

Arquitetura no ambiente Google Cloud Platform

Case técnico Enjoei

Luiz Guimarães de Melo Neto
Data Engineer

Arquitetura GCP

Análise Automatizada de Carrinhos de Usuários para Estratégias de Marketing no Varejo

1. Cenário e Necessidade

- Competição acirrada no varejo exige decisões baseadas em dados rápidos e precisos.
- Análises de carrinhos de usuários são essenciais para insights e estratégias de marketing.

2. Objetivo do Projeto

- Mensurar e agregar dados de carrinhos para facilitar análise.
- Gerar insights acionáveis para o time de marketing e BI.

3. Solução Proposta: Arquitetura Automatizada

- Coleta e análise diária dos dados de carrinhos dos clientes.
- Processamento e armazenamento em um fluxo automatizado.

Arquitetura GCP

4. Benefícios da Arquitetura

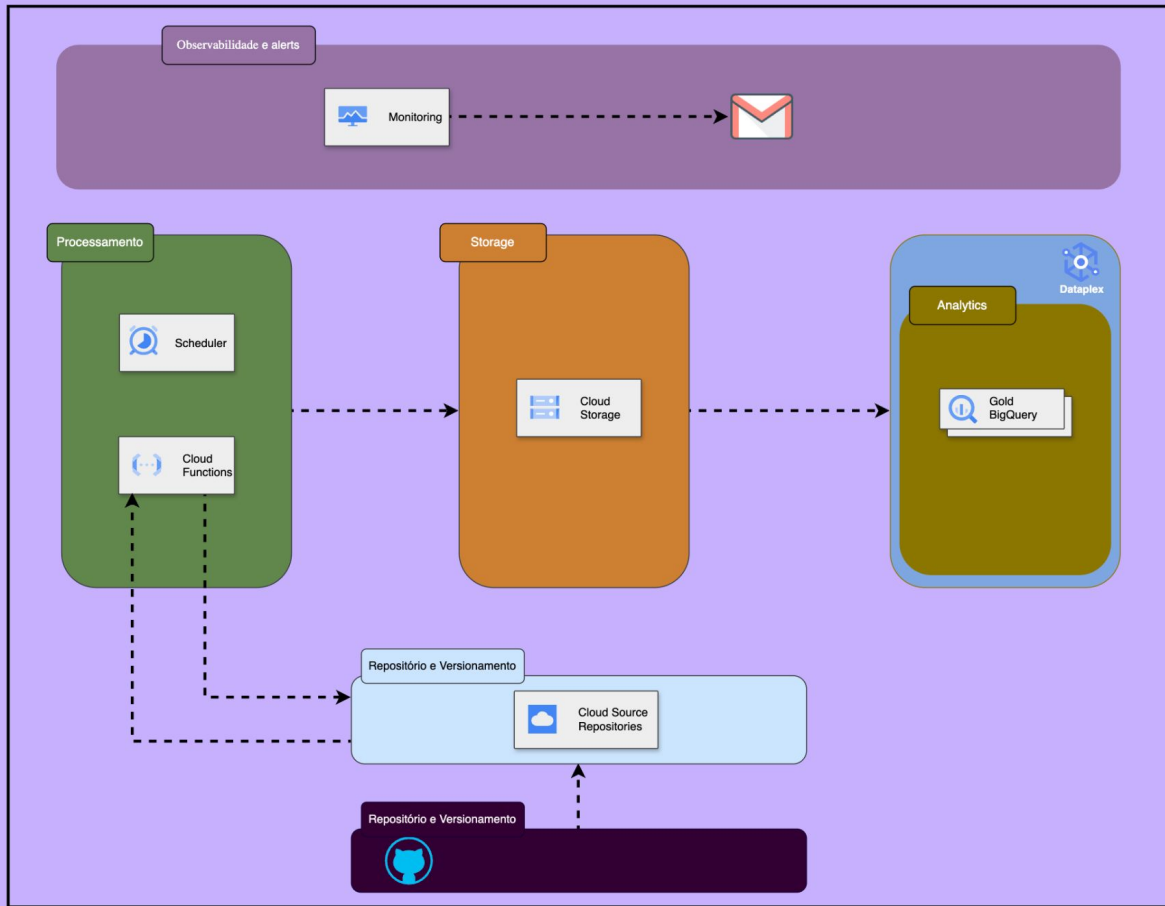
- Extração contínua e 100% confiável dos dados.
- Exportação automática para CSV no Google Cloud Storage.
- Integração fácil com dashboards para o time de BI.
- Arquitetura barata e totalmente escalável, adaptando-se às necessidades de negócio.
- Ambiente de insights mais frequentes e confiáveis para decisões ágeis e estratégias adaptáveis.

Overview da Arquitetura

Arquitetura utilizada para extração, limpeza e disponibilização dos dados de carrinhos dos usuários.



Arquitetura GCP



Serviços:



Cloud Functions

Utilizado para fazer o armazenamento do nosso código Python que fará a conexão com a API e retornará os dados.

Documentação: <https://cloud.google.com/functions/docs?hl=pt-br>



BigQuery

Utilizado para fazer o armazenamento, consulta e transformações dos dados vindos do Google Analytics.

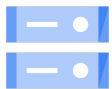
Documentação: <https://cloud.google.com/bigquery/docs?hl=pt-br>



Identity And Access Management

Responsável por toda a parte de permissões da nossa arquitetura.

Documentação: <https://cloud.google.com/iam/docs/overview?hl=pt-br>



Cloud Storage

Responsável por armazenar e disponibilizar os dados em formato CSV.

Documentação: <https://cloud.google.com/storage/docs?hl=pt-br>

Serviços:



Cloud Scheduler

O Cloud Scheduler é um serviço do Google Cloud para agendar e automatizar tarefas como cron jobs e chamadas de APIs em intervalos específicos.

Documentação: <https://cloud.google.com/scheduler/docs?hl=pt-br>



Cloud Monitoring

O Cloud Monitoring é um serviço do Google Cloud que coleta, visualiza e alerta sobre métricas de desempenho de recursos e aplicações em nuvem e on-premises.

Documentação: <https://cloud.google.com/monitoring?hl=pt-br>



Cloud Source Repositories

O Source Repositories é um serviço de repositórios Git totalmente gerenciado no Google Cloud, permitindo hospedagem, versionamento e colaboração em projetos de desenvolvimento diretamente na nuvem.

Documentação: <https://cloud.google.com/iam/docs/overview?hl=pt-br>

Premissas

- Definir as regras de localização e tempo de vida dos arquivos presentes no bucket;
- Definir quais contas terão acesso ao bucket e quais permissões serão dadas;
- Deverá também ser disponibilizada uma Service Account do GCP e a partir dela os dados serão enviados para o Cloud Storage;



Boas práticas + Pontos de atenção

- Atenção com as concessões de permissões no IAM. Todas devem ser bem avaliadas para que seja mantido a segurança dos serviços e garantir que ninguém sem o conhecimento devido altere algum processo;
- Quaisquer modificações no pipeline dos dados deve ser validada para que não haja quebra na ordem dos processos;

Pontos de melhorias

- Essa arquitetura foi pensada para garantir a eficiência da entrega com o menor custo possível, caso seja possível um custo maior poderíamos implantar as seguintes melhorias:
 - Utilizar uma máquina virtual com um Airflow para orquestração dos dados;
 - Para evitar que o processo seja feito via Python, poderíamos subir cada uma das requisição da API como uma tabela bruta no bigquery e fazer os cruzamentos via SQL, utilizando o Dataform ou as Schedules Querys.