----

1. As companies move past the experimental phase with Hadoop, many cite the need for additional capabilities, including _____   [d   ]
a) Improved data storage and information retrieval
b) Improved extract, transform and load features for data integration
c) Improved data warehousing functionality
d) Improved security, workload management, and SQL support

2. Point out the correct statement.   [b   ]
a) Hadoop do need specialized hardware to process the data
b) Hadoop 2.0 allows live stream processing of real-time data
c) In the Hadoop programming framework output files are divided into lines or records
d) None of the mentioned

3. According to analysts, for what can traditional IT systems provide a foundation when they're integrated with big data technologies like Hadoop?   [a   ]
a) Big data management and data mining
b) Data warehousing and business intelligence
c) Management of Hadoop clusters
d) Collecting and storing unstructured data

4. Hadoop is a framework that works with a variety of related tools. Common cohorts include _____   [a   ]
a) MapReduce, Hive and HBase
b) MapReduce, MySQL and Google Apps
c) MapReduce, Hummer and Iguana
d) MapReduce, Heron and Trumpet

5. Point out the wrong statement.   [c   ]
a) Hardtop processing capabilities are huge and its real advantage lies in the ability to process terabytes & petabytes of data
b) Hadoop uses a programming model called "MapReduce", all the programs should conform to this model in order to work on the Hadoop platform
c) The programming model, MapReduce, used by Hadoop is difficult to write and test
d) All of the mentioned

6. What was Hadoop named after?   [c   ]
a) Creator Doug Cutting's favorite circus act
b) Cutting's high school rock band
c) The toy elephant of Cutting's son
d) A sound Cutting's laptop made during Hadoop development

7. All of the following accurately describe Hadoop, EXCEPT _____   [b   ]
a) Open-source
b) Real-time
c) Java-based
d) Distributed computing approach

8. _____ can best be described as a programming model used to develop Hadoop-based applications that can process massive amounts of data.   [a   ]
a) MapReduce
b) Mahout
c) Oozie
d) All of the mentioned

9. _____ has the world's largest Hadoop cluster.   [c   ]
a) Apple
b) Datamatics
c) Facebook
d) None of the mentioned

10. Facebook Tackles Big Data With _____ based on Hadoop.   [a   ]
a) 'Project Prism'
b) 'Prism'
c) 'Project Big'
d) 'Project Data'

11) _____ is a platform for constructing data flows for extract, transform, and load (ETL) processing and analysis of large datasets.   [c   ]

a) Pig Latin
b) Oozie
c) Pig
d) Hive

12) Point out the correct statement.   [a   ]

a) Hive is not a relational database, but a query engine that supports the parts of SQL specific to querying data
b) Hive is a relational database with SQL support
c) Pig is a relational database with SQL support
d) All of the mentioned

13) _____ hides the limitations of Java behind a powerful and concise Clojure API for Cascading.[c]

a) Scalding
b) HCatalog
c) Cascalog
d) All of the mentioned

14) Hive also support custom extensions written in _____   [b   ]
a) C#
b) Java
c) C
d) C++

15) Point out the wrong statement.   [a   ]
a) Elastic MapReduce (EMR) is Facebook's packaged Hadoop offering
b) Amazon Web Service Elastic MapReduce (EMR) is Amazon's packaged Hadoop offering
c) Scalding is a Scala API on top of Cascading that removes most Java boilerplate
d) All of the mentioned

16) _____ is the most popular high-level Java API in Hadoop Ecosystem   [d   ]
a) Scalding
b) HCatalog
c) Cascalog
d) Cascading

17) _____ is general-purpose computing model and runtime system for distributed data analytics.
                                                                                    [a   ]
a) Mapreduce
b) Drill
c) Oozie
d) None of the mentioned

18) The Pig Latin scripting language is not only a higher-level data flow language but also has operators
similar to _____   [a   ]
a) SQL
b) JSON
c) XML
d) All of the mentioned

19) _____ jobs are optimized for scalability but not latency.   [d   ]
a) Mapreduce
b) Drill
c) Oozie
d) Hive

20) _____ is a framework for performing remote procedure calls and data serialization.[c   ]
a) Drill
b) BigTop
c) Avro
d) Chukwa

21) MongoDB is a _____ database.Ans. Document

22) _____has the world's largest Hadoop ClusterAns. Facebook

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

24) Hadoop named after _____Ans. Apache Hadoop

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

26) _____is also known as Hosted HypervisorAns.virtual machine monitor

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

physical

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

_____.Ans. Personal information

29) Reporting and visualization enables_____Ans. **data-driven strategies**

30) The significance of _____is to provide information about a dataset's characteristics and structure.Ans.Bigdata

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 JS 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)Js 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 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. Define Hadoop and its components.

Hadoop is an open-source framework. It is based on Java. It manages the storage and processing of large amounts of data for applications. The elements of Hadoop are:

- HDFS

- MapReduce

- YARN

- Hadoop Common

42. What is MapReduce?

MapReduce is a model for processing and creating big data across a distributed system.

43. What is HDFS? How does it work?

HDFS is the storage component of Hadoop and handles large files by distributing them.

44. Can you describe data serialization in big data?

Data serialization is the process of converting an object into a stream of bytes. It helps save or transmit more easily.

45. What is a distributed file system?

Distributed File System or DFS is a service that allows an organization server to save files distributed on multiple file servers or locations. It enhances accessibility, fault tolerance, and scalability rather than relying on a single centralized file server.

46. What are Apache Pig's basic operations?

Apache Pig is a high-level platform for analyzing and processing large datasets. Its primary operations are loading, filtering, transforming, and storing data.

47. Explain NoSQL databases in the context of big data.

NoSQL is a database infrastructure suitable for the heavy demands of big data.

48. What is a data warehouse?

A data warehouse is a repository wherein structured data is stored and managed. This enterprise system helps analyze and report structured and semi-structured data from various sources.

49. How does a columnar database work?

A columnar database organizes data by columns rather than rows, offering advantages in terms of storage efficiency and query performance.

50. What is Apache Hive? How is it used?

Apache Hive is a data warehouse infrastructure. It provides a SQL-like language (HiveQL) for querying and managing large datasets.