1. **What is NOSQL:**

**Answer**: NoSQL is an approach to [database](https://searchsqlserver.techtarget.com/definition/database) design that can accommodate a wide variety of data models, including key-value, document, columnar and graph formats. NoSQL, which stand for "not only [SQL](https://searchsqlserver.techtarget.com/definition/SQL)," is an alternative to traditional relational databases in which data is placed in tables and data [schema](https://searchsqlserver.techtarget.com/definition/schema) is carefully designed before the database is built. NoSQL databases are especially useful for working with large sets of distributed data.

1. **How does Data get stored in NOSQL DB?**

**Answer**: Every piece of data is stored in a JSON format.

1. **What is column family in HBASE?**

**Answer**:

Column families are the base storage mechanism in HBase.   An HBase table is comprised of one or more column families, each of which is stored in a separate set of region files sharing a common key.

1. **How many Max no.of columns can be added to HBase Table?**

**Answer**:

There is a limit to the number of column families in HBase. There is one Mem Store (It’s a write cache which stores new data before writing it into Hfiles) per Column Family, when one is full, they all flush.

The more you add column families there will be more Memstore created and Memstore flush will be more frequent. It will degrade the performance.

1. **Why columns are not defined at the time of table creation in HBase?**
2. **How does data get managed in HBase?**
3. **What happens internally when new data gets inserted into HBase table?**

**Advanced HBase:**

**Task 1:**

1. **NOSQL Database:**

**Answer**: NoSQL is an approach to [database](https://searchsqlserver.techtarget.com/definition/database) design that can accommodate a wide variety of data models, including key-value, document, columnar and graph formats. NoSQL, which stand for "not only [SQL](https://searchsqlserver.techtarget.com/definition/SQL)," is an alternative to traditional relational databases in which data is placed in tables and data [schema](https://searchsqlserver.techtarget.com/definition/schema) is carefully designed before the database is built. NoSQL databases are especially useful for working with large sets of distributed data.

1. **TYPES OF NOSQL Databases:**

There are 4 basic types of NoSQL databases:

1. Key-Value Store – It has a Big Hash Table of keys & values {Example- Riak, Amazon S3 (Dynamo)}
2. Document-based Store- It stores documents made up of tagged elements. {Example- CouchDB}
3. Column-based Store- Each storage block contains data from only one column, {Example- HBase, Cassandra}
4. Graph-based-A network database that uses edges and nodes to represent and store data. {Example- Neo4J}

# 3. CAP theorem

* Consistency: every read would get you the most recent write
* Availability: every node (if not failed) always executes queries
* Partition-tolerance: even if the connections between nodes are down, the other two (A & C) promises, are kept.

4.HBase – Architecture

1. In HBase, tables are split into regions and are served by the region servers. Regions are vertically divided by column families into “Stores”. Stores are saved as files in HDFS.
2. HBase has three major components: the client library, a master server, and region servers. Region servers can be added or removed as per requirement.

## **Master Server**

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.

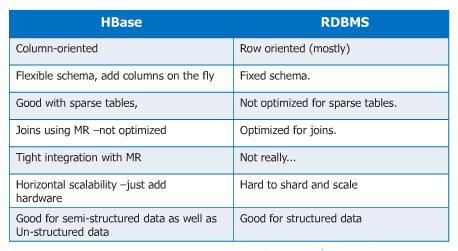
When we take a deeper look into the region server, it contain regions and stores as shown below:



## **Zookeeper**

* Zookeeper is an open-source project that provides services like maintaining configuration information, naming, providing distributed synchronization, etc.
* Zookeeper has ephemeral nodes representing different region servers. Master servers use these nodes to discover available servers.
* In addition to availability, the nodes are also used to track server failures or network partitions.
* Clients communicate with region servers via zookeeper.
* In pseudo and standalone modes, HBase itself will take care of zookeeper.

5.HBase vs RDBMS



# 