Developing Computational Thinking Skills in Society 5.0 Era

Credits : gifer.com

INTRODUCTION TO BIG DATA

What is Big Data?

What makes data “Big”?

OMITS WEBINAR - 11 SEPT 2020

Ferisa Tri Putri Prestasi

B.Sc of Mathematics - Machine Learning and Big Data Laboratory Institut Teknologi Sepuluh Nopember Surabaya

Head of Research Development and Knowledge Management of Data Science Indonesia East Java Chapter

Data Scientist of PT. Multipolar Technology Tbk

Content Creator and Public Relation of RMDS NarasioData (Data Science Training and Certification)

/ ferisaprestasi

/ in/ferisaprestasi

/ @ferisaprestasi

tableau public / profile/ferisa.tri.putri.prestasi

TABLE OF CONTENTS

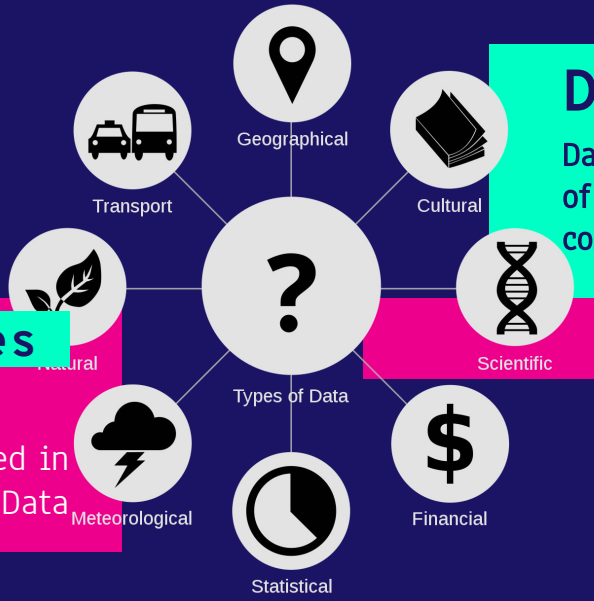
01

OVERVIEW OF BIG DATA

Let us begin by defining the term of Data

Data ?

Data are individual facts, statistics, or items of information, often numeric, that are collected through observation



Text **Number** **Images**

Audio **Video**

All facts and figures which can be stored in digital format can be term as Data

https://en.wikipedia.org/wiki/Data

5,000% growth

Fact

59T GB

From 2010 to 2020, the amount of data created, captured, copied, and consumed in the world **increased** from **1.2 trillion gigabytes**

https://www.forbes.com/sites/gilpress/2021/12/30/54-predictions-about-the-state-of-data-in-2021/

Example Source that causes Increasing Data

Social Media

Social Media and all network are generating data

Scientific Instruments

Collecting all sorts of data example data from universe

Mobile Device

Tracking all objects all the time from user and engagement

Sensor Technology and Network

Measuring all kinds of data such as from e-commerce, financial services, real time search,etc

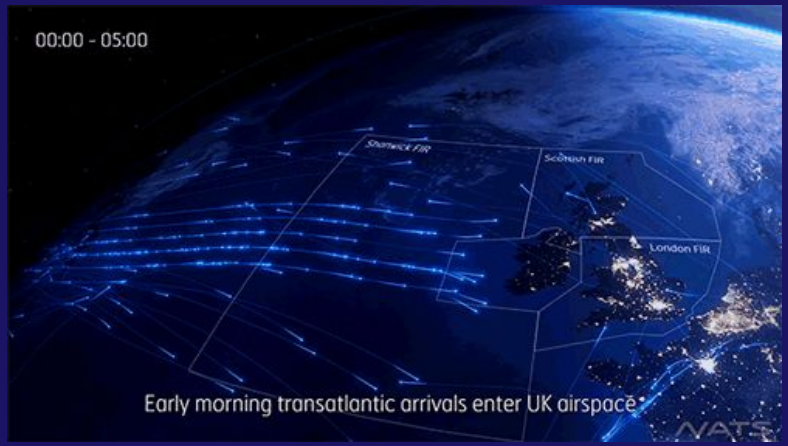
INTRODUCTION

No single standard definition! here is from Wikipedia:

"Big data is a field that treats ways to **analyze**, systematically extract information from, or otherwise deal with data sets that are **too large or complex** to be dealt with by traditional data-processing application software."

- 'Big-data' is similar to 'Small-data', but bigger

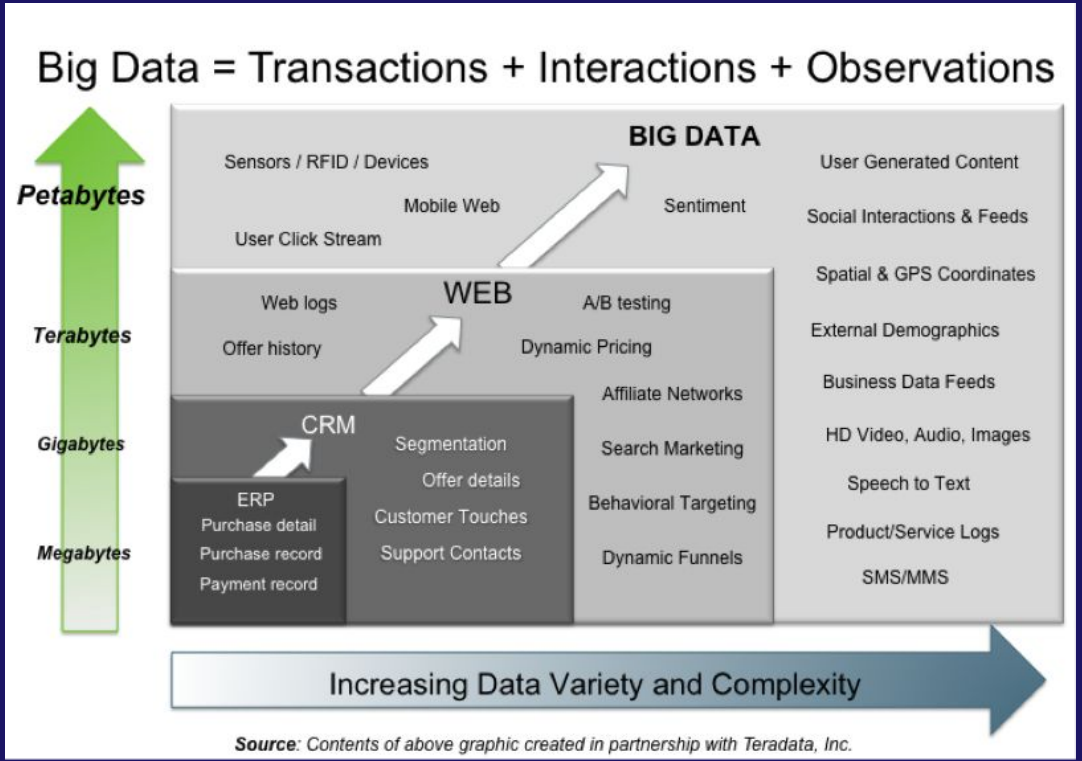
...but having data bigger consequently requires different approaches techniques, tools and architectures



Credits : gifer.com

https://en.wikipedia.org/wiki/Big_data

Big Data Context



https://www.wiideman.com/blog/tools-software/using-big-data-to-get-big-seo-results

“Data is the new science.
Big Data holds
the answers.”

Pat Gelsinger

CEO of Intel Corporation

02

BIG DATA CHARACTERISTICS

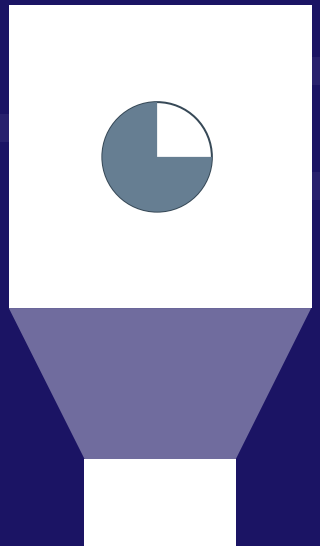
FEATURES OF THE TOPIC

Volume (Scale)

Large amounts of Data

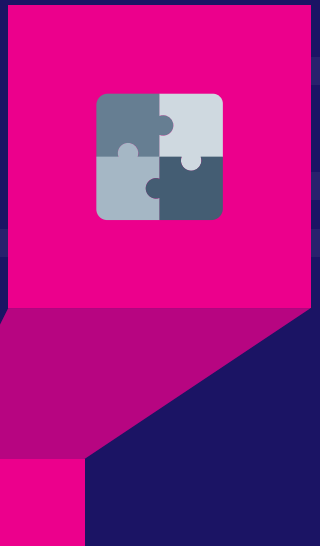
Velocity (Speed)

Needs to be analyzed **quickly**



Variety (Complexity)

Different types of structured and unstructured data



Volume (Scale)

Data Volume

- Growth 40% per year
- From 8 zettabytes (2016) to 44zb (2020)

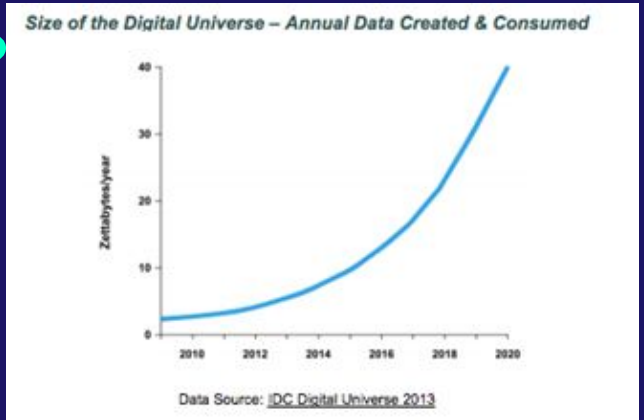
Data volume is increasing exponentially

Data Tweets per Day

Exponential increase in collected/generated data

Size of the Digital Universe

Exponential increase in collected/generated data



https://www.scnsoft.com/blog/what-is-big-data

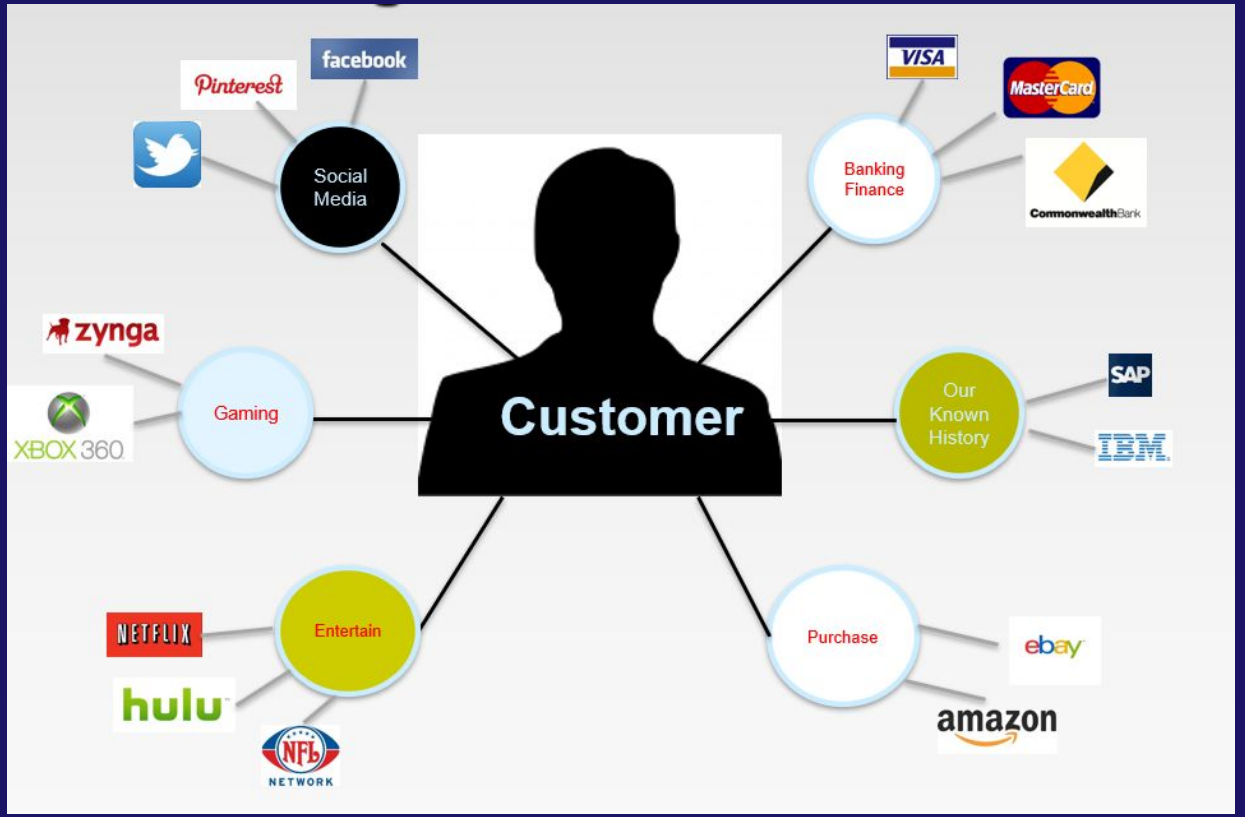
Variety (Complexity)

Different Types:

- Relational Data (Tables/Transaction/Legacy Data)
- Text Data (Web)
- Semi-structured Data (XML)
- Graph Data (Social Network, Semantic Web (RDF))
- Streaming Data (You can only scan the data once)
- A single application can be generating/collecting many types of data

Different Sources:

- Movie reviews from IMDB and Rotten Tomatoes
- Product reviews from different provider websites



https://www.scnsoft.com/blog/what-is-big-data

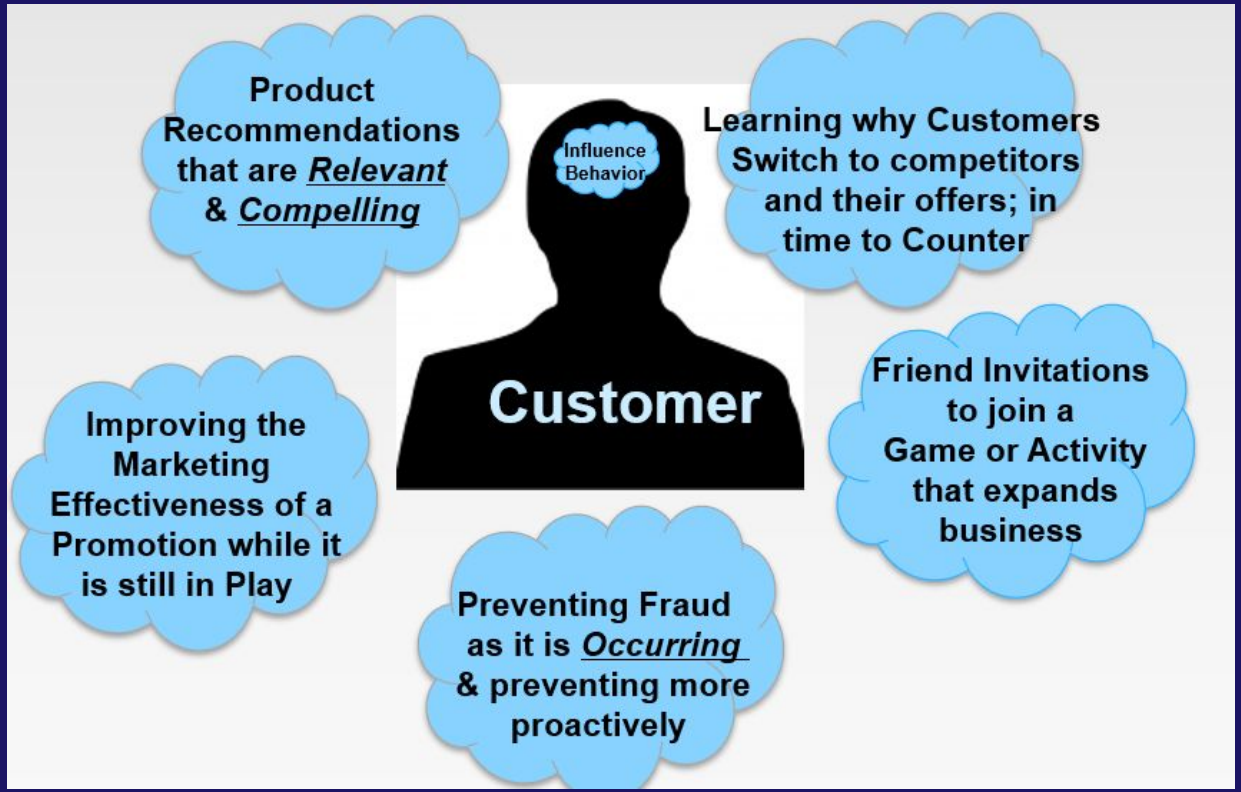
Velocity (Speed)

- Data is begin generated fast and need to be processed fast
- Online Data Analytics
- Late decisions missing opportunities
- Examples :

1. **E-Promotions**: Based on your current location, your purchase history, what you like send promotions right now for store next to you

2. **Healthcare monitoring**: sensors monitoring your activities and body any abnormal measurements require immediate reaction

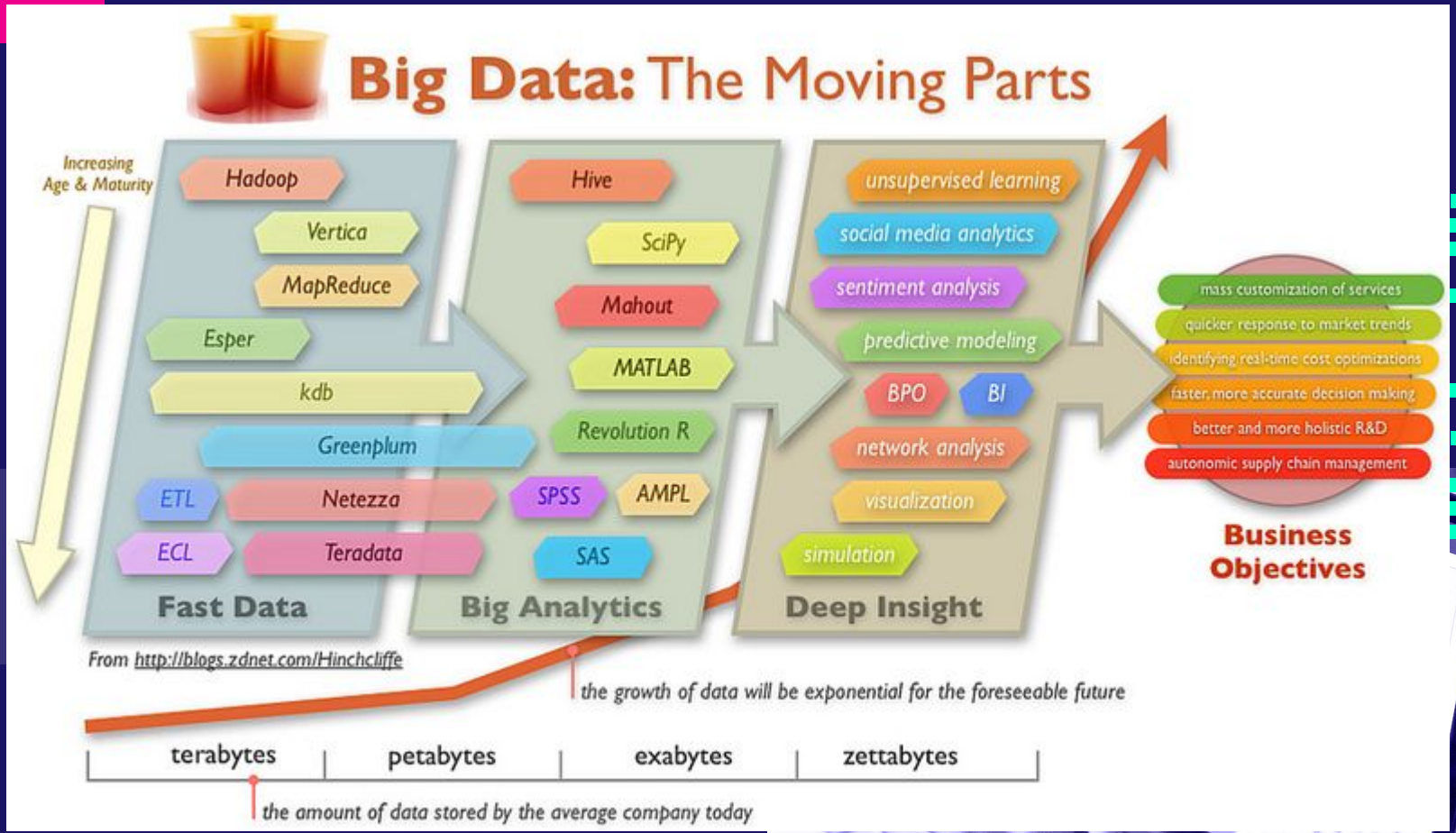
3. **Disaster management and response**



https://www.scnsoft.com/blog/what-is-big-data

03

BIG DATA TOOLS AND TECHNOLOGIES



https://whatsthebigdata.com/2012/12/18/big-data-the-moving-parts/

Top 15 Big Data Tools for Data Analysis

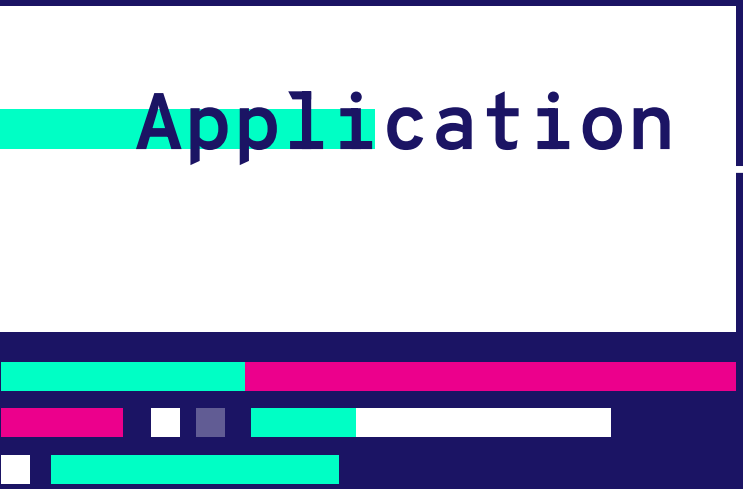
1. Xplenty
2. Adverity
3. Apache Hadoop
4. CDH (Cloudera Distribution for Hadoop)
5. Cassandra
6. Knime
7. Datawrapper
8. MongoDB
9. Lumify
10. HPCC
11. Storm
12. Apache SAMOA
13. Talend
14. Rapidminer
15. Qubole
16. Tableau
17. R

https://www.softwaretestinghelp.com/big-data-tools/

04

BIG DATA APPLICATIONS

BIG DATA APPLICATION FROM DIFFERENT INDUSTRIES



Customer Analytics

Demographic data, Transactional Data, Web Behaviour Data, Comments Data, Rate Product Data

Industrial Analytics

Historical Sensor data to foster proactive maintenance.

Business Process Analytics

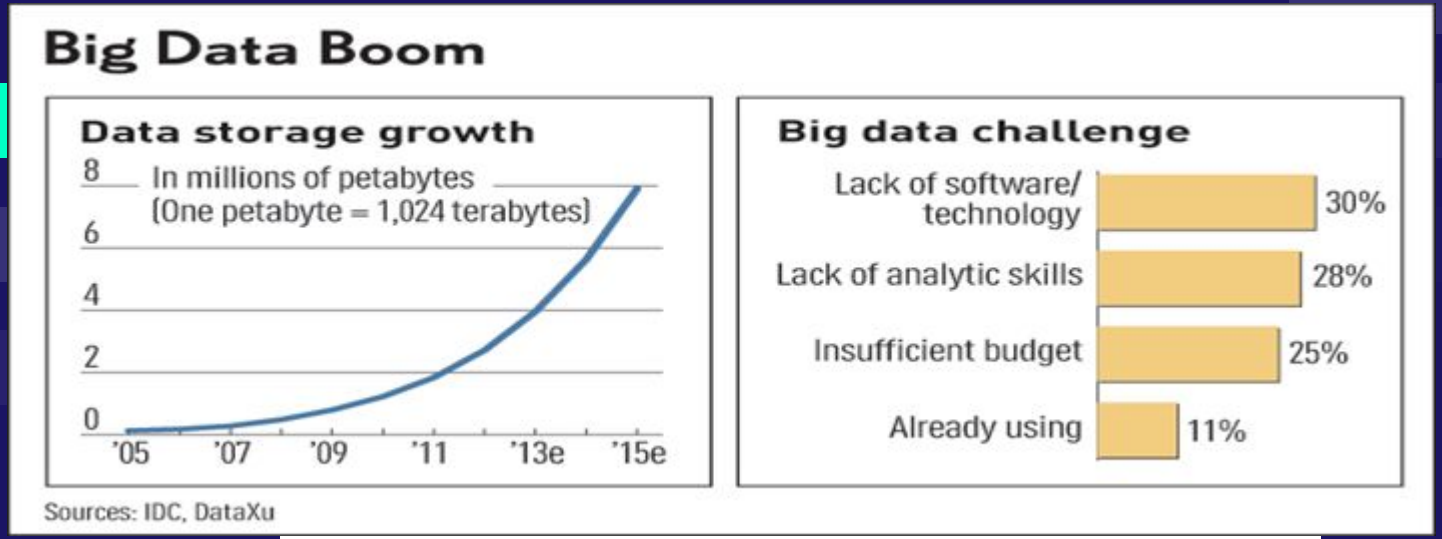
Telemetry data that comes from each truck in real time to identify a typical behavior of each driver.

Analytics for Fraud Detection

Customer banking history transactions

https://www.scnsoft.com/blog/what-is-big-data

Challenges in Handling Big Data



1. **The Bottleneck is in technology**

New architecture, algorithms, techniques are needed

2. **Also in skills**

Experts in using the new technology and dealing with big data

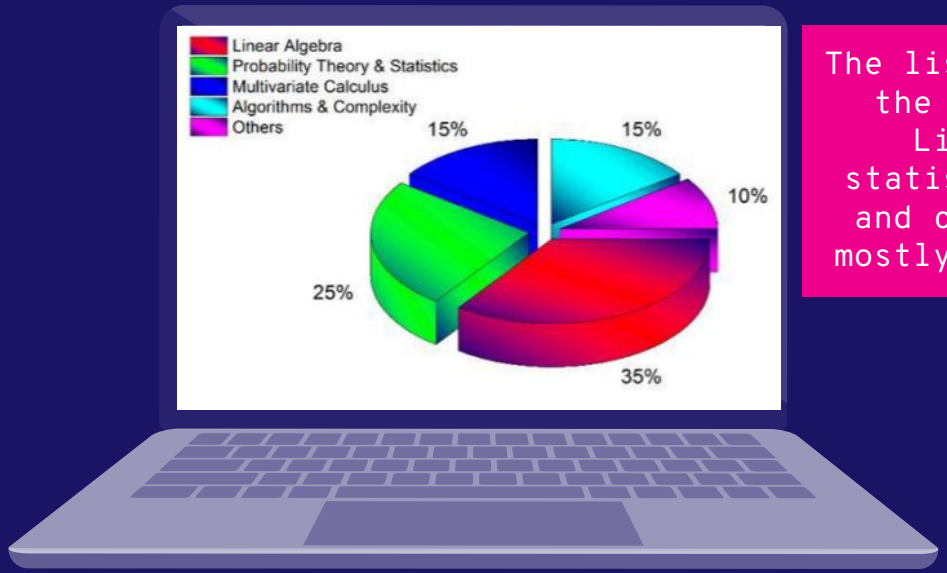
https://www.codelantic.com/why-big-data-is-growing-so-fast/

What can I prepare for answer the challenges?

1. Analytical Skills
2. Data Visualization Skills
3. Familiarity with Business Domain and Big Data Tools
4. Skills of Programming
5. Problem Solving Skills
6. SQL - Structured Query Language
7. Skills of Data Mining
8. Familiarity with Technologies
9. Familiarity With Public Cloud and Hybrid Clouds
10. Skills from Hands-on experience

https://www.upgrad.com/blog/big-data-skills/

KEY MATHEMATICAL CONCEPTS IN BIG DATA ANALYTICS



The list of mathematical methods mostly used in the analysis of big data, to name some are Linear Algebra, Probability theory and statistics, Multivariate calculus, Algorithms and complexity and the lists goes that would mostly represent any data facet mathematically

Ex : Linear Algebra concepts of optimization techniques used in machine learning

a. Matrix Operation
b. Vector Spaces
c. Eigen Values
d. Eigen Vector
e. Orthogonalization, etc

SQL

Programming

Data Mining

https://www.ijeat.org/wp-content/uploads/papers/v8i5S/E10430585S19.pdf

THANKS !

Do you have any questions?
Contact : linkedin.com/in/ferisaprestasi
Line : ferisatri

Credits : slidego