



WINTER – 2022 EXAMINATION
Model Answer

Subject Name: Advanced Database Management System

Subject Code: 22521

Important Instructions to examiners:

- 1) The answers should be examined by key words and not as word-to-word as given in the model answer scheme.
- 2) The model answer and the answer written by candidate may vary but the examiner may try to assess the understanding level of the candidate.
- 3) The language errors such as grammatical, spelling errors should not be given more Importance (Not applicable for subject English and Communication Skills).
- 4) While assessing figures, examiner may give credit for principal components indicated in the figure. The figures drawn by candidate and model answer may vary. The examiner may give credit for any equivalent figure drawn.
- 5) Credits may be given step wise for numerical problems. In some cases, the assumed constant values may vary and there may be some difference in the candidate's answers and model answer.
- 6) In case of some questions credit may be given by judgement on part of examiner of relevant answer based on candidate's understanding.
- 7) For programming language papers, credit may be given to any other program based on equivalent concept.
- 8) As per the policy decision of Maharashtra State Government, teaching in English/Marathi and Bilingual (English + Marathi) medium is introduced at first year of AICTE diploma Programme from academic year 2021-2022. Hence if the students in first year (first and second semesters) write answers in Marathi bilingual language (English + Marathi), the Examiner shall consider the same and assess the answer based on matching of concepts with model answer.

Q. No.	Sub Q. N.	Answer	Marking Scheme
1		Attempt any <u>FIVE</u> of the following:	10 M
	a)	State the use of concurrency control.	2 M
	Ans	It is a procedure in DBMS which helps us for the management of two simultaneous processes to execute without conflicts between each other, these conflicts occur in user systems. Concurrency can simply be said to be executing multiple transactions at a time required to increase time efficiency. If many transactions try to access the same data then inconsistency arises. Concurrency control required to maintain consistency • Waiting time will be decreased. • Response time will decrease. • Resource utilization will increase. • System performance & Efficiency is increased.	Any 2 uses- 2M



b)	Define Big Data.	2 M
Ans	Big Data contains a large amount of data that is not being processed by traditional storage or the processing unit. It is used by many multinational companies to process the data and business of many organizations. The data flow would exceed 150 exabytes per day before replication. There are five v's of Big Data that explains the characteristics. 5 V's of Big Data Volume (huge amount of data), Veracity (inconsistencies uncertainty of data), Variety (different formats of data from various sources), (extract useful data), Velocity (high speed of accumulation of data)	Correct definition-2M
c)	Give any four characteristics of XML.	2 M
Ans	XML separates data from HTML XML simplifies data sharing XML simplifies data transport XML simplifies Platform change XML increases data availability XML can be used to create new internet languages	four characteristic s of XML-2M
d)	Define Data Mart & Meta Data.	2 M
Ans	Data Mart: A data mart is a subset of a data warehouse focused on a particular part of business, department, or subject area. Data marts make specific data available to a defined group of users, which allows those users to quickly access critical information without wasting time searching through an entire data warehouse. Meta Data: Metadata in DBMS is characterized as data about data. It describes the context and information about data, the way that data is stored, various relations among data. Metadata in a relational DBMS stores data Constraints, Table Relationships, Data Types, Columns, Tables, and so on.	Data Mart- 1M & Meta Data- 1M
e)	Enlist any four types of join.	2 M
Ans	• Inner Join : Returns records that have matching values in both tables	Listing of 4 join types- 2M(1/2 M each)



```
anki — mongo — 80x55
> use GeeksforGeeks
switched to db GeeksforGeeks
> db.student.insertOne({
...   name : "Sumit",
...   age : 20,
...   branch : "CSE",
...   course : "C++ STL",
...   mode : "online",
...   paid : true,
...   amount : 1499
... })
{
  "acknowledged" : true,
  "insertedId" : ObjectId("5e548cdc92e6dfa3fc48ddae")
}
>
```

2. Read Operations –The Read operations are used to retrieve documents from collection, or in other words, read operations are used to query a collection document. You can perform read operation using the following method provided by the MongoDB:

Method	Description
db.collection.find()	It is used to retrieve documents from the collection.

Example: In this example, we are retrieving the details of students from the student collection using db.collection.find() method.

```
anki — mongo — 80x55
> use GeeksforGeeks
switched to db GeeksforGeeks
> db.student.find().pretty()
4
  {
    "_id" : ObjectId("5e548cdc92e6dfa3fc48ddae"),
    "name" : "Sumit",
    "age" : 20,
    "branch" : "CSE",
    "course" : "C++ STL",
    "mode" : "online",
    "paid" : true,
    "amount" : 1499
  }
3
4
  {
    "_id" : ObjectId("5e548d3192e6dfa3fc48ddaf"),
    "name" : "Sumit",
    "age" : 20,
    "branch" : "CSE",
    "course" : "C++ STL",
    "mode" : "online",
    "paid" : true,
    "amount" : 1499
  }
3
4
  {
    "_id" : ObjectId("5e548d3192e6dfa3fc48ddb8"),
    "name" : "Rohit",
    "age" : 21,
    "branch" : "CSE",
    "course" : "C++ STL",
    "mode" : "online",
    "paid" : true,
    "amount" : 1499
  }
3
>
```



3. Update Operations –The update operations are used to update or modify existing document in the collection. You can perform update operations using following methods provided by the MongoDB

Method	Description
db.collection.updateOne()	It is used to update a single document in the collection that satisfy the given criteria.
db.collection.updateMany()	It is used to update multiple documents in the collection that satisfy the given criteria.
db.collection.replaceOne()	It is used to replace single document in the collection that satisfy the given criteria.

Example: In this example, we are updating the year of course in all the documents student collection using db.collection.updateMany() method.

```
1. Open the MongoDB Shell (mongo) and connect to the database.
2. Use the use() command to select the database.
3. Use the updateMany() method to update the year of course in all the documents of the student collection.
4. Use the find() method to verify the update.

> use(student)
> updateMany({'year': 2019}, {'year': 2020})
{ "n": 3, "updatedExisting": true }
> find()
{ "_id": "1", "name": "Ravi", "year": 2020 }
{ "_id": "2", "name": "Smit", "year": 2020 }
{ "_id": "3", "name": "Aniket", "year": 2020 }
```

4. Delete Operations –The delete operation are used to delete or remove documents from a collection. You can perform delete operations using following methods provided by the MongoDB:

Method	Description
db.collection.deleteOne()	It is used to delete a single document from the collection that satisfy the given criteria.
db.collection.deleteMany()	It is used to delete multiple documents from the collection that satisfy the given criteria.

Example: In this example, we are deleting a document from the student collection using db.collection.deleteMany() method.



	<pre>anki — mongo — 80x56 > use GeeksforGeeks switched to db GeeksforGeeks > db.student.find().pretty() { "_id" : ObjectId("5e540d3192e6dfa3fc48ddaf"), "name" : "Sumit", "age" : 20, "branch" : 'CSE', "course" : 'C++ STL', "mode" : "online", "paid" : true, "amount" : 1499, "year" : 2020 } { "_id" : ObjectId("5e54103592e6dfa3fc48ddb1"), "name" : "Rohit", "age" : 21, "branch" : 'CSE', "course" : 'C++ STL', "mode" : "online", "paid" : true, "amount" : 1499 } > db.student.deleteMany({}) { "acknowledged" : true, "deletedCount" : 2 } > █</pre>	
d)	Explain features of R-programming along with its R-programming use in various applications.	4 M



Ans	1. Comprehensive Language-R is a comprehensive programming language meaning that it provides services for statistical modeling as well as for software development. 2. Provides a Wide Array of Packages-R is most widely used because of its availability of libraries. R has CRAN, which is a repository holding more than 10,000 packages. These packages appeal to every functionality and different fields that deal with data. Based on user requirements and preferences, packages provide different features to their users. 3. Possesses a Number of Graphical Libraries-The most important feature that sets it apart from other programming languages of Data Science is its rich collection of graphical libraries like ggplot2, plotly, etc. that are capable of making aesthetic and quality visualizations. 4. Open-source-R is an open-source programming language. This means that it is free of cost and requires no license. Furthermore, you can contribute towards the development of R, customize its packages and add more features. 5. Cross-Platform Compatibility-R supports cross-platform compatibility. It can run on any OS in any software environment. It can also be run on any hardware configuration without any extra workarounds. 6. Facilities for Various Industries-Almost every industry that makes use of data utilizes the R language. While only the academic areas made use of R in the past, it is now being heavily used in industries that require to mine insights from data. 7. Performs Fast Calculations-Through R, you can perform a wide variety of complex operations on vectors, arrays, data frames and other data objects of varying sizes. Furthermore, all these operations operate at a lightning speed. R provides various suites of operators to perform these miscellaneous calculations. 8. Integration with Other Technologies-R can be integrated with a number of different technologies, frameworks, software packages, and programming languages. It can be paired with Hadoop to use its distributed computing capabilities. It can also be integrated with programs in other programming languages like C++, Java, Python, and FORTRAN.	1 M for each feature (4 features are expected)
4.	Attempt any <u>THREE</u> of the following:	12 M
a)	Explain Mobile database with its need.	4 M
Ans	A Mobile database is a database that can be connected to a mobile computing device via a mobile network (or wireless network). Here the client and the server have wireless connections. In today's world, mobile computing is growing very rapidly, and it is huge potential in the field of the database. It will be applicable on different-different devices.	2 M for explanation and 2 M for definition

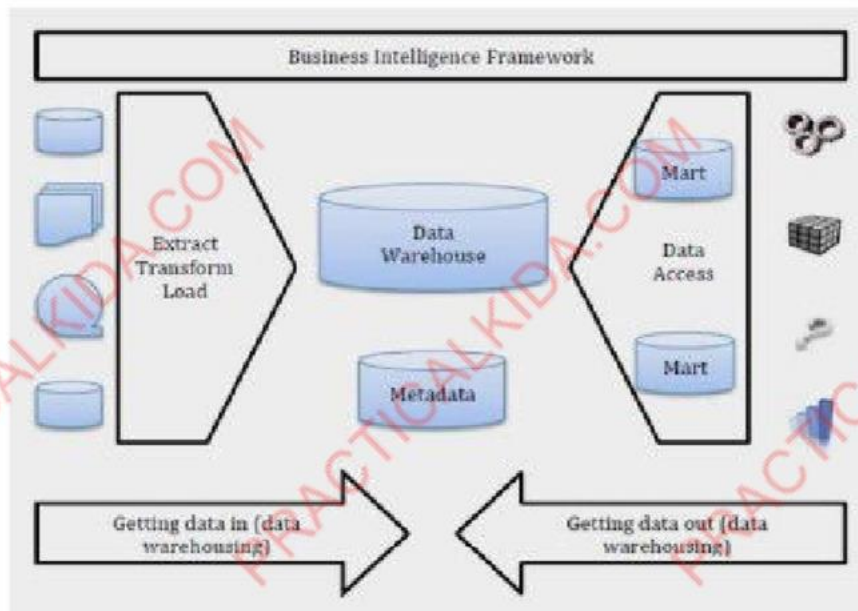


	like android based mobile databases, iOS based mobile databases, etc. C need of mobile database examples of databases are Couch base Lite, Object Box, etc. Need of Mobile database: • A cache is maintained to hold frequent and transactions so that they are due to connection failure. • As the use of laptops, mobile and PDAs is increasing to reside in the n system. • Mobile databases are physically separate from the central database serve • Mobile databases resided on mobile devices. • Mobile databases are capable of communicating with a central database s other mobile clients from remote sites. • With the help of a mobile database, mobile users must be able to work with wireless connection due to poor or even non-existent connections (disconn • A mobile database is used to analyze and manipulate data on mobile devi	
b)	Explain Three-Tier client server model.	4 M
Ans	Database Application Server Server Application Client User Client • The 3-Tier architecture contains another layer between the client and serv this architecture, client can't directly communicate with the server. • The application on the client-end interacts with an application server w further communicates with the database system. • End user has no idea about the existence of the database beyond the appl server. The database also has no idea about any other user beyond the appl • The 3-Tier architecture is used in case of large web application.	2 M for diagram and 2 M for explanation
c)	Describe Business Intelligence Framework & Features.	4 M
Ans	• Business intelligence (BI) refers to a technology-driven process that trans the data into actionable information. Organizations have a huge flow o coming from their customer end. • The data-driven decisions help in decision-making purposes in the organiz	2 M for BI framework



- It has a large dataset processed on relational databases.
- It represents results on dashboards and reports by charts and graphs with
- It depends on a small scale of past data, and there is no intelligence involved in business management has to decide based on the information.
- The quality of the solutions is volumetric and presents the accurate solution data visualizations.

and 2 M for features



Features of Business Intelligence:



		1. Data Warehouse-Data warehouse includes the process of collecting all the data needed. In addition, data warehouse provides a data storage environment where data onto multiple data sources will be ETLed(Extracted, Transformed, and Loaded), cleaned up, and stored on a specific topic, indicating powerful data integration and maintenance capabilities of BI. 2. Data Analysis-You need the ability of data analysis to aid in enterprise modeling. OLAP is a data analysis tool based on data warehouse environment. More precisely, it solves the shortcoming of low efficiency of multi-dimensional analysis based on OLTP analysis. Business intelligence (BI) leverages data analysis to form actionable insights that inform an organization's strategic and tactical business decisions. 3. Data Mining-In practical applications, data mining is also used to mine the data and predict the future. Data mining is a deeper level of active design and analysis of business data. It is a process of using knowledge discovery tools to extract previously unknown and potentially useful knowledge. It is an active method of automatic discovery. 4. Data Visualization-Data visualization can reflect business operations intuitively. Enterprises could conduct purposeful analysis on those abnormal data points to explore the possible causes. Then further make business decisions and strategies based on previous moves.	
	d)	Describe oracle cloud & its features.	4 M
	Ans	Oracle Cloud: • Oracle Cloud is one of the few cloud providers that can offer a complete set of cloud services to meet all your enterprise computing needs. • Use Oracle Infrastructure as a Service (IaaS) offerings to quickly set up virtual machines, storage, and networking capabilities you need to run just about any kind of workload. Your infrastructure is managed, hosted, and supported by Oracle. • Use Oracle Platform as a Service offerings to provision ready-to-run environments for your enterprise IT and development teams, so they can build and deploy applications, based on proven Oracle databases and application servers. • Use Oracle Software as a Service (SaaS) offerings to run your business from the Cloud. Oracle offers cloud-based solutions for Human Capital Management, Enterprise Resource Planning, Supply Chain Management, and many other business applications, all managed, hosted, and supported by Oracle. Features of oracle Cloud: • Oracle is completely scalable, as well as its style features a relatively distributed scale-out style with smart storage space, flash technology.	2 M for Oracle cloud and 2 M for features of oracle cloud



		• Oracle is the selection of the majority of successful companies and enterprises as a result of its reliable data source monitoring alternatives and alternatives. • Capacities to enhance efficiency. • On-Demand Throughput and Storage Provisioning.	
5.		Attempt any <u>TWO</u> of the following:	12 M
	a)	Explain flower expressions and Nested queries in XQuery.	6 M
Ans		The programming language XQuery defines FLWOR (pronounced 'flower') as an expression that supports iteration and binding of variables to intermediate results. FLWOR is an acronym: FOR, LET, WHERE, ORDER BY, RETURN. • For - selects a sequence of nodes • Let - binds a sequence to a variable • Where - filters the nodes • Order by - sorts the nodes • Return - what to return (gets evaluated once for every node) E.g: for \$x in doc("books.xml")/bookstore/book where \$x/price>30 order by \$x/title return \$x/title The for clause selects all book elements under the bookstore element into a variable called \$x. The where clause selects only book elements with a price element with a value greater than 30. The order by clause defines the sort-order. Will be sort by the title element. The return clause specifies what should be returned. Here it returns the title element. Nested XQuery Nested XQuery can be considered the multiple nested operation from the XQuery. In nested queries, a query is written inside a query. The result of inner query is used in execution of outer query For each book in the bibliography, list the title and authors, grouped inside a result element. E.g: <results> { for \$b in doc("bib.xml")/bib/book	3 M FLWOR Expressions explanation And 3 M explanation of Nested XQuery

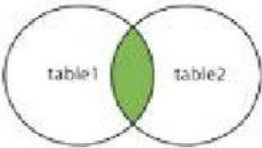
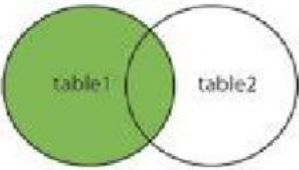
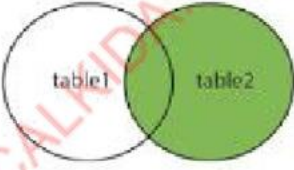
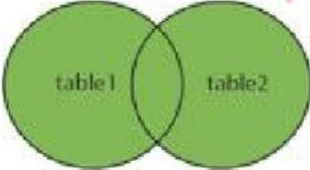


	<pre>return <result> { \$b/title } { for \$a in \$b/author return \$a } </result> } </results></pre>	
b)	Write query to execute find () function on collection: Inventory. (i) to display all document where status equals "A" and "G". (ii) to display all document in collection. (iii) to display all documents where quality is less than 30 and greater than 10.	6 M
Ans	Query :- <pre>db.inventory.insertMany([{ First_Name: "Radhika", Last_Name: "Sharma", qty: 30, e_mail: "radhika_sharma.123@gmail.com", status: "B" }, { First_Name: "Rachel", Last_Name: "Miller", qty: "15", e_mail: "Rachel_Christopher.123@gmail.com", status: "G" }, { First_Name: "Fathima", Last_Name: "Sheik", qty: "25", e_mail: "Fathima_Sheik.123@gmail.com", status: "A" }])</pre> i) db. inventory.find({ \$and : [{"status": "A"}, {"status": "G"}] }) ii) db.inventory.find()	2 M for each correct query



	iii) db.inventory.find({\$and: [{ "qty": { \$lt: 30 } }, { "qty": { \$gt: 10 } }] })	
c)	Explain inheritance for PL/SQL objects with examples	6 M
Ans	The object which inherits the behaviour of other objects is called as a Subtype and object which is being inherited is called as the Supertype. When an instance is created for the subtype, the instance inherits the behaviour of both the supertype and the subtype. A subtype cannot inherit the properties of two supertypes when there is no relation between them, similar to how a human child cannot have more than one father or mother. But the subtype created on a supertype can inherit the properties of that supertype and all its parent supertypes if present. A method in a supertype is by default overridden when the subtype provides a method with the same name and type, similar to its supertype. Example: - <pre>CREATE TYPE food_ot AS OBJECT (name VARCHAR2(100), food_group VARCHAR2 (50), grown_in VARCHAR2 (100)) NOT FINAL ;</pre> A very simple type: it contains just three attributes and no methods (procedures or functions). It also contains the NOT FINAL clause. This means, rather obviously, I am "not done." What that means in the context of object types is that I might want to create a subtype of this type. <pre>CREATE TYPE dessert_t UNDER food_ot (contains_chocolate CHAR (1), year_created NUMBER(4)) NOT FINAL ;</pre> Notice the UNDER food_ot syntax in place of the IS OBJECT, creating a new type "under" food, it clearly is an object type.	3 M Explanation And 3 M Example
6.	Attempt any <u>TWO</u> of the following:	12 M



	INNER JOIN  - Left Outer Join: Returns all records from the left table, and the matched records from the right table LEFT JOIN  - Right Outer Join: Returns all records from the right table, and the matched records from the left table RIGHT JOIN  - Full Join / Full Outer Join: Returns all records when there is a match in either or right table FULL OUTER JOIN 	
f)	Enlist Data Mining Techniques.	2 M
Ans	Following are the Data Mining Techniques: Classification, Clustering, Regression, Prediction, Association rules, Sequential pattern mining, Data cleaning, Data visualization, machine learning.	Any 4 -2M (1/2 M each)
g)	Write any four benefits of NOSQL.	2 M



a)	Define lock. Explain two phase locking protocol with neat example.	6 M
Ans	Locks: Lock is a data variable which is associated with a data item. Locks help synchronize access to the database items by concurrent transactions. All lock requests are made to the concurrency-control manager. Transactions pro TWO PHASE LOCKING PROTOCOL: - Two phase Locking protocol: which is also known as 2PL. Two phase locking protocol requires that each transaction issues lock and unlock requests in two phases. Growing phase: A transaction may obtain locks but may not release any lock. Shrinking phase: A transaction may release locks, but may not obtain any new locks. If the conversion is allowed, then upgrading of locks from S(A) to X(A) happens in growing phase and the downgrade of locks from X(A) to S(A) happens in shrinking phase. It is true that 2PL protocol offers serializability. However, it does not ensure that locks not happen. Example: -	1 M definition 3 M explanation 2 M example



		<table><tr><td></td><td>T1</td><td>T1</td></tr><tr><td>1</td><td>Lock-S(A)</td><td></td></tr><tr><td>2</td><td></td><td>Lock-S (A)</td></tr><tr><td>3</td><td>Lock- X(B)</td><td></td></tr><tr><td>4</td><td>---</td><td>---</td></tr><tr><td>5</td><td>Unlock(A)</td><td></td></tr><tr><td>6</td><td></td><td>Lock- X(C)</td></tr><tr><td>7</td><td>Unlock(B)</td><td></td></tr><tr><td>8</td><td></td><td>Unlock(A)</td></tr><tr><td>9</td><td></td><td>Unlock(C)</td></tr></table> This is just a skeleton transaction that shows how unlocking and locking work in PL. Note for: Transaction T1: • The growing Phase is from steps 1-3. • The shrinking Phase is from steps 5-7. • Lock Point at 3 Transaction T2: • The growing Phase is from steps 2-6. • The shrinking Phase is from steps 8-9. • Lock Point at 6		T1	T1	1	Lock-S(A)		2		Lock-S (A)	3	Lock- X(B)		4	---	---	5	Unlock(A)		6		Lock- X(C)	7	Unlock(B)		8		Unlock(A)	9		Unlock(C)	
	T1	T1																															
1	Lock-S(A)																																
2		Lock-S (A)																															
3	Lock- X(B)																																
4	---	---																															
5	Unlock(A)																																
6		Lock- X(C)																															
7	Unlock(B)																																
8		Unlock(A)																															
9		Unlock(C)																															
b)	Explain object & object identity. Write SQL query for the given table	<table><tr><td>Class: Employee</td></tr><tr><td>Name</td></tr><tr><td>Salary</td></tr><tr><td>Post</td></tr><tr><td>Gender</td></tr><tr><td>Store</td></tr><tr><td>Print</td></tr><tr><td>Update</td></tr></table>	Class: Employee	Name	Salary	Post	Gender	Store	Print	Update	6 M																						
Class: Employee																																	
Name																																	
Salary																																	
Post																																	
Gender																																	
Store																																	
Print																																	
Update																																	
Ans	SQL objects :- SQL objects are schemas, journals, catalogs, tables, aliases, views, indexes, constraints, triggers, sequences, stored procedures, user-defined functions, user-defined types, variables, and SQL packages. SQL creates and maintains these objects as system objects. Object identity :- Object identity is a fundamental object orientation concept. With object identity, objects can contain or refer to other objects.	3 M explanation 3 M for correct query																															



	Identity is a property of an object that distinguishes the object from all other objects in the application. Query for given table :- <pre>CREATE TYPE person AS object(Name varchar(10), Post varchar(30), Salary number(10), Gender varchar(7)) NOT FINAL METHOD print RETURNS text METHOD update RETURNS text; //method creation CREATE instance METHOD print RETURNS text FOR Person BEGIN RETURN self.name self.post self.salary self.gender; END CREATE instance METHOD update RETURNS text FOR Person BEGIN RETURN self.post="Manager"; END // for output SELECT Name, print FROM Person;</pre>	
c)	Consider following input data for your Map Reduce Program: Welcome to college class College is best	6 M



	College is bad Draw Map Reduce Architecture and explain its phases.	
Ans	The diagram illustrates the Map Reduce architecture with the following components and flow: Input: A yellow box containing the text: "Welcome to college", "Class", "College is best", "College is bad", "Bad". Input Splits: The input is divided into four pink boxes: "Welcome to college", "Class College is", "Best College is", and "Bad". Mapping: Each input split is processed by a mapper (green box) to produce key-value pairs: "Welcome to college" maps to "Welcome,1", "To, 1", "College,1". "Class College is" maps to "Class,1", "College,1". "Best College is" maps to "Best,1", "College,1". "Bad" maps to "bad,1". Shuffling: The mapped key-value pairs are shuffled into red boxes, grouped by their keys: - Key "bad": "bad,1" - Key "Class": "Class,1" - Key "Best": "Best,1" - Key "College": "College,1", "College,1", "College,1" - Key "is": "is,1", "is,1" - Key "to": "to,1" - Key "Welcome": "Welcome,1" Reducer: The shuffled data is processed by reducers (blue boxes) to produce intermediate values: - Key "bad": "bad,1" - Key "Class": "Class,1" - Key "Best": "Best,1" - Key "College": "College,1" - Key "is": "is,2" - Key "to": "to,1" - Key "Welcome": "Welcome,1" Final Output: A purple box containing the final aggregated results: "Bad1", "Class1", "Best1", "College1", "Is2", "To1", "Welcome1". Phases of Map Reduce are as follows:- Input Files: The data that is to be processed by the MapReduce task is stored in files. These input files are stored in the Hadoop Distributed File System. The file format is arbitrary, while the line-based log files and the binary format can also be used. Input Splits: An input in the MapReduce model is divided into small fixed-size parts called input splits. This part of the input is consumed by a single map. The input is generally a file or directory stored in the HDFS. Mapping: This is the first phase in the map-reduce program execution where the data of each split is passed line by line, to a mapper function to process it and produce the output values. Shuffling: It is a part of the output phase of Mapping where the relevant records are consolidated from the output. It consists of merging and sorting. So, all the key-value pairs which have the same keys are combined. In sorting, the inputs from the mapping step are taken and sorted. It returns key-value pairs, sorting the output. Reduce: All the values from the shuffling phase are combined and a single output is returned. Thus, summarizing the entire dataset.	3 M for map reduce architecture 3 M for explanation



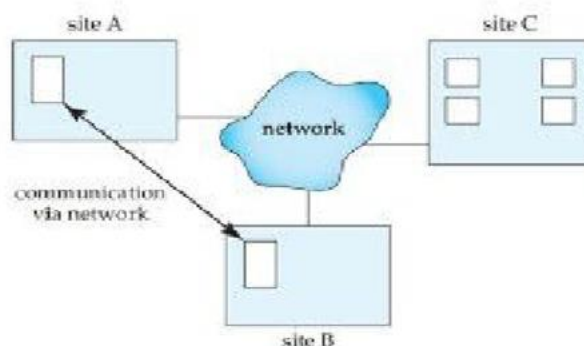
	Ans	Flexible Data Model: NoSQL databases are highly flexible as they can store and combine any type of data, both structured and unstructured, unlike relational databases that can store data in a structured way only. Evolving Data Model: NoSQL databases allow you to dynamically update the schema to evolve with changing requirements while ensuring that it would cause no interruption or downtime to the application. Elastic Scalability: NoSQL databases can scale to accommodate any type of data growth while maintaining a low cost. High Performance: NoSQL databases are built for great performance, measured in terms of both throughput (it is a measure of overall performance) and latency (it is the delay between request and actual response). Open-source: NoSQL databases don't require expensive licensing fees and can run on inexpensive hardware, rendering their deployment cost-effective.	Any 4 benefits-2M (1/2 M each)
2.		Attempt any <u>THREE</u> of the following:	12 M
	a)	Explain nested query with suitable example.	4 M
	Ans	• A nested query is a query that has another query embedded within it. • The embedded query is called a subquery. • A subquery typically appears within the WHERE clause of a query. • It can sometimes appear in the FROM clause or HAVING clause. • A nested query in SQL contains a query inside another query. • The result of the inner query will be used by the outer query. • Example: Select all Developers who earn more than all the Managers. SELECT * FROM employees WHERE role = 'Developer' AND salary > ALL (SELECT salary FROM employees WHERE role = 'Manager');	Explanation-3M Example-1M
	b)	Differentiate between SQL & NOSQL database system.	4 M



Ans			Any 4 points-4M
	SQL	NoSQL	
	Databases are categorized as Relational Database Management System (RDBMS).	NoSQL databases are categorized as Non-relational or distributed database system.	
	SQL databases have fixed or static predefined schema.	NoSQL databases have dynamic schema.	
	SQL databases display data in form of tables so it is known as table-based database.	NoSQL databases display data as collection of key-value pair, documents, graph databases or wide-column stores.	
	SQL databases are vertically scalable.	NoSQL databases are horizontally scalable.	
	SQL databases use a powerful language "Structured Query Language" to define and manipulate the data.	In NoSQL databases, collection documents are used to query the data, also called unstructured query language varies from database to database.	
	SQL databases are best suited for complex queries.	NoSQL databases are not so good for complex queries because these are not as powerful as SQL queries.	
	SQL databases are not best suited for hierarchical data storage.	NoSQL databases are best suited for hierarchical data storage.	
Ans	c) Explain distributed database in detail.		4 M
	- Distributed database is a database that is not limited to one computer system. It is like a database that consists of two or more files located in different computers or sites either on the same network or on an entirely different network. - These sites do not share any physical component. Distributed database is needed when a particular data in the database needs to be accessed by users globally. It needs to be handled in such a way that for a user it always looks like one single database. - By contrast, a Centralized database consists of a single database file located on one site using a single network.		Proper explanation with diagram-4M



- Logically connected single database and that will be dispersed to different nodes (connected by computer communication network).



Features of Distributed Databases

In general, distributed databases include the following features:

- Location independency:** Data is independently stored at multiple sites managed by independent Distributed database management systems (DDMS).
- Network linking:** All distributed databases in a collection are linked by a network and communicate with each other.
- Distributed query processing:** Distributed query processing is the procedure of answering queries (which means mainly read operations on large data sets) in a distributed environment.
- Hardware independent:** The different sites where data is stored are hardware independent. There is no physical contact between these distributed databases, which is accomplished often through virtualization.

Distributed transaction management: Distributed database provides a consistent distribution through commit protocols, distributed recovery methods, and distributed concurrency control techniques in case of many transaction failures.

Distributed Data Storage

Consider a relation r that is to be stored in the database. There are two approaches for storing this relation in the distributed database:

- Replication.** The system maintains several identical replicas (copies) of a relation, and stores each replica at a different site. The alternative to replication is to store only one copy of relation r .
- Fragmentation.** The system partitions the relation into several fragments and stores each fragment at a different site.



	d)	Explain Aggregation in MapReduce.	4 M
Ans	• Map-reduce is a data processing programming model that helps to perform operations on large data sets and produce aggregated results. • MongoDB provides the <code>mapReduce()</code> function to perform the map-reduce operations. This function has two main functions, i.e., map function and reduce function. • The map function is used to group all the data based on the key-value and the reduce function is used to perform operations on the mapped data. • So, the data is independently mapped and reduced in different spaces and then combined together in the function and the result will be saved to the specified collection. • This <code>mapReduce()</code> function generally operates on large data sets only. • Using Map Reduce you can perform aggregation operations such as max, min, etc. on the data using some key and it is similar to <code>groupBy</code> in SQL. • It performs operations on data independently and in parallel. In this example, we have five records from which we need to take out the maximum marks of each section and the keys are id, sec, marks. <pre>{ "id": 1, "sec": "A", "marks": 80 } { "id": 2, "sec": "A", "marks": 90 } { "id": 1, "sec": "B", "marks": 99 } { "id": 1, "sec": "B", "marks": 95 } { "id": 1, "sec": "C", "marks": 90 }</pre> Here we need to find the maximum marks in each section. So, our key by which we will group documents is the sec key and the value will be marks. Inside the map function, we use <code>emit(this.sec, this.marks)</code> function, and we will return the sec and marks of each record (document) from the emit function. This is similar to <code>groupBy</code> in MySQL. <pre>var map = function() { emit(this.sec, this.marks); }</pre> After iterating over each document the emit function will give back the data like this: <pre>{ "A": [80, 90], "B": [99, 90], "C": [90] }</pre> <pre>var reduce = function(sec, marks) { return Array.max(marks); }</pre>		Proper explanation with example-4M



	Here in reduce() function, we have reduced the records now we will output them a new collection. {out : "collectionName"} <pre>db.collectionName.mapReduce(map,reduce,{out : "collectionName"});</pre> In the above query we have already defined the map, reduce.	
3.	Attempt any <u>THREE</u> of the following:	12 M
a)	With neat diagram explain Data warehousing Lifecycle.	4 M
Ans	<pre>graph TD; RS((Requirement Specification)) --> DM((Data Modelling)); DM --> ETL((ETL Design and Development)); ETL --> OLAP((OLAP Cubes)); OLAP --> UI((UI Development)); UI --> Maint((Maintenance)); Maint --> Test((Test and Deployment)); Test --> RS; subgraph Center; DWC[(Data Warehousing Life Cycle)]; end</pre> 1. Requirement Specification: It is the first step in the development of the Data Warehouse and is done by business analysts. In this step, Business Analysts prepare business requirement specification documents. More than 50 requirements are collected from the client side and it takes 3-4 months to gather all the requirements. After the requirements are gathered, the data modeler recognizes the dimensions, facts & combinations based on the requirements. We can say that this is an overall blueprint of the data warehouse. But, this phase is more about determining business needs and placing them in the data warehouse. 2. Data Modelling: This is the second step in the development of the Data Warehouse. Data Modelling is the process of visualizing data distribution, designing databases by fulfilling the requirements to transform the data into a format that can be stored in the data warehouse. For example, whenever we are building a house, we put all the things in the correct position as specified in the blueprint.	1 M for diagram and 3 M for explanation



	3. ELT Design and Development: This is the third step in the development of Data Warehouse. ETL or Extract, Transfer, Load tool may extract data from various source systems and store it in a data lake. An ETL process can extract data from the lake, after that transform it and load it into a data warehouse for reporting. For optimal speeds, good visualization, and the ability to build replicable, and consistent data pipelines between all of the existing architecture and the new data warehouse, we need ELT tools. 4. OLAP Cubes: This is the fourth step in the development of the Data Warehouse. An OLAP cube, also known as a multidimensional cube or hypercube, is a structure that allows fast analysis of data according to the multiple dimensions that define a business problem. 5. UI Development: This is the fifth step in the development of the Data Warehouse. So far, the processes discussed have taken place at the back end. There is a need for a user interface for how the user and a computer system interact, in particular the use of input devices and software, to immediately access the data warehouse for analysis and generating reports. 6. Maintenance: This is the sixth step in the development of the Data Warehouse. In this phase, we can update or make changes to the schema and data warehouse's application domain or requirements. Data warehouse maintenance systems provide means to keep track of schema modifications as well, for instance schema modifications. At the schema level, we can perform operations for the Insert and change dimensions and categories. Changes are, for example, adding or deleting user-defined attributes. 7. Test and Deployment: This is often the ultimate step in the Data Warehouse development cycle. Businesses and organizations test data warehouses to see whether the required business problems are implemented successfully or not. Data warehouse testing involves the scrutiny of enormous volumes of data.	
b)	Explain Multimedia Databases.	4 M
Ans	Multimedia database is a collection of multimedia data which includes text, images, graphics (drawings, sketches), animations, audio, and video, among others. Multimedia databases have extensive amounts of data which can be multimedia and multisource. A framework which manages these multimedia databases and their different types of data, the data can be stored, utilized, and delivered in more than one way is known as a multimedia database management system. The multimedia database can be classified into three types. These types are: 1. Static media 2. Dynamic media 3. Dimensional media	4 M for explanation



	The contents of a multimedia database management system can be: 1. Media data: It is the actual data which represents an object. 2. Media format data: The information such as resolution, sampling rate, encoding system, etc. about the format of the media data under consideration undergoes acquisition, processing, and encoding is the media format data. 3. Media keyword data: Media keyword data are the keyword description related to the generation of data. This data is also known as content descriptive. Examples of content descriptive data are place, time, date of recording. 4. Media feature data: Media feature data contains data which is correlated dependent such as kind of texture, distribution of, and the different shapes in the data.									
c)	Explain with example any four operation with MongoDB.	4 M								
Ans	1. Create Operations –The create or insert operations are used to insert or new documents in the collection. If a collection does not exist, then it will create a new collection in the database. You can perform, create operations using following methods provided by the MongoDB: <table><tr><th>Method</th><th>Description</th></tr><tr><td>db.collection.insertOne()</td><td>It is used to insert a single document in collection.</td></tr><tr><td>db.collection.insertMany()</td><td>It is used to insert multiple documents in collection.</td></tr><tr><td>db.createCollection()</td><td>It is used to create an empty collection.</td></tr></table> Example: In this example, we are inserting details of a single student in the form of document in the student collection using db.collection.insertOne () method.	Method	Description	db.collection.insertOne()	It is used to insert a single document in collection.	db.collection.insertMany()	It is used to insert multiple documents in collection.	db.createCollection()	It is used to create an empty collection.	1 M for each correct operation (4 operation should explained)
Method	Description									
db.collection.insertOne()	It is used to insert a single document in collection.									
db.collection.insertMany()	It is used to insert multiple documents in collection.									
db.createCollection()	It is used to create an empty collection.									