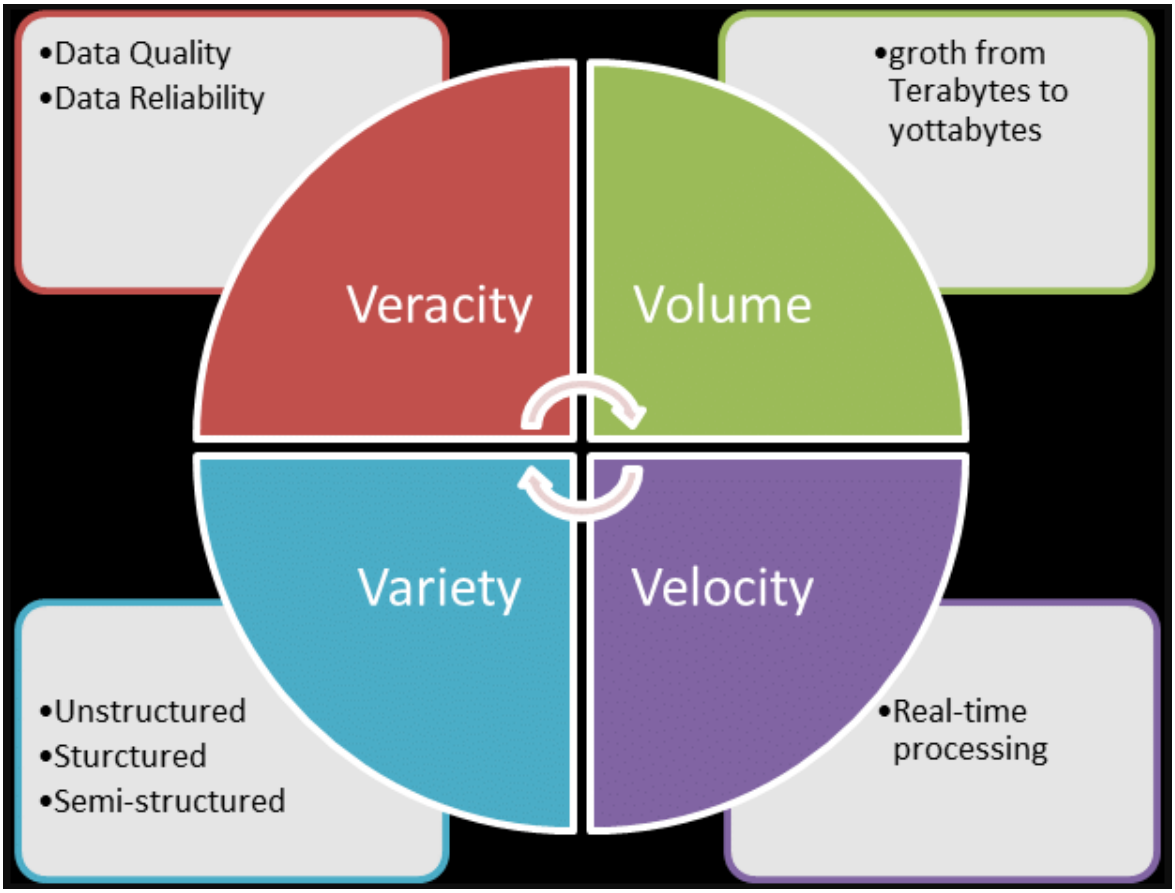
BIG DATA

Big Data refers to extremely large and complex datasets that traditional data processing tools cannot handle efficiently. It involves collecting, storing, processing, and analysing massive volumes of data to extract useful insights.

The 4 Vs of Big Data:

1. **Volume** – The amount of data generated (in terabytes, petabytes, etc.)
2. **Velocity** – The speed at which data is generated and processed
3. **Variety** – The different types and sources of data (text, images, videos, logs, etc.)
4. **Veracity** – The accuracy and trustworthiness of the data



1. Volume

   Refers to the massive amount of data generated every second. Example: Facebook generates terabytes of user data every day. Big Data deals with terabytes, petabytes, or even Exabyte of data.

2. Velocity

   Refers to how fast data is being generated, collected, and processed. Example: Stock market data, online transactions, or live sensor feeds. Big Data systems must handle real-time or near real-time data**.**

3. Variety

   Refers to the different forms of data: structured (databases), semi-structured (XML, JSON), and unstructured (videos, emails, images, audio). Big Data must handle diverse data formats from multiple sources

4. Veracity

   Refers to the trustworthiness of the data. Incomplete, inconsistent, or ambiguous data can lead to incorrect analysis. Big Data systems must ensure data cleansing and validation