SNS COLLEGE OF TECHNOLOGY

Coimbatore-35.
An Autonomous Institution

COURSE NAME : DATA ANALYTICS

II YEAR/ IV SEMESTER

UNIT – I INTRODUCTION

Topic: *Introduction to Data*

Dr.K.Sangeetha

HoD

Department of Computer Science and Engineering

Let's look at
Big Data
in a different way

Introduction/Data Analytics/Sangeetha K/CSE/SNSCT

What was the capacity of our first Digital storage?

1946 Williams Tube

• Professor Fredrick C. Williams and his colleagues from university of Manchester developed the first random access computer memory in 1946.

• He used a series of electrostatic cathode-ray tubes for digital storage.

• A storage of 1024 bits of information was successfully.
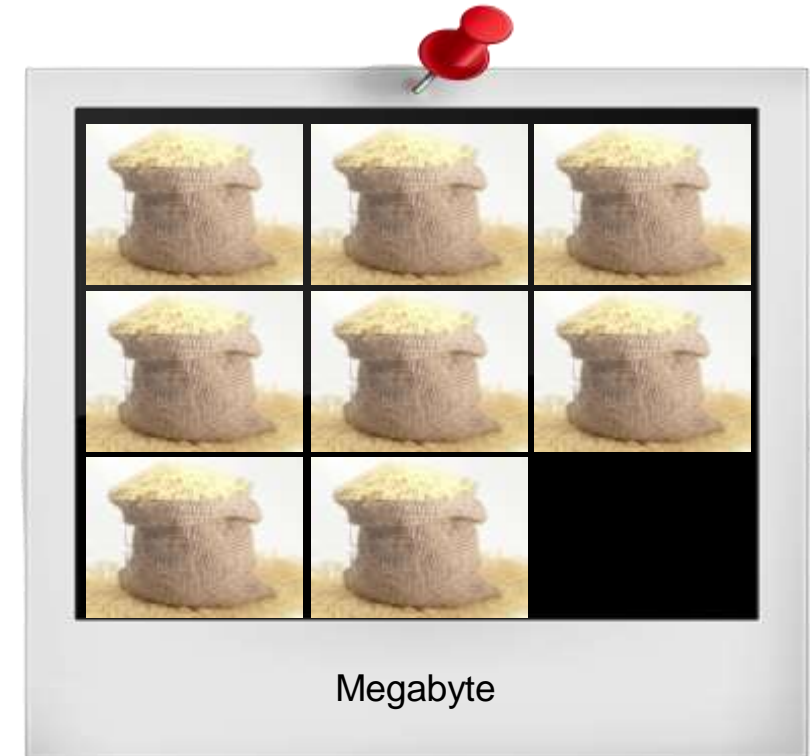
Byte   : one grain of rice

Byte

Byte   : one grain of rice
Kilobyte   : cup of rice

Kilobyte

Byte   : one grain of rice
Kilobyte   : cup of rice
Megabyte : 8 bags of rice

Megabyte

Byte   : one grain of rice
Kilobyte   : cup of rice
Megabyte : 8 bags of rice
Gigabyte   : 3 Semi trucks

Gigabyte

Byte   : one grain of rice
Kilobyte   : cup of rice
Megabyte : 8 bags of rice
Gigabyte : 3 Semi trucks
Terabyte   : 2 Container Ships

Terabyte

Byte   : one grain of rice
Kilobyte   : cup of rice
Megabyte : 8 bags of rice
Gigabyte   : 3 Semi trucks
Terabyte   : 2 Container Ships
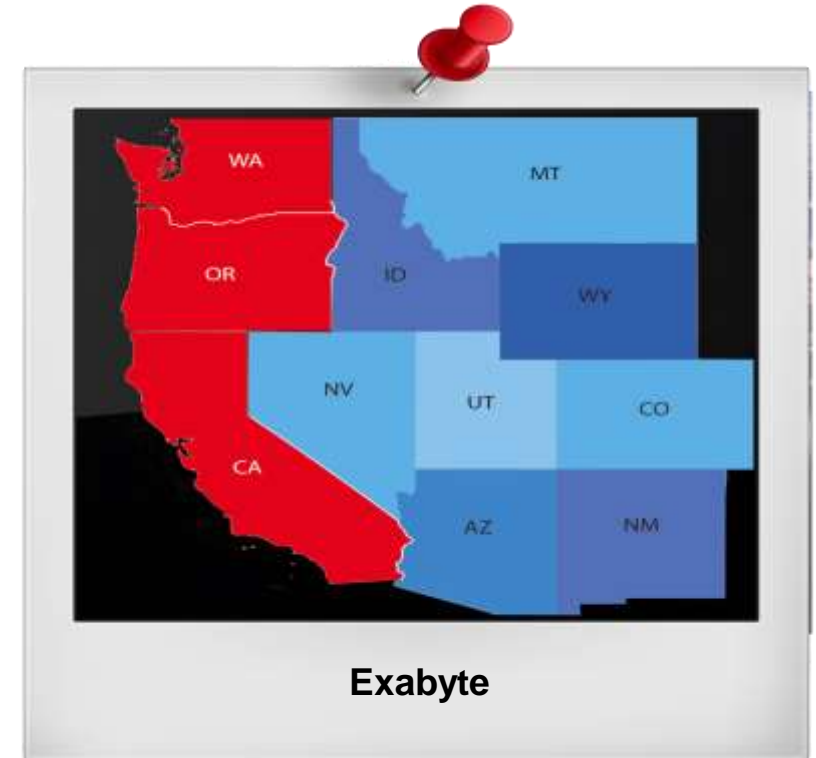**Petabyte   : Blankets Manhattan**

Petabyte

Byte　　　　: one grain of rice
Kilobyte　　: cup of rice
Megabyte : 8 bags of rice
Gigabyte　: 3 Semi trucks
Terabyte　: 2 Container Ships
Petabyte　: Blankets Manhattan
Exabyte　　: Blankets west coast states

Exabyte

Byte   : one grain of rice
Kilobyte   : cup of rice
Megabyte : 8 bags of rice
Gigabyte   : 3 Semi trucks
Terabyte   : 2 Container Ships
Petabyte   : Blankets Manhattan
Exabyte   : Blankets west coast states
Zettabyte : Fills the Pacific Ocean

Zettabyte

Byte : one grain of rice
Kilobyte : cup of rice
Megabyte : 8 bags of rice
Gigabyte : 3 Semi trucks
Terabyte : 2 Container Ships
Petabyte : Blankets Manhattan
Exabyte : Blankets west coast states
Zettabyte : Fills the Pacific Ocean

Yottabyte

Byte   : one grain of rice
Kilobyte   : cup of rice

Hobbyist

Megabyte : 8 bags of rice

Gigabyte : 3 Semi trucks

Terabyte   : 2 Container Ships

Petabyte   : Blankets Manhattan

Exabyte   : Blankets west coast states

Zettabyte : Fills the Pacific Ocean

Yottabyte : A EARTH SIZE RICE BALL!

Byte   : one grain of rice
Kilobyte   : cup of rice

Hobbyist

Megabyte : 8 bags of rice
Gigabyte   : 3 Semi trucks

Desktop

Terabyte   : 2 Container Ships

Petabyte   : Blankets Manhattan

Exabyte   : Blankets west coast states

Zettabyte : Fills the Pacific Ocean

Yottabyte : A EARTH SIZE RICE BALL!

Byte : one grain of rice		Hobbyist
Kilobyte : cup of rice		
Megabyte : 8 bags of rice		Desktop
Gigabyte : 3 Semi trucks		
Terabyte : 2 Container Ships		Internet
Petabyte : Blankets Manhattan		
Exabyte : Blankets west coast states		
Zettabyte : Fills the Pacific Ocean		
Yottabyte : A EARTH SIZE RICE BALL!		

Byte : one grain of rice	Hobbyist
Kilobyte : cup of rice	
Megabyte : 8 bags of rice	Desktop
Gigabyte : 3 Semi trucks	
Terabyte : 2 Container Ships	Internet
Petabyte : Blankets Manhattan	
Exabyte : Blankets west coast states	
Zettabyte : Fills the Pacific Ocean	
Yottabyte : A EARTH SIZE RICE BALL!	

Byte   : one grain of rice
Kilobyte : cup of rice
Megabyte : 8 bags of rice
Gigabyte : 3 Semi trucks
Terabyte : 2 Container Ships
Petabyte : Blankets Manhattan
Exabyte   : Blankets west coast states
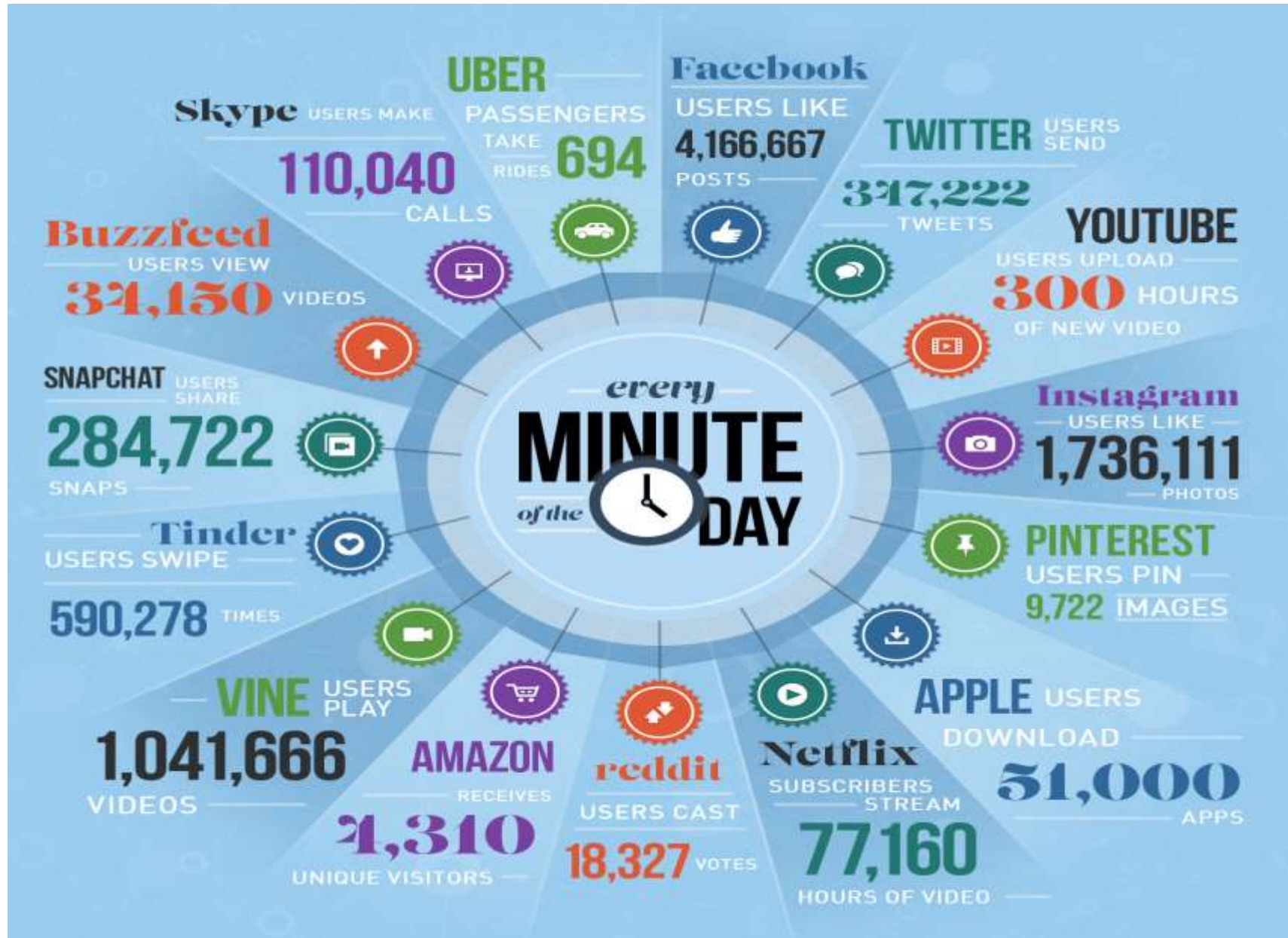Zettabyte : Fills the Pacific Ocean
Yottabyte : A EARTH SIZE RICE BALL!

Byte : one grain of rice		Hobbyist
Kilobyte : cup of rice		
Megabyte : 8 bags of rice		Desktop
Gigabyte : 3 Semi trucks		
Terabyte : 2 Container Ships		Internet
Petabyte : Blankets Manhattan		
Exabyte : Blankets west coast states		Big Data
Zettabyte : Fills the Pacific Ocean		

The Future?

Introduction/Data Analytics/Sangeetha K/CSE/SNSCT

PLANES:

Minute-to-minute monitoring

⛈ Weather

Detailed weather sensors are more accurate than the National Weather Service, telling the airline when to prepare de-icers and delay and cancel flights.

⚙ Parts

Even a five degree temperature variation may indicate a part needs to be replaced.

✈ Flight Plan

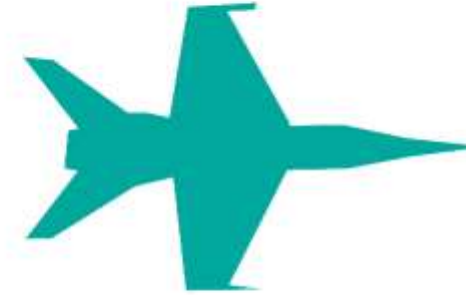Keeping an eye on a plane's path from the ground, and alerting pilots of any anomalies.

Introduction/Data Analytics/Sangeetha K/CSE/SNSCT

Traditional vs Big Data

Big data is a term for *data* sets that are **so** *large* **or complex** that **traditional** *data* **processing** **application** **software is inadequate** to deal with them.
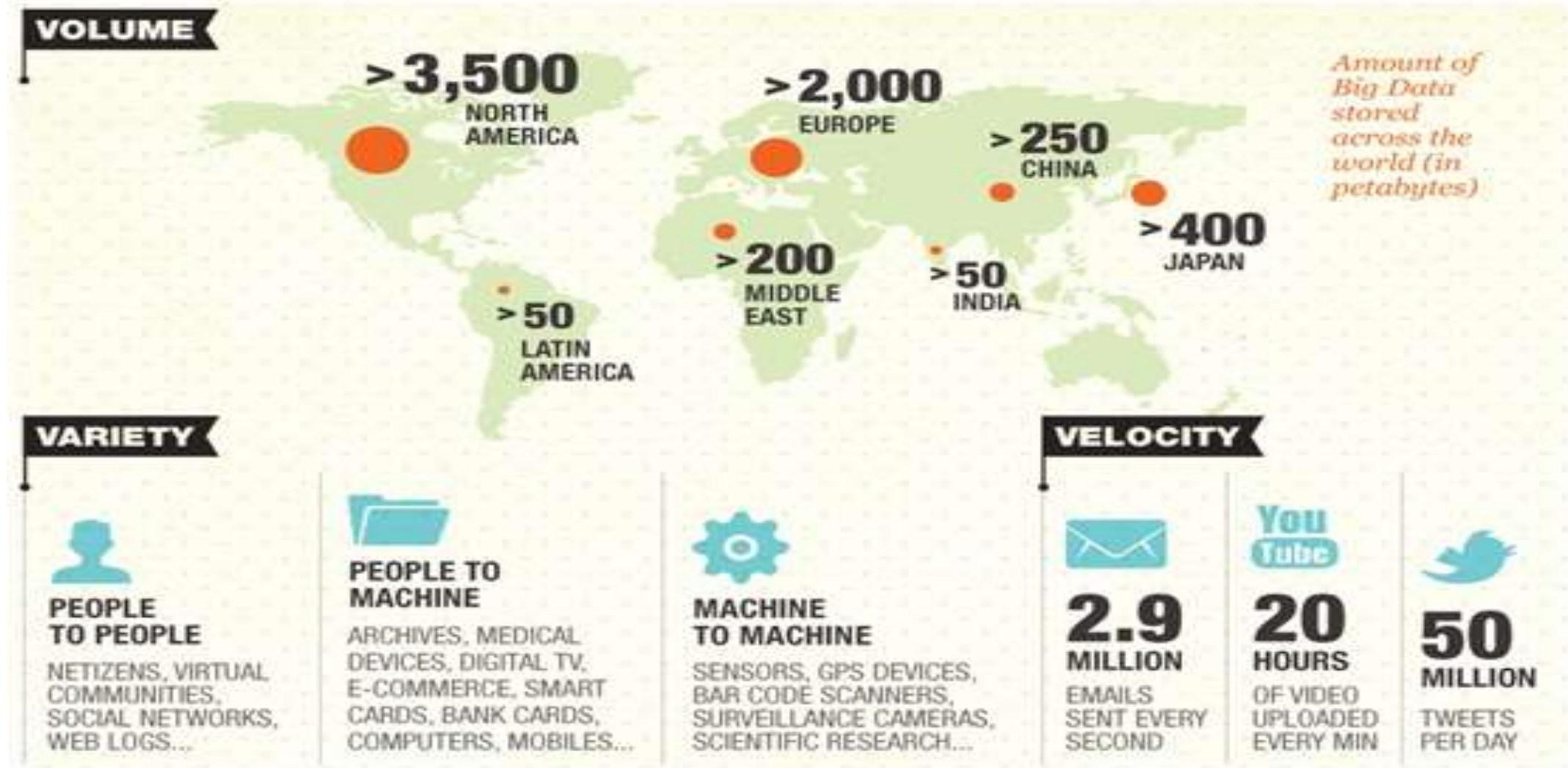
AMOUNT OF DATA (VOLUME)
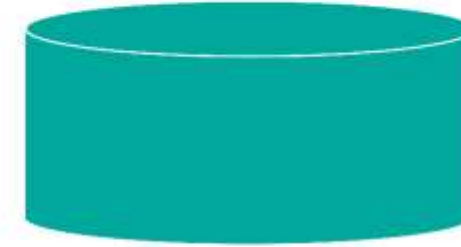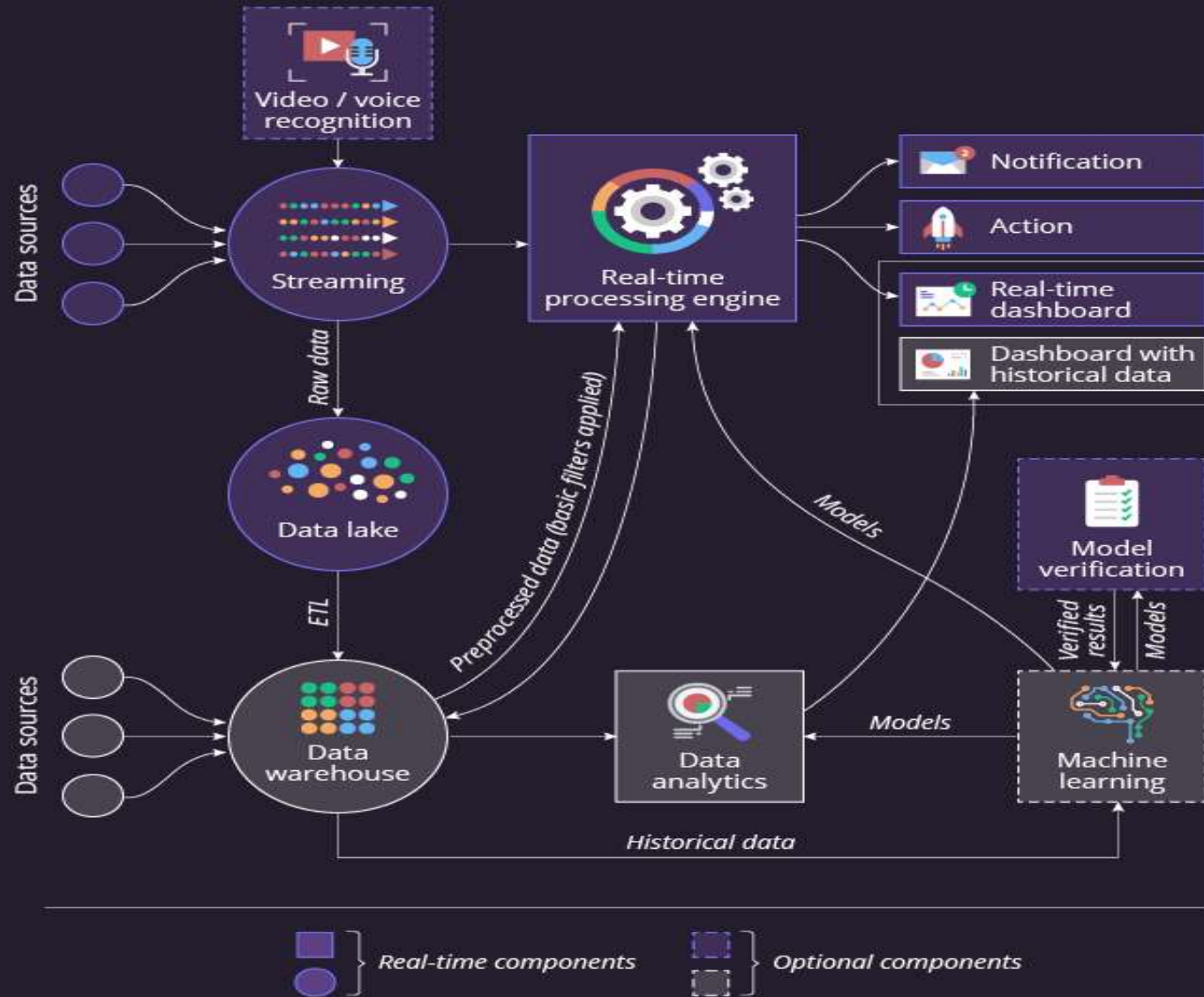
RATE OF DATA GENERATION AND TRANSMISSION (VELOCITY)

TYPES OF STRUCTURED AND UNSTRCTURED DATA (VARIETY)

Introduction/Data Analytics/Sangeetha K/CSE/SNSCT

Introduction/Data Analytics/Sangeetha K/CSE/SNSCT

REAL-TIME BIG DATA ANALYTICS ARCHITECTURE

Introduction/Data Analytics/Sangeetha K/CSE/SNSCT

Five Big Data use cases

Big Data Exploration
Find, visualize, understand all big data to improve business knowledge

Enhanced 360° View of the Customer
Achieve a unified view, incorporating many data sources

Security/Intelligence Extension
Lower risk, detect fraud and monitor security in real-time

Operations Analysis
Analyze a variety of machine data for improved business results

Data Warehouse Modernization
Integrate big data and data warehouse capabilities to increase operational efficiency

Data Science vs Big data Vs Data Analytics

Data science is the combination of statistics, mathematics, programming, problem-solving, capturing data in ingenious ways, the ability to look at things differently, and the activity of cleansing, preparing, and aligning data. This umbrella term includes various techniques that are used when extracting insights and information from data.

Big data refers to significant volumes of data that cannot be processed effectively with the **traditional applications** that are currently used. The processing of big data begins with raw data that isn't aggregated and is most often impossible to store in the memory of a single computer.

Data analytics is the science of examining raw data to reach **certain conclusions**.

References

- Joao Moreira, Andre Carvalho, Tomás Horvath – "A General Introduction to Data Analytics" Wiley -2018
- Dean J, ―Big Data, Data Mining and Machine learning, Wiley publications, 2014.
- Provost F and Fawcett T, ―Data Science for Business, O'Reilly Media Inc, 2013.
- Janert PK, ―Data Analysis with Open Source Tools, O'Reilly Media Inc, 2011.
- Weiss SM, Indurkhya N and Zhang T, ―Fundamentals of Predictive Text Mining, Springer-Verlag London Limited, 2010.
- Marz N and Warren J,- Big Data, Manning Publications,2015
- Runkler T A, - Data Analytics: Models and Algorithms for Intelligent data analysis, Springer, 2012