Screenshot 1 — Google Cloud Platform Dataproc

Google Cloud Platform · WordCount

Search products and resources

Google account: Brendan Russ (bvr6@pitt.edu)

Dataproc

- Clusters
- Jobs
- Workflows
- Autoscaling policies
- Component exchange
- Notebooks

← cluster-ad0a   ⊞ SUBMIT JOB   ↻ REFRESH   🗑 DELETE   ☰ VIEW LOGS

ⓘ For PD-Standard without local SSDs, we strongly recommend provisioning 1TB or larger to ensure consistently high I/O performance. See https://cloud.google.com/compute/docs/disks/performance for information on disk I/O performance.

Name	cluster-ad0a
Cluster UUID	d5bafe40-2bb5-446d-b0ca-551ac079dc44
Type	Dataproc Cluster
Status	✓ Running

MONITORING   JOBS   VM INSTANCES   CONFIGURATION   WEB INTERFACES

Filter instances

	Name ↑	Role	
✓	cluster-ad0a-m	Master	SSH ▾
✓	cluster-ad0a-w-0	Worker	
✓	cluster-ad0a-w-1	Worker	

Equivalent REST

10:05 PM 11/9/2020

Screenshot 2 — SSH terminal

bvr6@cluster-ad0a-m: ~/CloudLeast5 - Google Chrome

ssh.cloud.google.com/projects/wordcount-293603/zones/us-east4-c/instances/cluster-ad0a-m?authuser=1&hl=en_US&projectNumber=109974008917&useAdminProxy=true

```
Connected, host fingerprint: ssh-rsa 0 D6:AF:1A:F5:0C:29:01:79:9F:47:D0:61:6E:39
:B2:14:0A:B5:DA:A3:9D:C3:38:27:E6:2D:7E:63:6E:59:25:AD
Linux cluster-ad0a-m 5.8.0-0.bpo.2-amd64 #1 SMP Debian 5.8.10-1~bpo10+1 (2020-09
-26) x86_64

The programs included with the Debian GNU/Linux system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*/copyright.

Debian GNU/Linux comes with ABSOLUTELY NO WARRANTY, to the extent
permitted by applicable law.
Last login: Tue Nov 10 01:43:03 2020 from 35.235.241.32
bvr6@cluster-ad0a-m:~$ ls
CloudLeast5  CloudWordCount
bvr6@cluster-ad0a-m:~$ cd CloudLeast5
bvr6@cluster-ad0a-m:~/CloudLeast5$ ls
CloudLeast5.jar  bin  cloudLeast5  lib  src  wordData
bvr6@cluster-ad0a-m:~/CloudLeast5$ hadoop jar CloudLeast5.jar wordData wordOutput
20/11/10 03:07:18 INFO client.RMProxy: Connecting to ResourceManager at cluster-ad0
a-m/10.150.0.4:8032
20/11/10 03:07:19 INFO client.AHSProxy: Connecting to Application History server at
 cluster-ad0a-m/10.150.0.4:10200
20/11/10 03:07:19 INFO input.FileInputFormat: Total input files to process : 3
20/11/10 03:07:19 WARN hdfs.DataStreamer: Caught exception
java.lang.InterruptedException
        at java.lang.Object.wait(Native Method)
        at java.lang.Thread.join(Thread.java:1252)
        at java.lang.Thread.join(Thread.java:1326)
        at org.apache.hadoop.hdfs.DataStreamer.closeResponder(DataStreamer.java:980
)
        at org.apache.hadoop.hdfs.DataStreamer.endBlock(DataStreamer.java:630)
        at org.apache.hadoop.hdfs.DataStreamer.run(DataStreamer.java:807)
20/11/10 03:07:19 INFO mapreduce.JobSubmitter: number of splits:3
20/11/10 03:07:20 INFO Configuration.deprecation: yarn.resourcemanager.system-metri
cs-publisher.enabled is deprecated. Instead, use yarn.system-metrics-publisher.enab
led
20/11/10 03:07:20 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_16036
77175143_0033
20/11/10 03:07:20 INFO impl.YarnClientImpl: Submitted application application_16036
77175143_0033
20/11/10 03:07:20 INFO mapreduce.Job: The url to track the job: http://cluster-ad0a
-m:8088/proxy/application_1603677175143_0033/
```

10:07 PM 11/9/2020

```
20/11/10 03:07:20 INFO mapreduce.Job: Running job: job_1603677175143_0033
20/11/10 03:07:27 INFO mapreduce.Job: Job job_1603677175143_0033 running in uber mo
de : false
20/11/10 03:07:27 INFO mapreduce.Job:  map 0% reduce 0%
20/11/10 03:07:40 INFO mapreduce.Job:  map 100% reduce 0%
20/11/10 03:07:48 INFO mapreduce.Job:  map 100% reduce 20%
20/11/10 03:07:49 INFO mapreduce.Job:  map 100% reduce 60%
20/11/10 03:07:51 INFO mapreduce.Job:  map 100% reduce 80%
20/11/10 03:07:52 INFO mapreduce.Job:  map 100% reduce 100%
20/11/10 03:07:52 INFO mapreduce.Job: Job job_1603677175143_0033 completed successfully
20/11/10 03:07:52 INFO mapreduce.Job: Counters: 50
        File System Counters
                FILE: Number of bytes read=1410
                FILE: Number of bytes written=1666421
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=1027
                HDFS: Number of bytes written=156
                HDFS: Number of read operations=34
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=15
        Job Counters
                Killed reduce tasks=1
                Launched map tasks=3
                Launched reduce tasks=5
                Data-local map tasks=3
                Total time spent by all maps in occupied slots (ms)=116224
                Total time spent by all reduces in occupied slots (ms)=143328
                Total time spent by all map tasks (ms)=29056
                Total time spent by all reduce tasks (ms)=35832
                Total vcore-milliseconds taken by all map tasks=29056
                Total vcore-milliseconds taken by all reduce tasks=35832
                Total megabyte-milliseconds taken by all map tasks=59506688
                Total megabyte-milliseconds taken by all reduce tasks=73383936
        Map-Reduce Framework
                Map input records=3
                Map output records=116
                Map output bytes=1148
                Map output materialized bytes=1470
                Input split bytes=345
                Combine input records=0
```

10:08 PM
11/9/2020

```
                Data-local map tasks=3
                Total time spent by all maps in occupied slots (ms)=116224
                Total time spent by all reduces in occupied slots (ms)=143328
                Total time spent by all map tasks (ms)=29056
                Total time spent by all reduce tasks (ms)=35832
                Total vcore-milliseconds taken by all map tasks=29056
                Total vcore-milliseconds taken by all reduce tasks=35832
                Total megabyte-milliseconds taken by all map tasks=59506688
                Total megabyte-milliseconds taken by all reduce tasks=73383936
        Map-Reduce Framework
                Map input records=3
                Map output records=116
                Map output bytes=1148
                Map output materialized bytes=1470
                Input split bytes=345
                Combine input records=0
                Combine output records=0
                Reduce input groups=58
                Reduce shuffle bytes=1470
                Reduce input records=116
                Reduce output records=19
                Spilled Records=232
                Shuffled Maps =15
                Failed Shuffles=0
                Merged Map outputs=15
                GC time elapsed (ms)=1740
                CPU time spent (ms)=5260
                Physical memory (bytes) snapshot=2534617088
                Virtual memory (bytes) snapshot=27758739456
                Total committed heap usage (bytes)=2008547328
        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=682
        File Output Format Counters
                Bytes Written=156
bvr6@cluster-ad0a-m:~/CloudLeast5$
```

10:08 PM
11/9/2020

```
ovr6@cluster-ad0a-m:~/CloudLeast5$ hadoop fs -getmerge wordOutput collectedResults
ovr6@cluster-ad0a-m:~/CloudLeast5$ cat collectedResults
yep     18
her     3
security        2
with    1
soup    3
renew   2
told    1
the     4
client  3
event   2
was     1
simple  4
suspend 3
should  2
world   1
enough  5
government      3
of      2
threaten        1
ovr6@cluster-ad0a-m:~/CloudLeast5$ █
```