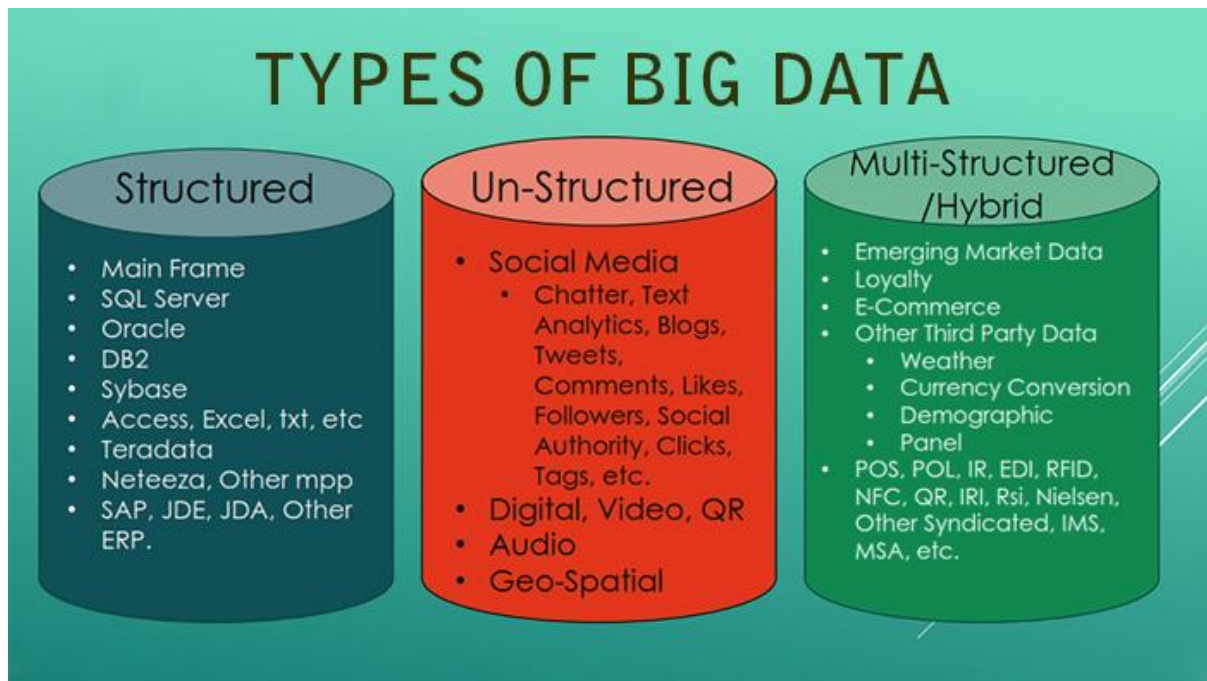


The challenges of Big Data

Big Data by its very nature is hugely challenging to work with. But the rewards of making sense of Big Data is hugely rewarding too. All Big Data can be categorized into:



- **Structured** –that which can be stored in rows and columns like relational data sets
- **Unstructured** – data that cannot be stored in rows and columns like video, images, etc.
- **Semi-structured** – data in XML that can be read by machines and human

There is a certain standardized process to work with Big Data which can be highlighted using the methodology of ETL.

ETL (Extract, Transform, Load)

Extract – getting the data from multiple sources

Transform – convert it to fit into the analytical needs

Load – get it into the right systems to derive value

Apache Hadoop is the most important framework for working with Big Data. The biggest strength of Hadoop is scalability. It can upgrade from working on a single node to thousands of nodes without any issue in a seamless manner.

The variety of Big Data means that we could be looking at data from videos, text, transactional data, sensor information, statistical data, social media conversations, search engine queries, ecommerce data, financial information, weather data, news updates, forum discussions, executive reports, and so on. Converting all this data into Business Intelligence is critical to an organization's success. Hadoop's strength lies in the fact that it is an open source platform and runs its operations on commodity hardware. Using this platform, it is possible to swiftly ingest,

process and store extremely huge amounts of data and deploy it wherever and whenever needed.