Assignment No 4 (Group B)

Problem Statement:- Write a simple program in SCALA using Apache Spark framework

Introduction:-

- SCALA:- Scala is a Java-based Hybrid programming language. It combines the features of functional-oriented and object-oriented programming language. It is used by integration with the Java Virtual machine and can compile the written code.
- Apache Spark is a lightning-fast cluster computing designed for fast computation. It was built on top of Hadoop MapReduce and it extends the MapReduce model to efficiently use more types of computations which includes Interactive Queries and Stream Processing

Features of Apache Spark

Apache Spark has following features.

- **Speed** − Spark helps to run an application in Hadoop cluster, up to 100 times faster in memory, and 10 times faster when running on disk. This is possible by reducing number of read/write operations to disk. It stores the intermediate processing data in memory.
- **Supports multiple languages** − Spark provides built-in APIs in Java, Scala, or Python. Therefore, you can write applications in different languages. Spark comes up with 80 high-level operators for interactive querying.
- **Advanced Analytics** − Spark not only supports ‘Map’ and ‘reduce’. It also supports SQL queries, Streaming data, Machine learning (ML), and Graph algorithms.

Installation of Scala:-

It is necessary to install Java before installing Scala

Step 1: Verifying Java Installation:-

```
$java -version
```

Step 2: Once Java is installed, we need to install Scala

```
$scala –version
```

Step 3: Downloading Scala

https://spark.apache.org/downloads.html

Step 4: Installing Scala

Extract the Scala tar file

Type the following command for extracting the Scala tar file.

```
$ tar xvf scala-2.11.6.tgz
```

Move Scala software files(c drive)

Set PATH for Scala

(Control panel-env variable-new-SPARK_HOME,Value-c:\spark

To set path-%SPARK_HOME%BIN)

Step 5: Verifying Scala Installation

```
$spark-shell
```

Things to note about Scala

- It is case sensitive
- If you are writing a program in Scala, you should save this program using ".scala"
- Scala execution starts from main() methods
- Any identifier name cannot begin with numbers. For example, variable name "123salary" is invalid.
- You can not use Scala reserved keywords for variable declarations or constant or any identifiers.

Variable declaration in Scala:-

In Scala, you can declare a variable using 'var' or 'val' keyword. The decision is based on whether it is a constant or a variable. If you use 'var' keyword, you define a variable as mutable variable. On the other hand, if you use 'val', you define it as immutable. Let's first declare a variable using "var" and then using "val".

1) Declare using var:-

```
var Var1 : String = "Ankit"
```

In the above Scala statement, you declare a mutable variable called "Var1" which takes a string value. You can also write the above statement without specifying the type of variable.

2) Declare using val:-

```
val Var2 : String = "Ankit"
```

In the above Scala statement, we have declared an immutable variable "Var2" which takes a string "Ankit". Try it for without specifying the type of variable.

Operations on variables:-

You can perform various operations on variables. There are various kinds of operators defined in Scala. For example: Arithmetic Operators, Relational Operators, Logical Operators, Bitwise Operators, Assignment Operators.

```
scala> var Var4 = 2
Output: Var4: Int = 2
scala> var Var5 = 3
Output: Var5: Int = 3
```

Apply '+' operator

```
Var4+Var5
Output:
res1: Int = 5
```

Apply "==" operator

```
Var4==Var5
Output:
res2: Boolean = false
```

The if-else expression in Scala:-

In Scala, if-else expression is used for conditional statements. You can write one or more conditions inside "if". Let's declare a variable called "Var3" with a value 1 and then compare "Var3" using if-else expression.

```
var Var3 =1
if (Var3 ==1){
 println("True")}else{
 println("False")}
Output: True
```

Iteration in Scala:-

Scala also has a FOR-loop which is the most widely used method for iteration. It has a simple syntax too.

```
for( a <- 1 to 10){
 println( "Value of a: " + a );
 }
Output:
Value of a: 1
Value of a: 2
Value of a: 3
Value of a: 4
Value of a: 5
Value of a: 6 till 10
```

Declare a simple function in Scala and call it by passing value:-

You can define a function in Scala using "def" keyword. Let's define a function called "mul2" which will take a number and multiply it by 10. You need to define the return type of function, if a function not returning any value you should use the "Unit" keyword.

In the below example, the function returns an integer value. Let's define the function "mul2":

```
def mul2(m: Int): Int = m * 10
Output: mul2: (m: Int)Int
Now let's pass a value 2 into mul2
mul2(2)
Output:
res9: Int = 20
```

Writing & Running a program in Scala using an editor:-

Let us start with a "Hello World!" program. It is a good simple way to understand how to write, compile and run codes in Scala.

```
object HelloWorld {
 def main(args: Array[String]) {
 println("Hello, world!")
 }
 }
```

Compile a Scala Program:-

To run any Scala program, you first need to compile it. “Scalac” is the compiler which takes source program as an argument and generates object files as output.

Let’s start compiling your “HelloWorld” program using the following steps:

1. For compiling it, you first need to paste this program into a text file then you need to save this program as HelloWorld.scala
2. Now you need change your working directory to the directory where your program is saved
3. After changing the directory you can compile the program by issuing the command.

```
scalac HelloWorld.scala
```

4. After compiling, you will get Helloworld.class as an output in the same directory. If you can see the file, you have successfully compiled the above program.

Running Scala Program:-

```
command::- scala HelloWorld
```

Advantages of using Scala for Apache Spark:-

- Working with Scala is more productive than working with Java
- Scala is faster than Python and R because it is compiled language
- Scala is a functional language

Comparing Scala, Java, Python and R APIs in Apache Spark:-

Metrics	Scala	Java	Python	R
Type	Compiled	Compiled	Interpreted	Interpreted
JVM based	Yes	Yes	No	No
Verbosity	Less	More	Less	Less
Code Length	Less	More	Less	Less
Productivity	High	Less	High	High
Scalability	High	High	Less	Less
OOPS Support	Yes	Yes	Yes	Yes

Question:-

1) Mention how Scala is different from Java
2) What is the difference between var and value in scala?
3) What are the advantages of Scala
4) Mention the different types of Scala literals
5) What is the difference between Scala and Java
6) Define types of Scala Identifiers