

Assignment 3.9

Explain the below concepts with an example in brief.

• Nosql Databases

NOSQL stands for Not Only SQL Database. It provides a DBMS system that does not use the conventional tabular relations used in RDBMS. It is an alternative to the RDBMS.

NoSQL is especially useful for use towards working with large datasets of distributed data. It is a complete different approach towards database design which can accommodate a wide variety of data model including columnar, key-value, graphs, documents.

It allows for great flexibility as well as operational performance.

• Types of Nosql Databases

There are mainly 4 types:

- a. **Key-Value** : Leverages hashtable key-values.
- b. **Column based** : Each store block contains data from one column.
- c. **Document based** : Documents made up of tagged elements.
- d. **Graph based** : Network based, edges and nodes to represent and store data.

• CAP Theorem

CAP theorem is basically a map of different database solutions depicting it's CAP properties that show us the tradeoff of choosing a particular database system.

CAP stands for Consistency, Availability and Partition. As per CAP theorem we cannot have a database system that is Consistent, has high availability and partition tolerant all at once. There has to be a tradeoff. It is denoted by a triangle.

Availability - A request made to the data store shall always eventually complete, no matter what pattern of failures have occurred.

Consistency - The results of 'earlier' writes are consistent and are always seen by 'later' reads.

Partition tolerance – In case of network issues, it might face refusal of the network to deliver any subset of the messages sent between nodes. Hence it should be tolerant.

• HBase Architecture

HBase architecture comprises of Regions and Stores.

- a. **Master Server** : It assigns regions to region servers with the help of zookeeper. It also facilitates load balancing by distributing loads among different region servers. It is responsible for schema and metastore operations.

- b. **Region Server** : Handles the data related operations, takes requests from clients. It is responsible only for the regions under itself. This is the workhorse of HBase.
- c. **Zookeeper** : This is a helper solution to the Master Server that helps in the synchronizations, maintaining configurations etc. It tracks server failures, availability etc.

● **HBase vs RDBMS**

HBASE	RDBMS
Column Oriented Schema	Row oriented Schema
Flexible Schema	Fixed Schema
No Query language available in particular	SQL
De-normalized data	Normalized data
Good for all kinds of data.	Good for structured data