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

1. If your application has a variable schema where each row is slightly different
2. If you find that your data is stored in collections, that is all keyed on the same value, meta data, message data or binary data that is all keyed on the same value
3. If you need random, real time read/write access to your Big Data.
4. If you need key based access to data when storing or retrieving.
5. If you have huge amount of data with existing Hadoop cluster.

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

HBASE can run in two modes

1. Standalone mode
2. Distributed mode
3. Pseudo Distributed mode.
4. Fully Distributed mode.

Standalone mode-

* By default Hbase runs in this mode
* This mode does not use HDFS and it uses the local system
* It runs all HBase daemons and a local Zookeeper all up in the same JVM

Distributed Mode-

Pseudo Distributed mode.

As pseudo means that not genuine so pseudo distributed means that it’s just imitation of the distributed mode

All daemons run on a single node. A pseudo-distributed mode is simply a distributed mode run on a single host. This is used for prototyping HBASE . And this should not be used for production and evaluating the HBASE performance.

Following are the xml configurations one has to make in order to setup HBASE in the Pseudo distributed mode.

<configuration>

...

<property>

<name>hbase.rootdir</name>

<value>hdfs://h-24-30.sfo.stumble.net:8020/hbase</value>

</property>

<property>

<name>hbase.cluster.distributed</name>

<value>true</value>

</property>

<property>

<name>hbase.zookeeper.quorum</name>

<value>h-24-30.sfo.stumble.net</value>

</property>

...

</configuration>

Fully Distributed Mode

Fully-distributed where the daemons are spread across all nodes in the cluster. This is the one which is used in the real time and for production of the Hadoop applications.

Following are the xml configurations one has to make in order to setup HBASE in the fully distributed mode.

<configuration>

...

<property>

<name>hbase.rootdir</name>

<value>hdfs://namenode.example.org:8020/hbase</value>

<description>The directory shared by RegionServers.

</description>

</property>

<property>

<name>hbase.cluster.distributed</name>

<value>true</value>

<description>The mode the cluster will be in. Possible values are

false: standalone and pseudo-distributed setups with managed Zookeeper

true: fully-distributed with unmanaged Zookeeper Quorum (see hbase-env.sh)

</description>

</property>

...

</configuration>

● Need and working of zookeeper in Hbase?

-Zookeeper is a cluster management service used for synchronization between masters & clients. Hadoop is a user of this service.

- Zookeeper maintains a live cluster state

-Hbase has go multiple masters and in order to maintain the function of the hmasters fluently zookeeper.

-The other common uses for ZooKeeper are storage of small bits of dynamic configuration data, locking, and group membership