

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot - 360001
Ph. No - (0281)2471645

2 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478



Lt. Shree Chimanbhai Shukla

MSCIT SEM 1 - MONGODB

Shree H.N.Shukla College Campus,
Street No. 2, Vaishali Nagar,
Nr. Amrapali Railway Crossing,
Raiya Road, Rajkot.
Ph. (0281)2440478, 2472590



Shree H.N.Shukla College Campus,
Street No. 3, Vaishali Nagar,
Nr. Amrapali Railway Crossing,
Raiya Road, Rajkot.
Ph. (0281)2471645

Website: www.hnsgroupofcolleges.org

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot – 360001
Ph. No - (0281)2471645

2 – Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

M.Sc.IT 1 MongoDB

Chapter 1:- * Unit 1:- Introduction to NoSQL Database

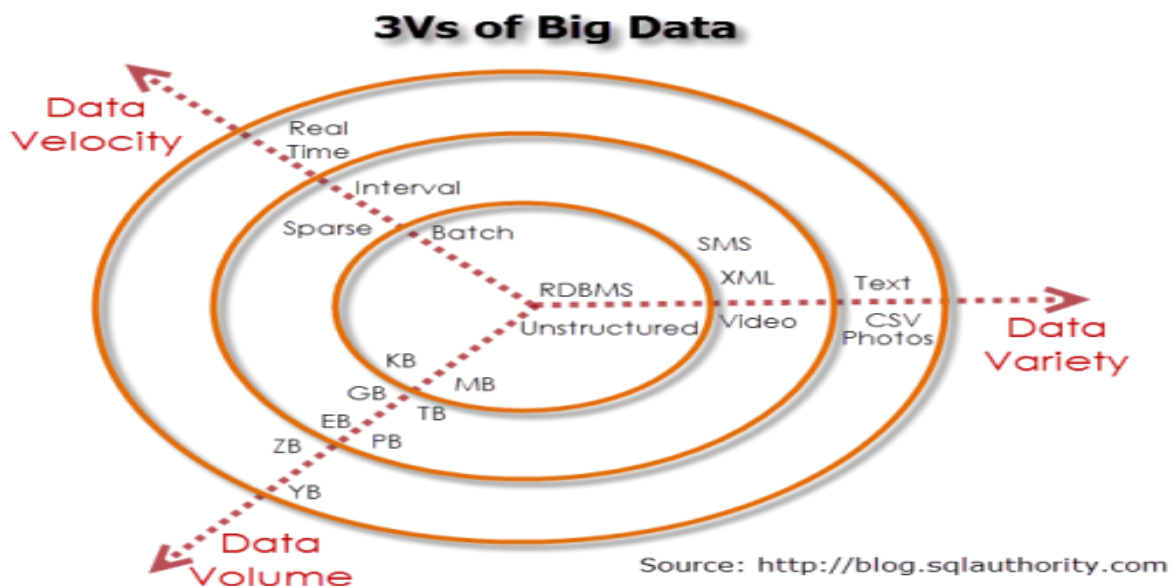
* What is Big Data?

Day by day, the amount of enterprise data has been increasing, and such large-scale data is referred to as **big data**.

It includes huge data and the concept of data **volume, variety and velocity**.

Big data is a field for analyze systematically extracted information from data set that is too large and complex.

Now a day's many services and analytical applications are based on big data. For handling big data, advance technology is requiring for data storage.



SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot - 360001
Ph. No - (0281)2471645

2 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

Question	Answer
What is Big data?	Large or Huge amount of data
3v stands for ?	Volume, Variety, Velocity
Large scale data is refers as _____	Big data
Big data is require in _____	Analysis and many services

* What is NoSQL?

NoSQL means "Not Only SQL".

It is alternate solution of traditional databases where you have to put the data in tables and the structure have to follow, but NoSQL database provides a mechanism for storage and retrieval of data **without any predefined schema for data insertion.**

NoSQL databases mostly used for big data storage. NoSQL is design specifically to manage big data, a task for which commonly used relational database management systems (RDBMSs) is not well suited.

A NoSQL (originally referring to "non SQL" or "non relational") database provides a mechanism for storage and retrieval of data.

NoSQL databases are increasingly **used in big data and real-time web applications.** NoSQL systems are also sometimes called "Not only SQL" to emphasize that they may support SQL-like query languages.

NoSQL databases can be easily installed in cheap commodity hardware clusters as transaction and data volumes increase. **This means that you can process and store more data at much less cost.**

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot - 360001
Ph. No - (0281)2471645

2 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

Question	Answer
NoSQL stands for ?	Not Only SQL
NoSQL design for ?	Handling large amount of data and database
NoSQL require schema ?	No
_____ is alternate solution of traditional databases	NoSQL
_____ provides a mechanism for storage and retrieval of data.	NoSQL Database
NoSQL databases are increasingly used in _____	big data and real-time web applications
you can process and store more data at much less cost with the help of _____	NoSQL Database
NoSQL also referred as	Non Relational
NoSQL databases can be easily installed in cheap commodity hardware clusters. Is it true?	Yes

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot - 360001
Ph. No - (0281)2471645

2 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

* NoSQL Features & Characteristics

The three main features of NoSQL databases are

1. Scale-out
2. Replication
3. Flexible data structure

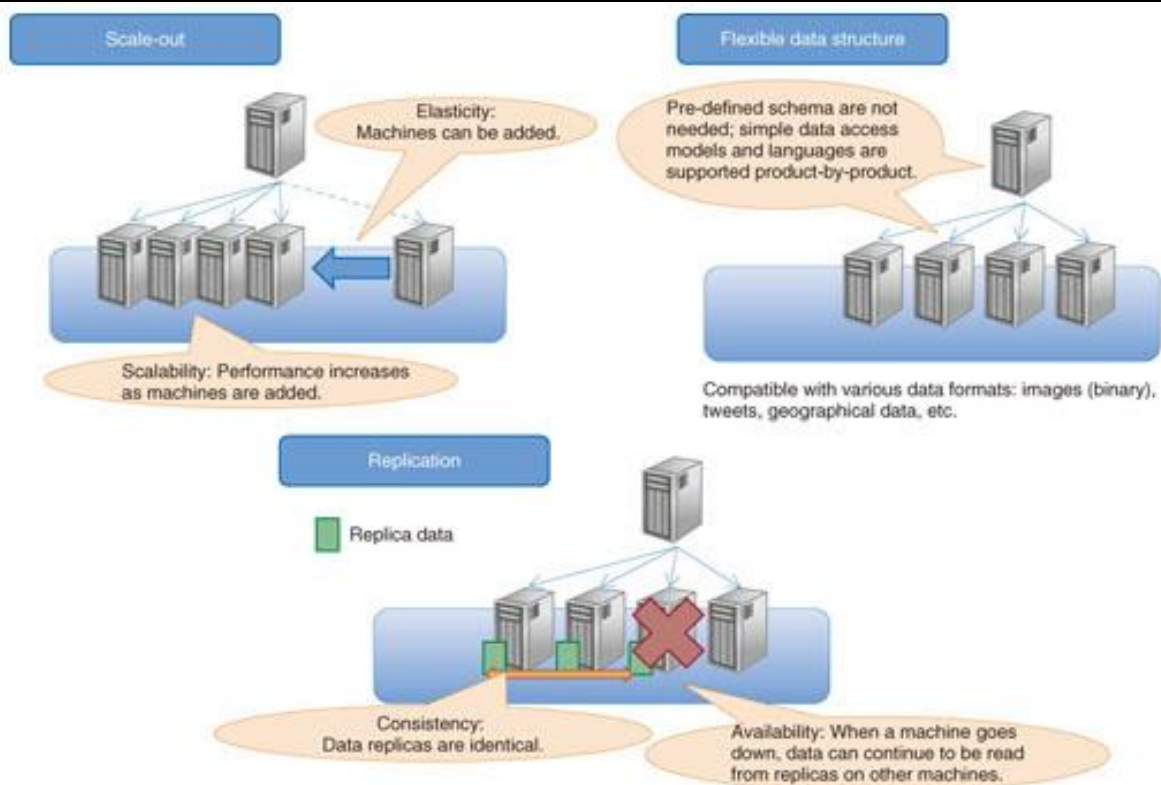


Fig. NoSQL database features and characteristics.

1. Scale-out

Scale-out refers to achieving **high performance** by **using many machines** in a distributed manner.

Distributing the data over a large number of machines enables scaling of the data set and distribution of the processing load.

A common feature of many NoSQL databases is that data is automatically distributed to new machines when they are added to the cluster, so **the performance is also improved**.

Scale-out is evaluated in terms of scalability and elasticity.

2. Replication

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot – 360001
Ph. No - (0281)2471645

2 – Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

Replication is the **copying of data** to achieve data redundancy and load distribution.

Even if data consistency has been lost among the replicas, it is eventually achieved: this is known as *eventual consistency*.

Replication is evaluated in terms of consistency and availability.

3. Flexibility

A flexible data structure means that there is no need to define a structure as a *database schema*.

Traditional RDBMSs require pre-defined schemas, and redefining them carries a high cost. NoSQL, on the other hand, does not require defined schemas, so users can store data with **various different structures in the same database table**.

However, most NoSQL databases do not support high-level query languages such as SQL, which is used by RDBMS, so products that support either simple relational operations or indexing have been released.

This feature is evaluated qualitatively.

4. Elastic Scalability

RDBMSs are not as easy to scale out on commodity clusters, whereas NoSQL databases are made for transparent **expansion**, taking advantage of new nodes.

These databases are designed for use with low-cost commodity hardware.

Scalability indicates how the performance of a NoSQL database cluster scales with the number of physical machines.

If performance improves as machines are added to a NoSQL database cluster, we can say that the **NoSQL product has high scalability**.

5. Big data applications

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot – 360001
Ph. No - (0281)2471645

2 – Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

While RDBMSs have grown to match the growing needs, but it's difficult to realistically use one RDBMS to manage such data volumes. These volumes are however easily handled by NoSQL databases.

6. Database administration

The best RDBMSs require the services of expensive administrators to design install and maintain the systems. On the other hand, NoSQL databases require much less **hands-on management**, with data distribution and **auto repair capabilities**, simplified data models and fewer tuning and administration requirements.

7. Economy

RDBMSs require installation of expensive storage systems and proprietary servers, while NoSQL databases can be easily installed in cheap commodity hardware clusters as transaction and data volumes increase. This means that **you can process and store more data at much less cost.**

8. Consistency

Consistency is a measure of the strength of data integrity. The parameters used to evaluate it include the number of replicas and the latency within the cluster. The consistency benchmarking software must check the consistency among replicas in an update-heavy workload, so data management of that workload, in which data should be updated, is necessary.

9. Availability

Availability refers to the ability of the overall system to continue operating during a network failure, called a **network partition**, or a physical machine failure.

High availability means that the **system can work without interruption** and without degraded performance, even when some machines go down and the network is partitioned or both.

Question

Answer

Main features of NoSQL ?

Scale-out, Replication, Flexible

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot - 360001
Ph. No - (0281)2471645

2 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

	structure
Replication means ?	Multiple copy of data
Scale-out means ?	Machines can be added to increase the performance at any time in the system
Eventual consistency means ?	Even if data consistency has been lost among the replicas, it is eventually achieved
_____ is a measure of the strength of data integrity.	Consistency
NoSQL can process and store more data at much less cost than RDBMS. Is it true?	True
_____ refers to the ability of the overall system to continue operating during a network failure, called a <i>network partition</i> , or a physical machine failure.	Availability
Flexibility means _____	No need to define structure or schema in advance
Elastic scalability means _____	Easy expansion or growth
Replication means _____	Multiple copies of data for load distribution

* History

The term *NoSQL* was used by "**Carlo Strozzi in 1998**" to name his lightweight database. It was NOSQL open source relational database that did not expose the standard SQL interface, but was still relational.

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot - 360001
Ph. No - (0281)2471645

2 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

Most of the early NoSQL systems did not attempt to provide (ACID) atomicity, consistency, isolation and durability guarantees that will provided here.

Based on 2014 revenue, the NoSQL market leaders are MarkLogic, MongoDB, and Datastax.

Based on 2015 popularity rankings, the most popular NoSQL databases are MongoDB, Apache Cassandra, and Redis.

Question	Answer
MongoDB is ?	NoSQL database
In 1998 who use the term NoSQL ?	Carlo Strozzi
NoREL refers as _____	No Relational
Early NoSQL systems support ACID?	No
The most popular NoSQL databases are...	Mongoddb, Apache Cassandra, Redis etc...

*Types of NoSQL databases

1. Key-value data stores:

Key-value (KV) stores use the associative array (also known as a map or dictionary) as their fundamental data model.

In this model, data is represented as a collection of key-value pairs, such that each possible key appears at most once in the collection.

The key-value model is one of the simplest data models, and richer data models are often implemented as an extension of it.

A key-value store, or key-value database, is a data storage model designed for storing, retrieving, and managing associative arrays, a data structure more commonly known as a dictionary or hash.



SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)

3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot – 360001
Ph. No - (0281)2471645

2 – Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

Dictionaries contain a collection of objects, or records, which in turn have many different fields within them, each containing data. These records are stored and retrieved using a key that **uniquely identifies the record**, and is used to **quickly find the data** within the database.

key-value systems offers considerable flexibility and more closely follows modern concepts like object-oriented programming.

Advantages:- efficiency, scalability, fault-tolerance. Data is stored sorted by key, very fast, very scalable, simple model, able to distribute horizontally

Disadvantage:- there is no client-server support built in to the library. Many data structures (objects) can't be easily modeled as key value pairs.

2. Document based data stores:-

A document-oriented database, or document store, is a computer program **designed for storing, retrieving, and managing document-oriented information**, also known as semi-structured data.

Document-oriented databases are **one of the main categories of NoSQL databases**, and the popularity of the term "**document-oriented database**" has grown with the use of the term NoSQL itself.

XML databases are a subclass of document-oriented databases that are optimized to work with XML documents. A document-oriented system depends on internal structure in the document in order to extract metadata that the database engine uses for further optimization.

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot – 360001
Ph. No - (0281)2471645

2 – Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

One of the other defining characteristics of a document-oriented database is that **the database offers an API or query language that retrieves documents based on their contents.**

Document databases contrast strongly with the traditional relational database (RDB). Relational databases generally store data in **separate tables** that are defined by the programmer, and a **single object may be spread across several tables.**

Document databases store all information for a given object in a single instance in the database, and every stored object can be different from every other.

Advantages:-

- unstructured data can be stored easily.
- json format easier to write.
- support rich queries and full indexes.
- supports Map-Reduce queries.

Disadvantage:-

In SQL database, the database administrators and analysts can carefully design the schema for the table; once the schema is in place, programs can only add records that match that schema. This **type of schema is not in document database.**

3) Graph based data stores

This kind of database is designed for data whose relations are **well represented as a graph** consisting of **elements interconnected with a finite number of relations** between them. The type of data could be social relations, public transport links, road maps or network topologies. It is based on the mathematical concept of graph theory.

Graph structures could be represented **in network model databases** from the late 1960s.

A number of improvements to graph databases appeared year by year and generate index web pages. In the late 2000s, commercial transactional graph databases such as oracle became available.

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)

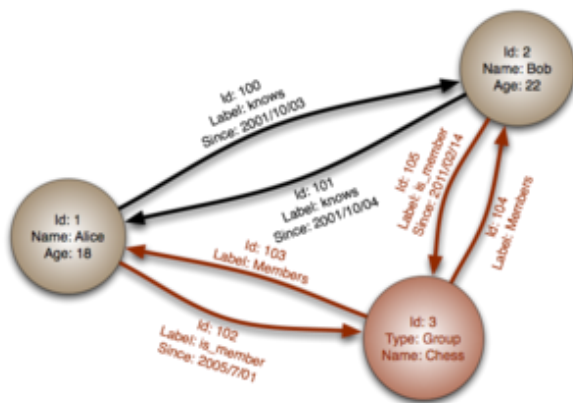


3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot – 360001
Ph. No - (0281)2471645

2 – Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

In computing, a graph database is a database that uses graph structures for semantic queries with **nodes, edges and properties to represent and store data.**

A key concept of the system is the *graph* (or *edge* or *relationship*), which **directly relates data items in the store.**



Graph databases employ nodes, properties, and edges.

Graph databases are based on graph theory. Graph databases employ nodes, edges and properties.

- **Nodes represent** entities such as people, businesses, accounts, or any other item you might want to keep track of. They are roughly the equivalent of the *record*, *relation* or *row* in a relational database, or the *document* in a document database.
- **Edges, also** known as *graphs* or *relationships*, are the lines that connect nodes to other nodes; they represent the relationship between them. Meaningful patterns emerge when examining the connections and interconnections of nodes, properties, and edges.
- **Properties** are relevant information that relate to nodes.

4) Column based data store:-

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot – 360001
Ph. No - (0281)2471645

2 – Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

These databases **organize the data in tables, similar to an RDBMS**, however, they store the content by columns instead of rows. They are **good for data warehousing applications**. Examples of column-based databases are Hbase, Cassandra, Hypertable, and so on. A column of a distributed data store is a NoSQL object of the lowest level in a key space.

It is a tuple (a key-value pair) consisting of three elements:

- **Unique name:** Used to reference the column
- **Value:** The content of the column. It can have different types.
- **Timestamp:** The system timestamp used to determine the valid content.

The column is used as a store for the value and has a timestamp that is used to differentiate the valid content from old ones.

The primary benefit you can get by storing data in a column-oriented database is that some of your queries could become really fast.

You probably wouldn't choose to use a column oriented database for a typical transactional application. You might **choose to use another database like RDBMS or NoSQL for your transactional load** with regular exports to the column oriented database for analytics.

Advantages:-

Not all the operations can be done online
Get and Put return result in milliseconds.
more efficient than row or document store

Question	Answer
Key-value stores use ?	Associative array
Key-value data model is like ?	Map or dictionary
Why Key-value data model is used ?	For fast data retrieval
Limitations of Key-value model is ?	Data cannot easily model in this structure
Advantage of Document based data store ?	Semi structured or Unstructured data can easily stored
What is collection ?	Group of documents

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot - 360001
Ph. No - (0281)2471645

2 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

Graph base data store is like ?	Network model database
Node means ?	Any entity
Edge means ?	Graph of Relationship
Column base data store is like ?	Tables similar to RDBMS
Unique name is used for ____	Reference the column
What is Timestamp?	Timestamp used to determine the valid content
Give the definition of Edge.	Edge is the lines that connect nodes to other nodes.

* **Locally hosted database**

A local database is one that is local to your application only. It support SQL Server CE (Compact Edition) format.

A "local database" is a database that is stored on the local PC where an application resides.

But in Distributed Database data can be store like in different system and user actually have to access the data like from request and response manner because user may not know the location of data in the computer.

Advantages:-

- 1) In a distributed database, data can be stored in different systems like personal computers.
- 2) A user doesn't know where the data is located physically. Database presents the data to the user as if it were located locally.
- 3) Database can be accessed over different networks.
- 4) Data can be joined and updated from different tables which are located on different machines.
- 5) Even if a system fails the integrity of the distributed database is maintained.
- 6) A distributed database is secure.

Disadvantages:-

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot – 360001
Ph. No - (0281)2471645

2 – Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

- 1) Since the data is accessed from a remote system, performance is reduced.
- 2) Static SQL cannot be used.
- 3) Network traffic is increased in a distributed database.
- 4) Database optimization is difficult in a distributed database.
- 5) Different data formats are used in different systems.
- 6) Different DBMS products are used in different systems which increases in complexity of the system.
- 7) Managing system catalog is a difficult task.
- 8) While recovering a failed system, the DBMS has to make sure that the recovered system is consistent with other systems.
- 9) Managing distributed deadlock is a difficult task.

Database as a service (DBaaS)

Some cloud platforms offer options for using a database as a service, without physically launching a virtual machine instance for the database.

In such a configuration, application owners do not have to install and maintain the database themselves.

Instead, the database service provider have responsibility for installing and maintaining the database, and application owners pay according to their usage.

For example, Amazon Web Services provides three database services as part of its cloud offering: SimpleDB, a NoSQL key-value store; Amazon Relational Database Service, a SQL-based database service with a MySQL interface; and DynamoDB. Similarly, Microsoft offers the Azure SQL Database service as part of its cloud offering.

Architecture and common characteristics

Most database services offer web-based consoles, which the end user can use to provision and configure database instances.



SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)

3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot – 360001
Ph. No - (0281)2471645

2 – Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

For example, the Amazon Web Services web-console enables users to launch database instances, create snapshots (similar to backups) of databases, and monitor database statistics.

Database services consist of a database-manager component, which controls the underlying database instances using a service API. The service API is exposed to the end user, and permits users to perform maintenance and scaling operations on their database instances.

For example, the Amazon Relational Database Service's service API enables creating a database instance, modifying the resources available to a database instance, deleting a database instance, creating a snapshot (similar to a backup) of a database, and restoring a database from a snapshot.

Database services make the underlying software-stack transparent to the user - the stack typically includes the operating system, the database and third-party software used to manage the database.

The service provider is responsible for installing, patching and updating the software stack and ensuring the overall health and performance of the database.

Database services take care of scalability and high availability of the database.

Because DBaaS solutions can reduce entry barrier costs, small businesses that found the costs of in-house database solutions too prohibitive might welcome cloud-based database solutions because of their affordability and flexibility.

The DBaaS model can also help reduce data and database redundancy and improve overall Quality of Service.

Database configuration installation can be done by db itself. It provide high scalability and high availability with proper maintenance of data.

Some db itself provides its own DbaaS like oracle provides 11g and 12c as cloud services.
Most db services offer web based console.

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot - 360001
Ph. No - (0281)2471645

2 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

Some db manager components provide many API for db services. For example create instance, create snapshot, monitoring, auto backup, security, reports, are the different services.

Dbaas solutions can reduces the costing with cloud based db so very much affordable and flexibility for small businesses.

The Dbaas model can also help for reduce the data and db redundancy and improve overall quality of service.

Question	Answer
Local hosted database store the data ?	Data stored in local area
Distributed Database store the data ?	Data can be stored in different system
DBaaS stands for ?	Database as a service
DBaaS model is helpful for ?	Reduce data redundancy and costing
A local database is one that is _____ to your application only.	Local
Database services take care of _____ of the database	scalability and high availability
The DBaaS model helps to _____.	reduce data and database redundancy and improve overall Quality of Service
Database configuration installation can be done by _____	database itself
Most Database provides web based console services. true or false	True
Who is responsible for installing,	The service provider

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot – 360001
Ph. No - (0281)2471645

2 – Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

patching and updating and ensuring
the performance of the database.

* Unit 2: Introduction to MongoDB

* MongoDB Concept

MongoDB is an open-source report database, and heading NoSQL database. MongoDB is written in C, C++ and JavaScript

MongoDB is a cross-stage, archive arranged database that gives, superior, high accessibility, and simple adaptability. MongoDB deals with idea of accumulation and report.

Databases:-

Database is a physical compartment for storing data. Every database gets its own particular set of documents on the record framework. A single MongoDB server commonly has number of databases. The databases act as containers of data and they are independent of each other. A MongoDB database contains one or more collections.

Collections:-

Collections are a gathering of MongoDB documents. It has similar concept like a table in RDBMS. It exists inside a database. There is no need to define structure like tables of RDBMS.

Document:-

A document stored in a collection is a unit of data. A document contains a set of fields or key-value pairs. The keys are strings, the values can be of various types: strings, integers, floats, etc. You can even store a document as the value of a field in another document.

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot – 360001
Ph. No - (0281)2471645

2 – Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

* Advantage of MongoDB over RDBMS:-

- MongoDB is record database in which one holds different files and records. Number of fields, substance and size of the record might be contrast starting with one record then onto the next.
- Structure is clear.
- No complex joins.
- Easy inquiry capacity.
- MongoDB helps dynamic inquiries on records utilizing a record based inquiry language that is almost as capable as SQL.
- Tuning is good.
- Simplicity of scale-out: MongoDB is not difficult to scale.
- Transformation/ mapping is not required.
- Utilizes inward memory for empowering quicker get of data.

RDBMS	MongoDB
Database	Database
Table	Collection
Row	Document
Column	Field
Table Join	Embedded Documents
Primary Key	Primary Key (Default key _id provided by mongodb itself)
Database Server and Client	
Mysqld/Oracle	Mongod
mysql/sqlplus	Mongo

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot – 360001
Ph. No - (0281)2471645

2 – Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

<u>Question</u>	<u>Answer</u>
Mongoddb Collection in RDBMS known as ?	Table
Collection is used for ?	Collecting or gathering mongoddb documents
Mongoddb documents contains ?	Set of fields or key value pair data.
Mongoddb is used for ?	Starting mongoddb server

* Downloading Installing and running MongoDB:-

* Install MongoDB On Windows

(1) To install the MongoDB on windows, first download the latest release of MongoDB from <http://www.mongodb.org/downloads>

(2) Make sure you get correct version of MongoDB depending upon your windows version. To get your windows version open command prompt and execute following command

```
C:\>wmic os get osarchitecture
```

```
OSArchitecture
```

```
64-bit
```

```
C:\>
```

32-bit versions of MongoDB only support databases smaller than 2GB and suitable only for testing and evaluation purposes.

(3) Now extract your downloaded file to c:\ drive or any other location. Make sure name of the extracted folder is is the version of MongoDB download.

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot – 360001
Ph. No - (0281)2471645

2 – Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

(4) In case you have extracted the mongodb at different location, then go to that path by using command **cd FOLDER/DIR** and now run the above given process.

(5) MongoDB requires a data folder to store its files. The default location for the MongoDB data directory is `c:\data\db`. So you need to create this folder using the Command Prompt. Execute the following command sequence

```
C:\>md data  
  
C:\>cd data  
  
C:\data\>md db
```

(6) If you have installed the MongoDB at different location, then you need to specify any alternate path for `\data\db` by setting the path `dbpath` in `mongod.exe`. For the same issue following commands

In command prompt navigate to the bin directory present into the mongodb installation folder. Suppose my installation folder is **D:\set up\mongodb**

```
C:\Users\XYZ>d:  
  
D:\>cd "set up"  
  
D:\set up>cd mongodb  
  
D:\set up\mongodb>cd bin  
  
D:\set up\mongodb\bin>mongod.exe --dbpath "d:\set up\mongodb\data"
```

This will show **waiting for connections** message on the console output indicates that the `mongod.exe` process is running successfully.

(7) Now to run the mongodb you need to open another command prompt and issue the following command `c:\mongodb`

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot - 360001
Ph. No - (0281)2471645

2 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

```
D:\set up\mongodb\bin>mongo.exe
```

```
MongoDB shell version: 2.4.6
```

```
connecting to: test
```

```
>db.test.save( { a: 1 } )
```

```
>db.test.find()
```

```
Output .....
```

> This will show that mongodb is installed and run successfully. Next time when you run mongodb you need to issue only commands

*** Install MongoDB on Ubuntu**

(1) Run the following command to import the MongoDB public GPG Key:

```
sudo apt-key adv --keyserver hkp://keyserver.ubuntu.com:80 --recv 7F0CEB10
```

(2) Create a /etc/apt/sources.list.d/mongodb.list file using the following command.

```
echo 'deb http://downloads-distro.mongodb.org/repo/ubuntu-upstart dist 10gen' | sudo tee /etc/apt/sources.list.d/mongodb.list
```

(3) Now issue the following command to update the repository:

```
sudo apt-get update
```

(4) Now install the MongoDB by using following command:

```
apt-get install mongodb-10gen=2.2.3
```

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot - 360001
Ph. No - (0281)2471645

2 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

(5) In the above installation 2.2.3 is currently released mongodb version. Make sure to install latest version always. Now mongodb is installed successfully.

Start MongoDB

```
sudo service mongodb start
```

Stop MongoDB

```
sudo service mongodb stop
```

Restart MongoDB

```
sudo service mongodb restart
```

To use mongodb run the following command

```
mongo
```

This will connect you to running mongod instance.

*** Configuring MongoDB**

There are some default configuration setting for running MongoDB but we can set them also manually by following ways.

There are two ways to configure MongoDB.

- (1) We can modify them by command-line parameters
- (2) By using File-based configurations.

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot – 360001
Ph. No - (0281)2471645

2 – Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

(1) Command Line parameters

We can override the default MongoDB settings by passing command-line parameters to the mongod program.

For example, the next command tells MongoDB to use C:\mongodb_data as data directory by sending it as a --dbpath argument:

```
C:\>mongodb\bin\mongod --dbpath C:\mongodb_data
```

The following table lists some useful command-line parameters and their functions:

Parameter	What it does
--dbpath	Path to the directory for storing data files.
--bind_ip	IP address that the mongod server will listen on, default is 127.0.0.1.
--port	Port address that mongod will listen on, default is 27017.
--logpath	Full file path to the log file where the MongoDB messages will be written. By default all messages are written to standard output.
--logappend	Setting this option to true appends the messages at the end of the log file. Setting it to false overwrites the log.

(2) File based configuration

An alternative to sending all those command-line parameters to mongod manually is to put the required configuration settings in a file and then pass the path of the file as a --config option.

For example, consider the following sample configuration file:

```
dbpath = D:\mongodb_data
```

```
logpath = D:\mongodb.log
```

```
logappend = true
```

We store this file to a location, say C:\mongodb.conf. Now, to start MongoDB with these settings, we have to enter the next command in the CMD prompt:

```
C:\>mongodb\bin\mongod --config C:\mongodb.conf
```


SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot – 360001
Ph. No - (0281)2471645

2 – Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

mongod will be loaded with these configuration settings. Note that file-based parameters are the same as those for command-line options.

*** Stopping MongoDB**

There are several ways you can shutdown a running MongoDB server.

In the terminal window (or CMD prompt window in case you are on Windows) running the mongod process, hit Ctrl + C . This will signal the server to do a clean shutdown, flush, and close its data files.

From the mongo shell

From the mongo interactive shell, you can issue a shutdownServer() command , causing mongod to terminate:

```
>use admin
```

```
switched to db admin
```

```
>db.shutdownServer()
```

*** MongoDB Statistics**

To get stats about mongod server type the command **db.stats()** in mongod client. This will show the database, name, number of collection and documents in the database. Output the command is shown below:

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot - 360001
Ph. No - (0281)2471645

2 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

```
C:\Windows\system32\cmd.exe - mongo.exe
> db.stats()
{
  "db" : "test",
  "collections" : 3,
  "objects" : 5,
  "avgObjSize" : 39.2,
  "dataSize" : 196,
  "storageSize" : 12288,
  "numExtents" : 3,
  "indexes" : 1,
  "indexSize" : 8176,
  "fileSize" : 201326592,
  "nsSizeMB" : 16,
  "dataFileVersion" : {
    "major" : 4,
    "minor" : 5
  },
  "ok" : 1
}
```

* MongoDB Help

To get list of commands type **db.help()** in mongodb client. This will give you list of commands as follows:

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot - 360001
Ph. No - (0281)2471645

2 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

```
C:\Windows\system32\cmd.exe - mongo.exe
D:\set up\mongodb\bin>mongo.exe
MongoDB shell version: 2.4.6
connecting to: test
> db.help()
DB methods:
  db.addUser(userDocument)
  db.adminCommand(nameOrDocument) - switches to 'admin' db, and runs command [ just calls db.runCommand(...) ]
  db.auth(username, password)
  db.cloneDatabase(fromhost)
  db.commandHelp(name) returns the help for the command
  db.copyDatabase(fromdb, todb, fromhost)
  db.createCollection(name, < size : ..., capped : ..., max : ... > )
  db.currentOp() displays currently executing operations in the db
  db.dropDatabase()
  db.eval(func, args) run code server-side
  db.fsyncLock() flush data to disk and lock server for backups
  db.fsyncUnlock() unlocks server following a db.fsyncLock()
  db.getCollection(cname) same as db['cname'] or db.cname
  db.getCollectionNames()
  db.getLastError() - just returns the err msg string
  db.getLastErrorObj() - return full status object
  db.getMongo() get the server connection object
  db.getMongo().setSlaveOk() allow queries on a replication slave server
  db.getName()
  db.getPrevError()
  db.getProfilingLevel() - deprecated
  db.getProfilingStatus() - returns if profiling is on and slow threshold
  db.getReplicationInfo()
  db.getSiblingDB(name) get the db at the same server as this one
  db.hostInfo() get details about the server's host
  db.isMaster() check replica primary status
  db.killOp(opid) kills the current operation in the db
  db.listCommands() lists all the db commands
  db.loadServerScripts() loads all the scripts in db.system.js
  db.logout()
  db.printCollectionStats()
  db.printReplicationInfo()
  db.printShardingStatus()
  db.printSlaveReplicationInfo()
  db.removeUser(username)
  db.repairDatabase()
  db.resetError()
  db.runCommand(cmdObj) run a database command.  if cmdObj is a string, turns it into { cmdObj : 1 }
  db.serverStatus()
  db.setProfilingLevel(level,<slows>) 0=off 1=slow 2=all
  db.setVerboseShell(flag) display extra information in shell output
  db.shutdownServer()
  db.stats()
  db.version() current version of the server
```

* Installing the PHP MongoDB driver on Unix

In a Unix-based system, the PHP driver for MongoDB can be installed using the PECL Program (PECL - PHP Extension Community Library)

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot – 360001
Ph. No - (0281)2471645

2 – Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

You need to have it installed on your machine, which can be done by using the following command:

sudo pecl install mongo

When the installation is finished, edit the php.ini file (usually found at /etc/php.ini) to add the line:

```
extension=mongo.so  
and then restart Apache.
```

In case you don't have pecl installed on your machine, you can download the driver source code from GitHub, build it, and install it manually:

```
$ tar zxvf mongodb-mongodb-php-driver-<commit_id>.tar.gz  
$ cd mongodb-mongodb-php-driver-<commit_id>  
$ phpize  
$ ./configure  
$ sudo make install
```

Check out the Mongo driver installation page

<http://www.php.net/manual/en/mongo.installation.php> on the PHP official website to get operating system specific detailed information.

*** Installing the PHP driver for MongoDB on windows**

Steps:-

Step – 1 : Download the ZIP archive <http://downloads.mongodb.org/mongo-latest-php5.2vc6ts.zip> on your machine and extract it.

Step – 2 : Copy the php_mongo.dll file from the extracted folder to the PHP extension directory; this is usually the folder name ext inside your PHP installation.

Step -3 : Open the php.ini file inside your PHP installation and add the following line:
extension=php_mongo.dll

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot - 360001
Ph. No - (0281)2471645

2 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

Step - 4 : Save the file and close it. Restart the Apache web server.

Step - 5 : Open up your text editor and add the following code to a new file:

```
<?php  
phpinfo();
```

Step - 6 : Save the file as phpinfo.php inside the DocumentRoot of the Apache web server (the htdocs folder).

Step - 7 : Execute the phpinfo.php script in your browser

(<http://localhost/phpinfo.php>). Scroll down to find the section mongo to see all the MongoDB driver-specific information.

mongo		
MongoDB Support		enabled
Version		1.0.11
Directive	Local Value	Master Value
mongo.allow_empty_keys	0	0
mongo.allow_persistent	1	1
mongo.auto_reconnect	1	1
mongo.chunk_size	262144	262144
mongo.cmd	\$	\$
mongo.default_host	localhost	localhost
mongo.default_port	27017	27017
mongo.long_as_object	0	0
mongo.native_long	0	0
mongo.utf8	1	1

Congratulations! You have successfully installed the PHP driver for MongoDB.

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot – 360001
Ph. No - (0281)2471645

2 – Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

* Connecting to the MongoDB Server from PHP

Use the API provided by the PHP-MongoDB driver to create a connection to the Mongo server from a PHP script.

Creating a connection to the MongoDB server from PHP

Step – 1 : Open up your text editor and add the following code in a new file:

```
<?php
try{
$mongo = new MongoClient(); //create a connection to MongoDB
$databases = $mongo->listDBs(); //List all databases
echo '<pre>';
print_r($databases);
$mongo->close();
} catch(MongoConnectionException $e) {

//handle connection error

die($e->getMessage());
}
```

Step -2 : Save the file as test_connection.php under the DocumentRoot of your web server.

Step – 3 : Open up your browser, and execute the script by going to the location

http://localhost/test_connection.php:

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot - 360001
Ph. No - (0281)2471645

2 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

Configuring the PHP MongoDB Connection

When no parameter is passed to the constructor of the Mongo class, it connects to the Mongo server running on localhost, on **port 27017** (or whatever value is specified for mongo.default_host and mongo.default_port in php.ini).

If we want to connect to a server running on a different host and/or port, **we can pass the connection string (mongodb://<hostname>:<port_number>)** as the \$server parameter to the Mongo constructor.

For example, to connect to a Mongo server listening on port 8888, we will type the following command:

```
$mongo = new Mongo($server="mongodb://localhost:8888");
```

The Data Model and Working with Data

With MongoDB, you may embed related data in a single structure or document.

These schema are generally known as “denormalized” models, and take advantage of MongoDB’s rich documents. Consider the following diagram:



SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



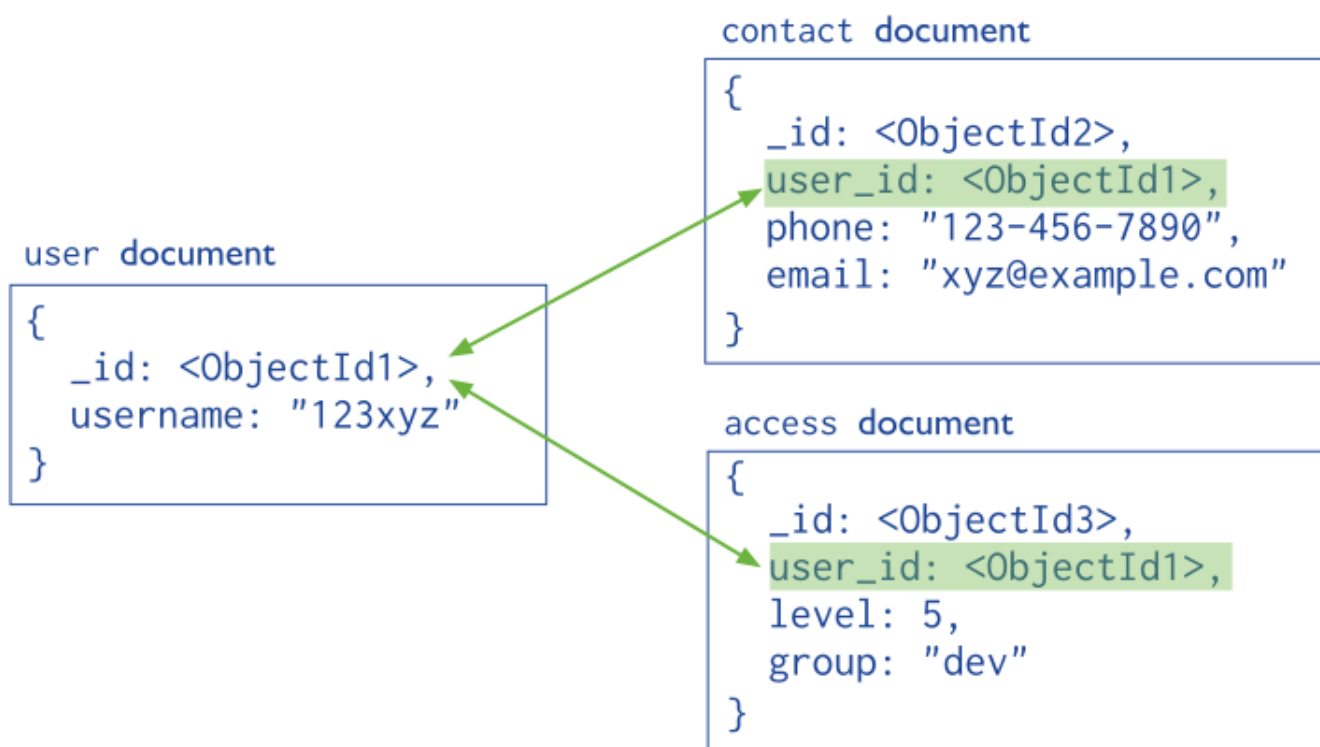
3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot - 360001
Ph. No - (0281)2471645

2 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

Embedded data models allow applications to store related pieces of information in the same database record.

As a result, applications may need to issue fewer queries and updates to complete common operations.

Normalized Data Models



Document databases such as *MongoDB* use *JSON* documents in order to store records, just as tables and rows store records in a relational database.

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot - 360001
Ph. No - (0281)2471645

2 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

Question	Answer
Which command is used to know the os architecture of your computer ?	wmic os get osarchitecture in command prompt
Which folder you can create in c: drive to keep the data of mongodb	data folder in c: drive
To start the mongodb in unix which command is used ?	Sudo service mongodb start
Which is configuration file of php ?	Php.ini
Db.help() display ?	List of commands
--dbpath command is used for ?	Path to the directory for storing the data files
Which is default port of mongod server	27017
Which dll file you require to configure php with mongodb	Php_mongo.dll
Mongodb support which data model	Embedded data model

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot – 360001
Ph. No - (0281)2471645

2 – Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

Chapter 2

Unit 1 - Learning MongoDB by implementing web Application & Using Mongodb with relational Databases

(1) Explain Collection and different commands.

(1) createCollection() :-

It is used to create the collection in database.

Syntax:- db.createCollection(name,options)

Ex:- db.createCollection(“mytable”)

The second argument is optional in this method.

(2) showCollections() :-

It will display all the collections of current database.

(3) insert() :-

It is used to insert document (record) in the collection.

Ex:- db.mytable.insert({“no” :001})

(4) drop() :-

It is used to drop a collection from the database.

Syntax:- db.collectionname.drop()

Ex:- db.mytable.drop()

drop method return either true or false. If the collection dropped successfully then returns true otherwise return false.

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot – 360001
Ph. No - (0281)2471645

2 – Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

(2) MongoDB Datatype.

- String : This is most commonly used data type to store the data. String in mongodb must be UTF-8 valid.
- Integer : This type is used to store a numerical value. Integer can be 32 bit or 64 bit depending upon your server.
- Boolean : This type is used to store a Boolean (true/ false) value.
- Double : This type is used to store floating point values.
- Min/ Max keys : This type is used to compare a value against the lowest and highest BSON elements.
- Arrays : This type is used to store arrays or list or multiple values into one key.
- Timestamp : This can be handy for recording when a document has been modified or added.
- Object : This data type is used for embedded documents.
- Null : This type is used to store a Null value.
- Symbol : This data type is used identically to a string however, it's generally reserved for languages that use a specific symbol type.
- Date : This data type is used to store the current date or time in UNIX time format. You can specify your own date time by creating object of Date and passing day, month, year into it.
- Object ID : This data type is used to store the document's ID.
- Binary data : This data type is used to store binary data.
- Code : This data type is used to store JavaScript code into document.
- Regular expression : This data type is used to store regular expression.

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot – 360001
Ph. No - (0281)2471645

2 – Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

(3) Explain document and different commands.

In MongoDB the document is inserted into collection into any database.

(1) insert() :-

For insert the document this method is used.

Syntax :- db.collectionname.insert(document name)

Ex:- db.stud.insert (

```
{  
  Rno: "1",  
  Name: "mayur",  
  Course:"bca",  
  Detail:  
    {Duration: "6 month"  
      Trainer: "jigneshsir"},  
  Batch: "2016"  
}
```

)

(2) find() :-

For display the documents you have to use find method.

Ex:- db.stud.find()

(3) update():-

To update the document or to modify existing document of a collection, update() is used.

Syntax:- db.collectionname.update(selection criteria, updatedata)

Ex:- db.stud.update({'course':'bca'}, {\$set: {'course':'bscit'}})

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot – 360001
Ph. No - (0281)2471645

2 – Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

(4) array() :-

To insert multiple document in a collection you have to pass an array of documents to the insert method.

```
Ex:- var allcourse=[
    {
        Rno:1
        Name: "mayur"
        Course: "bca"
        Detail:
            {Duration: "6 month"
              Trainer: "jigneshsir"},
        Batch: "2016"
    }
    {
        Rno:2
        Name: "nikhil"
        Course: "bca"
        Detail:
            {Duration: "6 month"
              Trainer: "jigneshsir"},
        Batch: "2016"
    }
    {
        Rno:3
        Name: "nikesh"
        Course: "bca"
        Detail:
            {Duration: "6 month"
              Trainer: "jigneshsir"},
        Batch: "2016"
    }
];
```

(5) Insert an array :-



SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)

3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot - 360001
Ph. No - (0281)2471645

2 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

To insert an array in to document you have to use insert method with array name as a parameter. It is also known as bulk insert.

Ex:- db.stud.insert(allcourse);

(6) delete():-

To delete the document from collection, remove() is used.
It include the following parameters.

- 1) deletion criteria:- the criteria according to documents will be remove.
- 2) justOne :- it remove only one document when set to true or 1.

Ex:- db.stud.remove({rno: 1})
db.stud.remove ({type: "programming language"}, 1)

If you want to remove all the documents from the collection then pass empty {} as a parameter of remove method.

Ex:- db.stud.remove({})

(7) find().limit() :-

To display all records or to display limited records find() & limit() used.

The find() return all the document but if you want to display the single document then you have to pass {} with find().

db.stud.find({})

The limit() is used for retrieve one or two document and field from the collection. It accept number argument.

Syntax:- db.collectionname.find().limit(number)

Example:- db.stud.find().limit(2)



SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)

3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot - 360001
Ph. No - (0281)2471645

2 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

(8) skip() :-

It is used to skip the document from starting of the document as per given number argument.

Syntax:- `db.collectionname.find().limit(number).skip(number)`

Example :- `db.stud.find().limit(1).skip(1)`

(9) sort() :-

It is used to sort the documents in the collections.

Here the sorting order is specified as 1 or -1. Here 1 is used for ascending order
And -1 is used for descending order.

Example:- `db.stud.find().sort({“course” :-1 })`

(10) pretty() :-

This method is used to display the result with formatting in proper way.

Syntax:- `db.collectionname.find().pretty()`

Example:- `db.stud.find().pretty()`

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot - 360001
Ph. No - (0281)2471645

2 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

Which method is used to create collection?	<code>createCollection()</code>
Which method is used to insert document in the collection? Give the syntax.	<code>Insert()</code> <code>Insert({"column name":data})</code>
How can you delete or remove the collection from database?	<code>Drop()</code>
Which data type is used to store multiple values?	<code>Array</code>
Which data type is used to store document ID.	<code>Object ID</code>
Which data type is used to record document added and modification time?	<code>Timestamp</code>
Which method is used to display all document?	<code>find()</code>
Which method is used to display limited document?	<code>find().limit()</code>
Which method is used to display document with proper format?	<code>find().pretty()</code>

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot - 360001
Ph. No - (0281)2471645

2 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

(4) DIFFERENT RDBMS WHERE CAUSE EQUIVALENTS IN MONGODB

<u>Operation</u>	<u>Syntax</u>	<u>Example</u>	<u>RDBMS Equivalent</u>
Equality	{<key>:<value>}	db.mycol.find ({"by":"tutorialspoint"}) .pretty()	where by = 'tutorialspoint'
Less Than	{<key>:{\$lt:<value>}}	db.mycol.find ({"likes":{\$lt:50}}).pretty ()	where likes < 50
Less Than Equals	{<key>:{\$lte:<value>}}	db.mycol.find ({"likes":{\$lte:50}}) .pretty()	where likes <=50
Greater Than	{<key>:{\$gt:<value>}}	db.mycol.find ({"likes":{\$gt:50}}) .pretty()	where likes > 50
Greater Than Equals	{<key>:{\$gte:<value> }}	db.mycol.find ({"likes":{\$gte:50}}) .pretty()	where likes >=50
Not Equals	{<key>:{\$ne:<value>}}	db.mycol.find ({"likes":{\$ne:50}}) .pretty()	where likes !=50

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot – 360001
Ph. No - (0281)2471645

2 – Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

(5) Create database, collection and connection with mongodb in php.

```
<?php
```

```
$m=new MongoClient();  
echo "connection mongo client sucessfully";
```

```
$db=$m->mydb;  
echo "Database mydb is selected";
```

```
$collection=$db->createCollection("my col");  
echo "collection created successfully";
```

```
?>
```

(6) Test Connection

```
<?php
```

```
try  
{  
$mongo = new MongoClient(); //create a connection to MongoDB
```

```
$databases = $mongo->listDBs(); //List all databases  
echo '<pre>';  
print_r($databases);  
$mongo->close();  
}
```

```
catch(MongoConnectionException $e)  
{  
die($e->getMessage()); //handle connection error  
}
```

Question

Answer

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot - 360001
Ph. No - (0281)2471645

2 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

Give the syntax of less than cause in mongodb. Give the example also.	<code>{{(<key>:{\$lt:<value>}} db.result.find({"percentage":{\$lt:60}}).pretty()</code>
Which command is used for display the message?	<code>echo "message"</code>
Which function is used to create mongodb client in php?	<code>MongoClient()</code>
Which function is used to display all databases of mongodb?	<code>listDBs()</code>
Which method is used to end the connection of mongodb and php?	<code>close()</code>

(7) Insert document in emp collection in hns database via php.

```
<?php
    $m=new MongoClient();
    echo "connection to database sucessfully";
    $db=$m->mydb;
    echo "Database mydb is selected";
    $collection=$db->mycol;
    echo "collection created successfully";

    $document=array(
        "title"=>"MongoDB",
        "description"=>"database",
        "likes"=>"100",
        "url"=>"http://www.hnshukla.com",
        "by"=>"hns",
    );

    $collection->insert($document);
    echo "document inserted successfully";
?>
```

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot - 360001
Ph. No - (0281)2471645

2 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

(8) Display document from emp collection from hns database.

```
<?php
try
{
    $connection = new MongoClient();
    $database = $connection->selectDB('hns');
    $collection = $database->selectCollection('emp');
}

catch(MongoConnectionException $e)
{
    die("Failed to connect to database ".$e->getMessage());
}

$cursor = $collection->find();

?>
<?php
echo "<table border=1>";
echo "<tr><th>EMPNO</th><th>NAME</th><th>SALARY</th>";

while ($cursor->hasNext()):$data = $cursor->getNext();

    echo "<tr><td>";

    echo $data['empno'];
    echo "</td><td>";
    echo $data['empnm'];
    echo "</td><td>";
    echo $data['salary'];

    echo "</td></tr>";

endwhile;
echo "</table>";
?>
```

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot – 360001
Ph. No - (0281)2471645

2 – Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

(9) Update document from emp collection from hns database.

```
<?php
    $m=new MongoClient();
    echo "connection to database sucessfully";

    $db=$m->hns;
    echo "Database is selected";

    $collection=$db->emp;
    echo "collection created successfully";

    $collection->update(array("empnm"=>"mayur"),
    array('$set'=>array("empnm"=>"hemang")));

    echo "document updated successfully";

    $cursor=$collection->find();
    foreach($cursor as $document)
    {
    echo $document["empnm"]."\n";
    }
?>
```

(10) Delete document from emp collection from hns database.

```
<?php
$m= new MongoClient();
echo "connection to database sucessfully";

$db=$m->hns;
echo "database selected";

$collection = $db->emp;
echo "collection selected sucessfully";

$collection->remove(array"empnm"=>"hemang"));
```

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot - 360001
Ph. No - (0281)2471645

2 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

```
echo "documents deleted successfully";
```

```
$cursor =$collection->find();
```

```
echo "Updated document";
```

```
foreach ($cursor as $document)
```

```
{  
    echo $document["empnm"] . "\n";
```

```
}  
?>
```

(11) Insert data into database from textbox.

Ans:-

empno :-	<input type="text"/>	<input type="button" value="Insert"/>
empname:-	<input type="text"/>	
empsalary:-	<input type="text"/>	

```
<html>  
    <body>  
<form  
action="MongoConnection.php"method="post">
```

```
Enter employee no:- <input type="text" name="no"/><br>  
Enter employee name:- <input type="text" name="nm"/><br>  
Enter employee salary:- <input type="text" name="sal"/><br>
```

```
<input type="submit" name ="insert" value="Insert"/>
```

```
</form>  
    </body>
```

```
</html>
```

```
-----  
<?php  
$no=$_POST['no'];  
$nm=$_POST['nm'];  
$sal=$_POST['sal'];
```

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot - 360001
Ph. No - (0281)2471645

2 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

```
$m = new MongoClient();  
$db = $m->hns;  
echo "Connection successfully";
```

```
$coll = $db->emp;  
echo "Database is found is here";
```

```
$data = array('empno' => $no,'empnm' => $nm,'salary'=> $sal);
```

```
$coll->insert($data);  
echo "Insert Successfully";
```

```
header('Location:/mscit1/displaydata.php');  
?>
```

(12) Example of Blog Post.

```
<html>  
<head>
```

```
<title>Blog Post Creator</title>
```

```
</head>  
<body>
```

```
<h1>Blog Post Creator</h1>
```

```
<form action="blogpostconn.php" method="post">
```

```
<h3>Title</h3>
```

```
<p>
```

```
<input type="text" name="title" id="title/">
```

```
</p>
```

```
<h3>Content</h3>
```

```
<textarea name="content" rows="20"></textarea>
```

```
<p>
```

```
<input type="submit" name="btn_submit" value="Save"/>
```

```
</p>
```

```
</form>
```

```
</body>
```

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot - 360001
Ph. No - (0281)2471645

2 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

</html>

<?php

```
try {
    $connection = new MongoClient();
        $database = $connection->selectDB('myblogsite');
    $collection = $database->selectCollection('articles');

    $article = array(
        'title' => $_POST['title'],
        'content' => $_POST['content'],
        'saved_at' => new MongoDate() // $date
    );

        $collection->insert($article);
    }

    catch(MongoConnectionException $e)
    {
        die("Failed to connect to database ".$e->getMessage());
    }
    catch(MongoException $e)
    {
        die('Failed to insert data '.$e->getMessage());
    }
}
```

?>

<html>

<body>

Article saved. _id:<?php echo \$article['_id'];?>.

Write another one?

</body>

</html>

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot – 360001
Ph. No - (0281)2471645

2 – Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

<u>Question</u>	<u>Answer</u>
How can you create the collection student in mydb database?	<code>\$collection=\$mydb->student;</code>
How can you select the collection student from mydb database?	<code>\$collection=\$mydb ->selectCollection('student');</code>
Which method is used to print the message and exit from the current php script.	<code>die()</code> method
Which method is used to delete the document from collection?	<code>remove()</code>
Which method is used to fetch system error message in catch block?	<code>getMessage()</code>

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot – 360001
Ph. No - (0281)2471645

2 – Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

(13) Explain Mongodb and RDBMS together.

Any relational database has typical style and design that shows number of tables and the relationship between these tables.

Mongodb is document database which is based on collection in any database.

In Mongodb there is no fixed structure or no any complex join like RDBMS.

Mongodb support big data, mobile and social infrastructure, user data management, content management and data hub.

Mongodb support array as first class object of a document so we can easily handle many to one or many to many relationship.

Mongodb also support embedded document and nested document.

Mongodb also support denormalization of your data.

In RDBMS mysql is somehow similar to Mongodb.

Mongodb and mysql are both capable of representing rich data model.

Where mysql use fixed design table but Mongodb has design free document.

Mysql and Mongodb both support B-tree index with similar behavior.

Mongodb and mysql both support replication.

Mongodb also support reference document so one document may also a reference to another document.

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot – 360001
Ph. No - (0281)2471645

2 – Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

(14) Explain how to manage a relationship between documents in MongoDB.

Mongodb relationship shows that how different documents are related to each other logically.

This relationship can be design through referenced relationship and embedded relationship.

This relationship can be one to one, one to many, many to one or many to many.

For example:-

- suppose we have blog application then each post has a title that is one to one relationship.
- Each author and blog is one to many relationships.
- And any post has many tags and the tags refer to many post so that is many to many relationship.

Reference Relationship:-

In the reference relationship method, it includes standard relationship both the client and location documents will be kept independent.

The client document contents a field that will reference the location document's id field.

Mongodb support the structure data format into collections of JSON document.

The JSON document is useful for data management and include human readable data format.

A JSON support rich and more flexible data structure then tables.

If we use JSON document in our database then we don't need any object relational mapped between our database and the application.

JSON stands for Java Script Object Notation.

It is open standard, lightweight, language-independent data interchange format based on JavaScript. JSON makes it easy to share data between devices like clients and servers.

Embedded Relationship:-

In the embedded method, we will embed the address document inside the client document.

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot - 360001
Ph. No - (0281)2471645

2 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

In this method, all the related information keeps in a single document, which makes it simple to recover and keep up-to-date. The entire document can be recovered in a single query.

<u>Question</u>	<u>Answer</u>
Mongodb support embedded document. True or False	True
Mongodb support de normalization of your data and replication. True or False	True
Mongodb and Mysql both database capable or representing rich data model. True or False	True
Mongodb support reference document. True or False	True
Mongodb require fix structure like RDBMS. True or False	False
What is embedded relationship?	All related information keeps in a single document. It embed the address document inside the client document.
Why embedded relationship is require?	It is helpful to recover document with single query and maintain up to date data.
What is Reference relationship?	It keeps independent documents from client. the location reference will be provide to client with document's id.
Give the full form of JSON.	Java Script Object Notation

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot – 360001
Ph. No - (0281)2471645

2 – Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

Chapter-3 **Session Management**

- Understanding HTTP sessions.
- Understanding PHP native session handling.
- Implementing session handling with MongoDB.
- Putting Session Manager.
- Building user authentication module, creating login, logout and user profile.

*** Explain HTTP sessions.**

- HTTP session is a sequence of network request-response transaction.
- An http client request by establishing a transmission control port on a server.
- Http server listen client requests from that port and after receive the request server send back some response message.
- The http support transport layer and the url of a page contains the domain name and the port number.
- An http is stateless protocol and it does not require http server about each user for the duration of multiple request.
- Some web applications implements states or server side session using web forms, http cookies or http hidden variables.
- HTTP sessions is an industry standard feature that allows Web servers to maintain user identity and to store user-specific data during multiple request/response interactions between a client application and a Web application.
- HTTP sessions preserves:
 - Information about the session itself (session identifier, creation time, time last accessed, etc.)
 - Contextual information about the user (client login state, for example, plus whatever else the Web application needs to save)
- HTTP sessions are useful for load balancing.

*** Explain PHP session handling.**

- In PHP session handler a special class is used for session handling.
- There are 7 method which are used for internal session handling like open, close, read, write, destroyed, GC and create_sid.
- Session.save-handler class will wrap the internal save handler and configure the directive which files by default.
- When a plain instance(object) of session handler is set as a save handler using session_set_save_handler() then it will wrap the current save handlers.

Methods:-

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot – 360001
Ph. No - (0281)2471645

2 – Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

1. Session handler_create_session_sid():- it return a new session Id
2. Session handler_open():- it initialize a session .
3. Session handler_close():-it close the session.
4. Session handler_read():- it read the session data.
5. Session handler_write():- it write the session data.
6. Session handler_gc():- it clean up old session.
7. Session handler_destroyed():- it destroyed a session.

* Understanding PHP native session handling:

Php handle on session natively here it configure the session on server (php.ini file) and not in the application.

So if we have several website on one server then we only need to configure the server for the changes to be applied on all the websites.

There is different libraries can be use for native session handling, here you can also create a new library that starts a native php session and allows to store and retrieve different things from it.

With the release of PHP4, session management was introduced as an extension to the PHP language. PHP provides several session-related functions, and developing applications that use PHP sessions is straightforward.

The three important features of session management are mostly taken care of by the PHP scripting engine.

How to use PHP sessions?, How sessions are started and ended ? and How session variables are used?. We list the PHP functions for building session-based web applications. Because not all browsers support cookies, and some users actively disable them.

When a user first enters the session-based application by making a request to a page that starts a session, PHP generates a session ID and creates a file that stores the session-related variables.

PHP sets a cookie to hold the session ID in the response the script generates. The browser then records the cookie and includes it in subsequent requests.

Between the browser and the server when initial requests are made to a session-based application The out-of-the-box configuration of PHP session management uses disk-based files to store session variables.

Using files as the session store is enough for most applications in which the numbers of concurrent sessions are limited.

Starting a Session PHP provides a session_start() function that creates a new session and subsequently identifies and establishes an existing one. The first time a PHP script calls session_start(), a session identifier is generated, and, by default, a Set-Cookie header field is included in the response.

The response sets up a session cookie in the browser with the name PHPSESSID and the value of the session identifier.

The PHP session management automatically includes the cookie without the need to call to the setcookie() or header() functions.

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot – 360001
Ph. No - (0281)2471645

2 – Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

The session identifier (ID) is a random string of 32 hexadecimal digits, such as fcc17f071bca9bf7f85ca281094390b4. As with other cookies, the value of the session ID is made available to PHP scripts in the \$HTTP_COOKIE_VARS associative array and in the \$PHPSESSID variable.

When a new session is started, PHP creates a session file. With the default configuration, session files are written in the /tmp directory using the session identifier, prefixed with sess_, for the filename.

However, if the identified session file can't be found, session_start() creates an empty session file.

Session Variables need to be registered with the session_register() function that's used in a session. If a session has not been initialized, the session_register() function calls session_start() to open the session file.

Once registered, session variables are made persistent and are available to scripts that initialize the session. PHP tracks the values of session variables and saves their values to the session file: there is no need to explicitly save a session variable before a script ends.

* Implementing Session handling with MongoDB

To implement session with MongoDB you have to use connect framework and express framework. First of all you have to connect-mongodb.

API

MongoDBStore

This module exports a single function which takes an instance of connect (or Express) and returns a MongoDBStore class that can be used to store sessions in MongoDB.

The MongoDBStore class has 2 required options:

- (1) uri: a MongoDB connection string
- (2) collection: the MongoDB collection to store sessions in.

If you pass in an instance of the [express-session module](#) the MongoDBStore class will enable you to store your Express sessions in MongoDB.

In many circumstances, connect-mongo will not be the only part of your application which need a connection to a MongoDB database. It could be interesting to re-use an existing connection.

Alternatively, you can configure connect-mongo to establish a new connection.

A Session is a way to store the information that is used across multiple page not stored in user Computer.

The web server does not know who you are or what you do because the http is having stateless protocol and so does not maintain the state

Session is used to store the information in the format of session variables.



SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)

3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot – 360001
Ph. No - (0281)2471645

2 – Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

Here user can store the information that is used for multiple page like username, back color, etc...

The session is available by default until the user close the browser.

If you want to store information permanent then you can store it in database.

In php session_start () function is used to start a session.

You have to write this function very first thing in your document.

If you want to store session variable value then a globle variable \$_session is used.

If you want to change the session variable value then you have to just override it.

If you want to remove all globle session variable & destroyed the session then you have to used session_unset() & session_destroyed().

Ex1 :-

Website Counter

```
<?php
    Session_start();
    If(isset($_SESSION['counter']))
    {
        $_SESSION['counter']+=1;
    }
    Else
    {
        $_SESSION['counter']=1;
    }
    $msg="you have visited this page".$_SESSION['counter'];
    $msg="in this session";
?>
<html>
<head>
    <title> counter</title>
</head>
<body>
    <?php echo ($msg); ?>
</body>
</html>
```

Ex-2:-

```
<html>
    <body>
        <?php
            //remove all session variable
            Session_unset();
            //destroyed the session
            Session_destoryed();
```


SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot – 360001
Ph. No - (0281)2471645

2 – Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

?>
</body>
</html>

Question	Answer
What is session?	Session is connectivity between the server and the client. The total period of time devoted to an activity.
Is HTTP stateless protocol?	Yes
What is the use of \$_SESSION[]?	It is used to store session data in key-value pairs using the \$_SESSION[] it is super global array.
What is HTTP session?	HTTP session is a sequence of network request-response transaction.
What is session management?	It is a process of secure handling multiple requests to a web based application or service from a single user or entity.
What is session_set_save_handler() ?	It is used to set the user level session storage functions which are used for storing and retrieving data associated with session.
Explain php native session handling?	Internal session handling for php by using several methods.
_____ method is used to start a session.	session_start()

*** Explain Session Manager.**

Session Management implements the exchange mechanism between the user and the application. The session id is used for remember the user in application or in web application.

There are many different frameworks which provide their own session management features like J2EE, Asp.Net, PHP etc.

There are different framework available for session management which have its own advantages and disadvantages.

Session Manager:-

Atlas:-

Mongodb Atlas provides all the features that have made the monogodb as NO SQL database.

It provides data model flexibility, scalability & combine them with optimizations and automated best practice to guaranty of performance and security.

Advantages:-

Session Manager is a simple powerful extension that makes it quick and easy to save, update, remove, and restore sets of tabs in any browser. It create **sessions** for daily routines for time saving and easiness of work.

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot – 360001
Ph. No - (0281)2471645

2 – Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

In PHP session management will be performed by different functions.

A session is a way to store information (in variables) to be used across multiple pages. Unlike a cookie, the information is not stored on the user's computer.

Session variables remains until the user closes the browser. So; Session variables hold information about one single user, and are available to all pages in one application. If you need a permanent storage, you may want to store the data in a database.

A session is started with the `session_start()` function. Session variables are set with the PHP global variable: `$_SESSION`.

To remove all global session variables and destroy the session, use `session_unset()` and `session_destroy()`

session handling mechanism of PHP to use a MongoDB database for managing sessions instead of using the filesystem. Before we proceed into implementation, we are going to briefly cover the basics, mainly the `session_set_save_handler()` function.

The `session_set_save_handler()` function allows us to define our own functions for storing and retrieving session data.

The function takes six arguments, each one being the name of a callback function. This is what the method signature looks like:

```
bool session_set_save_handler(callback $open, callback $close, callback $read, callback $write, callback $destroy, callback $gc)
```

Session Manager is a simple yet powerful extension that makes it quick and easy to save, update, remove, and restore sets of tabs. Common uses include: - Creating sessions for daily routines: pages you open in the morning, afternoon and evening.

Amazon EC2(Amazon Elastic Compute Cloud) Simple Systems Manager (SSM) is an Amazon Web Services tool that allows an IT professional to automatically configure virtual servers in a cloud or in on-premises data center.

AWS Systems Manager Parameter Store consists of standard and advanced parameters. Standard parameters are available at no additional charge.

IAM Role:- IAM authorizes to start a session for an EC2 instance (IAM policy). The administrator uses the AWS Management Console or the terminal (AWS CLI and additional plugin required) to start a session via the Systems Manager. ... The Session Manager sends audit logs to CloudWatch Logs or S3.

SSM Role:- AWS Systems Manager Session Manager. Session Manager is a fully managed AWS Systems Manager capability that lets you manage your Amazon EC2 instances, on-premises instances, and virtual machines (VMs) through an interactive one-click browser-based shell or through the AWS CLI.

AWS Systems Manager Agent (SSM Agent) is Amazon software that can be installed and configured on an EC2 instance, an on-premises server, or a virtual machine (VM). SSM Agent makes it possible for Systems Manager to update, manage, and configure these resources.

Different Tools:

Website: www.hnsgroupofcolleges.org

Page 58 of 122

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot – 360001
Ph. No - (0281)2471645

2 – Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

(1) Windows terminal Session Management Tool

(2) AWS Amazon Web Services

(3) PuTTY Session Manager

(4) Huawei Unified Session Manager



SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)

3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot – 360001
Ph. No - (0281)2471645

2 – Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

Chapter 4

Aggregation Queries

- (1) Generating sample data
- (2) Understanding Map Reduce.
- (3) Performing Map Reduce in Monod and PHP, Aggregation using group()
- (4) Listing distinct values for field.
- (5) Counting documents with count()

Q-1 Explain aggregation in Mongodb.

Ans:-

Once you have data stored in Mongodb then you may want to do more than just data retrieval.

You may want to analysis or display the data in any interesting way at that time Mongodb provides the aggregation framework, aggregation map command, map reduce support etc.

Aggregation operations process the data record and return computed results.

The aggregation framework transfer and combined the documents in a collection.

It means that aggregation operation is one-group values from multiple document together and return one single result.

In SQL count(*) and group by is equivalent to Mongodb aggregation.

The aggregate() is used for aggregation in Mongodb.

Syntax:- db.collectionname.aggregate(pipeline, operations)

Example:- db.stud.aggregate([{\$group: {_id: '\$gender' total: {\$num:1}}}]])

Q-2 Explain aggregation expressions.

Ans:-

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot – 360001
Ph. No - (0281)2471645

2 – Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

Expression	Description	Example
\$sum	Sums up the defined value from all documents in the collection.	db.mycol.aggregate({{\$group : { _id : "\$by_user", num_tutorial : {\$sum : "\$likes"}}}})
\$avg	Calculates the average of all given values from all documents in the collection.	db.mycol.aggregate({{\$group : { _id : "\$by_user", num_tutorial : {\$avg : "\$likes"}}}})
\$min	Gets the minimum of the corresponding values from all documents in the collection.	db.mycol.aggregate({{\$group : { id : "\$by_user", num_tutorial : {\$min : "\$likes"}}}})
\$max	Gets the maximum of the corresponding values from all documents in the collection.	db.mycol.aggregate({{\$group : { id : "\$by_user", num_tutorial : {\$max : "\$likes"}}}})
\$push	Inserts the value to an array in the resulting document.	db.mycol.aggregate({{\$group : { id : "\$by_user", url : {\$push: "\$url"}}}})
\$addToSet	Inserts the value to an array in the resulting document but does not create duplicates.	db.mycol.aggregate({{\$group : { id : "\$by_user", url : {\$addToSet : "\$url"}}}})
\$first	Gets the first document from the source documents according to the grouping. Typically this makes only sense together with some previously applied "\$sort"-stage.	db.mycol.aggregate({{\$group : { _id : "\$by_user", first_url : {\$first : "\$url"}}}})
\$last	Gets the last document from the source documents according to the grouping. Typically this makes only sense together with some previously applied "\$sort"-stage.	db.mycol.aggregate({{\$group : { _id : "\$by_user", last_url : {\$last : "\$url"}}}})

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot - 360001
Ph. No - (0281)2471645

2 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

Question	Answer
Why aggregation is used?	For data retrieval, analysis, and display processed data
Aggregations is equivalent to ???	Sql count(*) and group by
Which function is used for aggregation in Mongoddb?	aggregate()
Explain \$sum().	It is used to sums up the defined value from all documents in the collection
Explain \$push()	It is used to inserts the value to an array in the resulting document.
Explain \$addToSet()	It is used to inserts the value to an array in the resulting document but does not create duplicates.
Explain \$last()	It is used to get the last document from the source documents as per group.
Explain \$max()	It is used to get the maximum of the corresponding values from all documents in the collection.

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot – 360001
Ph. No - (0281)2471645

2 – Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

Q-3 Explain aggregation pipelines & Map reduce concept.

Ans:-

Aggregation pipelines give you a way to transfer and combined document in your collection.

It is very similar to unix pipeline command.

In the pipeline concept, you can send the output from one command to another as input and then go on.

The aggregate help to take an array of pipeline operator in the cell.

Mongodb also support the same concept in aggregation framework where a set of possible stage and set of document used for producing a result and then the result will be found as JSON document at the end of application pipeline.

There are different possible stages in aggregation framework, which are as under.

- (1) \$ project - It is used to select some specific field from a collection.
- (2) \$ match - This is filtering operation for reduce the amount of document which will be given as input to the next stage.
- (3) \$group - It is used for simple count operations where the actual aggregation done.
- (4) \$sort – It is used to sort the documents.
- (5) \$skip – It is used to skip the documents from given amount of documents.
- (6) \$ limit – It is used to limit the amount of documents by the given number starting from current position.
- (7) \$ unwind – It is used to unwind the document data using array. This operation will be undone the joint data and keep the documents individual. It will increase the amount of documents for the next stage.

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot – 360001
Ph. No - (0281)2471645

2 – Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

Question	Answer
What is Aggregation pipelines?	It is a way to transfer and combined documents in your collection.
Why aggregate is used?	Aggregate is used to take an array of pipeline operator in the cell.
\$ project is used for?	It is used for select some specific field from a collection.
\$ match is used for?	It is filtering operation for reduce the amount of document which will be given as input to the next stage.
\$skip is used for?	It is used to skip the documents from given amount of documents.
\$limit is used for?	– It is used to limit the amount of documents by the given number starting from current position.

Q-4 Map Reduce.

Ans:-

Map reduce is powerful and flexible tool for aggregating data.

It can be solve some problems that are too complex for aggregation framework.

Map reduce is two step approach to data processing it means first you have to map the data and then after reduce the data. In MongoDB we can used the map reduce command on a collection.

Map reduce use java script as its query language so it can perform complex logics.

Map reduce is slow so it should not be used for real time data analysis.

Map reduce is generally use for processing the data in large volume of data sets.

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot – 360001
Ph. No - (0281)2471645

2 – Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

Map reduce can work with multiple servers and splits a problems into chunks & send it to different machines, when all machine are finished its work then they merge all the pieces of the solution back into a full solution.

Map reduce is useful and powerful but it is complex.

The map reduce function first queries the collection then maps the result documents to emit key-value pairs which is then reduce based on the keys that have multiple value.

MapReduce Command

Following is the syntax of the basic mapReduce command –

```
>db.collection.mapReduce(  
  function() {emit(key,value);}, //map function  
  function(key,values) {return reduceFunction}, { //reduce function  
    out: collection,  
    query: document,  
    sort: document,  
    limit: number  
  } )
```

The map-reduce function first queries the collection, then maps the result documents to emit key-value pairs which is then reduced based on the keys that have multiple values.

In the above syntax

- **map** is a javascript function that maps a value with a key and emits a key-value pair
- **reduce** is a javascript function that reduces or groups all the documents having the same key
- **out** specifies the location of the map-reduce query result
- **query** specifies the optional selection criteria for selecting documents
- **sort** specifies the optional sort criteria
- **limit** specifies the optional maximum number of documents to be returned

Using MapReduce

Consider the following document structure storing user posts. The document stores user_name of the user and the status of post.

```
{
```

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot – 360001
Ph. No - (0281)2471645

2 – Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

```
"post_text": "tutorialspoint is an awesome website for tutorials",  
"user_name": "mark",  
"status":"active"  
}
```

Now, we will use a mapReduce function on our **posts** collection to select all the active posts, group them on the basis of user_name and then count the number of posts by each user using the following code –

```
>db.posts.mapReduce  
(  
  function() { emit(this.user_id,1); },  
  
  function(key, values) {return Array.sum(values)},  
  {  
    query:{status:"active"},  
    out:"post_total"  
  }  
)
```

The above mapReduce query outputs the following result –

```
{  
  "result" : "post_total",  
  "timeMillis" : 9,  
  "counts" : {  
    "input" : 4,  
    "emit" : 4,  
    "reduce" : 2,  
    "output" : 2  
  },  
  "ok" : 1,  
}
```

The result shows that a total of 4 documents matched the query (status:"active"), the map function emitted 4 documents with key-value pairs and finally the reduce function grouped mapped documents having the same keys into 2.

To see the result of this mapReduce query use the find operator –

```
>db.posts.mapReduce(  
  function() { emit(this.user_id,1); },  
  function(key, values) {return Array.sum(values)},  
  {  
    query:{status:"active"},  
    out:"post_total"  
  }  
)
```

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot – 360001
Ph. No - (0281)2471645

2 – Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

```
function() { emit(this.user_id,1); },  
function(key, values) {return Array.sum(values)}, {  
  query:{status:"active"},  
  out:"post_total"  
}  
).find()
```

The above query gives the following result which indicates that both users **tom** and **mark** have two posts in active states –

```
{ "_id" : "tom", "value" : 2 }  
{ "_id" : "mark", "value" : 2 }
```

In similar manner, MapReduce queries can be used to construct large complex aggregation queries. The use of custom Javascript functions makes usage of MapReduce very flexible and powerful.

Map-reduce is a data processing paradigm for condensing large volumes of data into useful *aggregated* results. For map-reduce operations, MongoDB provides the mapReduce database command.

Consider the following map-reduce operation:

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)

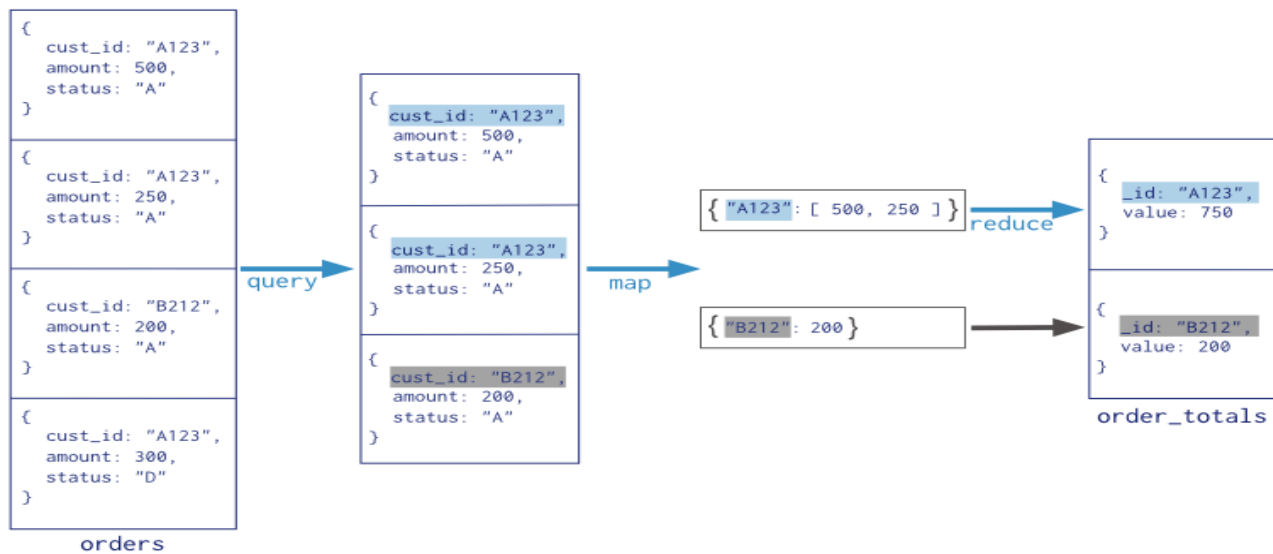


3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot - 360001
Ph. No - (0281)2471645

2 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

Collection

```
db.orders.mapReduce(  
  map    → function() { emit( this.cust_id, this.amount ); },  
  reduce → function(key, values) { return Array.sum( values ) },  
  query  → {  
    query: { status: "A" },  
    out: "order_totals"  
  }  
)
```



```
db.orders.aggregate(  
  [  
    { $match: { status: "A" } },  
    { $group: { _id: "$cust_id", total: { $sum: "$amount" } } },  
    { $sort: { total: -1 } }  
  ],  
  {  
    explain: true  
  }  
)
```

In this map-reduce operation, MongoDB applies the *map* phase to each input document (i.e. the documents in the collection that match the query condition). The map function emits key-value pairs. For those keys that have multiple values, MongoDB applies the *reduce* phase, which collects and condenses the aggregated data. MongoDB then stores the results in a collection. Optionally, the output of the reduce function may pass through a *finalize* function to further condense or process the results of the aggregation.

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot - 360001
Ph. No - (0281)2471645

2 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

Example of Map Reduce

```
db.books.insert( {
    "title" : "MongoDB: The Definitive Guide",
    "published" : "2013-05-23",
    "authors": [
        { "firstName" : "BRIJESH", "lastName" : "PATEL" }
    ],
    "categories" : [ "Databases", "NoSQL", "Programming" ],
    "publisher" : { "name" : "NACHIKETA" },
    "price" : 32.99
} )
db.books.insert( {
    "title" : "MongoDB in Action",
    "published" : "2011-12-16",
    "authors": [
        { "firstName" : "VISHAL", "lastName" : "RANPARA" }
    ],
    "categories" : [ "Databases", "NoSQL", "Programming" ],
    "publisher" : { "name" : "JAMNADAS" },
    "price" : 30.83
} )
***
db.books.insert( {
    "title" : "MongoDB in Action 2",
    "published" : "2011-12-16",
    "authors": [
        { "firstName" : "JIGNESH", "lastName" : "THANKI" }
    ],
    "categories" : [ "Databases", "NoSQL", "Programming" ],
    "publisher" : { "name" : "JAMNADAS" },
    "price" : 35.30
} )
***
db.books.insert( {
```

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot - 360001
Ph. No - (0281)2471645

2 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

```
"title" : "Scaling MongoDB",  
"published" : "2011-03-07",  
"authors": [  
  { "firstName" : "NIKHIL", "lastName" : "DAVE" }  
],  
"categories" : [ "Databases", "NoSQL" ],  
"publisher" : { "name" : "NIRAV PRAKASHAN" },  
"price" : 25.30  
})
```

COMMAND:-

```
db.runCommand( {  
  mapReduce: "books",  
  map: function() {  
    for (var index = 0; index < this.authors.length; ++index) {  
      var author = this.authors[ index ];  
      emit( author.firstName + " " + author.lastName, 1 );  
    }  
  },  
  reduce: function(author, counters) {  
    count = 0;  
    for (var index = 0; index < counters.length; ++index)  
    {  
      count += counters[index];  
    }  
    return count;  
  },  
  out: { inline: 1 }  
})
```

***OUTPUT

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot - 360001
Ph. No - (0281)2471645

2 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

```
{
  "results" : [
    {
      "_id" : "BRIJESH PATEL",
      "value" : 1
    },
    {
      "_id" : "BRIJESH PATEL",
      "value" : 3
    },
    {
      "_id" : "NIKHIL DAVE",
      "value" : 2
    },
    {
      "_id" : "JIGNESH THANKI",
      "value" : 1
    },
    {
      "_id" : "VISHAL RANPARA",
      "value" : 1
    }
  ],
  "counts" : {
    "input" : 8,
    "emit" : 11,
    "reduce" : 2,
    "output" : 8
  },
  "ok" : 1
}
```

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot – 360001
Ph. No - (0281)2471645

2 – Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

Question	Answer
What is the full form or MR.	Map Reduce
_____ is a javascript function that maps a value with a key and emits a key-value pair	map
_____ is a data processing paradigm for condensing large volumes of data into useful <i>aggregated</i> results.	Map-reduce

Q-5 create hns database and create website collection with 10 documents and perform different aggregate function on it.

Use hns

```
Db.website.insert({"id":1, "domain name":"test1.com", "hosting":"google.com"})
```

Id	Domain name	Hosting
1	Test1.com	www.google.com
2	Test2.com	www.amazon.com
3	Test3.com	www.amazon.com
4	Test4.com	www.rediff.com
5	Test5.com	www.w3school.com
6	Test6.com	www.google.com
7	Test7.com	www.amazon.com
8	Test8.com	www.google.com
9	Test9.com	www.rediff.com
10	Test10.com	www.w3school.com

(1) Group):-

```
Db.website.aggregate(  
{  
$group: {_id:"$hosting", $total:{$sum:1}}  
}  
);
```


SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot – 360001
Ph. No - (0281)2471645

2 – Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

(2) sorting() :-

```
Db.website.aggregate(  
{  
$group: {_id:"$hosting", total: {$sum:1}}  
},  
{  
$sort: {total:-1}  
}  
);
```

Where -1 is for descending and +1 is for ascending order.

(3) match():-

```
Db.website.aggregate(  
{  
$match: {hosting: www.amazon.com}  
},  
{  
$group: {_id:"$hosting", total: {$sum:1}}  
}  
);
```

(4) unwind() :-

It is used to unwind document data using array. It is helpful to separate the document.

Inventory document

Id 1

Item abc

Size S,M,L

```
Db.inventory.aggregate([ {unwind:"$size" }])
```

Output:-

```
{“id”:1, “item”: “abc”, “size”:”S”}  
{“id”:1, “item”: “abc”, “size”:”M”}
```

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot – 360001
Ph. No - (0281)2471645

2 – Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

```
{“id”:1, “item”: “abc”, “size”:”L”}
```

(5) skip() :-

Syntax:- { \$skip:<positive integer> }

Db.artical.agggregate

```
{  
  Skip:5  
}
```

Output:- it will skip the first 5 documents passed to it by the pipeline.

(6) \$out():-

It takes the document return by the pipeline and write them to the specified collection.

Syntax:- { \$out: “<output – collection>” }

This method create a new collection in the database if it does not exist.

Example:-

Document book

Id	Title	Author	Copies
1	Abc	X	2
2	Xyz	Y	2
3	Pqr	X	1
4	Xyz	Z	1
5	Zzz	X	3

Db.book.agggregate

```
([ { $group: { _id:”author”, book: { $push:”$title” } } }, { $out:”$author” } ])
```

Output:-

```
{“_id”: “x”, “books” : [“a”,”b”]}
```



SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)

3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot – 360001
Ph. No - (0281)2471645

2 – Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

```
{“_id”: “y”, “books” :[“c”, “d”, “e”]}
```

(7) \$project():-

It is used to pass the specified field to the next stage in the pipeline.

Syntax:- { \$project: { <specifications> } }

```
Db.books.aggregate({ $project: { title:1, author:1 } })
```

Consider a book collection with the following document...

```
{  
Id-1, title –abc123, isbn-0001122223333, author-last zzz first aaa, copies -3}
```

Output:-

```
{“id”:1, “title”:"abc123", “author”:{“last”:"zzz", “first”:"aaa"}}
```

(8) \$limit():-

It is used to limit the no of documents passed to the next stage in the pipeline.

The \$limit stage has the following prototype form:

```
{ $limit: <positive integer> }
```

\$limit takes a positive integer that specifies the maximum number of documents to pass along.

Example

```
db.article.aggregate({ $limit : 5 } );
```

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot – 360001
Ph. No - (0281)2471645

2 – Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

This operation returns only the first 5 documents passed to it from by the pipeline. \$limit has no effect on the content of the documents it passes.

(9) count()

db.collection.count(*query*, *options*)

Returns the count of documents that would match a [find\(\)](#) query. The db.collection.count() method does not perform the [find\(\)](#) operation but instead counts and returns the number of results that match a query.

Parameter	Type	Description
query	document	The query selection criteria.
options	document	Optional. Extra options for modifying the count.

The options document contains the following fields:

Field	Type	Description
limit	Integer	Optional. The maximum number of documents to count.
skip	Integer	Optional. The number of documents to skip before counting.
hint	string or document	Optional. An index name hint or specification for the query.

To count the number of all documents in the orders collection, use the following operation:

db.orders.count()

This operation is equivalent to the following:

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot – 360001
Ph. No - (0281)2471645

2 – Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

```
db.orders.find().count()
```

output:- 4

Count all Documents that Match a Query

Count the number of the documents in the orders collection with the field ord_dt greater than new Date('01/01/2012'):

```
db.orders.count( { ord_dt: { $gt: new Date('01/01/2012') } } )
```

The query is equivalent to the following:

```
db.orders.find( { ord_dt: { $gt: new Date('01/01/2012') } } ).count()
```

(10) Distinct()

```
db.collection.distinct(field, query)
```

Finds the distinct values for a specified field across a single collection and returns the results in an array.

Parameter	Type	Description
field	string	The field for which to return distinct values.
query	document	A query that specifies the documents from which to retrieve the distinct values.

The db.collection.distinct() method provides a wrapper around the distinct command. Results must not be larger than the maximum *BSON size*.

Array Fields

If the value of the specified field is an array, db.collection.distinct() considers each element of the array as a separate value.



SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)

3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot – 360001
Ph. No - (0281)2471645

2 – Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

For instance, if a field has as its value [1, [1], 1], then `db.collection.distinct()` considers 1, [1], and 1 as separate values.

For an example, see *Return Distinct Values for an Array Field*.

Index Use

When possible, `db.collection.distinct()` operations can use indexes.

Indexes can **also cover `db.collection.distinct()` operations.**

Examples

The examples use the inventory collection that contains the following documents:

```
{ "_id": 1, "dept": "A", "item": { "sku": "111", "color": "red" }, "sizes": [ "S", "M" ] }
{ "_id": 2, "dept": "A", "item": { "sku": "111", "color": "blue" }, "sizes": [ "M", "L" ] }
{ "_id": 3, "dept": "B", "item": { "sku": "222", "color": "blue" }, "sizes": "S" }
{ "_id": 4, "dept": "A", "item": { "sku": "333", "color": "black" }, "sizes": [ "S" ] }
```

Return Distinct Values for a Field

The following example returns the distinct values for the field `dept` from all documents in the inventory collection:

```
db.inventory.distinct( "dept" )
```

The method returns the following array of distinct `dept` values:

```
[ "A", "B" ]
```

Return Distinct Values for an Embedded Field

The following example returns the distinct values for the field `sku`, embedded in the `item` field, from all documents in the inventory collection:

```
db.inventory.distinct( "item.sku" )
```

The method returns the following array of distinct `sku` values:

```
[ "111", "222", "333" ]
```



SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)

3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot – 360001
Ph. No - (0281)2471645

2 – Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

Return Distinct Values for an Array Field

The following example returns the distinct values for the field sizes from all documents in the inventory collection:

```
db.inventory.distinct( "sizes" )
```

The method returns the following array of distinct sizes values:

```
[ "M", "S", "L" ]
```

Specify Query with distinct

The following example returns the distinct values for the field sku, embedded in the item field, from the documents whose dept is equal to "A":

```
db.inventory.distinct( "item.sku", { dept: "A" } )
```

The method returns the following array of distinct sku values:

```
[ "111", "333" ]
```

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot – 360001
Ph. No - (0281)2471645

2 – Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

Chapter-5

Handling Files with GridFS

- **Handling Files with GridFS**
- What is Grid?
- Storing files in GridFS
- Serving files from GridFS
- Reading files in chunks

- **Database Management**
- Database Administration
- Optimization
- Replication
- shade

Q - 1 Explain Grid.

A Grid class is representation of a simple grid interface.

Gridfs is utility for storing and retrieving files from the database.

Syntax:- class Grid(db, fs name)

This Grid class have 2 arguments

1 db(db):- a database instance to interact with

2 [FS name](string):- it is optional argument that provide root collection for gridFS.

This class return a grid as a output.

Grid support different function to perform different operation like.

(1) put() :- it puts binary data to the grid.

(2) get():- It gets binary data to grid.

(3) delete():- it delete the file from the grid.

(1) put():-

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot – 360001
Ph. No - (0281)2471645

2 – Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

It is used to put the binary data to the grid.

Syntax:- put(data[,optional],callback)

Here in the argument

Data is buffer data with binary data.

[optional] argument include different options for file like, _id, root, context type, chunk size or metadata.

With the callback will be called after this method is executed.

(2) get():-

This method is used to get binary data to the grid.

Syntax:- get(id, callback)

Here the 2 arguments have

Id is used for any file.

With the callback will be called after this method is executed.

(3) delete():-

It is used to delete the file from the grid.

Syntax:- delete(id, callback)

Id is used for any file.

With the callback will be called after this method is executed.

Q-2 what is grid fs?

- GridFS is the mongodb specification for storing and retrieving large files such as images, audio files, video, sound recorder etc.
- It is a file system to store the files but its data is stored within mongodb collection.
- GridFS has a capability to store files even greater than its document size.
- GridFS by default use two collection.
- FS.Files and FS.chunks to storing the data in grid files system.
- GridFS divided a file into chunks and store each chunks and store each chunks of data in a separate document with maximum size 255kb.
- The GridFS is used for storing the files larger than 16 MB.
- You can use GridFS to store as many files as needed in a directory.
- The GridFS is also helpful to access a portion of large file without loading the whole file into memory.
- GridFS is also used when you want to keep your file and metadata automatically display number of system.

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot – 360001
Ph. No - (0281)2471645

2 – Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

- GridFS support two collection to store file metadata and the chunks
- Each chunks is identified by its unique object id field.
- Hear the FS.Files work as a parent document and the FS chunks document can link with files_id field to its parent.

Advantage:-

1. GridFS can simply your stack.
2. GridFS can used instead of a separate tool for file storage.
3. GridFS is helpful and easy in fail over and scale out situation.
4. GridFS support large number of files to be store in the same directory.

Disadvantage:-

1. The performance is slower for accessing the files.
2. You can only modify the document by deleting them and resaving again.
3. It cannot lock all the chunks in a files at the same time.

GridFS is not used in 2 situation:-

1. If the files are smaller than 16 MB.
2. If you need to update content of the entire file automatically.

• Use GridFS:-

- To perform the GridFS operation for storing and retrieving files you have to use.

1. A mongodb driver
2. Mongo files command line tools.

• GridFS collection:-

- the GridFS store files in 2 collection

 1. the chunks collection(binary chunks)
 2. the file collection (metadata)

- GridFS place the collection in a common bucket which by default start with FS name.

- Ex:- FS.files & FS.chunks

When to Use GridFS

In MongoDB, use GridFS for storing files larger than 16 MB.

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot – 360001
Ph. No - (0281)2471645

2 – Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

In some situations, storing large files may be more efficient in a MongoDB database than on a system-level file system.

- If your file system limits the number of files in a directory, you can use GridFS to store as many files as needed.
- When you want to access information from portions of large files without having to load whole files into memory, you can use GridFS to recall sections of files without reading the entire file into memory.
- When you want to keep your files and metadata automatically synced and deployed across a number of systems and facilities, you can use GridFS. When using geographically distributed replica sets, MongoDB can distribute files and their metadata automatically to a number of mongod instances and facilities.

Do not use GridFS if you need to update the content of the entire file atomically. As an alternative, you can store multiple versions of each file and specify the current version of the file in the metadata. You can update the metadata field that indicates “latest” status in an atomic update after uploading the new version of the file, and later remove previous versions if needed.

Furthermore, if your files are all smaller than 16 MB BSON Document Size limit, consider storing the file manually within a single document instead of using GridFS. You may use the BinData data type to store the binary data.

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot – 360001
Ph. No - (0281)2471645

2 – Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

Question	Answer
Which class is represent simple grid?	Grid class
Put() method is used for _____	It is used to put binary data to the grid.
Fs stands for _____	File system
Gridfs utility is used for _____	Storing and retrieving large files from the database.
Gridfs by default have ___collection	Two
Maximum size of chunk is ?	255 kb
Gridfs is used for storing the files larger than _____	16 MB
Gridfs is not used if _____	The files are smaller than 16 MB and if you need to update content of the entire file automatically.
Gridfs is used in fail over and scale out situation. True or False	True
Give the default name of Gridfs collection	Fs.files and fs.chunks

To store and retrieve files using *GridFS*, use either of the following:

Use GridFS files

- A MongoDB driver.
- The mongofiles command-line tool.

Use GridFS Collections

GridFS stores files in two collections:

- chunks stores the binary chunks.
- files stores the file's metadata.

GridFS places the collections in a common bucket by prefixing each with the bucket name. By default, *GridFS* uses two collections with a bucket named fs:

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot – 360001
Ph. No - (0281)2471645

2 – Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

- fs.files
- fs.chunks

You can choose a different bucket name, as well as create multiple buckets in a single database. The full collection name, which includes the bucket name, is subject to the namespace length limit.

The chunks Collection

Each document in the chunks collection represents a distinct chunk of a file as represented in *GridFS*. Documents in this collection have the following form:

```
{  
  "_id" : <ObjectId>,  
  "files_id" : <ObjectId>,  
  "n" : <num>,  
  "data" : <binary>  
}
```

A document from the chunks collection contains the following fields:

chunks._id

The unique *ObjectId* of the chunk.

chunks.files_id

The _id of the “parent” document, as specified in the files collection.

chunks.n

The sequence number of the chunk. GridFS numbers all chunks, starting with 0.

chunks.data

The chunk’s payload as a *BSON* Binary type.

The files Collection

Each document in the files collection represents a file in [GridFS](#). Consider a document in the files collection, which has the following form:

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot - 360001
Ph. No - (0281)2471645

2 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

```
{
  "_id" : <ObjectId>,
  "length" : <num>,
  "chunkSize" : <num>,
  "uploadDate" : <timestamp>,
  "md5" : <hash>,
  "filename" : <string>,
  "contentType" : <string>,
  "aliases" : <string array>,
  "metadata" : <dataObject>,
}
```

Documents in the files collection contain some or all of the following fields:

files._id

The unique identifier for this document. The `_id` is of the data type you chose for the original document. The default type for MongoDB documents is *BSON ObjectId*.

files.length

The size of the document in bytes.

files.chunkSize

The size of each chunk in **bytes**. GridFS divides the document into chunks of size `chunkSize`, except for the last, which is only as large as needed. The default size is 255 kilobytes (kB).

Changed in version 2.4.10: The default chunk size changed from 256 kB to 255 kB.

files.uploadDate

The date the document was first stored by GridFS. This value has the Date type.

files.md5

An MD5 hash of the complete file returned by the `filemd5` command. This value has the String type.

files.filename

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot – 360001
Ph. No - (0281)2471645

2 – Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

Optional. A human-readable name for the GridFS file.

files.contentType

Optional. A valid MIME type for the GridFS file.

files.aliases

Optional. An array of alias strings.

files.metadata

Optional. Any additional information you want to store.

Applications may create additional arbitrary fields.

Question	Answer
Gridfs places the collections in ____	Store collections in a common bucket. default bucket names fs: ____
What is chunks._id.?	It is unique objectid of the chunk.
What is chunks.n?	It is sequence number of chunks. starting from 0.
What is files._id?	The unique identifier for this document.

MongoDB GridFS Indexes

GridFS uses indexes on each of the chunks and files collections for efficiency. *Drivers* that confirm to the GridFS specification automatically create these indexes for convenience. You can also create any additional indexes as desired to suit your application's needs.

There are two types of Indexes.

1. The chunks Index
2. The file index

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot – 360001
Ph. No - (0281)2471645

2 – Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

The chunks Index

GridFS uses a *unique, compound* index on the *chunks* collection using the *files_id* and *n* fields. This allows for efficient retrieval of chunks, as demonstrated in the following example:

```
db.fs.chunks.find( { files_id: myFileID } ).sort( { n: 1 } )
```

Drivers that conform to the *GridFS* specification will automatically ensure that this index exists before read and write operations.

If this index does not exist, you can issue the following operation to create it using the mongo shell:

```
db.fs.chunks.createIndex( { files_id: 1, n: 1 }, { unique: true } );
```

The files Index

GridFS uses an *index* on the *files* collection using the *filename* and *uploadDate* fields. This index allows for efficient retrieval of files, as shown in this example:

```
db.fs.files.find( { filename: myFileName } ).sort( { uploadDate: 1 } )
```

Drivers that conform to the *GridFS* specification will automatically ensure that this index exists before read and write operations.

If this index does not exist, you can issue the following operation to create it using the mongo shell:

```
db.fs.files.createIndex( { filename: 1, uploadDate: 1 } );
```

[1] The use of the term *chunks* in the context of *GridFS* is not related to the use of the term *chunks* in the context of sharding.

Q-3 Explain gridFS command

- gridFS support all the command of grid class which are as under.

 1. List:- It display the list of all files.
 2. Search:- It search al files you can added file name as sub string for the search operation.
 3. Put:- It added a file with name.
 4. Get:- It used to get a file with file name.
 5. Delete :- It delete all files or file with file name.

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot – 360001
Ph. No - (0281)2471645

2 – Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

Q4 Explain storing files, reading files and serving file with gridFS

- Mongofiles command is used to perform different operation with gridFS like storing and retrieving files from local system to gridFS.
- Mongofiles is a separate command so you have to perform the mongofiles command by opening command prompt.
- Here you have a command prompt for server(mongod), one command prompt for client(mongo) and one command prompt for mongofiles.
- Mongofiles include following commands.

--help:-

- Ex:- mongofile --help
- o/p :-return all supported command with syntax.

--version:-

- Ex:-this command is used to return the version of mongofiles.

List:-

- It will display the file list of mongo connected to local host.
- It also display the size of file.
- Sy:- mongofiles –d<dbname>list
- Ex:- mongofiles –d test list.

Put:-

- If you want to add or store the file into grideFS then put command is used.
- You can store different types of files like text files, image file, mp3, etc...
- First you have to put some text file for example “demo.txt” into bin directory now if you want to add this file into grideFS then follow the following syntax.
- Sy:- mongofile –d<db name> put<file name>
- Ex:- mongofile –d test put demo.txt.
- o/p :- add file demo.txt connected to localhost.

```
mongofiles.exe -d gridfs put song.mp3
```

Find:-

- Display files with the help of find command.
- If you want to display FS.files in the client windows then put this command
- `db.fs.files.find().pretty()` in the client dos screen.
- o/p:- -id, chunks size, upload data, length, md5 and filename will display as a output.

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot – 360001
Ph. No - (0281)2471645

2 – Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

- If u want to display the information about files from chunks collection then you have to write the command in the client dos screen like
- `db.fs.chunks.find().pretty()`
- It display the output like `_id`, `file_id`, `n` and `data`.

Delete:-

- If you want to delete a file of grideFS then you have to use the delete command in the mongofiles command prompt.
- Sy:- `mongofiles -d<dbname> delete<file name>`
- Ex:- `mongofiles -d hns delete demo.txt`
- o/p:- connected to localhost successfully delete all instance of “demo.txt” from grideFS.
`db.fs.files.remove({})`
`db.fs.chunks.remove({})`
`db.repairDatabase()` (cleans out the memory)

Get command & mongofile:-

- The get command is used to retrieve the file from grideFS to the local system.
- It does not display the content of file directly.
- Sy:- `mongofiles -d<db name> get<file name>`
- Ex:- `mongofiles -d hns get demo.txt`

<?php

```
$c=new MongoClient();
```

```
$db=$c->hns;
```

```
$gets=$db->getgrideFS();
```

```
Echo $gets->findOne(array("filename"=>"demo.txt"))->getBytes();    ?>
```

Note:- The mongofiles utility makes it possible to manipulate files stored in your MongoDB instance in GridFS objects from the command line. It is particularly useful as it provides an interface between objects stored in your file system and GridFS.

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot – 360001
Ph. No - (0281)2471645

2 – Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

| Question | Answer |
|---|--|
| There are ____types of indexes in gridfs indexes. | Two |
| What is chunks index? | It is unique index of the chunks for easy retrieval of chunks. |
| What is file index? | It is the index of the files used in gridfs. |
| Gridfs support all the commands of Grid class. True or False. | True |
| Mongofiles command is used for_____ | This command is used to perform storing and retrieval files from local system to gridfs. |
| --help command is used for_____ | It returns all supported command with syntax. |
| List command is used for_____ | It is used to display the file list of mongo connected to local host. |
| Get command is used for_____ | It is used to retrieve the file from gridfs to the local system. |
| | |

Sharding GridFS

There are two collections to consider with gridfs - files and chunks. If you need to shard a GridFS data store,

use the chunks collection setting { files_id : 1, n : 1 } or { files_id : 1 } as the shard key index.

files_id is an objectid. When you use sharding concept then the files collection use the _id field, possibly in combination with an application field.

You cannot use Hashed Sharding when sharding the chunks collection.

The files collection is small and only contains metadata.

Leaving files unsharded allows all the file metadata documents to live on the primary shard.

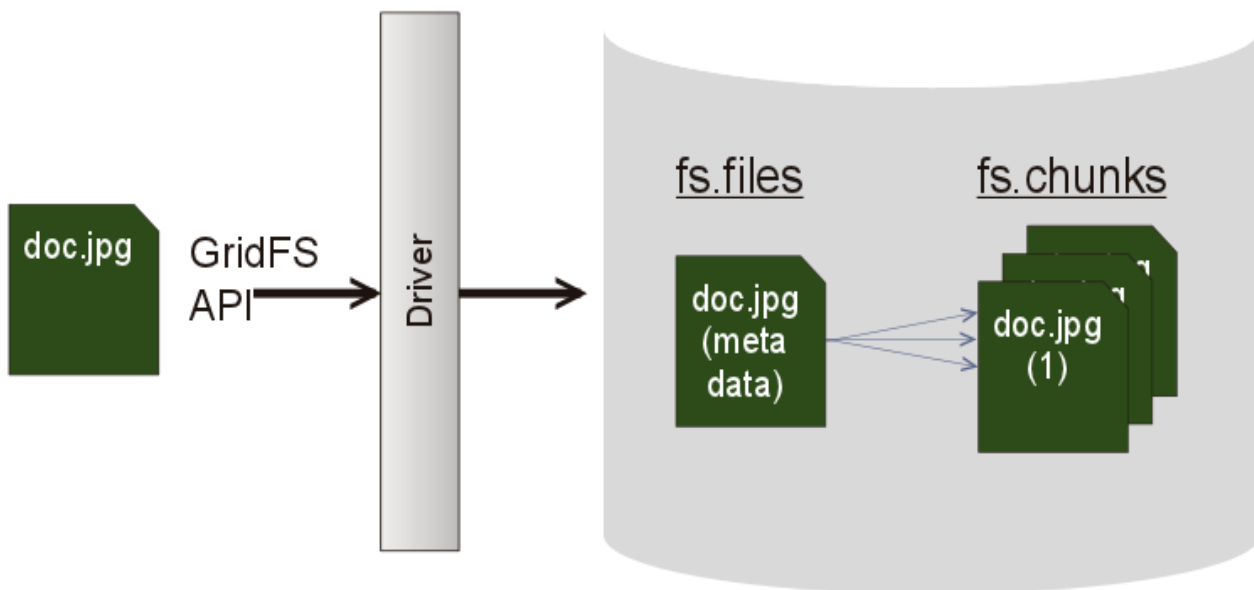
SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot - 360001
Ph. No - (0281)2471645

2 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478



Possible commands include:

- list - list all files; 'filename' is an optional prefix which listed filenames must begin with...
- search - search all files; 'filename' is a substring which listed filenames must contain
- put - add a file with filename 'filename'
- get - get a file with filename 'filename'
- get_id - get a file with the given '_id'
- delete - delete all files with filename 'filename'
- delete_id - delete a file with the given '_id'

(1) C:\MongoDB>mongofiles --help

Output:-

general options:

- /help print usage
- /version print the tool version and exit

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot - 360001
Ph. No - (0281)2471645

2 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

verbosity options:

/v, /verbose more detailed log output (include multiple times for more verbosity, e.g. -vvvvv)
/quiet hide all log output

connection options:

/h, /host: mongodb host to connect to
(setname/host1,host2 for replica sets)
/port: server port (can also use --host hostname:port)

authentication options:

/u, /username: username for authentication
/p, /password: password for authentication
/authenticationDatabase: database that holds the user's credentials
/authenticationMechanism: authentication mechanism to use

storage options:

/d, /db: database to use (default is 'test')
/l, /local: local filename for put|get
/t, /type: content/MIME type for put (optional)
/r, /replace remove other files with same name after put
/prefix: GridFS prefix to use (default is 'fs')
/writeConcern: write concern options e.g. --writeConcern majority,
--writeConcern '{w: 3, wtimeout: 500, fsync: true, j: true}' (defaults to

(2) C:\MongoDB>mongofiles --version

Output:-

mongofiles version: r3.0.12-15-gc0066f6

git version: 550d8c0009702cb8a651e94a0cd6034e276ada84

(3) C:\MongoDB>mongofiles list

Output:-

2016-10-01T14:03:55.385+0530 connected to: localhost

demo.txt 18

(4) C:\MongoDB>mongofiles put demo.txt

Output:-

2016-10-01T14:13:19.914+0530 connected to: localhost

added file: demo.txt

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot - 360001
Ph. No - (0281)2471645

2 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

(5) db.fs.files.find().pretty() -----type this command to the client dos screen

Output:-

```
{
  "_id" : ObjectId("57ef77278076990700000001"),
  "chunkSize" : 261120,
  "uploadDate" : ISODate("2016-10-01T08:43:19.917Z"),
  "length" : 18,
  "md5" : "cd0f7e804d7dcc24240ed8d41750faf6",
  "filename" : "demo.txt"
}
{
  "_id" : ObjectId("57ef77748076990de8000001"),
  "chunkSize" : 261120,
  "uploadDate" : ISODate("2016-10-01T08:44:36.731Z"),
  "length" : 18,
  "md5" : "cd0f7e804d7dcc24240ed8d41750faf6",
  "filename" : "demo.txt"
}
```

(6) db.fs.chunks.find() .pretty() -----type this command to the mongoclient dos screen

Output:-

```
{
  "_id" : ObjectId("57ef77278076990700000002"),
  "files_id" : ObjectId("57ef77278076990700000001"),
  "n" : 0,
  "data" : BinData(0,"aGkNCmhvdyANCmFyZQ0KeW91")
}
{
  "_id" : ObjectId("57ef77748076990de8000002"),
  "files_id" : ObjectId("57ef77748076990de8000001"),
  "n" : 0,
  "data" : BinData(0,"aGkNCmhvdyANCmFyZQ0KeW91")
}
```

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot – 360001
Ph. No - (0281)2471645

2 – Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

(7) C:\MongoDB>mongofiles -d hns delete demo.txt

Output:-

2016-10-03T15:07:16.385+0530 connected to: localhost
successfully deleted all instances of 'demo.txt' from GridFS

Storing Files:-

You use the storeUpload() function to store files into your database with GridFS. This function takes two parameters:

one indicates the fieldname of the file to be uploaded, and the other can optionally be used to specify the file's metadata.

Once used, the function reports back the _id of the file stored.

The following simple code example shows how to use the storeUpload() function:

```
// Connect to the database
$c = new MongoClient();
// Select the name of the database
$db = $c->contacts;
// Define the GridFS class to ensure we can handle the files
$gridFS = $db->getGridFS();
// Specify the HTML field's name attribute
$file = 'fileField';
// Specify the file's metadata (optional)
$metadata = array('uploadDate' => date());
// Upload the file into the database
$id = $gridFS->storeUpload($file,$metadata);
```

Update Files:-

Note:- Same code as above but just use update() method instead of storeUpload()

```
// Specify the metadata to be added
$metadata = array('$set' => array("Tag" => "Avatar"));
// Specify the search criteria to which to apply the metadata
$criteria = array('_id' => $id);
// Insert the metadata
$db->grid->update($criteria, $metadata);
```

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot – 360001
Ph. No - (0281)2471645

2 – Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

Retrieving Files:-

Of course, the ability to store your files in a database wouldn't do you any good if you weren't able to retrieve these files later.

Retrieving files is as easy as storing them.

The following example shows how to retrieve the filenames stored, which you accomplish using the `getFilename()` function:

```
// Connect to the database
$c = new MongoClient();
$db = $c->contacts;
// Initialize GridFS
$gridFS = $db->getGridFS();
// Find all files in the GridFS storage and store under the $cursor parameter
$cursor = $gridFS->find();
// Return all the names of the files
foreach ($cursor as $object) {
echo "Filename:". $object->getFilename();
}
```

Deleting Data:-

You can ensure that any previously stored data is removed by using the `delete()` function. This function takes one parameter: the `_id` of the file itself.

The following example illustrates how to use the `delete()` function to delete the file matching the Object ID `4c555c70be90968001080000`:

```
// Connect to the database
$c = new MongoClient();
// Specify the database name
$db = $c->contacts;
// Initialize GridFS
$gridFS = $db->getGridFS();
// Specify the file via it's ID
$id = new MongoClient('4c555c70be90968001080000');
$file = $gridFS->findOne(array('_id' => $id));
// Remove the file using the remove() function
$gridFS->delete($id);
```

Note:- <https://pymongo.readthedocs.io/en/3.6.0/api/gridfs/index.html>

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot – 360001
Ph. No - (0281)2471645

2 – Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

| Question | Answer |
|---|--|
| What is Sharding? | Sharding is the process of storing data records across multiple machines and it is mongodb's approach to meeting the demands of data growth. |
| storeUpload() method is used for _____ | Storing the data. |
| Delete_id() include which parameter? | Id |
| Search command is used for _____ | Search all files. |
| \$cursor = \$gridFS->find();
This line is used for _____ | Find all files in the GridFS storage and store under the \$cursor parameter |
| get_id is used for _____ | It is used to get a file with the given '_id' |

Q5 Explain database administration in mongodb.

- Mongodb provides different functionality for database administration. This functionality are used to keep the track and control on the database resource.
- The database administration provides the operation and maintains of mongodb instance.
- Database administration provide configuration operation and deployment process and many different task which are as under.
 1. Administration
 2. Deployment
 3. Performance & Monitoring
 4. Optimization
 5. Replication etc....

Administration:-

- Every database management system support the administration of database.
- It means that hear all the data and different use are managed properly.
- The administration task include new user creation, security, implementation of database model, modulating, backup & recovery of data.

Deployment:-

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot – 360001
Ph. No - (0281)2471645

2 – Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

- It is a one type of installation or implementation process where user allows to write different script code for your data and database.
- Hear the query can be performed in database.
- The indexing, database designing and modules are implemented hear.

Performance & monitoring:-

- Database administration (DBA) must retrieve the performance of database using different tools.
- DBA generate the report for database performance.
- DBA is also responsible for generate and create different user roles and perform basic database administrative task.

Q-6 Explain mongodb database administration basic task.

- In mongodb there are different basic task for database administration which are as under.

Starting and stopping mongodb server & client:-

- Hear you have to start the server first with mongod.exe file.
- You have to start client by using mongo.exe.
- There are different command line option used with mongod command like.
- --help
- --dbpath
- --port
- --config etc...
- If you want to stop mongodb server then you have to use the shutdown command.
- Db.shutdownServer()

File based configuration:-

- Mongodb support reading configuration information from a file the name of this file is mongodb.config.
- The configuration is automatically done by mongodb server.

Monitoring:-

- Hear the mongodb database perform will be checked by monitoring functionality.
- Accessing the data in memory is fast and accessing the date on disk is slow so hear you have to monitor how mongodb interact with disk and memory.

Server state:-

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot – 360001
Ph. No - (0281)2471645

2 – Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

- The server status command is used to provide a static information about mongodb server.

Security:-

- Hear you ensure that your system and data is secure with mongodb server and environment.
- You can setup the firewall also for allow mongodb to connect with internet network and client only.
- Mongodb support authentication like username and password for query connections.
- You can used the command like
 - --bind_it
 - --nohttpinterface
 - --noinet
 - --noscripting

Configuration & Maintenance:

- Hear you can perform different management operation, configuration performance and analysis task.

Login or Logfiles:-

- The log file are used to record all the changes in the database.
- By default mongod send its logs to STD out.
- Most init script use --logpath option to send logs to a file.

Backup & Recovery:-

- It is important to take regular backup of your system so if any failure occurs then you can restore your data back.
- You can use mongodump command for backup purpose and mongorestore command for recovery purpose.

Authentication Task:-

- Each database in mongodb have number of user so it provides security for mongodb communication for reading and writing client and server.
- You can use the command show users and show roles in admin database.

Q.7 write a note on optimization.

Ans: --> optimization is and act or process to make something is fully perfect.
--> It is a simple method to perform a design or a system as effective as possible.
-->there are many factors which effect the db performance and responsiveness.

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot - 360001
Ph. No - (0281)2471645

2 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

-->optimization is used to analyze the db its performance, indexing, query structure, data model, application design, operational factor, architecture, system configuration etc.

-->optimization is used with mongo db including different types of category like.

- 1) Analyzing mongodb (analyze number of connection, hardware, software etc)
- 2) To check query performance in mongodb (with db.profiler)
- 3) To check performance of current operation (with the help of methods)
- 4) Optimize query performance in mongodb (with indexing)
- 5) Optimize different design notes with (dynamic schema, dynamic data types, case sensitivity)

1) Analyzing mongodb

-->you can developed and operate the application with mongo db then some time you may need analyze the performance of application and its db.

-->here you need to process for function of db.number of open db connection, hardware, software, configuration etc.

2) To check mongodb performance & 3) check performance of current operation

--> There are different techniques for evaluate the performance of mongo db with different operations.

--> Mongo db provide **db profiler** that show characteristic of each operation of db.

--> Profiler use to locate different queries for write operation those are running.

--> Profiler provides two different methods which are as under.

a) Db.currentop ()

-->this method is used for display the record of current operation running on mongodb instance.

b) Db.explain ()

-->the explain command provides detail information about a given query.path in short it return the information stats of query execution.

Example

>use stud

>db.currentop ();

>db.mscit.find ({"rno":1}).explain ("execution stats");

4) Optimize query performance

-->the optimization is very useful for faster processing of multiple queries or a query on multiple fields here it can allow to create the index for searching operations.

-->by using the index structure the query **execution will be fast** and the index contents very small space in memory.

Syntax

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot - 360001
Ph. No - (0281)2471645

2 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

Db. <collname>.createindex ({field name});

Example

db.mscit.createindex ({"name":1});

5) Optimize different design notes

-->during the design notes in optimization technique include different types of data related functionality.

-->it include 3 types of functionality which are as under.

a) Dynamic schema

-->mongodb provides dynamic schema types db and data show there is no need to define the fix structure.

-->due to dynamic schema you can get the benefit of polymorphism and iterative development.

b) Case sensitive string

-->mongo db screen are case sensitive show STUD & stud is different.

c) Data type sensitive field

-->mongo db support different data types as per it value given by the user.

Example

```
{"rno":1} //number
```

```
{"rno":"1"} //string
```

Q.8 Explain mongodb replication.

Ans :

--> replications the process or a part of backup and recovery concept.

-->it is multiple copies of your data or db which is useful at the time of failover conditions.

-->replication protect the db from lost of data.

-->replication is the process of synchronizing the data across multiple servers.

-->replication allows you to recover the data from hardware failure or some software service damage.

-->replication provides additional copies of data so it can be use for disaster recovery, b'up and reporting.

-->replication provides multiple copies of data on different servers and it will automatically get the updates of changes.

Replication advantage

1) Its keep your data safe.

2) The data can be available at any time.

3) It provides disaster recovery.

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)

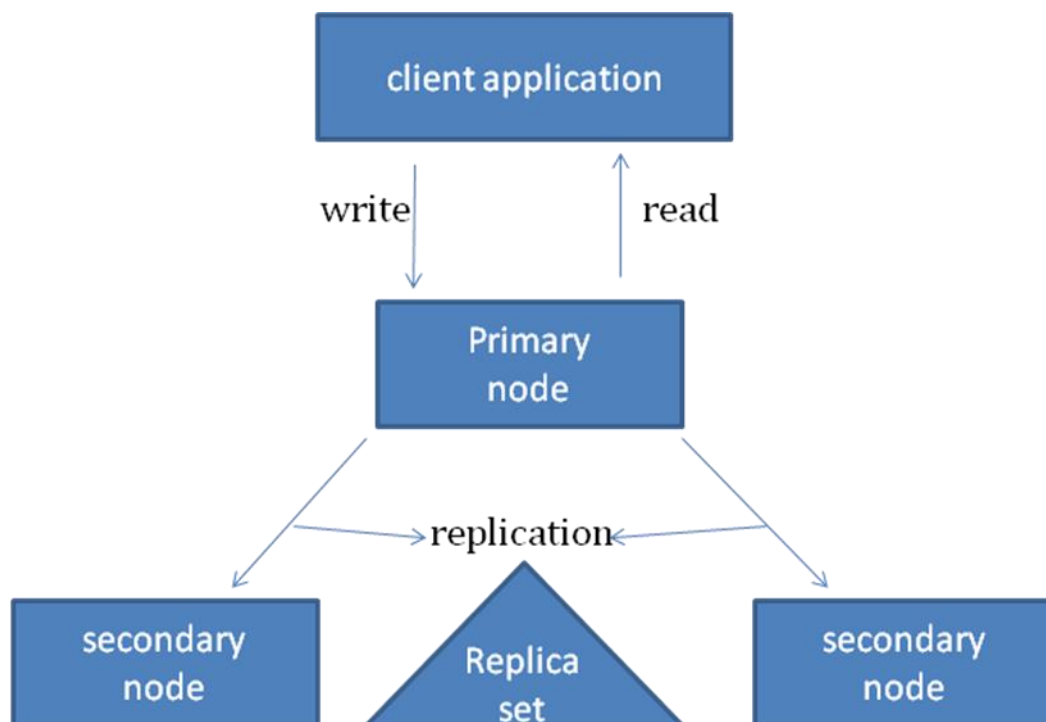


3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot - 360001
Ph. No - (0281)2471645

2 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

- 4) It cannot recover to down the server for maintains.
- 5) It provides scaling of data.
- 6) It is helpful for backup and recovery.

Replication work in mongodb Diagram



-->in mongodb replication is done by using replica set.

-->replica set is group of mongod instance.

-->in replica set one node is primary node and all other nodes are secondary node.

Primary node

-->this node receive all writing operation from other instances.

Secondary node

-->this node contains all other operation from primary node so they have the same data set of primary node.

Rule for replica set

-->replica set can have only one primary node.

-->replica set is a group of two or more nodes.

-->all data replicates from primary to secondary.

-->at the time of failover of primary node automatically another node can selected and that secondary node worked as primary node.

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot – 360001
Ph. No - (0281)2471645

2 – Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

Features of replica set

- >it provides cluster for number of nodes.
- >any one node can be primary and remaining are secondary.
- >all writing operations goes throw primary node.
- >provide automatic failure.
- >provide automatic recovery.

Steps for replication

Step 1: start up a mongo shell with the **--nodb** option, it means that this shell is not connected with any mongod.

(Shutdown the running server and start mongod server by using replica set command)

Step 2: write the following command.

```
>mongod--port27017-db path "c:/data" -- replicaset stud-demo
```

Or if you want to create replica set then write the command

```
Replicas=new replica set demo ({"nodes":3})
```

-->it will start mongod instance with stud-demo name

Or

It start replica set demo with one primary and two secondary node.

Q . 9 Explain sharding

Ans:

- >Sharding is the process of storing data records across multiple machines.
- >In mongodb the day by day size of the data increase so a single machine may not be sufficient to store the data.
- >sharding is used to solve the problem with scaling and accept multiple machines to support data growth and reading or writing operation.

Why sharding

- >in replication all writing operation performs in master node.
- >single replica set has limitation of 12 node.
- >memory cannot be large enough when archive data set is big.

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot - 360001
Ph. No - (0281)2471645

2 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

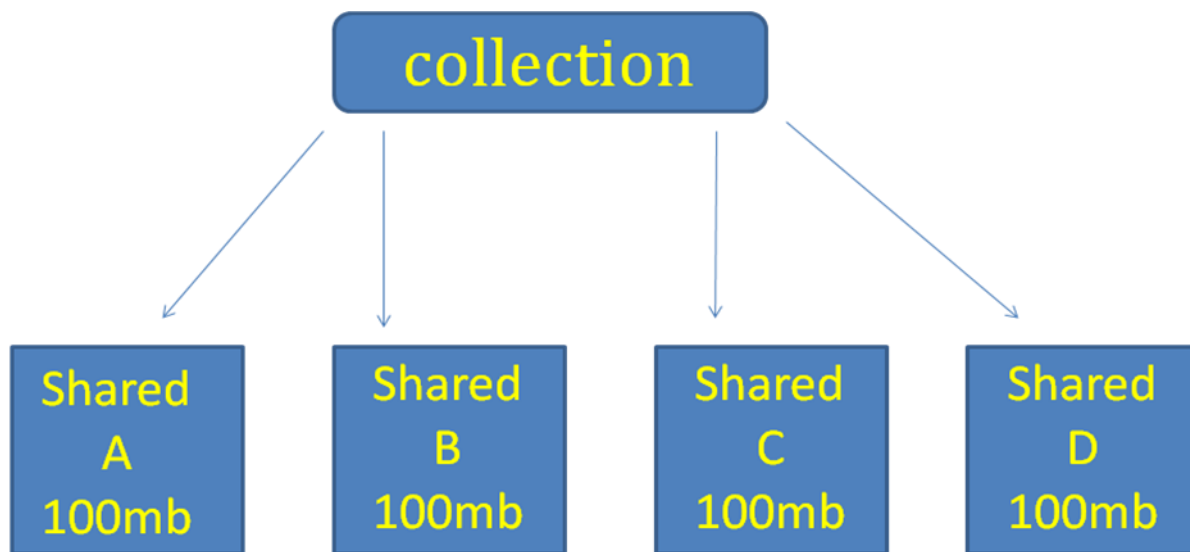
-->mongodb use sharding to support deployment with very large data set and reading and writing operation.

-->sharding provides the data set and distribution of data over multiple server (shard).

-->each shard is independent data base.

-->only local disk is not capable to store big data so sharding is require.

Diagram



-->shard are used to store the data in production environment each shard is a separate replica set.

-->in mongodb you can use query router and config server to perform operation and clustering of Meta data.

Sharding characteristic

-->sharding reduce the load.

-->sharding reduce the number of operation.

-->sharding reduce the amount of data for each server needs to store.

Components of sharding

-->there are 3 components of sharding.

- 1) Shard
- 2) Query router
- 3) Config server

1) Shard

-->it is used to store the data.

-->a mongodb shard cluster distributes data across one or more shards.

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot – 360001
Ph. No - (0281)2471645

2 – Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

-->each shard is deploys as a mongodb replica set.
-->the replica set store some part of the total data.

2) query router (mongos router)

-->the query router is also known as mongos instances.
-->each shard contains part of the cluster data then also you need one interface for operation of cluster as a whole.
-->the mongos process is a router that directs all reads and write to the appropriate shard.
-->it is the interface between client application and sharding operation then it returns the results to the client.
-->sharding contents more than one query router to divide client request load.
-->client can send request to one query router.

3) Config server

-->the config server is used to store the cluster meta data.
-->config server to store the shard clusters state.
-->this meta data include the global cluster configuration the location of each database, collection, the particular range of data and the history of the data across the shard.
-->every time mongos process is started then the mongos fetches a copy of the meta data from the config server.

-->you can get an overall view of you cluster by running sh. Status().
-->it will give you a summary of your shards, data base and collections.

Diagram of sharding cluster with 3 config server

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
 Nr. Amrapali Railway Crossing,
 Raiya road,
 Rajkot - 360001
 Ph. No - (0281)2471645

2 - Vaishalinagar
 Nr. Amrapali Railway Crossing,
 Raiya Road,
 Rajkot - 360001
 Ph.No - (0281)2440478

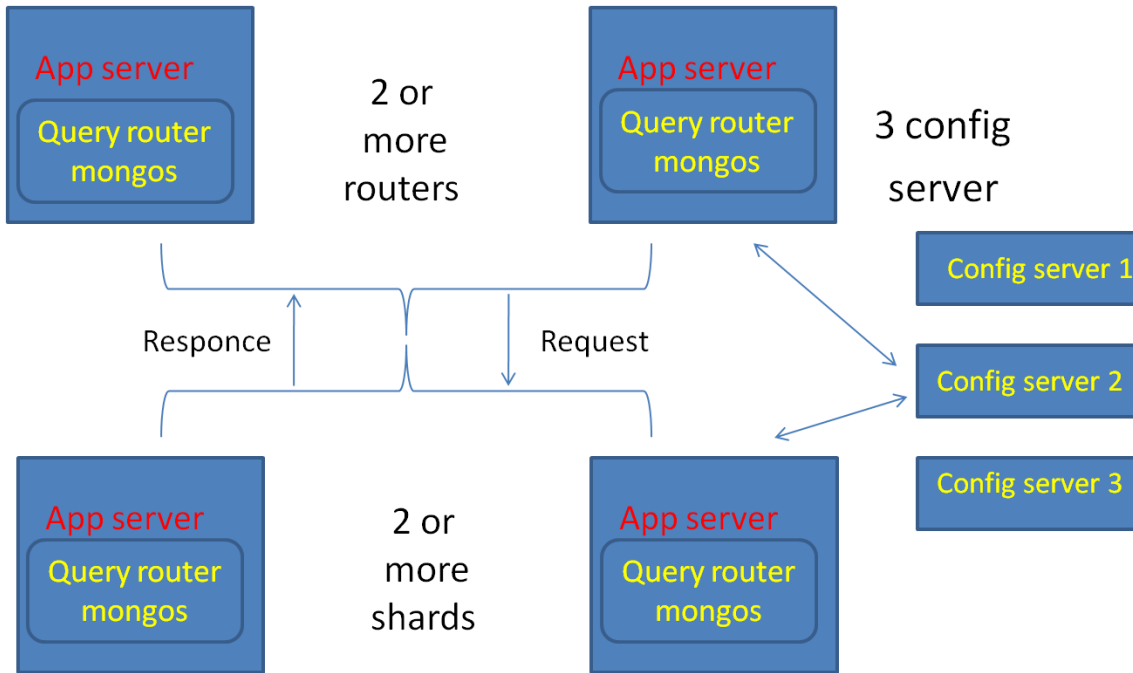
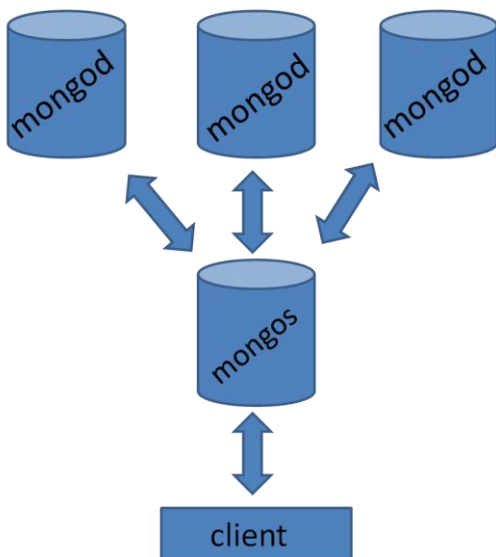


Diagram of sharding

Shared client connection



SHREE H. N. SHUKLA GROUP OF COLLEGES

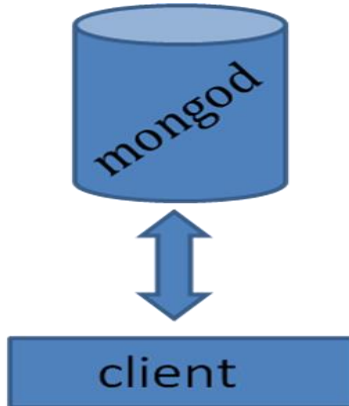
(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot – 360001
Ph. No - (0281)2471645

2 – Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

No shared client connection



| Question | Answer |
|-------------------------|---|
| What is deployment? | It one type of installation or implementation process. |
| What is optimization? | It is an act or process to make something is fully perfect for performance, indexing, data model, analysis. |
| DBA stands for_____ | Database Administration |
| What is replication? | It is the copy of data for backup and recovery purpose. |
| What is primary node? | Primary node received all writing operation from other instances. |
| What is secondary node? | These nodes contains all other operation from primary node so they have the same data set of primary node. |
| Why sharding? | Sharding is used to store data day by day data increased so sharding solve the problem of increasing data by accepting multiple machines. to store bigdata sharding is require. |
| | |

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot – 360001
Ph. No - (0281)2471645

2 – Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

Some IMP topics.

CHAPTER - 1

(1) Difference between SQL and NoSQL

Ever since the 1970s, SQL databases have been an integral part of the IT infrastructure of organizations. And today MySQL, an RDBMS based SQL implementation for the web, now powers very large-scale websites like Google, Facebook, Twitter and even YouTube. In fact, MySQL is the world's most popular database and remains so because of its open source nature.

while

Technology changes rapidly and now the new buzzword in the database world is NoSQL. The market is a formidable one with projected growth forecast to reach \$3.4 Billion in 2020, representing a compound annual growth rate (CAGR) of 21% for the period 2015 – 2020.

| SQL | NO SQL |
|--|---|
| Since 1970 sql databases have been integral part of it infrastructure. | New technology |
| MySQL is world's most popular database and open source nature. | New technology that cover 21% market for period of 2015-2020 |
| It involve Structure Query Language. | it doesn't involve the Structured Query Language. |
| Oracle, MySQL, SQL Server are the Example of SQL. | MongoDB, CouchDB, BigTable are the example of NoSQL. |
| Relational Databases (SQL) | Non-relational Databases (NoSQL) |
| Well established set of standards. | NoSQL is still a relatively young technology without the set of standards |
| Some SQL are costly. | Lower cost, easier scalability, and open source features make NoSQL an appealing option for many companies looking to integrate in Big Data |
| A relational database structures data into tables and rows | MongoDB structures data into collections of JSON documents. JSON is a self-describing, human readable data format. Originally designed for lightweight exchanges between browser and server |
| | |

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot – 360001
Ph. No - (0281)2471645

2 – Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

NoSQL

The points below highlight some of the biggest advantages and disadvantages of NoSQL.

Advantages

- Non-Relational means table-less: NoSQL databases are non-relational, hence, very different from SQL databases. This means they are easier to manage and they provide a higher level of flexibility with newer data models.
- Mostly Open Source and Low-Cost: The open source nature of NoSQL databases makes them an appealing solution for smaller organizations with limited budgets. The top NoSQL databases on the market today (MongoDB, MarkLogic, Couchbase, CloudDB, and Amazon’s Dynamo DB) allow for rapid processing of real-time Big Data applications in ways that are affordable.
- Easier scalability through support for Map Reduce: NoSQL database experts often use elastic scalability as a major selling point of NoSQL. NoSQL databases are designed to function on full throttle even with low-cost hardware.
- No need to develop a detailed database model: The non-relational nature of a NoSQL database allows database architects to quickly create a database without needing to develop a detailed (fine-grained) database model. This saves a lot of development time.

Disadvantages

- Community not as well defined: While its continuing to grow rapidly, the NoSQL community is relatively new and lacks the maturity of the MySQL user base. Obviously, NoSQL is growing rapidly, but for now MySQL is hard to beat for its vast network of highly experienced end users.
- Lack of reporting tools: A major problem with NoSQL databases is the lack of reporting tools for analysis and performance testing. However, with MySQL, you can find a wide array of reporting tools to help you prove your application’s validity.
- Lack of standardization: In order for NoSQL to grow, it needs a standard query language like SQL. This is a major issue highlighted by researchers at Microsoft, who claim that NoSQL’s lack of standardization can cause a problem during migration. Besides this, standardization is important for the database industry to unify itself in the future.

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot – 360001
Ph. No - (0281)2471645

2 – Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

(2) History of NoSQL

The term NoSQL was used by Carlo Strozzi in 1998 to name his lightweight, Strozzi NoSQL open-source relational database that did not expose the standard SQL interface, but was still relational.

His NoSQL RDBMS is distinct from the circa-2009 general concept of NoSQL databases. Strozzi suggests that, as the current NoSQL movement "departs from the relational model altogether; it should therefore have been called more appropriately 'NoREL'", referring to 'No Relational'.

Johan Oskarsson reintroduced the term NoSQL in early 2009 when he organized an event to discuss "open source distributed, non relational databases".

The name attempted to label the emergence of an increasing number of non-relational, distributed data stores, including open source clones of Google's BigTable/MapReduce and Amazon's Dynamo.

Most of the early NoSQL systems did not attempt to provide (ACID)atomicity, consistency, isolation and durability guarantees, contrary to the prevailing practice among relational database systems.

Based on 2014 revenue, the NoSQL market leaders are MarkLogic, MongoDB, and Datastax.

Based on 2015 popularity rankings, the most popular NoSQL databases are MongoDB,

Apache Cassandra, and Redis.

CHAPTER - 2

(1) `_id`

BSON short for Binary JSON, is a binary-encoded serialization of JSON-like documents.

The `_id` field is primary key for every document. It's called `_id` and is also accessible via `id`

That section is just indicating that the automatically generated ObjectIDs start with a timestamp so it's possible to sort your documents automatically. This is pretty cool since the `_id` is automatically indexed in every collection.

A BSON ObjectID is a 12-byte value consisting of a 4-byte timestamp (seconds since epoch), a 3-byte machine id, a 2-byte process id, and a 3-byte counter. Note that the timestamp and counter fields must be stored big endian unlike the rest of BSON.

10 down vote The `_id` field is the default field for Bson ObjectId's and it is, by default, indexed.

`_id` and `id` are not the same. You may also choose to add a field called `id`

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot – 360001
Ph. No - (0281)2471645

2 – Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

An ObjectId is a 12-byte BSON type having the following structure –

- The first 4 bytes representing the seconds since the unix epoch
- The next 3 bytes are the machine identifier
- The next 2 bytes consists of process id
- The last 3 bytes are a random counter value

MongoDB uses ObjectIds as the default value of `_id` field of each document, which is generated while the creation of any document. The complex combination of ObjectId makes all the `_id` fields unique.

Creating New ObjectId

To generate a new ObjectId use the following code –

```
>newObjectId = ObjectId()
```

The above statement returned the following uniquely generated id –

```
ObjectId("5349b4ddd2781d08c09890f3")
```

(2) runCommand()

Definition

```
db.runCommand(command)
```

Provides a helper to run specified database commands. This is the preferred method to issue database commands, as it provides a consistent interface between the shell and drivers.

| Parameter | Type | Description |
|-----------|-----------------------|---|
| command | document
or string | “A database command, specified either in document form or as a string. If specified as a string, <code>db.runCommand()</code> transforms the string into a document.” |

New in version 2.6: To specify a time limit in milliseconds, see `Terminate Running Operations`.

Behavior

`db.runCommand()` runs the command in the context of the current database. Some commands are only applicable in the context of the admin database, and you must change your db object to before running these commands.

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot – 360001
Ph. No - (0281)2471645

2 – Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

(3) Relational Model

Fundamental Differences The immediate and fundamental difference between MongoDB and an RDBMS is the underlying data model. A relational database structures data into tables and rows, while MongoDB structures data into collections of JSON documents. ... JSON is also a natural data format for use in the application layer.

Mapping Relational Databases and SQL to MongoDB

MongoDB is an open source document-oriented NoSQL database which stores data in the form of JSON-like objects. It has emerged as one of the leading databases due to its dynamic schema, high scalability, optimal query performance, faster indexing and an active user community.

Mapping Tables, Rows and Columns

Each database in MongoDB consists of collections which are equivalent to an RDBMS database consisting of SQL tables. Each collection stores data in the form of documents which is equivalent to tables storing data in rows. While a row stores data in its set of columns, a document has a JSON-like structure (known as BSON in MongoDB)

```
{
  "_id": ObjectId("5146bb52d8524270060001f3"),
  "age": 25,
  "city": "Los Angeles",
  "email": "mark@abc.com",
  "user_name": "Mark Hanks"
}
```

This document is equivalent to a single row in RDBMS.

Dynamic Schema

Different documents within a collection can have different schemas. The fields can be easily added, removed and modified anytime.

Mapping Joins and Relationships

Relationships in RDBMS are achieved using primary and foreign key relationships and querying those using joins. There is no such straightforward mapping in MongoDB but the relationships here are designed using embedded and linking documents.

Linking Documents

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot – 360001
Ph. No - (0281)2471645

2 – Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

MongoDB, the relations and their corresponding operations have to be taken care manually (for example, through code) as no foreign key constraints and rules apply.

Embedding Documents

If the data to be embedded is expected to grow larger in size, it is better to use Linking approach rather than Embedded approach to avoid the document becoming too large. Embedded approach is generally used in cases where a limited amount of information (like address in our example) has to be embedded.

Mapping Chart

RDBMS

| | |
|---------------|---------------------------------|
| RDBMS | MongoDB |
| Table | Collection |
| Row | Document |
| Column | Field |
| Relationships | Linking and Embedding Documents |

For MongoDB queries, let us assume a collection users with document structure as follows:

```
1{
2"_id": ObjectId("5146bb52d8524270060001f3"),
3"post_text":"This is a sample post" ,
4"user_name": "mark",
5"post_privacy": "public",
6"post_likes_count": 0
7}
Create
```

In MongoDB, there is no need to explicitly create the collection structure.

```
SQL: CREATE TABLE `posts` (`id` int(11) NOT NULL AUTO_INCREMENT, `post_text` varchar(500)
NOT NULL, `user_name` varchar(20) NOT NULL, `post_privacy` varchar(10) NOT
NULL, `post_likes_count` int(11) NOT NULL, PRIMARY KEY (`id`))
```

```
MongoDB: db.createCollection("posts")
```

Insert

To insert a document in MongoDB, we use the insert method which takes an object with key value pairs as its input

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot - 360001
Ph. No - (0281)2471645

2 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

SQL: INSERT INTO `posts` (`id`, `post_text`, `user_name`, `post_privacy`, `post_likes_count`)VALUES (NULL, 'This is a sample post', 'mark', 'public', '0');

MongoDB: db.posts.insert({user_name:"mark", post_text:"This is a sample post", post_privacy:"public", post_likes_count:0})

Read

MongoDB uses the find method which is equivalent to the SELECT command in SQL.

SQL: SELECT * FROM `posts`

MongoDB: db.posts.find()

Update

The first parameter to the update method specifies the criteria to select the documents. The second parameter specifies the actual update operation to be performed.

SQL: UPDATE posts SET post_privacy = "private" WHERE user_name='mark'

MongoDB: db.posts.update({user_name:"mark"},{\$set:{post_privacy:"private"}},{multi:true})

Remove

Removing documents is quite simple and similar to SQL.

SQL: DELETE FROM posts WHERE user_name='mark'

MongoDB: db.posts.remove({user_name:"mark"})



SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)

3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot – 360001
Ph. No - (0281)2471645

2 – Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

CHAPTER - 4

(1) Web Analytics using MongoDB

Real-Time Analytics

Most companies use analytics. Many can act on data from months, weeks, or even days ago. But few can respond to changes minute by minute, or second by second.

With MongoDB, analyze any data in place and in real time. Faster. With less money.

MongoDB is typically used for real-time analytics. Example applications include:

Financial Services, Government, High Tech, Retail

Analyze ticks, tweets, satellite imagery, weather trends, and any other type of data to inform trading algorithms in real time.

Identify social program fraud within seconds based on program history, citizen profile, and geospatial data.

Identify unique individuals across any type of device, browser or app and use a holistic behavioral model to advertise to them.

Set up a digital geo-fence around your brick-and-mortar locations to push in-store incentives to shoppers in real time.

Real-Time Analytics is Hard

Can't Stay Ahead. You need to account for many types of data, including unstructured and semi-structured data. And new sources present themselves unpredictably. Relational databases aren't capable of handling this, which leaves you hamstrung.

Can't Scale. You need to analyze terabytes or petabytes of data. You need sub-second response times. That's a lot more than a single server can handle. Relational databases weren't designed for this.

Batch. Batch processes are the right approach for some jobs. But in many cases, you need to analyze rapidly changing, multi-structured data in real time. You don't have the luxury of lengthy ETL processes to cleanse data for later.

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot – 360001
Ph. No - (0281)2471645

2 – Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

MongoDB Makes It Easy

Do the Impossible. MongoDB can incorporate any kind of data – any structure, any format, any source – no matter how often it changes. Your analytical engines can be comprehensive and real-time.

Scale Big. MongoDB is built to scale out on commodity hardware, in your data center or in the cloud. And without complex hardware or extra software. This shouldn't be hard, and with MongoDB, it isn't.

Real Time. MongoDB can analyze data of any structure directly within the database, giving you results in real time, and without expensive data warehouse loads.

Why Other Databases Fall Short

Most databases make you chose between a flexible data model, low latency at scale, and powerful access. But increasingly you need all three at the same time.

- Rigid Schemas. You should be able to analyze unstructured, semi-structured, and polymorphic data. And it should be easy to add new data. But this data doesn't belong in relational rows and columns. Plus, relational schemas are hard to change incrementally, especially without impacting performance or taking the database offline.
- Scaling Problems. Relational databases were designed for single-server configurations, not for horizontal scale-out. They were meant to serve 100s of ops per second, not 100,000s of ops per second. Even with a lot of engineering hours, custom sharding layers, and caches, scaling an RDBMS is hard at best and impossible at worst.
- Takes Too Long. Analyzing data in real time requires a break from the familiar ETL and data warehouse approach. You don't have time for lengthy load schedules, or to build new query models. You need to run aggregation queries against variably structured data. And you should be able to do so in place, in real time.

How MongoDB Makes it Easy

Organizations are using MongoDB for analytics because it lets them store any kind of data, analyze it in real time, and change the schema as they go.

- New Data. MongoDB's document model enables you to store and process data of any structure: events, time series data, geospatial coordinates, text and binary data, and anything else. You can adapt the structure of a document's schema just by adding new fields, making it simple to bring in new data as it becomes available.
- Horizontal Scalability. MongoDB's automatic sharding distributes data across fleets of commodity servers, with complete application transparency. With multiple options for scaling – including range-based, hash-based and location-aware sharding – MongoDB can support thousands of nodes, petabytes of data, and hundreds of thousands of ops per second without requiring you to build custom partitioning and caching layers.
- Powerful Analytics, In Place, In Real Time. With rich index and query support – including secondary, geospatial and text search indexes – as well as the aggregation framework and native MapReduce, MongoDB can run complex ad-hoc analytics and reporting in place.

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot – 360001
Ph. No - (0281)2471645

2 – Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

Extracting analytics data with MapReduce

The log contains raw data about page visits, but we need to extract some meaningful information out of it. For example, it might be useful to know how many times a page has been viewed over a certain time period, or what is the average response time for a page. It is possible to do so by applying MapReduce on the log. In the next example, we are going to do just that.

The map function and reduce functions are used for analysis of data where most of the unnecessary data will be removed. It also improves response time.

It is not generally a good idea to calculate analytics using such MapReduce in real time, especially if you are running a website that enjoys heavy user traffic. The log will be very large and constantly growing, so running MapReduce on it would take time (MapReduce jobs are known to be consistent and continuous, but their speed depends on several factors).

If we run the page view calculation program it will take a long time to load the page. Rather you should run processes in the background like cron job that execute the MapReduce jobs, stores the results in a collection, and have the analytics page simply read from that collection. This way we can keep the performance at an optimum level.

Logging with MongoDB

The log contains raw data about page visits, but we need to extract some meaningful information out of it.

For example, it might be useful to know how many times a page has been viewed over a certain time period, or what is the average response time for a page. It is possible to do so by applying MapReduce on the log.

The map function and reduce functions are used for analysis of data where most of the unnecessary data will be removed. It also improves response time. It is not generally a good idea to calculate analytics using such MapReduce in real time, especially if you are running a website that enjoys heavy user traffic.

The log will be very large and constantly growing, so running MapReduce on it would take time (MapReduce jobs are known to be consistent and continuous, but their speed depends on several factors).

By default, mongod sends its logs to stdout. Most init scripts use the --logpath option to send logs to a file.

If you are debugging a specific issue with your application, there are a couple of options for getting more info from the logs.

If you have multiple MongoDB instances on a single machine (say, a mongod and a mongos), make sure that their logs are stored in separate files. Make sure that you know where the logs are and have read access to the files.

MongoDB spits out a lot of log messages, but please do not run with the --quiet option.

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot – 360001
Ph. No - (0281)2471645

2 – Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

there is enough info for basic debugging (why is this slow, why isn't this starting up, etc.), but **the log does not take up too much space.**

Leaving the **log level at the default** is usually perfect: but, **you can change** the log level, either by restarting MongoDB with more v's or running the **setParameter command**:

```
> db.adminCommand({"setParameter" : 1, "logLevel" : 3})
```

Remember to turn log level back down to 0, or your logs may be needlessly noisy. You can turn log level up to 5, at which point mongod will print out almost every action it takes, including the contents of every request handled.

This can cause a lot of IO as mongod writes everything to the log file, which can slow down a busy system. **Turning on profiling is a better option if you need to see every operation as it's happening.**

By default, MongoDB logs information about queries that take longer than 100 milliseconds to run. If 100 ms is too short or too long for your application, you can change timing with `setProfilingLevel`:

Syntax:- `db.setProfilingLevel(<level>, <options>)`

`db.setProfilingLevel(1, { slowms: 20, sampleRate: 0.42 })`

You can also set this parameter by restarting MongoDB with the `--slowms` option.

| Level | Description |
|-------|--|
| 0 | The profiler is off and does not collect any data. This is the default profiler level. |
| 1 | The profiler collects data for operations that take longer than the value of <code>slowms</code> . |
| 2 | The profiler collects data for all operations. |

Extracting analytics data with MapReduce

What is MapReduce?

MapReduce is a software framework and programming model used for processing huge amounts of data. **MapReduce** program work in two phases, namely, Map and Reduce. Map tasks deal with splitting and mapping of data while Reduce tasks shuffle and reduce the data.

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



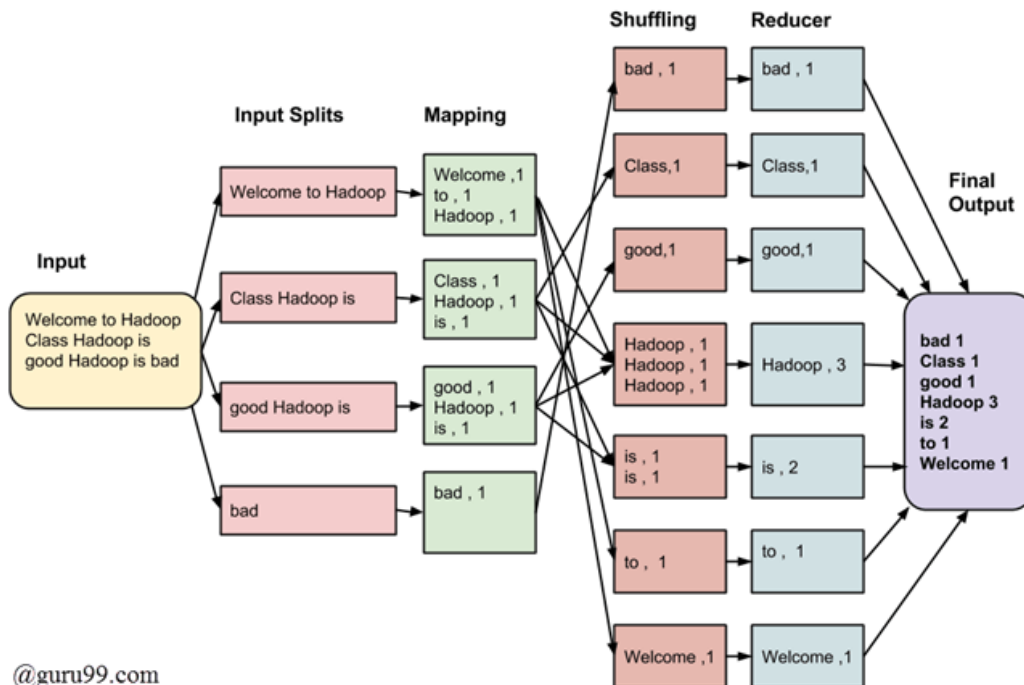
3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot – 360001
Ph. No - (0281)2471645

2 – Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

Hadoop is capable of running MapReduce programs written in various languages: Java, Ruby, Python, and C++. MapReduce programs are parallel in nature, thus are very useful for performing large-scale data analysis using multiple machines in the cluster.

The input to each phase is **key-value** pairs. In addition, every programmer needs to specify two functions: **map function** and **reduce function**.

The whole process goes through four phases of execution namely, splitting, mapping, shuffling, and reducing.



@guru99.com

MongoDB NoSQL database is used in the big data stack for storing and retrieving one item at a time from large datasets whereas Hadoop is used for processing these large data sets.

Hadoop is the way to go for organizations that do not want to add load to their primary storage system and want to write distributed jobs that perform well.

Traditional relational databases can work until datasets were being increased in megabytes and gigabytes, called “Big Data” then old technologies getting unfeasible. When it came to data storage and retrieval, these technologies simply down under the burden of such huge amounts of data.

Hadoop, Hive and Hbase, these popular technologies now have the capability of handling large sets of raw unstructured data, efficiently, as well as economically.

“Not Only SQL” or NoSQL databases, provide the facilitates the storage and retrieval of data in the loser consistency model along with added benefits like horizontal scaling, better availability and quicker access.

With its implementation in over five hundred top notch organizations across the globe, MongoDB certainly has emerged as the most popular NoSql databases amongst all.

However, there are various metrics like Google searches and the number of employment opportunities for Hadoop and MongoDB professionals that give a good idea of the popularity of these technologies. later MongoDB have

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot – 360001
Ph. No - (0281)2471645

2 – Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

successfully developed a MongoDB connector for Hadoop that facilitates enhanced integration combined with ease in execution of various tasks.

Integration of MongoDB with Hadoop for, offline analytics.

- The MongoDB-Hadoop connector uses the authority of Hadoop’s MapReduce to live application data in MongoDB by extracting values from Big Data – speedily as well as efficiently.
- The MongoDB-Hadoop connector projects it as ‘Hadoop compatible file system’ and MapReduce jobs can now be read directly from MongoDB, without being copied to the HDFS.
- The “necessity” of scanning entire collections has been eliminated as MapReduce jobs can pass queries by means of filters and can bind MongoDB’s indexing abilities like text search, compound, array, Geo-spatial and sparse indexes.
- Reading and writing back results from Hadoop jobs back to MongoDB in order to support queries and real time operational processes.

Scope of application - Hadoop and MongoDB

1) MongoDB is used for the operational part – as a real time data store.

2) Hadoop is used primarily for offline analysis and processing of batch data.

When it comes to analyzing data, the inbuilt aggregation features incorporated in MongoDB hold good in the majority of situations. However, there are cases that require a higher degree of data aggregation. Under such circumstances, Hadoop provides a powerful support for complex analytics.

- a) Hadoop, by means of single or multiple MapReduce jobs processes the data extracted from MongoDB. It is also possible to pull data from other locations in these MapReduce jobs in order to formulate a multi data solution.
- b) The results received from MapReduce jobs can be written back to MongoDB and they can be used for analysis and querying as and when required.
- c) MongoDB applications can thus make use of the data from batch analytics with a view of handing over to the end user or to facilitate other features down the line.

In a usual production environment, application data with their specific functionality and language may exist in more than one data store. Under such complex situations, Hadoop is used as an integrated source for data - as well as a data warehouse.

- a) MapReduce jobs transfer MongoDB data to Hadoop.
- b) As soon as the data from MongoDB and other sources is available in Hadoop, the datasets can be queried.
- c) At this stage data analysts can use Pig or MapReduce for querying large datasets that includes data from MongoDB.

Now a days MongoDB has emerged as the most preferred choice of developers. From the perspective of NoSQL databases, engineers at MongoDB have successfully integrated it with Hadoop. The MongoDB Hadoop variation is extremely effective in solving quite a few architectural problems pertaining to data warehousing, processing, data retrieval and aggregating.

MongoDB Real Time Analytics

MongoDB wasn’t originally developed for analytics however with data growing at an exponential rate and the need for having real-time capabilities, in terms of monitoring updates or availability of data, becoming more and more fundamental, it needed these features to survive.

MongoDB has grown immensely with time to counter such requirements and hence now it supports a lot of analytics capabilities directly built-in the database.

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot – 360001
Ph. No - (0281)2471645

2 – Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

There are mainly two methods to perform analytics using MongoDB:

Method 1: Replicating A MongoDB Database Into A SQL Database

Method 2: Data Virtualization

Method 1: Replicating A MongoDB Database Into A SQL Database

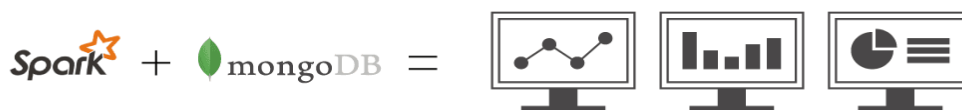
Replicating data into a SQL database allows users to keep on using MongoDB as their production database and use the relational format to analyze data with ease. SQL can now be used on this relational version of MongoDB data. This allows users to access and manipulate data with easiness and combine data from multiple tables using indexes to perform understanding analysis.

SQL brings in a lot of conveniences when working with lengthy aggregations and complex data joins. However, data replication is not as easy as it sounds. This requires an ETL job which might be complicated as it requires transferring data from a NoSQL environment to a SQL environment. These ETL jobs also need external hardware and support of data engineers & analysts to work properly.

Method 2: Data Virtualization

Data virtualization is a method that can be used for MongoDB real time analytics. This method is the ideal solution to counter the limitations of replicating databases.

Various tools provide an interactive & user-friendly interface. These tools can be connected with MongoDB with ease and allow the users to query or manipulate their data stored in MongoDB. Users can now develop visualizations and perform real-time analysis in just a few clicks making use of smart & easy to use dashboards and customer-facing reports. The advantage here is that it doesn't require any additional hardware or tedious ETL jobs to analyze data.



One such tool is [Apache Spark](#). MongoDB supports this popular framework that is loved by data scientists, engineers, & analysts. MongoDB provides powerful large-scale analytics features. These allow users to perform analysis within the platform by converting data into visualizations along with a parallel query execution engine to boost the performance.

MongoDB also supports a SQL-based BI connector that allows users to explore their MongoDB data using different business intelligence tools such as Looker, Microsoft Power BI, and others.

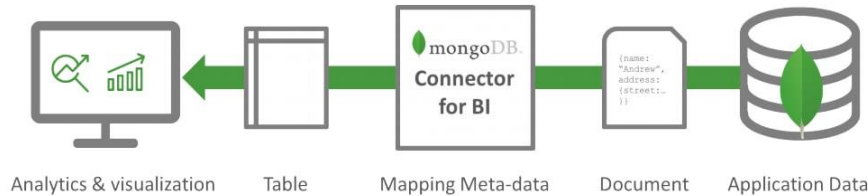
SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot – 360001
Ph. No - (0281)2471645

2 – Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478



Advantages Of MongoDB Real Time Analytics

- **Ad-hoc Queries:** MongoDB supports ad-hoc querying. It is very flexible and supports all different kinds of data.
- **Powerful Analytics:** MongoDB supports real-time analytics with a wide variety of data. It allows performing analysis on primary data, secondary data, and even on text searches. It has strong integrations with aggregation frameworks & the MapReduce paradigm.
- **Speed:** MongoDB being a document-oriented database, allows you to query data quickly. Its rich indexing capabilities allow it to perform way faster than a relational database.
- **Easy Setup:** MongoDB can be set up easily on any system.
- **Data Adaptability:** A NoSQL system like MongoDB supports a wide variety of data such as text data, geospatial data, etc. It provides an ultra-flexible data model making it easier to incorporate data and making adjustments for better performance.
- **Scalability:** NoSQL databases are built to scale. MongoDB's sharding capability allows it to distribute data across datasets, servers, etc. This gives it an unlimited growth capability and a higher production rate than a relational database.
- **Real-Time:** With MongoDB, you can analyze data of any structure within the database and get real-time results without costly data warehouse loads.

Disadvantages Of MongoDB Real Time Analytics

- **Memory Constraints:** MongoDB leads to unnecessary usage of memory. It stores every key-value pair and hence suffers from duplication of values.
- **No Support For Joins:** MongoDB doesn't support joins. Joins are implemented using programming languages such as Java, however, this makes the querying complex & hampers the performance.
- **No Referential Integrity (RI):** These are the defined and validated relations between different pieces of data. RI helps to keep the information consistent and adds another layer of validation underneath the programmatic one.

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot - 360001
Ph. No - (0281)2471645

2 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478



MBZ-003-1081003

Seat No. _____

M. Sc. (I.T.) (Sem. I) (CBCS) Examination

December - 2016

CS-03 : NoSQL Database : MongoDB

Faculty Code : 003

Subject Code : 1081003

Time : 2½ Hours]

[Total Marks : 70

- 1 (a) Objective type questions : 4
- ~~(1)~~ What is nosql?
 - (2) Most NoSQL databases support automatic _____ meaning that you get high availability and disaster recovery.
 - ~~(3)~~ Full form of nosql
 - ~~(4)~~ NoSQL databases is used mainly for handling large volumes of _____ data.
- (b) Answer in brief : (any one) 2
- (1) Difference between sql and nosql.
 - ~~(2)~~ How to install php driver on windows platform?
- (c) Answer in detail : (any one) 3
- ~~(1)~~ Explain characteristics of nosql.
 - (2) Explain history of nosql.
- (d) Write a note on : (any one) 5
- ~~(1)~~ What are the advantage and disadvantage of nosql? Explain in detail.
 - (2) Explain database, collection and document.

MBZ-003-1081003]

1

[Contd...

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot - 360001
Ph. No - (0281)2471645

2 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

- 2 (a) Objective type questions : 4
- (1) Which type of DBMS is MongoDB?
 - (2) What is the difference between MongoDB and MySQL?
 - (3) MongoDB database uses _____ for storing records.
 - (4) What is the default port number of mongodb?
- (b) Answer in brief : (any one) 2
- (1) Explain the use of `_id` key in mongodb.
 - (2) What is the use of `runCommand()`?
- (c) Answer in detail : (any one) 3
- (1) Explain following :
 - (i) `selectCollection()`
 - (ii) `selectDB()`
 - (iii) `getNext()`
 - (2) Explain `sort`, `limit` and `skip` function with example.
- (d) Write a note on : (any one) 5
- (1) Explain `insert()` and `remove()` with example.
 - (2) Explain `find` method with all operators.
- 3 (a) Objective type questions : 4
- (1) The session ID is a _____ bit character string.
 - (2) Which method is used to start a session?
 - (3) By default PHP uses _____ based session handling.
 - (4) `Session_start()` must send session data to the web browser via _____.
- (b) Answer in brief : (any one) 2
- (1) Limitations of native PHP session handling.
 - (2) Explain the session garbage collector.

MBZ-003-1081003]

2

[Contd...

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot - 360001
Ph. No - (0281)2471645

2 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

- (c) Answer in detail : (any **one**) 3
- (1) Explain Problems of using MongoDB and RDBMS together.
 - (2) What is HTTP Session?
- (d) Write a note on : (any **one**) 5
- (1) Write a code to use session in login and logout page.
 - (2) Explain session_set_save_handler in detail.
- 4 (a) Objective type questions : 4
- (1) MongoDB uses _____, a functional programming paradigm to perform aggregation and batch processing of data.
 - (2) Sorting by natural order means?
 - (3) reduce() takes the result array of _____ as its input and combines all the elements in that array into a single element by performing some operation.
 - (4) The concept MapReduce has been derived from which programming paradigm?
- (b) Answer in brief : (any **one**) 2
- (1) Explain count(), distinct() with example
 - (2) Explain emit().
- (c) Answer in detail : (any **one**) 3
- (1) How we can apply mapreduce in mongodb? Explain with example.
 - (2) Explain difference between mapreduce and group().
- (d) Write a note on : (any **one**) 5
- (1) What is MapReduce? Explain in detail.
 - (2) Explain group() function.

MBZ-003-1081003]

3

[Contd...

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot - 360001
Ph. No - (0281)2471645

2 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

- 5 (a) Objective type questions : 4
- (1) What is the size of chunk file?
 - (2) What is the maximum size of mongodb document?
 - (3) Which command is used to backing up a stud database?
 - (4) The mongoimport utility can load data in which three file formats?
- (b) Answer in brief : (any one) 2
- (1) What is chunk file?
 - (2) Explain getGridFS().
- (c) Answer in detail : (any one) 3
- (1) Write a short note on replication.
 - (2) What is sharding? Explain horizontal partitioning.
- (d) Write a note on : (any one) 5
- (1) What is GridFS? Explain advantages of GridFS.
 - (2) Explain mongorestore with all options.

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot - 360001
Ph. No - (0281)2471645

2 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478



MG-003-1081003

Seat No. _____

M. Sc. (I.T.) (Sem. I) (CBCS) Examination

January - 2018

CS - 03 : NoSQL Database : MongoDB

Faculty Code : 003

Subject Code : 1081003

Time : $2\frac{1}{2}$ Hours]

[Total Marks : 70

Instruction : Write answers of all the questions in main answer-sheet.

- 1 (A) Attempt the following objective questions : 4
- (1) What is collection?
 - (2) Give the full form of NoSQL.
 - (3) MongoDB is _____ type of NoSQL database.
 - (4) What is database-as-a-service?
- (B) Attempt any **one** of the following : 2
- (1) List out major types of NoSQL database.
 - (2) Explain history of NoSQL.
- (C) Attempt any **one** of the following : 3
- (1) List out the steps for installing php driver for mongodb on windows platform.
 - (2) Define objective of NoSQL.
- (D) Attempt any **one** of the following : 5
- (1) Explain concept of data modeling and working with data.
 - (2) Explain collection and document in detail.
- 2 (A) Attempt the following objective questions : 4
- (1) Which command displays the list of all databases?
 - (2) _____ stores the number of documents in MongoDB.
 - (3) db.collection.remove({}) is used for _____.
 - (4) What is the default port number of MongoDB? ϕ

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot - 360001
Ph. No - (0281)2471645

2 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

- (B) Attempt any **one** of the following : 2
(1) Explain limit and skip function.
(2) Explain create Collection() method.
- (C) Attempt any **one** of the following : 3
(1) Explain sort() method with example.
(2) Compare Relational database with NoSQL database.
- (D) Attempt any **one** of the following : 5
(1) Explain insert() and update() with example.
(2) Explain find() and pretty() methods with example and output.
- 3 (A) Attempt the following objective questions : 4
(1) _____ method is used to start a session.
(2) What is session?
(3) Use of \$_SESSION[].
(4) Is HTTP stateless protocol?
- (B) Attempt any **one** of the following : 2
(1) What is HTTP session?
(2) Explain concept of session management.
- (C) Attempt any **one** of the following : 3
(1) Explain session manager class.
(2) Briefly explain session_set_save handler() function.
- (D) Attempt any **one** of the following : 5
(1) Explain php native session handling with example.
(2) Write a php program to create login application with session using MongoDB.
- 4 (A) Attempt the following objective questions : 4
(1) What is reduce phase?
(2) The _____ phase breakdown the task into smaller subtasks and execute them to produce intermediate results.
(3) group() method can not be applied on a key having more than _____ distinct values.
(4) What do you mean by web analytics?
- (B) Attempt any **one** of the following : 2
(1) What is real time analytics? Explain it.
(2) What is MapReduce? Explain it.

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot - 360001
Ph. No - (0281)2471645

2 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

- (C) Attempt any **one** of the following : 3
- (1) Explain concept of extracting analytics data with mapreduce.
- (2) Differentiate MapReduce and group().
- (D) Attempt any **one** of the following : 5
- (1) Explain db.collection.distinct() and db.collection.count() with example.
- (2) Write a php code for count number of employee per department with using mapreduce in mongodb.
- 5 (A) Attempt the following objective questions : 4
- (1) In replica set, minimum _____ nodes are required.
- (2) _____ refers to the process of splitting data up and storing different portions of the data on different machines.
- (3) By default GridFS limits chunk size to _____ kb.
- (4) What is Grid?
- (B) Attempt any **one** of the following : 2
- (1) Explain files and chunks collection.
- (2) What is Replication? Explain it.
- (C) Attempt any **one** of the following : 3
- (1) Define explain() method.
- (2) Write a code for reading files from GridFS.
- (D) Attempt any **one** of the following : 5
- (1) Write a program for storing files in GridFS
- (2) Explain concept of sharding in detail.
-

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot - 360001
Ph. No - (0281)2471645

2 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478



PBV-003-1081003 Seat No. _____

M. Sc. (I.T.) (Sem. I) (CBCS) Examination

November / December - 2018

CS - 03 : NoSQL Database : MongoDB

Faculty Code : 003

Subject Code : 1081003

Time : $2\frac{1}{2}$ Hours]

[Total Marks : 70

Instruction : Write answers of all the questions in main answer sheet.

- 1 (a) Attempt the following objective questions. 4
- (1) What is document ?
 - (2) What is the default port address of the MongoDB ?
 - (3) Define NoSQL.
 - (4) What is scalability ?
- (b) Attempt any one of the following : 2
- (1) Define Database and Collection in MongoDB.
 - (2) Write a note on advantages of NoSQL.
- (c) Attempt any one of the following : 3
- (1) Write a note on data model and working with data.
 - (2) What is database-as-a-service ? Explain it.
- (d) Attempt any one of the following : 5
- (1) List out types of NoSQL databases and explain it.
 - (2) Explain characteristics of NoSQL.
- 2 (a) Attempt the following objective questions. 4
- (1) How does a new database get created in MongoDB ?
 - (2) What is relational model ?
 - (3) In MongoDB, the _____ method is used to add or insert new documents into a collection in your database.
 - (4) To check the created collections, _____ command is used.

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot – 360001
Ph. No - (0281)2471645

2 – Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

- (b) Attempt any one of the following : 2
- (1) Explain findOne() method.
 - (2) Explain sort() method.
- (c) Attempt any one of the following : 3
- (1) Explain insert() method with example.
 - (2) Write a note on managing relationships between documents.
- (d) Attempt any one of the following : 5
- (1) Explain update() with example.
 - (2) Explain find() method in detail.
- 3 (a) Attempt the following objective questions. 4
- (1) What is HTTP session ?
 - (2) Is HTTP stateless protocol ?
 - (3) What is session_id() ?
 - (4) Define session_destroy().
- (b) Attempt any one of the following : 2
- (1) Write a note on limitations of php native session handling.
 - (2) Advantages of HTTP session.
- (c) Attempt any one of the following : 3
- (1) Write a code for php native session handling.
 - (2) What is session manager ? Explain it.
- (d) Attempt any one of the following : 5
- (1) Explain session_set_save_handler() with its arguments.
 - (2) Write a code to create login, logout and user profile page using session.
- 4 (a) Attempt the following objective questions. 4
- (1) What is real time analytics ?
 - (2) What is map phase ?
 - (3) The concept of MapReduce has been derived from _____ programming paradigm.
 - (4) The result of group() method has to be less than _____ in size.

SHREE H. N. SHUKLA GROUP OF COLLEGES

(AFFILIATED TO SAURASHTRA UNIVERSITY)



3 - Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya road,
Rajkot – 360001
Ph. No - (0281)2471645

2 – Vaishalinagar
Nr. Amrapali Railway Crossing,
Raiya Road,
Rajkot - 360001
Ph.No - (0281)2440478

- (b) Attempt any one of the following : 2
- (1) Explain runCommand().
 - (2) Explain emit() method.
- (c) Attempt any one of the following : 3
- (1) Write a note on difference between MapReduce and group().
 - (2) Write a code for listing distinct values for specific field.
- (d) Attempt any one of the following : 5
- (1) Explain concept of MapReduce in detail.
 - (2) Write a code to find department wise maximum salary using MongoDB.
- 5 (a) Attempt the following objective questions. 4
- (1) In replication, _____ operations are performed on slave.
 - (2) The _____ collection of GridFS will contain the file name, type, size, upload time etc. information.
 - (3) What is optimization ?
 - (4) What is GridFS ?
- (b) Attempt any one of the following : 2
- (1) Explain files and chunks collection.
 - (2) Explain getGridFS() method.
- (c) Attempt any one of the following : 3
- (1) Write a code for uploading file in GridFS.
 - (2) Explain the concept of replication.
- (d) Attempt any one of the following : 5
- (1) What is sharding ? Explain it.
 - (2) Write a code reading files in chunks from GridFS.