Atmiya University
Faculty of Science,
Department of Computer Science & I.T.

Subject Name: **21UFSDE309 Data Science Using Python**

By: Dr. Hiren Kavathiya

What is Data Science?

- **Data Science** is the area of study which involves extracting insights from vast amounts of data using various scientific methods, algorithms, and processes. It helps you to discover hidden patterns from the raw data. The term Data Science has emerged because of the evolution of mathematical statistics, data analysis, and big data
- Data Science is an interdisciplinary field that allows you to extract knowledge from structured or unstructured data. Data science enables you to translate a business problem into a research project and then translate it back into a practical solution.

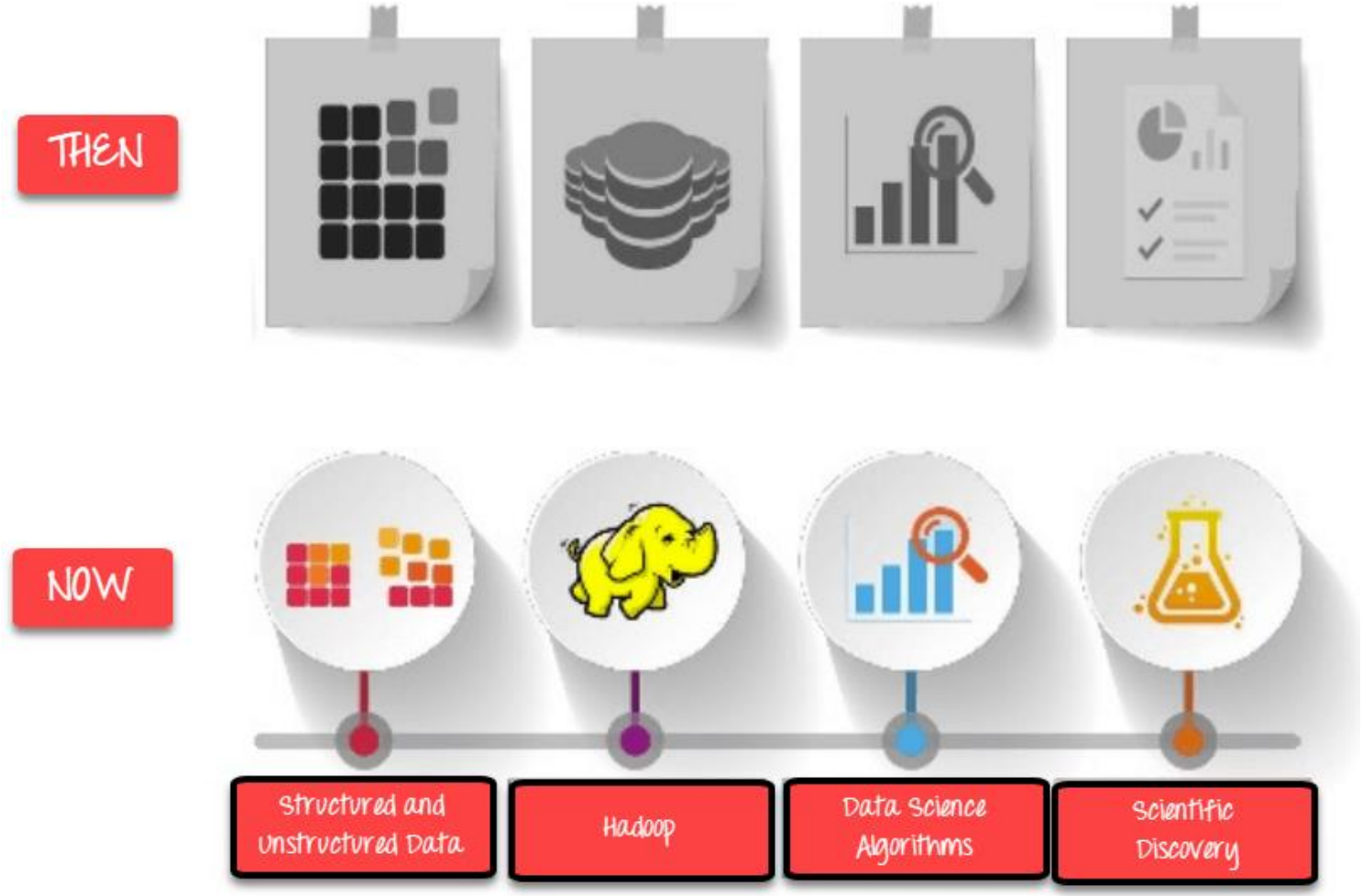
Why Data Science?

- Here are significant advantages of using Data Analytics Technology:
  - Data is the oil for today's world. With the right tools, technologies, algorithms, we can use data and convert it into a distinct business advantage
  - Data Science can help you to detect fraud using advanced machine learning algorithms
  - It helps you to prevent any significant monetary losses
  - Allows to build intelligence ability in machines

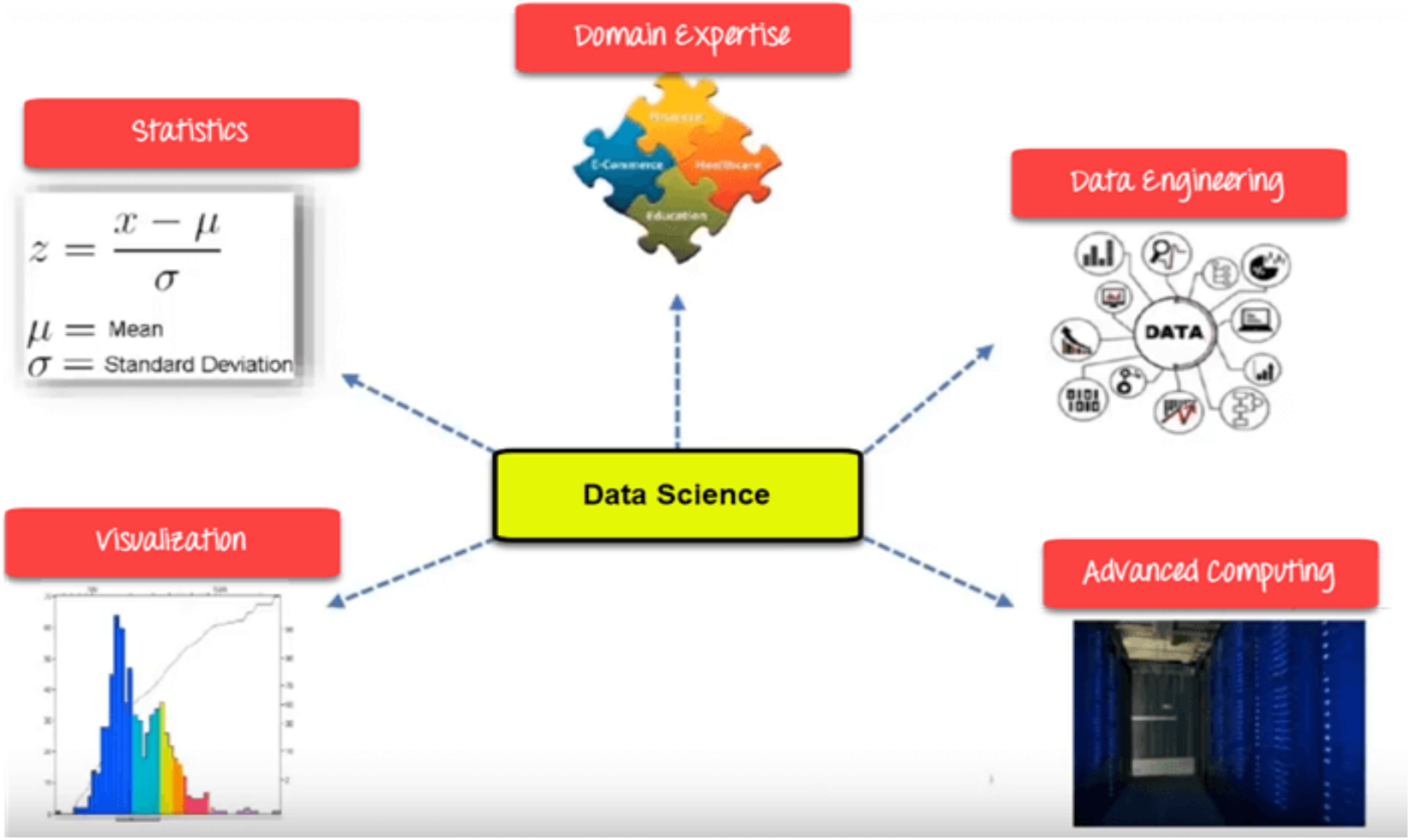
Why Data Science?

- You can perform sentiment analysis to gauge customer brand loyalty
- It enables you to take better and faster decisions
- It helps you to recommend the right product to the right customer to enhance your business

Why Data Science?



Data Science Components



Data Science Components

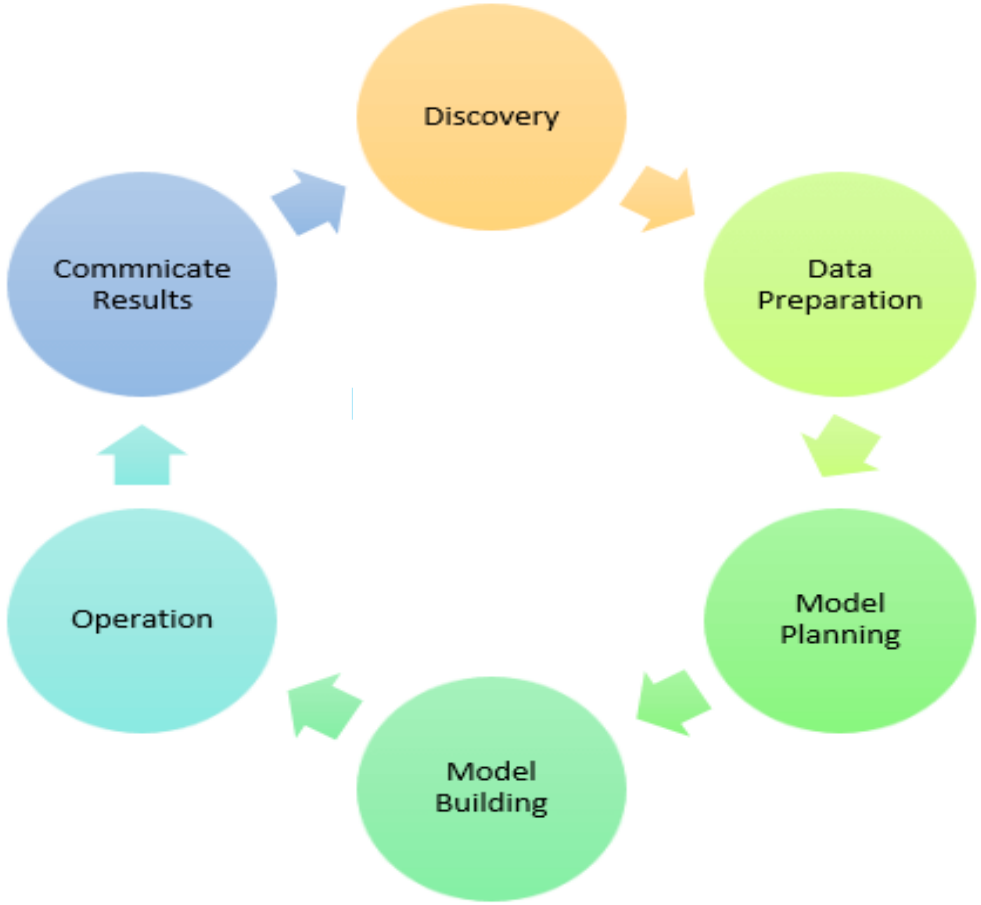
- **Statistics:**
  - Statistics is the most critical unit of Data Science basics, and it is the method or science of collecting and analyzing numerical data in large quantities to get useful insights.
- **Visualization:**
  - Visualization technique helps you access huge amounts of data in easy to understand and digestible visuals.
- **Machine Learning:**
  - Machine Learning explores the building and study of algorithms that learn to make predictions about unforeseen/future data.

Data Science Components

- **Deep Learning:**
  - Deep Learning method is new machine learning research where the algorithm selects the analysis model to follow.

Data Science Process

- Now in this Data Science , we will learn the Data Science Pr



Data Science Process

1. Discovery:

- Discovery step involves acquiring data from all the identified internal & external sources, which helps you answer the business question.
- The data can be:
  - Logs from webservers
  - Data gathered from social media
  - Census datasets
  - Data streamed from online sources using APIs

Data Science Process

- **2. Preparation:**
  - Data can have many inconsistencies like missing values, blank columns, an incorrect data format, which needs to be cleaned. You need to process, explore, and condition data before modelling. The cleaner your data, the better are your predictions.
- **3. Model Planning:**
  - In this stage, you need to determine the method and technique to draw the relation between input variables. Planning for a model is performed by using different statistical formulas and visualization tools. SQL analysis services, R, and SAS/access are some of the tools used for this purpose.

Data Science Process

- **4. Model Building:**
  - In this step, the actual model building process starts. Here, Data scientist distributes datasets for training and testing. Techniques like association, classification, and clustering are applied to the training data set. The model, once prepared, is tested against the "testing" dataset.
- **5. Operationalize:**
  - You deliver the final baselined model with reports, code, and technical documents in this stage. Model is deployed into a real-time production environment after thorough testing.

Data Science Process

6. Communicate Results

- In this stage, the key findings are communicated to all stakeholders. This helps you decide if the project results are a success or a failure based on the inputs from the model.

Data Science Jobs Roles

- Most prominent Data Scientist job titles are:
  - Data Scientist
  - Data Engineer
  - Data Analyst
  - Statistician
  - Data Architect
  - Data Admin
  - Business Analyst
  - Data/Analytics Manager

Data Science Jobs Roles

Data Scientist:

- **Role:** A Data Scientist is a professional who manages enormous amounts of data to come up with compelling business visions by using various tools, techniques, methodologies, algorithms, etc.
- **Languages**: R, SAS, Python, SQL, Hive, Matlab, Pig, Spark

Data Engineer:

- **Role**: The role of a data engineer is of working with large amounts of data. He develops, constructs, tests, and maintains architectures like large scale processing systems and databases.
- **Languages**: SQL, Hive, R, SAS, Matlab, Python, Java, Ruby, C + +, and Perl

Data Science Jobs Roles

- **Data Analyst:**
  - **Role**: A data analyst is responsible for mining vast amounts of data. They will look for relationships, patterns, trends in data. Later he or she will deliver compelling reporting and visualization for analyzing the data to take the most viable business decisions.
  - **Languages**: R, Python, HTML, JS, C, C+ + , SQL
- **Statistician:**
  - **Role**: The statistician collects, analyses, and understands qualitative and quantitative data using statistical theories and methods.
  - **Languages**: SQL, R, Matlab, Tableau, Python, Perl, Spark, and Hive

Data Science Jobs Roles

- **Data Administrator:**
  - **Role**: Data admin should ensure that the database is accessible to all relevant users. He also ensures that it is performing correctly and keeps it safe from hacking.
  - **Languages**: Ruby on Rails, SQL, Java, C#, and Python
- **Business Analyst:**
  - **Role**: This professional needs to improve business processes. He/she is an intermediary between the business executive team and the IT department.
  - **Languages**: SQL, Tableau, Power BI and, Python

Tools for Data Science

Data Analysis	Data Warehousing	Data Visualization	Machine Learning
R, **Spark, Python and SAS**	Hadoop, SQL, Hive	R, Tableau, Raw	Spark, Azure ML studio

Difference Between Data Science with BI (Business Intelligence)

Parameters	Business Intelligence	Data Science
Perception	Looking Backward	Looking Forward
Data Sources	Structured Data. Mostly SQL, but some time Data Warehouse	Structured and Unstructured data.Like logs, SQL, NoSQL, or text
Approach	Statistics & Visualization	Statistics, Machine Learning, and Graph
Emphasis	Past & Present	Analysis & Neuro-linguistic Programming
Tools	Pentaho, Microsoft Bl, QlikView,	R, TensorFlow

Applications of Data Science

- Some application of Data Science are:
- **Internet Search:**
  - Google search uses Data science technology to search for a specific result within a fraction of a second
- **Recommendation Systems:**
  - To create a recommendation system. For example, "suggested friends" on Facebook or suggested videos" on YouTube, everything is done with the help of Data Science.
- **Image & Speech Recognition:**
  - Speech recognizes systems like Siri, Google Assistant, and Alexa run on the Data science technique. Moreover, Facebook recognizes your friend when you upload a photo with them, with the help of Data Science.

Applications of Data Science

- **Gaming world:**
  - EA Sports, Sony, Nintendo are using Data science technology. This enhances your gaming experience. Games are now developed using Machine Learning techniques, and they can update themselves when you move to higher levels.
- **Online Price Comparison:**
  - PriceRunner, Junglee, Shopzilla work on the Data science mechanism. Here, data is fetched from the relevant websites using APIs.

Introduction of Python

- Python is a high-level, interpreted, interactive and object-oriented scripting language.
- Python is designed to be highly readable. It uses English keywords frequently where as other languages use punctuation, and it has fewer syntactical constructions than other languages.

Introduction of Python

- **Python is Interpreted**: Python is processed at runtime by the interpreter. You do not need to compile your program before executing it. This is similar to PERL and PHP.
- **Python is Interactive**: You can actually sit at a Python prompt and interact with the interpreter directly to write your programs.
- **Python is Object-Oriented**: Python supports Object-Oriented style or technique of programming that encapsulates code within objects.

Introduction of Python

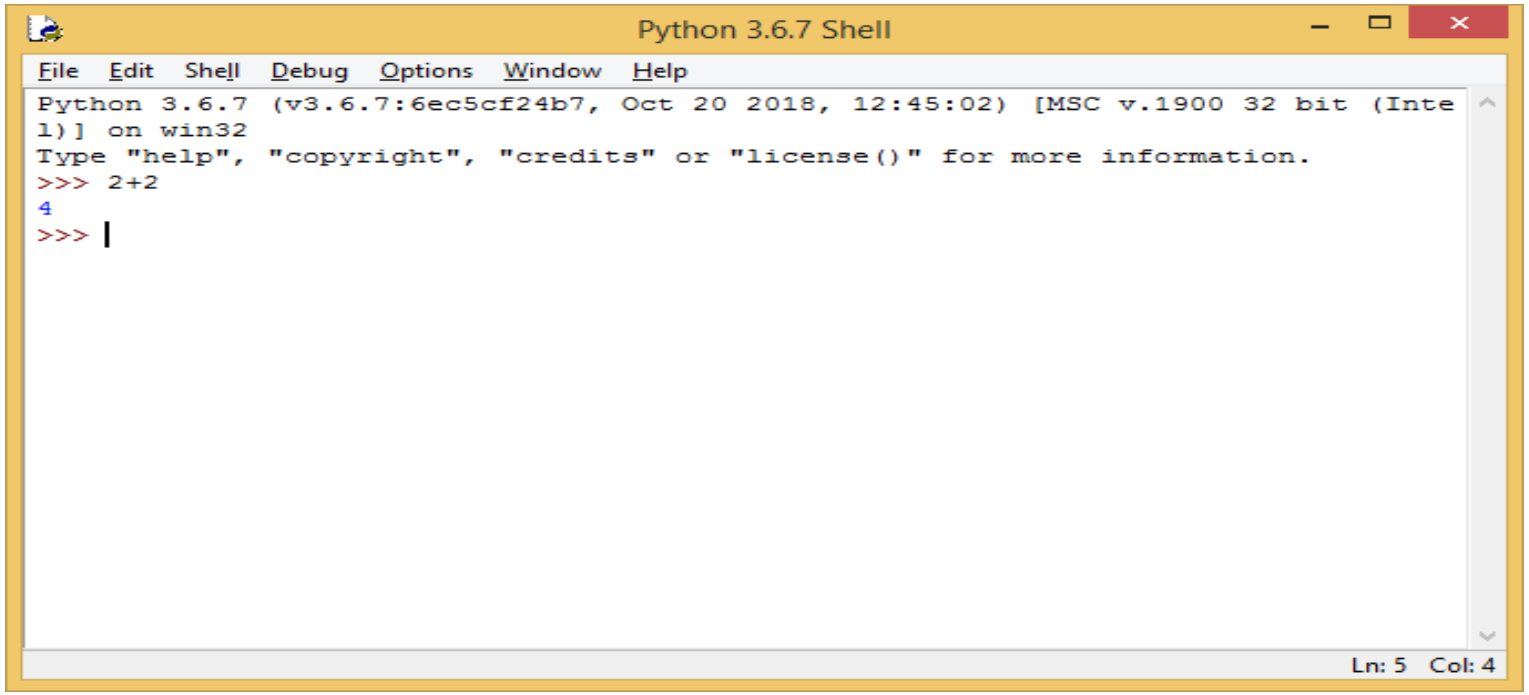
- **Python is a Beginner's Language**: Python is a great language for the beginner-level programmers and supports the development of a wide range of applications from simple text processing to WWW browsers to games.
- Python is an example of a high-level language; other high-level languages you might have heard of are C++, PHP, Pascal, C#, and Java.
- Thus, programs written in a high-level language have to be translated into something more suitable before they can run.

Introduction of Python

- Almost all programs are written in high-level languages because of their advantages. It is much easier to program in a high-level language so programs take less time to write, they are shorter and easier to read, and they are more likely to be correct.
- Second, high-level languages are portable, meaning that they can run on different kinds of computers with few or no modifications.

Introduction of Python

- The engine that translates and runs Python is called the Python Interpreter. There are two ways to use it:
  - **immediate mode** and **script mode**. In immediate mode, you type Python expressions into the Python Interpreter window, and the interpreter immediately shows the result:



Introduction of Python

- The **>>> is called the Python prompt**. The interpreter uses the prompt to indicate that it is ready for instructions.

Python History

- Invented at the ***National Research Institute for Mathematics and Computer Science in the Netherlands***, early 90s by **Guido Van Rossum**
- Python was conceived in the late 1980s and its implementation was started in December 1989
- Python is derived from many other languages, including ABC, Modula-3, C, C++, Algol-68, SmallTalk, Unix shell, and other scripting languages.
- Open sourced from beginning
- Considered a scripting language, but is much more Scalable, object oriented and functional from the beginning
- Used by Google from the beginning

Python History

- Author is Guido Van Rossum
- Convinced in Late 1980
- Finally released in 1991
- Start implementation in December 1989 at CWI in Netherland.
- Successor of ABC Programming language.
- Name came from BBC's TV Show – 'Monty Python's Flying Circus'

Features of Python

- ***Easy-to-learn***: Python has few keywords, simple structure, and a clearly defined syntax.
- ***Easy-to-read:*** Python code is more clearly defined and visible to the eyes.
- ***Easy-to-maintain:*** Python's source code is fairly easy-to-maintain.
- ***A broad standard library:*** Python's bulk of the library is very portable and cross- platform compatible on UNIX, Windows, and Macintosh.
- ***Interactive Mode:*** Python has support for an interactive mode which allows interactive testing and debugging of snippets of code.
- ***Portable:*** Python can run on a wide variety of hardware platforms and has the same interface on all platforms.

Features of Python

- ***Extendable:*** You can add low-level modules to the Python interpreter. These modules enable programmers to add to or customize their tools to be more efficient.
- ***Databases:*** Python provides interfaces to all major commercial databases.
- ***GUI Programming:*** Python supports GUI applications that can be created and ported to many system calls, libraries, and windows systems, such as Windows MFC, Macintosh, and the X Window system of UNIX.
- ***Scalable***: Python provides a better structure and support for large programs than shell scripting.

Installing Python

- Python is pre-installed on most Unix systems, including Linux and MAC OS X
- But for Windows Operating Systems, user can download from the https://www.python.org/downloads/
- Form the above link download latest version of python IDE and install.