FirstRanker.com
Firstranker's choice
www.FirstRanker.com www.FirstRanker.com

Code No: RT32052 **R13** SET - 1

III B. Tech II Semester Regular Examinations, April - 2017
DATA WARE HOUSING AND MINING
(Common to Computer Science Engineering and Information Technology)

Time: 3 hours Max. Marks: 70

Note: 1. Question Paper consists of two parts (**Part-A** and **Part-B**)
2. Answering the question in **Part-A** is compulsory
3. Answer any **THREE** Questions from **Part-B**

*****

PART –A

1	a)	Mention some application areas that use data warehousing technologies.	[3M]
	b)	What do you understand by the statement: "The data stored in a data warehouse is Non-volatile in nature"?	[4M]
	c)	What is the major difference between Star schema and Snowflake schema?	[3M]
	d)	What is meant by Data Discrimination?	[4M]
	e)	What is a Cluster in data mining?	[4M]
	f)	What is Concept Description?	[4M]

PART –B

2	a)	Define the terms OLTP and OLAP?	[3+3]
	b)	What are the major distinguishing features between OLTP and OLAP systems?	[10M]
3	a)	Explain the sequence of steps that are followed in Knowledge Discovery in Databases (KDD).	[8M]
	b)	Explain the various coupling schemes that can be used during the integration process of a Data mining System with a Data warehouse.	[8M]
4	a)	What are the different OLAP operations on multidimensional data?	[8M]
	b)	Define a measure. What are the different categories of measures?	[8M]
5	a)	Explain Hunt's algorithm for building a Decision Tree.	[10M]
	b)	Write the bisecting k-means algorithm with an example.	[6M]
6	a)	Define i) Maximum Frequent Item Sets ii) Closed Frequent Item Sets.	[3+3]
	b)	Explain the different attribute types that are used in attribute test condition in the Decision Tree.	[10M]
7	a)	In the context of utility, explain how Cluster Analysis helps in analyzing the data objects.	[4M]
	b)	Explain the different types of Clusterings.	[12M]

*****

www.FirstRanker.com

FirstRanker.com
Firstranker's choice
www.FirstRanker.com www.FirstRanker.com

Code No: RT32052 **R13** SET - 2

III B. Tech II Semester Regular Examinations, April - 2017
DATA WARE HOUSING AND MINING
(Common to Computer Science Engineering and Information Technology)

Time: 3 hours Max. Marks: 70

Note: 1. Question Paper consists of two parts (**Part-A** and **Part-B**)
2. Answering the question in **Part-A** is compulsory
3. Answer any **THREE** Questions from **Part-B**

*****

PART –A

1	a)	What is the main disadvantage of Snowflake schema?	[3M]
	b)	What is meant by Data Characterization?	[4M]
	c)	Mention some application areas that use data mining techniques.	[3M]
	d)	What is meant by "noise" in data pre-processing?	[4M]
	e)	What is Descriptive Mining?	[4M]
	f)	What is a Data Cube?	[4M]

PART –B

2	a)	Write in detail about the classification of Data Mining Systems?	[8M]
	b)	Explain the various data smoothing techniques that are used to handle noisy data.	[8M]
3		Explain the various methods that are used in Discretization and Concept Hierarchy Generation for numerical data.	[16M]
4	a)	We can get the background knowledge for a given set of data by using Concept Hierarchies. In this context, explain the four major sets of Concept Hierarchies.	[8M]
	b)	Explain the different schema models that can be built by using dimension and fact tables.	[8M]
5	a)	What are the issues of Decision Tree Induction?	[6M]
	b)	Explain the different attribute types that are used in attribute test condition in the Decision Tree.	[10M]
6	a)	State the Apriori Principle.	[4M]
	b)	Explain the important characteristics of decision tree induction algorithms.	[12M]
7	a)	Explain the DBSCAN algorithm in detail.	[10M]
	b)	Mention the time and space complexity of the DBSCAN algorithm.	[6M]

*****

www.FirstRanker.com

FirstRanker.com
Firstranker's choice
www.FirstRanker.com www.FirstRanker.com

Code No: RT32052 **R13** SET - 3

III B. Tech II Semester Regular Examinations, April - 2017
DATA WARE HOUSING AND MINING
(Common to Computer Science Engineering and Information Technology)

Time: 3 hours Max. Marks: 70

Note: 1. Question Paper consists of two parts (**Part-A** and **Part-B**)
2. Answering the question in **Part-A** is compulsory
3. Answer any **THREE** Questions from **Part-B**

*****

PART –A

1	a)	What is meant by the "Time-Variant" feature of a Data Warehouse?	[3M]
	b)	What is the main advantage of using Multidimensional OLAP (MOLAP) Servers?	[4M]
	c)	What is the difference between "Information Processing" and "Analytical Processing"?	[4M]
	d)	What is meant by Data Discrimination?	[4M]
	e)	Mention any two Density-Based methods used in Clustering Analysis.	[3M]
	f)	What is Predictive Mining?	[4M]

PART –B

2	a)	Explain the various pattern interestingness measures.	[8M]
	b)	List and Explain the different schemas that can be built using dimension tables and fact tables.	[8M]
3	a)	Discuss about various data warehouse models from the architecture point of view.	[8M]
	b)	What are the different types of OLAP servers give example of each.	[8M]
4	a)	Explain the various features of a Data warehouse.	[6M]
	b)	Explain the various data reduction techniques give advantages of each.	[10M]
5	a)	Explain the two strategies for avoiding model over fitting in the context of Decision Tree Induction.	[4+4]
	b)	Discuss the methods that are commonly used to evaluate the performance of a Classifier.	[8M]
6	a)	Give the formal definitions of the support and confidence metrics in Association Analysis.	[4+4]
	b)	Explain Over fitting due to Presence of Noise.	[4M]
	c)	Explain Over fitting due to Lack of Representative Samples.	[4M]
7	a)	Write the basic Agglomerative Hierarchical Clustering algorithm.	[6M]
	b)	Mention the time and space complexity for the basic Agglomerative Hierarchical Clustering algorithm.	[4M]
	c)	Discuss the key issues in Hierarchical Clustering.	[6M]

*****

www.FirstRanker.com

FirstRanker.com
Firstranker's choice
www.FirstRanker.com www.FirstRanker.com

Code No: RT32052 **R13** SET - 4

III B. Tech II Semester Regular Examinations, April - 2017
DATA WARE HOUSING AND MINING
(Common to Computer Science Engineering and Information Technology)

Time: 3 hours Max. Marks: 70

Note: 1. Question Paper consists of two parts (**Part-A** and **Part-B**)
2. Answering the question in **Part-A** is compulsory
3. Answer any **THREE** Questions from **Part-B**

*****

PART –A

1	a)	What are outliers in data mining?	[3M]
	b)	What is a Concept Hierarchy?	[4M]
	c)	What is meant by Data Characterization?	[4M]
	d)	Define a Datamart.	[4M]
	e)	What is meant by Pattern Evaluation in data mining?	[4M]
	f)	Give the definition of the term Classification in data mining.	[3M]

PART –B

2	a)	What is meant by i) Descriptive Mining ii) Predictive Mining. Compare them.	[3+3]
	b)	Explain the different OLAP operations on multidimensional data.	[10M]
3	a)	What is the need for Data Preprocessing? List various techniques.	[4M]
	b)	Explain the different techniques that are used to handle noisy data.	[8M]
	c)	Write notes on various performance issues that are encountered in Data Mining.	[4M]
4	a)	Explain the different types of OLAP Servers.	[8M]
	b)	Explain the major issues that are encountered in Data mining.	[8M]
5	a)	What is meant by i) Model Underfitting ii) Model Overfitting? Compare them	[6+6]
	b)	Explain the bootstrap approach in Classification process.	[4M]
6		Explain the FP-Tree Representation. Also, explain how the frequent item set is generated using FP- growth algorithm.	[8+8]
7	a)	In the context of understanding, explain how Clustering Analysis helps in analyzing and describing the data objects in real world.	[4M]
	b)	Write the basic k-means algorithm.	[4M]
	c)	Mention the time and space complexity for the basic k-means algorithm.	[4M]

*****

www.FirstRanker.com