Data Sciences using Python (05101305)

Dr. Kamini Solanki, (Associate Professor)
Parul Institute of Computer Application - BCA

CHAPTER-1

Overview of Data Science

Introduction of Data Science

What is Data Science and few definition of Data Science

- Data Science mean study of Data, analyze, visualize and perform some mathematical operation on the data.
- Here Data mean data of any XYZ Company, Data of any organization or any huge Data.
- So Basically Data mining and Data Science both are differ. Data mining is subset of Data Science.
- It involves many different disciplines like mathematical and statistical modelling, extracting data from it source and applying data visualization techniques.
- Is use to work for structured or un-structured data both.
- It is an automated way to analyze huge amounts of data and extract any kind of information on any data.

Cont...

- It is a new discipline that combines the aspects of statistics, mathematics, programming, and visualization to turn data into information.
- Data science is a multi-disciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from structured and unstructured data.
- Data science uses most powerful hardware, programming systems, and efficient algorithms for solving data related problems. So, Data Science is the future of artificial intelligence.
- Data Scientists collect, explore, analyze, and visualize data. They apply mathematical and statistical models to find patterns and solutions in the data.

Cont...

So we can say Data Science is all about

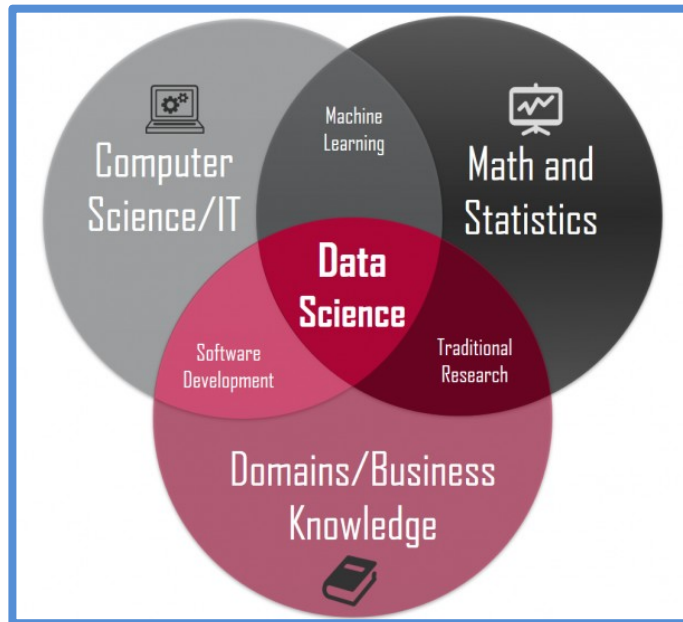Image Source : Javatpoint.com

Cont...
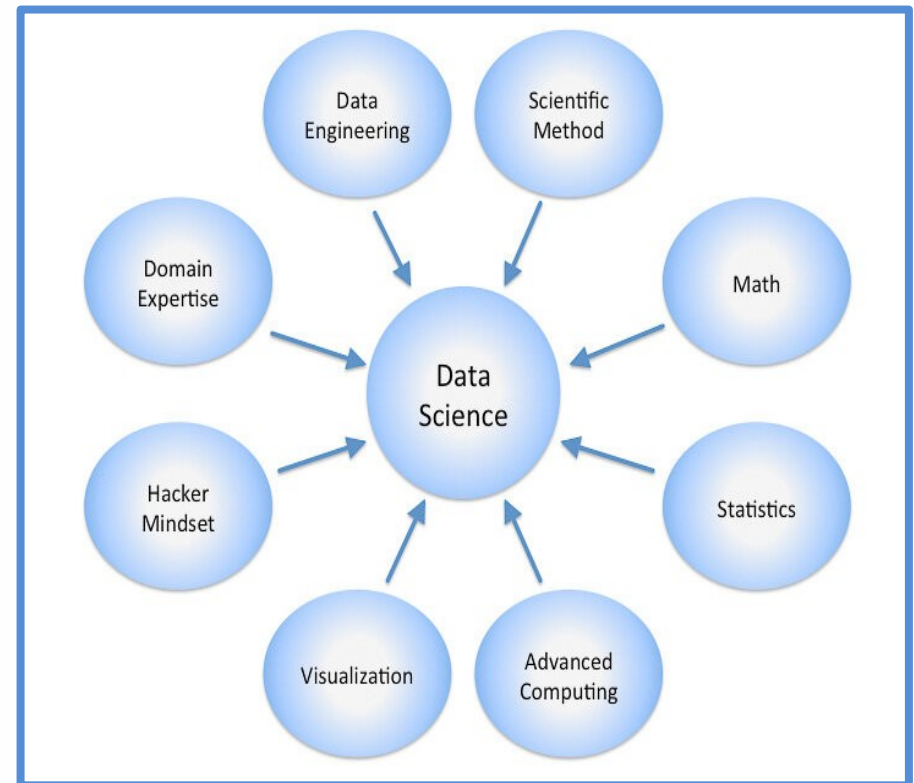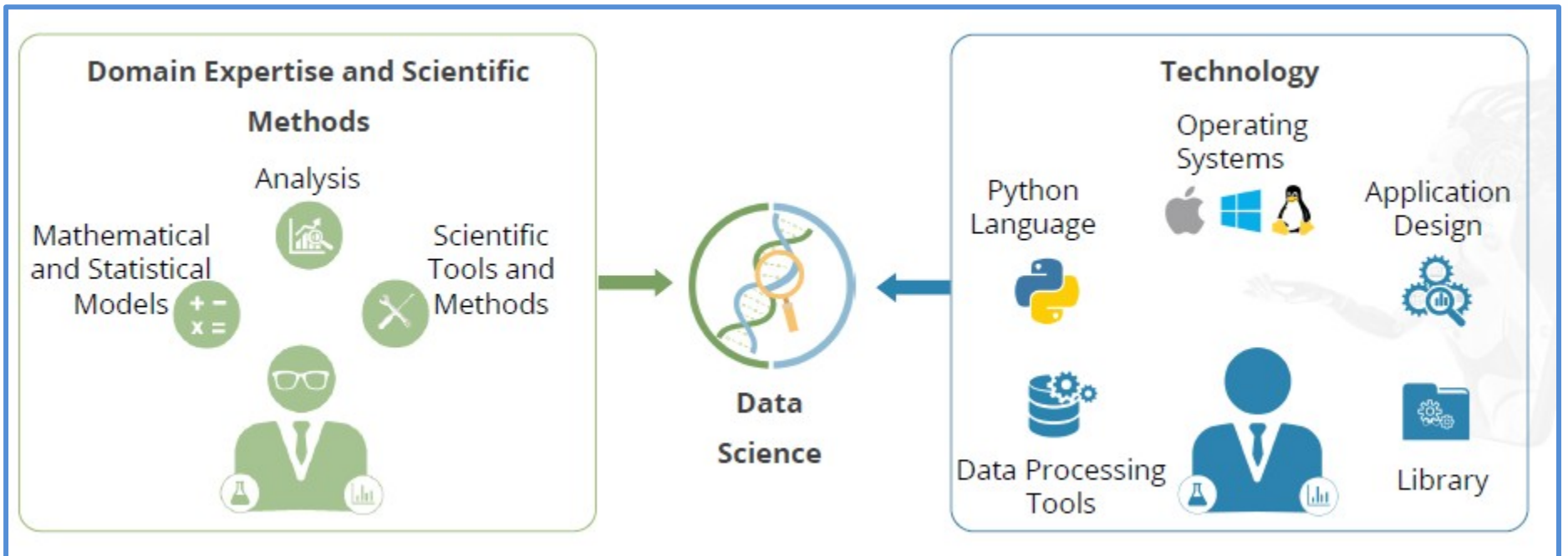

Image Source : towardsdatascience.com

Image Source: innoarchitech.com/

Cont...

When you combine domain expertise and scientific methods with technology, you get Data Science.
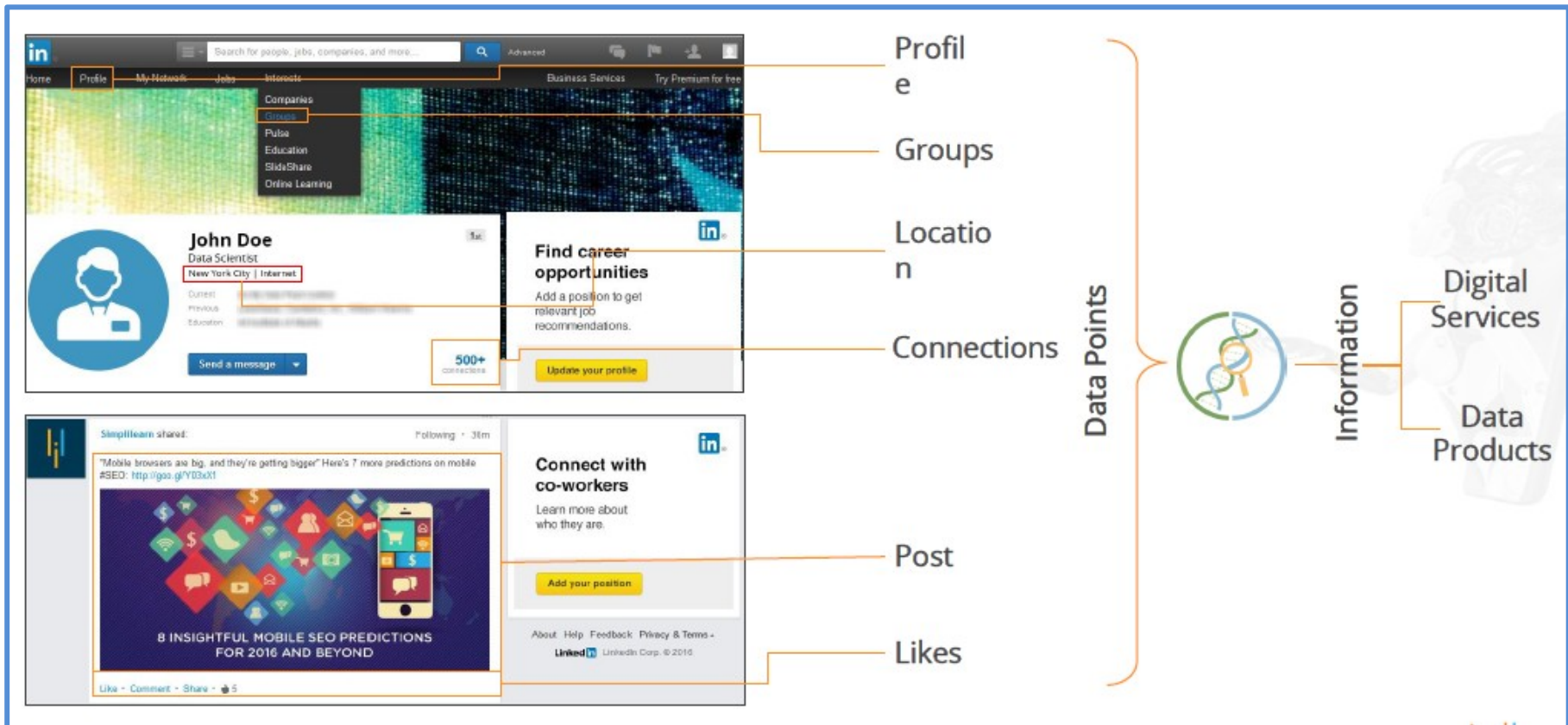

Image Source : Simplilearn

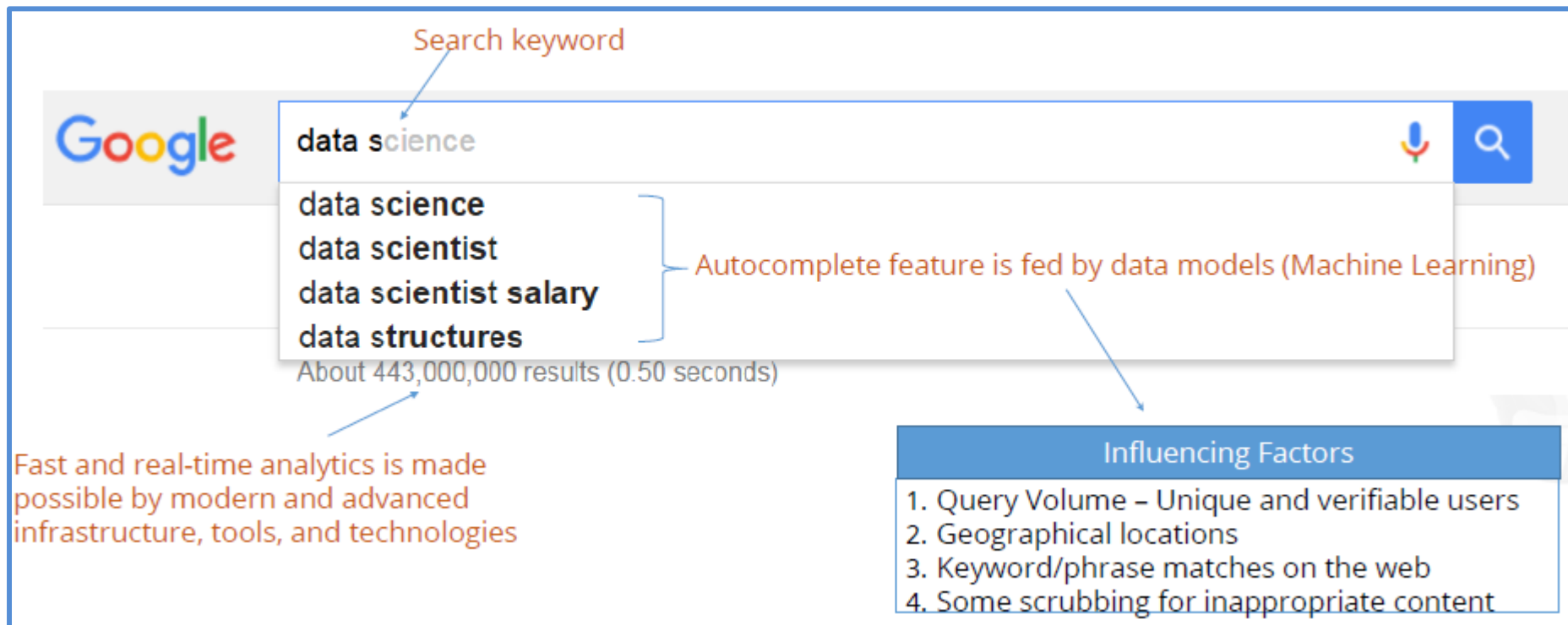# Different Sectors Using Data Science

Image Source : Simplilearn

LinkedIn using Data science

Image Source : Simplilearn

Data Science for search engine

Image Source : Simplilearn

Data science Healthcare

Wearable devices use Data Science to analyze data gathered by their biometric sensors.

Wearable device

Biometric data transfer

IoT Gateway

Data transfer to Servers

Enterprise Infrastructure

Make informed decisions

Engagement Dashboard

Data Analytics

Image Source : Simplilearn
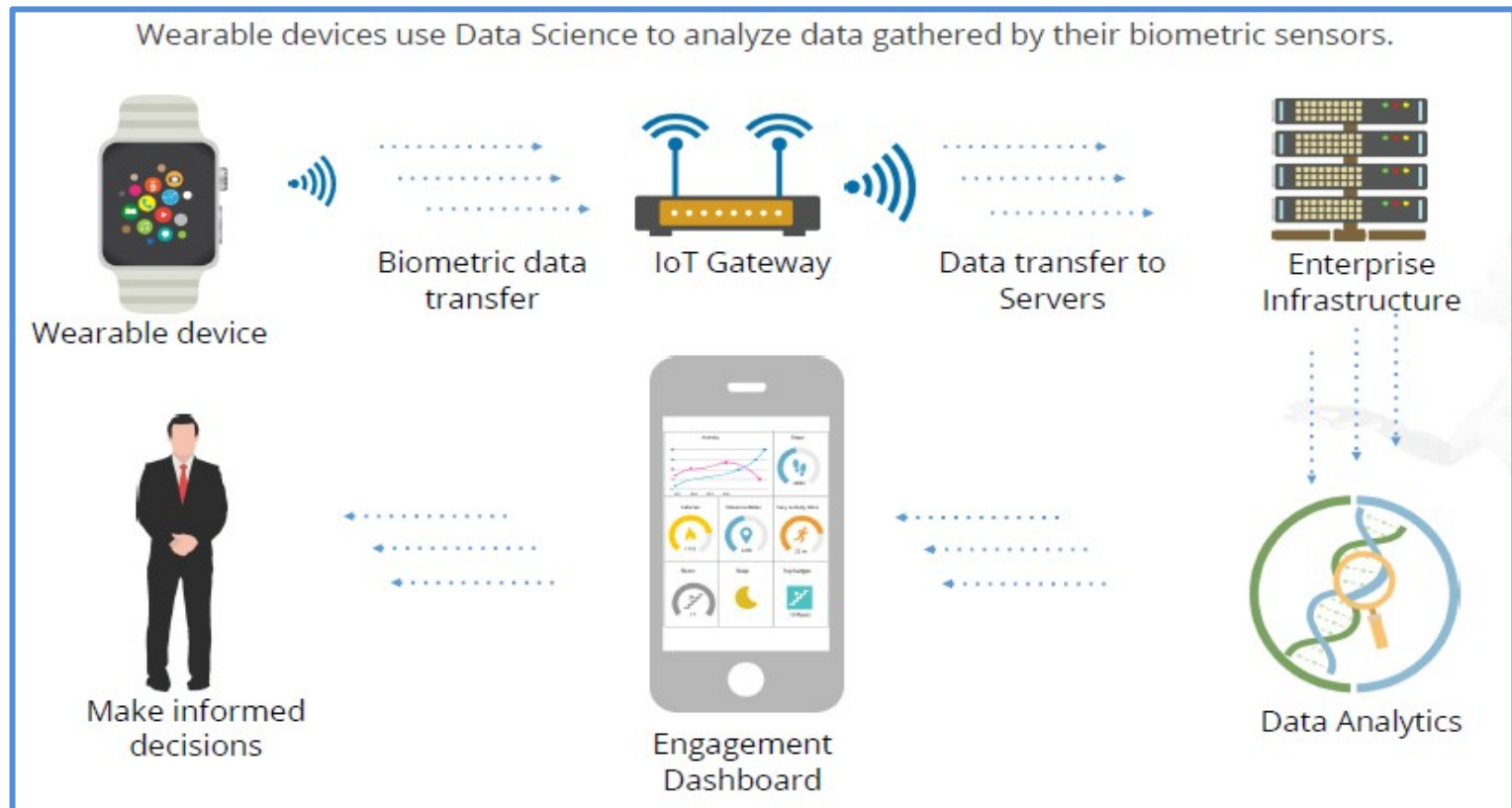
Data science in Finance sector

A loan manager can easily access and sift through a loan applicant's financial details using Data Science.

Loan Applicant

Loan Application Portal

Data transfer to Servers

Enterprise Infrastructure

APPROVED

DENIED

Make informed decisions

Credit report, Credit history, Approved amount, Risk etc.

Engagement Dashboard

Data Analytics

Image Source : Simplilearn

Data science for Public Sector

Image Source : Simplilearn

Parul® University

Challenge for Data Scientist or Data Analytics

- Biggest challenge for Data Scientist or Data Analytics is to collect the data.
- Because all the organization store its data in different format like some of are store in Data base, or excel file or in CSV file or in large cluster.
- Its become harder to analyze or arrange structure and un structured data.
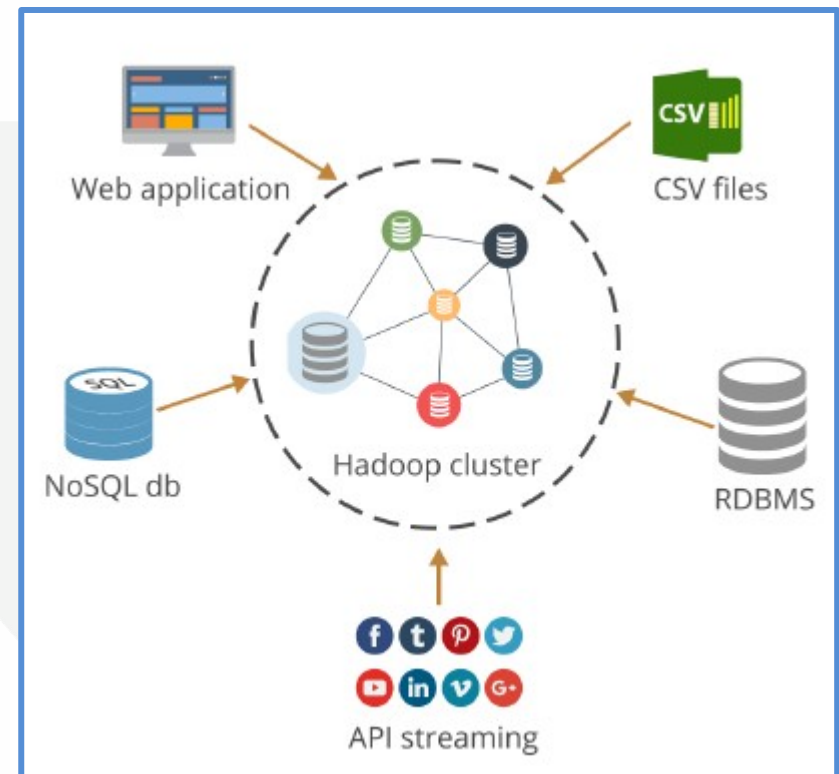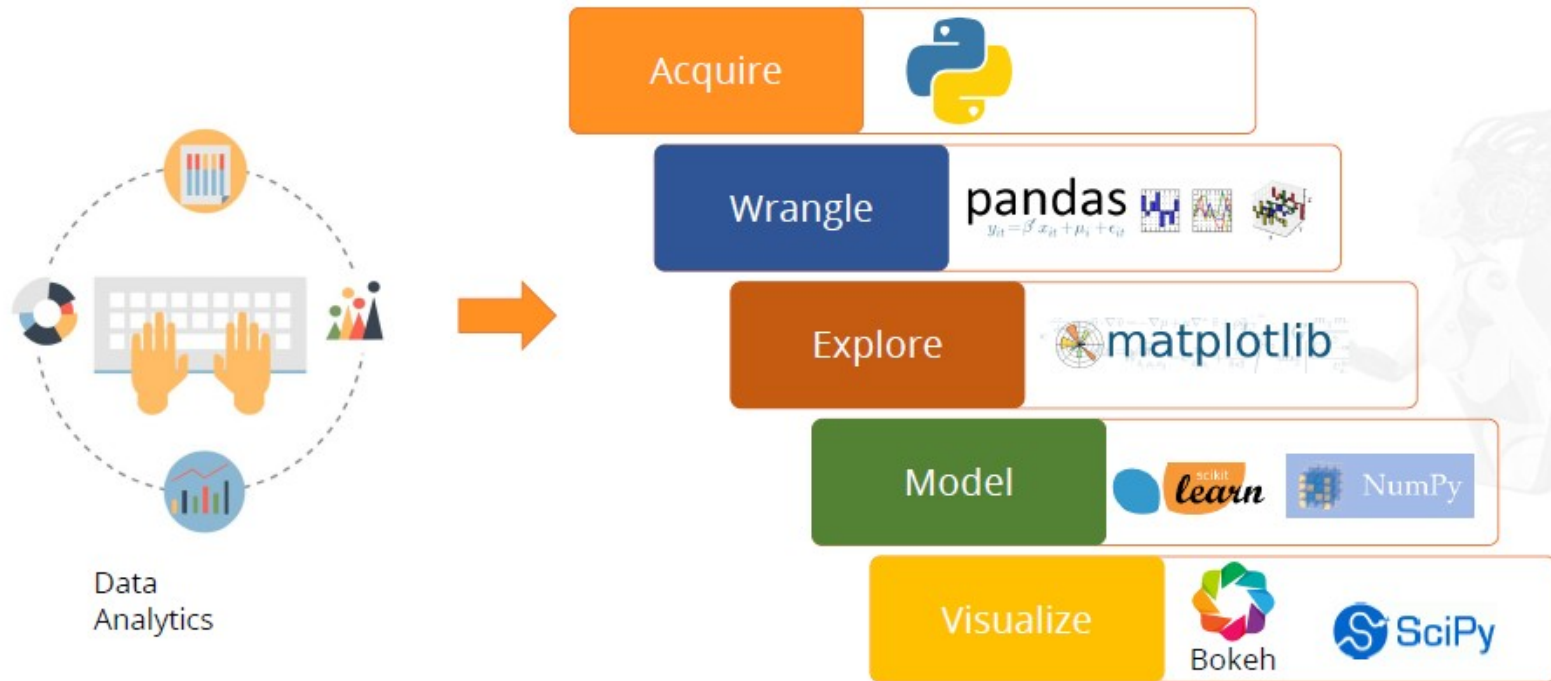

Image Source : Simplilearn

Purpose and Components of Python

Why Python is use for Data Science ?

- It a open source so there is no need to pay single penny for use it. One can download it install it and use it.
- For implementing the python program lots of IDLE is available even one can implement the program on google colab also.
- It provide numbers of freely available library file for different purpose in data science.
- Deep learning frameworks available with Python APIs even if some person wants to implement some system for fraud detection than also python provide NLP based algorithm.

Cont...

Python deals with each stage of data analytics efficiently by applying different libraries and packages.

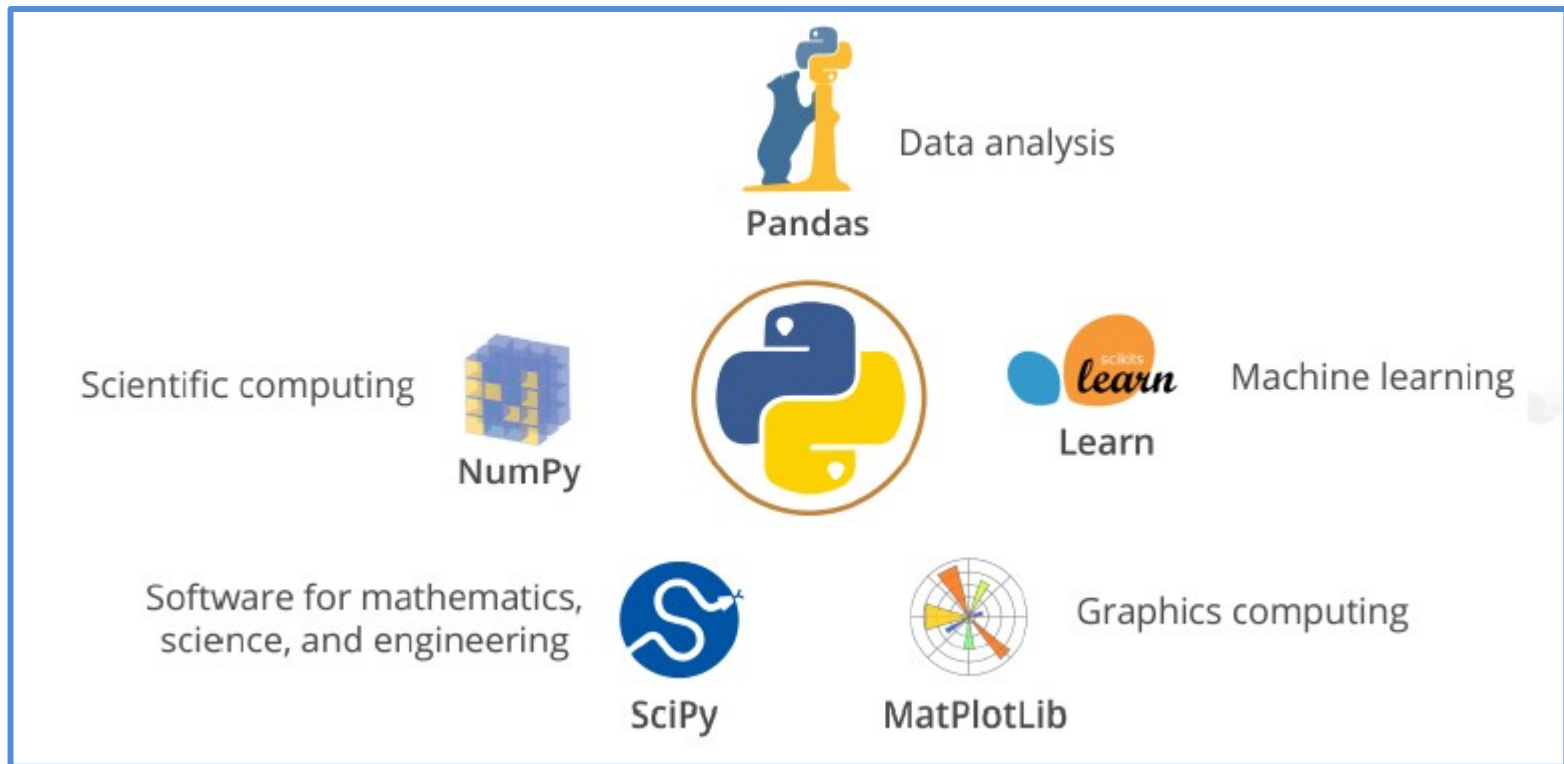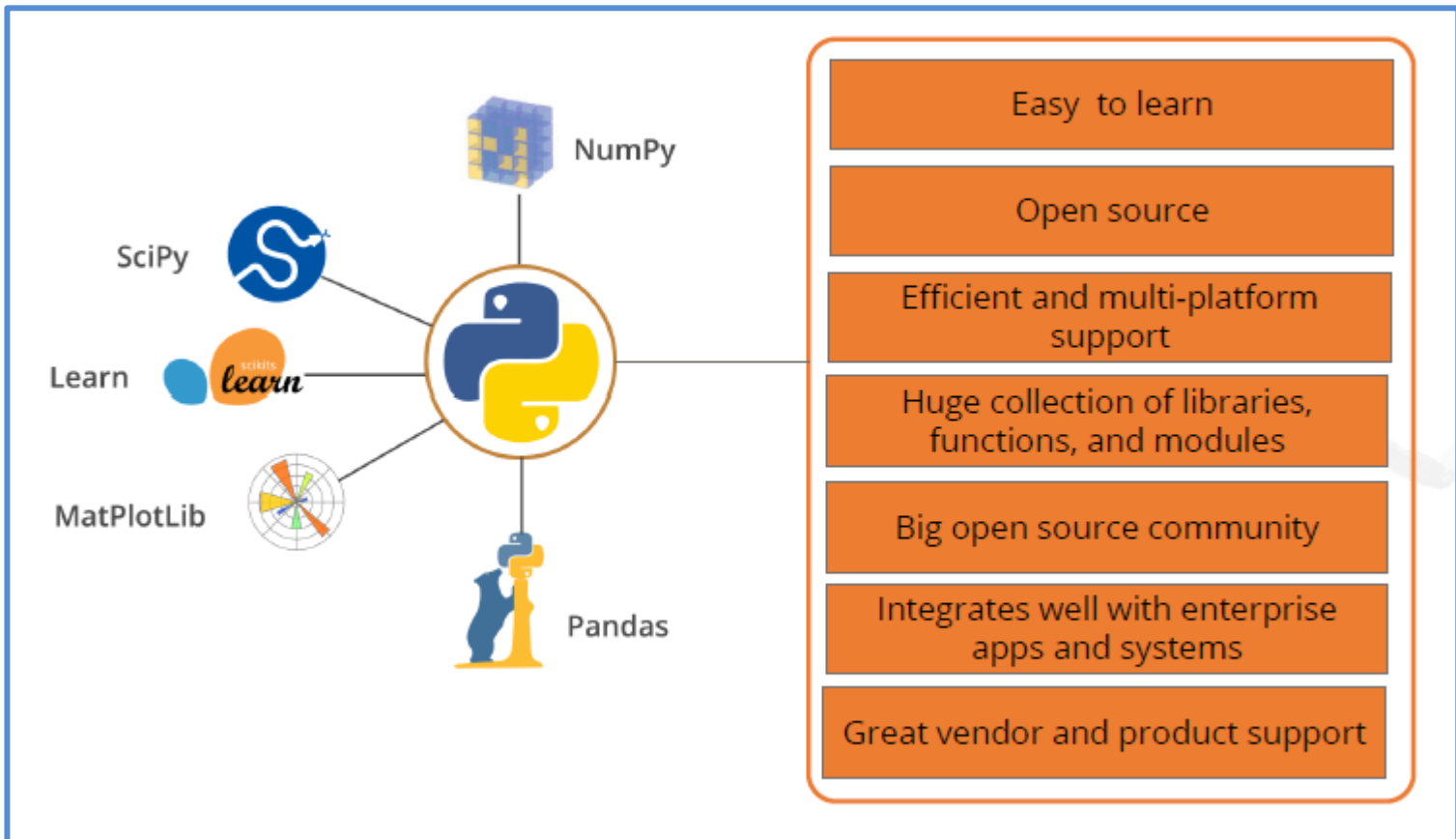- Acquire — python
- Wrangle — pandas $y_{it} = \beta' x_{it} + \mu_i + e_{it}$
- Explore — matplotlib
- Model — scikit learn, NumPy
- Visualize — Bokeh, SciPy

Data Analytics

Image Source : Simplilearn

Cont...

Image Source : Simplilearn

Benefits of Python

NumPy

SciPy

Learn

MatPlotLib

Pandas

- Easy to learn
- Open source
- Efficient and multi-platform support
- Huge collection of libraries, functions, and modules
- Big open source community
- Integrates well with enterprise apps and systems
- Great vendor and product support

Cont...

Python is supported by well-established data platforms and processing frameworks that help analyze data in a simple and efficient way.

Data Scientist

Python — Tool for data analysis and processing
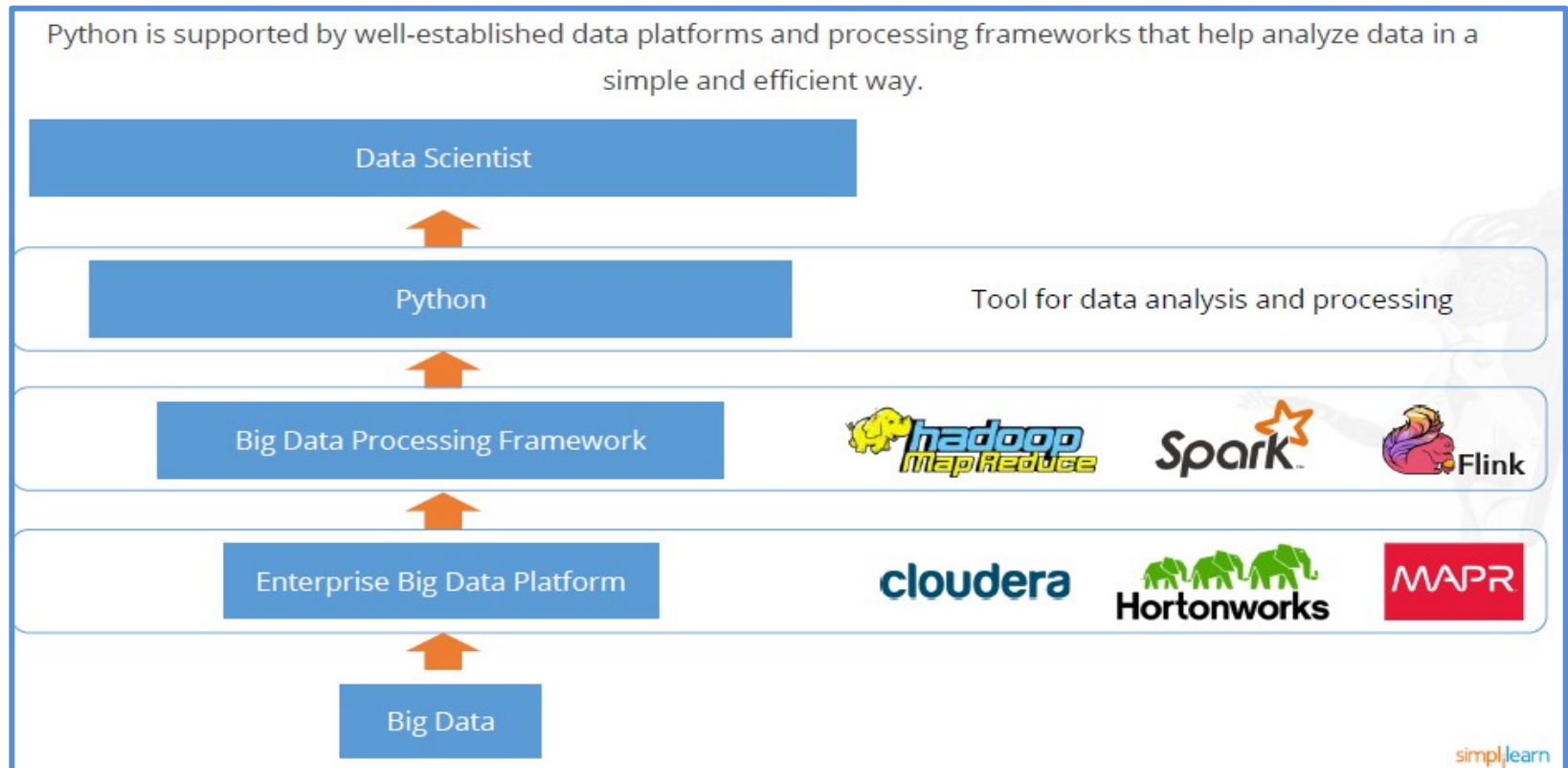
Big Data Processing Framework

Enterprise Big Data Platform

Big Data

Image Source : Simplilearn

DIGITAL LEARNING CONTENT

Parul® University

www.paruluniversity.ac.in