AdminManual Installation

Installing Hive

You can install a stable release of Hive by downloading and unpacking a tarball, or you can download the source code and build Hive using Maven (release 0.13 and later) or Ant (release 0.12 and earlier).

Hive installation has these requirements:

- Java 1.7 (preferred).
  Note: Hive versions 1.2 onward require Java 1.7 or newer. Hive versions 0.14 to 1.1 work with Java 1.6, but prefer 1.7. Users are strongly advised to start moving to Java 1.8 (see HIVE-8607).
- Hadoop 2.x (preferred), 1.x (not supported by Hive 2.0.0 onward).
  Hive versions up to 0.13 also supported Hadoop 0.20.x, 0.23.x.
- Hive is commonly used in production Linux and Windows environment. Mac is a commonly used development environment. The instructions in this document are applicable to Linux and Mac. Using it on Windows would require slightly different steps.

Installing from a Tarball

Start by downloading the most recent stable release of Hive from one of the Apache download mirrors (see Hive Releases).

Next you need to unpack the tarball. This will result in the creation of a subdirectory named `hive-x.y.z` (where `x.y.z` is the release number):

```
$ tar -xzvf hive-x.y.z.tar.gz
```

Set the environment variable `HIVE_HOME` to point to the installation directory:

```
$ cd hive-x.y.z
$ export HIVE_HOME={{pwd}}
```

Finally, add `$HIVE_HOME/bin` to your `PATH`:

```
$ export PATH=$HIVE_HOME/bin:$PATH
```

Installing from Source Code (Hive 1.2.0 and Later)

Version information

To build Hive 1.2.0 and later releases with Apache Maven, see Getting Started: Building Hive from Source. You will need Java 1.7 or newer.

Installing from Source Code (Hive 0.13.0 and Later)

Version information

To build Hive 0.13.0 and later releases with Apache Maven, see Getting Started: Building Hive from Source.

Installing from Source Code (Hive 0.12.0 and Earlier)

Version information

This section describes installation for Hive 0.12.0 and earlier releases, which use Apache Ant to build Hive.

Installing Hive is simple and only requires having Java 1.6 and Ant installed on your machine (for Hive 0.12 and earlier).

Hive is available via SVN at http://svn.apache.org/repos/asf/hive/branches. You can download it by running the following command.

```
$ svn co http://svn.apache.org/repos/asf/hive/branches/branch-#.# hive
```

where `#.#` is the Hive release number. For release 0.8.1, use "`branch-0.8-r2`".

To build Hive, execute the following command on the base directory:

```
$ ant package
```

It will create the subdirectory build/dist with the following contents:

- README.txt: readme file.
- bin/: directory containing all the shell scripts
- lib/: directory containing all required jar files
- conf/: directory with configuration files
- examples/: directory with sample input and query files

Subdirectory build/dist should contain all the files necessary to run Hive. You can run it from there or copy it to a different location, if you prefer.

In order to run Hive, you must have Hadoop in your path or have defined the environment variable HADOOP_HOME with the Hadoop installation directory.

Moreover, we strongly advise users to create the HDFS directories /tmp and /user/hive/warehouse (also known as hive.metastore.warehouse.dir) and set them chmod g+w before tables are created in Hive.

Next Steps

You can begin using Hive as soon as it is installed, although you will probably want to configure it first.

Hive CLI and Beeline CLI

To use the Hive command line interface (CLI) go to the Hive home directory and execute the following command:

```
$ bin/hive
```

The Hive home directory is the one with the contents of build/dist for Hive 0.12 and earlier; for Hive 0.13 and later it is packaging/target/apache-hive-*<release_string>*-bin/apache-hive-*<release_string>*-bin/.

HiveServer2 (introduced in Hive 0.11) has a new CLI called Beeline (see Beeline – New Command Line Shell). To use Beeline, execute the following command in the Hive home directory:

```
$ bin/beeline
```

Hive Metastore

Metadata is stored in an embedded Derby database whose disk storage location is determined by the Hive configuration variable named javax.jdo.option.ConnectionURL. By default, this location is ./metastore_db (see conf/hive-default.xml).

Using Derby in embedded mode allows at most one user at a time. To configure Derby to run in server mode, see Hive Using Derby in Server Mode.

To configure a database other than Derby for the Hive metastore, see Hive Metastore Administration.

Next Step: Configuring Hive.

HCatalog and WebHCat

HCatalog

Version

HCatalog is installed with Hive, starting with Hive release 0.11.0.

If you install Hive from the binary tarball, the `hcat` command is available in the `hcatalog/bin` directory. However, most `hcat` commands can be issued as `hive` commands except for "`hcat -g`" and "`hcat -p`". Note that the `hcat` command uses the `-p` flag for permissions but `hive` uses it to specify a port number. The HCatalog CLI is documented here and the Hive CLI is documented here.

HCatalog installation is documented here.

WebHCat (Templeton)

Version

WebHCat is installed with Hive, starting with Hive release 0.11.0.

If you install Hive from the binary tarball, the WebHCat server command `webhcat_server.sh` is in the `hcatalog/sbin` directory.

WebHCat installation is documented here.