Q. Code: 250130

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

B.E. / B.TECH. DEGREE EXAMINATIONS, MAY 2023

Sixth Semester

AD18602 – BIG DATA ANALYTICS AND VISUALIZATION

(Artificial Intelligence and Data Science)

(Regulation 2018/2018A)

TIME: 3 HOURS **MAX. MARKS: 100**

COURSE OUTCOMES	STATEMENT	RBT LEVEL
CO 1	Apply analytics to big data applications.	**3**
CO 2	Understand the hadoop frameworks.	**2**
CO 3	Use MongoDB and Cassandra for big data storage and retrieval.	**3**
CO 4	Work with hadoop ecosystem tools such as MapReduce, Hive, Pig.	**3**
CO 5	Visualize data to transform into information for further analysis.	**3**

PART- A (10 x 2 = 20 Marks)

(Answer all Questions)

		CO	RBT LEVEL
1.	Can the same visualization tool that we run over conventional data warehouse, be used in big data environment?	**1**	**2**
2.	Differentiate Traditional Business Intelligence (BI) and Big Data.	**1**	**2**
3.	What are the core aspects of Hadoop?	**2**	**1**
4.	What is the difference between replication and sharding?	**2**	**1**
5.	Write the query to conform the existence of database created in MongoDB shell.	**3**	**1**
6.	Define anti-entropy and read repair in Cassandra.	**3**	**1**
7.	Write a MapReduce program to count the occurrences of similar words in a file using combiner for optimization.	**4**	**3**
8.	Create a database “Students” with comments and database properties in a HIVE shell.	**4**	**2**
9.	What is a good visualization?	**5**	**1**
10.	Define probability density function.	**5**	**1**

PART- B (5 x 14 = 70 Marks)

			Marks	CO	RBT LEVEL
11. (a)	**(i)**	Explain the classification of analytics	**(9)**	**1**	**3**
	(ii)	Explain the different challenges faced in big data.	**(5)**	**1**	**3**
		(OR)			
(b)		Explain different terminologies used in big data environment.	**(14)**	**1**	**3**
12. (a)	**(i)**	Illustrate the YARN architecture with diagram.	**(8)**	**2**	**3**
	(ii)	Explain in detail how YARN takes Hadoop beyond batch.	**(6)**	**2**	**3**

(OR)

(b)	Explain Hadoop Distributed File System in detail with diagrams.	**(14)**	**2**	**3**

13. (a)	To create a collection by the name "alphabets" and insert documents in it containing two fields "_id" and "alphabet". The values stored in the "alphabet" field should be "a","b","c","d" etc, with one value stored per document. There should be 26 documents in all. We need to use cursor in MongoDB to iterate through the "alphabet" collection.	**(14)**	**3**	**3**

(OR)

(b)	**(i)**	Explain features of Cassandra in detail.	**(7)**	**3**	**3**
	(ii)	Explain CRUD operations briefly.	**(7)**	**3**	**3**

14. (a)	Write a MapReduce program to count the occurrence of similar word in a file. Use partitioner to partition key based on alphabets. Input Data: Welcome to Hadoop Session Introduction to Hadoop Introducing Hive Hive Session Pig Session	**(14)**	**4**	**3**

(OR)

(b)	Explain briefly Hive architecture, Hive Datatypes, Hive file format and Hive query language.	**(14)**	**4**	**3**

15. (a)	Explain Visualization plots with a suitable example.	**(14)**	**5**	**3**

(OR)

(b)	**(i)**	Explain the Gestalt principles of perception	**(8)**	**5**	**3**
	(ii)	Explain the Visualization tools used in Python	**(6)**	**5**	**3**

PART- C (1 x 10 = 10 Marks)

(Q.No.16 is compulsory)

		Marks	CO	RBT LEVEL
16.	Develop a MapReduce program for the given list of employees to sort data based on the salary using the total order sorting.	**(10)**	**4**	**5**

Input data:

Name	Designation	Department	Annual Salary
Tomas	Paramedic	Fire	91080
Tim	Lieutenant	Fire	114846
Eric	Sergeant	Police	104628
Luis	Police officer	Police	96060
Marie	Clerk	Police	53076

****