

SQL Replicate en Google Cloud

Para implementar algo similar a SQL Replication en Google Cloud, se utilizan varias herramientas que emulan el comportamiento del publicador, suscriptor y distribuidor.

A continuación, se explica cómo se integran estos conceptos en la arquitectura en Google Datastream para la replicación:

1. Publicador (Publisher)

El publicador es la fuente de datos, en este caso es la base de datos MySQL que contiene los datos provenientes de los archivos CSV. Esta base de datos se llena con datos mediante un proceso pipeline que lee los CSV desde Cloud Storage y los carga en la base de datos.

- En este caso, el Publicador sería la base de datos MySQL en Cloud SQL que almacena los datos de los archivos CSV.

2. Distribuidor (Distributor)

El distribuidor es el encargado de gestionar la replicación, asegurando que los datos del Publicador se envíen de manera correcta a los Suscriptores. En Google Cloud el servicio que puede cumplir este rol es Google Datastream.

- Datastream es el componente que actúa como el distribuidor. Monitorea la base de datos MySQL para detectar en tiempo real mediante CDC. Posteriormente, Datastream se encarga de enviar estos cambios a BigQuery, donde se almacenan y se actualizan de manera continua.

3. Suscriptor (Subscriber)

El suscriptor es el destino donde se reciben los datos replicados. En este caso, el suscriptor es BigQuery.

- BigQuery es el suscriptor, ya que recibe los datos replicados desde la base de datos MySQL a través de Datastream. Los datos se almacenan en tablas y se actualizan automáticamente conforme cambien en la base de datos fuente.

4. Artículo (Article)

Los artículos son los objetos individuales en la base de datos MySQL.

- Google SQL nos permite generar sistemas de gestión de bases de datos, en este caso, el gestor de base de datos es MySQL que alberga diferentes tablas que van a ser replicadas en BigQuery.

¿Cómo encajan estos roles en el flujo?

1. Publicados: Base de datos Cloud SQL (MySQL) que almacena los datos importados desde los archivos CSV.
2. Distribuidor: Google Datastream actúa como distribuidor, capturando los cambios en la base de datos MySQL y distribuyéndolos a BigQuery.
3. Suscriptor: Base de datos que contiene la replicación de los datos CSV en donde la data está situada en MySQL.