HADOOP ASSIGNMENT II

Name - Alvyn Abranches

Roll No - 1

PIG

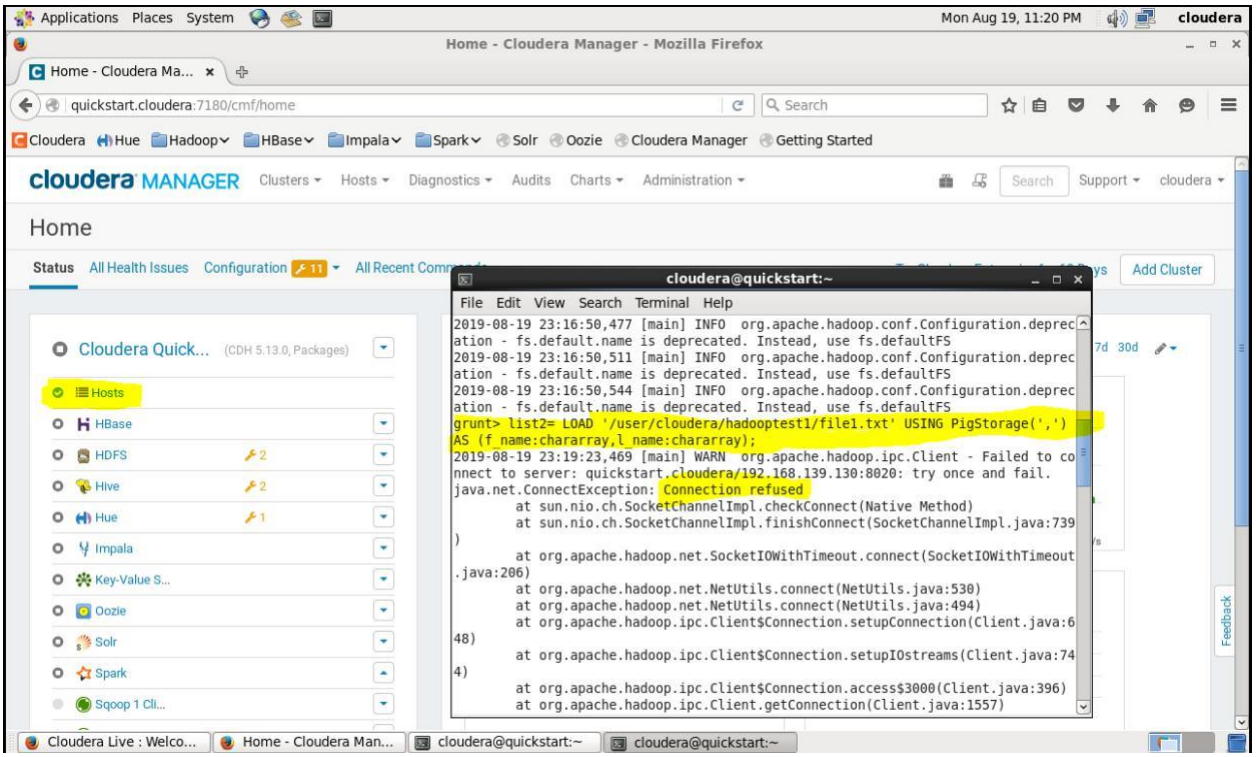
Pig is high-level programming language useful for analyzing large dataset

Pig has a nested relational model. In Pig programs are written in

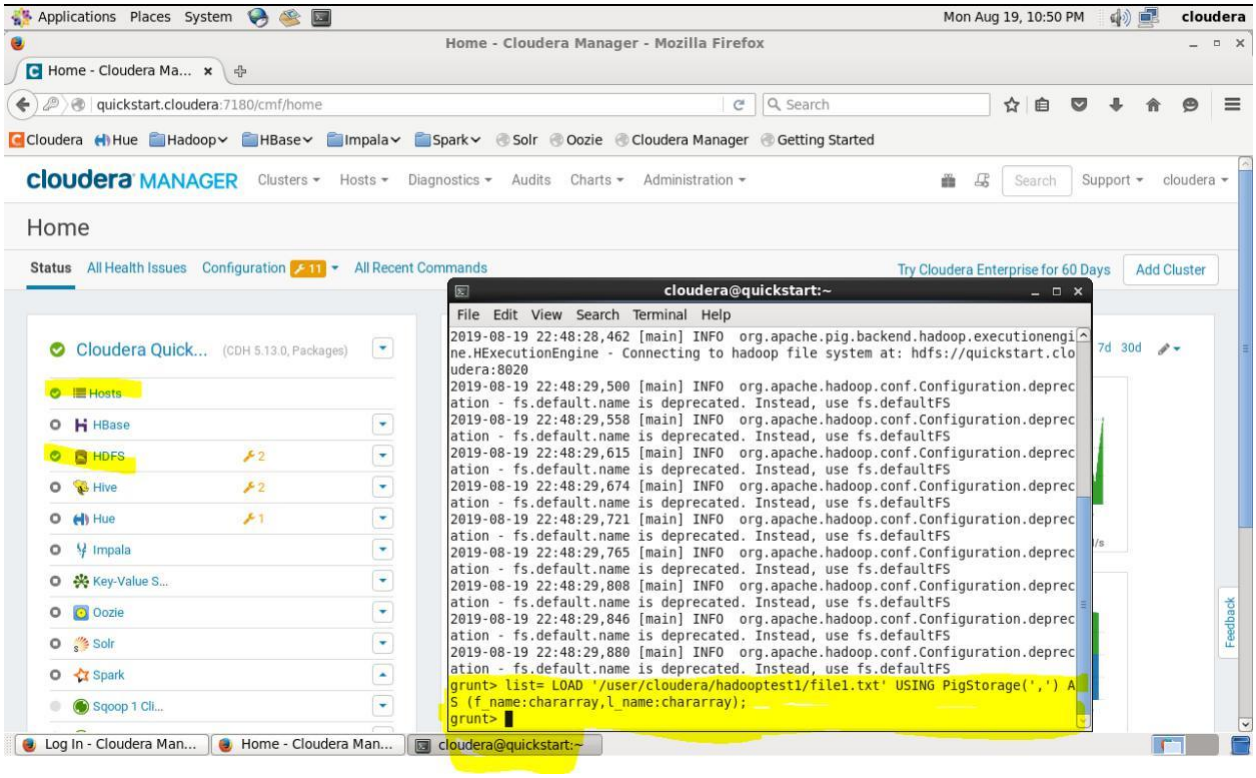
language pig latin Uses Extract Transform,Load(ETL)

Using pig Latin we can perform MapReduce task easily without writing comlex codes in java. Pig Latin is SQL-like query language

➔ When all services are stop



➔ When HDFS is start



⇨ Therefore , Pig depends on HDFS.

HIVE

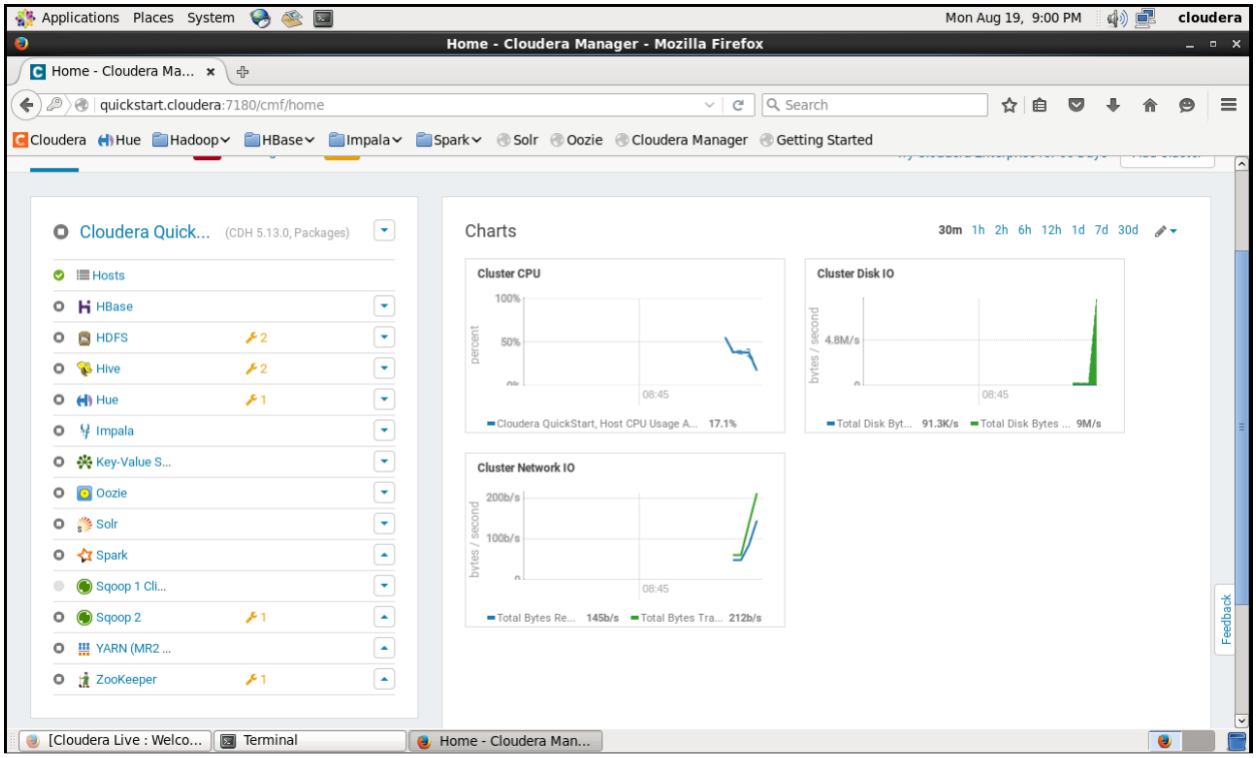
Hive is a data warehose software built on top of hadoop for providing query and analysis.

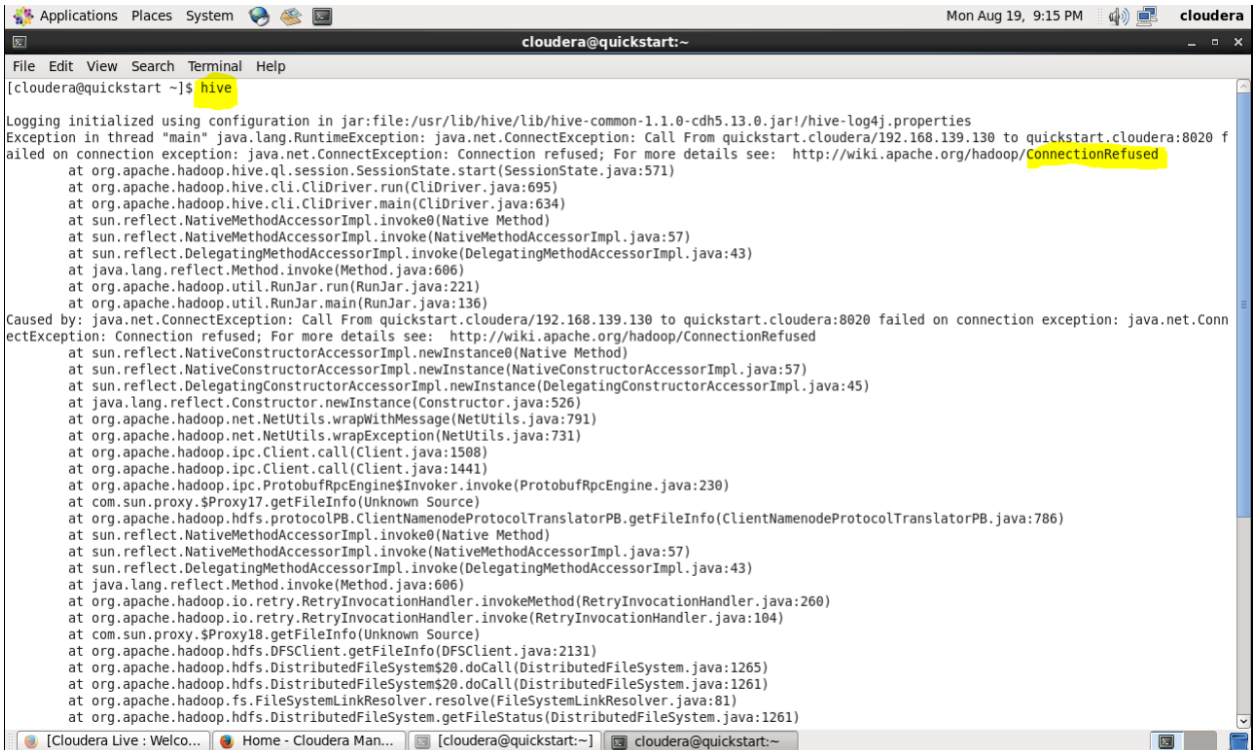
Hive provides SQL-like queries(HiveQL) which are implicitly converted intpo MapReduceor Spark jobs.

Built-in user definrd functions to manipulate dates stringsand others.

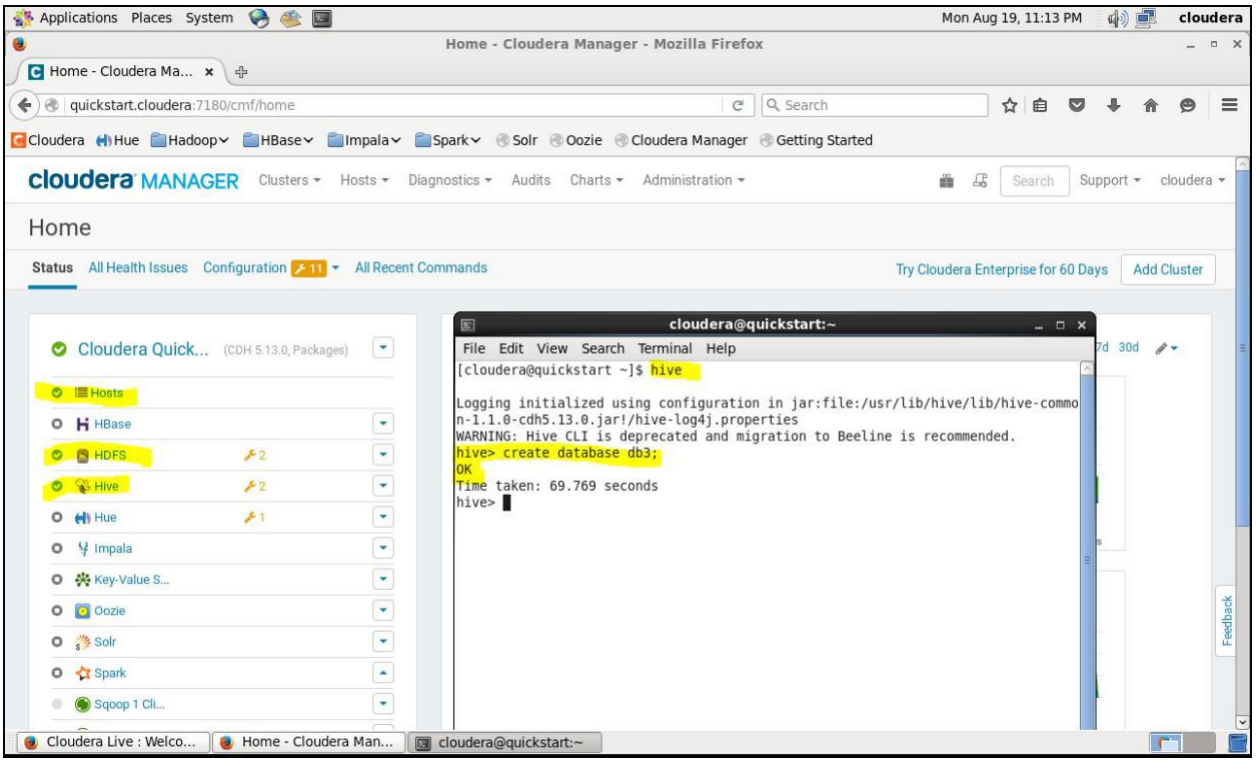
Metadata storage in a relational database management system, significantly reducing the timimg to perform semantic checks during query execution.

➔ When all services are stop





➔ When HDFS is started



⇨ Therefore, Hive is dependent on HDFS.

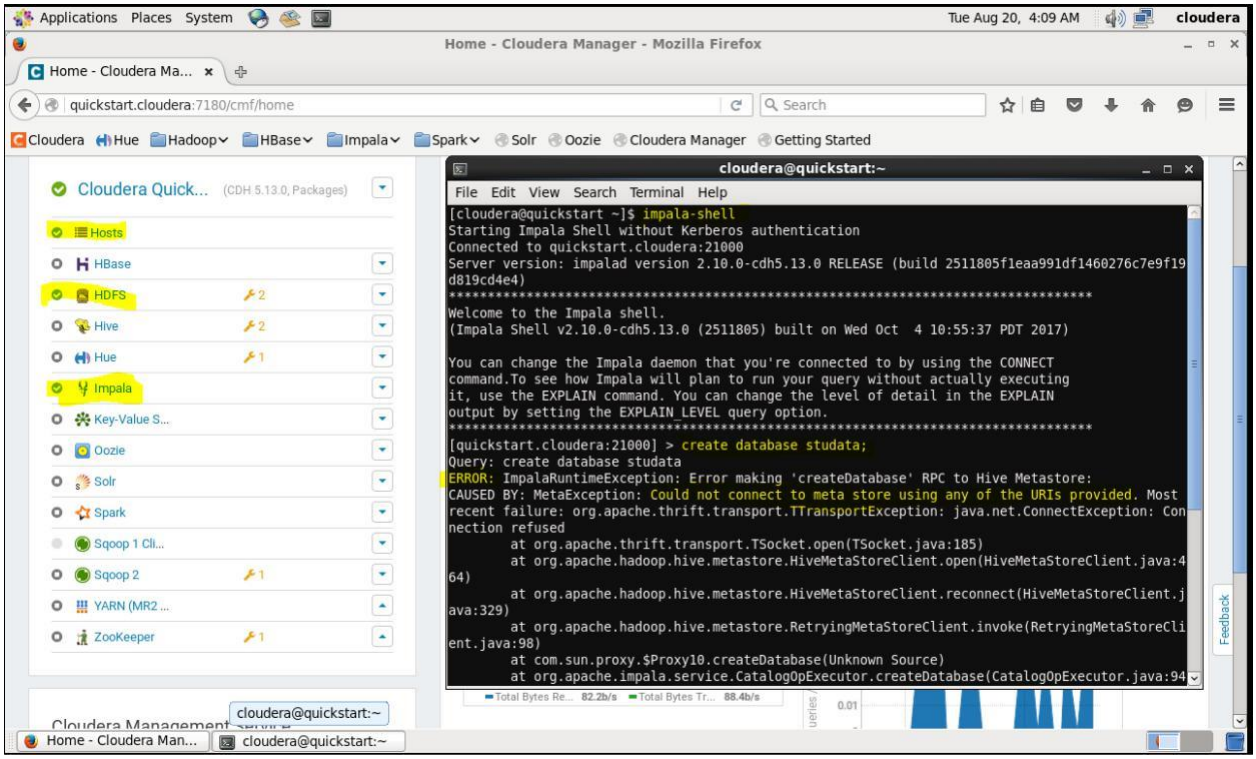
IMPALA

Impala is MPP (Massive Parallel Processing) SQL query engine for processing huge volume of data stored in stored in hadoop cluster.

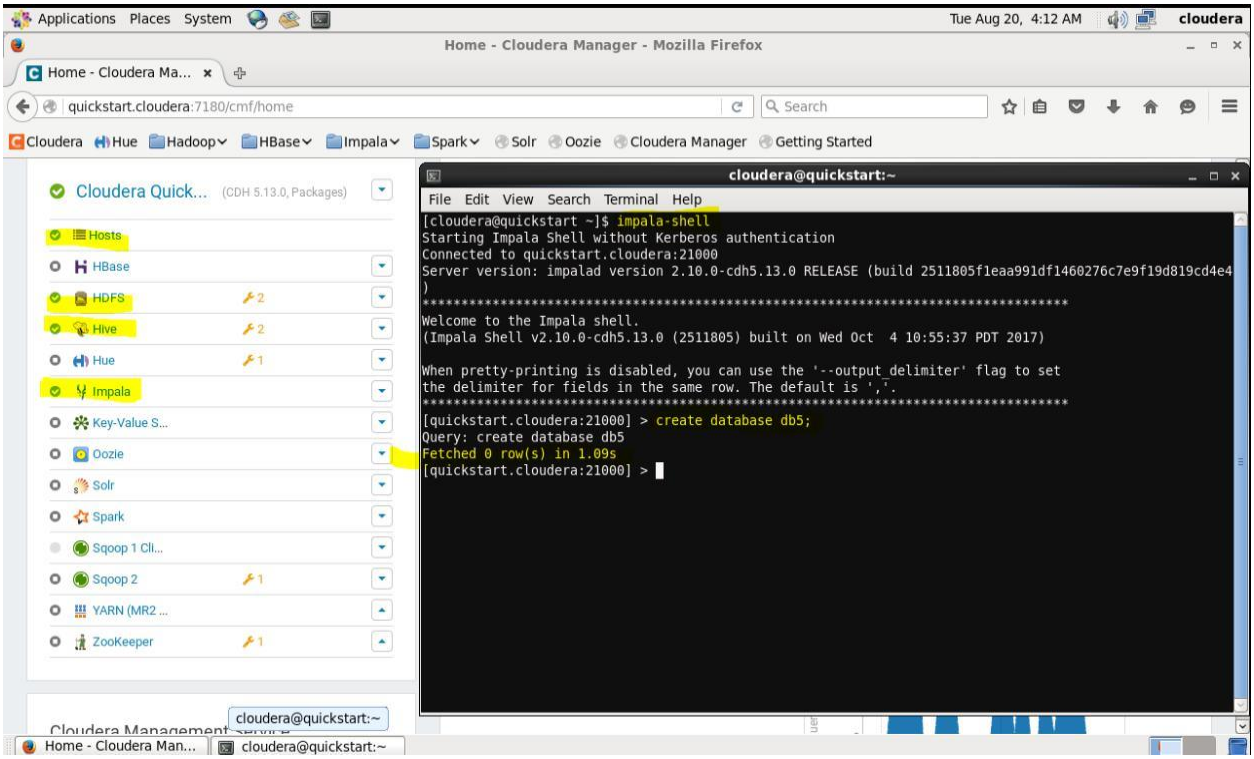
With impala ,user can communicate with HDFS or Hbase using SQL queries in faster way than sql engine like HIVE.

Impala is pioneering the use of the Parquet file format,a columnar storage layout that is optimized for large scale queries typical in data warehouse scenario.

➔ When HDFS is running but HIVE is not running



➔ When HDFS and HIVE both are running



- Therefore ,Impala depends on Hive and HDFS.

HBASE

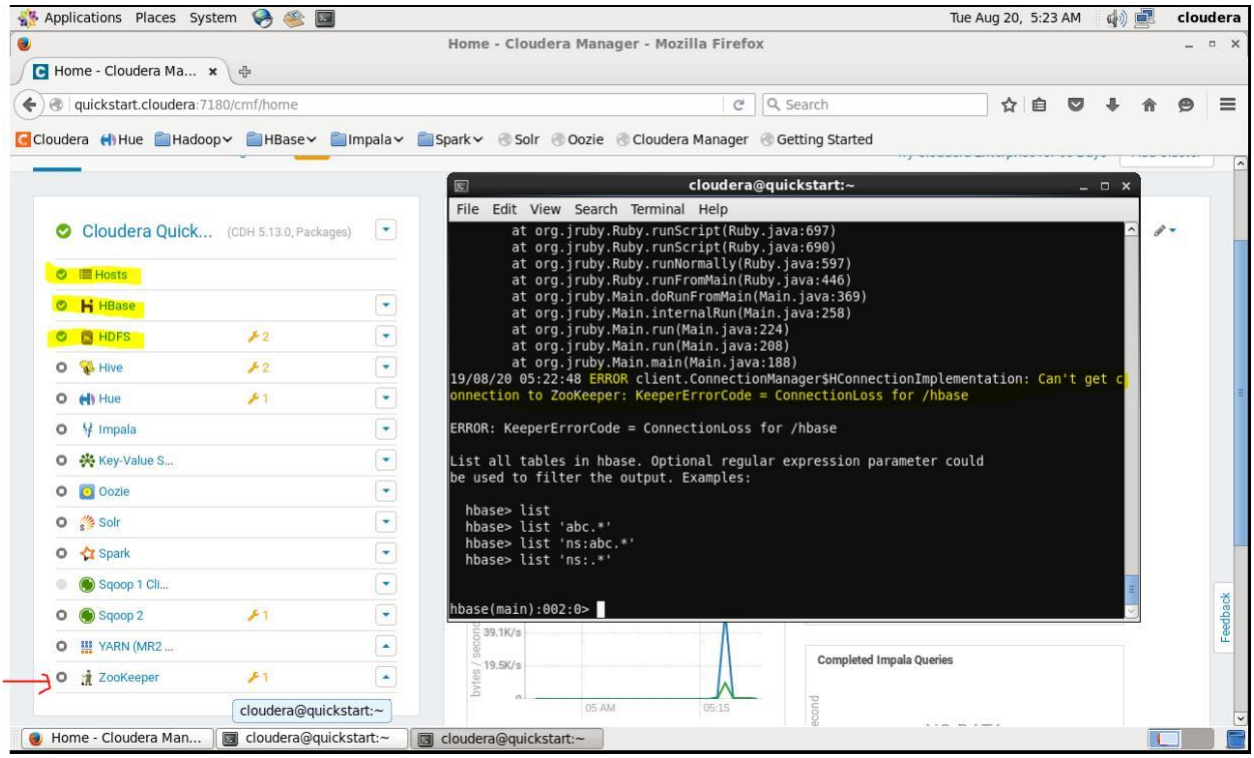
HBase is a database built on top of the HDFS.

HBase provides faster lookups for largers tables.

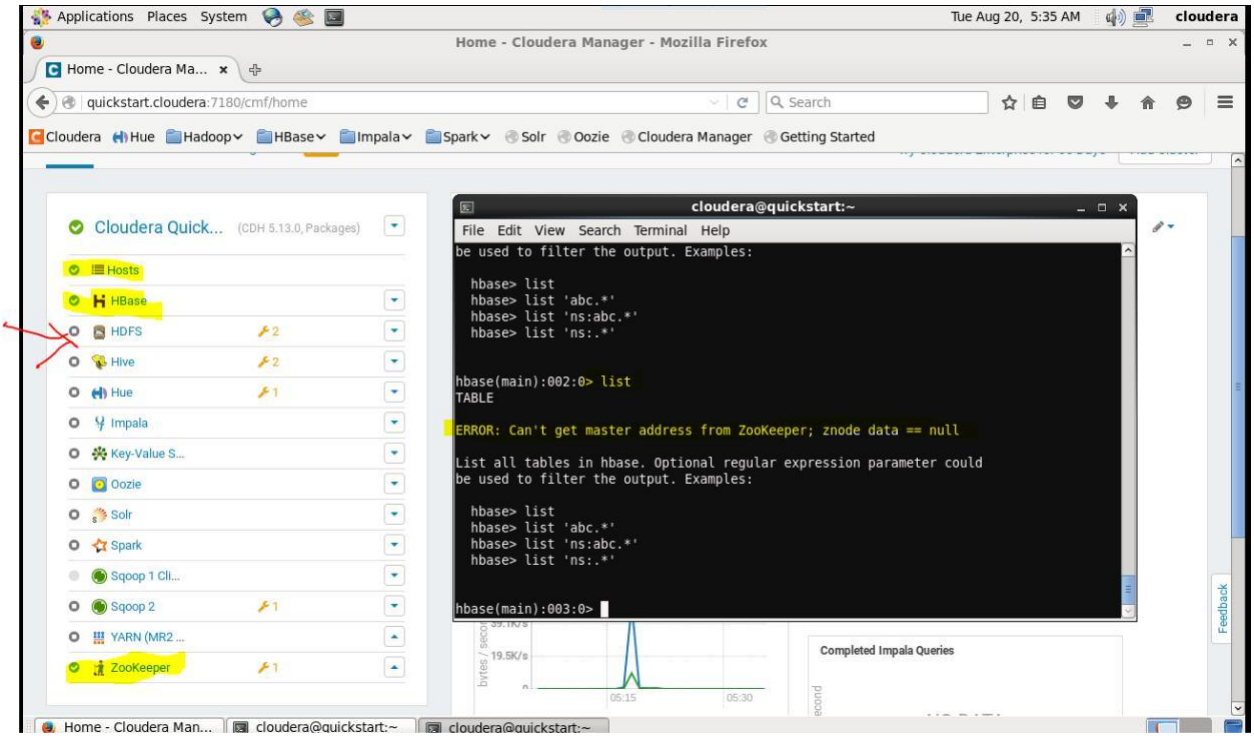
HBase is used to habe random , real-time read/write access to Big data.

HBase is a non-relational database modeled after googles bigtable.

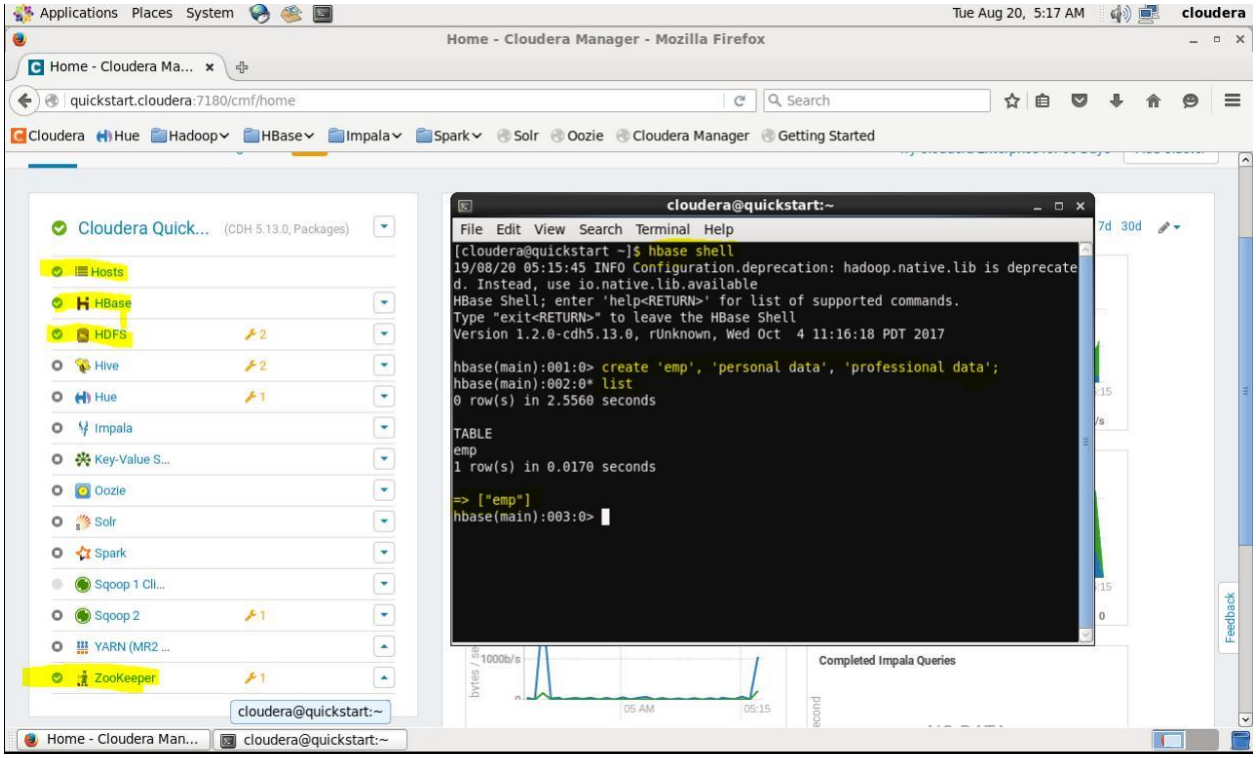
➔ When HBase , HDFS are running but ZooKeeper is not running



➔ When HBASE and ZOOKEEPER are running but HDFS is not running



➔ When HBase , HDFS and ZooKeeper are running



⇨ Therefore ,HBase depends on HDFS and Zookeeper.