**ASSIGNMENT 13.6**

**1. What kind of data Hbase can store.**

HBase is a column-oriented database that’s an open-source implementation of Google’s Big Table storage architecture.

It can manage structured and semi-structured data and has some built-in features such as scalability, versioning, compression and garbage collection.

 In a row-oriented data store, a row is a unit of data that is read or written together. In a column-oriented data store, the data in a column is stored together and hence quickly retrieved.

**2. Explain the important terms in hbase architecture.**

The important COMPONENTS in hbase architecture:

**HMaster**

**HRegionServer**

The HMaster in the HBase is responsible for

* Performing Administration
* Managing and Monitoring the Cluster
* Assigning Regions to the Region Servers
* Controlling the Load Balancing and Failover

The HRegionServer perform the following work

* Hosting and managing Regions
* Splitting the Regions automatically
* Handling the read/write requests
* Communicating with the Clients directly

**3. What is zookeeper? What happens if the zookeeper service is stopped?**

ZooKeeper is a centralized service for maintaining configuration information, naming, providing distributed synchronization, and providing group services. All of these kinds of services are used in some form or another by distributed applications. Each time they are implemented there is a lot of work that goes into fixing the bugs and race conditions that are inevitable.

ZooKeeper is a high-performance coordination service for distributed applications(like HBase).

It exposes common services like naming, configuration management, synchronization, and group services, in a simple interface so you don't have to write them from scratch.

HBase relies completely on Zookeeper.

If the zookeeper service is stopped, Hbase will not function properly.

**4. Is it necessary that region server be located on all DataNodes?**

We don't have to because its relatively benign since HBase will have most data available on the local datanode. rows are inserted into HBase tables and are getting stored as regions in different region server. So, the region server stores the dataSimilarly in terms of Hadoop, data is stored in the data nodes present in the hadoop cluster.

**5. How will you implement joins in Hbase?**

The two primary strategies are either **denormalizing** the data upon writing to HBase, or to have **lookup tables** and do the join between HBase tables in your application or MapReduce code (and as RDBMS' demonstrate, there are several strategies for this depending on the size of the tables, e.g., nested loops vs. hash-joins).