



GDG DevFest Taipei 2021



AGENDA

- Apocalypse 🤖
- Migrate to Where 🤔
- Save Yourself 🚚
- Save All 🚢🚢🚢
- Is Everyone Onboard 📝
- Reborn - SOP 📄📄📄
- Watch Out ⚠️



Apocalypse

What Happened

Google Trends

“

Google is announcing data storage policy changes for its education customers.

— 2021.02 Google Workspace

“

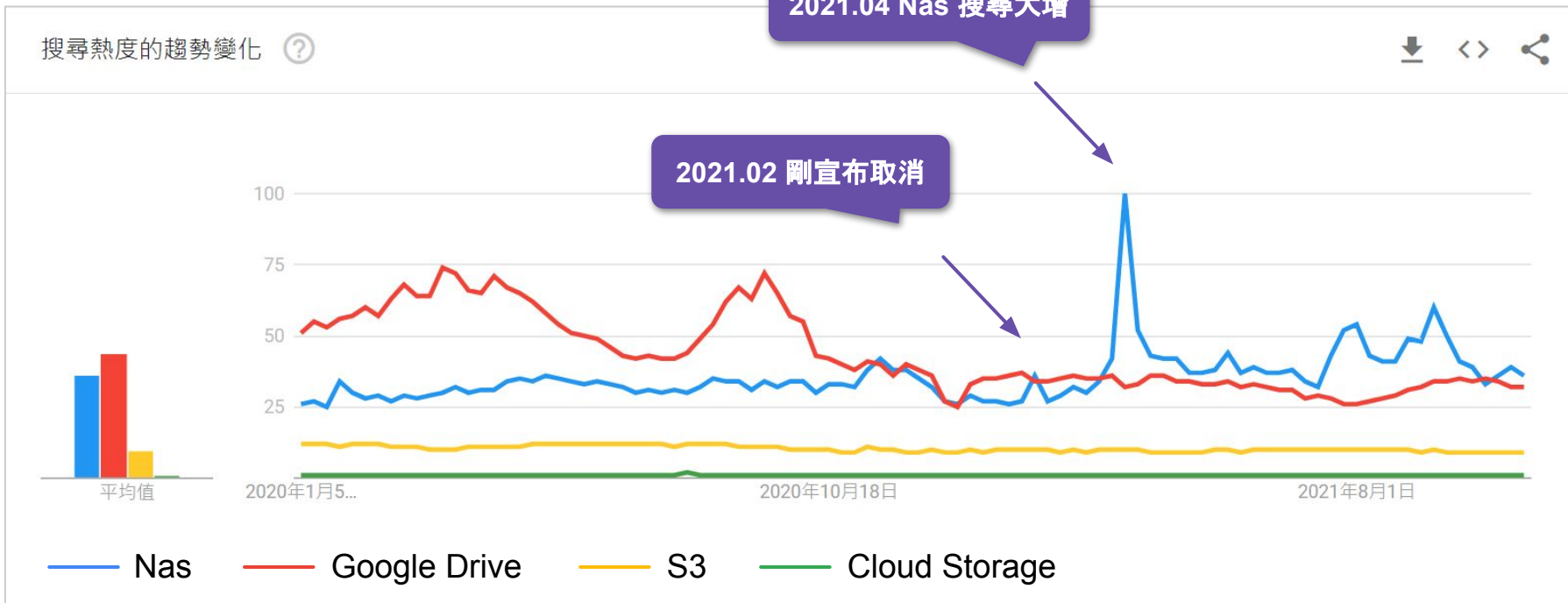
Google drive 教育版不再「無限儲存」，學校免費上限為 100TB



Apocalypse

What Happened

Google Trends



Migrate to Where

So Many Options

What do I need

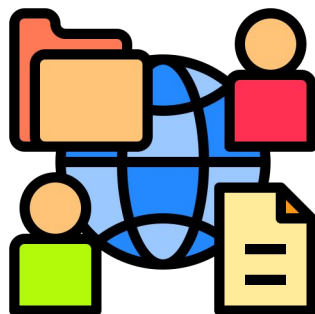


Migrate to Where

So Many Options

What do I need

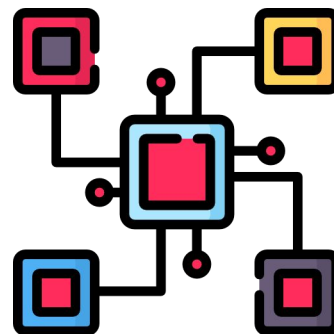
使用情境



費用&容量



其他工具整合



Migrate to Where

So Many Options

What do I need

Reference

- [12大「雲端儲存」價格容量大車拼, 看看誰「俗又大碗」!](#)
- [2021年雲端硬碟八大比較, 你喜歡哪一種?](#)
- [【雲端硬碟推薦】2021雲端空間比較、費用及評價, 看這篇就 夠了](#)
- [2021 最新5 家雲端硬碟比較推薦: Google Drive、iCloud、Dropbox、OneDrive、MEGA](#)
- [全球這5大雲端硬碟Cloud Drive到底誰最好用? Google、Amazon、Dropbox...](#)

Save Yourself

Solution

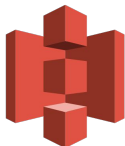
Architecture

Prerequisites

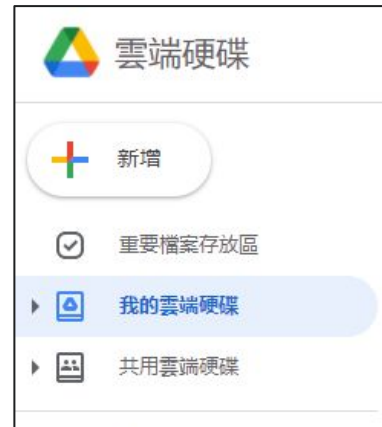
我就是那種濫善用無限空間的夫壞蛋好寶寶

- 只是備分, 321原則的一部分
- 沒有協作的場景
- 平常也不會存取

那我幹嘛不用 Cloud Storage



研究所兩年
+ 工作一年半的結
果



Save Yourself

Solution

Architecture

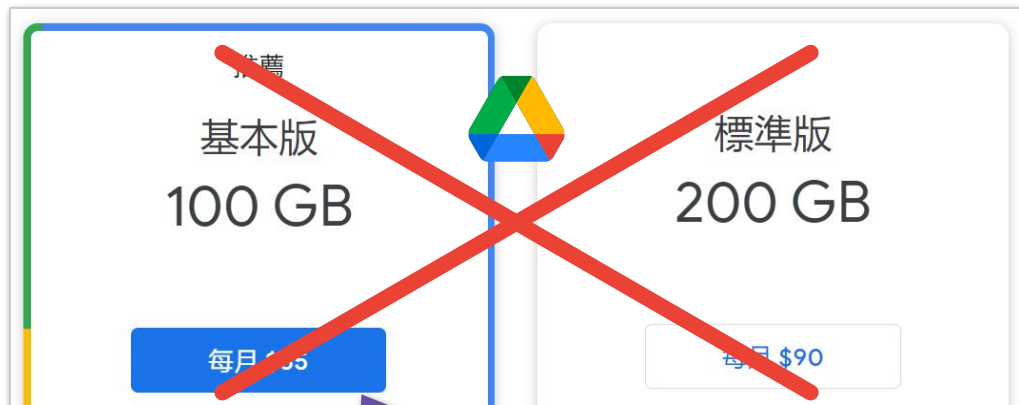
Prerequisites



OS: 怎麼不選 AWS S3 或是 Azure Blob Storage

因為敝司是 ~~GCP 代理商~~。因為今天是 ~~GDC~~。
如果你也是 Cloud Cloud 的開發者，會發現許多 Best Practice 都有 Cloud Storage 的身影

- 用多少付多少
- 權限管理更有彈性
- 靜態網頁部署
- 多種儲存空間級別
- 物件版本、生命週期
- 儲存在台灣
- 耐用性(durability)、可用性(availability)



Google One 最小
就要買到 100GB

Save Yourself

Solution

Architecture

Prerequisites

North America	South America	Europe	Asia	Australia
Location	Standard Storage (per GB per Month)	Nearline Storage (per GB per Month)	Coldline Storage (per GB per Month)	Archive Storage (per GB per Month)
Taiwan (asia-east1)	\$0.020	\$0.010	\$0.005	\$0.0015



許多儲存空間級別可以選擇。

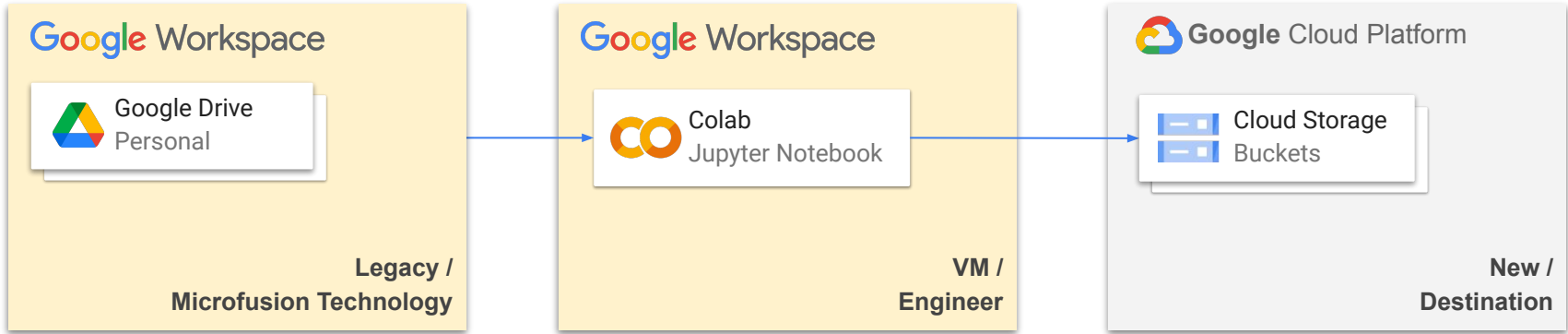
價格很便宜 (相對直接買 Google Drive 來做長時間的備分存放。)

Save Yourself

Solution

Architecture

Prerequisites



1

3

2

Save Yourself

Solution

Architecture

Prerequisites

Step 01. 把 Google Drive mount 到你的環境。



```
from google.colab import drive
mount = '/content/gdrive01' # you can change the path of mount location
print("Colab: mounting Google drive on ", mount)
drive.mount(mount)
```

Save Yourself

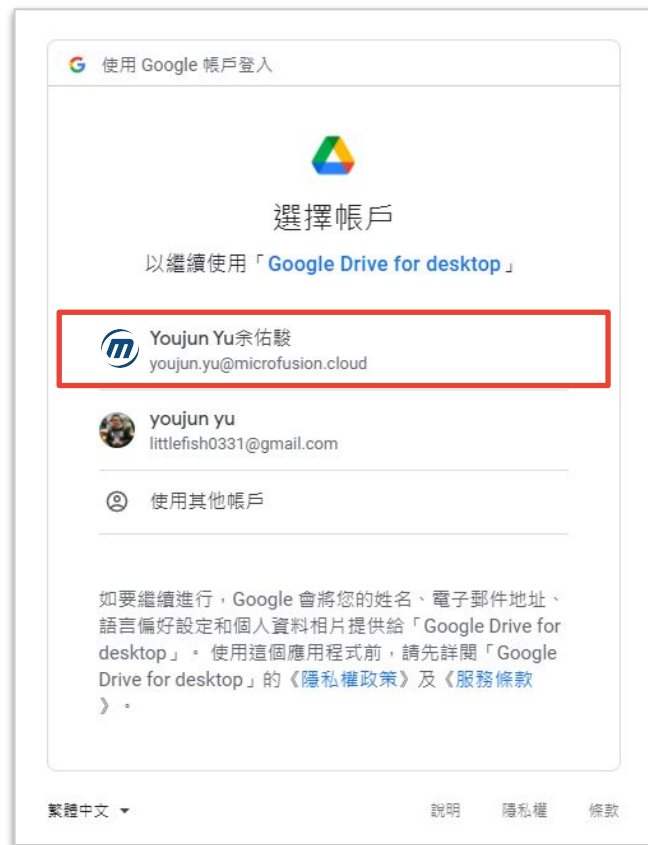
Solution

Architecture

Prerequisites

Step 01. 把 Google Drive mount 到你的環境。

選擇有權限讀取來源 Google Drive 的帳號



Save Yourself

Solution

Architecture

Prerequisites

Step 02. 建立 GCS 的 bucket

記得 bucket name 要全球唯一

Cloud Storage

← Create a bucket

! Name your bucket

Pick a **globally unique**, permanent name. [Naming guidelines](#)

demo4gdg

! That bucket name is taken. Try another.

▼ LABELS (OPTIONAL)

CONTINUE

• Choose where to store your data

This permanent choice defines the geographic placement of your data and affects cost, performance, and availability. [Learn more](#)

Location type

Multi-region
Highest availability across largest area

Dual-region
High availability and low latency across 2 regions

Region
Lowest latency within a single region

Location

asia-east1 (Taiwan)

Release Notes

Save Yourself

Solution

Architecture

Prerequisites

Step 03. 設定 IAM 有權限讀取。

- Storage Admin
- Viewer (可以不用)

The screenshot shows the Google Cloud IAM & Admin console. The left sidebar is titled 'IAM & Admin' and lists various services: IAM (selected), Identity & Organization, Policy Troubleshooter, Policy Analyzer, Organization Policies, Service Accounts, Workload Identity Federat..., and Labels. The main content area is titled 'IAM' and has 'ADD' and 'REMOVE' buttons. Below this, there are tabs for 'PERMISSIONS' (selected) and 'RECOMMENDATIONS HISTORY'. The main heading is 'Permissions for project "google-free-tier"'. A note states: 'These permissions affect this project and all of its resources. [Learn more](#)'. Below this, there are 'View By:' buttons for 'PRINCIPALS' (selected) and 'ROLES'. A search filter is set to 'littlefish0331'. A table lists the permissions for the principal 'littlefish0331@gmail.com':

Type	Principal	Name	Role
<input type="checkbox"/>	littlefish0331@gmail.com		Storage Admin
<input type="checkbox"/>			Viewer

Save Yourself

Solution

Architecture

Prerequisites

完成 Colab 環境與 GCS bucket 設定



```
from google.colab import auth
auth.authenticate_user()

project_id = 'savvy-generator-332202' # replace to your bucket name
!gcloud config set project {project_id}
!gcloud config list
!gsutil ls
```

Save Yourself

Solution

Architecture

Prerequisites

上傳檔案

```

# folder you want to migrate and bucket
copy_foler = "/content/gdrive01/MyDrive/Short-Talk/20211127-DevFest\\'21/demo4gdg-01"
bucket_name = 'demo4gdg'


# check the command
!echo gsutil -m cp -r {copy_foler} gs://{bucket_name}/


# execute
!gsutil -m cp -r {copy_foler}/ gs://{bucket_name}/
```

My Drive > Short-Talk > 20211127-DevFest'21 > demo4gdg-01 ▾

Name ↑

 2022_template.pptx


 demo excel.xlsx

 demo word.docx


 linktr-youjun.png


 Maroon 5 - Beautiful Mistakes ft. Megan Thee Stallion (Official Music Video)


 Maroon 5 - Beautiful Mistakes ft. Megan Thee Stallion (Official Music Video)

 youjun-cv.pdf

Cloud Storage

 [Browser](#)

 [Monitoring](#)

 [Settings](#)

← Bucket details

demo4gdg





Location	Storage class	Public access	Protection
asia-east1 (Taiwan)	Standard	 Subject to object ACLs	None

[OBJECTS](#) [CONFIGURATION](#) [PERMISSIONS](#) [PROTECTION](#) [LIFECYCLE](#)

Buckets > demo4gdg > demo4gdg-01 

[UPLOAD FILES](#) [UPLOAD FOLDER](#) [CREATE FOLDER](#) [MANAGE HOLDS](#) [DOWNLOAD](#) [DELETE](#)

Filter by name prefix only ▾  **Filter** Filter objects and folders

<input type="checkbox"/>	Name	Size	Type
<input type="checkbox"/>	 2022_template.pptx	112.1 KB	application/vnd.openxmlformats-officedocument.presentationml.presentation
<input type="checkbox"/>	 Maroon 5 - Beautiful Mistakes ft. M	3.5 MB	audio/mpeg
<input type="checkbox"/>	 Maroon 5 - Beautiful Mistakes ft. M	18.7 MB	video/mp4
<input type="checkbox"/>	 demo excel.xlsx	9.6 KB	application/vnd.openxmlformats-officedocument.spreadsheetml.sheet
<input type="checkbox"/>	 demo word.docx	12.2 KB	application/vnd.openxmlformats-officedocument.wordprocessingml.document
<input type="checkbox"/>	 linktr-youjun.png	113.7 KB	image/png
<input type="checkbox"/>	 youjun-cv.pdf	179.8 KB	application/pdf

Save All

Problem

Solution

欸!? 那是不是就是「MyDrive/*」催下去就好了

```
# folder you want to migrate and bucket
copy_foler = "/content/gdrive01/MyDrive/*"
bucket_name = 'demo4gdg'

# check the command
!echo gsutil -m cp -r {copy_foler} gs://{bucket_name}/

# execute
!gsutil -m cp -r {copy_foler}/ gs://{bucket_name}/
```

Save All

Problem

Solution

欸!? 那是不是就是「MyDrive/*」催下去就好了

D*mn! We got problem!

```
# folder you want to copy
copy_foler = "/content/gdrive01/MyDrive/Short-Talk/20211127-DevFest' 21/demo4gdg-02"
bucket_name = 'demo4gdg'

# check the command
!echo gsutil -m cp -r {copy_foler} gs://{bucket_name}/

# execute
!gsutil -m cp -r {copy_foler} gs://{bucket_name}/
```

```
demo4gdg-demo02 ☆
檔案 編輯 檢視畫面 插入 執行階段 工具 說明 上次儲存時間: 上午7:04
+ 程式碼 + 文字
Upload Google Drive File to Cloud Storage
[ ] 1 bucket_name = 'demo4gdg'
2 copy_foler = "/content/gdrive01/MyDrive/Short-Talk/20211127-DevFest' 21/demo4gdg-02"
3
4 # check the command
5 !echo gsutil -m cp -r {copy_foler} gs://{bucket_name}/
6
7 # execute
8 !gsutil -m cp -r {copy_foler} gs://{bucket_name}/

gsutil -m cp -r /content/gdrive01/MyDrive/Short-Talk/20211127-DevFest' 21/demo4gdg-02 gs://demo4gdg/
CommandException: Error opening file "file:///content/gdrive01/MyDrive/Short-Talk/20211127-DevFest' 21/demo4gdg-02/demo-doc/gdoc": [Error]
CommandException: Error opening file "file:///content/gdrive01/MyDrive/Short-Talk/20211127-DevFest' 21/demo4gdg-02/demo-site/site": [Error]
CommandException: Error opening file "file:///content/gdrive01/MyDrive/Short-Talk/20211127-DevFest' 21/demo4gdg-02/demo-form/form": [Error]
CommandException: Error opening file "file:///content/gdrive01/MyDrive/Short-Talk/20211127-DevFest' 21/demo4gdg-02/demo-sheet/sheet": [Error]
CommandException: Error opening file "file:///content/gdrive01/MyDrive/Short-Talk/20211127-DevFest' 21/demo4gdg-02/demo-slide/slides": [Error]
Copying file:///content/gdrive01/MyDrive/Short-Talk/20211127-DevFest' 21/demo4gdg-02/demo-txt.txt [Content-Type=text/plain]...
CommandException: 5 files/objects could not be transferred.
```

Save All

Problem

Solution

Well, Google Workspace Document, 也就是 Google Drive 裡面那些原生的檔案。

- 特殊的 file extension
- 只是連結
- 即使 mount, 也無法直接讀取*
- 只能下載, 再上傳!

下載還要轉檔



* Google Sheet 可以透過套件, 但是一般的 read.csv 會失敗。

Save All

Problem

Solution

[Google Workspace and Drive MIME Types | Google Drive API](#)
[Google Workspace documents and corresponding export MIME types](#)

```
from google.colab import auth
auth.authenticate_user()
from googleapiclient.discovery import build
drive_service = build('drive', 'v3')

# !pip install requires.io
import io
from googleapiclient.http import MediaIoBaseDownload

# Google Slide 的 URL
# https://docs.google.com/presentation/d/1aeyyl0kK22KtlxkUljCFRAog1h_kd-H0hZBKDPyg43w/edit?usp=sharing
file_id = '1aeyyl0kK22KtlxkUljCFRAog1h_kd-H0hZBKDPyg43w'
request = drive_service.files().export_media(fileId = file_id, mimeType='application/pdf')

fh = io.FileIO("/content/silde-can-download.pdf", 'wb')
downloader = MediaIoBaseDownload(fh, request)
done = False
while done is False:
    status, done = downloader.next_chunk()
    print("Download %d%%." % int(status.progress() * 100))
```

Is Everyone Onboard

We Need a Checklist !!

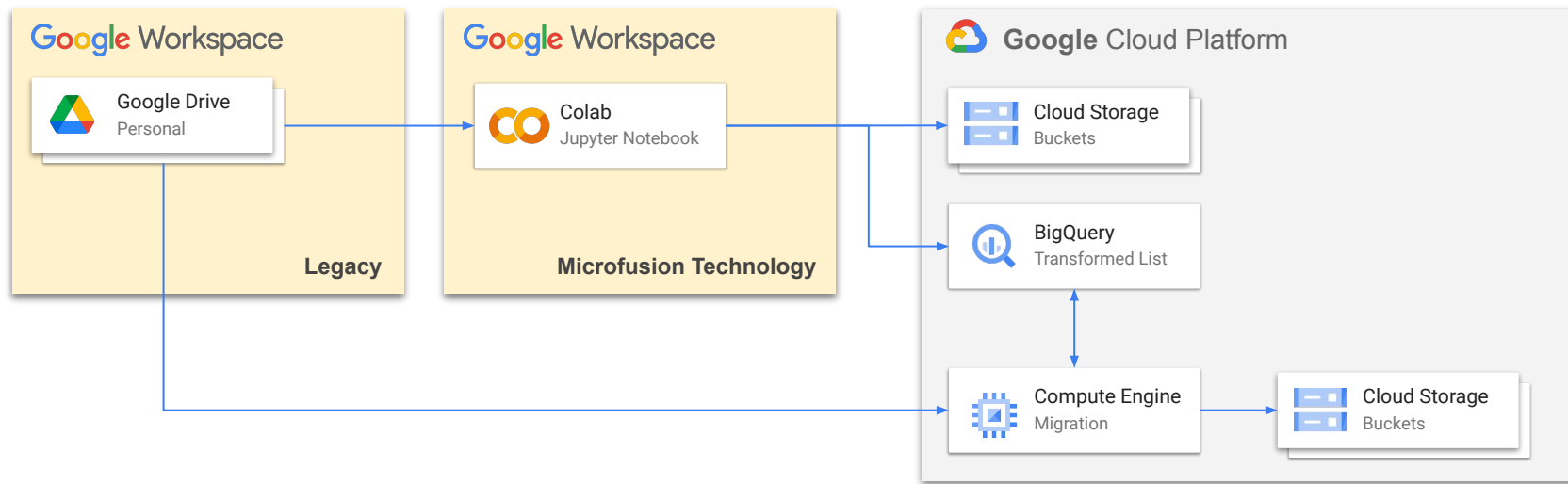
- 哪些檔案
- 檔案類型
- 檔案大小
- 檔案最後變動時間
- ...

Solution

- ls -lR
- du -a
- [Google Drive API](#)



Is Everyone Onboard



BigQuery 免費使用額度綽綽有餘，只是拿來存取一個清單

Computer Engine 可以開爆 → 平行 or 多台處理(不同 folder)。

Watch Out !

Careful

Conclusion

Colab 本身使用的限制

- 12小時的連續使用時間
- 資源規格

小型資料(5TB 以內)

- 用 Colab 即可
- 注意檔案大小

大型資料(超過5TB) or 公司的Google Drive

- 建議還是用 GCE, 開好一點的 VM
- 一定要做清單確認
- 可以先小規模搬遷
- 搬遷後, 讓 Google Drive 的資料先保留三個月

搬遷有風險, 執行前請詳閱操作說明書

About Me



余佑駿 Youjun
Cloud Architect



【經歷】

- 2021 宏庭股份有限公司
GCP解決方案部 雲端架構師
- 2019 國家高速網路與計算中心 資料科學家
- 2018 國泰世華銀行客戶經營部資料分析科 實習
- 2017 國立政治大學統計所
- 2012 國立清華大學數學系



履歷、專案
社群分享



我目前公司
上雲需求歡迎

GDG DevFest
Taipei 2021



Facebook / littlefish0331



Linkedin / you-jun-yu



Github / littlefish0331



littlefish0331@gmail.com





Thank You

Q & A time

* 特別感謝 宏庭科技GWS專家 吳孫銘Sam 的協助

