**Week 4 – Install Hive and Test**

**Assumptions**

1. Hadoop HDFS stand-alone system built and running
2. Remote ssh access is available to the system
3. JPS shows hadoop services are running
4. Hadoop user is logged into the Hadoop/HDFS virtual machine

**Introduction**

Apache Hive is a framework for warehousing data in the Hadoop ecosystem. It is another tool for managing and understanding huge volumes of data. It was specifically designed for analysts with strong SQL skills but poor Java programming skills. Like SQL, it is not ideal for every data science scenario. It is not optimal for machine-learning algorithms. But, it is good for less complex analysis of extremely large data sets. Finally, because SQL is ubiquitous within businesses, Hive can integrate well into business processes and existing data (White, 2015).

**Installation**

Hive is available from <http://hive.apache.org/downloads.html>

Execute the following commands -

* wget <http://apache.arvixe.com/hive/hive-1.2.1/apache-hive-1.2.1-bin.tar.gz>
* tar xzvf apache-hive-1.2.1-bin.tar.gz
* nano .bashrc (vi or vim are also acceptable alternatives to file editors on linux)
  + Add/update the following lines to your bash profile script

export HIVE\_HOME=/home/hadoop/apache-hive-1.2.1-bin

export PATH=$PATH:$HIVE\_HOME/bin

* Save the file and exit to the command prompt
* source .bashrc

**Execution**

* hive

You should see a similar output to the following –

Logging initialized using configuration in jar:file:/home/hadoop/apache-hive-1.2.1-bin/lib/hive-common-1.2.1.jar!/hive-log4j.properties

hive>

**Test**

hive> SHOW TABLES;

You should see an output that lets you know that there are no tables available

OK

Time taken: 0.08 seconds, Fetched: 0 row(s)

hive>

hive> CREATE TABLE pokes (foo INT, bar STRING);

hive> CREATE TABLE invites (foo INT, bar STRING) PARTITIONED BY (ds STRING);

hive> SHOW TABLES;

OK

invites

pokes

Time taken: 0.08 seconds, Fetched: 2 row(s)

hive>

hive> DROP TABLE pokes;

hive> SHOW TABLES;

OK

invites

Time taken: 0.08 seconds, Fetched: 1 row(s)

hive>

**Additional Tests**

Please refer to the apache Hive Getting Started guide here for more tests.

<https://cwiki.apache.org/confluence/display/Hive/GettingStarted#GettingStarted-CreatingHiveTables>