

Program 4 Report

Yangxiao Wang

Contents

1	Documentation	2
2	Source code	2
2.1	ShortBFS.java	2
3	Execution output	5
3.1	Output	5
4	Discussions	6
5	Lab Sessions 4	8
5.1	Lab4A	8
5.2	Lab4A Output	8
5.3	Lab4B	11
5.4	Lab4B Output	12

1 Documentation

In this project, we are using spark to implement Breadth First Search algorithm. My approach followed the provided descriptions in the assignment PDF. It first read the whole graph into the JavaPairRDD with String and customized object Data. Data contains all the attributes of a vertex. Then I started the while loop with the source vertex active, in the while loop:

- For every active vertex, update its neighbors' distance to source.
- For every vertex has a new and shorter distance, active this vertex.

2 Source code

2.1 ShortBFS.java

```
1 import org.apache.spark.SparkConf;
2 import org.apache.spark.api.java.JavaPairRDD;
3 import org.apache.spark.api.java.JavaRDD;
4 import org.apache.spark.api.java.JavaSparkContext;
5 import scala.Tuple2;
6
7 import java.io.Serializable;
8 import java.util.ArrayList;
9 import java.util.List;
10
11 /**
12  * @author Yangxiao on 11/20/2018.
13  */
14
15 public class ShortBFS
16 {
17
18     public static final String ACTIVE = "ACTIVE";
19     public static final String INACTIVE = "INACTIVE";
20
21     public static void main(String[] args)
22     {
23         // start Sparks and read a given input file
24         String inputFile = args[0];
25         String start = args[1];
26         String end = args[2];
27         SparkConf conf = new SparkConf().setAppName("BFS-based Shortest Path Search");
28         JavaSparkContext jsc = new JavaSparkContext(conf);
29         JavaRDD<String> lines = jsc.textFile(inputFile);
30         // now start a timer
31         long startTime = System.currentTimeMillis();
32         JavaPairRDD<String, Data> network =
33         lines.mapToPair(line ->
34         {
35             int indE = line.indexOf("=");
36             String vID = line.substring(0, indE);
37             String[] lists = line.substring(indE + 1).split("; ");
38             int n = lists.length;
39             List<Tuple2<String, Integer>> neighbors = new ArrayList<>();
40             for (int i = 0; i < n; i++)
41             {
42                 String[] parts = lists[i].split(",");
43                 neighbors.add(new Tuple2<>(parts[0], Integer.parseInt(parts[1])));
44             }
45             Data data1;
46             if (vID.equals(start))
47             {
48                 data1 = new Data(neighbors, 0, 0, ACTIVE);
49             }
50             else
51             {
```

```

52         data1 = new Data(neighbors, Integer.MAX_VALUE, Integer.MAX_VALUE, INACTIVE)↵
53     };
54     return new Tuple2<>(vID, data1);
55 }
56 );
57 System.out.println("Count = " + network.count() + "\n");
58
59 while (network.filter(f ->
60 {
61     if (f._2.status.equals(ACTIVE))
62     {
63         return true;
64     }
65     else
66     {
67         return false;
68     }
69 }
70 ).count() > 0)
71 {
72     JavaPairRDD<String, Data> propagatedNetwork = network.flatMapToPair(vertex ->
73     {
74         // If a vertex is "ACTIVE", create Tuple2( neighbor, new Data() ) for
75         // each neighbor where Data should include a new distance to this neighbor.
76         // Add each Tuple2 to a list. Dont forget this vertex itself back to the
77         // list. Return all the list items.
78         List<Tuple2<String, Data>> list = new ArrayList<>();
79         list.add(new Tuple2<>(vertex._1,
80         new Data(vertex._2.neighbors, vertex._2.distance, vertex._2.prev,
81         INACTIVE)));
82         if (vertex._2.status.equals(ACTIVE))
83         {
84             for (Tuple2<String, Integer> neighbor : vertex._2.neighbors)
85             {
86                 list.add(new Tuple2<>(neighbor._1,
87                 new Data(new ArrayList<>(), (neighbor._2 + vertex._2.distance),
88                 Integer.MAX_VALUE, INACTIVE)));
89             }
90         }
91         return list.iterator();
92     }
93 );
94
95 network = propagatedNetwork.reduceByKey((k1, k2) ->
96 {
97     // For each key, (i.e, each vertex), find the shortest distance and
98     // update this vertex Data attribute.
99     List<Tuple2<String, Integer>> neighbors =
100     k1.neighbors.size() == 0 ? k2.neighbors : k1.neighbors;
101     int dis = Math.min(k1.distance, k2.distance);
102     int prev = Math.min(k1.prev, k2.prev);
103     return new Data(neighbors, dis, prev, INACTIVE);
104 }
105 );
106
107 network = network.mapValues(value ->
108 {
109     // If a vertex new distance is shorter than prev, activate this vertex
110     // status and replace prev with the new distance.
111     if (value.distance < value.prev)
112     {
113         return new Data(value.neighbors, value.distance, value.distance, ACTIVE)↵
114     }
115     return value;
116 }
117 );
118 }
119 List<Data> res = network.lookup(end);
120 System.out.println("from " + start + " to " + end + " takes distance = " + res.get↵

```

```

121         (0).distance);
122         System.out.println("Time = " + (System.currentTimeMillis() - startTime));
123     }
124
125     /**
126     * Vertex Attributes
127     */
128     static class Data implements Serializable
129     {
130         List<Tuple2<String, Integer>> neighbors; // <neighbor0, weight0>, ...
131         String status; // "INACTIVE" or "ACTIVE"
132         Integer distance; // the distance so far from source to this vertex
133         Integer prev; // the distance calculated in the previous iteration
134
135         public Data()
136         {
137             neighbors = new ArrayList<>();
138             status = "INACTIVE";
139             distance = 0;
140         }
141
142         public Data(List<Tuple2<String, Integer>> neighbors,
143                     Integer dist, Integer prev, String status)
144         {
145             if (neighbors != null)
146             {
147                 this.neighbors = new ArrayList<>(neighbors);
148             }
149             else
150             {
151                 this.neighbors = new ArrayList<>();
152             }
153             this.distance = dist;
154             this.prev = prev;
155             this.status = status;
156         }
157     }

```

3 Execution output

3.1 Output

```
2018-11-27 12:42:31 INFO TaskSchedulerImpl:54 - Removed TaskSet 252.0, whose tasks have all completed, from pool
2018-11-27 12:42:31 INFO DAGScheduler:54 - ResultStage 252 (lookup at ShortBFS.java:99) finished in 2.288 s
2018-11-27 12:42:31 INFO DAGScheduler:54 - Job 22 finished: lookup at ShortBFS.java:99, took 2.293663 s
from 0 to 1500 takes distance = 41
Time = 147700
2018-11-27 12:42:31 INFO SparkContext:54 - Invoking stop() from shutdown hook
2018-11-27 12:42:31 INFO BlockManagerInfo:54 - Removed broadcast_43_piece0 on v0242p.host.s.uw.edu:36926 in memory (size: 2.4 KB, free: 366.2 MB)
2018-11-27 12:42:31 INFO BlockManagerInfo:54 - Removed broadcast_43_piece0 on 172.28.203.198:41883 in memory (size: 2.4 KB, free: 366.2 MB)
2018-11-27 12:42:31 INFO AbstractConnector:318 - Stopped Spark@67af833b(HTTP/1.1,[http/1.1]){0.0.0.0:4042}
2018-11-27 12:42:31 INFO SparkUI:54 - Stopped Spark web UI at http://v0242p.host.s.uw.edu:4042
2018-11-27 12:42:31 INFO StandaloneSchedulerBackend:54 - Shutting down all executors
2018-11-27 12:42:31 INFO CoarseGrainedSchedulerBackend$DriverEndpoint:54 - Asking each executor to shut down
2018-11-27 12:42:31 INFO MapOutputTrackerMasterEndpoint:54 - MapOutputTrackerMasterEndpoint stopped!
2018-11-27 12:42:31 INFO MemoryStore:54 - MemoryStore cleared
2018-11-27 12:42:31 INFO BlockManager:54 - BlockManager stopped
2018-11-27 12:42:31 INFO BlockManagerMaster:54 - BlockManagerMaster stopped
2018-11-27 12:42:31 INFO OutputCommitCoordinator$OutputCommitCoordinatorEndpoint:54 - OutputCommitCoordinator stopped!
2018-11-27 12:42:31 INFO SparkContext:54 - Successfully stopped SparkContext
2018-11-27 12:42:31 INFO ShutdownHookManager:54 - Shutdown hook called
2018-11-27 12:42:31 INFO ShutdownHookManager:54 - Deleting directory /tmp/spark-700dcf8b-a72b-4a49-898d-0597c28406ac
2018-11-27 12:42:31 INFO ShutdownHookManager:54 - Deleting directory /tmp/spark-f9eb5600-8aef-4f79-8b6d-9df9fef0bc98
```

Figure 1: 0-1500

```
2018-11-27 12:11:06 INFO DAGScheduler:54 - ResultStage 252 (lookup at ShortBFS.java:99) finished in 2.878 s
2018-11-27 12:11:06 INFO DAGScheduler:54 - Job 22 finished: lookup at ShortBFS.java:99, took 2.882568 s
from 0 to 2999 takes distance = 32
Time = 153992
2018-11-27 12:11:06 INFO SparkContext:54 - Invoking stop() from shutdown hook
2018-11-27 12:11:06 INFO AbstractConnector:318 - Stopped Spark@b93aad(HTTP/1.1,[http/1.1]){0.0.0.0:4040}
2018-11-27 12:11:06 INFO SparkUI:54 - Stopped Spark web UI at http://v0242p.host.s.uw.edu:4040
2018-11-27 12:11:06 INFO StandaloneSchedulerBackend:54 - Shutting down all executors
2018-11-27 12:11:06 INFO CoarseGrainedSchedulerBackend$DriverEndpoint:54 - Asking each executor to shut down
2018-11-27 12:11:06 INFO MapOutputTrackerMasterEndpoint:54 - MapOutputTrackerMasterEndpoint stopped!
2018-11-27 12:11:06 INFO MemoryStore:54 - MemoryStore cleared
2018-11-27 12:11:06 INFO BlockManager:54 - BlockManager stopped
2018-11-27 12:11:06 INFO BlockManagerMaster:54 - BlockManagerMaster stopped
2018-11-27 12:11:06 INFO OutputCommitCoordinator$OutputCommitCoordinatorEndpoint:54 - OutputCommitCoordinator stopped!
2018-11-27 12:11:06 INFO SparkContext:54 - Successfully stopped SparkContext
2018-11-27 12:11:06 INFO ShutdownHookManager:54 - Shutdown hook called
2018-11-27 12:11:06 INFO ShutdownHookManager:54 - Deleting directory /tmp/spark-dfdb726f-3ed3-4083-b4fb-916ad12bed35
2018-11-27 12:11:06 INFO ShutdownHookManager:54 - Deleting directory /tmp/spark-67b8201d-7d0f-4a73-9b5b-bd74e35969c9
```

Figure 2: 0-2999 with 12 cores

```
2018-11-26 17:03:40 INFO TaskSchedulerImpl:54 - Removed taskset 252.0, whose tasks have all completed, from pool
2018-11-26 17:03:40 INFO DAGScheduler:54 - ResultStage 252 (collect at ShortBFS.java:96) finished in 10.422 s
2018-11-26 17:03:40 INFO DAGScheduler:54 - Job 22 finished: collect at ShortBFS.java:96, took 10.425772 s
2018-11-26 17:03:40 INFO BlockManagerInfo:54 - Removed taskresult_85 on 172.28.203.196:38253 in memory (size: 24.3 MB, free: 366.3 MB)
Size = 3000
from 0 to 2999 takes distance = 32
Time = 272355
2018-11-26 17:03:40 INFO SparkContext:54 - Invoking stop() from shutdown hook
2018-11-26 17:03:40 INFO AbstractConnector:318 - Stopped Spark@b93aad(HTTP/1.1,[http/1.1]){0.0.0.0:4040}
2018-11-26 17:03:40 INFO SparkUI:54 - Stopped Spark web UI at http://v0242p.host.s.uw.edu:4040
2018-11-26 17:03:40 INFO StandaloneSchedulerBackend:54 - Shutting down all executors
2018-11-26 17:03:40 INFO CoarseGrainedSchedulerBackend$DriverEndpoint:54 - Asking each executor to shut down
2018-11-26 17:03:40 INFO MapOutputTrackerMasterEndpoint:54 - MapOutputTrackerMasterEndpoint stopped!
2018-11-26 17:03:40 INFO MemoryStore:54 - MemoryStore cleared
2018-11-26 17:03:40 INFO BlockManager:54 - BlockManager stopped
2018-11-26 17:03:40 INFO BlockManagerMaster:54 - BlockManagerMaster stopped
2018-11-26 17:03:40 INFO OutputCommitCoordinator$OutputCommitCoordinatorEndpoint:54 - OutputCommitCoordinator stopped!
2018-11-26 17:03:40 INFO SparkContext:54 - Successfully stopped SparkContext
2018-11-26 17:03:40 INFO ShutdownHookManager:54 - Shutdown hook called
2018-11-26 17:03:40 INFO ShutdownHookManager:54 - Deleting directory /tmp/spark-630b8467-d64b-47cd-95aa-c3c411f374ce
2018-11-26 17:03:40 INFO ShutdownHookManager:54 - Deleting directory /tmp/spark-745f5357-6786-4eb3-9171-6c5d2bbc6d8b
```

Figure 3: 0-2999 with 1 core

# of Comp nodes	Avg # of cores per node	Total # of cores	Elapsed time (sec)	Performance improvement
1	1	1	272.355	1
1	2	2	221.677	1.228612
1	3	3	234.051	1.163657
2	1	2	162.512	1.675907
2	2	4	192.527	1.414633
2	3	6	171.951	1.583911
4	1	4	163.253	1.6683
4	2	8	197.543	1.378712
4	3	12	147.7	1.843974272

4 Discussions

- Programmability Pros: the original algorithm is easy to implement, it first transfer input file to JavaPairRDD with String and Data. Then it find Active nodes' neighbors, and calculate the distance. And update its distance. These processes are straightforward and easy to implemented. Furthermore, the data exchanging and parallel part are already taken cared by Spark, unlike MPI, I do not need write code for them.
- Programmability Cons: The Data object is not optimized for this BFS search, and spark api is just generally not easy to understand such as the *flatMapTopair* and *reduceByKey*.
- Execution performance Pros: In general, the most time consuming part in Spark is collect operation. Fortunately, my code does not have any collect operations so the performance should be fast with the RDD transformation operations.
- Execution performance Cons: The RDD transformation processes contain duplicated loops which could slow the execution performance. For example, *network* to *propagatedNetwork* with traversal through all items in *network*, then *propagatedNetwork* reduced to *network* with traversal through duplicated items in *propagatedNetwork*, and loop through *network* to update distance. To improve this procedure, the updating part can be moved into the reduce part, since only reduced item has updated distance.
- Improvement:

```

1 network = propagatedNetwork.reduceByKey((k1, k2) ->
2 {
3     // For each key, (i.e, each vertex), find the shortest distance and
4     // update this vertex Data attribute.
5     List<Tuple2<String, Integer>> neighbors =
6     k1.neighbors.size() == 0 ? k2.neighbors : k1.neighbors;
7     int dis = Math.min(k1.distance, k2.distance);
8     int prev = Math.min(k1.prev, k2.prev);
9
10    String status = (k1.status.equals(ACTIVE) || k2.status.equals(ACTIVE)) ? ACTIVE :↵
        INACTIVE;
11
12    if (dis < prev || status.equals(ACTIVE))
13    {
14        return new Data(neighbors, dis, dis, ACTIVE);
15    }
16    else
17    {
18        return new Data(neighbors, dis, prev, INACTIVE);
19    }
20 }
21 );

```

Check and update the distance while reducing the duplicated keys might improve the performance.

5 Lab Sessions 4

5.1 Lab4A

```
1 import org.apache.spark.SparkConf;
2 import org.apache.spark.api.java.JavaRDD;
3 import org.apache.spark.api.java.JavaSparkContext;
4
5 import java.io.Serializable;
6 import java.util.Arrays;
7 import java.util.Comparator;
8
9 public class MyClass
10 {
11     public static void main(String[] args)
12     {
13         // a driver program
14         // initialize Spark Context
15         SparkConf conf = new SparkConf().setAppName("My Driver");
16         JavaSparkContext sc = new JavaSparkContext(conf);
17
18         // read data from a file
19         JavaRDD<String> document = sc.textFile("sample.txt");
20         // read data from another data structure
21         JavaRDD<Integer> numbers = sc.parallelize(Arrays.asList(0, 1, 2, 3, 4, 5));
22
23         // apply transformations/actions to RDD
24         System.out.println("#words = " +
25             document.flatMap(s -> Arrays.asList(s.split(" ")).iterator()).count());
26         System.out.println("max = " + numbers.max(new MyClassMax()));
27
28         sc.stop(); // stop Spark Context
29     }
30 }
31
32 class MyClassMax implements Serializable, Comparator<Integer>
33 {
34     @Override
35     public int compare(Integer o1, Integer o2)
36     {
37         return Integer.compare(o1, o2);
38     }
39 }
```

5.2 Lab4A Output

```
#words = 11
max = 5
```

```
1 [wyxiao_css534@cssmp11 spark_programs]$ spark-submit --class MyClass --master local --
  MyClass.jar
2 2018-11-21 20:14:02 WARN NativeCodeLoader:62 - Unable to load native-hadoop library --
  for your platform... using builtin-java classes where applicable
3 2018-11-21 20:14:02 INFO SparkContext:54 - Running Spark version 2.3.1
4 2018-11-21 20:14:02 INFO SparkContext:54 - Submitted application: My Driver
5 2018-11-21 20:14:02 INFO SecurityManager:54 - Changing view acls to: wyxiao_css534
6 2018-11-21 20:14:02 INFO SecurityManager:54 - Changing modify acls to: wyxiao_css534
7 2018-11-21 20:14:02 INFO SecurityManager:54 - Changing view acls groups to:
8 2018-11-21 20:14:02 INFO SecurityManager:54 - Changing modify acls groups to:
9 2018-11-21 20:14:02 INFO SecurityManager:54 - SecurityManager: authentication --
  disabled; ui acls disabled; users with view permissions: Set(wyxiao_css534); --
```

```

groups with view permissions: Set(); users with modify permissions: Set(←
wuxiao_css534); groups with modify permissions: Set()
10 2018-11-21 20:14:03 INFO  Utils:54 - Successfully started service 'sparkDriver' on ←
port 33246.
11 2018-11-21 20:14:03 INFO  SparkEnv:54 - Registering MapOutputTracker
12 2018-11-21 20:14:03 INFO  SparkEnv:54 - Registering BlockManagerMaster
13 2018-11-21 20:14:03 INFO  BlockManagerMasterEndpoint:54 - Using org.apache.spark.←
storage.DefaultTopologyMapper for getting topology information
14 2018-11-21 20:14:03 INFO  BlockManagerMasterEndpoint:54 - BlockManagerMasterEndpoint ←
up
15 2018-11-21 20:14:03 INFO  DiskBlockManager:54 - Created local directory at /tmp/←
blockmgr-887fd971-00d2-43b5-afd9-ad8826788558
16 2018-11-21 20:14:03 INFO  MemoryStore:54 - MemoryStore started with capacity 366.3 MB
17 2018-11-21 20:14:03 INFO  SparkEnv:54 - Registering OutputCommitCoordinator
18 2018-11-21 20:14:03 INFO  log:192 - Logging initialized @1672ms
19 2018-11-21 20:14:03 INFO  Server:346 - jetty-9.3.z-SNAPSHOT
20 2018-11-21 20:14:03 INFO  Server:414 - Started @1749ms
21 2018-11-21 20:14:03 INFO  AbstractConnector:278 - Started ServerConnector@5dda6f9{HTTP←
/1.1,[http/1.1]}{0.0.0.0:4040}
22 2018-11-21 20:14:03 INFO  Utils:54 - Successfully started service 'SparkUI' on port ←
4040.
23 2018-11-21 20:14:03 INFO  ContextHandler:781 - Started o.s.j.s.←
ServletContextHandler@6bc28a83{/jobs,null,AVAILABLE,@Spark}2018-11-21 20:14:03 INFO←
ContextHandler:781 - Started o.s.j.s.ServletContextHandler@384fc774{/jobs/json,←
null,AVAILABLE,@Spark}
24 2018-11-21 20:14:03 INFO  ContextHandler:781 - Started o.s.j.s.←
ServletContextHandler@e3cee7b{/jobs/job,null,AVAILABLE,@Spark}
25 2018-11-21 20:14:03 INFO  ContextHandler:781 - Started o.s.j.s.←
ServletContextHandler@6b9267b{/jobs/job/json,null,AVAILABLE,@Spark}
26 2018-11-21 20:14:03 INFO  ContextHandler:781 - Started o.s.j.s.←
ServletContextHandler@408b35bf{/stages,null,AVAILABLE,@Spark}
27 2018-11-21 20:14:03 INFO  ContextHandler:781 - Started o.s.j.s.←
ServletContextHandler@29ad44e3{/stages/json,null,AVAILABLE,@Spark}
28 2018-11-21 20:14:03 INFO  ContextHandler:781 - Started o.s.j.s.←
ServletContextHandler@15bcf458{/stages/stage,null,AVAILABLE,@Spark}
29 2018-11-21 20:14:03 INFO  ContextHandler:781 - Started o.s.j.s.←
ServletContextHandler@fac80{/stages/stage/json,null,AVAILABLE,@Spark}
30 2018-11-21 20:14:03 INFO  ContextHandler:781 - Started o.s.j.s.←
ServletContextHandler@726386ed{/stages/pool,null,AVAILABLE,@Spark}
31 2018-11-21 20:14:03 INFO  ContextHandler:781 - Started o.s.j.s.←
ServletContextHandler@649f2009{/stages/pool/json,null,AVAILABLE,@Spark}
32 2018-11-21 20:14:03 INFO  ContextHandler:781 - Started o.s.j.s.←
ServletContextHandler@14bb2297{/storage,null,AVAILABLE,@Spark}
33 2018-11-21 20:14:03 INFO  ContextHandler:781 - Started o.s.j.s.←
ServletContextHandler@69adf72c{/storage/json,null,AVAILABLE,@Spark}
34 2018-11-21 20:14:03 INFO  ContextHandler:781 - Started o.s.j.s.←
ServletContextHandler@797501a{/storage/rdd,null,AVAILABLE,@Spark}
35 2018-11-21 20:14:03 INFO  ContextHandler:781 - Started o.s.j.s.←
ServletContextHandler@1a15b789{/storage/rdd/json,null,AVAILABLE,@Spark}
36 2018-11-21 20:14:03 INFO  ContextHandler:781 - Started o.s.j.s.←
ServletContextHandler@57f791c6{/environment,null,AVAILABLE,@Spark}
37 2018-11-21 20:14:03 INFO  ContextHandler:781 - Started o.s.j.s.←
ServletContextHandler@51650883{/environment/json,null,AVAILABLE,@Spark}
38 2018-11-21 20:14:03 INFO  ContextHandler:781 - Started o.s.j.s.←
ServletContextHandler@6c4f9535{/executors,null,AVAILABLE,@Spark}
39 2018-11-21 20:14:03 INFO  ContextHandler:781 - Started o.s.j.s.←
ServletContextHandler@5bd1ceca{/executors/json,null,AVAILABLE,@Spark}
40 2018-11-21 20:14:03 INFO  ContextHandler:781 - Started o.s.j.s.←
ServletContextHandler@30c31dd7{/executors/threadDump,null,AVAILABLE,@Spark}
41 2018-11-21 20:14:03 INFO  ContextHandler:781 - Started o.s.j.s.←
ServletContextHandler@499b2a5c{/executors/threadDump/json,null,AVAILABLE,@Spark}
42 2018-11-21 20:14:03 INFO  ContextHandler:781 - Started o.s.j.s.←
ServletContextHandler@596df867{/static,null,AVAILABLE,@Spark}
43 2018-11-21 20:14:03 INFO  ContextHandler:781 - Started o.s.j.s.←
ServletContextHandler@3e587920{/null,AVAILABLE,@Spark}
44 2018-11-21 20:14:03 INFO  ContextHandler:781 - Started o.s.j.s.←
ServletContextHandler@2ef8a8c3{/api,null,AVAILABLE,@Spark} 2018-11-21 20:14:03 INFO←
ContextHandler:781 - Started o.s.j.s.ServletContextHandler@7544a1e4{/jobs/job/←
kill,null,AVAILABLE,@Spark}
45 2018-11-21 20:14:03 INFO  ContextHandler:781 - Started o.s.j.s.←
ServletContextHandler@70e0accd{/stages/stage/kill,null,AVAILABLE,@Spark}

```



```

46 2018-11-21 20:14:03 INFO SparkUI:54 - Bound SparkUI to 0.0.0.0, and started at http://v0242p.host.s.uw.edu:4040
47 2018-11-21 20:14:03 INFO SparkContext:54 - Added JAR file:/home/wyxiao_css534/←
    spark_programs/MyClass.jar at spark://v0242p.host.s.uw.edu:33246/jars/MyClass.jar ←
    with timestamp 1542860043497
48 2018-11-21 20:14:03 INFO Executor:54 - Starting executor ID driver on host localhost
49 2018-11-21 20:14:03 INFO Utils:54 - Successfully started service 'org.apache.spark.←
    network.netty.NettyBlockTransferService' on port 41795.
50 2018-11-21 20:14:03 INFO NettyBlockTransferService:54 - Server created on v0242p.host←
    .s.uw.edu:41795
51 2018-11-21 20:14:03 INFO BlockManager:54 - Using org.apache.spark.storage.←
    RandomBlockReplicationPolicy for block replication policy
52 2018-11-21 20:14:03 INFO BlockManagerMaster:54 - Registering BlockManager ←
    BlockManagerId(driver, v0242p.host.s.uw.edu, 41795, None)
53 2018-11-21 20:14:03 INFO BlockManagerMasterEndpoint:54 - Registering block manager ←
    v0242p.host.s.uw.edu:41795 with 366.3 MB RAM, BlockManagerId(driver, v0242p.host.s.←
    uw.edu, 41795, None)
54 2018-11-21 20:14:03 INFO BlockManagerMaster:54 - Registered BlockManager ←
    BlockManagerId(driver, v0242p.host.s.uw.edu, 41795, None)
55 2018-11-21 20:14:03 INFO BlockManager:54 - Initialized BlockManager: BlockManagerId(←
    driver, v0242p.host.s.uw.edu, 41795, None)
56 2018-11-21 20:14:03 INFO ContextHandler:781 - Started o.s.j.s.←
    ServletContextHandler@6ce90bc5{/metrics/json,null,AVAILABLE,@Spark}
57 2018-11-21 20:14:04 INFO MemoryStore:54 - Block broadcast_0 stored as values in ←
    memory (estimated size 236.7 KB, free 366.1 MB)
58 2018-11-21 20:14:04 INFO MemoryStore:54 - Block broadcast_0_piece0 stored as bytes in←
    memory (estimated size 22.9 KB, free 366.0 MB)
59 2018-11-21 20:14:04 INFO BlockManagerInfo:54 - Added broadcast_0_piece0 in memory on ←
    v0242p.host.s.uw.edu:41795 (size: 22.9 KB, free: 366.3 MB)
60 2018-11-21 20:14:04 INFO SparkContext:54 - Created broadcast 0 from textFile at ←
    MyClass.java:20
61 2018-11-21 20:14:04 INFO FileInputFormat:249 - Total input paths to process : 1
62 2018-11-21 20:14:04 INFO SparkContext:54 - Starting job: count at MyClass.java:26
63 2018-11-21 20:14:04 INFO DAGScheduler:54 - Got job 0 (count at MyClass.java:26) with ←
    1 output partitions
64 2018-11-21 20:14:04 INFO DAGScheduler:54 - Final stage: ResultStage 0 (count at ←
    MyClass.java:26)
65 2018-11-21 20:14:04 INFO DAGScheduler:54 - Parents of final stage: List()
66 2018-11-21 20:14:04 INFO DAGScheduler:54 - Missing parents: List()
67 2018-11-21 20:14:04 INFO DAGScheduler:54 - Submitting ResultStage 0 (MapPartitionsRDD←
    [3] at flatMap at MyClass.java:26), which has no missing parents
68 2018-11-21 20:14:04 INFO MemoryStore:54 - Block broadcast_1 stored as values in ←
    memory (estimated size 4.3 KB, free 366.0 MB)
69 2018-11-21 20:14:04 INFO MemoryStore:54 - Block broadcast_1_piece0 stored as bytes in←
    memory (estimated size 2.5 KB, free 366.0 MB)
70 2018-11-21 20:14:04 INFO BlockManagerInfo:54 - Added broadcast_1_piece0 in memory on ←
    v0242p.host.s.uw.edu:41795 (size: 2.5 KB, free: 366.3 MB)
71 2018-11-21 20:14:04 INFO SparkContext:54 - Created broadcast 1 from broadcast at ←
    DAGScheduler.scala:1039
72 2018-11-21 20:14:04 INFO DAGScheduler:54 - Submitting 1 missing tasks from ←
    ResultStage 0 (MapPartitionsRDD[3] at flatMap at MyClass.java:26) (first 15 tasks ←
    are for partitions Vector(0))
73 2018-11-21 20:14:04 INFO TaskSchedulerImpl:54 - Adding task set 0.0 with 1 tasks
74 2018-11-21 20:14:04 INFO TaskSetManager:54 - Starting task 0.0 in stage 0.0 (TID 0, ←
    localhost, executor driver, partition 0, PROCESS_LOCAL, 7897 bytes)
75 2018-11-21 20:14:04 INFO Executor:54 - Running task 0.0 in stage 0.0 (TID 0)
76 2018-11-21 20:14:04 INFO Executor:54 - Fetching spark://v0242p.host.s.uw.edu:33246/←
    jars/MyClass.jar with timestamp 1542860043497
77 2018-11-21 20:14:04 INFO TransportClientFactory:267 - Successfully created connection←
    to v0242p.host.s.uw.edu/172.28.203.196:33246 after 25 ms (0 ms spent in bootstraps←
    )
78 2018-11-21 20:14:04 INFO Utils:54 - Fetching spark://v0242p.host.s.uw.edu:33246/jars/←
    MyClass.jar to /tmp/spark-9f7bf541-b76b-4c3c-b7f3-a2e1e79d1f56/userFiles-0a84da4c-←
    a028-481b-8821-9493efcb083e/fetchFileTemp1534358842930517119.tmp
79 2018-11-21 20:14:04 INFO Executor:54 - Adding file:/tmp/spark-9f7bf541-b76b-4c3c-b7f3←
    -a2e1e79d1f56/userFiles-0a84da4c-a028-481b-8821-9493efcb083e/MyClass.jar to class ←
    loader
80 2018-11-21 20:14:04 INFO HadoopRDD:54 - Input split: file:/home/wyxiao_css534/←
    spark_programs/sample.txt:0+66
81 2018-11-21 20:14:04 INFO Executor:54 - Finished task 0.0 in stage 0.0 (TID 0). 875 ←
    bytes result sent to driver

```

```

82 2018-11-21 20:14:04 INFO TaskSetManager:54 - Finished task 0.0 in stage 0.0 (TID 0) ←
    in 216 ms on localhost (executor driver) (1/1)
83 2018-11-21 20:14:04 INFO TaskSchedulerImpl:54 - Removed TaskSet 0.0, whose tasks have←
    all completed, from pool
84 2018-11-21 20:14:04 INFO DAGScheduler:54 - ResultStage 0 (count at MyClass.java:26) ←
    finished in 0.309 s
85 2018-11-21 20:14:04 INFO DAGScheduler:54 - Job 0 finished: count at MyClass.java:26, ←
    took 0.358323 s
86 #words = 11
87 2018-11-21 20:14:04 INFO SparkContext:54 - Starting job: max at MyClass.java:27
88 2018-11-21 20:14:04 INFO DAGScheduler:54 - Got job 1 (max at MyClass.java:27) with 1 ←
    output partitions
89 2018-11-21 20:14:04 INFO DAGScheduler:54 - Final stage: ResultStage 1 (max at MyClass←
    .java:27)
90 2018-11-21 20:14:04 INFO DAGScheduler:54 - Parents of final stage: List()
91 2018-11-21 20:14:04 INFO DAGScheduler:54 - Missing parents: List()
92 2018-11-21 20:14:04 INFO DAGScheduler:54 - Submitting ResultStage 1 (←
    ParallelCollectionRDD[2] at parallelize at MyClass.java:22), which has no missing ←
    parents
93 2018-11-21 20:14:04 INFO MemoryStore:54 - Block broadcast_2 stored as values in ←
    memory (estimated size 1864.0 B, free 366.0 MB)
94 2018-11-21 20:14:04 INFO MemoryStore:54 - Block broadcast_2_piece0 stored as bytes in←
    memory (estimated size 1236.0 B, free 366.0 MB)
95 2018-11-21 20:14:04 INFO BlockManagerInfo:54 - Added broadcast_2_piece0 in memory on ←
    v0242p.host.s.uw.edu:41795 (size: 1236.0 B, free: 366.3 MB)
96 2018-11-21 20:14:04 INFO SparkContext:54 - Created broadcast 2 from broadcast at ←
    DAGScheduler.scala:1039
97 2018-11-21 20:14:04 INFO DAGScheduler:54 - Submitting 1 missing tasks from ←
    ResultStage 1 (ParallelCollectionRDD[2] at parallelize at MyClass.java:22) (first ←
    15 tasks are for partitions Vector(0))
98 2018-11-21 20:14:04 INFO TaskSchedulerImpl:54 - Adding task set 1.0 with 1 tasks
99 2018-11-21 20:14:04 INFO TaskSetManager:54 - Starting task 0.0 in stage 1.0 (TID 1, ←
    localhost, executor driver, partition 0, PROCESS_LOCAL, 7910 bytes)
100 2018-11-21 20:14:04 INFO Executor:54 - Running task 0.0 in stage 1.0 (TID 1)
101 2018-11-21 20:14:04 INFO Executor:54 - Finished task 0.0 in stage 1.0 (TID 1). 824 ←
    bytes result sent to driver
102 2018-11-21 20:14:04 INFO TaskSetManager:54 - Finished task 0.0 in stage 1.0 (TID 1) ←
    in 16 ms on localhost (executor driver) (1/1)
103 2018-11-21 20:14:04 INFO TaskSchedulerImpl:54 - Removed TaskSet 1.0, whose tasks have←
    all completed, from pool
104 2018-11-21 20:14:04 INFO DAGScheduler:54 - ResultStage 1 (max at MyClass.java:27) ←
    finished in 0.025 s
105 2018-11-21 20:14:04 INFO DAGScheduler:54 - Job 1 finished: max at MyClass.java:27, ←
    took 0.028227 s
106 max = 5
107 2018-11-21 20:14:04 INFO AbstractConnector:318 - Stopped Spark@5dda6f9{HTTP/1.1,[http←
    /1.1]}{0.0.0.0:4040}
108 2018-11-21 20:14:04 INFO SparkUI:54 - Stopped Spark web UI at http://v0242p.host.s.uw←
    .edu:4040
109 2018-11-21 20:14:04 INFO MapOutputTrackerMasterEndpoint:54 - ←
    MapOutputTrackerMasterEndpoint stopped!
110 2018-11-21 20:14:04 INFO MemoryStore:54 - MemoryStore cleared
111 2018-11-21 20:14:04 INFO BlockManager:54 - BlockManager stopped
112 2018-11-21 20:14:04 INFO BlockManagerMaster:54 - BlockManagerMaster stopped
113 2018-11-21 20:14:04 INFO OutputCommitCoordinator$OutputCommitCoordinatorEndpoint:54 ←
    OutputCommitCoordinator stopped!
114 2018-11-21 20:14:04 INFO SparkContext:54 - Successfully stopped SparkContext
115 2018-11-21 20:14:04 INFO ShutdownHookManager:54 - Shutdown hook called
116 2018-11-21 20:14:04 INFO ShutdownHookManager:54 - Deleting directory /tmp/spark-94←
    ef4c29-a6a1-4ff9-91da-613fdb927e0
117 2018-11-21 20:14:04 INFO ShutdownHookManager:54 - Deleting directory /tmp/spark-9←
    f7bf541-b76b-4c3c-b7f3-a2e1e79d1f56
118 [wyxiao_css534@cssmpi1 spark_programs]$

```

5.3 Lab4B

```
1 import org.apache.spark.SparkConf;
```

```

2 import org.apache.spark.api.java.JavaRDD;
3 import org.apache.spark.api.java.JavaSparkContext;
4 import org.apache.spark.api.java.function.FlatMapFunction;
5
6 import java.util.Arrays;
7 import java.util.Iterator;
8
9 public class JavaWordCountFlatMap
10 {
11
12     public static void main(String[] args)
13     {
14         // configure spark
15         SparkConf sparkConf = new SparkConf().setAppName("Java Word Count FlatMap")
16             .setMaster("local[2]").set("spark.executor.memory", "2g");
17         // start a spark context
18         JavaSparkContext sc = new JavaSparkContext(sparkConf);
19
20         // provide path to input text file
21         String path = "sample.txt";
22
23         // read text file to RDD
24         JavaRDD<String> lines = sc.textFile(path);
25
26         // Java 8 with lambdas: split the input string into words
27         JavaRDD<String> words = lines.flatMap(new FlatMapFunction<String, String>()
28         {
29             @Override
30             public Iterator<String> call(String s) throws Exception
31             {
32                 return Arrays.asList( s.split( " " ) ).iterator();
33             }
34         });
35
36         // print #words
37         System.out.println("#words = " + words.count());
38     }
39 }
40
41 }

```

5.4 Lab4B Output

#words = 11

```

1 [wyxiao_css534@cssmpi1 spark_programs]$ spark-submit --class JavaWordCountFlatMap --\
  master local JavaWordCountFlatMap.jar 2018-11-21 21:02:31 WARN NativeCodeLoader:62\
  - Unable to load native-hadoop library for your platform... using builtin-java \
  classes where applicable
2 2018-11-21 21:02:31 INFO SparkContext:54 - Running Spark version 2.3.1
3 2018-11-21 21:02:31 INFO SparkContext:54 - Submitted application: Java Word Count \
  FlatMap
4 2018-11-21 21:02:31 INFO SecurityManager:54 - Changing view acls to: wyxiao_css534
5 2018-11-21 21:02:31 INFO SecurityManager:54 - Changing modify acls to: wyxiao_css534
6 2018-11-21 21:02:31 INFO SecurityManager:54 - Changing view acls groups to:
7 2018-11-21 21:02:31 INFO SecurityManager:54 - Changing modify acls groups to:
8 2018-11-21 21:02:31 INFO SecurityManager:54 - SecurityManager: authentication \
  disabled; ui acls disabled; users with view permissions: Set(wyxiao_css534); \
  groups with view permissions: Set(); users with modify permissions: Set(\
  wyxiao_css534); groups with modify permissions: Set()
9 2018-11-21 21:02:31 INFO Utils:54 - Successfully started service 'sparkDriver' on \
  port 40558.
10 2018-11-21 21:02:31 INFO SparkEnv:54 - Registering MapOutputTracker
11 2018-11-21 21:02:31 INFO SparkEnv:54 - Registering BlockManagerMaster
12 2018-11-21 21:02:31 INFO BlockManagerMasterEndpoint:54 - Using org.apache.spark.\
  storage.DefaultTopologyMapper for getting topology information

```

```

13 2018-11-21 21:02:31 INFO BlockManagerMasterEndpoint:54 - BlockManagerMasterEndpoint ←
up
14 2018-11-21 21:02:31 INFO DiskBlockManager:54 - Created local directory at /tmp/←
blockmgr-b5047bb6-6a3e-4533-836a-6ce1f34a75a5
15 2018-11-21 21:02:31 INFO MemoryStore:54 - MemoryStore started with capacity 366.3 MB
16 2018-11-21 21:02:31 INFO SparkEnv:54 - Registering OutputCommitCoordinator
17 2018-11-21 21:02:31 INFO log:192 - Logging initialized @1648ms
18 2018-11-21 21:02:32 INFO Server:346 - jetty-9.3.z-SNAPSHOT
19 2018-11-21 21:02:32 INFO Server:414 - Started @1725ms
20 2018-11-21 21:02:32 INFO AbstractConnector:278 - Started ServerConnector@54afd745{←
HTTP/1.1,[http/1.1]}{0.0.0.0:4040}
21 2018-11-21 21:02:32 INFO Utils:54 - Successfully started service 'SparkUI' on port ←
4040.
22 2018-11-21 21:02:32 INFO ContextHandler:781 - Started o.s.j.s.←
ServletContextHandler@13579834{/jobs,null,AVAILABLE,@Spark}2018-11-21 21:02:32 INFO←
ContextHandler:781 - Started o.s.j.s.ServletContextHandler@71e9a896{/jobs/json,←
null,AVAILABLE,@Spark}
23 2018-11-21 21:02:32 INFO ContextHandler:781 - Started o.s.j.s.←
ServletContextHandler@6b9267b{/jobs/job,null,AVAILABLE,@Spark}
24 2018-11-21 21:02:32 INFO ContextHandler:781 - Started o.s.j.s.←
ServletContextHandler@29ad44e3{/jobs/job/json,null,AVAILABLE,@Spark}
25 2018-11-21 21:02:32 INFO ContextHandler:781 - Started o.s.j.s.←
ServletContextHandler@15bcf458{/stages,null,AVAILABLE,@Spark}
26 2018-11-21 21:02:32 INFO ContextHandler:781 - Started o.s.j.s.←
ServletContextHandler@5af9926a{/stages/json,null,AVAILABLE,@Spark}
27 2018-11-21 21:02:32 INFO ContextHandler:781 - Started o.s.j.s.←
ServletContextHandler@43c67247{/stages/stage,null,AVAILABLE,@Spark}
28 2018-11-21 21:02:32 INFO ContextHandler:781 - Started o.s.j.s.←
ServletContextHandler@649f2009{/stages/stage/json,null,AVAILABLE,@Spark}
29 2018-11-21 21:02:32 INFO ContextHandler:781 - Started o.s.j.s.←
ServletContextHandler@14bb2297{/stages/pool,null,AVAILABLE,@Spark}
30 2018-11-21 21:02:32 INFO ContextHandler:781 - Started o.s.j.s.←
ServletContextHandler@69adf72c{/stages/pool/json,null,AVAILABLE,@Spark}
31 2018-11-21 21:02:32 INFO ContextHandler:781 - Started o.s.j.s.←
ServletContextHandler@797501a{/storage,null,AVAILABLE,@Spark}
32 2018-11-21 21:02:32 INFO ContextHandler:781 - Started o.s.j.s.←
ServletContextHandler@1a15b789{/storage/json,null,AVAILABLE,@Spark}
33 2018-11-21 21:02:32 INFO ContextHandler:781 - Started o.s.j.s.←
ServletContextHandler@57f791c6{/storage/rdd,null,AVAILABLE,@Spark}
34 2018-11-21 21:02:32 INFO ContextHandler:781 - Started o.s.j.s.←
ServletContextHandler@51650883{/storage/rdd/json,null,AVAILABLE,@Spark}
35 2018-11-21 21:02:32 INFO ContextHandler:781 - Started o.s.j.s.←
ServletContextHandler@6c4f9535{/environment,null,AVAILABLE,@Spark}
36 2018-11-21 21:02:32 INFO ContextHandler:781 - Started o.s.j.s.←
ServletContextHandler@5bd1ceca{/environment/json,null,AVAILABLE,@Spark}
37 2018-11-21 21:02:32 INFO ContextHandler:781 - Started o.s.j.s.←
ServletContextHandler@30c31dd7{/executors,null,AVAILABLE,@Spark}
38 2018-11-21 21:02:32 INFO ContextHandler:781 - Started o.s.j.s.←
ServletContextHandler@499b2a5c{/executors/json,null,AVAILABLE,@Spark}
39 2018-11-21 21:02:32 INFO ContextHandler:781 - Started o.s.j.s.←
ServletContextHandler@596df867{/executors/threadDump,null,AVAILABLE,@Spark}
40 2018-11-21 21:02:32 INFO ContextHandler:781 - Started o.s.j.s.←
ServletContextHandler@c1fca1e{/executors/threadDump/json,null,AVAILABLE,@Spark}
41 2018-11-21 21:02:32 INFO ContextHandler:781 - Started o.s.j.s.←
ServletContextHandler@241a53ef{/static,null,AVAILABLE,@Spark}
42 2018-11-21 21:02:32 INFO ContextHandler:781 - Started o.s.j.s.←
ServletContextHandler@24f43aa3{/null,AVAILABLE,@Spark}
43 2018-11-21 21:02:32 INFO ContextHandler:781 - Started o.s.j.s.←
ServletContextHandler@63fd4873{/api,null,AVAILABLE,@Spark} 2018-11-21 21:02:32 INFO←
ContextHandler:781 - Started o.s.j.s.ServletContextHandler@7957dc72{/jobs/job/←
kill,null,AVAILABLE,@Spark}
44 2018-11-21 21:02:32 INFO ContextHandler:781 - Started o.s.j.s.←
ServletContextHandler@6ab72419{/stages/stage/kill,null,AVAILABLE,@Spark}
45 2018-11-21 21:02:32 INFO SparkUI:54 - Bound SparkUI to 0.0.0.0, and started at http:←
//v0242p.host.s.uw.edu:4040
46 2018-11-21 21:02:32 INFO SparkContext:54 - Added JAR file:/home/wyxiao_css534/←
spark_programs/JavaWordCountFlatMap.jar at spark://v0242p.host.s.uw.edu:40558/jars/←
JavaWordCountFlatMap.jar with timestamp 1542862952155
47 2018-11-21 21:02:32 INFO Executor:54 - Starting executor ID driver on host localhost
48 2018-11-21 21:02:32 INFO Utils:54 - Successfully started service 'org.apache.spark.←
network.netty.NettyBlockTransferService' on port 37979.

```

```

49 2018-11-21 21:02:32 INFO NettyBlockTransferService:54 - Server created on v0242p.host.s.uw.edu:37979
50 2018-11-21 21:02:32 INFO BlockManager:54 - Using org.apache.spark.storage.RandomBlockReplicationPolicy for block replication policy
51 2018-11-21 21:02:32 INFO BlockManagerMaster:54 - Registering BlockManager BlockManagerId(driver, v0242p.host.s.uw.edu, 37979, None)
52 2018-11-21 21:02:32 INFO BlockManagerMasterEndpoint:54 - Registering block manager v0242p.host.s.uw.edu:37979 with 366.3 MB RAM, BlockManagerId(driver, v0242p.host.s.uw.edu, 37979, None)
53 2018-11-21 21:02:32 INFO BlockManagerMaster:54 - Registered BlockManager BlockManagerId(driver, v0242p.host.s.uw.edu, 37979, None)
54 2018-11-21 21:02:32 INFO BlockManager:54 - Initialized BlockManager: BlockManagerId(driver, v0242p.host.s.uw.edu, 37979, None)
55 2018-11-21 21:02:32 INFO ContextHandler:781 - Started o.s.j.s.ServletContextHandler@1a5b8489{/metrics/json,null,AVAILABLE,@Spark}
56 2018-11-21 21:02:32 INFO MemoryStore:54 - Block broadcast_0 stored as values in memory (estimated size 236.7 KB, free 366.1 MB)
57 2018-11-21 21:02:33 INFO MemoryStore:54 - Block broadcast_0_piece0 stored as bytes in memory (estimated size 22.9 KB, free 366.0 MB)
58 2018-11-21 21:02:33 INFO BlockManagerInfo:54 - Added broadcast_0_piece0 in memory on v0242p.host.s.uw.edu:37979 (size: 22.9 KB, free: 366.3 MB)
59 2018-11-21 21:02:33 INFO SparkContext:54 - Created broadcast 0 from textFile at JavaWordCountFlatMap.java:24
60 2018-11-21 21:02:33 INFO FileInputFormat:249 - Total input paths to process : 1
61 2018-11-21 21:02:33 INFO SparkContext:54 - Starting job: count at JavaWordCountFlatMap.java:35
62 2018-11-21 21:02:33 INFO DAGScheduler:54 - Got job 0 (count at JavaWordCountFlatMap.java:35) with 1 output partitions
63 2018-11-21 21:02:33 INFO DAGScheduler:54 - Final stage: ResultStage 0 (count at JavaWordCountFlatMap.java:35)
64 2018-11-21 21:02:33 INFO DAGScheduler:54 - Parents of final stage: List()
65 2018-11-21 21:02:33 INFO DAGScheduler:54 - Missing parents: List()
66 2018-11-21 21:02:33 INFO DAGScheduler:54 - Submitting ResultStage 0 (MapPartitionsRDD[2] at flatMap at JavaWordCountFlatMap.java:27), which has no missing parents
67 2018-11-21 21:02:33 INFO MemoryStore:54 - Block broadcast_1 stored as values in memory (estimated size 3.8 KB, free 366.0 MB)
68 2018-11-21 21:02:33 INFO MemoryStore:54 - Block broadcast_1_piece0 stored as bytes in memory (estimated size 2.2 KB, free 366.0 MB)
69 2018-11-21 21:02:33 INFO BlockManagerInfo:54 - Added broadcast_1_piece0 in memory on v0242p.host.s.uw.edu:37979 (size: 2.2 KB, free: 366.3 MB)
70 2018-11-21 21:02:33 INFO SparkContext:54 - Created broadcast 1 from broadcast at DAGScheduler.scala:1039
71 2018-11-21 21:02:33 INFO DAGScheduler:54 - Submitting 1 missing tasks from ResultStage 0 (MapPartitionsRDD[2] at flatMap at JavaWordCountFlatMap.java:27) (first 15 tasks are for partitions Vector(0))
72 2018-11-21 21:02:33 INFO TaskSchedulerImpl:54 - Adding task set 0.0 with 1 tasks
73 2018-11-21 21:02:33 INFO TaskSetManager:54 - Starting task 0.0 in stage 0.0 (TID 0, localhost, executor driver, partition 0, PROCESS_LOCAL, 7897 bytes)
74 2018-11-21 21:02:33 INFO Executor:54 - Running task 0.0 in stage 0.0 (TID 0)
75 2018-11-21 21:02:33 INFO Executor:54 - Fetching spark://v0242p.host.s.uw.edu:40558/jars/JavaWordCountFlatMap.jar with timestamp 1542862952155
76 2018-11-21 21:02:33 INFO TransportClientFactory:267 - Successfully created connection to v0242p.host.s.uw.edu/172.28.203.196:40558 after 26 ms (0 ms spent in bootstraps)
77 2018-11-21 21:02:33 INFO Utils:54 - Fetching spark://v0242p.host.s.uw.edu:40558/jars/JavaWordCountFlatMap.jar to /tmp/spark-8becd335-04e5-4888-b961-1a19c4a5e0a1/userFiles-140397a7-8224-4238-acf5-bc870b789791/fetchFileTemp3321203663210588973.tmp
78 2018-11-21 21:02:33 INFO Executor:54 - Adding file:/tmp/spark-8becd335-04e5-4888-b961-1a19c4a5e0a1/userFiles-140397a7-8224-4238-acf5-bc870b789791/JavaWordCountFlatMap.jar to class loader
79 2018-11-21 21:02:33 INFO HadoopRDD:54 - Input split: file:/home/wyxiao_css534/spark_programs/sample.txt:0+66
80 2018-11-21 21:02:33 INFO Executor:54 - Finished task 0.0 in stage 0.0 (TID 0). 875 bytes result sent to driver
81 2018-11-21 21:02:33 INFO TaskSetManager:54 - Finished task 0.0 in stage 0.0 (TID 0) in 220 ms on localhost (executor driver) (1/1)
82 2018-11-21 21:02:33 INFO TaskSchedulerImpl:54 - Removed TaskSet 0.0, whose tasks have all completed, from pool
83 2018-11-21 21:02:33 INFO DAGScheduler:54 - ResultStage 0 (count at JavaWordCountFlatMap.java:35) finished in 0.314 s
84 2018-11-21 21:02:33 INFO DAGScheduler:54 - Job 0 finished: count at

```

```

      JavaWordCountFlatMap.java:35, took 0.359184 s
85 #words = 11
86 2018-11-21 21:02:33 INFO  SparkContext:54 - Invoking stop() from shutdown hook
87 2018-11-21 21:02:33 INFO  AbstractConnector:318 - Stopped Spark@54afd745{HTTP/1.1,[↵
      http/1.1]}{0.0.0.0:4040}
88 2018-11-21 21:02:33 INFO  SparkUI:54 - Stopped Spark web UI at http://v0242p.host.s.uw↵
      .edu:4040
89 2018-11-21 21:02:33 INFO  MapOutputTrackerMasterEndpoint:54 - ↵
      MapOutputTrackerMasterEndpoint stopped!
90 2018-11-21 21:02:33 INFO  MemoryStore:54 - MemoryStore cleared
91 2018-11-21 21:02:33 INFO  BlockManager:54 - BlockManager stopped
92 2018-11-21 21:02:33 INFO  BlockManagerMaster:54 - BlockManagerMaster stopped
93 2018-11-21 21:02:33 INFO  OutputCommitCoordinator$OutputCommitCoordinatorEndpoint:54 ↵
      OutputCommitCoordinator stopped!
94 2018-11-21 21:02:33 INFO  SparkContext:54 - Successfully stopped SparkContext
95 2018-11-21 21:02:33 INFO  ShutdownHookManager:54 - Shutdown hook called
96 2018-11-21 21:02:33 INFO  ShutdownHookManager:54 - Deleting directory /tmp/spark-8↵
      becd335-04e5-4888-b961-1a19c4a5e0a1
97 2018-11-21 21:02:33 INFO  ShutdownHookManager:54 - Deleting directory /tmp/spark-75↵
      dc5a29-2d6d-4ba3-9fdd-df843839578e
98 [wyxiao_css534@cssmpi1 spark_programs]$

```