Big Data Summer Training

BigData Analytics-Python Programming

Prepared and Presenting By: Amrit Chhetri (Certified BigData Analyst),
Principal Techno-Functional Consultant and Principal IT Security Consultant

About Me :

- Me:
  - I'm Amrit Chhetri from Bara Mangwa, West Bengal, India, a beautiful Village/Place in Darjeeling.
  - I am CSCU, CEH, CHFI,CPT, CAD, CPD, IOT & BigData Analyst(University of California), Information Security Specialist(Open University, UK) and Machine Learning Enthusiast (University of California[USA] and Open University[UK]), Certified Cyber Physical System Exert(Open University[UK]) and Certified Smart City Expert.
- Current Position:
  - Principal IT Security Consultant/Instructor, Principal Forensics Investigator and Principal Techno-Functional Consultant/Instructor .
  - BigData Consultant to KnowledgeLab.
- Experiences:
  - I was J2EE Developer and BI System Architect/Designer of DSS for APL and Disney World
  - I have played the role of BI Evangelist and Pre-Sales Head for BI System* from OST
  - I have worked as Business Intelligence Consultant for national and multi-national companies including HSBC, APL, Disney, Fidedality , LG(India) , Fidelity, BOR(currently ICICI), Reliance Power. * *Top 5 Indian BI System (by NASSCOM)*

Data Science and BigData Processing with Python Training Session-I

Agendas/Modules-Session-I:

- Python Overview
- Advantages/Disadvantages
- Python Interpreter and PATH
- Python Console and Python IDLE
- Python Coding on LINUX
- Steps to run Python on Eclipse(Windows)
- Writing to Screen
- Python Editors and IDEs
- Python Identifiers, Variables and Keywords
- Sample Codes

Python Overview :

- Python is an interpreted, Objected Oriented, high-level and Interactive Programming Language.
- Python was created by Guido Van Rossum in late eighties.
- It is an Open Source Programming Platform.
- The official guides and other resources are available at http://python.org .
- Python Programming is massively on
  - BigData Processing
  - IT Security and Penetration Testing
  - Computer Forensics
  - Application Development

Advantages/Disadvantages :

- Advantages:
  - Python is easy to learn and write Scripting Language.
  - Python has rich set of built-in functions and third-party modules.

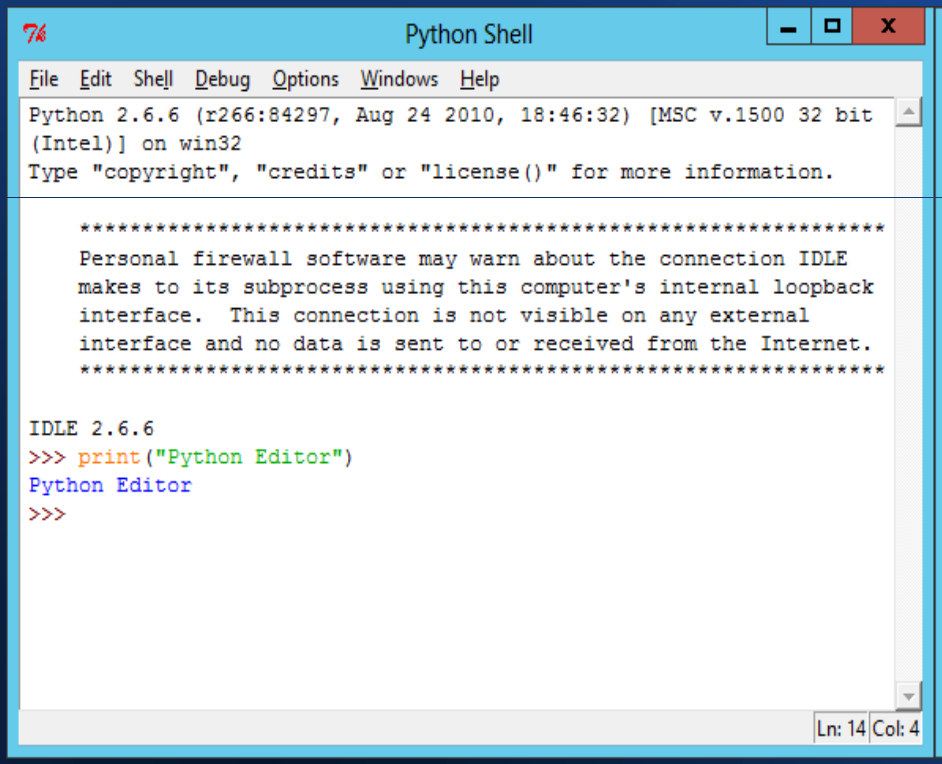
- Disadvantages:
  - Python code is slower than compiled languages like C/C++ and Java.
  - Python code for mobile apps requires additional run-time engine.

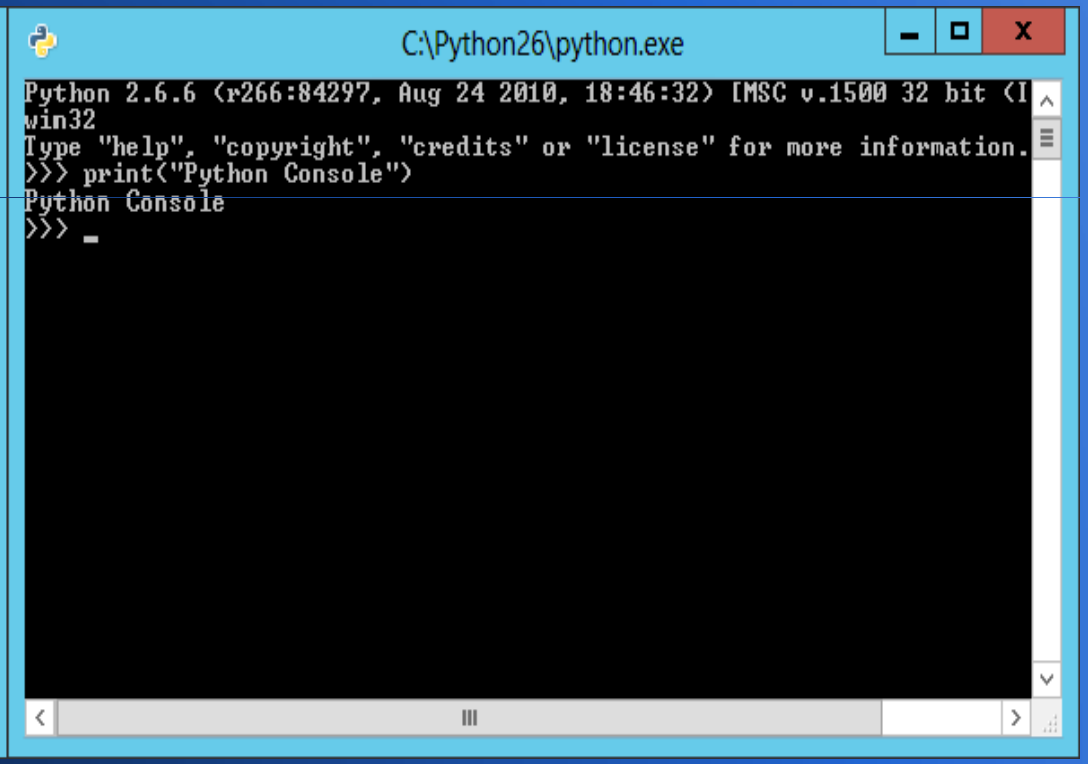
Python Interpreter and PATH :

- Python interpreter is provided by python.exe on Windows System and python binary on Linux System.
- In Windows, Python is configured in Environmental Properties.
- In Linux, Python PATH is configured inside /.bashrc or etc/profile.
- Python code is saved with an extension of .py and it is executed by Python Interpreter.

Python Console and Python IDLE :

- In Windows, Python installation gives Command Line and GUI Programming interface.
- IDLE and Python(Command Line) are two default programming interfaces.





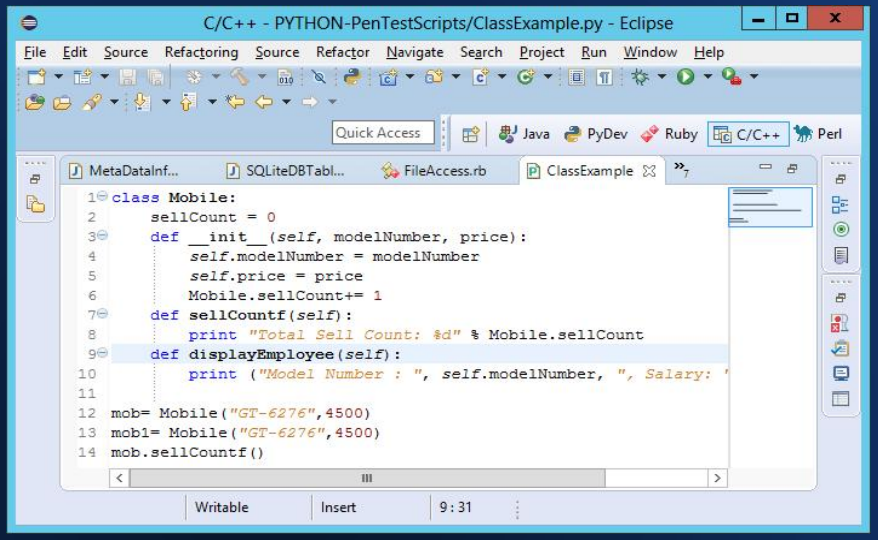
Python Coding on LINUX:

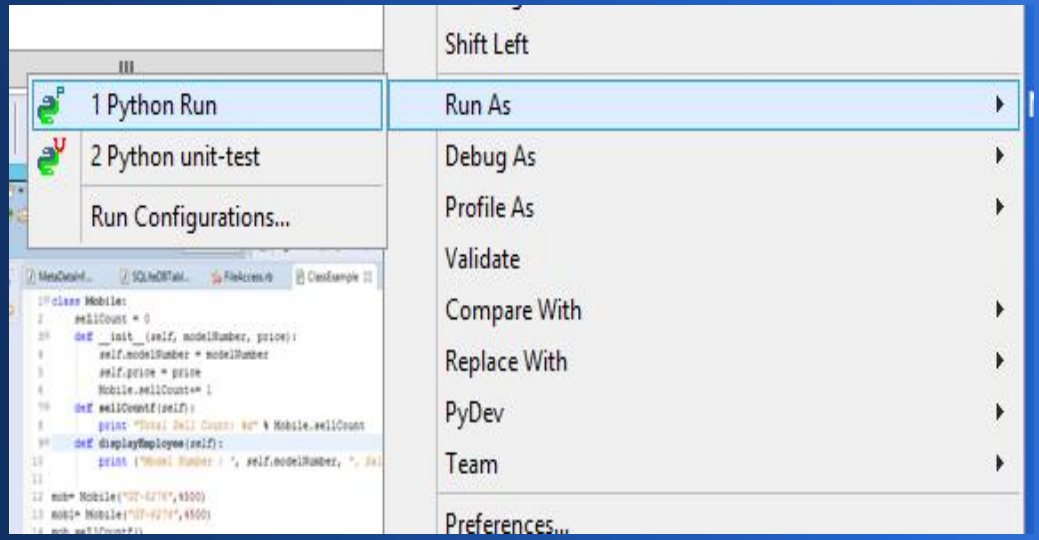
- Follow steps below to run Python Code on Linux:
  - Open Linux Terminal and type ‘sudo gedit Sample.py‘
  - Type 'python Sample.py' to run the code

- Follow steps to run Python using Eclipse:
  - Install JDK on Linux System
  - Install Eclipse and configure Pydev plugin from http://pydev.org
  - Create Pydev Project and create new Python file
  - Click over the file and click on ‘Run As Python’

Steps to run Python on Eclipse (Windows) :

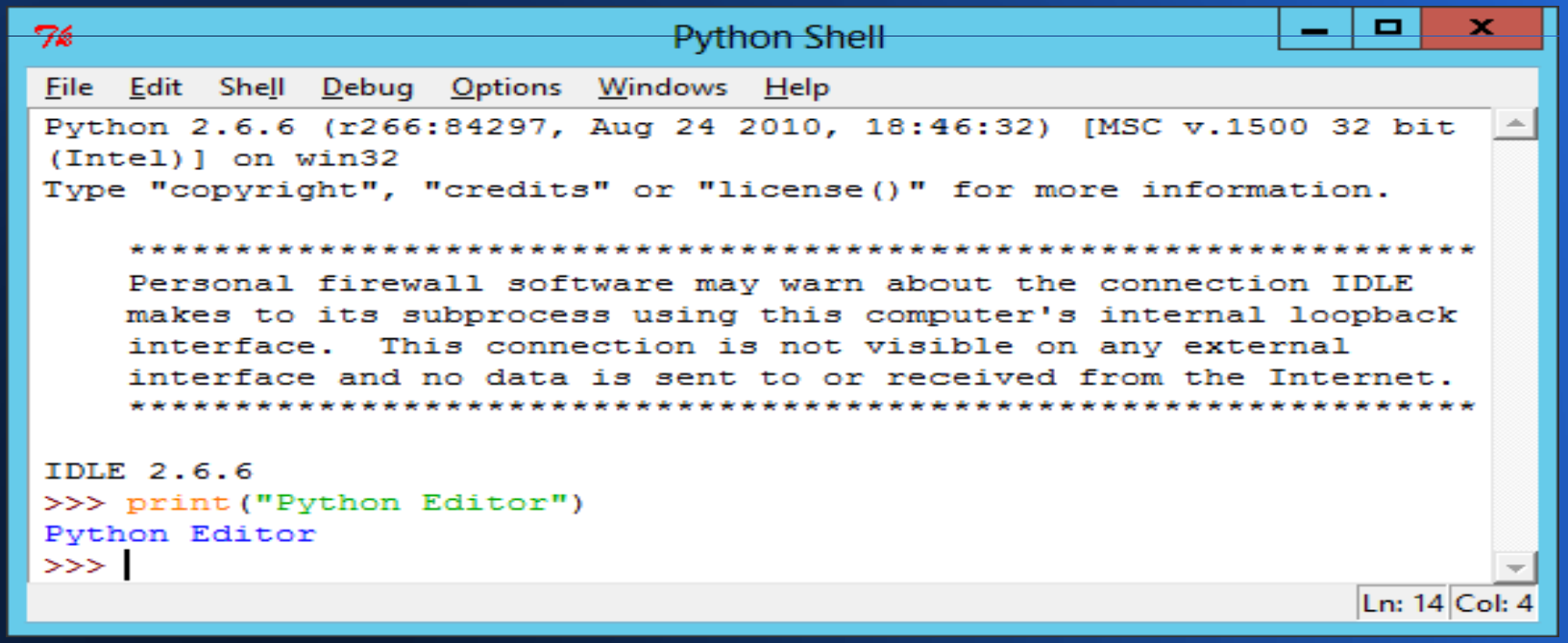
- Download and unzip Eclipse inside folder.
- Download python from http://python.org and install Python.
- Open Eclipse and configure Pydev for Python Interpreter.
- Create new file and save it with .py extension.
- Write and run the code with 'Run As-Python Run' option.





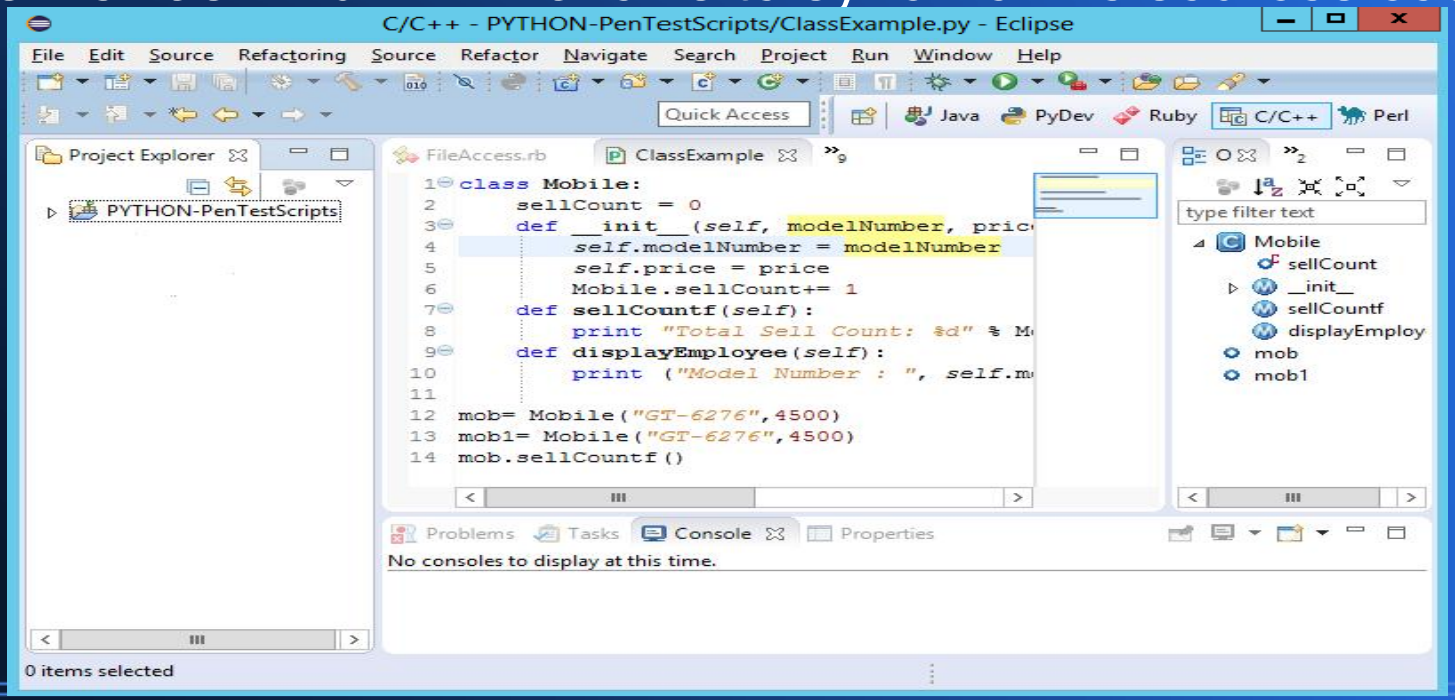
Writing to Screen:

- Python print() function writes to Console- Windows Dos or Linux Terminal.
- It takes values to print as arguments as comma separated values,
- Formatting is also supported while printing the value.
- '%X' is used as the place holder for different types of values.



Python Editors and IDEs:

- Third-Party IDE gives Code Editor, Compiler and Debugger environment .
- Netbeans, Eclipse, Geany are common Python IDE with all common features .
- Eclipse is open source IDE and Netbeans is from Oracle. Eclipse works on Windows as well as Linux .IDE allows to synchronize sources code repository .



Python Identifiers, Variables and Keywords:

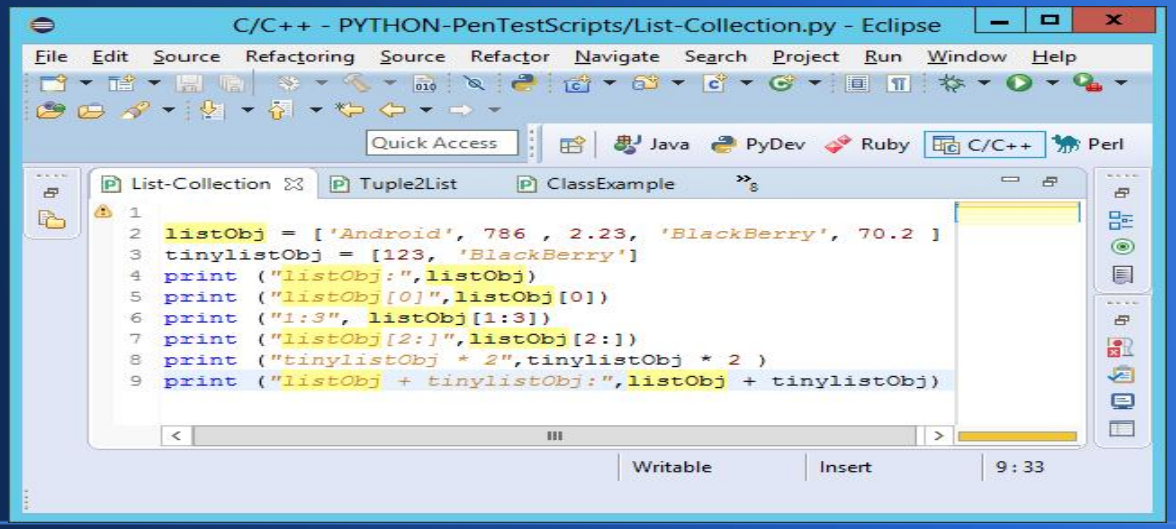
- Identifier is the smallest unit of Python Code .
- Variable is the named address space in the memory .
- Keyword are the reserved word and it has special meaning to Python Interpreter .
- Data types is assigned as the values assigned to the variable .
- Data types are assigned while interpreting or executing the codes.
- The values computed by the expression in right-hand side is assigned to the variable in the right-hand side .

Sample Codes:

- Code:

```
listObj = ['Android', 786 , 2.23, 'BlackBerry', 70.2 ]
tinylistObj = [123, 'BlackBerry']
print ("listObj:",listObj)
print ("listObj[0]",listObj[0])
print ("1:3", listObj[1:3])
print ("tinylistObj * 2",tinylistObj * 2 )
```

- Screenshot:



Data Science and BigData Processing with Python Training Session-II

Agendas/Modules:

- Revision of Session-I :Theory and Labs
- Python Identifier
- Python Key Words
- Python Indentation
- Python Operators
- Python Functions

Revision of Session-Theory and Labs:

- Python is interpreted, object oriented, interactive programming language.
- It is open source programming language and http://python.org is official site.
- Guido Van Rossum is the main brain behind Python Language.
- Eclipse, Netbeans, IDLE, Anaconda, Geany are common Python IDEs.
- Import statement is used to import external modules.

Python Identifiers:

- A Python Identifier is a name given to identify a variable, function, class, module or other
- Python des not support punctuation characters such as @, $ and % within identifier
- Identifier started with _ is meant to be private
- numberOfCoffeeShops=20, is an example of integer identifier

Python Key Words:

- The words which has special meaning to Python Interpreter.
- Key Word is the Reserved Word and it can't be used as an identifier .
- All in-built functions are reserved keywords.
- print, and, def, for, break and continue are examples of reserved keyword
- Example of def Key Word:

```
def sum(x,y):
        return (x+y)
num1=int(input(“Enter The First Number”))
num2=int(input(“Enter The Second Number”))
print(“Sum of “+ str(num1) +” and “ + str(num2) +”:”+sum(num1,num2))
```

Python Indentation:

- Python Codes does not use braces
- The blocks of codes are denoted by line intendation
- Example

```
def sum(number1, number2):
    result=0
    if number1 >=number2:
        result=number1+number2
    else:
        result=number1+number2*2
  return result
print(sum(23,34))
```

Python Operators:

- Operators are special symbols and they have special meaning to Python Interpreter.
- There are different types of Operators and they are
  - Arithmetic Operators(7) :+,-,*,/,%,**, //
  - Assignment Operators(7):+=,-=,*=,/=,%=,**=, //=
  - Comparison (i.e., Relational) Operators (6) : >,<,>=,<=,==,!= or <>
  - Logical Operators(3): and, or , not
  - Bitwise Operators (6): &,|,^,~, <<,>>
  - Membership or Identity Operators(2) :in , not in

Python Functions:

- Function is program construct that defines a unit task
- A function is defined using def keyword
- The value is returned with return keyword
- Python function can take arguments and values are passed by reference.
- Function Example:

```
def sum(number1, number2):
    result=0
   if number1 >=number2:
       result=number1+number2
   else:
       result=number1+number2*2
   return result
print(sum(23,34))
```

Data Science and BigData Processing with Python Training Session-III

Agendas/Modules:

- Revision of Session-III:Theory and Labs
- User Inputs Statement
- Control Statements
- Control Statement- if-elif-else
- Looping - for loop
- Looping - while loop

Revision of Day 2-Theory and Labs:

- No braces are used in Python Codes, indentations are to create blocks of code .
- Python Code is saved as .py file.
- PyDev is the eclipse plugin for Python coding.

User Inputs Statement- input() Function :

- input() or raw_inut() function is used to accept user's inputs
- The argument value to input() is the message to display
- The value returned by input() is always string
- In-built functions are used to covert it into different data-types

Control Statements:

- Control statements are used to control the flow of code execution
- Control statement are given written anywhere inside the code
- Every control statement has condition to check and values to compute
- if-elif-else, for and while are condition-checking statement

Control Statement- if-elif-else :

- if-elif-else statement checks the condition at the beginning
- The checks continue until and unless one of them becomes false
- else part is executed if conditions become false
- Example:

Control Statement - break and continue :

- break and continue are used in-process check of if-elif-else, for and while
- break terminates the inner-most loop and continues the outer one
- break statement is generally after a child-condition

Looping- for Loop :

- for is used iterate through a set of values for a given range-condition
- A variable is used in for loop to hold locally iterated values
- in() is used for range-condition and its second index is exclusive
- Example:

Looping - while loop:

- for is also used iterate through a set of values but its local variable is defined explicit
- A variable is used in for loop to hold locally interated value
- in() is used for range-condition and its second index is exclusive

Data Science and BigData Processing with Python Training Session-IV

Agendas/Modules:

- Revision of Session-III:Theory and Labs
- Type-Casting-int() and str()
- Common Data Structure
- Working with List and Tuples
- Debugging Python Code

Revision of Day 2 and 3-Theory and Labs:

- Input() is used for user inputs
- Return type of function is defined in function delaration
- In Eclipse Ctrl+F11 is executed to run program

Type Casting-int() and str() :

- input() or raw_inut() function is used to accept user's inputs
- Int() function is used to convert string value into int
- str() is used to convert int string form
- ** is exponential operator and // is division operator

Common Data Structure:

- Python has in-built data structure and it supports external one like ctype
- List, Tuple, Dictionary are common Data Structure of Python
- The elements of data structure are iterated using while or for loop
- In Python, colon(:) is used for function declaration and control statements. One tab is given in the immediate next line to indicate the next-block of code.

Working with List and Tuples:

- Python has in-built data structure and it support external one like ctype
- List, Tuple, Dictionary are common Data Structure of Python
- The elements of data structure are iterated using while or for loop
- The index of in() operator is always exclusive i.e last value is not used
- In Python, : is used for function definition, looping and one tab is given in the immediate next line.

Configuring PyDev with Eclipse:

- Install Eclipse on PC
- Install PyDev Plugin on Eclipse using http://pydev.org/updates
- Install Python 2.7
- Go to Preferences-> PyDev-> Interpreter->Python Interpreter and select Python.exe.
- Create new PyDev Project and write the codes!
- Wow! Eclipse is ready for Python Coding and Debugging.

Debugging Python Code:

- Debugging is the process of running code in controlled fashion.
- Debugging helps in resolving issues or error on codes.
- Debugging helps in knowing values of values during the different

Data Science and BigData Processing with Python Training Session-V

Agendas/Modules:

- BigData MapReduce -Introduction
- Python MapReduce Module
- MapReduce using MrJob
- Accessing Database Table-mysqldb
- Accessing NonSQL using Python

THANK YOU ALL