

TASK 4 - Big Query

1. Buat sebuah bucket di GCS. upload beberapa file ke bucket tersebut!

- a. Pilih menu create pada halaman buckets, create bucket dengan memperhatikan region yang kita pilih yang nantinya disamakan dengan region saat kita membuat dataset tempat table yang akan kita buat, karena region merupakan lokasi data center kita menyimpan data yang kita upload di GCP.

The screenshot shows the Google Cloud Storage Buckets page. The 'CREATE' button is circled in red. Below the main content area, a table lists the created bucket 'bucket-alta-de-4-yovinasilvia' with its location 'asia-southeast1' also circled in red. A notification at the bottom states 'Created bucket bucket-alta-de-4-yovinasilvia'.

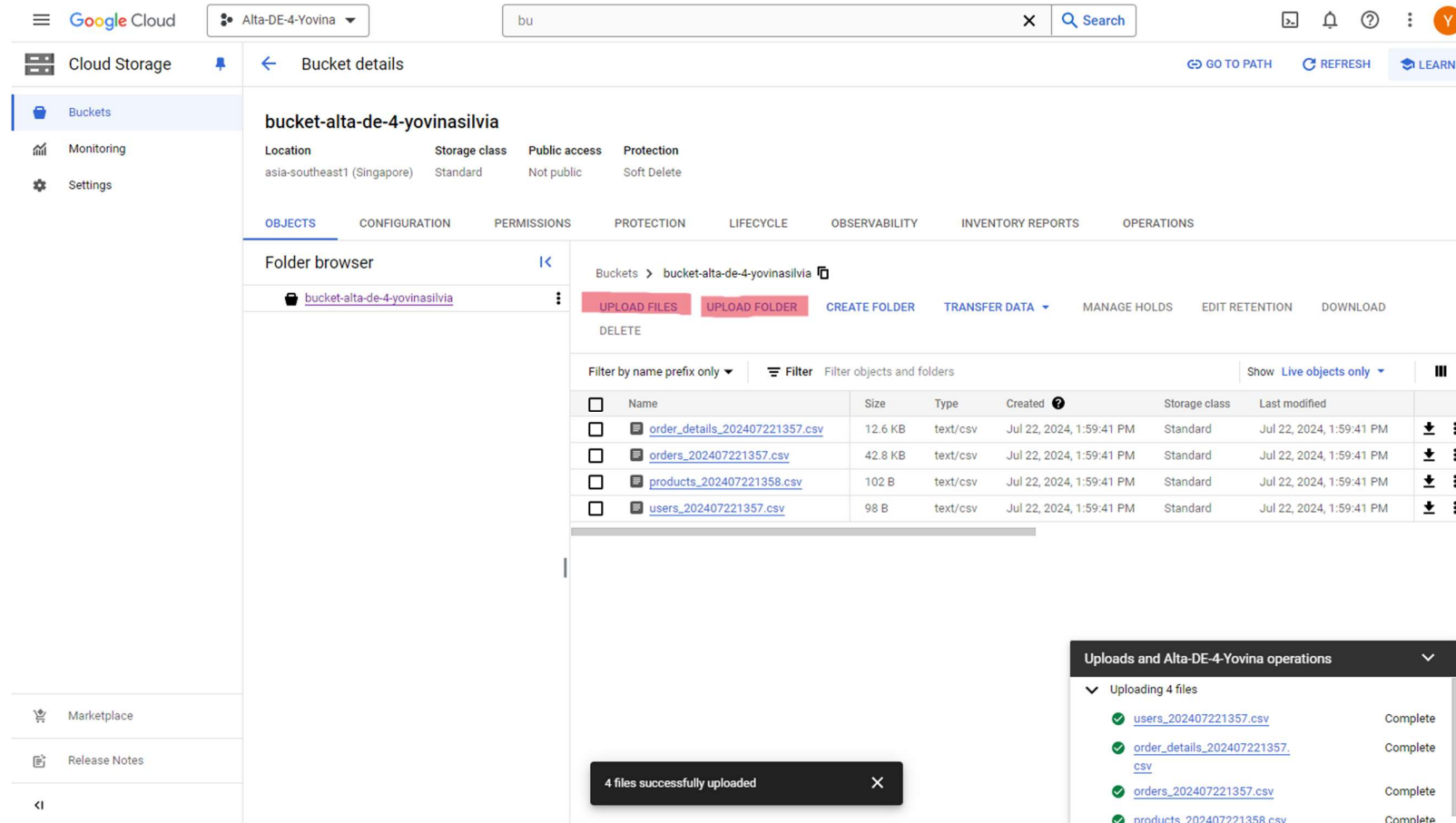
Name	Created	Location type	Location	Default storage class	Last modified	Public access	Access control
bucket-alta-de-4-yovinasilvia	Jul 22, 2024, 1:36:57 PM	Region	asia-southeast1	Standard	Jul 22, 2024, 1:36:57 PM	Not public	Uniform

DATA ENGINEER BATCH 4

Mentee: Yovina Silvia

Mentor: Bilal Benefit

- b. Upload file dari local komputer kita bisa dengan drag and drop dari penyimpanan local atau pilih menu upload pada halaman GCP.



The screenshot displays the Google Cloud Storage console for a bucket named 'bucket-alta-de-4-yovinasilvia'. The interface includes a sidebar with navigation options like Buckets, Monitoring, and Settings. The main area shows bucket details such as Location (asia-southeast1), Storage class (Standard), Public access (Not public), and Protection (Soft Delete). Below this, there are tabs for OBJECTS, CONFIGURATION, PERMISSIONS, PROTECTION, LIFECYCLE, OBSERVABILITY, INVENTORY REPORTS, and OPERATIONS. The OBJECTS tab is active, showing a folder browser view with a table of objects. The table lists four CSV files: 'order_details_202407221357.csv', 'orders_202407221357.csv', 'products_202407221358.csv', and 'users_202407221357.csv'. All files are 12.6 KB in size and were created on Jul 22, 2024, at 1:59:41 PM. A notification at the bottom center states '4 files successfully uploaded'. On the bottom right, a panel titled 'Uploads and Alta-DE-4-Yovina operations' shows the upload progress for the four files, all marked as 'Complete'.

Name	Size	Type	Created	Storage class	Last modified
order_details_202407221357.csv	12.6 KB	text/csv	Jul 22, 2024, 1:59:41 PM	Standard	Jul 22, 2024, 1:59:41 PM
orders_202407221357.csv	42.8 KB	text/csv	Jul 22, 2024, 1:59:41 PM	Standard	Jul 22, 2024, 1:59:41 PM
products_202407221358.csv	102 B	text/csv	Jul 22, 2024, 1:59:41 PM	Standard	Jul 22, 2024, 1:59:41 PM
users_202407221357.csv	98 B	text/csv	Jul 22, 2024, 1:59:41 PM	Standard	Jul 22, 2024, 1:59:41 PM

DATA ENGINEER BATCH 4

Mentee: Yovina Silvia

Mentor: Bilal Benefit

2. Hapus file yang sudah diupload

a. Centang pada sisi kiri file yang ingin dihapus, lalu pilih delete

Google Cloud Alta-DE-4-Yovina bu Search

Bucket details

bucket-alta-de-4-yovinasilvia

Location: asia-southeast1 (Singapore) Storage class: Standard Public access: Not public Protection: Soft Delete

OBJECTS CONFIGURATION PERMISSIONS PROTECTION LIFECYCLE OBSERVABILITY INVENTORY REPORTS OPERATIONS

Folder browser

bucket-alta-de-4-yovinasilvia

UPLOAD FILES UPLOAD FOLDER CREATE FOLDER TRANSFER DATA MANAGE HOLDS EDIT RETENTION DOWNLOAD DELETE

Filter by name prefix only Filter objects and folders Show Live objects only

	Name	Size	Type	Created	Storage class	Last modified	Public access	Version history	Encryption	Object retention
<input checked="" type="checkbox"/>	order_details_202407221357.csv	12.6 KB	text/csv	Jul 22, 2024, 1:59:41 PM	Standard	Jul 22, 2024, 1:59:41 PM	Not public	—	Google-managed	—
<input type="checkbox"/>	orders_202407221357.csv	42.8 KB	text/csv	Jul 22, 2024, 1:59:41 PM	Standard	Jul 22, 2024, 1:59:41 PM	Not public	—	Google-managed	—
<input type="checkbox"/>	products_202407221358.csv	102 B	text/csv	Jul 22, 2024, 1:59:41 PM	Standard	Jul 22, 2024, 1:59:41 PM	Not public	—	Google-managed	—
<input type="checkbox"/>	users_202407221357.csv	98 B	text/csv	Jul 22, 2024, 1:59:41 PM	Standard	Jul 22, 2024, 1:59:41 PM	Not public	—	Google-managed	—

DATA ENGINEER BATCH 4

Mentee: Yovina Silvia

Mentor: Bilal Benefit

c. Pilih Delete pada menu pop up yang muncul

The screenshot shows the Google Cloud Storage console interface. At the top, the Google Cloud logo and user profile 'Alta-DE-4-Yovina' are visible. The search bar contains 'bu'. The main heading is 'Bucket details' for 'bucket-alta-de-4-yovinasilvia'. Below this, a table shows bucket metadata: Location (asia-southeast1 (Singapore)), Storage class (Standard), Public access (Not public), and Protection (Soft Delete). A navigation bar includes tabs for OBJECTS, CONFIGURATION, PERMISSIONS, PROTECTION, LIFECYCLE, OBSERVABILITY, INVENTORY REPORTS, and OPERATIONS. The 'OBJECTS' tab is active, showing a 'Folder browser' on the left with a list of objects: 'bucket-alta-de-4-yovinasilvia', 'order_details_202407221357', 'orders_202407221357.csv', 'products_202407221358.csv', and 'users_202407221357.csv'. The first object is selected. A 'Delete 1 object?' dialog box is centered on the screen, displaying a warning: 'Because a soft-delete policy is enabled for this bucket, deleted objects will be kept in a soft-deleted state for 7 days before being permanently deleted. While objects are in a soft-deleted state, they can be restored.' The dialog has 'CANCEL' and 'DELETE' buttons. In the background, a table of object details is partially visible, showing columns for Name, Public access, Version history, Encryption, and Object retention.

Name	Public access	Version history	Encryption	Object retention
order_details_202407221357	Not public	—	Google-managed	—
orders_202407221357.csv	Not public	—	Google-managed	—
products_202407221358.csv	Not public	—	Google-managed	—
users_202407221357.csv	Not public	—	Google-managed	—

DATA ENGINEER BATCH 4

Mentee: Yovina Silvia

Mentor: Bilal Benefit

d. Kita sudah berhasil menghapus file dalam bucket

Google Cloud Alta-DE-4-Yovina bu X Search

Bucket details

bucket-alta-de-4-yovinasilvia

Location: asia-southeast1 (Singapore) Storage class: Standard Public access: Not public Protection: Soft Delete

OBJECTS CONFIGURATION PERMISSIONS PROTECTION LIFECYCLE OBSERVABILITY INVENTORY REPORTS OPERATIONS

Folder browser

bucket-alta-de-4-yovinasilvia

Buckets > bucket-alta-de-4-yovinasilvia

UPLOAD FILES UPLOAD FOLDER CREATE FOLDER TRANSFER DATA MANAGE HOLDS EDIT RETENTION DOWNLOAD DELETE

Filter by name prefix only Filter Filter objects and folders

<input type="checkbox"/>	Name	Size	Type	Created	Storage class	Last modified	Public a
<input type="checkbox"/>	orders_202407221357.csv	42.8 KB	text/csv	Jul 22, 2024, 1:59:41 PM	Standard	Jul 22, 2024, 1:59:41 PM	Not pub
<input type="checkbox"/>	products_202407221358.csv	102 B	text/csv	Jul 22, 2024, 1:59:41 PM	Standard	Jul 22, 2024, 1:59:41 PM	Not pub
<input type="checkbox"/>	users_202407221357.csv	98 B	text/csv	Jul 22, 2024, 1:59:41 PM	Standard	Jul 22, 2024, 1:59:41 PM	Not pub

bucket-alta-de-4-yovinasilvia/order_details_202407221357.csv deleted X

DATA ENGINEER BATCH 4

Mentee: Yovina Silvia

Mentor: Bilal Benefit



3. Lakukan eksplorasi sample data wikipedia (<https://console.cloud.google.com/bigquery?kipedia&page=table>) dengan menggunakan big query!

Menampilkan dari table wikipedia untuk data dengan title 'Time Lord' berdasarkan timestamp paling besar/baru dengan limit data yang ditampilkan sebanyak 100 baris.

querynya:

```
select * from `bigquery-public-data.samples.wikipedia`  
where title = 'Time Lord'  
order by timestamp desc  
limit 100;
```

The screenshot shows the Google Cloud BigQuery console interface. At the top, there's a header with Google Cloud logo, user profile 'Alta-DE-4-Yovina', and a search bar. Below the header, the 'Explorer' panel on the left shows a tree view of resources, including 'samples' and 'Models (1)'. The main panel displays a query titled 'Untitled query' with the following SQL code:

```
1 select * from `bigquery-public-data.samples.wikipedia`  
2 where title = 'Time Lord'  
3 order by timestamp desc  
4 limit 100;
```

Below the query editor, the 'Query results' section shows a table with 11 rows of data. The table has columns: Row, title, id, language, wp_namespace, is_redirect, revision_id, contributor_ip, contributor_id, contributor_username, timestamp, is_minor, is_bot, reversion_id, and comment. The 'title' column for all rows is 'Time Lord'. The 'timestamp' column shows descending values from 1264876232 to 1264503467. The 'comment' column contains various notes related to the article's history.

Row	title	id	language	wp_namespace	is_redirect	revision_id	contributor_ip	contributor_id	contributor_username	timestamp	is_minor	is_bot	reversion_id	comment
1	Time Lord	75912		0	null	340939490	98.218.46.72	null	null	1264876232	null	null	null	/* Technology */
2	Time Lord	75912		0	null	340819666	null	1806135	J Greb	1264812426	null	null	null	Avoid redirect...
3	Time Lord	75912		0	null	340814031	null	3208432	SeanNovack	1264810389	null	null	null	Undid last revision: "Eart Like" generally refers to atmosphere and biosphie Star Trek would call it an "M-class" planet. Please
4	Time Lord	75912		0	null	340765462	null	5581759	Jammy0002	1264792117	null	null	null	/* Culture and society */ is not red and orange, or huge either. (Source: The of Time Part Two)
5	Time Lord	75912		0	null	340722637	193.203.84.34	null	null	1264776264	null	null	null	/* Regeneration */
6	Time Lord	75912		0	null	340722169	193.203.84.34	null	null	1264776079	null	null	null	/* Regeneration */
7	Time Lord	75912		0	null	340395063	66.16.135.172	null	null	1264623931	null	null	null	/* Regeneration */
8	Time Lord	75912		0	null	340394930	66.16.135.172	null	null	1264623888	null	null	null	/* Regeneration */
9	Time Lord	75912		0	null	340394608	66.16.135.172	null	null	1264623775	null	null	null	/* Regeneration */
10	Time Lord	75912		0	null	340112173	null	11168701	Letsy2	1264503533	true	null	null	...even if I am an Americ
11	Time Lord	75912		0	null	340112048	null	11168701	Letsy2	1264503467	true	null	null	I always thought that Do

At the bottom of the results table, it says 'Results per page: 50' and '1 - 50 of 100'. Below the results, there's a 'Job history' section with a 'REFRESH' button.

DATA ENGINEER BATCH 4

Mentee: Yovina Silvia

Mentor: Bilal Benefit



4. **Munculkan jumlah kontribusi dari masing-masing contributor_ip. urutkan dari kontribusi terbesar ke kontribusi terkecil**
Querynya:

```
select
  contributor_ip,
  count(*) as contribution_count
from
  `bigquery-public-data.samples.wikipedia`
group by
  contributor_ip
order by
  contribution_count DESC
limit 100;
```

DATA ENGINEER BATCH 4

Mentee: Yovina Silva

Mentor: Bilal Benefit



Free trial status: Rp4,923,900.00 credit and 91 days remaining. Activate your full account to get unlimited access to all of Google Cloud—use any remaining credits, then pay only for what you use.

DISMISS **ACTIVATE**

Google Cloud Alta-DE-4-Yovina Search (/) for resources, docs, products, and more Search

Explorer + ADD

Viewing resources. SHOW STARRED ONLY

- fdi_drug
- fdi_food
- fdi_banks
- fec
- samples
 - Models (1)
 - github_nested
 - github_timeline
 - gsod
 - nativity
 - shakespeare
 - trigrams
 - wikipedia

SHOW MORE

SUMMARY

Nothing currently selected

Untitled query RUN SAVE DOWNLOAD SHARE SCHEDULE MORE

```
1 select
2   contributor_ip,
3   count(*) as contribution_count
4 from
5   `bigquery-public-data.samples.wikipedia`
6 group by
7   contributor_ip
8 order by
9   contribution_count DESC
10 limit 100;
```

Query completed.

Press Alt+F1 for Accessibility Options.

Query results SAVE RESULTS EXPLORE DATA

JOB INFORMATION RESULTS CHART JSON EXECUTION DETAILS EXECUTION GRAPH

Row	contributor_ip	contribution_count
1	null	237051825
2	Template namespace initialisat...	48803
3	Conversion script	31744
4	217.129.67.28	24431
5	204.153.84.10	23215
6	68.39.174.238	22614
7	202.156.6.54	19444
8	208.81.184.4	16427
9	131.107.0.80	14623
10	131.107.0.73	14431
11	156.110.24.142	12649
12	222.251.209.94	12259
13	69.138.229.246	11622
14	206.47.220.230	10351

Results per page: 50 1 - 50 of 100

Job history REFRESH