**ASSIGNMENT 31.1**

1. **Differences between HBASE and HDFS.**

* HDFS is a distributed file system that is suitable for storing large files whereas HBase is a database built on the top of HDFS.
* HDFS doesn’t support fast individual record lookups while hbase provides fast lookups.
* HDFS provides high latency batch processing and hbase provides low latency access to single rows from billions of records
* Hdfs provides only sequential access of data while hbase internally uses hash tables and provides random access.

1. **List and explain the main components of HBASE.**

HBase has 3 major components:

* **the client library**
* **master server**
* Assigns regions to the region servers and takes the help of Apache ZooKeeper for this task.
* Handles load balancing of the regions across region servers.
* It unloads the busy servers and shifts the regions to less occupied servers.
* Maintains the state of the cluster by negotiating the load balancing.
* Is responsible for schema changes and other metadata operations such as creation of tables and column families.
* **region servers**
* Handle read and write requests for all the regions under it.
* Decide the size of the region by following the region size thresholds.
* Region servers can be added or removed as per requirement.

1. **Does Hbase support sql?**

Hbase supports NOSQL and the features of the nosql are Generic data model

Heterogeneous containers, including sets, maps, and arrays Dynamic type discovery and conversion.

• Data is stored in single tables as compared to joining multiple tables