

databricks Assignment_45.3 (Scala) Import Notebook

```
import org.apache.spark.{SparkConf, SparkContext}

object DataFilter {

  def main(args: Array[String]): Unit = {

    // create Spark context with Spark configuration
    val sc = new SparkContext(new SparkConf().setAppName("AV_Log_Processing").setMaster("local[*]"))

    // Read text file in spark RDD
    val textFile = sc.textFile("input.txt");

    val splitRdd = textFile.map { s =>
      val a = s.split("[ ]")
      val date = Array(a(0) + " " + a(1))
      (date ++ a.takeRight(10)).mkString("\t")
    }
    // RDD[ Array[ String ] ]

    // printing values
    splitRdd.foreach(println)

    // how to store split values in different column and write it into file
  }

}

import org.apache.spark.{SparkConf, SparkContext}
defined object DataFilter

import org.apache.spark.{SparkConf, SparkContext}
val textFile = sc.textFile("input1.txt");
val lineLengths = textFile.map(s => s.length)
val totalLength = lineLengths.reduce((a, b) => a + b)

@org.apache.hadoop.mapred.InvalidInputException: Input path does not exist: /input1.txt
```

Assignment_45.3.html Assignment_45 (1).html Assignment_45.html Show all

11:46 PM 1/2/2019