

Cloudera Semantic Search user guide for Cloudera on cloud environments (Preview)

Date published: Jul 1, 2026

Date modified: Aug 14, 2026

This document has been released as part of a technical preview for features described herein. Technical preview components are provided as a convenience to our customers for their evaluation and trial usage. These components are provided 'as is' without warranty or support. Further, Cloudera assumes no liability for the usage of technical preview components, which should be used by customers at their own risk.

Legal Notice

© Cloudera Inc. 2026. 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.

This document has been released as part of a technical preview for features described herein. Technical preview components are provided as a convenience to our customers for their evaluation and trial usage. These components are provided 'as is' without warranty or support. Further, Cloudera assumes no liability for the usage of technical preview components, which should be used by customers at their own risk.

Contents

Legal Notice	2
Contents	3
Cloudera Semantic Search Overview	4
Background information	4
System Requirements	7
Hardware requirements	7
Cluster requirements	7
Software requirements	8
Getting Started	10
Preparation for Cloudera Semantic Search cluster deployment	10
Prerequisites	10
Deploying a Cloudera Semantic Search cluster on a Cloudera on cloud environment	10
Validating the Cloudera Semantic Search deployment	11
Managing the Cloudera Semantic Search cluster	13
Limitations	14
FAQs and Troubleshooting	15
How to get host details	15
The OpenSearch Dashboard UI is not accessible through Knox	16
How to run the commands using the OpenSearch dashboard	16
How to add Machine Learning, Ingest, or Coordinator roles to the Cloudera Semantic Search cluster	22
Only one OpenSearch service definition available in Ranger for multiple Cloudera Semantic Search (Data Hub) clusters on a single environment	24
Workload user is unable to request the cluster status	24
Appendix	26

This document has been released as part of a technical preview for features described herein. Technical preview components are provided as a convenience to our customers for their evaluation and trial usage. These components are provided 'as is' without warranty or support. Further, Cloudera assumes no liability for the usage of technical preview components, which should be used by customers at their own risk.

Note:

This release of the Cloudera Semantic Search service for Cloudera on cloud deployments is currently available as a technical preview. Technical preview releases are considered developmental and are not intended for use in production environments.

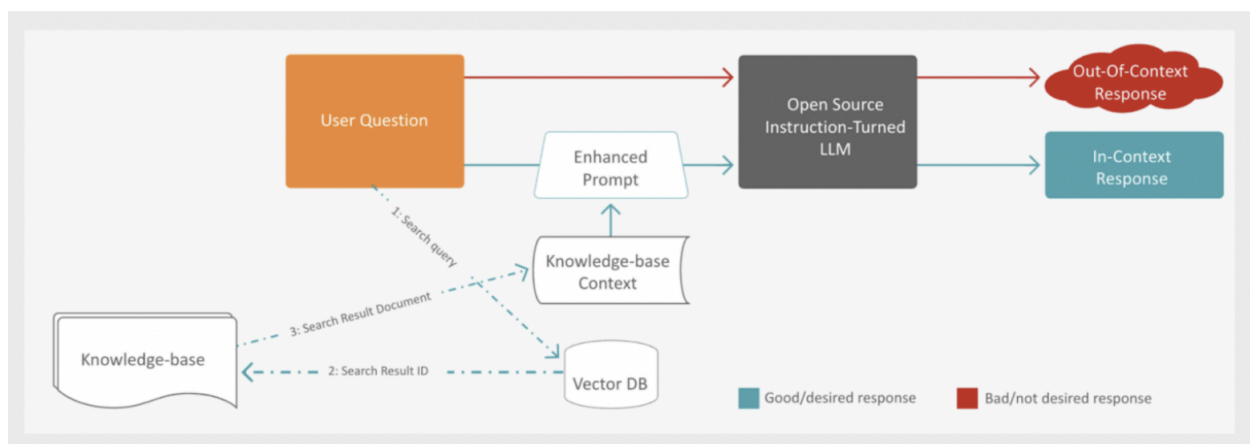
Cloudera Semantic Search Overview

Background information

Cloudera Semantic Search is the strategic evolution of our text search solution into a highly scalable, hybrid multimodal Vector Database and Semantic retrieval platform powered by OpenSearch. It acts as the secure "memory" and retrieval foundation for the entire Cloudera portfolio, transforming how enterprises power Retrieval-Augmented Generation (RAG) and autonomous AI agents, along with anomaly detection for time-series data. This functionality supports indexing, K-NN search, and Neural search. The underlying approach leverages an embedding model to generate vector representations (embeddings) of documents or knowledge sources. These vectors are stored and indexed, enabling searching and retrieval to find documents semantically similar to an input query. Furthermore, Neural search extends K-NN search by supporting various sorts of searches, such as semantic, multimodal, and hybrid search, using pre-trained, custom, or external foundation models.

Workflow

The following diagram illustrates the Retrieval Augmented Generation (RAG) workflow:



This document has been released as part of a technical preview for features described herein. Technical preview components are provided as a convenience to our customers for their evaluation and trial usage. These components are provided 'as is' without warranty or support. Further, Cloudera assumes no liability for the usage of technical preview components, which should be used by customers at their own risk.

The following typical Retrieval Augmented Generation workflow is valid for a search functionality:

1. Indexing of documents

- An embedding model is selected to convert structured and unstructured data into k-NN vectors for indexing in the Cloudera Semantic Search service.

Note:

The term "document" refers to a data unit and does not imply only text. The document could be either structured or unstructured data.

- The embedding model generates vector representations (embeddings) of the document collection.
- The vectors are stored and indexed for fast search and retrieval.

2. Search and retrieval

- A user enters an input query.
- The same embedding model used for indexing the documents generates the embedding for the input query.
- This embedding, which is a vector, is passed to the semantic search engine.
- The semantic search engine searches across the stored vector representations of documents and identifies the top-k, which is the top 1 or more documents that are semantically closest to the input embedding.
- The search engine returns the resultset, which can then be used to retrieve the source documents.

For more information, see [Deploying an LLM ChatBot Augmented with Enterprise Data](#).

Use cases

Semantic search is a crucial technology in generative artificial intelligence (GenAI) use cases. Using RAG, enterprises can enhance the functionalities of large language models (LLMs), such as ChatGPT, with enterprise-specific knowledge. Support for an RAG architecture generally involves retrieving documents that are semantically similar to a given input. For example, the Cloudera Large Language Model (LLM) Applied Machine Learning Prototypes (AMPs), augmented with enterprise data, leverage semantic search and are built around an RAG (Retrieval Augmented Generation) architecture.

Semantic search use cases can provide more than use cases built around the RAG architecture and ChatBots, including recommendation engines, sentiment analysis, and various automated categorization tasks, such as customer feedback, product reviews, and news articles.

This document has been released as part of a technical preview for features described herein. Technical preview components are provided as a convenience to our customers for their evaluation and trial usage. These components are provided 'as is' without warranty or support. Further, Cloudera assumes no liability for the usage of technical preview components, which should be used by customers at their own risk.

Having the semantic search engine and the underlying vector store capability as part of Cloudera helps maintain your data gravity within Cloudera. This means that you can leverage semantic search against the entire data estate, including all knowledge sources, such as websites, documentation, customer support systems, and internal data systems, such as data lake houses, and Slack channels of your organization, without having to transfer this data or its vectorized representation outside the organization. This approach ensures that the security and authorization levels of the source data remain intact.

Cloudera Semantic Search includes vector search functionality, which is exposed as part of the Cloudera service, including the following functionalities:

- Indexing
Indexes are the mechanism for storing and searching vector data.
- K-NN search
- Neural search.
Neural search builds on K-NN search functionality by allowing users to use pre-trained models, upload custom models, or connect to foundation models hosted on external platforms. Various sorts of searches are supported, such as semantic, multimodal, and hybrid search.

This document has been released as part of a technical preview for features described herein. Technical preview components are provided as a convenience to our customers for their evaluation and trial usage. These components are provided 'as is' without warranty or support. Further, Cloudera assumes no liability for the usage of technical preview components, which should be used by customers at their own risk.

System Requirements

Refer to the following topics to learn more about the hardware, operating system, and database requirements to install and deploy Cloudera Semantic Search in your Cloudera on cloud environments.

Hardware requirements

The recommended way to use Cloudera Semantic Search on cloud is to deploy a Data Hub cluster using the default cluster configurations. This approach ensures that hardware is pre-selected, eliminating the need for manual hardware planning.

Cluster requirements

Learn about the various cluster setup requirements to install and deploy Cloudera Semantic Search in your Cloudera on cloud environments.

AWS

Node Type	Instance Type	Storage	Node Count
Gateway	m5.2xlarge	Standard	1
Master	m5.2xlarge	gp3	3
Worker	m5.2xlarge	gp3	3
Ingest	c7i.2xlarge	gp3	0
Coordinator	r7i.2xlarge	gp3	0
Compute	r7i.2xlarge	gp3	0

Azure

Node Type	Instance Type	Storage	Node Count
Gateway	Standard_D8s_v3	StandardSSD_LRS	1
Master	Standard_D8s_v3	Premium_LRS	3
Worker	Standard_D8s_v3	Premium_LRS	3
Ingest	Standard_F8s_v2	Premium_LRS	0

This document has been released as part of a technical preview for features described herein. Technical preview components are provided as a convenience to our customers for their evaluation and trial usage. These components are provided 'as is' without warranty or support. Further, Cloudera assumes no liability for the usage of technical preview components, which should be used by customers at their own risk.

Coordinator	Standard_E8s_v3	Premium_LRS	0
Compute	Standard_E8s_v3	Premium_LRS	0

Software requirements

Learn about the operating system, database, and third-party software requirements for installing and deploying Cloudera Semantic Search in your Cloudera on cloud environments.

Table 2.2 Software configurations

	Supported versions	Note
Cloudera Runtime	Cloudera on cloud 7.3.2.100	Currently, it supports only the 7.3.2.100 version.
Cloudera Manager	7.13.2.100	
Operating System	RHEL 9	Supported on the operating system with the same minor versions for the Cloudera installation, with which the Cloudera Semantic Search feature is installed.
JDK	Java 21	Support for the Cloudera install, with which the Cloudera Semantic Search feature is installed.
Search engines	• Lucene: The native search library, offering an HNSW implementation with efficient filtering capabilities. • NMSLIB (Non-Metric Space Library): A legacy	Currently, Cloudera Semantic Search does not support the Faiss (Facebook AI Similarity Search) engine.

This document has been released as part of a technical preview for features described herein. Technical preview components are provided as a convenience to our customers for their evaluation and trial usage. These components are provided 'as is' without warranty or support. Further, Cloudera assumes no liability for the usage of technical preview components, which should be used by customers at their own risk.

	implementation of HNSW (now deprecated).	
--	--	--

This document has been released as part of a technical preview for features described herein. Technical preview components are provided as a convenience to our customers for their evaluation and trial usage. These components are provided 'as is' without warranty or support. Further, Cloudera assumes no liability for the usage of technical preview components, which should be used by customers at their own risk.

Getting Started

Note:

This release of the Cloudera Semantic Search service for Cloudera on cloud deployments is currently available as a technical preview. Technical preview releases are considered developmental and are not intended for use in production environments.

Preparation for Cloudera Semantic Search cluster deployment

Learn about the instructions on how to deploy a Cloudera Semantic Search cluster on your Cloudera on cloud environments.

Prerequisites

Before deploying Cloudera Semantic Search in your Cloudera on cloud environments, you must meet the following prerequisites.

- Verify that an active cloud provider account is established with Amazon Web Services (AWS) and Microsoft Azure.
- Ensure that a designated environment has been successfully registered within the Cloudera platform.
- A valid set of cloud credentials is required to facilitate a secure connection to the specified cloud provider environment.

Deploying a Cloudera Semantic Search cluster on a Cloudera on cloud environment

Learn how to deploy a Cloudera Semantic Search cluster on your Cloudera on cloud environment.

Steps

1. Create a cluster definition on your Cloudera on cloud environment, for example AWS or Azure.
 - a. To create a cluster definition on AWS, see [Creating a cluster from a definition on AWS](#).

This document has been released as part of a technical preview for features described herein. Technical preview components are provided as a convenience to our customers for their evaluation and trial usage. These components are provided 'as is' without warranty or support. Further, Cloudera assumes no liability for the usage of technical preview components, which should be used by customers at their own risk.

Note: From the “Cluster Definition” drop-down list, use the following cluster definition:

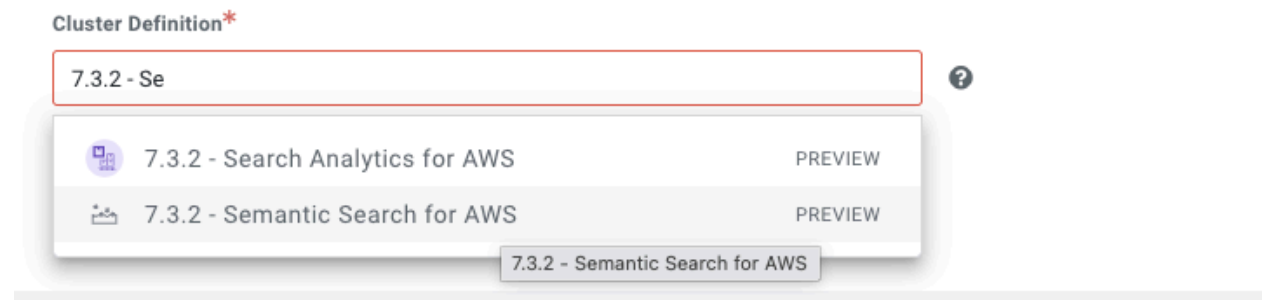
7.3.2 - Semantic Search for <Cloud Provider>

For example,

7.3.2 - Semantic Search for AWS

Services

Select the Cluster Definition option to create your cluster quickly by using one of the prescriptive cluster definitions.



General Settings

- b. To create a cluster definition on Azure, see [Creating a cluster from a definition on Azure](#).

Post requisites

1. Enable user authentication on the Cloudera Semantic Search cluster. For more information, see [User Authentication on Cloudera Semantic Search Cluster](#) document, which covers Kerberos, basic, and certificate-based authentication along with creating and managing Admin certificates.
2. Enable the Ranger plugin for Cloudera Semantic Search cluster, see [The Ranger plugin support for the Cloudera Semantic Search service](#) document.
3. Ensure that you perform the workaround mentioned in [Workload user is unable to request the cluster status](#) section.

Validating the Cloudera Semantic Search deployment

Learn how to validate the Cloudera Semantic Search deployment on your Cloudera on cloud environment.

This document has been released as part of a technical preview for features described herein. Technical preview components are provided as a convenience to our customers for their evaluation and trial usage. These components are provided 'as is' without warranty or support. Further, Cloudera assumes no liability for the usage of technical preview components, which should be used by customers at their own risk.

About this task

The default Cloudera Semantic Search administrator user is `opensearch`, when Kerberos is enabled. However, if you enable basic authentication, the default administrator user is `admin`.

Before you begin

- The Cloudera Semantic Search cluster must be deployed.

Steps

1. Log in to one of the Master nodes.
2. Use the following “curl” command to check whether the Cloudera Semantic Search is installed correctly for the respective nodes.
 - a. When the basic authentication is enabled.

Note: Copy the

`/var/lib/cloudera-scm-agent/agent-cert/cm-auto-global_cacerts.pem` file to the working directory.

None

```
curl https://<Master_Host_Name>:<Master_Port>/_cat/nodes?v --cacert  
cm-auto-global_cacerts.pem -u admin:admin
```

- b. When the Kerberos authentication is enabled.

Note: Copy the

`/var/run/cloudera-scm-agent/process/<master-process>/opensearch.keytab` file to the working directory.

None

```
kinit -kt opensearch.keytab opensearch/<Master_Host_Name>  
  
curl --cacert cm-auto-global_cacerts.pem --negotiate -u :  
https://<Master_Host_Name>:<Master_Port>/_cat/nodes?v
```

Expected output:

The output contains basic host information and statistics such as IP address, CPU load, and so on.

This document has been released as part of a technical preview for features described herein. Technical preview components are provided as a convenience to our customers for their evaluation and trial usage. These components are provided ‘as is’ without warranty or support. Further, Cloudera assumes no liability for the usage of technical preview components, which should be used by customers at their own risk.

3. Use the Semantic Search dashboard to verify the installation. For more information, see [The OpenSearch Dashboard UI is not accessible through Knox](#) in the FAQs section.

Managing the Cloudera Semantic Search cluster

Learn how to manage a Cloudera Semantic Search cluster from your Cloudera on cloud environment.

Managing the Cloudera Semantic Search cluster is performed through the Cloudera Data Hub platform. Because the cluster operates as a Data Hub deployment, standard cluster lifecycle management operations apply directly.

- **Stopping and Starting:** You can stop the cluster to manage compute costs or perform maintenance, and restart it when needed.
- **Deleting:** Use this operation to permanently remove the cluster when it is no longer required.
- **Retrying:** In the event of deployment or operation failures, you can retry cluster operations to recover the service state.

For comprehensive procedures and specific steps, please refer to the official [Managing Data Hub clusters documentation](#).

This document has been released as part of a technical preview for features described herein. Technical preview components are provided as a convenience to our customers for their evaluation and trial usage. These components are provided 'as is' without warranty or support. Further, Cloudera assumes no liability for the usage of technical preview components, which should be used by customers at their own risk.

Limitations

This topic details the known limitations and constraints of Cloudera Semantic Search.

- When multiple Semantic Search clusters exist on a single environment, only one OpenSearch service definition is available on the Ranger Admin UI.
- Security guarantees are not provided.
- ARM architecture is not supported.
- Knox does not support proxying OpenSearch Dashboards.
- Neural Search model registration requires internet access to download PyTorch updates from the OpenSearch model repository. You can download these updates manually if you prefer not to connect the cluster to the internet. This feature enables Private AI, which ensures that your data remains secure within your organization.
- Cloudera Runtime version 7.3.2.100 is the only supported version.

This document has been released as part of a technical preview for features described herein. Technical preview components are provided as a convenience to our customers for their evaluation and trial usage. These components are provided 'as is' without warranty or support. Further, Cloudera assumes no liability for the usage of technical preview components, which should be used by customers at their own risk.

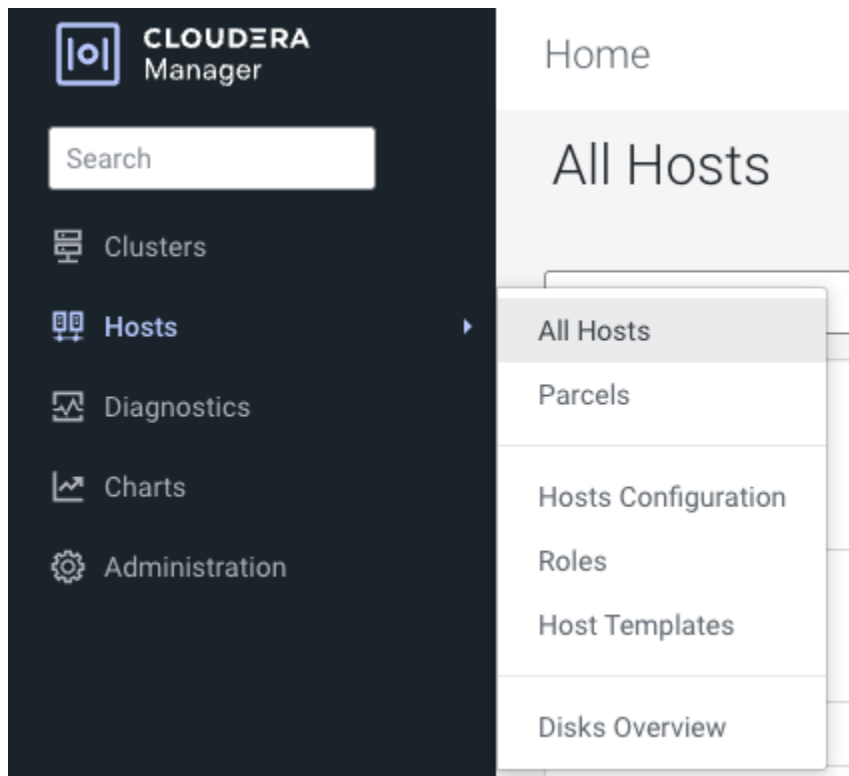
FAQs and Troubleshooting

A collection of frequently asked questions and debugging steps while installing and deploying the Cloudera Semantic Search.

How to get host details

Refer to the following steps to get the host details.

1. Log in to the Cloudera Manager as an administrator.
2. Go to the Hosts/All Host menu



3. On the All Hosts page you can find the cluster hosts
4. Copy the hostname.

This document has been released as part of a technical preview for features described herein. Technical preview components are provided as a convenience to our customers for their evaluation and trial usage. These components are provided 'as is' without warranty or support. Further, Cloudera assumes no liability for the usage of technical preview components, which should be used by customers at their own risk.

All Hosts Configuration Inspect All Hosts

Q Search Filters Last Updated: Jul 22, 1:05:51 PM UTC

Filters

- STATUS
 - Good Health 8
- CLUSTERS
- CORES
- COMMISSION STATE
- LAST HEARTBEAT
- LOAD (1 MINUTE)
- LOAD (5 MINUTES)
- LOAD (15 MINUTES)
- MAINTENANCE MODE

Actions for Selected

Columns: 12 Selected

☐	Status	Name	IP	IPv6	Roles	Tags	Commission State	Last Heartbeat	Load Average	Disk Usage
☐	✓	test-css-tpayer3-data0.css-732.xcu2-8y8x.dev.cldr.work	10.124.200.61		2 Roles	1 Tag	Commissioned	2.91s	0.15 0.12 0.05	82.2 GiB / 816.1 GiB
☐	✓	test-css-tpayer3-data1.css-732.xcu2-8y8x.dev.cldr.work	10.124.200.65		2 Roles	1 Tag	Commissioned	2.65s	0.12 0.09 0.05	82 GiB / 816.1 GiB
☐	✓	test-css-tpayer3-data2.css-732.xcu2-8y8x.dev.cldr.work	10.124.200.60		2 Roles	1 Tag	Commissioned	887ms	0.12 0.14 0.09	82.2 GiB / 816.1 GiB
☐	✓	test-css-tpayer3-gateway0.css-732.xcu2-8y8x.dev.cldr.work	10.124.200.6		9 Roles	1 Tag	Commissioned	227ms	0.42 0.57 0.59	83.7 GiB / 620.8 GiB
☐	✓	test-css-tpayer3-ingest0.css-732.xcu2-8y8x.dev.cldr.work	10.124.200.67		2 Roles	1 Tag	Commissioned	1.07s	0.06 0.07 0.02	81.5 GiB / 422.9 GiB
☐	✓	test-css-tpayer3-manager0.css-732.xcu2-8y8x.dev.cldr.work	10.124.200.64		2 Roles	1 Tag	Commissioned	246ms	0.03 0.05 0.03	81.5 GiB / 422.9 GiB
☐	✓	test-css-tpayer3-manager1.css-732.xcu2-8y8x.dev.cldr.work	10.124.200.62		2 Roles	1 Tag	Commissioned	58ms	0.03 0.04 0.00	81.5 GiB / 422.9 GiB
☐	✓	test-css-tpayer3-manager2.css-732.xcu2-8y8x.dev.cldr.work	10.124.200.63		2 Roles	1 Tag	Commissioned	2.54s	0.01 0.04 0.00	81.5 GiB / 422.9 GiB

The OpenSearch Dashboard UI is not accessible through Knox

Problem description

In the Cloudera Search Service Technical Preview release, the OpenSearch Dashboard UI does not support Knox-based access. Attempting to reach the dashboard UI through Knox results in an inaccessible endpoint.

Workaround

Use port forwarding to access the Dashboards UI. Run the following command to set up port forwarding:

```
ssh -i ~/.ssh/<ssh-key> -L 5601:<gateway-host-FQDN>:5601
ccloudbreak@<gateway-host-FQDN>
```

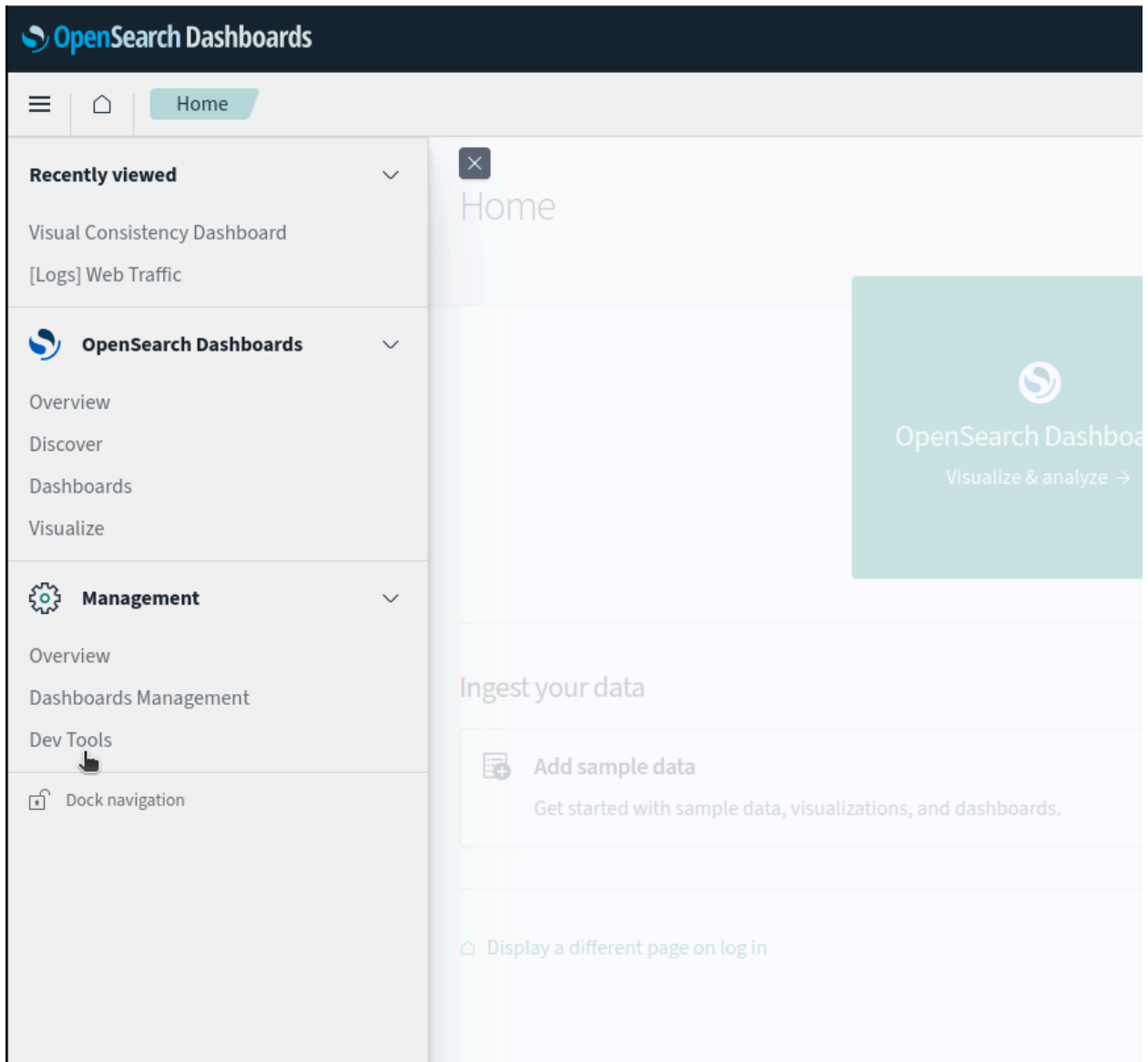
After the port forwarding session is established, access the Dashboards UI at

<https://localhost:5601> in a web browser.

How to run the commands using the OpenSearch dashboard

1. Navigate to the dev tools console from the dropdown menu on the left (or use the link on the right).

This document has been released as part of a technical preview for features described herein. Technical preview components are provided as a convenience to our customers for their evaluation and trial usage. These components are provided 'as is' without warranty or support. Further, Cloudera assumes no liability for the usage of technical preview components, which should be used by customers at their own risk.



Important: When you create an index for Approximate Nearest Neighbor searches, you must use the Lucene engine exclusively. The Faiss engine is not supported and might cause cluster instability.

2. Go to <https://docs.opensearch.org/3.6/vector-search/vector-search-techniques/approximate-knn/>

Enter the code block into the console. Select the first row and click Run to create the index.

This document has been released as part of a technical preview for features described herein. Technical preview components are provided as a convenience to our customers for their evaluation and trial usage. These components are provided 'as is' without warranty or support. Further, Cloudera assumes no liability for the usage of technical preview components, which should be used by customers at their own risk.

OpenSearch Dashboards



Dev Tools

Console

HistorySettingsHelpExportImport

1PUT my-knn-index-1

2{

3 "settings": {

4 "index": {

5 "knn": true,

6 "knn.algo_param.ef_search": 100

7 }

8 },

9 "mappings": {

10 "properties": {

11 "my_vector1": {

12 "type": "knn_vector",

13 "dimension": 2,

14 "space_type": "l2",

15 "method": {

16 "name": "hnsw",

17 "engine": "lucene",

18 "parameters": {

19 "ef_construction": 128,

20 "m": 24

21 }

22 }

23 },

24 "my_vector2": {

25 "type": "knn_vector",

26 "dimension": 4,

27 "space_type": "innerproduct",

28 "method": {

29 "name": "hnsw",

30 "engine": "lucene",

31 "parameters": {

32 "ef_construction": 256,

33 "m": 48

34 }

35 }

36 }

37 }

38 }

39 }

1- {

2 "ackr

3 "shar

4 "inde

5- }

This document has been released as part of a technical preview for features described herein. Technical preview components are provided as a convenience to our customers for their evaluation and trial usage. These components are provided 'as is' without warranty or support. Further, Cloudera assumes no liability for the usage of technical preview components, which should be used by customers at their own risk.

```
200 - OK 105 ms
1  #! Deprecation: [index.store.hybrid.mmap.extensions] setting was deprecated in OpenSearch and will be
   removed in a future release! See the breaking changes documentation for the next major version.
2  {
3    "acknowledged": true,
4    "shards_acknowledged": true,
5    "index": "my-knn-index-1"
6  }
```

3. Insert data.

```
Console
History Settings Help Export Import
1  POST _bulk
2  { "index": { "_index": "my-knn-index-1", "_id": "1" } }
3  { "my_vector1": [1.5, 2.5], "price": 12.2 }
4  { "index": { "_index": "my-knn-index-1", "_id": "2" } }
5  { "my_vector1": [2.5, 3.5], "price": 7.1 }
6  { "index": { "_index": "my-knn-index-1", "_id": "3" } }
7  { "my_vector1": [3.5, 4.5], "price": 12.9 }
8  { "index": { "_index": "my-knn-index-1", "_id": "4" } }
9  { "my_vector1": [5.5, 6.5], "price": 1.2 }
10 { "index": { "_index": "my-knn-index-1", "_id": "5" } }
11 { "my_vector1": [4.5, 5.5], "price": 3.7 }
12 { "index": { "_index": "my-knn-index-1", "_id": "6" } }
13 { "my_vector2": [1.5, 5.5, 4.5, 6.4], "price": 10.3 }
14 { "index": { "_index": "my-knn-index-1", "_id": "7" } }
15 { "my_vector2": [2.5, 3.5, 5.6, 6.7], "price": 5.5 }
16 { "index": { "_index": "my-knn-index-1", "_id": "8" } }
17 { "my_vector2": [4.5, 5.5, 6.7, 3.7], "price": 4.4 }
18 { "index": { "_index": "my-knn-index-1", "_id": "9" } }
19 { "my_vector2": [1.5, 5.5, 4.5, 6.4], "price": 8.9 }
20
```

This document has been released as part of a technical preview for features described herein. Technical preview components are provided as a convenience to our customers for their evaluation and trial usage. These components are provided 'as is' without warranty or support. Further, Cloudera assumes no liability for the usage of technical preview components, which should be used by customers at their own risk.

200 - OK63 ms

```
{
  "took": 42,
  "errors": false,
  "items": [
    {
      "index": {
        "_index": "my-knn-index-1",
        "_id": "1",
        "_version": 1,
        "result": "created",
        "_shards": {
          "total": 2,
          "successful": 1,
          "failed": 0
        },
        "_seq_no": 0,
        "_primary_term": 1,
        "status": 201
      }
    },
    {
      "index": {
        "_index": "my-knn-index-1",
        "_id": "2",
        "_version": 1,
        "result": "created",
        "_shards": {
          "total": 2,
          "successful": 1,
          "failed": 0
        },
        "_seq_no": 1,
        "_primary_term": 1,
        "status": 201
      }
    },
    {
      "index": {
        "_index": "my-knn-index-1",
        "_id": "3",
        "_version": 1
```

4. Run a k-nn query as shown below.

Console

History Settings Help Export Import

```
1 GET my-knn-index-1/_search
2 {
3   "size": 2,
4   "query": {
5     "knn": {
6       "my_vector2": {
7         "vector": [2, 3, 5, 6],
8         "k": 2
9       }
10    }
11  }
12 }
```

This document has been released as part of a technical preview for features described herein. Technical preview components are provided as a convenience to our customers for their evaluation and trial usage. These components are provided 'as is' without warranty or support. Further, Cloudera assumes no liability for the usage of technical preview components, which should be used by customers at their own risk.

```
1 {
2   "took": 24,
3   "timed_out": false,
4   "_shards": {
5     "total": 1,
6     "successful": 1,
7     "skipped": 0,
8     "failed": 0
9   },
10  "hits": {
11    "total": {
12      "value": 2,
13      "relation": "eq"
14    },
15    "max_score": 84.7,
16    "hits": [
17      {
18        "_index": "my-knn-index-1",
19        "_id": "7",
20        "_score": 84.7,
21        "_source": {
22          "my_vector2": [
23            2.5,
24            3.5,
25            5.6,
26            6.7
27          ],
28          "price": 5.5
29        }
30      },
31      {
32        "_index": "my-knn-index-1",
33        "_id": "8",
34        "_score": 82.2,
35        "_source": {
36          "my_vector2": [
37            4.5,
38            5.5,
39            6.7,
40            3.7
41          ],
42          "price": 4.4
43        }
44      }
45    ]
46  }
47 }
```

This document has been released as part of a technical preview for features described herein. Technical preview components are provided as a convenience to our customers for their evaluation and trial usage. These components are provided 'as is' without warranty or support. Further, Cloudera assumes no liability for the usage of technical preview components, which should be used by customers at their own risk.

How to add Machine Learning, Ingest, or Coordinator roles to the Cloudera Semantic Search cluster

The Cloudera Semantic Search cluster does not have Ingest, Coordinator and Machine Learning Roles by default, however these roles can be added on demand by scaling the associated node groups.

The Cloudera Semantic Search cluster has Coordinator, Ingest and Worker Node Groups with default 0 node count. These node groups can be scaled up on demand when the associated roles are in need.

Prerequisites

Ensure that the CDP CLI is installed on your machine. For more information, see [Installing Beta CDP CLI](#).

Steps

1. Launch CDP CLI.
2. Add the Coordinator role by scaling the Coordinator node group.

None

```
cdp datahub scale-cluster --cluster-name <cluster-name> --instance-group-name  
coordinator --instance-group-desired-count 1
```

3. Add Ingest Role by scaling the Ingest node group.

None

```
cdp datahub scale-cluster --cluster-name <cluster-name> --instance-group-name  
ingest --instance-group-desired-count 1
```

4. Add ML Role by scaling the Worker node group.

This document has been released as part of a technical preview for features described herein. Technical preview components are provided as a convenience to our customers for their evaluation and trial usage. These components are provided 'as is' without warranty or support. Further, Cloudera assumes no liability for the usage of technical preview components, which should be used by customers at their own risk.

None

```
cdp datahub scale-cluster --cluster-name <cluster-name> --instance-group-name  
worker --instance-group-desired-count 1
```

5. After performing the steps given above, go to **Cloudera Manager > OpenSearch** service and click **Actions > Start** to restart the service and any other stale services.

Only one OpenSearch service definition available in Ranger for multiple Cloudera Semantic Search (Data Hub) clusters on a single environment

Problem description

When multiple Cloudera Semantic Search clusters exist on a single environment, only one OpenSearch service definition is available on the Ranger Admin UI. The service definition might not correspond to the cluster endpoint you are using, which can result in incorrect access control policies being applied to your OpenSearch resources.

Workaround

Verify that the OpenSearch URL in the Ranger Admin UI service definition corresponds to the cluster endpoint you are using. To verify the service definition:

1. Log in to Ranger Admin UI.
2. Navigate to the OpenSearch service definition.
3. Confirm that the OpenSearch URL matches the endpoint of the Cloudera Semantic Search cluster you intend to manage.
4. If the URL does not match, update the service definition to point to the correct cluster endpoint.

This document has been released as part of a technical preview for features described herein. Technical preview components are provided as a convenience to our customers for their evaluation and trial usage. These components are provided 'as is' without warranty or support. Further, Cloudera assumes no liability for the usage of technical preview components, which should be used by customers at their own risk.

Workload user is unable to request the cluster status

Problem description

When a Cloudera workload user tries to access a cluster monitoring endpoint through Knox Token Integration, the request fails with a `security_exception` error. The `ranger_authorized_role` does not include the required cluster monitoring permissions (`cluster:monitor/main` and `cluster:monitor/health`), causing the request to return a 403 Forbidden response.

Workaround

1. Log in to the cluster host as an admin or a super admin user.
2. Add the required cluster monitoring permissions to the `ranger_authorized_role`:

```
None
curl -k -u admin:admin \
  -X PATCH
  "https://<css-node>:9200/_plugins/_security/api/roles/ranger_authorized_role" \
  -H "Content-Type: application/json" \
  -d '[
    {
      "op": "add",
      "path": "/cluster_permissions",
      "value": ["cluster:monitor/main", "cluster:monitor/health"]
    }
  ]'
```

After running the command, the workload user can access the cluster status endpoint successfully.

Appendix

1. To enable role-based Document-Level Security (DLS) in OpenSearch, see [Configuring Document-Level Security for Cloudera Semantic Search](#) document.
2. To check the usage of the `securityadmin.sh` script while working with the security plugin, see [Applying changes to the security configuration files](#) document.
3. Client-side examples that demonstrate how to connect to Cloudera Semantic Search and authenticate with Kerberos, see [Client-side Examples](#) document.
4. To check how to build charts and add them to the respective dashboards, see [Adding Cloudera Semantic Search chart to the Cloudera Manager](#) document.
5. To know how to integrate Knox with Cloudera Semantic Search, see [Integrating Knox with Cloudera Semantic Search](#) document.

This document has been released as part of a technical preview for features described herein. Technical preview components are provided as a convenience to our customers for their evaluation and trial usage. These components are provided 'as is' without warranty or support. Further, Cloudera assumes no liability for the usage of technical preview components, which should be used by customers at their own risk.