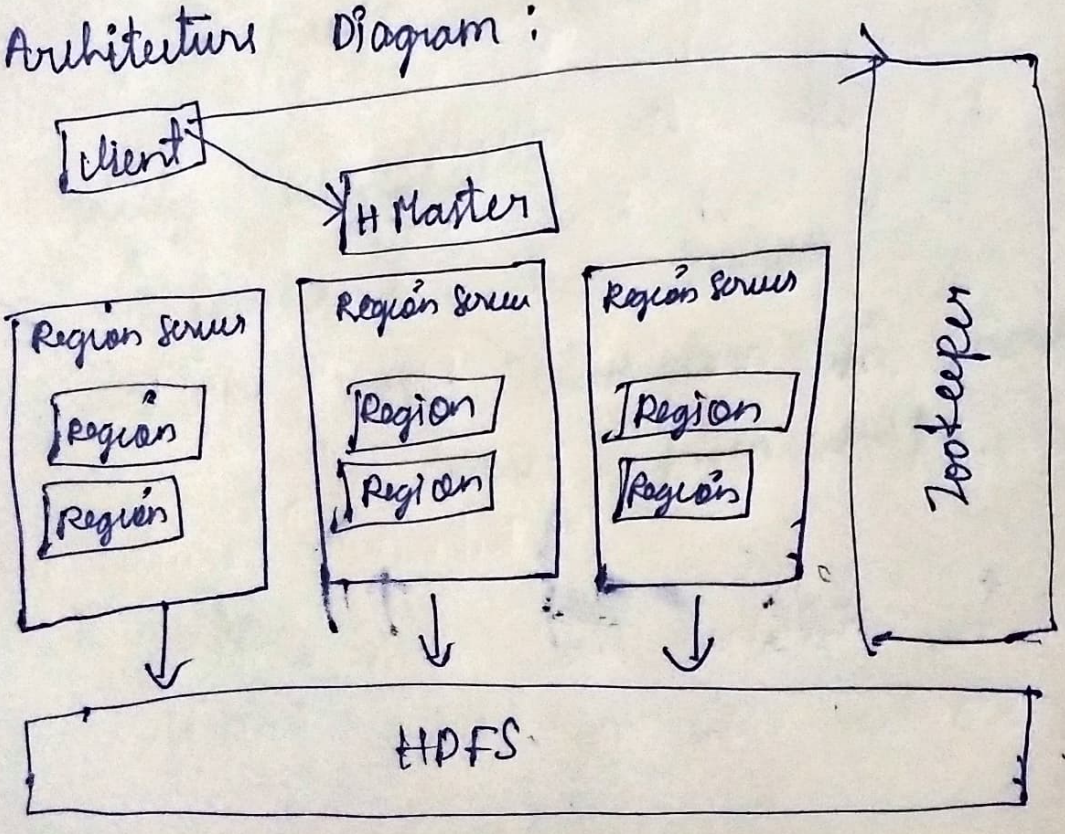
H-Base :

> It is an open source, non-relational, distributed database designed to handle large volumes of data.

> It's Build on top of the Hadoop Distributed File System (HDFS) and is modeled after Google's Big table.

> It can store massive amounts of data from terabytes to petabytes. It is column oriented & horizontally Scalable.

Architecture Diagram :



It consists of HBase 3 components:

1. HMaster:

> The implementation of Master Server in HBase is HMaster.

> It is a process in which regions are assigned to region server as well as DDL operations.

> It monitors all region server instances present in the cluster.

2. Region Server:

> HBase Tables are divided horizontally by row key range into regions.

> Regions are the basic building elements of Hbase cluster that consists of the distribution of tables.

> Region server runs on HDFS Data node which is present in Hadoop cluster. (size = 256 MB).

3. Zookeeper:

> It is like a coordinator in HBase.

> It provides services like maintaining information, naming, providing distributed synchronization, servers failure, etc...;

There are two types of data storage mediums:

1. Row-Oriented:

> The data is stored and retrieved one row at a time.

> This could lead to several problems; suppose we want only some part of the data from the row but according to this approach you have to retrieve the complete row even if you don't need it.

> Eg: MySQL, PostgreSQL.

> Less efficient in the case when we perform operations on a complex database.

2. Column-Oriented:

> The data is stored and retrieved based on the columns.

> In the column-oriented approach we can filter out the data which are required to us from the whole set of data with the help of corresponding columns.

> Here the read & write operation are slower.

> Eg : HBase, Cassandra,

> It is efficient while performing operations on the entire database.

Advantages:

* Can store large data sets
* Database can be shared
* High availability
* Cost-effective.

Dis-Advantages:

* No support SQL structure
* No transaction support
* Sorted only on key
* Memory issues on the cluster.

Applications: Real-time analytics, social media applications, OLTP, IOT applications.