

COLLEGE OF ENGINEERING ,PUNE.

(An Autonomous Institute of Government of Maharashtra, Pune - 411005)

End Semester Examination

Sub: (CT-410) Advanced DBMS

Year : B.Tech(I.T)

Duration: 3 hrs

Academic Year : 2012-13

Max. Marks : 50

- Instructions:** 1. All questions are compulsory.
2. Figures to right indicate full marks.

- Q. 1**
- a** Briefly describe and compare the Client-Server and Collaborating Servers architectures. 3
- b** With pipelined parallelism, it is often a good idea to perform several operations in a pipeline on a Single processor, even when many processors are available. 3
- a. Explain why.
- b. Would the arguments you advanced in part a hold if the machine has a shared-memory architecture? Explain why or why not.
- c. Would the arguments in part a hold with independent parallelism?
- c** Consider a relation that is fragmented horizontally by *plant_number*: 4
- employee (name, address, salary, plant_number)*
- Assume that each fragment has two replicas: one stored at the Delhi site and one stored locally at the plant site. Describe a good processing strategy for the following queries entered at Pune site.
- a. Find all employees at the Mumbai plant.
- b. Find the average salary of all employees.
- c. Find the highest-paid employee at each of the following sites: Calcutta, Channai, Bengaluru, and Ahmadabad.
- d. Find the lowest-paid employee in the company.
- Q. 2**
- a** Compare the relative merits of centralized and hierarchical deadlock detection in a distributed DBMS. 2
- b** What is a *phantom deadlock* ? Give an example. 2
- c** Given that LDAP functionality can be implemented on top of a database system, what is the need for the LDAP standard? 2

d Suppose you have a database with two tables with the following schema:

4

T1 : (A int PRIMARY KEY, B int, C varchar)

T2 : (D int REFERENCES T1.A, E varchar)

And suppose you are running the following query:

```
SELECT T1.A,COUNT(*)
FROM T1,T2
AND T1.A = T2.D
AND T1.B > n
GROUP BY T1.A
```

Your system has the following configurations:

Parameter	Description	Value
T1	Size of table T1	300 MB
T2	Size of table T2	600 MB
M	Size of memory of a single node	100 MB
T1	Number of records in T1	10 million
int	Size of integer, in bytes	4
T	Disk throughput, in bytes / sec	20 MB/sec
N	Network transmit rate, in bytes / sec	10 MB/sec
f	Fraction of tuples selected by T1.B > n predicate	0.5

Suppose you run the query on a single node. Assuming there are no indices, and the tables are not sorted, indicate what join algorithm would be most efficient, and estimate the time to process the query, ignoring disk seeks and CPU costs. Include a brief justification or calculation.

- Q. 3**
- a Differentiate OLTP with OLAP 3
 - b Write short notes on multidimensional data model 3
 - c List and explain the OLAP operations in multidimensional data model? 4

Q. 4 a *Transaction Id* *Items*

4

234145	{X, Z}
543789	{U, V, W, X, Y, Z}
965157	{U, W, Y}
489651	{V, X, Z}
48965	{U, Y}
325687	{W, X, Y}
147895	{X, Y, Z}
617851	{U, Z}
824697	{V, Y}
102458	{V, W, X}

Using the given data, for each of the rules given below, compute support and confidence.

$V \rightarrow W$

$X \rightarrow Z$

$U \rightarrow Y$

$Z \rightarrow V$

$W \rightarrow U$

$Y \rightarrow X$

OR

Given the table below:

4

Age	Income	Class
30	high	no
35	high	yes
40	medium	yes
40	low	yes
40	low	no
35	low	yes
30	medium	no
30	low	yes
30	medium	yes
35	medium	yes
35	high	yes
40	medium	no

Build a decision tree for the above training data. Perform a multi-way split for Income, and use Gain Ratio (using entropy) as the criteria to judge the goodness of a split. Show the decisions for internal nodes and class labels at the leaf nodes.

3/4

- b** How is a data warehouse different from a database 2
 - c** Differentiate Agglomerative and Divisive Hierarchical Clustering 2
 - d** Describe Tree pruning methods. 2
- Q. 5 a** Explain the difference between a system crash and a 'disaster'. Workflow systems require concurrency and recovery like database systems. List 3 reasons why simple RDBMS cannot be applied using 2PL, WAL and 2PC. 5

OR

Discuss the drawbacks in ensuring serializability for long duration transactions. Is a high-performance transaction system necessarily a real-time system? Give reasons.

- b** Explain how a TP monitor manages memory and processor resources more effectively than a typical OS. What is the purpose of compensating transactions? 5