Course Code : CDT 305-2 **KQLR/RS – 24 / 1592**

Fifth Semester B. Tech. (Computer Science and Engineering / Data Science) Examination

DATA WAREHOUSING AND BUSINESS INTELLIGENCE

Time : 3 Hours] [Max. Marks : 60

Instructions to Candidates :—

(1) Assume suitable data wherever necessary.

(2) All questions carry marks as indicated.

(3) Use graph paper to plot the graphs.

1. (a) Assume you are the owner of an automobile store. What are the key questions that BI will help you answer ? 4(CO2)

(b) What is metadata and what are different categories of metadata ? 4(CO2)

(c) Differentiate between Enterprise Resource Planning and Business Intelligence. 2(CO2)

2. (a) Mr. Smith manages a small product distribution company. Because the business is growing fast, Mr. Smith recognizes that it is time to manage the vast information pool to help guide the accelerating growth. Mr. Smith, who is familiar with spreadsheet software, currently employs a small sales force of four people. He asks you to develop a data warehouse application prototype that will enable him to study sales figure by year, region, salesperson and product.

(1) Identify the appropriate fact table components.

(2) Identify the appropriate dimension tables.

(3) Draw a star schema diagram for this data warehouse.

(4) Identify the attributes for dimension tables. 6(CO1)

(b) Explain the following types of Dimension tables with suitable examples :

(a) SCD (Slowly Changing Dimension).

(b) RCD (Rapidly Changing Dimension).

(c) Role Playing Dimension. 4(CO1)

3. (a) Elaborate the following approaches for designing a data warehouse :

(a) Ralph Kimball's Approach.

(b) Inmon's Approach. 4(CO1)

(b) What is data profiling ? List at least four different types of analysis which are conducted as a part of data profiling. 4(CO1)

(c) Justify the importance of data staging area in data warehouse architecture. 2(CO1)

4. (a) In a hypothetical sample of 20 people the amounts of money with them were found to be as follows :

114, 108, 100, 98, 101, 109, 117, 119, 126, 131, 136, 143, 156, 168, 182, 195, 207, 219, 235, 118.

Draw the histograms of the frequency distribution using :

(i) Equi-depth binning.

(ii) Equi-width binning.

Class Interval	Frequency
50 – 100	2
100 – 150	11
150 – 200	4
200 – 250	3

4(CO3)

(b) Suppose that the data for analysis includes the attribute age. The age values for the data tuples are (in increasing order) 13, 15, 16, 16, 19, 20, 20, 21, 22, 22, 25, 25, 25, 25, 30, 33, 33, 35, 35, 35, 35, 36, 40, 45, 46, 52, 70.

(a) What is the mean of the data ? What is the median ?

(b) What is the midrange of the data ?

(c) Can you find the first quartile (Q1) and the third quartile (Q3) of the data ?

(d) Give the five-number summary of the data.

(e) Show a boxplot of the data.

(f) How is a quantile-quantile plot different from a quantile plot ? 6(CO3)

5. (a) List and elaborate the critical focus area of enterprise reporting characteristics in OLAP world. 4(CO1)

(b) "The primary differences between the two is that dashboards monitor the performance of operational processes whereas Scorecards chart the progress of tactical and Strategic goals." Explain giving an example. 4(CO2)

(c) Define Key Performance Indicator (KPI). Why is it required ? 2(CO2)

6. (a) What are the main Hadoop Components ? What function do they perform ? 5(CO4)

(b) What is NoSQL ? How does it fit into the Big Data analytics picture ? 5(CO4)