MCS-221

MASTER OF COMPUTER APPLICATIONS (NEW) (MCA-NEW)

Term-End Examination

December, 2021

MCS-221 : DATA WAREHOUSING AND DATA MINING

Time : 3 hours *Maximum Marks : 100*

(Weightage : 70%)

Note : *Question no.* ***1*** *is* ***compulsory****. Answer any* ***three*** *questions from the rest.*

1. (a) Define a Data Warehouse. List and explain the four characteristics of a Data Warehouse. *10*

(b) Explain the following OLAP architectures and draw their architectural diagram : *10*

(i) Multidimensional Online Analytical Processing (MOLAP)

(ii) Hybrid Online Analytical Processing (HOLAP)

IGNOUAssignmentGuru.com

(c) Define "data cleaning" which is a data preprocessing technique. In this context, explain the concept of Noisy data cleaning along with some suitable examples. *10*

(d) What is a Decision Tree ? How is it useful in classification ? With the help of an example, explain the process of construction of a decision tree and its representation. *10*

2. (a) What is Text Mining ? Where is it used ? Explain any two text mining techniques with the help of a suitable example for each. *10*

(b) In the context of mining multimedia data on the web, explain the following terms : *10*

(i) Page Rank

(ii) Hits

(iii) Page Layout Analysis

(iv) Vision Page Segmentation

3. (a) Define a Data Lake. Explain the step-by-step process of creating a data lake. *10*

(b) With the help of an example, explain the star schema dimensional model. *10*

IGNOUAssignmentGuru.com

4. (a) What is Metadata ? What are its contents ? Justify how metadata can be an important component in Data Warehousing ? Also mention its types. *10*

(b) Explain the three components of ETL. Also mention how to improve the ETL performance. *10*

5. Write short notes on any ***four*** of the following : *4×5=20*

(a) Data Marts

(b) Association Rule Generation

(c) Apriori Algorithm

(d) Clustering and its Methods

(e) Bayes' Theorem

IGNOUAssignmentGuru.com

No. of Printed Pages : 3

MCS-221

MASTER OF COMPUTER APPLICATION (MCA) (NEW)

Term-End Examination

June, 2022

MCS-221 : DATA WAREHOUSING AND DATA MINING

Time : 3 Hours *Maximum Marks : 100*

Weightage : 70%

Note : (*i*) *Question No.* ***1*** *is compulsory.*

(*ii*) *Attempt any* ***three*** *questions from the rest.*

---

1. (a) What is Association Rule Mining ? With reference to this, explain the concepts given below : 10

   (i) Support Count

IGNOUAssignmentGuru.com

(ii) Frequent Item set

(iii) Association Rule

(iv) Rule Evaluation metrics

(b) Define classification. With the help of an example for each, explain the following classification models : 10

(i) Descriptive modeling

(ii) Predictive modeling

(c) Describe the snowflake schema multidimensional modeling technique. Give an example and illustrate it. List its pros and cons. 10

(d) What is complex data modeling ? Where is this used ? Explain "Anchor" complex data model. 10

2. (a) List and describe all the four types of OLAP operations. 10

(b) Define text mining. Mention any *three* applications of it. What are the various techniques used to analyze the web-usage patterns ? 10

IGNOUAssignmentGuru.com

3. (a) Define a data warehouse. List and explain all the 3 types of data warehouses. 10

   (b) What is data transformation ? List and explain all the data transformation steps. 10

4. (a) What is data mining ? List out the key differences between data warehousing and data mining. Mention any *four* applications of data mining. 10

   (b) What is ETL ? Why do you need ETL in data warehouses ? Also, explain how ETL works in Data Warehousing ? 10

5. Write short notes on any ***four*** of the following : 4×5=20

   (a) OLTP

   (b) Data Marts

   (c) Web structure mining

   (d) Cloud Data Warehousing

   (e) Data Granularity

IGNOUAssignmentGuru.com

MCS-221

MASTER OF COMPUTER APPLICATIONS (NEW) (MCA-NEW)

Term-End Examination

December, 2022

MCS-221 : DATA WAREHOUSING AND DATA MINING

Time : 3 hours *Maximum Marks : 100*

(Weightage : 70%)

Note : *Question no.* ***1*** *is* ***compulsory****. Answer any* ***three*** *questions from the rest.*

1. (a) Define Dimensional Modelling. With reference to this, define the terms Facts, Fact Table, Dimensions and Dimensional Table. Give an example use-case and derive fact table and dimension tables. *10*

(b) With the help of a Data Warehouse Architecture diagram, explain the following components and their significance : *10*

(i) ETL

(ii) Metadata

(iii) Data Warehouse Access Tools

(iv) Data Warehouse Reporting Layer

IGNOUAssignmentGuru.com

(c) Define a Decision Tree. With the help of an example, explain the construction and representation of decision tree. Also, mention its strengths and weaknesses. *10*

(d) Discuss the following categories of Data Mining Issues : *10*

(i) Mining Methodology and User Iteration Issues

(ii) Performance-based Issues

(iii) Diverse Data Types Issues

2. (a) Define Data Cleaning. Explain the ways and means of handling the Missing Values and Noisy Data while data preprocessing. *10*

(b) Write and explain the K-means algorithm for clustering. How does it work ? *10*

3. (a) Why does dimensionality reduction of text need to be done ? Explain Tokenization process and Vector from text approach, with the help of a suitable example for each. *10*

(b) Enumerate the best practices for Data Warehouse Architecture. *5*

(c) Describe the following types of data marts : *5*

(i) Dependent data marts

(ii) Independent data marts

IGNOUAssignmentGuru.com

4. (a) With the help of an example use-case, explain the Snowflake schema. List its advantages and disadvantages. *10*

(b) Define Web Mining. What kind of tasks can be performed using it ? Discuss its features and also a few applications. *10*

5. Write short notes on the following : *4×5=20*

(a) Data Transformation (with reference to data preprocessing)

(b) Cloud Data Warehousing

(c) Data Lake and its Architecture

(d) Data Warehouse Automation

IGNOUAssignmentGuru.com

No. of Printed Pages : 3

MCS-221

MASTER OF COMPUTER APPLICATIONS (MCA-NEW)

Term-End Examination

June, 2023

MCS-221 : DATA WAREHOUSING AND DATA MINING

Time : 3 Hours *Maximum Marks : 100*

Weightage : 70%

Note : *(i) Question No. 1 is compulsory.*

(ii) Answer any ***three*** *questions from the rest.*

---

1. (a) Discuss the following data preprocessing stages briefly : 10

(i) Data Cleaning

(ii) Data Integration

(iii) Data Reduction

(iv) Data Transformation

P. T. O.

IGNOUAssignmentGuru.com

(b) Define Aggregate fact tables and derived dimension tables. What are their significance ? Give an example. Also mention their advantages and disadvantages. 10

(c) Enumerate the key challenges in data warehouse design. 5

(d) Differentiate between a data lake and a data warehouse. 5

(e) Write and explain the Apriori algorithm to identify the most frequently occurring elements and meaningful associations in a dataset. 10

2. (a) What is Cluster Analysis ? How is this used in Data Mining ? Give an example. Also mention few applications of cluster analysis in data mining. 10

(b) List and discuss various types of Web-mining. 10

3. (a) Explain the following techniques for Dimensionality Reduction : 10

(i) Feature Selection

(ii) Feature Extraction

(b) Discuss the layered implementation of ETL in a Data Warehouse. 10

IGNOUAssignmentGuru.com

4. (a) Define OLAP. Differentiate between Multicube and Hypercube. Mention the applications of OLAP reporting system. 10

   (b) List and explain the following types of Data Warehouses : 10

   (i) Enterprise Data Warehouse (EDW)

   (ii) Operational Data Store

5. Write short notes on the following : 4×5=20

   (a) ELT *vs.* ETL

   (b) Data Marts

   (c) Applictions of Data Mining

   (d) OLAP data cube operations

IGNOUAssignmentGuru.com