Enrollment No._____

NAVRACHANA UNIVERSITY
a UGC recognized University

School: School of Engineering and Technology
Program/s: B.Tech(CSE)
Year: 3rd **Semester:** 6th
Examination: End Semester Examination
Examination year: May – 2023

Course Code: CS415 **Course Name:** Big Data Analytics
Date: 19/05/2023 **Total Marks:** 40
Time: 02:00 am to 04:00 am **Total Pages:** 03

Instructions:

a) Attempt All the Questions
b) No Calculator is allowed

Q. No.	Details	Marks	CO	BTLO
	Attempt All the Questions(1 marks each)	40		
Q.1	Which of the following is NOT a characteristic of big data? a. Volume b. Velocity c. Variety d. Validation		CO2	BTLO2
Q.2	Which of the following is an example of structured data? a. Emails b. Tweets c. Sensor data d. Sales transactions		CO3	BTLO3
Q.3	Which of the following is an example of semi-structured data? a. XML files b. CSV files c. Relational databases d. JSON files		CO4	BTLO4
Q.4	Which of the following is an example of unstructured data? A. Videos B. Emails C. JSON files D. CSV files		CO5	BTLO5
Q.5	Which of the following is NOT a popular big data storage technology? A. Hadoop Distributed File System (HDFS) B. Amazon S3 C. MySQL D. d. Google Cloud Storage		CO1	BTLO1

Q.6	What is the primary reason why traditional data processing tools cannot handle big data? A. They cannot store large amounts of data B. They cannot process data fast enough C. They do not support distributed computing D. d. They cannot analyze complex data types		CO2	BT
Q.7	Which of the following is NOT a step in the big data analytics process? A. Data collection B. Data integration C. Data analysis D. d. Data encryption		CO3	BTLO3
Q.8	Which of the following is NOT a type of big data analysis? A. Descriptive analysis B. Diagnostic analysis C. Predictive analysis D. d. Proactive analysis		CO4	BTLO4
Q.9	What is the purpose of data mining in big data analytics? A. To find patterns in data B. To store data in a database C. To encrypt data for security D. d. To visualize data for decision-making		CO5	BTLO5
Q.10	What is the purpose of machine learning in big data analytics? A. To find patterns in data B. To store data in a database C. To encrypt data for security D. d. To visualize data for decision-making		CO5	BTLO5
Q.11	Which of the following is an example of supervised machine learning? A. Clustering B. Association rule learning C. Decision tree D. d. Principal component analysis		CO2	BTLO5
Q.12	Which of the following is an example of a data preprocessing technique? A. Principal component analysis B. Random forest C. Logistic regression D. d. Gradient descent		CO2	BTLO3
Q.13	Which of the following is NOT a data integration technique? A. Extract, Transform, Load (ETL) B. Extract, Load, Transform (ELT)		CO3	BTLO3

	C. Extract, Transform, Convert (ETC) D. d. Extract, Convert, Load (ECL)		
Q.14	Which of the following is a popular big data processing framework? A. Excel B. Python C. Apache Hadoop D. d. SQL Server	CO3	BTLO3
Q.15	Which of the following is a popular big data streaming platform? A. Apache Kafka B. MySQL C. Hadoop Distributed File System (HDFS) D. d. Amazon S3	CO2	BTLO2
Q.16	Which of the following is a popular big data NoSQL database? A. Oracle Database B. SQL Server C. MongoDB D. MySQL	CO2	BTLO2
Q.17	What is the primary advantage of using Hadoop for big data processing? A. Hadoop can store and process structured data B. Hadoop is faster than traditional data processing tools C. Hadoop can process data in parallel across multiple machines D. d. Hadoop is easier to use than traditional data processing tools	CO4	BTLO1
Q.18	Which of the following is a popular big data machine learning algorithm? A. K-Nearest Neighbors B. Support Vector Machines C. Naive Bayes D. d. All of the above	CO4	BTLO1
Q.19	What is the purpose of data sampling in big data analytics? A. To select a random subset of data for analysis B. To eliminate outliers from the data C. To preprocess data before analysis D. d. To aggregate data for faster processing	CO4	BTLO1
Q.20	Which of the following is a popular big data distributed computing framework? A. Apache Hadoop B. Apache Spark C. Apache Flink D. Apache Cassandra	CO5	BTLO2
Q.21	Which of the following is NOT a popular big data analytics use case? A. Fraud detection	CO5	BTLO4

	B. Recommendation systems C. Sentiment analysis D. d. Image processing			
Q.22	Which of the following is NOT a big data governance concern? A. Data quality B. Data privacy C. Data security D. d. Data visualization		CO5	BTLO4
Q.23	Which AWS service allows you to process and analyze streaming data in real-time? A. Amazon S3 B. Amazon Kinesis C. AWS Glue D. d. Amazon Redshift		CO2	BTLO4
Q.24	Which AWS service provides a managed Apache Spark environment for big data processing? A. Amazon EMR B. Amazon Kinesis C. AWS Glue D. d. Amazon Redshift		CO1	BTLO4
Q.25	Which AWS service is used for interactive querying and analysis of big data stored in S3? A. Amazon Redshift B. Amazon Athena C. AWS Glue D. Amazon QuickSight		CO1	BTLO4
Q.26	Which AWS service is a fully-managed, highly scalable, and secure data lake solution? A. Amazon S3 B. AWS Glue C. Amazon Redshift D. d. Amazon Elasticache		CO1	BTLO4
Q.27	Which AWS service allows you to build and train machine learning models on big data? A. Amazon Sagemaker B. Amazon Athena C. AWS Glue D. d. Amazon Redshift		CO2	BTLO4

28	Which AWS service allows you to run SQL queries on your data stored in S3? A. Amazon Redshift B. Amazon Athena C. AWS Glue D. d. Amazon EMR		CO4	BTLO4
Q.29	Which AWS service enables real-time processing of streaming data at a massive scale? A. Amazon S3 B. Amazon Kinesis Data Streams C. Amazon RDS D. d. Amazon Elastic MapReduce		CO6	BTLO4
Q.30	Which AWS service provides a fully-managed Apache Hadoop framework? A. Amazon S3 B. AWS Glue C. Amazon Redshift D. d. Amazon EMR		CO6	BTLO4
Q.31	Which AWS service provides a fully-managed Elastic search service for searching and analyzing data? A. Amazon S3 B. Amazon Kinesis C. AWS Glue D. d. Amazon Elastic search Service		CO6	BTLO4
Q.32	Which AWS service is used to collect, process, and analyze large volumes of IoT data? A. Amazon S3 B. Amazon Kinesis C. AWS Glue D. d. Amazon QuickSight		CO1	BTLO1
Q.33	Which AWS service is a fully-managed data warehouse solution? A. Amazon S3 B. Amazon Redshift C. AWS Glue D. d. Amazon Athena		CO4	BTLO4
Q.34	Which of the following is a popular data visualization library in Python used in Big Data Analytics? A. Matplotlib B. Seaborn C. Plotly D. d. All of the above		CO4	BTLO4

Q.35	Which of the following is a commonly used technique for handling imbalanced datasets in Big Data Analytics? A. Oversampling B. Undersampling C. SMOTE (Synthetic Minority Over-sampling Technique) D. d. All of the above		CO3	BTLO5
Q.36	Which of the following is a common method of handling missing data in Big Data Analytics? A. Dropping the missing data rows or columns B. Filling the missing data with the mean or median C. Interpolating the missing data values D. All of the above		CO3	BTLO5
Q.37	What is the primary purpose of data preprocessing in Big Data Analytics? A. To clean and transform raw data B. To visualize data for insights C. To apply machine learning algorithms D. d. To store data in a database		CO3	BTLO5
Q.38	Which of the following is NOT a commonly used data visualization tool in Big Data Analytics? A. Tableau B. Power BI C. Matplotlib D. d. Excel		CO4	BTLO1
Q.39	What is the purpose of dimensionality reduction in Big Data Analytics? A. To remove redundant or irrelevant features B. To increase the dimensionality of the data C. To create new features from existing ones D. d. To validate the accuracy of the data		CO4	BTLO1
Q.40	Which of the following is a commonly used technique for text mining in Big Data Analytics? A. Natural Language Processing (NLP) B. Decision trees C. Support Vector Machines (SVM) D. d. Linear regression		CO5	BTLO1

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