Download and install oracle vm and set-up cloudera vm.

**Prerequisites**

- Install oracle virtual box. (https://www.virtualbox.org/)

- Install cloudera virtual machine into the local. (https://www.cloudera.com/)

**Verify database**

Command: show databases;

**Dive into database created using command:**

use database\_name

**Verify tables in the database using show tables;**

select \* from customer\_orders

calculate the maximum cost for each category using group by clause on the loaded data using command:

select category, max(cost) from customer\_orders group by category;

Let's try to execute the above command in both hive and impala and calculate the execution time using Hue.

As we all know Impala sit's directly on the hdfs whereas hive sits on the hdfs -> map-reducer. So, the execution will be faster using impala when compared to the hive.