




BOTICÁRIO - DATA ENGINEERING

Éricka Caracas de Castro Santos

Objetivo



Definir arquitetura de referência com
tecnologias de nuvem pública

Requisitos

Permear as camadas de ingestão, processamento, armazenamento, consumo, análise, segurança e governança

Substituição gradativa do cenário on-premises atual

Incorporação de componentes e tecnologias que permitam a analisarmos dados em tempo real

Arquitetura que atenda diversos fins como Analytics, Data Science, API's

Arquitetura de Dados

Google Cloud Platform

Vantagens

- Alta escalabilidade
- Gerenciamento remoto
- Compartilhamento e processamento de dados em tempo real
- Machine learning / IA

Desvantagens

- Dependência de conexão à internet
- Despesas inesperadas e custos contínuos
- Menos controle sobre a infraestrutura
- Riscos de segurança relacionados à privacidade de dados

Fonte de dados

SAP HANA



Arquivos



REAL TIME DATA

Dispositivos



Pagamentos



Vendas



Ingestão e Processamento



Cloud Data Fusion



Dataproc



Cloud Functions



Cloud Storage

Streaming (tempo real)

Serviço de Mensagens assíncrono e escalonável



Pub/Sub



Processamento

Dataflow



Armazenamento



BigQuery

Camadas

RAW

Dados brutos sem tratamento ou transformação.

Ideal separar em camada de dados sensíveis e não sensíveis para conformidade com LGPD e GDPR.

TRUSTED

Dados tratados, transformados, padronizados e confiáveis para uso. Ideal separar em camada de dados sensíveis e não sensíveis para conformidade com LGPD e GDPR.

REFINED

Dados refinados e enriquecidos com regras de negócio para atender diversas áreas específicas. Ideal separar em camada de dados sensíveis e não sensíveis para conformidade com LGPD e GDPR.

Governança



Data Catalog



Cloud IAM



Cloud Billing

Disponibilização

Ambiente para análises exploratórias



sandbox_planejamento



sandbox_comercial



sandbox_juridico



sandbox_suporte



Visualização



Vertex AI



Looker Studio



Power BI



Tableau



Google Sheets



Segurança e Monitoramento



Cloud Monitoring



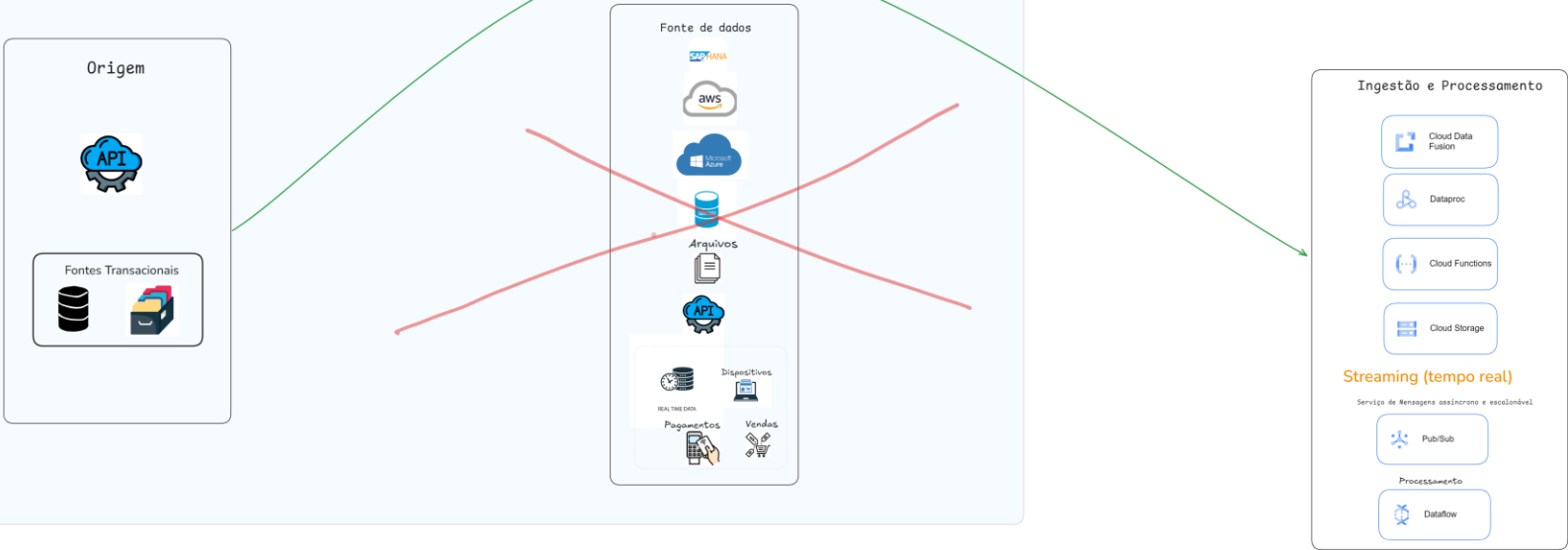
Cloud Audit Logs



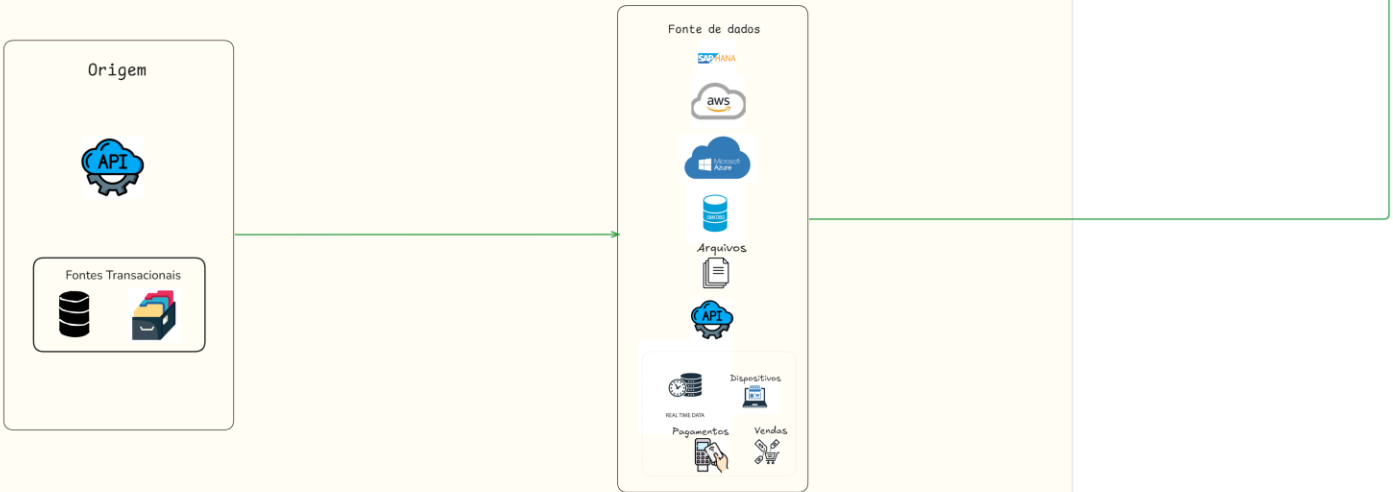
Secret Manager

Substituição gradativa do cenário on-premises atual

Dados Atuais



Dados Históricos



Importante

Para garantir o bom funcionamento da arquitetura de referência é necessária uma forte cultura de dados na empresa, adoção e padronização do modelo adotado por todos os funcionários.

Práticas como workshops recorrentes sobre como otimizar custos de consultas, conformidade com LGPD/GDPR e documentação robusta com responsabilidade de todos resultarão em ótimo desempenho e resultados.

Contato

Éricka Caracas

ericka.caracas@gmail.com

<https://github.com/erickacaracas>

[linkedin.com/in/erickacaracas/](https://www.linkedin.com/in/erickacaracas/)

Obrigada ! 