**Session 21**

**Assignment 1 Question**

*Sample file:*

*………………………….*

**Hello World**

**It’s a sunny day**

**<blank-line>**

**When will it rain?**

**<blank-line>**

**Will it rain today!**

**Code:**

**package** com.spark.core.prctice

**import** org.apache.spark.SparkConf

**import** org.apache.spark.SparkContext

**object** AdvancedSparkAssignment1 **extends** App {

**var** conf=**new** SparkConf().setMaster("local").setAppName("AdvanceSparkAssignment1")

**var** sc=**new** SparkContext(conf)

**var** blank\_lines = sc.accumulator(0, "count\_blank\_lines")

**val** Rdd1=sc.textFile("C:\\SparkInputs\\accumlators.txt", 3).foreach{line=>**if** (line.length()==0) blank\_lines+=1}

println("########## START ##########")

println(":::::::::Total Blank Lines::::::::")

println(s" Blank Line =${blank\_lines.value}");

}

Result:

7/07/28 16:07:24 INFO Executor: Finished task 2.0 in stage 0.0 (TID 2). 2044 bytes result sent to driver

17/07/28 16:07:24 INFO TaskSetManager: Finished task 2.0 in stage 0.0 (TID 2) in 23 ms on localhost (3/3)

17/07/28 16:07:24 INFO DAGScheduler: ResultStage 0 (foreach at AdvancedSparkAssignment1.scala:11) finished in 0.118 s

17/07/28 16:07:24 INFO TaskSchedulerImpl: Removed TaskSet 0.0, whose tasks have all completed, from pool

17/07/28 16:07:24 INFO DAGScheduler: Job 0 finished: foreach at AdvancedSparkAssignment1.scala:11, took 0.210805 s

########## START ##########

:::::::::Total Blank Lines::::::::

Blank Line =2

17/07/28 16:07:24 INFO SparkContext: Invoking stop() from shutdown hook

17/07/28 16:07:24 INFO SparkUI: Stopped Spark web UI at http://19.236.244.139:4040

17/07/28 16:07:24 INFO MapOutputTrackerMasterEndpoint: MapOutputTrackerMasterEndpoint stopped!

17/07/28 16:07:24 INFO MemoryStore: MemoryStore cleared

17/07/28 16:07:24 INFO BlockManager: BlockManager stopped

17/07/28 16:07:24 INFO BlockManagerMaster: BlockManagerMaster stopped

17/07/28 16:07:24 INFO OutputCommitCoordinator$OutputCommitCoordinatorEndpoint: OutputCommitCoordinator stopped!

17/07/28 16:07:24 INFO SparkContext: Successfully stopped SparkContext

17/07/28 16:07:24 INFO ShutdownHookManager: Shutdown hook called

17/07/28 16:07:24 INFO ShutdownHookManager: Deleting directory C:\Users\ngajapat\AppData\Local\Temp\spark-a432e142-c694-4d49-8452-6783e451b91e

17/07/28 16:07:24 INFO RemoteActorRefProvider$RemotingTerminator: Shutting down remote daemon.

