Installing HCatalog

Table of contents

1 Server Installation

Prerequisites

- Machine on which the server can be installed - this should have access to the hadoop cluster and to a mysql db
- MySQL db
- Hadoop cluster
- In a secure environment, Unix user that the server will run as, and an associated kerberos service principal and keytabs.
- *hcatalog*. & *hcatalog-server* .rpm packages.

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

Thrift Server Install

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

RPM installation will create a headless user named "hcat" on the server machine if it doesn't exist. Server will run as this user.

Download the MySQL Java connector libraries on *hcatsvr.acme.com*. in some directory. We will refer to this directory as dbroot. You can obtain these from [MySQL's download site](#).

If using an rpm, install appropriate rpms:

```
rpm -ivh hcatalog-version.rpm hcatalog-server-version.rpm
```

Database Setup

Select a machine to install the database on. This need not be the same machine as the Thrift server. 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 -h hcatdb.acme.com -p

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;
```

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

Thrift server config

Now you need to edit your `/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	*dbpassword* 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 *SVRHOST* with the name of the machine you are installing the Thrift server on.
hive.metastore.sasl.enabled	Set to false by default. Set to true if its a secure environment.
hive.metastore.kerberos.keytab.file	The path to the Kerberos keytab file containg the metastore thrift server's service principal. Need to set only in secure enviroment.
hive.metastore.kerberos.principal	The service principal for the metastore thrift server. You can reference your host as _HOST and it will be replaced with actual hostname. Need to set only in secure environment.

You can now procede to starting the server.

2 Starting the Server

```
sudo service start hcatalog-server
```

---

3 Logging

Server activity logs and gc logs are located in `/var/log/hcat_server`. Logging configuration is located at `/etc/hcatalog/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

```
sudo service stop hcatalog-server
```

5 Client Install

```
rpm -ivh hcatalog-version.rpm
```

Now you need to edit your `/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 common 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 wish your Thrift server to use by replacing *SVRHOST* with the name of the machine you are installing the Thrift server on.
hive.metastore.sasl.enabled	Set to false by default. Set to true if its a secure environment.
hive.metastore.kerberos.principal	The service principal for the metastore thrift server. You can reference your host as _HOST and it will be replaced with actual hostname. Need to set only in secure environment.

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