

MID-1 BDA Bit Bank

UNIT-1

1	What is the term used for a collection of large, complex data sets that cannot be processed using traditional data processing tools? a) Big Data b) Small Data c) Medium Data d) Mini Data	[]
2	Which of the following is not one of the four V's of Big Data? a) Velocity b) Volume c) Variety d) Value	[]
3	What is the process of transforming structured and unstructured data into a format that can be easily analyzed? a) Data Mining b) Data Warehousing c) Data Integration d) Data Processing	[]
4	Which of the following is a tool used for processing and analyzing Big Data? a) Hadoop b) MySQL c) PostgreSQL d) Oracle	[]
5	What is the process of examining large and varied data sets to uncover hidden patterns, unknown correlations, market trends, customer preferences, and other useful information? a) Data Mining b) Data Warehousing c) Data Integration d) Data Processing	[]
6	Which of the following is not a common challenge associated with Big Data? a) Data Quality b) Data Integration c) Data Privacy d) Data Duplication	[]
7	Which of the following is a technique used to extract meaningful insights from data sets that are too large or complex to be processed by traditional data processing tools? a) Business Intelligence b) Machine Learning c) Artificial Intelligence d) Data Science	[]
8	What is the process of storing and managing data in a way that allows for efficient retrieval and analysis? a) Data Warehousing b) Data Mining c) Data Integration d) Data Processing	[]
9	Which of the following is a common programming language used for Big Data processing? a) C++ b) Java c) Python d) All of the above	[]
10	What is the process of combining data from multiple sources into a single, unified view? a) Data Mining b) Data Warehousing c) Data Integration d) Data Processing	[]
11	What is the process of cleaning and transforming data before it is used for analysis? a) Data Mining b) Data Warehousing c) Data Integration d) Data Preprocessing	[]
12	Data in ____ bytes size is called big data a)Meta b) Giga c)Tera d)Peta	[]
13	Transaction of data of the bank is a type of. a)Unstructured data b)Structured data c)Both a and b d)None of the above	[]
14	_____ is a collection of data that is used in volume, yet growing exponentially with time a) Big Database b)Big DBMS c)Big Datafile d)Big Data	[]
15	HDFS is a _____ system a) distributed processing b) Distributed Storage c) GFS d) AFS	[]

UNIT-II

1	Which of the following is a popular NoSQL database used for Big Data processing? a) MySQL b) PostgreSQL c) Oracle d) MongoDB	[]
2	Which of the following is a popular open-source platform used for real-time data processing and analytics? a) Apache Kafka b) Apache Hadoop c) Apache Spark d) Apache Storm	[]
3	Which of the following is not a common data storage technology used for Big Data processing? a) Hadoop Distributed File System (HDFS) b) Cassandra c) MySQL d) Amazon S3	[]
4	Which of the following is not a common type of machine learning algorithm? a) Supervised Learning b) Unsupervised Learning c) Reinforcement Learning d) Decision Learning	[]
5	_____ characteristic of data the term, 'Big Data' dealing with its retention. a)velocity b)volume c)Varity c)volatility	[]
6	What is the major advantages of storing data in block size 128MB. What is the major advantages of storing data in block size 128MB a)It saves disk seek time b)It saves disk processing time c)It saves disk access time d)It saves disk latency time	[]
7	HDFS performs replication, although it results in data redundancy? a)True b)False	[]
8	Which of the following is the Single Point of Failure a) NameNode b)Secondary NameNode c)DataNode d)None	[]
9	NameNode tries to keep the first copy of data nearest to the client machine a) Always true b) Always false c)True if the client machine is the part of the cluster d)True if the client machine is not the part of the cluster	[]
10	Which of the following is not a strong feature for nosql databases? Scalability Relational data Faster data access than RDBMS. Data easily held across multiple servers	[]
11	Which of the following is not a NoSQL database? a)Cassandra b)MongoDB c) SQL Server d)None of the above	[]
12	Which of the following are the simplest NoSQL databases? a)Key-value Stores b)Document Databases c)Wide-column d)All of the above	[]
13	Which of the following is a strong feature for NoSQL databases? a)Higher Scalability b)Cost Effective c)Support flexible schema d)All of the above	[]
14	NoSQL can be referred to as _____ a) No SQL b) Only SQL c) Not Only SQL d) SQL Undefined	[]
15	The core principle of NoSQL is _____ a)High availability b)Low availability c)both High & Low availability d) None of above	[]
16	NoSQL databases is used mainly for handling large volumes of _____ data.	[]

	a)Structured b) unstructured c)semi-structured d) None of above	
17	NoSQL databases are increasingly used in _____ a)Big data b)Real-time web applications c)Both Big data & Real-time web applications d)None of the above	[]
18	Which of the following is a document-oriented NoSQL Database? a)CouchDB b)OrientDB c)FlockDB d)Hypertable	[]
19	This command is used to list all the files. a) hdfs dfs -ls b)hdfs dfs -list c)hdfs dfs -R d) hdfs dfs -ls	[]
20	This command is used to display content of a file. a) hdfs dfs -ls b)hdfs dfs -list c) hdfs dfs -cat d) hdfs dfs -ls	[]

UNIT-III

1	Identify among the options below which is general-purpose computing model and runtime system for Distributed Data Analytics. a) HDFS b)MapReduce c)Oozie d)All of the above	[]
2	Mapper class is a) Static type b)Generic type c)Abstract type d) Final	[]
3	Fixed-size pieces of MapReduce job is known as _____ a)Splits b)Tasks c) Maps d)Records	[]
4	The output of map tasks is written in? a) Local disk b) File system c) HDFS d) Secondary storage	[]
5	Hadoop Common Package contains? a) msi files b) war files c) exe files d)jar files	[]
6	Identify the slave node among the following. a)Job node b)Data node c)Task node d)Name node	[]
7	Identify the node which acts as a checkpoint node in HDFS. a)Secondary Name node b) Secondary data node c) Name node d)Data node	[]
8	What is the minimum amount of data that a disk can read or write in HDFS? a) Byte size b)Block size c)Heap d)None of the above	[]
9	Which of the following is the daemon of Hadoop? a)NameNode b) Node manager c) DataNode d) All of the above	[]
10	Which of the following property gets configured on mapred-site.xml ? a)Replication factor b)Java Environment variables c)Directory names to store hdfs files d)Which framework to use with MapReduce jobs	[]
11	Which statement is false about Hadoop: a) It runs with commodity hardware b)It is a part of the Apache project sponsored by the ASF c)It is best for real-time streaming of data d)It can handle any type of data	[]
12	Which of the following is true about Hadoop? a) Hadoop MapReduce processes the data at very low latency b) Hadoop can't handle structured data c)Hadoop is schema-less – we don't need to define a schema before writing the data d)None of the above	[]

13	Which Hadoop configuration file contains the setting for HDFS daemons? a) yarn-site.xml b) hdfs-site.xml c) mapred-site.xml d) None of the above	[]
14	In HDFS, data node sends frequent heartbeats to name node a)True b) False	[]
15	Clients connect to _____ for I/O a) NameNode b) DataNode c)Secondary Node d)None of the Above	[]