

Documentação Técnica: Guia de ELT no BigQuery

1. Resumo Executivo

Este guia descreve o processo de **ELT (Extract, Load, Transform)** de dados extraídos de uma **API**, carregados em um **Cloud Storage Bucket** e, posteriormente, processados no **Google BigQuery**. O objetivo é padronizar a ingestão e transformação de dados, garantindo um fluxo robusto, escalável e de fácil manutenção para qualquer membro da equipe.

2. Prioridades

- **Extração da API:** Conectar e extrair dados de uma fonte externa (API) de forma automatizada e programática. Esta é a etapa inicial e de alta prioridade.
- **Carga no Bucket:** Carregar os dados brutos e sem transformação em um *bucket* do Cloud Storage. Este passo garante que tenhamos uma cópia fiel e intacta dos dados de origem para futuras auditorias ou reprocessamentos.
- **Transformação no BigQuery:** Utilizar o poder de processamento do BigQuery para limpar, validar e enriquecer os dados. Esta é a etapa final, onde os dados são preparados para consumo analítico.

3. Responsabilidades e Histórico

- **Responsável:** Clauderson Branco Xavier
- **Ações:** Configuração do *bucket* de destino e criação dos *scripts* de transformação no BigQuery.
- **Data de Atualização:** 24/09/2025

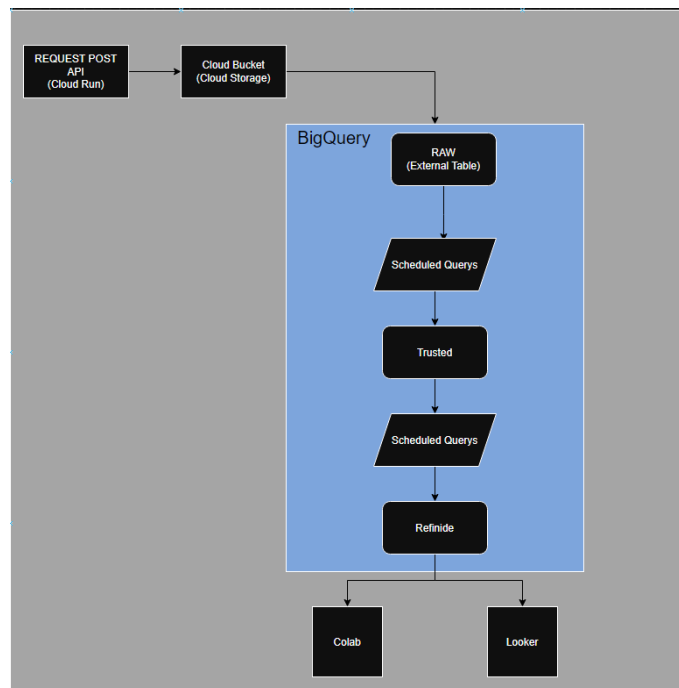
4. Diagramas de Arquitetura e Fluxo de Dados

4.1 Fluxo de Dados ELT

1. **Extract (Extração):** Um processo (por exemplo, um script Python ou uma função Cloud) faz uma chamada à API para extrair os dados.
2. **Load (Carga):** Os dados brutos, no formato original da API (por exemplo, JSON ou CSV), são carregados diretamente em um *bucket* do Cloud Storage.

3. **Transform (Transformação):** O BigQuery acessa os dados diretamente do *bucket* (usando tabelas externas) e executa *queries* SQL para limpá-los, padronizá-los e, finalmente, carregá-los em tabelas finais prontas para análise.

4.2 Diagrama de fluxo



5. Guia de Implementação Passo a Passo

5.1. Pré-requisitos

- Acesso e credenciais para a API de origem.
- Acesso ao projeto **Google Cloud** com permissões para **Cloud Storage** e **BigQuery**.
- Ferramenta para orquestração da extração (por exemplo, Cloud Run, ou um script local).

5.2. Etapas do Processo

1. **Extração:** Crie um script para chamar a API e obter os dados.
2. **Carregamento:** Configure o script para salvar os dados brutos em um *bucket* no Cloud Storage (Preferencialmente em *.parquet*).
3. **Transformação:**
 - Crie uma **Tabela Externa** no BigQuery apontando para o arquivo no *bucket*.

- Crie *queries* SQL que leiam da Tabela Externa, apliquem as transformações (limpeza, filtros, etc.) e insiram os dados em uma tabela final, por exemplo, na sua camada **trusted** depois crie views para a **refined**.

6. Segurança

- **Credenciais da API:** Armazene as chaves ou tokens da API em um local seguro, como o **Secrets Manager** do Google Cloud, e **nunca** as inclua diretamente no código.
- **Permissões:** Garanta que a conta de serviço que acessa a API e o *bucket* tenha as permissões mínimas necessárias. O BigQuery precisa de permissão de leitura no *bucket*.

7. Registro de Decisões Arquiteturais

- **Decisão:** Utilizar o Cloud Storage como ponto intermediário de carga dos dados brutos.
- **Justificativa:** A carga intermediária no *bucket* garante um backup dos dados brutos, desacopla o processo de extração do de transformação e facilita o reprocessamento dos dados, caso necessário.

8. Changelog

- **v1.0.0 (Data):** Implementação inicial do fluxo de ELT com API, Cloud Storage e BigQuery.

9. Branch Trabalhada

Abaixo segue alguns Pull requests de branches trabalhadas trabalhadas.

[34-t15---fazer-ingestão-dos-dados-no-bq](#)

