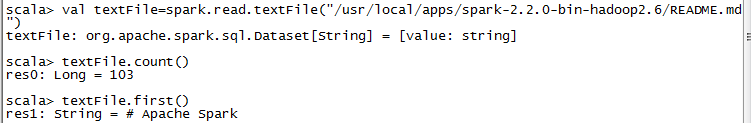
## Quick Start



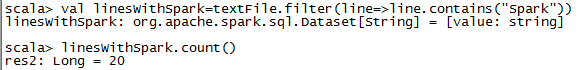












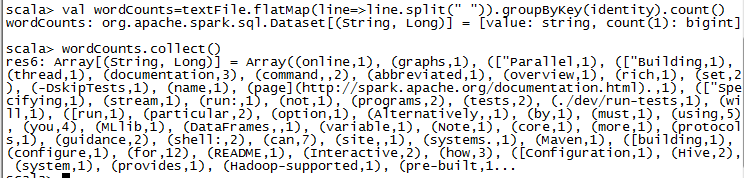






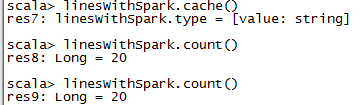








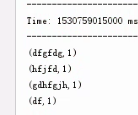




## Spark Streaming

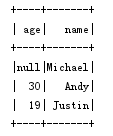
**package** scdx.spark  
**import** org.apache.spark.SparkConf  
**import** org.apache.spark.streaming.{Seconds, StreamingContext}  
  
**object** StreamingDemo {  
 **def** main(args: Array[String]): Unit = {  
 **val** conf = **new** SparkConf().setAppName(**"StreamingDemo01"**).setMaster(**"local[4]"**)  
 **val** streamingContext=**new** StreamingContext(conf,*Seconds*(3))  
 **val** inputStream=streamingContext.socketTextStream(**"192.168.253.101"**,8888)  
 **val** outputStream=inputStream.flatMap(\_.split(**" "**)).map((\_,1)).reduceByKey(\_+\_)  
  
 outputStream.print()  
 streamingContext.start()  
 streamingContext.awaitTermination()  
  
  
 }





## Spark SQL

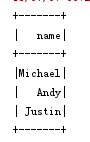
**val** *spark* = SparkSession  
 .*builder*()  
 .master(**"local"**)  
 .appName(**"Spark SQL basic example"**)  
 .config(**"spark.some.config.option"**, **"5"**)  
 .getOrCreate()  
  
**val** *df* = *spark*.read.json(**"D:\\scdx\\spark.bigdata.mr\\src\\main\\resources\\people.json"**)  
*df*.show()



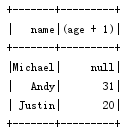
*df*.printSchema()



*df*.select(**"name"**).show()



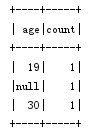
*df*.select(**$"name"**, **$"age"**+1 ).show()



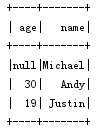
*df*.filter(**$"age"** > 21).show()



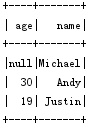
*df*.groupBy(**"age"**).count().show()



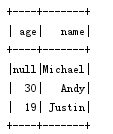
*df*.createOrReplaceTempView(**"people"**)  
**val** *sqlDF* = *spark*.sql(**"SELECT \* FROM people"**)  
*sqlDF*.show()



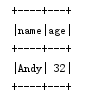
*df*.createGlobalTempView(**"people"**)  
*spark*.sql(**"SELECT \* FROM global\_temp.people"**).show()



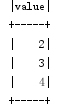
*spark*.newSession().sql(**"SELECT \* FROM global\_temp.people"**).show()



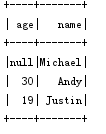
**case class** Person(name: String, age: Long)  
**val** *caseClassDS* = *Seq*(*Person*(**"Andy"**, 32)).toDS()  
*caseClassDS*.show()



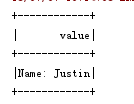
**val** *primitiveDS* = *Seq*(1, 2, 3).toDS()  
*primitiveDS*.map(\_ + 1).collect()  
*primitiveDS*.map(\_ + 1).show()



**val** *path* = **"D:\\scdx\\spark.bigdata.mr\\src\\main\\resources\\people.json"  
val** *peopleDS* = *spark*.read.json(*path*).as[Person]  
*peopleDS*.show()



**val** *peopleDF* = *spark*.sparkContext  
 .textFile(**"D:\\scdx\\spark.bigdata.mr\\src\\main\\resources\\people.txt"**)  
 .map(\_.split(**","**))  
 .map(attributes => *Person*(attributes(0), attributes(1).trim.toInt))  
 .toDF()  
  
*peopleDF*.createOrReplaceTempView(**"people"**)  
**val** *teenagersDF* = *spark*.sql(**"SELECT name, age FROM people WHERE age BETWEEN 13 AND 19"**)  
*teenagersDF*.map(teenager => **"Name: "** + teenager(0)).show()



*teenagersDF*.map(teenager => **"Name: "** + teenager.getAs[String](**"name"**)).show()

