

Cloudera Runtime 7.3.2

Cloudera Semantic Search Overview

Date published: 2026-05-11

Date modified: 2026-08-14

CLOUDERA

<https://docs.cloudera.com/>

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.

Contents

Introduction to Cloudera Semantic Search [Technical Preview].....	4
--	----------

Introduction to Cloudera Semantic Search [Technical Preview]

Cloudera Semantic Search provides vector database and semantic retrieval capabilities for Cloudera on cloud environments to power AI-driven applications and anomaly detection.

Overview

Cloudera Semantic Search integrates hybrid multimodal search functionality into the enterprise data estate. It serves as the retrieval foundation for autonomous AI agents and Retrieval-Augmented Generation (RAG) by transforming knowledge sources into searchable vector representations. This service maintains data gravity by enabling searches across internal documentation, support systems, and data lakehouses without transferring information outside the organization.

[../images/CSS-RAGFlow.mp4](#)

Core functionalities

The service includes the following primary features for managing vector data:

- Indexing: Mechanism for storing and searching high-dimensional vector data.
- K-NN search: Functionality to find the top documents semantically closest to an input query.
- Neural search: Extension of K-NN that supports semantic, multimodal, and hybrid searches using foundation models.

Retrieval-Augmented Generation workflow

The search functionality follows a structured process to enhance large language models with enterprise-specific context:

1. Indexing: An embedding model is selected to convert structured and unstructured data into k-NN vectors for indexing in the Cloudera Semantic Search service.
2. Querying: The same model generates an embedding for a user input query.
3. Retrieval: The search engine identifies documents semantically similar to the query embedding and returns the resultset.

Component roles

The service uses several node types to manage operations and data:

- Cluster_manager nodes: Manages cluster state, tracks node joins, and allocates shards.
- Data nodes: Store and search data while performing indexing and aggregation operations.
- Coordinating nodes: Delegate client requests to shards and aggregate final results.
- ML nodes: Execute machine learning tasks and resource calculations for AI models.
- Ingest nodes: Pre-process and transform data through an ingest pipeline before indexing.
- Dashboard nodes: Provide a visual interface to interact with and monitor search data.



Important: This release of Cloudera Semantic Search for Cloudera on cloud is a technical preview and is not intended for production environments.

Additional details

For more information on Cloudera Semantic Search, see [Cloudera Semantic Search user guide for Cloudera on cloud environments](#), technical preview documentation.