**Big Data**:

Data which are very large in size is called Big Data.

Normally we work on data of size MB(WordDoc ,Excel) or maximum GB(Movies, Codes) but data in Peta bytes i.e. 10^15 byte size is called Big Data.

## **3V's of Big Data**

1. **Velocity:** The data is increasing at a very fast rate. It is estimated that the volume of data will double in every 2 years.
2. **Variety:** Now a days data are not stored in rows and column. Data is structured as well as unstructured. Log file, CCTV footage is unstructured data. Data which can be saved in tables are structured data like the transaction data of the bank.
3. **Volume:** The amount of data which we deal with is of very large size of Peta bytes.

## **Issues**

Huge amount of unstructured data which needs to be stored, processed and analyzed.

## **Solution**

**Storage:** This huge amount of data, Hadoop uses HDFS (Hadoop Distributed File System) which uses commodity hardware to form clusters and store data in a distributed fashion. It works on Write once, read many times principle.

**Processing:** Map Reduce paradigm is applied to data distributed over network to find the required output.

**Analyze:** Pig, Hive can be used to analyze the data.

**Cost:** Hadoop is open source so the cost is no more an issue.