# Administer and monitor

## GenAI

NetApp
October 06, 2025

# Table of Contents

# Administer and monitor

## Manage the GenAI infrastructure

You can view details about your deployed GenAI RAG infrastructure or remove the chatbot infrastructure if you no longer need it.

### View information about the infrastructure

You can view information about the chatbot infrastructure.

**Steps**

1. Log in to Workload Factory using one of the console experiences.
2. In the AI workloads tile, select **Deploy & manage**.
3. Select the **Infrastructure** menu.
4. View information about the infrastructure, which includes details about the following components:
   - AWS settings
   - Infrastructure settings
   - The AI engine
   - The vector database

### Remove the infrastructure

If you no longer need the chatbot infrastructure that you deployed for one or more chatbots, you can remove it from Workload Factory.

> All chatbots that have been deployed on this infrastructure will be disabled and all chat history will be deleted.

This operation only removes the links to the AI infrastructure from Workload Factory; it does not remove all the components from AWS. You'll need to manually delete the following infrastructure components from AWS:

- The VM instance
- Private endpoints
- The volume on the FSx for ONTAP file system that contains the AI databases
- The IAM role
- The policy
- The security group

**Steps**

1. Log in to Workload Factory using one of the console experiences.
2. In the AI workloads tile, select **Deploy & manage**.
3. Select the **Infrastructure** menu.
4. Select ••• and select **Remove chatbot infrastructure**.

5. Confirm that you want to delete the infrastructure and select **Remove**.

**Result**

The chatbot infrastructure components are removed from Workload Factory.

# Manage GenAI knowledge bases

After you create a knowledge base, you can view the knowledge base details, modify the knowledge base, integrate additional data sources, or delete the knowledge base.

## View information about a knowledge base

You can view information about the settings for a knowledge base and the data source that are integrated.

**Steps**

1. Log in to Workload Factory using one of the console experiences.
2. In the AI workloads tile, select **Deploy & manage**.
3. Select the knowledge base that you want to view.

   If defined, the conversation starters that are currently being used display in the right pane.

4. To view knowledge base details, select ••• and select **Manage knowledge base**.

   This page displays the published status, embedding status of the data sources, embedding mode, the list of all embedded data sources, and more.

   The **Actions** menu enables you to manage the knowledge base if you want to make any changes.

## Edit a knowledge base

You can update a knowledge base by changing some settings, or you can add or remove data sources.

Each time you add, modify, or remove data sources from the knowledge base, you must sync the data source so that it is re-indexed to the knowledge base. Syncing is incremental, so Amazon Bedrock only processes the objects in your FSx for ONTAP volume that have been added, modified, or deleted since the last sync.

**Steps**

1. Log in to Workload Factory using one of the console experiences.
2. In the AI workloads tile, select **Deploy & manage**.
3. From the Knowledge bases inventory page, select the knowledge base that you want to update.
4. Select ••• and select **Manage knowledge base**.

   This page displays the published status, embedding status of the data sources, embedding mode, the list of all embedded data sources, and more.

5. Select the **Actions** menu and select **Edit knowledge base**.
6. In the Edit knowledge base page, you can change the knowledge base name, description, embedding model, chat model, feature enablement, choose whether conversation starters are created automatically or manually, and the snapshot policy used for the volume that contains the knowledge base.

If you use Manual mode for conversation starters, you can change conversation starters here as well.

> Every knowledge base scan, which includes embedding, costs. If data guardrails is enabled after a knowledge base has been created, then the knowledge base gets scanned again and incurs costs. Similarly, if you change chat models, GenAI will re-scan the associated data sources (incurring a cost).

7. Select **Save** after you have made your changes.

## Protect a knowledge base with snapshots

You can protect your knowledge base data by taking and restoring snapshots of your knowledge base volumes. You can restore from a snapshot to revert to the previous version of the knowledge base at any time.

Snapshots can be faster and more storage-efficient than backups, and enable you to protect each knowledge base using a different protection policy. Some of the scenarios where snapshots can be useful are:

- Accidental data loss or corruption
- Recovering from incorrect data being ingested into the knowledge base
- Testing different data sources or chunking strategies, and quickly reverting when the testing is complete

### Take a snapshot of a knowledge base volume

You can save the state of a knowledge base by taking a manual snapshot of the knowledge base volume.

#### Steps

1. Log in to Workload Factory using one of the console experiences.
2. In the AI workloads tile, select **Deploy & manage**.
3. From the Knowledge bases inventory page, select the knowledge base that you want to protect.
4. Select ••• and select **Manage knowledge base**.

   This page displays the published status, embedding status of the data sources, embedding mode, the list of all embedded data sources, and more.

5. Select the **Actions** menu and select **Snapshot > Create new snapshot**.
6. Optionally, select **Define snapshot name** and enter a custom name for the snapshot.

   Defining a custom name can help you better determine the contents of a snapshot if you need to restore it in the future.

7. Select **Create**.

   A snapshot of the knowledge base is created.

### Restore a snapshot of a knowledge base volume

You can restore a manual or scheduled snapshot of a knowledge base volume at any time.

You cannot restore a snapshot using the Generative AI workloads UI if the database stored on the volume is corrupt or has been deleted. As a workaround, you can restore the snapshot using the ONTAP CLI on the ONTAP cluster where the volume is hosted.

**Steps**

1. Log in to Workload Factory using one of the console experiences.
2. In the AI workloads tile, select **Deploy & manage**.
3. From the Knowledge bases inventory page, select the knowledge base that you want to restore.
4. Select ••• and select **Manage knowledge base**.

   This page displays the published status, embedding status of the data sources, embedding mode, the list of all embedded data sources, and more.

5. Select the **Actions** menu and select **Snapshot > Restore snapshot**.

   The snapshot selection dialog appears, where you can see a list of the snapshots that have been created for this knowledge base.

6. (Optional) Deselect the **Pause running and scheduled scans after restoring the snapshot** option if you want scheduled and currently running data source scans to continue after the snapshot is restored.

   This option is enabled by default to ensure that a scan doesn't happen while the knowledge base is in a partially restored state, or that a scan doesn't update a freshly restored knowledge base with older data.

7. Select the snapshot you want to restore from the list.
8. Select **Restore**.

### Clone a knowledge base

You can create a new knowledge base from a knowledge base snapshot. This is useful if the original knowledge base is corrupted or lost.

**Steps**

1. Log in to Workload Factory using one of the console experiences.
2. In the AI workloads tile, select **Deploy & manage**.
3. From the Knowledge bases inventory page, select the knowledge base that you want to restore.
4. Select ••• and select **Manage knowledge base**.

   This page displays the published status, embedding status of the data sources, embedding mode, the list of all embedded data sources, and more.

5. Select the **Actions** menu and select **Snapshot > Clone knowledge base**.

   The clone dialog appears.

6. Optionally, deselect the **Pause running and scheduled scans after cloning the snapshot** option if you want scheduled and currently running data source scans to continue after the snapshot is cloned.

   This option is enabled by default to ensure that a scan doesn't happen while the knowledge base is in a partially restored state, or that a scan doesn't update a freshly restored knowledge base with older data.

7. Select the snapshot you want to clone from the list.
8. Select **Continue**.
9. Enter a name for the new knowledge base.
10. Choose a filesystem SVM and volume name for the new knowledge base.
11. Select **Clone**.

## Add additional data sources to a knowledge base

You can embed additional data sources in your knowledge base to populate it with additional organization data.

**Steps**

1. Log in to Workload Factory using one of the console experiences.
2. In the AI workloads tile, select **Deploy & manage**.
3. From the Knowledge bases inventory page, select the knowledge base where you want to add the data source.
4. Select ••• and select **Add data source**.
5. Select the type of data source you want to add:
   - Add FSx for ONTAP file system (use files from an existing FSx for ONTAP volume)
   - Add file system (use files from a generic SMB or NFS share)

Unresolved directive in knowledge-base/manage-knowledgebase.adoc - include::_include/add-data-source-kb.adoc[]

**Result**

The data source is integrated into your knowledge base.

## Synchronize your data sources with a knowledge base

Data sources are synchronized with the associated knowledge base automatically once a day so that any data source changes are reflected in the chatbot. If you make changes to any of your data sources and you'd like to synchronize the data immediately, you can perform an on-demand synchronization.

Syncing is incremental, so Amazon Bedrock only processes the objects in your data sources that have been added, modified, or deleted since the last sync.

**Steps**

1. Log in to Workload Factory using one of the console experiences.
2. In the AI workloads tile, select **Deploy & manage**.
3. From the Knowledge bases inventory page, select the knowledge base that you want to synchronize.
4. Select ••• and select **Manage knowledge base**.
5. Select the **Actions** menu and select **Scan now**.

   You'll see a message that your data sources are being scanned, and a final message when the scan is complete.

**Result**

The knowledge base is synchronized with the attached data sources and any active chatbot will start using the newest information from your data sources.

### Pause or resume a scheduled synchronization

If you want to pause or resume the next synchronization (scan) of the data sources, you can do so at any time. You might need to pause the next scheduled synchronization if you are going to make changes to a data source and don't want the synchronization happening during the change window.

**Steps**

1. Log in to Workload Factory using one of the console experiences.
2. In the AI workloads tile, select **Deploy & manage**.
3. From the Knowledge bases & Connectors menu, select the knowledge base for which you want to pause or resume scans.
4. Select ••• and select **Manage knowledge base**.
5. Select the **Actions** menu and select **Scan > Pause scheduled scan** or **Scan > Resume scheduled scan**.

   You'll see a message that the next scheduled scan has either been paused or resumed.

## Evaluate chat models before creating a knowledge base

You can evaluate the available foundational chat models before creating a knowledge base so you can see which model works best for your implementation. Since model support varies by AWS region, refer to this AWS documentation page to verify which models you can use in the regions where you plan to deploy your knowledge base.

> This functionality is available only when no knowledge bases have been created — when no knowledge bases exist in the Knowledge bases inventory page.

**Steps**

1. Log in to Workload Factory using one of the console experiences.
2. In the AI workloads tile, select **Deploy & manage**.
3. From the Knowledge bases inventory page, you'll see the option to select the chat model on the right side of the page for the Chatbot.
4. Select the chat model from the list and enter a set of questions in the prompt area to see how the chatbot responds.
5. Try multiple models to see which model is best for your implementation.

**Result**

Use that chat model when you create your knowledge base.

## Unpublish your knowledge base

After you've published your knowledge base so that it can be integrated with a chatbot application, you can unpublish it if you want to disable the chatbot application from accessing the knowledge base.

Unpublishing the knowledge base stops any chat applications from working. The unique API endpoint at which the knowledge base was accessible is disabled.

**Steps**

1. Log in to Workload Factory using one of the console experiences.
2. In the AI workloads tile, select **Deploy & manage**.
3. From the Knowledge bases inventory page, select the knowledge base that you want to unpublish.
4. Select ••• and select **Manage knowledge base**.

   This page displays the published status, embedding status of the data sources, embedding mode, and the list of all embedded data sources.

5. Select the **Actions** menu and select **Unpublish**.

**Result**

The knowledge base is disabled and is no longer accessible by a chatbot application.

## Delete a knowledge base

If you no longer need a knowledge base, you can delete it. When you delete a knowledge base, it is removed from Workload Factory and the volume that contains the knowledge base is deleted. Any applications or chatbots that are using the knowledge base will stop working. Deleting a knowledge base is not reversible.

When you delete a knowledge base, you should also disassociate the knowledge base from any agents it is associated with to fully delete all resources associated with the knowledge base.

**Steps**

1. Log in to Workload Factory using one of the console experiences.
2. In the AI workloads tile, select **Deploy & manage**.
3. From the Knowledge bases inventory page, select the knowledge base that you want to delete.
4. Select ••• and select **Manage knowledge base**.
5. Select the **Actions** menu and select **Delete knowledge base**.
6. In the Delete knowledge base dialog, confirm that you want to delete it and select **Delete**.

**Result**

The knowledge base is removed from Workload Factory and its associated volume is deleted.

# Manage Amazon Q Business connectors

After you create a connector for Amazon Q Business, you can view the connector details, modify the connector, integrate additional data sources, or delete the connector.

## View information about a connector

You can view information about the settings for a connector and the data sources that are integrated.

**Steps**

1. Log in to Workload Factory using one of the console experiences.
2. In the AI workloads tile, select **Deploy & manage**.
3. From the Knowledge bases & Connectors inventory page, select the connector that you want to view.

4. To view connector details, select ••• and select **Manage connector**.

   This page displays the published status, embedding status of the data sources, embedding mode, the list of all embedded data sources, and more.

   The **Actions** menu enables you to manage the connector if you want to make any changes.

## Edit a connector

You can update a connector by changing some settings, or you can add or remove data sources.

Each time you add, modify, or remove data sources from the connector, GenAI needs to send the data source information to Amazon Q Business so that it is re-indexed. Syncing is incremental, so Amazon Q Business only processes the objects in your FSx for ONTAP volume that have been added, modified, or deleted since the last sync.

**Steps**

1. Log in to Workload Factory using one of the console experiences.
2. In the AI workloads tile, select **Deploy & manage**.
3. From the Knowledge bases & Connectors inventory page, select the connector that you want to update.
4. Select ••• and select **Manage connector**.

   This page displays the published status, embedding status of the data sources, embedding mode, the list of all embedded data sources, and more.

5. Select the **Actions** menu and select **Edit connector**.
6. In the Edit connector page, you can change the connector name, description, embedding model, data guardrails enablement, and the snapshot policy used for the volume that contains the connector.

   Every data source scan, which includes embedding, incurs a cost. If you enable data guardrails after a connector has been created, then the data source gets scanned again and incurs costs.

7. Select **Save** after you have made changes.

## Add additional data sources to a connector

You can embed additional data sources in your connector to populate it with additional organization data.

**Steps**

1. Log in to Workload Factory using one of the console experiences.
2. In the AI workloads tile, select **Deploy & manage**.
3. From the Knowledge bases & Connectors inventory page, select the connector where you want to add the data source.
4. Select ••• and select **Add data source**.
5. Select the type of data source you want to add:
   - Add FSx for ONTAP file system (use files from an existing FSx for ONTAP volume)
   - Add file system (use files from a generic SMB or NFS share)

Unresolved directive in connector/manage-connector.adoc - include::_include/add-data-source-kb.adoc[]

**Result**

The data source is integrated into your connector.

## Synchronize your data sources with a connector

Data sources are synchronized with the associated connector automatically once a day so that any data source changes are reflected in Amazon Q Business. If you make changes to any of your data sources and you'd like to synchronize (scan) the data immediately, you can perform an on-demand synchronization.

Syncing is incremental, so Amazon Q Business only processes the objects in your data sources that have been added, modified, or deleted since the last sync.

**Steps**

1. Log in to Workload Factory using one of the console experiences.
2. In the AI workloads tile, select **Deploy & manage**.
3. From the Knowledge bases & Connectors menu, select the connector that you want to synchronize.
4. Select ••• and select **Manage connector**.
5. Select the **Actions** menu and select **Scan now**.

   You'll see a message that your data sources are being scanned, and a final message when the scan is complete.

**Result**

The connector is synchronized with the attached data sources and Amazon Q Business will start using the newest information from your data sources.

### Pause or resume a scheduled synchronization

If you want to pause or resume the next synchronization (scan) of the data sources, you can do so at any time. You might need to pause the next scheduled synchronization if you are going to make changes to a data source and don't want the synchronization happening during the change window.

**Steps**

1. Log in to Workload Factory using one of the console experiences.
2. In the AI workloads tile, select **Deploy & manage**.
3. From the connector inventory page, select the connector for which you want to pause or resume scans.
4. Select ••• and select **Manage connector**.
5. Select the **Actions** menu and select **Scan > Pause scheduled scan** or **Scan > Resume scheduled scan**.

   You'll see a message that the next scheduled scan has either been paused or resumed.

## Delete a connector

If you no longer need a connector, you can delete it. When you delete a connector, it is removed from Workload Factory and the volume that contains the connector is deleted. Deleting a connector is not reversible.

When you delete a connector, you should also disassociate the connector from any agents it is associated with to fully delete all resources associated with the connector.

**Steps**

1. Log in to Workload Factory using one of the console experiences.
2. In the AI workloads tile, select **Deploy & manage**.
3. From the Knowledge bases & Connectors inventory page, select the connector that you want to delete.
4. Select ••• and select **Manage connector**.
5. Select the **Actions** menu and select **Delete connector**.
6. In the Delete connector dialog, confirm that you want to delete it and select **Delete**.

**Result**

The connector is removed from Workload Factory and its associated volume is deleted.

# Manage GenAI data sources

After you create a knowledge base or a connector using data sources on your FSx for ONTAP file system, you can view the data source details, update or change the data source contents, edit data source settings, or delete the data source.

## View information about a data source

You can view information about the contents of a data source and you can view its embedding status with the knowledge base or connector. Since data sources are associated with a knowledge base or connector, you'll need to choose the knowledge base or connector first before you can view the data source details.

**Steps**

1. Log in to Workload Factory using one of the console experiences.
2. In the AI workloads tile, select **Deploy & manage**.
3. Select the knowledge base or connector where the data source resides, and then select ••• and select **Manage knowledge base** or **Manage connector**.

   The bottom part of the page lists the associated data sources.

4. Expand each row by selecting the ⌄ to view detailed information about each data source, such as the FSx for ONTAP file system, the volume, and the path where the data source resides.

   It also lists the embedding information and whether that data source is currently embedded in the knowledge base or connector.

## Edit data source settings

You can edit information about a data source that you've integrated with a knowledge base or connector. Most of the information is fixed after you've added a data source, but you can make changes to the some of the configuration (such as chunking definition or permission awareness).

**Steps**

1. Log in to Workload Factory using one of the console experiences.

2. In the AI workloads tile, select **Deploy & manage**.
3. From the knowledge bases inventory page, select the knowledge base where the data source resides, and then select ••• and select **Manage knowledge base**.

   The bottom part of the page lists the data sources that are part of this knowledge base.
4. In the row of the data source that you want to edit, select ••• and select **Edit data source**.
5. In the Edit data source page, select ⌄ to expand the row for chunk definition.
6. Update the settings for the chunking strategy and configuration, and permission awareness (for SMB volumes) and select **Save**.

**Result**

The data source settings are updated and the AI system synchronizes the data source so that it is re-indexed to the knowledge base.

## Update the contents of an existing data source

You can change the contents of a data source at any time to add or update your organizational data. If this data source is being actively used in a knowledge base, you must sync the data source so that it is re-indexed to the knowledge base. Syncing is incremental, so Amazon Bedrock only processes the objects in your FSx for ONTAP volume that have been added, modified, or deleted since the last sync.

Data sources are synchronized with the knowledge base automatically once a day so that any data source changes are reflected in the chatbot. If you make changes to a data source and you'd like to synchronize the data immediately, you can perform an on-demand synchronization.

## Delete a data source

If you no longer need a data source to be part of your knowledge base, you can delete it.

**Steps**

1. Log in to Workload Factory using one of the console experiences.
2. In the AI workloads tile, select **Deploy & manage**.
3. From the knowledge base inventory page, select the knowledge base where the data source resides, and then select ••• and select **Manage knowledge base**.

   The bottom part of the page lists the data sources that are part of this knowledge base.
4. In the row of the data source that you want to delete, select ••• and select **Delete data source**.
5. In the Delete data source dialog, confirm that you want to delete it and select **Confirm**.

**Result**

The data source is removed from the knowledge base and the AI system removes the indexed information about this data source from the knowledge base. Any information from that data source will no longer be available to chatbots that are using the knowledge base.

# Monitor workload operations with Tracker in NetApp Workload Factory

Monitor and track the execution of workload operations and monitor task progress with Tracker in NetApp Workload Factory.

**About this task**

NetApp Workload Factory provides Tracker, a monitoring feature, so you can monitor and track the progress and status of workload operations, review details for operation tasks and subtasks, and diagnose any issues or failures.

Several actions are available in Tracker. You can filter jobs by time frame (last 24 hours, 7 days, 14 days, or 30 days), workload, status, and user; find jobs using the search function; and download the jobs table as a CSV file. You can refresh Tracker at any time, and quickly retry a failed operation or edit parameters for a failed operation and try the operation again.

Tracker supports two levels of monitoring depending on the operation. Each task, such as file system deployment, displays the task description, status, start time, task duration, user, region, proxy resource, task ID, and all related sub tasks. You can view API responses to understand what happened during the operation.

**Tracker task levels with examples**

- Level 1 (task): Tracks file system deployment.
- Level 2 (sub task): Tracks the sub tasks related to the file system deployment.

**Operation status**

Operation status in Tracker is as follows *in progress*, *success*, and *failed*.

**Operation frequency**

Operation frequency is based on the job type and the job schedule.

**Events retention**

Events are retained in the user interface for 30 days.

## Track and monitor operations

Track and monitor operations in the Workload Factory console with Tracker.

**Steps**

1. Log in using one of the console experiences.
2. From the workload menu, select **Administration** and then select **Tracker**.
3. In Tracker, use the filters or search to narrow job results. You can also download a jobs report.

## View API request

View the API request in the Codebox for a task in Tracker.

**Steps**

1. In Tracker, select a task.
2. Select the actions menu and then select **View API request**.

## Retry a failed operation

Retry a failed operation in Tracker. You can also copy the error message of a failed operation.

You can retry a failed operation up to 10 times.

**Steps**

1. In Tracker, select a failed operation.
2. Select the actions menu and then select **Retry**.

**Result**

The operation is re-initiated.

## Edit and retry a failed operation

Edit the parameters of the failed operation and retry the operation outside Tracker.

**Steps**

1. In Tracker, select a failed operation.
2. Select the actions menu and then select **Edit and retry**.

   You are redirected to the operation page where you can edit the parameters and retry the operation.

**Result**

The operation is re-initiated. Go to Tracker to view the status of the operation.

## Copyright information



## Trademark information