**HOMEWORK ASSIGNMENT: MODULE 1**

1. Install Scala on windows or Linux (Ubuntu VM)

2. Set up the Eclipse IDE for Scala

3. Refresh Java OOPs Concepts

Answer:

I have already installed IntelliJ in my system as part of our project requirement. Hence I would prefer to go with IntelliJ rather than installing Eclipse on my system.

**HOMEWORK ASSIGNMENT: MODULE 2**

1. How to sorts the elements in an array?

<https://github.com/snkrishnan1/SparkScalaAssignments/blob/master/Assignments/src/SortArray.scala>

Output:

<https://github.com/snkrishnan1/SparkScalaAssignments/blob/master/OtherDocuments/Assignment2/SortArray1.png>

<https://github.com/snkrishnan1/SparkScalaAssignments/blob/master/OtherDocuments/Assignment2/SortArray2.png>

1. Find the particular word count in the sentence.

<https://github.com/snkrishnan1/SparkScalaAssignments/blob/master/Assignments/src/WordCountInSentence.scala>

Output:

<https://github.com/snkrishnan1/SparkScalaAssignments/blob/master/OtherDocuments/Assignment2/WordCountInSentence.png>

1. Write the Scala program for finding the GREATEST COMMON DIVISOR (GCD) for two numbers using while loop?

<https://github.com/snkrishnan1/SparkScalaAssignments/blob/master/Assignments/src/GCD.scala>

Output:

<https://github.com/snkrishnan1/SparkScalaAssignments/blob/master/OtherDocuments/Assignment2/GCD.png>

1. Write the Scala function to get odd values and even values in a List?

<https://github.com/snkrishnan1/SparkScalaAssignments/blob/master/Assignments/src/OddEvenNumbers.scala>

Output:

<https://github.com/snkrishnan1/SparkScalaAssignments/blob/master/OtherDocuments/Assignment2/Odd-Even-Numbers.png>

5. Write the code for merging Array with ArrayBuffer?

<https://github.com/snkrishnan1/SparkScalaAssignments/blob/master/Assignments/src/MergeArrayWithArrayBuffer.scala>

Output:

<https://github.com/snkrishnan1/SparkScalaAssignments/blob/master/OtherDocuments/Assignment2/MergeArrayWithArrayBuffer.png>

**HOMEWORK ASSIGNMENT: MODULE 3**

Complete the following tasks:

1. Create a Map

 Remove existing key from the Map

 Add the new key and value to the existing Map

<https://github.com/snkrishnan1/SparkScalaAssignments/blob/master/Assignments/src/MapAddRemoveKeyValues.scala>

Output:

<https://github.com/snkrishnan1/SparkScalaAssignments/blob/master/OtherDocuments/Assignment3/Map%20Add%20Remove%20Key.png>

<https://github.com/snkrishnan1/SparkScalaAssignments/blob/master/OtherDocuments/Assignment3/Map-Add-Reduce.png>

2. Create an Employee class with method increment and currentSal and use a read-only property salary.

<https://github.com/snkrishnan1/SparkScalaAssignments/blob/master/Assignments/src/EmployeeSalary.scala>

Output:

3. Write a Scala program to add all the elements in the Array by using reduceLeft function.

<https://github.com/snkrishnan1/SparkScalaAssignments/blob/master/Assignments/src/ReduceLeftToAddAllElements.scala>

Output:

<https://github.com/snkrishnan1/SparkScalaAssignments/blob/master/OtherDocuments/Assignment3/ReduceLeftToAddAllElements.png>

<https://github.com/snkrishnan1/SparkScalaAssignments/blob/master/OtherDocuments/Assignment3/reduceLeft%20function.png>

4. Please do a sample implementation of Multiple Inheritance in Scala.

<https://github.com/snkrishnan1/SparkScalaAssignments/blob/master/Assignments/src/MultipleInheritance.scala>

Output:

<https://github.com/snkrishnan1/SparkScalaAssignments/blob/master/OtherDocuments/Assignment3/MultipleInheritance.png>

**HOMEWORK ASSIGNMENT: MODULE 4**

Complete the following tasks:

1. Install the Spark in Ubuntu VM

2. Refresh all the concepts of Scala

3. Refresh SBT Concepts

4. Write the spark program to count the error messages in log file

<https://github.com/snkrishnan1/SparkScalaAssignments/blob/master/Assignments/src/ErrorCount.scala>

Output:

<https://github.com/snkrishnan1/SparkScalaAssignments/blob/master/OtherDocuments/Assignment4/ErrorCount.png>

**HOMEWORK ASSIGNMENT: MODULE 5**

Complete the following tasks:

1. Write a spark code to count the words in a file.

<https://github.com/snkrishnan1/SparkScalaAssignments/blob/master/Assignments/src/WordCountInFile.scala>

Output:

<https://github.com/snkrishnan1/SparkScalaAssignments/blob/master/OtherDocuments/Assignment5/WordCountInFile.png>

2. Write a spark program which reads the files from HDFS, calculate the Maximum salary (Create a file in local system according to the requirement and save it in HDFS).

<https://github.com/snkrishnan1/SparkScalaAssignments/blob/master/Assignments/src/EmployeeSalary.scala>

Output:

<https://github.com/snkrishnan1/SparkScalaAssignments/blob/master/OtherDocuments/Assignment5/CopyFileToHDFSFromLocal.png>

<https://github.com/snkrishnan1/SparkScalaAssignments/blob/master/OtherDocuments/Assignment5/CalculateMaxSalary.png>

**HOMEWORK ASSIGNMENT: MODULE 6**

Complete the following tasks:

1. Create RDDs to filter each line for the keyword “Spark”

<https://github.com/snkrishnan1/SparkScalaAssignments/blob/master/Assignments/src/FilterEachLineForSparkWord.scala>

Output:

<https://github.com/snkrishnan1/SparkScalaAssignments/blob/master/OtherDocuments/Assignment6/FilterEachIineForSparkKeyword.png>

2. Perform a WordCount of particular word in a sentence (pg1661.txt), i.e., so the results are (K, V) pairs of (word, count)

<https://github.com/snkrishnan1/SparkScalaAssignments/blob/master/Assignments/src/WordCountKeyPair.scala>

Output:

<https://github.com/snkrishnan1/SparkScalaAssignments/blob/master/OtherDocuments/Assignment6/WordCountKeyPair.png>

3. How to join the two RDDs

<https://github.com/snkrishnan1/SparkScalaAssignments/blob/master/Assignments/src/Join2RDD.scala>

Output:

<https://github.com/snkrishnan1/SparkScalaAssignments/blob/master/OtherDocuments/Assignment6/Join2RDD.png>

**HOMEWORK ASSIGNMENT: MODULE 7**

Complete the following tasks:

Use the spark-shell to run the following spark shell commands

Dataset:

The column names are given below

id,actorname,genre,noofmovies,hits

1001,Johnny Lever,Commedian,210,190 1002,Amit Mahesh,New Comer,2,1 1003,Salman Khan,SuperStar,300,291 1004,Johnny Depp,SuperStar,289,270 1005,Mallika Sherawat,Actress,20,10 1006,Amitabh Bachan,SuperStar,350,300 1007,Micheal Bijlu,New Comer,4,1 1008,Rocky Angelo,New Comer,3,1 1009,Rajani Kanth,SuperStar,400,399 1010,Arnold,SuperStar,261,242

1. Please construct the dataset as given below using Seq and case class ActorDetail.

2. Construct a data frame using the above dataset by creating 2 partitions (use parallelize method).

3. Show the all the details of the actDetailsDF and also show only the first two records of the actorDetails dataset.

4. How do you print the schema of the actor details dataframe?

5. Register the DataFrame as temp table and only select the rows for whom the no of hits are more than 250 and print the output.

<https://github.com/snkrishnan1/SparkScalaAssignments/blob/master/Assignments/src/ActorDetails.scala>

Output:

<https://github.com/snkrishnan1/SparkScalaAssignments/blob/master/OtherDocuments/Assignment7/ActorsDetails%20-%20Execution.png>

**HOMEWORK ASSIGNMENT: MODULE 8**