IERG4330/ESTR4316/IEMS5730 Spring 2022
Homework #1 (K8s)

Release date: Jan 24, 2021

Due date: Feb 13, 2021 (Sun) 11:59pm.

The solution will be posted soon after the deadline. No late homework will be accepted!

Every Student MUST include the following statement, together with his/her signature in the submitted homework.

I declare that the assignment submitted on the Blackboard system is original except for source material explicitly acknowledged and that the same or related material has not been previously submitted for another course. I also acknowledge that I am aware of University policy and regulations on honesty in academic work, and of the disciplinary guidelines and procedures applicable to breaches of such policy and regulations, as contained in the website http://www.cuhk.edu.hk/policy/academichonesty/.

Signed (Student Chan) Date: 5-2-2022

Name Chan Kai Yin SID 1155124983

Submission notice:

- Submit your homework via the blackboard system
- Only the following students are required to submit this assignment:
  - Students who HAVE taken IERG4300/ESTR4300/IEMS5709
  - IERG4330/ ESTR431 students who have been granted the prerequisite exemption.

General homework policies:

A student may discuss the problems with others. However, the work a student turns in must be created COMPLETELY by oneself ALONE. A student may not share ANY written work or pictures, nor may one copy answers from any source other than one's own brain.

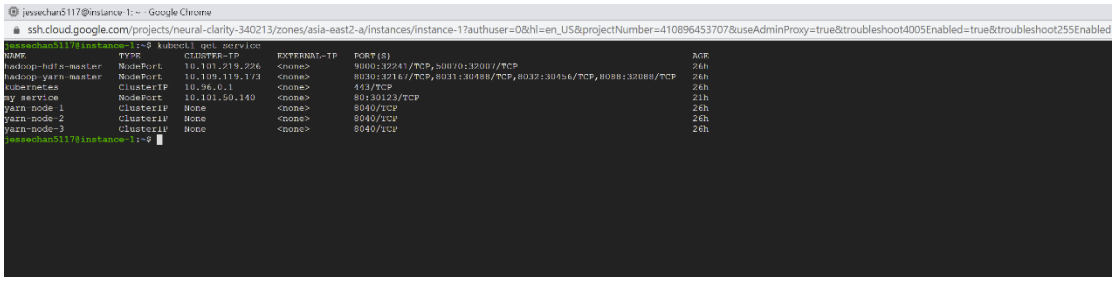
Each student **MUST LIST** on the homework paper the **name of every person he/she has discussed or worked with**. If the answer includes content from any other source, the student **MUST STATE THE SOURCE.** Failure to do so is cheating and will result in sanctions. Copying answers from someone else is cheating even if one lists their name(s) on the homework.

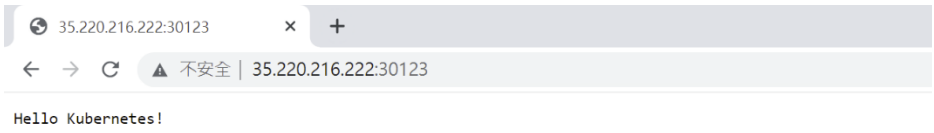
If there is information you need to solve a problem but the information is not stated in the problem, try to find the data somewhere. If you cannot find it, state what data you need, make a reasonable estimate of its value and justify any assumptions you make. You will be graded not only on whether your answer is correct but also on whether you have done an intelligent analysis.

Q1

a.)

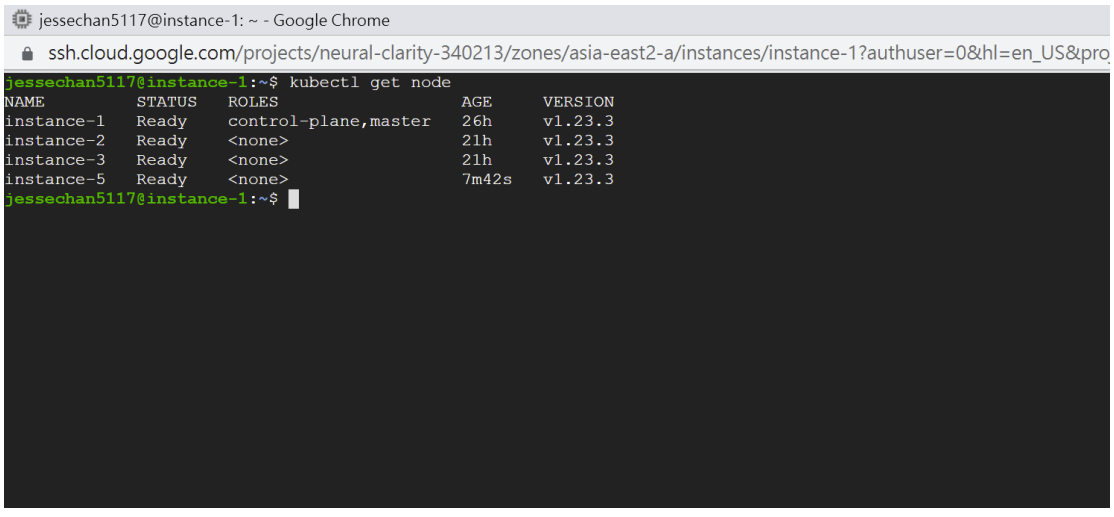
Applying the hello-world-demo.yaml





b.)

Multi-node Kubernetes Cluster Setup:



Applying the hadoop.yaml

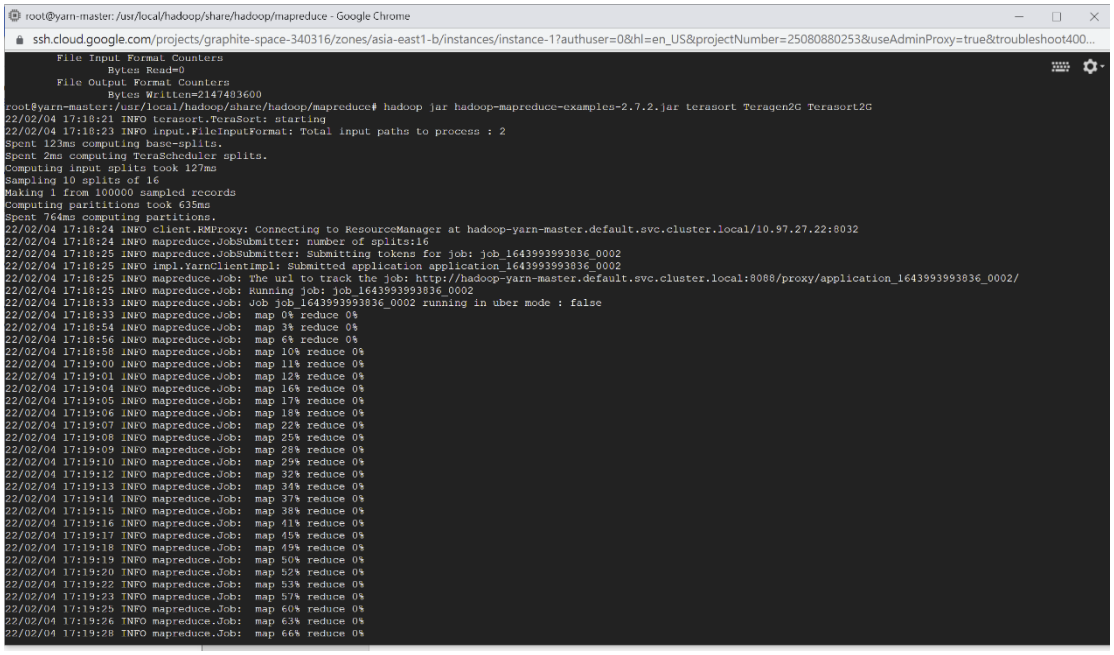
```
jessechan5117@instance-1:~$ kubectl get pod
NAME                READY   STATUS    RESTARTS   AGE
hadoop-datanode-1   1/1     Running   0          2m40s
hadoop-datanode-2   1/1     Running   0          2m40s
hadoop-datanode-3   1/1     Running   0          2m40s
hdfs-master         1/1     Running   0          2m40s
hello-world-9gcz4   1/1     Running   0          4m49s
hello-world-9tlf2   1/1     Running   0          4m59s
yarn-master         1/1     Running   0          2m40s
yarn-node-1         1/1     Running   0          2m40s
yarn-node-2         1/1     Running   0          2m40s
yarn-node-3         1/1     Running   0          2m40s
jessechan5117@instance-1:~$
```

2GB Teragen:

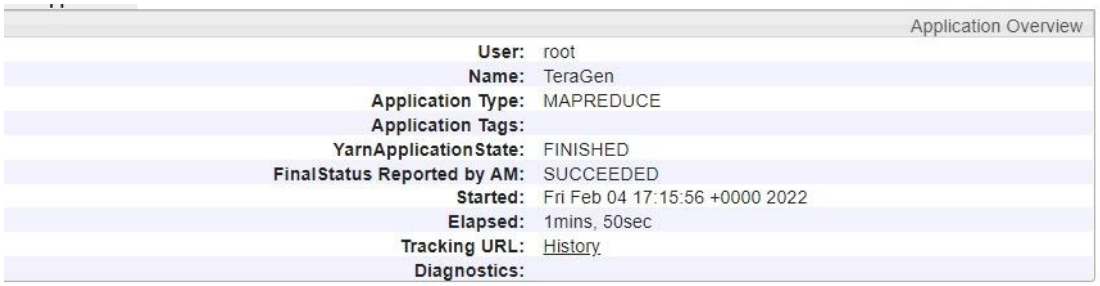
ssh.cloud.google.com/projects/neural-clarity-340213/zones/asia-east2-a/instances/instance-1?authuser=0&hl=en_US&projectNumber=410896453707&useAdmir

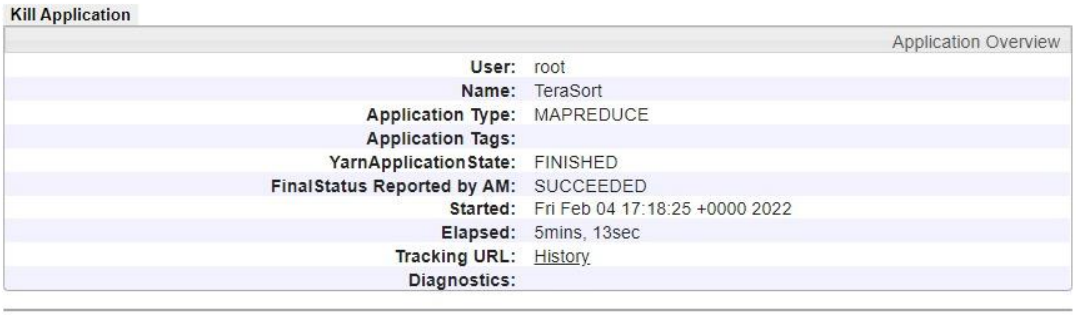
```
22/02/04 16:26:47 INFO mapreduce.JobSubmitter: number of splits:2
22/02/04 16:26:48 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1643991689410_0001
22/02/04 16:26:48 INFO impl.YarnClientImpl: Submitted application application_1643991689410_0001
22/02/04 16:26:48 INFO mapreduce.Job: The url to track the job: http://hadoop-yarn-master.default.svc.cluster.local:8088/proxy/ap
plication_1643991689410_0001/
22/02/04 16:26:48 INFO mapreduce.Job: Running job: job_1643991689410_0001
22/02/04 16:26:58 INFO mapreduce.Job: Job job_1643991689410_0001 running in uber mode : false
22/02/04 16:26:58 INFO mapreduce.Job:  map 0% reduce 0%
22/02/04 16:27:09 INFO mapreduce.Job:  map 9% reduce 0%
22/02/04 16:27:11 INFO mapreduce.Job:  map 17% reduce 0%
22/02/04 16:27:12 INFO mapreduce.Job:  map 21% reduce 0%
22/02/04 16:27:14 INFO mapreduce.Job:  map 27% reduce 0%
22/02/04 16:27:15 INFO mapreduce.Job:  map 31% reduce 0%
22/02/04 16:27:18 INFO mapreduce.Job:  map 37% reduce 0%
22/02/04 16:27:19 INFO mapreduce.Job:  map 42% reduce 0%
22/02/04 16:27:21 INFO mapreduce.Job:  map 47% reduce 0%
22/02/04 16:27:22 INFO mapreduce.Job:  map 50% reduce 0%
22/02/04 16:27:23 INFO mapreduce.Job:  map 54% reduce 0%
22/02/04 16:27:25 INFO mapreduce.Job:  map 55% reduce 0%
22/02/04 16:27:26 INFO mapreduce.Job:  map 58% reduce 0%
22/02/04 16:27:28 INFO mapreduce.Job:  map 64% reduce 0%
22/02/04 16:27:29 INFO mapreduce.Job:  map 69% reduce 0%
22/02/04 16:27:31 INFO mapreduce.Job:  map 73% reduce 0%
22/02/04 16:27:32 INFO mapreduce.Job:  map 75% reduce 0%
22/02/04 16:27:34 INFO mapreduce.Job:  map 76% reduce 0%
22/02/04 16:27:38 INFO mapreduce.Job:  map 77% reduce 0%
22/02/04 16:27:45 INFO mapreduce.Job:  map 78% reduce 0%
22/02/04 16:27:49 INFO mapreduce.Job:  map 80% reduce 0%
22/02/04 16:27:52 INFO mapreduce.Job:  map 81% reduce 0%
22/02/04 16:27:55 INFO mapreduce.Job:  map 84% reduce 0%
22/02/04 16:27:58 INFO mapreduce.Job:  map 86% reduce 0%
22/02/04 16:28:02 INFO mapreduce.Job:  map 87% reduce 0%
22/02/04 16:28:05 INFO mapreduce.Job:  map 90% reduce 0%
22/02/04 16:28:09 INFO mapreduce.Job:  map 91% reduce 0%
22/02/04 16:28:15 INFO mapreduce.Job:  map 94% reduce 0%
22/02/04 16:28:18 INFO mapreduce.Job:  map 96% reduce 0%
22/02/04 16:28:22 INFO mapreduce.Job:  map 98% reduce 0%
22/02/04 16:28:26 INFO mapreduce.Job:  map 99% reduce 0%
22/02/04 16:28:29 INFO mapreduce.Job:  map 100% reduce 0%
22/02/04 16:28:31 INFO mapreduce.Job: Job job_1643991689410_0001 completed successfully
22/02/04 16:28:32 INFO mapreduce.Job: Counters: 31
        File System Counters
                FILE: Number of bytes read=0
                FILE: Number of bytes written=235076
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=167
                HDFS: Number of bytes written=2147483600
                HDFS: Number of read operations=8
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=4
        Job Counters
                Launched map tasks=2
                Other local map tasks=2
                Total time spent by all maps in occupied slots (ms)=174982
                Total time spent by all reduces in occupied slots (ms)=0
                Total time spent by all map tasks (ms)=174982
                Total vcore-milliseconds taken by all map tasks=174982
                Total megabyte-milliseconds taken by all map tasks=179181568
        Map-Reduce Framework
                Map input records=21474836
                Map output records=21474836
                Input split bytes=167
                Spilled Records=0
                Failed Shuffles=0
                Merged Map outputs=0
                GC time elapsed (ms)=1744
                CPU time spent (ms)=44350
                Physical memory (bytes) snapshot=409341952
                Virtual memory (bytes) snapshot=1750532096
                Total committed heap usage (bytes)=257425408
        org.apache.hadoop.examples.terasort.TeraGen$Counters
                CHECKSUM=46124753271996946
        File Input Format Counters
                Bytes Read=0
        File Output Format Counters
                Bytes Written=2147483600
root@yarn-master:/usr/local/hadoop/share/hadoop/mapreduce#
```

2G TeraSort:



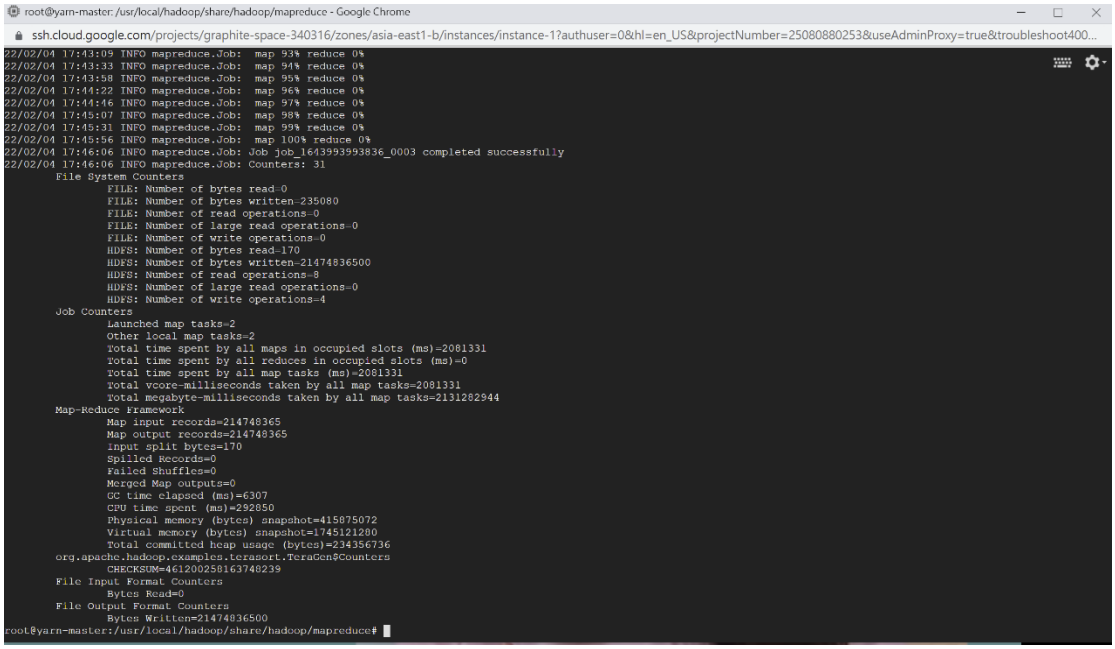
Time:



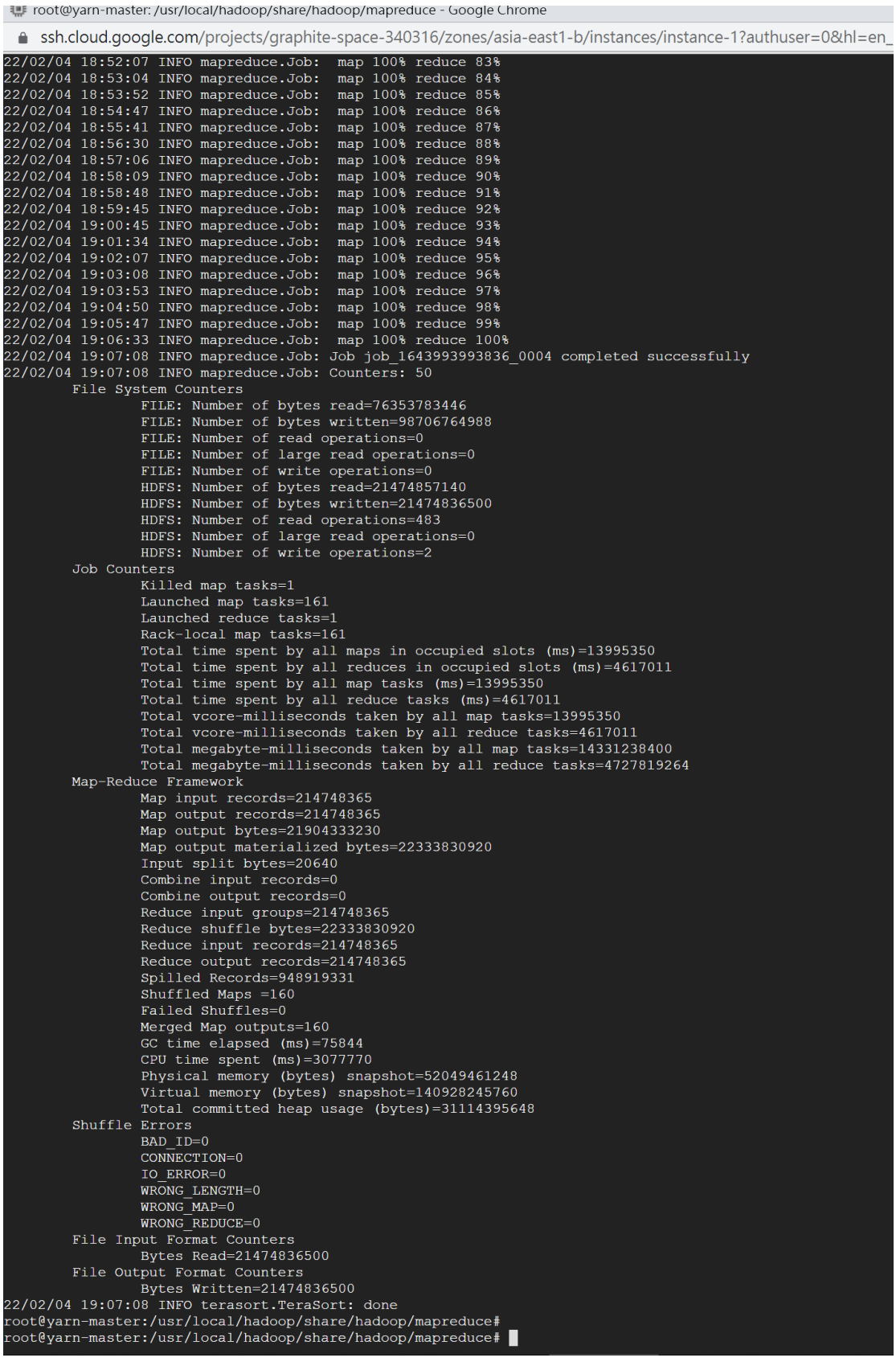


Total: ~7 mins

20G TeraGen:


root@yarn-master: /usr/local/hadoop/share/hadoop/mapreduce - Google Chrome
ssh.cloud.google.com/projects/graphite-space-340316/zones/asia-east1-b/instances/instance-1?authuser=0&hl=en_US&projectNumber=25080880253&useAdminProxy=true&troubleshoot400...
22/02/04 17:43:09 INFO mapreduce.Job: map 93% reduce 0%
22/02/04 17:43:33 INFO mapreduce.Job: map 94% reduce 0%
22/02/04 17:43:58 INFO mapreduce.Job: map 95% reduce 0%
22/02/04 17:44:22 INFO mapreduce.Job: map 96% reduce 0%
22/02/04 17:44:46 INFO mapreduce.Job: map 97% reduce 0%
22/02/04 17:45:07 INFO mapreduce.Job: map 98% reduce 0%
22/02/04 17:45:31 INFO mapreduce.Job: map 99% reduce 0%
22/02/04 17:45:56 INFO mapreduce.Job: map 100% reduce 0%
22/02/04 17:46:06 INFO mapreduce.Job: Job job_1643993993836_0003 completed successfully
22/02/04 17:46:06 INFO mapreduce.Job: Counters: 31
File System Counters
FILE: Number of bytes read=0
FILE: Number of bytes written=235080
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=170
HDFS: Number of bytes written=21474836500
HDFS: Number of read operations=8
HDFS: Number of large read operations=0
HDFS: Number of write operations=4
Job Counters
Launched map tasks=2
Other local map tasks=2
Total time spent by all maps in occupied slots (ms)=2081331
Total time spent by all reduces in occupied slots (ms)=0
Total time spent by all map tasks (ms)=2081331
Total vcore-milliseconds taken by all map tasks=2081331
Total megabyte-milliseconds taken by all map tasks=2131282944
Map-Reduce Framework
Map input records=214748365
Map output records=214748365
Input split bytes=170
Spilled Records=0
Failed Shuffles=0
Merged Map outputs=0
GC time elapsed (ms)=6307
CPU time spent (ms)=292850
Physical memory (bytes) snapshot=415875072
Virtual memory (bytes) snapshot=1745121280
Total committed heap usage (bytes)=234356736
org.apache.hadoop.examples.terasort.TeraGen$Counters
CHECKSUM=461200250163748239
File Input Format Counters
Bytes Read=0
File Output Format Counters
Bytes Written=21474836500
root@yarn-master:/usr/local/hadoop/share/hadoop/mapreduce#

20G TeraSort:



Time:

Kill Application

Application Overview	
User:	root
Name:	TeraGen
Application Type:	MAPREDUCE
Application Tags:	
YarnApplicationState:	FINISHED
FinalStatus Reported by AM:	SUCCEEDED
Started:	Fri Feb 04 17:25:44 +0000 2022
Elapsed:	20mins, 19sec
Tracking URL:	History
Diagnostics:	

Kill Application

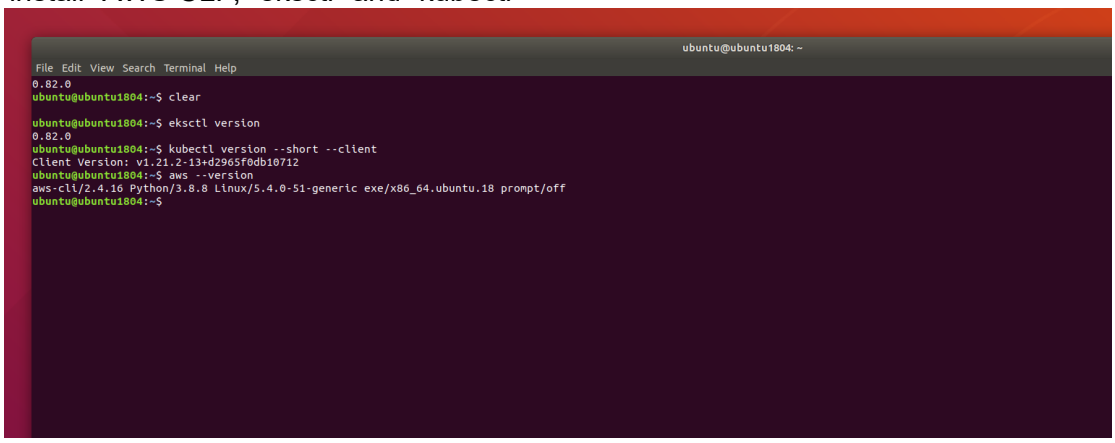
Application Overview	
User:	root
Name:	TeraSort
Application Type:	MAPREDUCE
Application Tags:	
YarnApplicationState:	FINISHED
FinalStatus Reported by AM:	SUCCEEDED
Started:	Fri Feb 04 17:48:31 +0000 2022
Elapsed:	1hrs, 18mins, 34sec
Tracking URL:	History
Diagnostics:	

Application Metrics

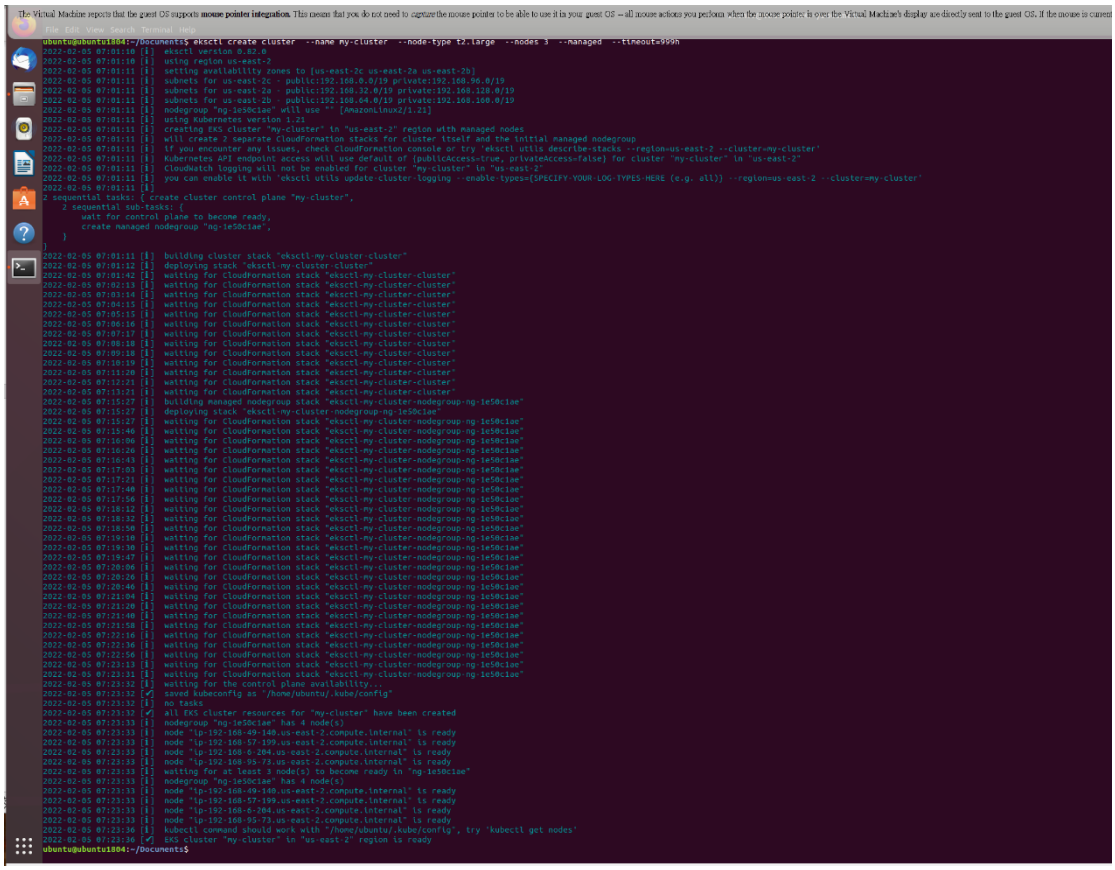
Total: ~1hr 38mins

c.)

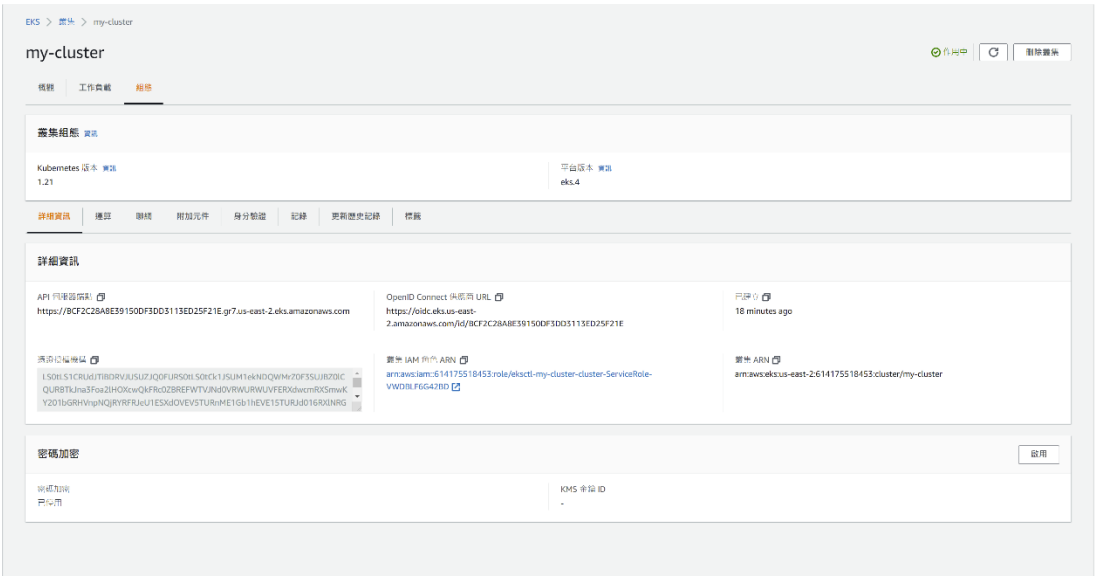
install “AWS CLI”, “eksctl” and “kubectl”

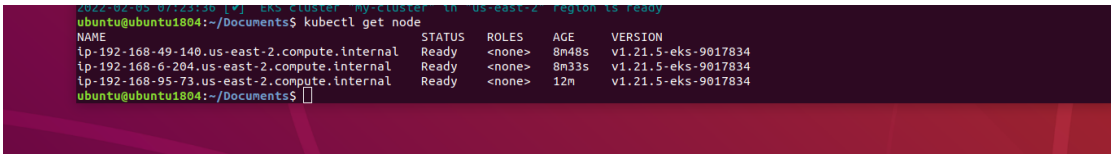


Create a serverless Kubernetes cluster in Amazon EKS:

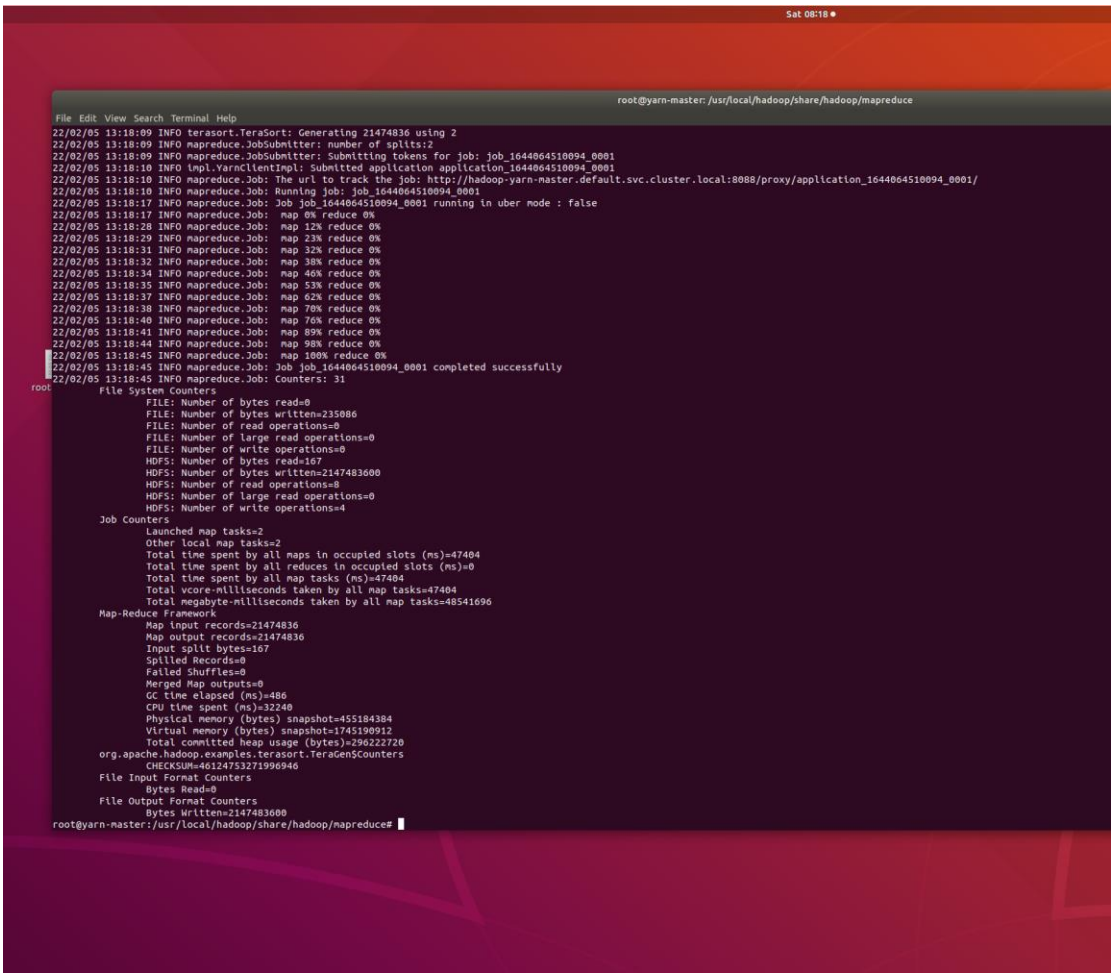


EKS:

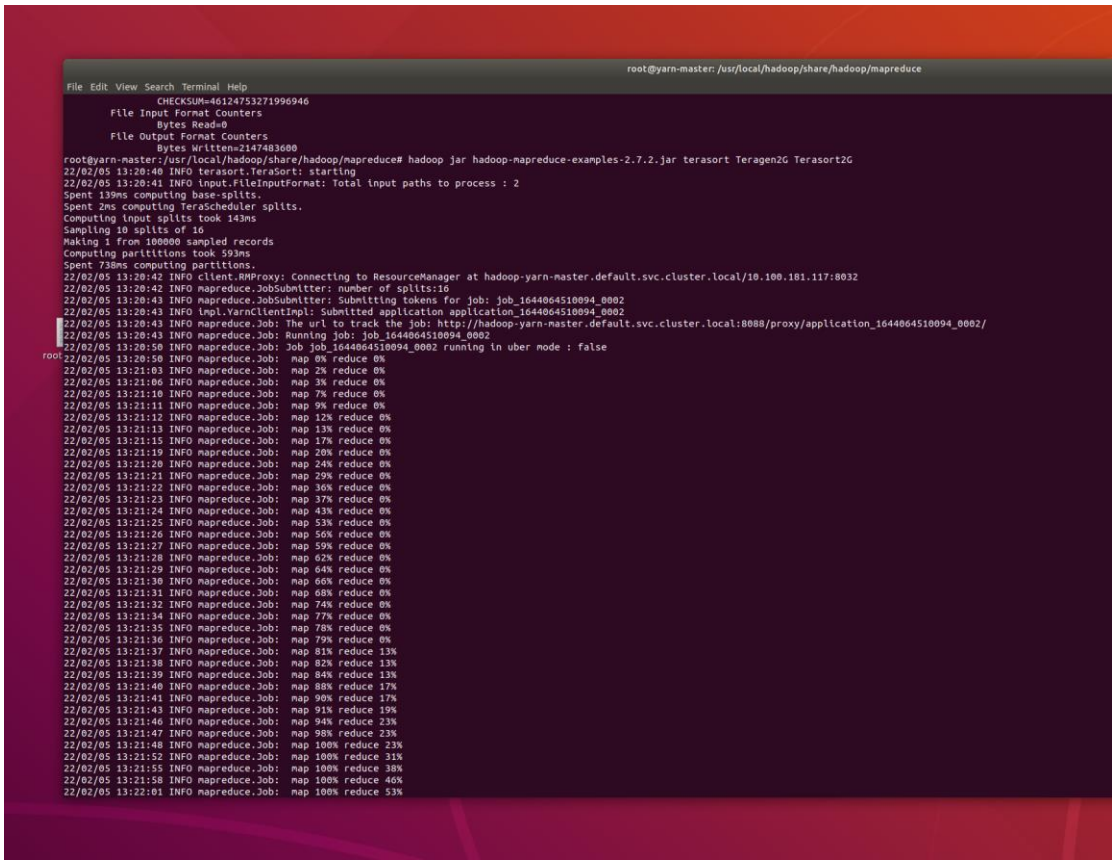

EKS > 叢集 > my-cluster
my-cluster
概觀
工作負載
組態
叢集組態 資訊
Kubernetes 版本 資訊
1.21
平台版本 資訊
eks.4
詳細資訊
連接
聯絡
附加元件
身分驗證
記錄
更新歷史記錄
標籤
詳細資訊
API 伺服器端點
https://BCF2C28A8E39150DF3DD3113ED25F21E.gr7.us-east-2.eks.amazonaws.com
OpenID Connect 供應商 URL
https://oidc.eks.us-east-2.amazonaws.com/id/BCF2C28A8E39150DF3DD3113ED25F21E
18 minutes ago
叢集 IAM 角色 ARN
arn:aws:iam::614175518453:role/eksctl-my-cluster-cluster-ServiceRole-VWD8LF6G42BD
叢集 ARN
arn:aws:eks:us-east-2:614175518453:cluster/my-cluster
密碼加密
啟用
KMS 金鑰 ID


ubuntu@ubuntu1804:~/Documents$ kubectl get node
NAME STATUS ROLES AGE VERSION
ip-192-168-49-140.us-east-2.compute.internal Ready <none> 8m48s v1.21.5-eks-9017834
ip-192-168-6-204.us-east-2.compute.internal Ready <none> 8m33s v1.21.5-eks-9017834
ip-192-168-95-73.us-east-2.compute.internal Ready <none> 12m v1.21.5-eks-9017834
ubuntu@ubuntu1804:~/Documents$

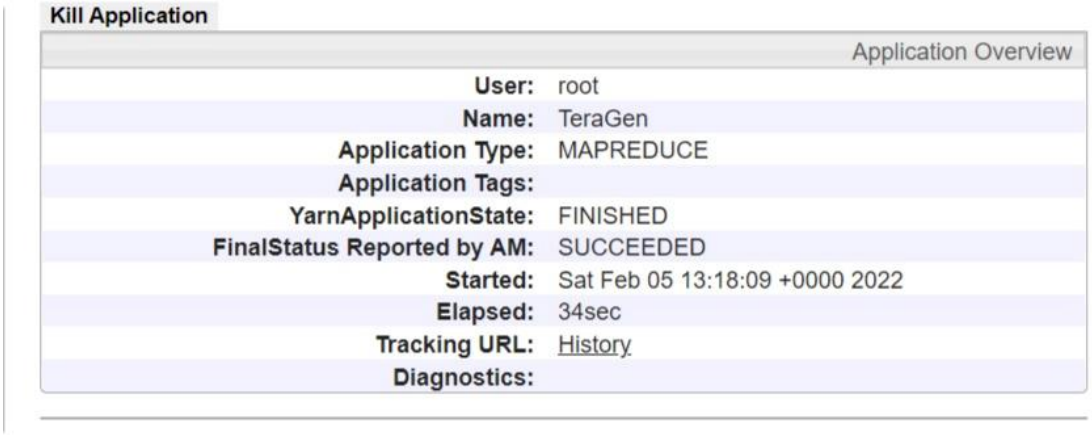
TeraGen 2G:

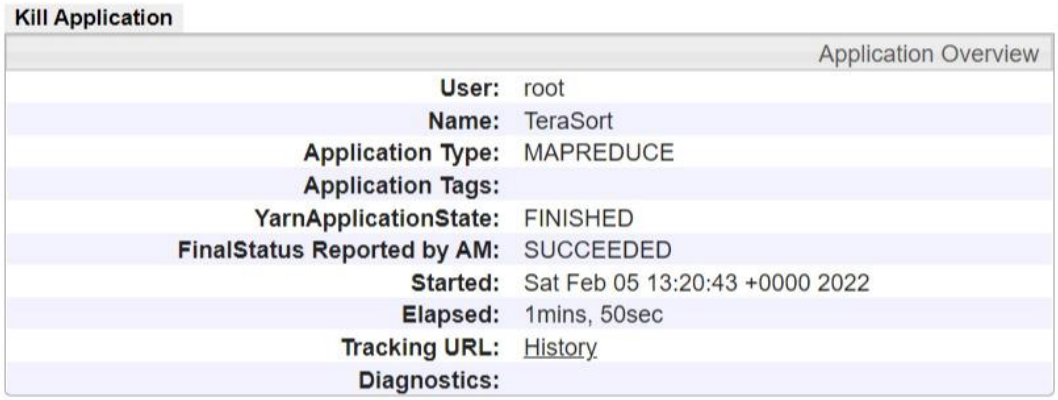

Sat 08:18
root@yarn-master: /usr/local/hadoop/share/hadoop/mapreduce
File Edit View Search Terminal Help
22/02/05 13:18:09 INFO terasort.TeraSort: Generating 21474836 using 2
22/02/05 13:18:09 INFO mapreduce.JobSubmitter: number of splits:2
22/02/05 13:18:09 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1644064510094_0001
22/02/05 13:18:10 INFO impl.YarnClientImpl: Submitted application application_1644064510094_0001
22/02/05 13:18:10 INFO mapreduce.Job: The url to track the job: http://hadoop-yarn-master.default.svc.cluster.local:8088/proxy/application_1644064510094_0001/
22/02/05 13:18:10 INFO mapreduce.Job: Running job: job_1644064510094_0001
22/02/05 13:18:17 INFO mapreduce.Job: Job job_1644064510094_0001 running in uber mode : false
22/02/05 13:18:17 INFO mapreduce.Job: map 0% reduce 0%
22/02/05 13:18:28 INFO mapreduce.Job: map 12% reduce 0%
22/02/05 13:18:29 INFO mapreduce.Job: map 23% reduce 0%
22/02/05 13:18:31 INFO mapreduce.Job: map 32% reduce 0%
22/02/05 13:18:32 INFO mapreduce.Job: map 38% reduce 0%
22/02/05 13:18:34 INFO mapreduce.Job: map 46% reduce 0%
22/02/05 13:18:35 INFO mapreduce.Job: map 53% reduce 0%
22/02/05 13:18:37 INFO mapreduce.Job: map 62% reduce 0%
22/02/05 13:18:38 INFO mapreduce.Job: map 70% reduce 0%
22/02/05 13:18:40 INFO mapreduce.Job: map 76% reduce 0%
22/02/05 13:18:41 INFO mapreduce.Job: map 89% reduce 0%
22/02/05 13:18:44 INFO mapreduce.Job: map 98% reduce 0%
22/02/05 13:18:45 INFO mapreduce.Job: map 100% reduce 0%
22/02/05 13:18:45 INFO mapreduce.Job: Job job_1644064510094_0001 completed successfully
22/02/05 13:18:45 INFO mapreduce.Job: Counters: 31
File System Counters
FILE: Number of bytes read=0
FILE: Number of bytes written=235086
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=167
HDFS: Number of bytes written=2147483600
HDFS: Number of read operations=8
HDFS: Number of large read operations=0
HDFS: Number of write operations=4
Job Counters
Launched map tasks=2
Other local map tasks=2
Total time spent by all maps in occupied slots (ms)=47404
Total time spent by all reduces in occupied slots (ms)=0
Total time spent by all map tasks (ms)=47404
Total vcore-milliseconds taken by all map tasks=47404
Total megabyte-milliseconds taken by all map tasks=48541696
Map-Reduce Framework
Map input records=21474836
Map output records=21474836
Input split bytes=167
Spilled Records=0
Failed Shuffles=0
Merged Map outputs=0
GC time elapsed (ms)=486
CPU time spent (ms)=32240
Physical memory (bytes) snapshot=455184384
Virtual memory (bytes) snapshot=1745190912
Total committed heap usage (bytes)=296222720
org.apache.hadoop.examples.terasort.TeraGen$Counters
CHECKSUM=46124753271996946
File Input Format Counters
Bytes Read=0
File Output Format Counters
Bytes Written=2147483600
root@yarn-master:/usr/local/hadoop/share/hadoop/mapreduce#

TeraSort2G:



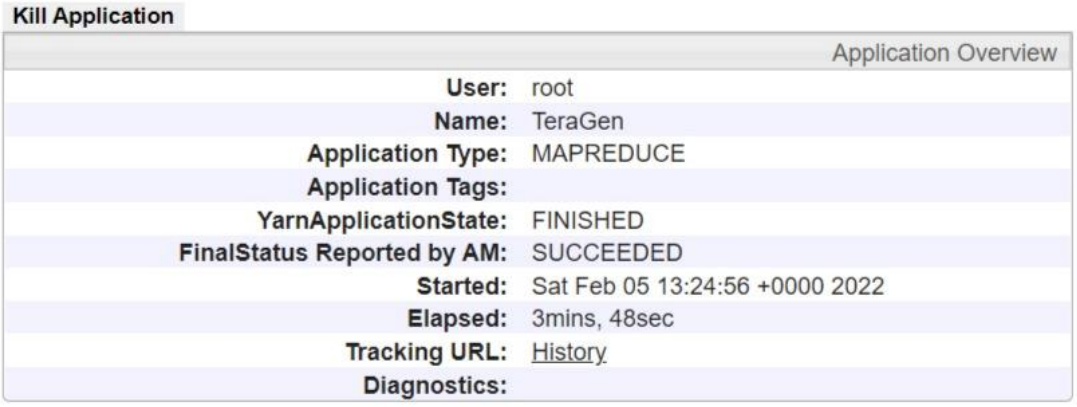
Time:

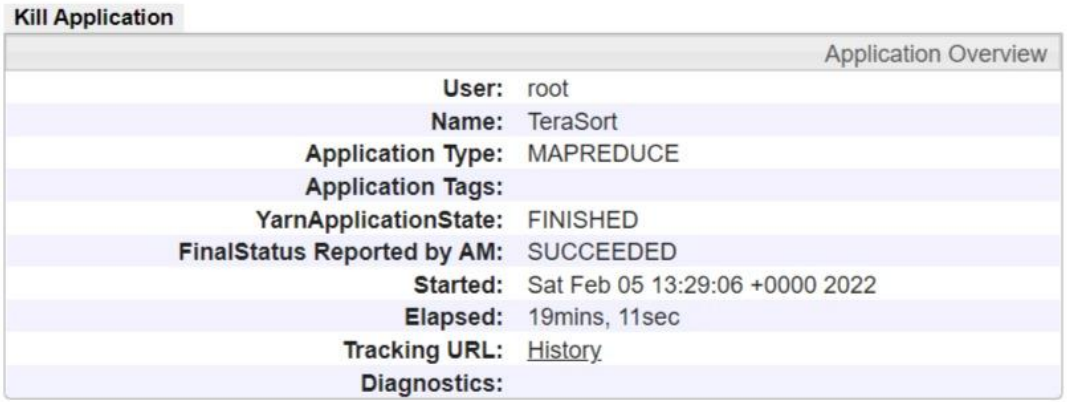




Total: ~2min

For the 20G:





Time: ~23 mins

d.)

Using Time cost on (Teragen + Terasort) of 20G for comparison:

	Hadoop	Hadoop Over K8s	Hadoop Over Serverless Kubernetes
Time	22 mins	1hr 38mins	23 mins

Normal Hadoop setup have the best performance. Because all the data are transmitted under the same network and the delay is low.

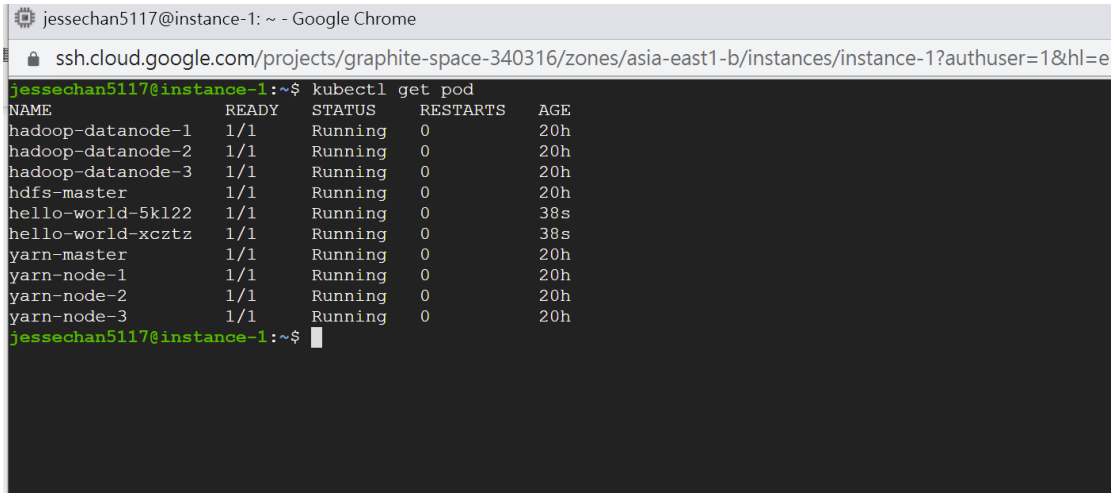
Hadoop over K8s are having worst performance because of the delay of the data transition are much longer under different network. Because of the limitation (disk storage in east server) of free trail account, it is forced to set up the instances with different regions and it further increase the delay.

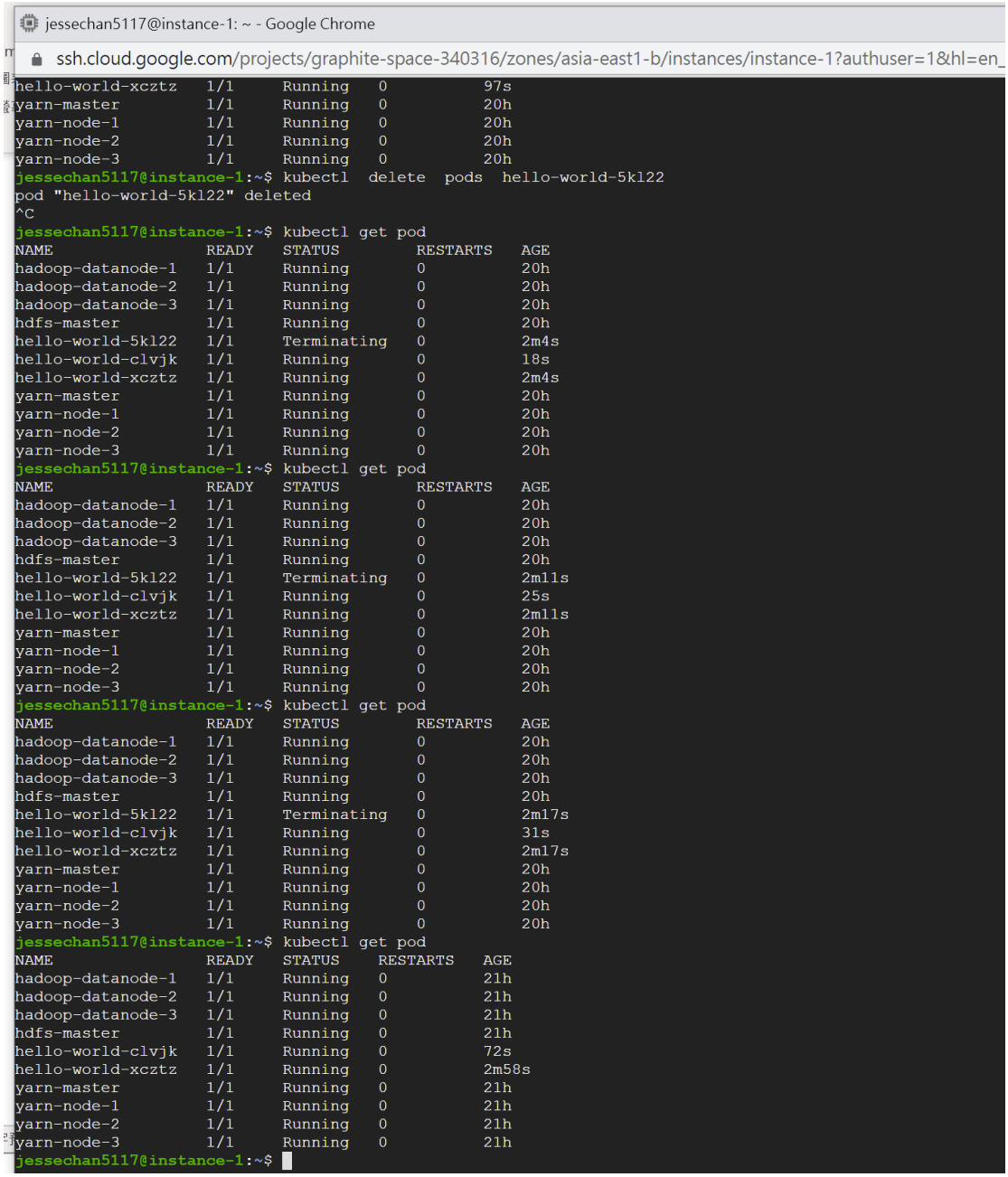
Performance of Hadoop Over Serverless Kubernetes are similar to normal Hadoop setup. It is because the AWS have optimized the network and data flow between different service.

e.)

i)

2 Hello World pod running:





Kubernetes will create a new pod (clvjk) for the application and the killed pod will first terminating for a while and got deleted.

ii)

Running 20G Terasort for kill test:

```
root@yarn-master:/usr/local/hadoop/share/hadoop/mapreduce# hadoop jar hadoop-mapreduce-examples-2.7.2.jar terasort Teragen20G Terasort20G_Killtest
22/02/05 14:04:57 INFO terasort.TeraSort: starting
22/02/05 14:04:58 INFO input.FileInputFormat: Total input paths to process : 2
Spent 172ms computing base splits.
Spent 3ms computing TeraScheduler splits.
Computing input splits took 175ms
Sampling 10 splits of 160
Making 1 from 100000 sampled records
Computing parititions took 674ms
Spent 852ms computing partitions.
22/02/05 14:04:59 INFO client.RMProxy: Connecting to ResourceManager at hadoop-yarn-master.default.svc.cluster.local/10.97.27.22:8032
22/02/05 14:04:59 INFO mapreduce.JobSubmitter: number of splits:160
22/02/05 14:05:00 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1643993993836_0005
22/02/05 14:05:00 INFO impl.YarnClientImpl: Submitted application application_1643993993836_0005
22/02/05 14:05:00 INFO mapreduce.Job: The url to track the job: http://hadoop-yarn-master.default.svc.cluster.local:8088/proxy/application_1643993993836_0005/
22/02/05 14:05:00 INFO mapreduce.Job: Running job: job_1643993993836_0005
22/02/05 14:05:07 INFO mapreduce.Job: Job job_1643993993836_0005 running in uber mode : false
22/02/05 14:05:07 INFO mapreduce.Job:  map 0% reduce 0%
22/02/05 14:05:34 INFO mapreduce.Job:  map 1% reduce 0%
22/02/05 14:05:43 INFO mapreduce.Job:  map 2% reduce 0%
22/02/05 14:05:50 INFO mapreduce.Job:  map 3% reduce 0%
22/02/05 14:05:53 INFO mapreduce.Job:  map 4% reduce 0%
22/02/05 14:05:57 INFO mapreduce.Job:  map 5% reduce 0%
22/02/05 14:06:01 INFO mapreduce.Job:  map 6% reduce 0%
22/02/05 14:06:06 INFO mapreduce.Job:  map 7% reduce 0%
22/02/05 14:06:14 INFO mapreduce.Job:  map 8% reduce 0%
22/02/05 14:06:19 INFO mapreduce.Job:  map 9% reduce 0%
```

Delete a yarn-node:

```
22/02/05 14:10:07 INFO mapreduce.Job:  map 38% reduce 7%
22/02/05 14:10:16 INFO mapreduce.Job:  map 39% reduce 7%
22/02/05 14:10:31 INFO mapreduce.Job:  map 40% reduce 7%
22/02/05 14:10:42 INFO mapreduce.Job:  map 41% reduce 7%
22/02/05 14:10:53 INFO mapreduce.Job:  map 42% reduce 7%
22/02/05 14:11:02 INFO mapreduce.Job:  map 43% reduce 7%
22/02/05 14:11:11 INFO mapreduce.Job:  map 44% reduce 7%
22/02/05 14:11:21 INFO mapreduce.Job:  map 45% reduce 7%
22/02/05 14:11:33 INFO mapreduce.Job:  map 46% reduce 7%
22/02/05 14:11:37 INFO mapreduce.Job: Task Id : attempt_1643993993836_0005_m_000057_0, Status : FAILED
Container launch failed for container_1643993993836_0005_01_000060 : java.net.UnknownHostException: Invalid host name: local host is: (unknown); destination host is: "
yarn-node-1":38589; java.net.UnknownHostException; For more details see:  http://wiki.apache.org/hadoop/UnknownHost
        at sun.reflect.GeneratedConstructorAccessor57.newInstance(Unknown Source)
        at sun.reflect.DelegatingConstructorAccessorImpl.newInstance(DelegatingConstructorAccessorImpl.java:45)
        at java.lang.reflect.Constructor.newInstance(Constructor.java:526)
        at org.apache.hadoop.net.NetUtils.wrapWithMessage(NetUtils.java:792)
        at org.apache.hadoop.net.NetUtils.wrapException(NetUtils.java:744)
        at org.apache.hadoop.ipc.Client$Connection.<init>(Client.java:409)
        at org.apache.hadoop.ipc.Client.getConnection(Client.java:1518)
        at org.apache.hadoop.ipc.Client.call(Client.java:1451)
        at org.apache.hadoop.ipc.Client.call(Client.java:1412)
        at org.apache.hadoop.ipc.ProtobufRpcEngine$Invoker.invoke(ProtobufRpcEngine.java:229)
        at com.sun.proxy.$Proxy80.startContainers(Unknown Source)
        at org.apache.hadoop.yarn.api.impl.pb.client.ContainerManagementProtocolPBClientImpl.startContainers(ContainerManagementProtocolPBClientImpl.java:96)
        at sun.reflect.GeneratedMethodAccessor14.invoke(Unknown Source)
        at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
        at java.lang.reflect.Method.invoke(Method.java:606)
        at org.apache.hadoop.io.retry.RetryInvocationHandler.invokeMethod(RetryInvocationHandler.java:191)
        at org.apache.hadoop.io.retry.RetryInvocationHandler.invoke(RetryInvocationHandler.java:102)
        at com.sun.proxy.$Proxy81.startContainers(Unknown Source)
        at org.apache.hadoop.mapreduce.v2.app.launcher.ContainerLauncherImpl$Container.launch(ContainerLauncherImpl.java:151)
        at org.apache.hadoop.mapreduce.v2.app.launcher.ContainerLauncherImpl$EventProcessor.run(ContainerLauncherImpl.java:375)
        at java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecutor.java:1145)
        at java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecutor.java:615)
        at java.lang.Thread.run(Thread.java:745)
Caused by: java.net.UnknownHostException
        ... 18 more

22/02/05 14:11:45 INFO mapreduce.Job:  map 47% reduce 7%
22/02/05 14:11:58 INFO mapreduce.Job:  map 48% reduce 7%
22/02/05 14:12:09 INFO mapreduce.Job:  map 49% reduce 7%
22/02/05 14:12:17 INFO mapreduce.Job:  map 50% reduce 7%
```

Datanode Information

In operation

Node	Last contact	Admin State	Capacity	Used	Non DFS Used	Remaining	Blocks	Block pool used	Failed Volumes	Version
hadoop-datanode-3:50010 (10.39.0.1:50010)	2	In Service	96.75 GB	18.65 GB	33.42 GB	44.68 GB	157	18.65 GB (19.27%)	0	2.7.2
hadoop-datanode-1:50010 (10.44.0.1:50010)	226	In Service	96.75 GB	22.68 GB	4.74 GB	69.33 GB	191	22.68 GB (23.44%)	0	2.7.2
hadoop-datanode-2:50010 (10.39.0.3:50010)	0	In Service	96.75 GB	25.2 GB	26.88 GB	44.67 GB	212	25.2 GB (26.05%)	0	2.7.2

Decomissioning

Error occurs, the data node-1 is lost contact and become a dead node. But the map-reduce job is continued because of the yarn will handle the fault-tolerance, backup

the data missed and keep on the map-reduce job.