

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

```
KING: 243
LEAR: 236
DRAMATIS: 1
PERSONAE: 1
king: 65
of: 447
Britain: 2
OF: 15
FRANCE: 10
DUKE: 3
BURGUNDY: 8
CORNWALL: 63
ALBANY: 67
EARL: 2
KENT: 156
GLOUCESTER: 141
EDGAR: 126
son: 29
to: 438
Gloucester: 26
EDMUND: 99
bastard: 7
CURAN: 6
a: 366
courtier: 1
Old: 13
Man: 11
```

Word Count Output

cloud-hundred-1234-bucket

Location

us (multiple regions in United States)

Storage class

Standard

Public access

Not public

Protection

None

OBJECTS

CONFIGURATION

PERMISSION

PROTECTION

LIFECYCLE

OBSERVABILITY

INVENTORY REPORTS

Buckets > cloud-hundred-1234-bucket

UPLOAD FILES

UPLOAD FOLDER

CREATE FOLDER

TRANSFER DATA

MANAGE HOLDS

EDIT RETENTION

DOWNLOAD




DELETE

Filter by name prefix only

Filter

Filter objects and folders

Show deleted data

	Name	Size	Type	Created	Storage class	Last modified	Public access	Version history	Encrypti
<input type="checkbox"/>	 result_outputs-00000-of-00001.txt	44.1 KB	text/plain	17 Feb 2024, 22:50:09	Standard	17 Feb 2024, 22:50:09	Not public	—	Google-  

Word Count Output in Bucket

3. MINST

🔍

Your BigQuery projects will have new capabilities after February 14, 2024. Services and roles will be enabled automatically to help with these changes. [Learn more](#)

DISMISS

Explorer

+ ADD

⏪

🔍

Type to search

🔍

Viewing resources.

SHOW STARRED ONLY

▼ storied-program-413500

▶ External connections

▼ MNIST

▶ Images

SHOW MORE

SUMMARY

▼

Nothing currently selected

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

Upload a local file

LAUNCH THIS GUIDE

Google Drive

Google storage service

LAUNCH THIS GUIDE

Google Cloud Storage

Google object storage service

LAUNCH THIS GUIDE

☑ Show welcome page on startup

Job history

Images created.

GO TO TABLE

×

REFRESH

⬆

