

WEBINAR

Modernize seu Data Warehouse

Uma jornada de migração para a
plataforma Databricks Lakehouse



Iniciaremos em breve

- A sessão será gravada
- Adicione suas perguntas no Q&A e interaja no chat
- Usem o botão "Reações" do Zoom
- Faça parte do nosso grupo no Linkedin Lakehouse Brasil





WEBINAR

Modernize seu Data Warehouse

Uma jornada de migração para a
plataforma Databricks Lakehouse

30 de outubro de 2024

Palestrantes:



Luis Assunção
DATABRICKS



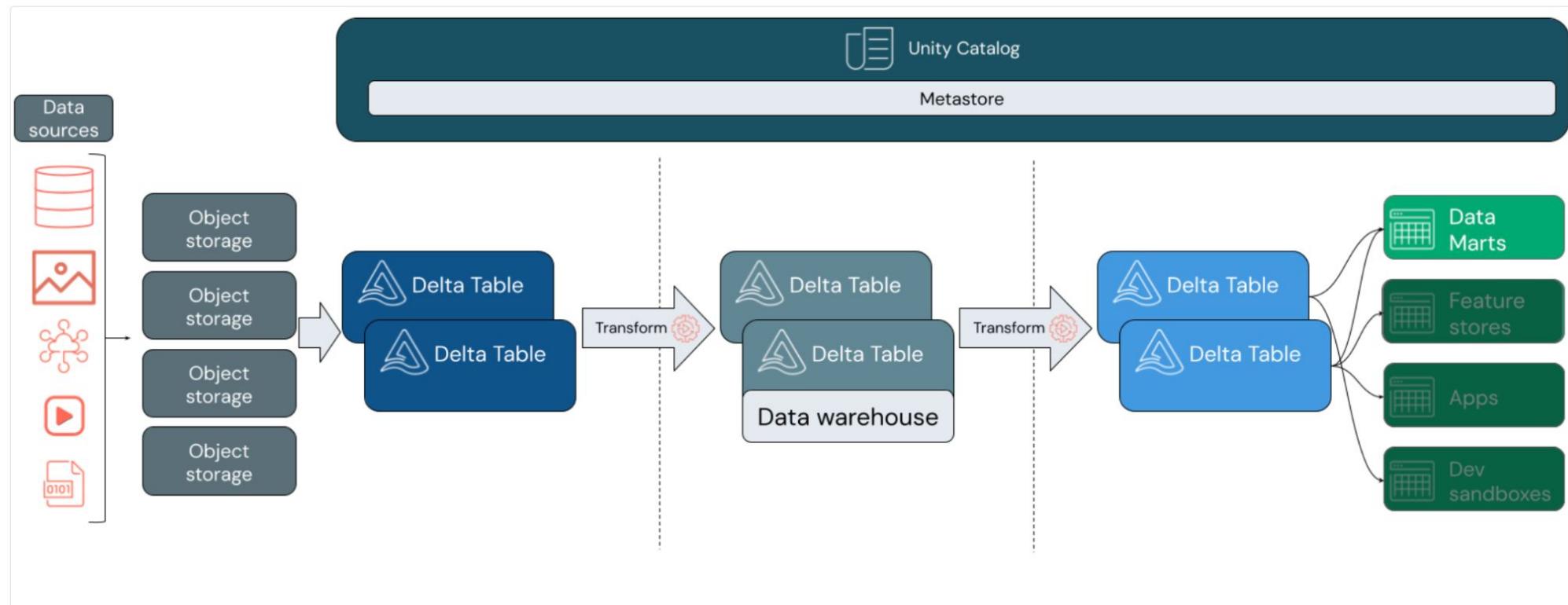
Luiz Carrosoni
DATABRICKS

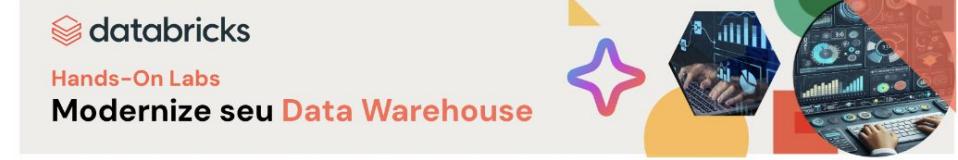




A arquitetura **LAKEHOUSE** e o **Databricks SQL** trazem recursos de Data Warehouse para sua Arquitetura de Dados.

Usando estruturas de dados, relações e ferramentas de gerenciamento familiares, você pode modelar um Data Warehouse de alto desempenho e econômico que pode ser executado diretamente em seu Data Lake.





Material disponível no repositório:



Databricks-BR

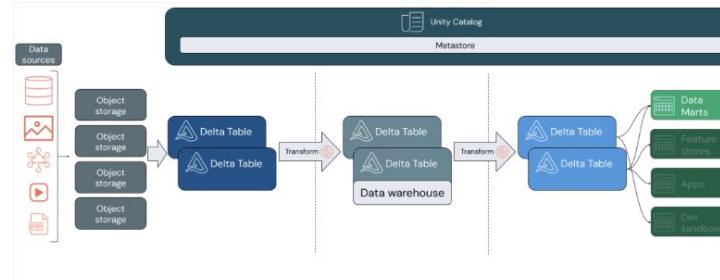


github.com/Databricks-BR/dw

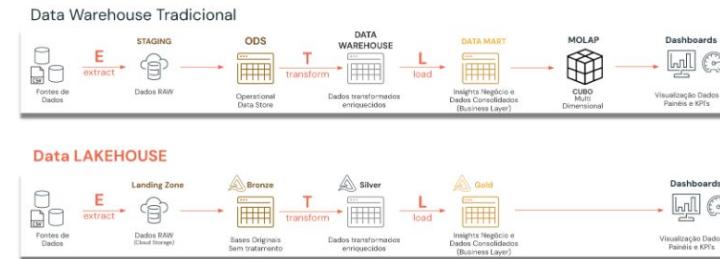
Databricks Labs - MODERNIZE SEU DATA WAREHOUSE

Treinamento na plataforma Databricks com foco nas funcionalidades em DATA WAREHOUSE.

A arquitetura LAKEHOUSE e o Databricks SQL trazem clouds data warehousing recursos para o seu data lake. Usando estruturas de dados, relações e ferramentas de gerenciamento familiares, você pode modelar um data warehouse de alto desempenho e econômico que pode ser executado diretamente em seu data lake.



Tal como acontece com um Data Warehouse tradicional, você modela os dados de acordo com os requisitos de negócios e depois os fornece aos usuários finais para análises e relatórios. Ao contrário de um data warehouse tradicional, é possível simplificar e unificar a Arquitetura de Dados, trazendo maior agilidade e colaboração na equipe de Dados (engenheiros, cientistas e analistas), com toda a Governança, Segurança e Rastreabilidade, garantindo a integração com o ecossistema de soluções de dados, incluindo as diversas ferramentas de Visualização de Dados.



Referências

- [O que é um Data Warehousing no Databricks](#)
- [Página principal do Databrick SQL](#)
- [SQL Language reference - Inglês](#)
- [SQL Language reference - Português](#)



DATA
WAREHOUSE



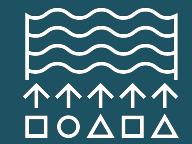
Lakehouse

Diferenciais

Modelagem
de Dados



Data
Quality



Eficiência
de Custos



Performance



Data
Visualization



Proteção
dos dados



Governança

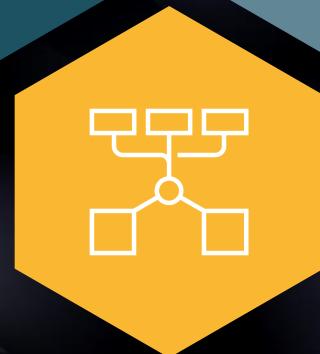


Estratégias
de Migração



Gen AI

DATA WAREHOUSE



Camadas de Transformação de Dados

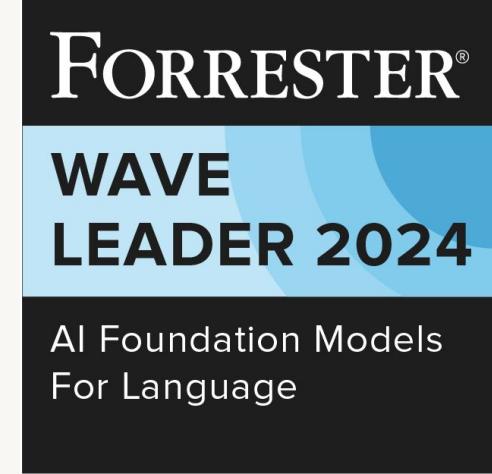
Data Warehouse Tradicional



Data LAKEHOUSE

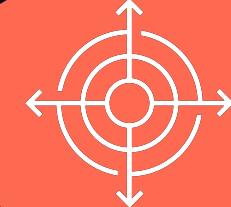
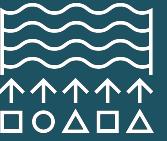


6x Líder em <6 Meses!



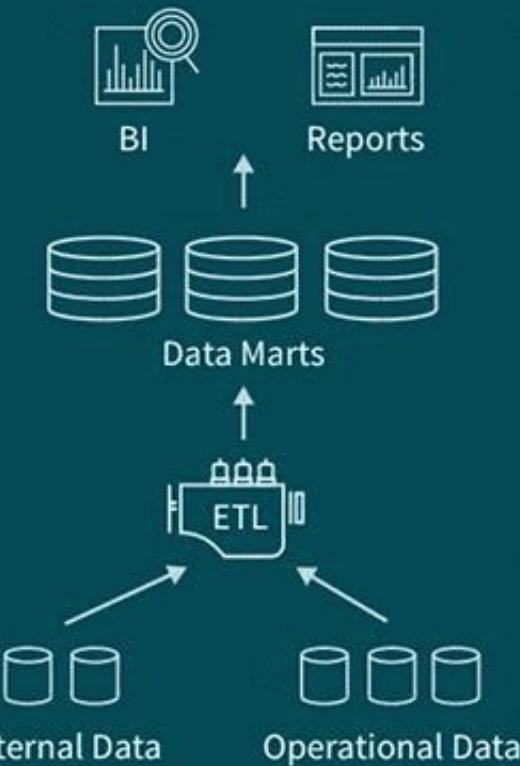
Evolução DATA WAREHOUSE

Lakehouse



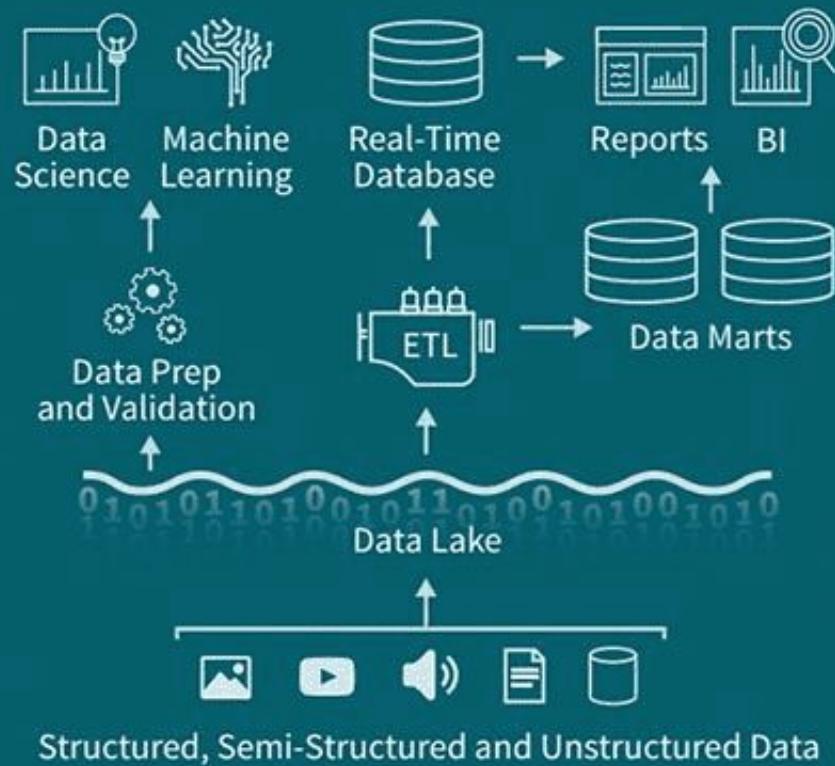
Late 1980's

Data Warehouse



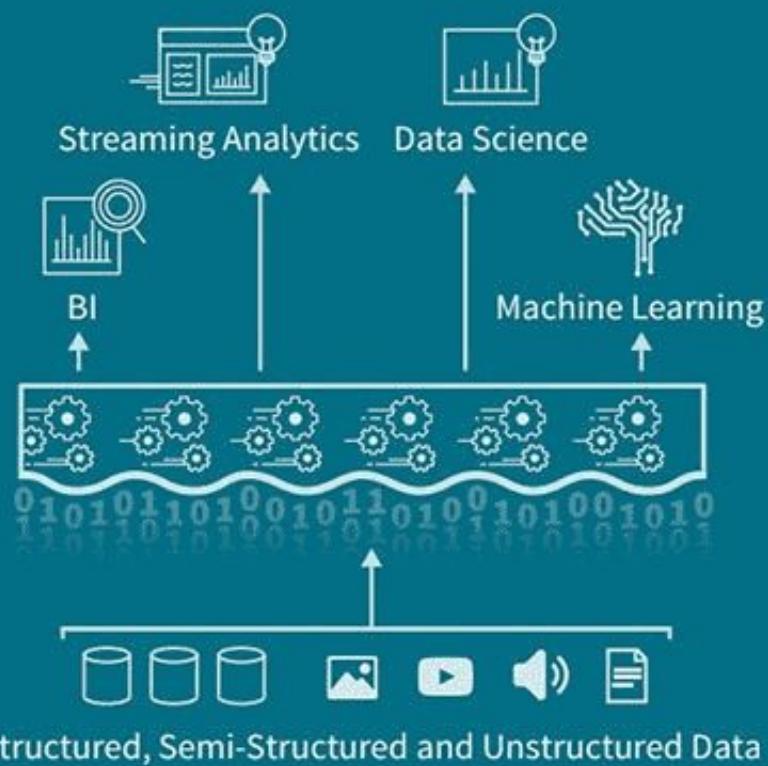
2011

Data Lake



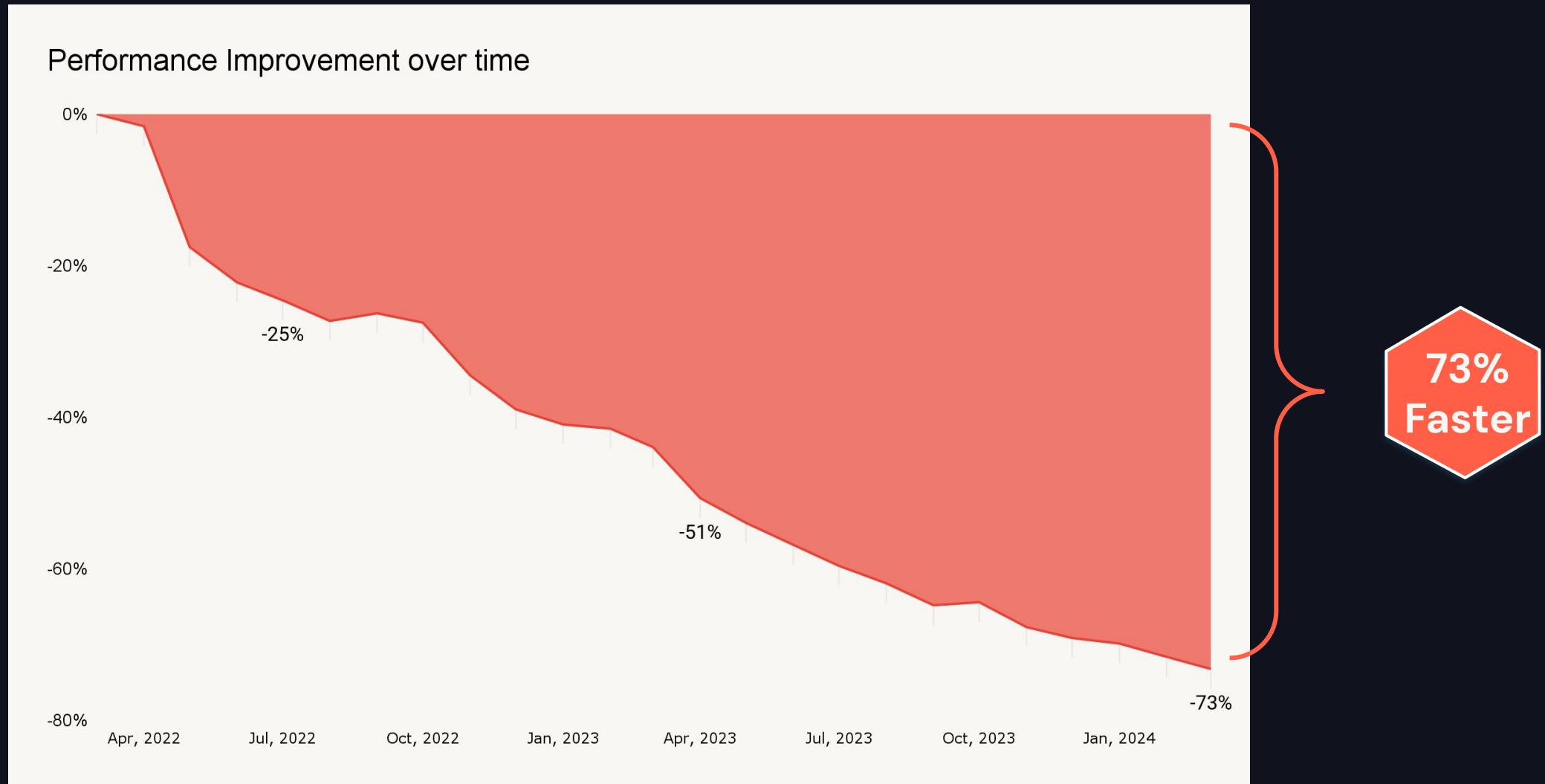
2020

Lakehouse



SOURCE: DATABRICKS

BI queries de Cliente - melhora de 73% nos últimos 2 anos



Predictive Optimization

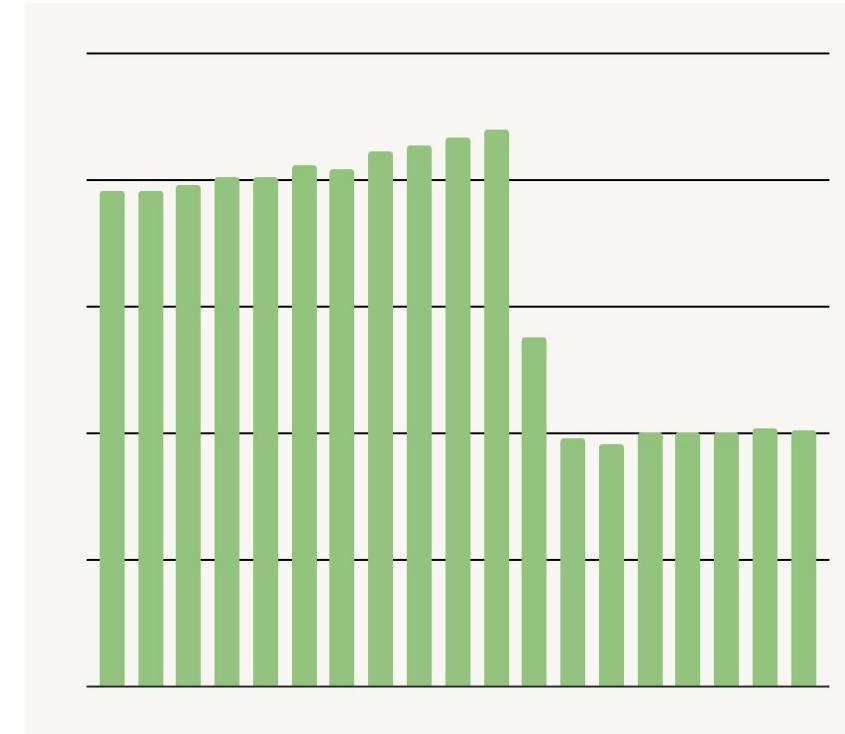
Layouts de tabelas Delta otimizados por IA para melhor relação custo-desempenho



Responsável por:

OPTIMIZE, VACUUM, ANALYZE, Liquid clustering

Os dados são otimizados e agrupados de forma inteligente para tornar as consultas mais rápidas e reduzir os custos de armazenamento



Liquid Clustering

Alto desempenho e fácil de usar

Chega de usar Partições.

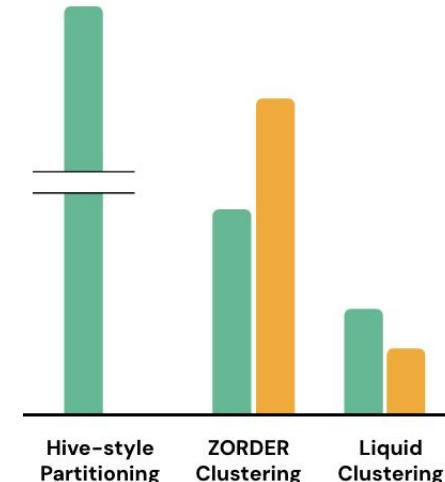
```
OPTIMIZE minha_tabela  
ZORDER BY (data, cpf, sku_id)
```



Faster, Cheaper

Ingestion and Clustering

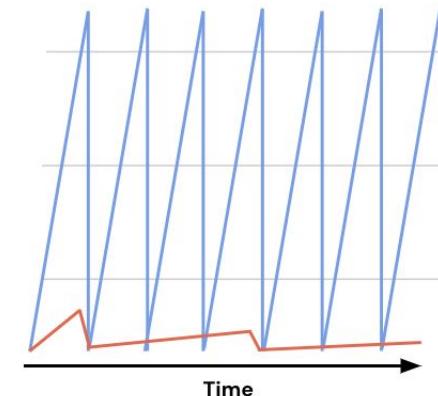
Time Write amplification (bytes)



Read Performance

Bytes scanned for Point Queries
Lower is better

ZORDER Liquid



Lakehouse Federation

Descubra, Explore e Governe todos os seus dados



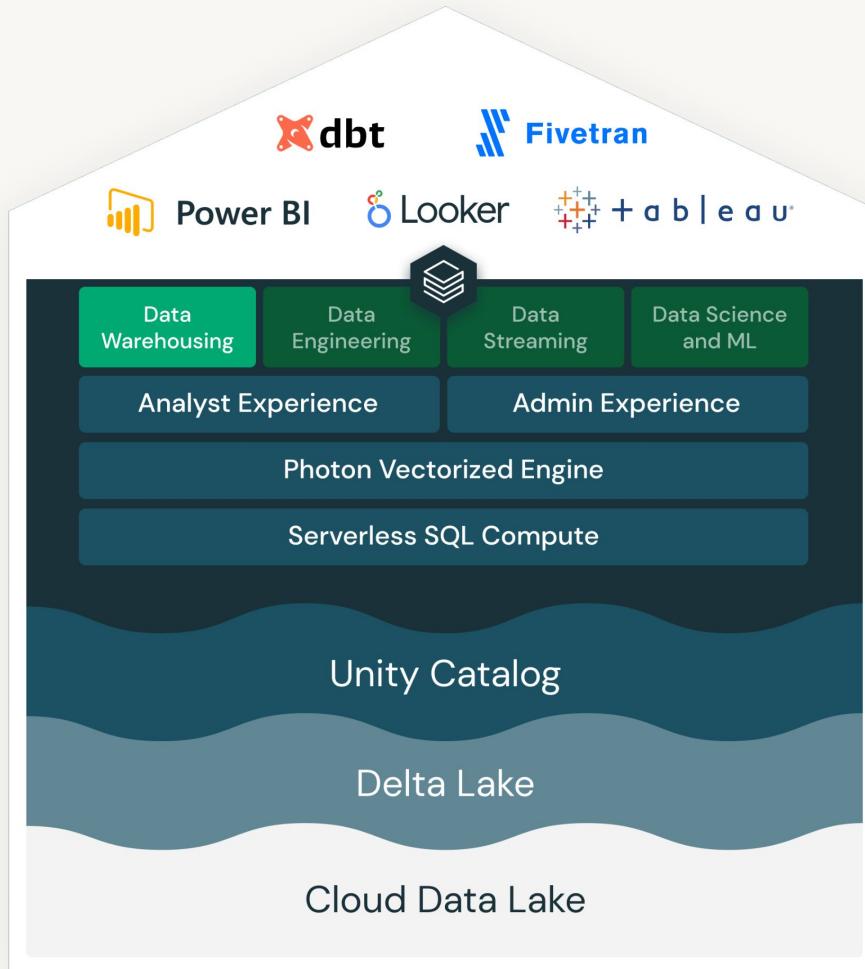
DIFERENCIAIS

Lakehouse



O melhor DW é o Lakehouse

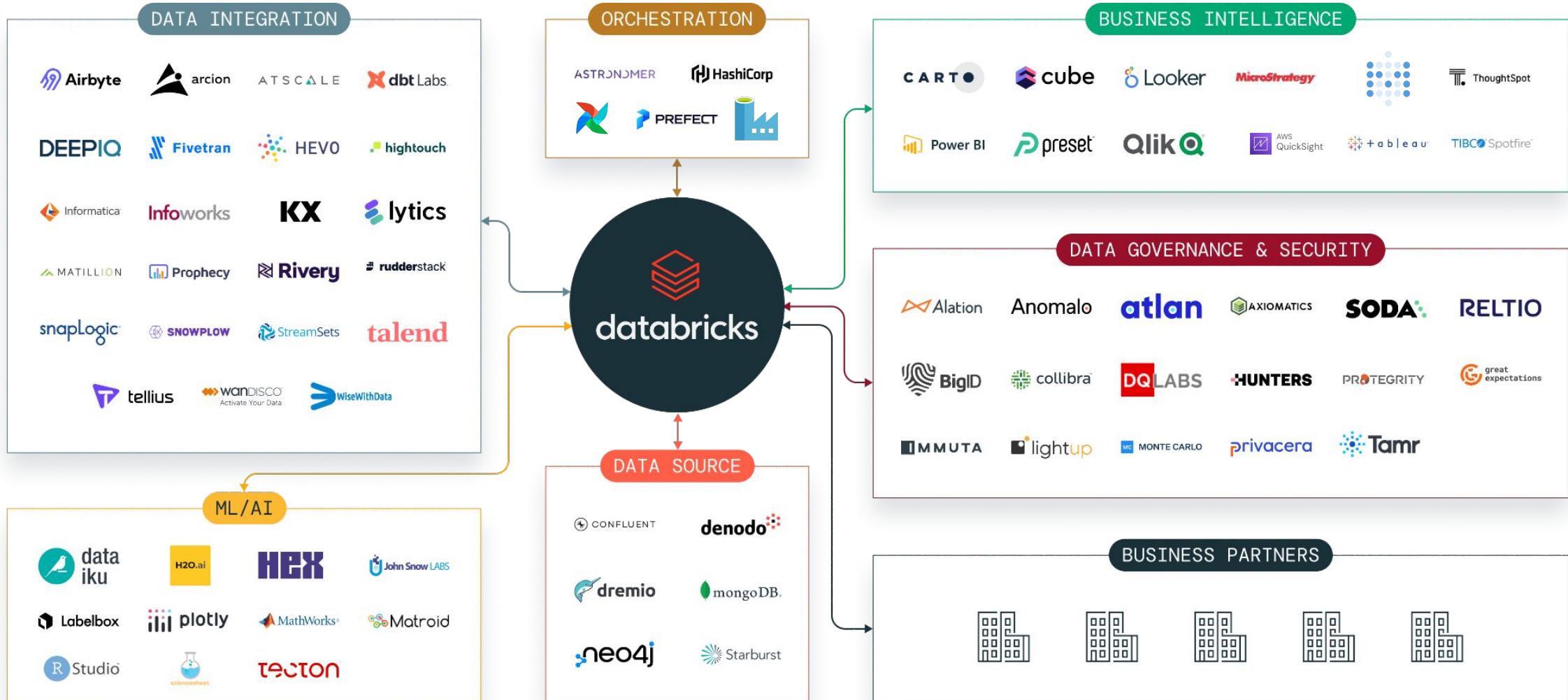
Potencializado pelo *Databricks SQL*



- Excelente **integração** com o Ecossistema
- **Facilidade** de Uso
- **Desempenho** e Escalabilidade
- **Governança** de Dados centralizada



Databricks possui um ecossistema moderno de dados





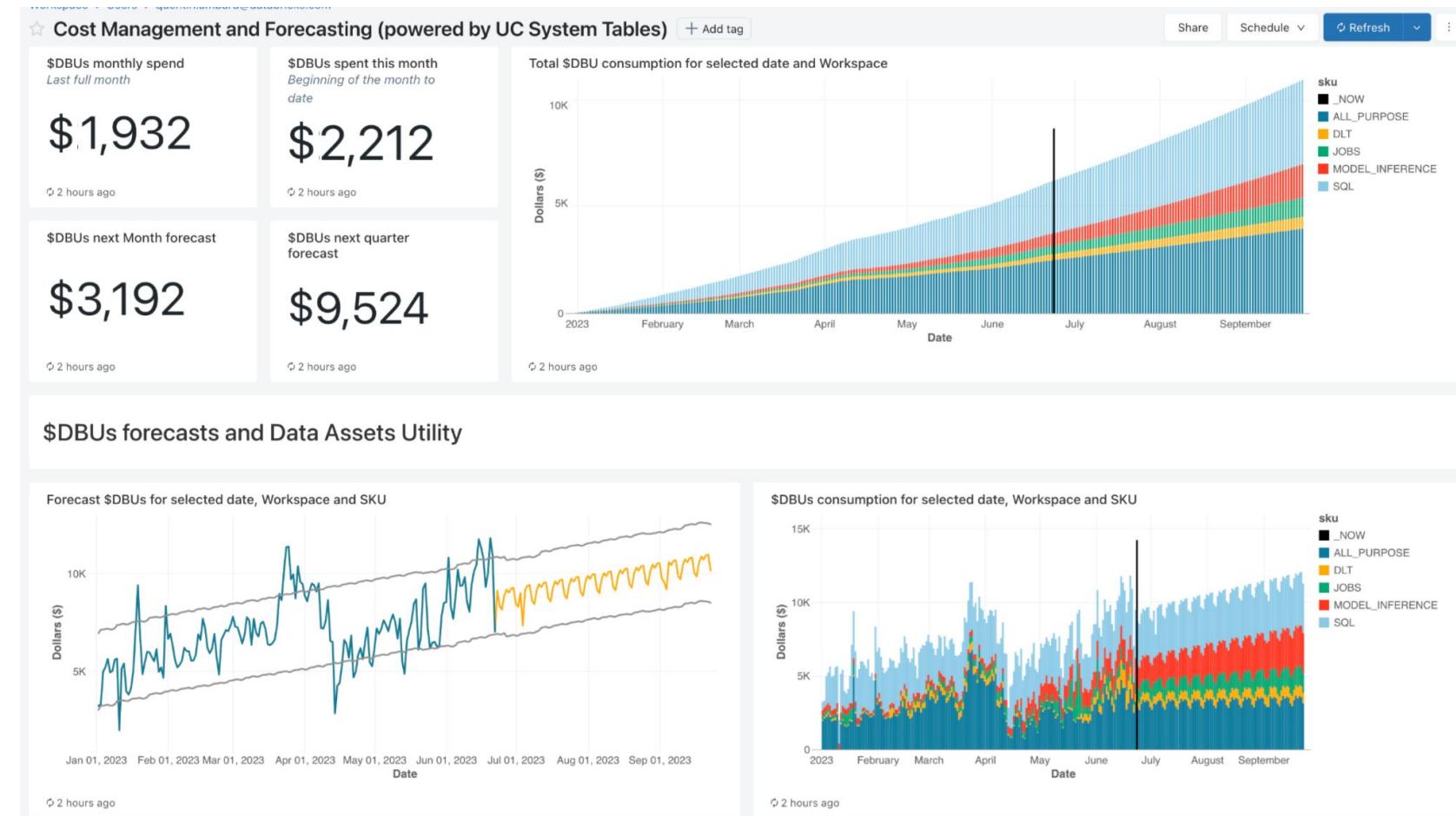
Eficiência
de Custos



Lakehouse

Eficiência de Custos

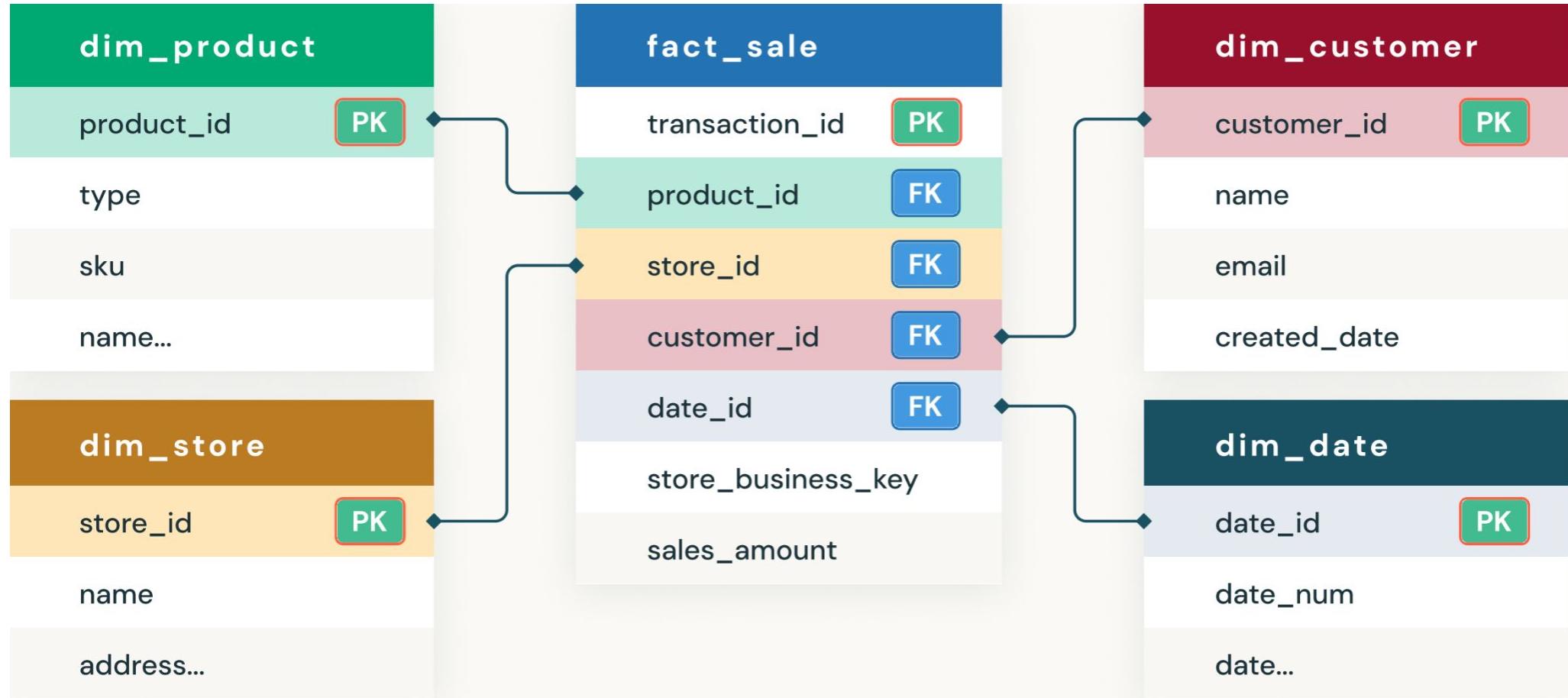
- Auto Scale
- Auto Terminate
- Clusters TAGs
- Regras de Acesso
- FinOps
- System Tables
- Alertas de Uso
- Controle de Budget



Modelagem de Dados



Modelagem Dimensional – STAR SCHEMA



Modelagem Dimensional – STAR SCHEMA

fact_sale	
transaction_id	PK
product_id	FK
store_id	FK
customer_id	FK
date_id	FK
store_business_key	
sales_amount	

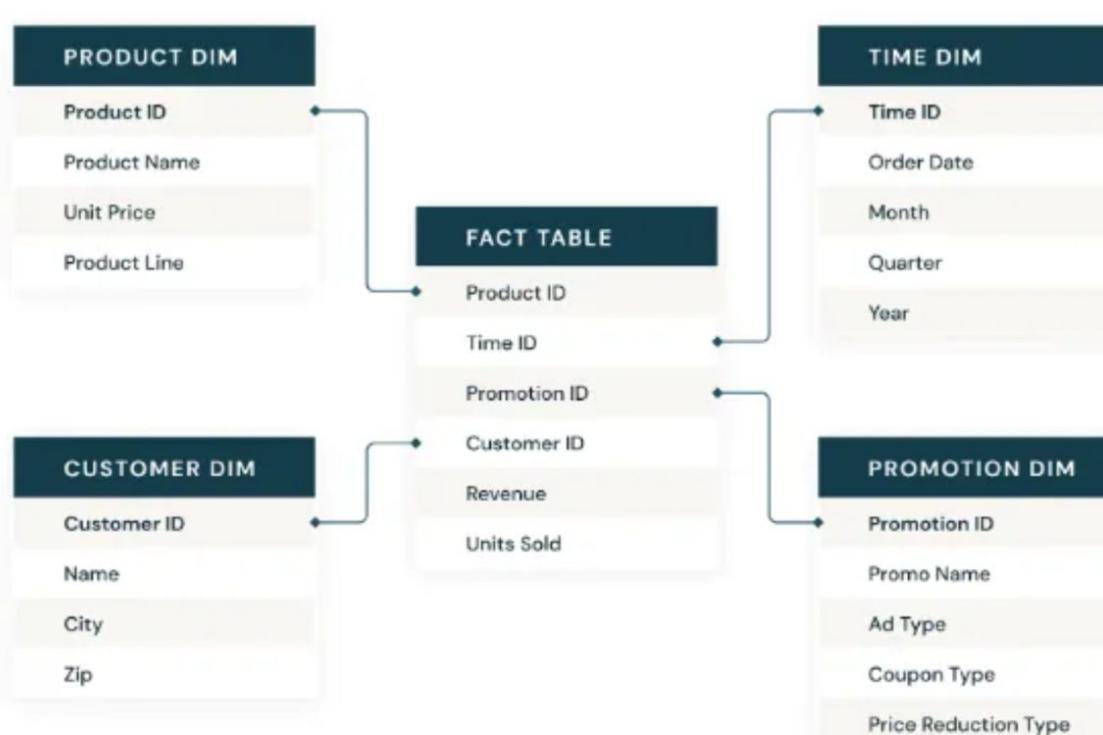
```
CREATE OR REPLACE TABLE fact_sales (
    transaction_id BIGINT PRIMARY KEY,
    product_id BIGINT NOT NULL CONSTRAINT dim_product_fk
        FOREIGN KEY REFERENCES dim_product,
    store_id BIGINT NOT NULL CONSTRAINT dim_store_fk
        FOREIGN KEY REFERENCES dim_store,
    customer_id BIGINT NOT NULL CONSTRAINT dim_customer_fk
        FOREIGN KEY REFERENCES dim_customer,
    date_id BIGINT NOT NULL CONSTRAINT dim_date_fk
        FOREIGN KEY REFERENCES dim_date,
    store_business_key STRING,
    sales_amount DOUBLE
);
```



OBT

Star schema

One big table



"Tabelão desnormalizado" →

ONE BIG TABLE
Product ID
Product Name
Unit Price
Product Line
Time ID
Order Date
Month
Quarter
Year
Promotion ID
Promo Name
Ad Type
Coupon Type
Price Reduction Type
Customer ID
Name
City
Zip



Lakehouse

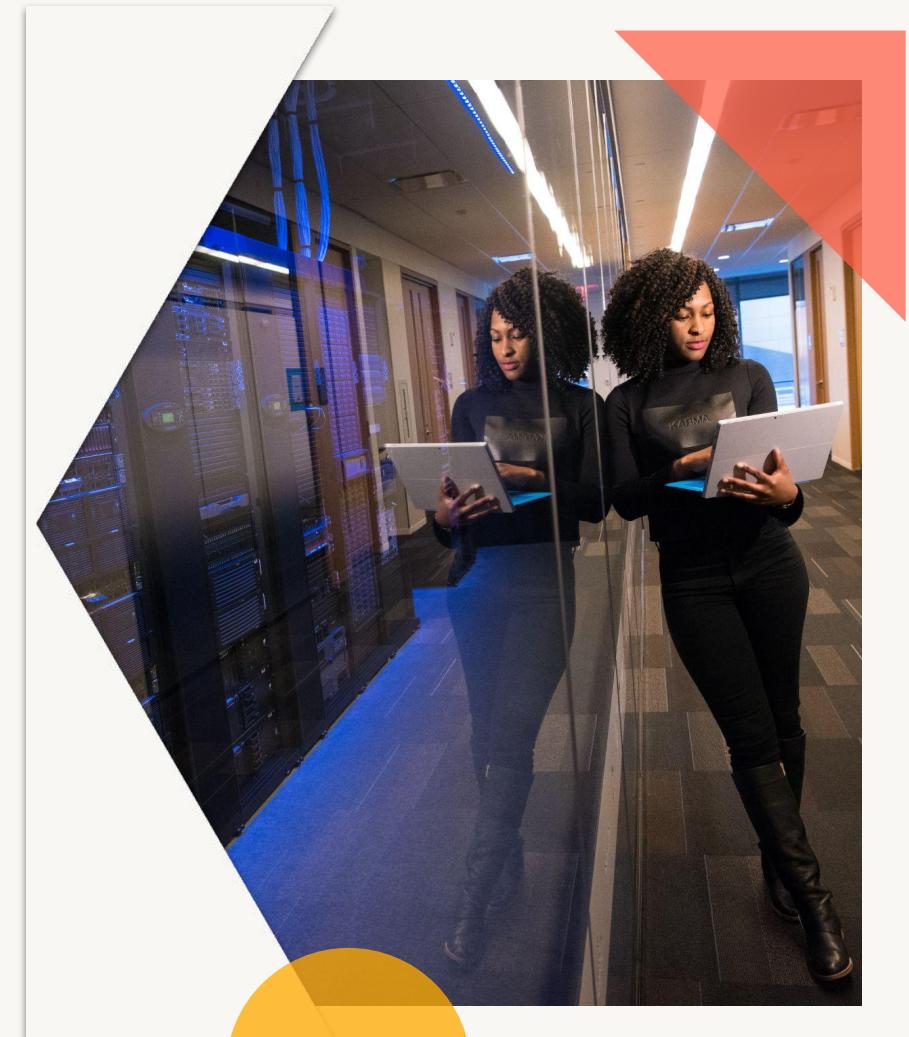
Data Quality

Data Quality

Garantindo a Qualidade dos Dados

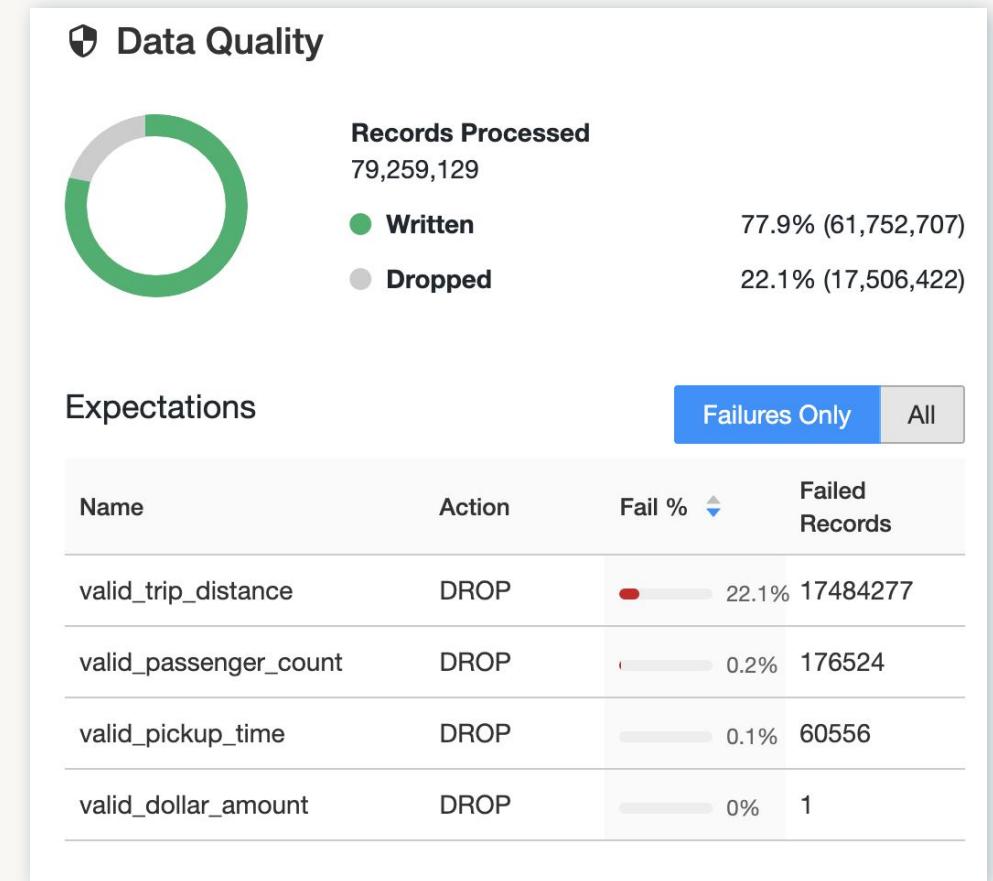


Como automatizar testes de
qualidade de dados?



Validação e monitoramento de **qualidade de dados**

```
CREATE STREAMING LIVE TABLE fire_account_bronze AS
(
    CONSTRAINT valid_account_open_dt
    EXPECT (account_open_dt is not null
            and (account_close_dt > account_open_dt))
    ON VIOLATION DROP ROW
    ...
)
```



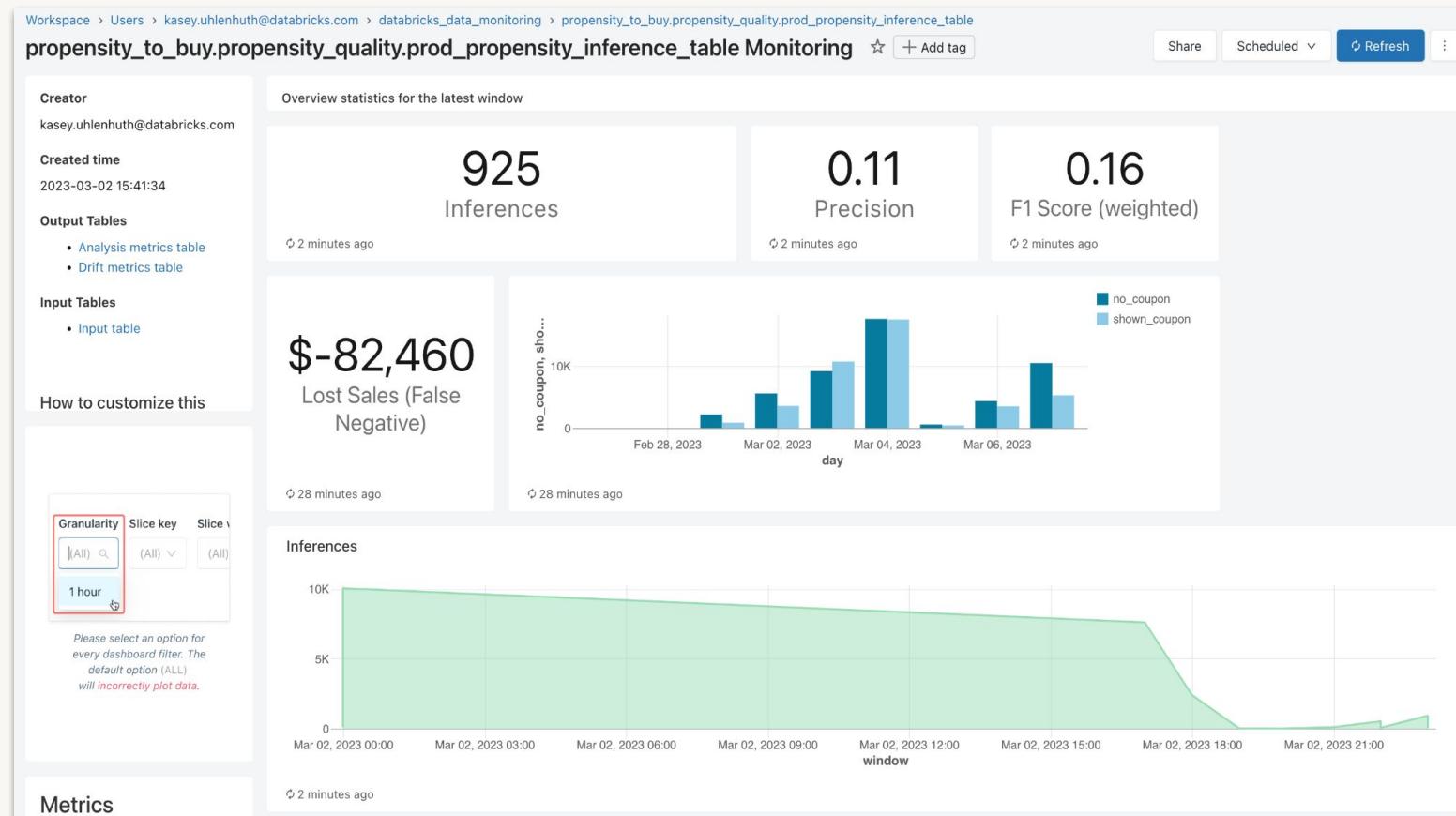
Framework: **DELTA LIVE TABLES**



Lakehouse Monitoring

Insights automatizados e métricas prontas para uso em pipelines de dados e ML

- Totalmente gerenciado
- Fácil configuração
- Métricas prontas
- Solução unificada





Performance



Performance

Duração da coluna (segundos, menor é melhor)



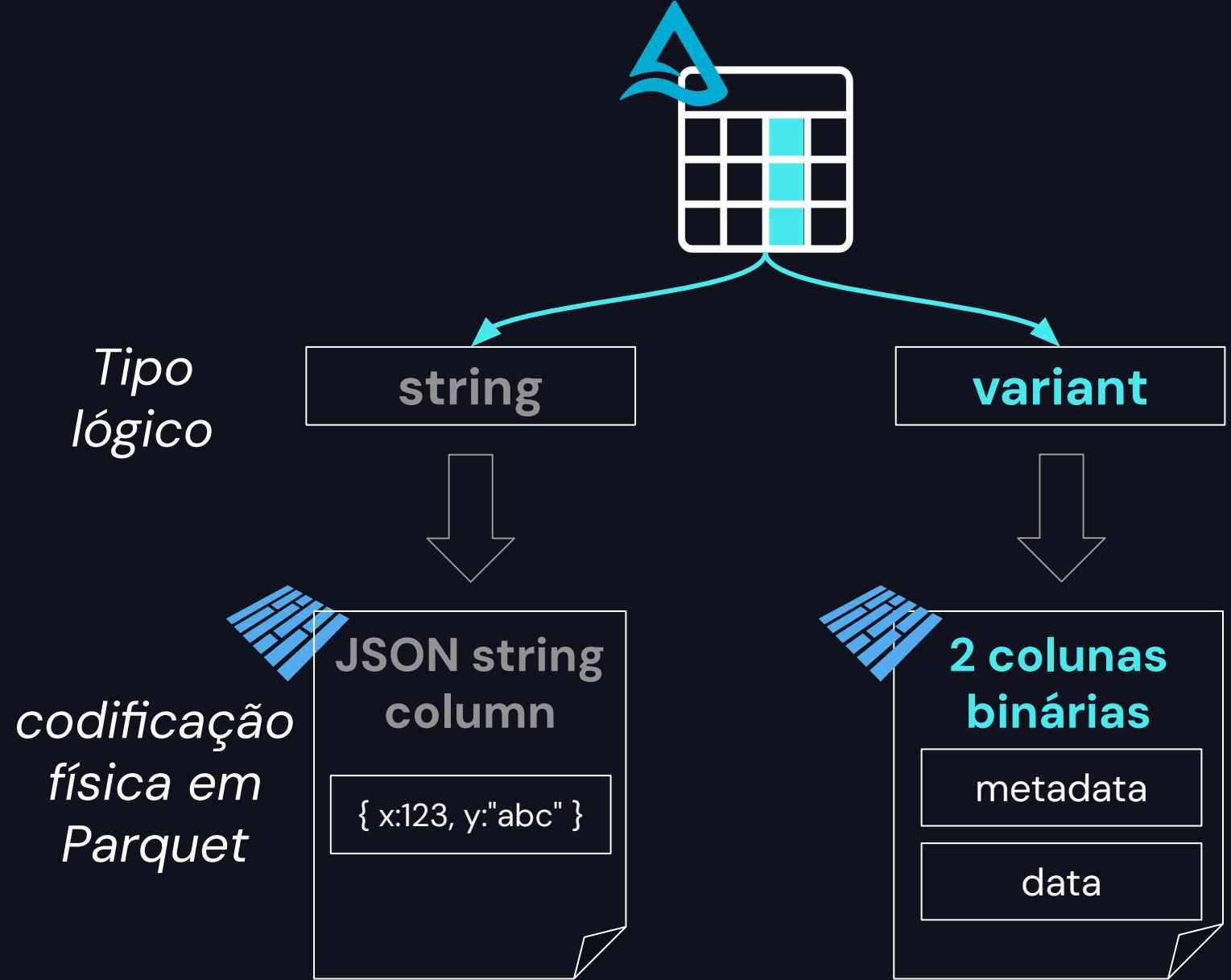
Consultas de BI simultâneas



Variant Type

Novo tipo de dado para armazenar dados semi-estruturados em uma coluna.

Ideia principal: Em vez de salvar o json como **string**, armazenar com código binário para facilitar a busca de informações.



Variant Type é fácil de utilizar

Funções em Spark para fazer o parsing dos dados

Sintaxe de Spark SQL para extrair campos da coluna variante

```
INSERT INTO variant_tbl (event_data)
VALUES (PARSE_JSON('{"level": "warning",
"message": "invalid request",
"user_agent": "Mozilla/5.0 ..."}'));
```



```
SELECT * FROM variant_tbl
WHERE variant_get(event_data, '$.user_agent', 'string')
like '%Mozilla%'
```



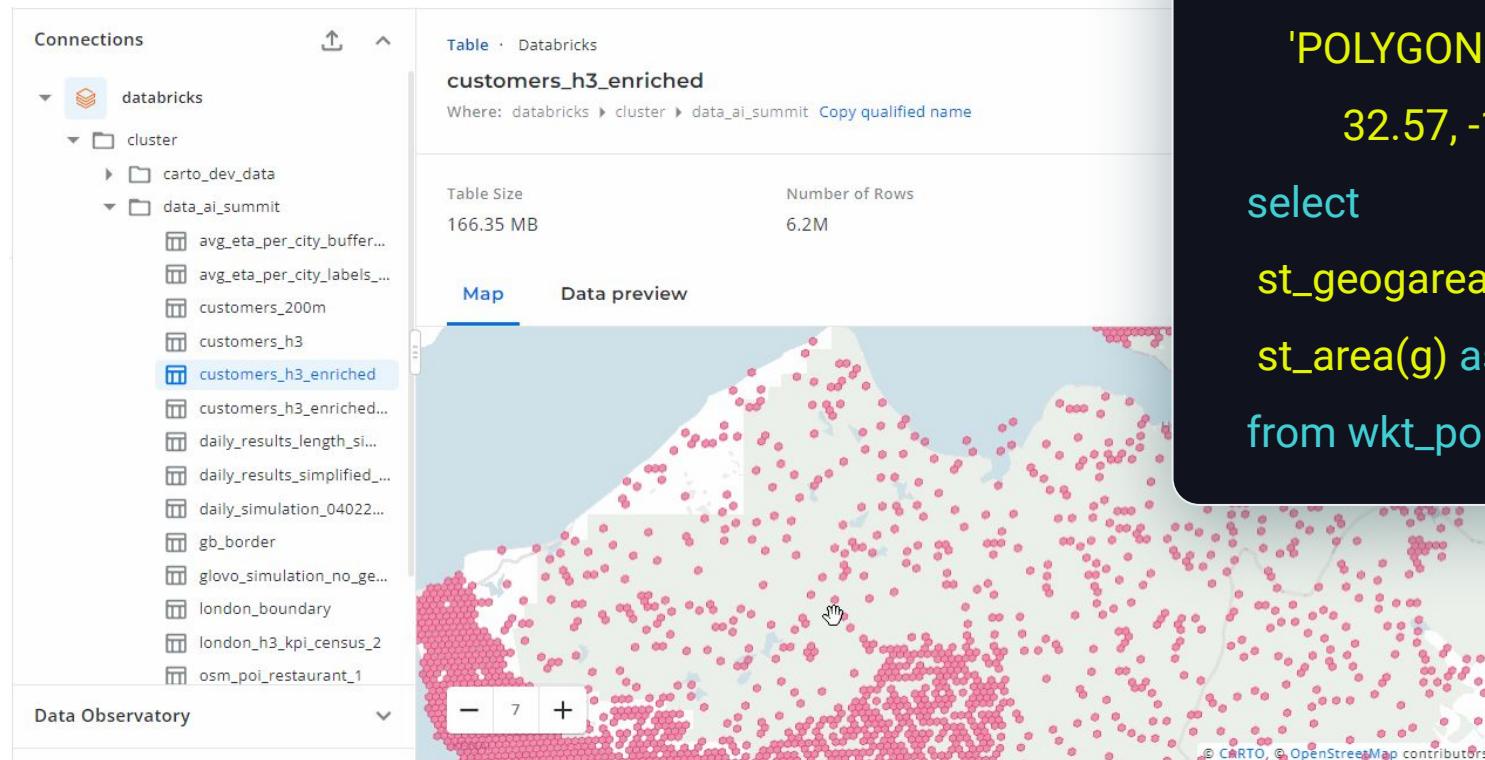
GEO Spatial SQL

Performance para análises Geoespaciais

60+ Funções Geo

Tipos Geométricos e Geográficos

H3- indexing



with wkt_poly as (select

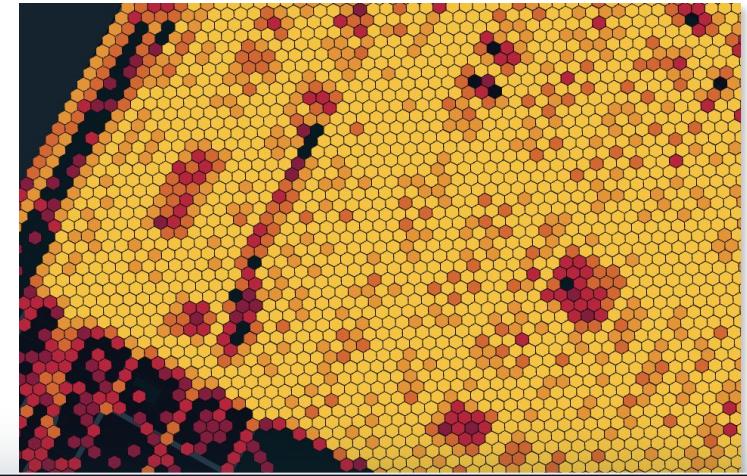
```
'POLYGON((-115.42 32.57, -115.42 32.57,-115.42  
32.57,-115.42 32.57, -115.42 32.57))' as g )
```

select

```
st_geogarea(g) as dbx_area_meters,
```

```
st_area(g) as dbx_area_units
```

from wkt_poly

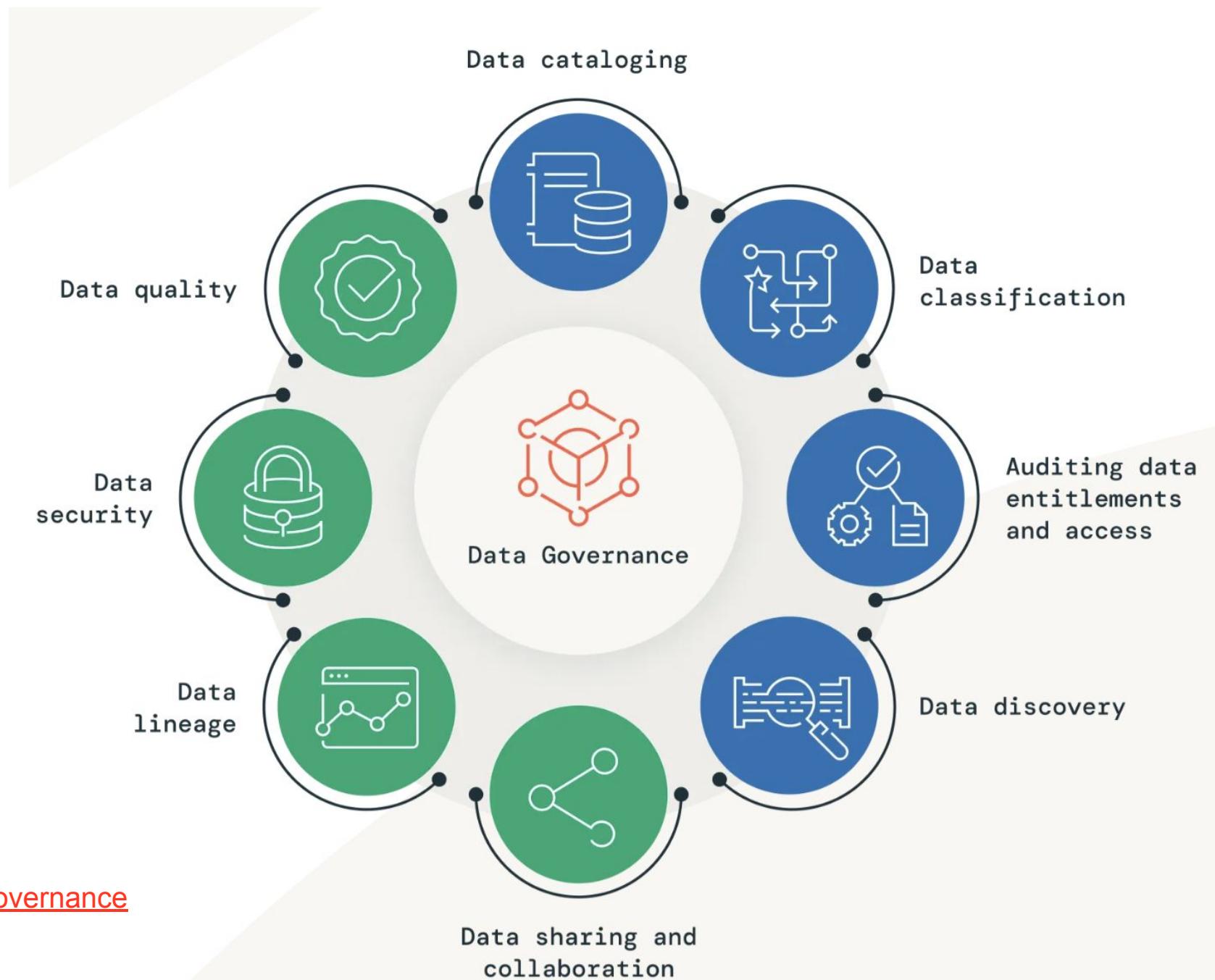
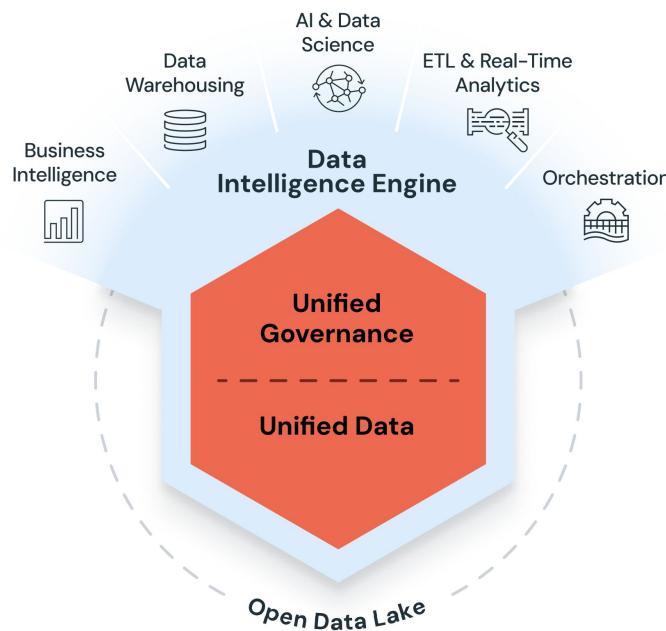


Lakehouse



Governança

Governança de Dados



www.databricks.com/discover/data-governance

+ New

Workspace

Recents

Catalog

Workflows

Compute

SQL

SQL Editor

Queries

Dashboards

Alerts

Query History

SQL Warehouses

Data Engineering

Job Runs

Data Ingestion

Delta Live Tables

Machine Learning

Playground

Experiments

Catalog Explorer

unity-catalog-demo



Send feedback

+ Add

Browse DBFS

Serverless Shared...

Serverless

S

Catalog

data_warehouse

- ▼ In my org
 - ▼ data_warehouse
 - comercial
 - dim_customer
 - dim_date
 - dim_product
 - dim_store
 - fact_sales
 - contabil
 - default
 - financeiro
 - fiscal
 - information_schema
 - marketing
 - rh

Delta Sharing

Catalogs > data_warehouse > comercial >

data_warehouse.comercial.fact_sales

Owner: luis.assuncao@databricks.com

Popularity: Size: 2.4KiB, Last 1 file Updated: horas

Tags: [Add tags](#)

Tabela Fato com os Dados de Venda, com indicadores (KPI) da área Comercial

Columns Sample Data Details Permissions History Lineage Insights Quality

Filter columns...

[View relationships](#) [AI generate](#)

Column	Type	Comment	Tags
transaction_id	bigint	+	+
date_id	bigint	+	+
customer_id	bigint	+	+
product_id	bigint	+	+
store_id	bigint	+	+
store_business_key	string	+	+
sales_amount	double	+	+

Controle de acesso centralizado

Conceda e gerencie centralmente as permissões de acesso

Usando comandos ANSI SQL DCL

```
GRANT <privilege> ON <securable_type>  
<securable_name> TO `<principal>`
```

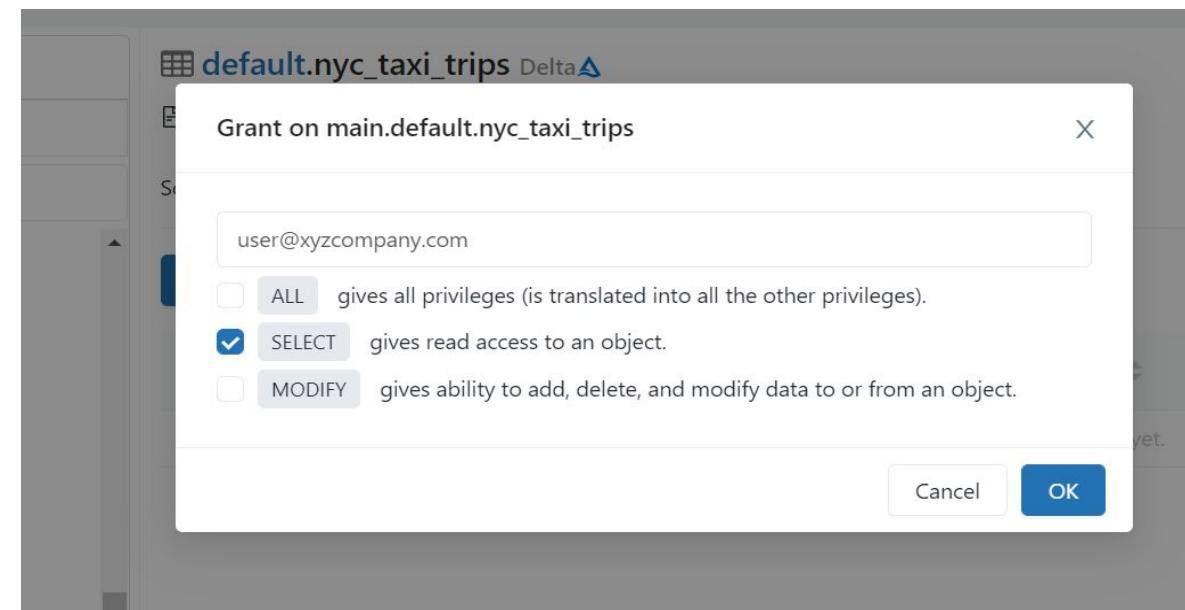
```
GRANT SELECT ON iot.events TO engineers
```

Nível de
Permissão
(privilégios)

'Table'= collection of
files in S3/ADLS

Nomes dos grupos
sincronizados com
seu provedor de
identidades (IAM)

Usando a Interface Visual (UI)



Data insights

Catalogs > catalog_A > schema_A >

catalog_A.schema_A.orders

Owner: name@domain.com Edit Popularity: 3.0 Comment: Add comment

Columns Sample data Details Permissions History Lineage Insights

Show only assets I have access to

Captured usage of this table over the last 30 days

Explanation on how the usage is captured

Frequent notebooks

- Weekly forecast user@databricks.com
- Q3 forecast user@databricks.com
- Throughput estimations user@databricks.com
- Purchasing procedures user@databricks.com
- Operational expenses user@databricks.com

Frequent dashboards

- Safety Stock Planning user@databricks.com
- Weekly forecast user@databricks.com
- Inventory planning user@databricks.com
- Supply/demand planning user@databricks.com
- Customer analysis user@databricks.com

Frequent joined tables

- customers catalog.schema
- loyalty_segments catalog.schema
- orders catalog.schema
- promo_prices catalog.schema
- purchase_orders catalog.schema

Frequent users

- user1@databricks.com Queried this table 38 times
- user2@databricks.com Queried this table 38 times
- user3@databricks.com Queried this table 38 times
- user4@databricks.com Queried this table 38 times
- user5@databricks.com Queried this table 38 times

Frequent queries

Query	Code Preview	Add	Copy
Inventory past 24 hours user@databricks.com	SELECT COUNT_IF('profile' IS NOT NULL)/COUNT(*)*100 'profile', COUNT_IF('profile' IS NOT ...)		
Sales past 24 hours user@databricks.com	SELECT COUNT_IF('profile' IS NOT NULL)/COUNT(*)*100 'profile', COUNT_IF('profile' IS NOT ...)		
New query (20) user@databricks.com	SELECT COUNT_IF('profile' IS NOT NULL)/COUNT(*)*100 'profile', COUNT_IF('profile' IS NOT ...)		
Operational expenses last month user@databricks.com	SELECT COUNT_IF('profile' IS NOT NULL)/COUNT(*)*100 'profile', COUNT_IF('profile' IS NOT ...)		
New query (18)			

Updated 1 min ago... Refresh

Delta sharing

Storage credentials

External locations

Linhagem de Dados





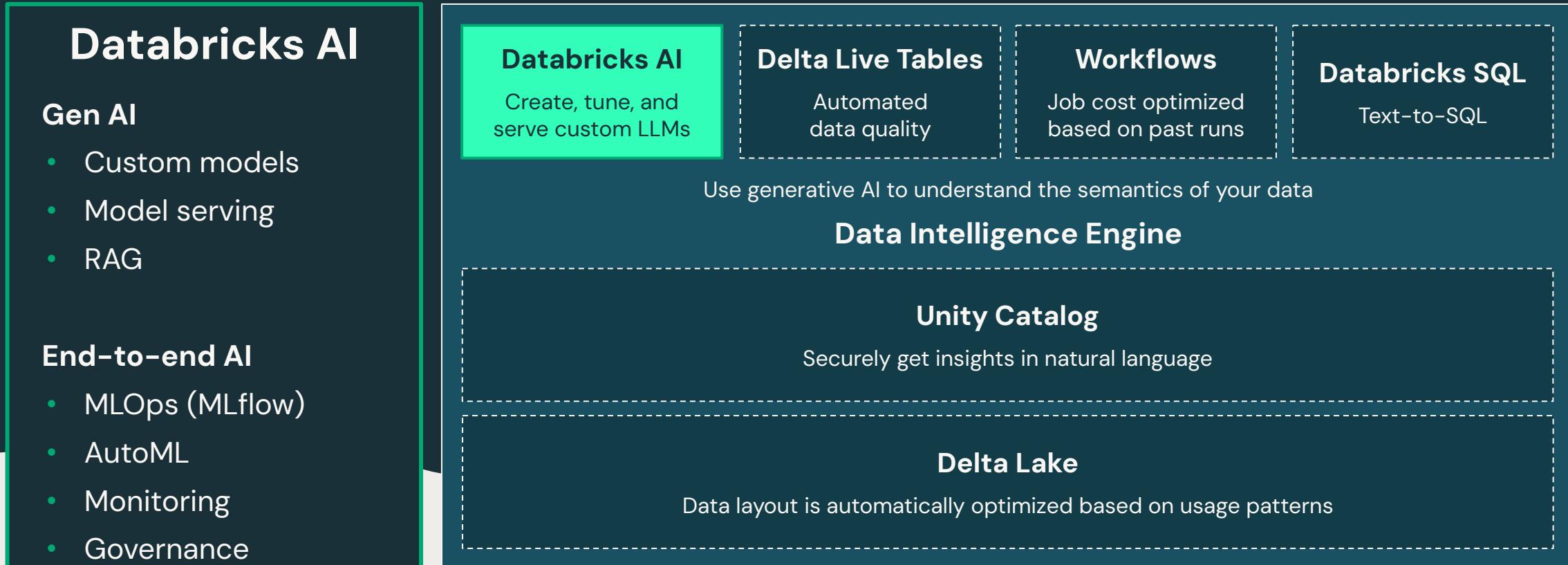
Lakehouse



Gen AI



Databricks Data Intelligence Platform



Open Data Lake

All Raw Data
(Logs, Texts, Audio, Video, Images)

Databricks Assistant

IA com entendimento semântico dos seus dados— nativo nos Notebooks, SQL editor e file editor

The screenshot shows the Databricks SQL Editor interface. On the left, the sidebar includes options like New, Workspace, Recents, Data, Workflows, Compute, SQL, SQL Editor (which is selected), Queries, Dashboards, Alerts, Query History, SQL Warehouses, Data Engineering, Delta Live Tables, Machine Learning, Experiments, Features, Models, Serving, Marketplace, and Partner Connect. The URL in the browser bar is <https://e2-dogfood.staging.cloud.databricks.com/marketplace?o=6051921418418893>.

The main area is titled "Assistant". It contains a section about the Assistant's capabilities: "The Assistant can accelerate your work by diagnosing errors, suggesting code or queries, and answering questions. Mistakes are possible, so be sure to check facts and share your feedback. (More info)". Below this is a search bar with the placeholder "Start with picking the tables to query?" and a dropdown menu showing "Top10", "New query", and "New query". The dropdown menu also includes a "Save" button.

In the query editor, the text "1 | SELECT or Option+Space(+Shift for AI) or browse" is displayed. The results pane below is currently empty, showing a large blue cube icon and the text "Run a query to show the results".

Intelligent Table Search

Pesquisa contextual e precisa + sugestão

The screenshot shows the Databricks interface with a search overlay. The search bar contains the query "What table should I use for revenue?". A suggestion card appears, stating "The **revenue_data** table is a good option." Below the card, there is a table listing three tables: "revenue Inventory", "Streaming Sales", and "Repairs". Each table entry includes a preview icon, the name, the schema, the last modified date, and the author's email.

Table	Schema	Last modified	Author
revenue Inventory	production/accounts/	Aug 25, 2022	user@domain.com
Streaming Sales	production/accounts/	Aug 26, 2022	matthew.macwan@nba.com
Repairs	production/accounts/	Aug 26, 2022	joshua.rodriguez@nba.com

Aceleradores – Go To Market

Data LAKEHOUSE



Data LAKEHOUSE + Genie Spaces



Genie



Ask your question...





Genie Data Room

databricks

Quick search... 88K

A

Análise e Exploração DW (1)

EDA DW

Edit Share

Chats

Tables

Instructions

Back to list

bronze_dolar

dolar_ano

dolar_mes

Aa dolar_dia

dolar_fechamento

dolar_abertura

dolar_max

dolar_min

silver_empresas

bronze_estabelecimentos

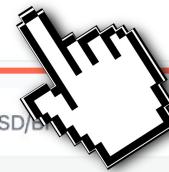
LA Qual a média do valor do dólar mensal ?

This query calculates the average closing value of the USD/BRL exchange rate for each month. It groups the data by year and month, then averages the closing values.

dolar_ano	dolar_mes	media_fechamento
2022	1	5,53
2022	2	5,19
2022	3	4,97
2022	4	4,74
2022	5	4,95
2022	6	5,06
2022	7	5,37
2022	8	5,14
2022	9	5,24
2022	10	5,25
2022	11	5,28
18 rows		

Hide generated code

```
SELECT
    dolar_ano,
    dolar_mes,
    AVG(dolar_fechamento) as media_fechamento
FROM
    openfinance.internal_data.bronze_dolar
GROUP BY
    dolar_ano,
    dolar_mes
ORDER BY
    dolar_ano,
    dolar_mes
```





Genie Data Room

databricks

bronze_dolar

silver_empresas

bronze_estabelecimentos

A

Análise e Exploração DW (1)
EDA DW

Edit Share

Chats

Tables

Instructions

Quick search... 36K

Qual a média do valor do dólar mensal ?

This query calculates the average closing value of the USD/BND exchange rate for each month. It groups the data by year and month, then averages the closing rate within each group. The results are ordered by year and month.

dolar_ano	dolar_mes	media_fechamento
2.022	1	5,53
2.022	2	5,19
2.022	3	4,97
2.022	4	4,74
2.022	5	4,95
2.022	6	5,06
2.022	7	5,37
2.022	8	5,14
2.022	9	5,24
2.022	10	5,25
2.022	11	5,29

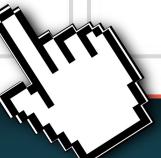
Show generated code

Quick actions

Auto visualize Bar chart Line chart Pie chart

Quick actions

Auto visualize Bar chart Line chart Pie chart



Quick search... 38K

A

Análise e Exploração DW (1)

EDA DW

[Edit](#) [Share](#)[Chats](#)[Tables](#)[Instructions](#)[Back to list](#)

bronze_dolar

silver_empresas

bronze_estabelecimentos

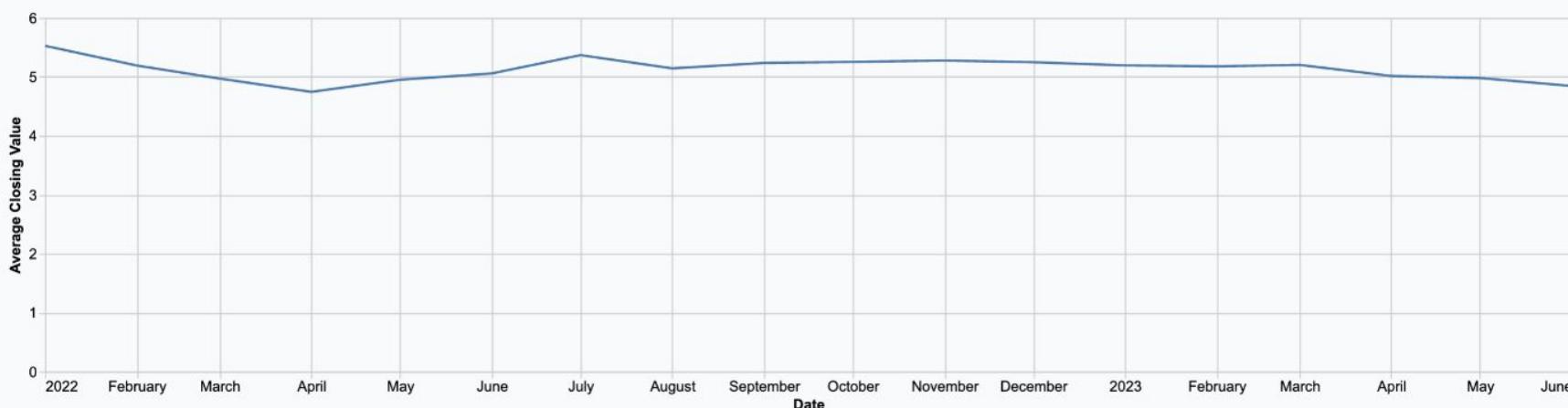
LA Qual a média do valor do dólar mensal ?

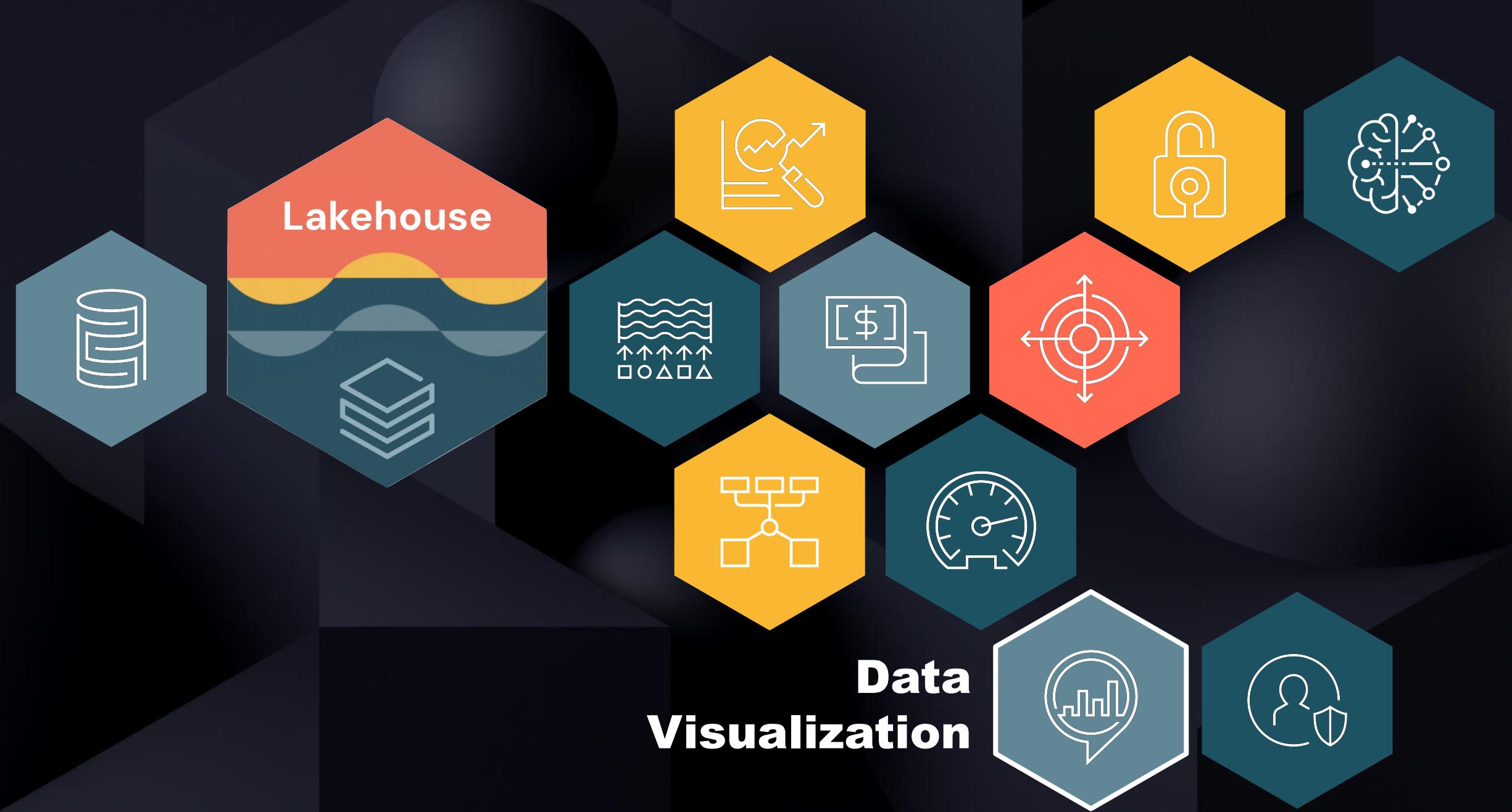
This query calculates the average closing value of the USD/BND exchange rate for each month. It groups the data by year and month, then averages the closing rate within each group. The results are ordered by year and month.

dolar_ano	dolar_mes	media_fechamento
2.022	1	5,53
2.022	2	5,19
2.022	3	4,97
2.022	4	4,74
2.022	5	4,95
2.022	6	5,06
2.022	7	5,37
2.022	8	5,14
2.022	9	5,24
2.022	10	5,25
2.022	11	5,28
18 rows		

[Show generated code](#)

LA Visualize

[Show generated code](#)



Databricks AI/BI

Intelligent analytics for real-world data



table

nento is ready!

based on this dataset using
the chart you'd like to see.
in the right panel.

crie um gráfico de



faturamento

ano

mes

regiao

faturamento

pis

cofins

icms

notas_entradas

notas_saida

to	pis	cofins			notas_saida
5578	16546	76367	458204	14301	1000
6574	3683	16997	124646	3183	354
1128	4947	22834	137003	4276	504
5220	1724	7957	45087	1490	469
0384	16578	76512	459069	14328	890
6666	6666	61752	666615	5616	666

Submit

table

nento is ready!

based on this dataset using
the chart you'd like to see.
in the right panel.

crie um gráfico de barras do faturamento mensal



to	pis	cofins	icms	notas_entradas	notas_saida
5578	16546	76367	458204	14301	1000
6574	3683	16997	124646	3183	354
1128	4947	22834	137003	4276	504
5220	1724	7957	45087	1490	469
0384	16578	76512	459069	14328	890
6666	6666	6666	6666	6666	666

table

nento is ready!

based on this dataset using
the chart you'd like to see.
in the right panel.

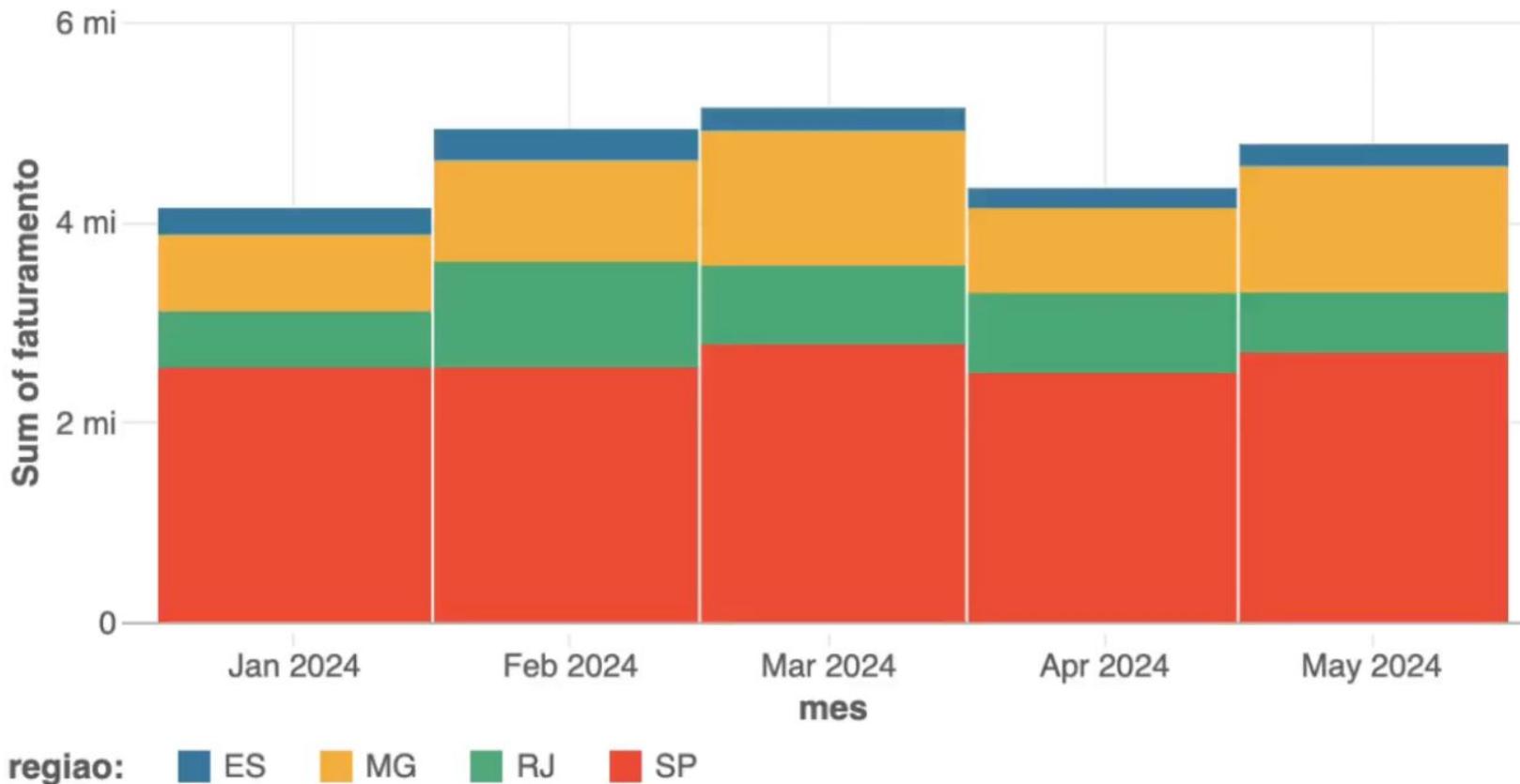
crie um gráfico de barras do faturamento mensal

Accept

Reject

C

Monthly Revenue



to

pis

cofins

icms

notas_entradas

notas_saida

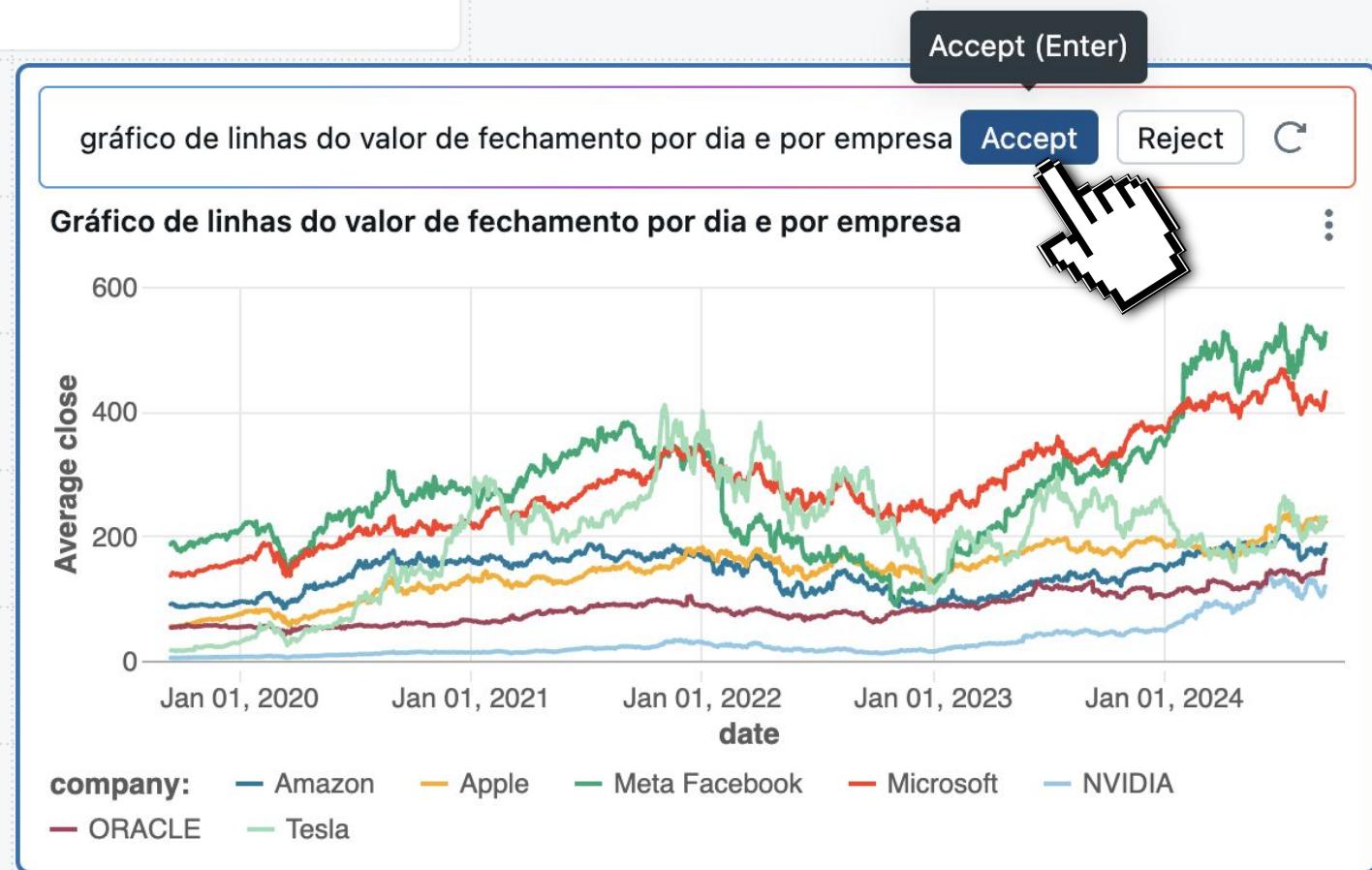


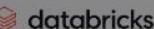
Dashboard for stock_bigtech table

The dashboard you created for luis_assuncao.genie_aibi.stock_bigtech is ready!

The widget to the right of this one allows you to generate a chart based on this dataset using natural language. Click the widget to start typing a description of the chart you'd like to see. Alternatively, you can edit any selected widget using the controls in the right panel.

You can also edit, copy, or delete any widget on the canvas.



 **databricks**

Search data, notebooks, recents, and more... + P

e2-field-eng-west 

+ New Send feedback databricks_stock_market_big_tech Published C 40s ago Schedule Share ⋮

Workspace Recents Catalog Workflows Compute

SQL

SQL Editor Queries Dashboards Genie Alerts Query History SQL Warehouses

Data Engineering

Job Runs Data Ingestion Pipelines

Machine Learning

Playground Experiments Features Models Serving

Marketplace Partner Connect

company

Stock Market Dashboard

Gráfico de linhas do valor de fechamento por dia e por empresa

Average close

Oct 01, 2019 Apr 01, 2024

company

- Amazon
- Apple
- Meta Facebook
- Microsoft
- NVIDIA
- ORACLE
- Tesla

Volume de ações por dia e por empresa

Sum of volume

Oct 01, 2019 Apr 01, 2024

date

company

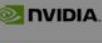
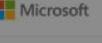
- Amazon
- Apple
- Meta Facebook
- Microsoft
- NVIDIA
- ORACLE
- Tesla

Copy embed code

You can copy the code below and paste it into your website. Note that users will still need to log in, and that your dashboard needs to be published in order for them to view it.

```
<iframe src="https://e2-demo-field-eng.cloud.databricks.com/embed/dashboardsv3/01ef756abde9176387c7a1e6e7a8245d?o=1444828305810485" width="100%" height="600" frameborder="0"></iframe>
```

Published dashboards are **allowed** to be embedded in any domain.

logo	company	stock	min_close	max_close	percentual_variacao
	Apple	AAPL	54.42	234.82	331.50
	NVIDIA	NVDA	4.29	135.58	3060.37
	Microsoft	MSFT	134.65	467.56	247.24

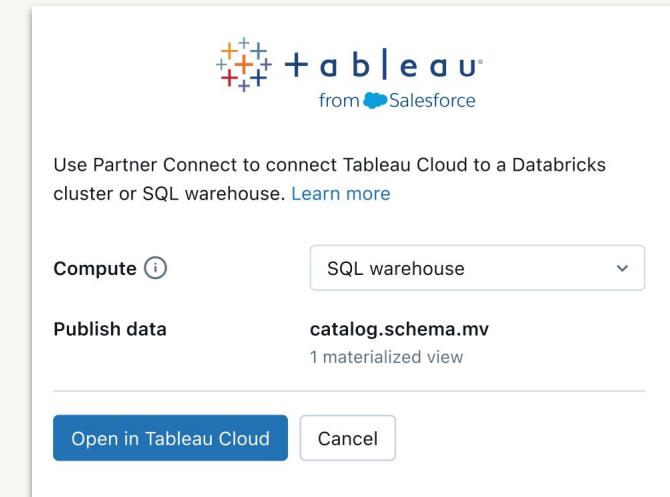
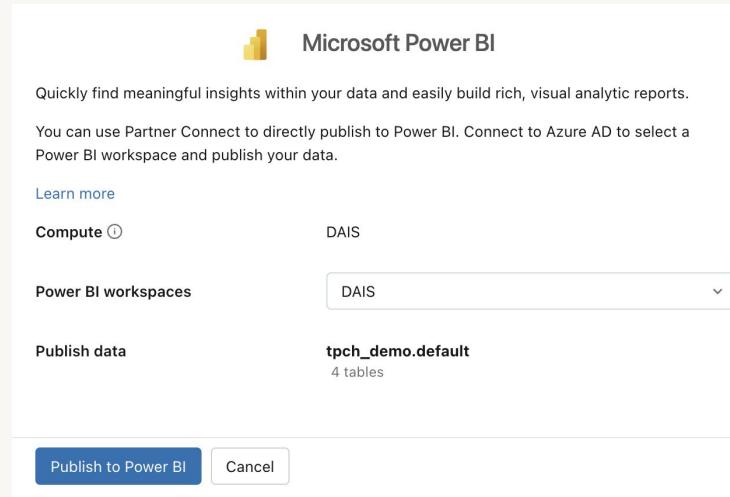
Integração com Power BI & Tableau

Integração perfeita de catálogo e sincronização de modelo de dados

Integração Power BI

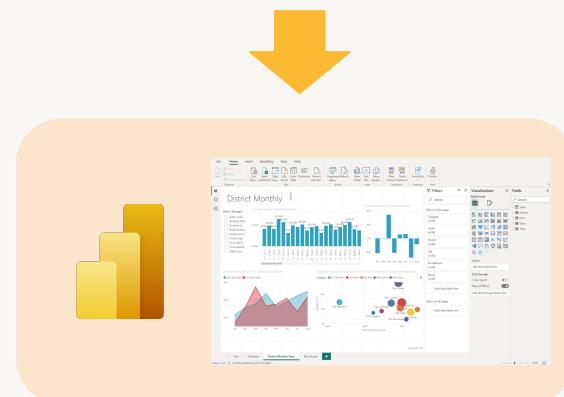
Publique conjuntos de dados do UC na interface do usuário do Databricks direto Power BI Online.

Sincronize esquemas inteiros, incluindo relacionamentos de tabelas (PK/FK), para economizar tempo.



Integração Tableau

Explore facilmente os conjuntos de dados do Unity Catalog no Tableau Online com um único clique no Data Explorer.



Lakehouse



Proteção
dos dados

Proteção de Dados

Lei de Proteção de Dados sensíveis

Anonimização:

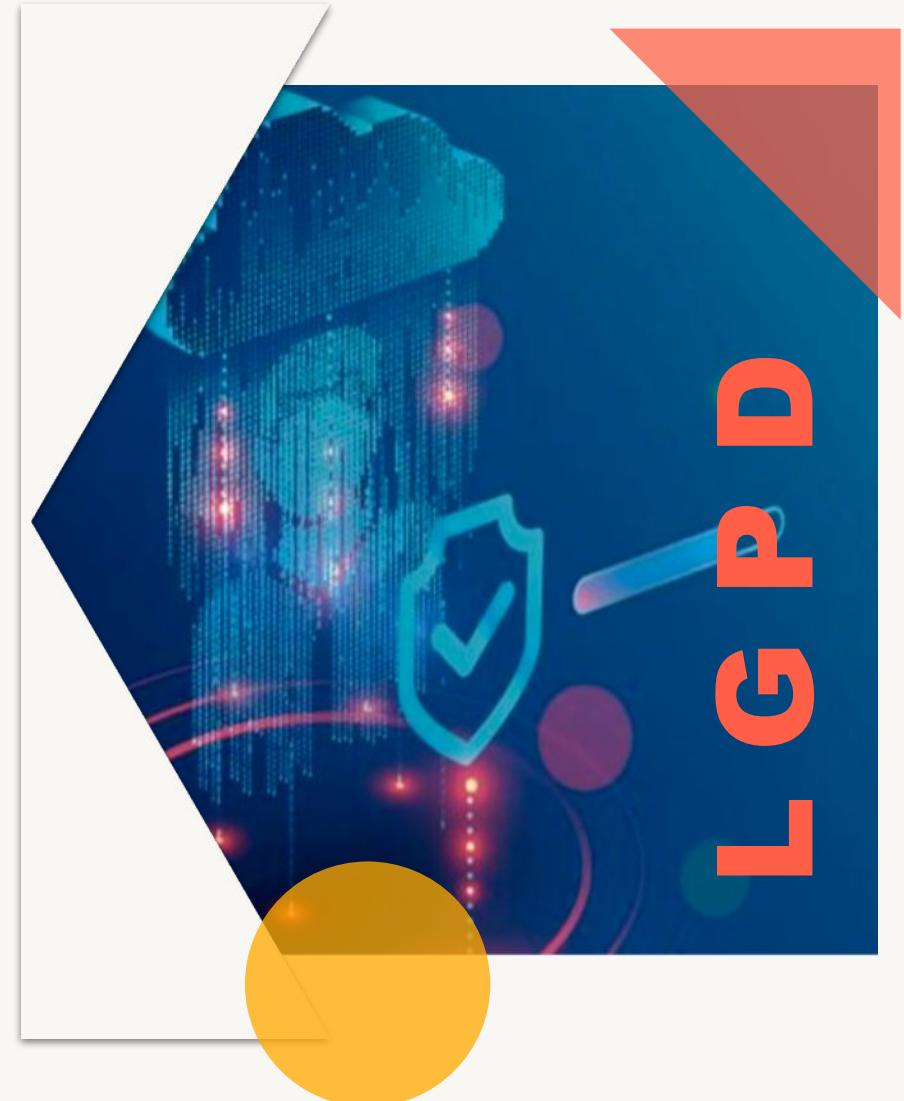
Mascaramento e/ou Criptografia

Controle de Acesso: Governança

Rastreabilidade: Auditoria - LOGs

RLS – Row Level Security

CLS – Column Level Security



Exemplo de máscara de coluna:

```
CREATE FUNCTION cpf_mask(cpf STRING)
    RETURN
        CASE WHEN is_member('Recursos_Humanos')
            THEN cpf
            ELSE '*****' END;
```

```
CREATE TABLE funcionarios (
    nome_funcionario STRING,
    cpf STRING MASK cpf_mask);
```

<https://learn.microsoft.com/pt-br/azure/databricks/data-governance/unity-catalog/row-and-column-filters>



Lakehouse

Estratégias de Migração

E como realizar a Migração ?

Será que vale a pena migrar os pipelines como estão (AS-IS)?

Estratégia LIFT & SHIFT ?

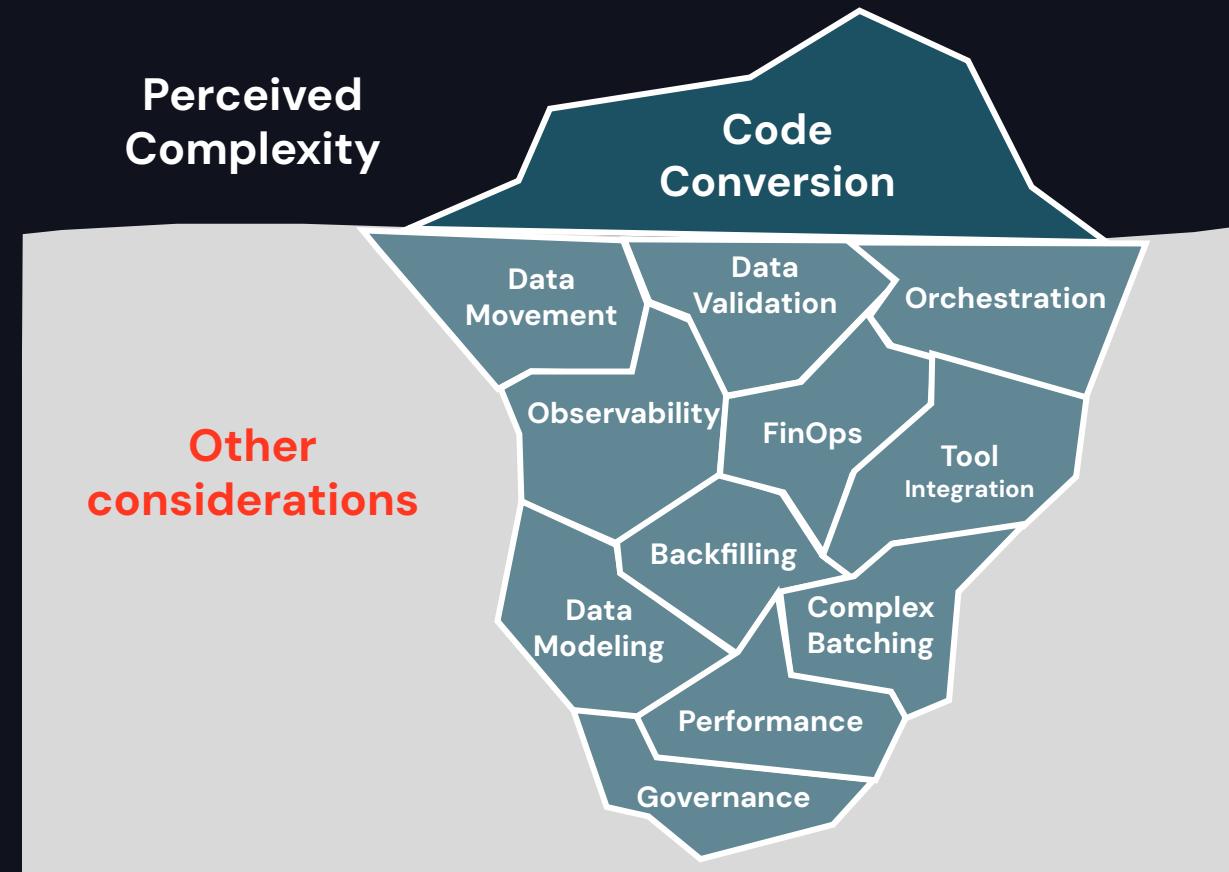
OU refatorar o código seguindo
as boas práticas e
buscando melhoria de performance?



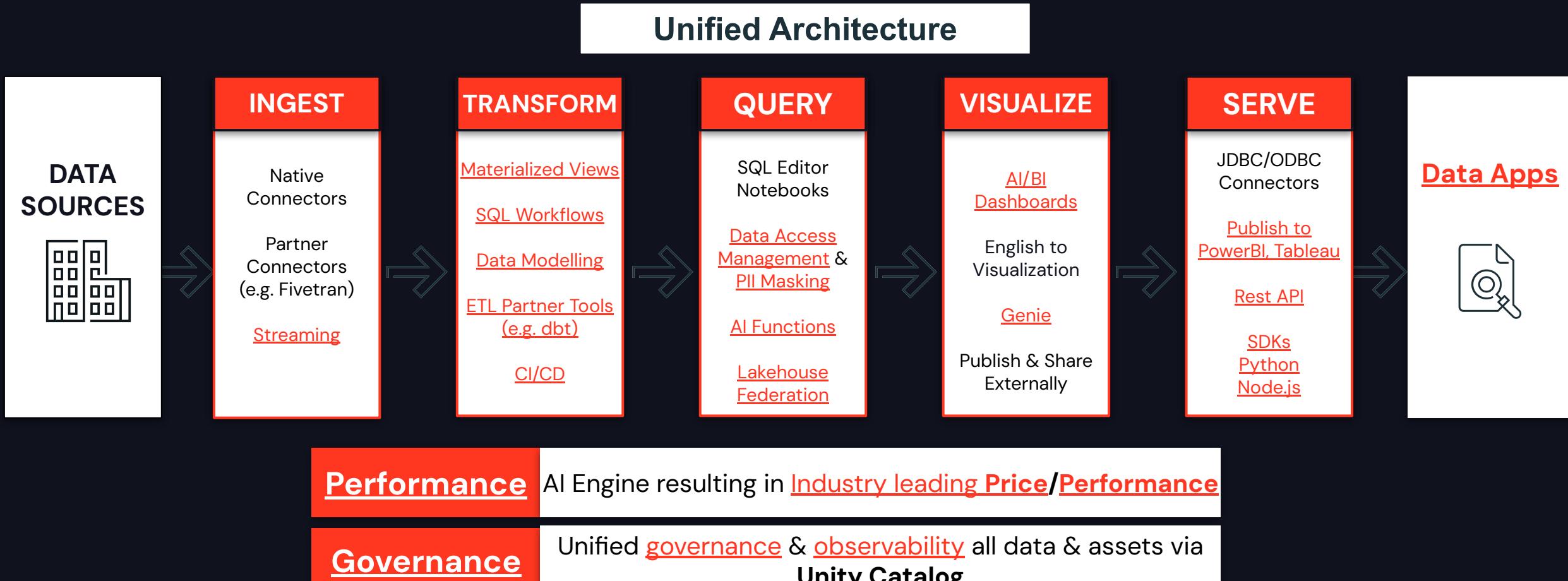
Componentes de uma migração

Não é somente conversão de código

- Movimentação de dados / Ingestão
- Governança
- Modelagem de Dados
- SQL/Código/DDL
- Integração com Ferramentas
- FinOps



Solução Completa de DW (com blogs)



[DELTA UNIFORM](#)



UNITY CATALOG



DELTA SHARING



MARKETPLACE



Tradução de Código

Abordagens

- Open Source:
 - SQLGlot (MIT License) com suporte a Databricks SQL
 - Remorph ([Databricks Labs](#)) extende o SQLGlot com novos métodos
- Soluções licenciadas:
- Field Accelerators:
 - Com abordagens de Gen-AI



Announcing Brickbuilder
Solutions for Migrations



by [Michael Lumb](#)
August 11, 2022 in [Company Blog](#)



Validação de Dados

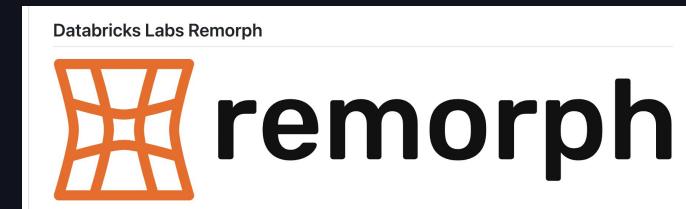
Verificando a consistência da informação

- **Open Source:**
 - **Datacompy (Apache)**
 - **Datafold data-diff (MIT):**
 - Também com versão paga e interface
 - **Remorph Reconcile (Databricks Labs):**
 - Comparativo In-database (hashing)
 - Integrado com AI/BI Dashboards
 - Thresholds e Aggregations
 - Plugin UDF para tipos complexos
- Soluções de SI

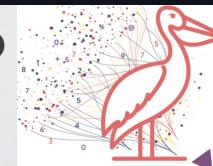
Capital One
https://www.capitalone.com › Tech › Blog › Open Source

DataComPy Open Source Comparison Project

ORG_ID	SUB_PLAN	CREATED_AT
	65%	79%
	647 rows different =	790 rows different =
Primary Key	Staging	Production
100000043	Team	Team
100000060	NULL	Individual
100000113	NULL	Individual
100000029	Team	Individual
100000110	Business	Business



Pelican: A Smarter Way to Validate Data
Powered by Onix & Datametica



Inscreva-se para os próximos webinars

Do ZERO ao GenAI

Databricks como seu
atalho para o futuro

13 DE NOVEMBRO

10:00 às 11:00



Mergulhe no Unity Catalog

Uma jornada prática para
governança de dados e IA

26 DE NOVEMBRO

10:00 às 11:00



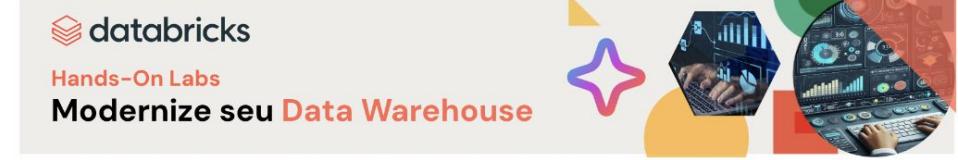
Delta Lake Uniform:

Unificando os formatos de
Tabelas Iceberg, Hudi e
Delta Lake

12 DE DEZEMBRO

10:00 às 11:00





Material disponível no repositório:



Databricks-BR

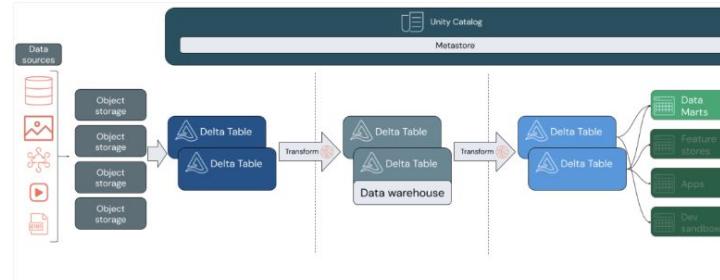


github.com/Databricks-BR/dw

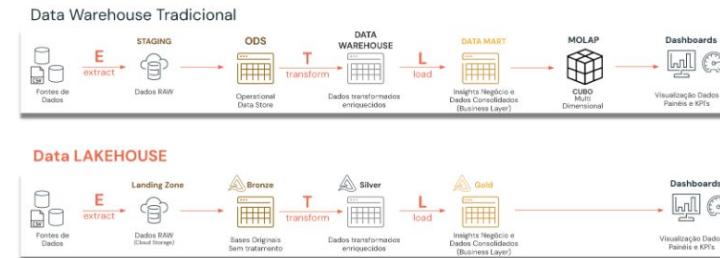
Databricks Labs - MODERNIZE SEU DATA WAREHOUSE

Treinamento na plataforma Databricks com foco nas funcionalidades em DATA WAREHOUSE.

A arquitetura LAKEHOUSE e o Databricks SQL trazem clouds data warehousing recursos para o seu data lake. Usando estruturas de dados, relações e ferramentas de gerenciamento familiares, você pode modelar um data warehouse de alto desempenho e econômico que pode ser executado diretamente em seu data lake.



Tal como acontece com um Data Warehouse tradicional, você modela os dados de acordo com os requisitos de negócios e depois os fornece aos usuários finais para análises e relatórios. Ao contrário de um data warehouse tradicional, é possível simplificar e unificar a Arquitetura de Dados, trazendo maior agilidade e colaboração na equipe de Dados (engenheiros, cientistas e analistas), com toda a Governança, Segurança e Rastreabilidade, garantindo a integração com o ecossistema de soluções de dados, incluindo as diversas ferramentas de Visualização de Dados.



Referências

- [O que é um Data Warehousing no Databricks](#)
- [Página principal do Databrick SQL](#)
- [SQL Language reference - Inglês](#)
- [SQL Language reference - Português](#)

Obrigado!