Tony Nguyen

Dr. Shawn Bowers

CPSC 324 01

12 February 2024

Homework 1

1. **Step 1**

a. *Question a*

- IAM (Cloud Identity and Access Management) is used to inspect and modify these roles and permissions.
- Three types of IAM basic roles
    - Roles/Viewers
        - Read-only actions
        - Do not affect state
        - E.g., viewing but not modifying the existing resources or data
    - Roles/Editor
        - All viewer permission
        - Also permit actions that modify state
    - Roles/Owner
        - All editor and viewer permissions
        - Also allows managing roles and permissions for a project and setting up billing for a project

b. *Question b*

- Storage Admin

  - Grant full control of objects and buckets
  - Only applied to an individual bucket and objects within it
- Storage Object Admin
  - Grant full control of objects
  - Including listing, creating, viewing, and deleting
  - Note that only objects are granted access, not the entire bucket
- Storage Folder Admin
  - In Beta version
  - Grant full control over folders and objects
  - Including listing, creating, viewing, and deleting
  - Higher privilege than Storage Object Admin but still lower than Storage Admin

c. *Question c*

- Cloud Translation API
  - Integrate text translation into a website or application
  - $45/hour use, which is quite expensive
  - I realize that, based on my personal experience, Google's translation service is getting smarter and more reliable. The wording and semantics sound more natural.
- Cloud Vision API
  - Integrate the Google Vision feature
  - Able to label images, faces, logos, and landmarks, OCR, and more into applications

- Potentially easier to use compared to Keras OCR as this API leverages a wide range of pre-trained data
- 3.5 cents per count

- Vertex AI API
  - Powerful tool to train custom machine learning models with minimal machine learning expertise and efforts
  - Leverage Gemini, a generative AI model, to power the API
  - Has the ability to recognize and understand essentially any inputs

2. **Step 2**

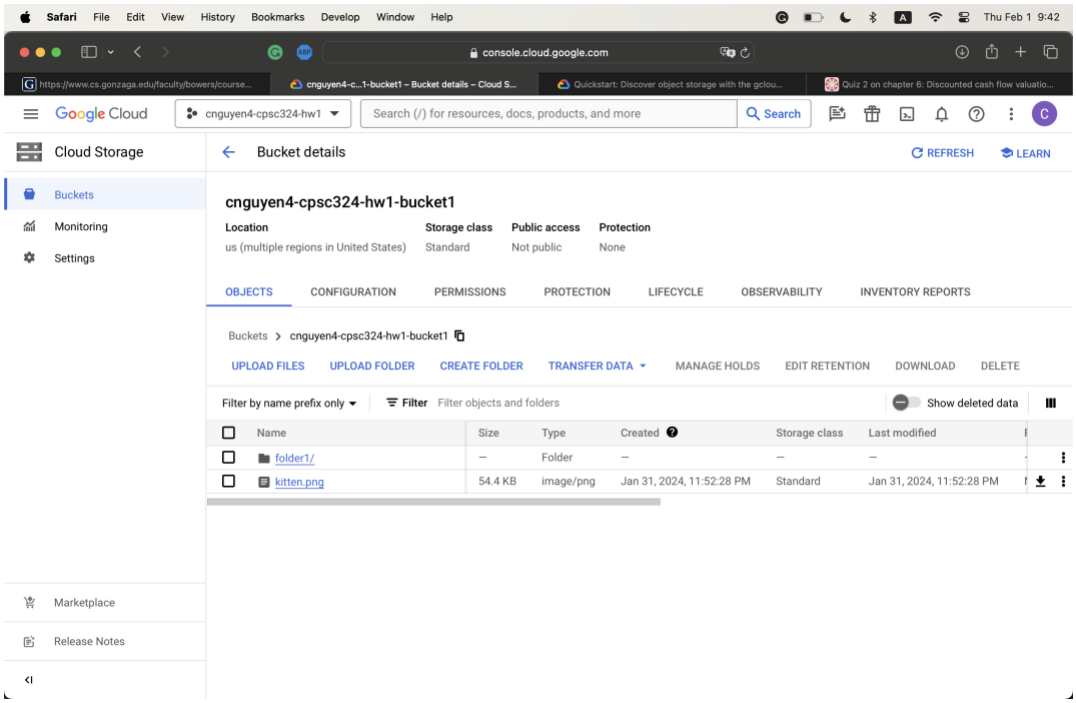
```
(cpsc322) tony@Tonys-MacBook-Pro CPSC324 % gcloud -h
Usage: gcloud [optional flags] <group | command>
  group may be           access-approval | access-context-manager |
                         active-directory | ai | ai-platform | alloydb |
                         anthos | api-gateway | apigee | app | artifacts |
                         asset | assured | auth | backup-dr | batch | bigtable |
                         billing | bms | builds | certificate-manager |
                         cloud-shell | components | composer | compute |
                         config | container | data-catalog |
                         database-migration | dataflow | dataplex | dataproc |
                         datastore | datastream | deploy | deployment-manager |
                         dns | domains | edge-cache | edge-cloud | emulators |
                         endpoints | essential-contacts | eventarc | filestore |
                         firebase | firestore | functions | healthcare | iam |
                         iap | identity | ids | immersive-stream |
                         infra-manager | iot | kms | logging | looker |
                         memcache | metastore | ml | ml-engine | monitoring |
                         netapp | network-connectivity | network-management |
                         network-security | network-services | notebooks |
                         org-policies | organizations | policy-intelligence |
                         policy-troubleshoot | privateca | projects | publicca |
                         pubsub | recaptcha | recommender | redis |
                         resource-manager | resource-settings | run | scc |
                         scheduler | secrets | service-directory | services |
                         source | spanner | sql | storage | tasks |
                         telco-automation | topic | transcoder | transfer |
                         vmware | workbench | workflows | workspace-add-ons |
                         workstations
  command may be         cheat-sheet | docker | feedback | help | info | init |
                         survey | version

For detailed information on this command and its flags, run:
  gcloud --help
(cpsc322) tony@Tonys-MacBook-Pro CPSC324 %
```

3. **Step 3**

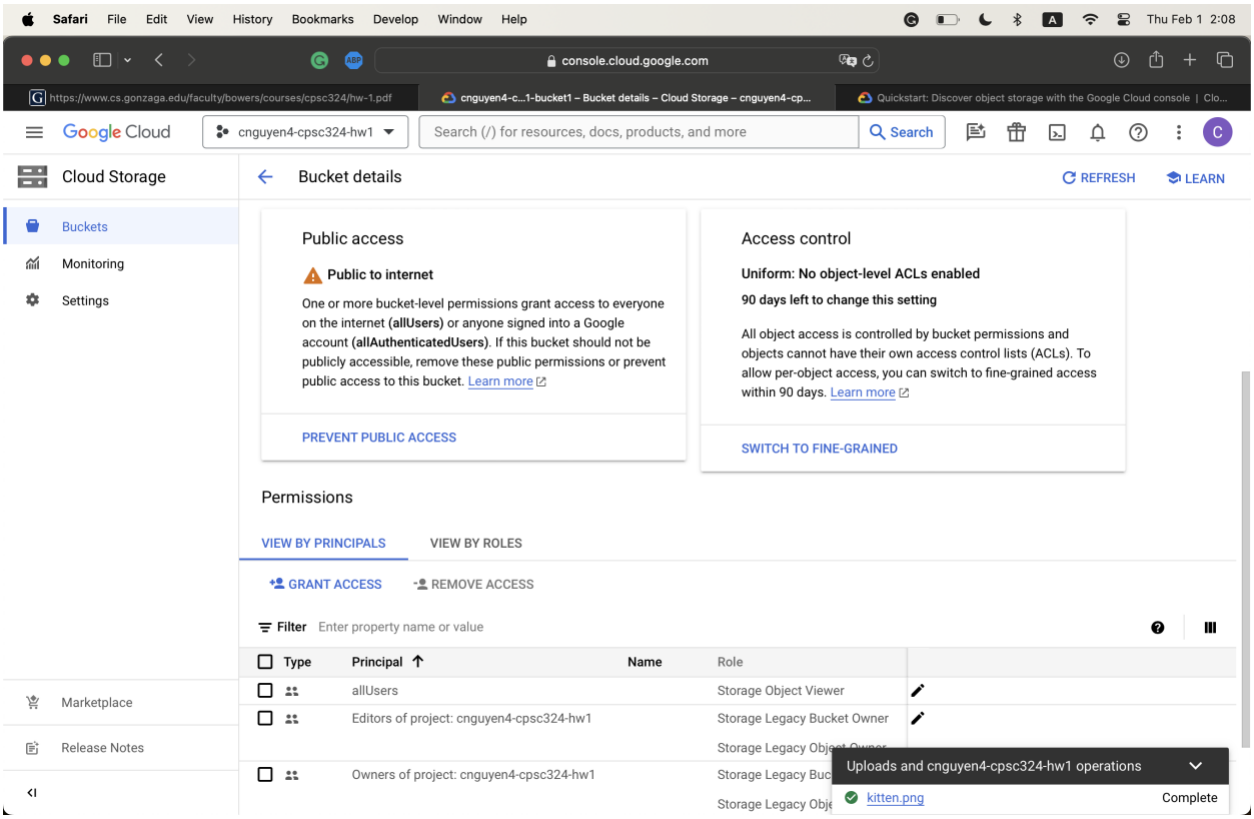
- Commands
  - Initialize the gcloud CLI

        - gcloud init
    - Create a new bucket
        - Buckets are the basic containers that hold data
        - gcloud storage buckets create gs://cnguyen4-cpsc324-hw1-bucket1 --uniform-bucket-level-access
    - Copy an image from a location on your local machine to a bucket
        - gcloud storage cp Downloads/kitten.png gs://cnguyen4-cpsc324-hw1-bucket1
    - Download an object from your bucket
        - gcloud storage cp gs://cnguyen4-cpsc324-hw1-bucket1/kitten.png Downloads/kitten.png
        - Note that this is similar to uploading a resource locally to Google Cloud
    - Copy an object to a folder in the bucket
        - gcloud storage cp gs://cnguyen4-cpsc324-hw1-bucket1/kitten.png gs://cnguyen4-cpsc324-hw1-bucket1/folder1/kitten3.png
        - Note the same structure
    - List operation
        - gcloud storage ls gs://cnguyen4-cpsc324-hw1-bucket1/
        - gcloud storage ls gs://cnguyen4-cpsc324-hw1-bucket1/ --long
        - Similar to the Linux command line
        - Note that the --*long* tag returns detailed image information
    - Make the objects publicly available

        - gcloud storage buckets add-iam-policy-binding gs://cnguyen4-cpsc324-hw1-bucket1/ --member=allUsers --role=roles/storage.objectViewer
        - gcloud storage buckets remove-iam-policy-binding gs://cnguyen4-cpsc324-hw1-bucket1/ --member=allUsers --role=roles/storage.objectViewer
        - Note that adding and removing have similar structures
        - If we want to add and/or remove a specific user, set --*member=user:jane@gmail.com*. Similar structure above
    - Delete an object
        - gcloud storage rm gs://cnguyen4-cpsc324-hw1-bucket1/kitten.png
    - Clean up
        - gcloud storage rm gs://cnguyen4-cpsc324-hw1-bucket1/ --recursive
- Screenshot before deleting



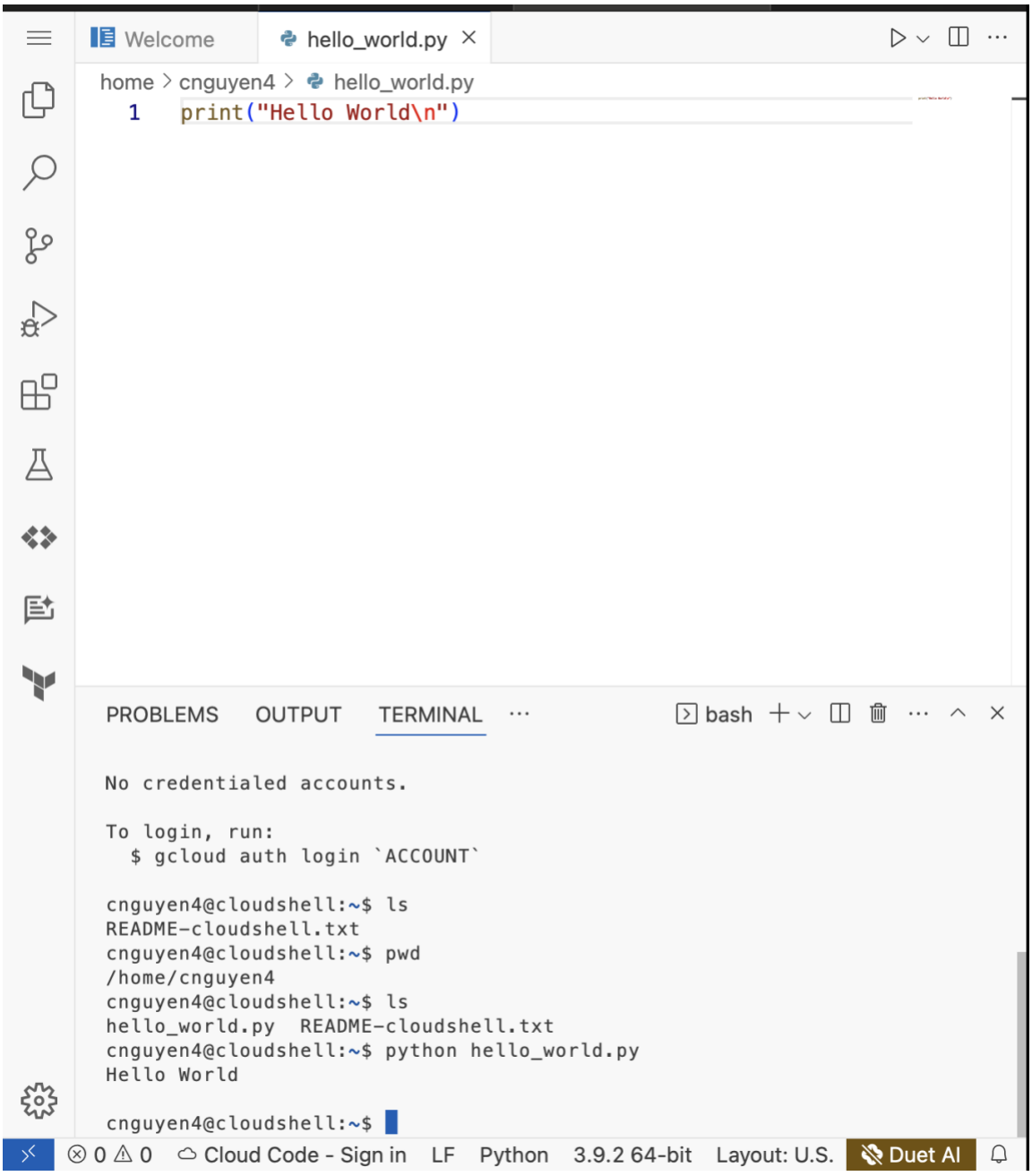
4. Step 4



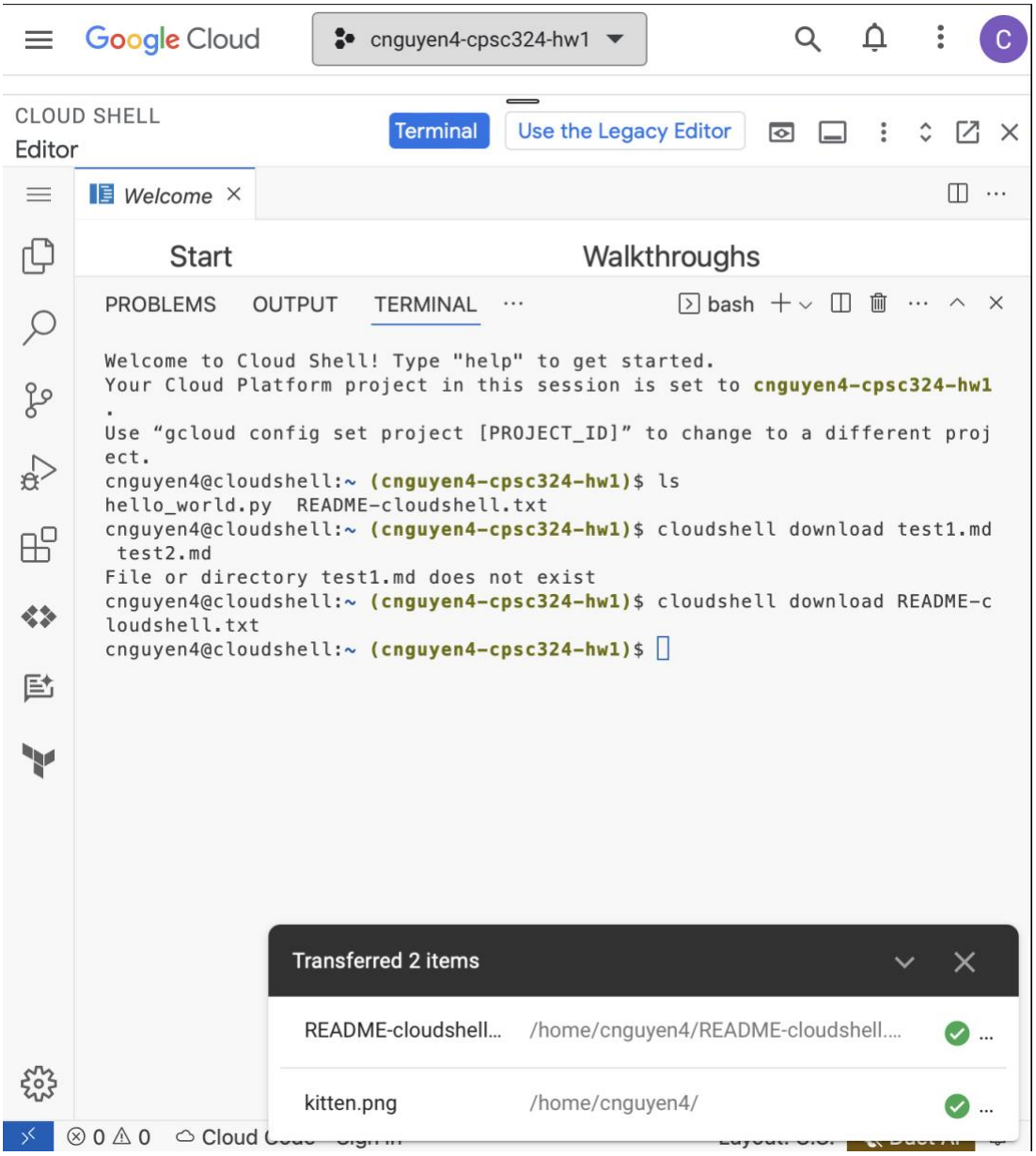
- https://storage.cloud.google.com/cnguyen4-cpsc324-hw1-bucket1/kitten.png?authuser=2.

5. Step 5

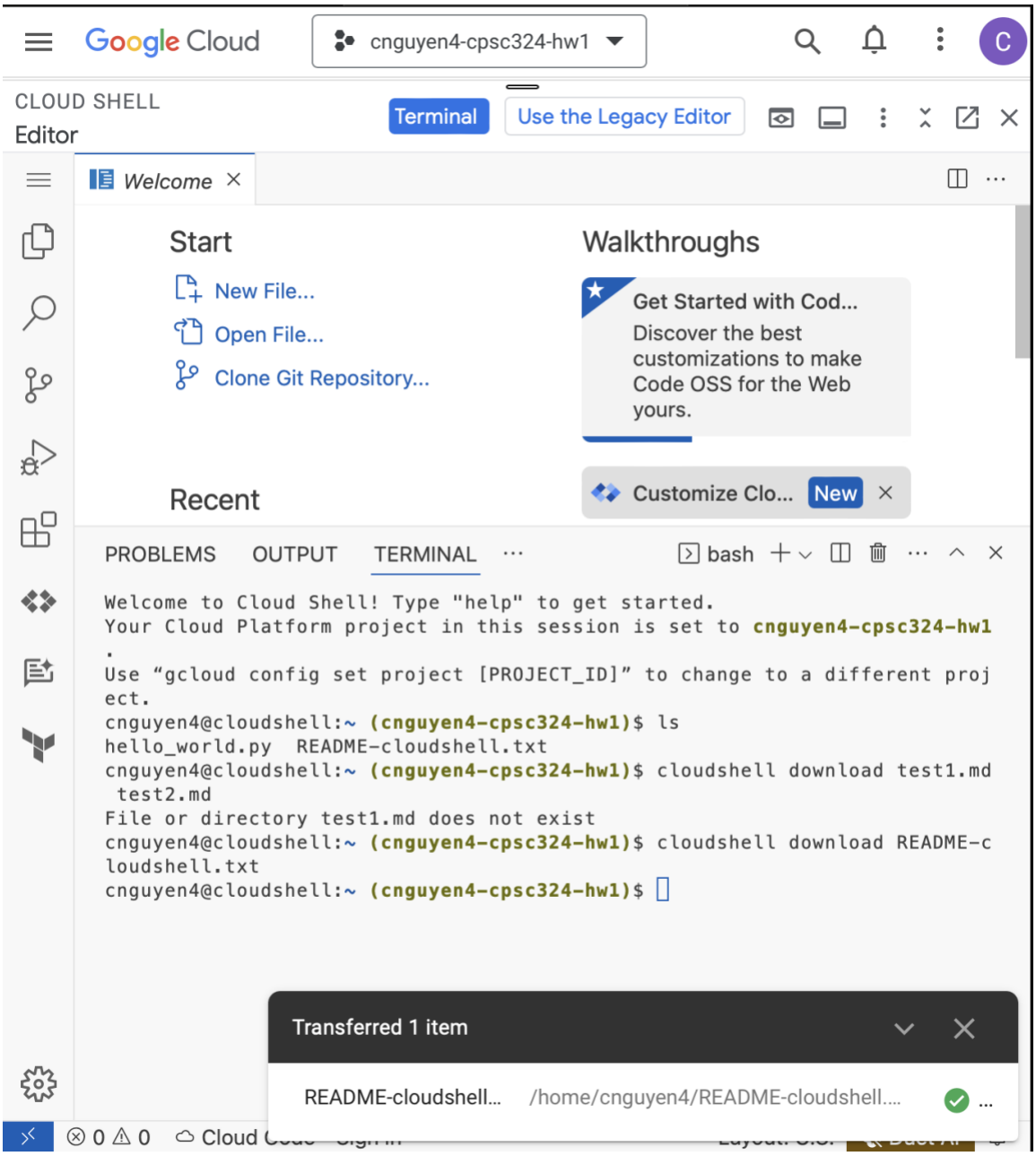
- Ephemeral mode means that files created during a given session are deleted after it ends
- Using Cloud Shell to run a Python Script



- Using Cloud Shell to upload a file

- 

- Using Cloud Shell to download a file
  - cloudshell download *file_name*



- Using a local terminal to transfer/download files from Cloud Shell to the Local environment
    - gcloud cloud-shell scp cloudshell:/home/cnguyen4/kitten.png localhost:kitten.png
    - Note that this command should be run within the target director=]
    - Tag cloudshell and localhost can be interchanged to depict the side of transfer

```
(cpsc322) tony@Tonys-MacBook-Pro hw1-tonixsmm % gcloud cloud-shell scp cloudshell:/home/cnguyen4/kitten.png localhost:kitten.png
kitten.png                                                    100%   54KB 481.7KB/s   00:00
(cpsc322) tony@Tonys-MacBook-Pro hw1-tonixsmm % ls
hw1-lab-notebook.docx   kitten.png              ~$1-lab-notebook.docx
jobs_in_data.csv        ufo_sightings.zip
(cpsc322) tony@Tonys-MacBook-Pro hw1-tonixsmm %
```

6. Step 6

```
(cpsc322) tony@Tonys-MacBook-Pro hw1-tonixsmm % gcloud storage cp jobs_in_data.csv gs://cnguyen4-cpsc324-hw1-bucket1
Copying file://jobs_in_data.csv to gs://cnguyen4-cpsc324-hw1-bucket1/jobs_in_data.csv
  Completed files 1/1 | 1.1MiB/1.1MiB
(cpsc322) tony@Tonys-MacBook-Pro hw1-tonixsmm % gcloud storage ls gs://cnguyen4-cpsc324-hw1-bucket1 --long
   1104818  2024-02-12T18:49:35Z  gs://cnguyen4-cpsc324-hw1-bucket1/jobs_in_data.csv
TOTAL: 1 objects, 1104818 bytes (1.05MiB)
(cpsc322) tony@Tonys-MacBook-Pro hw1-tonixsmm % gcloud storage cp ufo_sightings.csv gs://cnguyen4-cpsc324-hw1-bucket1
Copying file://ufo_sightings.csv to gs://cnguyen4-cpsc324-hw1-bucket1/ufo_sightings.csv
  Completed files 1/1 | 14.6MiB/14.6MiB
(cpsc322) tony@Tonys-MacBook-Pro hw1-tonixsmm % gcloud storage ls gs://cnguyen4-cpsc324-hw1-bucket1 --long
   1104818  2024-02-12T18:49:35Z  gs://cnguyen4-cpsc324-hw1-bucket1/jobs_in_data.csv
  15349323  2024-02-12T18:52:01Z  gs://cnguyen4-cpsc324-hw1-bucket1/ufo_sightings.csv
TOTAL: 2 objects, 16454141 bytes (15.69MiB)
(cpsc322) tony@Tonys-MacBook-Pro hw1-tonixsmm % gcloud storage cp Home_and_Kitchen_5.json gs://cnguyen4-cpsc324-hw1-bucket1
WARNING: Parallel composite upload was turned ON to get the best performance on
uploading large objects. If you would like to opt-out and instead
perform a normal upload, run: `gcloud config set
storage/parallel_composite_upload_enabled False` If you would like to
disable this warning, run: `gcloud config set
storage/parallel_composite_upload_enabled True` Note that with
parallel composite uploads, your object might be uploaded as a
composite object (https://cloud.google.com/storage/docs/composite-
objects), which means that any user who downloads your object will
need to use crc32c checksums to verify data integrity. gcloud storage
is capable of computing crc32c checksums, but this might pose a
problem for other clients.

Copying file://Home_and_Kitchen_5.json to gs://cnguyen4-cpsc324-hw1-bucket1/Home_and_Kitchen_5.json
  Completed files 32/1 | 3.2GiB/3.2GiB | 683.8kiB/s

Average throughput: 4.0MiB/s
(cpsc322) tony@Tonys-MacBook-Pro hw1-tonixsmm % gcloud storage ls gs://cnguyen4-cpsc324-hw1-bucket1 --long
3480167832  2024-02-12T19:06:30Z  gs://cnguyen4-cpsc324-hw1-bucket1/Home_and_Kitchen_5.json
   1104818  2024-02-12T18:49:35Z  gs://cnguyen4-cpsc324-hw1-bucket1/jobs_in_data.csv
  15349323  2024-02-12T18:52:01Z  gs://cnguyen4-cpsc324-hw1-bucket1/ufo_sightings.csv
TOTAL: 3 objects, 3496621973 bytes (3.26GiB)
(cpsc322) tony@Tonys-MacBook-Pro hw1-tonixsmm %
```

7. Step 7

- Download and unzip the file

```
cnguyen4@cloudshell:~ (cnguyen4-cpsc324-hw1)$ wget -v https://datarepo.en
g.ucsd.edu/mcauley_group/data/amazon_v2/categoryFilesSmall/Home_and_Kitch
en_5.json.gz
--2024-02-12 21:53:14--  https://datarepo.eng.ucsd.edu/mcauley_group/data
/amazon_v2/categoryFilesSmall/Home_and_Kitchen_5.json.gz
Resolving datarepo.eng.ucsd.edu (datarepo.eng.ucsd.edu)... 132.239.8.30
Connecting to datarepo.eng.ucsd.edu (datarepo.eng.ucsd.edu)|132.239.8.30|
:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 953876340 (910M) [application/x-gzip]
Saving to: ‘Home_and_Kitchen_5.json.gz’

Home_and_Kitchen_5 100%[=============>] 909.69M  45.9MB/s    in 16s

2024-02-12 21:53:30 (57.2 MB/s) - ‘Home_and_Kitchen_5.json.gz’ saved [953
876340/953876340]

cnguyen4@cloudshell:~ (cnguyen4-cpsc324-hw1)$ gunzip Home_and_Kitchen_5.j
son.gz
```

- Create a new bucket and upload it

```
cnguyen4@cloudshell:~ (cnguyen4-cpsc324-hw1)$ gcloud storage buckets crea
te gs://cnguyen4-cpsc324-hw1-bucket1 --uniform-bucket-level-access
Creating gs://cnguyen4-cpsc324-hw1-bucket1/...
cnguyen4@cloudshell:~ (cnguyen4-cpsc324-hw1)$ gcloud storage cp Home_and_
Kitchen_5.json gs://cnguyen4-cpsc324-hw1-bucket1
WARNING: Parallel composite upload was turned ON to get the best performa
nce on
uploading large objects. If you would like to opt-out and instead
perform a normal upload, run: `gcloud config set
storage/parallel_composite_upload_enabled False` If you would like to
disable this warning, run: `gcloud config set
storage/parallel_composite_upload_enabled True` Note that with
parallel composite uploads, your object might be uploaded as a
composite object (https://cloud.google.com/storage/docs/composite-
objects), which means that any user who downloads your object will
need to use crc32c checksums to verify data integrity. gcloud storage
is capable of computing crc32c checksums, but this might pose a
problem for other clients.

Copying file://Home_and_Kitchen_5.json to gs://cnguyen4-cpsc324-hw1-bucke
t1/Home_and_Kitchen_5.json
  Completed files 32/1 | 3.2GiB/3.2GiB | 27.8MiB/s

Average throughput: 66.9MiB/s
cnguyen4@cloudshell:~ (cnguyen4-cpsc324-hw1)$
```

8. **Step 8**

```
(cpsc322) (cpsc322) tony@Tonys-MacBook-Pro hw1-tonixsmm % python hw1_create_bucket.py
Bucket cnguyen4_cpsc324_hw1_bucket1 created.
(cpsc322) (cpsc322) tony@Tonys-MacBook-Pro hw1-tonixsmm % gcloud storage ls
gs://cnguyen4_cpsc324_hw1_bucket1/
(cpsc322) (cpsc322) tony@Tonys-MacBook-Pro hw1-tonixsmm %
```

9. Step 9

```
(cpsc322) (cpsc322) tony@Tonys-MacBook-Pro hw1-tonixsmm % python hw1_gcs_upload.py
Enter a name for the new bucket: cnguyen4_cpsc324_hw1_bucket1
Creating bucket cnguyen4_cpsc324_hw1_bucket1...
Bucket cnguyen4_cpsc324_hw1_bucket1 created.
Enter the name of the file to upload: ufo_sightings.csv
Uploading file ufo_sightings.csv...
File ufo_sightings.csv uploaded.
<Blob: cnguyen4_cpsc324_hw1_bucket1, ufo_sightings.csv, 1707802996213099>
```

```
(cpsc322) (cpsc322) tony@Tonys-MacBook-Pro hw1-tonixsmm % python hw1_gcs_upload.py
Enter a name for the new bucket: cnguyen4_cpsc324_hw1_bucket1
Bucket cnguyen4_cpsc324_hw1_bucket1 already exists.
Enter the name of the file to upload: jobs_in_data.csv
Uploading file jobs_in_data.csv...
File jobs_in_data.csv uploaded.
<Blob: cnguyen4_cpsc324_hw1_bucket1, jobs_in_data.csv, 1707803930396373>
<Blob: cnguyen4_cpsc324_hw1_bucket1, ufo_sightings.csv, 1707802996213099>
(cpsc322) (cpsc322) tony@Tonys-MacBook-Pro hw1-tonixsmm %
```

```
(cpsc322) (cpsc322) tony@Tonys-MacBook-Pro hw1-tonixsmm % python hw1_gcs_upload.py
Enter a name for the new bucket: cnguyen4_cpsc324_hw1_bucket2
Creating bucket cnguyen4_cpsc324_hw1_bucket2...
Bucket cnguyen4_cpsc324_hw1_bucket2 created.
Enter the name of the file to upload: ufo_sightings.csv
Uploading file ufo_sightings.csv...
File ufo_sightings.csv uploaded.
<Blob: cnguyen4_cpsc324_hw1_bucket2, ufo_sightings.csv, 1707804066814712>
(cpsc322) (cpsc322) tony@Tonys-MacBook-Pro hw1-tonixsmm % python hw1_gcs_upload.py
Enter a name for the new bucket: cnguyen4_cpsc324_hw1_bucket2
Bucket cnguyen4_cpsc324_hw1_bucket2 already exists.
Enter the name of the file to upload: jobs_in_data.csv
Uploading file jobs_in_data.csv...
File jobs_in_data.csv uploaded.
<Blob: cnguyen4_cpsc324_hw1_bucket2, jobs_in_data.csv, 1707804088495421>
<Blob: cnguyen4_cpsc324_hw1_bucket2, ufo_sightings.csv, 1707804066814712>
```