

- ① NO-SQL - schema flexibility, dynamic schemas, replication, simple relationship, replication, integrity, horizontal scalability, distributed tuples, semi structured data.

Issues in NOSQL

- ① Standardisation is not there
- ② Processing and difficult in executing complex queries.

Characteristics

- ① NOSQL - non relation database storage system that flexible data model.

eg Key-Value Pair
(Cassandra (Facebook))

HBase

JSON

Key-Value Pair.

- ② NOSQL - no fixed schema such as table or join.
data replication is possible.
data store is fault tolerant.

Q Explain the following

1. ACID properties in SQL
2. Trigger, view and schedule.

① ACID properties in SQL.

SQL Database Transaction exhibit ACID properties.

ACID - Atomicity, consistency, isolation and duration.

① Atomicity - all operations in transaction must be completed, if interrupted it must be undone (rollback)

② Consistency - The transaction must follow integrity constraint and follow constraint principle.

Eg A student's marks even when it is updated and even when result is announced, it should remain same.

③ Isolation - if there are two transactions in database it should be isolated from each other and done separately.

④ Durability - transaction must persist once completed.

2. Trigger, view scheduler.

1. Trigger - special stored procedure.

Trigger executes, when specific action occurs in database.

Eg. INSERT, UPDATE, DELETE.

2. View - refers to logical construct used in query statement.

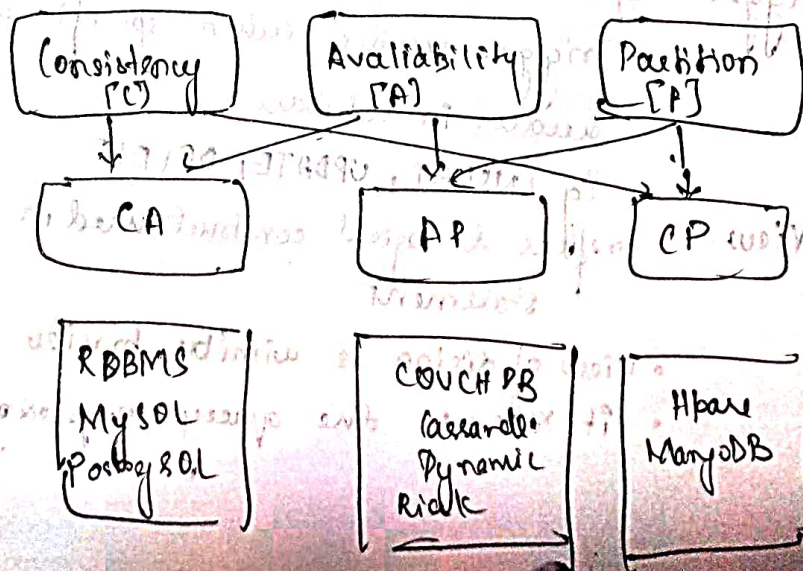
- View of a view is similar to view of table
- it reduces the query complexity.

- ③ Scheduler - means chronological sequence of instruction which executes concurrently.
- all instruction of transaction is included.
 - order of instruction is maintained.
 - executes multiple transaction in allotted time interval.

③ Explain CAP Theorem [6M].

Among C, A, P, atleast two are present for the application / service / process.

- **Consistency** - in distributed database all nodes observe same data at same time.
- **Availability** - means during the transaction, all fields ^{values} must be available in different partitions of database so each request receives a response on success as well as failure.
- **Partition** - means division of large database into different database without affecting the operations on them by adopting specified procedure.
- **Partition tolerance** - refers to continuation of operation even in case of message loss / node failure or node not reachable.



① Explain BASE properties. [3M].

BASE is flexible model for nosql database.

BA - means Basic Availability

S - means soft

E - means eventual consistency

① BA - ensures by distribution of shards across many data nodes with high degree of replication.

② Soft - ensures processing even in the presence of inconsistencies but achieving consistency eventually.

③ Eventual consistency - means consistency requirement in NOSQL database meeting at some point of time in future.

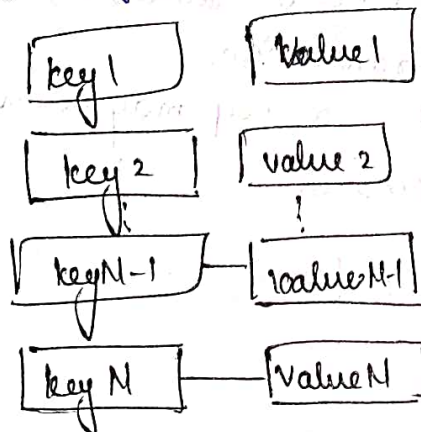
⑤ Explain Key Value Datastore its advantages and limitations [4M]

- The simplest way to represent a scheme-less data store is key value pairs.

- Data retrieval is fast in key-value pair datastore.

- Key maps to large data string.

- Key value store access use a primary key for accessing the values.



Key
"Priya"

Value.
"Category: Student; class:
B.Tech, Semester VIII;
Branch: Engineering;
MOBILE: 9999876532;"

Advantages of Key Value Store.

1. Data store can store any data in field value.
2. A query just requests the values and returns the value as a single item.
3. Key-value store is consistent.
4. Key-value datastore may be hierarchy so may be ordered key-value store.
5. Returned values on queries can be converted into list, table - columns, data from fields and columns.
6. Have i) scalability ii) reliability iii) portability iv) low operational cost.
7. The key can be synthetic or auto generated.

Limitations of Key value architecture.

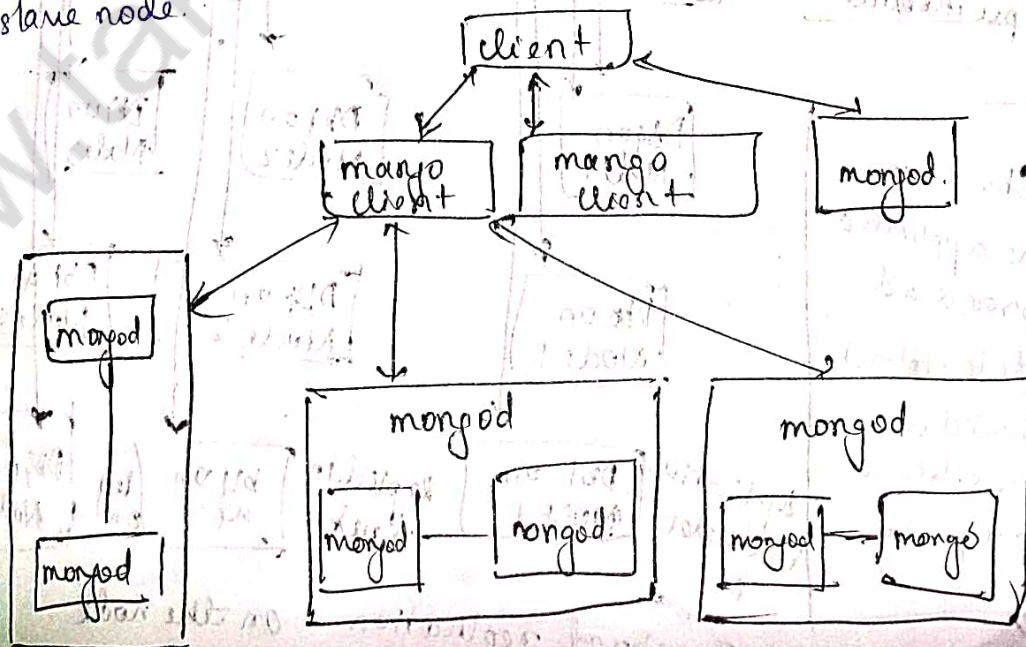
1. No indexes are maintained on values.
2. It does not provide traditional database capabilities such as ACID transaction.
3. Maintaining unique values as key may become difficult when of data increases.
4. Queries can not be performed on individual values.

Q. What are characteristics of Big Data NOSQL collection. [6M]

1. High and easy scalability.
 - Horizontal scalability - by add more machines at data nodes into the pool of resources.
 - scales out using multi-utility cloud services.
2. Support of replication.
 - ensures high availability, partition, reliability
3. Distributable.
 - enhances performance and throughput.
4. Usage of NOSQL server - less expensive, automatic repair, easier distribution.
5. Usage of open source tool - cheap and open source while RDBMS are expensive, big server storage system.
6. Support of schema-less model - no predefined schema.
7. Support of integrated caching. - supports caching in memory. increases the output performance.
8. No inflexibility they are not flexible like SQL have no structured way of storing and manipulating data.

Q. Explain master slave distribution model [4M]

A node server as primary node and other nodes are slave node.



- Master directs slave
- slave nodes replicate on multiple slave server in (MSD) model.
- even updation happens in master it also happens in slave.
- uses slave for read operation.

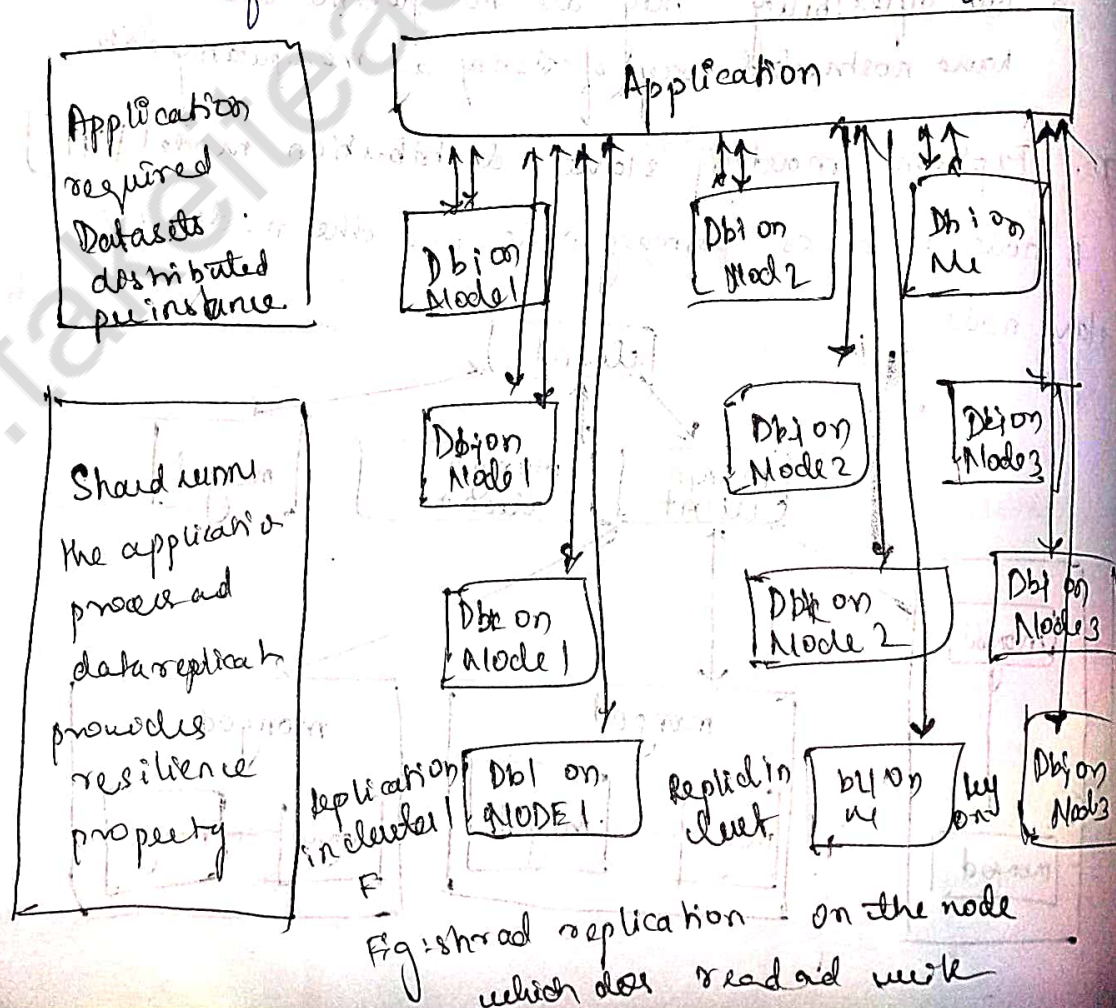
Master Slave Replication. Processing performance decrease to replicated MSD.

Complexity cluster based processing

b. Peer to Peer Distribution Model (P2P)

Peer to peer distribution model and replication shows the following characteristics.

- All replications nodes accept read request
- send the request
- All replicas function equally
- Node failure do not cause write capability



- 8) List and explain any 10 features of MongoDB (M).
- MongoDB is open source DBMS.
 - Mongo program creates and manage database.
 - MongoDB manages collection and document store.
 - Characteristics i) NO SQL ii) distributed iii) open-source.
 - iv) non-relational v) Scalable vi) flexible vii) Indexed.
 - ix) multi-master x) fault-tolerant.

Features of MongoDB.

- ① Distributed DB - makes it highly available and provides horizontal scalability.
- ② No complex Joins.
- ③ Querying, indexing and real time aggregation allows accessing and analyzing the data efficiently.
- ④ Deep query ability - supports dynamic queries on documents using a document based query language that is nearly as powerful as SQL.
- ⑤ MongoDB is a document data store in which collection holds different document. Data store is in form of JSON document.
- ⑥ Storing of data is flexible: data store consists of JSON document. The fields in one document can vary from other. data structure can be changed over time.
- ⑦ Storing of document in JSON format.
- ⑧ Document model is well defined. Structure of document is clear. Document is the unit of storing data in MongoDB database.

⑨ Collection stores a numbers of Many DB documents. Collection exists with single DB to achieve a single purpose. Collection may store document that do not have same fields.

⑩ Indexes on any field

⑪ Atomic operation

⑫ Fast in place of update

⑬ Conversion / mapping.

⑭ How does replication achieved in Many DB [6M].

- Replication ensure high availability in Big Data

- MongoDB replicate with help of replica set

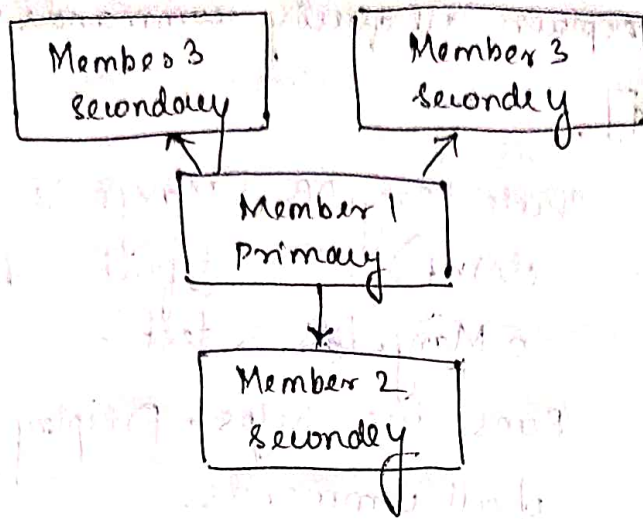
Replica set → MongoDB is a group of mongod (Many DB server) processes that store same dataset

- A replica set usually has minimum three nodes,

- one of them is called primary, receives all write operation.

- all other nodes are secondary

- a primary node can choose among the secondary node at the time of automatic failover or maintenance



Auto-sharding is method of distributing data across multiple machine in a distributed application environment

Commands	Description
rs.initiate()	To initiate new replica set
rs.config()	To check replica set configuration
rs.status()	To check the status of replica set
rs.add()	To add member of replica set

Vertical scaling by increase the resource of single machine

Horizontal scaling can be achieved by sharding mechanism - more database server can be added to support data growth demand & more read and write operation.

A shard stores lesser data than the actual data
handle less number of operations in a single machine

⑩ List and explain all query commands in MongoDB [6M].

1. `Mongo`. Skute MongoDB (Mongo is MongoDB client). The default database in MongoDB is test
2. `db.help()` Runs the help. Displays list of all commands.
3. `db.state()` get status of about MongoDB server.
4. `use <database name>` Creates database
5. `Db` Outputs the name of existing database.
6. `Dbs` gets list all databases.
7. `db.dropDatabase()` Drops a database.
8. `db.database name.insert()` creates a collection using insert
9. `db.<database name>.find()` Views all documents in collection
10. `db.<database name>.update()` Update document
11. `db.<database name>.remove()` Delete a document

① List and explain features and components of Cassandra DB [BM].

Features of Cassandra.

1. Maximizes the number of writes - writes are not very costly
2. Maximizes data duplication.
3. Does not support join, group by, OR clause and aggregation.
4. Uses classes consisting of ordered keys and semi-structured data storage system.
5. Is fast and easily scalable with write operation spread across the cluster.
6. Is distributed DBMS designed for handling a high volume of unstructured data across multiple cloud servers.
7. Has peer to peer distribution in system across the India.

Components of Cassandra:

Node	Place where data stores for processing
Data center	Collection of many related nodes
Cluster	Collection of many data center
Commit log	Used for crash recovery, each write operation is written in commit log
Memtable	Memory resistant data structure after write in commit log, data written in memtable temporarily

SST Table

When mem table reaches a certain threshold, data flush into SST Table.

Bloom Filter

Fast and memory efficient, probabilistic data structure to find whether an element is present in a set. Bloom filters are added after every query.

(12) Explain

a. Consistency Command. [4M]

Consistency command shows the current consistency level.

1. ALL: High consistent. A write must be written to commit log and memtable on all replica nodes in a cluster.
2. EACH - QUORAM: A write must be written in commit log and memtable on all replica nodes in datacenter.
3. LOCAL - QUORAM: A write must be written in commit log and memtable on replica node in same cluster.
4. ONE: A write must be written to commit log and memtable, of a least one replica node in cluster.
5. LOCAL-ONE: Same as one but least two or three replica nodes, respectively.

6. ANY: A write must be written to at least one node.

7. LOCAL-ONE: A write must be written for at least one replica node in local data center.

8. SERIAL: linearizable consistency to prevent unconditional update.

9. LOCAL-SERIAL: same as serial but restricted to local data center.

