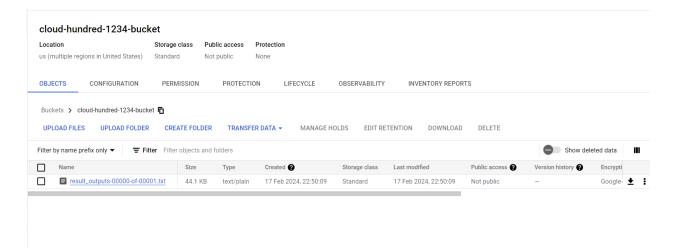Cloud Computing – Lab

1. Description of wordcount2.py

What wordcount2.py does is take text as input and parses it into exact words. In essence, it performs word counting on a file. The imports are necessary modules for building data processing pipelines. The 'WordExtractingDoFn' class is used to process lines of text in the input file and parses it into words. The 'def run' function will run the main class using the Apache pipelines. The argument parser parses arguments to specify input and output file locations. Then, the pipeline is constructed and configured. With pipeline running, the program reads from input file and saves it. Then, it will begin extracting and counting. It will split each line into separate words using the 'WordExtractingDoFn' class. Next, it converts all words to lowercase and filters words starting from a-f. It counts the occurrence of each words using 'CombinePerKey' transform. Once it finishes the extraction process, it will begin some formatting. It will format word count into strings. Finally, it writes the final words into output files using 'WriteToText' transform.

2. WordCount

```
taimour_arshad12@cloudshell:~ (cloud-thousand-12345)$ PROJECT=$(gcloud config list project --format "value(core.project)")
taimour_arshad12@cloudshell:~ (cloud-thousand-12345)$ echo $PROJECT
cloud-thousand-12345
taimour_arshad12@cloudshell:~ (cloud-thousand-12345)$ BUCKET=gs://$PROJECT-bucket
taimour_arshad12@cloudshell:~ (cloud-thousand-12345)$ echo $BUCKET
gs://cloud-thousand-12345-bucket
taimour_arshad12@cloudshell:~ (cloud-thousand-12345)$ []
```

Command lines

Word Count Output

Word Count Output in Bucket

3. MINST

Explorer   + ADD

Type to search

Viewing resources.
SHOW STARRED ONLY

⌄ storied-program-413500
  ▷ 🔗 External connections
  ⌄ 🔲 MNIST
      📊 Images
        SHOW MORE

SUMMARY

Nothing currently selected

‹ ✅ New code manageme ›

Try the Google Trends demo query
This simple query generates the top search terms in the US from the Google Trends public data set.

🔍 OPEN THIS QUERY   VIEW DATASET

Add your own data

⬆️ **Local file**	△ **Google Drive**	▦ **Google Cloud Storage**
Upload a local file	Google storage service	Google object storage service
LAUNCH THIS GUIDE	LAUNCH THIS GUIDE	LAUNCH THIS GUIDE

☑ Show welcome page on startup

Job history   ⟳ REFRESH

Images created.   GO TO TABLE   ✕

CLOUD SHELL

Read from Pub
Running
0 sec
1 stage

toDict
Running
5 sec
1 stage

Prediction
Running
17 min 34 sec
1 stage

to byte
Running
0 sec
1 stage

to Pub
Running
0 sec