Cloudera Data Flow

Slack to Milvus

Date published: 2021-04-06
Date modified: 2025-09-30

https://docs.cloudera.com/

Legal Notice

Contents

ReadyFlow overview: Slack to Milvus [Technical Preview]

You can use the Slack to Milvus [Technical Preview] ReadyFlow to consume messages from a Slack channel, vectorize them using a HuggingFace model and write the results to Milvus.

This ReadyFlow reads messages from a Slack channel, vectorizes the data using a HuggingFace embedding model, and stores the results in Milvus vector DB. The default HuggingFace model is 'all-MiniLM-L12-v2'. A Slack bot access token and a Milvus access token are required to run this flow. Define a KPI on the failure_WriteToMilvus connection to monitor failed write operations.

Note: This ReadyFlow is considered Technical Preview and is not designed for production use. The flow uses Python processors and must be run in NiFi 2.x. To ensure correct functionality, make sure your Milvus collection has been created with the same dimension as the 'all-MiniLM-L12-v2' model (384).

The ReadyFlow handles the Slack text messages only, it does not handle attachments in the message.

Slack to Milvus [Technical Preview] ReadyFlow details	
Source	Slack
Source Format	Slack
Destination	Milvus
Destination Format	Vector DB

Prerequisites

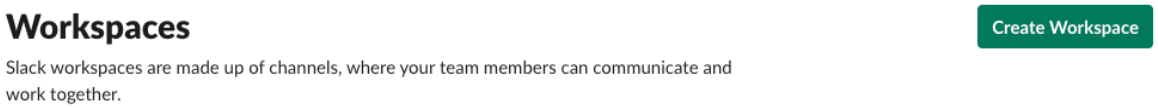
Learn how to collect the information you need to deploy the Slack to Milvus [Technical Preview] ReadyFlow, and meet other prerequisites.

For your data ingest source

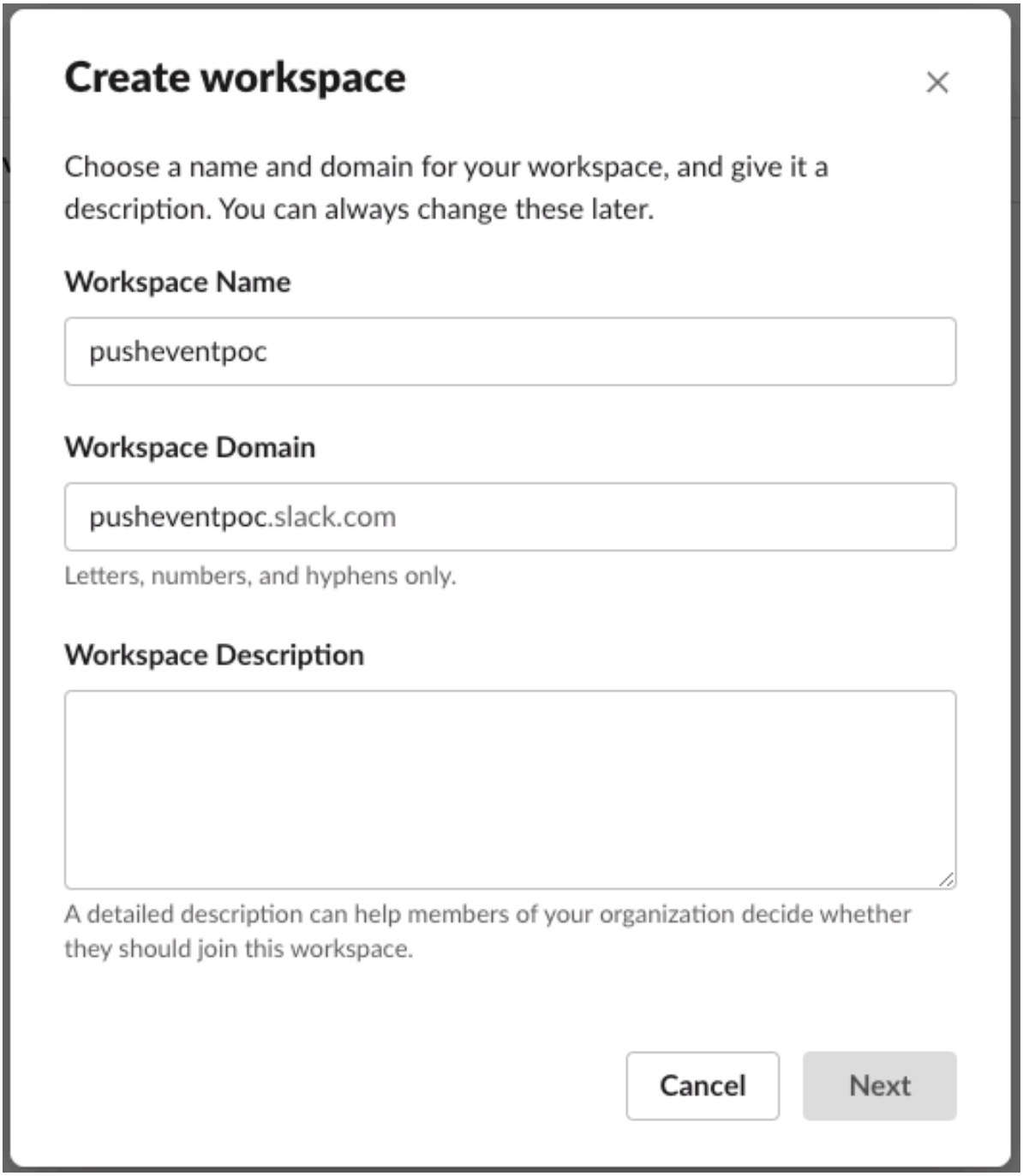
- You have created a Slack workspace and channel.
  1. In Slack Workspace Directory click Manage Organization.

  2. Click Create Workspace.

  3. Provide Workspace Name and Workspace Domain. You may leave Workspace Description empty.



- You have your Slack User or Bot Access Token.
- You have configured the necessary scopes for your token access.

For more information on getting started with Slack Enterprise Grid sandboxes, see the Slack documentation.

For Cloudera Data Flow

- You have enabled Cloudera Data Flow for an environment.

  For information on how to enable Cloudera Data Flow for an environment, see Enabling Cloudera Data Flow for an Environment.
- You have created a Machine User to use as the Cloudera Workload User.

- You have given the Cloudera Workload User the EnvironmentUser role.

  1. From the Management Console, go to the environment for which Cloudera Data Flow is enabled.
  2. From the Actions drop down, click Manage Access.
  3. Identify the user you want to use as a Workload User.

     Note:

     The Cloudera Workload User can be a machine user or your own user name. It is best practice to create a dedicated Machine user for this.
  4. Give that user EnvironmentUser role.
- You have synchronized your user to the Cloudera on cloud environment that you enabled for Cloudera Data Flow.

  For information on how to synchronize your user to FreeIPA, see Performing User Sync.
- You have granted your Cloudera user the DFCatalogAdmin and DFFlowAdmin roles to enable your user to add the ReadyFlow to the Catalog and deploy the flow definition.

  1. Give a user permission to add the ReadyFlow to the Catalog.
     a. From the Management Console, click User Management.
     b. Enter the name of the user or group you wish to authorize in the Search field.
     c. Select the user or group from the list that displays.
     d. Click Roles Update Roles .
     e. From Update Roles, select DFCatalogAdmin and click Update.

     Note: If the ReadyFlow is already in the Catalog, then you can give your user just the DFCatalogViewer role.
  2. Give your user or group permission to deploy flow definitions.
     a. From the Management Console, click Environments to display the Environment List page.
     b. Select the environment to which you want your user or group to deploy flow definitions.
     c. Click Actions Manage Access to display the Environment Access page.
     d. Enter the name of your user or group you wish to authorize in the Search field.
     e. Select your user or group and click Update Roles.
     f. Select DFFlowAdmin from the list of roles.
     g. Click Update Roles.
  3. Give your user or group access to the Project where the ReadyFlow will be deployed.
     a. Go to Data Flow Projects .
     b. Select the project where you want to manage access rights and click More Manage Access .
  4. Start typing the name of the user or group you want to add and select them from the list.
  5. Select the Resource Roles you want to grant.
  6. Click Update Roles.
  7. Click Synchronize Users.

For your data ingest target

- Your Milvus version is 2.4.4.
- You have the Milvus access token.
- You have the name of the destination Milvus collection.
- You have the URI of the destination Milvus instance.

- Your Milvus collection schema has the following field names and field types:
  - id *(INT64)* You must enable Auto ID on this field.
  - text_embedding *(FLOAT_VECTOR)*
  - text *(VARCHAR)*
  - author *(VARCHAR)*
  - channel *(VARCHAR)*
- The ‘text_embedding’ field in your destination Milvus collection is configured with the same dimensions as the default HuggingFace 'all-MiniLM-L12-v2' model (384).
- The text, author, and channel VARCHAR fields have been configured with Max Length values large enough to accommodate their respective values in your PDFs.

Related Concepts

Required parameters

Required parameters

When deploying the Slack to Milvus [Technical Preview] ReadyFlow, you have to provide the following parameters. Use the information you collected in *Prerequisites*.

Table 1: Slack to Milvus [Technical Preview] ReadyFlow configuration parameters

Parameter name	Description
HuggingFace Embedding Model	Specify the HuggingFace model name to use for embedding the data. The default model is 'all-MiniLM-L12-v2'.
Milvus Access Token	Specify the access token for the destination Milvus instance.
Milvus Collection Name	Specify the name of the destination Milvus collection.
Milvus Instance URI	Specify the URI of the destination Milvus instance.
Slack Channel ID	Specify the Slack channel ID.
Slack User or Bot Access Token	Specify the Slack user or bot access token.

Related Concepts

Prerequisites

Related Information

Deploying a ReadyFlow