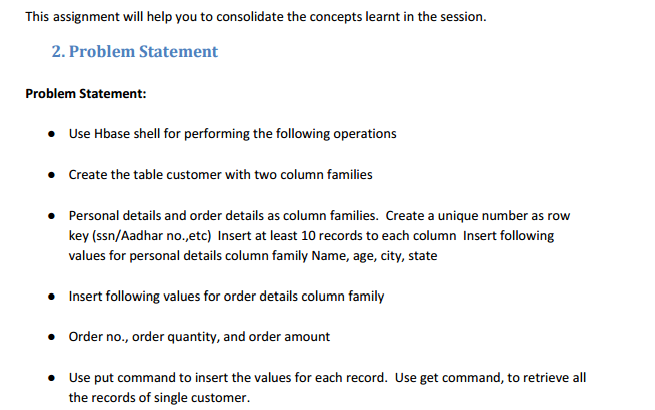
Assignment-31.2:



**When should we use HBASE, list some of the scenarios for the same in realtime**

* NoSQL DB, HBase were provided with good functionalities.
* but it is still not a ‘Fit for All’ solution.
* The key areas provided before finalizing HBase for application are.

***1. Data volume:***

* Data volume is the most common point to be considered.
* One should have peta bytes of data which need to be processed in a distributed environment.
* For a small amount of data storage and processing will happen inside a single node.
* Mean while it will keep other nodes idle.
* There might be misuse of technology framework.

**2. Application Types:**

* HBase is not best suited for transactional applications for large volume like MapReduce jobs, relational analytics, etc.
* when one is provded with a variable schema with different rows.
* It will support a key dependent access to your stored data.

**3. Hardware environment:**

* HBase will run on top of HDFS.
* HDFS will work efficiently using a large number of nodes
* And a minimum of 5 nodes were needed.
* When there is a good hardware support, then HBase will be a good selection.

**4. No requirement of relational features:**

* All the application should not be in any need for RDBMS features like transaction, triggers, complex query, complex joins etc.
* When one is capable of building their own application without using any of these features, then one may go for HBase.

**5. Quick access to data:**

* When one needs real time data access then HBase will be best suited.
* A large tables can be stored with multi structured data.
* ‘flashback’ support will be provided to the queries.
* So it will be suitable for data fetching for particular instance of time.

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

HBase will be of two run modes:

* 1. **Standalone HBase**

### Distributed

**1. Standalone mode:**

* Standalone mode will act as a default mode.
* “Quick Start - Standalone HBase” section.
* In standalone mode, HBase will not go for HDFS – rather will use the local filesystem instead
* all HBase daemons and a local ZooKeeper will run in same JVM.
* Zookeeper gather some well known port so clients may have frequent conversation with the HBase.

**2. Distributed mode:**

* Distributed mode can be subdivided into distributed
* but all the daemons will run on a single node
* pseudo-distributed
* fully-distributed
* all the daemons will spread across all the nodes inside the cluster

Distributed mode can be subdivided into 2 types and they are:

1. Pseudo-distributed - where all the daemons will run on a single node.
2. Fully-distributed - where all the daemons will be spread across all nodes inside the cluster.

**2.1 Pseudo-distributed mode:**

* pseudo-distributed mode - fully-distributed mode which will run on a single host.
* One can use this configuration for testing and prototyping on HBase.
* One cannot use this configuration for production or for evaluating the HBase performance.

**2.2 Fully-distributed mode:**

* In case of production environment the distributed mode will be more appropriate.
* Because distributed mode will have multiple instances of HBase daemons which will run on a multiple servers inside the cluster.

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

### Distributed HBase will depend on Zookeeper for the purpose of cluster configuration as well management.

### Apache HBase, ZooKeeper coordinates, communicates, and shares state will be between Master and Region Server.

### HBase will design policy using ZooKeeper only for transient data that is, for coordination and state communication.

### If the HBase’s ZooKeeper data is removed, only the transient operations will get affected

### data can be written and read to/from the HBase.

* Distributed Apache HBase (TM) installation will depend upon running the ZooKeeper cluster.
* Participating nodes and clients need to able to access the running ZooKeeper ensemble.
* Apache HBase by default will manage ZooKeeper "cluster".
* It will start and stop ZooKeeper ensemble as part of HBase start/stop process.
* This variable,by defaults be true,and will tell the HBase whether to start/stop the ZooKeeper ensemble servers as part of HBase start/stop.