

1. What does big data analytics mean? (15 Marks)

The term Big data analytics refers to the strategy of analyzing large volumes of data, or big data. The large amount of data which gathered from a wide variety of sources, including social networks, videos, digital images, sensors, and sales transaction records is called Big Data. The main purpose in analyzing all this data is to uncover patterns and connections that might otherwise be invisible, and that might provide valuable insights about the users who created it. Through this insight, businesses may be able to gain an edge over their rivals and make superior business decisions

2. Why is big data analytics important? (5 Marks)

Most important advantage of Big Data analysis is, it helps organizations harness their data and use it to identify new opportunities. With the help of this, companies lead to smarter business moves, more efficient operations, higher profits, and happier customers.

3. List some tools used for big data (20 Marks)

There are various tools in Big Data technology which are deployed for importing, sorting, and analyzing data. List of some tools are as follows:

- Apache Hive
- Apache Spark
- MongoDB
- MapReduce
- Apache Sqoop
- Cassandra
- Apache Flume
- Apache Pig
- Apache Splunk
- Apache Hadoop

4. What is data cleansing? (5 Marks)

Data cleansing it is also known as Data scrubbing, it is a process of removing data which incorrect, duplicated or corrupted. This process is used for enhancing the data quality by eliminating errors and irregularities.

5. What is the K-mean algorithm? (5 Marks)

K-mean is a partitioning technique in which objects are categorized into K groups. In this algorithm, the clusters are spherical with the data points aligned around that cluster, and the variance of the clusters is similar to one another.

6. What is the difference between data mining and data analysis? (10 Marks)

Data mining:

- A hypothesis is not required in Data Mining

- Data mining demands clean and well-documented data
- Results of Data mining are not easy to interpret
- Data mining algorithms automatically develop an equation

Data Analysis:

- Data analysis begins with a hypothesis
- Data analysis involves data cleaning, therefore, it does not require clean and well-documented data.
- Data analysts interpret results and present it to the stakeholders
- In Data analysis we have to develop own equations

7. What are the most common analytical technique categories? **(10 Marks)**

Most of the widely used analytical techniques falls into one of the following categories:

- Statistical methods
- Forecasting
- Regression analysis
- Database querying
- Data warehouse
- Machine learning and data mining

8. What is machine learning? **(10 Marks)**

Machine learning is a category of an algorithm that helps software applications to become more accurate in predicting outcomes without being explicitly programmed. The basic concept of machine learning is to build algorithms that can receive input data and use statistical analysis to predict an output while updating outputs as new data becomes available.

It enables the computers or the machines to make data-driven decisions rather than being explicitly programmed for carrying out a certain task.

9. What is the difference between data mining and data profiling? **(10 Marks)**

The main difference between data mining and data profiling is as follows:

- **Data profiling:** It targets the instant analysis of individual attributes like price vary, distinct price and their frequency, an incidence of null values, data type, length, etc.
- **Data mining:** It focuses on dependencies, sequence discovery, relation holding between several attributes, cluster analysis, detection of unusual records etc.

10. What are an Eigenvalue and Eigenvector? **(10 Marks)**

These both the values are used for understanding linear transformations. In data analysis, we usually calculate the eigenvectors for a correlation or covariance matrix. Eigenvectors are nothing but the directions along which a particular linear transformation acts by flipping, compressing or stretching. Eigenvalue can be referred to as the strength of the transformation in the direction of eigenvector or the factor by which the compression occurs.