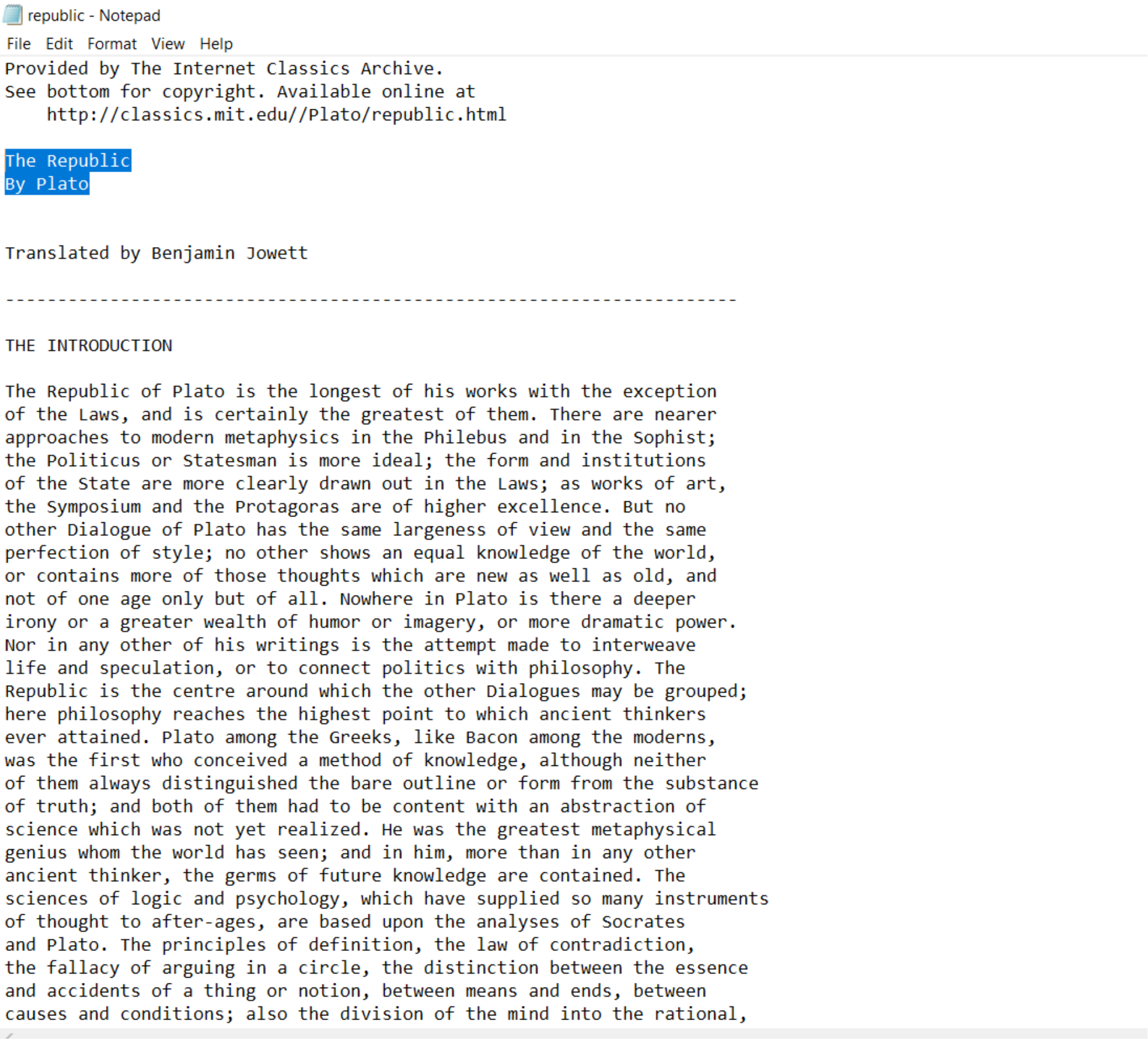
Yogeshwaran S

Case Study 8

Module 8: Data Flow and Data Processing

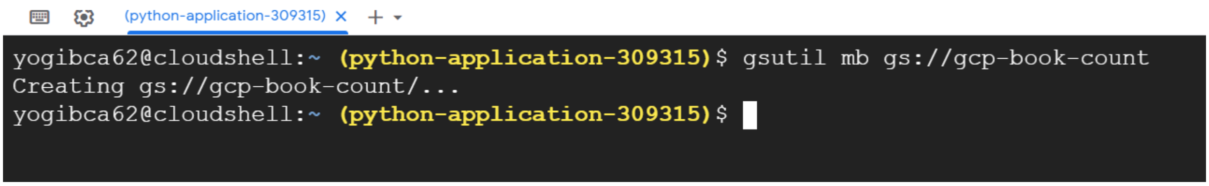
Step 1 –

I create a txt file of a book called “The Republic”

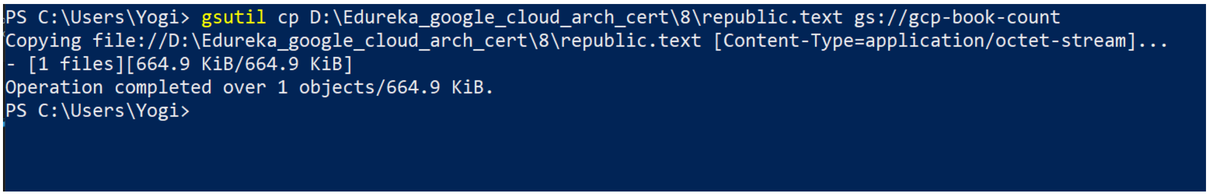


```
Provided by The Internet Classics Archive.
See bottom for copyright. Available online at
    http://classics.mit.edu//Plato/republic.html

The Republic
By Plato


Translated by Benjamin Jowett

----------------------------------------------------------------------

THE INTRODUCTION

The Republic of Plato is the longest of his works with the exception
of the Laws, and is certainly the greatest of them. There are nearer
approaches to modern metaphysics in the Philebus and in the Sophist;
the Politicus or Statesman is more ideal; the form and institutions
of the State are more clearly drawn out in the Laws; as works of art,
the Symposium and the Protagoras are of higher excellence. But no
other Dialogue of Plato has the same largeness of view and the same
perfection of style; no other shows an equal knowledge of the world,
or contains more of those thoughts which are new as well as old, and
not of one age only but of all. Nowhere in Plato is there a deeper
irony or a greater wealth of humor or imagery, or more dramatic power.
Nor in any other of his writings is the attempt made to interweave
life and speculation, or to connect politics with philosophy. The
Republic is the centre around which the other Dialogues may be grouped;
here philosophy reaches the highest point to which ancient thinkers
ever attained. Plato among the Greeks, like Bacon among the moderns,
was the first who conceived a method of knowledge, although neither
of them always distinguished the bare outline or form from the substance
of truth; and both of them had to be content with an abstraction of
science which was not yet realized. He was the greatest metaphysical
genius whom the world has seen; and in him, more than in any other
ancient thinker, the germs of future knowledge are contained. The
sciences of logic and psychology, which have supplied so many instruments
of thought to after-ages, are based upon the analyses of Socrates
and Plato. The principles of definition, the law of contradiction,
the fallacy of arguing in a circle, the distinction between the essence
and accidents of a thing or notion, between means and ends, between
causes and conditions; also the division of the mind into the rational,
```

Step 2 create a bucket called “book” on gcp and move this txt file

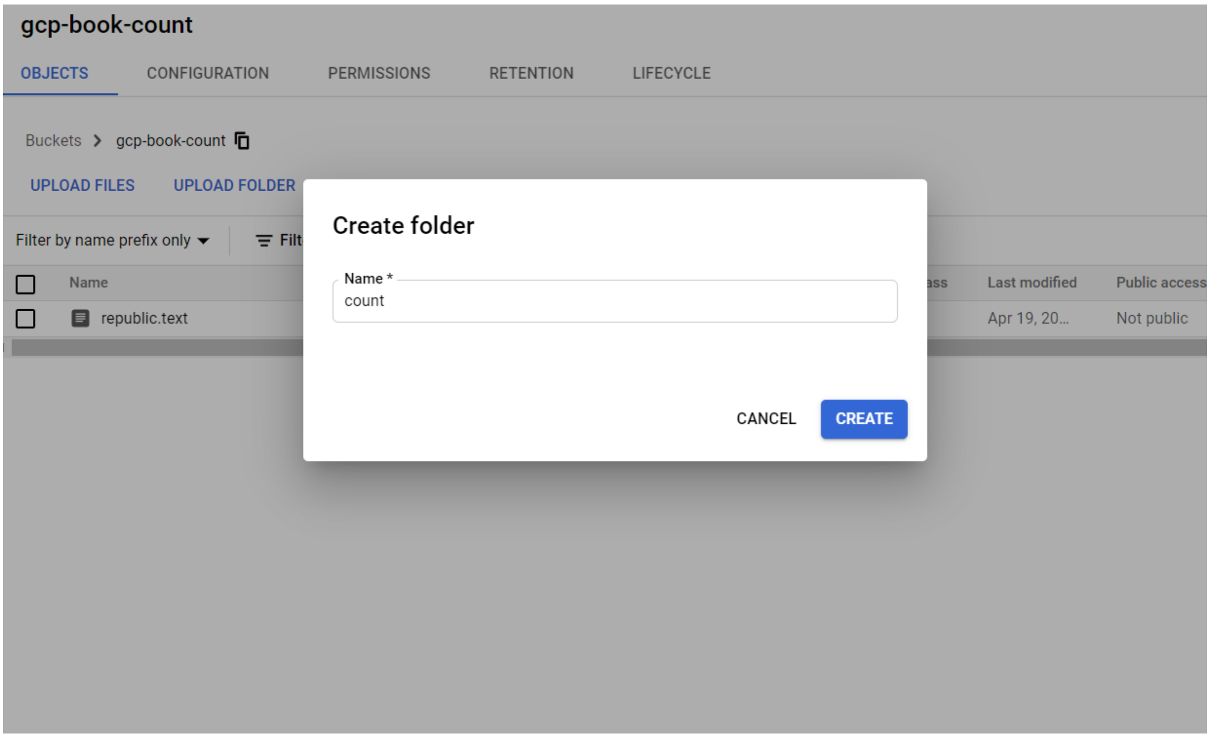


```
yogibca62@cloudshell:~ (python-application-309315)$ gsutil mb gs://gcp-book-count
Creating gs://gcp-book-count/...
yogibca62@cloudshell:~ (python-application-309315)$
```



```
PS C:\Users\Yogi> gsutil cp D:\Edureka_google_cloud_arch_cert\8\republic.text gs://gcp-book-count
Copying file://D:\Edureka_google_cloud_arch_cert\8\republic.text [Content-Type=application/octet-stream]...
- [1 files][664.9 KiB/664.9 KiB]
Operation completed over 1 objects/664.9 KiB.
PS C:\Users\Yogi>
```

Step 3 create a output folder and temp folder inside the bucket



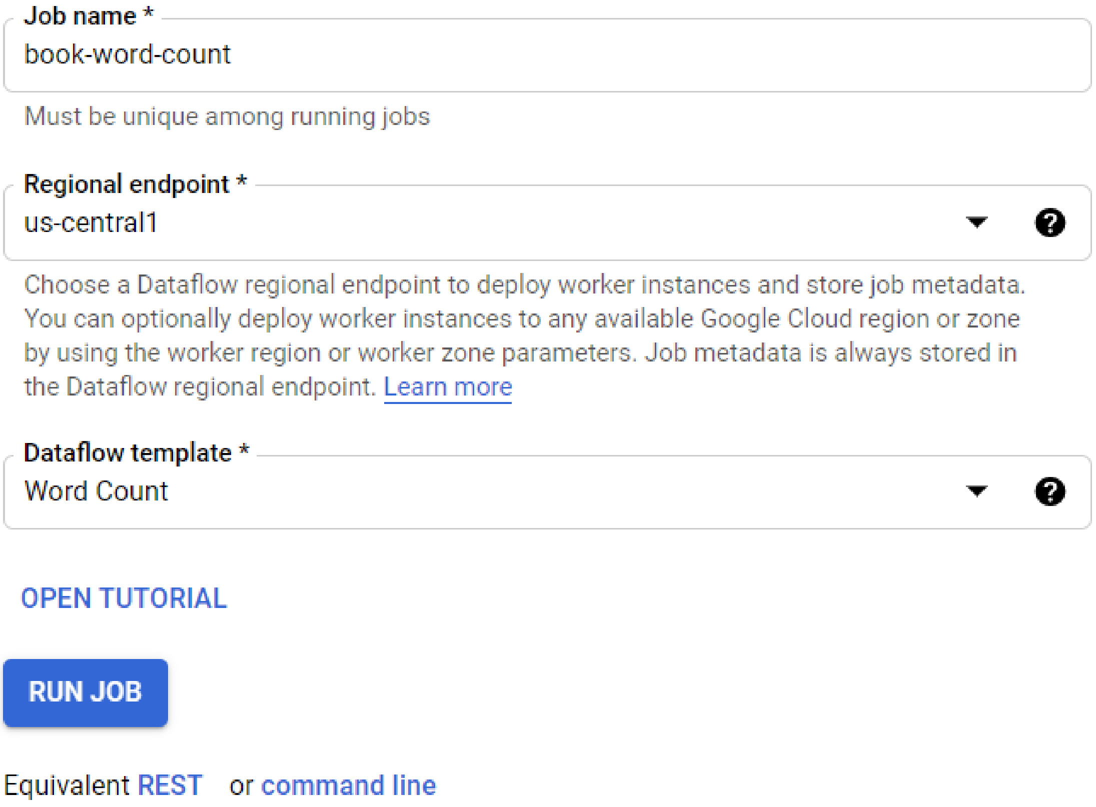
Name	Last modified	Public access
republic.text	Apr 19, 20...	Not public

Create folder

Name * count

Step 4 create a dataflow

I select word count template from dataflow



Job name *

book-word-count

Must be unique among running jobs

Regional endpoint *

us-central1

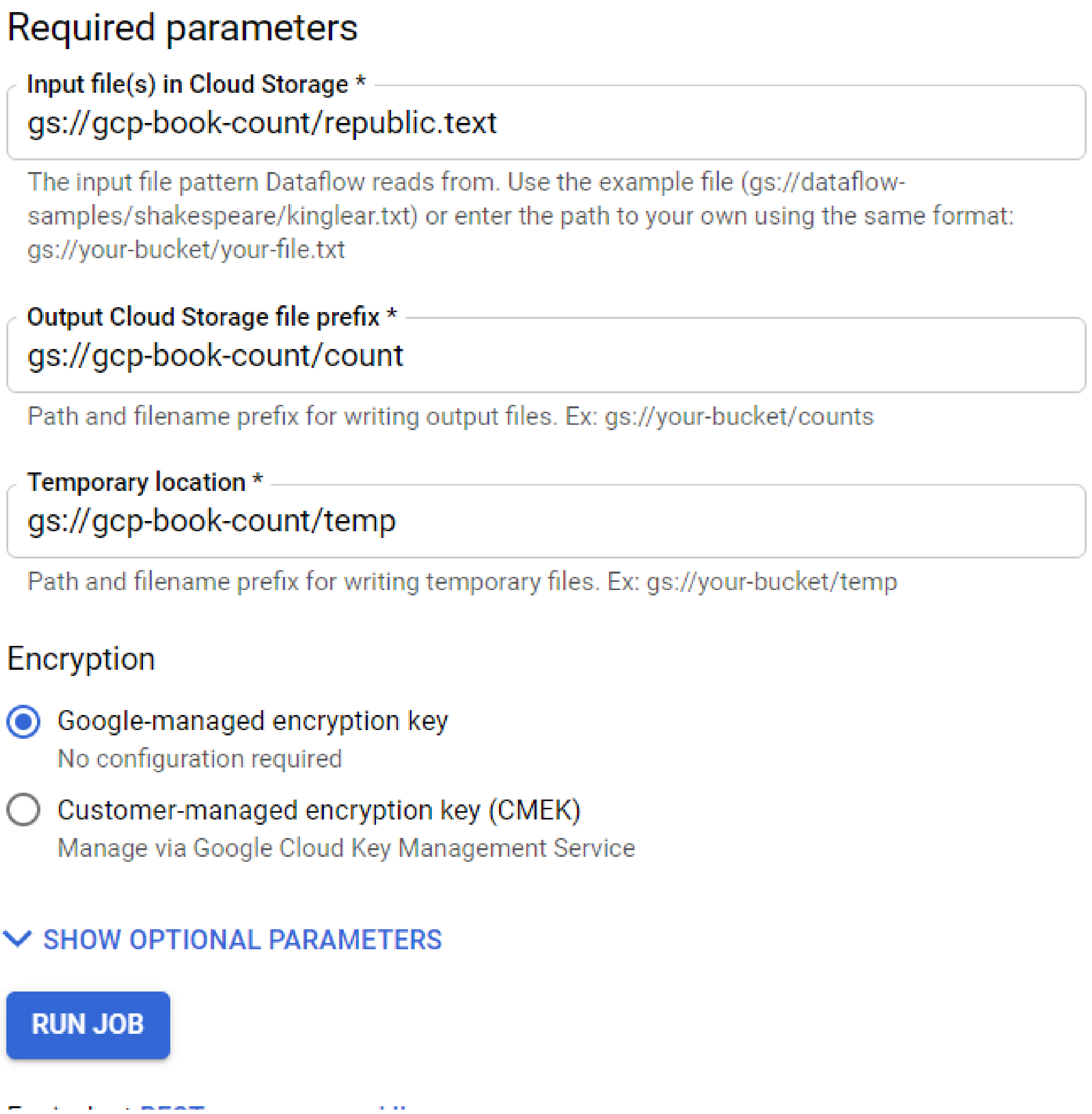
Choose a Dataflow regional endpoint to deploy worker instances and store job metadata. You can optionally deploy worker instances to any available Google Cloud region or zone by using the worker region or worker zone parameters. Job metadata is always stored in the Dataflow regional endpoint. Learn more

Dataflow template *

Word Count

Step 5

I fill necessary things



Required parameters

Input file(s) in Cloud Storage *
gs://gcp-book-count/republic.text

The input file pattern Dataflow reads from. Use the example file (gs://dataflow-samples/shakespeare/kinglear.txt) or enter the path to your own using the same format: gs://your-bucket/your-file.txt

Output Cloud Storage file prefix *
gs://gcp-book-count/count

Path and filename prefix for writing output files. Ex: gs://your-bucket/counts

Temporary location *
gs://gcp-book-count/temp

Path and filename prefix for writing temporary files. Ex: gs://your-bucket/temp

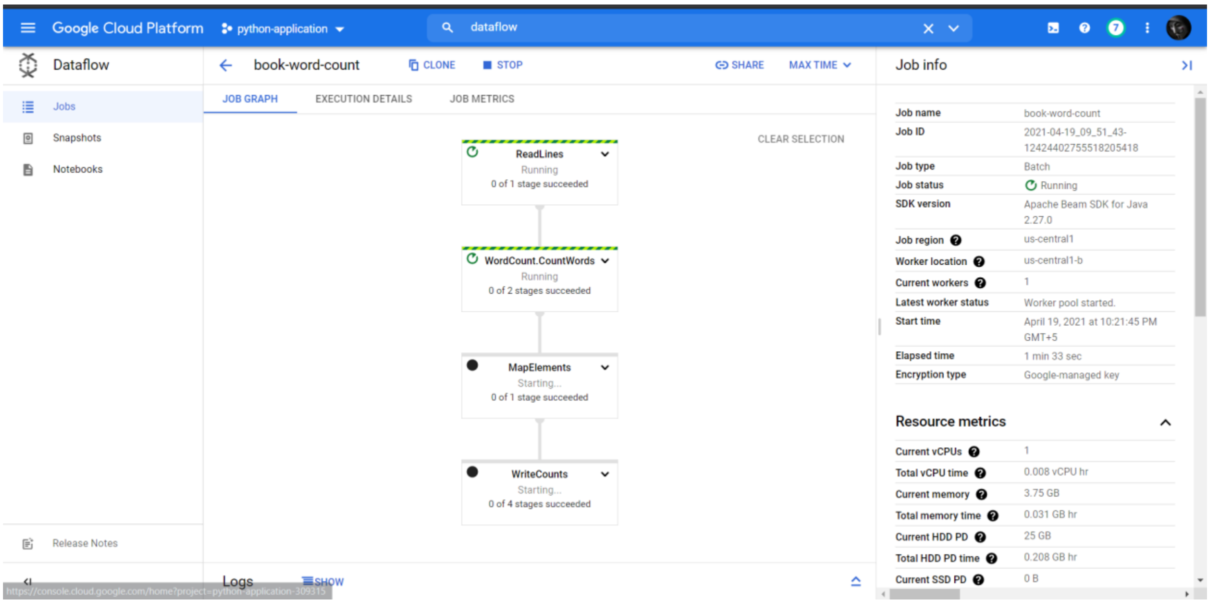
Encryption

● Google-managed encryption key
No configuration required

○ Customer-managed encryption key (CMEK)
Manage via Google Cloud Key Management Service

Step 6

Job ready



ReadLines
Running
0 of 1 stage succeeded

WordCount.CountWords
Running
0 of 2 stages succeeded

MapElements
Starting...
0 of 1 stage succeeded

WriteCounts
Starting...
0 of 4 stages succeeded

Job info

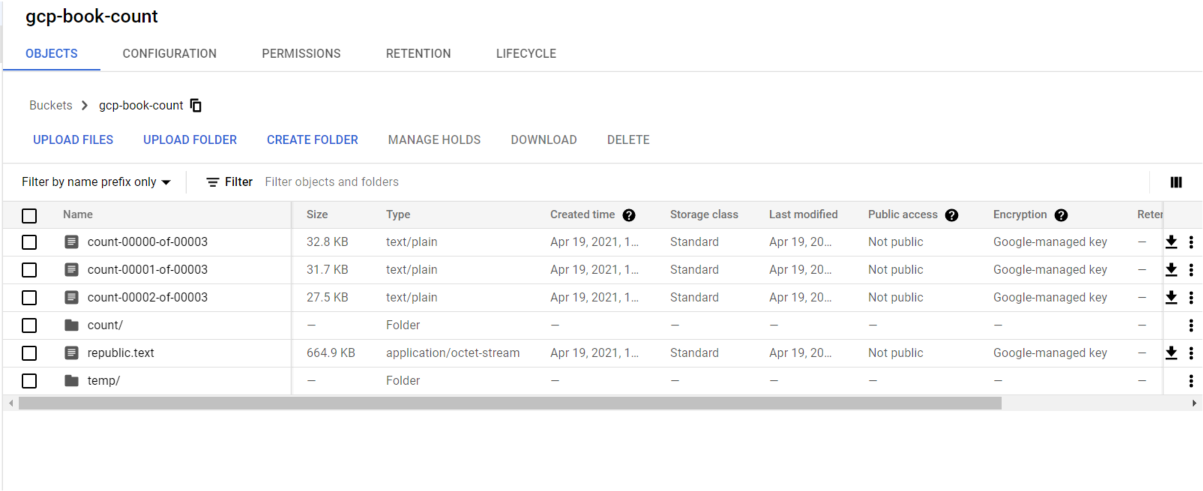
Job name	book-word-count
Job ID	2021-04-19_09_51_43-12424402755518205418
Job type	Batch
Job status	Running
SDK version	Apache Beam SDK for Java 2.27.0
Job region	us-central1
Worker location	us-central1-b
Current workers	1
Latest worker status	Worker pool started.
Start time	April 19, 2021 at 10:21:45 PM GMT+5
Elapsed time	1 min 33 sec
Encryption type	Google-managed key

Resource metrics

Current vCPUs	1
Total vCPU time	0.008 vCPU hr
Current memory	3.75 GB
Total memory time	0.031 GB hr
Current HDD PD	25 GB
Total HDD PD time	0.208 GB hr
Current SSD PD	0 B

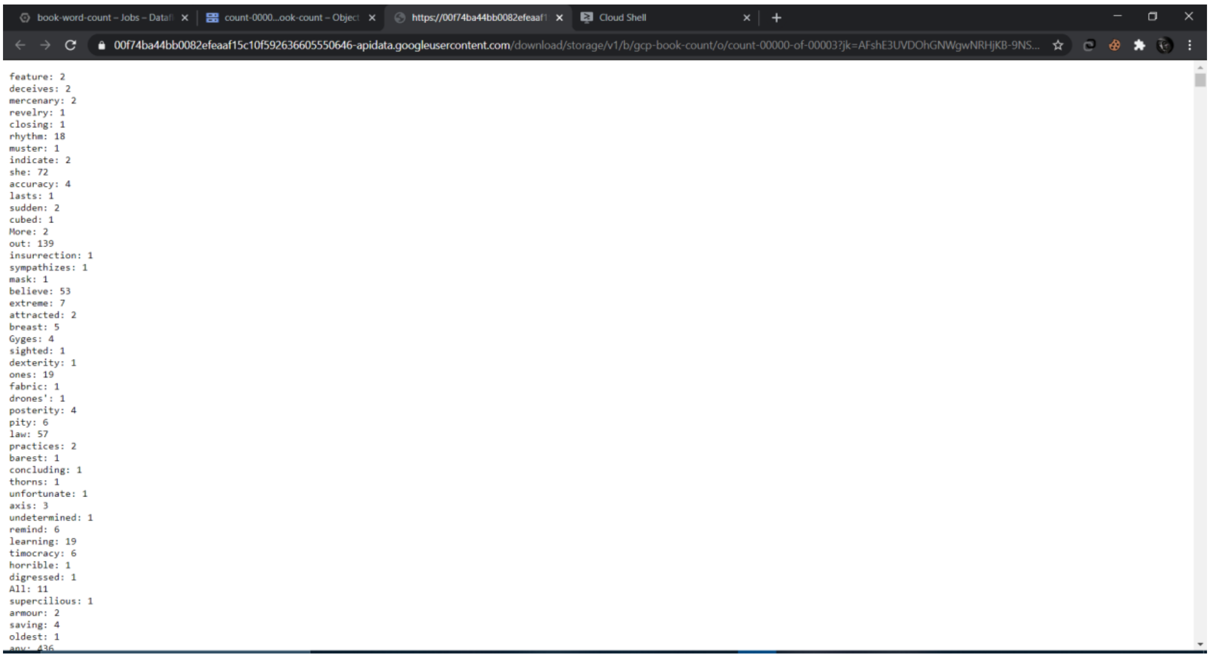
Step 7

Now I can see the outputs in my bucket



□	Name	Size	Type	Created time	Storage class	Last modified	Public access	Encryption	Reter
□	count-00000-of-00003	32.8 KB	text/plain	Apr 19, 2021, 1...	Standard	Apr 19, 20...	Not public	Google-managed key	–
□	count-00001-of-00003	31.7 KB	text/plain	Apr 19, 2021, 1...	Standard	Apr 19, 20...	Not public	Google-managed key	–
□	count-00002-of-00003	27.5 KB	text/plain	Apr 19, 2021, 1...	Standard	Apr 19, 20...	Not public	Google-managed key	–
□	count/	–	Folder	–	–	–	–	–	–
□	republic.text	664.9 KB	application/octet-stream	Apr 19, 2021, 1...	Standard	Apr 19, 20...	Not public	Google-managed key	–
□	temp/	–	Folder	–	–	–	–	–	–

Step 8


book-word-count – Jobs – Dataf
count-0000...ook-count – Object
https://00f74ba44bb0082efeaaf1
Cloud Shell
00f74ba44bb0082efeaaf15c10f592636605550646-apidata.googleusercontent.com/download/storage/v1/b/gcp-book-count/o/count-00000-of-00003?jk=AFshE3UVDOhGNWgwNRHjKB-9NS...

```
feature: 2
deceives: 2
mercenary: 2
revelry: 1
closing: 1
rhythm: 18
muster: 1
indicate: 2
she: 72
accuracy: 4
lasts: 1
sudden: 2
cubed: 1
More: 2
out: 139
insurrection: 1
sympathizes: 1
mask: 1
believe: 53
extreme: 7
attracted: 2
breast: 5
Gyges: 4
sighted: 1
dexterity: 1
ones: 19
fabric: 1
drones': 1
posterity: 4
pity: 6
law: 57
practices: 2
barest: 1
concluding: 1
thorns: 1
unfortunate: 1
axis: 3
undetermined: 1
remind: 6
learning: 19
timocracy: 6
horrible: 1
digressed: 1
All: 11
supercilious: 1
armour: 2
saving: 4
oldest: 1
any: 436
```

Output is here and I will attach my output file also from this report