

Module 1: Introduction to Big Data and Hadoop - Multiple Choice Questions

Questions:

1. Which of the following is NOT one of the 5 V's of Big Data? a) Volume b) Velocity c) Variety d) Visibility
2. Hadoop is primarily designed to handle which type of data? a) Small, structured datasets b) Large, unstructured datasets c) Real-time streaming data d) Relational database data
3. Which component of the Hadoop ecosystem is responsible for storing large files across a cluster of commodity hardware? a) MapReduce b) YARN c) HDFS d) Hive
4. In HDFS, what is the role of the NameNode? a) Stores actual data blocks b) Manages the file system metadata c) Performs data replication d) Executes MapReduce tasks
5. Which MapReduce phase is responsible for sorting and combining intermediate data? a) Mapper b) Reducer c) Shuffle and Sort d) Partitioner
6. What is the primary advantage of using Hadoop for Big Data processing? a) It is easy to learn and use b) It provides real-time data processing c) It can scale to handle massive datasets on commodity hardware d) It supports SQL-like queries for data analysis
7. Which of the following is NOT a core component of the Hadoop ecosystem? a) HDFS b) MapReduce c) Spark d) YARN
8. In the context of Big Data, what does "Velocity" refer to? a) The speed at which data is generated and processed b) The variety of data types being handled c) The accuracy and reliability of data d) The total volume of data being stored
9. What is the purpose of data replication in HDFS? a) To improve data security b) To increase data processing speed c) To ensure fault tolerance and data availability d) To reduce storage costs
10. Which of the following best describes the MapReduce paradigm? a) A distributed file system for storing large datasets b) A programming model for parallel processing of Big Data c) A resource management framework for Hadoop clusters d) A SQL-like query language for analyzing data in Hadoop

Answers:

1. d) Visibility
2. b) Large, unstructured datasets
3. c) HDFS
4. b) Manages the file system metadata
5. c) Shuffle and Sort
6. c) It can scale to handle massive datasets on commodity hardware
7. c) Spark
8. a) The speed at which data is generated and processed

9. c) To ensure fault tolerance and data availability
10. b) A programming model for parallel processing of Big Data