UNIVERSITY OF PUNE

[4364]-777

B. E. (*Computer Engg*)(SEM-II) Examination - 2013

Advanced Databases (2008 Course)

[Time: *3* Hours]   **[Max. Marks: 100]**

Instructions:

1   *Answer any three questions from each section.*

2   *Answers to the **two sections** should be written in **separate answer-books**.*

3   *Neat diagrams must be drawn wherever necessary.*

4   *Black figures to the right indicate full marks.*

5   *Use of logarithmic tables, slide rule, Mollier charts, electronic pocket calculator and steam tables is allowed.*

6   *Assume suitable data, if necessary.*

SECTION -I

Q.1   A   For each of the three partitioning techniques, namely round-robin,   6
               hash partitioning, and range partitioning, give an example of a
               query for which that partitioning technique would provide the
               fastest response.

        B   Write a short note on Parallel query optimization.   5

        C   What is interquery parallelism? Explain cache coherency problem   6
               and protocol available to guarantee cache coherency.

OR

Q.2	A	Explain parallel External sort-Merge with example.	6
	B	Explain interoperation parallelism suitable example.	6
	C	What factor could result in skew when a relation is partitioned on one of its attributes by Hash partitioning and Range partitioning. In each case, what can be done to reduce the skew?	5

Q. 3	A	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. Describe the techniques or protocol used to ensure this property in detail.	7
	B	Explain semijoin strategy along with example.	6
	C	Write short note on LDAP.	4

OR

Q. 4	A	Explain deadlock handling with respect to distributed database.	6
	B	Explain the difference between data replication in a distributed system and the maintenance of a remote backup site.	2
	C	What are the different approaches to store a relation in the distributed database. Explain them in brief.	6
	D	Write short note on multidatabase system.	3

| Q. 5 | A | Write short notes on: | 8 |

    i)   SOAP.
    ii)   XML DTD.

| | B | Explain the structure of XML data with example. | 8 |

OR

| Q. 6 | A | Explain the following with respect to web architecture | 8 |

    i)   Web server.
    ii)   Common gateway interface.
    iii)   Cookie.
    iv)   Uniform Resource Locator.

| | B | Which are different parsers for XML? Explain them in detail. | 8 |

SECTION II

Q. 7 A What are different data cleaning methods? 8

B Explain architecture of data warehouse with a neat diagram. 6

C A data warehouse can be modeled by either a star schema or a snowflake schema. Briefly describe the similarities and the differences of the two models, and then analyze their advantages and disadvantages with regard to one another. Give your opinion of which might be more empirically useful and state the reasons behind your answer. 4

OR

Q. 8 A Explain indexing OLAP data with example 6

B Explain the following operation on the multidimensional data with example. 6

i) Roll up and drill down.   ii) Slicing & dicing

C Explain three different data warehouse models. 6

Q. 9 A Consider following training set. 8

Class Label	A	B	C
C1	S	Y	X
C1	B	Y	X
C1	B	R	X
C1	S	R	X
C2	S	B	X
C2	B	B	Z
C2	B	Y	Z
C2	B	B	X
C2	S	Y	Z

Construct decision tree based on above training set using ID3.

B Explain K mean algorithm with example. Also state it weakness 8

OR

Q. 10   A   A database has 5 transactions. Let $min\ sup = 0.6$ and   8

$min\ conf = 0.8.$

Customer	Date	Items bought
100	10/15	{I,P,A,D,B,C}
200	10/15	{D,A,E,F}
300	10/16	{C,D,B,E}
400	10/18	{B,A,C,K,D}
500	10/19	{A,G,T,C}

    i)   List the frequent k-itemset for the largest k,
    ii)   List all the strong association rules (with support and confidence)

    B   Explain in detail classification and prediction. What is the   8
difference between them.

Q. 11   A   What do you mean by relevance ranking? Explain TF/IDF   8
methods of ranking.

    B   Explain the following: 8

    i)   Inverted Index
    ii)   Ontology
    iii)   Stop Words.
    iv)   Random walk

OR

Q. 12   A   What is page ranking and popularity ranking? Explain in brief.   8

    B   Explain the following terms   8

    i)   Web crawlers.
    ii)   Homonyms
    iii)   Vector space model
    iv)   Synonyms