Trino Overview

Date published:
Date modified:

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

Trino overview.....4
Trino architecture.....4
Trino query execution.....4

Trino overview

Trino is a fast, distributed SQL query engine that reads and joins data from many remote data sources in a single query, without first copying the data into a central store.

Trino does not store data. It connects to your data sources, reads the data from them, and runs your SQL queries across a group of nodes that work at the same time. This design lets you query large datasets quickly and combine data from different systems in one query.

In Cloudera Data Warehouse, you run Trino as a Virtual Warehouse. The Virtual Warehouse provides the compute resources that run your queries against the data sources you configure.

Trino provides the following benefits:

- Reads data directly from the source, so you do not move or duplicate data.
- Joins data from different systems, such as a database and a data lake, in one query.
- Runs queries in parallel across many workers for fast results.
- Scales compute up or down based on the query workload.

Trino architecture

A Trino cluster contains one coordinator and one or more workers that together plan queries, read data from the data sources, and return results in parallel.

A Trino Virtual Warehouse in Cloudera Data Warehouse is a cluster made up of two types of nodes:

Coordinator

The coordinator is the node that the client connects to. It reads the SQL query, plans the work, and divides the work into smaller tasks for the workers. When the workers finish, the coordinator joins their results and returns the final result to the client. Each cluster has one coordinator.

Workers

The workers do the actual work. Each worker runs the tasks it receives, reads data from the data sources, and shares results with the other workers. In Cloudera Data Warehouse, the number of workers can increase or decrease automatically based on how busy the warehouse is.

Trino query execution

Trino runs a query by breaking it into stages, tasks, and splits that run in parallel across the worker nodes, and then combines the partial results into one final result.

When you run a query, Trino breaks it into smaller pieces of work that run at the same time across the cluster. Trino uses the following units of work:

- **Statement** - The SQL text that you send to the coordinator.
- **Query** - The plan that the coordinator builds from the statement.
- **Stage** - A part of the query plan. Stages pass data to one another.
- **Task** - A part of a stage that runs on a worker. One stage runs as many tasks across the workers.
- **Split** - A piece of the source data that a task reads. Splits let many tasks work in parallel.

After the tasks finish, the coordinator collects the results from all the workers, joins them, and sends the final result to the client.