01/06/17 FN

Reg. No. : | | | | | | | | | | | | |

Question Paper Code : 72059

B.E./B.Tech. DEGREE EXAMINATION, APRIL/MAY 2017.

Sixth/Seventh Semester

Information Technology

IT 6702 — DATA WAREHOUSING AND DATA MINING

(Common to Computer Science and Engineering)

(Regulations 2013)

Time : Three hours Maximum : 100 marks

Answer ALL questions.

PART A — (10 × 2 = 20 marks)

1. How is a data warehouse different from database? How they are similar?
2. What is data discretization?
3. List the distinct features of OLTP with OLAP.
4. What is multidimensional data model? Give example.
5. Why we need data transformation? Mention the ways by which data can be transformed.
6. List the five primitives for specification of a data mining task.
7. How do you evaluate accuracy of a classifier?
8. What is lazy learner? Give an example.
9. What is meant by K-Nearest Neighbor Algorithm?
10. List the some applications of data mining.

PART B — (5 × 16 = 80 marks)

11. (a) Explain mapping data warehouse with multi processor architecture with the concept of parallelism and data partitioning. (16)

Or

(b) Discuss Data Extraction, Clean up and transformation tools with meta data management. (16)

12. (a) Explain different categories of OLAP tools with diagram. (16)

Or

(b) (i) Summarize Multi dimensional data model. (8)

(ii) Discuss about Cognous Impromptu. (8)

13. (a) Why do we need to preprocess data? What are the different forms of preprocessing? (16)

Or

(b) Describe in detail data mining functionalities and the different kinds of patterns can be mined. (16)

14. (a) Discuss the single dimensional Boolean association rule mining for transaction database. (16)

Or

(b) Discuss about constraint based association rule mining with examples and state how association mining to correlation analysis is dealt with. (16)

15. (a) Describe different types of data in Cluster Analysis. (16)

Or

(b) Describe different Hierarchical methods in cluster analysis. (16)

Reg. No. :

Question Paper Code : 41301

27/04/18 AN

B.E./B.Tech. DEGREE EXAMINATION, APRIL/MAY 2018
Seventh Semester
Information Technology
IT6702 – DATA WAREHOUSING AND DATA MINING
(Common to : Computer Science and Engineering)
(Regulations 2013)

Time : Three Hours | Maximum : 100 Marks

Answer ALL questions

PART – A **(10×2=20 Marks)**

1. Define data warehouse and data mart.
2. What is the use of metadata in data warehouse ?
3. Differentiate OLTP and OLAP systems.
4. List out the three distinct types of reporting.
5. What are the major tasks done in data pre-processing ?
6. Consider that the minimum and maximum values for the attribute "salary" are 12,000 and 98,000 respectively and the mapping range of salary is [0.0, 1.0]. Find the transformation for the salary 73,600 using Min-Max Normalization.
7. How does a Support Vector Machine (SVM) find a hyperplane ?
8. What are eager learners and lazy learners ?
9. Define outliers. How will you determine outliers in the data ?
10. List out the difference between characterization and clustering.

PART – B **(5×16=80 Marks)**

11. a) i) Explain the 3-tier data warehouse architecture and its various components. **(12)**

ii) Give a short description about shared-memory and shared-disk architecture. **(4)**

(OR)

b) Design a star-schema, snow-flake schema and fact-constellation schema for the data warehouse that consists of the following four dimensions : (Time, Item, Branch and Location). Include the appropriate measures required for the schemas. **(16)**

12. a) Explain the characteristics of OLAP cube and various OLAP operations with an example. **(16)**

(OR)

b) i) Briefly Outline the special reporting options provided by Cognos Impromptu for improving the value of distributed standard reports. **(8)**

ii) Give a short description on the guidelines for selecting OLAP systems. **(8)**

13. a) i) Describe the key functionalities of a data mining system and give examples of each datamining functionality using a real-life database that you are familiar with. **(12)**

ii) How will you handle missing value in a dataset before mining process ? **(4)**

(OR)

b) i) Describe the differences between the following approaches for the integration of a data mining system with a database or data warehouse system : no coupling, loose coupling, semi-tight coupling, and tight coupling. State which approach you think is the most popular and outline the reason. **(12)**

ii) Consider the following data for the attribute AGE : 4, 8, 21, 5, 21, 24, 34, 28, 25. Perform smoothing by bin means and bin boundaries using a bin depth of 3. **(4)**

14. a) i) A database has following four transactions as listed below in the Table 14.i. Let minimum_support = 50% and minimum_confidence = 80%. Find the frequent Item-set using Apriori algorithm. **(6)**

Table 14.i

TID	Items_Purchased
T101	A, C, D
T102	B, C, E
T103	A, B, C, E
T104	B, E

ii) Describe the process of generating association rules from the identified frequent itemsets. **(3)**

iii) Elucidate the steps in the decision tree classification process with an example. **(7)**

(OR)

b) i) Explain the process of multi-layer feed-forward neural network classification using back propagation learning. **(10)**

ii) Why naive Bayesian classification is called naive ? Briefly outline the major ideas of naive Bayesian classification. **(6)**

15. a) i) Explain k-means algorithm with an example. Describe the pros and cons of k-means in comparison with the k-medoids algorithm . **(10)**

ii) Consider that the data mining task is to cluster the following eight points A1, A2, A3, B1, B2, B3, C1 and C2 (with (x; y) representing location) into three clusters. A1(2; 10); A2(2;5); A3(8;4); B1(5;8); B2(7;5); B3(6;4); C1(1;2); C2(4;9):

The distance function is Euclidean distance. Suppose initially we assign A1, B1 and C1 as the center of each cluster, respectively. Use the k-means algorithm to show the three cluster centers after the first round of execution and the final three clusters. **(6)**

(OR)

b) i) Briefly outline the key features of various clustering methods with an example. **(12)**

ii) Describe the various trends in Data Mining. **(4)**

———————

11/04/19

Reg. No. : | | | | | | | | | | | | |

Question Paper Code : 53241

B.E./B.Tech. DEGREE EXAMINATIONS, APRIL/MAY 2019.

Sixth/Seventh Semester

Information Technology

IT 6702 — DATA WAREHOUSING AND DATA MINING

(Common to Computer Science and Engineering)

(Regulation 2013)

(Also Common to PTIT 6702 – Data Warehousing and Data Mining for B.E. (Part-Time) – Computer Science and Engineering – Regulation 2014)

Time : Three hours | Maximum : 100 marks

Answer ALL questions.

PART A — (10 × 2 = 20 marks)

1. Outline the need for data cleaning.
2. What is data transformation ?
3. Outline the need for a multidimensional data model.
4. What is Multirelational OLAP?
5. Define Data mining.
6. What is Pattern evaluation?
7. Why tree pruning is useful in decision tree induction?
8. What is association based classification?
9. Present an outline of hierarchical clustering.
10. What is an outlier?

PART B — (5 × 13 = 65 marks)

11. (a) Outline the characteristics of a data warehouse and write a note on Meta data. (13)

Or

(b) Explain the ETL process in data warehousing and outline its need. (13)

12. (a) What is OLAP? Outline the OLAP operations with an example.

Or

(b) Explain the following with an example :

(i) Star Schema (4)

(ii) Snowflake Schema (5)

(iii) Fact constellation Schema. (4)

13. (a) Present an outline of the following :

(i) Rough Set approach (6)

(ii) Fuzzy Set approaches. (7)

Or

(b) Outline the schemes for integration of a data mining system with a data warehouse. (13)

14. (a) Outline the Naive Bayes algorithm with an example. (13)

Or

(b) What is support vector machine? Outline the working principle of support vector machine with an example.

15. (a) (i) Outline the requirement of Clustering in Data Mining. (7)

(ii) Specify some important application of Cluster Analysis. (6)

Or

(b) Outline the steps in K Means clustering algorithm with an example. (13)

PART C — (1 × 15 = 15 marks)

16. (a) Apply the Apriori algorithm for mining frequent itemsets from the following dataset :

TID	Items Purchased
1	BREAD, BUTTER, MILK
2	TEA, CAKE
3	BREAD, BUTTER, MILK, CAKE
4	BREAD, BUTTER
5	BUTTER, MILK
6	BREAD, BUTTER, MILK
7	TEA, CAKE
8	BREAD, BUTTER, MILK
9	JUICE, CAKE
10	BREAD, MILK

Use 0.3 (30%) for the value of minimum support.

Or

(b) Outline the steps in classification by back propagation with an example. (15)

Reg. No. : | | | | | | | | | | | | |

Question Paper Code : 50769

B.E./B.Tech. DEGREE EXAMINATION, NOVEMBER/DECEMBER 2017
Seventh Semester
Information Technology
IT 6701 – INFORMATION MANAGEMENT
(Regulations 2013)

Time : Three Hours | Maximum : 100 Marks

Answer ALL questions

PART – A **(10×2=20 Marks)**

1. What is Hadoop HDFS ?
2. Define MapReduce.
3. Define Program Security.
4. Write short notes on Firewalls.
5. What is Data Synchronization ?
6. How to Manage Data Quality ?
7. Explain briefly Conceptual Design.
8. Write short notes on Granularity of Contents.
9. What is Sensitive data ?
10. Define Big Data.

PART – B **(16×5=80 Marks)**

11. A) Elaborate on Database Design and Modelling.

(OR)

B) Discuss Trends in Big Data Systems in detail.

12. A) Explain in detail Network Security.

(OR)

B) Discuss the concepts of Data Privacy in detail.

13. A) Elaborate on Master Data Management (MDM) with suitable examples.

(OR)

B) Explain the concepts of Data Governance in detail.

14. A) Explain Navigation Systems and Labelling Systems in detail.

(OR)

B) Discuss about Principles of Information Architecture and Framework in detail

15. A) List out and explain the major functions of Data Administration in detail.

(OR)

B) Elaborate on Data retention policies.

Reg. No. :

Question Paper Code : 91775

B.E./B.Tech. DEGREE EXAMINATIONS, NOVEMBER/DECEMBER 2019

Seventh Semester

Information Technology

IT 6701 – INFORMATION MANAGEMENT

(Regulations 2013)

Time : Three Hours | Maximum : 100 Marks

Answer ALL questions

PART – A **(10×2=20 Marks)**

1. Bringout various applications of map reduce.
2. What is stored procedure ? Give an example.
3. List the various types of intrusion detection system.
4. List various levels of operating system protection.
5. Why synchronization is needed for data governance ?
6. What are the key benefits and limitations of MDM ?
7. Write notes on granularity of contents.
8. What are the responsibilities of an information architect ?
9. How do you estimate life cycle management costs ? Give an example.
10. What are the challenges in data administration ?

PART – B **(5×13=65 Marks)**

11. a) What is data modelling ? Discuss elaborately about conceptual, logical and physical data models with a neat block diagram. **(13)**

(OR)

b) What is Hadoop and why it is used ? Describe Hadoop Distributed Fire System (HDFS) architecture with a neat block diagram. **(13)**

12. a) Discuss elaborately about the various types of firewalls with necessary diagrams. **(13)**

(OR)

b) Explain in detail about the various levels of operating system protections. **(13)**

13. a) Elaborate in detail the architectural dimensions of MDM. **(13)**

(OR)

b) Elaborate data governance framework in detail. **(13)**

14. a) Explain in detail :

i) Navigation system. **(7)**

ii) Labelling system. **(6)**

(OR)

b) i) Discuss about principles of information architecture and framework in detail. **(10)**

ii) List different organization schemes used in an organization system. **(3)**

15. a) Explain the purpose, scope and policies of data retention with an example. **(13)**

(OR)

b) Detail different testing methods and delivering methods in big data application. **(13)**

PART – C **(1×15=15 Marks)**

16. a) For a banking application, discuss how various types of noSQL databases are involved to support the banking system. **(15)**

(OR)

b) Create a dataset for rate of petrol consumption using HiveQL. Explain the following SELECT statement syntax for the rate of petrol consumption dataset created.

a) Computing with columns **(4)**

b) WHERE clauses **(4)**

c) HAVING clauses **(4)**

d) GROUP BY clauses. **(3)**