**Practical 10**

**Aim: Using Impala Tool**

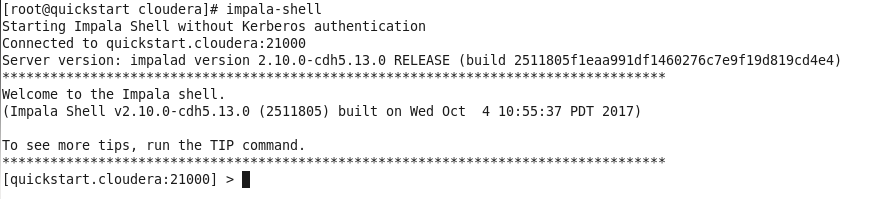
**What is Impala?**

Impala is a MPP (Massive Parallel Processing) SQL query engine for processing huge volumes of data that is stored in a 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.

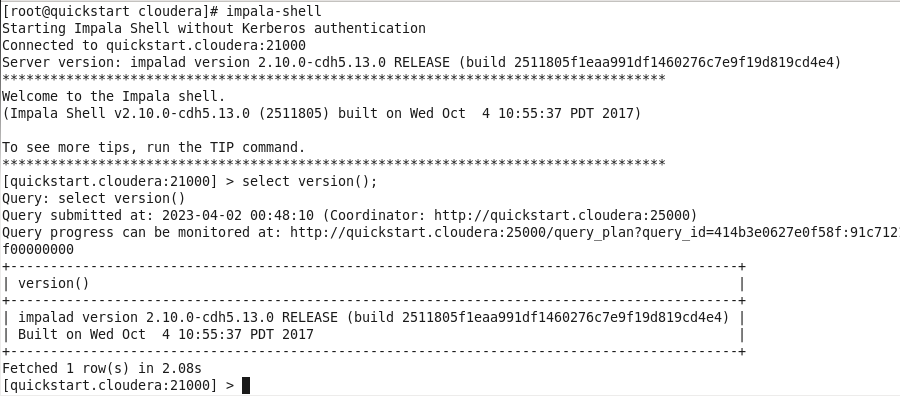
In other words, Impala is the highest performing SQL engine (giving RDBMS-like experience) which provides the fastest way to access data that is stored in a Hadoop Distributed File System.

1. Start the Impala shell

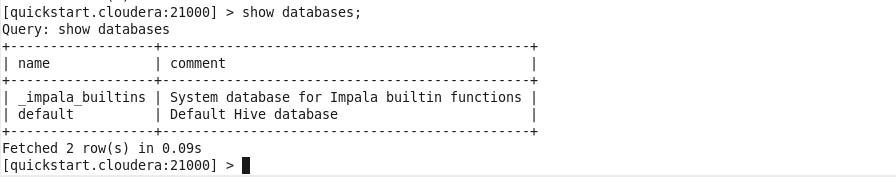
Use with --quiet to disable logs



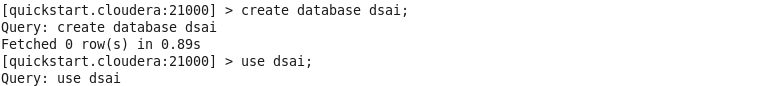
1. Display the version



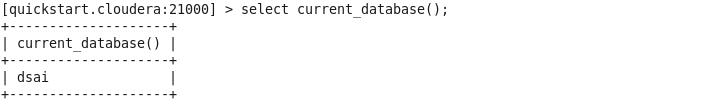
1. Show all databases



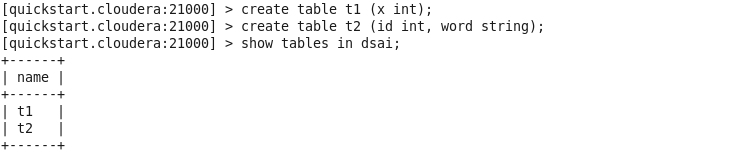
1. Creating a database and Using a database



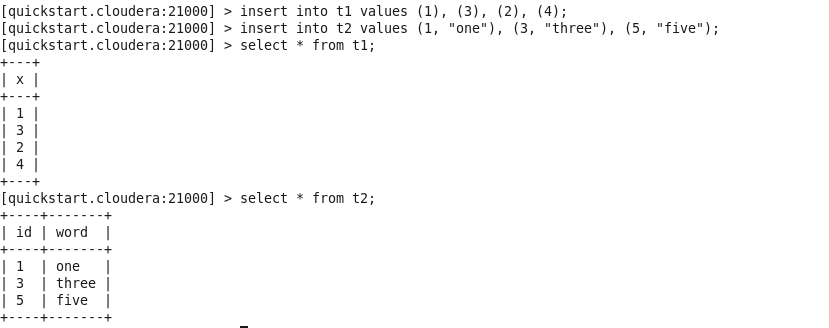
1. Display the current database



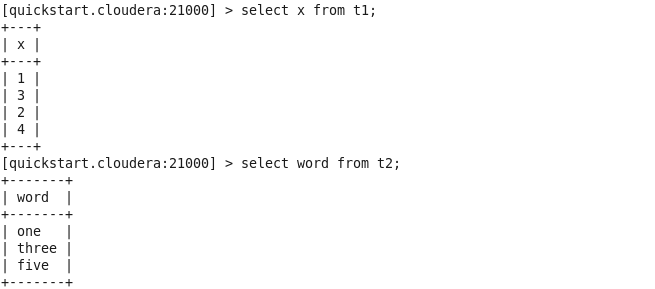
1. Creating a Table and Displaying list of tables in the database



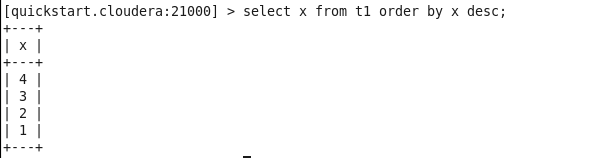
1. Adding data into tables and Displaying all data from tables



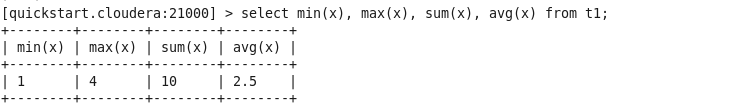
1. Display a column’s data from table



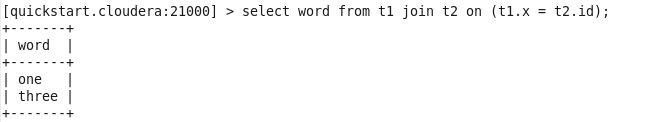
1. Using orderby



1. Using min, max, sum, average



1. Performing join



1. Creating External Table

