

Spark Streaming Case Study

First Part - You have to create a Spark Application which streams data from a file on local directory on your machine and does the word count on the fly. The word should be done by the spark application in such a way that as soon as you drop the file in your local directory, your spark application should immediately do the word count for you.

Solution to First Part

The source code for the first part is given below. We use `textFileStream` method of the spark context to monitor a directory for the presence of a file. Once the file is available, the application does the wordcount on it.

```
package Stream

import org.apache.spark.{SparkConf, SparkContext}
import org.apache.spark.streaming.{Seconds, StreamingContext}

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

    if(args.length < 1)
    {
      println("Usage: StreamingCaseStudy <local_dir_to_monitor>")
      return
    }

    Task1(args.head)

  }

  def Task1(dirname: String): Unit = {

    val conf = new
SparkConf().setMaster("local[2]").setAppName("SparkSteamingExample")
    val sc = new SparkContext(conf)
    sc.setLogLevel("WARN")
    val ssc = new StreamingContext(sc, Seconds(3))

    val lines = ssc.textFileStream("file://" + dirname)

    val words = lines.flatMap(_.split(" "))
    val wordCounts = words.map(x => (x, 1)).reduceByKey(_ + _)

    wordCounts.print()
    println("Monitoring directory " + dirname + " for file to do wordcount")

    ssc.start()
    ssc.awaitTermination()

  }
}
```

The program output is given below. Following are the steps.

1. Create a new directory called `streamdata` where we will create a new file on which application would run the wordcount
2. We create jar of the application and launch it using `spark-submit` utility on the VM. We also provide the name of the directory to monitor as part of the command line argument
3. We create a file `Test.txt` under `streamdata` directory. We see that Spark Application prints the wordcount of the file

Create new directory

```
[acadgild.mmisra ~]$ mkdir streamdata
[acadgild.mmisra ~]$ cd streamdata/
[acadgild.mmisra streamdata]$ ls
[acadgild.mmisra streamdata]$
[acadgild.mmisra streamdata]$
```

Launch Spark application

```
[acadgild.mmisra ~]$ spark-submit --class Stream.StreamingCaseStudy --deploy-mode
client \
> --master spark://${SPARK_MASTER_IP}:${SPARK_MASTER_PORT} /home/acadgild/spark1.jar
/home/acadgild/streamdata
18/09/26 06:59:07 INFO spark.SparkContext: Running Spark version 2.2.1
18/09/26 06:59:07 WARN util.NativeCodeLoader: Unable to load native-hadoop library for
your platform... using builtin-java classes where applicable
18/09/26 06:59:08 WARN util.Utils: Your hostname, localhost.localdomain resolves to a
loopback address: 127.0.0.1; using 192.168.0.107 instead (on interface eth16)
18/09/26 06:59:08 WARN util.Utils: Set SPARK_LOCAL_IP if you need to bind to another
address
18/09/26 06:59:08 INFO spark.SparkContext: Submitted application: SparkSteamingExample
18/09/26 06:59:08 INFO spark.SecurityManager: Changing view acls to: acadgild
18/09/26 06:59:08 INFO spark.SecurityManager: Changing modify acls to: acadgild
18/09/26 06:59:08 INFO spark.SecurityManager: Changing view acls groups to:
18/09/26 06:59:08 INFO spark.SecurityManager: Changing modify acls groups to:
18/09/26 06:59:08 INFO spark.SecurityManager: SecurityManager: authentication
disabled; ui acls disabled; users with view permissions: Set(acadgild); groups with
view permissions: Set(); users with modify permissions: Set(acadgild); groups with
modify permissions: Set()
18/09/26 06:59:08 INFO util.Utils: Successfully started service 'sparkDriver' on port
37807.
18/09/26 06:59:08 INFO spark.SparkEnv: Registering MapOutputTracker
18/09/26 06:59:08 INFO spark.SparkEnv: Registering BlockManagerMaster
18/09/26 06:59:08 INFO storage.BlockManagerMasterEndpoint: Using
org.apache.spark.storage.DefaultTopologyMapper for getting topology information
18/09/26 06:59:08 INFO storage.BlockManagerMasterEndpoint: BlockManagerMasterEndpoint
up
18/09/26 06:59:08 INFO storage.DiskBlockManager: Created local directory at
/tmp/blockmgr-681522af-77e3-41e6-821f-80046ce67202
18/09/26 06:59:08 INFO memory.MemoryStore: MemoryStore started with capacity 366.3 MB
18/09/26 06:59:08 INFO spark.SparkEnv: Registering OutputCommitCoordinator
18/09/26 06:59:09 INFO util.log: Logging initialized @2784ms
18/09/26 06:59:09 INFO server.Server: jetty-9.3.z-SNAPSHOT
18/09/26 06:59:09 INFO server.Server: Started @2928ms
18/09/26 06:59:09 INFO server.AbstractConnector: Started
ServerConnector@60ec3b6e{HTTP/1.1,[http/1.1]}{0.0.0.0:4040}
18/09/26 06:59:09 INFO util.Utils: Successfully started service 'SparkUI' on port
4040.
18/09/26 06:59:09 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@7646731d{/jobs,null,AVAILABLE,@Spark}
18/09/26 06:59:09 INFO handler.ContextHandler: Started
```

```
o.s.j.s.ServletContextHandler@797501a{/jobs/json,null,AVAILABLE,@Spark}
18/09/26 06:59:09 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@57f791c6{/jobs/job,null,AVAILABLE,@Spark}
18/09/26 06:59:09 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@5bd1ceca{/jobs/job/json,null,AVAILABLE,@Spark}
18/09/26 06:59:09 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@499b2a5c{/stages,null,AVAILABLE,@Spark}
18/09/26 06:59:09 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@clfca1e{/stages/json,null,AVAILABLE,@Spark}
18/09/26 06:59:09 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@344344fa{/stages/stage,null,AVAILABLE,@Spark}
18/09/26 06:59:09 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@285f09de{/stages/stage/json,null,AVAILABLE,@Spark}
18/09/26 06:59:09 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@31500940{/stages/pool,null,AVAILABLE,@Spark}
18/09/26 06:59:09 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@48e64352{/stages/pool/json,null,AVAILABLE,@Spark}
18/09/26 06:59:09 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@4362d7df{/storage,null,AVAILABLE,@Spark}
18/09/26 06:59:09 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@1c25b8a7{/storage/json,null,AVAILABLE,@Spark}
18/09/26 06:59:09 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@750fe12e{/storage/rdd,null,AVAILABLE,@Spark}
18/09/26 06:59:09 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@3e587920{/storage/rdd/json,null,AVAILABLE,@Spark}
18/09/26 06:59:09 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@24f43aa3{/environment,null,AVAILABLE,@Spark}
18/09/26 06:59:09 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@1e11bc55{/environment/json,null,AVAILABLE,@Spark}
18/09/26 06:59:09 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@70e0accd{/executors,null,AVAILABLE,@Spark}
18/09/26 06:59:09 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@6ab72419{/executors/json,null,AVAILABLE,@Spark}
18/09/26 06:59:09 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@4fdfa676{/executors/threadDump,null,AVAILABLE,@Spark}
18/09/26 06:59:09 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@5be82d43{/executors/threadDump/json,null,AVAILABLE,@Spark}
18/09/26 06:59:09 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@345e5a17{/static,null,AVAILABLE,@Spark}
18/09/26 06:59:09 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@1e886a5b{/,null,AVAILABLE,@Spark}
18/09/26 06:59:09 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@6e33c391{/api,null,AVAILABLE,@Spark}
18/09/26 06:59:09 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@213e3629{/jobs/job/kill,null,AVAILABLE,@Spark}
18/09/26 06:59:09 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@2a7b6f69{/stages/stage/kill,null,AVAILABLE,@Spark}
18/09/26 06:59:09 INFO ui.SparkUI: Bound SparkUI to 0.0.0.0, and started at
http://192.168.0.107:4040
18/09/26 06:59:09 INFO spark.SparkContext: Added JAR file:/home/acadgild/spark1.jar at
spark://192.168.0.107:37807/jars/spark1.jar with timestamp 1537925349497
18/09/26 06:59:09 INFO executor.Executor: Starting executor ID driver on host
localhost
18/09/26 06:59:09 INFO util.Utils: Successfully started service
'org.apache.spark.network.netty.NettyBlockTransferService' on port 44005.
18/09/26 06:59:09 INFO netty.NettyBlockTransferService: Server created on
192.168.0.107:44005
18/09/26 06:59:09 INFO storage.BlockManager: Using
org.apache.spark.storage.RandomBlockReplicationPolicy for block replication policy
18/09/26 06:59:09 INFO storage.BlockManagerMaster: Registering BlockManager
BlockManagerId(driver, 192.168.0.107, 44005, None)
18/09/26 06:59:09 INFO storage.BlockManagerMasterEndpoint: Registering block manager
```

```
192.168.0.107:44005 with 366.3 MB RAM, BlockManagerId(driver, 192.168.0.107, 44005,
None)
18/09/26 06:59:09 INFO storage.BlockManagerMaster: Registered BlockManager
BlockManagerId(driver, 192.168.0.107, 44005, None)
18/09/26 06:59:09 INFO storage.BlockManager: Initialized BlockManager:
BlockManagerId(driver, 192.168.0.107, 44005, None)
18/09/26 06:59:10 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@375b5b7f{/metrics/json,null,AVAILABLE,@Spark}
Monitoring directory /home/acadgild/streamdata for file to do wordcount
-----
Time: 1537925352000 ms
-----
```

```
-----
Time: 1537925355000 ms
-----
```

```
-----
Time: 1537925358000 ms
-----
```

Create a new file Test.txt with following content in the directory being monitored

```
[acadgild.mmisra streamdata]$ vim Test.txt
[acadgild.mmisra streamdata]$ cat Test.txt
This is a test to check spark streaming wordcount
[acadgild.mmisra streamdata]$
```

spark application output

```
-----
Time: 1537925385000 ms
-----
```

```
-----
Time: 1537925388000 ms
-----
```

```
(is,1)
(check,1)
(,1)
(This,1)
(streaming,1)
(test,1)
(spark,1)
(wordcount,1)
(a,1)
(to,1)
...
```

```
-----
Time: 1537925391000 ms
-----
```

```
-----
Time: 1537925394000 ms
-----
```

```
^CYou have mail in /var/spool/mail/acadgild
[acadgild.mmisra ~]$
```

Conclusion – We see that application is working as expected and doing the word count.

Second Part - In this part, you will have to create a Spark Application which should do the following

1. Pick up a file from the local directory and do the word count
2. Then in the same Spark Application, write the code to put the same file on HDFS.
3. Then in same Spark Application, do the word count of the file copied on HDFS in step 2
4. Lastly, compare the word count of step 1 and 2. Both should match, other throw an error

Solution to Second Part

The source code for the second part is given below. The program does the following

1. Take local file name and HDFS path location as command line arguments
2. There is a single method called wordcount which can be invoked either on local or HDFS file. It returns a List of (word,count). We compare the Lists to be equal for both local and HDFS file to see if the word counts match
3. There is a method to copy the file from local to HDFS using RDD.saveAsTextFile method

```
package Stream

import org.apache.spark.sql.SparkSession
import org.apache.spark.{SparkConf, SparkContext}

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

    val localfile = "file://" + args(0)
    val hdfsfile = args(1)

    if(args.length != 2)
    {
      println("usage: <local_file_name> <hdfs_path>")
    }
    val spark = SparkSession.builder().master(master="local")
      .appName(name="spark word count example")
      .config("spark.some.config.option", "some-value").getOrCreate()
    spark.sparkContext.setLogLevel("WARN")

    import spark.implicits._

    println("Doing wordcount for file " + localfile)
    val localwc = wordcount(spark.sparkContext, localfile)

    // print the wordcount
    println(localwc)

    // copy the file to HDFS
    copyFileToHDFS(spark.sparkContext, localfile, hdfsfile)

    // do wordcount on the HDFS file
    println("Doing wordcount for file " + hdfsfile)
    val hdfs wc = wordcount(spark.sparkContext, hdfsfile)

    println(hdfs wc)
```

```

// compare the lists for both local and hdfs wordcount
if(localwc.equals(hdfswc)==true)
{
    println("wordcount matches")
}
else
{
    println("wordcount does not match")
}

}

def copyFileToHDFS(sc:SparkContext,local:String,hdfs:String)={

    // create RDD for local file
    println("copying " + local + " to " + hdfs)
    val fileRDD = sc.textFile(local ,1)
    // save the RDD to HDFS
    fileRDD.saveAsTextFile(hdfs)
    println("copying done")

}

def wordcount(sc:SparkContext, name:String )=
{

    val fileRDD = sc.textFile(name ,1)
    // split the lines into words

    // we split the lines into word, map each word to a tuple of (word,1) and then do
countByKey
    val ans = fileRDD.flatMap(x=>x.split(" ")).map(p=>(p,1)).countByKey

    //ans.map(x=>println(x))
    ans.toList

}

}

```

Program Output

```
• MobaXterm 10.4 •
(SSH client, X-server and networking tools)

→ SSH session to acadgild@192.168.56.2
  • SSH compression : v
  • SSH-browser      : v
  • X11-forwarding   : v (remote display is forwarded through SSH)
  • DISPLAY          : v (automatically set on remote server)

→ For more info, ctrl+click on help or visit our website
```

```
Last login: Wed Sep 26 06:46:17 2018 from 192.168.56.1
[acadgild.mmisra ~]$ spark-submit --class Stream.SparkWordCount --deploy-mode client \
> --master spark://${SPARK_MASTER_IP}:${SPARK_MASTER_PORT} /home/acadgild/spark1.jar \
> /home/acadgild/story.txt /files/story
18/09/26 09:12:35 INFO spark.SparkContext: Running Spark version 2.2.1
18/09/26 09:12:36 WARN util.NativeCodeLoader: Unable to load native-hadoop library for
your platform... using builtin-java classes where applicable
18/09/26 09:12:36 WARN util.Utils: Your hostname, localhost.localdomain resolves to a
loopback address: 127.0.0.1; using 192.168.0.107 instead (on interface eth16)
18/09/26 09:12:36 WARN util.Utils: Set SPARK_LOCAL_IP if you need to bind to another
address
18/09/26 09:12:36 INFO spark.SparkContext: Submitted application: spark word count
example
18/09/26 09:12:36 INFO spark.SecurityManager: Changing view acls to: acadgild
18/09/26 09:12:36 INFO spark.SecurityManager: Changing modify acls to: acadgild
18/09/26 09:12:36 INFO spark.SecurityManager: Changing view acls groups to:
18/09/26 09:12:36 INFO spark.SecurityManager: Changing modify acls groups to:
18/09/26 09:12:36 INFO spark.SecurityManager: SecurityManager: authentication
disabled; ui acls disabled; users with view permissions: Set(acadgild); groups with
view permissions: Set(); users with modify permissions: Set(acadgild); groups with
modify permissions: Set()
18/09/26 09:12:37 INFO util.Utils: Successfully started service 'sparkDriver' on port
33746.
18/09/26 09:12:37 INFO spark.SparkEnv: Registering MapOutputTracker
18/09/26 09:12:37 INFO spark.SparkEnv: Registering BlockManagerMaster
18/09/26 09:12:37 INFO storage.BlockManagerMasterEndpoint: Using
org.apache.spark.storage.DefaultTopologyMapper for getting topology information
18/09/26 09:12:37 INFO storage.BlockManagerMasterEndpoint: BlockManagerMasterEndpoint
up
18/09/26 09:12:37 INFO storage.DiskBlockManager: Created local directory at
/tmp/blockmgr-57aaa03b-9f0d-4715-8016-a4bdbcf4d4e1a
18/09/26 09:12:37 INFO memory.MemoryStore: MemoryStore started with capacity 366.3 MB
18/09/26 09:12:37 INFO spark.SparkEnv: Registering OutputCommitCoordinator
18/09/26 09:12:37 INFO util.log: Logging initialized @2486ms
18/09/26 09:12:37 INFO server.Server: jetty-9.3.z-SNAPSHOT
18/09/26 09:12:37 INFO server.Server: Started @2597ms
18/09/26 09:12:37 INFO server.AbstractConnector: Started
ServerConnector@5941b6df{HTTP/1.1,[http/1.1]}{0.0.0.0:4040}
18/09/26 09:12:37 INFO util.Utils: Successfully started service 'SparkUI' on port
4040.
18/09/26 09:12:37 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@408b35bf{/jobs,null,AVAILABLE,@Spark}
18/09/26 09:12:37 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@615f972{/jobs/json,null,AVAILABLE,@Spark}
18/09/26 09:12:37 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@73393584{/jobs/job,null,AVAILABLE,@Spark}
18/09/26 09:12:37 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@48e64352{/jobs/job/json,null,AVAILABLE,@Spark}
```

18/09/26 09:12:37 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@4362d7df{/stages,null,AVAILABLE,@Spark}
18/09/26 09:12:37 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@1c25b8a7{/stages/json,null,AVAILABLE,@Spark}
18/09/26 09:12:37 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@750fe12e{/stages/stage,null,AVAILABLE,@Spark}
18/09/26 09:12:37 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@24f43aa3{/stages/stage/json,null,AVAILABLE,@Spark}
18/09/26 09:12:37 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@1e11bc55{/stages/pool,null,AVAILABLE,@Spark}
18/09/26 09:12:37 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@70e0accd{/stages/pool/json,null,AVAILABLE,@Spark}
18/09/26 09:12:37 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@6ab72419{/storage,null,AVAILABLE,@Spark}
18/09/26 09:12:37 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@4fdfa676{/storage/json,null,AVAILABLE,@Spark}
18/09/26 09:12:37 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@5be82d43{/storage/rdd,null,AVAILABLE,@Spark}
18/09/26 09:12:37 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@345e5a17{/storage/rdd/json,null,AVAILABLE,@Spark}
18/09/26 09:12:37 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@443dbe42{/environment,null,AVAILABLE,@Spark}
18/09/26 09:12:37 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@1734f68{/environment/json,null,AVAILABLE,@Spark}
18/09/26 09:12:37 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@5ed190be{/executors,null,AVAILABLE,@Spark}
18/09/26 09:12:37 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@5bbc9f97{/executors/json,null,AVAILABLE,@Spark}
18/09/26 09:12:37 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@41382722{/executors/threadDump,null,AVAILABLE,@Spark}
18/09/26 09:12:37 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@425357dd{/executors/threadDump/json,null,AVAILABLE,@Spark}
18/09/26 09:12:37 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@210386e0{/static,null,AVAILABLE,@Spark}
18/09/26 09:12:37 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@7db0565c{/,null,AVAILABLE,@Spark}
18/09/26 09:12:37 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@52each4b{/api,null,AVAILABLE,@Spark}
18/09/26 09:12:37 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@cc62a3b{/jobs/job/kill,null,AVAILABLE,@Spark}
18/09/26 09:12:37 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@29539e36{/stages/stage/kill,null,AVAILABLE,@Spark}
18/09/26 09:12:37 INFO ui.SparkUI: Bound SparkUI to 0.0.0.0, and started at
http://192.168.0.107:4040
18/09/26 09:12:37 INFO spark.SparkContext: Added JAR file:/home/acadgild/spark1.jar at
spark://192.168.0.107:33746/jars/spark1.jar with timestamp 1537933357678
18/09/26 09:12:37 INFO executor.Executor: Starting executor ID driver on host
localhost
18/09/26 09:12:37 INFO util.Utils: Successfully started service
'org.apache.spark.network.netty.NettyBlockTransferService' on port 36693.
18/09/26 09:12:37 INFO netty.NettyBlockTransferService: Server created on
192.168.0.107:36693
18/09/26 09:12:37 INFO storage.BlockManager: Using
org.apache.spark.storage.RandomBlockReplicationPolicy for block replication policy
18/09/26 09:12:37 INFO storage.BlockManagerMaster: Registering BlockManager
BlockManagerId(driver, 192.168.0.107, 36693, None)
18/09/26 09:12:37 INFO storage.BlockManagerMasterEndpoint: Registering block manager
192.168.0.107:36693 with 366.3 MB RAM, BlockManagerId(driver, 192.168.0.107, 36693,
None)
18/09/26 09:12:37 INFO storage.BlockManagerMaster: Registered BlockManager
BlockManagerId(driver, 192.168.0.107, 36693, None)
18/09/26 09:12:37 INFO storage.BlockManager: Initialized BlockManager:


```

BlockManagerId(driver, 192.168.0.107, 36693, None)
18/09/26 09:12:38 INFO handler.ContextHandler: Started
o.s.j.s.ServletContextHandler@70325d20{/metrics/json,null,AVAILABLE,@Spark}
Doing wordcount for file file:///home/acadgild/story.txt
List((used,1), (don't,1), (I'm,1), (health,1), (One,1), (enjoying,1), (please,1),
(,11), (able,1), (for,3), (tricks,1), (any,2), (inviting,1), (Once,1), (in,2),
(have,5), (offered,1), (is,3), (How,1), (hungry,1), (said,2), (given,1),
(Humiliated,,1), (himself,1), (After,2), (Don't,1), (dinner.,1), (but,3), (She,2),
(Stork,1), (can't,1), (what,1), (if,1), (good,1), (problem,1), (big,1), (felt,1),
(house.,1), (him,1), (stammered,,1), (served,1), (had,1), (not.,1), (soup?,1),
(Fox,1), (bill,,3), (invited,1), (before,1), (just,2), (saw,1), (left,2), (it,2),
(fox.,2), (finishing,1), (you,,1), (a,10), (upset,,1), (accepted,1), (as,1),
(stomach,1), (him.,1), (surprise,1), (Clever,1), (disappointment,,1), (animals,1),
(troubling,1), (running.,1), (sweetly,1), (thanking,1), (I,3), (feast,4), (that,2),
(time,,1), (out,1), (cannot,1), (upon,1), (replied,2), (happily,1), (to,6),
(stork,,1), (now.,1), (ache.,1), (met,1), (you,2), (soup,6), (StoryThe,1), (was,3),
(long,3), (licked,1), (stork's,1), (there,1), (sorry.,1), (The,6), (promised,,1),
(on,1), (friend.,1), (asked,2), (came,,1), (poor,1), (my,1), (gain,1), (after,1),
(take,1), (good,,1), (Soon,,1), (bowl!,1), (reached,1), (Oh,2), (befriended,1),
(acted,1), (house,1), (he,6), (obviously,1), (ashamed.,1), (kitchen,,1),
(mischievous,1), (lived,1), (her,,1), (their,1), (He,3), (not,2), (with,5), (from,1),
(To,1), (hers,,1), (easily,2), (soup!,1), (enjoy,1), (place.,1), (both,1), (plate.,1),
(touched,1), (invitation.,1), (could,3), (make,1), (fox,,1), (tip,1), (remembered,1),
(its,2), (them.,1), (cunning,1), (say,1), (place,2), (brought,1), (dear,,1),
(sorry,1), (them,,1), (her,3), (shallow,1), (Cunning,1), (playing,1), (soup.,2),
(better,1), (fox's,1), (I'd,1), (more,1), (When,1), (As,1), (arrived,1), (narrow,1),
(offer.,1), (very,4), (pleasantries,,1), (it?,1), (jar,1), (some,2), (day,3), (fox,8),
(exchanging,1), (tummy,1), (like,2), (other,1), (of,4), (stork.,1), (and,12),
(neck.,1), (speak,1), (stork,11), (trust,,1), (went,1), (leave,1), (the,34))
copying file:///home/acadgild/story.txt to /files/story
copying done
Doing wordcount for file /files/story
List((used,1), (don't,1), (I'm,1), (health,1), (One,1), (enjoying,1), (please,1),
(,11), (able,1), (for,3), (tricks,1), (any,2), (inviting,1), (Once,1), (in,2),
(have,5), (offered,1), (is,3), (How,1), (hungry,1), (said,2), (given,1),
(Humiliated,,1), (himself,1), (After,2), (Don't,1), (dinner.,1), (but,3), (She,2),
(Stork,1), (can't,1), (what,1), (if,1), (good,1), (problem,1), (big,1), (felt,1),
(house.,1), (him,1), (stammered,,1), (served,1), (had,1), (not.,1), (soup?,1),
(Fox,1), (bill,,3), (invited,1), (before,1), (just,2), (saw,1), (left,2), (it,2),
(fox.,2), (finishing,1), (you,,1), (a,10), (upset,,1), (accepted,1), (as,1),
(stomach,1), (him.,1), (surprise,1), (Clever,1), (disappointment,,1), (animals,1),
(troubling,1), (running.,1), (sweetly,1), (thanking,1), (I,3), (feast,4), (that,2),
(time,,1), (out,1), (cannot,1), (upon,1), (replied,2), (happily,1), (to,6),
(stork,,1), (now.,1), (ache.,1), (met,1), (you,2), (soup,6), (StoryThe,1), (was,3),
(long,3), (licked,1), (stork's,1), (there,1), (sorry.,1), (The,6), (promised,,1),
(on,1), (friend.,1), (asked,2), (came,,1), (poor,1), (my,1), (gain,1), (after,1),
(take,1), (good,,1), (Soon,,1), (bowl!,1), (reached,1), (Oh,2), (befriended,1),
(acted,1), (house,1), (he,6), (obviously,1), (ashamed.,1), (kitchen,,1),
(mischievous,1), (lived,1), (her,,1), (their,1), (He,3), (not,2), (with,5), (from,1),
(To,1), (hers,,1), (easily,2), (soup!,1), (enjoy,1), (place.,1), (both,1), (plate.,1),
(touched,1), (invitation.,1), (could,3), (make,1), (fox,,1), (tip,1), (remembered,1),
(its,2), (them.,1), (cunning,1), (say,1), (place,2), (brought,1), (dear,,1),
(sorry,1), (them,,1), (her,3), (shallow,1), (Cunning,1), (playing,1), (soup.,2),
(better,1), (fox's,1), (I'd,1), (more,1), (When,1), (As,1), (arrived,1), (narrow,1),
(offer.,1), (very,4), (pleasantries,,1), (it?,1), (jar,1), (some,2), (day,3), (fox,8),
(exchanging,1), (tummy,1), (like,2), (other,1), (of,4), (stork.,1), (and,12),
(neck.,1), (speak,1), (stork,11), (trust,,1), (went,1), (leave,1), (the,34))
wordcount matches
[acadgild.mmisra ~]$

[acadgild.mmisra ~]$ hadoop fs -ls /files

```

18/09/26 09:12:25 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

Found 9 items

-rw-r--r--	1	acadgild supergroup	1042	2018-09-14 11:41	/files/Sports_data.txt
drwxr-xr-x	-	acadgild supergroup	0	2018-07-19 16:38	/files/assignment_1
drwxr-xr-x	-	acadgild supergroup	0	2018-07-19 16:41	/files/assignment_2
-rw-r--r--	1	acadgild supergroup	317	2018-07-22 22:21	/files/custompiginput.txt
-rw-r--r--	1	acadgild supergroup	145	2018-08-06 15:14	/files/example.txt
-rw-r--r--	1	acadgild supergroup	2283410	2018-07-19 21:35	/files/movies.csv
-rw-r--r--	1	acadgild supergroup	709550327	2018-07-19 16:06	/files/ratings.csv
-rw-r--r--	1	acadgild supergroup	336	2018-07-19 21:50	/files/test.txt

You have mail in /var/spool/mail/acadgild

after running the program

[acadgild.mmisra ~]\$ hadoop fs -ls /files

18/09/26 09:18:47 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

Found 10 items

-rw-r--r--	1	acadgild supergroup	1042	2018-09-14 11:41	/files/Sports_data.txt
drwxr-xr-x	-	acadgild supergroup	0	2018-07-19 16:38	/files/assignment_1
drwxr-xr-x	-	acadgild supergroup	0	2018-07-19 16:41	/files/assignment_2
-rw-r--r--	1	acadgild supergroup	317	2018-07-22 22:21	/files/custompiginput.txt
-rw-r--r--	1	acadgild supergroup	145	2018-08-06 15:14	/files/example.txt
-rw-r--r--	1	acadgild supergroup	2283410	2018-07-19 21:35	/files/movies.csv
-rw-r--r--	1	acadgild supergroup	709550327	2018-07-19 16:06	/files/ratings.csv
drwxr-xr-x	-	acadgild supergroup	0	2018-09-26 09:12	/files/story
-rw-r--r--	1	acadgild supergroup	336	2018-07-19 21:50	/files/test.txt

You have mail in /var/spool/mail/acadgild

[acadgild.mmisra ~]\$ hadoop fs -ls /files/story

18/09/26 09:18:57 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

Found 2 items

-rw-r--r--	1	acadgild supergroup	0	2018-09-26 09:12	/files/story/_SUCCESS
-rw-r--r--	1	acadgild supergroup	1763	2018-09-26 09:12	/files/story/part-00000

[acadgild.mmisra ~]\$