Cloudera AI

Cloudera AI Workbenches (on premises)

Date published: 2020-07-16
Date modified: 2025-10-31

https://docs.cloudera.com/

Legal Notice

Contents

Provisioning a Cloudera AI Workbench

In Cloudera AI on premises, the Cloudera AI Workbench provides a space for the data scientists' work. After your Administrator has created or given you access to an environment, you can set up a workbench.

Before you begin

- The first user to access the Cloudera AI Workbench after it is created must have the EnvironmentAdmin role assigned.
- If you have the cert-manager setup in your cluster, the configuration to enable the Certification Manager has to be completed prior to provisioning a Cloudera AI Workbench. For details, see Configuring cluster issuer for cert-manager. These steps are required only if you have the cert-manager setup.

  During the workspace installation process, Cloudera AI validates the cluster issuer configurations. If a cluster issuer is properly labeled and annotated, it will be utilized to sign certificates for both the workbench and workload domains.
- The Certification Manager has a limitation that restricts common names to a maximum of 64 characters. To address this, the static subdomain feature must be used to configure a shorter Cloudera AI subdomain. During the workbench provisioning workflow, preflight checks are conducted to ensure that the Cloudera AI Workbench domain name complies with this character limit.

Procedure

1. Log in to the Cloudera on premises web interface using your corporate credentials or other credentials that you received from your Cloudera administrator.
2. Click Cloudera AI Workbenches.
3. Click Provision Workbench. The Provision Workbench panel displays.
4. On Provision Workbench panel, fill out the following fields:
   a) Workbench Name - Give the Cloudera AI Workbench a name. For example, test-cml. Do not use capital letters in the workbench name.
   b) Select Environment - From the dropdown menu, select the environment where the Cloudera AI Workbench must be provisioned. If you do not have any environments available to you in the dropdown, contact your Cloudera administrator to gain access.
   c) Namespace - Enter the namespace to use for the Cloudera AI Workbench.
   d) NFS Server - Select Internal to use an NFS server that is integrated into the Kubernetes cluster. This is the recommended selection at this time.

      The path to the internal NFS server is already set in the environment.
5. In the Production Cloudera AI option enable the following features:
   a) Enable Governance - Select this checkbox to enable advanced lineage and governance features.

      Governance Principal Name - If the Enable Governance checkbox is selected, you can use the default value of mlgov, or enter an alternative name. The alternative name must be present in your environment and be given permissions in Ranger to allow the Cloudera AI Governance service deliver events to Atlas.
   b) Enable Model Metrics - Selec this checkbox to enable exporting metrics for models to a PostgreSQL database.

6. In the Other Settings option enable the following features:
   a) Enable Ring Fencing - Select this checkbox to enable ring fencing, so that all Cloudera AI infrastructure pods for the workbench will be exclusively scheduled on the dedicated Cloudera AI nodes, ensuring complete resource isolation.

      For more details on Ring Fencing, see *Enabling Ring Fencing in Cloudera AI Workbench*.
   b) Enable TLS - Select this checkbox to enable https access to the workbench.

      To enable TLS, follow the guidelines in Deploying an Cloudera AI Workbench with support for TLS.

      Upload the certificates to Management Console Settings CA Certificates Miscellaneous . Alternatively, you can configure the Certification Manager to automatically generate certificates for the Cloudera AI infrastructure and workload subdomains. For more details, see Certification Manager service for increased security.
   c) Enable Monitoring - Select this checkbox for administrators (users with the EnvironmentAdmin role) to use a Grafana dashboard to monitor resource usage in the provisioned workbench.
   d) Cloudera AI Static Subdomain - Specify a custom name for the workbench endpoint, which is also used for the URLs of models, applications, and experiments. Only one workbench with the specific subdomain endpoint name can be running at a time. You can create a wildcard certificate for this endpoint in advance. The workbench name has this format: <static subdomain name>.<application domain>

      Note: The endpoint name can have a maximum of 15 characters, using alphanumerics and hyphen or underscore only, and must start and end with an alphanumeric character.
7. Click the Provision Workbench button. The new workbench provisioning process takes several minutes.

Results

Once ring fencing is enabled, all Cloudera AI infrastructure pods for the workbench will be exclusively scheduled on the dedicated Cloudera AI nodes, ensuring complete resource isolation.

What to do next

After the workbench is provisioned, you can log in by clicking the workbench name on the Cloudera AI Workbenches page. The first user to log in must be the administrator.

Test backing up of the Cloudera AI Workbench. Ensure that the backup completes successfully, and then ensure you have a process to back up the workbench at regular intervals.

Related Information

Monitoring Cloudera AI Workbenches
Deploying a Cloudera AI Workbench with support for TLS
Removing Cloudera AI Workbenches
Backing up Cloudera AI Workbenches

Enabling Ring Fencing in Cloudera AI Workbench

The Ring Fencing feature is available from Cloudera AI 1.5.5 SP1 or higher releases. Ring fencing ensures that Cloudera AI infrastructure pods are exclusively scheduled on designated Cloudera AI nodes within the Kubernetes cluster.

About this task

In on premises environments, Cloudera AI infrastructure services share cluster resources with Kubernetes system components, other platform services, and user workloads. When CPU bursting is enabled, resource-intensive users or system workloads can consume additional resources, potentially preempting Cloudera AI services and making the application temporarily inaccessible.

Ring fencing ensures that Cloudera AI infrastructure pods are exclusively scheduled on designated Cloudera AI nodes within the Kubernetes cluster. This isolation is implemented through the following mechanisms:

- Kubernetes taints and tolerations – used to prevent non-Cloudera AI workloads from being scheduled on dedicated nodes.
- Node affinity – used to ensure Cloudera AI pods are scheduled only on the intended nodes.

Together, these mechanisms guarantee that Cloudera AI workloads are scheduled only on nodes dedicated to Cloudera AI.

Before you begin

To use ring fencing:

- The cluster must be configured for ring fencing, ensuring a dedicated set of nodes is allocated for Cloudera AI infrastructure workloads.

  To enable ring fencing for one workbench, the nodes dedicated for Cloudera AI infrastructure must have the following minimum resource requirements:

  - A minimum of 32 CPU cores
  - A minimum of 60 GiB memory
- The designated nodes must be configured with the appropriate labels and taints.

  Apply labels and taints to the nodes for Cloudera AI infrastructure scheduling by using the following command:

  Replace [***NODE NAME***] with the actual node name in the cluster.

  ```
  kubectl taint nodes [***NODE NAME***] cml-infra=true:NoSchedule
  kubectl label nodes [***NODE NAME***] cml-infra-node="true" --overwrite
  ```
- A rolling restart of Cloudera Embedded Container Service is required to migrate non-Cloudera AI infrastructure workloads from nodes dedicated to Cloudera AI infrastructure, relocating them to non-Cloudera AI infrastructure nodes. This process is recommended in the following scenarios:

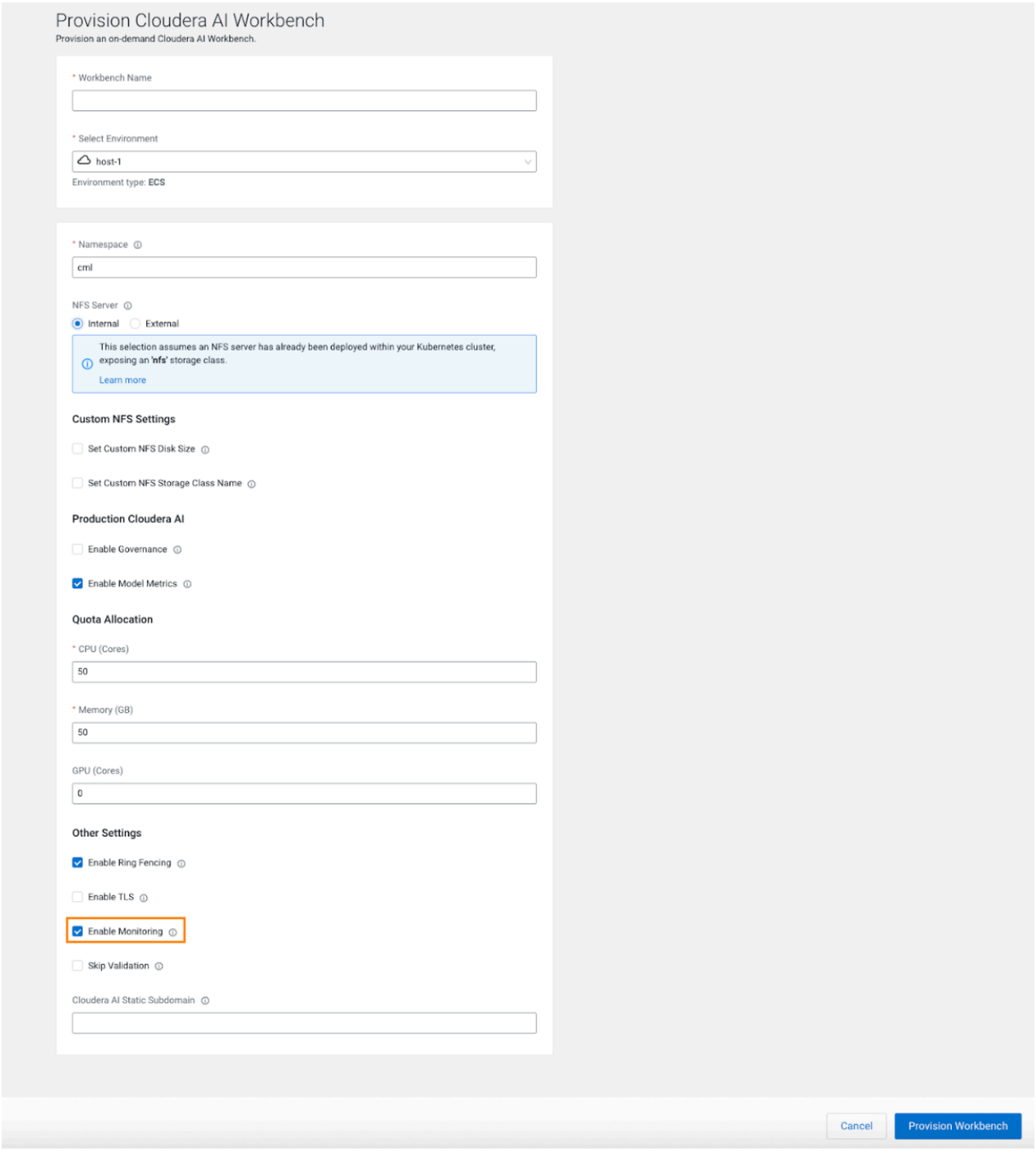
  - Initially, when tainting the nodes.
  - Subsequently, with each addition of new nodes (that is, applying new taints and labels to the nodes).
- For Openshift Container Platform installations, ensure nodes are tainted before installing cpd-pvc. Afterward, a manual rolling restart of the deployment is necessary.

Procedure

Enable Ring Fencing as part of provisioning a workbench.

Ring fencing can only be enabled during the creation of a workbench. If the cluster is configured for ring fencing, the enabling option appears during the workbench creation process.

Figure 1: Enabling Ring Fencing



Once ring fencing is enabled, all Cloudera AI infrastructure pods for that workbench will be exclusively scheduled on the dedicated Cloudera AI nodes, ensuring complete resource isolation.

To verify if Ring Fencing is enabled for a workbench, select 'View Workbench Details' from the 'Actions' menu.

Monitoring Cloudera AI Workbenches

This topic shows you how to monitor resource usage on your Cloudera AI Workbenches.

About this task

Cloudera AI leverages Prometheus and Grafana to provide a dashboard that allows you to monitor how CPU, memory, storage, and other resources are being consumed by Cloudera AI Workbenches. Prometheus is an internal data source that is auto-populated with resource consumption data for each workbench. Grafana is a monitoring dashboard that allows you to create visualizations for resource consumption data from Prometheus.

Each Cloudera AI Workbench has its own Grafana dashboard.

Before you begin

Required Role: EnvironmentAdmin

You need the EnvironmentAdmin to view the Workbench details page.

Procedure

1. Log in to the web interface.
2. Click Cloudera AI Workbenches.
3. For the workbench you want to monitor, click Actions Open Grafana .

Results

Cloudera AI provides you with several default Grafana dashboards:

- K8s Cluster: Shows cluster health, deployments, and pods
- K8s Containers: Shows pod info, cpu and memory usage
- K8s Node: Shows node CPU and memory usage, disk usage and network conditions
- Models: Shows response times, requests per second, CPU and memory usage for model replicas.

You might choose to add new dashboards or create more panels for other metrics. For more information, see the *Grafana documentation*.

What to do next

Note: Prometheus captures data for the previous 30 days.

Related Information

Monitoring and Alerts

Removing Cloudera AI Workbenches

This topic describes how to remove an existing Cloudera AI Workbench and clean up any cloud resources associated with the workbench. Currently, only Cloudera users with both the MLAdmin role and the EnvironmentAdmin account role can remove workbenches.

Procedure

1. Log in to the web interface.
2. Click Cloudera AI Workbenches.

3. Click on the Actions icon and select Remove Workbench.
   a) Force Delete - This property is not required by default. You should first attempt to remove your workbench with this property disabled.

      Enabling this property deletes the workbench from Cloudera but does not guarantee that the underlying kubernetes resources used by the workbench are cleaned up properly. Go to you kuberknetes administration console to make sure that the resources have been successfully deleted.
4. Click OK to confirm.

Upgrading Cloudera AI Workbenches and supported upgrade paths

Learn how to upgrade Cloudera AI Workbenches and the supported upgrade paths.

How to upgrade Cloudera AI Workbenches

1. Log in to Cloudera AI web interface as an Administrator.
2. On the **Workbench** page, click Upgrade corresponding to the workbench name.

Supported upgrade paths

The following table lists the supported upgrade paths for upgrading Cloudera AI Workbenches:

Important: If you upgrade to version 1.5.5, Cloudera recommends to upgrade to 1.5.5 SP11 or higher versions.

Current Machine Learning version	Target Machine Learning or Cloudera AI version
1.5.3	1.5.5 CHF1 1.5.5 SP1 (recommended)
1.5.4 1.5.4 SP2 1.5.4 SP1 1.5.4 CHF3 1.5.4 CHF1	1.5.5 CHF1 1.5.5 SP1 (recommended)

Backups for Cloudera AI Workbenches

Cloudera AI enables the efficient creation of machine learning projects, jobs, experiments, machine learning models, and applications within workbenches. The data and metadata of these artifacts are stored in different types of storage systems in on premises environments or in external NFS-backed workbenches outside of an on premises environment.

You can back up a Cloudera AI Workbench, and restore it at a later time. The backup preserves all files, models, applications and other assets within the workbench. However, for external NFS-based workbenches, files are not automatically backed up by Cloudera AI. All workbench backups are accessible through the Workbench Backup Catalog UI.

The Backup and Restore feature allows you to keep your machine learning artifacts safe by backing up all data, except files in external NFS-backed workbenches, to protect against potential disasters. If your Cloudera AI Workbench is backed up, this feature enables you to restore the saved data allowing you to recover your Cloudera AI artifacts exactly as they were at the time of the backup. Administrators can use the Backup and Restore feature to perform on-demand backups of Cloudera AI Workbenches. During the backup process, core services running in

the workbench are temporarily shut down to ensure the consistency of the backup data. To minimize disruptions, Cloudera recommends to schedule backups during off-peak hours.

The duration of the workbench backup process depends on the volume of data being copied. The backup procedure involves transferring data from both block volumes and internal NFS, with the time required to back up NFS typically being the most significant factor. Cloudera recommends to regularly back up Cloudera AI Workbenches to ensure data presercvation.

The time required to backup NFS largely depends on the volume of data, as well as the type and number of files. Based on the data size, you can configure a timeout value during the backup process. The status of ongoing and previous backups can be monitored trough the Cloudera AI Workbench UI and the Backup Catalog UI.

There is no restriction on the number of backups that can be created, and the backup snapshots are retained indefinitely within the underlying on premises cluster as long as the original workbench (from which this backup was taken from) is not deleted. Cloudera AI Workbench backup details are stored in the Workbench Backup Catalog UI within the Cloudera AI Control Plane. These entries can be listed, viewed, deleted or restored as needed.

Note: Deleting workbench backups from the UI is not supported.

Restoring a backup overwrites the existing Cloudera AI Workbench, from which this backup was taken, with the data from the backup. During this process, all the projects, jobs, applications, and other assets, that existed during the backup are automatically available in the new workbench. The restoration process involves overwriting the current workbench, followed by launching restore jobs to recreate storage volumes from the backup snapshots. The restore process takes longer than a standard workbench provisioning operation due to the additional steps involved in copying data from the backup to the new storage volumes. Restores are always performed as full-copy operations, which include the restoration of metadata files (such as the Sense database) and project files, as well as a complete restoration of data at the storage level. The restoration time is primarily influenced by the NFS restoration process, which typically takes at least as long as the original backup of the file system. Additionally, the restored workbench is always created using the latest version of the Cloudera AI software, which may differ from the version of the original workbench that was backed up.

Restoring the workbench creates a workbench based on the control plane version. The data will be restored to the state when the backup was created but the workbench version will be the one supported by control plane. Consequently, if the control plane is version 1.5.5, reverting to a previous version is not possible by design. Backups are exclusively intended for disaster recovery, enabling data restoration but not version rollback.

Note: Additionally, restoration cannot be performed across different environments or clusters.

Note: Cloudera AI Workbench Backup and Restore feature is available on both Cloudera Embedded Container Service and OCP. It can be accessed through the Cloudera AI UI and the CDP CLI.

Workbench backup and restore prerequisites

To backup and restore workbenches, check that the following prerequisites are satisfied.

The following prerequisites apply to Backup functionality.

- Backup is enabled only for workbenches that are successfully installed.
- All workloads (sessions, jobs, applications, models) shall be turned off manually by the user before taking backup. This will ensure consistency in the backup data.
- Core services running in the workbench are shut down during the backup process. So, during the backup process, the workbench will not be accessible to the customer.
- It is recommended that backups are taken during off-peak hours to minimize user impacts.
- Time to backup is proportional to the amount of data present in the workbench. So, give sufficient timeout when triggering backup.
- Backup is supported for both external and internal NFS-backed workbenches.

- Make sure enough disk space is available for taking workbench backup. Workbench backup generally takes an equivalent amount of storage space compared to the workbench itself.

The following prerequisites apply to Restore functionality.

- Workbench restore does not create a new workbench. It will instead replace the running workbench with an older backup.
- Restore process, overwrites the existing workbench with one of the older backups. This means that anything on the running workbench which is not backed up will get lost forever. So, make sure you really want to restore an older version of the workbench. If you want to save the current state before restore, you can do so by first taking a new backup and then proceeding with the restore.
- All workloads (sessions, jobs, applications, models) shall be turned off manually by the user before triggering restore.
- All workloads (sessions, jobs, applications, models) shall be turned on manually by the user after restore has completed.
- Time to restore is proportional to the amount of data present in the backup. In general, restoration will take at most 12 hours to complete.
- Always make sure that any ongoing backup for a workbench is completed (by looking at workbench status and backup event logs), before triggering restore for it.

Backing up a Cloudera AI Workbench

Backing up an Cloudera AI Workbench preserves all files, models, applications, and other assets in the workbench, although files in external NFS-backed workbenches are not backed up by Cloudera AI automatically.

Procedure

1. In the Workbenches UI, find the workbench to back up. The workbench must be in the Installation completed state, otherwise backup is disabled.
2. Enter the workbench, and manually stop all workloads (sessions, jobs, applications, and models).

   For external NFS backed workbenches, manually back up the configured external NFS data to another location. This manual backup of the NFS data will be used when this particular backup is restored in future. Ignore this step if the workbench is configured with internal NFS, as internal NFS data is backed up and restored automatically by Cloudera AI.
3. In the Actions menu for that workbench, select Backup Workbench.
4. In the Backup Workbench modal, enter a Backup Name to identify the workbench, and enter an appropriate timeout value.
5. Click Backup to start the process.

Results

The workbench shuts down, and the backup process begins. The workbench state changes to reflect the ongoing backup progress. If necessary, click Cancel to cancel the backup process. The backup process can take some time to complete, depending on the amount of data to copy.

Note: The default timeout is 12 hours. The estimated time to complete a backup (from the cloud provider) is now periodically added to the event logs.

What to do next

Monitoring event logs

You can monitor the progress of the backup process by checking the event logs. In the Actions menu for the workbench, click View Event Logs, and then on the Events & Logs tab, click View Event Logs again for the latest backup event.

When the backup process completes, the workbench enters the Installation completed state again.

If there were issues during backup, appropriate error messages will be displayed in the event logs. However the workbench will recover from failure and will be reverted back to the original state when backup was triggered.

Cleaning up and backing up the Cloudera AI Workbench database manually

With the help of the script you can manually back up and clean up the Cloudera AI Workbench database.

- Configure the Workbench kubeconfig by setting it as an environment variable for the script to access. Use the following command:

  ```
  export KUBECONFIG=<your kubeconfig>
  ```

  The script prompts you to specify a directory to store the backup file, define the retention time (entries older than this will be cleaned up), and optionally provide a Workbench name.

The result is an .sql file as a snapshot of the Cloudera AI database.

Note: This script is not intended for restoring or reverting your Cloudera AI database backup. To back up and restore your Workbench, including the database, please refer to: Restoring a Cloudera AI Workbench .

```
#!/bin/bash

            DB_NAME="sense"
            # Ask for database configuration
            read -p "Enter backup directory: " BACKUP_DIR
            while true; do
            read -p "Enter retention time (e.g., '30 days', '2 months'): "
 RETENTION_TIME

            # Validate format: number + space + interval unit (singular or
 plural)
            if [[ "$RETENTION_TIME" =~ ^[1-9][0-9]*\ (day|days|month|months|
year|years|hour|hours|minute|minutes|second|seconds)$ ]]; then
            break
            else
            echo "Invalid retention time format. Please use formats like
'30 days', '2 months', or '1 year'."
            fi
            done
            read -p "If on premises, enter workbench name: " INPUT_NAMESPACE
            K8S_NAMESPACE=${INPUT_NAMESPACE:-mlx}

            DB_POD="db-0"

            # Table, timestamp column, and optional extra condition
            TABLES=(
            "dashboards:created_at:id NOT IN (SELECT DISTINCT current_dashb
oard_id FROM applications WHERE applications.deleted_at IS NULL)"
            "dashboard_pods:created_at:status != 'running'"
            "model_deployments:created_at:stopped_at IS NOT NULL"
            "user_events:created_at:"
            )

            # Ensure backup directory exists
            if ! mkdir -p "$BACKUP_DIR"; then
            echo "Error: Failed to create or access backup directory '$BACKU
P_DIR'. Check your permissions." >&2
            exit 1
            fi
```

```
            echo "Backup directory verified: $BACKUP_DIR"

            # Create a timestamped backup file
            BACKUP_FILE="$BACKUP_DIR/${DB_NAME}_backup_$(date +"%Y%m%d%H%M%
S").dump"
            echo "Backup file path: $BACKUP_FILE"

            # Inside the pod: create the dump file
            kubectl exec -n "$K8S_NAMESPACE" -c db "$DB_POD"-- pg_dump -d "$
DB_NAME" -F c -f /tmp/db_backup.dump

            # Then copy it to local
            kubectl cp -c db "$K8S_NAMESPACE/$DB_POD:/tmp/db_backup.dump"
"$BACKUP_FILE"

            if [ $? -eq 0 ]; then
            echo "Backup created successfully: $BACKUP_FILE"
            else
            echo "Backup failed!" >&2
            exit 1
            fi

            # Prompt the user for confirmation before cleanup
            read -p "Are you sure you want to delete old entries from the
 database? (yes/no): " CONFIRM
            if [[ "$CONFIRM" != "yes" ]]; then
            echo "Cleanup aborted by user."
            exit 0
            fi

            # Clean up old entries from multiple tables
            for ENTRY in "${TABLES[@]}"; do
            IFS=":" read -r TABLE TIMESTAMP_COLUMN EXTRA_CONDITION <<< "$EN
TRY"

            # Build base DELETE query
            DELETE_QUERY="DELETE FROM $TABLE WHERE $TIMESTAMP_COLUMN < NOW()
 - INTERVAL '$RETENTION_TIME'"

            # Append extra condition if present
            if [ -n "$EXTRA_CONDITION" ]; then
            DELETE_QUERY+=" AND $EXTRA_CONDITION"
            fi

            echo "Executing cleanup query on $TABLE: $DELETE_QUERY"

            # Execute the cleanup query using kubectl exec
            kubectl exec -n "$K8S_NAMESPACE" -c db "$DB_POD" -- psql -d "$D
B_NAME" -c "$DELETE_QUERY"

            if [ $? -eq 0 ]; then
            echo "Old entries deleted successfully from $TABLE."
            else
            echo "Error: Failed to delete old entries from $TABLE. Cleanup p
rocess interrupted — not all tables were processed." >&2
            exit 1
            fi
            done
```

Restoring a Cloudera AI Workbench

Restoring a backup overwrites the existing Cloudera AI Workbench (from which the backup was taken from) and automatically imports the restored data. All of the projects, jobs, applications and so on in the original workbench are recreated in the new one.

About this task

Note: Restoring a workbench is a non-reversible operation. The restore process overwrites the existing workpace with older backup data. Any data in the running workbench that is not backed up will be lost. To save the current state, take a new backup before proceeding with the restore operation.

Procedure

1. In the Workbench Backups UI, find the workbench to restore. You can search for the workbench name or CRN. There can be multiple backups for a given workbench.
2. Enter the workbench, and manually stop all workloads (sessions, jobs, applications, and models).

   For external NFS backed workbenches, copy the manual backup of external NFS data (corresponding to this particular backup) to the configured external NFS export. Ignore this step if the workbench is configured with internal NFS, as internal NFS data is backed up and restored automatically by Cloudera AI.
3. Look for the backup to restore, and click Restore. The restore process starts, and the workplace states changes to Creating Workbench.

Results

The restore process can take some time, depending on the amount of data to copy. When it is complete, you can find the restored workbench in the Workbenches UI.

Note: If there is an issue during the restore process, the event log will show the relevant error messages. In case of error, the workbench will not recover from the failure automatically and will not revert back to the original state prior to the restore operation.

What to do next

Monitoring event logs

You can monitor the progress of the backup process by checking the event logs. In the Actions menu for the workbench, click View Event Logs, and then on the Events & Logs tab, click View Event Logs again for the latest backup event.

When the backup process completes, the workbench enters the installation completed state again.

If there were issues during backup, appropriate error messages will be displayed in the event logs. However the workbench will recover from failure and will be reverted back to the original state when backup was triggered.