Reg. No.										

MANIPAL INSTITUTE OF TECHNOLOGY

MANIPAL

(A constituent unit of MAHE, Manipal)

V SEMESTER B.TECH. COMPUTER SCIENCE AND ENGINEERING

END SEMESTER EXAMINATIONS, NOV 2019

SUBJECT: BUSINESS INTELLIGENCE AND ITS APPLICATIONS (CSE 4024)

REVISED CREDIT SYSTEM
(22-11-2019)

Time: 3 Hours MAX. MARKS: 50

Instructions to Candidates:

- Answer **ALL** the questions.
- Missing data may be suitably assumed.

1A. How to extract information from stored unstructured data and semi-structured data? Explain. **4M**

1B. Briefly explain the comparison of features of OLTP and OLAP in-terms of: **3M**

a) Sources of data
b) Data Contents
c) Space requirements
d) Backup and Recovery
e) Operations, Indexes and Joins
f) Derived data and aggregates

1C. Consider a data warehouse that has three dimensions say A1, A2, A3 wherein each dimension has six attributes. If I start with the cube [A1, A2, A3]. **3M**

i) What would be the number of inner cuboids for any face of the cube?
ii) How many cuboids will be there in total in the cube?
iii) What would be the number of outer cuboids for any face of the cube?

2A. What is Howard Dresner definition of Business Intelligence? List and explain the features of Business Intelligence? **3M**

2B. Draw a schematic representation of BI component framework. In which layer business values and meta data management falls into. Explain. **4M**

2C. What are the best practices adapted in BI? Explain. Which disciplines in BI complement following statements? **3M**

a) To identify, understand and discern data
b) To gain insight and foresight
c) To consolidate data from disparate data sources

3A. What is William H Inmon definition of data warehouse? Explain the common approaches of Data integration. **3M**

3B. Explain performance measurement. Describe the characteristics of a business metrics. **3M**

3C. What are dashboards? How to create a dashboards. Give the differences between scorecards and dashboards. **4M**

4A. Draw the ER diagram for the following: You set up a database company, Art-Base that builds a product for art galleries. The core of this product is a database with a schema that captures all the information that galleries need to maintain. **3M**

Galleries keep information about artists, their names (which are unique), birthplaces, age, and style of art. For each piece of artwork, the artist, the year it was made, its unique title, its type of art (e.g., painting, lithograph, sculpture, photograph), and its price must be stored. Pieces of art work are also classified into groups of various kinds, for example, portraits, still life's, works by Picasso, or works of the 19th century; a given piece may belong to more than one group. Each group is identified by a unique name (like those just given) that describes the group. Finally, galleries keep information about customers. For each customer, galleries keep that person's unique name, address, total amount of dollars spent in the gallery (very important!), and the artists and groups of art that the customer tends to like.

4B. The Company you work for wants to digitize their time cards. You have been asked to design the database for submitting and approving time cards. **5M**

A timecard should have hours worked and date submitted. Each timecard is associated with exactly one employee. Each timecard should have a unique id. Each timecard has a status: it is either approved, not approved, or pending. Each employee has a unique id. Each employee has a name and address. Each employee submits a time card every pay period, i.e., In one year, they will submit multiple time cards. Each employee either has direct deposit or physical check as their method of payment. Each employee is associated with exactly one manager. Each manager has a unique id and a name. Each manager is in charge of multiple employees. Each manager approves time cards for multiple employees.

i. Draw the conceptual data model for the above problem.
ii. Draw the logical data model for the above problem.
iii. Draw the physical data model for the above problem.
iv. Convert the logical data model to star schema.
v. Identify the RCD in above problem and tackle it in star schema

4C. Discuss all the different types of containers used in control flow of SSIS. **2M**

5A. Suppose that a data warehouse for Big University consists of the following four dimensions: student, course, semester, and instructor, and two measures count and avg_grade. When at the lowest conceptual level (e.g., for a given student, course, semester, and instructor combination), the avg_grade measure stores the actual course grade of the student. At higher conceptual levels, avg_grade stores the average grade for the given combination. **5M**

i. Draw a snowflake schema diagram for the data warehouse.

ii. Starting with the base cuboid [student, course, semester, instructor], what specific OLAP operations should one perform in order to list the average grade of CS courses for each Big University student.

iii. If each dimension has five levels (including all), such as "student<major<status<university<all", how many cuboids will this cube contain (including the base and apex cuboids)?

iv. Identify SCD and explain how to tackle them.

v. Identify the different types of fact in snowflake schema.

5B. What are the long term benefits when enterprise reporting is built on culture of fact based decision making? Write a balanced scorecard for improving the performance of craft brewery. Show the tabular representation of objectives, measures, targets and initiative for all the perspectives. **5M**