See [https://hive.apache.org](https://hive.apache.org/)

Paul posted an example of how he got it to run.  Take a look at it if you need help.  Thanks again Paul.

Here is a helpful tutorial on using and installing hive.

<http://www.dezyre.com/hadoop-tutorial/hive-tutorial>

Hey everybody,

I got a little carried away. I installed Hive based on Tom White's book "Hadoop: The Definitive Guide" and some online research. I think I am not the only one who has that book. It's pretty good, but Hadoop has changed. You can't use it as a cookbook.

One mistake I had to troubleshoot is that you need HDFS running when you start Hive. Neither White nor the Apache website I looked at explicitly state that. You get the ConnectionRefused error if you don't. Because Hive has the same HDFS://localhost:9000 property that HBase does, there has to be a service responding on port 9000.

The only other thing I did wrong was to not recognize the Hive prompt. It starts a new line with  
hive>. This looks like a contination of whatever is on the line above it. So, I restarted a working VM a few times.

Hadoop Hive

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).

Hive is available from <http://hive.apache.org/downloads.html>. The latest release is Hive 1.2.1. Its archive is an 8.31 MB file and it downloaded in 20 seconds. Installation also requires adding the hive path variables to .bashrc:

[hadoop@localhost ~]$ wget <http://apache.claz.org/hive/hive-1.2.1/apache-hive-1.2.1-bin.tar.gz>

[hadoop@localhost ~]$ tar xzvf apache-hive-1.2.1-bin.tar.gz

[hadoop@localhost ~]$ nano .bashrc

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

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

[hadoop@localhost ~]$ source .bashrc

[hadoop@localhost bin]$ start-dfs.sh

Starting namenodes on [localhost]

localhost: starting namenode, logging to /home/hadoop/hadoop/logs/hadoop-hadoop-namenode-localhost.localdomain.out

localhost: starting datanode, logging to /home/hadoop/hadoop/logs/hadoop-hadoop-datanode-localhost.localdomain.out

Starting secondary namenodes [0.0.0.0]

0.0.0.0: starting secondarynamenode, logging to /home/hadoop/hadoop/logs/hadoop-hadoop-secondarynamenode-localhost.localdomain.out

[hadoop@localhost bin]$ jps

3714 NameNode

4117 Jps

4008 SecondaryNameNode

[hadoop@localhost bin]$ hive

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> SHOW TABLES;

OK

Time taken: 0.345 seconds

hive> quit;

[hadoop@localhost bin]$