Quick Revise

Multiple-Choice Questions

Q1. Which of the following is not a characteristic of Big Data?

a. Volume
b. Variability
c. Variety
d. Velocity

Ans. The correct option is b.

Q2. Who among the following do you think would be able to deal with the growing number of data sources efficiently?

a. Business developer
b. Data scientist
c. Sales executive
d. Web designer

Ans. The correct option is b.

Q3. Which one of the following is not an example of external datasources?

a. Data from CRM
b. Data from Web logs
c. Data from government sources
d. Data from market surveys

Ans. The correct option is a.

Q4. Which of the following does not belong to the traditional database technology?

a. RDBMS
b. DBMS
c. Flat files
d. NoSQL

Ans. The correct option is d.

Q5. If a Big Data analyst were to analyze data from a database of call logs provided by a telecom service provider, which element of Big Data would he be dealing with?

a. Volume
b. Variable
c. Variety
d. Velocity

Ans. The correct option is a.

Q6. Some people call this data as "structured but not relational." Which data are we talking about?

a. Structured data
b. Unstructured data
c. Semi-structured data
d. Mixed data

Ans. The correct option is c.

Q7. The data generated from a GPS satellite and Web logs is classified as _____.

a. Structured data
b. Unstructured data
c. Both structured and unstructured data
d. Semi-structured data

Ans. The correct option is d.

Q8. The data being captured can be in any form or structure. Which characteristic of Big Dat are we talking about?

a. Volume
b. Velocity
c. Variety
d. Value

Ans. The correct option is c.

Subjective Questions

Q1. List and discuss the four elements of Big Data.

Ans. Big Data primarily consists of the following four elements:

- **Volume**—Volume is the amount of data generated by organizations or individua Today, the volume of data in most organizations is approaching around exabytes. So experts predict the volume of data to reach zetabytes in the coming years. Organizatio are doing their best to handle this ever-increasing volume of data. For example, Goo Inc. processes around 20 petabytes of data, and Twitter feeds generate around 8 teraby of data every day.

- **Velocity**—Velocity describes the rate at which data is generated, captured, and shared. Enterprises can capitalize on data only if it is captured and shared in real time. Information processing systems such as CRM and Enterprise Resource Planning (ERP) face problems associated with data, which keeps adding up but cannot be processed quickly. These systems are able to attend data in batches every few hours; however, even this time lag causes the data to lose its importance as new data is constantly being generated. For example, eBay analyzes around 5 million transactions per day in real time to detect and prevent frauds arising from the use of PayPal.
- **Variety**—We all know that data is being generated at a very fast pace. Now, this data is generated from different types of sources, such as internal, external, social, and behavioral, and comes in different formats, such as images, text, videos, etc. Even a single source can generate data in varied formats; for example, GPS and social networking sites, such as Facebook, produce data of all types, including text, images, videos, etc.
- **Veracity**—Veracity generally refers to the uncertainty of data, i.e., whether the obtained data is correct or consistent. Out of the huge amount of data that is generated in almost every process, only the data that is correct and consistent can be used for further analysis. Data when processed becomes information; however, a lot of effort goes in processing the data. Big Data, especially in the unstructured and semi-structured forms, is messy in nature, and it takes a good amount of time and expertise to clean that data and make it suitable for analysis.

Q2. As an HR manager of a company providing Big Data solutions to clients, what characteristics would you look for while recruiting a potential candidate for the position of a data analyst?

Ans. A Big Data analyst should be a well-trained professional who is able to collect data from different sources, organize it in a suitable form, and analyze it to generate the desired results. A Big Data analyst should have the following technical and soft skills:

Technical Skills:

- Understanding of Hadoop, Hive, and MapReduce
- Knowledge of natural language processing
- Knowledge of statistical analysis and analytical tools
- Knowledge of conceptual and predictive modeling

Soft Skills:

- Strong written and verbal communication skills
- Analytical ability
- Basic understanding of how a business works

Q3. You are planning the marketing strategy for a new product in your company. Identify and list some limitations of structured data related to this work.

Ans. Structured data has certain limitations associated with it, when it comes to product marketing and advertising.

Some of these limitations are:

- Marketing strategies do not provide space for predefined rules
- There is hardly any definite relation between product sales and marketing strategies
- Structured data cannot find complex correlation patterns
- Behavioral analysis is not possible from structured data