Unit - V

NoSQL :-
(1998 Carlo Strozzi)

- NoSQL stands for Not SQL or Not only SQL
- It is just apposite of SQL.
- The idea of NoSQL founded in 1998. with term light weight, schema less. by carlo strozzi
- It is non-relational database management system.
- It does not require fixed schema
- It avoids joins.
- It is easy to scale.
- It is greater flexibility then SQL because it store data structured, semi-structured and unstructured in the form of graph, chart, object, etc
- Lack of standardization of query languages.
- The purpose of using NoSQL database is to hold distributed data store with homogeneous data storage.
- It is used for big data or real-time web application.
  ex- MongoDB, BigTable, RevanDB, Redis, Canandra

Brief History of NOSQL Databases :-

i) 1998 - Carlo Strozzi
Carlo Strozzi used the term NoSQL for his light weight, open source relational database.

ii) 2000 - Neo4j
Graph database Neo4j is launched

iii) 2004
Google Big table is launched

iv) 2005
Couch DB is launched.

v) 2007
The research paper on Amazon Dynamo is launched.

vi) 2008.
Facebook open sources the Cassandra Project

vii) 2009
The term NoSQL was Re-introduced

Feature of NO SQL :-

i) Non-relational :- • NO SQL database never follows the relational model.
- never provides tables with flat fixed columns records.
- Does not require object relational mapping & data Normalisation.
- NO complex features like query language, referential integrity, joins & acid properties.

ii) Schema-free :- • NO SQL databases are either schema free or have relaxed schema.
- Do not require any sort of definition of the schema of the data.
- Offers heterogeneous structure of data in the same domain.
- ~~Simple API~~

iii) Simple API :- • Offers easy to use interface for storage and querying data provided.
- API allows low level data manipulation & selection method.

iv) Distributed :- (NO SQL shared nothing) :-
- Multiple NO SQL databases can be executed in the distributed fashion.
- Offer autoscaling and fail over capabilities.
- Only providing eventual consistency.

Types of NOSQL database

1. Key value stores
2. Column-oriented database
3. Document based database
4. Graph based database

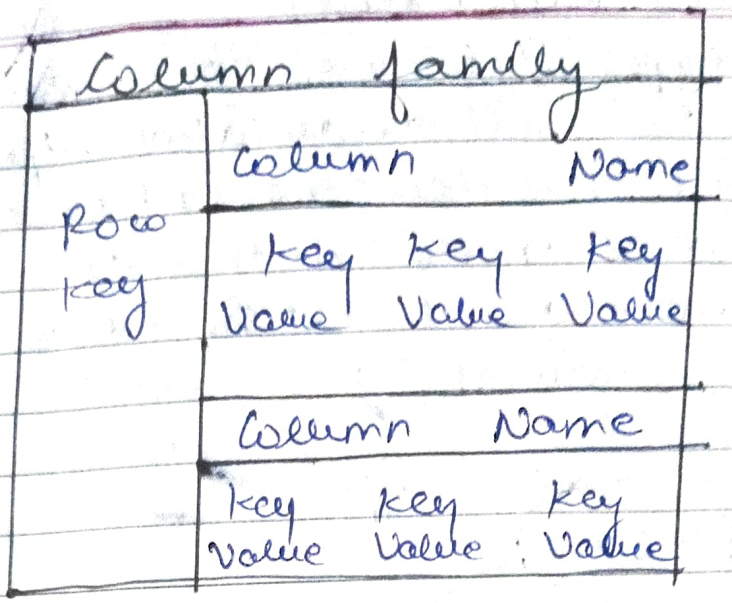
1) Key value stores :-
Data is stored in key-value pairs. where storage database stores data as a hash table where each key is unique & the value can be (JSON), (BLOB), (STRING) etc.

Key	Value
Name	Joe
age	42
Occupation	teacher
Height	175 cm

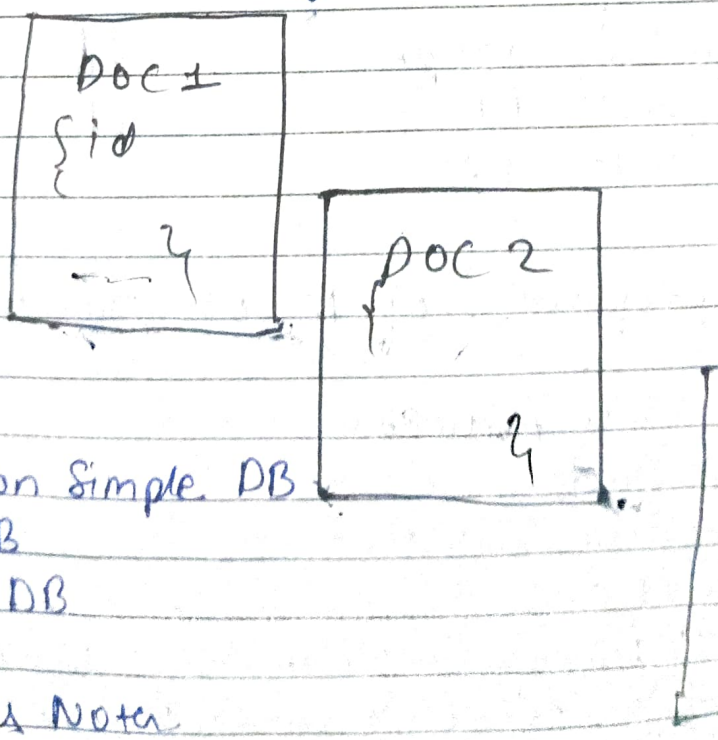
eg→ Redis, Dynamo, Riak.

2) Column-oriented database : It works on columns & are based on Bigtable paper by Google, every column is treated separately.
- Every column is treated separately values of single column database are stored contiguously.

eg→ HBASE, Cassandra & Hypertable



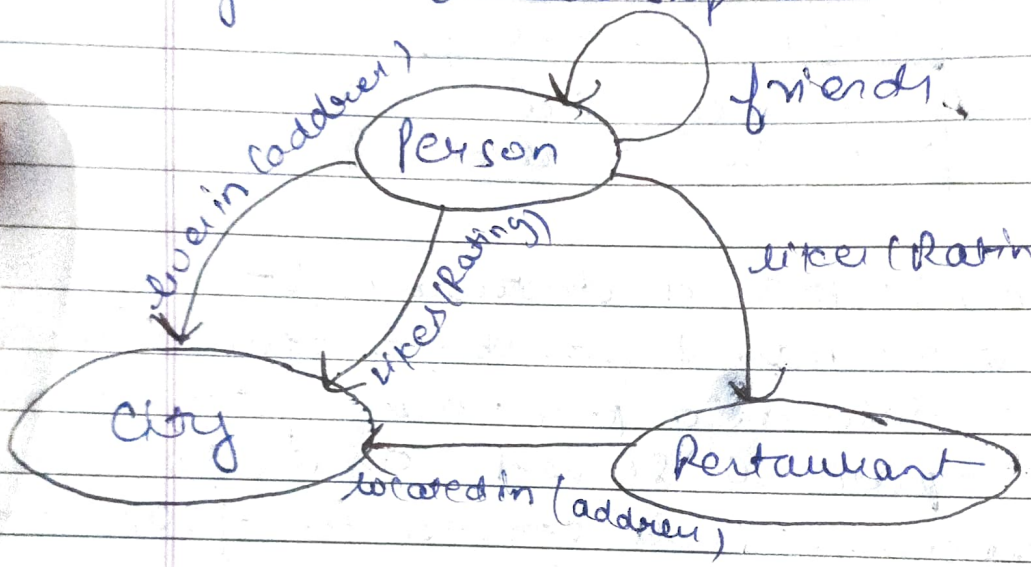
3) Document Based database :-
Document - Based database stored and retrieve data as a key value pair. What value part is stored as a document. The document is stored in json or ~~xml~~ xml format.



eg:- Amazon Simple DB
Couch DB
mongo DB
Riak
Lotus Notes

(4) Graph Based Database:-

Graph type Database stores the entities as well as relationship b/w those entities. The entity is stored as a node with the relationship as edges. An edge gives a relationship b/w nodes.



ex:- Neo4j
Infinite graph
Orient DB
flock DB.

⇒ BASE properties of NoSQL Database:-

Basically Available
Soft state
Eventually consistent.

- In ACID properties belongs to RDBMS
- BASE properties belongs to NoSQL

i) Basically Available :— The system works all the time available in case of failure.

ii) Soft state :- The data source doesn't have to be consistent all the time.

iii) Eventually consistent :- Not immediately consistent, but at some point data will consistent & No guarantee about when this will be occur.

CAP Theorem :-

C → Consistency
A → Availability
P → Partition tolerance

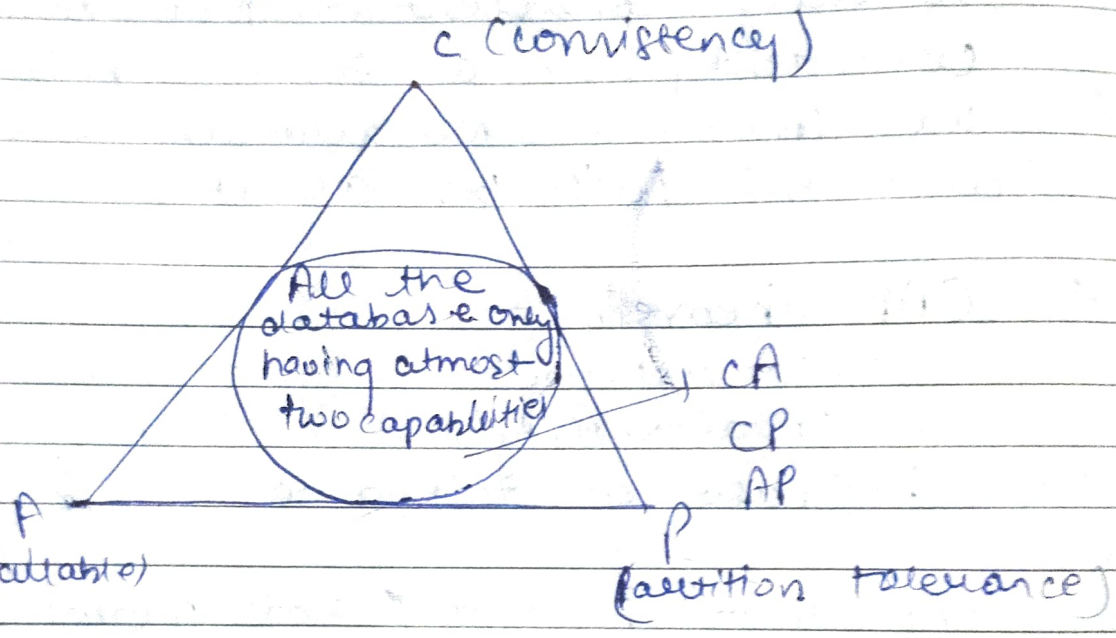
→ CAP theorem state that network shared datasystem can only strongly support two of the following three categories.

i) Consistency :- Consistency means that the node will have the same copies of a replicated data item visible for various transaction.

- Consistency refers to every client having the same view of data.

ii) Availability :- Every node (on either side of a network partition) must be able to respond in a reasonable amount of time.

iii) Partition tolerance :- It means that the system can continue operating even if the network connecting the nodes has a fault.



a) CA (Consistency or Availability) :
System respond last updated data & promise higher Availability.
eg :- RDBMS, Postgre SQL)

b) CP (Consistency Partition tolerance) :-
System can be distributed & promise to respond last updated data.
eg :- HBase, Mongo DB, Redis.

c) AP (Availability Partition tolerance) :-
System can be distributed & promise to has higher Availability
eg :- Cassandra, couch DB, Riak, Dynamo

RDBMS VS Mongo DB :-

RDBMS	Mongo DB.
Database	Database
Table	Collection.
Tuple / Row	Document
Column	field
Table join	embedded

The most popular database NOSQL database is an opensource, document oriented database, simple mongo DB document structure.