

Source Installation

Table of contents

1 Server Installation from Source.....	2
2 Starting the Server.....	6
3 Logging.....	6
4 Stopping the Server.....	6
5 Client Installation.....	6

1 Server Installation from Source

Prerequisites

- Machine to build the installation tar on
- Machine on which the server can be installed - this should have access to the Hadoop cluster in question, and be accessible from the machines you launch jobs from
- an RDBMS - we recommend MySQL and provide instructions for it
- Hadoop cluster
- Unix user that the server will run as, and, if you are running your cluster in secure mode, an associated Kerberos service principal and keytabs.

Throughout these instructions when you see a word in *italics* it indicates a place where you should replace the word with a locally appropriate value such as a hostname or password.

Building a tarball

If you downloaded HCatalog from Apache or another site as a source release, you will need to first build a tarball to install. You can tell if you have a source release by looking at the name of the object you downloaded. If it is named `hcatalog-src-0.4.0-incubating.tar.gz` (notice the **src** in the name) then you have a source release.

If you do not already have Apache Ant installed on your machine, you will need to obtain it. You can get it from the [Apache Ant website](#). Once you download it, you will need to unpack it somewhere on your machine. The directory where you unpack it will be referred to as *ant_home* in this document.

If you do not already have Apache Forrest installed on your machine, you will need to obtain it. You can get it from the [Apache Forrest website](#). Once you download it, you will need to unpack it somewhere on your machine. The directory where you unpack it will be referred to as *forrest_home* in this document.

To produce a tarball from this do the following:

Create a directory to expand the source release in. Copy the source release to that directory and unpack it.

```
mkdir /tmp/hcat_source_release
cp hcatalog-src-0.4.0-incubating.tar.gz /tmp/
hcat_source_release
cd /tmp/hcat_source_release
tar xzf hcatalog-src-0.4.0-incubating.tar.gz
```

Change directories into the unpacked source release and build the installation tarball.

```
cd hcatalog-src-0.4.0-incubating
```

```
ant_home/bin/ant -Dhcatalog.version=0.4.0 -
Dforrest.home=forrest_home tar
```

The tarball for installation should now be at `build/hcatalog-0.4.0.tar.gz`

Database Setup

If you do not already have Hive installed with MySQL, the following will walk you through how to do so. If you have already set this up, you can skip this step.

Select a machine to install the database on. This need not be the same machine as the Thrift server, which we will set up later. For large clusters we recommend that they not be the same machine. For the purposes of these instructions we will refer to this machine as *hivedb.acme.com*

Install MySQL server on *hivedb.acme.com*. You can obtain packages for MySQL from [MySQL's download site](#). We have developed and tested with versions 5.1.46 and 5.1.48. We suggest you use these versions or later. Once you have MySQL up and running, use the `mysql` command line tool to add the `hive` user and `hivemetastoredb` database. You will need to pick a password for your `hive` user, and replace *dbpassword* in the following commands with it.

```
mysql -u root
mysql> CREATE USER 'hive'@'hivedb.acme.com' IDENTIFIED BY
' dbpassword' ;
mysql> CREATE DATABASE hivemetastoredb DEFAULT CHARACTER SET
latin1 DEFAULT COLLATE latin1_swedish_ci;
mysql> GRANT ALL PRIVILEGES ON hivemetastoredb.* TO
'hive'@'hivedb.acme.com' WITH GRANT OPTION;
mysql> flush privileges;
mysql> quit;
```

Use the database installation script found in the Hive package to create the database. `hive_home` in the line below refers to the directory where you have installed Hive. If you are using Hive rpms, then this will be `/usr/lib/hive`.

```
mysql -u hive -D hivemetastoredb -hhivedb.acme.com -p < hive_home/
scripts/metastore/upgrade/mysql/hive-schema-0.9.0.mysql.sql
```

Thrift Server Setup

If you do not already have Hive running a metastore server using Thrift, you can use the following instructions to setup and run one. You may skip this step if you already are using a Hive metastore server.

Select a machine to install your Thrift server on. For smaller and test installations this can be the same machine as the database. For the purposes of these instructions we will refer to this machine as *hcatsvr.acme.com*.

If you have not already done so, install Hive 0.9 on this machine. You can use the [binary distributions](#) provided by Hive or rpms available from [Apache Bigtop](#). If you use the Apache Hive binary distribution, select a directory, henceforth referred to as *hive_home*, and untar the distribution there. If you use the rpms, *hive_home* will be `/usr/lib/hive`.

Install the MySQL Java connector libraries on *hcatsvr.acme.com*. You can obtain these from [MySQL's download site](#).

Select a user to run the Thrift server as. This user should not be a human user, and must be able to act as a proxy for other users. We suggest the name "hive" for the user. Throughout the rest of this documentation we will refer to this user as *hive*. If necessary, add the user to *hcatsvr.acme.com*.

Select a *root* directory for your installation of HCatalog. This directory must be owned by the *hive* user. We recommend `/usr/local/hive`. If necessary, create the directory. You will need to be the *hive* user for the operations described in the remainder of this Thrift Server Setup section.

Copy the HCatalog installation tarball into a temporary directory, and untar it. Then change directories into the new distribution and run the HCatalog server installation script. You will need to know the directory you chose as *root* and the directory you installed the MySQL Java connector libraries into (referred to in the command below as *dbroot*). You will also need your *hadoop_home*, the directory where you have Hadoop installed, and the port number you wish HCatalog to operate on which you will use to set *portnum*.

```
tar xzf hcatalog-0.4.0.tar.gz
cd hcatalog-0.4.0
share/hcatalog/scripts/hcat_server_install.sh -r root -d dbroot -h hadoop_home -p portnum
```

Now you need to edit your *hive_home/conf/hive-site.xml* file. Open this file in your favorite text editor. The following table shows the values you need to configure.

Parameter	Value to Set it to
hive.metastore.local	false
javax.jdo.option.ConnectionURL	jdbc:mysql:// <i>hostname</i> /hivemetastoredb?createDatabaseIfNotExist=true where <i>hostname</i> is the name of the machine you installed MySQL on.
javax.jdo.option.ConnectionDriverName	com.mysql.jdbc.Driver

Parameter	Value to Set it to
javax.jdo.option.ConnectionUserName	hive
javax.jdo.option.ConnectionPassword	<i>dbpassword</i> value you used in setting up the MySQL server above.
hive.semantic.analyzer.factory.impl	org.apache.hcatalog.cli.HCatSemanticAnalyzerFactory
hadoop.clientside.fs.operations	true
hive.metastore.warehouse.dir	<p>The directory can be a URI or an absolute file path. If it is an absolute file path, it will be resolved to a URI by the metastore:</p> <ul style="list-style-type: none"> -- If default hdfs was specified in core-site.xml, path resolves to HDFS location. -- Otherwise, path is resolved as local file: URI. <p>This setting becomes effective when creating new tables (it takes precedence over default DBS.DB_LOCATION_URI at the time of table creation).</p> <p>You only need to set this if you have not yet configured Hive to run on your system.</p>
hive.metastore.uris	thrift:// <i>hostname</i> : <i>portnum</i> where <i>hostname</i> is the name of the machine hosting the Thrift server, and <i>portnum</i> is the port number used above in the installation script.
hive.metastore.execute.setugi	true
hive.metastore.sasl.enabled	Set to true if you are using kerberos security with your Hadoop cluster, false otherwise.
hive.metastore.kerberos.keytab.file	The path to the Kerberos keytab file containing the metastore Thrift server's service principal. Only required if you set hive.metastore.sasl.enabled above to true.
hive.metastore.kerberos.principal	The service principal for the metastore Thrift server. You can reference your host as <code>_HOST</code> and it will be replaced with your actual hostname. Only required if you set hive.metastore.sasl.enabled above to true.

You can now proceed to starting the server.

2 Starting the Server

To start your server, HCatalog needs to know where Hive is installed. This is communicated by setting the environment variable `HIVE_HOME` to the location you installed Hive. Start the HCatalog server by switching directories to `root` and invoking `HIVE_HOME=hive_home sbin/hcat_server.sh start`

3 Logging

Server activity logs are located in `root/var/log/hcat_server`. Logging configuration is located at `root/conf/log4j.properties`. Server logging uses `DailyRollingFileAppender` by default. It will generate a new file per day and does not expire old log files automatically.

4 Stopping the Server

To stop the HCatalog server, change directories to the `root` directory and invoking `HIVE_HOME=hive_home sbin/hcat_server.sh stop`

5 Client Installation

Select a `root` directory for your installation of HCatalog client. We recommend `/usr/local/hcat`. If necessary, create the directory.

Copy the HCatalog installation tarball into a temporary directory, and untar it.

```
tar zxf hcatalog-0.4.0.tar.gz
```

Now you need to edit your `hive_home/conf/hive-site.xml` file. You can use the same file as on the server **except the value of `javax.jdo.option.ConnectionPassword` should be removed**. This avoids having the password available in plain text on all of your clients.

The HCatalog command line interface (CLI) can now be invoked as `HIVE_HOME=hive_home root/bin/hcat`.