🚀 Automatización de Carga Incremental con Cloud Run y Cloud Scheduler

Este documento describe el proceso completo para automatizar una carga incremental de datos en BigQuery, utilizando un Job de Cloud Run ejecutado diariamente mediante Cloud Scheduler.

# 📆 Contexto

La carga incremental permite actualizar tablas del Data Warehouse con nuevos datos provenientes de archivos cargados en Cloud Storage. Estos archivos se suben como tablas temporales a BigQuery y luego se comparan con las tablas principales (`dim\_\*` y `fact\_review`) para insertar solo los registros nuevos.

Tablas involucradas:  
- dim\_business  
- dim\_users  
- dim\_states  
- dim\_cities  
- dim\_locations  
- fact\_review

Objetivo: Evitar duplicados e insertar únicamente registros nuevos, validando datos y tipos, y limpiando las tablas temporales al finalizar.

# ⚙️ Lógica de la Carga Incremental

Pasos realizados por el script:  
1. Lectura de tablas temporales cargadas en BigQuery.  
2. Comparación con las tablas definitivas usando claves primarias para detectar nuevos registros.  
3. Inserción de datos nuevos en las tablas del DW.  
4. Validación de tipos de datos  
5. Eliminación de las tablas temporales una vez finalizada la carga.

# ⚖️ Configuración del Job en Cloud Run

Se creó un Job en Cloud Run llamado:

***etl-incremental-job***

***Región:  
us-central1***

***Comando para listar los Jobs:  
gcloud run jobs list --region=us-central1***

# 🔐 Cuenta de Servicio y Permisos

***Cuenta usada por Cloud Scheduler para invocar el Job:  
scheduler-run-invoker@shining-rampart-455602-a7.iam.gserviceaccount.com***

***Permiso otorgado:  
gcloud run jobs add-iam-policy-binding etl-incremental-job \  
 --region=us-central1 \  
 --member="serviceAccount:scheduler-run-invoker@shining-rampart-455602-a7.iam.gserviceaccount.com" \  
 --role="roles/run.invoker"***

# 🗓️ Crear Job en Cloud Scheduler

***Job creado para ejecutar diariamente el Job de Cloud Run:  
  
gcloud scheduler jobs create http scheduler-etl-incremental \  
 --project=shining-rampart-455602-a7 \  
 --location=us-central1 \  
 --schedule="0 8 \* \* \*" \  
 --http-method=POST \  
 --uri="https://us-central1-run.googleapis.com/apis/run.googleapis.com/v1/namespaces/shining-rampart-455602-a7/jobs/etl-incremental-job:run" \  
 --oauth-service-account-email=scheduler-run-invoker@shining-rampart-455602-a7.iam.gserviceaccount.com \  
 --time-zone="America/Argentina/Buenos\_Aires"***

# 🔧 Probar y Ejecutar Manualmente

***Ejecutar manualmente el Job:  
gcloud scheduler jobs run scheduler-etl-incremental \  
 --location=us-central1  
  
Ver historial de ejecuciones:  
gcloud run jobs executions list \  
 --job=etl-incremental-job \  
 --region=us-central1***

Todo listo para ejecutar el ETL incremental todos los días automáticamente con confiabilidad en la nube