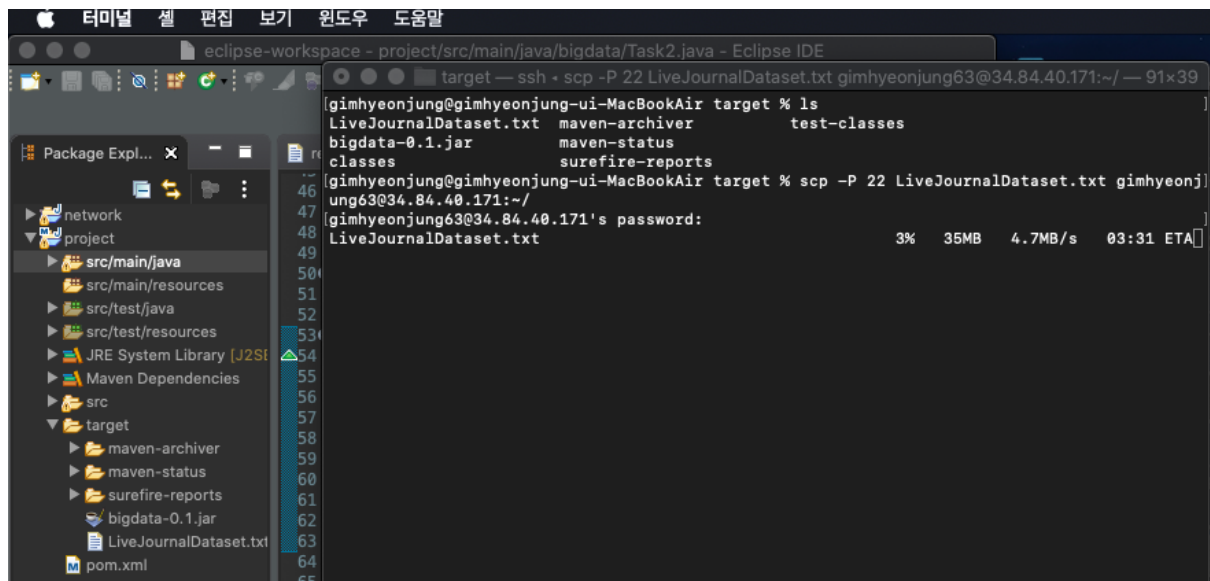


Graph Mining 보고서

20153167 김현중

1. LiveJournal Dataset HDFS 업로드 과정 및 결과

프로젝트의 target 폴더에 dataset인 LiveJournalDataset.txt를 다운로드해준다.
다운로드 해준 후 ssh 서버의 master node로 LiveJournalDataset.txt를 옮겨주는 작업을 진행 한다.



The screenshot shows the Eclipse IDE interface. On the left, the Package Explorer displays the project structure, including 'src/main/java', 'src/main/resources', 'src/test/java', 'src/test/resources', 'JRE System Library [J2SE 1.8]', 'Maven Dependencies', 'src', 'target', 'maven-archiver', 'maven-status', 'surefire-reports', 'bigdata-0.1.jar', 'LiveJournalDataset.txt', and 'pom.xml'. The 'target' folder is expanded, showing its contents. The main editor area displays a terminal window with the following commands and output:

```
target — ssh + scp -P 22 LiveJournalDataset.txt gimhyeonjung63@34.84.40.171:~/ — 91x39
gimhyeonjung@gimhyeonjung-ui-MacBookAir target % ls
LiveJournalDataset.txt  maven-archiver  test-classes
bigdata-0.1.jar         maven-status
classes                 surefire-reports
gimhyeonjung@gimhyeonjung-ui-MacBookAir target % scp -P 22 LiveJournalDataset.txt gimhyeonj
ung63@34.84.40.171:~/
[gimhyeonjung63@34.84.40.171's password:
LiveJournalDataset.txt                               3%  35MB  4.7MB/s  03:31 ETA]
```

ssh 서버를 실행시키게 되면 LiveJournalDataset.txt를 성공적으로 전송받은 것을 확인 할 수 있다. master node에 위치한 LiveJournalDataset.txt를 하둡파일시스템(hdfs)에 -put 명령어를 통해 업로드 해준다.

hdfs dfs -ls 명령어를 통해 hdfs에 데이터셋이 성공적으로 업로드 됐음을 확인할 수 있다.

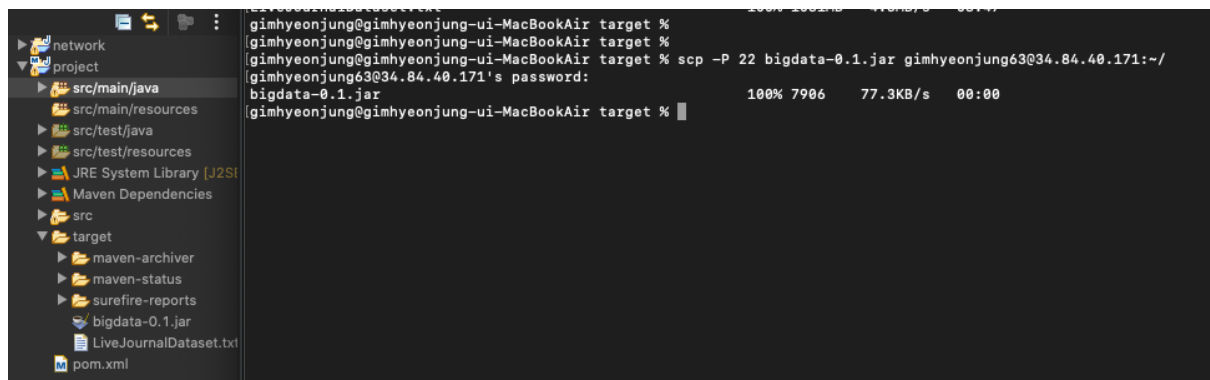
```

[gimhyeonjung63@kmucluster-m:~$ ls
bigdata-0.1.jar LiveJournalDataset.txt
[gimhyeonjung63@kmucluster-m:~$ hdfs dfs -ls
Found 1 items
drwxr-xr-x - gimhyeonjung63 hadoop 0 2020-06-26 05:23 .sparkStaging
[gimhyeonjung63@kmucluster-m:~$ hdfs dfs -put LiveJournalDataset.txt
[gimhyeonjung63@kmucluster-m:~$ hdfs dfs -ls
Found 2 items
drwxr-xr-x - gimhyeonjung63 hadoop 0 2020-06-26 05:23 .sparkStaging
-rw-r--r-- 2 gimhyeonjung63 hadoop 1080597849 2020-06-26 10:23 LiveJournalDataset.txt

```

2. Task1 실행 과정 및 결과

프로젝트의 RunAs -> install maven을 통하여 target 폴더에 jar파일을 생성해준다.
 생성된 jar파일을 ssh 서버의 master node로 옮겨주는 작업을 실행한다



```

gimhyeonjung@gimhyeonjung-ui-MacBookAir target %
gimhyeonjung@gimhyeonjung-ui-MacBookAir target %
gimhyeonjung@gimhyeonjung-ui-MacBookAir target % scp -P 22 bigdata-0.1.jar gimhyeonjung63@34.84.40.171:~/
gimhyeonjung63@34.84.40.171's password:
bigdata-0.1.jar
gimhyeonjung@gimhyeonjung-ui-MacBookAir target %
100% 7906 77.3KB/s 00:00

```

ssh 서버에 jar 파일을 확인할 수 있다.

```

[gimhyeonjung63@kmucluster-m:~$ ls
bigdata-0.1.jar LiveJournalDataset.txt
[gimhyeonjung63@kmucluster-m:~$ hdfs dfs -ls
Found 4 items
drwxr-xr-x - gimhyeonjung63 hadoop 0 2020-06-26 05:23 .sparkStaging
-rw-r--r-- 2 gimhyeonjung63 hadoop 1080597849 2020-06-26 10:23 LiveJournalDataset.txt

```

입력파일 명(LiveJournalDataset.txt)을 인자로 주고 jar 파일의 task1을 실행한다.

```

[gimhyeonjung63@kmucluster-m:~$ hadoop jar bigdata-0.1.jar bigdata.Task1 -Dmapreduce.job.reduces=5 LiveJournalDataset.txt

```

```
gimhyeonjung@kmucluster:~$ hadoop jar bigdata-0.1.jar bigdata.Task1 -Dmapreduce.job.reduces=5 LiveJournalDataset.txt
20/06/26 10:24:37 INFO client.RMProxy: Connecting to ResourceManager at kmucluster-m/10.146.0.2:8032
20/06/26 10:24:37 INFO client.AHSProxy: Connecting to Application History server at kmucluster-m/10.146.0.2:10200
20/06/26 10:24:38 INFO input.FileInputFormat: Total input files to process : 1
20/06/26 10:24:38 INFO mapreduce.JobSubmitter: number of splits:8
20/06/26 10:24:38 INFO Configuration.deprecation: yarn.resourcemanager.system-metrics-publisher.enabled is deprecated. Instead, use yarn.system-metrics-publisher.enabled
20/06/26 10:24:38 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1591965586205_0014
20/06/26 10:24:39 INFO impl.YarnClientImpl: Submitted application application_1591965586205_0014
20/06/26 10:24:39 INFO mapreduce.Job: The url to track the job: http://kmucluster-m:8088/proxy/application_1591965586205_0014/
20/06/26 10:24:39 INFO mapreduce.Job: Running job: job_1591965586205_0014
20/06/26 10:24:47 INFO mapreduce.Job: Job job_1591965586205_0014 running in uber mode : false
20/06/26 10:25:08 INFO mapreduce.Job: map 0% reduce 0%
20/06/26 10:25:09 INFO mapreduce.Job: map 15% reduce 0%
20/06/26 10:25:09 INFO mapreduce.Job: map 18% reduce 0%
20/06/26 10:25:10 INFO mapreduce.Job: map 27% reduce 0%
20/06/26 10:25:11 INFO mapreduce.Job: map 42% reduce 0%
20/06/26 10:25:14 INFO mapreduce.Job: map 44% reduce 0%
20/06/26 10:25:15 INFO mapreduce.Job: map 47% reduce 0%
20/06/26 10:25:16 INFO mapreduce.Job: map 53% reduce 0%
20/06/26 10:25:17 INFO mapreduce.Job: map 63% reduce 0%
20/06/26 10:25:21 INFO mapreduce.Job: map 64% reduce 0%
20/06/26 10:25:22 INFO mapreduce.Job: map 67% reduce 0%
20/06/26 10:25:32 INFO mapreduce.Job: map 72% reduce 0%
20/06/26 10:25:35 INFO mapreduce.Job: map 75% reduce 0%
20/06/26 10:25:49 INFO mapreduce.Job: map 83% reduce 0%
20/06/26 10:25:50 INFO mapreduce.Job: map 88% reduce 0%
20/06/26 10:25:51 INFO mapreduce.Job: map 92% reduce 0%
20/06/26 10:25:54 INFO mapreduce.Job: map 100% reduce 0%
20/06/26 10:26:13 INFO mapreduce.Job: map 100% reduce 14%
20/06/26 10:26:14 INFO mapreduce.Job: map 100% reduce 14%
20/06/26 10:26:16 INFO mapreduce.Job: map 100% reduce 73%
20/06/26 10:26:19 INFO mapreduce.Job: map 100% reduce 77%
20/06/26 10:26:20 INFO mapreduce.Job: map 100% reduce 81%
20/06/26 10:26:22 INFO mapreduce.Job: map 100% reduce 85%
20/06/26 10:26:26 INFO mapreduce.Job: map 100% reduce 89%
20/06/26 10:26:28 INFO mapreduce.Job: map 100% reduce 93%
20/06/26 10:26:32 INFO mapreduce.Job: map 100% reduce 97%
20/06/26 10:26:33 INFO mapreduce.Job: map 100% reduce 100%
20/06/26 10:26:34 INFO mapreduce.Job: Job job_1591965586205_0014 completed successfully
20/06/26 10:26:34 INFO mapreduce.Job: Counters: 49
    File System Counters
      FILE: Number of bytes read=778739234
      FILE: Number of bytes written=1466187508
      FILE: Number of read operations=0
      FILE: Number of large read operations=0
      FILE: Number of write operations=0
      HDFS: Number of bytes read=1088627537
      HDFS: Number of bytes written=627700937
      HDFS: Number of read operations=49
      HDFS: Number of large read operations=0
      HDFS: Number of write operations=15
    Job Counters
      Launched map tasks=8
      Launched reduce tasks=5
      Data-local map tasks=8
      Total time spent by all maps in occupied slots (ms)=1854820
      Total time spent by all reduces in occupied slots (ms)=641164
      Total time spent by all map tasks (ms)=463795
      Total time spent by all reduce tasks (ms)=160291
      Total vcore-milliseconds taken by all map tasks=463795
      Total vcore-milliseconds taken by all reduce tasks=160291
      Total megabyte-milliseconds taken by all map tasks=949667840
      Total megabyte-milliseconds taken by all reduce tasks=328275968

-----
Map-Reduce Framework
  Map input records=68993773
  Map output records=68475391
  Map output bytes=547803128
  Map output materialized bytes=684754150
  Input split bytes=1016
  Combine input records=0
  Combine output records=0
  Reduce input groups=2704435
  Reduce shuffle bytes=684754150
  Reduce input records=68475391
  Reduce output records=42081237
  Spilled Records=146348158
  Shuffled Maps =40
  Failed Shuffles=0
  Merged Map outputs=40
  GC time elapsed (ms)=9864
  CPU time spent (ms)=421360
  Physical memory (bytes) snapshot=10622083072
  Virtual memory (bytes) snapshot=45663784960
  Total committed heap usage (bytes)=10065281024

Shuffle Errors
  BAD_ID=0
  CONNECTION=0
  IO_ERROR=0
  WRONG_LENGTH=0
  WRONG_MAP=0
  WRONG_REDUCE=0

File Input Format Counters
  Bytes Read=1088626521
File Output Format Counters
  Bytes Written=627700937
  Bytes Read=0
  Bytes Written=0
```

Yarn Resource manager를 통하여 세부적인 실행과정 확인 가능

Show 20 entries																	Search:		
ID	User	Name	Application Type	Queue	Application Priority	StartTime	FinishTime	State	FinalStatus	Running Containers	Allocated CPU VCores	Allocated Memory MB	Reserved CPU VCores	Reserved Memory MB	% of Queue	% of Cluster	Progress	Tracking UI	Blacklisted Nodes
application_1591965586205_0014	gimhyeonjung63	bigdata-0.1.jar	MAPREDUCE	default	0	Fri Jun 26 19:24:39 +0900	N/A	RUNNING	UNDEFINED	8	8	16384	0	0	88.9	88.9	<div></div>	ApplicationMaster	0

complete-agency-256107 > kmuccluster

complete-agency-256107 > kmuccluster

hadoop

MapReduce Application application_1591965586205_0014

Active Jobs

Show 20 entries

Job ID	Name	State	Map Progress	Maps Total	Maps Completed	Reduce Progress	Reduces Total	Reduces Completed
job_1591965586205_0014	bigdata-0.1.jar	RUNNING		8	0		5	0

Showing 1 to 1 of 1 entries

First Previous 1 Next Last

실행완료 되었을 시 UI

complete-agency-256107 > kmuccluster

hadoop

MapReduce Job job_1591965586205_0014

Job Overview

Job Name: bigdata-0.1.jar
 User Name: gimhyeonjung63
 Queue: default
 State: SUCCEEDED
 Uberized: false
 Submitted: Fri Jun 26 10:24:39 UTC 2020
 Started: Fri Jun 26 10:24:45 UTC 2020
 Finished: Fri Jun 26 10:26:32 UTC 2020
 Elapsed: 1 mins, 46 sec

Diagnostics:

Average Map Time: 57sec
 Average Shuffle Time: 5sec
 Average Merge Time: 75sec
 Average Reduce Time: 19sec

ApplicationMaster	Attempt Number	Start Time	Node	Logs
1		Fri Jun 26 10:24:40 UTC 2020	kmuccluster-w-1.asia-northeast1-a.c.complete-agency-256107.internal.8042	/gateway/default/yarn/logs

Task Type	Total	Complete
Map	8	8
Reduce	5	5

Attempt Type	Failed	Killed	Successful
Maps	0	0	8
Reduces	0	0	5

출력결과 확인

```
[gimhyeonjung63@kmcluster-m:~]$ hdfs dfs -ls Task1.txt
Found 6 items
-rw-r--r-- 2 gimhyeonjung63 hadoop          0 2020-06-26 10:26 Task1.txt/_SUCCESS
-rw-r--r-- 2 gimhyeonjung63 hadoop 124942915 2020-06-26 10:26 Task1.txt/part-r-00000
-rw-r--r-- 2 gimhyeonjung63 hadoop 125667038 2020-06-26 10:26 Task1.txt/part-r-00001
-rw-r--r-- 2 gimhyeonjung63 hadoop 125149683 2020-06-26 10:26 Task1.txt/part-r-00002
-rw-r--r-- 2 gimhyeonjung63 hadoop 125277593 2020-06-26 10:26 Task1.txt/part-r-00003
-rw-r--r-- 2 gimhyeonjung63 hadoop 126663708 2020-06-26 10:26 Task1.txt/part-r-00004
[gimhyeonjung63@kmcluster-m:~]$ hdfs dfs -cat Task1.txt/*
0      1
0      2
0      3
0      4
0      5
0      6
0      7
0      8
0      9
0     10
0     11
0     12
0     13
0     14
0     15
0     16
0     17
0     18
0     19
0     20
0     21
0     22
0     23
0     24
0     25
0     26
0     27
0     28
0     29
0     30
0     31
0     32
0     33
0     34
0     35
0     36
0     37
0     38
0     39
0     40
0     41
0     42
0     43
0     44
0     45
0     46
0     47
```

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

```
10225 220001
10225 938570
10225 1277562
10225 669184
10225 521965
10225 1287967
10225 1379429
10225 2636152
10225 2636153
10225 2636154
10225 778968
10225 2636155
10225 2636156
10225 2636157
10225 310001
10225 77336
10225 383839
10225 2636150
10225 2636151
10230 209843
10230 1172803
10230 194001
10230 2636172
10230 2636173
10230 2636174
10230 2636175
10230 292108
10230 2050423
10230 1157167
10230 1083743
10235 502535
10235 210816
10235 967693
10235 27909
10235 524814
10235 554113
10235 274703
10235 119820
10235 70033
10235 991642
10235 417043
10235 538013
10235 480657
10235 476305
10235 64535
10235 550544
10235 206618
10235 52505
10235 805782
10235 24603
10235 16284
10235 502554
10235 985104
10235 498200
10235 945582
10235 177315
```

.....

.....

.....

.....

.....

3. Task2 실행 과정 및 결과

입력파일 명(Task1.txt)을 인자로 주고 jar 파일의 task2를 실행한다.

```
(gimhyeonjung63@kmucluster-m:~)$ hadoop jar bigdata-0.1.jar bigdata.Task2 -Dmapreduce.job.reduces=5 Task1.txt
```

```

gimhyeonjung@kmucluster-m:~$ hadoop jar bigdata-0.1.jar bigdata.Task2 -Dmapreduce.job.reduces=5 Task1.txt
20/06/26 10:31:28 INFO client.RMPProxy: Connecting to ResourceManager at kmucluster-m/10.146.0.2:8032
20/06/26 10:31:29 INFO client.AMRProxy: Connecting to ApplicationHistory server at kmucluster-m/10.146.0.2:10200
20/06/26 10:31:29 INFO input.FileInputFormat: Total input files to process : 5
20/06/26 10:31:29 INFO mapreduce.JobSubmitter: number of splits:5
20/06/26 10:31:29 INFO Configuration.deprecation: yarn.resourcemanager.system-metrics-publisher.enabled is deprecated. Instead, use yarn.system-metrics-publisher.enabled
20/06/26 10:31:30 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1591965586205_0015
20/06/26 10:31:30 INFO impl.YarnClientImpl: Submitted application application_1591965586205_0015
20/06/26 10:31:30 INFO mapreduce.Job: The url to track the job: http://kmucluster-m:8080/proxy/application_1591965586205_0015/
20/06/26 10:31:30 INFO mapreduce.Job: Running job: job_1591965586205_0015
20/06/26 10:31:37 INFO mapreduce.Job: Job job_1591965586205_0015 running in uber mode : false
20/06/26 10:31:37 INFO mapreduce.Job: map 0% reduce 0%
20/06/26 10:31:57 INFO mapreduce.Job: map 16% reduce 0%
20/06/26 10:32:00 INFO mapreduce.Job: map 35% reduce 0%
20/06/26 10:32:06 INFO mapreduce.Job: map 48% reduce 0%
20/06/26 10:32:21 INFO mapreduce.Job: map 45% reduce 0%
20/06/26 10:32:27 INFO mapreduce.Job: map 58% reduce 0%
20/06/26 10:32:42 INFO mapreduce.Job: map 68% reduce 0%
20/06/26 10:32:45 INFO mapreduce.Job: map 61% reduce 0%
20/06/26 10:32:48 INFO mapreduce.Job: map 64% reduce 0%
20/06/26 10:32:51 INFO mapreduce.Job: map 73% reduce 0%
20/06/26 10:32:57 INFO mapreduce.Job: map 79% reduce 0%
20/06/26 10:32:59 INFO mapreduce.Job: map 88% reduce 0%
20/06/26 10:33:25 INFO mapreduce.Job: map 84% reduce 0%
20/06/26 10:33:31 INFO mapreduce.Job: map 89% reduce 0%
20/06/26 10:33:37 INFO mapreduce.Job: map 94% reduce 0%
20/06/26 10:33:39 INFO mapreduce.Job: map 96% reduce 0%
20/06/26 10:33:40 INFO mapreduce.Job: map 100% reduce 0%
20/06/26 10:33:58 INFO mapreduce.Job: map 100% reduce 16%
20/06/26 10:33:59 INFO mapreduce.Job: map 100% reduce 44%
20/06/26 10:34:01 INFO mapreduce.Job: map 100% reduce 72%
20/06/26 10:34:02 INFO mapreduce.Job: map 100% reduce 74%
20/06/26 10:34:05 INFO mapreduce.Job: map 100% reduce 82%
20/06/26 10:34:07 INFO mapreduce.Job: map 100% reduce 89%
20/06/26 10:34:11 INFO mapreduce.Job: map 100% reduce 94%
20/06/26 10:34:12 INFO mapreduce.Job: map 100% reduce 97%
20/06/26 10:34:13 INFO mapreduce.Job: map 100% reduce 100%
20/06/26 10:34:13 INFO mapreduce.Job: Job job_1591965586205_0015 completed successfully
20/06/26 10:34:13 INFO mapreduce.Job: Counters: 50

File System Counters
  FILE: Number of bytes read=1714139474
  FILE: Number of bytes written=257315625
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=627701572
  HDFS: Number of bytes written=49276650
  HDFS: Number of read operations=46
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=15

Job Counters
  Launched map tasks=5
  Launched reduce tasks=5
  Data-local map tasks=4
  Rack-local map tasks=1
  Total time spent by all maps in occupied slots (ms)=2079948
  Total time spent by all reduces in occupied slots (ms)=538536
  Total time spent by all map tasks (ms)=519987
  Total time spent by all reduce tasks (ms)=134634
  Total vcore-milliseconds taken by all map tasks=519987
  Total vcore-milliseconds taken by all reduce tasks=134634
  Total megabyte-milliseconds taken by all map tasks=1864933376
  Total megabyte-milliseconds taken by all reduce tasks=275738432

-----
Map-Reduce Framework
  Map input records=42851237
  Map output records=85782474
  Map output bytes=685619792
  Map output materialized bytes=857824890
  Input split bytes=635
  Combine input records=0
  Combine output records=0
  Reduce input groups=4846609
  Reduce shuffle bytes=857824890
  Reduce input records=85782474
  Reduce output records=4846609
  Spilled Records=257107422
  Shuffled Maps =25
  Failed Shuffles=0
  Merged Map outputs=25
  GC time elapsed (ms)=5555
  CPU time spent (ms)=475418
  Physical memory (bytes) snapshot=7285388096
  Virtual memory (bytes) snapshot=35063742464
  Total committed heap usage (bytes)=7828684928

Shuffle Errors
  BAD_ID=0
  CONNECTION=0
  IO_ERROR=0
  WRONG_LENGTH=0
  WRONG_MAP=0
  WRONG_REDUCE=0

File Input Format Counters
  Bytes Read=627700937
File Output Format Counters
  Bytes Written=49276650

```

Yarn Resource manager를 통하여 세부적인 실행과정 확인 가능

ID	User	Name	Application Type	Queue	Application Priority	StartTime	FinishTime	State	FinalStatus	Running Containers	Allocated CPU VCores	Allocated Memory MB	Reserved CPU VCores	Reserved Memory MB	% of Queue	% of Cluster	Progress	Tracking UI	Blacklisted Nodes
application_1591965586205_0015	gimhyeonjung63	bigdata-0.1.jar	MAPREDUCE	default	0	Fri Jun 26 19:31:30 +0900 2020	N/A	RUNNING	UNDEFINED	6	6	12288	0	0	66.7	66.7	<div></div>	ApplicationMaster	0

complete-agency-256107 > kmucuster

hadoop

MapReduce Application application_1591965586205_0015

Logged in as: dr.who

Cluster

Application

About Jobs

Tools

Active Jobs

Show 20 entries

Job ID	Name	State	Map Progress	Maps Total	Maps Completed	Reduce Progress	Reduces Total	Reduces Completed
job_1591965586205_0015	bigdata-0.1.jar	RUNNING		5	5		5	0

Showing 1 to 1 of 1 entries

First Previous 1 Next Last

실행완료 되었을 시 UI

complete-agency-256107 > kmucuster

hadoop

MapReduce Job job_1591965586205_0015

Logged in as: dr.who

Application

Job

Overview

Counters

Configuration

Map tasks

Reduce tasks

Tools

Job Overview

Job Name: bigdata-0.1.jar

User Name: gimhyeonjung63

Queue: default

Status: SUCCEEDED

Uberized: false

Submitted: Fri Jun 26 10:31:30 UTC 2020

Started: Fri Jun 26 10:31:36 UTC 2020

Finished: Fri Jun 26 10:34:11 UTC 2020

Elapsed: 2mins, 35sec

Diagnostics:

Average Map Time: 1mins, 43sec

Average Shuffle Time: 5sec

Average Merge Time: 8sec

Average Reduce Time: 12sec

ApplicationMaster	Attempt Number	Start Time	Node	Logs
1		Fri Jun 26 10:31:31 UTC 2020	kmucuster-w-0-asia-northeast1-a-c-complete-agency-256107.internal.8042	/gateway/default/yarn/logs

Task Type	Total	Complete
Map	5	5
Reduce	5	5
Attempt Type	Failed	Killed
Maps	0	5
Reduces	0	5

Show 20 entries Search:																		
ID	User	Name	Application Type	Queue	Application Priority	StartTime	FinishTime	State	FinalStatus	Running Containers	Allocated CPU VCoers	Allocated Memory MB	Reserved CPU VCoers	Reserved Memory MB	% of Queue	% of Cluster	Progress	Tracking UI
application_1591965586205_0015	gimhyeonjung63	bigdata-0.1.jar	MAPREDUCE	default	0	Fri Jun 26 19:31:30 +0900 2020	Fri Jun 26 19:34:11 +0900 2020	FINISHED	SUCCEEDED	N/A	N/A	N/A	N/A	N/A	0.0	0.0	<div></div>	History
																		0

출력결과 확인

```
gimhyeonjung63@kmucluster-m:~$ hdfs dfs -ls
Found 4 items
drwxr-xr-x  - gimhyeonjung63 hadoop          0 2020-06-26 05:23 .sparkStaging
-rw-r--r--  2 gimhyeonjung63 hadoop 1088597849 2020-06-26 10:23 LiveJournalDataset.txt
drwxr-xr-x  - gimhyeonjung63 hadoop          0 2020-06-26 10:26 Task1.txt
drwxr-xr-x  - gimhyeonjung63 hadoop          0 2020-06-26 10:34 Task2.txt
gimhyeonjung63@kmucluster-m:~$ hdfs dfs -ls Task2.txt
Found 6 items
-rw-r--r--  2 gimhyeonjung63 hadoop          0 2020-06-26 10:34 Task2.txt/_SUCCESS
-rw-r--r--  2 gimhyeonjung63 hadoop 9854888 2020-06-26 10:34 Task2.txt/part-r-00000
-rw-r--r--  2 gimhyeonjung63 hadoop 9855769 2020-06-26 10:34 Task2.txt/part-r-00001
-rw-r--r--  2 gimhyeonjung63 hadoop 9854846 2020-06-26 10:34 Task2.txt/part-r-00002
-rw-r--r--  2 gimhyeonjung63 hadoop 9855724 2020-06-26 10:34 Task2.txt/part-r-00003
-rw-r--r--  2 gimhyeonjung63 hadoop 9855423 2020-06-26 10:34 Task2.txt/part-r-00004
gimhyeonjung63@kmucluster-m:~$ hdfs dfs -cat Task2.txt/*
0      94
5      109
10     97
15     25
20     77
25     9
30     19
35     9
40     56
45     25
50     106
55     16
60     4
65     39
70     21
75     89
80     23
85     22
90     72
95     152
100    79
105    97
110    180
115    28
120    64
125    204
130    54
135    184
140    84
145    29
150    87
155    9
160    16
165    32
170    26
175    1
180    150
185    76
190    155
195    77
200    9
205    31
210    102
215    30
220    25
225    7
230    21
235    67
```

.....

.....

.....

.....

.....

.....

.....

.....

.....

```

8400 52
8405 58
8410 27
8415 12
8420 36
8425 43
8430 35
8435 50
8440 25
8445 8
8450 15
8455 99
8460 11
8465 50
8470 7
8475 67
8480 10
8485 28
8490 104
8495 7
8500 13
8505 1
8510 78
8515 110
8520 99
8525 9
8530 75
8535 1
8540 44
8545 22
8550 20
8555 8
8560 57
8565 157
8570 126
8575 61
8580 46
8585 42
8590 73
8595 32
8600 2
8605 66
8610 36
8615 60
8620 6
8625 3
8630 6
8635 121
8640 179
8645 17
8650 19
8655 11
8660 35
8665 2
8670 326
8675 859
8680 46
8685 596
8690 14
8695 246
8700 78
8705 93
8710 97

```

.....

.....

.....

.....

.....

.....

Task1, Task2 완료 후 hdfs 구조

```

gimhyeonjung63@kmucluster-m:~$ ls
bigdata-0.1.jar  LiveJournalDataset.txt
gimhyeonjung63@kmucluster-m:~$ hdfs dfs -ls
Found 4 items
drwxr-xr-x - gimhyeonjung63 hadoop 0 2020-06-26 05:23 .sparkStaging
-rw-r--r-- 2 gimhyeonjung63 hadoop 1080597849 2020-06-26 10:23 LiveJournalDataset.txt
drwxr-xr-x - gimhyeonjung63 hadoop 0 2020-06-26 10:26 Task1.txt
drwxr-xr-x - gimhyeonjung63 hadoop 0 2020-06-26 10:34 Task2.txt

```