



**NAVRACHANA
UNIVERSITY**

a UGC recognized University

School: School of Engineering and Technology
Program/s: Computer Science & Engineering
Year: 2nd **Semester:** 7th
Examination: End Semester Examination
Examination year: December - 2021

Course Code: CS415 **Course Name:** Big Data Analytics & Data Science
Date: 02/12/2021
Time: 02:30 pm to 04:30 pm

Total Marks: 40
Total Pages: 2

Instructions:

- Write each answer on a new page.
- Use of a calculator is permitted/not permitted.
- * COs=Course Outcome mapping. # BTL=Bloom's Taxonomy Level mapping

Q. No.	Details	Marks	COs*	BTL#
Q.1	Multiple choice questions.			
1.	_____ controlled the distribution of application requests to compute resources in cluster. a) Task Tracker c) Resource Manager b) Job Tracker d) Node Manager	10	CO1	BT1
2.	_____ task is to consolidated relevant records from mapping phase output. a) Shuffling c) Sorting b) Reducing d) Mapping		CO2	BT2
3.	_____ is the node where data is presented in advance before any processing takes place. a) Data Node c) Mapper b) Name Node d) Payload		CO3	BT3
4.	_____ is the execution of a mapper or a reducer on a slice of data. a) Task c) Job b) Task Attempt d) Master		CO3	BT2
5.	_____ supports both real time as well as batch processing. a) Apache Hadoop c) Apache spark b) Apache Pig d) Apache Hive		CO4	BT3
6.	_____ uses dynamic schema. a) Mongo dB c) MySQL b) DBMS d) RDBMS		CO4	BT4
7.	_____ uses multi-query approach thereby reducing the length of codes to great extent. a) Apache Pig c) Map Reduce b) Apache Storm d) Apache Spark		CO4	BT3
8.	_____ provides a full stack analytics platform but its visualization capabilities provide a simple to use drag and drop interface which allow chart or complex graphics. a) Sisense c) Data Wrapping b) Plotly d) High Charts		CO5	BT5
9.	_____ is responsible for mining vast amounts of data. a) Data Engineer c) Data Analyst b) Satisfaction d) Data Administrator		CO3	BT3
10.	_____ is the product of Google's brain team coming together for the purpose of advancing machine learning.		CO5	BT5

- a) Tensor Flow c) Keras
 b) Tableau d) Rapid Miner

Q.2	Fill in the blanks/Short Answers/True or False	5		
	1. _____ is the node where Job Tracker runs and which accepts job request from clients.		C01	BT1
	2. Flink has the ability to recover from faults easily. (True/False)		C02	BT2
	3. _____ is a data warehouse software project built on the top of apache Hadoop for providing data query and analysis.		C03	BT3
	4. The full form of CDH_____.		C04	BT2
	5. Instant visual feedback is feature of _____.		C05	BT3
Q.3	Explain about the working of map reduce.	5	C03	BT3
	Or			
	Explain Hadoop daemons in detail.		C04	BT2
Q.4	Explain the difference between apache pig and apache Cassandra.	5	C04	BT4
	Or			
	Explain the commands used in the installation of hive.		C04	BT4 BT6
Q.5	Write about Pseudo distributed mode and Fully distributed mode of operations in hadoop.	5	C04	BT2 BT4
Q.6	What is tableau? How can to differentiate from data wrapper.	5	C05	BT4 BT5 BT6
Q.7	Explain the process involved in data science.	5	C02	BT3
	Or		C05	BT4
	Explain Geospatial and multivariate display.		C04	BT4
			C05	

*****End of Question Paper*****