#TheStudyCircle

Your station for Big Data Technologies

What is Big Data

- 4v of Big data : Volume, Velocity, Variety, Value
- Revolution of Cheap hardware
- Huge processing capacity at lower cost
- Data Collection
- Data Transfer
- Data processing
- Data storage
- Data Analytics (Value from collected data)

Volume

Not the case of **“BIG DATA”**

- File size is less than 300 MB
- Enough time to process this volume (more than 20 minutes)

Case of **“BIG DATA”**

- File size is more than 300 MB to some Gigabytes
- Traditional .NET or Java batch jobs are incapable of handling this volume even with parallel threads
- Need faster processing times (few seconds)

Velocity

Not the case of **“BIG DATA”**

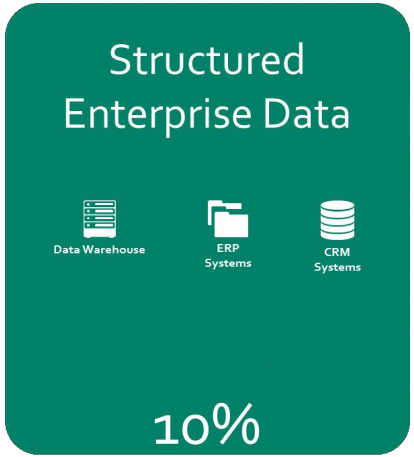
- 1-2 million requests taking over less than of 20-30 sec (maximum 1 min) is not the case of big data
- These cases can be handled with increased capacity of servers and their number

Case of **“BIG DATA”**

- More than 1-2 million requests taking over more than 1 minutes is the case of big data
- These case tend not be solved even with increased capacity or numbers.
- Time is there to introduce “BIG DATA”

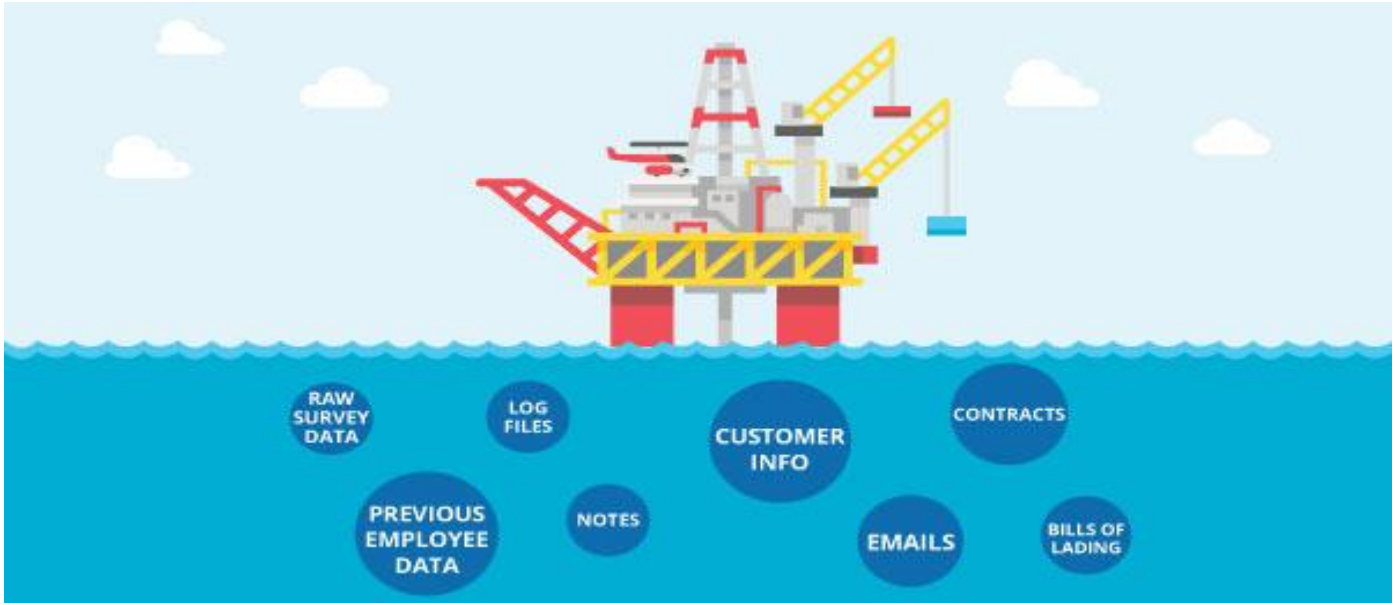
Variety in data

- Data collected in in huge volume is precondition but also of multiple **data types**
- Data is JSON, CSV, Text file, Clickstreams and even unstructured data



Value

- Data collected in in huge volume is precondition but also has some value present in it
- Data is new oil. So it need to be extracted and processed to get the value from it



Classification of application

Complexity and chances of Data growth (Type ||| being highest)

Type of Application	Indicators	Examples
Type I	1. No Capacity Issues 2. No Chances of exponential data growth 3. Less/No value in generated data	1.Reference Data Web services 2.Small time back office applications 3. Small capacity batches
Type II	1. No Capacity Issues in day to day transactions 2. No chances of exponential data growth 3. Considerable amount of **valuable** data generation (File size > 300 MB)	1. Money transfer applications. 2. Account operations 3. Account update systems 4. Operations team day to applications.
Type III	1. Capacity issue 2. Data growth is exponential 3. Extremely valuable data generated from applications operations	1. Recommendation systems 2. Aggregator services 3. Customer Insights applications 4. Production and operations monitoring applications