

CASO DE SUCESSO

UNITY CATALOG

DATABRICKS



/ Migração e otimização de Lakehouse usando o Unity Catalog

Este projeto simplificou as operações, reduziu custos e aproveitou as capacidades avançadas do **Unity Catalog** para aprimorar a **governança e a segurança dos dados**.



DESAFIOS

Gestão Complexa de Dados

- / A cliente gerenciava um extenso e complexo Data Lakehouse, composto por **5 workspaces** (3 Premium e 2 Standard), enfrentando desafios operacionais significativos com mais de **1000 execuções diárias de jobs e mais de 100 usuários**.

/ Configuração Complexa

O Apache Ranger foi utilizado para **maskamento de dados**, rodando em um cluster do Azure Kubernetes Service com suporte do PostgreSQL, para armazenar regras de acesso. Sua configuração era complexa, e **qualquer falha poderia expor dados sensíveis**.

/ Criação Desorganizada de Dados

Com a liberdade que os usuários tinham para criarem tabelas, houve uma desorganização significativa na categorização e contexto dos dados, o que complicou a descoberta de informações.

/ Custos de Armazenamento Desnecessários

Arquivos e tabelas de usuários com pouco conhecimento da plataforma eram salvos no DBFS, resultando em custos desnecessários de armazenamento, já que não havia gerenciamento sobre quem e como o DBFS estava sendo utilizado.

/ Armazenamento Inadequado de Metadados

O uso de um metastore externo, baseado em MySQL, enfrentou **escassez de memória**, afetando o **desempenho** e a **escalabilidade**.

/ Workspace Databricks Isolado

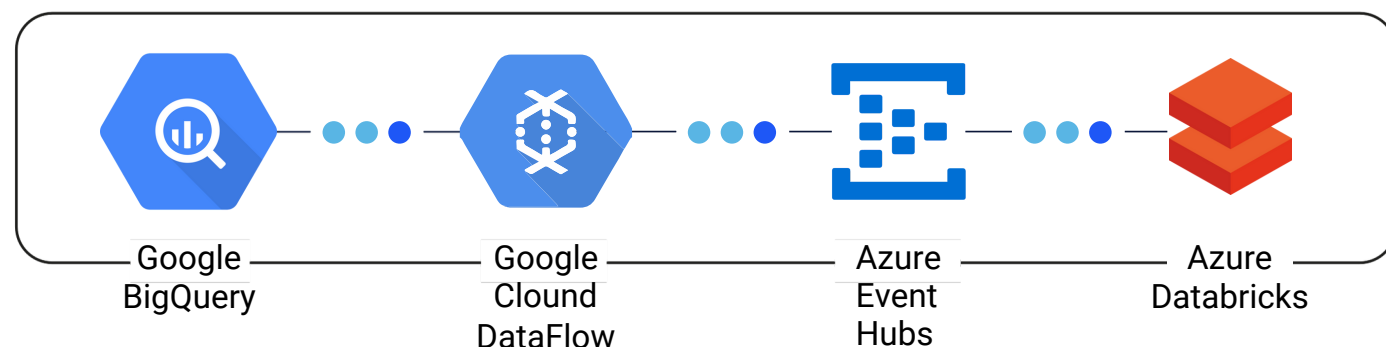
Sempre que alguém externo precisava acessar dados dentro da plataforma, um workspace Databricks isolado era criado com tabelas clonadas e anonimizadas, acessadas por meio de terminais virtuais gerenciados pela empresa.

/ Controle de Acesso Inconsistente

O acesso ao armazenamento era concedido por meio de chaves do Storage Access Keys, fornecendo **acesso total ao armazenamento** e configurado via Init Script.

/ Problemas em Ambientes Multi-Cloud

Como se tratava de um ambiente multi-cloud, algumas tabelas de outras ferramentas eram acessadas. No entanto, para evitar sobrecarregar as ferramentas e abusar do uso das conexões, fluxos foram criados para enviar os dados dessas ferramentas para as plataformas na nuvem correspondentes, **aumentando a complexidade na obtenção desses dados**. Toda essa integração foi feita via código:



SOLUÇÃO

A migração e otimização do Data Lakehouse da cliente melhorou a governança de dados e aumentou a eficiência geral.



BENEFÍCIOS

/ Migração Eficiente de Dados

Ao habilitar logs e analisar o uso das tabelas, a cliente reduziu o número de tabelas de mais de **6000 para aproximadamente 600**. Essa redução **diminuiu o esforço de migração e reduziu os custos de armazenamento**.

/ Controle de Acesso Simplificado

O gerenciamento centralizado dos controles de acesso garantiu consistência e segurança, com quaisquer alterações sendo automaticamente sincronizadas entre todos os workspaces.

/ Melhoria na Organização dos Dados

O Unity Catalog foi estruturado em três camadas, permitindo uma melhor categorização e facilitando o gerenciamento de tabelas e bancos de dados, organizados por departamento ou camadas da arquitetura Medallion.

/ Redução da Complexidade na Ingestão de Dados

A substituição de configurações anteriores pela Federação Lakehouse simplificou a arquitetura e reduziu os custos operacionais.

/ Gerenciamento Centralizado de Metadados

O metastore centralizado do Unity, eliminou a necessidade de bancos de dados adicionais para metadados de tabelas, simplificando o acesso em múltiplos workspaces.

/ Acesso Seguro ao Armazenamento

A transição para Identidades Gerenciadas para o acesso ao armazenamento aumentou significativamente a segurança e facilitou o gerenciamento.



BENEFÍCIOS

/ Mascaramento de Dados Simplificado

As novas funcionalidades de mascaramento e controle de acesso do sistema reduziram a necessidade de configurações complexas e serviços externos.

/ Armazenamento de Dados com Melhor Custo-Benefício

A implementação de Volumes direcionou os uploads de dados para opções de armazenamento mais baratas, reduzindo custos e eliminando a desorganização.

/ Compartilhamento Eficiente de Dados com Pessoas Externas

O compartilhamento de dados com externos foi otimizado usando o Delta Sharing, melhorando o controle de acesso sem a necessidade de ambientes personalizados.

/ Rastreamento Rápido da Origem dos Dados

A interface de linhagem do Unity Catalog facilitou o rastreamento rápido das origens dos dados, aumentando a produtividade e a governança de dados.



BIG NUMBERS

Redução do número de tabelas

ANTES

- Mais de

6000



DEPOIS

- Aproximadamente

600

- 5 workspaces enfrentando desafios operacionais significativos com mais de 1000 execuções diárias de jobs e mais de 100 usuários.



- Organizados por departamento ou camadas da arquitetura Medallion.

