Apache Hive

Functional Area: High Level( Integrated) Systems for data warehousing

Overview: Hive is a data warehousing infrastructure based on Hadoop. Hive provides a simple query language, Hive QL, which is based on SQL and which enables user easy data summarization, ad-hoc querying and analysis of large volumes of data. In addition Hive also support plug-ins to custom mappers and reducers to do more sophisticated analysis

Data Units:

The Hive data is organized into

1. Databases
2. Tables
3. Partition
4. Buckets( or clusters)

**Apache Hive Installation on Ubuntu:**

**Installation Prerequisite:**

- Install Java and Hadoop on the machine

- Created hduser as a dedicated hadoop system user.

**Installation**

-- Installed hadoop in /home/hduser/hadoop folder. In the tutorial I would be installing hive in /usr/lib/hive folder.

-- download hive stable version from this link.  
http://mirror.tcpdiag.net/apache/hive/stable/  
  
-- enter into the directory where the stable version is downloaded. The file by default would be downloaded in the downloads directory.

*$ cd ~/Downloads*

-- Unzip the tar file

*# tar xzf apache-hive-0.13.1-bin.tar.gz*  
  
-- Create directory

*#mkdir /usr/lib/hive*

*-- Move hive-0.13.0 to hive*

*# mv hive-0.13.0 /usr/lib/hive/hive-0.13.0  
(Exit from root to hduser by using command: su hduser or exit)*

-- Set the HIVE\_HOME path in bashrc file  
 To open the bashrc file use this command

*hduser@system\_name:$ gedit ~/.bashrc*

-- In the bashrc file add the following lines

e*xport HIVE\_HOME=/usr/lib/hive/hive-0.10.0*

*export PATH=$PATH:$HIVE\_HOME/bin  
--* We have installed hive, type hive in the command line and you can see the hive shell

$ hive

hive>

Links and Resources:

<https://cwiki.apache.org/confluence/display/Hive/Tutorial>

Hadoop tutorial: Apche Hive video  
<https://www.youtube.com/watch?v=Pn7Sp2-hUXE>

http://www.orzota.com/hive-tutorial-for-beginners/  
http://www.slideshare.net/martyhall/hadoop-tutorial-hive