Q1. Explain data structure of big data.

Ans. Big data is made up of structured, semi-structured, and unstructured data, and is organized in a collection of data values, relationships and functions.

1. Structured data:

- It's the data which follows a pre-defined format and thus, is straightforward to analyze.
- It conforms to a tabular format together with relationships between different rows and columns.
- You can think of SQL databases as a common example.
- Structured data relies on how data could be stored, processed as well as, accessed.

2. Unstructured data:

- This type of big data comes with unknown form and cannot be stored in traditional ways and cannot be analyzed unless it's transformed into a structured format.
- You can think of multimedia content like audios, videos, images as examples of unstructured data.

3. Semi-structured data.

- It is a type of big data that doesn't conform with a formal structure of data models.
- But it comes with some kinds of organizational tags or other markers that help to separate semantic elements, as well as, enforce hierarchies of fields and records within that data.

Q2. Explain applications of Hadoop.

Ans. Hadoop is primarily used for storing and processing large volumes of data (big data) across a cluster of computers, making it ideal for applications like data analytics, data warehousing, etc.

1. **Marketing Analytics:**
   - Analyzing customer behavior patterns from patterns from website logs, social media interactions, and purchase history to understand market trends and target campaigns effectively.

2. **Risk management:**
   - Identifying potential risks on financial data by analyzing large datasets of transactions and market trends.

3. **Scientific Research:**
   - Analyzing large datasets generated from experiments in fields like genomics, climate science, and astronomy.

4. **Log Analysis:**
   - Processing and analyzing large volumes of system logs to identify potential issues, security threats and performance bottlenecks.

Q3. Explain Hadoop Ecosystem.

Ans. Hadoop ecosystem is a platform or a suite which provides various services to solve the big data problems.

- It includes Apache projects and and various commercial tools and solutions.
- There are four major elements of Hadoop :-

1. HDFS
- HDFS is the primary or major component of Hadoop ecosystem and is responsible for storing large data sets of structured or unstructured data across various nodes and thereby maintaining the metadata in the form of log files.

2. YARN
- Yet another Resource Negotiator, as the name implies, YARN is the one who helps to manage the resources across the clusters.

3. Map Reduce
- By making the use of distributed and parallel algorithms, MapReduce makes it possible to carry over the processing's logic and helps to write applications which transform big data sets into a manageable one.

4. Hadoop Common Utilities
- Hadoop common utilities are nothing but our Java library and java files or we can say the java scripts that we need for all the other components present on a Hadoop cluster.