**Anuj Tewatia**

**Big Data**

**05/12/2018**

What is big data?

* **Big data is a huge volume of data, being large and complex, which cannot be saved and processed easily with traditional methods and systems. Such as RDBMS. Data coming from various sources, in structured and unstructured format can be classified in big data.**

What is Hadoop?

* **Hadoop is a set of open source procedures and program, which is used for big data operations, it consists of several modules, and each module organizes the parallel processing of big data and server processors.**

What is HDFS?

**HDFS is a one of module of Hadoop, which is known as distributed file system.  It allows data to be stored in an easily accessible format. Storing and managing the data across all clusters.**

What is MapReduce?

* **Mapreduce is also one of module of Hadoop, which process the stored data from HDFS. It processes the data parallelly in distributed environment. MapReduce contains the summary of each server.**

Why we need cloud?

* **It is not easy to perform the analysis on the big data on single server, hence most of the organizations uses the different cloud services for processing the big data analysis.**