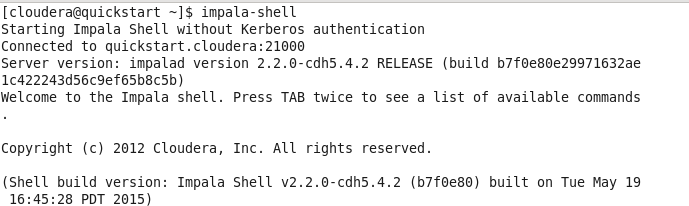
**Aim** : Commands in impala

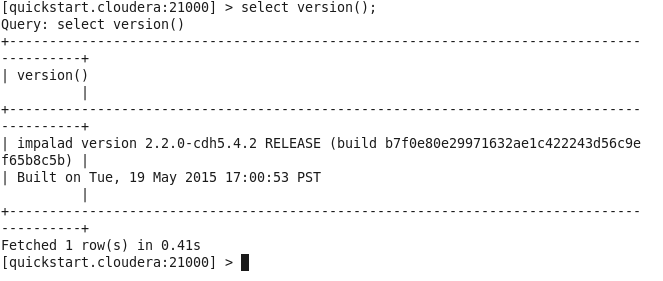
Impala is a MPP (Massive Parallel Processing) SQL query engine for processing huge volumes of data that is stored in Hadoop cluster. It is an open source software which is written in C++ and Java. It provides high performance and low latency compared to other SQL engines for Hadoop.

In other words, Impala is the highest performing SQL engine (giving RDBMS-like experience) which provides the fastest way to access data that is stored in Hadoop Distributed File System.

**impala-shell**

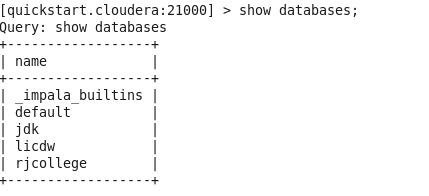


**select version();**

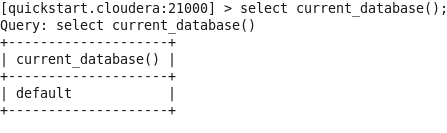


**Show databases:**This command will list existing databases in impala.

**show databases;**

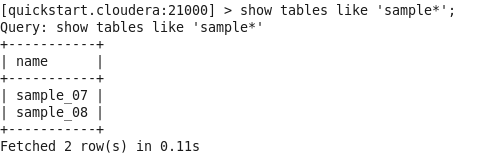


**select current\_database();**

  
  
**show tables**;

  
  
**show tables in jdk;**



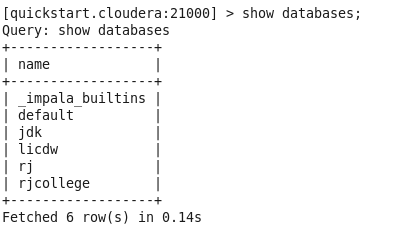
**show tables like 'sample\*';**  


**Create a database:**This command is used to create a database.

**Syntax:** create database <database\_name>

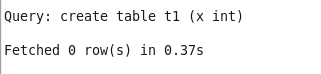
**create database RJ;**

  
  
**show databases;**



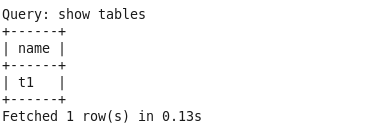
**use RJ;**  

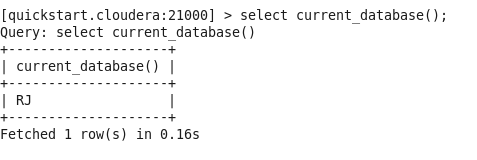

**Create table:**This command is used to create tables.

**Syntax:** create table <table\_name>(field\_name1 data\_type, field\_name2 data\_type);  
**create table t1 (x int);**  


**Show tables is used to view all the tables in current database**

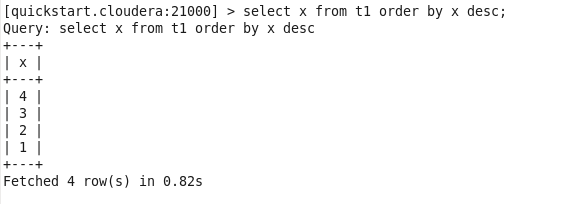
**show tables;**

  
  
**select current\_database();**



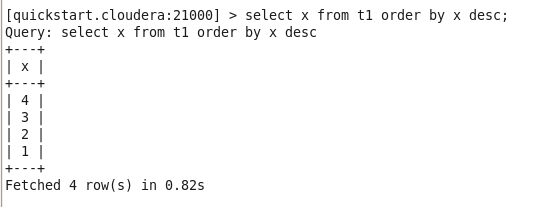
**Insert:**This command is used to insert records in either overwrite or append record in tables.

**Syntax:** Insert into <database\_name>.<table\_name> values(field\_value1, field\_value2)  
**insert into t1 values (1),(3),(2),(4);**



**Order by : This is used to sort the data**

**select x from t1 order by x desc;**



**Aggregate functions:**

**select min(x), max(x), sum(x), avg(x) from t1;**

