

JAMIA HAMDARD
Department of Computer Science and Engineering
School of Engineering Science and Technology
New Dehi-110062
Odd Semester Examination-2024
Programme Name: Bachelor of Computer Application (BCA)

Paper Name: Introduction to Data Science and Big Data (PE-2)
Time: 2.30 Hours

Paper Code: BCA PE523/ BSC PE523
Max. Marks: 60

SECTION A (20x1 = 20 Marks)				
Q.No	All questions are compulsory	Marks	CO	BL
1.	NoSQL databases typically require a predefined schema for data storage, similar to relational Databases.	1	CO1	L1
2.	One of the key benefits of NoSQL databases is their ability to enforce strict data consistency in distributed systems.	1	CO1	L2
3.	MapReduce is a two-step process consisting of the "Map" phase and the "Reduce" phase.	1	CO3	L1
4.	HDFS uses a master-slave architecture, where the NameNode is the master and manages metadata, while the DataNodes store the actual data.	1	CO2	L2
5.	HBase supports strong consistency, meaning that once data is written, it is immediately visible to all clients.	1	CO1	L1
6.	NoSQL is often used for storing unstructured data in Big Data environments.	1	CO2	L2
7.	By creating multiple copies of data enhance data replication reliability in HDFS?	1	CO3	L3
8.	A Hadoop cluster is running slowly due to frequent garbage collection because of Network issues.	1	CO2	L1
9.	By parallel processing across nodes Apache Pig handle large datasets in a distributed environment.	1	CO1	L1
10.	AES encryption is commonly used to enforce encryption in Apache Spark jobs.	1	CO2	L2
11.	Which of the following challenges is specifically associated with Big Data's velocity? (A) Ensuring data accuracy (B) Handling the speed at which data is generated (C) Reducing data storage requirements (D) Visualizing the data	1	CO1	L1
12.	Which type of data does the variety aspect of Big Data primarily address? (A) Structured (B) Unstructured (C) Both structured and unstructured (D) Neither	1	CO2	L2
13.	Which command is used to list the files in a Hadoop directory? (A) hdfs dfs -ls (B) hdfs dfs -rm (C) hdfs dfs -put (D) hdfs dfs -copyFromLocal	1	CO3	L2
14.	Which of the following is NOT one of the 3Vs of Big Data? (A) Volume (B) Velocity (C) Variety (D) Validation	1	CO3	L1
15.	What is the primary purpose of HDFS in Big Data storage? (A) To store relational data (B) To store large files across multiple Machines (C) To store in-memory data (D) To compress files	1	CO2	L2
16.	In Hadoop's HDFS, each data block is by default replicated across _____ nodes to ensure fault tolerance.	1	CO3	L3
17.	Apache Hive converts HiveQL queries into _____ jobs, which are executed on the Hadoop cluster to process large-scale data.	1	CO1	L2

18.	In the MapReduce framework, the _____ phase aggregates the intermediate output produced by the Map tasks and produces the final result.	1	CO2	L1
19.	_____ is an example of a widely used document-based NoSQL database, where data is stored in JSON-like format.	1	CO1	L1
20.	Apache Pig is primarily used to process large datasets on Hadoop using a scripting language called _____.	1	CO2	L2

Section B Attempt any four questions out of six (4*5=20 Marks)				
Q.No	Questions	Marks	CO	BL
1.	How is Big Data being utilized in the pharmaceutical industry to enhance drug development and patient outcomes?	5	CO2	L2
2.	What is data replication in Hadoop and how does it contribute to fault tolerance in HDFS?	5	CO4	L3
3.	What is Hbase and how does it differ from traditional relational databases?	5	CO3	L3
4.	Write short notes on Avro file based data structure and Hadoop I/O.	5	CO1	L1
5.	Explain credit risk management and challenges to successful credit risk management.	5	CO4	L4
6.	What are the different cases for fraud detection using big data analytics?	5	CO5	L5

Section C (2x10 = 20 Marks) Attempt any two questions out of four				
Q.No	Questions	Marks	CO	BL
1.	Discuss the four main types of NoSQL databases. Explain their internal structure and typical use cases, and compare them to traditional relational databases.	10	CO2	L2
2.	Describe the working of the MapReduce framework in Hadoop. How does it perform parallel data processing? Explain with an example.	10	CO3	L3
3.	Describe Cassandra data model and discuss the characteristic of Cassandra data model. What are the use cases of Cassandra data model.	10	CO5	L2
4.	Discuss in detail YARN and it's role in Hadoop.	10	CO4	L4