Birla Institute of Technology & Science, Pilani
Work-Integrated Learning Programmes Division
First Semester 2018-2019
Comprehensive Examination (EC-3 Regular)

Course No. : CSI ZG515
Course Title : DATA WAREHOUSING
Nature of Exam : Open Book
Weightage : 50%
Duration : 3 Hours
Date of Exam : 25/11/2018 (AN)

No. of Pages = 1
No. of Questions = 7

1. Please follow all the *Instructions to Candidates* given on the cover page of the answer book.
2. Assumptions made if any, should be stated clearly at the beginning of your answer.

Q.1. For following statements, indicate True or False with proper justification:
(a) It is a good practice to drop the indexes before the initial load.
(b) The choice of index type depends on cardinality.
(c) The importance of metadata is the same for data warehouse and an operational system.
(d) Backing up the data warehouse is not necessary because you can recover data from the source systems.
(e) Sorts and merges of files are common in staging area. [5]

Q.2. Consider a data warehouse, where the fact data is calculated to be 48GB of data per year, and 5 years' worth of data are to be kept online. The data is to be partitioned by month and four concurrent queries are to be allowed. Compute the partition size, Temporary Space and Space Required for this scenario. [5]

Q.3. In a real-time data warehouse, suggest a way of handling changes to dimensions in real time. [5]

Q.4. Compare & contrast outriggers & mini-dimensions. Give situations under which you would prefer an outrigger to a mini-dimension & vice versa with example. [5]

Q.5 (a) Why do we need a separate Data Warehouse? What data is stored in a warehouse? Discuss with example how do we represent this data? [8]

Q.5 (b) What is slicing and dicing? Explain with real time usage and business reasons of it's use. [2]

Q.6 (a) What is meant by the selectivity for a column in a physical table? What type of indexing technique is suitable for low-selectivity data? Explain with example. [3]

Q.6 (b) Why dimension tables are wide and the fact table is deep? Explain with example. [3]

Q.6 (c) Consider the following relation Cars: [4]

Brand	Type	Color	Risk
Maruti	Swift	Grey	Low
Maruti	Swift	Red	Medium
Mahindra	XUV	Black	Medium
Renault	Pulse	Black	Low

(i) Construct a bitmap index for the attributes Brand and Color for this table.
(ii) Indicate how these two bitmap-indices can be used to answer the query: Give the total number of red Maruti cars with a medium risk score.

Q.7. Suppose that a data warehouse for Big-University consists of the following four dimensions: student, course, semester, and instructor, and a measure avg_grade. At the lowest conceptual level the avg_grade stores the actual course grade of the student in a semester for a course. At higher conceptual levels, avg_grade stores the average grade for the given combination.
(a) Draw a snowflake schema diagram for the data warehouse.
(b) Starting with the base cuboid [student, course, semester, instructor], what specific OLAP operations are needed to get the average grade of CS students in 2018. [6 + 4 = 10]