, click the name of an Edge site.
      » The Edge site page appears.
2. In the top right corner, click **Actions** → **Edit**.
   The **Edit Edge site** wizard starts.
3. Enter the required information.

Field	Description
Name	The name of the Edge site. Use a meaningful name, for example NetherlandsDataCentre1. Do not use spaces or special characters. This field is mandatory and the name must be globally unique.
Description	The description of the Edge site. We recommend to put at least basic location information of the Edge site. This field is mandatory.

4. Click **Save**.
   » The Edge sites overview appears with the new name and description.

Update Edge user password

When you download the Edge site installer, a dedicated user is created in Collibra Data Intelligence Cloud. This user has always "Edge" as first name and the "Edge site name" as last name.

A user will be created for each Edge site. This user is deleted when you delete the Edge site.

> **Note** The Edge user must have the Connect Edge to Collibra global permission.

Steps

1. Reset the password of the Edge user in Collibra.
2. Connect to the Edge master node via SSH.
3. Run the following script: `/usr/local/bin/edge update-dgc-creds -i`
4. Enter the username and new password of the Edge user.

Update the outbound proxy configuration

If you have to change the outbound proxy configuration of a running Edge site, you can use Collibra's outbound proxy update script.

Steps

1. Find the **proxy.properties** file on the server that you used during the configuration of the outbound proxy.
2. Update the file with the new property values and save the file.
3. Go to **/usr/local/bin** and run the following command:

```
./edge update-outbound-proxy -u /path/to/proxy.properties
```

Help file of the script

```
$ /usr/local/bin/edge update-outbound-proxy --help
  Collibra Edge Utility for updating Outbound Proxy settings.
  Usage:
         edge update-outbound-proxy.sh -h|--help
         edge update-outbound-proxy.sh -g|--generate-template
<filename>
         edge update-outbound-proxy.sh -u|--update-outbound-
proxy <filename>


  -h|--help                   - Show help
  -g|--generate-template      - generate template file for
proxy properties in <filename>
  -u|--update-outbound-proxy  - update outbound-proxy secret
based on proxy properties <filename>
```

Back up and restore an Edge site

To avoid losing your Edge site configurations, you can back up an Edge site and later restore it, for example when you want to reinstall an Edge site with a new installer.

> **Note** For privacy reasons, Edge site backups remain in your personal environment and are not sent to the cloud.

Back up an Edge site

On the server that runs your Edge site, execute the following command:

```
~$ ./edge backup -o /<path to folder where you want to save the backup file>/<backup-name>.yaml
```

» Edge creates a backup of your Edge site in the selected folder of the command.

> **Tip** If the edge command is not available, you first have to download it.

On the server from which you manage your EKS cluster, execute the following commands:

```
~$ ./edge backup -o /<path to folder where you want to save the backup file>/<backup-name>.yaml
```

» Edge creates a backup of your Edge site in the defined folder of the last command.

> **Tip** If the edge command is not available, you first have to download it.

Restore an Edge site

1. Optionally, download a new Edge installer.
1. Open an Edge site.
   a. In the main menu, click ⠿, then ⚙ **Settings**.
      » The Collibra settings page opens.
   b. Click **Edge**
      » The Edge sites overview appears.
   c. In the Edge site overview, click the name of an Edge site.
      » The Edge site page appears.
2. In the top right corner, click **Actions** → **Redownload installer**.
   » A new Edge installer is downloaded.
2. Run the Edge installer and add the backup file as a parameter:

```
install-master.sh properties.yaml -r registries.yaml -b /<path
to backup file>/edge-backup.yaml
```

1. Optionally, download a new Edge installer.
1. Open an Edge site.
   a. In the main menu, click ⠿, then ⚙ **Settings**.
      » The Collibra settings page opens.
   b. Click **Edge**
      » The Edge sites overview appears.
   c. In the Edge site overview, click the name of an Edge site.
      » The Edge site page appears.
2. In the top right corner, click **Actions** → **Redownload installer**.
   » A new Edge installer is downloaded.
2. Run the Edge installer and add the backup file as a parameter:

```
./run-installer-job.sh properties.yaml --repositories
repositories.json --set collibra_
edge.collibra.minio.persistence.size=120Gi -b /<path to backup
file>/edge-backup.yaml
```

Delete an Edge site

You can delete an Edge site if you no longer need it.

Prerequisites

- You have created an Edge site.
- You have a global role that has the System administration global permission.
- You have a global role that has the Manage Edge sites global permission.

Steps

1. Open an Edge site.
   a. In the main menu, click , then **Settings**.
      » The Collibra settings page opens.
   b. Click **Edge**
      » The Edge sites overview appears.
   c. In the Edge site overview, click the name of an Edge site.
      » The Edge site page appears.
2. In the top right corner, click **Actions** → **Delete Edge site**.
   » The Delete Edge site wizard starts.
3. Click **Delete Edge site**.
   » The Edge sites overview appears, without the deleted Edge site.
4. On the server that hosts the Edge site, go to **/usr/local/bin** where you can find the uninstall script **uninstall-edge.sh**, then run one of the following commands:

	Command
Delete Edge site but keep its data. The data consists of drivers, required files for capabilities, and data that was saved by Edge capabilities	`/usr/local/bin/uninstall-edge.sh`

	Command
Delete Edge site and its data.	`/usr/local/bin/uninstall-edge.sh --remove-local-data`
Delete Edge site without confirmation request, for example if you want to delete the site via a script. You can use this in combination with removing the site data.	`/usr/local/bin/uninstall-edge.sh --remove-local-data --force`

Warning When you delete an Edge site, the EBS volumes containing the data are also removed. If you like to keep your data, first back up these EBS volumes.

1. Open an Edge site.
   a. In the main menu, click ⠿, then ⚙ **Settings**.
      » The Collibra settings page opens.
   b. Click **Edge**
      » The Edge sites overview appears.
   c. In the Edge site overview, click the name of an Edge site.
      » The Edge site page appears.
2. In the top right corner, click **Actions** → **Delete Edge site**.
   » The Delete Edge site wizard starts.
3. Click **Delete Edge site**.
   » The Edge sites overview appears, without the deleted Edge site.
4. On the server from which you manage your EKS cluster, run this command:

```
<extracted installer>/resources/tools/installer-job/uninstall-edge-on-managed-k8s.sh
```

Troubleshooting Edge

In this section, you find some articles that help you to troubleshoot Edge issues.

General troubleshooting Edge

The following table shows how to solve issues you encountered while working with Edge. Select the tab of your installation type, k3s or EKS.

Issue	Proposed solution
You get the following error message: `Out of disk space`	You have to restart the Kubernetes pod in Edge. Run the following command: `sudo /usr/local/bin/kubectl delete pod <pod_name> --namespace <pod_namespace>` **Tip** For more information about Pods and namespaces, see the Kubernetes documentation.

You get one of the following error message:
• `Cannot allocate memory`
• `Error syncing pod`
	You have to restart Edge.
	Do the following:
	1. Run the following commands to remove all workflows:
	`sudo /usr/local/bin/kubectl delete --all workflows --namespace=collibra-capabilities`
	`sudo /usr/local/bin/kubectl delete --all workflows --namespace=collibra-fast`
	2. Run the following command to reboot Edge:
	`sudo reboot`

Issue	Proposed solution
You get the following error message: `Out of disk space`	You have to restart the Kubernetes pod in Edge. Run the following command: `kubectl delete pod <pod_name> --namespace <pod_namespace>` **Tip** For more information about Pods and namespaces, see the Kubernetes documentation.

You get one of the following error message:
 	
• `Cannot allocate memory`
• `Error syncing pod`
	You have to restart Edge.
	
	Do the following:
	
	1. Run the following commands to remove all workflows:
	
	`kubectl delete --all workflows --namespace=collibra-capabilities`
	
	`kubectl delete --all workflows --namespace=collibra-fast`
	
	2. Run the following command to reboot Edge:
	
	`sudo reboot`

Edge logging

When you encounter an issue in Edge, you can use diagnostic files and log files to provide a lot of indications about the issue. If you want to report a problem to Collibra support, you can include the diagnostics or log files in the support ticket. As a result, Collibra support will be able to determine what went wrong and find a solution to your issue.

You can create a diagnostics file which is a ZIP file with log files and information about the environment in which you install an Edge site. Edge also generates two types of log files that are not included in the diagnostics file:

- Edge infrastructure log files, which are sent to Collibra Data Intelligence Cloud immediately upon creation.
- Metadata connector log files, which can only be stored locally.

Edge diagnostics file

The Edge diagnostics file is a ZIP file that is created by running the diagnostics script in the Edge site installer folder. The diagnostics script checks amongst others:

- Your operating system setup
- Your firewall settings
- Connectivity information
- Edge cluster logs.

You can send the diagnostics file to Collibra support when you have an issue with the Edge site installation.

Edge infrastructure log files

Edge infrastructure logs contain Edge infrastructure information, for example Edge status updates, capability information. The logs can be used by Collibra Support to help solve a

general issue with Edge. The log files do not contain any database content or private information.

By default, the Edge infrastructure log files are always enabled on an information level. You can enable debug level logging per specific capability when you add or edit an Edge capability. As a result, Edge sends infrastructure log with more information about that capability to Collibra Data Intelligence Cloud. Edge infrastructure log files can contain the following information:

- Job execution phases
- The Edge status
- Service updates
- System upgrades

These log files can only be accessed by Collibra Support.

> **Note** By default, **Debug** logging for an Edge capability is set to `False`. We highly recommend only enabling the **Debug** logging for an Edge capability if an issue arises.

Metadata connector log files

Metadata connector log files contain the logs of the JDBC connections between the Edge capability and your data source. These log files can be used by Collibra Support to help solve issues with processing or accessing data. The log files may contain information about your data source.

For security reasons, these log files are not automatically sent to your Collibra Data Intelligence Cloud environment. You can, however, create the log files, save and review them locally and then attach them to a Collibra support ticket.

Edge system monitoring

The system monitoring, executed via SignalFX, sends the following information to your Collibra environment:

- CPU usage
- Memory usage
- Network statistics

Collibra Support can then analyze this information to troubleshoot potential anomalies. These data are only available to Collibra personnel.

Verbosity log levels

The verbosity log levels indicate how much information you want to see in the Catalog Connector log files. You can change the verbosity log levels in the Edge capability for which you want to create logs. The following verbosity log levels are available:

Verbosity log level	Description
No logging	The Catalog Connector logs are not created. This is the default.
Low	The Catalog Connector logs contain the following: • All connection query logs • Any errors
Middle	The Catalog Connector logs include the Low logs and: • All cache queries • Additional information about the request
High	The Catalog Connector logs include the Middle logs and: • The body of the request • The response

Create an Edge diagnostics file

You can create an Edge diagnostics file to check issues with the Edge site installation in your environment.

Prerequisites

- You have created an Edge site.
- You have downloaded the Edge installer.

Steps

Edge site is not yet installed

Running the diagnostics script without an installed Edge site, checks if your system meets all requirements to install the Edge site.

1. Extract the Edge installer.
2. On the command line, go to the folder with the extracted files.
3. In this folder, go to **resources/tools**.
4. Run the following command to create the diagnostics file:

   ```
   edge-diagnostics.sh --diag-file <file name>.tgz
   ```

   » A TGZ file with the given file name is created and contains all Edge diagnostics file.

Edge site is already installed

On the command line, run the following command to create the diagnostics file:

```
edge diagnostics --diag-file <file name>.tgz
```

» A TGZ file with the given file name is created and contains all Edge diagnostics file.

> **Tip** If the edge command is not available, you first have to download it.

What's next?

You can send the diagnostics file to Collibra support to help you resolve your installation issues.

Create Metadata connector log files

If you have an issue with a JDBC connection, for example while registering a data source via Edge, you can create the Metadata connector log files and then save and review them locally. If you create a support ticket, you can attach the reviewed Metadata connector log files to Collibra support to help you with your issue.

Prerequisites

- You have a global role that has the System administration global permission.
- You have a global role with the Manage connections and capabilities global permission, for example Edge integration engineer.
- You have created and installed an Edge site.
- You have created a JDBC connection and an Edge capability.

Steps

1. Edit the Edge capability that contains the JDBC connection for which you want to create a log file.
    a. Click the name of the Edge capability to open it.
    b. Click **Edit**.
    c. In the **General** section, click the **Log level** drop-down menu.
    d. Select the log verbosity level.

    > **Tip** The level must be at least *low*.

    e. Click **Save**.
        » The fields become read-only.
2. Click **Run** to rerun the Edge capability.
3. Contact Collibra support to request the Edge job ID of the Edge capability.
4. Run the following command:

```
./edge catalog-connector --jobid <Edge job ID> --dst <path to destination>
```

» The log file is created and stored in the predefined destination.

> **Tip** If the edge command is not available, you first have to download it.

Prerequisites

- You have a Linux host with kubectl access to your EKS installation.
- You have mc (minio client) installed in /usr/local/bin:

```
sudo curl -L "https://dl.min.io/client/mc/release/linux-
amd64/mc" -o /usr/local/bin/mc
sudo chmod +x /usr/local/bin/mc
```

Steps

Execute the following commands:

```
kubectl -n collibra-edge port-forward service/minio 9000:9000 &

MC_ACCESSKEY="$(kubectl get secrets edge-minio -n collibra-edge
-o jsonpath="{.data.accesskey}" | base64 --decode)"

MC_SECRETKEY="$(kubectl get secrets edge-minio -n collibra-edge
-o jsonpath="{.data.secretkey}" | base64 --decode)"

export MC_HOST_edge="http://${MC_ACCESSKEY}:${MC_
SECRETKEY}@localhost:9000"

mc cp --quiet --recursive edge/cdata/<jobId> <destination_
directory>

pkill -f "port-forward"
```

Enable debug logging for Edge infrastructure logs

By default, the Edge infrastructure logs are always enabled on an information level. If you have an issue with Edge in general, you can enable Edge to create Edge infrastructure debug log files and send them to Collibra Data Intelligence Cloud. Collibra support uses these log files to solve Edge issues.

Prerequisites

- You have a global role that has the System administration global permission.
- You have a global role with the Manage connections and capabilities global permission, for example Edge integration engineer.
- You have created and installed an Edge site.
- You have created a JDBC connection and an Edge capability.

Steps

1. In the main menu, click , then **Settings**.
   » The Collibra settings page opens.
2. On the **Settings** page, click **Edge**.
   » The Edge sites overview appears.
3. Click the site that runs the capability with issues.
   » The site details page appears.
4. On the **Capabilities** tab, click the name of the Edge capability.
5. Click **Edit**.
6. In the **General** section, click the **Debug** drop-down menu and select *true*.

   > **Note** This field is by default set to *false*. If you set it to *true*, it will automatically revert to *false* after 24 hours.

7. Click **Save**.
   » The fields become read-only.

8. Click **Run** to rerun the Edge capability.
   » The log files are automatically sent to Collibra Data Intelligence Cloud.

Disable OpenTelemetry

If you reinstalled an Edge site with a new version it is possible that the new setup is not working due to a missing network connectivity. You would have to request for OpenTelemetry to be added again. If this request takes days to be completed, you may want to disable OpenTelemetry to still have a running Edge site.

Disable OpenTelemetry at installation time

Add the flag `--disable-otel` when you run the installation script.

```
sudo sh install-master.sh --storage-path /var/edge/storage
properties.yaml \
 --disable-otel \
 -r registries.yaml
```

```
./run-installer-job.sh properties.yaml --repositories
repositories.json \
 --set collibra_edge.collibra.minio.persistence.size=120Gi \
 --disable-otel
```

Edge FAQ

The following table contains the most frequently asked questions about Edge that were not answered anywhere else in the Edge documentation.

Question	Answer
Who benefits from using Edge?	All customers who want to ingest data into Collibra Data Intelligence Cloud benefit from Edge. Some of the benefits for using Edge are: • Data is processed in the customer's secure environment and only the process results are sent to Collibra Data Intelligence Cloud. • Edge can automatically anonymize sensitive profiling data before sending it to Collibra Data Intelligence Cloud. • Edge can automatically classify the metadata and send the classification results together with the profiling results to Collibra Data Intelligence Cloud. • Edge enables better profiling performance, because data no longer has to be copied or moved.

Question	Answer

Does Edge replace the Jobserver?	Customers can choose between Edge and Jobserver.
	
	The main differences between Edge and Jobserver are the following:
	
	• Edge is based on Kubernetes, a distributed runtime, which means:
	  ◦ It can scale out.
	  ◦ It offers built in resource management.
	  ◦ It has out of the box high availability.
	  ◦ It has reliable delivery of results to Collibra Data Intelligence Cloud.
	• Edge is a Collibra service compatible with on-premises as well as cloud environments.
	• Edge offers continuous delivery of capability types and updates will be installed on a regular basis.
	• Edge updates are included in Collibra Data Intelligence Cloud releases.
	
	Jobserver features correspond to Edge capabilities, each one is developed and deployed independently of one another. In the future, we will provide a script for migrating features from Jobserver to Edge where applicable.

Question	Answer
Can Edge use Kubernetes provided by a Cloud vendor, for example GKE, AKS, Amazon EKS.	When the Edge site is installed in a Cloud environment, it does not use a managed Kubernetes provided by the Cloud vendor, because Kubernetes is already included in the Edge site installation process. Collibra manually manages the cluster on top of EC2 or similar machines in other platforms. In the first releases, we cannot benefit from seamless integration of various Cloud services offered by those platforms, for example, embedded authentication, auto-scaling and databases.
Can Edge be installed on top of a customer's provided Kubernetes cluster?	You can now install Edge on EKS.
How does Edge connect to Collibra Data Intelligence Cloud?	An Edge site is installed in the customer's environment, close to the data source. The Edge site communicates to Collibra Data Intelligence Cloud and other 3rd party systems using an HTTPS connection. **Note** The connection between an Edge site and Collibra Data Intelligence Cloud is outbound-only.
Who controls Edge?	Edge is controlled by the customer through local access via the Collibra Data Intelligence Cloud user interface. You can also use local access via the Linux shell for advanced troubleshooting when Edge is unable to connect.
Can an Edge site connect to more than one Collibra environment?	No. Every Edge site belongs and authenticates to only one Collibra Data Intelligence Cloud environment.

Question	Answer
Can Edge use customer-provided certificates to connect to Collibra Data Intelligence Cloud?	Currently, we do not support this. Edge is a Collibra product that can run on the customer's on-premises or cloud environment. The authentication between the Edge site and Collibra Data Intelligence Cloud is controlled and secured by Collibra. The keys and credentials are generated when you install the Edge site.
Does Edge support mTLS when connecting to Collibra Data Intelligence Cloud?	Currently, we do not support this.
Is Edge horizontally scalable?	Currently, Edge is not horizontally scalable. You cannot add more nodes.