6] Big Data Fundamentals with PySpark (M2 Programming in PySpark RDD’s )

Pair RDDs in PySpark (M2 -> SM3)

1) Download RAR on STAGING\_AREA

2) Extract the RAR

3) Put the \*.ipynb files in UBUNTU\_HOME/test-jupyter/P2/M2/sm3

4) Import the notes in Jupyter Notebook

5) Follow the instructions and give the solutions

* Code/Dataset

day3/3\_PairRDDsinPySpark.zip

TimeLine = 20 Minutes

7] Big Data Fundamentals with PySpark (M2 Programming in PySpark RDD’s )

Advanced RDD Actions (M2 -> SM4)

1) Download RAR on STAGING\_AREA

2) Extract the RAR

3) Put the \*.ipynb files in UBUNTU\_HOME/test-jupyter/P2/M2/sm4

4) Import the notes in Jupyter Notebook

5) Follow the instructions and give the solutions

* Code/Dataset

day3/4\_AdvancedRddActions.zip

TimeLine = 30 Minutes

8] PySpark SQL & DataFrames

(M3->sm1)

1) Download RAR on STAGING\_AREA

2) Extract the RAR

3) Put the \*.ipynb files in UBUNTU\_HOME/test-jupyter/P2/M3/sm1

4) Import the notes in Jupyter Notebook

5) Follow the instructions and give the solutions

* Code/Dataset

data/1\_AbstractingDatawithDataFrames.zip

TimeLine = 15 Minutes

9] PySpark Sql and DataFrame

(M3->sm2)

1) Download RAR on STAGING\_AREA

2) Extract the RAR

3) Put the \*.ipynb files in UBUNTU\_HOME/test-jupyter/P2/M3/sm2

4) Import the notes in Jupyter Notebook

5) Follow the instructions and give the solutions

* Code/Dataset

data/2\_OperatingonDataFramesinPySpark.zip

TimeLine = 25 Minutes