TELANGANA UNIVERSITY
S.S.R. DEGREE COLLEGE, NIZAMABAD (C.C:5029)
VI SEMESTER INTERNAL ASSESSMENT II EXAMINATIONS
DATA SCIENCE (BIG DATA) QUESTION BANK

----

1. A _____ node acts as the Slave and is responsible for executing a Task assigned to it by the
JobTracker.   [c   ]
a) MapReduce
b) Mapper
c) TaskTracker
d) JobTracker

2. Point out the correct statement.   [a   ]
a) MapReduce tries to place the data and the compute as close as possible
b) Map Task in MapReduce is performed using the Mapper() function
c) Reduce Task in MapReduce is performed using the Map() function
d) All of the mentioned

3. _____ part of the MapReduce is responsible for processing one or more chunks of data and
producing the output results.   [a   ]
a) Maptask
b) Mapper
c) Task execution
d) All of the mentioned

4. _____ function is responsible for consolidating the results produced by each of the Map()
functions/tasks.   [a   ]
a) Reduce
b) Map
c) Reducer
d) All of the mentioned

advertisement

5. Point out the wrong statement.   [d   ]
a) A MapReduce job usually splits the input data-set into independent chunks which are processed by
the map tasks in a completely parallel manner
b) The MapReduce framework operates exclusively on <key, value> pairs
c) Applications typically implement the Mapper and Reducer interfaces to provide the map and
reduce methods
d) None of the mentioned

6. Although the Hadoop framework is implemented in Java, MapReduce applications need not be written in _____   [a   ]
a) Java
b) C
c) C#
d) None of the mentioned

7. _____ is a utility which allows users to create and run jobs with any executables as the mapper and/or the reducer.   [b   ]
a) HadoopStrdata
b) Hadoop Streaming
c) Hadoop Stream
d) None of the mentioned

8. _____ maps input key/value pairs to a set of intermediate key/value pairs.   [a   ]
a) Mapper
b) Reducer
c) Both Mapper and Reducer
d) None of the mentioned

9. The number of maps is usually driven by the total size of _____   [a   ]
a) inputs
b) outputs
c) tasks
d) None of the mentioned

10. _____ is the default Partitioner for partitioning key space.   [c   ]
a) HashPar
b) Partitioner
c) HashPartitioner
d) None of the mentioned

11. Running a _____ program involves running mapping tasks on many or all of the nodes in our cluster.   [a   ]
a) MapReduce
b) Map
c) Reducer
d) All of the mentioned

12. Mapper implementations are passed the JobConf for the job via the _____ method.
[b   ]
a) JobConfigure.configure
b) JobConfigurable.configure
c) JobConfigurable.configurable
d) None of the mentioned

13. Point out the correct statement.   [d   ]
a) Applications can use the Reporter to report progress
b) The HadoopMapReduce framework spawns one map task for each InputSplit generated by the InputFormat for the job
c) The intermediate, sorted outputs are always stored in a simple (key-len, key, value-len, value) format
d) All of the mentioned

14. Input to the _____ is the sorted output of the mappers.   [a   ]
a) Reducer
b) Mapper
c) Shuffle
d) All of the mentioned

15. The right number of reduces seems to be _____   [d   ]
a) 0.90
b) 0.80
c) 0.36
d) 0.95

16. Point out the wrong statement.   [a   ]
a) Reducer has 2 primary phases
b) Increasing the number of reduces increases the framework overhead, but increases load balancing and lowers the cost of failures
c) It is legal to set the number of reduce-tasks to zero if no reduction is desired
d) The framework groups Reducer inputs by keys (since different mappers may have output the same key) in the sort stage

17. The output of the _____ is not sorted in the Mapreduce framework for Hadoop.[d   ]
a) Mapper
b) Cascader
c) Scalding
d) None of the mentioned

18. Which of the following phases occur simultaneously?   [a   ]
a) Shuffle and Sort
b) Reduce and Sort
c) Shuffle and Map
d) All of the mentioned

19. Mapper and Reducer implementations can use the _____ to report progress or just indicate that
they are alive.   [c   ]
a) Partitioner
b) OutputCollector
c) Reporter
d) All of the mentioned

20. _____ is a generalization of the facility provided by the MapReduce framework to collect data
output by the Mapper or the Reducer.   [b   ]
a) Partitioner
b) OutputCollector
c) Reporter
d) All of the mentioned

21) The _____infrastructure of a big data is based on a distributed computing model.

Ans. physical

22) Security infrastructure refers the data about your constituents needs to be protected to

_____.   Ans. Personal information

23) Reporting and visualization enables_____ Ans. data-driven strategies

24) The significance of _____is to provide information about a dataset's characteristics

and structure. Ans.Bigdata

25) MongoDB support cross platform and is written in _____ language. Ans. C++ language

26)_____is also known as Hosted Hypervisor Ans.virtual machine monitor

27) 27MongoDB is a _____ database. Ans. Document

28)_____has the world's largest Hadoop Cluster   Ans. Facebook

29) Facebook Tackles Big Data with _____ based on HadoopAns.Project Prism

30) Hadoop named after _____   Ans. Apache Hadoop

31)_____ property of window sets or returns the fact in the status bar of a window.
Ans. Status

32) for-loop has a combination of ____, ____ and ____ in single statement.
Ans. initialization, condition, iteration

33) A JavaScript object is an entity having ___ and ___.
Ans. state, behavior

34) ____ method of window object display the alert box containing message with OK button.
Ans. alert ()

35) An object can group data together with ____ needed to manipulate it.
Ans. functions

36)JavaScript supports 2 types of objects; ____ and ___objects.
Ans. built-in, user defined

37)Math, String, Array, Date are examples of ____objects.
Ans. built in

38)When an HTML document is loaded into a ___, it becomes a document object.
Ans. web browser

39)____ defines logical structure of document.
Ans. Document object

40)____ method of window object calls a function or evaluates an expression after a specified number of milliseconds.
Ans.setTimeout ()

41. What is big data? Why is it important?

Big data is a large set of data that cannot be managed by normal software. It comprises audio, text, video, websites, and multimedia content. Big data is important because it helps make informed decisions, improves the efficiency of operations, and predicts risks and failures even before they arise.

42. Can you explain the 5 Vs of big data?

The five Vs of Big Data are:

Volume: Amount of data stored in a data warehouse.

- Velocity: It's the speed at which data is produced in real-time.

- Variety: Big data consists of a variety of data sets, like structured, semi-structured, and unstructured data.

- Veracity: The reliability or the quality of data.

- Value: Raw data is useless for any organization, but once it is transformed into valuable insights, its value increases for any organization.

43. What are the differences between big data and traditional data processing systems?

Traditional data processing systems are designed for structured data and operate within defined limits. In contrast, big data systems handle large amounts of both structured and unstructured data, leveraging distributed computing and storage for scalability.

44. How does big data drive decision-making in modern businesses?

Big data helps in decision-making by providing actionable insights from large datasets. It enables data-driven strategies and predictive analytics and enhances the understanding of customer behavior, market trends, and operational efficiency.

45. What are some common challenges faced in big data analysis?

Challenges include managing data volume, velocity, and variety, ensuring data quality, addressing security concerns, handling real-time processing, and dealing with the complexities of distributed computing environments.

46. How do big data and data analytics differ?

Big data processes large datasets, while data analytics focuses on extracting insights from data. Big data includes storage and processing, while data analytics focuses on statistical analysis.

47. Can you name various big data technologies and platforms?

Some big data technologies include:

- Hadoop

- Apache Spark

- Apache Flink

- NoSQL databases (e.g., MongoDB)

The popular platforms are Apache HBase and Apache Kafka.

48. How is data privacy managed in big data?

Data privacy is managed through encryption, access controls, anonymization techniques, and compliance with regulations such as GDPR. Privacy-preserving methods like differential privacy are also employed.

49. What role does big data play in AI and ML?

Big data provides the vast datasets needed for training machine learning models. It enhances AI capabilities by enabling deep learning algorithms to analyze large volumes of data.

50. How does big data impact cloud computing?

Big data impacts cloud computing by offering storage and processing capabilities. Cloud platforms like AWS, Azure, and Google Cloud offer big data services.