

Slack to Pinecone

Date published: 2021-04-06

Date modified: 2024-06-03

CLOUDERA

Legal Notice

© Cloudera Inc. 2024. All rights reserved.

The documentation is and contains Cloudera proprietary information protected by copyright and other intellectual property rights. No license under copyright or any other intellectual property right is granted herein.

Unless otherwise noted, scripts and sample code are licensed under the Apache License, Version 2.0.

Copyright information for Cloudera software may be found within the documentation accompanying each component in a particular release.

Cloudera software includes software from various open source or other third party projects, and may be released under the Apache Software License 2.0 (“ASLv2”), the Affero General Public License version 3 (AGPLv3), or other license terms. Other software included may be released under the terms of alternative open source licenses. Please review the license and notice files accompanying the software for additional licensing information.

Please visit the Cloudera software product page for more information on Cloudera software. For more information on Cloudera support services, please visit either the Support or Sales page. Feel free to contact us directly to discuss your specific needs.

Cloudera reserves the right to change any products at any time, and without notice. Cloudera assumes no responsibility nor liability arising from the use of products, except as expressly agreed to in writing by Cloudera.

Cloudera, Cloudera Altus, HUE, Impala, Cloudera Impala, and other Cloudera marks are registered or unregistered trademarks in the United States and other countries. All other trademarks are the property of their respective owners.

Disclaimer: EXCEPT AS EXPRESSLY PROVIDED IN A WRITTEN AGREEMENT WITH CLOUDERA, CLOUDERA DOES NOT MAKE NOR GIVE ANY REPRESENTATION, WARRANTY, NOR COVENANT OF ANY KIND, WHETHER EXPRESS OR IMPLIED, IN CONNECTION WITH CLOUDERA TECHNOLOGY OR RELATED SUPPORT PROVIDED IN CONNECTION THEREWITH. CLOUDERA DOES NOT WARRANT THAT CLOUDERA PRODUCTS NOR SOFTWARE WILL OPERATE UNINTERRUPTED NOR THAT IT WILL BE FREE FROM DEFECTS NOR ERRORS, THAT IT WILL PROTECT YOUR DATA FROM LOSS, CORRUPTION NOR UNAVAILABILITY, NOR THAT IT WILL MEET ALL OF CUSTOMER’S BUSINESS REQUIREMENTS. WITHOUT LIMITING THE FOREGOING, AND TO THE MAXIMUM EXTENT PERMITTED BY APPLICABLE LAW, CLOUDERA EXPRESSLY DISCLAIMS ANY AND ALL IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO IMPLIED WARRANTIES OF MERCHANTABILITY, QUALITY, NON-INFRINGEMENT, TITLE, AND FITNESS FOR A PARTICULAR PURPOSE AND ANY REPRESENTATION, WARRANTY, OR COVENANT BASED ON COURSE OF DEALING OR USAGE IN TRADE.

Contents

ReadyFlow overview: Slack to Pinecone.....	4
Prerequisites.....	4
List of required configuration parameters for the Slack to Pinecone ReadyFlow.....	8

ReadyFlow overview: Slack to Pinecone

You can use the Slack to Pinecone ReadyFlow to consume messages from a Slack channel, vectorize them using an OpenAI model and write the results to Pinecone.

This ReadyFlow reads messages from a Slack channel, vectorizes the data using an OpenAI embedding model, and stores the results in Pinecone vector DB. The default OpenAI model is 'text-embedding-ada-002'. A Slack bot access token, an OpenAI API key and a Pinecone API key are required to run this flow. Define a KPI on the failure_WriteToPinecone connection to monitor failed write operations.



Note: This ReadyFlow uses Python processors and must be run in NiFi 2.x. To ensure correct functionality, make sure your Pinecone index has been created with the same dimension as the 'text-embedding-ada-002' model (1536). The ReadyFlow handles the Slack text messages only, it does not handle attachments in the message.

Slack to Pinecone ReadyFlow details	
Source	Slack
Source Format	Slack
Destination	Pinecone
Destination Format	Vector DB

Prerequisites

Learn how to collect the information you need to deploy the Slack to Pinecone 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.

Workspaces

Slack workspaces are made up of channels, where your team members can communicate and work together.

Create Workspace

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

Create workspace ✕

Choose a name and domain for your workspace, and give it a description. You can always change these later.

Workspace Name

Workspace Domain

Letters, numbers, and hyphens only.

Workspace Description

A detailed description can help members of your organization decide whether they should join this workspace.

- 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 DataFlow

- You have enabled DataFlow for an environment.
For information on how to enable DataFlow for an environment, see [Enabling DataFlow for an Environment](#).
- You have created a Machine User to use as the CDP Workload User.

- You have given the CDP Workload User the EnvironmentUser role.
 1. From the Management Console, go to the environment for which DataFlow 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 CDP 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 CDP Public Cloud environment that you enabled for DataFlow.

For information on how to synchronize your user to FreeIPA, see [Performing User Sync](#).

- You have granted your CDP 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 DataFlow 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

- You have your OpenAI API key.
- You have your Pinecone environment and index into which you want to ingest data.
- Your Pinecone index was configured with the 'text-embedding-ada-002 model'.
- You have your Pinecone API key.

Related Concepts

[List of required configuration parameters for the Slack to Pinecone ReadyFlow](#)

List of required configuration parameters for the Slack to Pinecone ReadyFlow

When deploying the Slack to Pinecone ReadyFlow, you have to provide the following parameters. Use the information you collected in *Prerequisites*.

Table 1: Slack to Pinecone ReadyFlow configuration parameters

Parameter name	Description
OpenAI API Key	Specify the API key used to authenticate to OpenAI.
Pinecone API Key	Specify the API key used to authenticate to Pinecone.
Pinecone Environment	Specify the name of the Pinecone environment.
Pinecone Index	Specify the name of the Pinecone index. To ensure correct functionality of this ReadyFlow, make sure your Pinecone index has been created with the same dimension as the 'text-embedding-ada-002' model (1536).
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](#)