

Total No. of Questions :12]

SEAT No. :

P860

[Total No. of Pages : 4

[4659]-240

B.E. (Computer Engineering)

d - ADVANCED DATABASES

(2008 Pattern) (Elective-III) (Semester-II)

Time : 3 Hours]

[Max. Marks : 100

Instructions to the candidates:

- 1) Answer three questions from section I and three question from sectionII.*
- 2) Answers to the two sections should be written in separate books.*
- 3) Neat diagrams must be drawn wherever necessary.*
- 4) 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

- Q1)** a) For each of the three partitioning techniques, namely round- robin, hash partitioning, and range partitioning, give an example of a query for which that partitioning technique would provide the fastest response. **[6]**
- b) Write a short note on parallel hash join. **[5]**
- c) What is interquery parallelism? Explain cache coherency problem and protocol available to guarantee cache coherency. **[6]**

OR

- Q2)** a) Explain parallel External sort- merge. **[6]**
- b) Explain design issues in parallel database system. **[6]**
- c) What factors 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]**

P.T.O.

- Q3)** 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 technique or protocol used to ensure this property in detail. [7]
- b) Explain how the following differ: Fragmentation transparency, replication transparency, and location transparency. [6]
- c) Describe and compare homogeneous and heterogeneous databases with respect to distributed databases. [4]

OR

- Q4)** a) Explain the following with respect to robustness of distributed databases.
 i) Co-ordinator selection.
 ii) Majority -based approach. [8]
- b) Explain network partition problem w.r.t to distributed databases. [5]
- c) What are the different approaches to store a relation in the distributed database. Explain them in brief. [4]
- Q5)** a) Write short notes on: [8]
 i) SOAP.
 ii) Client - Server architecture.
- b) Explain the structure of XML data with example. [8]

OR

- Q6)** a) Explain the following with respect to web architecture . [8]
 i) Web server.
 ii) Common gateway interface.
 iii) Cookie.
 iv) Uniform resource locator.
- b) Explain XML parsers in detail. [8]

SECTION -II

- Q7)** a) What is data warehouse? What is the difference between data warehouse and operational database system? [9]
- b) Explain the following: [8]
 i) Data cube
 ii) OLAP

OR

- Q8)** a) Discuss the different data smoothing techniques. [8]
 b) Suppose a group of 12 sales price records has been sorted as follows: [4]
 5, 10, 11, 13, 15, 35, 50, 55, 72, 92, 204, 215
 Partition them into three bins by each of the following methods:
 i) Equal -frequency (equidepth) partitioning
 ii) Equal - width partitioning
- c) Write a short note on data mart. [5]
- Q9)** a) Explain in detail classification and prediction. [8]
 b) Explain Apriori algorithm with example. [9]

OR

- Q10)**a) Consider following training set. [9]

Day	Out look	Temperature	Humidity	Wind	Play tennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

Write ID3 classification algorithm. Construct a decision tree based on above training set using ID3.

- b) What are Bayesian classifiers? [2]
 c) Explain K mean algorithm with example. [6]

- Q11)** a) What is relevance feedback? Explain in brief. [4]
b) Explain in detail information retrieval and structured data. [6]
c) Explain in detail popularity ranking. [6]

OR

- Q12)** a) What is the difference between a false positive and a false drop? If it is essential that no relevant information be missed by an information retrieval query, is it acceptable to have either false positives or false drops? Why? [4]
b) Explain in detail Web search Engines. [6]
c) Explain following terms with examples w.r.t to IR. [6]
i) Homonyms
ii) Synonyms
iii) Proximity

□□□