

1.

Which language is used to retrieve data from text files in 'acquiring data' step in data science process?

Single choice.

(0.5 Points)

Structure query language

Scripting language

Both structure query language and scripting language

Assembly level language

2.

_____ is the simplest operation applied to node networks to get best possible route from base location to target location .

Single choice.

(0.5 Points)

optimal path

shortest path query

path query

Euclidean distance function

3.

Which one of the following is not an axes of data variety?

Single choice.

(0.5 Points)

Structural variety

Availability variety

Semantic Variety

Medium Variety

4.

Which of the following is NOT a correct statement

Single choice.

(0.5 Points)

Streaming data management systems cannot be separated from real-time processing of data.

A stream is defined as a possibly unbounded sequence of data items or records.

Each data is generally timestamped and in some cases geo-tagged.

Stream of data need be related to or correlated with each other in a data stream.

5.

The goal of big data modelling is to explore the _____

Single choice.

(0.5 Points)

Size of the data

Nature of the data

Type of the data

Volume of the data

6.

Counting number of words in a file using Map Reduce does NOT involve which of these steps

Single choice.

(0.5 Points)

- a. Map
- b. Shuffle and sort
- c. Reduce
- d. Merge and Sort

7.

Which of the following is NOT a benefit of distributed file system.

Single choice.

(0.5 Points)

High concurrency

Data scalability

Fault tolerance

High data redundancy

8.

In JSON the right format writing name and value pair is

Single choice.

(0.5 Points)

'name: value'

name='value'

name="value"

"name": "value"

9.

Veracity with respect to Big data is? (choose multiple correct answers)

Multiple choice.

(0.5 Points)

Quality

Validity

Volatility

Volume

10.

Which one of the following is the example of NOSQL databases?

Single choice.

(0.5 Points)

Neo4j

mongoDB

Cassandra

All of the above

11.

Which of the statement is correct?

Single choice.

(0.5 Points)

To compute the average we select a cell and enter =Average and then we select the data that we want to compute the average from.

To compute the average we select the data that we want to compute then in the cell we enter =Average.

To compute the average we enter =Average in new cell and then we select the data that we want to compute the average from and we select a cell.

12.

In classification process, the goal is to predict the _____ of data.

Single choice.

(0.5 Points)

Range

Numeric value

Category or class

Format

13.

To locate text data , search engine uses some form of _____ and _____ .

Single choice.

(0.5 Points)

Hide, seek model

vector model ,similarity search (NOT SURE)

scalar model ,similarity search

Retrieval model, value pair

14.

Which one of the following is NOT the component of data science

Single choice.

(0.5 Points)

People

Process

Programmability

Property

15.

Scale-out way of data scalability in big data results in _____ issue.

Single choice.

(0.5 Points)

Volume

Security

Retrieving

Storing

16.

There are many similar functions defined and used for different things. A popular similarity measure is the cosine function, which measures the cosine function of the angle between these two vectors. The intuition is that if the vectors are identical, then

Single choice.

(0.5 Points)

the angle between them is right angle.

the cosine function evaluates to zero.

the angle between them is zero.

the cosine function evaluates to odd value.

17.

In Regression process, the goal is to predict the _____ of data.

Single choice.

(0.5 Points)

Range

Numeric value

Type

Format

18.

_____ is one the spreadsheet application in OS.

Single choice.

(0.5 Points)

calc1

oocalc

ocalc

Oocalcc

19.

With Data > Filter command,

Single choice.

(0.5 Points)

we can reset and remove disappeared data from the spreadsheet.

we can only reset disappeared data from the spreadsheet.

we can only remove disappeared data from the spreadsheet.

we can filter and move disappeared data from the spreadsheet.

20.

Distributed file system works on the principle .

(i) Data replication

(ii) Data partitioning

Single choice.

(0.5 Points)

Only i

Only ii

Both i and ii

Neither i and ii

21.

Extracting a sub-array of pixels from an image is an example for _____ data operation.

Single choice.

(0.5 Points)

Single

Collections

Double

Structured

22.

_____ is one of the V's in Big data processing

Single choice.

(0.5 Points)

Validity

Volume

Versatile

Veer

23.

_____ provides numerical value to describe your data.

Single choice.

(0.5 Points)

Summary statistics

Analytical programming

Querying the data

Business data

24.

_____ are the applications which made Big data market more valuable

Single choice.

(0.5 Points)

Recommendation engines

Sentimental analysis

Mobile Advertising

All of the above

25.

The _____ is most important ingredient for successful data science program in an industry

Single choice.

(0.5 Points)

Organizational policy

Organizational buy-in

Organizational Employees

Organizational Product

26.

With the help of _____ command we can see the content of the directory.

Single choice.

(0.5 Points)

ds

ls

dir

list

27.

_____ technique is used commonly for dimensionality reduction.

Single choice.

(0.5 Points)

Principal Component Analysis (PCA)

Property Component Analysis (PCA)

Software Component Analysis (SCA)

Module Component Analysis (MCA)

28.

A common convention in many areas of Computer Science we use log to the base 2 instead of log to the base 10, when many important numbers are powers of two. In reality, log to the base two of x is the same number as log to the base ten of x times log to the base two of ten. The second number that is log to the base two of ten is a constant. So the relative score of IDF

Single choice.

(0.5 Points)

is lower when base is two.

does change because of the base we use.

is higher when base is ten.

does not change regardless of the base we use

29.

Text data is form of _____ data.

Single choice.

(0.5 Points)

structured

unstructured

relational

vectored

30.

The count of each item occurring in a frequency matrix, which is generated from document (text file) is:

Single choice.

(0.5 Points)

1

one raise to power of 2

10

ten raise to power of 2

31.

The goal of big data management is to figure out _____

Single choice.

(0.5 Points)

Infrastructure support for the data

Application support for the data

Service support of the data

Monetary support for the data

32.

There are two main JSON structure compose :

Single choice.

(0.5 Points)

name and values

arrays and objects

keys and values

none of the above

33.

Adding more memory like RAM happens in _____ scaling

Single choice.

(0.5 Points)

Vertical scaling

Horizontal scaling

Top-down scaling

Bottom-up scaling

34.

Data ingestion is the part of _____

(i) Big data management

(ii) Big data modelling

Single choice.

(0.5 Points)

Only (i)

Only (ii)

Both (i) and (ii)

Neither (i) nor (ii)

35.

The full form of JSON is _____

Single choice.

(0.5 Points)

Java Standard Output Notation

JavaScript Output Name

JavaScript Object Notation

Java Source Open Network

36.

Two techniques that led to the Big data era are ((choose two correct answers))

Multiple choice.

(0.5 Points)

Data Torrent

Cloud computing/On-demand computing

Development of industrialization

Development of new programming languages

37.

Data variety is equal to _____ in Big data.

Single choice.

(0.5 Points)

Simplicity of the data

Voluminous data

Complexity of the data

Same kind of data

38.

The particular database and the set of related programs which helps in implementation and updates in queries is called as

Single choice.

(0.5 Points)

implemented applications

backup applications

utility programs

database application

39.

What does the term NVMe stands for, which is a storage type used in Big data

Single choice.

(0.5 Points)

Non-Volatile Memory EROM

Non-Volatile Memory EPROM

Non-Volatile Memory Express

New Volatile Memory Express

40.

One of the measurement parameter used in the measuring the efficiency of data operations in big data management is

Single choice.

(0.5 Points)

Mean square error

Speed

Parallelism

Accuracy