Spark core and Spark Collections are Mandatory for Big data Engineer

Spark Transformation and Action

Dataframe

Diff Bt Mapreduce and Spark?

**Ans:** MR is I/O driven

Spark is Memory driven

Apache spark is inmemory distributed processing framework on top HDFS ,s3

there are bunch of APIs to process the data they are called as Transformation and action .they are also known as core spark.

Tightly integrated with programming languages such as scala python,java ,r

to be proficient in spark you need to learn one of the programming languages preferably scala

you often hear about yarn,mesos they are just frameworks to run the jobs.

Developers do not need to worry about it.

**REPL** – Read,Evaluate,Print and Evaluate

Scala.press tab to navigate the package in REPL

type help we will get a list of commands.

:Quit we will come out of repl

**SBT**

Simple Build Tool

it will resolve the dependencies of our application

it will help to build the application in jar file

Scala is a JVM based object oriented and functional programming language

we have to use either val or var before variable name

val means immutable

var means mutable

if I want to override the val variable it will throw error

we have to specify “” for string in scala

**Ex:** val a =100

a.toString //similar byte code will genrate internally

all primitive classes are called as value classes.

All java packages can be accessed in scala .

Play is a framework in scala to develop web application

**Basic programming context:**

val a=0 //implicitly create a class through command line

pritnln(“value of a is “ + a)

Commands are start with colons

**Ex:**

:help

for multi line coding in command line

**Ex:**

val a=0;println(“value of a is” + a)

One of the feature in scala is we can assign block of code to a variable

**Ex:**

val c={

val a=(math.random \* 100).toInt

val b=(math.random \*100).toInt

a=b

}

**Ex:2**

val c={

val a=(math.random \* 100).toInt

val b=(math.random \*100).toInt

a=b

“HelloWorld”

**}**

**Ex:3**

val c={

val a=(math.random \* 100).toInt

val b=(math.random \*100).toInt

**}**

This above code will return Unit()

Unit is a kind of void in java it will return nothing.