Shivaji University , Kolhapur

Question Bank For Mar 2022 (Summer) Examination

Subject Code **83858**   Subject Name : **Advanced Database Systems**

Unit I: Parallel and Distributed Databases

1. What are parallel systems? Explain Speedup and Scaleup in parallel systems with the help of diagram.
2. What are parallel systems? Explain parallel database architectures in detail with diagrams.
3. Compare all parallel database system architectures.
4. Explain data partitioning techniques used in parallel databases. Also give comparison between data partitioning techniques.
5. What is parallel database system? Explain data partitioning techniques used in parallel databases.
6. Explain pipelined parallelism and independent parallelism in detail.
7. Explain fragment-and-replicate join with diagrams.
8. Explain design of parallel systems.
9. Explain how to store data in distributed database systems.
10. What is distributed database systems? Explain horizontal and vertical fragmentation used in distributed database systems.
11. Explain distributed transactions in detail.
12. What are the different types of distributed database system? Explain Semijoin strategy used in distributed query processing.
13. Explain two phase commit (2PC) protocol in brief. Also explain how 2PC protocol handles failure of a participating site and failure of a coordinator.

UNIT 2: Advanced SQL

1. Explain various numeric and string functions used in oracle with syntax and example.

2. Describe oracle sequence. Explain sequence in Oracle with syntax and example.

3. What is trigger? What is its purpose? Explain types of trigger?

4. What is trigger? Explain in detail, syntax to create trigger in oracle.

5. What is stored procedure in PL/SQL? Give its advantages. Explain in detail, syntax to create stored procedure in PL/SQL.

6. Explain in detail, syntax to create stored procedure in PL/SQL. Write PL/SQL Procedure to find factorial of given number.

7. Explain in detail, syntax to create function in PL/SQL. Write PL/SQL function to find factorial of given number.

8. What is procedural SQL? What are its advantages? Give difference between stored procedure and function in PL/SQL.

9. What is cursor in PL/SQL? Describe types of cursor. Also describe cursor attributes in detail.

10. Give difference between SQL and host language. What is embedded SQL and how it is used?

UNIT 3: NoSQL Database Management

1. What is NoSQL? Explain types of NoSQL databases in detail.

2. Explain sorted ordered column oriented stores. Also explain HBase and Hypertable.

3. Explain key/value stores in detail. Also explain cassendra.

4. Explain document databases in detail. Also explain MongoDB and CouchDB.

5. What is MongoDB? Explain its features, advantages and disadvantages.

6. Describe MongoDB. Explain following various commands of MongoDB with syntax and example.

    A. Use
    B. Insert()
    C. Find()
    D. Save()

7. What is MongoDB? What is collection and document used in MongoDB ?

Consider the following collection

student (rollno, name, marks).

Solve following Queries using MongoDB

1. Create collection student (1M)

2. Insert 2 documents in student collection (1M)

3. Display all documents of student collection. (1M)

4. Display students having marks greater than 70. (2M)

5. Display only one document of student collection. (1M)

8. What is CouchDB? Give difference between MongoDB and CouchDB.

UNIT NO 04

1. Explain data-information in decision making cycle
2. How to maintain security for Database Administration?
3. What is need of database in an organization?
4. Explain special considerations aspects of a DBMS into an organization.
5. Explain role of the database in an organization development.
6. Explain in brief evolution of the database administration function.
7. Explain DA and DBA characteristics
8. Explain DA and DBA activities
9. Summaries DBA activities with suitable diagram
10. What are the desired DBA skills ?
11. Explain DBA activities and services.
12. Explain DBA's technical role .
13. Differentiate Security Vulnerability & Security Measures
14. Different Database Administration Tools.
15. Explain the difference between Data And Information. Give some examples of raw data and information.
16. Define Dirty Data and identify some of its sources.

UNIT NO 05

1. What is Business Intelligence ? explain in brief.
2. How the Data Warehouses is useful for Data Analysis ?
3. Explain Business Intelligence Architecture with suitable diagram.
4. Explain Decision Support Data.
5. What is Online Analytical Processing ?
6. What is Star Schemas Explain steps for implementing a Warehouse?
7. What is Data Mining Explain with suitable example ?
8. Explain in brief SQL Extensions for OLAP
9. What is Materialized Views ?
10. What is FireBase ?
11. What are Decision Support Systems?Explain how OLAP tools can be used for data analysis ?
12. Compare OLIP and OLAP.
13. What is the role of the metadata repository in a data warehouse? How does it differ from a catalog ie relational DBMS ?

UNIT NO 06

1. What is mathematical model ? Explain different classes of model.
2. Enlist and explain Different Mathematical models for decision making
3. What is Data preparation for data mining ?
4. What is Data exploration ?
5. What are different applications of data mining ?
6. Explain in details data mining process
7. What is Data validation ?
8. What is Data transformation ?
9. What is data reduction ?
10. Differentiate Univariate analysis & Bivariate analysis
11. Explain Multivariate Analysis.
12. What is Association rule mining ? Explain Single-dimension association rules
13. Explain Apriori Algorithm.
14. Explain general Association Rules
15. What is data-mining? Explain the analysis methodologies ?

MCQ All UNITS

Sr.No.	Questions	Answers
1.	When a participating site S_1 decides to commit the transaction T upon receiving the <prepare T> message from the coordinator, it sends _____ message to the coordinator site. A) <prepare T> B) <commit T> C) <ready T> D) <abort T>	C
2.	A heterogeneous distributed database is which of the following? A) The same DBMS is used at each location and data are not distributed across all nodes. B) The same DBMS is used at each location and data are distributed across all nodes. C) A different DBMS is used at each location and data are not distributed across all nodes. D) A different DBMS is used at each location and data are distributed across all nodes.	D
3.	Which of the following parallel database architecture is mainly used by distributed database system? A) Shared Memory B) Shared Disk C) Shared Nothing D) Hierarchical	C
4.	_____ refers to the execution of a single query in parallel on multiple processors and disks. A) Interquery parallelism B) Intraquery parallelism C) Interdependent parallelism D) None of the above.	B

5.	A semijoin is which of the following? A) Only the joining attributes are sent from one site to another and then all of the rows are returned. B) All of the attributes are sent from one site to another and then only the required rows are returned. C) Only the joining attributes are sent from one site to another and then only the required rows are returned. D) All of the attributes are sent from one site to another and then only the required rows are returned.	C
6.	What is the difference between PL/SQL Function and PL/SQL Procedure? A) PL/SQL function may or may not return the value whereas PL/SQL Procedure must have to return the value. B) PL/SQL Procedure may or may not return the value whereas PL/SQL Function must have to return the value. C) PL/SQL Function may or may not return the function whereas PL/SQL Procedure must have to return the function. D) None of the above	B
7.	How many types of PL/SQL Cursor are there? A) 1 B) 2 C) 3 D) 4	B
8.	_____ is a popular, open-source, sorted ordered column-family store that is modeled on the ideas proposed by Google's Bigtable. A) HBase B) Hypertable C) Cloudata D) None of the above	A

9.	_____ has properties of both Google Bigtable and Amazon Dynamo. A) Voldemort B) Cassendra C) Riak D) None of the above	B
10.	What kind of database MongoDB is? A) Graph Oriented B) Document Oriented C) Key Value Pair D) Column Based	B
11.	_id in MongoDB is a _____ bytes hexadecimal number which assures the uniqueness of every document. A) 12 B) 13 C) 14 D) None of the above	A
12.	Poor data administration can lead to which of the following? A) A single definition of the same data entity & Missing data elements B) Familiarity with existing data C) All A,B,D D) Missing data elements	D
13.	Point out the wrong statement. A) BI is a category of database software that provides an interface to help users quickly and interactively scrutinize the results in a variety of dimensions of the data & Data is factual information for analysis B) BI is a category of database software that provides an interface to help users quickly and interactively scrutinize the results in a variety of dimensions of the data C) Customer relationship management (CRM) entails all aspects of interaction that a company has with its customer D) None of the mentioned	B
14.	Which of the following does not form part of BI Stack in SQL Server? A) OIBEE	D

	B) OSSIS C) OBSAS D) OBIEE	
15.	Which of the following is the preferred way to recover a database after a system failure? A) Rollback & Roll forward B) Switch to duplicate database & Roll forward C) Switch to duplicate database D) Reprocess transactions	C
16.	KPI stands for? A) Key Performance Indicators B) Key Performance Identifer C) Key Processes Identifer D) Key Processes Indicators	A
17.	A shared lock allows which of the following types of transactions to occur? A) Delete& Insert B) Read & Insert C) Read D) Update & Insert	C
18.	This is an XML-based metalanguage developed by the Business Process Management Initiative (BPMI) as a means of modeling business processes, much as XML is, itself, a metalanguage with the ability to model enterprise data. A) BizTalk B) BPML C) e-biz D). ebXML	B
19.	Performance analysis and tuning is which of the following? A) Undertaken one time at the time of a DBMS installation B) Undertaken one time at the time of the implementation of a new application C) Undertaken as an ongoing part of the managing & backup of a database D) Undertaken as an ongoing part of managing a database	D
20.	Patterns that can be discovered from a given database are which type… A) More than one type B) Multiple type always C) One type only D) No specific type	A
21.	Firms that are engaged in sentiment mining are analyzing data collected from?	A

	A) social media sites. B) in-depth interviews. C) focus groups. D) experiments.	
22.	——— is the out put of KDD… A) Query B) Useful Information C) Data D) information	B
23.	Which of the following applied on warehouse? A) write only B) read only C) both a & b D) none of these	B
24.	The transaction log includes which of the following? A) The essential data of the record &The before-image of a record B) The before and after-image of a record & The after-image of a record C) The before and after-image of a record D) The essential data of the record	D
25.	In which of the following architectures memory bus is not a bottleneck? A) Shared memory and shared disk B) Shared disk and shared nothing C) Shared memory and shared nothing D) All of the above	B
26.	Storing a separate copy of the database at multiple locations is which of the following? A) Data Replication B) Horizontal Partitioning C) Vertical Partitioning D) Horizontal and Vertical Partitioning	A
27.	Which of the following parallel database architecture is mainly used by distributed database system? A) Shared Memory B) Shared Disk C) Shared Nothing D) Hierarchical	C
28.	In _____, different queries or transactions execute in parallel	A

	with one another. A) Interquery parallelism B) Intraquery parallelism C) Independent parallelism D) None of the above.	
29.	If we are to ensure atomicity, all the sites in which a transaction T executed must agree on the final outcome of the execution. T must either commit at all sites, or it must abort at all sites. To ensure this property, the transaction coordinator of T must execute a _____ A) Commit B) Commit protocol. C) Rollback D) None of the above	B
30.	If the speed of a parallel system is N when the larger system has N times the resources of the smaller system, then the speedup is _____ A) Linear Speedup B) Sublinear Speedup C) Superlinear Speedup D) None of the above	A
31.	How can we specify a row-level trigger? A) Using ON ROW B) Using FOR EACH COL C) Using FOR EACH ROW D) Using OR ROW	C
32.	A _____ is created to hold the output of an SQL statement that may return two or more rows (but could return 0 or only one row). A) Implicit cursor B) Explicit cursor C) Both A and B D) None of the above	B
33.	Which statement is incorrect A) Stored procedure may return a value and function must return a value.	C

	B) Function has only IN parameter.	
	C) Try and Catch can be used with both stored procedure and function.	
	D) Stored procedure has IN and OUT parameter.	
34.	What is the aim of NoSQL? A) NoSQL is not suitable for storing structured data. B) NoSQL provides an alternative to SQL databases to store textual data. C) NoSQL is a new data format to store large datasets. D) None of the above.	C
35.	_____ can be used for batch processing of data and aggregation operations. A) Hive B) MapReduce C) Oozie D) None of the mentioned	B
36.	What kind of database MongoDB is? A) Graph Oriented B) Document Oriented C) Key Value Pair D) Column Based	B
37.	What is frequent pattern growth? A) Same as frequent itemset mining B) Use of hashing to make discovery of frequent itemsets more efficient C) Mining of frequent itemsets without candidate generation D) None of these	C
38.	When do you consider an association rule interesting? A) If it only satisfies min_support B) If it only satisfies min_confidence C) If it satisfies both min_support and min_confidence D) There are other measures to check so	C
39.	What techniques can be used to improve the efficiency of apriori algorithm? A) hash based techniques B) transaction reduction C) Partitioning D) All of these	D

40.	Which of the following is a descriptive model? A) Regression B) Regression & Association Rule C) Association rules. D) Sequence discovery.	D
41.	What are closed frequent itemsets? A) A closed itemset B) A frequent itemset C) An itemset which is both closed and frequent D) None of these	C
42.	—— is not a data mining functionality? A) Clustering and Analysis B) Selection and interpretation C) Classification and regression D) Characterization and Discrimination	B
43.	Which of the following is best practices for firebase? A) Avoid Nesting Data B) Denormalize Data C) Both A and B D)None of the above	C
44.	Founders of Firebase? A) Google B) James Tamplin C) Donald D & Chamberlin D. Raymond F. Boyce	B
45.	Which of the following is true? A) Both apriori and FP-Growth uses horizontal data format B) Both apriori and FP-Growth uses vertical data format C) Both a and b D)None of these	A
46.	Capability of data mining is to build _____ models. A) predictive B) imperative C) interrogative D) retrospective	A
47.	The apriori algorithm works in a ..and ..fashion? A) top-down and depth-first B) top-down and breath-first C) bottom-up and depth-first D) bottom-up and breath-first	D
48.	Removing duplicate records is a process called .... recovery A) data pruning B) data cleaning	C

	C) data cleansing D) data cleaning & data pruning	
49.	Which of the following are interestingness measures for association rules? A) recall B) lift C) accuracy D) compactness	B
50.	The important metrics in measuring the efficiency of the parallel database system is A) Scaleup B) Speedup C) Both A and B D) None of the above	C
51.	Which of the following helps in implementing interoperation parallelism? A) Interdependent parallelism B) Intraquery parallelism C) Interquery parallelism D) Pipelined parallelism	D
52.	A homogeneous distributed database is which of the following? A) The same DBMS is used at each location and data are not distributed across all nodes. B) The same DBMS is used at each location and data are distributed across all nodes. C) A different DBMS is used at each location and data are not distributed across all nodes. D) A different DBMS is used at each location and data are distributed across all nodes.	B
53.	A distributed database is which of the following? A) A single logical database that is spread to multiple locations and is interconnected by a network B) A loose collection of file that is spread to multiple locations and is interconnected by a network	A

	C) A single logical database that is limited to one location.	
	D) A loose collection of file that is limited to one location.	
54.	If the speed of a parallel system is N when the larger system has N times the resources of the smaller system, then the speedup is _____ A) Linear Speedup B) Sublinear Speedup C) Superlinear Speedup D) None of the above	A
55.	If we parallelize the execution of large number of small transactions in a database system, then which of the following would be improved? A) Response time B) Rotational latency C) Throughput D) All of the above	C
56.	Which of the following are benefits of Triggers? A) Automatic generation of derived column values. B) Auditing C) Creation of replica tables for backup purposes. D) All of the above	D
57.	An _____ is automatically created in procedural SQL when the SQL statement returns only one value. A) Implicit cursor B) Explicit cursor C) Both A and B D) None of the above	A
58.	Advantage of stored procedure A) Reduce network traffic and increase performance B) Re-use of code C) Security D) All of the above	D

59.	A collection and a document in MongoDB is equivalent to.............. concepts respectively. A) Table and Column B) Table and Row C) Column and Row D) Database and Table	B
60.	Which of the following format is supported by MongoDB ? A) XML B) BSON C) SQL D) All of the above	B
61.	Which of the following command in MongoDB will return all the posts with number of likes greater than 100 and less than 200, both inclusive? A) db.posts.find({ likes : { $gt : 100, $lt : 200 } }); B) db.posts.find({ likes : { $gte : 100, $lt : 200 } }); C) db.posts.find({ likes : { $gt : 100 , $lte : 200 } }); D) db.posts.find({ likes : { $gte : 100 , $lte : 200 } });	D
62.	Which of the following areas are affected by BI? A) Revenue B) CRM C) Sales D)CPM	B
63.	_____ is a system where operations like data extraction, transformation and loading operations are executed. A) Data staging B) Data integration C) ETL D) None of the above	A
64.	Data that are not of interest to the data mining task is called as .... A) Noisy data B) Missing data C) Changing data D) Irrelevant data	D
65.	What does Apriori algorithm do?	A

	A) It mines all frequent patterns through pruning rules with lesser support B) It mines all frequent patterns through pruning rules with higher support C) Both a and b D) None of these	
66.	BI In an Internet context, this is the practice of tailoring Web pages to individual users' characteristics or preferences. A) Web services B) customer-facing C) client/server D) personalization	D
67.	This is a broad category of applications and technologies for gathering, storing, analyzing, and providing access to data to help enterprise users make better business decisions. A). best practice B). data mart C). business information warehouse D). business intelligence	D
68.	_____ is a category of applications and technologies for presenting and analyzing corporate and external data. A) MIS B) DIS C) EIS D) CIS	C
69.	How many types of BI users are there? A) 2 B) 3 C) 4 D)5	C
70.	Locking may cause which of the following problems? A) Erroneous updates B) Deadlock C) Versioning D) All of the above.	B
71.	Point out the correct statement. A) OLAP is an umbrella term that refers to an assortment of software applications for analyzing an organization's raw data for intelligent decision making B) Business intelligence equips enterprises to gain business advantage from data C) BI makes an organization agile thereby giving it a lower edge in today's evolving market condition D) None of the mentioned	B
72.	Which of the following is the preferred way to recover a database	A

	after a transaction in progress terminates abnormally? A) Rollback B) Rollforward C) Switch to duplicate database D) Reprocess transactions	
73.	A data warehouse administrator is concerned with which of the following? A) The time to make a decision but not the typical roles of a database administrator B) The time to make a decision and the typical roles of a database administrator C) The typical roles of a data administrator and redesigning existing applications D) The typical roles of a database administrator and redesigning existing applications	B
74.	If both data and database administration exist in an organization, the database administrator is responsible for which of the following? A) Data modeling B) Database design C) Metadata D) All of the above.	B
75.	In which fragmentation, each fragment is a subset of the tuples in the original relations? A) Horizontal B) Vertical C) Both A and B D) None of the above	A