1. DATA SCIENCE FUNDAMENTALS

Table of Contents

1. Data Introduction 2
2. What is Data? 2
3. Information 2
4. Data Processing 2
5. Data, Information Examples 3
5.1 Example 3
5.2 Example 3
6. What is Data Science? 4
6.1 Definition 1 4
6.2 Definition 2 4
6.3 Scientist 4
6.4 Nature of Scientist 4
6.5 Process behind the scientist 4
7. BigData 6
8. Data Measurement table 6
9. Types of BigData 7
9.1 Structured 7
9.2 Semi structured 7
9.3 Unstructured data 7
10. V specialty in BigData 9
10.1 Volume 9
10.2 Variety 9
10.3 Velocity 9
11. Data Analytics 10
12. Different types of Analytics 10
12.1 Descriptive Analytics 10
12.2 Diagnostic Analytics 10
12.3 Predictive Analytics 10
12.4 Prescriptive Analytics 11

1. DATA SCIENCE FUNDAMENTALS

1. Data Introduction

- ✓ Currently we are living in the data world.
- ✓ Everyone is communicating by using devices and social networks, due to this huge amount of data is generating.
- ✓ All applications are generating data.
  - Ecommerce applications
  - Banking applications
  - Social network etc.

2. What is Data?

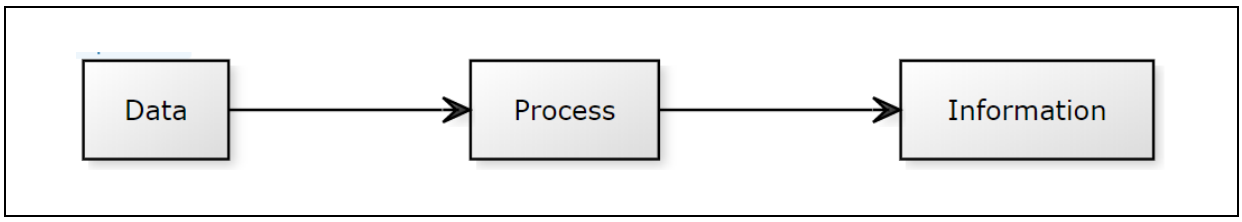
- ✓ Data is a collection of Facts.
- ✓ Facts,
  - Numbers
  - Alphabets
  - Alphanumeric
  - Symbols
  - Images
  - Audio
  - Video & etc

3. Information

- ✓ Data we may not understand clearly.
- ✓ The processed data is called as information.

4. Data Processing

- ✓ Converting the raw data into meaningful information

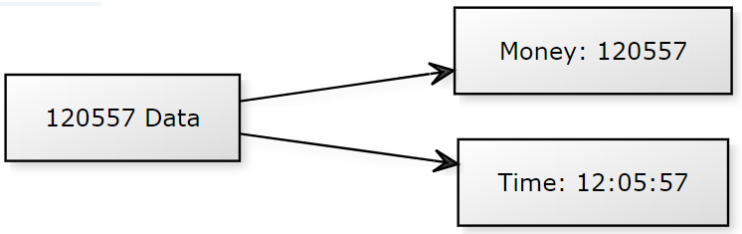


5. Data, Information Examples

5.1 Example

Data	Information
Daniel, Data Scientist, age is sweet sixteen, Bengaluru, India	Name : Daniel Profession : Data Scientist Age : 16 Location : Bengaluru Country : India

5.2 Example



6. What is Data Science?

6.1 Definition 1

- ✓ Data science is a combination of scientific methods, algorithms to extract knowledge and insights from data.

6.2 Definition 2

- ✓ Data science is all about how we take data,
  - Process data
  - Understand the past/present conditions
  - Decision making process
  - Predicting the future results
  - Create new industries/products condition

6.3 Scientist

- ✓ Scientist is a person who is having expert knowledge on specific scenarios.

6.4 Nature of Scientist

- ✓ A scientist used to conduct more research to achieve results.

6.5 Process behind the scientist

- ✓ In simple words a scientist is someone who,
  - Systematically gathers information
  - Doing research
  - Getting evidence
  - Test the evidence
  - Gaining knowledge
  - Understanding the knowledge.
  - Sharing knowledge

7. BigData

- ✓ Extremely large data sets is called as a BigData.

8. Data Measurement table

Data Measurement Chart
(Types of Various Units of Memory)

Bit	----	Single Binary Digit (1 or 0)
Byte	----	8 Bits
Kilo Byte	(KB)	1,024 Bytes
Mega Byte	(MB)	1,024 Kilobytes
Giga Byte	(GB)	1,024 Megabytes
Tera Byte	(TB)	1,024 Gigabytes
Peta Byte	(PB)	1,024 Terabytes
Exa Byte	(EB)	1,024 Petabytes
Zetta Byte	(ZB)	1,024 Exabytes
Yotta Byte	(YB)	1,024 Zettabyte

9. Types of BigData

- ✓ There are mainly 3 types of big data.
  - Structured
  - Semi-structured
  - Unstructured

9.1 Structured

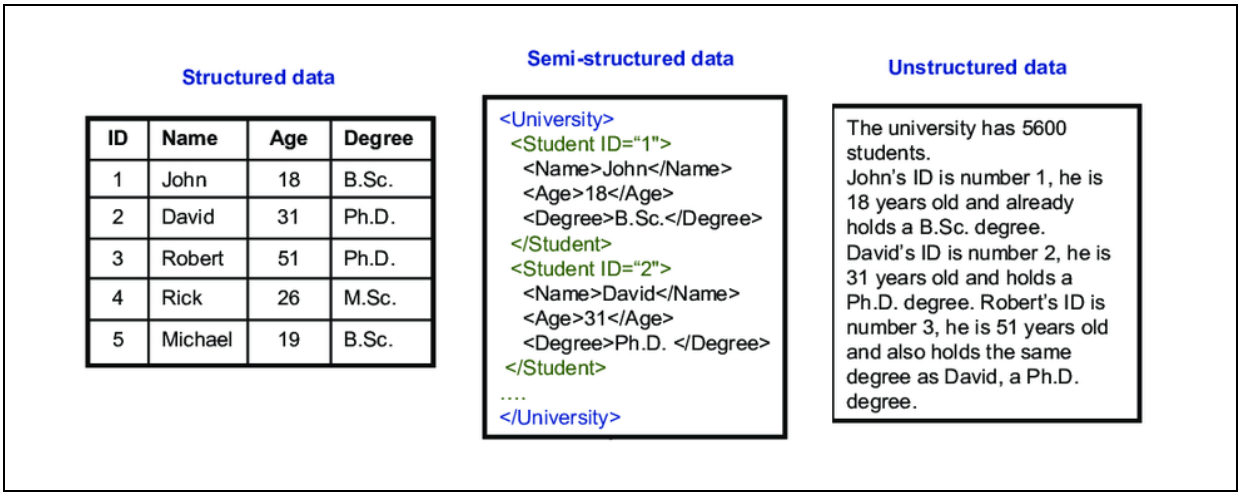
- ✓ Structured data will be stored in table formats like rows and columns.
- ✓ It is very easy to understand.
- ✓ Examples like, a table in database, excel file, csv file.

9.2 Semi structured

- ✓ Semi structured may not in proper format
- ✓ This type of data having some properties or tags.
- ✓ Example
  - Xml file
  - JSON file

9.3 Unstructured data

- ✓ Unstructured data have no proper format.
- ✓ In this world mostly unstructured data exits.
- ✓ Example
  - log file
  - audio, video
  - e-mail & etc


Structured data
ID Name Age Degree
1 John 18 B.Sc.
2 David 31 Ph.D.
3 Robert 51 Ph.D.
4 Rick 26 M.Sc.
5 Michael 19 B.Sc.
Semi-structured data
<University>
<Student ID="1">
<Name>John</Name>
<Age>18</Age>
<Degree>B.Sc.</Degree>
</Student>
<Student ID="2">
<Name>David</Name>
<Age>31</Age>
<Degree>Ph.D. </Degree>
</Student>
....
</University>
Unstructured data
The university has 5600 students.
John's ID is number 1, he is 18 years old and already holds a B.Sc. degree.
David's ID is number 2, he is 31 years old and holds a Ph.D. degree. Robert's ID is number 3, he is 51 years old and also holds the same degree as David, a Ph.D. degree.

10. V specialty in BigData

10.1 Volume

- ✓ Volume of the data means, size of the data.
- ✓ Based on size only we can measure as the data is in big data or not.

10.2 Variety

- ✓ Variety of data means, the generated data may existing in different formats like
- ✓ It can be,
  - Structured
  - Semi-structured
  - Unstructured.

10.3 Velocity

- ✓ Velocity of data means, speed of the data generating and processing from the source points.

In short

- Volume = Size
- Variety = Types
- Velocity = Speed

11. Data Analytics

- ✓ All companies are analysing big data to improve the business.
- ✓ From the result of analysis, companies used to reach customers to understand them clearly.

11.1 Every company's target

- ✓ Ultimate target of any company is, wanted to stand in best position, and should be competitive with other companies.

12. Different types of Analytics

- ✓ There are mainly four types of analytics,
    1. Descriptive Analytics
    2. Diagnostic Analytics
    3. Predictive Analytics
    4. Prescriptive Analytics

12.1 Descriptive Analytics

- ✓ Descriptive analytics helps to understand like what is happening?
- ✓ Descriptive analytics provide the information about happened situations.
- ✓ Examples
    - Amazon offered one product then how the customers are buying that specific product.

12.2 Diagnostic Analytics

- ✓ Diagnostic analytics helps to understand like why did it happen?
- ✓ It is used to do a deeper understanding about the data to find out the cause of events and behaviours.

12.3 Predictive Analytics

- ✓ Predictive analytics helps to understand like what will happen in future?
- ✓ It is used to understand the future expectations, trends and provide the information about what will be happen in future.

12.4 Prescriptive Analytics

- ✓ Prescriptive analytics helps to understand like what is the next best action?
- ✓ It is used to apply the best decision actions for improvement.