b. With a neat diagram, describe the steps in the MapReduce parallel flow data model. (10 Mar OR a. Write the Java code for MAP and REDUCE of word count problem. Describe the steps compiling and removing the MapReduce program. (10 Mar b. Briefly explain HDFS Name Node federation, NFS Gateway, Snapshots, Checkpoint a backups. (10 Mar Module-2 3 a. With neat diagrams, explain the Oozie DAG workflow and the types of nodes in workflow. (10 Mar b. Explain the features and benefits of apache HIVE in Hadoop. (10 Mar b. Explain the features and benefits of apache HIVE in Hadoop. (10 Mar COR 4 a. How do you run MapReduce and Message Passing Interface (MPI) on YARN architecture (08 Mar b. With neat diagram discuss the various frameworks that run under YARN. (08 Mar c. Discuss the various features of Hadoop YARN administration. (04 Mar b. Write any five Business Intelligence (BI) applications for various sectors. (10 Mar b. Explain the star schema of design of Data Ware Housing with an example. (07 Mar c. What is a confusion matrix? Explain. (03 Mar COR 6 a. Explain with diagram CRISP-DM data mining cycle. (10 Mar b. What do you understand by the term Data visualization? How is it important in Big D Analytics? (05 Mar c. Differentiate between Data Mining and Data Warehousing. (05 Mar c. Differentiate between Data Mining and Data Warehousing. (05 Mar c. Differentiate between Data Mining and Data Warehousing. (05 Mar c. Explain the design principles of an artificial neural network. (08 Mar b. List the advantages and disadvantages of a regression model. (06 Mar c. What is a splitting variable? Describe three criteria for choosing a splitting variable.	USN			17CS8
Time: 3 hrs. Max. Marks: 10 Note: Answer any FIVE full questions, choosing ONE full question from each module. 1 a. With a neat diagram explain the components of the Hadoop Distributed File Syst (HDFS). (00 Max) b. With a neat diagram, describe the steps in the MapReduce parallel flow data model. (10 Max) 0 0 (10 Max) 1 a. With neat diagram, explain the Oozie DAG workflow and the types of nodes in workflow. 0 (10 Max) (10 Max) 1 a. How do you run MapReduce and Message Passing Interface (MPI) on YARN architecture (08 Max) 1 Discuss the various features of Hado				23
Note: Answer any FIVE full questions, choosing ONE full question from each module. Module-1 a. With a neat diagram explain the components of the Hadoop Distributed File Syst (HDFS). (10 Maa) b. With a neat diagram, describe the steps in the MapReduce parallel flow data model. (10 Maa) OR a. Write the Java code for MAP and REDUCE of word count problem. Describe the steps compiling and removing the MapReduce program. (10 Maa) b. Briefly explain HDFS Name Node federation, NFS Gateway, Snapshots, Checkpoint 1 backups. (10 Maa) a. Write the features and benefits of apache HIVE in Hadoop. (10 Maa) b. Explain the features and benefits of apache HIVE in Hadoop. (10 Maa) Module-2 a. How do you run MapReduce and Message Passing Interface (MPI) on YARN architecture (08 Maa) b. With neat diagram discuss the various frameworks that run under YARN. (08 Maa) c. Discuss the various features of Hadoop YARN administration. (04 Maa) OR a. Write any five Business Intelligence (B1) applications for various sectors. (10 Maa) Module-1 5 a. Write any five Business Intelligence (B1) applications for various sectors. (10 Maa) OR OR <t< th=""><th></th><th></th><th>Big Data Analytics</th><th></th></t<>			Big Data Analytics	
Module-1 a. With a neat diagram explain the components of the Hadoop Distributed File Syst (HDFS). (10 Max) b. With a neat diagram, describe the steps in the MapReduce parallel flow data model. (10 Max) (HDFS). (10 Max) (ID Max) (10 Max) (ID Max) (ID Max) (ID Max) (ID Max) (ID Max) (ID Max) Briefly explain HDFS Name Node federation, NFS Gateway, Snapshots, Checkpoint is backups. (ID Max) (ID Max) (ID Max)	Tim	ne: 3	hrs. Max	Marks: 100
1 a. With a neat diagram explain the components of the Hadoop Distributed File Syst (HDFS). (10 Mar b. With a neat diagram, describe the steps in the MapReduce parallel flow data model. (10 Mar 0 2 a. Write the Java code for MAP and REDUCE of word count problem. Describe the steps compiling and removing the MapReduce program. (10 Mar b. Briefly explain HDFS Name Node federation, NFS Gateway, Snapshots, Checkpoint to backups. (10 Mar b. Briefly explain HDFS Name Node federation, NFS Gateway, Snapshots, Checkpoint to backups. (10 Mar a. With neat diagrams, explain the Oozie DAG workflow and the types of nodes in workflow. (10 Mar b. Explain the features and benefits of apache HIVE in Hadoop. (10 Mar b. Explain the features and benefits of apache HIVE in Hadoop. (10 Mar c. Discuss the various frameworks that run under YARN. (08 Mar b. With neat diagram discuss the various frameworks that run under YARN. (08 Mar c. Discuss the various features of Hadoop YARN administration. (04 Mar Module-3 5 a. Write any five Business Intelligence (BI) applications for various sectors. (10 Mar b. Explain the star schema of design of Data Ware Housing with an example. (07 Mar c. What is a confusion matrix? Explain. (03 Mar OR a. Explain with diagram CRISP-DM data mining cycle. (10 Mar b. What do you understand by the term Data visualization? How is it impor		N	ote: Answer any FIVE full questions, choosing ONE full question from each	module.
(HDFS). (10 Mar b. With a neat diagram, describe the steps in the MapReduce parallel flow data model. (10 Mar 0R 0 a. Write the Java code for MAP and REDUCE of word count problem. Describe the steps compiling and removing the MapReduce program. (10 Mar b. Briefly explain HDFS Name Node federation, NFS Gateway, Snapshots, Checkpoint is backups. (10 Mar a. With neat diagrams, explain the Oozie DAG workflow and the types of nodes in workflow. (10 Mar b. Explain the features and benefits of apache HIVE in Hadoop. (10 Mar b. Explain the features and benefits of apache HIVE in Hadoop. (10 Mar c. With neat diagram discuss the various frameworks that run under YARN. (08 Mar b. With neat diagram discuss the various frameworks that run under YARN. (08 Mar c. Discuss the various features of Hadoop YARN administration. (04 Mar c. Write any five Business Intelligence (B1) applications for various sectors. (10 Mar b. Explain the star schema of design of Data Ware Housing with an example. (07 Mar c. What is a confusion matrix? Explain. (03 Mar OR a. Explain with diagram CRISP-DM data mining cycle. (10 Mar b. What do you understand by the term Data visualization? How is it important in Big C Analytics? (05 Mar c.				
b. With a neat diagram, describe the steps in the MapReduce parallel flow data model. (10 Mar OR a. Write the Java code for MAP and REDUCE of word count problem. Describe the steps compiling and removing the MapReduce program. (10 Mar b. Briefly explain HDFS Name Node federation, NFS Gateway, Snapshots, Checkpoint a backups. (10 Mar Module-2 3 a. With neat diagrams, explain the Oozie DAG workflow and the types of nodes in workflow. (10 Mar b. Explain the features and benefits of apache HIVE in Hadoop. (10 Mar b. Explain the features and benefits of apache HIVE in Hadoop. (10 Mar c) Mar B. With neat diagram discuss the various frameworks that run under YARN. (08 Mar c) Discuss the various frameworks that run under YARN. (08 Mar c) Discuss the various frameworks that run under YARN. (04 Mar b. Write any five Business Intelligence (BI) applications for various sectors. (10 Mar b. Explain the star schema of design of Data Ware Housing with an example. (07 Mar c) What is a confusion matrix? Explain. (03 Mar C)	1	a.		1 File System (10 Marks
OR 2 a. Write the Java code for MAP and REDUCE of word count problem. Describe the steps compiling and removing the MapReduce program. (10 Main b. Briefly explain HDFS Name Node federation, NFS Gateway, Snapshots, Checkpoint is backups. (10 Main backups.) 3 a. With neat diagrams, explain the Oozie DAG workflow and the types of nodes in workflow. (10 Main b. Explain the features and benefits of apache HIVE in Hadoop. (10 Main b. Explain the features and benefits of apache HIVE in Hadoop. (10 Main b. Explain the features and benefits of apache HIVE in Hadoop. (10 Main b. Explain the features and benefits of apache HIVE in Hadoop. (10 Main b. Under MapReduce and Message Passing Interface (MPI) on YARN architecture (08 Main c. Discuss the various frameworks that run under YARN. (08 Main c. Discuss the various features of Hadoop YARN administration. (04 Main b. Explain the star schema of design of Data Ware Housing with an example. (10 Main b. Explain the star schema of design of Data Ware Housing with an example. (07 Main c. What is a confusion matrix? Explain. (03 Main c. Differentiate between Data Mining and Data Warehousing. (05 Main c. Differentiate between Data Mining and Data Warehousing. (05 Main c. Differentiate between Data Mining and Data Warehousing. (05 Main c. Differentiate between Data Mining and Data Warehousing. (05 Main c. Differentiate between Data Mining and Data Warehousing. (05 Main c. Differentiate between Data Mining and Data Warehousing. (06 Main c. What is a splitting variable? Describe three criteria for choosing a splitting variable. (06 Main c. What is a splitting variable? Describe three criteria for choosing a splitting variable. (06 Main c. What is a splitting variable? Describe three criteria for choosing a splitting variable. (06 Main c. What is a splitting variable? Describe three criteria for choosing a splitting variable. (06 Main c. What is		b.		odel.
 2 a. Write the Java code for MAP and REDUCE of word count problem. Describe the steps compiling and removing the MapReduce program. (10 Mai b. Briefly explain HDFS Name Node federation, NFS Gateway, Snapshots, Checkpoint is backups. (10 Mai backups. (1				(10 Marks
compiling and removing the MapReduce program. (10 Mail b. Briefly explain HDFS Name Node federation, NFS Gateway, Snapshots, Checkpoint i backups. (10 Mail 3 a. With neat diagrams, explain the Oozie DAG workflow and the types of nodes in workflow. (10 Mail b. Explain the features and benefits of apache HIVE in Hadoop. (10 Mail 6 a. How do you run MapReduce and Message Passing Interface (MPI) on YARN architecture (08 Mail (08 Mail c. Discuss the various features of Hadoop YARN administration. (04 Mail 6 a. Write any five Business Intelligence (BI) applications for various sectors. (10 Mail b. Explain the star schema of design of Data Ware Housing with an example. (07 Mail) c. What is a confusion matrix? Explain. (08 Mail) 6 a. Explain with diagram CRISP-DM data mining cycle. (10 Mail) b. What do you understand by the term Data visualization? How is it important in Big D (05 Mail) c. Differentiate between Data Mining and Data Warehousing. (05 Mail) c. Differentiate between Data Mining and Tate work. (08 Mail) do you understand by the term Data visualization? How is it important in Big D (05 Mail) do You understand by the term Data visualization? How is it important in Big D (05 Mail) do Star (05 Mail) (05 Mail) <t< td=""><td></td><td></td><td></td><td></td></t<>				
 b. Briefly explain HDFS Name Node federation, NFS Gateway, Snapshots, Checkpoint a backups. (10 Mar Module-2 3 a. With neat diagrams, explain the Oozie DAG workflow and the types of nodes in workflow. (10 Mar b. Explain the features and benefits of apache HIVE in Hadoop. (10 Mar C) 4 a. How do you run MapReduce and Message Passing Interface (MPI) on YARN architecture (08 Mar c) b. With neat diagram discuss the various frameworks that run under YARN. (08 Mar c) c. Discuss the various features of Hadoop YARN administration. (04 Mar c) 5 a. Write any five Business Intelligence (BI) applications for various sectors. (10 Mar c) b. What is a confusion matrix? Explain. (03 Mar c) c. What is a confusion matrix? Explain. (08 Mar c) d. Explain with diagram CRISP-DM data mining cycle. (10 Mar d) you understand by the term Data visualization? How is it important in Big D Analytics? (05 Mar c) d. Explain the design principles of an artificial neural network. (08 Mar c) d. Explain the design principles of an artificial neural network. (08 Mar c) d. Explain the design principles of an artificial neural network. (08 Mar c) d. Explain the design principles of an artificial neural network. (06 Mar c) d. Explain the design principles of an artificial neural network. (06 Mar c) d. What is a splitting variable? Describe three criteria for choosing a splitting variable. (06 Mar c) 	2	a.		the steps of (10 Marks
Module-2 3 a. With neat diagrams, explain the Oozie DAG workflow and the types of nodes in workflow. (10 Mar b. Explain the features and benefits of apache HIVE in Hadoop. (10 Mar OR 4 a. How do you run MapReduce and Message Passing Interface (MPI) on YARN architecture (08 Mar b. With neat diagram discuss the various frameworks that run under YARN. (08 Mar c. Discuss the various features of Hadoop YARN administration. (04 Mar b. Write any five Business Intelligence (BI) applications for various sectors. (10 Mar b. Explain the star schema of design of Data Ware Housing with an example. (07 Mar c. What is a confusion matrix? Explain. (03 Mar OR 6 a. Explain with diagram CRISP-DM data mining cycle. (10 Mar b. What do you understand by the term Data visualization? How is it important in Big D Analytics? (05 Mar c. Differentiate between Data Mining and Data Warehousing. (05 Mar d. Explain the design principles of an artificial neural network. (08 Mar b. List the advantages and disadvantages of a regression model. (06 Mar c. What is a splitting variable? Describe three criteria for choosing a splitting variable. (06 Mar		b.	Briefly explain HDFS Name Node federation, NFS Gateway, Snapshots, C	
 3 a. With neat diagrams, explain the Oozie DAG workflow and the types of nodes in workflow. (10 Mar b. Explain the features and benefits of apache HIVE in Hadoop. (10 Mar) 4 a. How do you run MapReduce and Message Passing Interface (MPI) on YARN architecture (08 Mar) b. With neat diagram discuss the various frameworks that run under YARN. (08 Mar) c. Discuss the various features of Hadoop YARN administration. (04 Mar) 6 a. Write any five Business Intelligence (BI) applications for various sectors. (10 Mar) c. What is a confusion matrix? Explain. (03 Mar) 6 a. Explain with diagram CRISP-DM data mining cycle. (10 Mar) b. What do you understand by the term Data visualization? How is it important in Big D Analytics? (05 Mar) c. Differentiate between Data Mining and Data Warehousing. (05 Mar) c. Differentiate between Data Mining and Data Warehousing. (05 Mar) c. What is a splitting variable? Describe three criteria for choosing a splitting variable. (06 Mar) 			backups.	(10 Marks
workflow. (10 Mail b. Explain the features and benefits of apache HIVE in Hadoop. (10 Mail OR 0 4 a. How do you run MapReduce and Message Passing Interface (MPI) on YARN architecture (08 Mail (08 Mail b. With neat diagram discuss the various frameworks that run under YARN. (08 Mail c. Discuss the various features of Hadoop YARN administration. (04 Mail b. Write any five Business Intelligence (BI) applications for various sectors. (10 Mail b. Explain the star schema of design of Data Ware Housing with an example. (07 Mail) c. What is a confusion matrix? Explain. (03 Mail) b. What do you understand by the term Data visualization? How is it important in Big D (05 Mail) c. Differentiate between Data Mining and Data Warehousing. (05 Mail) c. Differentiate between Data Mining and Data Warehousing. (06 Mail) d. List the advantages and disadvantages of a regression model. (06 Mail) c. What is a splitting variable? Describe three criteria for choosing a splitting variable. (06 Mail)				
b. Explain the features and benefits of apache HIVE in Hadoop. (10 Mar OR 4 a. How do you run MapReduce and Message Passing Interface (MPI) on YARN architecture (08 Mar b. With neat diagram discuss the various frameworks that run under YARN. (08 Mar c. Discuss the various features of Hadoop YARN administration. (04 Mar Module-3 5 a. Write any five Business Intelligence (BI) applications for various sectors. (10 Mar b. Explain the star schema of design of Data Ware Housing with an example. (07 Mar c. What is a confusion matrix? Explain. (03 Mar OR 6 a. Explain with diagram CRISP-DM data mining cycle. (10 Mar b. What do you understand by the term Data visualization? How is it important in Big D Analytics? (05 Mar c. Differentiate between Data Mining and Data Warehousing. (05 Mar c. What is a splitting variable? Of an artificial neural network. (08 Mar b. List the advantages and disadvantages of a regression model. (06 Mar c. What is a splitting variable? Describe three criteria for choosing a splitting variable. (06 Mar	3	a.		nodes in the (10 Marks
 4 a. How do you run MapReduce and Message Passing Interface (MPI) on YARN architecture (08 Marchitecture). With neat diagram discuss the various frameworks that run under YARN. (08 Marchitecture). Discuss the various features of Hadoop YARN administration. (04 Marchitecture). (04 Marchitecture). (04 Marchitecture). (04 Marchitecture). (05 Marchitecture). (06 Marchitecture).		b.		(10 Marks
 4 a. How do you run MapReduce and Message Passing Interface (MPI) on YARN architecture (08 Marchitecture). With neat diagram discuss the various frameworks that run under YARN. (08 Marchitecture). Discuss the various features of Hadoop YARN administration. (04 Marchitecture). (04 Marchitecture). (04 Marchitecture). (04 Marchitecture). (05 Marchitecture). (06 Marchitecture).			OR	
 b. With neat diagram discuss the various frameworks that run under YARN. (08 Mar. c. Discuss the various features of Hadoop YARN administration. (04 Mar. 6 a. Write any five Business Intelligence (BI) applications for various sectors. (10 Mar. c. What is a confusion matrix? Explain. (03 Mar. 6 a. Explain with diagram CRISP-DM data mining cycle. (10 Mar. b. What do you understand by the term Data visualization? How is it important in Big D Analytics? (05 Mar. c. Differentiate between Data Mining and Data Warehousing. (05 Mar. 7 a. Explain the design principles of an artificial neural network. (08 Mar. b. List the advantages and disadvantages of a regression model. (06 Mar. c. What is a splitting variable? Describe three criteria for choosing a splitting variable. (06 Mar. 	4	a.		
 c. Discuss the various features of Hadoop YARN administration. (04 Mathematical Module-3) 5 a. Write any five Business Intelligence (BI) applications for various sectors. (10 Mathematical Mathema		b.	With neat diagram discuss the various frameworks that run under YARN.	(08 Marks (08 Marks
 5 a. Write any five Business Intelligence (BI) applications for various sectors. (10 Mar. b. Explain the star schema of design of Data Ware Housing with an example. (07 Mar. c. What is a confusion matrix? Explain. (03 Mar. 6 a. Explain with diagram CRISP-DM data mining cycle. (10 Mar. b. What do you understand by the term Data visualization? How is it important in Big D Analytics? (05 Mar. c. Differentiate between Data Mining and Data Warehousing. (05 Mar. 7 a. Explain the design principles of an artificial neural network. (08 Mar. b. List the advantages and disadvantages of a regression model. (06 Mar. c. What is a splitting variable? Describe three criteria for choosing a splitting variable. (06 Mar. 		c.		(04 Marks
 b. Explain the star schema of design of Data Ware Housing with an example. (07 Material of Material Scheme Schem			Module-3	
 c. What is a confusion matrix? Explain. (03 Matrix) 6 a. Explain with diagram CRISP-DM data mining cycle. (10 Matrix) b. What do you understand by the term Data visualization? How is it important in Big D Analytics? (05 Matrix) c. Differentiate between Data Mining and Data Warehousing. (05 Matrix) 7 a. Explain the design principles of an artificial neural network. (08 Matrix) b. List the advantages and disadvantages of a regression model. (06 Matrix) c. What is a splitting variable? Describe three criteria for choosing a splitting variable. (06 Matrix) 	5			(10 Marks (07 Marks
 6 a. Explain with diagram CRISP-DM data mining cycle. (10 Man b. What do you understand by the term Data visualization? How is it important in Big D Analytics? (05 Man c. Differentiate between Data Mining and Data Warehousing. (05 Man <u>Module-4</u> 7 a. Explain the design principles of an artificial neural network. (08 Man b. List the advantages and disadvantages of a regression model. (06 Man c. What is a splitting variable? Describe three criteria for choosing a splitting variable. (06 Man 				(07 Marks) (03 Marks)
 6 a. Explain with diagram CRISP-DM data mining cycle. (10 Man b. What do you understand by the term Data visualization? How is it important in Big D Analytics? (05 Man c. Differentiate between Data Mining and Data Warehousing. (05 Man <u>Module-4</u> 7 a. Explain the design principles of an artificial neural network. (08 Man b. List the advantages and disadvantages of a regression model. (06 Man c. What is a splitting variable? Describe three criteria for choosing a splitting variable. (06 Man 		A		
 b. What do you understand by the term Data visualization? How is it important in Big D Analytics? (05 Mar c. Differentiate between Data Mining and Data Warehousing. (05 Mar 7 a. Explain the design principles of an artificial neural network. (08 Mar b. List the advantages and disadvantages of a regression model. (06 Mar c. What is a splitting variable? Describe three criteria for choosing a splitting variable. (06 Mar 	6	a.		(10 Mark
 c. Differentiate between Data Mining and Data Warehousing. (05 Manual Module-4) 7 a. Explain the design principles of an artificial neural network. (08 Manual b. List the advantages and disadvantages of a regression model. (06 Manual C. What is a splitting variable? Describe three criteria for choosing a splitting variable. (06 Manual C. What is a splitting variable? Describe three criteria for choosing a splitting variable. (06 Manual C. What is a splitting variable? Describe three criteria for choosing a splitting variable. (06 Manual C. What is a splitting variable? Describe three criteria for choosing a splitting variable. (06 Manual C. What is a splitting variable? Describe three criteria for choosing a splitting variable. (06 Manual C. What is a splitting variable? Describe three criteria for choosing a splitting variable. (06 Manual C. What is a splitting variable? Describe three criteria for choosing a splitting variable. (06 Manual C. What is a splitting variable? Describe three criteria for choosing a splitting variable. (06 Manual C. What is a splitting variable? Describe three criteria for choosing a splitting variable. (06 Manual C. What is a splitting variable? Describe three criteria for choosing a splitting variable. (06 Manual C. What is a splitting variable? Describe three criteria for choosing a splitting variable. (06 Manual C. What is a splitting variable? Describe three criteria for choosing a splitting variable. (06 Manual C. What is a splitting variable? Describe three criteria for choosing a splitting variable? (06 Manual C. What is a splitting variable? (06 Manual C. What		b.	What do you understand by the term Data visualization? How is it important	
Module-4 (08 Man 7 a. Explain the design principles of an artificial neural network. (08 Man b. List the advantages and disadvantages of a regression model. (06 Man c. What is a splitting variable? Describe three criteria for choosing a splitting variable. (06 Man		c.		(05 Marks) (05 Marks)
 7 a. Explain the design principles of an artificial neural network. (08 Man b. List the advantages and disadvantages of a regression model. (06 Man c. What is a splitting variable? Describe three criteria for choosing a splitting variable. (06 Man 			, Cas	
 b. List the advantages and disadvantages of a regression model. (06 Man c. What is a splitting variable? Describe three criteria for choosing a splitting variable. (06 Man 	7	a.		(08 Marks
(06 Ma)		b.	List the advantages and disadvantages of a regression model.	(06 Marks
1 of 2		c.	what is a splitting variable? Describe three criteria for choosing a splitting var	(06 Mark
l of 2				
			1 of 2	
			Cal	
As grant				8

(10 Marks)

OR

- Explain the design principles of an Artificial Neural Network. 8 a.
 - How does the apriori algorithm work? Apply the same for the following example. b.

110 1 400	s the aprior argonanti we
T _{ID}	List of Item – IDs
T ₁₀₀	I_1, I_2, I_5
T ₂₀₀	I_2, I_4
T ₃₀₀	I_2, I_3
T ₄₀₀	I_1, I_2, I_4
T ₅₀₀	I_1, I_3
T ₆₀₀	I_2, I_3
T ₇₀₀	I ₁ , I ₃
T ₈₀₀	I_1, I_2, I_3, I_5
T ₉₀₀	I_1, I_2, I_3

Assume the support count = 2.

(10 Marks)

(06 Marks)

Module-5

- 9 Compare text mining with data mining. a.
 - What is Naïve Baye's technique? Explain its model. b. (06 Marks) (08 Marks)
 - Explain steps in the text mining process and architecture. C.

OR

- 10 What is Web mining? Explain the different types of Web mining. (08 Marks) a. Write a short note on Social Network Analysis (SNA). Numerical examples on Naïve b. Baye's model, SYM and SNA (Rank calculation). (06 Marks)
 - Explain three types of Web mining. Use an appropriate flow diagram to represent the same. C.

(06 Marks)

2 of 2