1. Hbase can store data of semi structured type. Where the data will not have any specified schema.
2. The important terms in hbase architecture are Masterserver, Regions, RegionServer and zookeeper. MasterServer: the master server assigns regions to the region servers and takes the help of zookeeper. Handles load balancing of the regions across region servers. It unloads the busy servers and shifts the regions to less occupied servers. Regions: Regions are nothing but tables that are split up and spread across the region server 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. Zookeeper: Zookeeer is an open source project that provides service like maintaining configuration information, naming, providing distributed synchronization,etc
3. : Zookeeer is an open source project that provides service like maintaining configuration information, naming, providing distributed synchronization,etc. Cliets connot communicate with the regionServer if the zookeeper is failed.
4. No it is not necessary to have region server in all data nodes
5. Joins can be done by loading hbase data in hive or impala but it is suitable for small data set for large data set in mapreduce take Hbase table objects to take one table and by extending tablemapper use 2nd table.