1. What kind of data Hbase can store.

* HBase is a column-oriented database and the tables in it are sorted by row and is an implementation of Google’s Big Table storage architecture.
* Apache HBase is used to have random, real-time read/write access to Big Data.
* Apache HBase is a non-relational database modeled after Google's Bigtable. Bigtable acts up on Google File System, likewise Apache HBase works on top of Hadoop and HDFS.
* It can manage structured and semi-structured data and has some built-in features such as scalability, versioning, compression and garbage collection.

1. Explain the important terms in hbase architecture.

HBase has three major components: the client library, a master server, and region servers. Region servers can be added or removed as per requirement.

## **MasterServer**

The 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.

## **Regions**

## Regions are nothing but tables that are split up and spread across the region servers.

### **Region server**

The region servers have regions that -

* Communicate with the client and handle data-related operations.
* Handle read and write requests for all the regions under it.
* Decide the size of the region by following the region size thresholds.

The store contains memory store and HFiles. Memstore is just like a cache memory. Anything that is entered into the HBase is stored here initially. Later, the data is transferred and saved in Hfiles as blocks and the memstore is flushed.

1. What is zookeeper? What happens if the zookeeper service is stopped.

* HBase uses ZooKeeper as a distributed coordination service to maintain server state in the cluster.
* Zookeeper maintains which servers are alive and available, and provides server failure notification.
* Zookeeper uses consensus to guarantee common shared state And there should be three or five machines for consensus.
* If the zookeeper service is stopped, Hbase will not function properly.

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

No,it is not necessary that region server be located on all DataNodes since HBase will have most data available on the local datanode

1. 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)