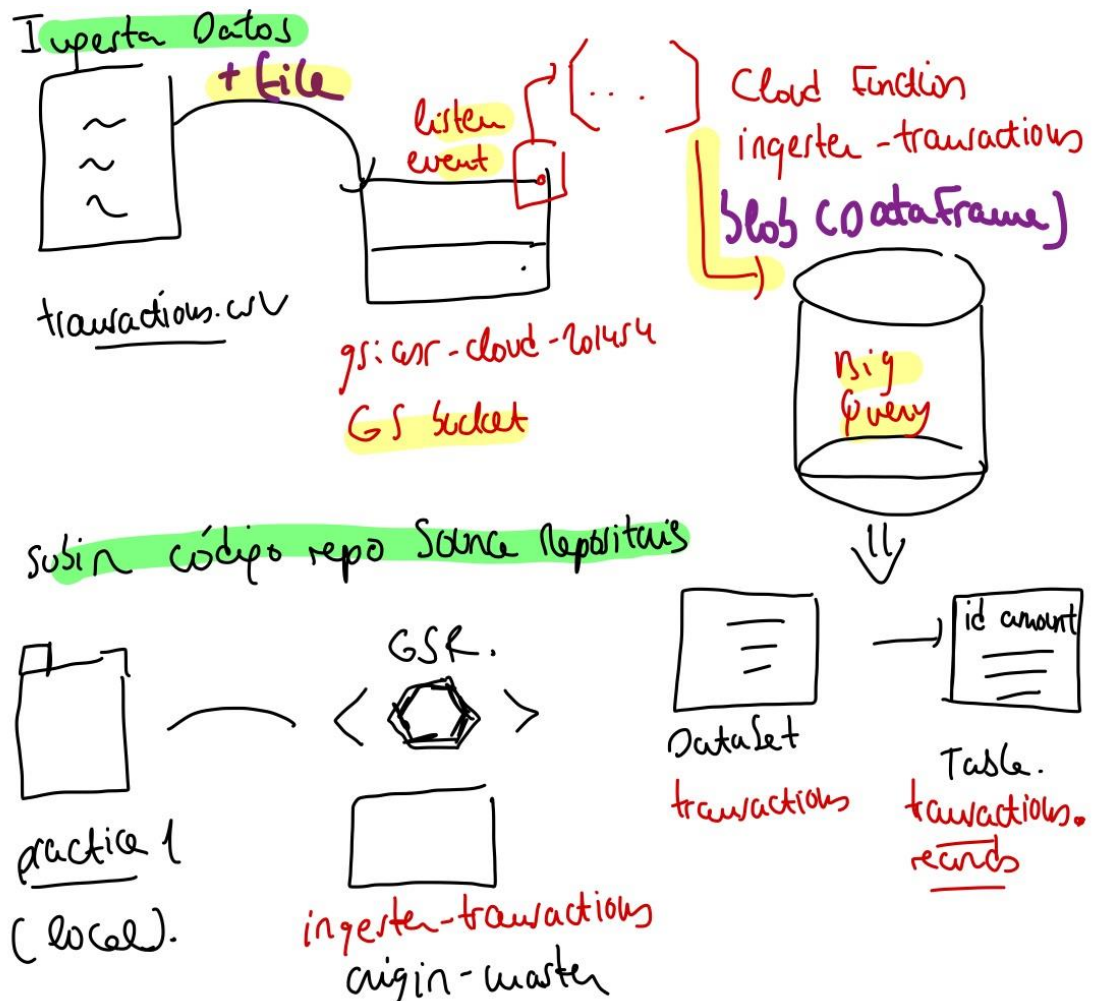




## PRACTICA 1 - DESPLIEGUE DE CLOUD FUNCTION DE INGESTA DE DATOS

ARQUITECTURA DE SERVICIOS DE RED - CLOUD

Teresa González García 1º MIT + ADE



- Crear CLOUD STORAGE BUCKET
- Crear Dataset BQ + tabla.
- Deployment.sh
  - Código ejecutable main.py funcion run endpoint Cloud Functions -> trigger que se levanta cada vez que le llegue un archivo. FILE = EVENT. Dentro del endpoint run (event, context)
  - Trigger – bucket. File evento = blob. FUNCION CARGAR BLOB EN BQ
  - Dentro de endpoint:
    - o Download blob archivo local
    - o Preparar blob con dataframe (filas, columnas) -> requirements.txt. STRING Y FLOAT
    - o Subir a BQ con formato apropiado.
- Crear CLOUD SOURCE REPOSITORY y subir todo el proyecto.
- Estimar costes proyecto 1700 transacciones/s -> procesado, almacenamiento Storage y Big Query.

## 1. Crear bucket

```
C:\Users\teete\AppData\Local\Google\Cloud SDK> gsutil mb -c standard gs://asr-cloud-201701454
Creating gs://asr-cloud-201701454/...
```

```
C:\Users\teete\AppData\Local\Google\Cloud SDK>gsutil ls
gs://artifacts.nifty-envoy-328012.appspot.com/
gs://asr-cloud-201701454/
```

Blob path gs://asr-cloud-201701454

Por defecto location US multiregion (la + barata, tienen muchos DC)

Inicialmente vacio

```
C:\Users\teete\AppData\Local\Google\Cloud SDK>gsutil ls -r gs://asr-cloud-201701454
```

**asr-cloud-201701454**

Ubicación	Clase de almacenamiento	Acceso público	Protección
us (varias regiones en Estados Unidos)	Standard	⚠ Sujeto a las LCA de objetos	Ninguna

**OBJETOS** CONFIGURACIÓN PERMISOS PROTECCIÓN CICLO DE VIDA

Depósitos > asr-cloud-201701454

[SUBIR ARCHIVOS](#) [SUBIR CARPETA](#) [CREAR CARPETA](#) [ADMINISTRAR CONSERVACIONES](#) [DESCARGAR](#) [BORRAR](#)

Filtrar solo por prefijo de nombre **Filtro** Filtrar objetos y carpetas

<input type="checkbox"/>	Nombre	Tamaño	Tipo	Fecha de creación	Clase de almacenamiento	Última modificación	Acceso público
No hay filas para mostrar							

## 2. Crear dataset transactions y tabla records BQ

Instalamos comandos Alpha para acceder a bq (una opción)

```
C:\Users\teete\AppData\Local\Google\Cloud SDK>gcloud alpha bq datasets create transactions --description 'Transactions'
Created dataset [transactions].
```

Por defecto location US multiregion (la + barata, tienen muchos DC).

Clave administrada por Google

nifty-envoy-328012

transactions

nifty-envoy-328012:transactions

"Transactions"

Información del conjunto de datos

ID de conjunto de datos	nifty-envoy-328012:transactions
Fecha de creación	6 nov. 2021 15:40:12
Vencimiento predeterminado de la tabla	Nunca
Última modificación	6 nov. 2021 15:40:12
Ubicación de los datos	US

Inicialmente vacio.

### 3. Ejecutar Deployment.sh

- Archivo bash ejecutable unix
- Crea bucket si no existe
- Crea BQ y tabla si no existe.
- Deploy CF definida
  - o Funcion entripoint (ingest\_transactions dentro del main.py)
  - o Runtime (python 3.8)
  - o Trigger-resource = BUCKET
  - o Trigger-event = SE GUARDA NUEVO ARCHIVO/OBJETO

```
#!/usr/bin/env bash

## !!! IMPORTANT:
## Uncomment the following line for the first running. After you've authenticated
## once, you won't need to run this line anymore:
# gcloud auth application-default login

# GS: Check if blob name exists, else it's created
gsutil ls -b gs://asr-cloud-201701454 || gsutil mb -l US gs://asr-cloud-201701454

# Copy the example transaction in the desired bucket:
gsutil cp transaction.csv gs://asr-cloud-201701454

# BQ: Check if dataset transactions exists, else it's created.
bq.cmd ls transactions || bq.cmd mk -d transactions

# BQ: Check if table records exists, else it's created
#ID table= dataset.table
bq.cmd ls transactions | grep records || bq.cmd mk --table transactions.records ID:STRING,AMOUNT:FLOAT

# Deployment CLOUD FUNCTION
: '
Definimos la cloud function que se EJECUTA=DEPLOY
nombre ingester-transactions
entry-point=funcion que se ejecuta en el codigo= ingest_transactions.
lenguaje py3.8
trigger-resource= CLOUD STORAGE BUCKET. SE ENCUENTRA ESCUCHANDO LOS EVENTOS DEL BUCKET.
trigger-event= finaliza la carga del objeto (blob). Cuando se guarde el archivo, se ejecutara el entripoint
memory que ocupa
timeout =duracion de la cloud function (como mucho). pasado ese tiempo finaliza.
'

gcloud functions deploy ingester-transactions \
  --entry-point=ingest_transaction \
  --runtime python38 \
  --trigger-resource asr-cloud-201701454 \
  --trigger-event google.storage.object.finalize \
  --memory 512MB \
  --timeout 60s
```

Importante bq.cmd en Windows, si no bash no reconoce el comando.

Se ejecuta

```
PS C:\Users\teete\PracticaCloud\Practica1> .\deployment.sh
```

## 4. RESULTADOS

### - Bucket creado

**asr-cloud-201701454**

Ubicación: us (varias regiones en Estados Unidos) | Clase de almacenamiento: Standard | Acceso público: ⚠ Sujeto a las LCA de objetos | Protección: Ninguna

**OBJETOS** | CONFIGURACIÓN | PERMISOS | PROTECCIÓN | CICLO DE VIDA

Depósitos > asr-cloud-201701454

[SUBIR ARCHIVOS](#) | [SUBIR CARPETA](#) | [CREAR CARPETA](#) | [ADMINISTRAR CONSERVACIONES](#) | [DESCARGAR](#) | [BORRAR](#)

Filtrar solo por prefijo de nombre ▾ | **Filtro** Filtrar objetos y carpetas | ☐ Mostrar datos borrados

<input type="checkbox"/>	Nombre	Tamaño	Tipo	Fecha de creación ?	Clase de almacenamiento	Última modificación	Acceso púb
<input type="checkbox"/>	transaction.csv	40 B	application/vnd.ms-excel	7 nov. 2021 14:09:17	Standard	7 nov. 2021 14:0...	No público

**OBJETO PUBLICADO** | HISTORIAL DE VERSIONES

[DESCARGAR](#) | [EDITAR METADATOS](#) | [MODIFICAR PERMISOS](#) | [BORRAR](#)

**Descripción general**

Tipo	application/vnd.ms-excel
Tamaño	40 B
Fecha y hora de creación	7 nov. 2021 14:09:17
Última modificación	7 nov. 2021 14:09:17
Clase de almacenamiento	Standard
Tiempo personalizado	—
URL pública ?	No aplicable
URL autenticada ?	<a href="https://storage.cloud.google.com/asr-cloud-201701454/transaction.csv?authuser=1">https://storage.cloud.google.com/asr-cloud-201701454/transaction.csv?authuser=1</a>
URI de gsutil ?	gs://asr-cloud-201701454/transaction.csv

**Permisos**

Acceso público	No público
----------------	------------

**Protección**

Estado de conservación	Ninguno
Historial de versiones ?	—
Política de retención	Ninguno
Tipo de encriptación	Google-managed key

### - Cloud Function creada y ejecutada.

Cloud Functions										NOTAS DE VERSIÓN
Funciones										
Filtro Filtrar funciones										
<input type="checkbox"/>	<input type="radio"/>	Nombre ↑	Región	Activador	Tiempo de ejecución	Memoria asignada	Función ejecutada	Última implementación	Authentication ?	Acciones
<input type="checkbox"/>	<input checked="" type="radio"/>	ingester-transactions	us-central1	Bucket: asr-cloud-201701454	Python 3.8	512 MB	ingest_transaction	7 nov. 2021 14:11:12		⋮

## Ejemplo logs Cloud Function

```
2021-11-07T17:55:00.942768631Z ingester-transactions br2hh9hi8yn0 Function execution started
2021-11-07T17:55:02.852Z ingester-transactions br2hh9hi8yn0 OpenBLAS WARNING - could not determine the L2 cache size on this system, assuming 256k
2021-11-07T17:55:05.326Z ingester-transactions br2hh9hi8yn0 Table ID BQ transactions.records
2021-11-07T17:55:05.813Z ingester-transactions br2hh9hi8yn0 [1] Bucket name: asr-cloud-201701454
2021-11-07T17:55:05.813Z ingester-transactions br2hh9hi8yn0 [1] Filename storage path: transaction.csv
2021-11-07T17:55:06.144Z ingester-transactions br2hh9hi8yn0 Downloaded storage object transaction.csv from bucket asr-cloud-201701454 to local file /tmp/transaction.csv.
2021-11-07T17:55:06.225Z ingester-transactions br2hh9hi8yn0 Data to be ingested:
2021-11-07T17:55:06.225Z ingester-transactions br2hh9hi8yn0 ID AMOUNT
2021-11-07T17:55:06.225Z ingester-transactions br2hh9hi8yn0 0 TRIGGERCLOUDBUILD 100000.55
2021-11-07T17:55:07.161770339Z ingester-transactions br2hh9hi8yn0 Function execution took 6222 ms, finished with status: 'ok'
```

## Info función

```
C:\Users\teete\AppData\Local\Google\Cloud SDK>gcloud functions list
NAME                STATUS  TRIGGER  REGION
ingester-transactions  ACTIVE  Event Trigger  us-central1

C:\Users\teete\AppData\Local\Google\Cloud SDK>gcloud functions describe ingester-transactions
availableMemoryMb: 512
buildId: b1fc2d20-6555-434d-ae96-e3fec6dd0423
buildName: projects/319729001445/locations/us-central1/builds/b1fc2d20-6555-434d-ae96-e3fec6dd0423
entryPoint: ingest_transaction
eventTrigger:
  eventType: google.storage.object.finalize
  failurePolicy: {}
  resource: projects/_/buckets/asr-cloud-201701454
  service: storage.googleapis.com
ingressSettings: ALLOW_ALL
labels:
  deployment-tool: cli-gcloud
name: projects/nifty-envoy-328012/locations/us-central1/functions/ingester-transactions
runtime: python38
serviceAccountEmail: nifty-envoy-328012@appspot.gserviceaccount.com
sourceUploadUrl: https://storage.googleapis.com/gcf-upload-us-central1-5952df30-c112-45c4-98e5-4d32efcb3aad/65f70770-11cc-4afe-ad9a-36b6de8d0486.zip
status: ACTIVE
timeout: 60s
updateTime: '2021-11-07T13:11:12.861Z'
versionId: '3'
```

- Tabla transactions.records con datos (dataframe)

▼ nifty-envoy-328012		records	
▼ transactions			
records			
		Esquema	Detalles
		<a href="#">Vista previa</a>	
Fila	ID	AMOUNT	
1	201701454teresa12	100000.55	

## 5. OTRAS PRUEBAS

Subimos otro archivo con otro nombre y + transacciones

```
C:\Users\teete\AppData\Local\Google\Cloud SDK> gsutil cp C:\Users\teete\PracticaCloud\Practical\transaction1.csv gs://asr-cloud-201701454
Copying file:///C:\Users\teete\PracticaCloud\Practical\transaction1.csv [Content-Type=application/vnd.ms-excel]...
/ [1 files][ 72.0 B/ 72.0 B]
Operation completed over 1 objects/72.0 B.
```

```
C:\Users\teete\AppData\Local\Google\Cloud SDK>gsutil ls gs://asr-cloud-201701454
gs://asr-cloud-201701454/transaction.csv
gs://asr-cloud-201701454/transaction1.csv
```

Se ha insertado en la tabla records los nuevos registros formateados (fila 2 y 3).

records

Esquema Detalles [Vista previa](#)

Fila	ID	AMOUNT
1	201701454teresa12	100000.55
2	201701454teresa12	100000.55
3	1234567teresaprueba	1000.6789

Se añaden más transacciones

```
C:\Users\teete\AppData\Local\Google\Cloud SDK>bq show transactions
Dataset nifty-envoy-328012:transactions

  Last modified      ACLs              Labels  Type
-----
06 Nov 15:40:12  Owners:
                  201701454@alu.comillas.edu,
                  projectOwners
Writers:
                  projectWriters
Readers:
                  projectReaders
                  DEFAULT

C:\Users\teete\AppData\Local\Google\Cloud SDK>bq show transactions.records
Table nifty-envoy-328012:transactions.records

  Last modified      Schema              Total Rows  Total Bytes  Expiration  Time Partitioning  Clustered Fields  Labels
-----
07 Nov 18:55:08  |- ID: string       7           191
                  |- AMOUNT: float
```

Ejemplo QUERY

```
C:\Users\teete\AppData\Local\Google\Cloud SDK> bq query select * from transactions.records
Waiting on bqjob_rcceb24a8e7d6cac_0000017d09512b08_1 ... (0s) Current status: DONE
```

```
+-----+-----+
|      ID      | AMOUNT |
+-----+-----+
| 1234567teresaprueba | 1000.6789 |
| 201701454teresa12   | 100000.55 |
| 201701454teresa12   | 100000.55 |
| 201701454teresa12   | 100000.55 |
| 201701454teresa12   | 100000.55 |
| 201701454teresa12   | 100000.55 |
| TRIGGERCLOUDBUILD   | 100000.55 |
+-----+-----+
```

Ordenar por Fecha

Filtrar consultas

loy

11:03  SELECT \* from transactions.records where ID="201701454teresa12"

11:02  SELECT \* from transactions.records

## 6. Subir código a repositorio

Creamos repo ingester-transactions

```
C:\Users\teete\AppData\Local\Google\Cloud SDK> gcloud source repos create ingester-transactions
Created [ingester-transactions].
WARNING: You may be billed for this repository. See https://cloud.google.com/source-repositories/docs/pricing for details.
```

Lista repos

```
C:\Users\teete\AppData\Local\Google\Cloud SDK>gcloud source repos list
REPO_NAME      PROJECT_ID      URL
ingester-transactions  nifty-envoy-328012  https://source.developers.google.com/p/nifty-envoy-328012/r/ingester-transactions
prototype      nifty-envoy-328012  https://source.developers.google.com/p/nifty-envoy-328012/r/prototype
```

Lo clonamos y copiamos los archivos de práctica 1

```
C:\Users\teete\PracticaCloud> gcloud source repos clone ingester-transactions
Cloning into 'C:\Users\teete\PracticaCloud\ingester-transactions'...
warning: You appear to have cloned an empty repository.
Project [nifty-envoy-328012] repository [ingester-transactions] was cloned to [C:\Users\teete\PracticaCloud\ingester-transactions].
```

```
C:\Users\teete\PracticaCloud\ingester-transactions>xcopy /E C:\Users\teete\PracticaCloud\Practica1 .
C:\Users\teete\PracticaCloud\Practica1\deployment.sh
C:\Users\teete\PracticaCloud\Practica1\main.py
C:\Users\teete\PracticaCloud\Practica1\Practica 1.docx
C:\Users\teete\PracticaCloud\Practica1\README.md
C:\Users\teete\PracticaCloud\Practica1\requirements.txt
C:\Users\teete\PracticaCloud\Practica1\transaction1.csv
6 archivo(s) copiado(s)
```

```
C:\Users\teete\PracticaCloud\ingester-transactions>dir
El volumen de la unidad C es Windows
El número de serie del volumen es: 160F-04D2

Directorio de C:\Users\teete\PracticaCloud\ingester-transactions

07/11/2021  15:07    <DIR>          .
07/11/2021  15:07    <DIR>          ..
07/11/2021  14:09                1.557 deployment.sh
07/11/2021  14:01                3.842 main.py
07/11/2021  14:04            300.445 Practica 1.docx
06/11/2021  14:33                692 README.md
06/11/2021  14:33                136 requirements.txt
07/11/2021  14:19                72 transaction1.csv
               6 archivos             306.744 bytes
               2 dirs      836.302.458.880 bytes libres
```



## Primer commit y push a la rama origin/master

```
C:\Users\teete\PracticaCloud\ingester-transactions>git add .

C:\Users\teete\PracticaCloud\ingester-transactions>git commit -m "Practica Cloud files added"
[master (root-commit) cca3b90] Practica Cloud files added
6 files changed, 190 insertions(+)
create mode 100644 Practica 1.docx
create mode 100644 README.md
create mode 100644 deployment.sh
create mode 100644 main.py
create mode 100644 requirements.txt
create mode 100644 transaction1.csv

C:\Users\teete\PracticaCloud\ingester-transactions>git push origin master
Enumerating objects: 8, done.
Counting objects: 100% (8/8), done.
Delta compression using up to 12 threads
Compressing objects: 100% (8/8), done.
Writing objects: 100% (8/8), 257.39 KiB | 16.09 MiB/s, done.
Total 8 (delta 0), reused 0 (delta 0), pack-reused 0
To https://source.developers.google.com/p/nifty-envoy-328012/r/ingester-transactions
* [new branch]      master -> master
```

Cloud Source Repositories

ingester-transactions > master

Archivos Resumen

Raíz del repositorio

- Practica 1.docx
- README.md
- deployment.sh
- main.py
- requirements.txt
- transaction1.csv

Introducción

En esta práctica vamos a preparar una Cloud Function que será responsable de "ingerir" ficheros de transacciones que se volcarán en un Bucket de Google Cloud Storage (cor exterior). Para ello necesitamos:

- Un bucket de comunicación
- Un ejemplo fichero que se volcará en el Bucket de comunicación
- Un dataset y una tabla en BigQuery donde volcar la información del archivo

La preparación de estos ingredientes ocurre en el fichero `deployment.sh`. El último paso de este script de despliegue es precisamente el despliegue de la Cloud Function en nu

Files

- Practica 1.docx
- README.md
- deployment.sh
- main.py
- requirements.txt
- transaction1.csv

Historial Capturas Puntos de registro

ID	Autor	Commit Date	Descripción
cca3b90	Teresa Gonzá...	2021-11-07 15:08	Practica Cloud files added

## 7. Estimar costes mensuales 1700 transacciones por segundo (almacenamiento y procesamiento)

### - CLOUD STORAGE

#### Almacenamiento

- Multi-region (US)
- Standard Storage

```
C:\Users\teete\PracticaCloud>gsutil du -sh gs://asr-cloud-201701454/transaction.csv
40 B gs://asr-cloud-201701454/transaction.csv
```

- Aproximadamente, cada archivo de transacción: 1 fila, 2 columnas son 40 bytes.

1700 transacciones/s -> En 1 segundo -> 68000 B= 66,40625 Kibibytes.

En un mes -> 178704000000 B = 174515625 Kibibytes = **166,431069374 Gibibytes**

## - BIG QUERY

### Almacenamiento

- Multi-region (US)
- Aproximadamente, cada archivo 27 bytes.

1700 transacciones/s -> En 1 segundo -> 45900 B = 44,8242188 Kibibytes.

En un mes -> 118972800000 B = 116184375 Kibibytes = **110,802054405212 Gibibytes**

### Extra: Procesamiento consultas (simples)

1 consulta -> 0.4 segundos y 27 B procesados/transacción

1700 transacciones/s -> En 1 segundo -> 45900 B = 44,8242188 Kibibytes.

En un mes -> 118972800000 B = 116184375 Kibibytes = **110,802054405212 Gibibytes**  
(active, long-term storage)

## - CLOUD FUNCTION (Procesamiento)

Tiempo ejecución -> 1.314 segundos. = 1314 ms

Tipo -> 512 MB (CPU 800 MHZ)

Número de invocaciones (<2 mill gratis):

1700 invocaciones /s -> 2592000 s (mes 30 días) \* 1700 = **4.406.400.000 invocaciones /mes**

## - COSTE TOTAL MENSUAL = 55323.70 \$

<div>Cloud Storage</div> <div>1x Standard Storage</div> <div>Location: United States</div> <div>Total Amount of Storage: 166.431 GiB</div> <div>Egress - Data moves within the same location: 0 GiB</div> <div>Always Free usage included: No</div> <div>USD 4.327</div> <div>BigQuery</div> <div>records</div> <div>Location: United States</div> <div>Active Storage 110.802 GiB</div> <div>Long-term Storage 110.802 GiB</div> <div>Queries 0.108 TiB</div> <div>USD 3.02</div>	<div>Cloud Functions</div> <div>ingest_transactions</div> <div>Invocations: 4,406,400,000</div> <div>GiB-seconds: 2,895,004,800 per month</div> <div>GHz-seconds: 4,632,007,680.000 per month</div> <div>Networking: 0 GiB per month</div> <div>USD 55,316.35</div> <div>A portion of your estimate fits within the Cloud Functions free tier. Please <a href="#">click here</a> for more detail.</div> <div>Total Estimated Cost: USD 55,323.70 per 1 month</div> <div>Estimate Currency</div> <div>USD - US Dollar</div>
--	--

Fuente: <https://cloud.google.com/products/calculator?hl=es>