

My Classroom > PySpark Certification Training Course

X PySpark Certification T...

Introduction to Python for Apache Spark > Presentation



Course Content

- Presentation
- Class 2 Recording
- Case Study I - Perform...
- Case Study I Solution
- Case Study II - Verifi...
- Case Study II Solution
- In-Class Demo - Input...
- Introduction to Python
- Functions, OOPs, Mod...

Personal Library

Presentation

3 / 141



COURSE OUTLINE //



MODULE 02

Introduction to Big Data Hadoop and Spark

Introduction to Python for Apache Spark

Functions, OOPs, Modules in Python

Deep Dive into Apache Spark Framework

Playing with Spark RDDs

Data Frames and Spark SQL

Machine Learning using Spark MLlib

Deep Dive into Spark MLlib

Understanding Apache Kafka and Apache Flume

Apache Spark Streaming - Processing Multiple Batches

Apache Spark Streaming - Data Sources

Implementing an End-to-End Project



1

31°C

Haze



Search



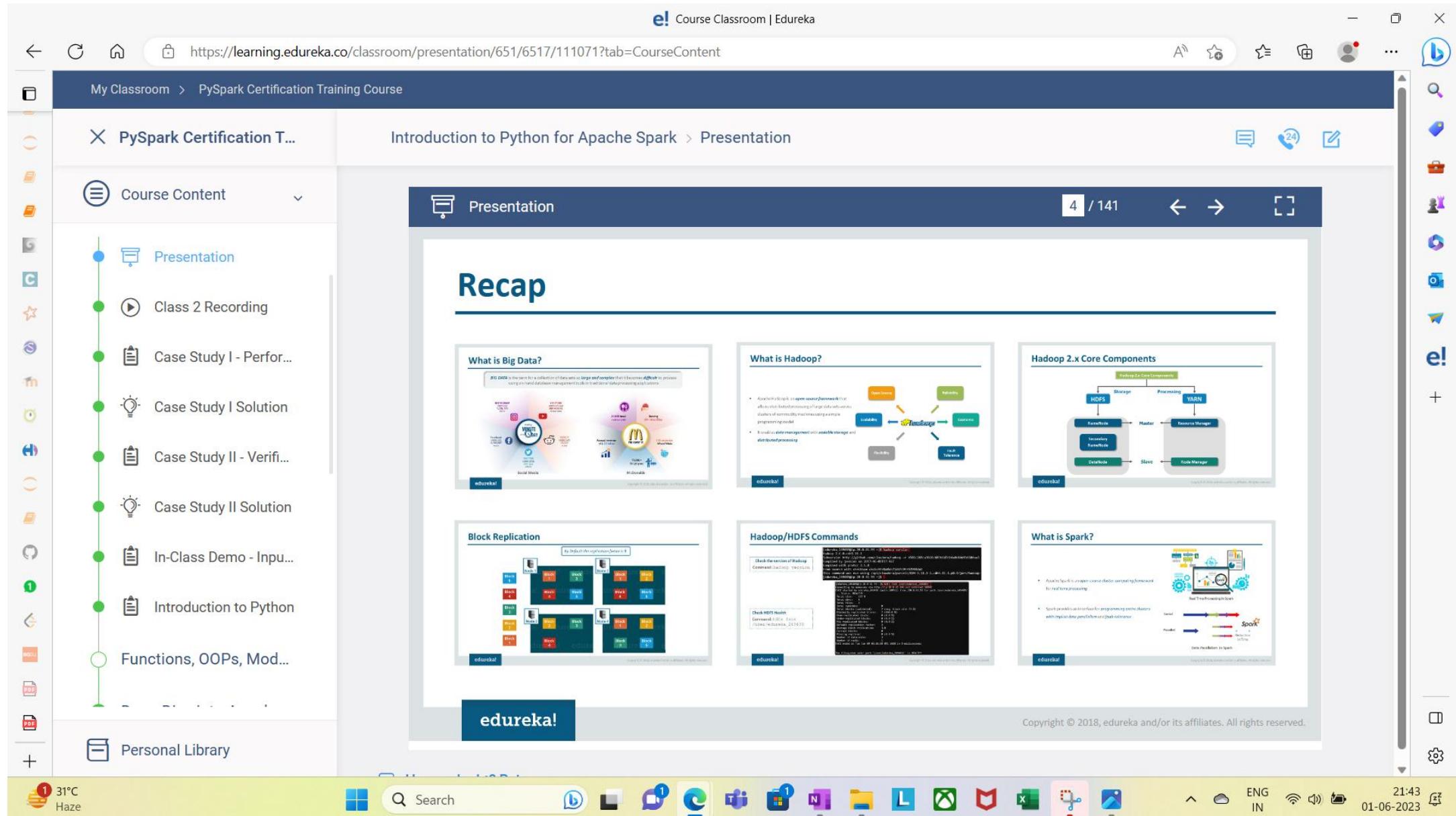
ENG

IN



21:43

01-06-2023



https://learning.edureka.co/classroom/presentation/651/6517/111071?tab=CourseContent

My Classroom > PySpark Certification Training Course

PySpark Certification T... Course Content

- Presentation
- Class 2 Recording
- Case Study I - Perform...
- Case Study I Solution
- Case Study II - Verifi...
- Case Study II Solution
- In-Class Demo - Input...
- Introduction to Python
- Functions, OOPs, Mod...

Introduction to Python for Apache Spark > Presentation

Presentation 5 / 141

Objectives

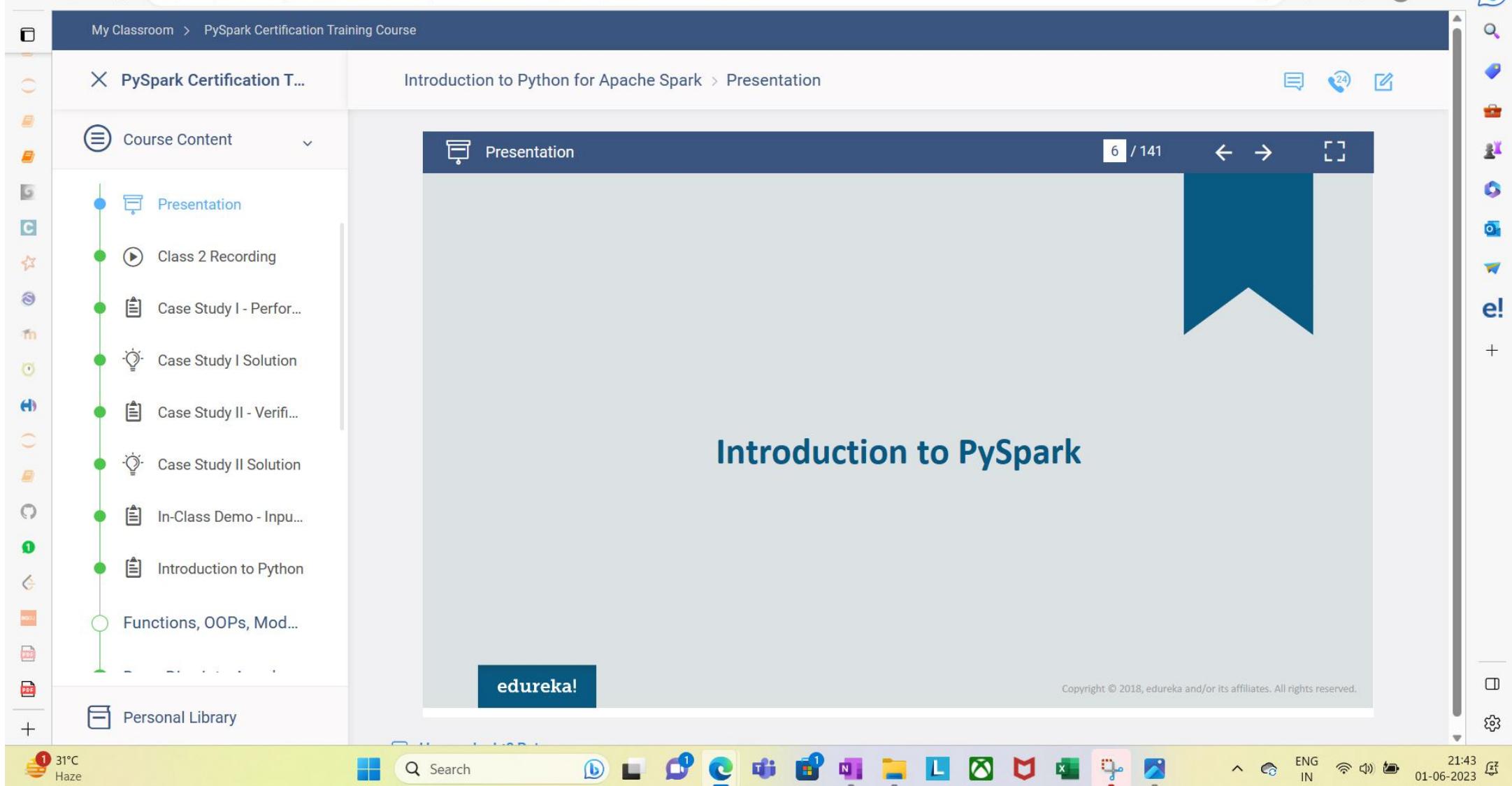
After completing this module, you should be able to:

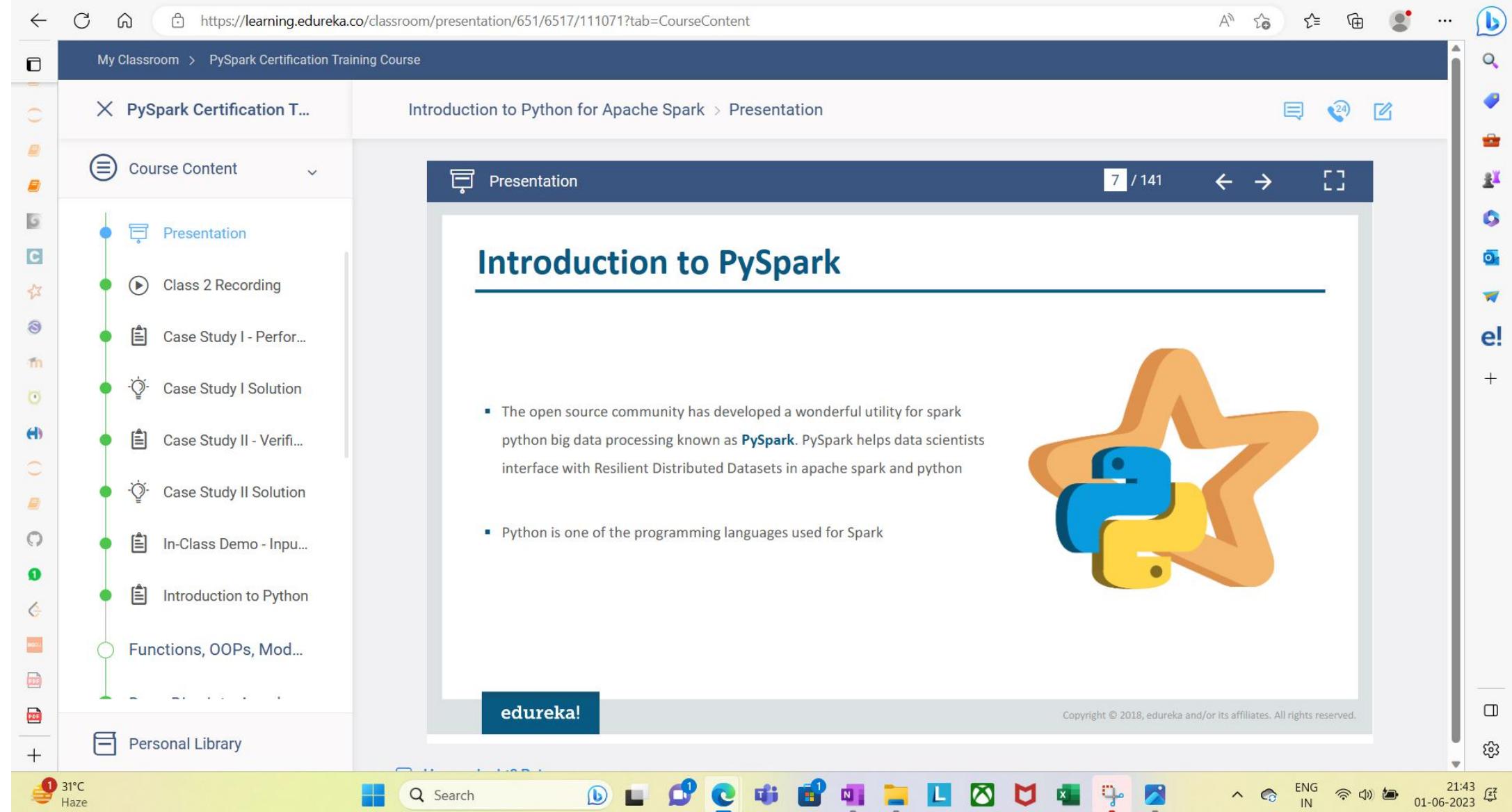
- Define Python
- Explain Numbers
- Work on Strings, Tuples, Lists, Dictionaries, and Sets
- Implement Operands and Expressions
- Write your First Python Program
- Understand Flow Control
- Take input from the user and perform operations on it



edureka!

Copyright © 2018, edureka and/or its affiliates. All rights reserved.







https://learning.edureka.co/classroom/presentation/651/6517/111071?tab=CourseContent



My Classroom > PySpark Certification Training Course

X PySpark Certification T...

Introduction to Python for Apache Spark > Presentation



Course Content

- Presentation
- Class 2 Recording
- Case Study I - Perform...
- Case Study I Solution
- Case Study II - Verifi...
- Case Study II Solution
- In-Class Demo - Input...
- Introduction to Python
- Functions, OOPs, Mod...

Personal Library

Presentation

8 / 141



Before begin with PySpark, let us learn about Python

edureka!

Copyright © 2018, edureka and/or its affiliates. All rights reserved.



31°C

Haze



Search



ENG

IN



01-06-2023

21:44

https://learning.edureka.co/classroom/presentation/651/6517/111071?tab=CourseContent

My Classroom > PySpark Certification Training Course

PySpark Certification T... Course Content

Presentation

Introduction to Python for Apache Spark > Presentation

9 / 141

Presentation

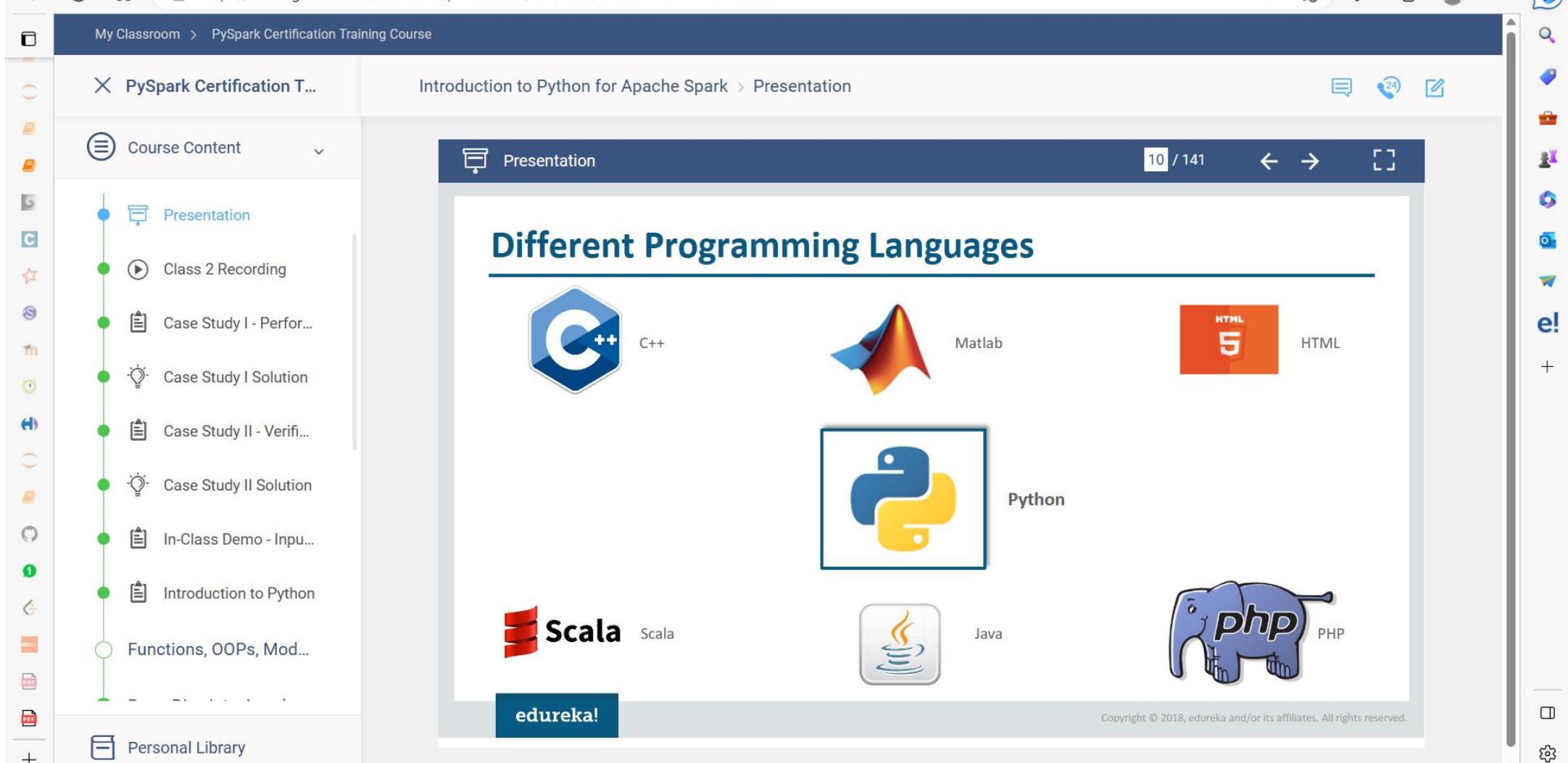
I have heard there are a lot of Programming languages?

Well, I will discuss few of the most famous ones

edureka!

Copyright © 2018, edureka and/or its affiliates. All rights reserved.

Personal Library



https://learning.edureka.co/classroom/presentation/651/6517/111071?tab=CourseContent

My Classroom > PySpark Certification Training Course

PySpark Certification T... Course Content

Presentation Class 2 Recording Case Study I - Perform... Case Study I Solution Case Study II - Verifi... Case Study II Solution In-Class Demo - Input... Introduction to Python Functions, OOPs, Mod... Personal Library

Introduction to Python for Apache Spark > Presentation

Presentation 11 / 141

Dave tell me which Programming language should I start with

It has to be Python, let me give you enough reasons to believe that

edureka!

Copyright © 2018, edureka and/or its affiliates. All rights reserved.

https://learning.edureka.co/classroom/presentation/651/6517/111071?tab=CourseContent

My Classroom > PySpark Certification Training Course

PySpark Certification T... Course Content

- Presentation
- Class 2 Recording
- Case Study I - Perform...
- Case Study I Solution
- Case Study II - Verifi...
- Case Study II Solution
- In-Class Demo - Input...
- Introduction to Python
- Functions, OOPs, Mod...

Introduction to Python for Apache Spark > Presentation

Presentation 12 / 141

Why Python?

Simple and Easy to Learn

Python is easy to understand, read and write too

Free and Open Source

Python is an example of a FLOSS (Free/Libre and Open Source Software) which means one can easily distribute copies of this software, read its source code, modify it etc.

High Level Language

One does not need to bother about the low-level details like memory allocation, etc. while writing Python Scripts

edureka!

Copyright © 2018, edureka and/or its affiliates. All rights reserved.

a=3
b=5
sum=a+b
print(sum)

compile

0100101011011
0111011000011
0110110101100
1010101011010

Personal Library

https://learning.edureka.co/classroom/presentation/651/6517/111071?tab=CourseContent

My Classroom > PySpark Certification Training Course

PySpark Certification T... Course Content

- Presentation
- Class 2 Recording
- Case Study I - Perform...
- Case Study I Solution
- Case Study II - Verifi...
- Case Study II Solution
- In-Class Demo - Inpu...
- Introduction to Python
- Functions, OOPs, Mod...

Introduction to Python for Apache Spark > Presentation

Presentation 13 / 141

Why Python?

Portable

Supported by many platforms like Linux, Windows, FreeBSD, Macintosh, Solaris, BeOS, OS/390, PlayStation, Windows CE etc.

Procedure Oriented **Object Oriented**

Supports different Programming Paradigm

Python supports object oriented programming as well as procedural programming

Extensible

Python code can invoke C and C++ libraries, can be called from C++ program, can integrate with Java and .NET components

edureka!

Copyright © 2018, edureka and/or its affiliates. All rights reserved.

31°C Haze

Search

Windows Start

Icons for various Microsoft applications: Edge, File Explorer, Task View, OneDrive, Mail, Calendar, OneNote, Powerpoint, Word, Excel, and others.

ENG IN

21:44 01-06-2023

https://learning.edureka.co/classroom/presentation/651/6517/111071?tab=CourseContent

My Classroom > PySpark Certification Training Course

PySpark Certification T... Course Content

- Presentation
- Class 2 Recording
- Case Study I - Perform...
- Case Study I Solution
- Case Study II - Verifi...
- Case Study II Solution
- In-Class Demo - Input...
- Introduction to Python
- Functions, OOPs, Mod...

Introduction to Python for Apache Spark > Presentation

Presentation 14 / 141

Why Python?

The diagram consists of five circular icons, each connected by a line to a text box describing a benefit of Python:

- Improved Programmer's Productivity (Icon: Shopping cart with code)
- Extensive Support Libraries (Icon: Document with code)
- Integration Feature (Icon: Wrench and screwdriver)
- Program Portability (Icon: Circular arrows)
- Object Oriented Language (Icon: Curly braces)
- Interpreted and Dynamic Structure (Icon: Delivery truck)

edureka!

Copyright © 2018, edureka and/or its affiliates. All rights reserved.

https://learning.edureka.co/classroom/presentation/651/6517/111071?tab=CourseContent

My Classroom > PySpark Certification Training Course

PySpark Certification T... Course Content

Presentation, Class 2 Recording, Case Study I - Perform..., Case Study I Solution, Case Study II - Verifi..., Case Study II Solution, In-Class Demo - Input..., Introduction to Python, Functions, OOPs, Mod...

Introduction to Python for Apache Spark > Presentation

Presentation 15 / 141

edureka!

Copyright © 2018, edureka and/or its affiliates. All rights reserved.

But, can you tell me, where is Python used?

Python is also widely used by a lot of companies, I will tell you about few giants

My Classroom > PySpark Certification Training Course

X PySpark Certification T...

Introduction to Python for Apache Spark > Presentation



Course Content

- Presentation
- Class 2 Recording
- Case Study I - Perform...
- Case Study I Solution
- Case Study II - Verifi...
- Case Study II Solution
- In-Class Demo - Inpu...
- Introduction to Python
- Functions, OOPs, Mod...

Personal Library

Who is using Python?

The popular YouTube video sharing system is written in Python



NASA uses Python for cryptography and intelligence analysis



Google makes extensive use of Python in its web search system



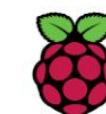
NETFLIX uses Python in its software infrastructure



Dropbox service codes both server and clients software primarily in Python



The Raspberry Pi single board computer promotes Python as its educational language

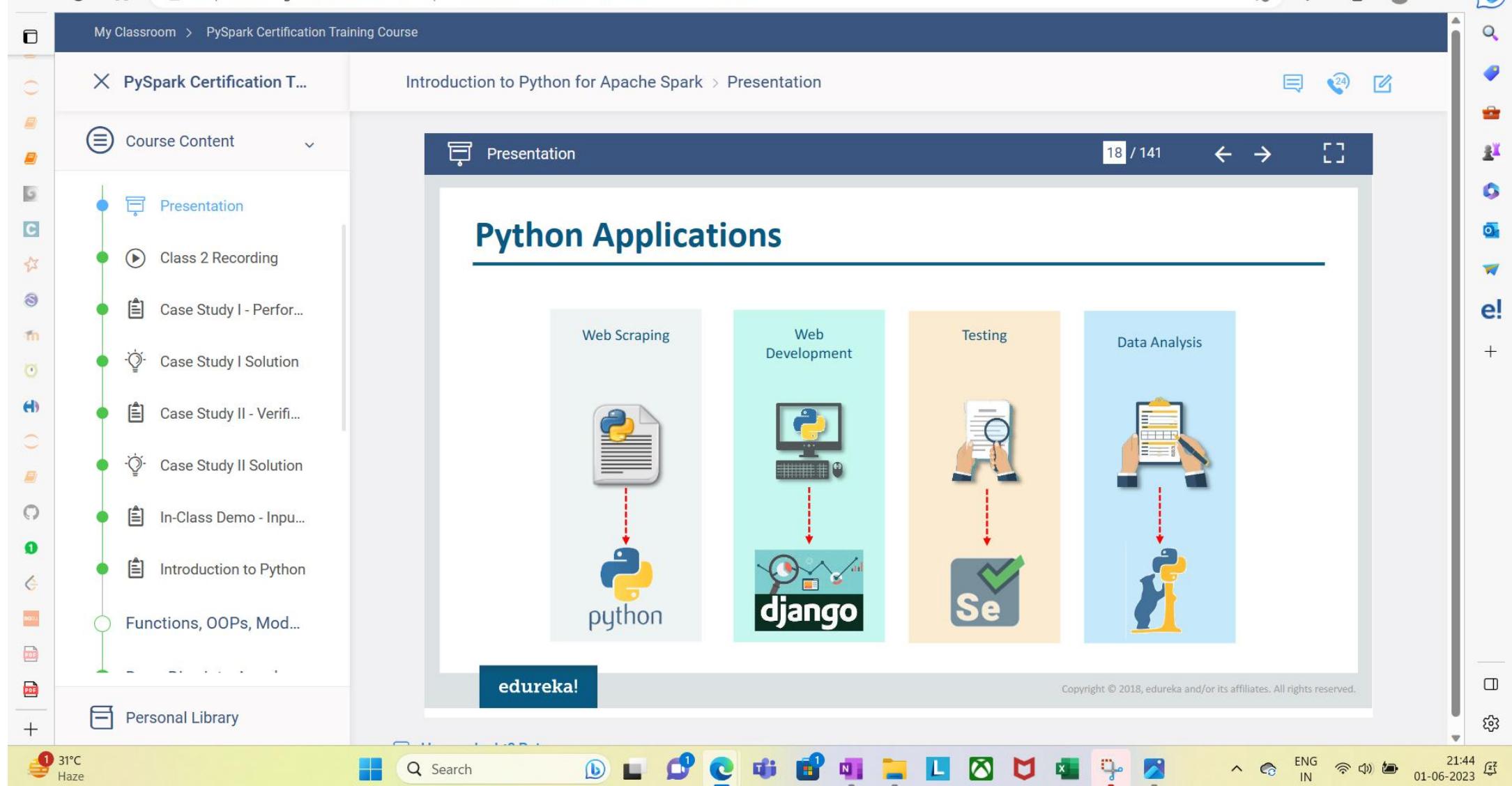


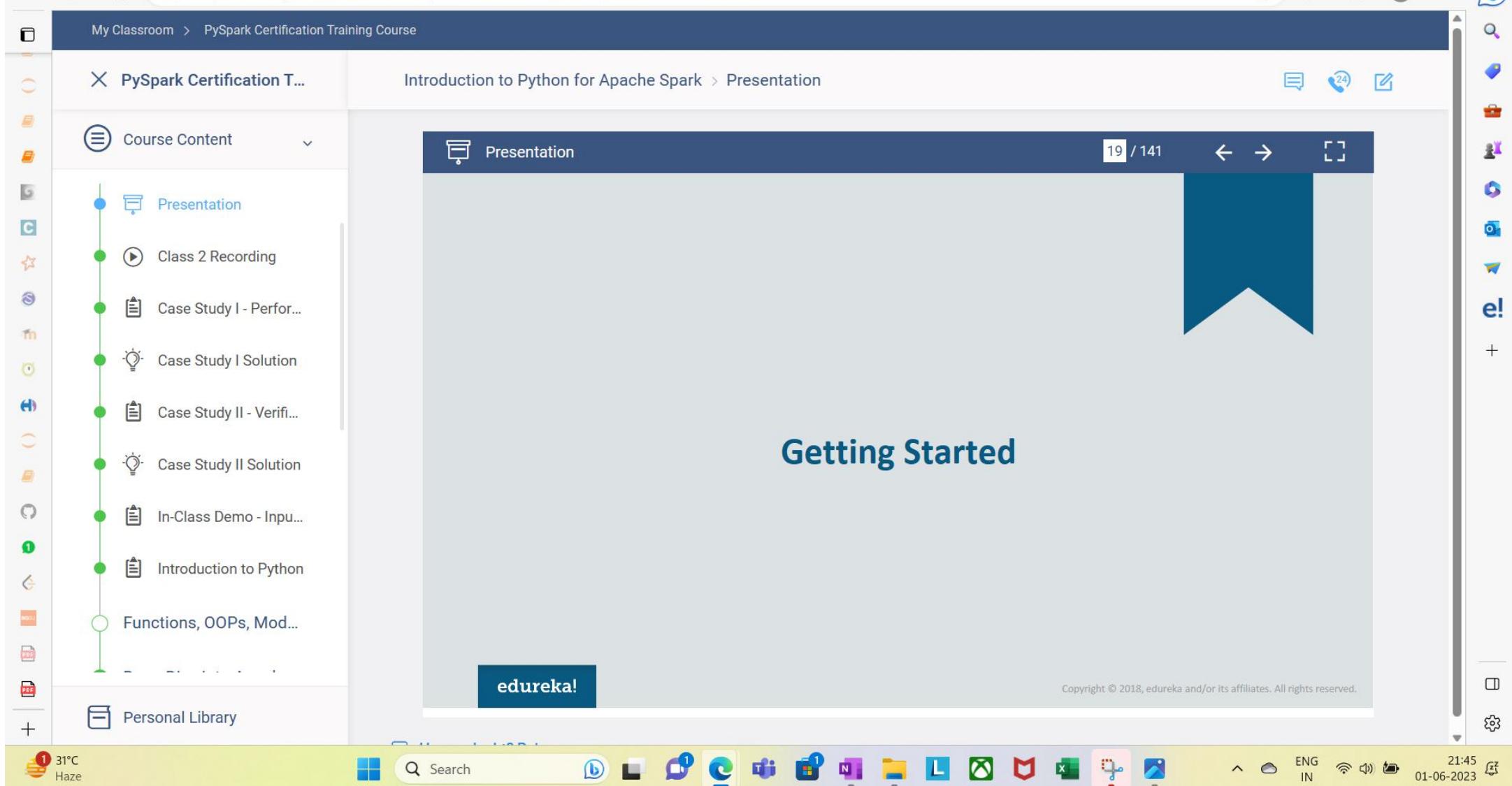
Copyright © 2018, edureka and/or its affiliates. All rights reserved.

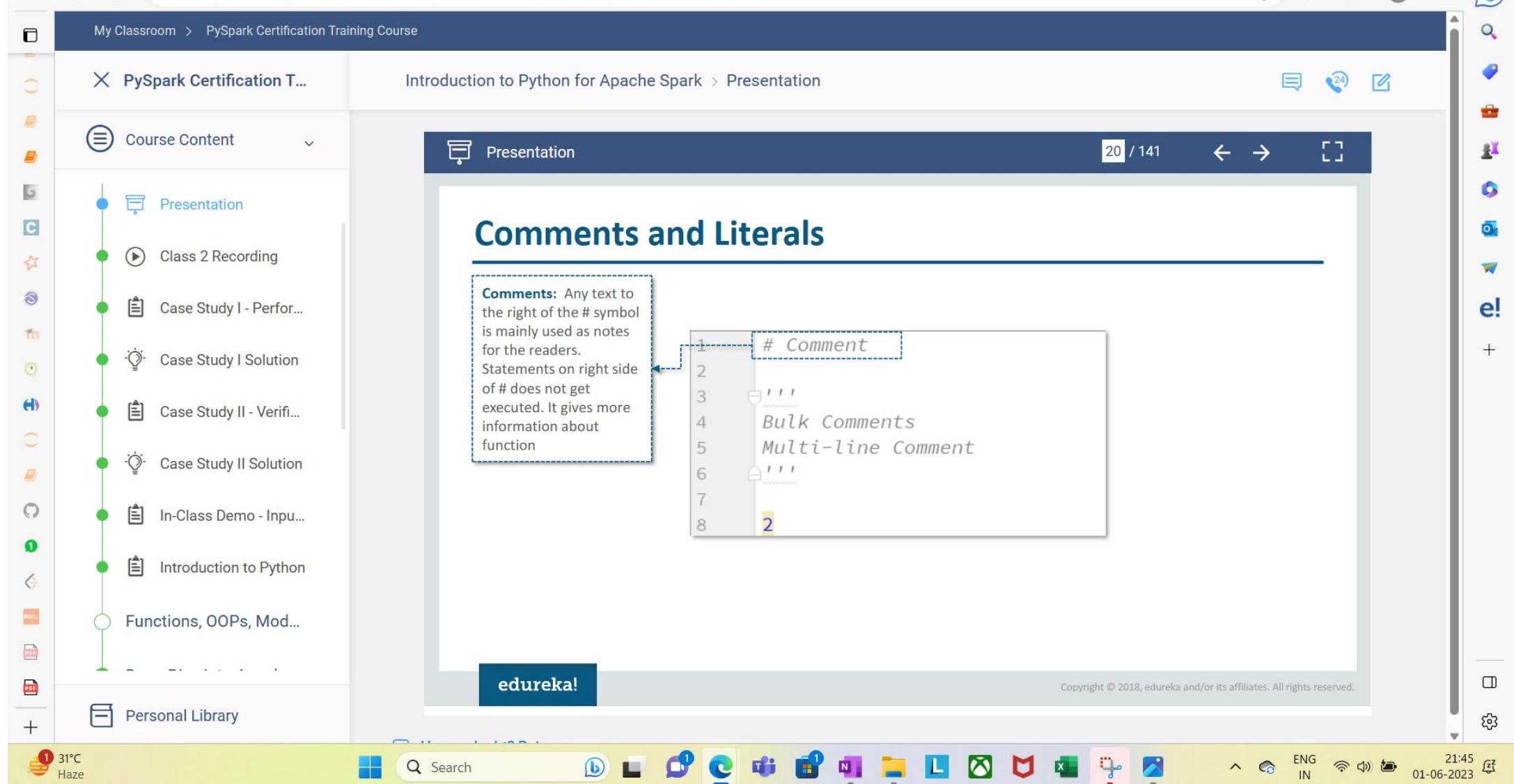
X PySpark Certification T...

Introduction to Python for Apache Spark > Presentation









https://learning.edureka.co/classroom/presentation/651/6517/111071?tab=CourseContent

My Classroom > PySpark Certification Training Course

PySpark Certification T... Course Content

Presentation Class 2 Recording Case Study I - Perform... Case Study I Solution Case Study II - Verifi... Case Study II Solution In-Class Demo - Input... Introduction to Python Functions, OOPs, Mod...

Personal Library

Introduction to Python for Apache Spark > Presentation

Presentation 21 / 141

Comments and Literals

Comments: Any text to the right of the # symbol is mainly used as notes for the readers. Statements on right side of # does not get executed. It gives more information about function

```
1 # Comment
2
3
4
5
6
7
8 2
```

Bulk Comments: Enclose the code in triple quoted strings ("'''")

edureka!

Copyright © 2018, edureka and/or its affiliates. All rights reserved.

https://learning.edureka.co/classroom/presentation/651/6517/111071?tab=CourseContent

My Classroom > PySpark Certification Training Course

PySpark Certification T... Course Content Presentation

Introduction to Python for Apache Spark > Presentation

22 / 141

Comments and Literals

Comments: Any text to the right of the # symbol is mainly used as notes for the readers. Statements on right side of # does not get executed. It gives more information about function

```
1 # Comment
2
3
4
5
6
7
8
```

Bulk Comments: Enclose the code in triple quoted strings ("")

Multi-line Comment

Literal Constants: Any number or character, or set of characters

edureka!

Copyright © 2018, edureka and/or its affiliates. All rights reserved.

Personal Library

https://learning.edureka.co/classroom/presentation/651/6517/111071?tab=CourseContent

My Classroom > PySpark Certification Training Course

PySpark Certification T... Course Content

Presentation, Class 2 Recording, Case Study I - Perform..., Case Study I Solution, Case Study II - Verifi..., Case Study II Solution, In-Class Demo - Input..., Introduction to Python, Functions, OOPs, Mod...

Introduction to Python for Apache Spark > Presentation

Presentation 23 / 141

Indentation

 No braces to indicate blocks of code for class and function definitions or flow control

 Blocks of code are denoted by line indentation, which is rigidly enforced

 The number of spaces in the indentation is variable, but all statements within the block must be indented the same amount

 Leading whitespace at the beginning of a logical line is used to compute the indentation level of the line, which in turn, is used to determine the grouping of statements

edureka!

Copyright © 2018, edureka and/or its affiliates. All rights reserved.

X PySpark Certification T...

Introduction to Python for Apache Spark > Presentation



Course Content

- Presentation
- Class 2 Recording
- Case Study I - Perform...
- Case Study I Solution
- Case Study II - Verifi...
- Case Study II Solution
- In-Class Demo - Input...
- Introduction to Python
- Functions, OOPs, Mod...

Personal Library

Presentation

24 / 141

Python Code Execution

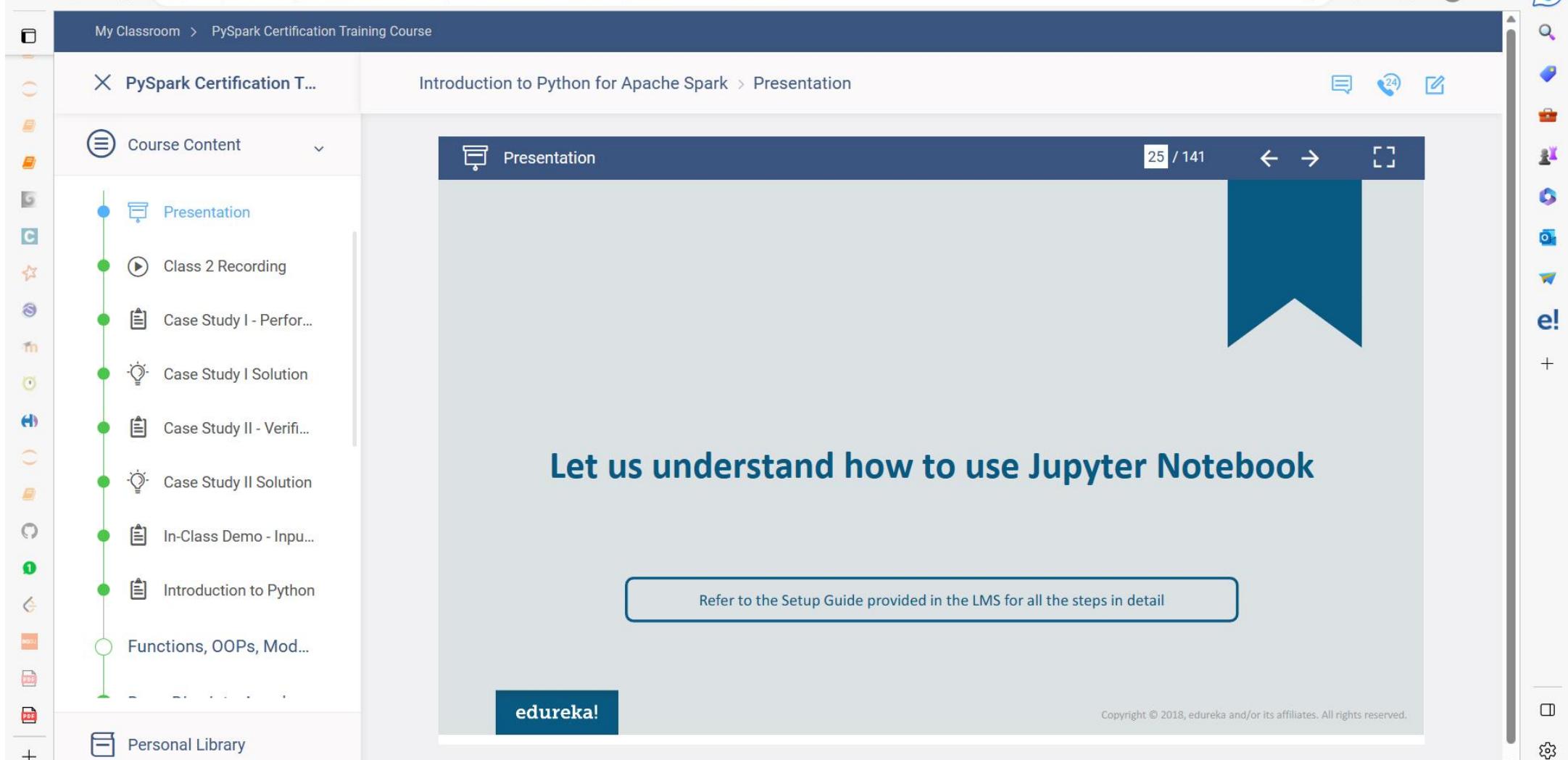
The structured code is written and saved with .py extension

The code is converted into Byte code for machine to understand

The Byte codes are executed

edureka!

Copyright © 2018, edureka and/or its affiliates. All rights reserved.



X PySpark Certification T...

Introduction to Python for Apache Spark > Presentation



Course Content

-  Presentation
-  Class 2 Recording
-  Case Study I - Perform...
-  Case Study I Solution
-  Case Study II - Verifi...
-  Case Study II Solution
-  In-Class Demo - Input...
-  Introduction to Python
-  Functions, OOPs, Mod...

Personal Library

Presentation

26 / 141



How to upload Files on Cloud Lab?

Refer to the Setup Guide provided in the LMS for all the steps in detail

edureka!

Copyright © 2018, edureka and/or its affiliates. All rights reserved.



1

31°C

Haze



Search



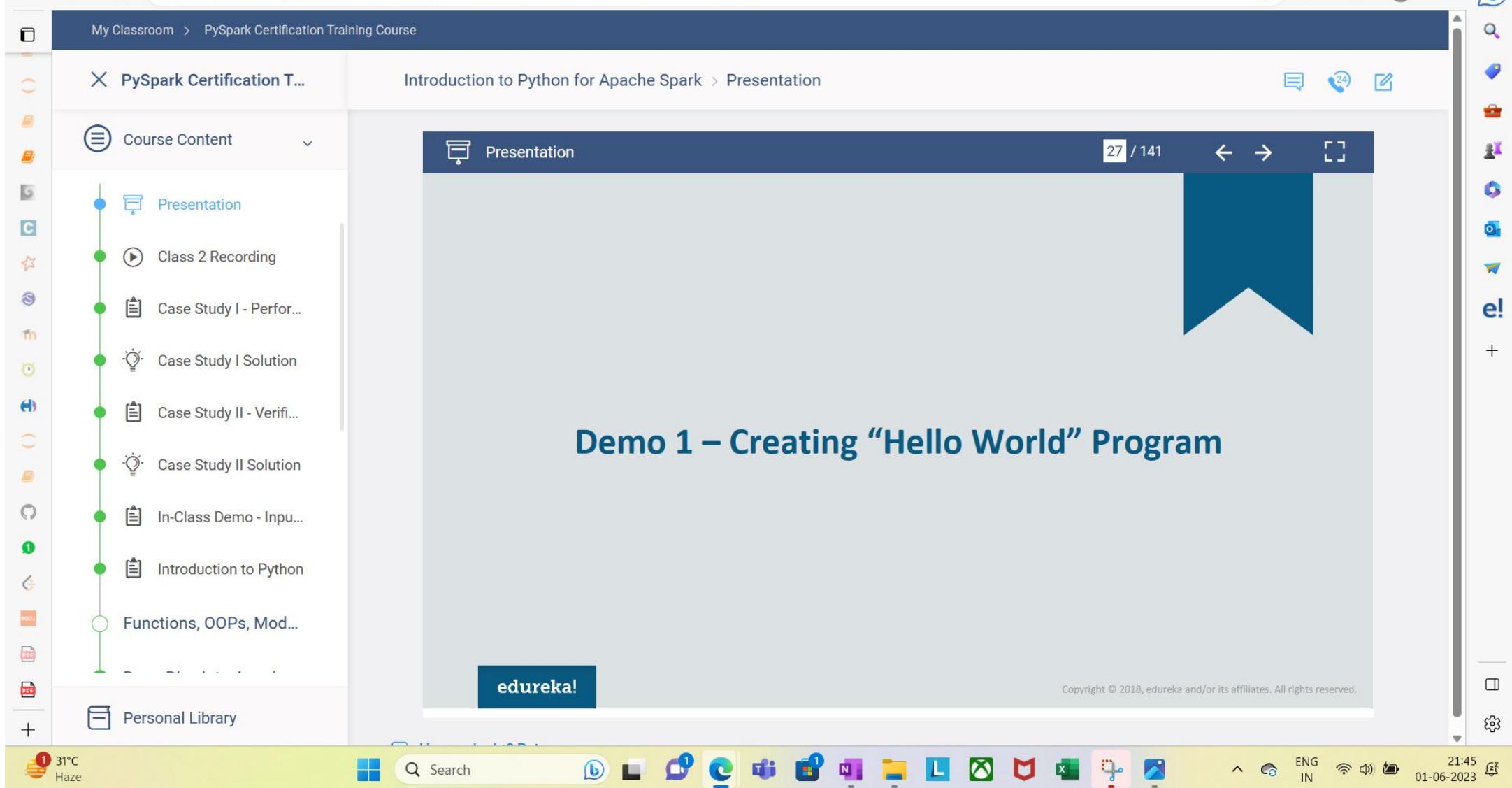
ENG

IN



21:45

01-06-2023



My Classroom > PySpark Certification Training Course

PySpark Certification T... Introduction to Python for Apache Spark > Presentation

Presentation 28 / 141 ← →

Creating “Hello World” Program

Every character in Python should be enclosed within single or double quotes

```
print('Hello World')
print("Welcome to Edureka")
```

Hello World
Welcome to Edureka

Output after running new.py

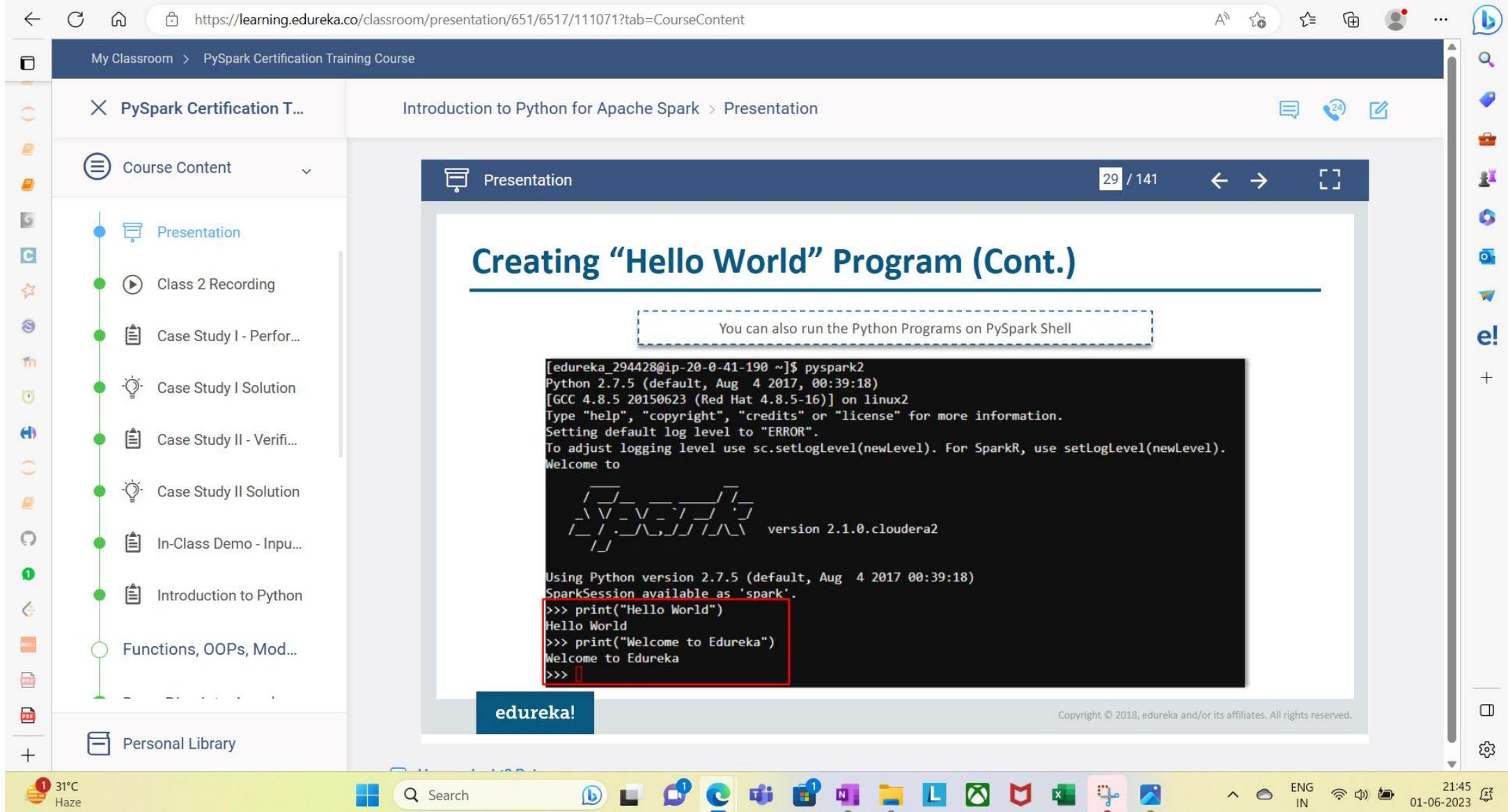
edureka!

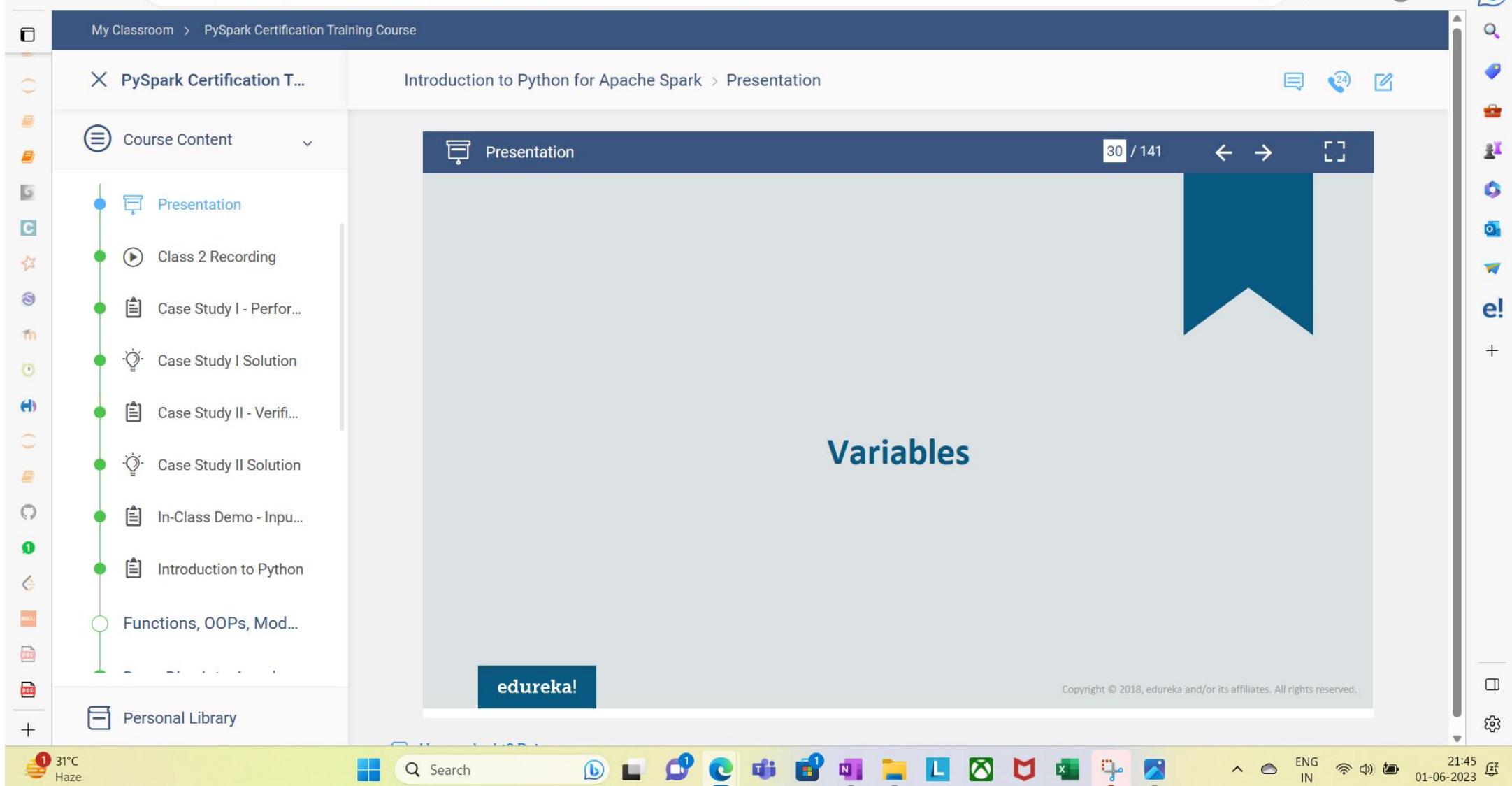
Copyright © 2018, edureka and/or its affiliates. All rights reserved.

Course Content

- Presentation
- Class 2 Recording
- Case Study I - Perform...
- Case Study I Solution
- Case Study II - Verifi...
- Case Study II Solution
- In-Class Demo - Inpu...
- Introduction to Python
- Functions, OOPs, Mod...

Personal Library





My Classroom > PySpark Certification Training Course

PySpark Certification T... Introduction to Python for Apache Spark > Presentation

Presentation 31 / 141

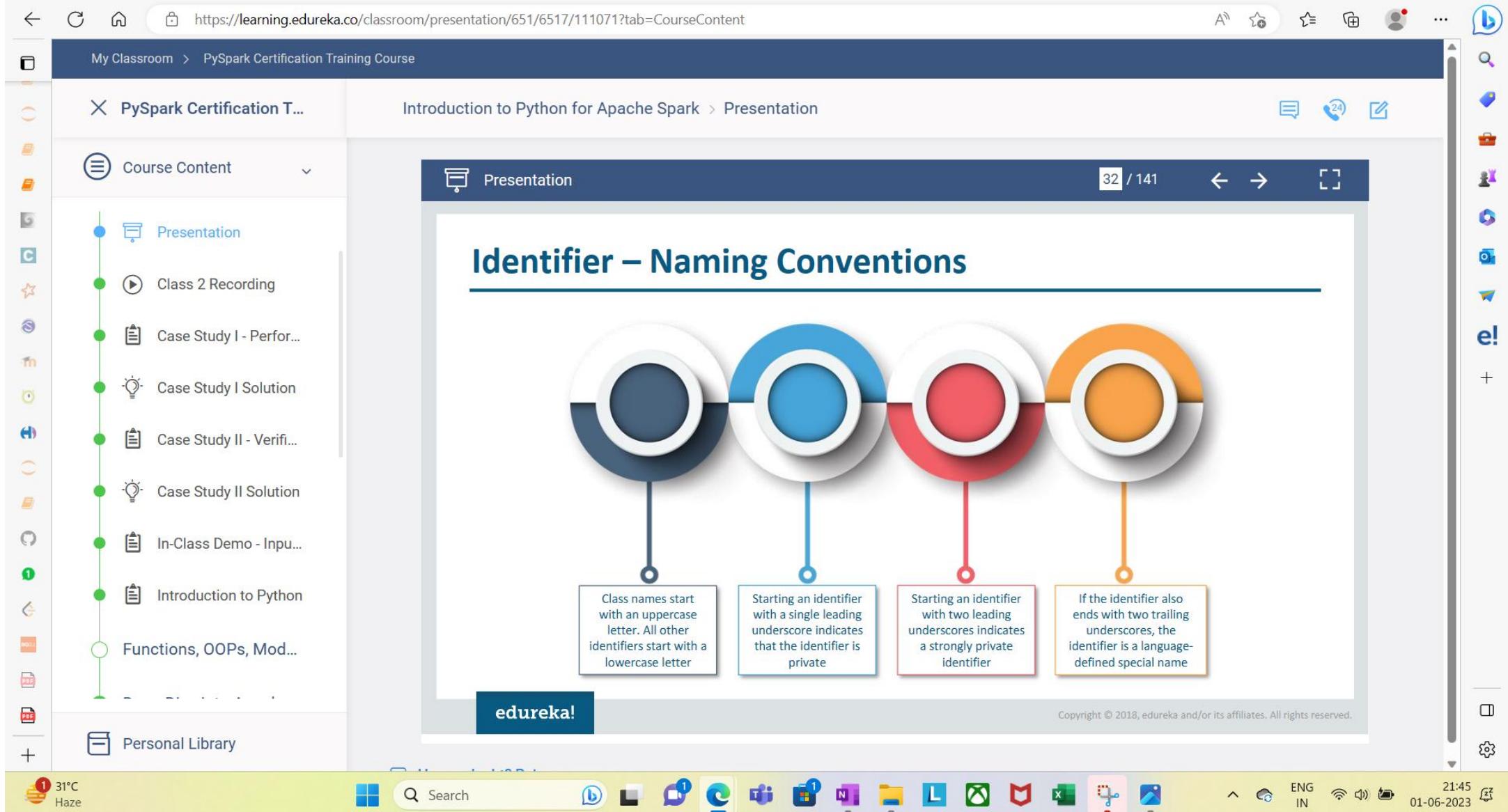
Identifier

- A Python Identifier is a name used to identify a variable, function, class, module or other object
- An identifier starts with a letter A to Z or a to z or an underscore (_) followed by zero or more letters, underscores and digits (0 to 9)
- Python is a case sensitive programming language
- Python does not allow special characters such as @, \$ and % within identifiers

edureka!

Copyright © 2018, edureka and/or its affiliates. All rights reserved.

Personal Library



https://learning.edureka.co/classroom/presentation/651/6517/111071?tab=CourseContent

My Classroom > PySpark Certification Training Course

PySpark Certification T... Course Content

- Presentation
- Class 2 Recording
- Case Study I - Perform...
- Case Study I Solution
- Case Study II - Verifi...
- Case Study II Solution
- In-Class Demo - Input...
- Introduction to Python
- Functions, OOPs, Mod...

Introduction to Python for Apache Spark > Presentation

Presentation 33 / 141

Variables

Variables are nothing but reserved memory locations to store values. This means that when you create a variable you reserve some space in memory

The diagram illustrates variable assignment in memory. On the left, code snippets are shown: `A = 16`, `B = 20`, and `C = 'edureka'`. Arrows point from each variable to a 3D cube representing memory. The cube has three stacked sections labeled `A=16`, `B=20`, and `C= 'edureka'`, with the word "Memory" repeated to its right.

edureka!

Copyright © 2018, edureka and/or its affiliates. All rights reserved.

31°C Haze Search

1 ENG IN 01-06-2023 21:45

My Classroom > PySpark Certification Training Course

PySpark Certification T... Introduction to Python for Apache Spark > Presentation

Presentation 34 / 141 ← →

Variables (Contd.)

Consider the below screenshot, it explains how to assign a value to a *Variable*

Assigning values 10 and edureka! to variables A and B respectively

Right click and click on 'Run File_name' to execute the code

```
A=10  
B='edureka!'  
print(A,B)
```

Output 10 edureka!

edureka!

Copyright © 2018, edureka and/or its affiliates. All rights reserved.



https://learning.edureka.co/classroom/presentation/651/6517/111071?tab=CourseContent



My Classroom > PySpark Certification Training Course

X PySpark Certification T...

Introduction to Python for Apache Spark > Presentation



Course Content

- Presentation
- Class 2 Recording
- Case Study I - Perform...
- Case Study I Solution
- Case Study II - Verifi...
- Case Study II Solution
- In-Class Demo - Input...
- Introduction to Python
- Functions, OOPs, Mod...

Personal Library

Presentation

35 / 141



Standard Data Types

edureka!

Copyright © 2018, edureka and/or its affiliates. All rights reserved.



1

31°C

Haze



Search



^

~

ENG

IN

Wi-Fi

Speaker

Screen

01-06-2023

21:45

X PySpark Certification T...

Introduction to Python for Apache Spark > Presentation

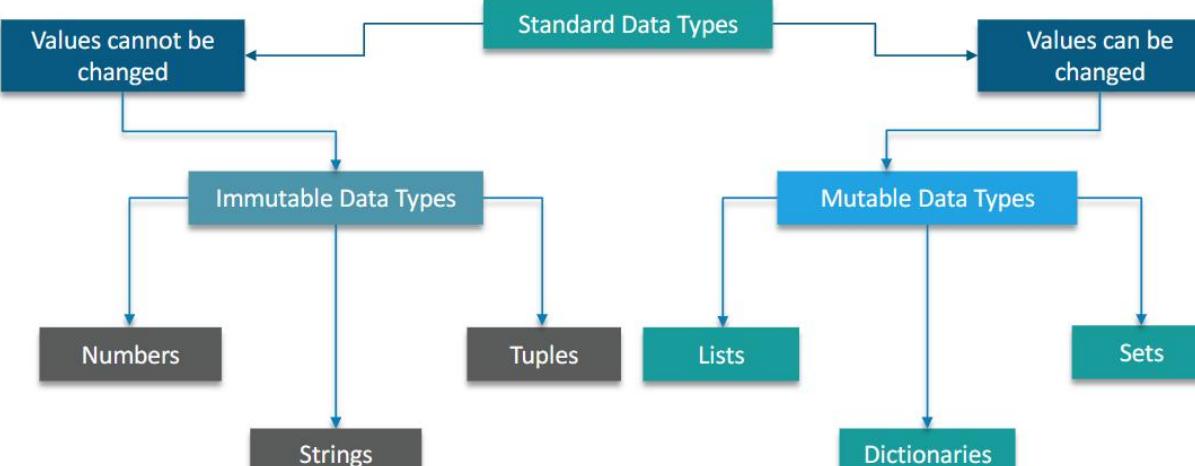


Course Content

- Presentation
- Class 2 Recording
- Case Study I - Perform...
- Case Study I Solution
- Case Study II - Verifi...
- Case Study II Solution
- In-Class Demo - Input...
- Introduction to Python
- Functions, OOPs, Mod...

Personal Library

Standard Data Types



edureka!

Copyright © 2018, edureka and/or its affiliates. All rights reserved.



My Classroom > PySpark Certification Training Course

PySpark Certification T... Introduction to Python for Apache Spark > Presentation

Presentation 37 / 141 ← →

Course Content

- Presentation
- Class 2 Recording
- Case Study I - Perform...
- Case Study I Solution
- Case Study II - Verifi...
- Case Study II Solution
- In-Class Demo - Input...
- Introduction to Python
- Functions, OOPs, Mod...

edureka!

Immutable Data Types

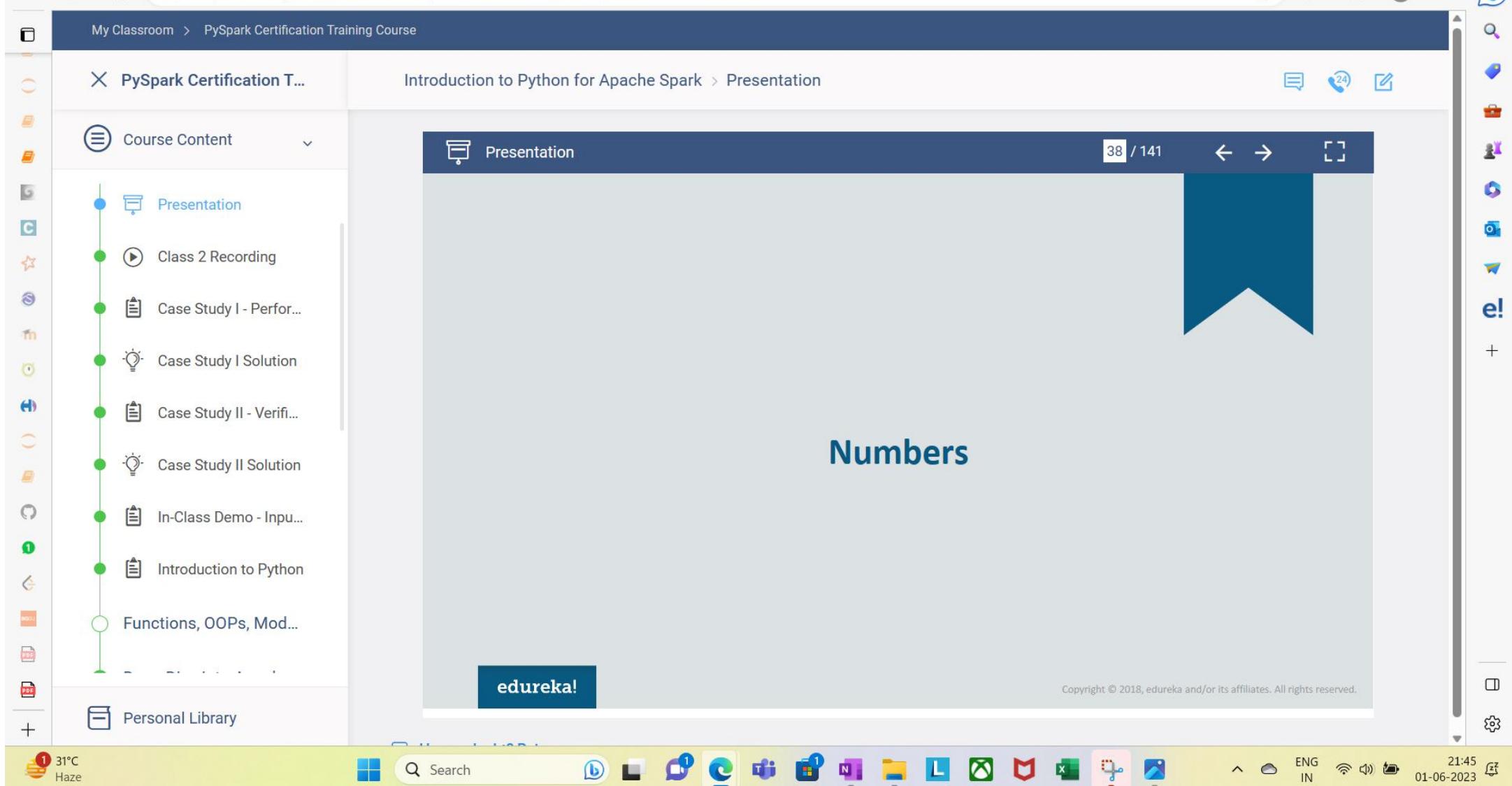
Numbers

Tuples

Strings

Copyright © 2018, edureka and/or its affiliates. All rights reserved.

```
graph TD; A[Immutable Data Types] --> B[Numbers]; A --> C[Tuples]; A --> D[Strings]
```



My Classroom > PySpark Certification Training Course

PySpark Certification T... Introduction to Python for Apache Spark > Presentation

Presentation 39 / 141 ← →

Numbers

Python supports three different numerical values

int (Signed Integers)

Complex Numbers

float (Real Numbers)

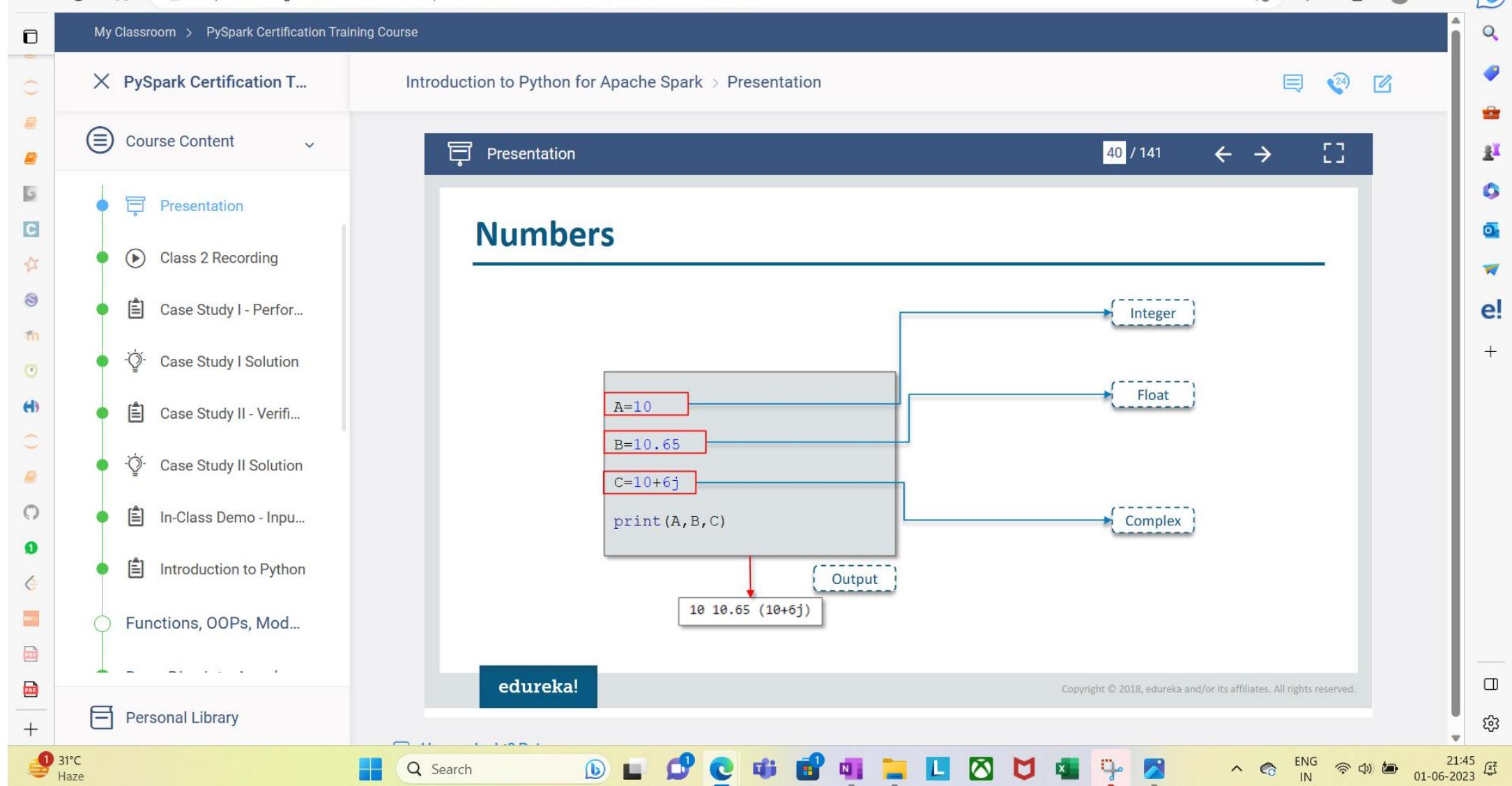
In Python you can represent **Numbers** in multiple ways:

- *Binary*
- *Octal*
- *Hexadecimal*

edureka!

Copyright © 2018, edureka and/or its affiliates. All rights reserved.

Personal Library



Course Classroom | Edureka

https://learning.edureka.co/classroom/presentation/651/6517/111071?tab=CourseContent

My Classroom > PySpark Certification Training Course

PySpark Certification T... Introduction to Python for Apache Spark > Presentation

Course Content

- Presentation
- Class 2 Recording
- Case Study I - Performance
- Case Study I Solution
- Case Study II - Verification
- Case Study II Solution
- In-Class Demo - Input
- Introduction to Python
- Functions, OOPs, Mod...

Presentation

41 / 141

Strings

edureka!

Copyright © 2018, edureka and/or its affiliates. All rights reserved.

Personal Library

31°C Haze

Search

1 24

L ENG IN 01-06-2023 21:45

https://learning.edureka.co/classroom/presentation/651/6517/111071?tab=CourseContent

My Classroom > PySpark Certification Training Course

PySpark Certification T... Course Content

- Presentation
- Class 2 Recording
- Case Study I - Perform...
- Case Study I Solution
- Case Study II - Verifi...
- Case Study II Solution
- In-Class Demo - Input...
- Introduction to Python
- Functions, OOPs, Mod...

Introduction to Python for Apache Spark > Presentation

Presentation 42 / 141

Strings

We can create them simply by enclosing characters in quotes

Example: string="Python"

OPERATIONS

Slicing	String[range]
Updating	String[range] + 'x'
Concatenation	String 1 + String 2
Repetition	String 1 * x
Membership	In, not in
Reverse	String [:-1]

edureka!

Copyright © 2018, edureka and/or its affiliates. All rights reserved.

31°C Haze Search

ENG IN 01-06-2023 21:45

My Classroom > PySpark Certification Training Course

PySpark Certification T... Introduction to Python for Apache Spark > Presentation

Presentation 43 / 141 ← →

Strings

Strings are amongst the most popular types in Python

```
str1='Welcome'  
str2="to"  
str3="""Edureka"""  
  
print(str1)  
print(str2)  
print(str3)
```

Single quote, double quotes or triple quotes can be used to show strings

Welcome
to
Edureka

edureka!

Copyright © 2018, edureka and/or its affiliates. All rights reserved.

Personal Library

X PySpark Certification T...

Introduction to Python for Apache Spark > Presentation



Course Content

- Presentation
- Class 2 Recording
- Case Study I - Perform...
- Case Study I Solution
- Case Study II - Verifi...
- Case Study II Solution
- In-Class Demo - Input...
- Introduction to Python
- Functions, OOPs, Mod...

Personal Library

Operations on Strings

```
string="Python"  
print(string)
```

Returns length of String

```
print(len(string))
```

Indexing and Slicing Operations

```
print(string[1:3])
```

Membership Checking

```
print('t' in string)
```

```
Python  
6  
yt  
True
```

Copyright © 2018, edureka and/or its affiliates. All rights reserved.

edureka!

https://learning.edureka.co/classroom/presentation/651/6517/111071?tab=CourseContent

My Classroom > PySpark Certification Training Course

PySpark Certification T... Course Content

Presentation Class 2 Recording Case Study I - Perform... Case Study I Solution Case Study II - Verifi... Case Study II Solution In-Class Demo - Input... Introduction to Python Functions, OOPs, Mod... Personal Library

Introduction to Python for Apache Spark > Presentation

Presentation 45 / 141

Operations on Strings (Contd.)

- String Formatting Operators
 - This operator is unique to Strings and makes up for the pack of having functions from C's printf() family

```
print("Welcome to %s"%(Python))
print("My name is %s and my age is
%d"%(Annie, 22))
```

Welcome to Python
My name is Annie and my age is 22

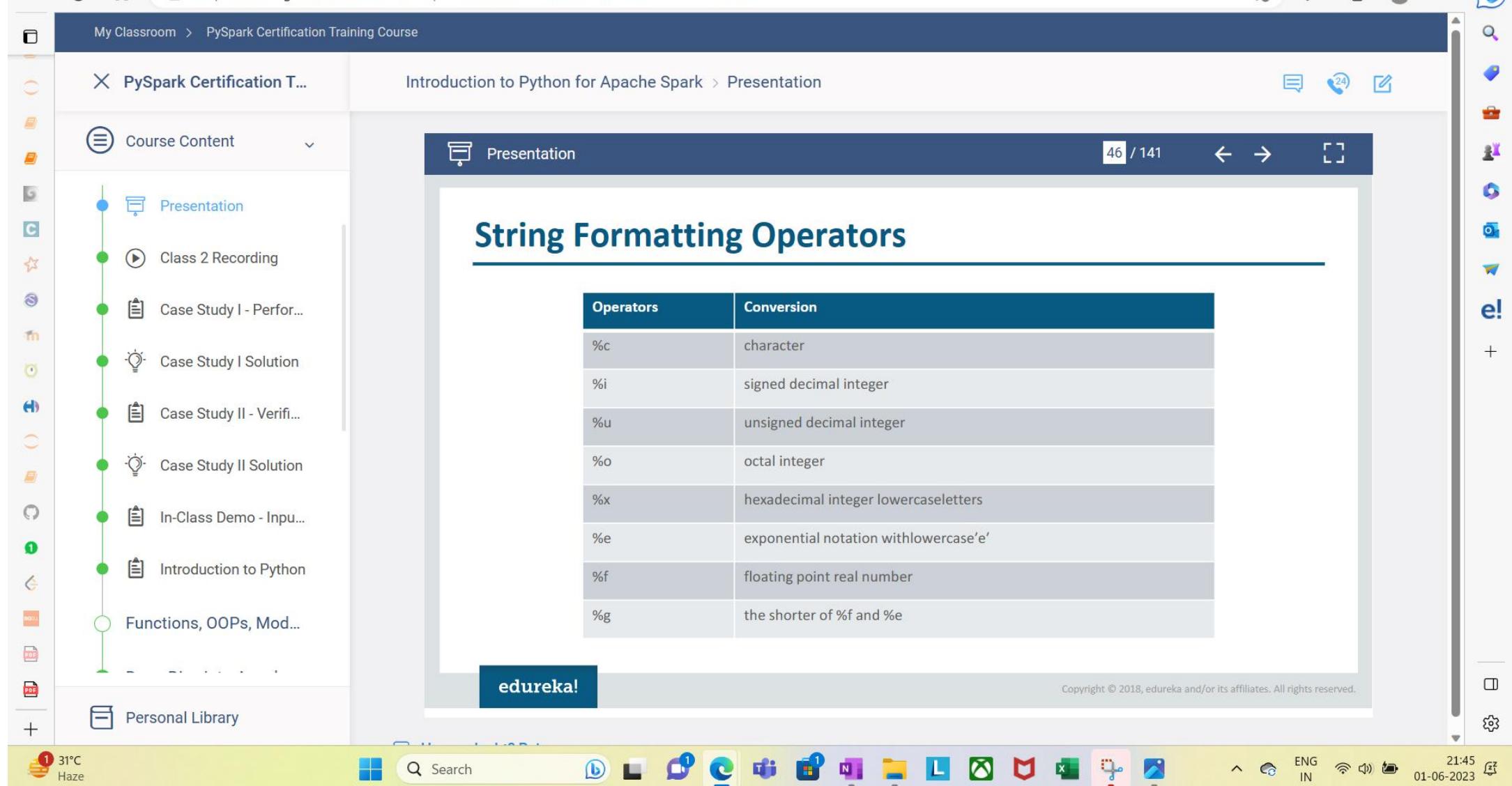
%c	%X
%s	%e
%i	%E
%d	%f
%u	%g
%o	%G
%x	

edureka!

Copyright © 2018, edureka and/or its affiliates. All rights reserved.

31°C Haze Search

ENG IN 01-06-2023 21:45



https://learning.edureka.co/classroom/presentation/651/6517/111071?tab=CourseContent

My Classroom > PySpark Certification Training Course

PySpark Certification T... Course Content

Presentation, Class 2 Recording, Case Study I - Perform..., Case Study I Solution, Case Study II - Verifi..., Case Study II Solution, In-Class Demo - Input..., Introduction to Python, Functions, OOPs, Mod...

Personal Library

Introduction to Python for Apache Spark > Presentation

Presentation 47 / 141

Built-in String Method

```
str='edureka'
print(str.capitalize())
print(str.count("ka",0,len(str)))
s=str.encode('utf-8','strict')
print(s)
```

Capitalizes the first letter of string
Counts how many times a string occurs within another string
Returns encoded string version of string

Edureka
1
b'edureka'

edureka!

Copyright © 2018, edureka and/or its affiliates. All rights reserved.

X PySpark Certification T...

Introduction to Python for Apache Spark > Presentation

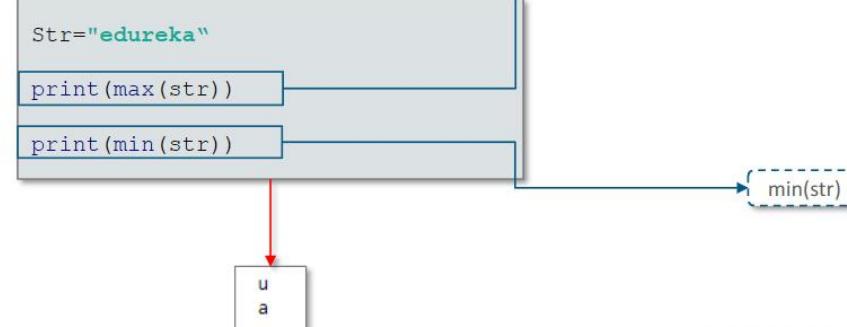


Course Content

- Presentation
- Class 2 Recording
- Case Study I - Perform...
- Case Study I Solution
- Case Study II - Verifi...
- Case Study II Solution
- In-Class Demo - Input...
- Introduction to Python
- Functions, OOPs, Mod...

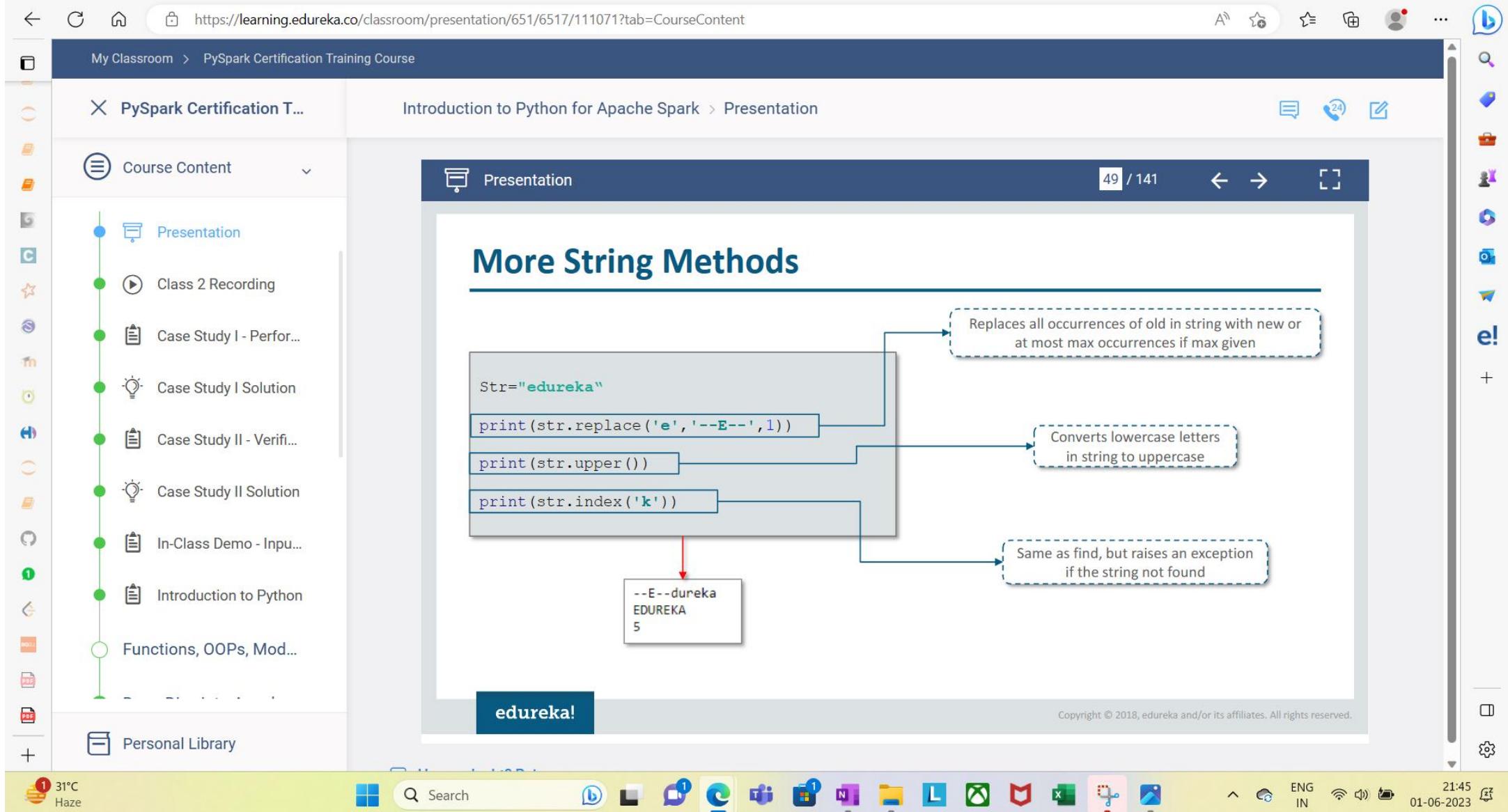
Personal Library

Built-in String Method (Contd.)



Note: As per ASCII values of Alphabets, min and max values from string are given

Copyright © 2018, edureka and/or its affiliates. All rights reserved.



My Classroom > PySpark Certification Training Course

PySpark Certification T... Course Content

Presentation Class 2 Recording Case Study I - Perform... Case Study I Solution Case Study II - Verifi... Case Study II Solution In-Class Demo - Input... Introduction to Python Functions, OOPs, Mod... Personal Library

Introduction to Python for Apache Spark > Presentation

Presentation 50 / 141

More String Methods

```
str1="Happy Learning"
print(str1[::-1])           ➔ Reverses the String
print(str1[2:7])            ➔ Slicing of String
print(str1.find('L'))       ➔ Returns the index at which the given letter is present in String
str2="Welcome to Edureka"
print(str1+str2)            ➔ Concatenation of String
print(str1*2)               ➔ Prints String 2 times
```

edureka!

Copyright © 2018, edureka and/or its affiliates. All rights reserved.

My Classroom > PySpark Certification Training Course

PySpark Certification T... Introduction to Python for Apache Spark > Presentation

Presentation 51 / 141 ← →

Course Content

- Presentation
- Class 2 Recording
- Case Study I - Perform...
- Case Study I Solution
- Case Study II - Verifi...
- Case Study II Solution
- In-Class Demo - Input...
- Introduction to Python
- Functions, OOPs, Mod...

Tuples

edureka!

Copyright © 2018, edureka and/or its affiliates. All rights reserved.

Personal Library



1

31°C

Haze



Search



ENG

IN

21:46

01-06-2023

My Classroom > PySpark Certification Training Course

PySpark Certification T... Introduction to Python for Apache Spark > Presentation

Presentation 52 / 141 ← →

Tuples

A Tuple is a sequence of immutable Python objects. Tuples are sequences, just like lists

Example: tuple=("Marketing","Sales")

The diagram illustrates various tuple operations:

- Slicing
- Delete Tuple
- Concatenation
- Updating Tuple
- Tuple Length
- Repetition

OPERATIONS

edureka!

Copyright © 2018, edureka and/or its affiliates. All rights reserved.

My Classroom > PySpark Certification Training Course

PySpark Certification T...

Course Content

- Presentation
- Class 2 Recording
- Case Study I - Perform...
- Case Study I Solution
- Case Study II - Verifi...
- Case Study II Solution
- In-Class Demo - Input...
- Introduction to Python
- Functions, OOPs, Mod...

Introduction to Python for Apache Spark > Presentation

Presentation

53 / 141 ← → ⟲ ⟳

Tuples - Built-in Tuple Functions

```
tup1= ("Hadoop", "Python", "Java")
print(len(tup1))
print(max(tup1))
print(min(tup1))
```

Gives the total length of the Tuple

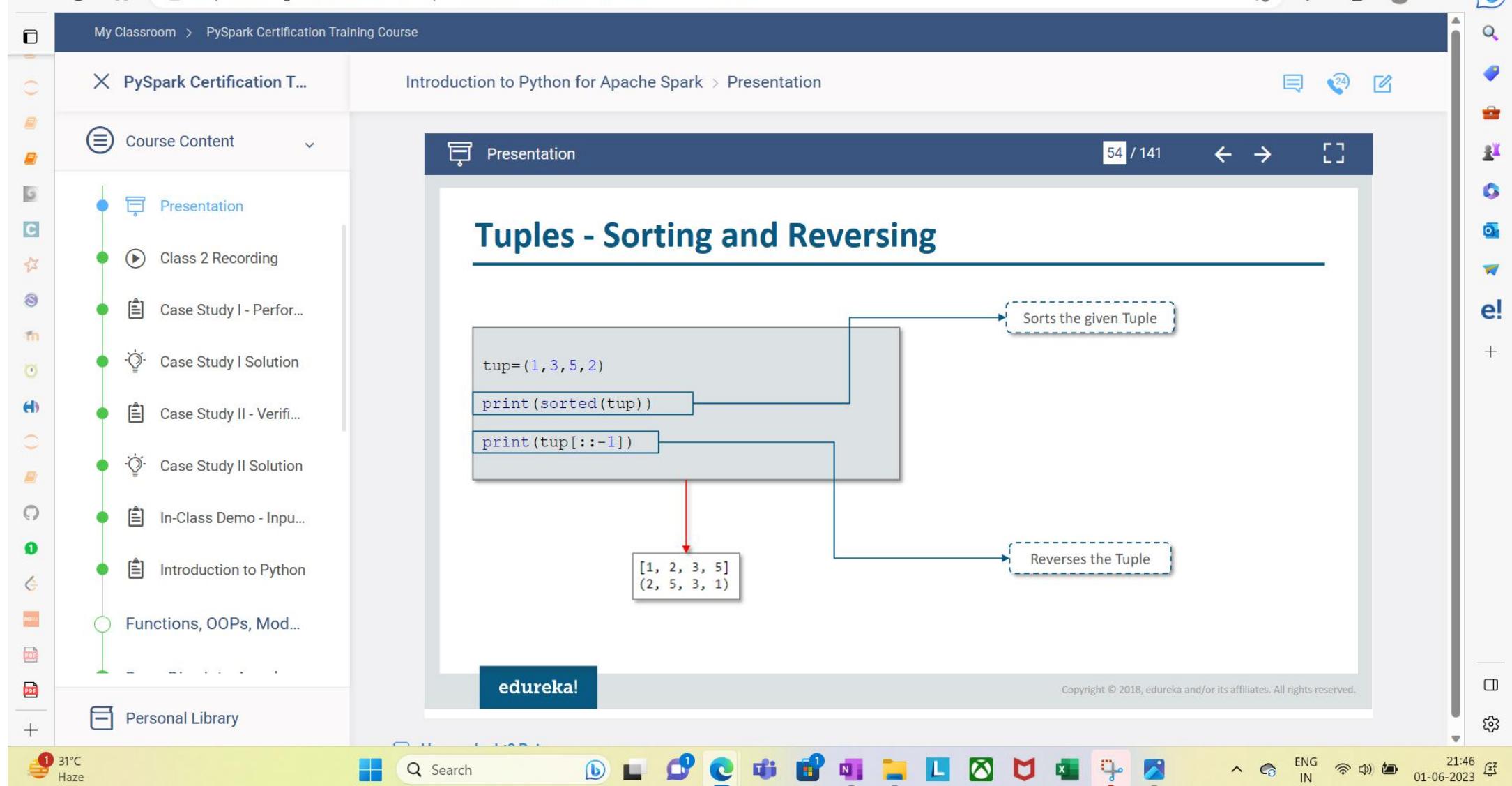
Returns item from the Tuple with max value

Returns item from the Tuple with min value

3
Python
Hadoop

edureka!

Copyright © 2018, edureka and/or its affiliates. All rights reserved.



X PySpark Certification T...

Introduction to Python for Apache Spark > Presentation



Course Content

- Presentation
- Class 2 Recording
- Case Study I - Perform...
- Case Study I Solution
- Case Study II - Verifi...
- Case Study II Solution
- In-Class Demo - Input...
- Introduction to Python
- Functions, OOPs, Mod...

Personal Library

Basic Tuple Operations

```
tup1= ("Hadoop", "Python", "Java")
print(len(tup1))
print(tup1*2)
print("Java" in tup1)
```

Shows Length of Tuple

Repetition

Membership Testing

```
3
('Hadoop', 'Python', 'Java', 'Hadoop', 'Python', 'Java')
True
```

X PySpark Certification T...

Introduction to Python for Apache Spark > Presentation



Course Content

- Presentation
- Class 2 Recording
- Case Study I - Perform...
- Case Study I Solution
- Case Study II - Verifi...
- Case Study II Solution
- In-Class Demo - Input...
- Introduction to Python
- Functions, OOPs, Mod...

Personal Library

More Operations Performed on Tuples

```
tup2=(1, 3, 5, 7)
tup3=(2, 4, 6, 8)
tup4=tup2+tup3
print(tup4)
```

```
del(tup2)
print(tup2)
```

```
(1, 3, 5, 7, 2, 4, 6, 8)
```

```
NameError: name 'tup2' is not defined
Traceback (most recent call last)
  5
  6 del(tup2)
----> 7 print(tup2)

NameError: name 'tup2' is not defined
```

Updating Tuple

Deleting elements

X PySpark Certification T...

Introduction to Python for Apache Spark > Presentation



Course Content

- Presentation
- Class 2 Recording
- Case Study I - Perform...
- Case Study I Solution
- Case Study II - Verifi...
- Case Study II Solution
- In-Class Demo - Inpu...
- Introduction to Python
- Functions, OOPs, Mod...

Personal Library

Difference between Lists and Tuples

```
list=[1,2,'Python','Android']
print(list)
list[2]='Data Science'
print(list)

tup=(1,2,3,'Data Science')
print(tup)

tup[1]='Python'
print(tup)
```

```
[1, 2, 'Python', 'Android']
[1, 2, 'Data Science', 'Android']
(1, 2, 3, 'Data Science')

Traceback (most recent call last)
<ipython-input-2-2838df9990ba> in <module>()
      8 print(tup)
      9
---> 10 tup[1]='Python'
     11 print(tup)
```

TypeError: 'tuple' object does not support item assignment

Lists are mutable whereas tuples
are immutable (we can not
change values in a tuple)

Copyright © 2018, edureka and/or its affiliates. All rights reserved.

edureka!

My Classroom > PySpark Certification Training Course

PySpark Certification T... Course Content

- Presentation
- Class 2 Recording
- Case Study I - Perform...
- Case Study I Solution
- Case Study II - Verifi...
- Case Study II Solution
- In-Class Demo - Input...
- Introduction to Python
- Functions, OOPs, Mod...

Introduction to Python for Apache Spark > Presentation

Presentation 58 / 141 ← →

A List inside A Tuple

```
tup=([1, 2, 3], [3, 4, 5], [5, 6, 7])
print(tup)
print(len(tup))
print(tup[0][0:2])
```

Lists are given as elements of a tuple

Slicing

([1, 2, 3], [3, 4, 5], [5, 6, 7])
3
[1, 2]

edureka!

Copyright © 2018, edureka and/or its affiliates. All rights reserved.

31°C Haze Search

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 59 60 61 62 63 64 65 66 67 68 69 69 70 71 72 73 74 75 76 77 78 79 79 80 81 82 83 84 85 86 87 88 89 89 90 91 92 93 94 95 96 97 98 99 99 100 101 102 103 104 105 106 107 108 109 109 110 111 112 113 114 115 116 117 118 119 119 120 121 122 123 124 125 126 127 128 129 129 130 131 132 133 134 135 136 137 138 139 139 140 141 141

ENG IN 21:46 01-06-2023

My Classroom > PySpark Certification Training Course

PySpark Certification T... Course Content

Presentation Class 2 Recording Case Study I - Perform... Case Study I Solution Case Study II - Verifi... Case Study II Solution In-Class Demo - Input... Introduction to Python Functions, OOPs, Mod...

Personal Library

Introduction to Python for Apache Spark > Presentation

Presentation 59 / 141

A List inside A Tuple

```
tup=([1,2,3],[3,4,5],[5,6,7])
print(tup)
print(len(tup))
print(tup[0][0:2])
```

Output: ([1, 2, 3], [3, 4, 5], [5, 6, 7])
3
[1, 2]

edureka!

Copyright © 2018, edureka and/or its affiliates. All rights reserved.

My Classroom > PySpark Certification Training Course

PySpark Certification T... Course Content

Presentation

Introduction to Python for Apache Spark > Presentation

60 / 141

A List inside A Tuple

```
tup = ([1, 2, 3], [3, 4, 5], [5, 6, 7])
print(tup)
print(len(tup))
print(tup[0][1])

#Updating Tuple
tup[0][1] = 90
print(tup)

#Deleting element in Tuple
del(tup[1][2])
print(tup)
```

As we are using lists inside a tuple, the overall tuple becomes *Mutable*

edureka!

Copyright © 2018, edureka and/or its affiliates. All rights reserved.

The screenshot shows a presentation slide titled "A List inside A Tuple". The slide contains Python code demonstrating tuples containing lists. Arrows point from the code to a printed output and a callout box explaining tuple mutability.

The code is:

```
tup = ([1, 2, 3], [3, 4, 5], [5, 6, 7])
print(tup)
print(len(tup))
print(tup[0][1])

#Updating Tuple
tup[0][1] = 90
print(tup)

#Deleting element in Tuple
del(tup[1][2])
print(tup)
```

The printed output is:

```
([1, 2, 3], [3, 4, 5], [5, 6, 7])
3
2
([1, 90, 3], [3, 4, 5], [5, 6, 7])
([1, 90, 3], [3, 4], [5, 6, 7])
```

A callout box states: "As we are using lists inside a tuple, the overall tuple becomes *Mutable*".

https://learning.edureka.co/classroom/presentation/651/6517/111071?tab=CourseContent

My Classroom > PySpark Certification Training Course

PySpark Certification T... Course Content

Presentation, Class 2 Recording, Case Study I - Perform..., Case Study I Solution, Case Study II - Verifi..., Case Study II Solution, In-Class Demo - Input..., Introduction to Python, Functions, OOPs, Mod...

Introduction to Python for Apache Spark > Presentation

Presentation 61 / 141

Converting Tuples into Lists

```
tuple1=(1,2,3,5,7,'a','b')
lst=list(tuple1)
print(lst)

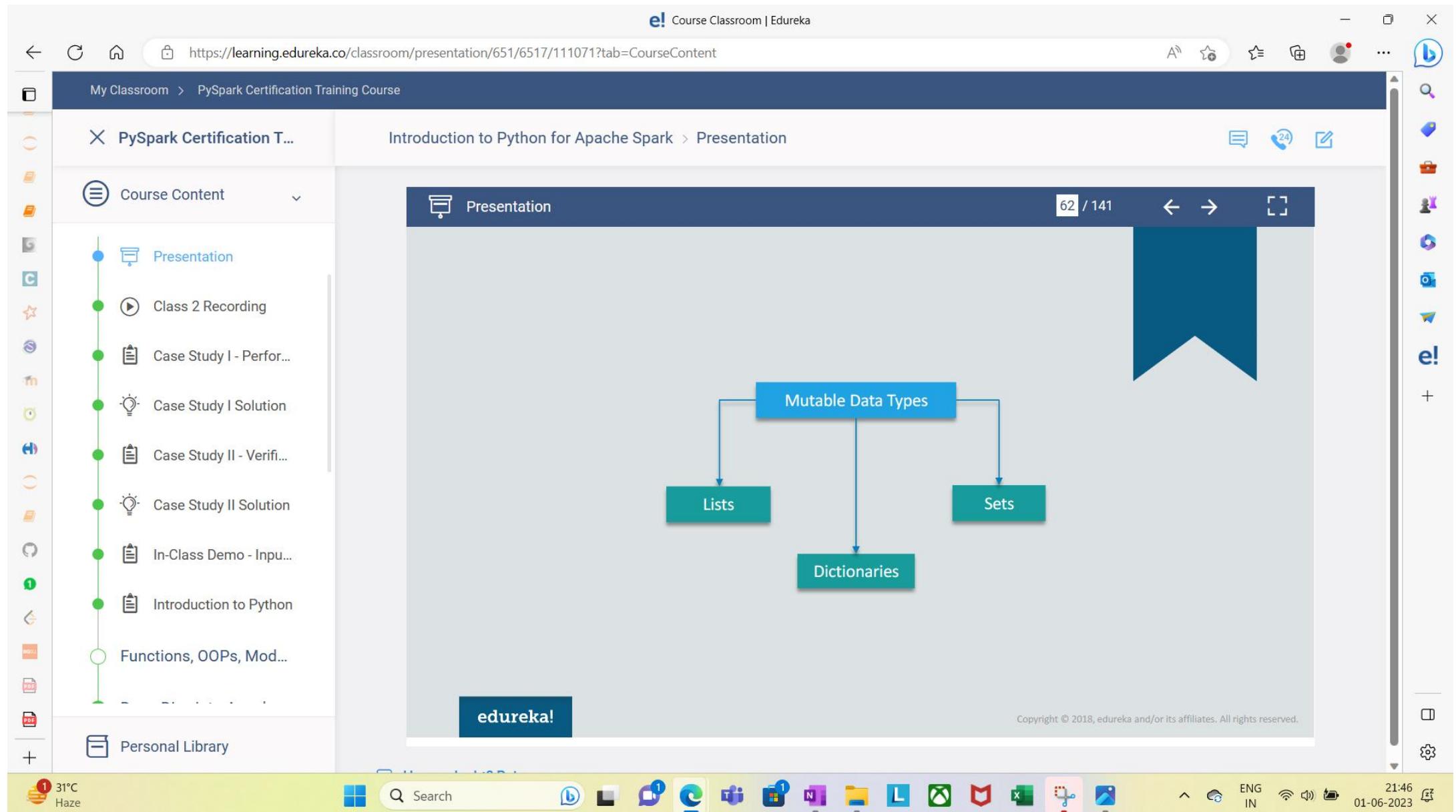
lst[1]='Python'
print(lst)

tuple2=tuple(lst)
print(tuple2)
```

Here, we are converting the tuple into a list, updating contents of the list and again converting the list into a tuple

edureka!

Copyright © 2018, edureka and/or its affiliates. All rights reserved.



https://learning.edureka.co/classroom/presentation/651/6517/111071?tab=CourseContent

My Classroom > PySpark Certification Training Course

PySpark Certification T... Course Content

Presentation, Class 2 Recording, Case Study I - Perform..., Case Study I Solution, Case Study II - Verifi..., Case Study II Solution, In-Class Demo - Input..., Introduction to Python, Functions, OOPs, Mod...

Introduction to Python for Apache Spark > Presentation

Presentation

64 / 141

Lists

List is the most versatile datatype available in Python, which can be written as a list of comma-separated values (items) between square brackets

Example list=["Marketing","Sales",8,11]

OPERATIONS

Slicing, Updating List, List Length, Repetition

Delete List Elements, Concatenation

edureka!

Copyright © 2018, edureka and/or its affiliates. All rights reserved.

X PySpark Certification T...

Introduction to Python for Apache Spark > Presentation



Course Content

- Presentation
- Class 2 Recording
- Case Study I - Perform...
- Case Study I Solution
- Case Study II - Verifi...
- Case Study II Solution
- In-Class Demo - Input...
- Introduction to Python
- Functions, OOPs, Mod...

Personal Library

Lists - Indexing and Slicing

```
list = ["Hadoop", "Python", "Android"]  
print(list[1])
```

list[i] returns the value at index i. Where i is an integer

```
print(list[0:2])
```

Returns values from index 0 to index 2 (returns values at index 0 and index 1)

```
print(list[-1])
```

The last element of any non- empty List is always list[-1]

```
Python  
['Hadoop', 'Python']  
Android
```

edureka!

Copyright © 2018, edureka and/or its affiliates. All rights reserved.

My Classroom > PySpark Certification Training Course

PySpark Certification T... Introduction to Python for Apache Spark > Presentation

Presentation 66 / 141 ← →

Lists - Updating and Deleting Elements

```
list=['Hadoop', 'Python', 'Android']
list[1]='Java'
print(list)

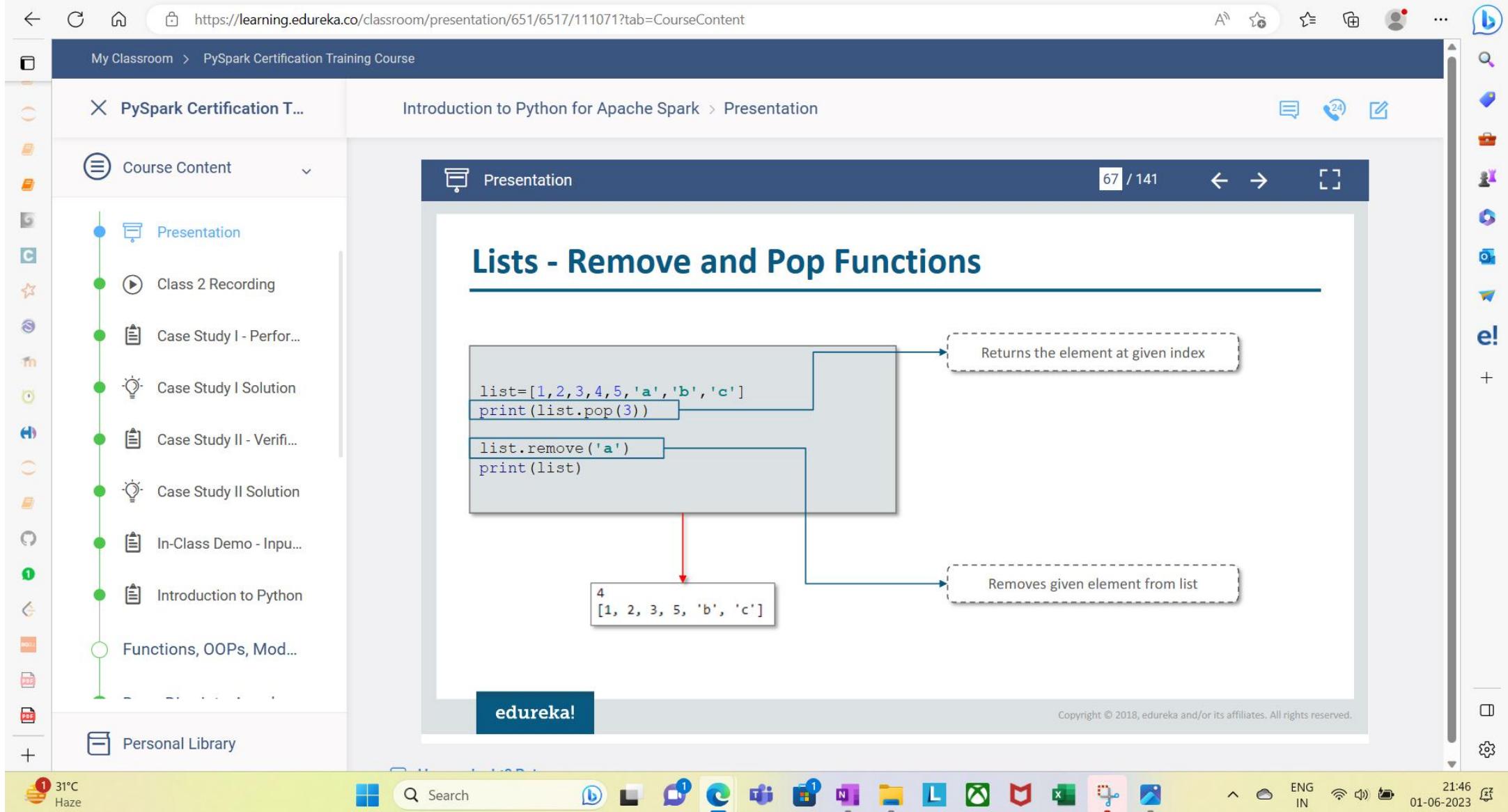
del(list[2])
print(list)
```

Updating element of a List
Deleting elements of List

['Hadoop', 'Java', 'Android']
['Hadoop', 'Java']

edureka!

Copyright © 2018, edureka and/or its affiliates. All rights reserved.



My Classroom > PySpark Certification Training Course

PySpark Certification T... Course Content

Presentation, Class 2 Recording, Case Study I - Perform..., Case Study I Solution, Case Study II - Verifi..., Case Study II Solution, In-Class Demo - Input..., Introduction to Python, Functions, OOPs, Mod...

Personal Library

Introduction to Python for Apache Spark > Presentation

Presentation 68 / 141 ← →

Lists - type()

```
list3=[1,2,5,'Python','Hadoop']
print(type(list3))
```

Returns the type of data type

```
print([x**2 for x in[1,2,3,4,5]])
```

Loop can be used in print statement for showing squares of elements of lists

<class 'list'>
[1, 4, 9, 16, 25]

edureka!

Copyright © 2018, edureka and/or its affiliates. All rights reserved.

X PySpark Certification T...

Introduction to Python for Apache Spark > Presentation



Course Content

- Presentation
- Class 2 Recording
- Case Study I - Perform...
- Case Study I Solution
- Case Study II - Verifi...
- Case Study II Solution
- In-Class Demo - Input...
- Introduction to Python
- Functions, OOPs, Mod...

Personal Library

Lists - Built-in Functions

```
list=[1,2,3]
list.append("Machine Learning")
print(list)

list.extend(['g','h'])
print(list)

list.insert(1,'Scripting')
print(list)

list.remove(3)
print(list)
```

Adds an item to the end of the List

Inserts many items at the end of list

Inserts an item at a given position

Removes an item from the List

edureka!

Copyright © 2018, edureka and/or its affiliates. All rights reserved.

https://learning.edureka.co/classroom/presentation/651/6517/111071?tab=CourseContent

My Classroom > PySpark Certification Training Course

PySpark Certification T... Course Content

Presentation, Class 2 Recording, Case Study I - Perform..., Case Study I Solution, Case Study II - Verifi..., Case Study II Solution, In-Class Demo - Input..., Introduction to Python, Functions, OOPs, Mod...

Personal Library

Introduction to Python for Apache Spark > Presentation

Presentation 70 / 141 ← →

Lists - Built-in Functions

```
list1=['Python', 'XYZ', 'ABC', 'PQR']
print(list1)
print(sorted(list1))
print(list1[::-1])
```

['Python', 'XYZ', 'ABC', 'PQR']
['ABC', 'PQR', 'Python', 'XYZ']
['PQR', 'ABC', 'XYZ', 'Python']

Returns Sorted List
Reverses the List

edureka!

Copyright © 2018, edureka and/or its affiliates. All rights reserved.

31°C Haze Search ENG IN 01-06-2023 21:46

X PySpark Certification T...

Introduction to Python for Apache Spark > Presentation



Course Content

- Presentation
- Class 2 Recording
- Case Study I - Perform...
- Case Study I Solution
- Case Study II - Verifi...
- Case Study II Solution
- In-Class Demo - Input...
- Introduction to Python
- Functions, OOPs, Mod...

Personal Library

A Tuple inside A List

```
list=[(1, 2, 3), ("Python", "Java")]  
print(list)
```

```
print(len(list))
```

```
print(list[1][0:1])
```

```
[(1, 2, 3), ('Python', 'Java')]  
2  
('Python',)
```

Tuples are given as elements in List

Slicing

My Classroom > PySpark Certification Training Course

PySpark Certification T... Course Content

Presentation Class 2 Recording Case Study I - Perform... Case Study I Solution Case Study II - Verifi... Case Study II Solution In-Class Demo - Input... Introduction to Python Functions, OOPs, Mod...

Personal Library

Introduction to Python for Apache Spark > Presentation

Presentation 72 / 141 ← →

A Tuple inside A List

```
list=[(1,2,3), ("Python", "Java")]
print(list)

print(len(list))

print(list[1][0:1])
```

The diagram shows a list `list` containing two elements: a tuple `(1, 2, 3)` and another tuple `("Python", "Java")`. The index `0` points to the first element (tuple), and index `1` points to the second element (tuple). The first element `(1, 2, 3)` is further expanded to show indices `0`, `1`, and `2` pointing to the individual integers. The second element `("Python", "Java")` is expanded to show indices `0` and `1` pointing to the individual strings.

edureka!

Copyright © 2018, edureka and/or its affiliates. All rights reserved.

31°C
Haze

Search

21:46
01-06-2023

https://learning.edureka.co/classroom/presentation/651/6517/111071?tab=CourseContent

My Classroom > PySpark Certification Training Course

PySpark Certification T... Course Content

- Presentation
- Class 2 Recording
- Case Study I - Perform...
- Case Study I Solution
- Case Study II - Verifi...
- Case Study II Solution
- In-Class Demo - Inpu...
- Introduction to Python
- Functions, OOPs, Mod...

Introduction to Python for Apache Spark > Presentation

Presentation 75 / 141

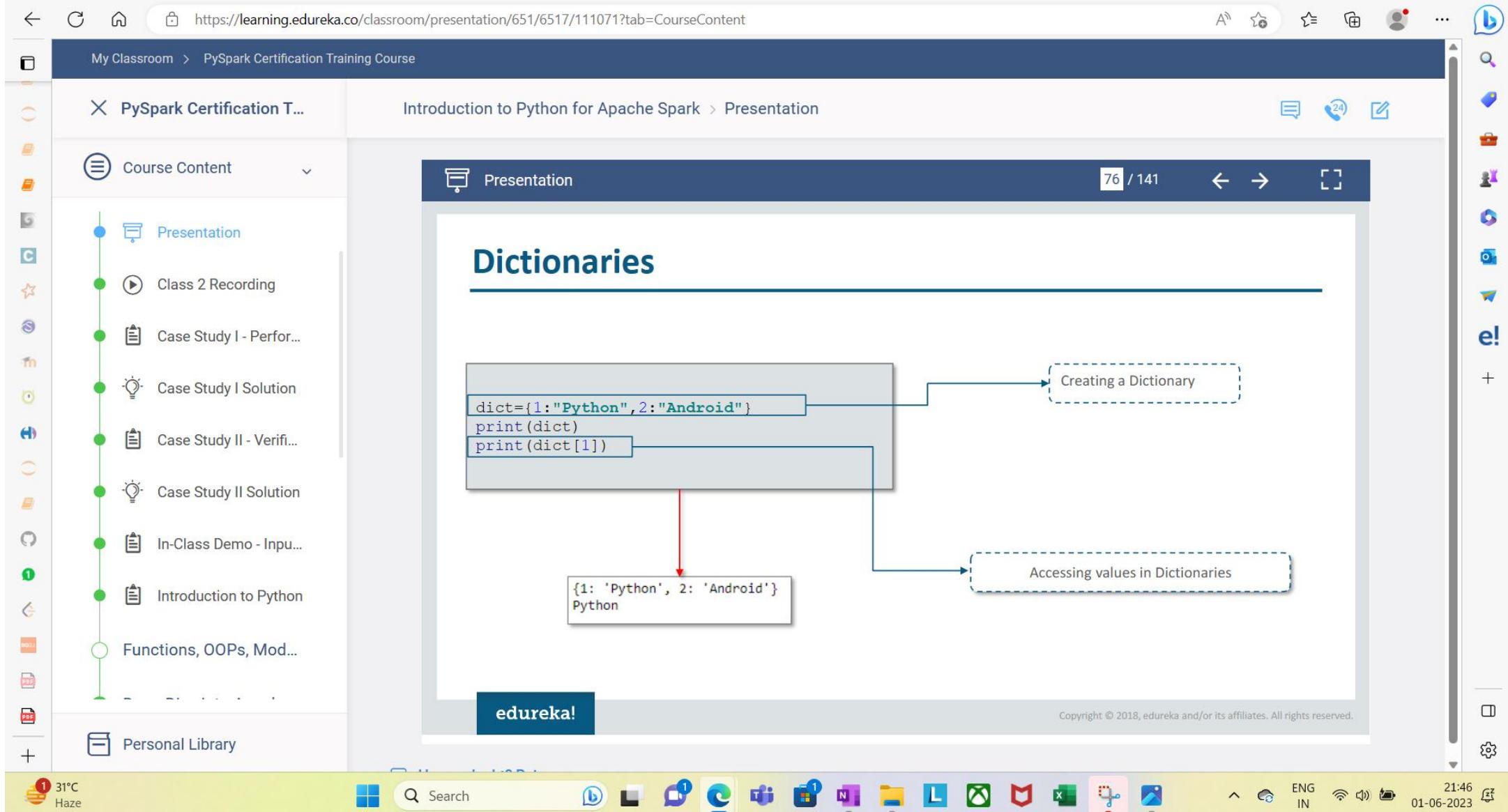
Dictionaries

Dictionary is an unordered collection of key-value pairs. It is generally used when we have a huge amount of data

Example: `dict={1:"Python"}`

edureka!

Copyright © 2018, edureka and/or its affiliates. All rights reserved.



My Classroom > PySpark Certification Training Course

PySpark Certification T... Introduction to Python for Apache Spark > Presentation

Presentation 77 / 141 ← →

Dictionaries - Updating and Deleting Elements

```
dict={1:"Python",2:"Android"}  
print(dict[1])  
  
dict[1]="Javascript"  
print(dict)  
  
del(dict[2])  
print(dict)
```

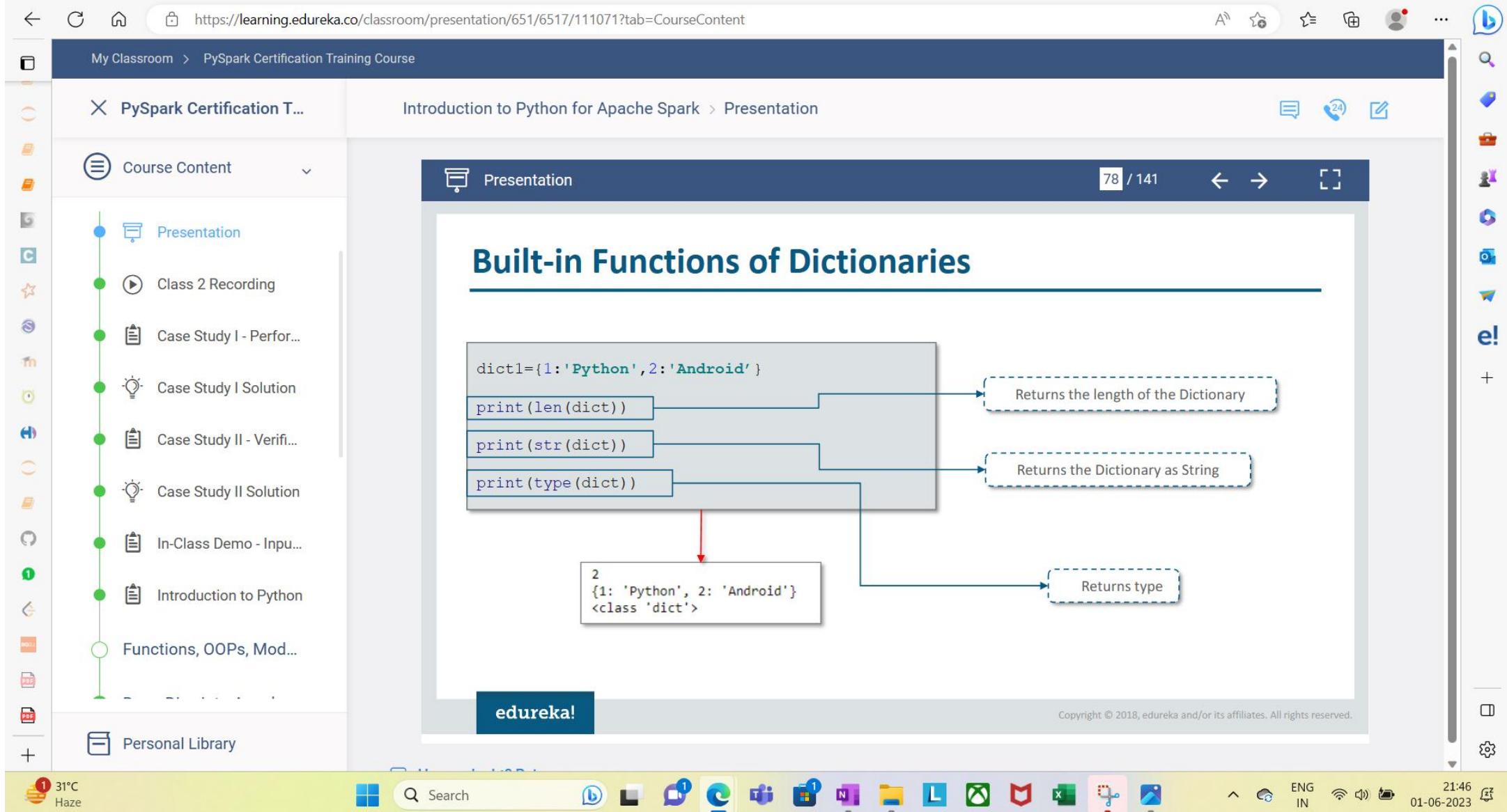
Updating elements
Deleting elements

{1: 'Python', 2: 'Android'}
{1: 'Javascript', 2: 'Android'}
{1: 'Javascript'}

edureka!

Copyright © 2018, edureka and/or its affiliates. All rights reserved.

31°C Haze Search





My Classroom > PySpark Certification Training Course

Introduction to Python for Apache Spark > Presentation



 Course Content

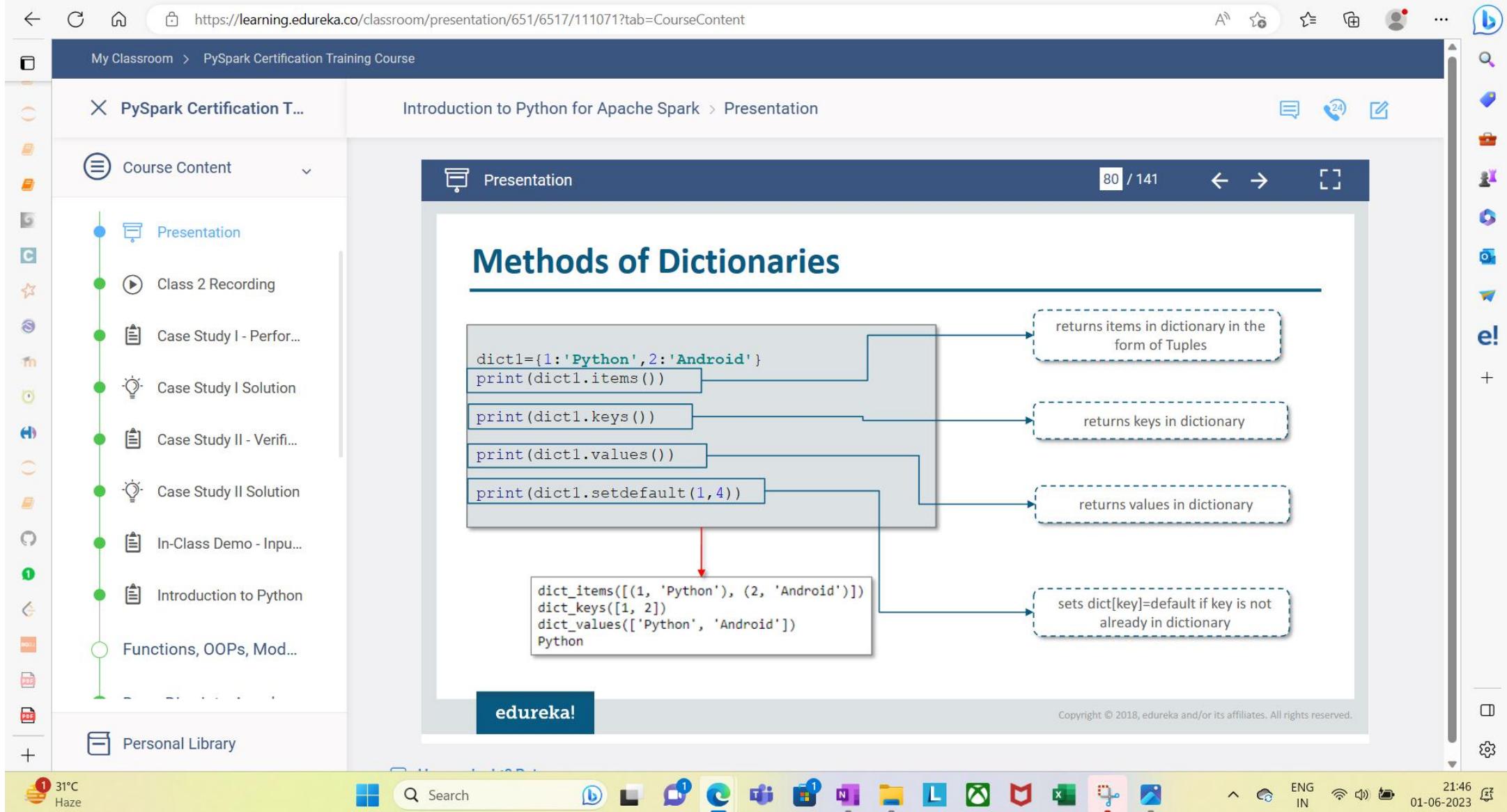
-  Presentation
 -  Class 2 Recording
 -  Case Study I - Perform...
 -  Case Study I Solution
 -  Case Study II - Verifi...
 -  Case Study II Solution
 -  In-Class Demo - Inpu...
 -  Introduction to Python
 -  Functions, OOPs, Mod...

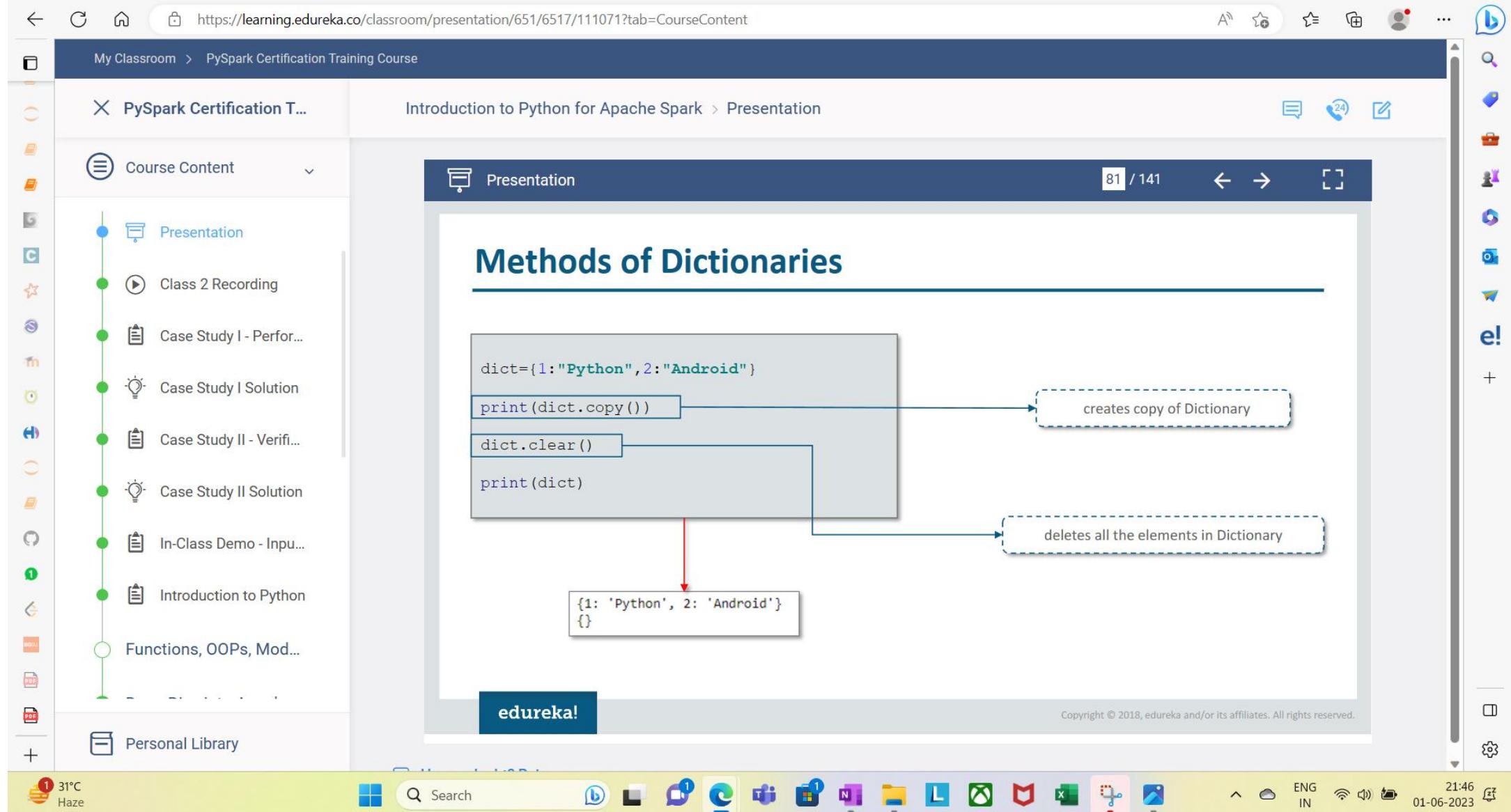
Built-in Functions of Dictionaries

```
rec = {'name': {'first': 'Bob', 'last': 'Smith'},  
      'jobs': ['dev', 'mgr'], 'age': 40.5}  
print(rec.get('name'))
```

Returns the value of
the key passed

```
{'first': 'Bob', 'last': 'Smith'}
```





My Classroom > PySpark Certification Training Course

PySpark Certification T... Course Content

Presentation

Introduction to Python for Apache Spark > Presentation

82 / 141

Sorting Keys for Loops

```
dic={3:'Python',1:'Java',2:'Big Data'}
ks=list(dic.keys())
print(ks)

sk=sorted(ks)
print(sk)

for key in sk:
    print(key,'=>',dic[key])
```

ks consists list of dictionary Keys

sk consists sorted keys of dictionary

Prints sorted keys with their respective values from dictionaries

[3, 1, 2]
[1, 2, 3]
1 => Java
2 => Big Data
3 => Python

edureka!

Copyright © 2018, edureka and/or its affiliates. All rights reserved.



My Classroom > PySpark Certification Training Course

X PySpark Certification T...

Introduction to Python for Apache Spark > Presentation



Course Content

- Presentation
- Class 2 Recording
- Case Study I - Perform...
- Case Study I Solution
- Case Study II - Verifi...
- Case Study II Solution
- In-Class Demo - Input...
- Introduction to Python
- Functions, OOPs, Mod...

Personal Library

Presentation

85 / 141



Sets

Set is an unordered collection of unique items. Set is defined by values separated by comma inside braces {}

Sets can also be created by calling the built-in set function:

```
1 | x = set('Welcome To Edureka')  
2 | print(x)
```

edureka!

Copyright © 2018, edureka and/or its affiliates. All rights reserved.



1

31°C

Haze



Search



ENG

IN



01-06-2023

21:46

My Classroom > PySpark Certification Training Course

PySpark Certification T... Course Content

- Presentation
- Class 2 Recording
- Case Study I - Perform...
- Case Study I Solution
- Case Study II - Verifi...
- Case Study II Solution
- In-Class Demo - Input...
- Introduction to Python
- Functions, OOPs, Mod...

Introduction to Python for Apache Spark > Presentation

Presentation 86 / 141 ← →

Set Operation - Union

Union of A and B is a set of all elements from both Sets. Union is performed using | operator. Consider the below example:

```
Union
1 A = {1, 2, 3, 4}
2 B = {3, 4, 5, 6}
3 print ( A | B)
output = {1, 2, 3, 4, 5, 6}
```

edureka!

Copyright © 2018, edureka and/or its affiliates. All rights reserved.

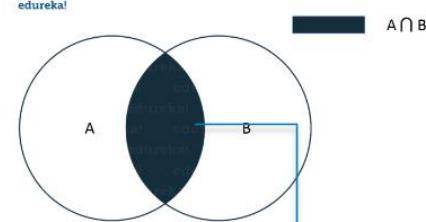
My Classroom > PySpark Certification Training Course

PySpark Certification T... Introduction to Python for Apache Spark > Presentation

Presentation 87 / 141 ← →

Set Operation - Intersection

Intersection of A and B is a set of elements that are common in both Sets. Intersection is performed using & operator



A \cap B

edureka!

edureka!

Copyright © 2018, edureka and/or its affiliates. All rights reserved.

```
1 A = {1, 2, 3, 4}
2 B = {3, 4, 5, 6}
3 print ( A & B )
```

Output = {3,4}

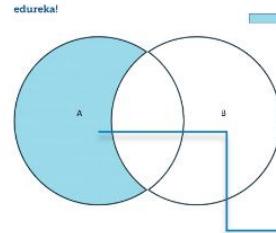
My Classroom > PySpark Certification Training Course

PySpark Certification T... Introduction to Python for Apache Spark > Presentation

Presentation 88 / 141 ← →

Set Operation - Difference

Difference of A and B ($A - B$) is a set of elements that are only in A but not in B. Similarly, $B - A$ is a set of element in B but not in A



edureka!

```
1 A = {1, 2, 3, 4, 5}
2 B = {4, 5, 6, 7, 8}
3 print(A - B)
```

Output = {1, 2, 3}

edureka!

Copyright © 2018, edureka and/or its affiliates. All rights reserved.

X PySpark Certification T...

Introduction to Python for Apache Spark > Presentation



Course Content

- Presentation
- Class 2 Recording
- Case Study I - Perform...
- Case Study I Solution
- Case Study II - Verifi...
- Case Study II Solution
- In-Class Demo - Input...
- Introduction to Python
- Functions, OOPs, Mod...

Personal Library

Operations Performed on Sets

```
set1={'a', 'b', 'd', 'o', 'v'}  
  
set2={'a', 'c', 'd', 'o', 'e'}  
  
print(set1|set2)  
print(set1&set2)  
print(set1-set2)
```

```
{'e', 'a', 'o', 'd', 'c', 'v', 'b'}  
{'d', 'o', 'a'}  
{'v', 'b'}
```

Union

Intersection

Difference

My Classroom > PySpark Certification Training Course

PySpark Certification T... Introduction to Python for Apache Spark > Presentation

Presentation 91 / 141 ← →

Operations Performed on Sets

```
s={1,2,3, 'a', 'b'}
s.add('c')
print(s)

s.remove(1)
print(s)

s.discard(3)
print(s)

s.pop()
print(s)

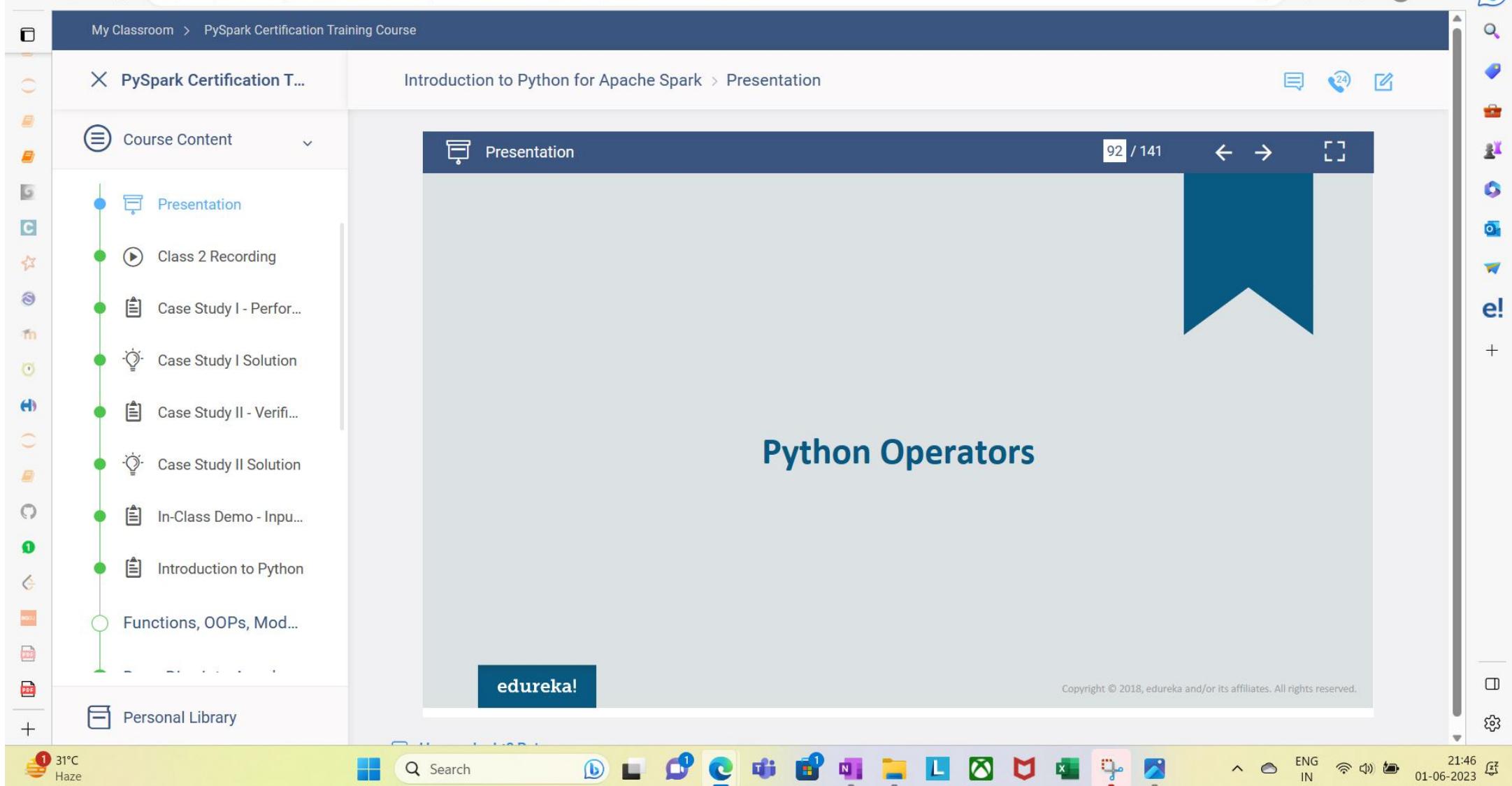
s.clear()
print(s)
```

{1, 2, 3, 'a', 'c', 'b'}
{2, 3, 'a', 'c', 'b'}
{2, 'a', 'c', 'b'}
'a', 'c', 'b'
set()

Adds element to set
Removes element from set
Removes element from set if present
Removes and returns an arbitrary element from list
Removes all element from set

edureka!

Copyright © 2018, edureka and/or its affiliates. All rights reserved.



https://learning.edureka.co/classroom/presentation/651/6517/111071?tab=CourseContent

My Classroom > PySpark Certification Training Course

PySpark Certification T... Course Content

- Presentation
- Class 2 Recording
- Case Study I - Perform...
- Case Study I Solution
- Case Study II - Verifi...
- Case Study II Solution
- In-Class Demo - Input...
- Introduction to Python
- Functions, OOPs, Mod...

Introduction to Python for Apache Spark > Presentation

Presentation 93 / 141 ← →

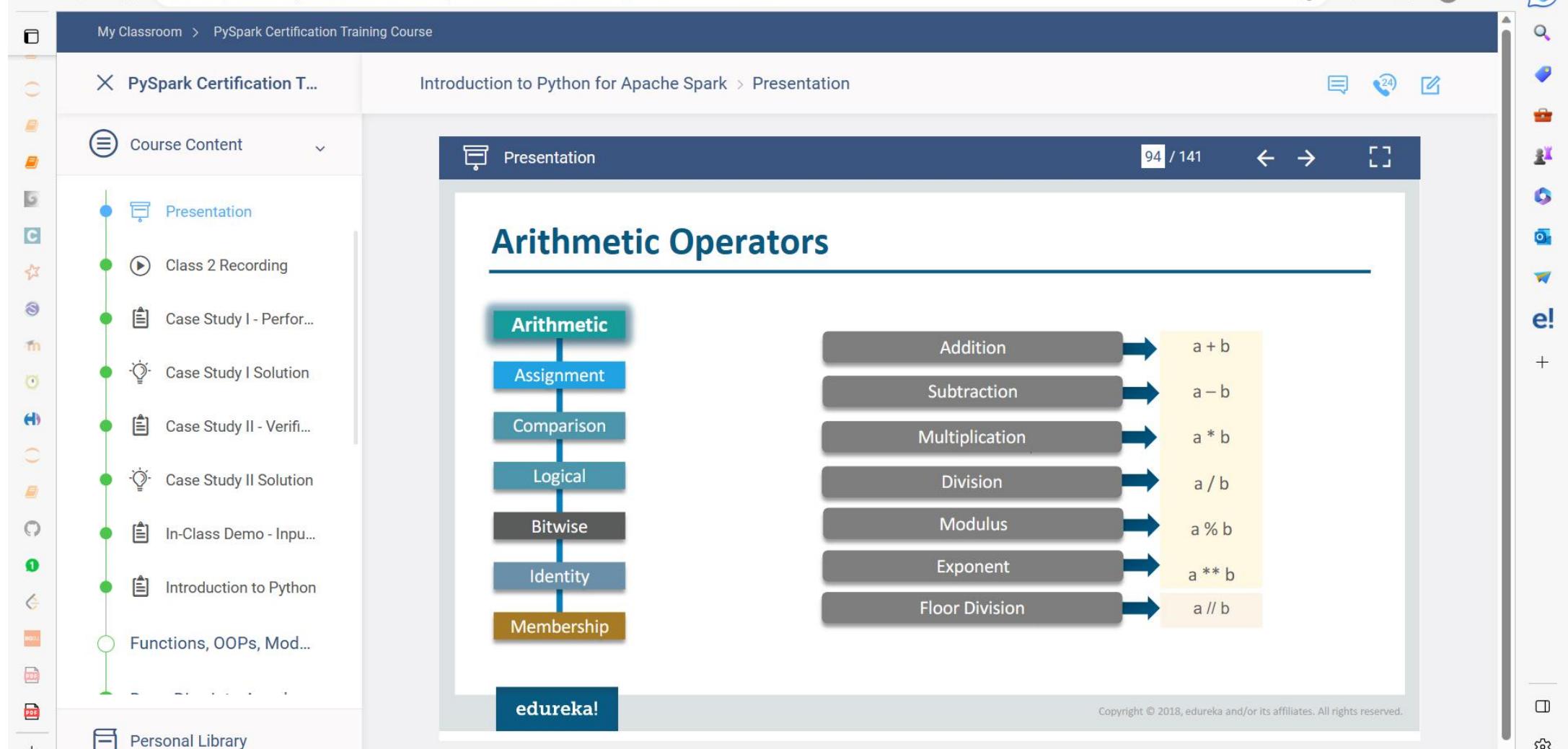
Operators

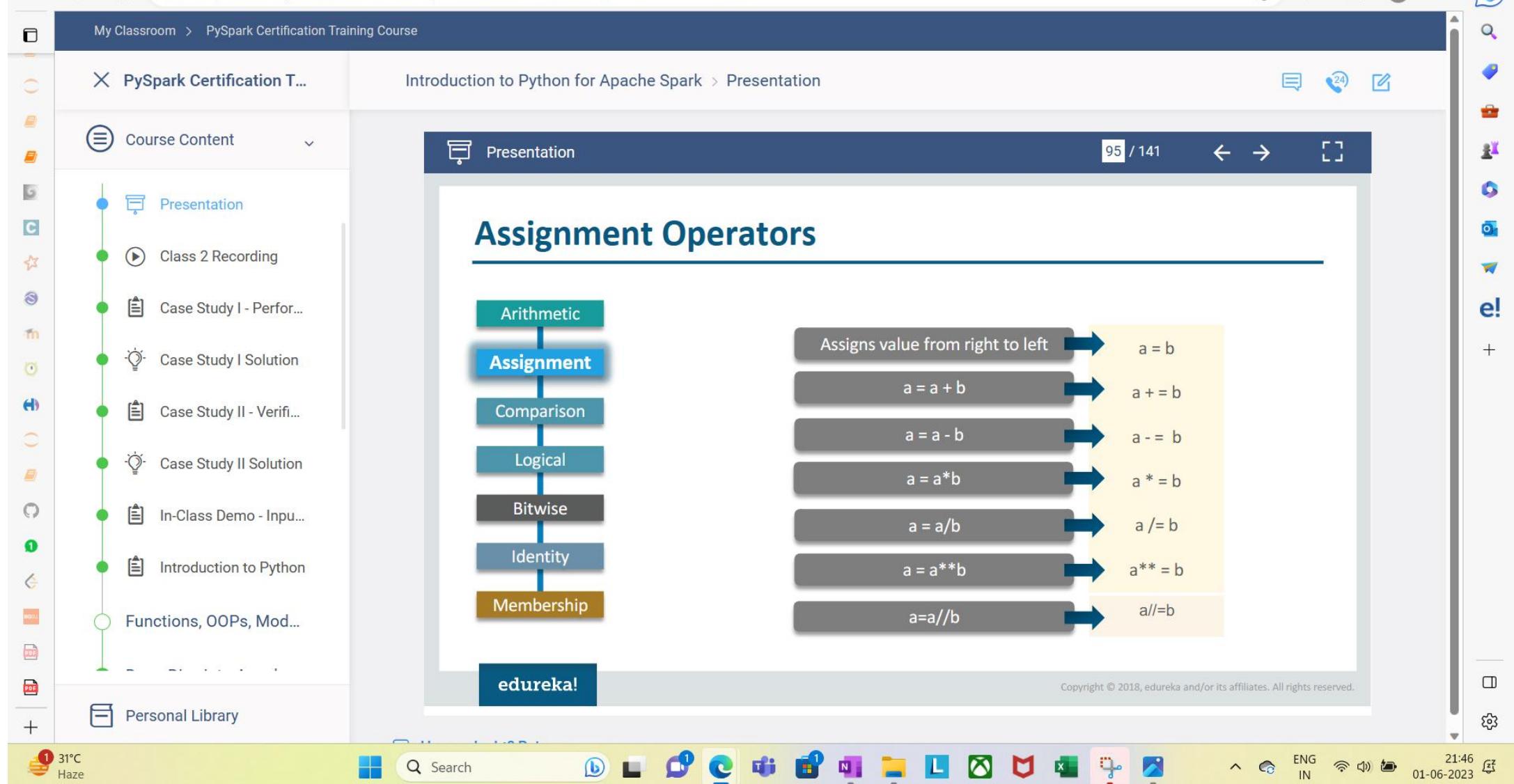
Operators are the constructs which can manipulate the values of the Operands. Consider the expression $2 + 3 = 5$, here 2 and 3 are Operands and + is called Operator

```
graph TD; Assignment --> Arithmetic; Assignment --> Comparison; Bitwise --> Bitwise; Membership --> Membership; Logical --> Logical; Identity --> Identity;
```

edureka!

Copyright © 2018, edureka and/or its affiliates. All rights reserved.





My Classroom > PySpark Certification Training Course

PySpark Certification T... Introduction to Python for Apache Spark > Presentation

Presentation 96 / 141 ← →

Comparison Operators

Arithmetic	Equal To	a == b
Assignment	Not Equal To	a != b
Comparison	Greater Than	a > b
Logical	Less Than	a < b
Bitwise	Greater Than Equal To	a >= b
Identity	Less Than Equal To	a <= b
Membership		

edureka!

Copyright © 2018, edureka and/or its affiliates. All rights reserved.

Personal Library

31°C Haze

Search

Cloud, Mail, Edge, OneDrive, Teams, Microsoft Store, File Explorer, LinkedIn, Excel, Word, Powerpoint, Paint, and other Microsoft Office icons.



e!

+

21:46

01-06-2023

My Classroom > PySpark Certification Training Course

PySpark Certification T... Introduction to Python for Apache Spark > Presentation

Presentation 97 / 141 ← →

Logical Operators

- Arithmetic
- Assignment
- Comparison
- Logical**
- Bitwise
- Identity
- Membership

 a and b Returns a if a is False, b otherwise

 a or b Returns b if b is False, a otherwise

 not a Returns True if a is True, False otherwise

edureka!

Copyright © 2018, edureka and/or its affiliates. All rights reserved.

My Classroom > PySpark Certification Training Course

X PySpark Certification T...

Introduction to Python for Apache Spark > Presentation



Course Content

- Presentation
- Class 2 Recording
- Case Study I - Perform...
- Case Study I Solution
- Case Study II - Verifi...
- Case Study II Solution
- In-Class Demo - Input...
- Introduction to Python
- Functions, OOPs, Mod...

Personal Library

Bitwise Operators

Arithmetic

Assignment

Comparison

Logical

Bitwise

Identity

Membership

Binary AND	→	$a \& b$
Binary OR	→	$a b$
Binary XOR	→	$a ^ b$
Binary NOT	→	$a \sim b$
Binary Left Shift	→	$a <<$
Binary Right Shift	→	$a >> b$

edureka!

Copyright © 2018, edureka and/or its affiliates. All rights reserved.



31°C

Haze



Search



https://learning.edureka.co/classroom/presentation/651/6517/111071?tab=CourseContent

My Classroom > PySpark Certification Training Course

PySpark Certification T... Introduction to Python for Apache Spark > Presentation

Course Content

- Presentation
- Class 2 Recording
- Case Study I - Perform...
- Case Study I Solution
- Case Study II - Verifi...
- Case Study II Solution
- In-Class Demo - Input...
- Introduction to Python
- Functions, OOPs, Mod...

Presentation

99 / 141

Identity Operators

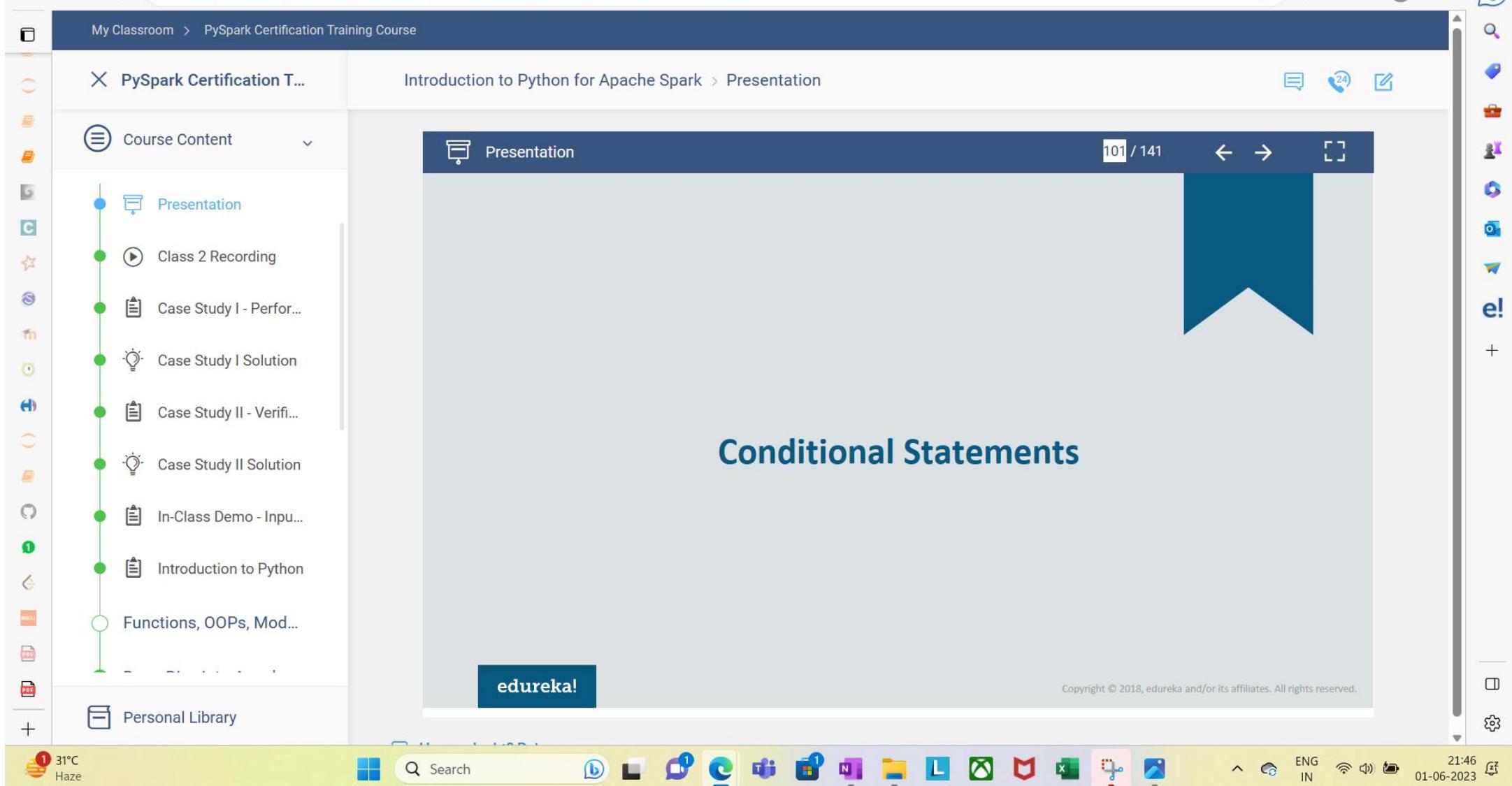
Arithmetic
Assignment
Comparison
Logical
Bitwise
Identity
Membership

is → Evaluates to TRUE if the variables on either side of the operator point to the same object and FALSE otherwise

is not → Evaluates to FALSE if the variables on either side of the operator point to the same object and TRUE otherwise

edureka!

Copyright © 2018, edureka and/or its affiliates. All rights reserved.



X PySpark Certification T...

Introduction to Python for Apache Spark > Presentation



Course Content

- Presentation
- Class 2 Recording
- Case Study I - Perform...
- Case Study I Solution
- Case Study II - Verifi...
- Case Study II Solution
- In-Class Demo - Input...
- Introduction to Python
- Functions, OOPs, Mod...

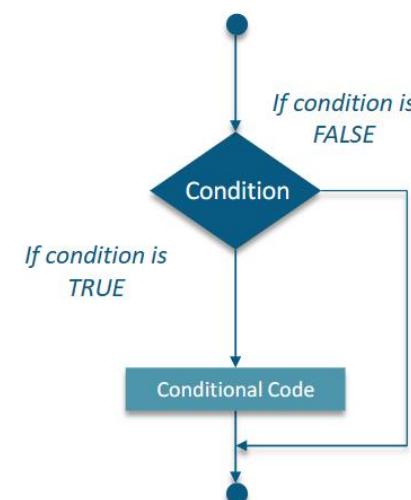
Personal Library

Conditional Statements

Conditional statements are used to execute a statement or a group of statements when some condition is true

Types of Conditional Statements

If Elif Else



edureka!

Copyright © 2018, edureka and/or its affiliates. All rights reserved.

My Classroom > PySpark Certification Training Course

PySpark Certification T... Course Content

- Presentation
- Class 2 Recording
- Case Study I - Perform...
- Case Study I Solution
- Case Study II - Verifi...
- Case Study II Solution
- In-Class Demo - Input...
- Introduction to Python
- Functions, OOPs, Mod...

Introduction to Python for Apache Spark > Presentation

Presentation 103 / 141 ← →

If Elif Else Statements

```
graph TD; Start((Start)) --> IfXLessY{If X<Y}; IfXLessY -- False --> ElifXMoreY{Elif X>Y}; IfXLessY -- True --> StatementA[Statement_a]; ElifXMoreY -- False --> ElseXEqualY{Else X=Y}; ElifXMoreY -- True --> StatementB[Statement_b]; ElseXEqualY --> StatementC[Statement_c]; StatementA --> Rest[Rest of the code]; StatementB --> Rest; StatementC --> Rest;
```

The flowchart illustrates the execution of an if-elif-else conditional statement. It begins at a red "Start" node, which leads to a diamond decision point labeled "If X<Y". From this point, two paths emerge: a green "False" path leading to another diamond decision point labeled "Elif X>Y", and a green "True" path leading directly to a dark grey rectangular box labeled "Statement_a". From the "Elif X>Y" diamond, a green "False" path leads to a third diamond decision point labeled "Else X=Y", while a green "True" path leads directly to a dark grey rectangular box labeled "Statement_b". Finally, all three statements ("Statement_a", "Statement_b", and "Statement_c") converge at a green rectangular box labeled "Rest of the code".

edureka!

Copyright © 2018, edureka and/or its affiliates. All rights reserved.

31°C Haze Search ENG IN 01-06-2023 21:46

My Classroom > PySpark Certification Training Course

PySpark Certification T...

Course Content

- Presentation
- Class 2 Recording
- Case Study I - Perform...
- Case Study I Solution
- Case Study II - Verifi...
- Case Study II Solution
- In-Class Demo - Input...
- Introduction to Python
- Functions, OOPs, Mod...

Introduction to Python for Apache Spark > Presentation

Presentation

104 / 141 ← →

If, Elif and Else Statements

Consider the syntax and example below:

Syntax:

```
if condition1:  
    statements  
  
elif condition2:  
    statements  
  
else:  
    statements
```

Example:

```
X=10  
Y=12  
  
if(X<Y):  
    print('X is less than Y')  
elif(X>Y):  
    print('X is greater than Y')  
else:  
    print('X and Y are equal')
```

X is less than Y

edureka!

Copyright © 2018, edureka and/or its affiliates. All rights reserved.

Course Classroom | Edureka

My Classroom > PySpark Certification Training Course

PySpark Certification T... Introduction to Python for Apache Spark > Presentation

Course Content

- Presentation
- Class 2 Recording
- Case Study I - Perform...
- Case Study I Solution
- Case Study II - Verifi...
- Case Study II Solution
- In-Class Demo - Inpu...
- Introduction to Python
- Functions, OOPs, Mod...

Personal Library

Presentation

105 / 141 ← → []

Loops

edureka!

Copyright © 2018, edureka and/or its affiliates. All rights reserved.

31°C Haze

Search

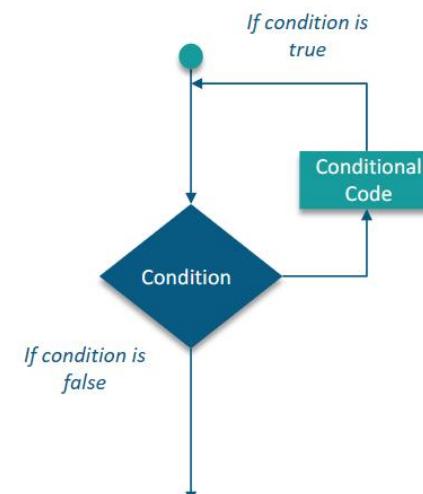
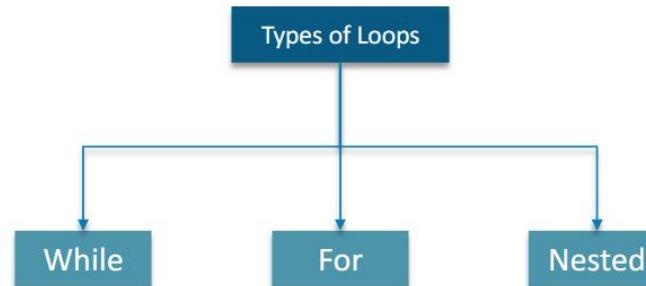
1 24

ENG IN

01-06-2023

Loops

A loop statement allows us to execute a statement or a group of statements multiple times



https://learning.edureka.co/classroom/presentation/651/6517/111071?tab=CourseContent

My Classroom > PySpark Certification Training Course

PySpark Certification T... Course Content

- Presentation
- Class 2 Recording
- Case Study I - Perform...
- Case Study I Solution
- Case Study II - Verifi...
- Case Study II Solution
- In-Class Demo - Input...
- Introduction to Python
- Functions, OOPs, Mod...

Introduction to Python for Apache Spark > Presentation

Presentation 107 / 141

While Loop

"While" loops are known as indefinite or conditional loops. They will keep iterating until certain conditions are met. There is no guarantee ahead of time regarding how many times the loop will iterate

Syntax:

```
1 | while expression:  
2 |     statements
```

```
graph TD; Start([Start]) --> Cond{While Condition?}; Cond -- True --> Body[Body of the loop]; Body --> Cond; Cond -- False --> Exit([Exit]);
```

edureka!

Copyright © 2018, edureka and/or its affiliates. All rights reserved.

My Classroom > PySpark Certification Training Course

PySpark Certification T... Course Content

Presentation

Introduction to Python for Apache Spark > Presentation

108 / 141

While Loop Example

```
count=0
while(count<5):
    print(count)
    count=count+1

print("Good bye!")
```

Condition

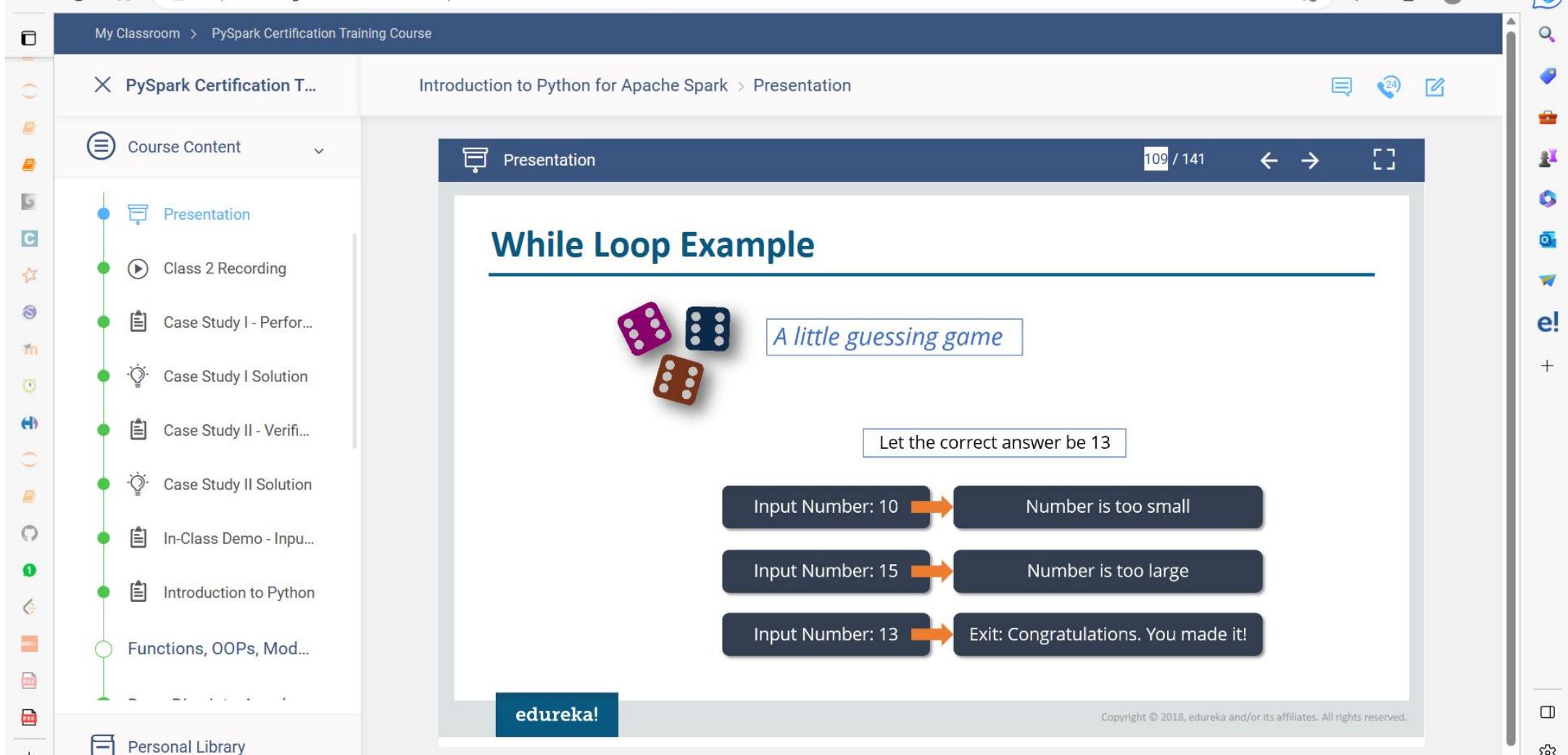
Conditional Code

Output – Prints all the integers between 0 and 5

0
1
2
3
4
Good bye!

edureka!

Copyright © 2018, edureka and/or its affiliates. All rights reserved.



X PySpark Certification T...

Introduction to Python for Apache Spark > Presentation



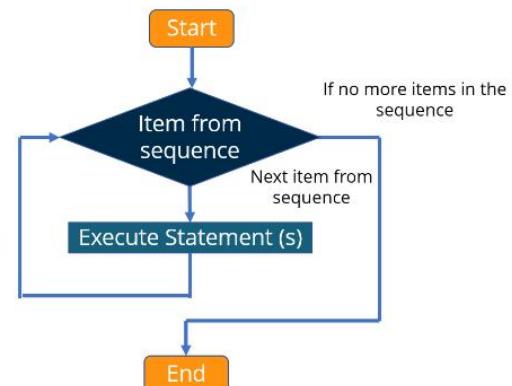
 Course Content

- Presentation
 - Class 2 Recording
 - Case Study I - Performance
 - Case Study I Solution
 - Case Study II - Verification
 - Case Study II Solution
 - In-Class Demo - Input
 - Introduction to Python
 - Functions, OOPs, Modules

For Loop

- “For” loop is a Python loop which repeats a group of statements a specified number of times.
The for loop provides a syntax where the following information is provided:
 - Boolean condition
 - The initial value of the counting variable
 - Incrementation of counting variable

```
1 for <variable> in <range>:  
2     stmt1  
3     stmt2  
4     ...  
5     stmtn
```



https://learning.edureka.co/classroom/presentation/651/6517/111071?tab=CourseContent

My Classroom > PySpark Certification Training Course

PySpark Certification T... Course Content

- Presentation
- Class 2 Recording
- Case Study I - Perform...
- Case Study I Solution
- Case Study II - Verifi...
- Case Study II Solution
- In-Class Demo - Inpu...
- Introduction to Python
- Functions, OOPs, Mod...

Introduction to Python for Apache Spark > Presentation

Presentation 111 / 141

For Loop Example

Difference between For and While loop is that, in While loop we don't know the amount of iterations, where as in For loop we are aware of how many times the block of code will be executed

```
fruits=['Banana', 'Apple', 'Grapes']  
  
for index in range(len(fruits)):  
    print(fruits[index])
```

edureka!

Copyright © 2018, edureka and/or its affiliates. All rights reserved.

https://learning.edureka.co/classroom/presentation/651/6517/111071?tab=CourseContent

My Classroom > PySpark Certification Training Course

PySpark Certification T... Course Content

- Presentation
- Class 2 Recording
- Case Study I - Perform...
- Case Study I Solution
- Case Study II - Verifi...
- Case Study II Solution
- In-Class Demo - Input...
- Introduction to Python
- Functions, OOPs, Mod...

Introduction to Python for Apache Spark > Presentation

Presentation 112 / 141 ← →

For Loop Example

We will be using For loop to write a program that calculates the factorial of any number

Factorial = $n(n-1)(n-2)\dots1$

Example

$$3! = 3(2)(1)$$
$$4! = 4(3)(2)(1)$$
$$5! = 5(4)(3)(2)(1)$$

edureka!

Copyright © 2018, edureka and/or its affiliates. All rights reserved.

My Classroom > PySpark Certification Training Course

PySpark Certification T...

Course Content

- Presentation
- Class 2 Recording
- Case Study I - Perform...
- Case Study I Solution
- Case Study II - Verifi...
- Case Study II Solution
- In-Class Demo - Input...
- Introduction to Python
- Functions, OOPs, Mod...

Introduction to Python for Apache Spark > Presentation

Presentation

113 / 141

Nested Loops

Nested Loop, basically means a loop inside a loop. It can be a For loop inside a While loop and vice-versa. It can also be a While loop inside a While loop or For loop inside a For loop

```
count=1
for i in range(10):
    print(str(i)*i)

    for j in range(0,i):
        count=count+1
```

For loop inside a For loop

1
22
333
4444
55555
666666
7777777
88888888
999999999

Output

edureka!

Copyright © 2018, edureka and/or its affiliates. All rights reserved.

X PySpark Certification T...

Introduction to Python for Apache Spark > Presentation

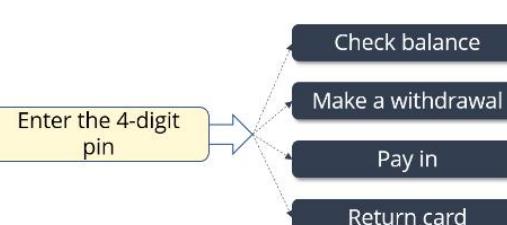


Course Content

- Presentation
- Class 2 Recording
- Case Study I - Perform...
- Case Study I Solution
- Case Study II - Verifi...
- Case Study II Solution
- In-Class Demo - Input...
- Introduction to Python
- Functions, OOPs, Mod...

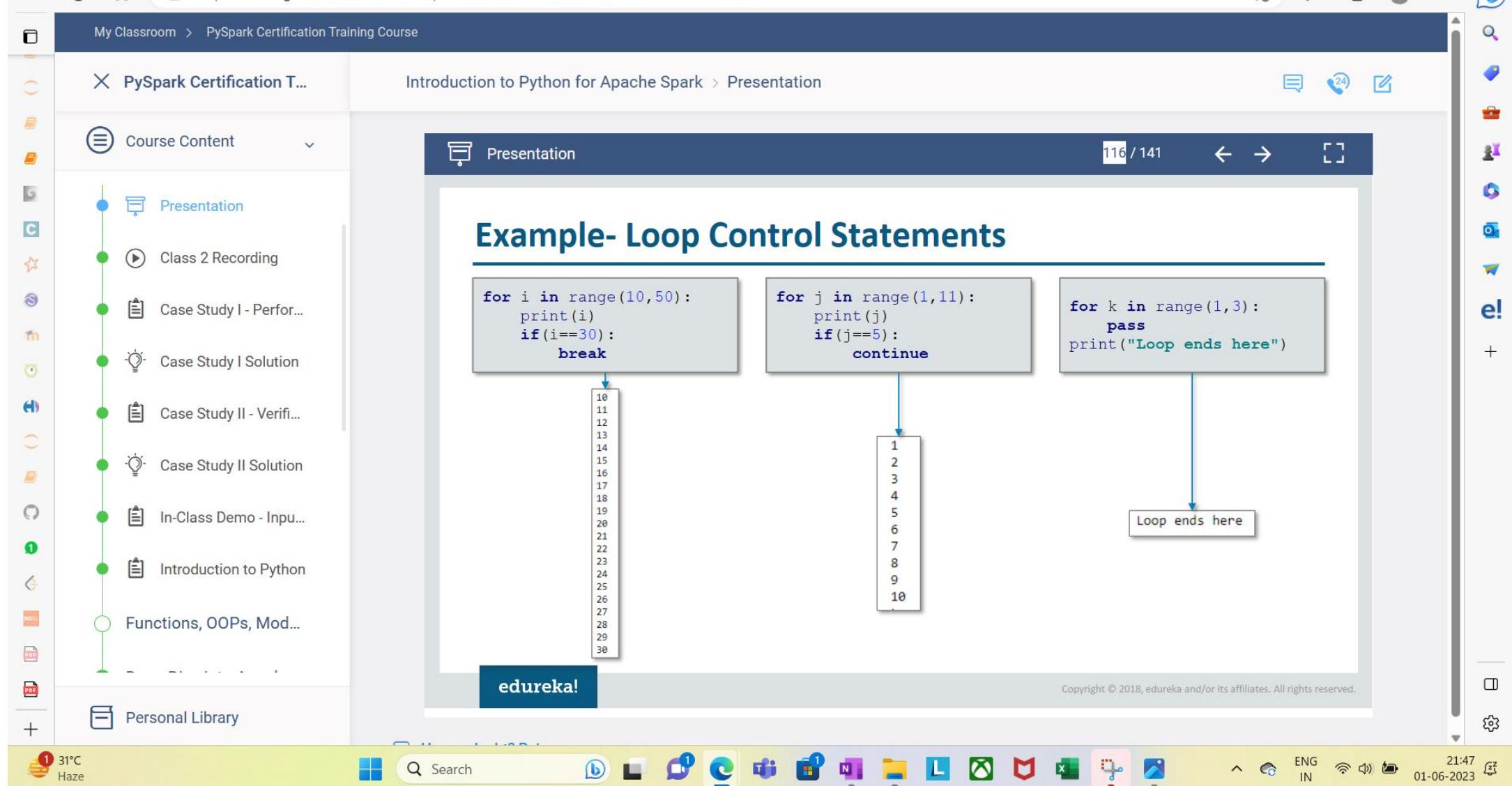
Personal Library

Nested Loops Example



edureka!

Copyright © 2018, edureka and/or its affiliates. All rights reserved.



https://learning.edureka.co/classroom/presentation/651/6517/111071?tab=CourseContent

My Classroom > PySpark Certification Training Course

PySpark Certification T... Introduction to Python for Apache Spark > Presentation

Presentation 117 / 141

Course Content

- Presentation
- Class 2 Recording
- Case Study I - Perform...
- Case Study I Solution
- Case Study II - Verifi...
- Case Study II Solution
- In-Class Demo - Inpu...
- Introduction to Python
- Functions, OOPs, Mod...

edureka!

Dave, I want to give an input to my program. How can I do that in Python?

Well, it's pretty easy. Just use the *input* function

Copyright © 2018, edureka and/or its affiliates. All rights reserved.



My Classroom > PySpark Certification Training Course

PySpark Certification T... Introduction to Python for Apache Spark > Presentation

Presentation 118 / 141 ← →

Reading Keyboard Input

Reading Keyboard Input

Python provides a built-in function to read a line of text from the standard function: `input`

We can take input from user using the `input()`

```
str=input("Enter your input")
print("Received input is: ",str)
```

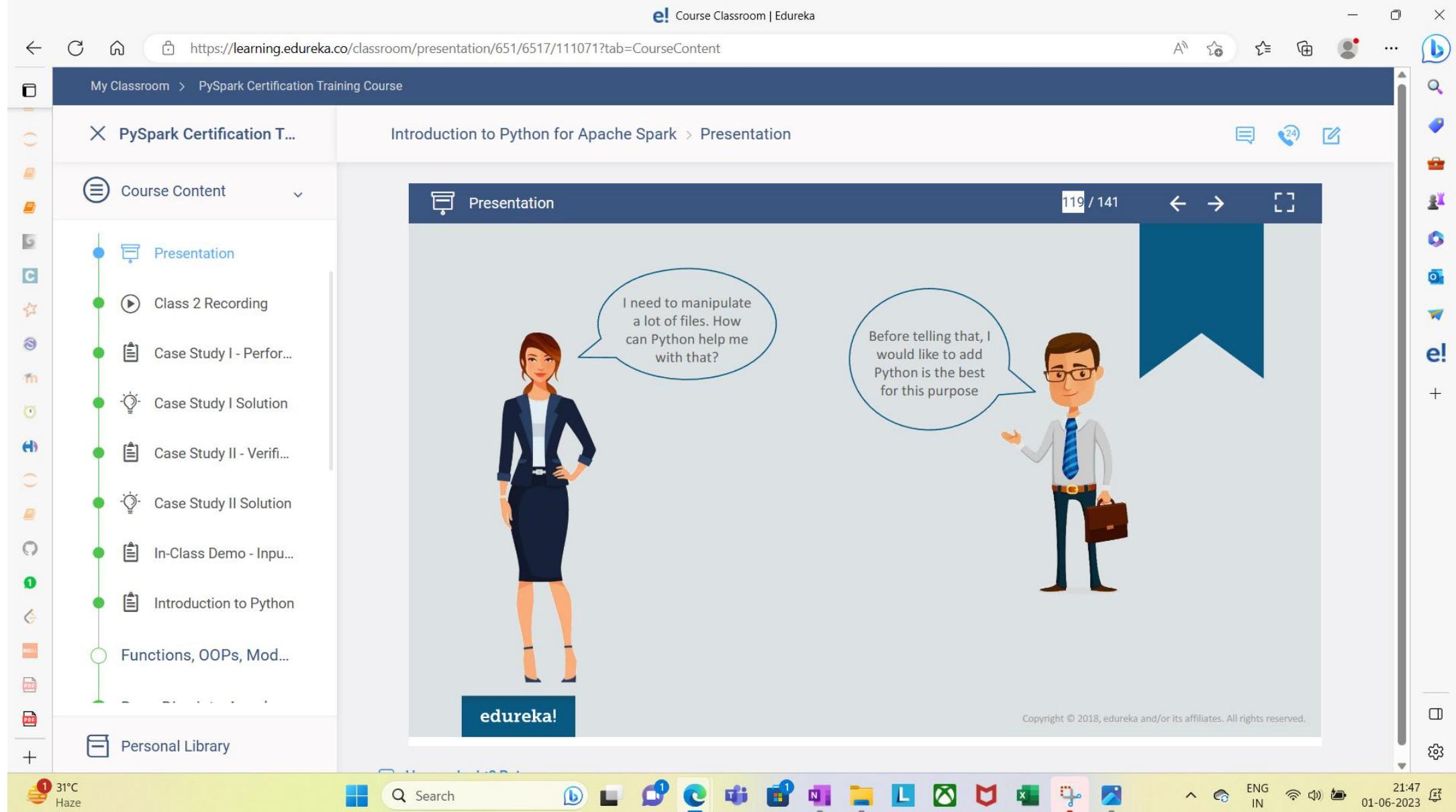
Enter your input String
Received input is: String

We print the input taken from the user

edureka!

Copyright © 2018, edureka and/or its affiliates. All rights reserved.

Personal Library



https://learning.edureka.co/classroom/presentation/651/6517/111071?tab=CourseContent

My Classroom > PySpark Certification Training Course

PySpark Certification T... Course Content

- Presentation
- Class 2 Recording
- Case Study I - Perform...
- Case Study I Solution
- Case Study II - Verifi...
- Case Study II Solution
- In-Class Demo - Input...
- Introduction to Python
- Functions, OOPs, Mod...

Introduction to Python for Apache Spark > Presentation

Presentation 120 / 141

Python Files Input/Output

Opening and Closing Files Writing and Reading Files Renaming Files

edureka!

Copyright © 2018, edureka and/or its affiliates. All rights reserved.

https://learning.edureka.co/classroom/presentation/651/6517/111071?tab=CourseContent

My Classroom > PySpark Certification Training Course

PySpark Certification T... Course Content

- Presentation
- Class 2 Recording
- Case Study I - Perform...
- Case Study I Solution
- Case Study II - Verifi...
- Case Study II Solution
- In-Class Demo - Input...
- Introduction to Python
- Functions, OOPs, Mod...

Introduction to Python for Apache Spark > Presentation

Presentation 121 / 141 ← →

Opening and Closing Files

Before reading and writing any data into a file, it is important to learn how to open and close a file

 Opening Files

Because unless you open a file, you can not write anything in a file or read anything from it

 Closing Files

And once you are done with reading or writing, you should close the file

edureka!

Copyright © 2018, edureka and/or its affiliates. All rights reserved.

https://learning.edureka.co/classroom/presentation/651/6517/111071?tab=CourseContent

My Classroom > PySpark Certification Training Course

PySpark Certification T... Course Content

- Presentation
- Class 2 Recording
- Case Study I - Perform...
- Case Study I Solution
- Case Study II - Verifi...
- Case Study II Solution
- In-Class Demo - Inpu...
- Introduction to Python
- Functions, OOPs, Mod...

Introduction to Python for Apache Spark > Presentation

Presentation 122 / 141

Open Function

- You can open Files using Python's built-in `open()` function

```
file_Object=open(file_name,[access_mode])
```

- Here are parameter details:
 - file_name:** The file_name argument is a string value that contains the name of the file that you want to access
 - access_mode:** The access_mode determines the mode in which the file has to be opened, i.e., read, write, append

edureka!

Copyright © 2018, edureka and/or its affiliates. All rights reserved.



My Classroom > PySpark Certification Training Course

X PySpark Certification T...

Introduction to Python for Apache Spark > Presentation



Course Content

- Presentation
- Class 2 Recording
- Case Study I - Perform...
- Case Study I Solution
- Case Study II - Verifi...
- Case Study II Solution
- In-Class Demo - Input...
- Introduction to Python
- Functions, OOPs, Mod...

Personal Library

Open Function – Access Modes (Contd.)

Modes	Description
a	Opens a file for appending
ab	Opens a file for appending in binary format
a+	Opens a file for both appending and reading
ab+	Opens a file for both appending and reading in binary format
w+	Opens a file for both writing and reading
wb+	Opens a file for both writing and reading in binary format

edureka!

Copyright © 2018, edureka and/or its affiliates. All rights reserved.



1

31°C

Haze



Search

ENG
IN

01-06-2023



21:47



My Classroom > PySpark Certification Training Course

X PySpark Certification T...

Introduction to Python for Apache Spark > Presentation



Course Content

- Presentation
- Class 2 Recording
- Case Study I - Perform...
- Case Study I Solution
- Case Study II - Verifi...
- Case Study II Solution
- In-Class Demo - Input...
- Introduction to Python
- Functions, OOPs, Mod...

Personal Library

Presentation

125 / 141



Writing Files

edureka!

Copyright © 2018, edureka and/or its affiliates. All rights reserved.



1

31°C

Haze



Search



ENG
IN



21:47
01-06-2023

https://learning.edureka.co/classroom/presentation/651/6517/111071?tab=CourseContent

My Classroom > PySpark Certification Training Course

PySpark Certification T... Course Content

- Presentation
- Class 2 Recording
- Case Study I - Perform...
- Case Study I Solution
- Case Study II - Verifi...
- Case Study II Solution
- In-Class Demo - Input...
- Introduction to Python
- Functions, OOPs, Mod...

Introduction to Python for Apache Spark > Presentation

Presentation 126 / 141

Writing Files

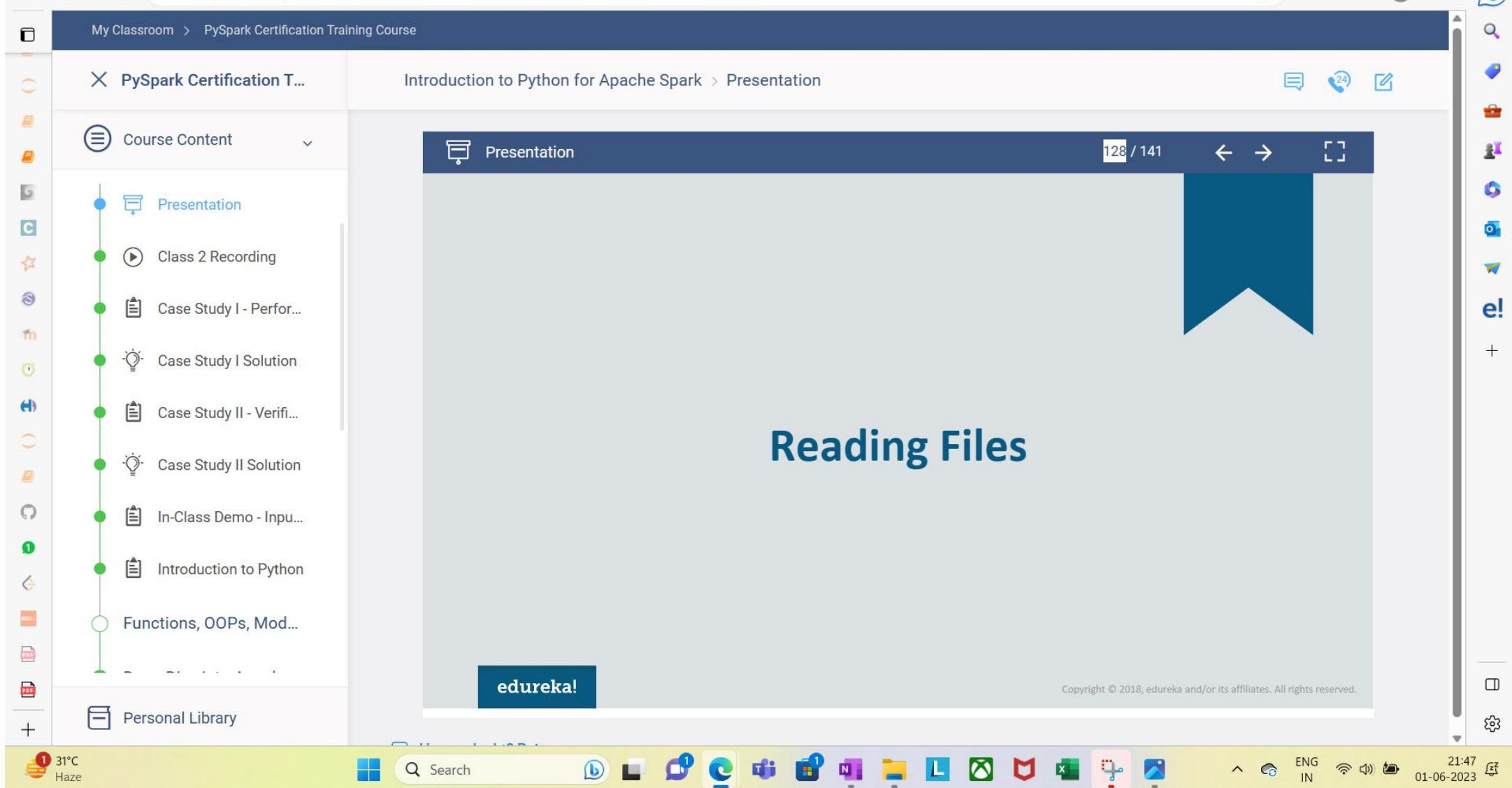
fileObject.write(string)

The `write()` method does not add a newline character '`\n`' to the end of the string



edureka!

Copyright © 2018, edureka and/or its affiliates. All rights reserved.



https://learning.edureka.co/classroom/presentation/651/6517/111071?tab=CourseContent

My Classroom > PySpark Certification Training Course

PySpark Certification T... Introduction to Python for Apache Spark > Presentation

Course Content

- Presentation
- Class 2 Recording
- Case Study I - Perform...
- Case Study I Solution
- Case Study II - Verifi...
- Case Study II Solution
- In-Class Demo - Input...
- Introduction to Python
- Functions, OOPs, Mod...

Personal Library

Presentation

129 / 141

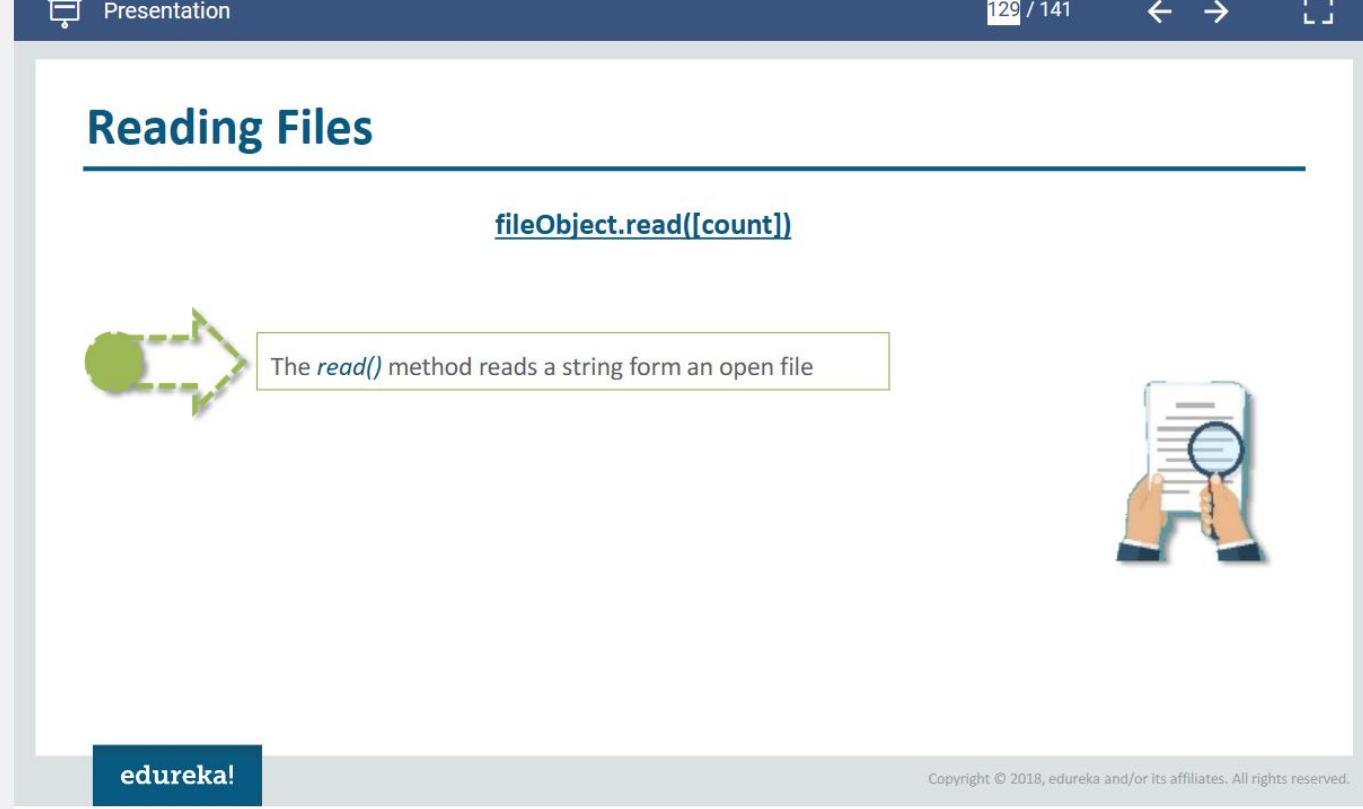
Reading Files

fileObject.read([count])

The `read()` method reads a string form an open file

edureka!

Copyright © 2018, edureka and/or its affiliates. All rights reserved.



My Classroom > PySpark Certification Training Course

PySpark Certification T... Introduction to Python for Apache Spark > Presentation

Course Content

- Presentation
- Class 2 Recording
- Case Study I - Perform...
- Case Study I Solution
- Case Study II - Verifi...
- Case Study II Solution
- In-Class Demo - Input...
- Introduction to Python
- Functions, OOPs, Mod...

Personal Library

Presentation

130 / 141

Reading Files

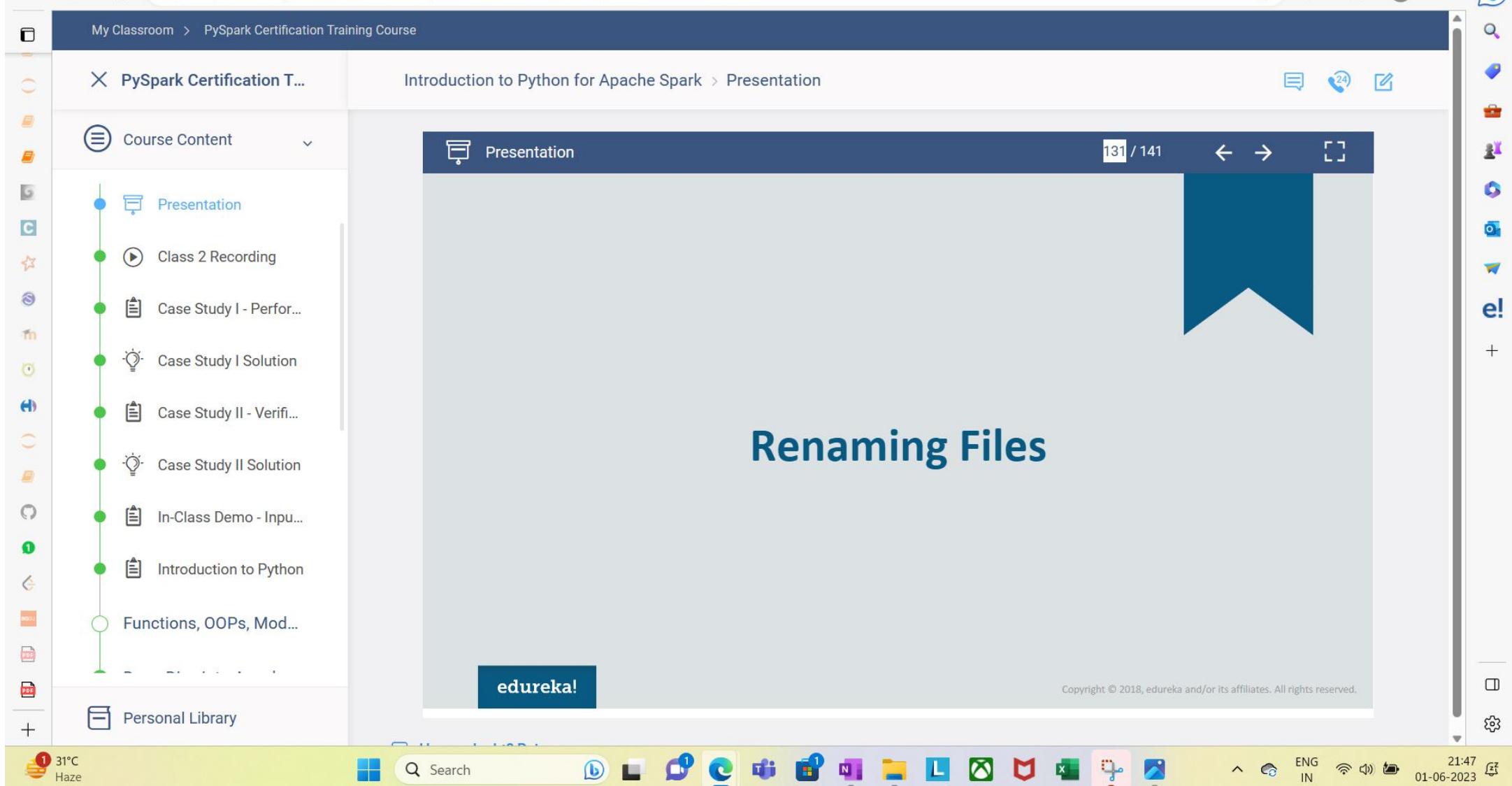
fileObject.read([count])

The `read()` method reads a string form an open file

It is important to note that Python strings can have binary data apart from text data

edureka!

Copyright © 2018, edureka and/or its affiliates. All rights reserved.



https://learning.edureka.co/classroom/presentation/651/6517/111071?tab=CourseContent

My Classroom > PySpark Certification Training Course

PySpark Certification T... Course Content

- Presentation
- Class 2 Recording
- Case Study I - Perform...
- Case Study I Solution
- Case Study II - Verifi...
- Case Study II Solution
- In-Class Demo - Inpu...
- Introduction to Python
- Functions, OOPs, Mod...

Introduction to Python for Apache Spark > Presentation

Presentation 132 / 141

Renaming Files

```
os.rename(current_file_name, new_file_name)
```

The `rename()` method takes two arguments, the current filename and the new filename

`rename()` is the method from `os` module, we going to learn it in module 5



edureka!

Copyright © 2018, edureka and/or its affiliates. All rights reserved.

e! Course Classroom | Edureka

https://learning.edureka.co/classroom/presentation/651/6517/111071?tab=CourseContent

My Classroom > PySpark Certification Training Course

X PySpark Certification T...

Introduction to Python for Apache Spark > Presentation

Course Content

- Presentation
- Class 2 Recording
- Case Study I - Perform...
- Case Study I Solution
- Case Study II - Verifi...
- Case Study II Solution
- In-Class Demo - Inpu...
- Introduction to Python
- Functions, OOPs, Mod...

Presentation

133 / 141

Deleting Files

edureka!

Copyright © 2018, edureka and/or its affiliates. All rights reserved.

Personal Library

31°C Haze

Search

1 24

ENG IN

01-06-2023

My Classroom > PySpark Certification Training Course

PySpark Certification T... Introduction to Python for Apache Spark > Presentation

Course Content

- Presentation
- Class 2 Recording
- Case Study I - Perform...
- Case Study I Solution
- Case Study II - Verifi...
- Case Study II Solution
- In-Class Demo - Input...
- Introduction to Python
- Functions, OOPs, Mod...

Deleting Files

os.remove(file_name)

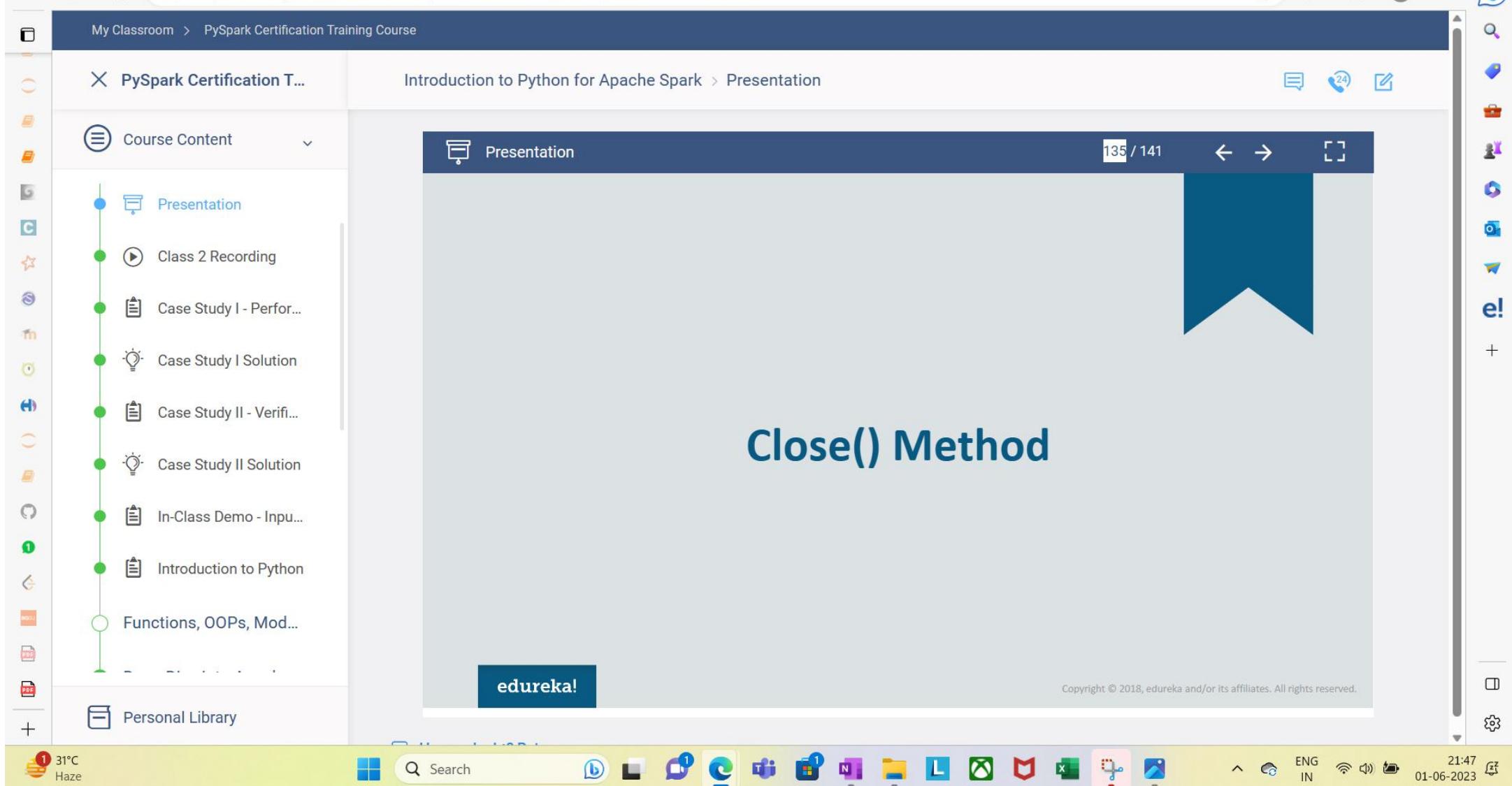
You can use the `remove()` method to delete files by supplying the name of the file to be deleted as an argument

`remove()` is the method from `os` module, we going to learn it in module 5

edureka!

Copyright © 2018, edureka and/or its affiliates. All rights reserved.





My Classroom > PySpark Certification Training Course

PySpark Certification T...

Course Content

- Presentation
- Class 2 Recording
- Case Study I - Perform...
- Case Study I Solution
- Case Study II - Verifi...
- Case Study II Solution
- In-Class Demo - Input...
- Introduction to Python
- Functions, OOPs, Mod...

Introduction to Python for Apache Spark > Presentation

Presentation

137 / 141

Summary

Different Programming Languages

edureka!

Standard Data Types

edureka!

Operators

edureka!

Conditional Statements

edureka!

Loops

edureka!

Python Files Input/Output

edureka!

edureka!

Copyright © 2018, edureka and/or its affiliates. All rights reserved.