Question 2: Cluster Screenshot:(block size: 128MB; single node with 4 cores; others are default)
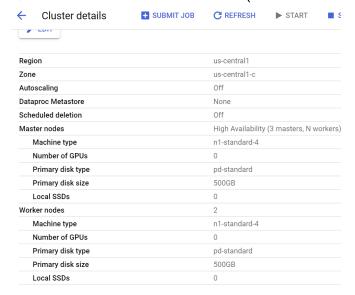
← Cluster details   ⊞ SUBMIT JOB   ⟳ REFRESH   ▶ START   ■ STOP   🗑 DELETE

✏ EDIT

Region	us-central1
Zone	us-central1-c
Autoscaling	Off
Dataproc Metastore	None
Scheduled deletion	Off
Master node	Single Node (1 master, 0 workers)
Machine type	n1-standard-4
Number of GPUs	0
Primary disk type	pd-standard
Primary disk size	500GB
Local SSDs	0
Secure Boot	Disabled
VTPM	Disabled
Integrity Monitoring	Disabled
Cloud Storage staging bucket	csee4121homework
Network	default
Network tags	None
Internal IP only	No
Image version ❓	2.0.37-debian10
Created	Apr 27, 2022, 1:45:05 PM
Optional components	JUPYTER
Properties	Show properties
Advanced security	Disabled
Labels	goog-datap... : cluster-e6... ⌄
Encryption type	Google-managed key

Cluster cluster-e60d properties

capacity-scheduler:yarn.scheduler.capacity.root.default.ordering-policy	fair
core:fs.gs.block.size	134217728

Job Properties(cores:4; Spark driver memory: 1GB; Spark executor memory: 5GB)

Job ID	job-51f5443d
Job UUID	985172af-17c2-4e53-8ac2-2a5ec0a8b6ea
Type	Dataproc Job
Status	✅ Succeeded

MONITORING   **CONFIGURATION**

✏ EDIT

Start time:	Apr 28, 2022, 12:37:25 AM
Elapsed time:	9 min 55 sec
Status:	Succeeded
Region	us-central1
Cluster	cluster-e60d
Job type	PySpark
Main python file	gs://csee4121homework/notebooks/jupyter/p1t2q2.py
Jar files	gs://csee4121/homework2/spark-xml_2.12-0.14.0.jar
Properties	
spark.executor.cores	4
spark.driver.cores	4
spark.executor.memory	5g
spark.driver.memory	1g
Labels	

Question 3: Cluster Screenshot: (3 nodes with 2 workers;others are default)

← Cluster details | ➕ SUBMIT JOB | ⟳ REFRESH | ▶ START | ■ S...

Region	us-central1
Zone	us-central1-c
Autoscaling	Off
Dataproc Metastore	None
Scheduled deletion	Off
Master nodes	High Availability (3 masters, N workers)
Machine type	n1-standard-4
Number of GPUs	0
Primary disk type	pd-standard
Primary disk size	500GB
Local SSDs	0
Worker nodes	2
Machine type	n1-standard-4
Number of GPUs	0
Primary disk type	pd-standard
Primary disk size	500GB
Local SSDs	0

Cluster cluster-1058 properties

capacity-scheduler:yarn.scheduler.capacity.root.default.ordering-policy	fair
core:fs.gs.block.size	134217728

Job Properties(cores:4; Spark driver memory: 1GB; Spark executor memory: 5GB)

MONITORING | CONFIGURATION

✏️ EDIT

Start time:	Apr 28, 2022, 1:15:22 AM
Elapsed time:	5 min 24 sec
Status:	Succeeded
Region	us-central1
Cluster	cluster-1058
Job type	PySpark
Main python file	gs://csee4121homework/notebooks/jupyter/p1t2q2.py
Jar files	gs://csee4121/homework2/spark-xml_2.12-0.14.0.jar
Properties	
spark.executor.cores	4
spark.driver.cores	4
spark.executor.memory	5g
spark.driver.memory	1g
Labels	

Question 4

Cluster Screenshot:(block size: 64MB; 3 nodes with 2 workers;others default)

← Cluster details   ＋ SUBMIT JOB   ⟳ REFRESH   ▶ START   ■ STOP   🗑 D

Region	us-central1
Zone	us-central1-a
Autoscaling	Off
Dataproc Metastore	None
Scheduled deletion	Off
Master nodes	High Availability (3 masters, N workers)
Machine type	n1-standard-4
Number of GPUs	0
Primary disk type	pd-standard
Primary disk size	500GB
Local SSDs	0
Worker nodes	2
Machine type	n1-standard-4
Number of GPUs	0
Primary disk type	pd-standard
Primary disk size	500GB
Local SSDs	0

Cluster cluster-7909 properties

hdfs:dfs.blocksize	67108864

Job Properties(cores:4; Spark driver memory: 1GB; Spark executor memory: 5GB)

← Job details   ⧉ CLONE   🗑 DELETE   ■ STOP   ⟳ REFRESH

MONITORING   CONFIGURATION

✏ EDIT

Start time:	Apr 28, 2022, 1:54:54 AM
Elapsed time:	5 min 43 sec
Status:	Succeeded
Region	us-central1
Cluster	cluster-7909
Job type	PySpark
Main python file	gs://csee4121homework/notebooks/jupyter/p1t2q2.py
Jar files	gs://csee4121/homework2/spark-xml_2.12-0.14.0.jar
Properties	
spark.executor.cores	4
spark.driver.cores	4
spark.executor.memory	5g
spark.driver.memory	1g
Labels	

EQUIVALENT REST

Question 5

Cluster Screenshot:(1 node with 2 workers; others are default)

← Cluster details   ➕ SUBMIT JOB   ↻ REFRESH   ▶ START   ■ STOP   🗑 DELETE

MONITORING	JOBS	VM INSTANCES	CONFIGURATION	WEB INTERFACES

✏ EDIT

Region	us-central1
Zone	us-central1-a
Autoscaling	Off
Dataproc Metastore	None
Scheduled deletion	Off
Master node	Standard (1 master, N workers)
Machine type	n1-standard-4
Number of GPUs	0
Primary disk type	pd-standard
Primary disk size	500GB
Local SSDs	0
Worker nodes	2
Machine type	n1-standard-4
Number of GPUs	0
Primary disk type	pd-standard
Primary disk size	500GB
Local SSDs	0

Cluster cluster-ba6a properties

capacity-scheduler:yarn.scheduler.capacity.root.default.ordering-policy	fair
core:fs.gs.block.size	134217728

Job Properties(without killing one worker)(cores:4; Spark driver memory: 5GB; Spark executor memory: 5GB):

← Job details   📋 CLONE   🗑 DELETE   ■ STOP   ↻ REFRESH

Job ID	job-5d0d54db
Job UUID	7bd61483-2130-45f3-8e34-1e26b59f487b
Type	Dataproc Job
Status	✔ Succeeded

MONITORING	CONFIGURATION

✏ EDIT

Start time:	Apr 28, 2022, 2:48:20 AM
Elapsed time:	1 hr 16 min
Status:	Succeeded
Region	us-central1
Cluster	cluster-ba6a
Job type	PySpark
Main python file	gs://csee4121homework/notebooks/jupyter/p1t2q5.py
Jar files	gs://csee4121/homework2/spark-xml_2.12-0.14.0.jar
Properties	
spark.executor.cores	4
spark.driver.cores	4
spark.executor.memory	5g
spark.driver.memory	5g
Labels	

Job Properties(after killing one worker)(cores:4; Spark driver memory: 5GB; Spark executor memory: 5GB):

← Job details   ⧉ CLONE   🗑 DELETE   ⬛ STOP   ⟳ REFRESH

Job ID	job-1def1185
Job UUID	517321f5-845e-426a-b267-b830b5ffecd5
Type	Dataproc Job
Status	✅ Succeeded

MONITORING   **CONFIGURATION**

✏ EDIT

Start time:	Apr 28, 2022, 4:41:48 AM
Elapsed time:	2 hr 26 min
Status:	Succeeded
Region	us-central1
Cluster	cluster-ba6a
Job type	PySpark
Main python file	gs://csee4121homework/notebooks/jupyter/p1t2q5.py
Jar files	gs://csee4121/homework2/spark-xml_2.12-0.14.0.jar
Properties	
spark.executor.cores	4
spark.driver.cores	4
spark.executor.memory	5g
spark.driver.memory	5g
Labels	

EQUIVALENT REST

Question 6

Cluster Screenshot:(1 node with 2 workers;replication factor = 1; others are default)

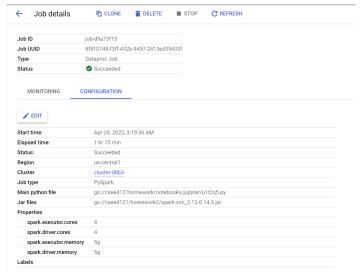← **Cluster details** ⊞ SUBMIT JOB ⟳ REFRESH ▶ START ■ STOP 🗑

MONITORING	JOBS	VM INSTANCES	CONFIGURATION	WEB INTERFACES

✏ EDIT

Region	us-central1
Zone	us-central1-a
Autoscaling	Off
Dataproc Metastore	None
Scheduled deletion	Off
Master node	Standard (1 master, N workers)
Machine type	n1-standard-4
Number of GPUs	0
Primary disk type	pd-standard
Primary disk size	500GB
Local SSDs	0
Worker nodes	2

Cluster cluster-d6b3 properties

capacity-scheduler:yarn.scheduler.capacity.root.default.ordering-policy	fair
core:fs.gs.block.size	134217728

Cluster cluster-d6b3 properties

hdfs:dfs.namenode.service.handler.count	10
hdfs:dfs.namenode.servicerpc-address	cluster-d6b3-m:8
hdfs:dfs.replication	1

Job Properties(without killing one worker)(cores:4; Spark driver memory: 5GB; Spark executor memory: 5GB):

← **Job details** ⧉ CLONE 🗑 DELETE ■ STOP ⟳ REFRESH

Job ID	job-d9a73f15
Job UUID	8f810748-f3ff-432b-9457-2813ed35433f
Type	Dataproc Job
Status	✓ Succeeded

MONITORING	CONFIGURATION

✏ EDIT

Start time:	Apr 28, 2022, 3:19:36 AM
Elapsed time:	1 hr 15 min
Status:	Succeeded
Region	us-central1
Cluster	cluster-d6b3
Job type	PySpark
Main python file	gs://csee4121homework/notebooks/jupyter/p1t2q5.py
Jar files	gs://csee4121/homework2/spark-xml_2.12-0.14.0.jar
Properties	
spark.executor.cores	4
spark.driver.cores	4
spark.executor.memory	5g
spark.driver.memory	5g
Labels	

Question 7(Cluster Screenshot:(block size: 64MB; 1 node with 2 workers; others are default))

← Cluster details   ➕ SUBMIT JOB   ↻ REFRESH   ▶ START   ■

MONITORING   JOBS   VM INSTANCES   **CONFIGURATION**   WEB II

✏ EDIT

Region	us-central1
Zone	us-central1-f
Autoscaling	Off
Dataproc Metastore	None
Scheduled deletion	Off
Master node	Standard (1 master, N workers)
Machine type	n1-standard-4
Number of GPUs	0
Primary disk type	pd-standard
Primary disk size	500GB
Local SSDs	0
Worker nodes	2
Machine type	n1-standard-4
Number of GPUs	0
Primary disk type	pd-standard
Primary disk size	500GB
Local SSDs	0

Cluster cluster-ce31 properties

hadoop-env:HADOOP_DATANODE_OPTS	-Xmx512m
hdfs:dfs.blocksize	67108864

Job Properties(without killing one worker)(Spark driver memory: cores:4; 5GB; Spark executor memory: 5GB):

← Job details   ⧉ CLONE   🗑 DELETE   ■ STOP   ↻ REFRESH

Job ID	job-46541f6b
Job UUID	fadb75c9-bf50-49aa-8b33-12c154494d0e
Type	Dataproc Job
Status	✅ Succeeded

MONITORING   **CONFIGURATION**

✏ EDIT

Start time:	Apr 28, 2022, 8:33:31 AM
Elapsed time:	1 hr 24 min
Status:	Succeeded
Region	us-central1
Cluster	cluster-ce31
Job type	PySpark
Main python file	gs://csee4121homework/notebooks/jupyter/p1t2q5.py
Jar files	gs://csee4121/homework2/spark-xml_2.12-0.14.0.jar
Properties	
spark.executor.cores	4
spark.driver.cores	4
spark.executor.memory	5g
spark.driver.memory	5g

Output   LINE WRAP: OFF