**Big Data**

**What is Big Data?**

Big data is data that contains greater **variety** arriving in increasing **volumes** and with ever-higher **velocity**. This is known as the three Vs.

Put simply, big data is larger, more complex data sets, especially from new data sources. These data sets are so voluminous that traditional data processing software just can’t manage them. But these massive volumes of data can be used to address business problems you wouldn’t have been able to tackle before.

Big data gives you new insights that open up new opportunities and business models. Getting started involves three key actions:

1. **Integrate**: Big data brings together data from many different sources and applications.
2. **Manage**: Big data requires storage. Your storage solution can be in the cloud, on-premises, or both. You can store your data in any form you want and bring your desired processing requirements and necessary process engines to those data sets on an on-demand basis.
3. **Analyze**: Your investment in big data pays off when you analyze and act on your data. Get new clarity with a visual analysis of your varied data sets. Explore the data further to make new discoveries. Share your findings with others. Build data models with [machine learning](https://lms.clarusway.com/mod/lesson/view.php?id=1053" \o "Machine Learning) and artificial intelligence. Put your data to work.

| **https://docs.google.com/uc?id=1_xE8K8J3Z-7Yc9ag6zsaVTuuj9dNJsjl** |
| --- |
| *Big Data Sources* |

Q: What are the **three V's** of big data?  
A: **Variety**: Refers to the different data types i.e. various data formats like text, audios, videos, etc.  
**Velocity** is the rate at which data grows. Social media contributes a major role in the velocity of growing data.  
**Volume** represents the volume i.e. amount of data that is growing at a high rate i.e. data volume in Petabytes(1 Petabytes = 1024 Terabytes).

**Creation of Big Data**

Data collection plays the most important role in the Big Data cycle. The Internet provides almost unlimited sources of data for a variety of topics. The importance of this area depends on the type of business, but traditional industries can acquire a diverse source of external data and combine those with their transactional data.

Your locations, your conversations from smartphones, weather conditions, bank account activities, stock market values, news values are some examples of big data sources.

**Challenges of Big Data**

While big data holds a lot of promise, it is not without its challenges.

Some of these challenges:

* **Quick Data Growth**: Data growing at such a quick rate is making it a challenge to find insights from it. There is more and more data generated every second from which the data that is actually relevant and useful has to be picked up for further analysis.
* **Storage**: Such a large amount of data is difficult to store and manage by organizations without appropriate tools and technologies.
* **Syncing Across Data Sources**: This implies that when organizations import data from different sources the data from one source might not be up to date as compared to the data from another source.
* **Security**: A huge amount of data in organizations can easily become a target for advanced persistent threats, so here lies another challenge for organizations to keep their data secure by proper authentication, data encryption, etc.
* **Unreliable Data**: We can’t deny the fact that big data can’t be 100 percent accurate. It might contain redundant or incomplete data, along with contradictions.

**Big Data Use Cases**

Big data can help you address a range of business activities, from customer experience to analytics. Here are just a few:

* **Product Development**: Companies like Netflix and P&G use big data to anticipate customer demand.
* **Predictive Maintenance**: Factors that can predict mechanical failures may be deeply buried in structured data, such as the year, make, and model of equipment, as well as in unstructured data that covers millions of log entries, sensor data, error messages, and engine temperature.
* **Customer Experience**: Big data enables you to gather data from social media, web visits, call logs, and other sources to improve the interaction experience and maximize the value delivered.
* **Fraud and Compliance**: Big data helps you identify patterns in data that indicate fraud and aggregate large volumes of information to make regulatory reporting much faster.
* [**Machine Learning**](https://lms.clarusway.com/mod/lesson/view.php?id=1053): You should have big data to train your [machine learning](https://lms.clarusway.com/mod/lesson/view.php?id=1053" \o "Machine Learning) models.
* **Operational Efficiency**: With big data, you can analyze and assess production, customer feedback and returns, and other factors to reduce outages and anticipate future demands.
* **Drive Innovation**: Big data can help you innovate by studying interdependencies among humans, institutions, entities, and processes and then determining new ways to use those insights.

| **https://docs.google.com/uc?id=1fK5f9v5fH6-dFmGOske8i1lABSLi7CL-** |
| --- |
| *Big Data application areas* |