1. Big Data with examples and type

Big Data: Vast and complex data sets that are hard to process and evaluate using traditional processing techniques are known to as "big data." These data sets could include a mixture of structured, unstructured, and semi-structured data types.

Examples:

1. Social media data: Sites like Facebook and Twitter produce huge amounts of information. This information, which is often unstructured, can contain user profiles, posts, comments, likes, likes, and the pages they follow.
2. Financial data: Financial firms produce and preserve enormous volumes of data, such as customer, market, and transactional data and financial analysis. Usually, this data is structured.
3. Medical Data: Medical imaging data, electronic health records of patients, and other healthcare-related data produce vast datasets. Typically, this data is structured or semi-structured.

Types of Big Data:

1. Structured Data: Data that is organized and follows a set format, such as that found in a database or spreadsheet is called structured data. It is simple to process and interpret this data.
2. Unstructured data: Text data in emails, blogs, social media posts, and customer reviews are examples of data that does not have a structured syntax and are thus known to as unstructured data. It’s hard to process this type of data with traditional tools.
3. Semi-structured data: Data that contains some structure, like data in XML or JSON formats, is called semi-structured data. Both traditional and modern tools can be used for processing this data.

2. 6 ‘V’s of Big Data (define each)

The 6 'V's of Big Data refer to the six characteristics that define large and complex data sets. They are:

1. Volume: The amount of information being produced, gathered, and stored for analysis is called data volume. The amount of data is increasing significantly because of emerging digital technology, making it more difficult to store, process, and analyze.
2. Velocity: The rate at which data is produced and handled is referred to as the velocity of data. Organizations must be able to process data quickly to get insights about its functioning and the results they are producing. As a result, data is generated and transmitted at a much faster rate.
3. Variety: When we talk about data, we're talking about the various forms and formats that are being produced, such as text, photos, audio, and video. With the advent of social media, it makes it even harder to process this data from millions of users.
4. Veracity: Data quality and accuracy are referred to as its veracity. Data inconsistencies and inaccuracies are more likely to increase as big data usage grows. Making informed decisions requires making sure the information is reliable.
5. Value: The usefulness/worthiness of the data is referred to as the data's value. The understanding and information that may be gained through analyzing data determine its worth. The ability of the organization to effectively handle and evaluate the data determines the value of the data.
6. Variability: Data inconsistency and irrational behavior are referred to as variability. Data can come from a variety of sources and have various formats, qualities, and structures.

3. Phases of Big Data analysis (discuss each)

Some of the crucial stages of big data analysis include:

1. Data Acquisition: Gathering and obtaining the data is the initial step in a Big Data analysis. Social media, detectors, and other equipment are just a few examples of the sources from which data can be gathered. Organizations should make sure that the data they gather is of good quality and reliability.
2. Data Cleaning: Following collection, the data must be cleaned and made ready for analysis. Data cleaning includes removing any repeated data, inconsistencies, and errors just to make it easier for processing.
3. Data Storage: Organizations must adopt specialized storage systems like Hadoop or NoSQL databases since big data is generally too massive to be kept in traditional database management systems.
4. Data processing: it includes using procedures and tools to obtain useful information from given datasets. It includes methods like statistical analysis, machine learning, and data mining.
5. Data visualization: After processing, the data must be shown in a way that is simple to explain and use. Charts, graphs, histograms, and dashboards are examples of data visualization tools that are used to make it easier for clients to understand.
6. Decision Making: This is the final phase of big data analysis which involves using the insights generated from the data to make decisions. This involves developing new products in the market and improving marketing techniques for wider outreach to people. The main moto is to use the drawn insights to increase company’s value.

4. Challenges in Big Data analysis (discuss each).

Here are some of the major challenges in Big Data analysis:

1. Data Quality: Big Data originates from many different sources, and its quality might vary greatly. Improper decision-making and erroneous conclusions might result from poor data quality. Making sure the data is of excellent quality requires performing processes like data cleaning and pre-processing.
2. Data Integration: Big Data frequently originates from numerous sources and takes varied forms. It can be difficult to integrate and combine numerous data sources, necessitating specialized tools and methods.
3. Scalability: The ability of a process to scale up to handle bigger data quantities is essential for big data analysis, which necessitates enormous computational resources. Hardware, software, and infrastructure that are specialized are needed for this.
4. Privacy and security: Big Data frequently includes sensitive information like corporate secrets or personal information. This data must be protected from illegal access and breaches, which calls for specialized tools and approaches.
5. Analysis and Interpretation: Big Data analysis calls for specialized knowledge and abilities to ensure that the data is appropriately examined and comprehended. This can be difficult, especially as the data's volume and complexity rise.
6. Cost: The specialized technology, software, and infrastructure needed to support big data research can be extremely pricey. To make sure that big data analysis is a worthwhile investment, organizations must carefully compare the costs and benefits.

References:

1. <https://bau.edu/blog/characteristics-of-big-data/#:~:text=Unstructured%20data%20accounts%20for%20the,databases%2C%20to%20name%20a%20few>.
2. <https://www.motivaction.nl/en/actualities/news/big-data-the-6-vs-you-need-to-look-at-for-important-insights#:~:text=Big%20data%20is%20best%20described,%2C%20value%2C%20veracity%20and%20variability>.
3. Lecure2 ppt
4. <https://intellipaat.com/blog/tutorial/data-analytics-tutorial/data-analytics-lifecycle/>
5. <https://www.upgrad.com/blog/major-challenges-of-big-data/#:~:text=This%20data%20needs%20to%20be,accumulating%20data%20from%20different%20sources>.