Started on Thursday, 27 May 2021, 10:09 AM
State Finished
Completed on Thursday, 27 May 2021, 10:53 AM
Time taken 43 mins 14 secs

Question **1**
Complete
Marked out of 1.00

Open source technology belongs to_____

- A. data perfect storm
- B. convergence perfect storm
- C. computing perfect storm
- D. none

Question **2**
Complete
Marked out of 1.00

Which of the advanced analytics have insights into all the products and all the transactions

- A. Broader insights
- B. Deeper insights
- C. Frictionless actions
- D. all

Question **3**
Complete
Marked out of 1.00

A voluminous amount of structured, semi-structured and unstructured data that has the potential to be mined for information.

- A. statistical data
- B. small data
- C. meta data
- D. big data

Question **4**

Complete

Marked out of 1.00

Which is not a data source provider

- A. Internet data
- B. supply chain data
- C. research work
- D. none

Question **5**

Complete

Marked out of 1.00

In _____ insights, operating a business globally is very complex.

- A. Deeper insights
- B. all
- C. Frictionless actions
- D. Broader insights

Question **6**

Complete

Marked out of 1.00

_____characteristic of big data, is the speed at which the data is generated and processed to meet the demands and challenges that lie in the path of growth and development.

- A. variety
- B. volume
- C. velocity
- D. veracity

Question **7**

Complete

Marked out of 1.00

The____ Function – It takes a set of data and converts it into another set of data, where individual elements are broken down into tuples

- A. Yarn
- B. HDFS
- C. Reduce
- D. Map

Question **8**

Complete

Marked out of 1.00

Select the type of data that can be structured easily ?

- A. directions to the shops
- B. Profile Photo
- C. Screenshots
- D. Date of Birth

Question **9**

Complete

Marked out of 1.00

Sources of unstructured data_____

- A. webpages
- B. images and videos
- C. surveys
- D. all

Question **10**

Complete

Marked out of 1.00

In _____, increased reliability and accuracy that will allow the deeper and broader insights to be automated into systematic actions.

- A. Frictionless actions
- B. Deeper insights
- C. all options
- D. Broader insights

Question **11**

Complete

Marked out of 2.00

The branch of data mining concerned with the prediction of future probabilities and trends.

- A. Behavioral Analytics
- B. Descriptive Analytics
- C. In-memory Analytics
- D. Predictive Analytics

Question **12**

Complete

Marked out of 2.00

Competitive and marketplace data, industry reports, consumer data, business data are example of _____source.

- A. internet data
- B. secondary research
- C. location data
- D. primary research

Question **13**

Complete

Marked out of 2.00

_____guide marketing strategies and tactics in cross channel life cycle marketing approach.

- A. all
- B. conversion loop
- C. stickiness and win-back loops
- D. repurchase loop

Question **14**

Complete

Marked out of 2.00

In _____traditional data management, analytics software and hardware technologies, open-source technology and commodity hardware are merged to create new alternatives for IT.

- A. Computing perfect storm
- B. Convergence perfect storm
- C. Data perfect storm
- D. all

Question **15**

Complete

Marked out of 2.00

Vendor catalogs, pricing and quality information are examples of _____resource.

- A. internet data
- B. primary research
- C. supply chain data
- D. location data

Question **16**

Complete

Marked out of 2.00

What are the five V's of Big Data?

1. Volume: The name 'Big Data' itself is related to a size which is enormous. Volume is a huge amount of data. Years of consumer records and transactions (150 billion +records per year).

2. Velocity: Velocity refers to the high speed of accumulation of data. In Big Data velocity data flows in from sources like machines, networks, social media, mobile phones etc. Dynamic transactions and social media info.

3. Variety: It refers to nature of data that is structured, semi-structured and unstructured data. It also refers to heterogeneous sources. Social media plus other unstructured data such as customer E-mails, call center conversations as well as transactional structured data

4. Veracity: It refers to inconsistencies and uncertainty in data, that is data which is available can sometimes get messy and quality and accuracy are difficult to control.

5. Value: After having the 4 V's into account there comes one more V which stands for Value!. The bulk of Data having no Value is of no good to the company, unless you turn it into something useful.

Question **17**

Complete

Marked out of 2.00

How big data and Hadoop are related to each other.

- There are many Big Data technologies that have been making an impact on the new technology stacks for handling Big Data, but Apache Hadoop is one technology that has been the darling of Big Data talk.
- Hadoop is an open-source platform for storage and processing of diverse data types that enables data-driven enterprises to rapidly derive the complete value from all their data.
- He explained the history and overview of Hadoop to us: The original creators of Hadoop are Doug Cutting (used to be at Yahoo! now at Cloudera). Doug and Mike were building a project called "Nutch" with the goal of creating a large Web index. They saw the MapReduce and GFS papers from Google, which were obviously super relevant to the problem.
- They integrated the concepts from MapReduce and GFS into Nutch; then later these two components were pulled out to form the genesis of the Hadoop project. The name "Hadoop" itself comes from Doug 's son, he just made the word up for a yellow plush elephant toy that he has.

- Hadoop gives organizations the flexibility to ask questions across their structured and unstructured data that were previously impossible to ask or solve

- The scale and variety of data have permanently overwhelmed the ability to cost-effectively extract value using traditional platforms.
- The scalability and elasticity of free, open-source Hadoop running on standard hardware allow organizations to hold onto more data than ever before.
- Hadoop handles a variety of workloads, including search, log processing, recommendation systems, data warehousing, and video/image analysis.

- Apache Hadoop is an open-source project administered by the Apache Software Foundation. The software was originally developed by the world 's largest Internet companies to capture and analyze the data that they generate
- NowApacheHadoopisaregisteredtrademarkoftheApacheSoftwareFoundation.
- Hadoop runs applications using the MapReduce algorithm, where the data is processed in parallel on different CPU nodes.
- In short, Hadoop framework is capable enough to develop applications capable of running on clusters of computers and they could perform complete statistical analysis for a huge amounts of data.
- Hadoop is an Apache open source framework WRITTEN IN JAVA that allows distributed processing oflargedatasetsacrossclustersofcomputersusingsimpleprogrammingmodels.
- A Hadoop frame-worked application works in an environment that provides distributed storage and computation across clusters of computers.
- Hadoop is designed to scale up from single server to thousands of machines, each offering local computation and storage.

The TWO CRITICAL COMPONENTS OF HADOOP are:

- **The Hadoop Distributed File System (HDFS):** HDFS is the storage system for a Hadoop cluster.
- **MapReduce:** Because Hadoop stores the entire dataset in small pieces across a collection of servers, analytical jobs can be distributed, in parallel, to each of the servers storing part of the data

Question **18**

Complete

Marked out of 2.00

What is Name Node?

NAME NODE:

- The Hadoop Distributed File System (HDFS) is based on the Google File System (GFS) and provides a distributed file system that is designed to run on large clusters (thousands of computers)of small computer machines in a reliable, fault-tolerant manner.
- HDFS uses a master/slave architecture where master consists of a single NameNode that manages the file system metadata and one or more slave Data Nodes that store the actual data.
- AfileinanHDFSnamespaceissplitintoseveralblocksandthoseblocksarestoredinaset of DataNodes. The NameNode determines the mapping of blocks to the DataNodes. The DataNodes takes care of read and write operation with the file system. They also take care of block creation, deletion and replication based on instruction given by NameNode.
- HDFSprovidesashelllikeanyotherfilesystemandalistofcommandsareavailableto interact with the file system. These shell commands will be covered in a separate chapter along with appropriate examples.
- **Name Node:** The NameNode is the center piece of an HDFS file system. It keeps the directory tree of all files in the file system, and tracks where across the cluster the file data is kept. It does not store the data of these files itself. Client applications talk to the NameNode whenever they wish to locate a file, or when they want to add/copy/move/delete a file. The NameNode is a Single Point of Failure for the HDFS Cluster.

Question **19**

Complete

Marked out of 2.00

What is difference between big data and data analytics?

Big Data:

1.Big Data can be considered as a Big Library where all the answers to all the questions are there but difficult to find the answers to your questions.
2.Big Data is a mostly unstructured set of data that has to be sorted out to find an answer to any question, and it is not very easy to process those enormous volumes of data.
3.In Big Data, one will need to use sophisticated technological tools such as automation tools or parallel computing tools to manage the Big Data because it is not easy to process the enormous volume of Big Data.
4.Big data is used by the industries such as banking industries, retail industries and many more.

Data Analytics:

1.Data Analytics looks like a book where you can find a solution to your problems.

2.In data analytics, one will find that the data will be already structured and it is easy to find an answer to a question.

3. In Data Analytics, one will use simple tools for statistical modelling and predictive modelling because the data to analyse is already structured and not complicated.

4.Data Analytics is mainly used by industries like IT Industries, Travel Industries, and Healthcare Industries.

The fundamental difference between Big Data and Data Analytics with an example.

Nature: Data Analytics like a book where you can find a solution to your problems, on the other hand, Big Data can be considered as a Big Library where all the answers to all the questions are there but difficult to find the answers to your questions.

Structure of Data: In data analytics, one will find that the data will be already structured and it is easy to find an answer to a question. But, on the other hand, Big Data is a mostly unstructured set of data that has to be sorted out to find an answer to any question, and it is not very easy to process those enormous volumes of data. Lots of filters have to be applied to find some meaning insight into Big Data.

Tools used in Big Data vs Data Analytics: In Data Analytics, one will use simple tools for statistical modelling and predictive modelling because the data to analyze is already structured and not complicated. In Big Data, one will need to use sophisticated technological tools such as automation tools or parallel computing tools to manage the Big Data because it is not easy to process the enormous volume of Big Data. More about Big Data Tools.

Type of Industry using Big Data and Data Analytics:

Data Analytics is mainly used by industries like IT Industries, Travel Industries, and Healthcare Industries. Data Analytics helps these industries to create new developments which are done by using historical data and analyzing past trends & patterns. Whereas, Big Data is used by industries such as banking industries, retail industries and many more. Big Data helps these industries in many ways to take some strategic business decisions.

Question **20**

Complete

Marked out of 2.00

What is Descriptive analytics?

Descriptive Analytics:

- Descriptive analytics is the interpretation of historical data to better understand changes that have occurred in a business.
- It describes the use of a range of historic data to draw comparisons.
- Most commonly reported financial metrics are a product of descriptive analytics, for example, year-over-year pricing changes, month-over-month sales growth, the number of users, or the total revenue per subscriber.
- It is a field of statistics that focuses on gathering and summarizing raw data to be easily interpreted.
- It concentrate on historical data, providing the context that is vital for understanding information and numbers.
- It ask about the past. They want to know what has been happening to the business and how this is likely to affect future sales.

Examples of descriptive analytics

Summarizing past events such as sales and operations data or marketing campaigns. Social media usage and engagement data such as Instagram or Facebook likes. Reporting general trends. Collating survey results

◄ Sample Quiz Activity

Jump to...