Reg No.



MANIPAL INSTITUTE OF TECHNOLOGY

## I SEMESTER M.TECH. (COMPUTER SCIENCE AND ENGINEERING / COMPUTER SCIENC AND INFORMATION SECURITY) END SEMESTER EXAMINATIONS, NOV 2017

SUBJECT: ADVANCED DATABASE SYSTEM [CSE 5102]

# REVISED CREDIT SYSTEM

## (18/11/2017)

#### TIME : 3 HOUR

### Instructions to the Candidates

- Answer ANY FIVE full Questions.
- Missing data can be suitably assumed
- 1A. Write a Java function using JDBC metadata features that takes a Result- Set as an input parameter, and prints out the result in tabular form, with appropriate names as column headings.
- 1B. Consider the following university placements schema. Write the structured query language queries for the following:

COMPANY (cname, clocation) STUDENT (srollno, sname, sdegree, slocation) INTERVIEW (cname, srollno, idate) OFFER (cname, srollno, osalary)

- i. Find the name of the students who have been interviewed but don't have any offer.
- ii. Find the name of all the students along with the name of the company if they have been offered a job by a company located in student's location.
- iii. Find the name of the companies that has interviewed every student in the college
- iv. Find the names of the student whose total offer salary is greater than 6 lakhs.
- 1C. Assume that only one tuple fits in a block and memory holds at most 3 blocks. Show the runs created on each pass of the sort-merge algorithm, when applied to sort the tuples on the first attribute: (kangaroo, 17), (wallaby, 21), (emu, 1), (wombat, 13), (platypus, 3), (lion, 8), (warthog, 4), (zebra, 11), (meerkat, 6), (hyena, 9), (hornbill, 2), (baboon, 12).
- 2A. Consider the following Schema of the University database. Write the query in relational algebra and evaluation plan for the following. Write a fully specified evaluation plan on an implementation for each of the algebra operations (i. and ii. ) involved.

classroom(<u>building</u>, <u>room number</u>, capacity) department(<u>dept name</u>, <u>building</u>, budget) course(<u>course id</u>, title, <u>dept name</u>, credits) instructor(<u>ID</u>, name, <u>dept name</u>, salary) section(<u>course id</u>, <u>sec id</u>, semester, year, building, room number, <u>time slot id</u>) teaches(<u>ID</u>, <u>course id</u>, <u>sec id</u>, semester, year) student(<u>ID</u>, name, <u>dept name</u>, tot cred) takes(<u>ID</u>, <u>course id</u>, <u>sec id</u>, semester, year, grade) advisor(<u>s ID</u>, <u>i ID</u>) time slot(time slot id, day, start time, end time) 3M

4M

MAX.MARKS: 50

prereq(course id, prereq id)

i. Find the names of all instructors in the Maths department, along with the titles of the courses that they teach.

ii. Find the names of all instructors in the Music department who have taught a course in 2009, along with the titles of the courses that they taught.3M

2B. Consider a database schema with a relation *Emp* whose attributes are as shown below, with types specified for multivalued attributes.

Emp = (ename, ChildrenSet multiset(Children), SkillSet multiset(Skills))
Children = (name, birthday)
Skills = (type, ExamSet setof(Exams))
Exams = (year, city)

Using the above schema, write the following queries in SQL.

- i. Find the names of all employees who have a child born on or after January 1, 2000.
- ii. Find those employees who took an examination for the skill type "typing" in the city "Dayton".

iii. List all skill types in the relation Emp.

2C. Consider the following relational schema.

employee (<u>person name</u>, street, city) works (<u>person name</u>, company name, salary) company (<u>company name</u>, city) manages (<u>person name</u>, manager name)

Give a schema definition in SQL corresponding to the relational schema, but using references to express foreign-key relationships for the following. i. employee ii. works iii. company iv. manages 4M

- 3A. Discuss the handling of skew with respect to the following.
  - i. Range partitioning
  - ii. Using Histogram

iii. Using virtual processor partitioning

- 3B. Explain the difference between data replication in a distributed system and the maintenance of a remote backup site. 3M
- 3C. "A data warehouse is a subject-oriented, integrated, time-variant, and nonvolatile collection of data in support of management's decision-making process." Justify your answer.
  4M
- 4A. Consider the sales multidimensional model represented by the star schema:



- i. Define the star schema in DMQL
- ii. Starting with the base cuboid [location, branch, time, item], what specific OLAP operations

3M

3M

should be performed

- a. to get the total amount collected by each branch in 2014?
- b. to get the no. of units sold with Panasonic brand under Kitchen\_Items type in each 4M location.
- 4B. Give the XML Schema for an XML representation of the following nested-relational schema:

Emp = (ename, ChildrenSet setof(Children), SkillsSet setof(Skills)) Children = (name, Birthday) Birthday = (day, month, year) Skills = (type, ExamsSet setof(Exams)) Exams = (year, city)

- 4C. How matrix multiplication can be done using natural join followed by grouping and aggregation? Give MapReduce implementation for the same. 3M
- 5A. What are Aggregates? Consider the following Data model oriented around a relational database (Fig. 1) convert to an aggregate data model.



3M

3M

- 5B. What is Document databases? With neat, explain Replica set configuration with higher 4M priority assigned to nodes in the same datacenter using MongoDB.
- 5C. What Is a Column-Family Data Store? Explain Cassandra's data model with column 3M families.