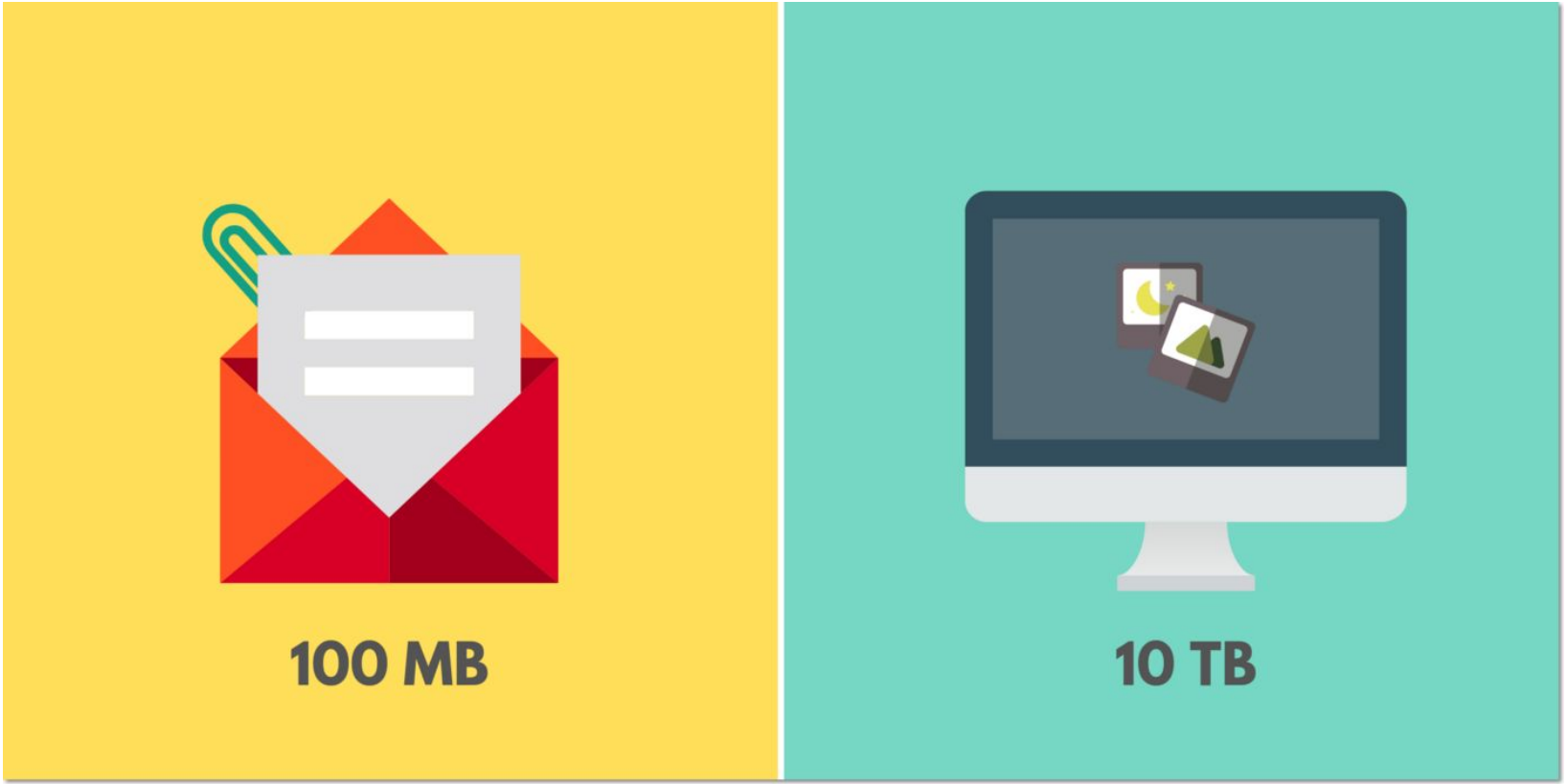

BIG DATA

What is Big Data?

Big Data refers to a huge volume of data, that cannot be stored and processed using the traditional computing approach within a given time frame.

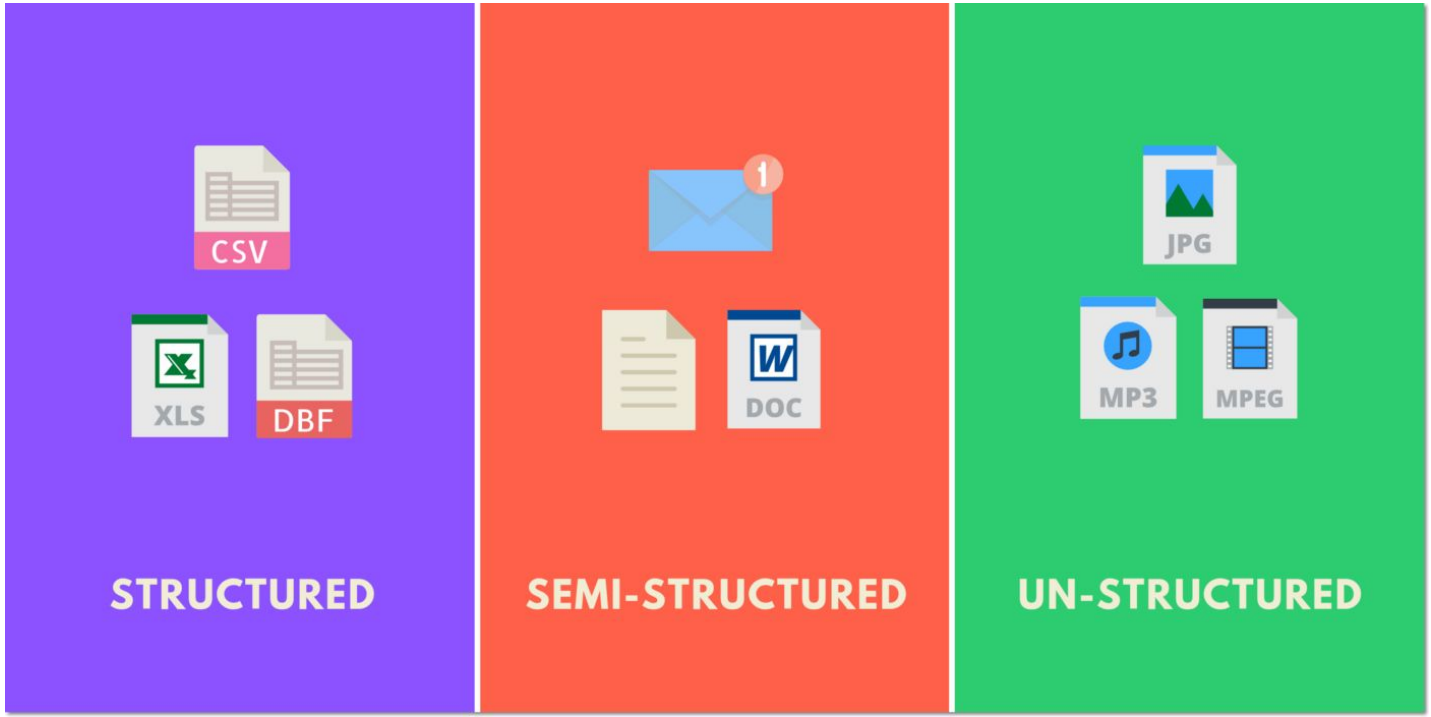
But how huge this data needs to be? To be termed as Big Data?



How Big Data is Classified ?

Big Data is classified into 3 different categories.

1. Structured Data
2. Semi-Structured Data
3. Unstructured Data



Structured Data refers to the data that has a proper structure associated with it. For example, the data that is present within the databases, the CSV files, and the excel spreadsheets can be referred to as Structured Data.

Semi-Structured Data refers to the data that does not have a proper structure associated with it. For example, the data that is present within the emails, the log files, and the word documents can be referred to as Semi-Structured Data.

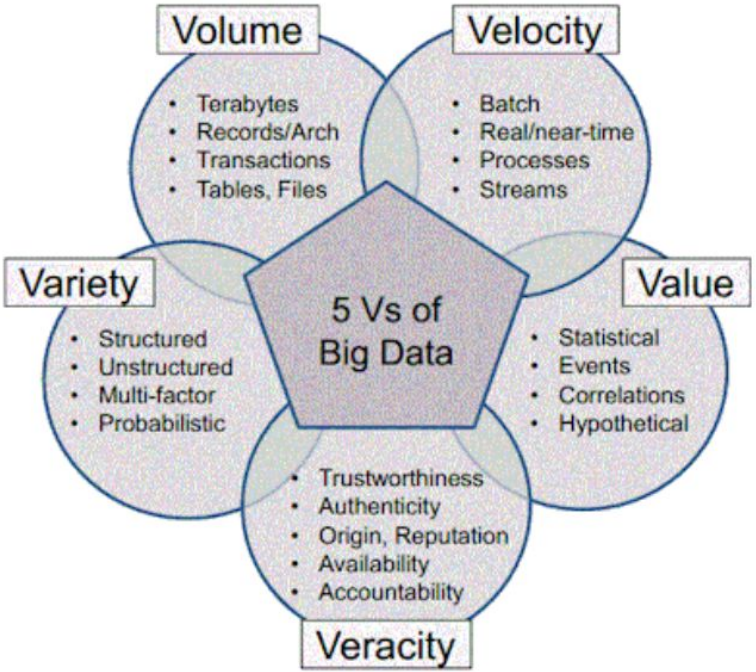
Un-Structured Data refers to the data that does not have any structure associated with it at all. For example, the image files, the audio files, and the video files can be referred to as Un-Structured Data.

Characteristics of Big Data

Big Data is categorized by 3 important characteristics.

1. Volume
2. Velocity
3. Variety


BIG DATA
Volume
Value
Veracity
Visualization
Variety
Velocity
Virality



- Volume- There is a lot of data!
- Velocity- The data rapidly accumulates. The data flow comes from social media, mobile devices, networks, and machines, in both a continuous and massive wave
- Variety - The data comes from many sources, both inside and outside of the organization
- Veracity- The data often contains inconsistencies, uncertainties, and duplication, a natural result of pulling the information from numerous and diverse sources
- Value - The data has no worth unless it can be analyzed, processed, and turned into something useful to the organization

Volume: Big data implies enormous volumes of data. It used to be employees created data. Now that data is generated by machines, networks and human interaction on systems like social media the volume of data to be analyzed is massive.

Value: Value measures the usefulness of data in making decisions. User can run certain queries against the data stored and thus can deduct important results from the filter data obtained and rank it according to the dimensions they require.

Variety: Variety refers to the many sources and types of data both structured and unstructured. We used to store data from sources like spreadsheets and databases. Now data comes in the form of emails, photos, videos, monitoring devices, PDFs, audio, etc. This variety of unstructured data creates problems for storage, mining and analyzing data.

Velocity: Big Data Velocity deals with the pace at which data flows in from sources like business processes, machines, networks and human interaction with things like social media sites, mobile devices, etc. The flow of data is massive and continuous. This real-time data can help researchers and businesses make valuable decisions that provide strategic competitive advantages and ROI if you are able to handle the velocity.

Veracity: Big Data Veracity refers to the biases, noise and abnormality in data. Veracity in data analysis is the biggest challenge when compares to things like volume and velocity.

Biggest Big Data Challenges


COLLECTION
STORAGE
NETWORK
RESEARCH
BIG DATA
CLOUD
TECHNOLOGY
ANALYSIS
VISUALIZATION
EXABITE
VOLUME

Insufficient knowledge

Expensive

Difficult to manage data quality

Security holes

Upscaling