

Big Data Analytics

Academic Year 2022-23

Agenda

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



Prof. Prakash Parmar

Data Engineer | Data Analyst

Assistant Professor

Vidyalankar Institute of Technology, Mumbai

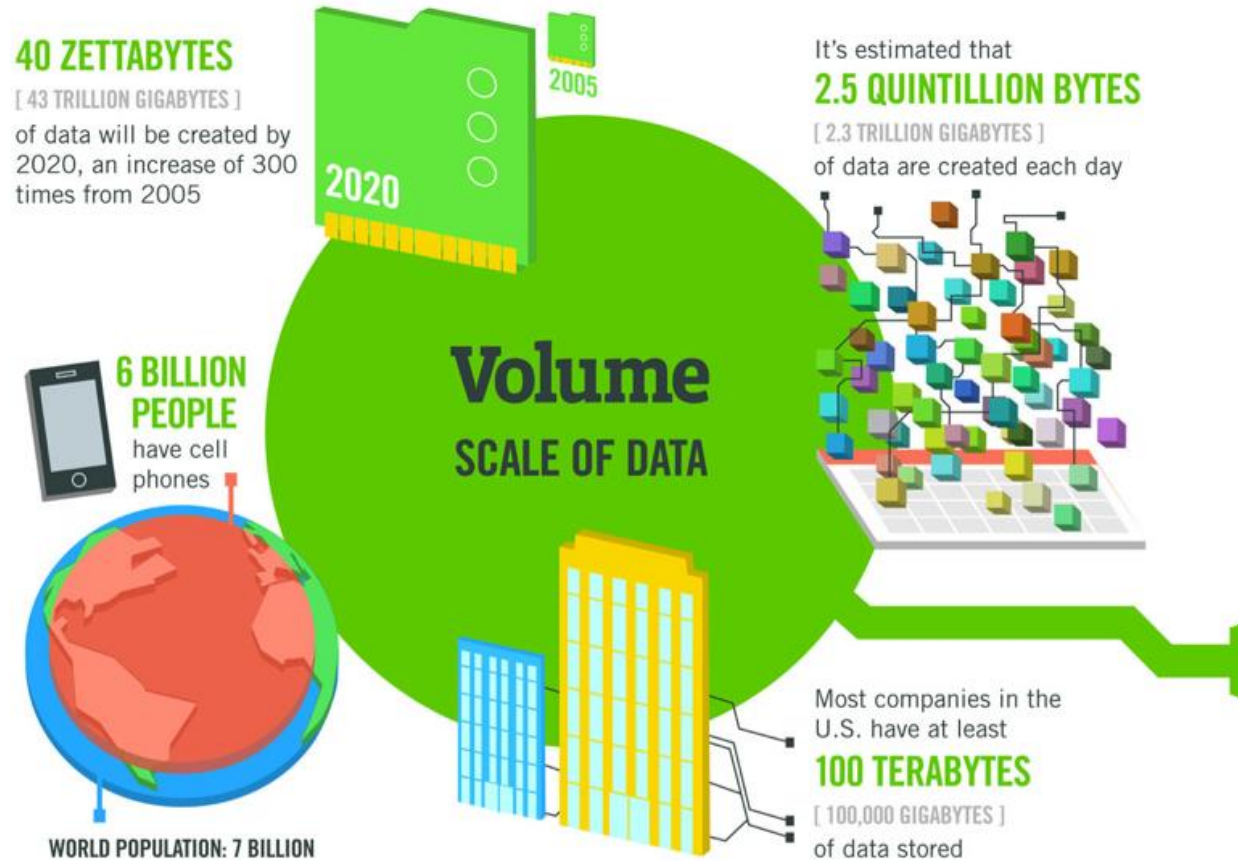
GATE Instructor: Vidyalankar Classes

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

Every 60 seconds



Data Growth in 2021

2 TRILLION

searches on Google by the end of 2021

1.134 TRILLION MB

volume of data created every day

3,026,626

emails sent every second, 67% of which are spam

278,108 PETABYTES

global IP data per month by the end of 2021

230,000

new malware versions created every day

82%

share of video in total global internet traffic at the end of 2021

Sources: TechJury, Internet Live Stats, Cisco, PurpleSec

Estimated Data Consumption from 2021 to 2024

Source: IDC / Statista



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 ?

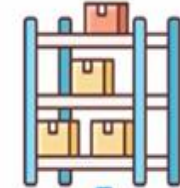


Big Data Sources



**generated large volume,
high velocity, in different
variety of data**

Storing



Processing



Analyzing

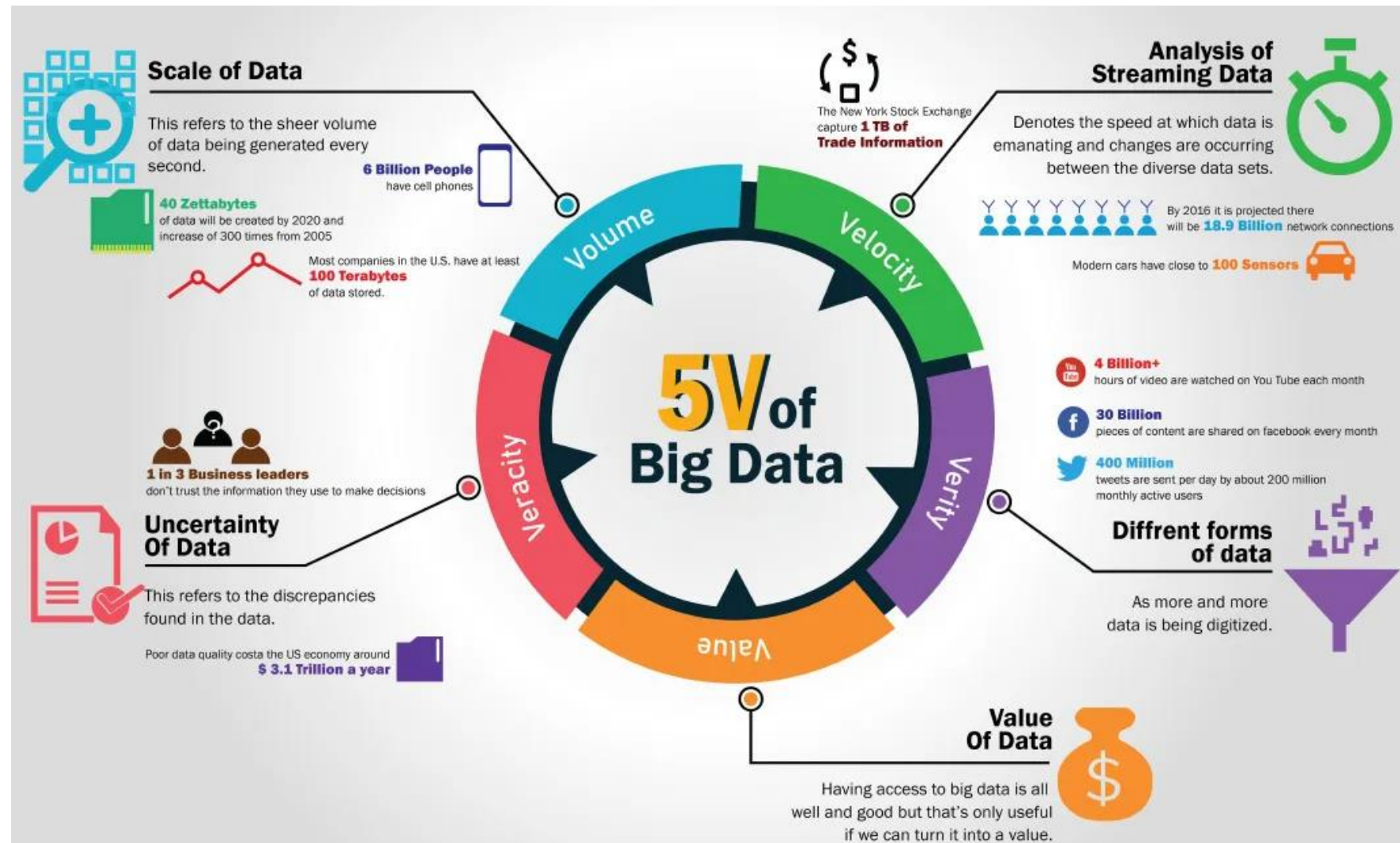


**Storing, processing and analyzing big data
became difficult using traditional tools**

Solution: required new tools who can solve these problems

“The large volume of data, generated with high velocity in different variety, which can not 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 5' V

Volume

(Amount of data)

Data volume very large, in PB or more
Global mobile traffic in 2016 is 6.2 EB
and in 2020 it is 40,000 EB

Velocity

(Speed of data generation)

Data generated with high velocity
Google search in 2018 is 32.8 million
and in 2020 it is 5.6 billion

Variety

(Diversity of data)

Data in different variety - structured , semi-structured, 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

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, 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

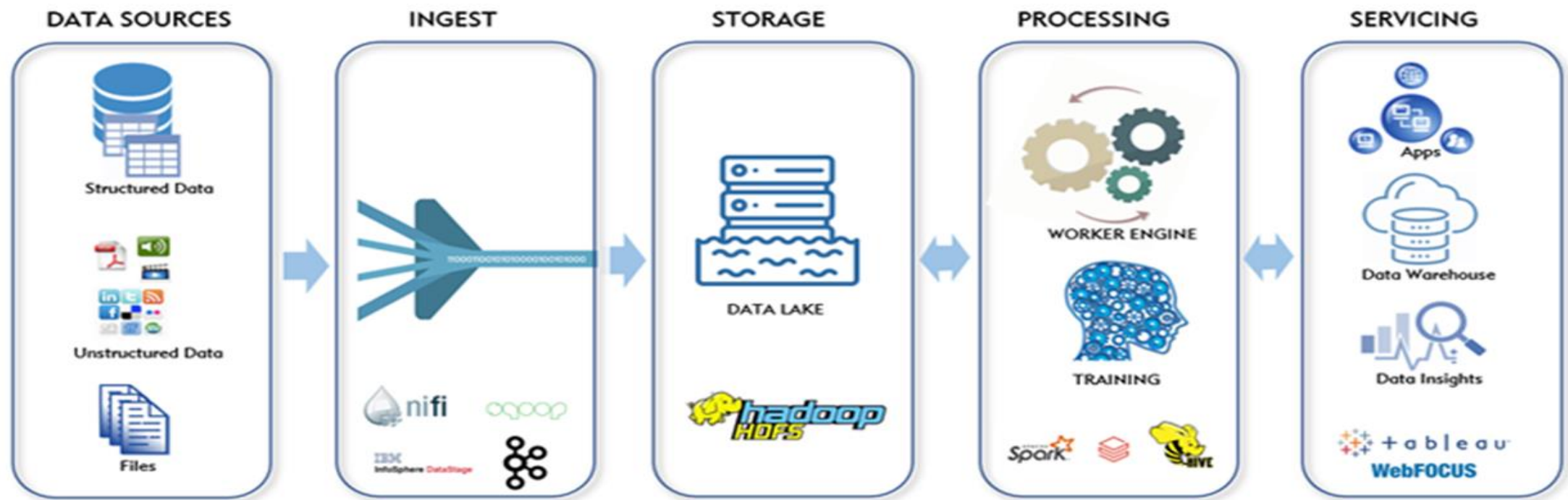
IoT Sensor data



Big Data Analytics ?

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

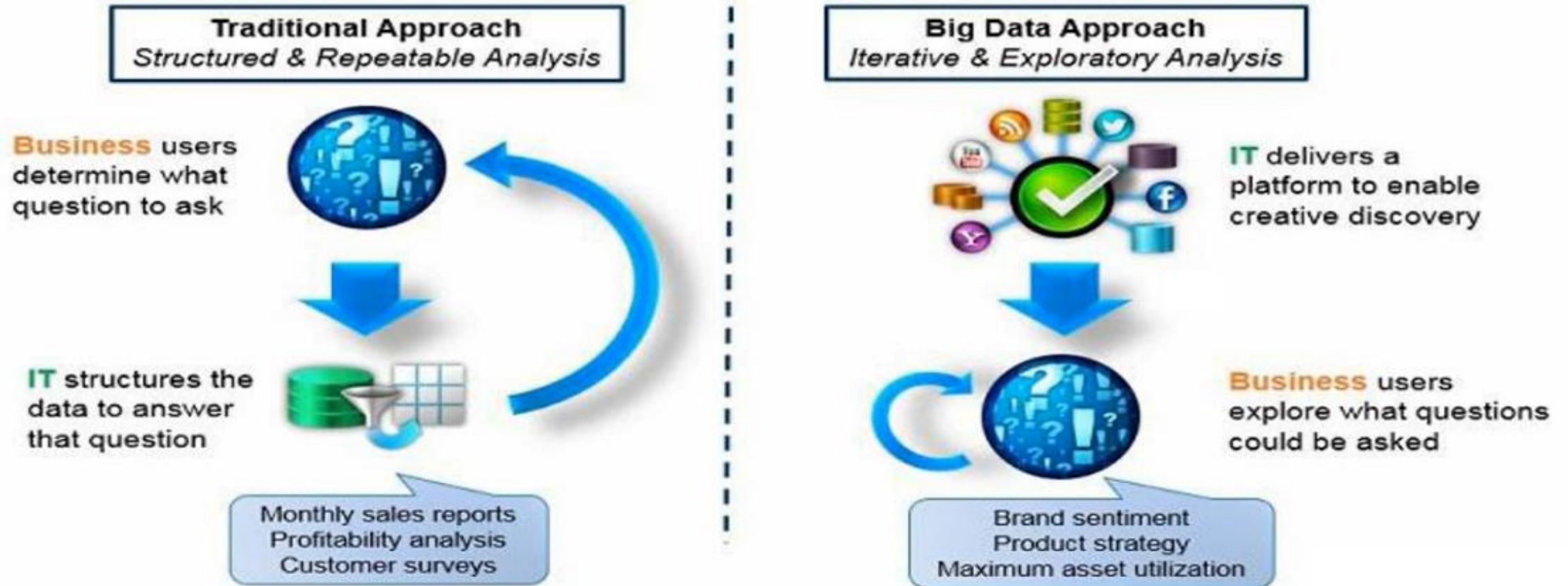
Big Data Analytics is a complete process of examining large sets of data through varied tools and processes in order to discover unknown patterns, hidden correlations, meaningful trends, and other insights for making data-driven decisions in the pursuit of better results.



Types of Analytics



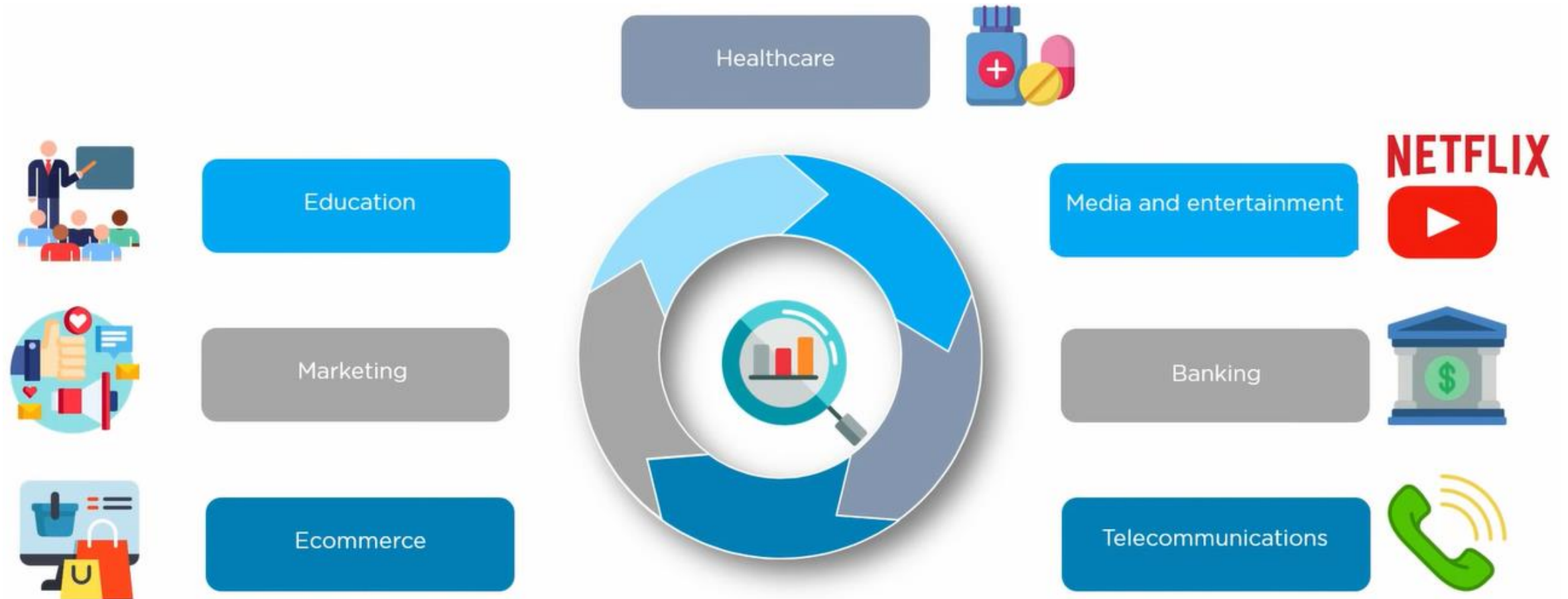
Traditional data v/s Big data business approach



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 per day or more.	But big data is generated more frequently mainly per seconds.
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 to process traditional data.	High system configuration is required to process big data.
5	Traditional data source is centralized, and it is managed in centralized form.	Big data source is distributed, and it is managed in distributed form.
6	Traditional data base tools are required to perform any data base operation.	Special kind of data base tools are required to perform any data base operation.
7	Its data model is strict schema based and it is static.	Its data model is flat schema based and it is dynamic.
8	data sources: ERP transaction data, CRM transaction data, financial data, organizational data, web transaction data etc.	Its data sources includes social media, device data, sensor data, video, images, audio etc.

Big Data Analytics Application



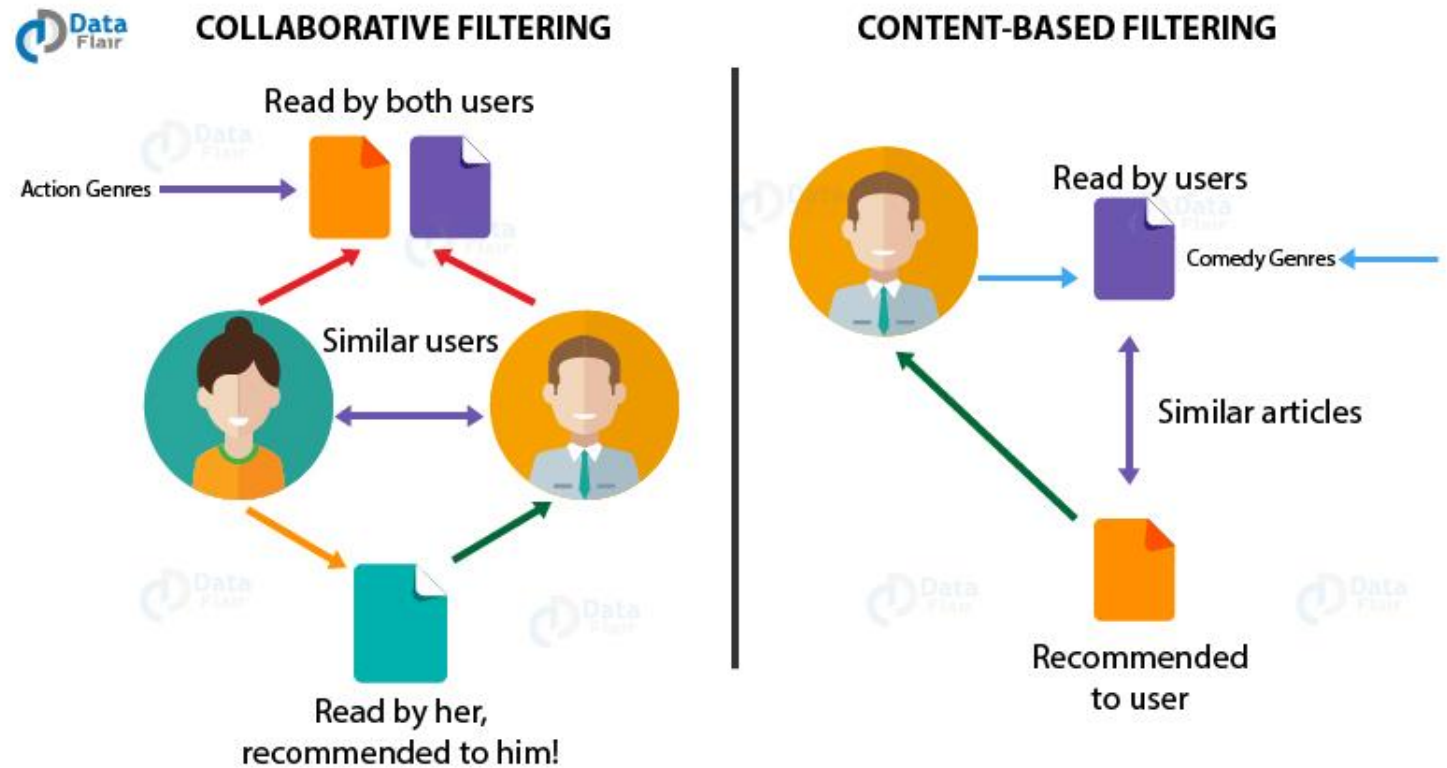
Big Data Analytics can help businesses identify new opportunities and the right strategic moves they need to make. Whether it is upselling to customers, improving productivity, marketing or improving customer satisfaction, the right application of Big Data solutions can help companies harness the power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



Use case: Big data in Netflix

Netflix implements data analytics models to discover customer behaviour 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

Delta airline uses analysis to improve customer experiences



They monitor tweets to find out their customers' experience regarding the journey, delays and so on

Airline identifies negative tweets and does the needful by upgrading the customer's ticket for the next journey if it is found out to be the airline's fault. This helps the airline build good customer relations.

Activate Windows
Go to Settings to activate Windows.

Use case: Rolls-Royce

Rolls-Royce manufactures massive jet engines. These engines are used by airlines and armed forces across the world



Rolls-Royce®



The company uses big data analytics to analyze how good the engine design is and if there has to be any more improvement

Big data analytics is used here in designing a product of higher quality

Activate Windows

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 the data from us as we browse the Web and show suggestions according to our preferences and interests.

Use case: Weather forecast

Challenge



It is very inconvenient when the weather changes suddenly. Imagine having storms, hurricanes, floods without a warning.

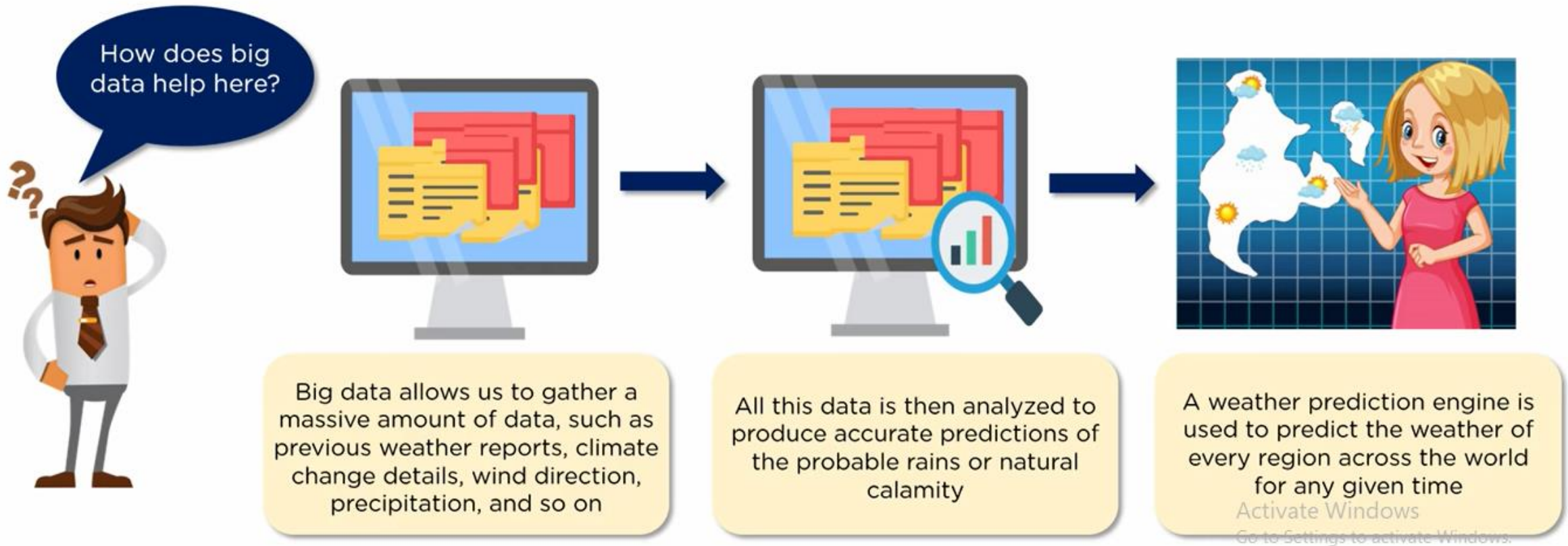
Solution



The solution is to design a system which predicts the weather conditions accurately. Big data is used in designing such a system

Activate Windows

Use case: Weather forecast



***Learn Fundamentals &
Enjoy Engineering***

Thank You



Prof. Prakash Parmar

Data Engineer | Data Analyst

Assistant Professor

Vidyalankar Institute of Technology, Mumbai

GATE Instructor: Vidyalankar Classes