

# 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 Joe) 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.

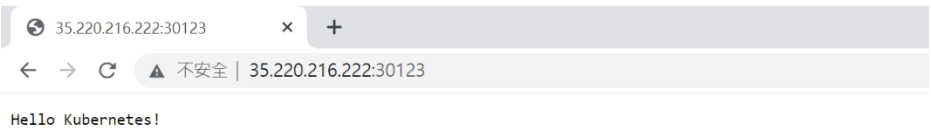
If there is information you need to solve a problem but the information is not stated in the problem, try to find the data somewhere. If you cannot find it, state what data you need, make a reasonable estimate of its value and justify any assumptions you make. You will be graded not only on whether your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

Q1

a.)

Applying the hello-world-demo.yaml

```
jessechan5117@instance-1:~$ kubectl get service
NAME          TYPE        CLUSTER-IP      EXTERNAL-IP      PORT(S)          AGE
redis-hdfs-master NodePort    10.101.219.226   <none>            9000:32241/TCP,50070:3200//TCP 26h
redis-paim-master NodePort    10.109.119.173   <none>            6030:3216/TCP,6031:30488/TCP,6032:30456/TCP,6038:32888/TCP 26h
kubernetes    ClusterIP   10.96.0.1        <none>            443/TCP           26h
my service    NodePort    10.101.50.140    <none>            80:30123/TCP       21h
paim-node-1   ClusterIP   None             <none>            8040/TCP           26h
paim-node-2   ClusterIP   None             <none>            8040/TCP           26h
paim-node-3   ClusterIP   None             <none>            8040/TCP           26h
jessechan5117@instance-1:~$
```



b.)

Multi-node Kubernetes Cluster Setup:

```
jessechan5117@instance-1:~$ kubectl get node
NAME          STATUS    ROLES          AGE    VERSION
instance-1    Ready     control-plane,master 26h    v1.23.3
instance-2    Ready     <none>          21h    v1.23.3
instance-3    Ready     <none>          21h    v1.23.3
instance-5    Ready     <none>          7m42s  v1.23.3
jessechan5117@instance-1:~$
```

## 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-9gez4                   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 16:27:59 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.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:

```
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-17authuser=08hl=en_US&projectNumber=25080880253&useAdminProxy=true&troubleshoot400...
File Input Format Counters
  Bytes Read=0
File Output Format Counters
  Bytes Written=2147403600
root@yarn-master: /usr/local/hadoop/share/hadoop/mapreduce# hadoop jar hadoop-mapreduce-examples-2.7.2.jar terasort TeraGen2G Terasort2G
22/02/04 17:18:21 INFO terasort.TeraSort: starting
22/02/04 17:18:23 INFO input.FileInputFormat: Total input paths to process : 2
Spent 123ms computing base-splits.
Spent 2ms computing TeraScheduler splits.
Computing input splits took 127ms
Sampling 10 splits of 16
Making 1 from 100000 sampled records
Computing partitions took 635ms
Spent 764ms computing partitions.
22/02/04 17:18:24 INFO client.RMProxy: Connecting to ResourceManager at hadoop-yarn-master.default.svc.cluster.local/10.97.27.22:8032
22/02/04 17:18:24 INFO mapreduce.JobSubmitter: number of splits:16
22/02/04 17:18:25 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1643993993036_0002
22/02/04 17:18:25 INFO impl.YarnClientImpl: Submitted application application_1643993993036_0002
22/02/04 17:18:25 INFO mapreduce.Job: The url to track the job: http://hadoop-yarn-master.default.svc.cluster.local:8088/proxy/application_1643993993036_0002/
22/02/04 17:18:25 INFO mapreduce.Job: Running job: job_1643993993036_0002
22/02/04 17:18:33 INFO mapreduce.Job: Job job_1643993993036_0002 running in uber mode : false
22/02/04 17:18:33 INFO mapreduce.Job: map 0% reduce 0%
22/02/04 17:18:34 INFO mapreduce.Job: map 3% reduce 0%
22/02/04 17:18:35 INFO mapreduce.Job: map 6% reduce 0%
22/02/04 17:18:36 INFO mapreduce.Job: map 10% reduce 0%
22/02/04 17:18:37 INFO mapreduce.Job: map 11% reduce 0%
22/02/04 17:18:38 INFO mapreduce.Job: map 12% reduce 0%
22/02/04 17:18:39 INFO mapreduce.Job: map 16% reduce 0%
22/02/04 17:18:40 INFO mapreduce.Job: map 17% reduce 0%
22/02/04 17:18:41 INFO mapreduce.Job: map 18% reduce 0%
22/02/04 17:18:42 INFO mapreduce.Job: map 22% reduce 0%
22/02/04 17:18:43 INFO mapreduce.Job: map 25% reduce 0%
22/02/04 17:18:44 INFO mapreduce.Job: map 28% reduce 0%
22/02/04 17:18:45 INFO mapreduce.Job: map 29% reduce 0%
22/02/04 17:18:46 INFO mapreduce.Job: map 32% reduce 0%
22/02/04 17:18:47 INFO mapreduce.Job: map 34% reduce 0%
22/02/04 17:18:48 INFO mapreduce.Job: map 37% reduce 0%
22/02/04 17:18:49 INFO mapreduce.Job: map 38% reduce 0%
22/02/04 17:18:50 INFO mapreduce.Job: map 41% reduce 0%
22/02/04 17:18:51 INFO mapreduce.Job: map 45% reduce 0%
22/02/04 17:18:52 INFO mapreduce.Job: map 49% reduce 0%
22/02/04 17:18:53 INFO mapreduce.Job: map 50% reduce 0%
22/02/04 17:18:54 INFO mapreduce.Job: map 52% reduce 0%
22/02/04 17:18:55 INFO mapreduce.Job: map 53% reduce 0%
22/02/04 17:18:56 INFO mapreduce.Job: map 57% reduce 0%
22/02/04 17:18:57 INFO mapreduce.Job: map 60% reduce 0%
22/02/04 17:18:58 INFO mapreduce.Job: map 63% reduce 0%
22/02/04 17:18:59 INFO mapreduce.Job: map 66% reduce 0%
```

## Time:

Application Overview	
User:	root
Name:	TeraGen
Application Type:	MAPREDUCE
Application Tags:	
YarnApplicationState:	FINISHED
FinalStatus Reported by AM:	SUCCEEDED
Started:	Fri Feb 04 17:15:56 +0000 2022
Elapsed:	1mins, 50sec
Tracking URL:	<a href="#">History</a>
Diagnostics:	

Application Overview	
User:	root
Name:	TeraSort
Application Type:	MAPREDUCE
Application Tags:	
YarnApplicationState:	FINISHED
FinalStatus Reported by AM:	SUCCEEDED
Started:	Fri Feb 04 17:18:25 +0000 2022
Elapsed:	5mins, 13sec
Tracking URL:	<a href="#">History</a>
Diagnostics:	

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=1745121200
  Total committed heap usage (bytes)=234356736
  org.apache.hadoop.examples.terasort.TeraGen$Counters
  CHECKSUM=461200258163748239
  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:

```
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_

22/02/04 18:52:07 INFO mapreduce.Job: map 100% reduce 83%
22/02/04 18:53:04 INFO mapreduce.Job: map 100% reduce 84%
22/02/04 18:53:52 INFO mapreduce.Job: map 100% reduce 85%
22/02/04 18:54:47 INFO mapreduce.Job: map 100% reduce 86%
22/02/04 18:55:41 INFO mapreduce.Job: map 100% reduce 87%
22/02/04 18:56:30 INFO mapreduce.Job: map 100% reduce 88%
22/02/04 18:57:06 INFO mapreduce.Job: map 100% reduce 89%
22/02/04 18:58:09 INFO mapreduce.Job: map 100% reduce 90%
22/02/04 18:58:48 INFO mapreduce.Job: map 100% reduce 91%
22/02/04 18:59:45 INFO mapreduce.Job: map 100% reduce 92%
22/02/04 19:00:45 INFO mapreduce.Job: map 100% reduce 93%
22/02/04 19:01:34 INFO mapreduce.Job: map 100% reduce 94%
22/02/04 19:02:07 INFO mapreduce.Job: map 100% reduce 95%
22/02/04 19:03:08 INFO mapreduce.Job: map 100% reduce 96%
22/02/04 19:03:53 INFO mapreduce.Job: map 100% reduce 97%
22/02/04 19:04:50 INFO mapreduce.Job: map 100% reduce 98%
22/02/04 19:05:47 INFO mapreduce.Job: map 100% reduce 99%
22/02/04 19:06:33 INFO mapreduce.Job: map 100% reduce 100%
22/02/04 19:07:08 INFO mapreduce.Job: Job job_1643993993836_0004 completed successfully
22/02/04 19:07:08 INFO mapreduce.Job: Counters: 50
  File System Counters
    FILE: Number of bytes read=76353783446
    FILE: Number of bytes written=98706764988
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=21474857140
    HDFS: Number of bytes written=21474836500
    HDFS: Number of read operations=483
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=2
  Job Counters
    Killed map tasks=1
    Launched map tasks=161
    Launched reduce tasks=1
    Rack-local map tasks=161
    Total time spent by all maps in occupied slots (ms)=13995350
    Total time spent by all reduces in occupied slots (ms)=4617011
    Total time spent by all map tasks (ms)=13995350
    Total time spent by all reduce tasks (ms)=4617011
    Total vcore-milliseconds taken by all map tasks=13995350
    Total vcore-milliseconds taken by all reduce tasks=4617011
    Total megabyte-milliseconds taken by all map tasks=14331238400
    Total megabyte-milliseconds taken by all reduce tasks=4727819264
  Map-Reduce Framework
    Map input records=214748365
    Map output records=214748365
    Map output bytes=21904333230
    Map output materialized bytes=22333830920
    Input split bytes=20640
    Combine input records=0
    Combine output records=0
    Reduce input groups=214748365
    Reduce shuffle bytes=22333830920
    Reduce input records=214748365
    Reduce output records=214748365
    Spilled Records=948919331
    Shuffled Maps =160
    Failed Shuffles=0
    Merged Map outputs=160
    GC time elapsed (ms)=75844
    CPU time spent (ms)=3077770
    Physical memory (bytes) snapshot=52049461248
    Virtual memory (bytes) snapshot=140928245760
    Total committed heap usage (bytes)=31114395648
  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=21474836500
  File Output Format Counters
    Bytes Written=21474836500
22/02/04 19:07:08 INFO terasort.TeraSort: done
root@yarn-master: /usr/local/hadoop/share/hadoop/mapreduce#
root@yarn-master: /usr/local/hadoop/share/hadoop/mapreduce#
```

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:	<a href="#">History</a>	
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:	<a href="#">History</a>	
Diagnostics:		

		Application Metrics
--	--	---------------------

Total: ~1hr 38mins

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

```
File Edit View Search Terminal Help
0.82.0
ubuntu@ubuntu1804:~$ clear

ubuntu@ubuntu1804:~$ eksctl version
0.82.0
ubuntu@ubuntu1804:~$ kubectrl version --short --client
Client Version: v1.21.2-13-d2965f0db10712
ubuntu@ubuntu1804:~$ aws --version
aws-cli/2.4.16 Python/3.8.8 Linux/x86_64 Ubuntu/18 prompt/off
ubuntu@ubuntu1804:~$
```

[illegible]



EKS > 集群 > my-cluster

## my-cluster

概觀 工作負載 組態

**叢集組態** 資訊

Kubernetes 版本 資訊	平台版本 資訊
1.21	eks.4

**詳細資訊**

API 伺服器資訊	OpenID Connect 供應商 URL	已建立
https://bcf2c28a8e91500f30d3113ed25f21e.gr7.us-east-2.eks.amazonaws.com	https://oidc.eks.us-east-2.amazonaws.com/id/bcf2c28a8e91500f30d3113ed25f21e	18 minutes ago
憑證記錄簿資訊	叢集 IAM 角色 ARN	叢集 ARN
LSOHLSTCLU61TBDRVUSU7JQOFURSOILS0CK1JSLM16ANDQW67OF3SUJ0Z0C QUR8TLma3FoaZi0HX0wC4PR0ZBREFTWYJN0DVRWURUVYPER0kucmRKS5mK Y2016GRWpWpKQ9RYRTJUTIESG0VEVSUR0HE1G01HEVE1STURJ0016R0J0NG	arn:aws:iam::6141755:role/eksctl-my-cluster-cluster-ServiceRole-VW03NFF6G4780	arn:aws:eks:us-east-2:6141755:18453:cluster/my-cluster

**密碼加密** 啟用

密鑰加密	KMS 密鑰 ID
已停用	-

```
2022-02-05 07:23:30 [?] EKS cluster 'my-cluster' in 'us-east-2' region is ready
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 32% 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 40% 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
root@yarn-master: /usr/local/hadoop/share/hadoop/mapreduce#
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=0
  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-millisecods taken by all map tasks=47404
  Total megabyte-millisecods 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)=29622720
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:

```
root@yarn-master: /usr/local/hadoop/share/hadoop/mapreduce
File Edit View Search Terminal Help
CHECKSUM=46124753271996946
File Input Format Counters
  Bytes Read=0
File Output Format Counters
  Bytes Written=2147483648
root@yarn-master: /usr/local/hadoop/share/hadoop/mapreduce# hadoop jar hadoop-mapreduce-examples-2.7.2.jar terasort TeraGen2G TeraSort2G
22/02/05 13:20:40 INFO terasort.TeraSort: starting
22/02/05 13:20:41 INFO input.FileInputFormat: Total input paths to process : 2
Spent 159ms computing base-splits.
Spent 2ms computing TeraScheduler splits.
Computing input splits took 143ms
Sampling 10 splits of 10
Making 1 from 100000 sampled records
Computing partitions took 593ms
Spent 738ms computing partitions.
22/02/05 13:20:42 INFO client.RMProxy: Connecting to ResourceManager at hadoop-yarn-master.default.svc.cluster.local/10.100.181.117:8032
22/02/05 13:20:42 INFO mapreduce.JobSubmitter: number of splits:16
22/02/05 13:20:43 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1644064510094_0002
22/02/05 13:20:43 INFO impl.YarnClientImpl: Submitted application application_1644064510094_0002
22/02/05 13:20:43 INFO mapreduce.Job: The url to track the job: http://hadoop-yarn-master.default.svc.cluster.local:8088/proxy/application_1644064510094_0002/
22/02/05 13:20:43 INFO mapreduce.Job: Running job: job_1644064510094_0002
22/02/05 13:20:50 INFO mapreduce.Job: Job job_1644064510094_0002 running in uber mode : false
root 22/02/05 13:20:50 INFO mapreduce.Job: map 0% reduce 0%
22/02/05 13:21:03 INFO mapreduce.Job: map 2% reduce 0%
22/02/05 13:21:06 INFO mapreduce.Job: map 3% reduce 0%
22/02/05 13:21:10 INFO mapreduce.Job: map 7% reduce 0%
22/02/05 13:21:11 INFO mapreduce.Job: map 9% reduce 0%
22/02/05 13:21:12 INFO mapreduce.Job: map 12% reduce 0%
22/02/05 13:21:13 INFO mapreduce.Job: map 13% reduce 0%
22/02/05 13:21:15 INFO mapreduce.Job: map 17% reduce 0%
22/02/05 13:21:19 INFO mapreduce.Job: map 20% reduce 0%
22/02/05 13:21:20 INFO mapreduce.Job: map 24% reduce 0%
22/02/05 13:21:21 INFO mapreduce.Job: map 29% reduce 0%
22/02/05 13:21:22 INFO mapreduce.Job: map 36% reduce 0%
22/02/05 13:21:23 INFO mapreduce.Job: map 37% reduce 0%
22/02/05 13:21:24 INFO mapreduce.Job: map 43% reduce 0%
22/02/05 13:21:25 INFO mapreduce.Job: map 53% reduce 0%
22/02/05 13:21:26 INFO mapreduce.Job: map 56% reduce 0%
22/02/05 13:21:27 INFO mapreduce.Job: map 59% reduce 0%
22/02/05 13:21:28 INFO mapreduce.Job: map 62% reduce 0%
22/02/05 13:21:29 INFO mapreduce.Job: map 64% reduce 0%
22/02/05 13:21:30 INFO mapreduce.Job: map 66% reduce 0%
22/02/05 13:21:31 INFO mapreduce.Job: map 68% reduce 0%
22/02/05 13:21:32 INFO mapreduce.Job: map 74% reduce 0%
22/02/05 13:21:34 INFO mapreduce.Job: map 77% reduce 0%
22/02/05 13:21:35 INFO mapreduce.Job: map 78% reduce 0%
22/02/05 13:21:36 INFO mapreduce.Job: map 79% reduce 0%
22/02/05 13:21:37 INFO mapreduce.Job: map 81% reduce 13%
22/02/05 13:21:38 INFO mapreduce.Job: map 82% reduce 13%
22/02/05 13:21:39 INFO mapreduce.Job: map 84% reduce 13%
22/02/05 13:21:40 INFO mapreduce.Job: map 88% reduce 17%
22/02/05 13:21:41 INFO mapreduce.Job: map 90% reduce 17%
22/02/05 13:21:43 INFO mapreduce.Job: map 91% reduce 19%
22/02/05 13:21:46 INFO mapreduce.Job: map 94% reduce 23%
22/02/05 13:21:47 INFO mapreduce.Job: map 98% reduce 23%
22/02/05 13:21:48 INFO mapreduce.Job: map 100% reduce 23%
22/02/05 13:21:52 INFO mapreduce.Job: map 100% reduce 31%
22/02/05 13:21:55 INFO mapreduce.Job: map 100% reduce 38%
22/02/05 13:21:58 INFO mapreduce.Job: map 100% reduce 46%
22/02/05 13:22:01 INFO mapreduce.Job: map 100% reduce 53%
```

Time:

Kill Application		Application Overview	
User:		root	
Name:		TeraGen	
Application Type:		MAPREDUCE	
Application Tags:			
YarnApplicationState:		FINISHED	
FinalStatus Reported by AM:		SUCCEEDED	
Started:		Sat Feb 05 13:18:09 +0000 2022	
Elapsed:		34sec	
Tracking URL:		<a href="#">History</a>	
Diagnostics:			

Kill Application	Application Overview
<b>User:</b>	root
<b>Name:</b>	TeraSort
<b>Application Type:</b>	MAPREDUCE
<b>Application Tags:</b>	
<b>YarnApplicationState:</b>	FINISHED
<b>FinalStatus Reported by AM:</b>	SUCCEEDED
<b>Started:</b>	Sat Feb 05 13:20:43 +0000 2022
<b>Elapsed:</b>	1mins, 50sec
<b>Tracking URL:</b>	<a href="#">History</a>
<b>Diagnostics:</b>	

Total: ~2min

For the 20G:

Kill Application	Application Overview
<b>User:</b>	root
<b>Name:</b>	TeraGen
<b>Application Type:</b>	MAPREDUCE
<b>Application Tags:</b>	
<b>YarnApplicationState:</b>	FINISHED
<b>FinalStatus Reported by AM:</b>	SUCCEEDED
<b>Started:</b>	Sat Feb 05 13:24:56 +0000 2022
<b>Elapsed:</b>	3mins, 48sec
<b>Tracking URL:</b>	<a href="#">History</a>
<b>Diagnostics:</b>	

Kill Application	Application Overview
<b>User:</b>	root
<b>Name:</b>	TeraSort
<b>Application Type:</b>	MAPREDUCE
<b>Application Tags:</b>	
<b>YarnApplicationState:</b>	FINISHED
<b>FinalStatus Reported by AM:</b>	SUCCEEDED
<b>Started:</b>	Sat Feb 05 13:29:06 +0000 2022
<b>Elapsed:</b>	19mins, 11sec
<b>Tracking URL:</b>	<a href="#">History</a>
<b>Diagnostics:</b>	

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:

```
jessechan5117@instance-1: ~ - Google Chrome
ssh.cloud.google.com/projects/graphite-space-340316/zones/asia-east1-b/instances/instance-1?authuser=1&hl=e
jessechan5117@instance-1:~$ kubectl get pod
NAME                READY   STATUS    RESTARTS   AGE
hadoop-datanode-1    1/1     Running   0           20h
hadoop-datanode-2    1/1     Running   0           20h
hadoop-datanode-3    1/1     Running   0           20h
hdfs-master          1/1     Running   0           20h
hello-world-5kl22    1/1     Running   0           38s
hello-world-xcztz    1/1     Running   0           38s
yarn-master          1/1     Running   0           20h
yarn-node-1          1/1     Running   0           20h
yarn-node-2          1/1     Running   0           20h
yarn-node-3          1/1     Running   0           20h
jessechan5117@instance-1:~$
```

```
jessechan5117@instance-1: ~ - Google Chrome
ssh.cloud.google.com/projects/graphite-space-340316/zones/asia-east1-b/instances/instance-1?authuser=1&hl=en_

hello-world-xcztz 1/1 Running 0 97s
yarn-master 1/1 Running 0 20h
yarn-node-1 1/1 Running 0 20h
yarn-node-2 1/1 Running 0 20h
yarn-node-3 1/1 Running 0 20h
jessechan5117@instance-1:~$ kubectl delete pods hello-world-5kl22
pod "hello-world-5kl22" deleted
^C
jessechan5117@instance-1:~$ kubectl get pod
NAME READY STATUS RESTARTS AGE
hadoop-datanode-1 1/1 Running 0 20h
hadoop-datanode-2 1/1 Running 0 20h
hadoop-datanode-3 1/1 Running 0 20h
hdfs-master 1/1 Running 0 20h
hello-world-5kl22 1/1 Terminating 0 2m4s
hello-world-clvjkl 1/1 Running 0 18s
hello-world-xcztz 1/1 Running 0 2m4s
yarn-master 1/1 Running 0 20h
yarn-node-1 1/1 Running 0 20h
yarn-node-2 1/1 Running 0 20h
yarn-node-3 1/1 Running 0 20h
jessechan5117@instance-1:~$ kubectl get pod
NAME READY STATUS RESTARTS AGE
hadoop-datanode-1 1/1 Running 0 20h
hadoop-datanode-2 1/1 Running 0 20h
hadoop-datanode-3 1/1 Running 0 20h
hdfs-master 1/1 Running 0 20h
hello-world-5kl22 1/1 Terminating 0 2m11s
hello-world-clvjkl 1/1 Running 0 25s
hello-world-xcztz 1/1 Running 0 2m11s
yarn-master 1/1 Running 0 20h
yarn-node-1 1/1 Running 0 20h
yarn-node-2 1/1 Running 0 20h
yarn-node-3 1/1 Running 0 20h
jessechan5117@instance-1:~$ kubectl get pod
NAME READY STATUS RESTARTS AGE
hadoop-datanode-1 1/1 Running 0 20h
hadoop-datanode-2 1/1 Running 0 20h
hadoop-datanode-3 1/1 Running 0 20h
hdfs-master 1/1 Running 0 20h
hello-world-5kl22 1/1 Terminating 0 2m17s
hello-world-clvjkl 1/1 Running 0 31s
hello-world-xcztz 1/1 Running 0 2m17s
yarn-master 1/1 Running 0 20h
yarn-node-1 1/1 Running 0 20h
yarn-node-2 1/1 Running 0 20h
yarn-node-3 1/1 Running 0 20h
jessechan5117@instance-1:~$ kubectl get pod
NAME READY STATUS RESTARTS AGE
hadoop-datanode-1 1/1 Running 0 21h
hadoop-datanode-2 1/1 Running 0 21h
hadoop-datanode-3 1/1 Running 0 21h
hdfs-master 1/1 Running 0 21h
hello-world-clvjkl 1/1 Running 0 72s
hello-world-xcztz 1/1 Running 0 2m58s
yarn-master 1/1 Running 0 21h
yarn-node-1 1/1 Running 0 21h
yarn-node-2 1/1 Running 0 21h
yarn-node-3 1/1 Running 0 21h
jessechan5117@instance-1:~$
```

Kubernetes will create a new pod (clvjkl) 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 14:04:47 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 partitions took 674ms
Spent 852ms computing partitions.
22/02/05 14:04:59 INFO client.RMPProxy: 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 14:10:17 INFO mapreduce.Job: map 37% reduce 7%
22/02/05 14:10:17 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-1738589; 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:752)
    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.$Proxy0.startContainers(Unknown Source)
    at org.apache.hadoop.yarn.api.impl.pb.client.ContainerManagementProtocolPBClientImpl.startContainers(ContainerManagementProtocolPBClientImpl.java:96)
    at sun.reflect.GeneratedMethodAccessor14.invoke(Unknown Source)
    at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
    at java.lang.reflect.Method.invoke(Method.java:606)
    at org.apache.hadoop.io.retry.RetryInvocationHandler.invokeMethod(RetryInvocationHandler.java:191)
    at org.apache.hadoop.io.retry.RetryInvocationHandler.invoke(RetryInvocationHandler.java:102)
    at com.sun.proxy.$Proxy0.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:1148)
    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.