

Installing HCatalog

Table of contents

1 Server Installation.....	2
2 Starting the Server.....	4
3 Logging.....	4
4 Stopping the Server.....	4
5 Client Install.....	4

1 Server Installation

Prerequisites

- 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
- MySQL db
- Hadoop cluster
- Unix user that the server will run as, and 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.

Database Setup

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 *hcatdb.acme.com*

Install MySQL server on *hcatdb.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'@'hcatdb.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'@'hcatdb.acme.com' WITH GRANT OPTION;
mysql> flush privileges;
mysql> quit;
```

In a temporary directory, untar the HCatalog artifact

```
tar xzf hcatalog-version.tar.gz
```

Use the database installation script found in the package to create the database

```
mysql -u hive -D hivemetastoredb -hhcatdb.acme.com -p < share/
hcatalog/hive/external/metastore/scripts/upgrade/mysql/hive-
schema-0.7.0.mysql.sql
```

Thrift Server Setup

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*.

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 "hcat" for the user. Throughout the rest of this documentation we will refer to this user as "hcat". 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 hcat user. We recommend `/usr/local/hcat`. If necessary, create the directory.

Download the HCatalog release 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-version.tar.gz cd hcatalog-version
share/hcatalog/scripts/hcat_server_install.sh -r root -d dbroot -
h hadoop_home -p portnum
```

Now you need to edit your *root/etc/hcatalog/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
javax.jdo.option.ConnectionURL	In the JDBC connection string, change DBHOSTNAME to the name of the machine you put the MySQL server on.
javax.jdo.option.ConnectionPassword	<i>dbpassword</i> value you used in setting up the MySQL server above
hive.metastore.warehouse.dir	The directory you want to use for the default database in your installation
hive.metastore.uris	You need to set the hostname to your Thrift server. Replace <i>SVRHOST</i> with the name of the machine

Parameter	Value to Set it to
	you are installing the Thrift server on. You can also change the port the Thrift server runs on by changing the default value of 3306.
hive.metastore.sasl.enabled	Set to true by default. Set to false if you do not wish to secure the thrift interface. This can be convenient for testing. We do not recommend turning this off in production.
hive.metastore.kerberos.keytab.file	The path to the Kerberos keytab file containing the metastore thrift server's service principal.
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

You can now proceed to starting the server.

2 Starting the Server

Start the HCatalog server by switching directories to *root* and invoking the start script `share/hcatalog/scripts/hcat_server_start.sh`

3 Logging

Server activity logs and gc 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 invoke the stop script `share/hcatalog/scripts/hcat_server_stop.sh`

5 Client Install

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

Download the HCatalog release into a temporary directory, and untar it.

```
tar xzf hcatalog-version.tar.gz
```

Now you need to edit your `root/etc/hcatalog/hive-site.xml` file. Open this file in your favorite text editor. The following table shows the values you need to configure. These

values should match the values set on the HCatalog server. Do **NOT** copy the configuration file from your server installation as that contains the password to your database, which you should not distribute to your clients.

Parameter	Value to Set it to
hive.metastore.warehouse.dir	The directory you want to use for the default database in your installation
hive.metastore.uris	You need to set the hostname with your Thrift server to use by replacing <i>SVRHOST</i> with the name of the machine you are installing the Thrift server on. You can also change the port the Thrift server runs on by changing the default value of 3306.

The HCatalog command line interface (CLI) can now be invoked as *root/bin/hcat*.