

Home

- [Apache Hive](#)
- [Hive Documentation](#)
 - [General Information about Hive](#)
 - [User Documentation](#)
 - [Administrator Documentation](#)
 - [HCatalog and WebHCat Documentation](#)
 - [Resources for Contributors](#)
- [Hive Versions and Branches](#)

Apache Hive

The **Apache Hive™** data warehouse software facilitates reading, writing, and managing large datasets residing in distributed storage and queried using SQL syntax.

Built on top of **Apache Hadoop™**, Hive provides the following features:

- Tools to enable easy access to data via SQL, thus enabling data warehousing tasks such as extract/transform/load (ETL), reporting, and data analysis.
- A mechanism to impose structure on a variety of data formats
- Access to files stored either directly in **Apache HDFS™** or in other data storage systems such as **Apache HBase™**
- Query execution via [Apache Tez™](#), [Apache Spark™](#), or MapReduce
- Procedural language with HPL-SQL
- Sub-second query retrieval via [Hive LLAP](#), [Apache YARN](#) and [Apache Slider](#).

Hive provides standard SQL functionality, including many of the later [SQL:2003](#), [SQL:2011](#), and [SQL:2016](#) features for analytics.

Hive's SQL can also be extended with user code via user defined functions (UDFs), user defined aggregates (UDAFs), and user defined table functions (UDTFs).

There is not a single "Hive format" in which data must be stored. Hive comes with built in connectors for comma and tab-separated values (CSV /TSV) text files, [Apache Parquet™](#), [Apache ORC™](#), and other formats. Users can extend Hive with connectors for other formats. Please see [File Formats](#) and [Hive SerDe](#) in the [Developer Guide](#) for details.

Hive is not designed for online transaction processing (OLTP) workloads. It is best used for traditional data warehousing tasks.

Hive is designed to maximize scalability (scale out with more machines added dynamically to the Hadoop cluster), performance, extensibility, fault-tolerance, and loose-coupling with its input formats.

Components of Hive include HCatalog and WebHCat.

- **HCatalog** is a table and storage management layer for Hadoop that enables users with different data processing tools — including Pig and MapReduce — to more easily read and write data on the grid.
- **WebHCat** provides a service that you can use to run Hadoop MapReduce (or YARN), Pig, Hive jobs. You can also perform Hive metadata operations using an HTTP (REST style) interface.

Hive Documentation

The links below provide access to the Apache Hive wiki documents. This list is not complete, but you can navigate through these wiki pages to find additional documents. For more information, please see the official [Hive website](#).

General Information about Hive

- [Getting Started](#)
- [Books about Hive](#)
- [Presentations and Papers about Hive](#)
- [Sites and Applications Powered by Hive](#)
- [Related Projects](#)
- [FAQ](#)
- [Hive Users Mailing List](#)
- [Hive IRC Channel: #hive on irc.freenode.net](#)
- [About This Wiki](#)

User Documentation

- Hive Tutorial
- Hive SQL Language Manual: Commands, CLIs, Data Types, DDL (create/drop/alter/truncate/show/describe), Statistics (analyze), Indexes, Archiving, DML (load/insert/update/delete/merge, import/export, explain plan), Queries (select), Operators and UDFs, Locks, Authorization
- File Formats and Compression: RCFile, Avro, ORC, Parquet; Compression, LZ0
- Procedural Language: Hive HPL/SQL
- Hive Configuration Properties
- Hive Clients
 - Hive Client (JDBC, ODBC, Thrift)
 - HiveServer2: Overview, HiveServer2 Client and Beeline, Hive Metrics
- Hive Web Interface
- Hive SerDes: Avro SerDe, Parquet SerDe, CSV SerDe, JSON SerDe
- Hive Accumulo Integration
- Hive HBase Integration
- Druid Integration
- Hive Transactions, Streaming Data Ingest, and Streaming Mutation API
- Hive Counters

Administrator Documentation

- Installing Hive
- Configuring Hive
- Setting Up Metastore
 - Hive Schema Tool
- Setting Up Hive Web Interface
- Setting Up Hive Server (JDBC, ODBC, Thrift, HiveServer2)
- Hive Replication
- Hive on Amazon Web Services
- Hive on Amazon Elastic MapReduce
- Hive on Spark: Getting Started

HCatalog and WebHCat Documentation

- HCatalog
- WebHCat (Templeton)

Resources for Contributors

- How to Contribute
- Hive Contributors Meetings
- Hive Developer Docs
 - Hive Developer Guide (code organization, compile and run Hive, unit tests, debug, pluggable interfaces)
 - Hive Developer FAQ (move files, build Hive, test Hive, MiniDriver and Beeline tests)
 - Plugin Developer Kit
 - Writing UDTFs
 - Hive APIs Overview
- Hive Testing Docs
 - FAQ: Testing
 - Developer Guide: Unit Tests
 - Unit Testing Hive SQL
 - Unit Test Parallel Execution
 - Tips for Adding New Tests
 - Hive PTest2 Infrastructure
 - Hive PreCommit Patch Testing
 - MiniDriver Tests
 - Running Yetus
 - MetaStore API Tests
- Hive Performance
- Hive Architecture Overview
- Hive Design Docs: Completed; In Progress; Proposed; Incomplete, Abandoned, Other
- Roadmap/Call to Add More Features
- Full-Text Search over All Hive Resources
- How to edit the website
- Becoming a Committer
- How to Commit

- [How to Release](#)
- [Project Bylaws](#)

Hive Versions and Branches

Recent versions of Hive are available on the [Downloads](#) page of the Hive website. For each version, the page provides the release date and a link to the change log. If you want a change log for an earlier version (or a development branch), use the [Configure Release Notes](#) page.

The [Apache Hive JIRA](#) keeps track of changes to Hive code, documentation, infrastructure, etc. The version number or branch for each resolved JIRA issue is shown in the "Fix Version/s" field in the Details section at the top of the issue page. For example, [HIVE-5107](#) has a fix version of 0.13.0.

Sometimes a version number changes before the release. When that happens, the original number might still be found in JIRA, wiki, and [mailing list](#) discussions. For example:

Release Number	Original Number
1.0.0	0.14.1
1.1.0	0.15.0
2.3.0	2.2.0

More information about Hive branches is available in How to Contribute: [Understanding Hive Branches](#).

Apache Hive, Apache Hadoop, Apache HBase, Apache HDFS, Apache, the Apache feather logo, and the Apache Hive project logo are trademarks of The Apache Software Foundation.