Name : Prajakta Keer
Roll No : 35261

classmate
Date
Page

ASSIGNMENT 4

Aim : Write an application using Hbase and HiveQL for flight information system.

Problem Statement : Write an application using Hbase and HiveQL for flight information system

a) Creating, dropping, altering the database tables
b) Creating external Hive table to connect to the Hbase for customer information table.
c) Load table with data insert new values and fields in table, join table with Hive.
d) Create index for flight information.
e) Find the average departure delay per day in 2008.

Objectives :
- To understand various NoSQL databases.
- To understand integration of NoSQL DB with Hadoop.
- To analyze performance of distributed processing with NoSQL.

Theory : **HBase Architecture** – It has 3 main components Hmaster, Region Server, Zookeeper. A distributed data store that can be scaled horizontally to 1000 of commodity servers and prototypes of index storage.

Data Models :

1) Tables are stored by row.
2) Table's schema only defines its column families.
   - Each family consits of any no. of columns
   - Each column consits of any no. of versions
   - Columns only exist when inserted, NULLs are free

33261

- Columns within a family are sorted and stored together

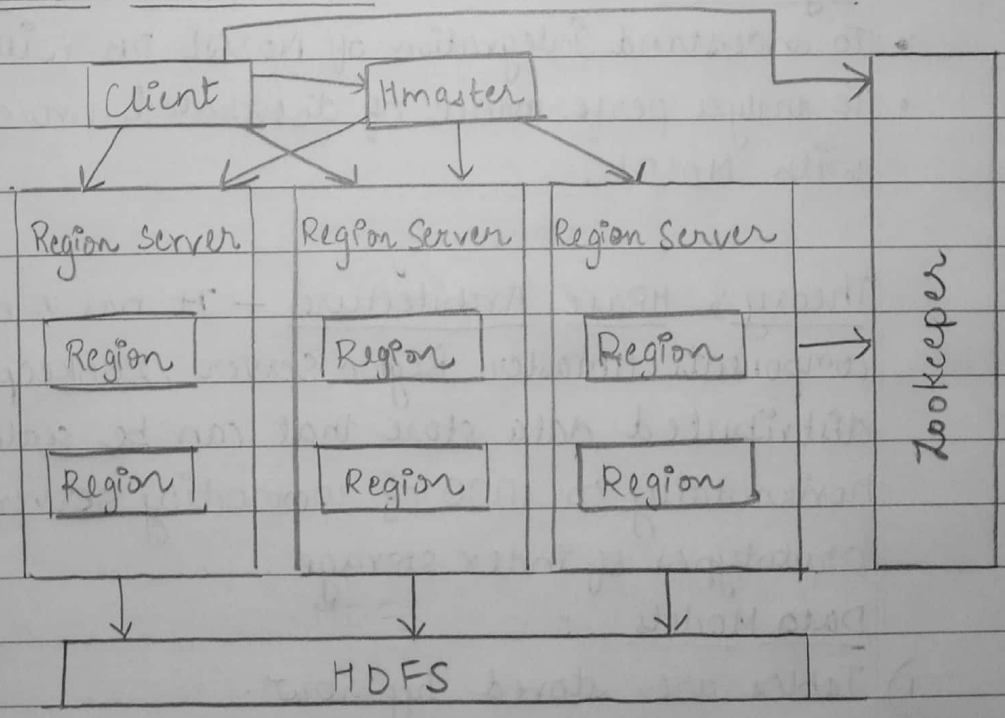
3) (Row, family, column, timestamp) → value

Members:

i) Master : • Responsible for monitoring region servers ② Load balancing for regions ③ Redirect client at region servers ④ The current SPOF.

ii) Regionserver : ① Serving requests to client ② Send heartbeat to master ③ Throughput and region members are scalable by region server.

HBase Architecture



Zookeeper – HBase depends on zookeeper and by default it manages a zookeeper instance as the authority on cluster state.

Roll No : 35251

classmate
Date
Page

Hive Architecture – It explains the flow of submission of query into Hive.
It support different type of clients such as

1) Thrift Server – It is cross language service provider platform that servers the request from all those programming language supports thrift.
2) JDBC Driver – It is used to establish a connection between Hive and Java application.
3) ODBC Driver – It allows the application that support the ODBC protocol to connect to Hive.

Conclusion : I learnt HBase and Hive architecture and implemented different queries.