Explain in brief

● When should we use HBASE, list some of the scenarios for the same in real time.

● What are the different modes in which Hbase can be run?

● Need and working of zookeeper in Hbase?

**When should we use HBASE, list some of the scenarios for the same in real time.**

Following are some of the key areas to be considered before finalizing HBase for your application.

**Data volume:** The volume of data is the most common point to be considered. You should have peta bytes of data to be processed in a distributed environment. Otherwise, for a small amount of data, it will be stored and processed in a single node, keeping other nodes idle. So, it will be a misuse of technology framework.

**Application Types:** HBase is not suitable for transactional applications, large volume MapReduce jobs, relational analytics, etc. It is preferred when you have a variable schema with slightly different rows. It is also suitable when you are going for a key dependent access to your stored data.

**Hardware environment:** HBase runs on top of HDFS. And HDFS works efficiently with a large number of nodes (minimum 5). So, if you have good hardware support, then HBase can be a good selection.

**No requirement of relational features:** Your application should not have any requirement for RDBMS features like transaction, triggers, complex query, complex joins etc. If you can build your application without these features, then go for HBase.

**Quick access to data:**If you need a random and real time access to your data, then HBase is a suitable candidate. It is also a perfect fit for storing large tables with multi structured data. It gives ‘flashback’ support to queries, which makes it more suitable for fetching data in a particular instance of time.

**What are the different modes in which Hbase can be run?**

There are three modes in which Hbase can run. They are:

1. Standalone mode
2. Pseudo distributed mode
3. Fully distributed mode

**Need and working of zookeeper in Hbase:**

HBase distributed application installation depends on a running ZooKeeper cluster. Apache HBase uses ZooKeeper to track the status of distributed data throughout the master and region servers with the help of **centralized configuration management** and **distributed mutex** mechanisms.

* 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.