DISTRIBUTED DATABASE MANAGEMENT SYSTEMS
(INF3132)

Time Allotted : 2½ hrs **Full Marks : 60**

Figures out of the right margin indicate full marks.

Candidates are required to answer Group A and any 4 (four) from Group B to E, taking one from each group.

Candidates are required to give answer in their own words as far as practicable.

Group – A

1. Answer any twelve: **12 × 1 = 12**

Choose the correct alternative for the following

(i) Fragmentation transparency ensures that:
(a) Users know where data fragments are stored.
(b) Users can query data without knowing how it's fragmented.
(c) Data is replicated at every site.
(d) Only vertical fragmentation is allowed.

(ii) A set of simple predicates is minimal if each of the simple predicate
(a) is relevant (b) is not relevant
(c)is complete (d) is disjoint

(iii) Sorting in query processing is commonly implemented using:
(a) Merge sort (b) Quick sort
(c) Bubble sort (d) Insertion sort

(iv) The hash join algorithm is particularly efficient when:
(a) Both relations are sorted (b) There is no index on the join attribute
(c) One relation is very large (d) Join attributes are numeric

(v) Which of the following is true about the NameNode in Hadoop.
(a) It stores the actual data in HDFS.
(b) It manages the metadata and file system namespace.
(c) It handles the execution of MapReduce jobs.
(d) It acts as a backup for the DataNodes.

(vi) Which of the following Hadoop components is responsible for executing MapReduce tasks.
(a) DataNode (b) JobTracker
(c) TaskTracker (d) ResourceManager

(vii) Which of the following is NOT an advantage of using Hadoop.
(a) Scalability (b) Fault tolerance
(c) High speed for small datasets (d) Ability to process large volumes of data

(viii) In HBase, what is a column family.
(a) A group of columns that are logically related and stored together on disk.
(b) A collection of tables within a single HBase instance.
(c) A unique identifier for rows in a table.
(d) A collection of row keys that are grouped together.

(ix) What is the default data model used by MongoDB.
(a) Relational tables with rows and columns (b) Graph-based data model
(c) Document-based data model (d) Key-value pair data model

(x) What is the "primary key" equivalent in MongoDB.
(a) Document ID (b) Collection Name
(c) Field Value (d) Index

Fill in the blanks with the correct word

(xi) The allocation schema specifies where each _____ of data is stored.

(xii) In **vertical fragmentation**, attributes are divided based on _____ requirements.

(xiii) **Derived horizontal fragmentation** is based on a relation that is _____ to another relation.

(xiv) The final step in query processing is query _____, which executes the plan chosen by the optimizer.

(xv) The ___ nested loop join reduces I/O by reading multiple tuples in a block at once.

Group - B

2. (a) What are the three levels of distribution transparency. Explain the two rules that a set of simple predicates must satisfy. *[(CO1)(CO3)(Understand/LOCQ)]*

(b) A distributed database system, have been fragmented into 3 fragments F1, F2, and F3. The fragments are allocated at 3 sites. The following applications are issued with same frequency.
P1 issued at site 1, reads 9 records of F1, updates 4 records of F1 and 5 records of F2
P2 issued at site 3 reads 14 records of F3 and 9 records of F2
P3 issued at site 2 reads 25 records of F2
Considering locality of references as objective, find out the optimal solution
[(CO3)(Evaluate/HOCQ)]

(3 + 2) + 7 = 12

3. (a) Consider a relation Employee (Emp-ID, Emp-Name, Salary, Branch) with branches located at Salt Lake, Park Street, and Rajarhat.
One application accesses data based on branch, while another application retrieves all records where Salary < 60000.
i) Design the set of simple predicates.
ii) Derive the minterm predicates.
iii) Identify the valid minterm predicates.
iv) Justify the completeness property with respect to the created fragments.

v) If a new application arrives that accesses tuples with respect to both Branch and Salary ≥ 60000, redesign the predicate set to preserve completeness. *[(CO3)(Create/HOCQ)]*

(b) Differentiate between primary and derived horizontal fragmentation with examples. *[(CO1)(Understand/LOCQ)]*

7 + 5 = 12

Group - C

4. (a) Explain the algorithm of 2PC protocol in distributed environment for bothCoordinator and Participants. *[(CO4)(Understand/LOCQ)]*

(b) Explain the different scenarios of node/site failures of both coordinator and participants in 2PC environment. *[(CO4)(Analyse/IOCQ)]*

6 + 6 = 12

5. (a) State 3 rules of query optimization *[(CO5)(Remember/LOCQ)]*

(b) Consider, two relations Student and Course. Student has 20000 records and it requires 300 blocks to store these records. Course has 5000 records, and blocks required is 100. In order to join the two tables, if nested loop join algorithm is used then calculate the number of block access required for best case and worst case. Which relation will be used as outer relation and which one as inner relation. Explain. *[(CO5)(Evaluate/HOCQ)]*

(c) What is the difference between nested loop join and block nested loop join algorithm. *[(CO5)(Analyse/IOCQ)]*

3 + 5 + 4 = 12

Group - D

6. (a) Describe with the help of a diagram the architecture of Hadoop Distributed File System. *[(CO2)(Understand/LOCQ)]*

(b) Suppose you have a word file with the following text.
"The world is going through a huge crisis. God save the world.
The world is beautiful" The file size is 110 MB. Explain how the file gets broken into input splits and explain the overall steps in mapper and reducer. *[(CO2)(Demonstrate/HOCQ)]*

5 + 7 = 12

7. (a) Explain how parallelism is encountered in Map Reduce paradigm.

(b) Using the Map Reduce paradigm compute the number of words startingwith vowel and number of words starting with consonant in the following text."There is a Workshop in HIT. The workshop is on Big Data Analytics. Heritage is in Kolkata." *[(CO2)(Apply/IOCQ)]*

4 + 8 = 12

Group - E

8. (a) Explain with an example how HBase is horizontally scalable *[(CO6)(Analyse/IOCQ)]*

(b) Explain the architecture of HBASE distributed database with light on Regions and HFiles. *[(CO6)(Analyse/IOCQ)]*

5 + 7 = 12

9. (a) Compare the ACID and BASE consistency model in the context of NoSQL databases. *[(CO6)(Analyse/IOCQ)]*

(b) Explain the challenges in migrating from RDBMS to NoSQL databases? *[(CO6)(Analyse/IOCQ)]*

(c) What are document and collection in the context of MongoDB database. State the advantages of MongoDB database. *[(CO6)(Understand/LOCQ)]*

4 + 3 + (2 + 3) = 12

Cognition Level	LOCQ	IOCQ	HOCQ
Percentage distribution	34.37	38.54	27.08