**Agenda**

1. Big Data ?

**Big Data Analytics**

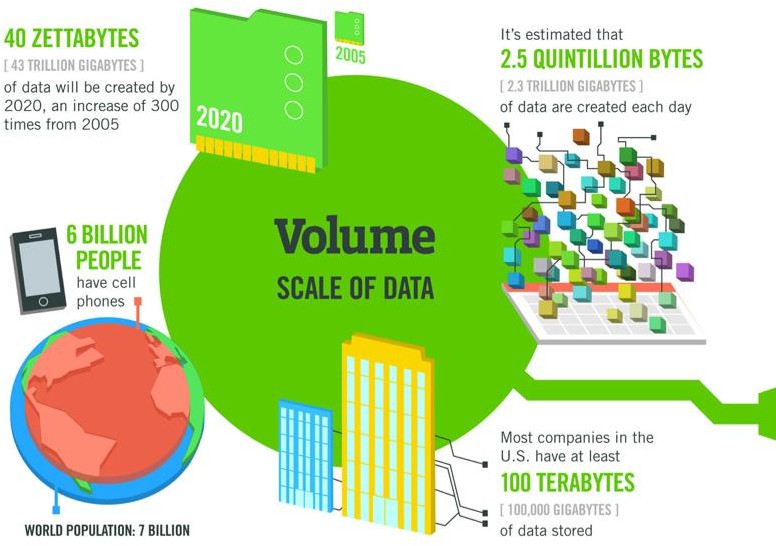
1. Characteristic of Big Data
2. Types of Big Data
3. Big Data Analytics
4. Traditional vs. Big Data business approach
5. Application of Big Data
6. Use cases of Big Data

# Big Data ?



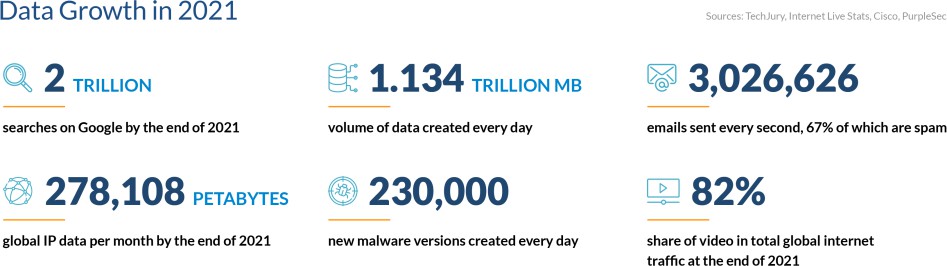
**Big Data ?**

* **Big Data means not Big or large, it is 1st characteristics of big data.**



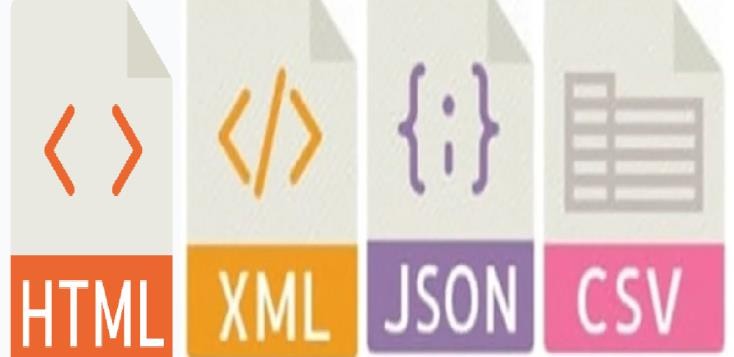
# Big Data ?

* **2nd characteristics of big data is velocity, data should be generated very high velocity.**

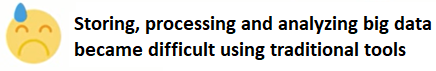
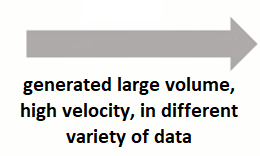
**Important Statistics about Data created every day**

# Big Data ?

* **3rd characteristics of big data is variety, means the different types of data- structured, unstructured and semi structured data.**



**Problems with Big Data ?**



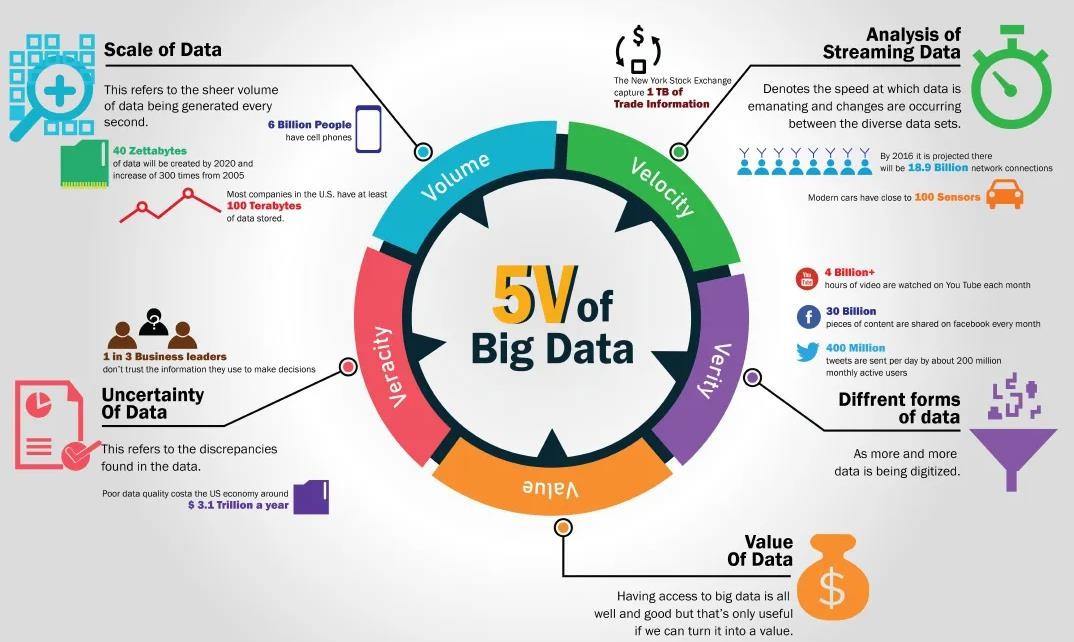
**Solution: required new tools who can solve these problems**

**Big Data?**

“The large volume of data, generated with high velocity in different variety, which cannot be stored,

processed and analyzed using traditional tools is known as Big data.”

**5v’s of Big Data**



**Big Data? (Characterizes of big data)**

It is basically a 5’ V

**Volume**

**(Amount of data)**

**Data volume is huge, in PB or more Global mobile traffic in 2016 was 6.2 EB and in 2020 it is 40,000 EB**

**Velocity**

**(Speed of data generation)**

**Data generated with high-velocity Google searches in 2018 was 32.8 million and in 2020 it was 5.6 billion**

**Variety**

**(Diversity of data)**

**Data in different varieties structured, semi-structured, and unstructured.**

**Audio, video, images, text, sensor data, etc.**

**Veracity**

## (Trustworthiness of data)

## Data should be trustworthy in terms of quality and accuracy

**Value**

**(Worth of data)**

**usefulness of gathered data for your business**

**Big Data Types**

**Structured Data**

**Predefined data models (databases)**

**Easy to search, text-based data.**

**Stored in:**

**Row and columns**

**Resides in:**

**Relational databases, table**

**(Oracle, MySQL, DB2 etc.)**

**Data warehouses**

**Examples:**

**Customer info, transaction info**

## Unstructured Data

## Characteristics

## No predefined data models

## Difficult to search

## Text, PDF, images, video, audio, etc.

## Stored in:

## Various forms

## Resides in:

## Application,

## data warehouses and lakes

## Examples:

## Documents, open-ended survey answers, social

## media data, images, audio, and video.Semi-Structured Data

## Characteristics

## Loosely organized, meta-level structures that can contain unstructured data HTML, XML, JSON, CSV

## Stored in:

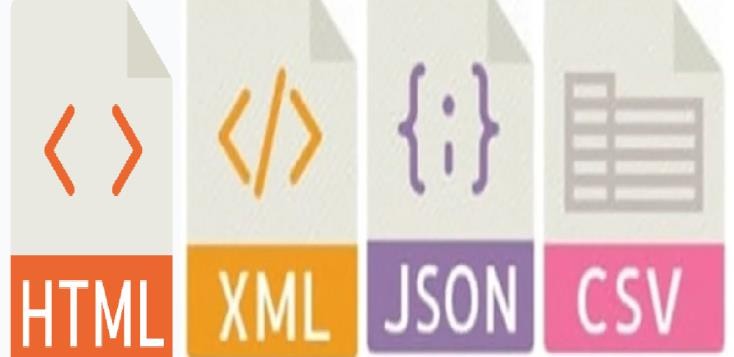
## Abstract forms

## Resides in:

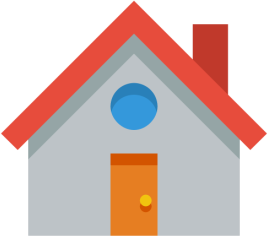
## Relational databases tagged text format

## Examples:

## Server logs, tweets organized by hashtags, emails sorting folders (inbox, sent, draft)

**Sources of Big data**





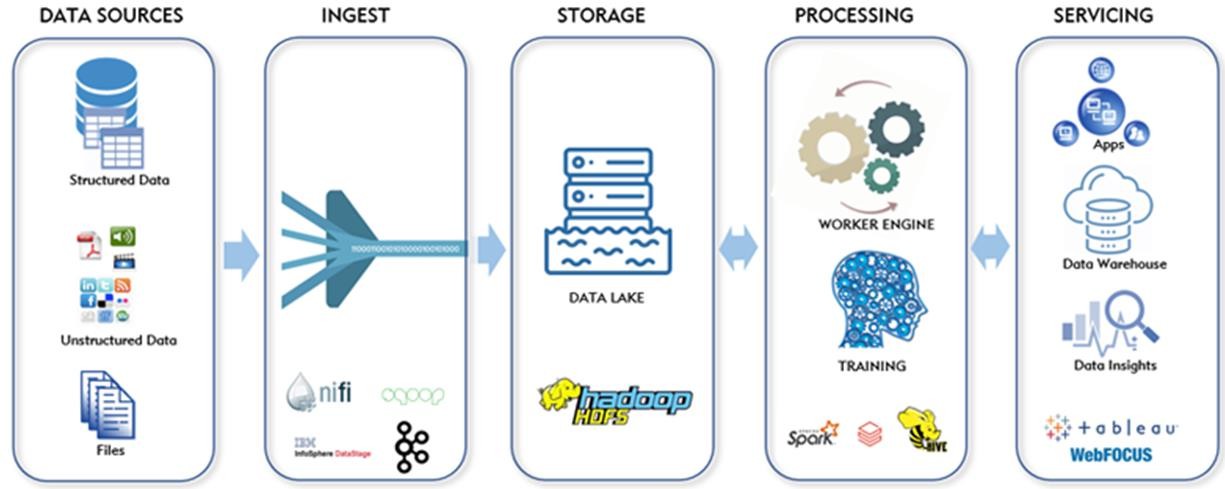
**Communication, Social media Transaction IoT Sensor data**

**Big Data**

**Big Data Analytics?**

In Simple words, Data Analytics is a process of collecting, organizing, and presenting data.

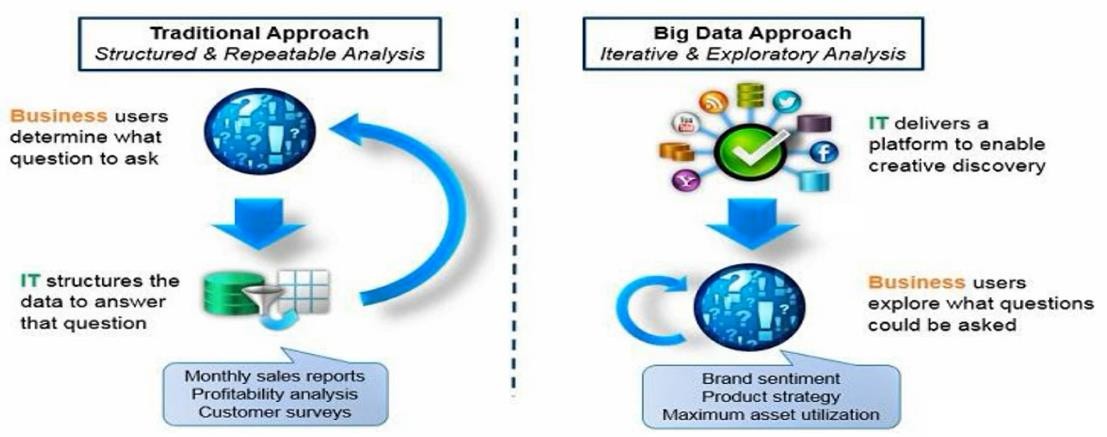
*Big Data Analytics is a complete process of examining large sets of data through varied tools and processes in order to discover unknown patterns, hidden correlations, meaningful trends, and other insights for making data-driven decisions in the pursuit of* better results.



**Types of Analytics**



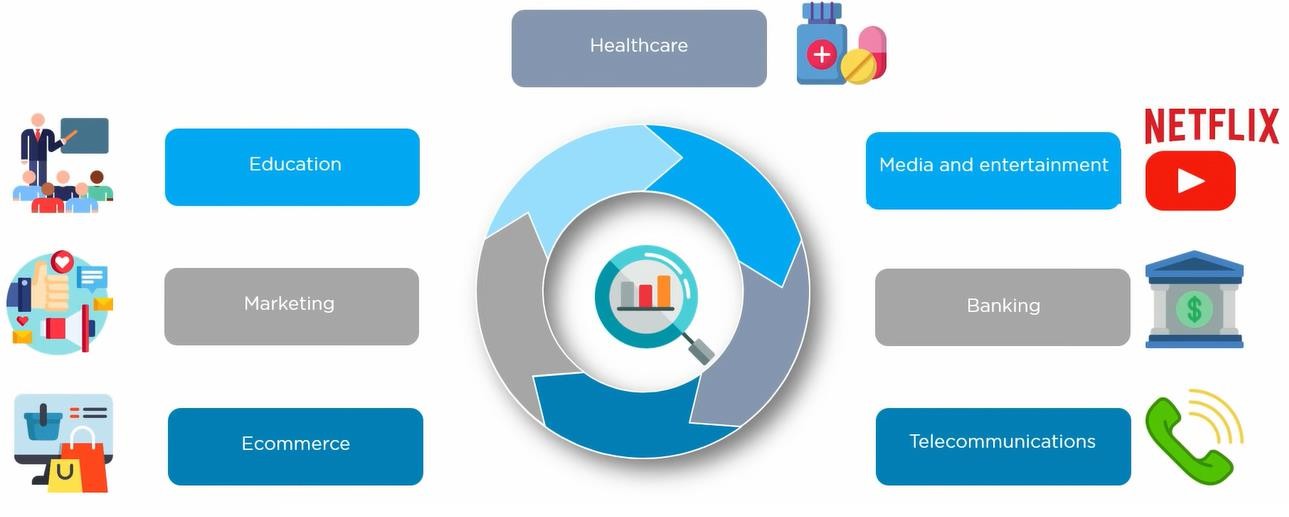
**Traditional data v/s Big data business approach**



**Traditional data v/s Big data**

|  |  |  |
| --- | --- | --- |
|  | **TRADITIONAL DATA** | **BIG DATA** |
| 1 | Its volume ranges from Gigabytes to Terabytes. | Its volume ranges from Petabytes to Zettabytes or Exabytes. |
| 2 | Traditional data is generated per hour or day or more. | But big data is generated more frequently mainly per second. |
| 2 | Traditional database system deals with structured data. | Big data system deals with structured, semi-structured, and unstructured data. |
| 4 | Normal system configuration is capable of processing traditional  data. | High system configuration is required to process big data. |
| 5 | Traditional data sources are centralized and managed in a centralized form. | A big data source is distributed, and it is managed in distributed form. |
| 6 | Traditional database tools are required to perform any database operation. | Special kinds of database tools are required to perform any database  operation. |
| 7 | Its data model is strictly schema-based and it is static. | Its data model is a flat schema and it is dynamic. |
| 8 | data sources: ERP transaction data, CRM transaction data, financial data, organizational data, web transaction data, etc. | Its data sources include social media, device data, sensor data, video, images, audio, etc. |

**Big Data Analytics Application**



**Big Data Analytics Uses**



BDA is used for product development and innovations



BDA is used for risk management

Big Data Analytics can help businesses identify new opportunities and the right strategic moves they need to make. Whether it is upselling to customers, improving productivity, marketing or improving customer satisfaction, the right application of Big Data solutions can help companies harness the power of data with scale.



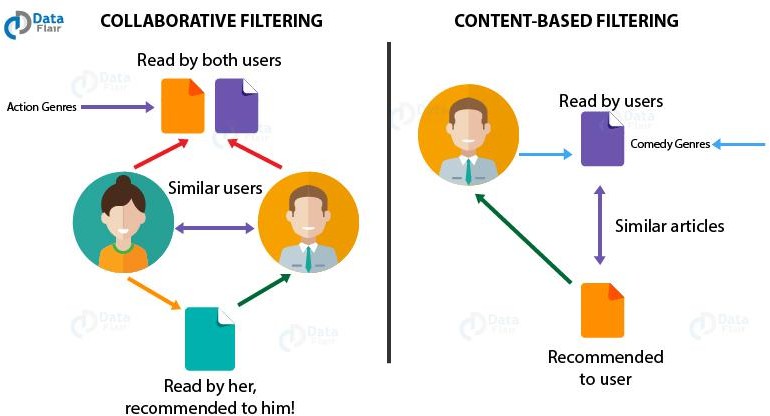
BDA is used to improve customer experience

Big Data Analytics



BDA is used in quicker and better decision making in organization

**Use case: Big data in Netflix**

Netflix implements data analytics models to discover customer behavior and buying patterns.

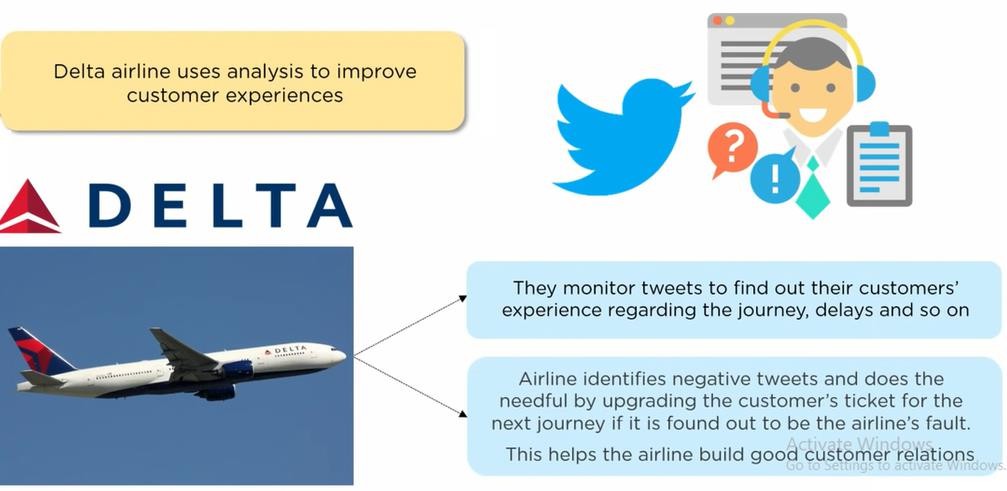
Then, using this information it recommends movies and TV shows to their customers. That is, it

analyses the customer’s choice and preferences and suggests shows and movies accordingly.

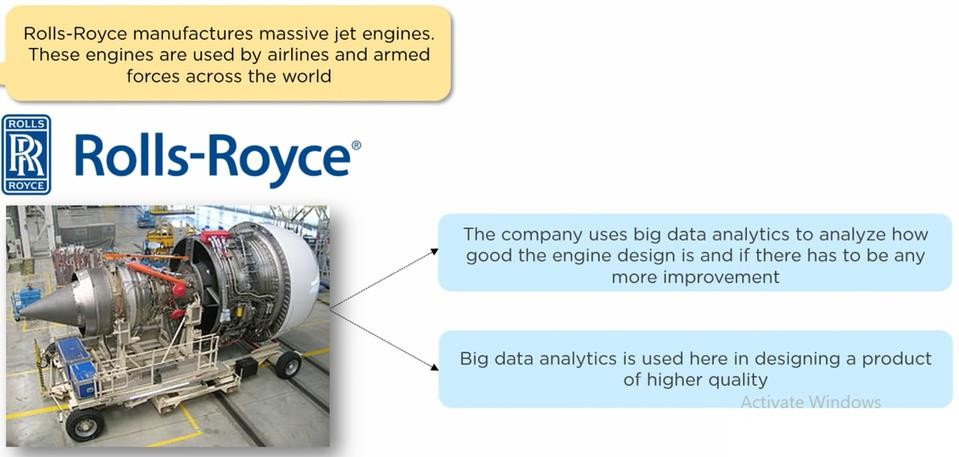
Netflix generally collects data, which is enough to create a detailed profile of its subscribers or customers. This profile helps them to know their customers better and, in the growth,

of the business. According to Netflix, around 75% of viewer activity is based on personalized recommendations.

**Use case: Delta Airline**



**Use case: Rolls-Royce**



**Use case: Big data at Google**

Google uses Big data to optimize and refine its core search and ad-serving algorithms. And Google continually develops new products and services that have Big data algorithms.

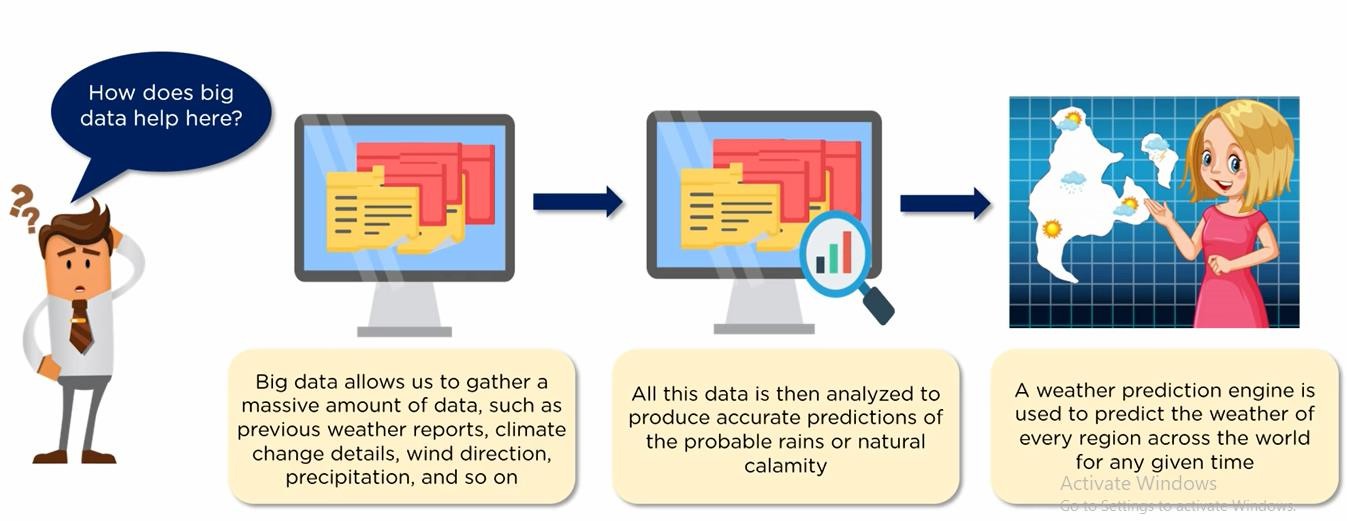
Google generally uses Big data from its Web index to initially match the queries with potentially useful results. It uses machine-learning algorithms to assess the reliability of data and then ranks the sites accordingly.

Google optimized its search engine to collect data from us as we browse the Web and show suggestions according to our preferences and interests.

**Use case: Weather forecast**



**Use case: Weather forecast**



**Big Data Analytics: Hadoop**

1. Big Data Challenges
2. Hadoop
3. Core Hadoop Components
4. Hadoop Ecosystem
5. Limitation of Hadoop