ASSIGNMENT9.3

Explain the below concepts with an example in brief.

● Nosql Databases

Relational databases were not designed to cope with the scale and agility challenges that face modern applications, nor were they built to take advantage of the commodity storage and processing power available today.

NoSQL encompasses a wide variety of different database technologies that were developed in response to the demands presented in building modern applications:

Features:-

- Generic Data Model -Heterogeneous containers, including sets, maps, and arrays
- Dynamic type discovery and conversion –NoSQL analytics systems support runtime type identification and conversion so that custom business logic can be used to dictate analytic treatment of variation.
- Non-relational and De-normalised- Data is stored in single tables as compared to joining multiple tables.
- Commodity hardware - Adding more of the economical servers allows NoSQL databases to scale to handle more data.
- Highly distributable - Distributed databases can store and process a set of information on more than one device.

● Types of Nosql Databases

There have been various approaches to classify NoSQL databases, each with different categories and subcategories, some of which overlap. What follows is a basic classification by data model, with examples:

Column:Accumulo,Cassandra,Druid,HBase,Vertica,SAP HANA

Document:Apache CouchDB,ArangoDB,Clusterpoint,Couchbase,Cosmos DB,HyperDex,IBM Domino,MarkLogic,MongoDB,OrientDB,Qizx,RethinkDB

Key-value:Aerospike,ArangoDB,Couchbase,Dynamo, FairComc-treeACE,FoundationDB,HyperDex,InfinityDB,MemcacheDB,MUMPS,Oracle NoSQL Database,OrientDB,Redis,Riak,Berkeley DB, SDBM/Flat Filedbm

Graph:AllegroGraph,ArangoDB,InfiniteGraph,Apache Giraph,MarkLogic,Neo4J,OrientDB,Virtuoso

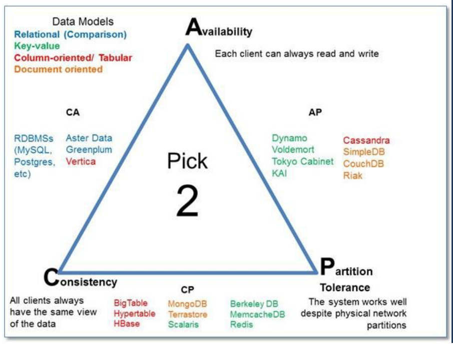
Multi-model:ArangoDB,Couchbase,FoundationDB,InfinityDB,MarkLogic,OrientDB

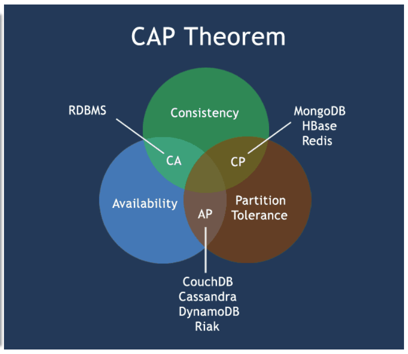
● CAP Theorem

•Consistency-This means that the data in the database remains consistent after the execution of an operation. For example after an update operation, all clients see the same data.

•Availability-This means that the system is always on (service guarantee availability), no downtime.

•Partition Tolerance-This means that the system continues to function even if the communication among the servers is unreliable, i.e. the servers may be partitioned into





● HBase Architecture

HBASE is a distributed column oriented database built on top of hadoop to provide real time access to Big Data. HBase is composed of three types of servers in a master slave type of architecture.

•Region servers serve data for reads and writes.

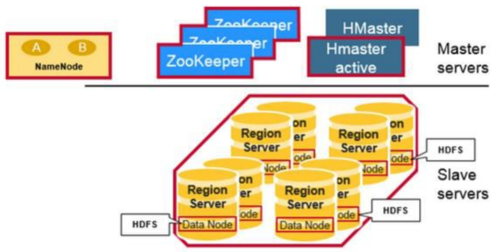
•HBase Master process handles the Region assignment, DDL (create, delete tables) operations

•Zookeeper maintains a live cluster state.

•The Hadoop Data Node stores the data that the Region Server is managing.

•All HBase data is stored in HDFS files.

•The Name Node maintains metadata information for all the physical data blocks that comprise the files.



● HBase vs RDBMS

RDBMS	HBASE
• RDBMS is row-oriented databases	HBase is a distributed, column-oriented data storage systen
• RDBMS tables have fixed-schema	Hbase tables do not have fixed-schema

• RDBMS tables guarantee ACID properties	Hbase tables guarantee consistency and partition tolerance
• RDBMS uses SQL (Structured query Langauge) to query the data	Hbase uses Java client API and Jruby