



CHALLENGE





Detrás de cada transacción de Bradescard, hay una persona con un sueño, un proyecto o una necesidad.

Cada pago, cada depósito, cada transferencia representa oportunidades que no podemos dejar pasar desapercibidas.

# Challenge

---



El reto consiste en garantizar monitoreo, validación y transformación de datos de manera automatizada, asegurando que cada transacción refleje información confiable y accionable desde su origen hasta los insights finales

# Nuestro manual

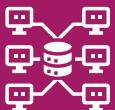
Introducción a Bradesco



SOFIPO y Marco Regulatorio en México



Metadata y Gobierno de Datos



Catálogo de Datos y Metadatos



Arquitectura Medallion



Propuesta, Timeline, Riesgos

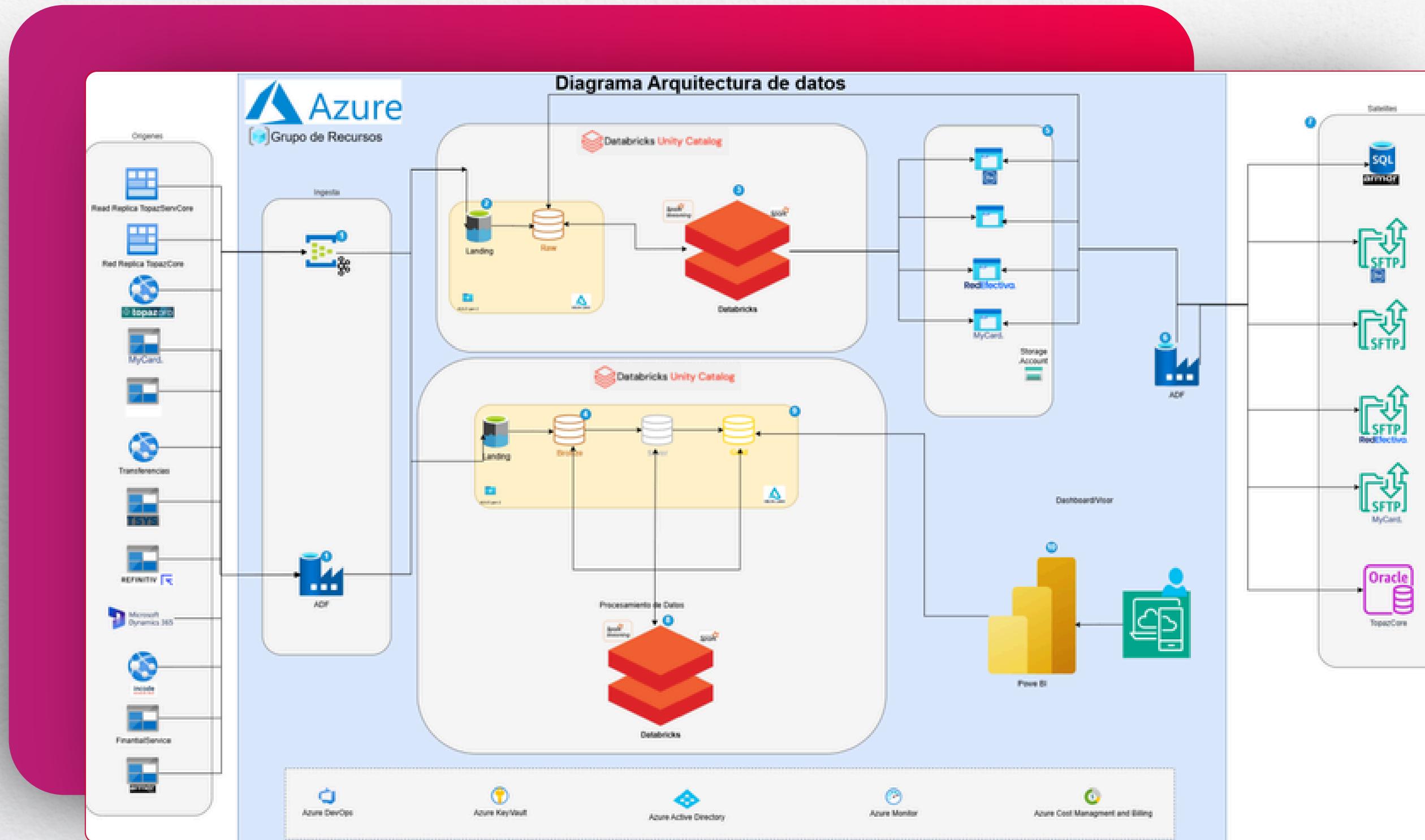


**Bradesco**  
Manual de  
Políticas y Prácticas

Maria Dávalos Castro | A04639016  
Lisbeth Vargas Corona | A01752595  
Andrea Ramos McC Gregor | A0445561  
Mariana Nieto Moreno | A04639071



# Arquitectura actual



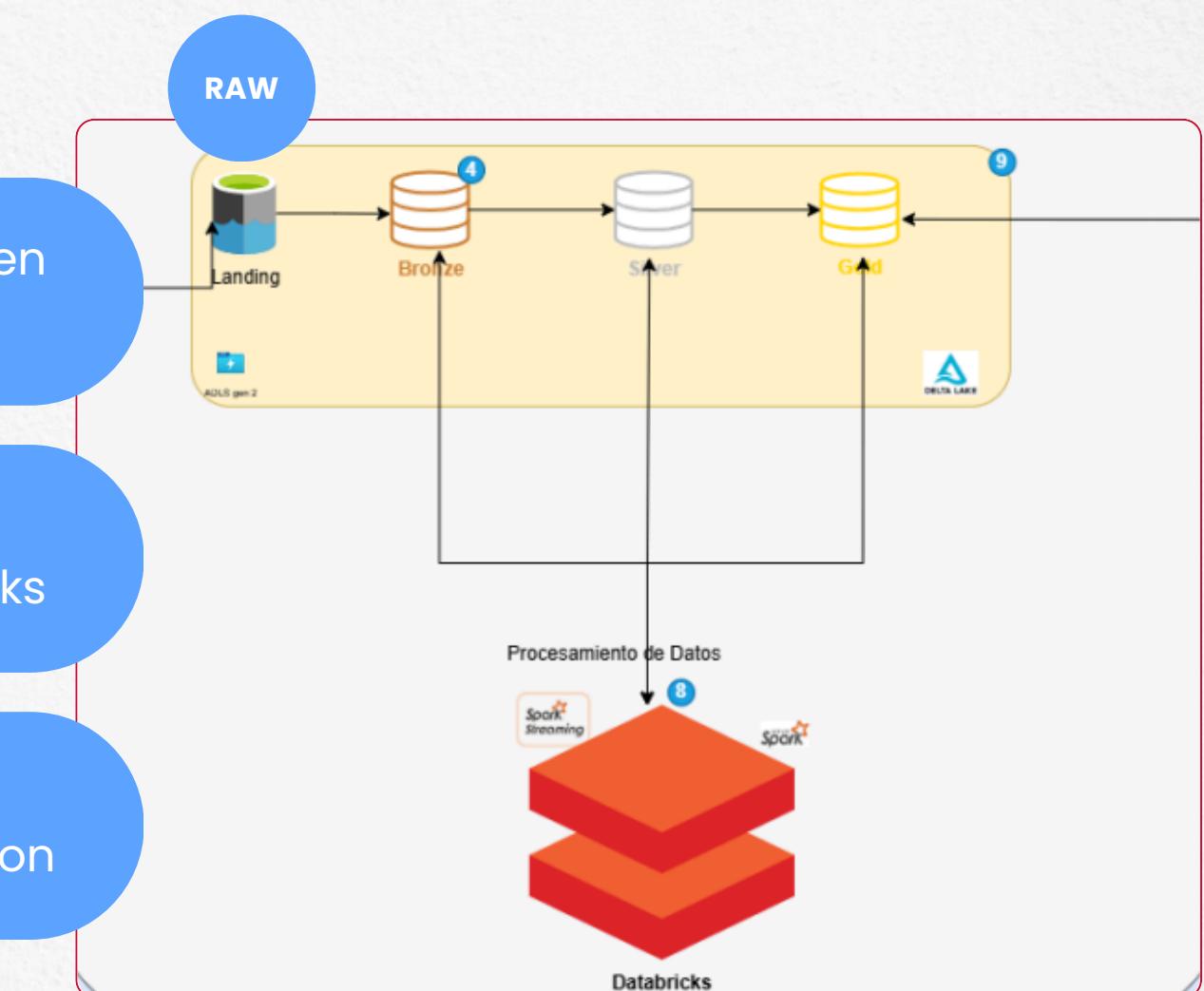
# RAW



Contiene los datos crudos tal como vienen de los sistemas fuente o APIs

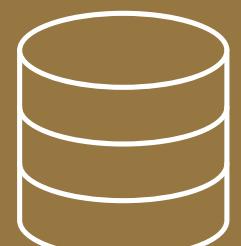
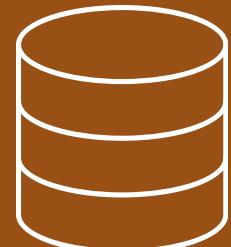
# Almacenados en Parquet (DT) para eficiencia y compatibilidad con Databricks

Funciona como respaldo y punto de partida para toda la arquitectura Medallion





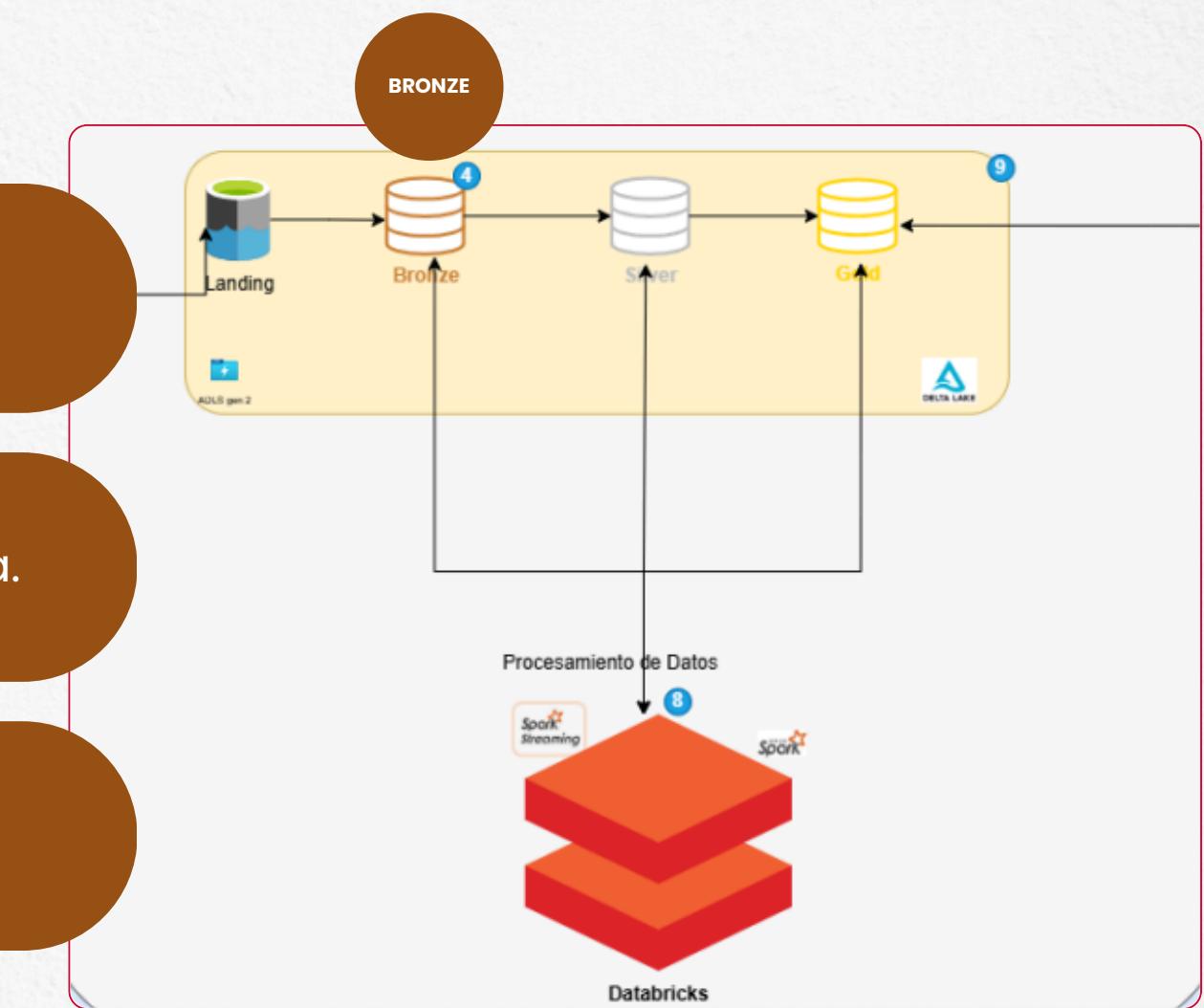
# BRONZE



Datos estructurados y versionados en Delta Tables.

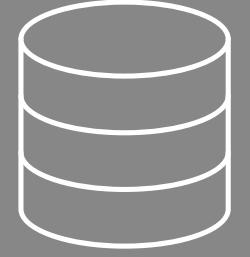
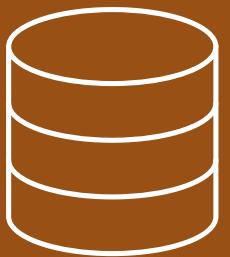
Fuente histórica confiable para auditoría.

Preparación inicial para limpieza y transformación.





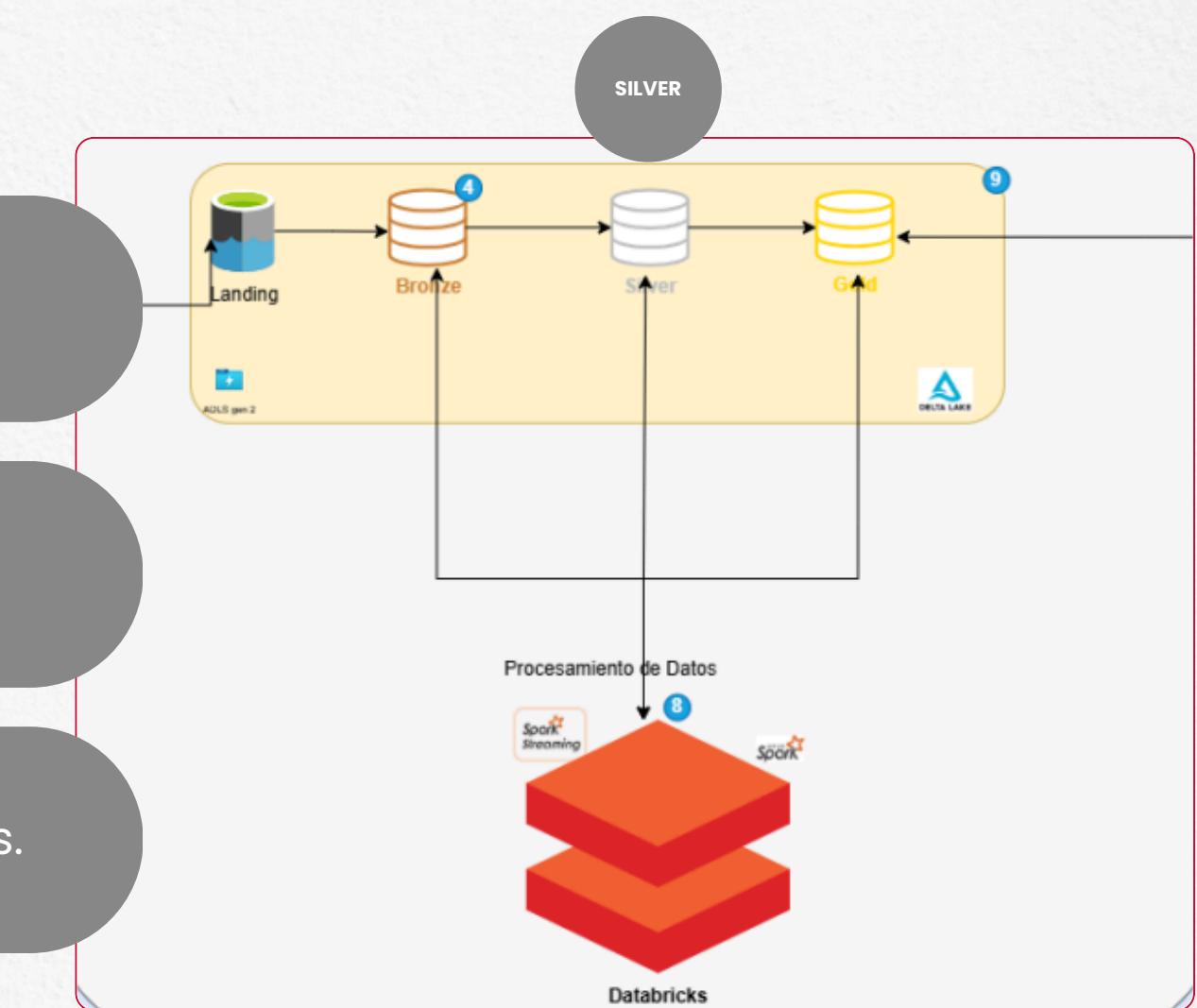
# SILVER



Datos limpios, validados y unificados.

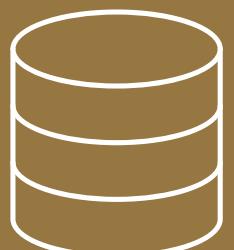
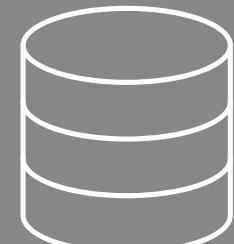
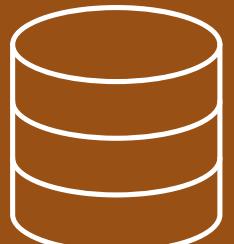
Integración de múltiples fuentes de manera consistente.

Listos para análisis y reportes confiables.





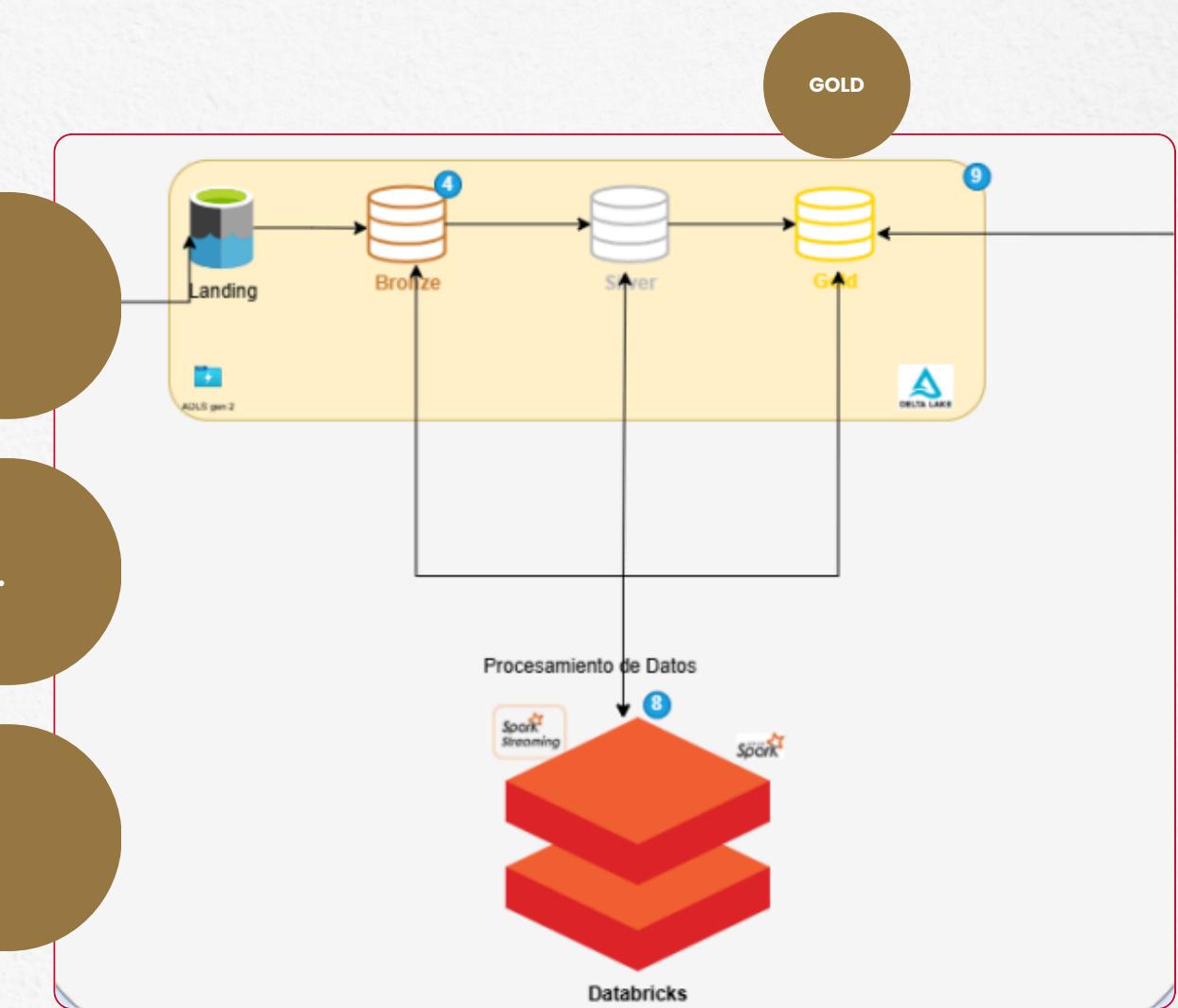
# GOLD

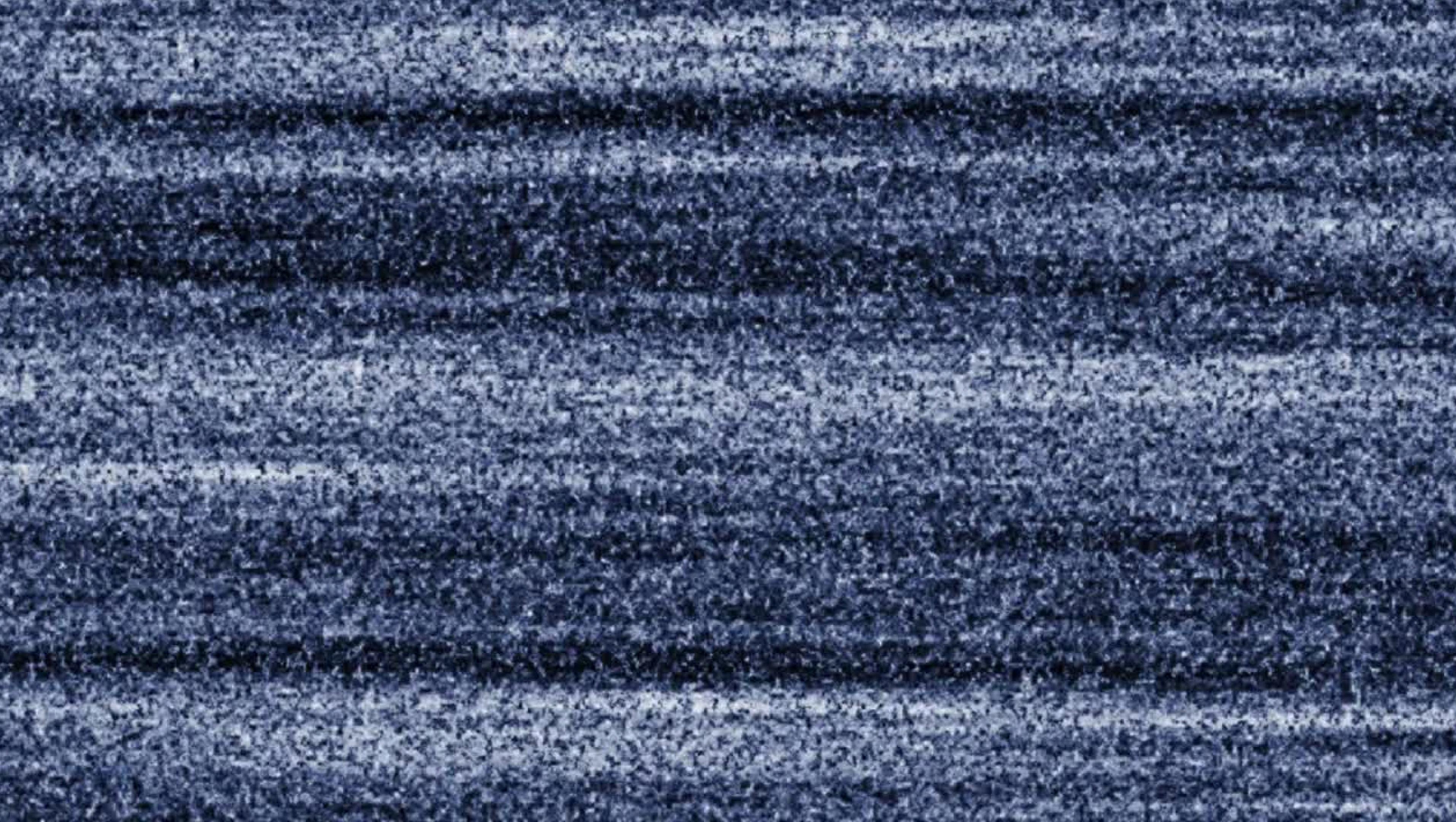


Datos refinados y agregados a nivel de negocio.

Alimenta dashboards y KPIs estratégicos.

Base de decisiones críticas y reportes de alto nivel.







## DATABASE CONNECTION ERROR

Failed to connect to required tables.

Error: 119 Unknown Database.

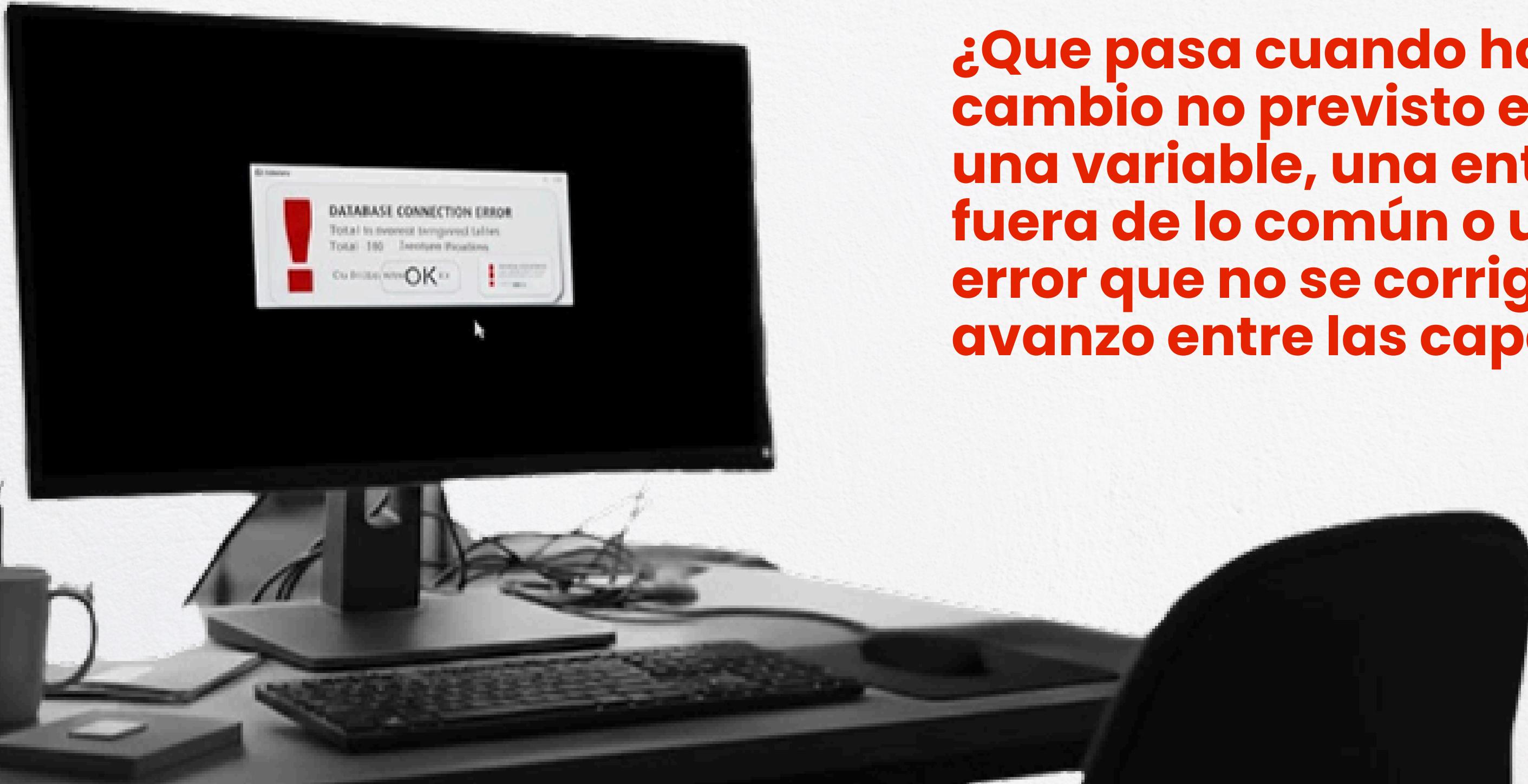
Error: Unknown Database

SQLSTATE[HY000]OKOKerror





**¿Que pasa cuando hay un cambio no previsto en una variable, una entrada fuera de lo común o un error que no se corrigió y avanza entre las capas?**





# NUESTRA Propuesta

Implementar **Delta Live Tables** sobre la arquitectura Medallion existente, garantizando que cada transacción, cada registro y cada reporte se genere a partir de datos frescos, limpios y confiables, automatizando la calidad y trazabilidad desde el primer byte hasta la decisión estratégica.

- Automatiza y asegura la calidad de los datos.
- Evita errores que lleguen a Silver y Gold.
- Escalable, auditabile y confiable para decisiones estratégicas.
- Posibilidad de alertas automáticas por correo o Teams ante fallos de calidad, mediante webhooks o integraciones con Power Automate.

LIVE TABLES PIPELINE

```
CREATE LIVE TABLE raw_data AS SELECT * FROM cloud_files("/data", "json")  
  
CREATE LIVE TABLE clean_data AS SELECT ... FROM LIVE.raw_data  
  
@dlt.table  
def scored_records():  
    return read("clean_data").map(lambda d: model_score(d))
```

# Dashboards

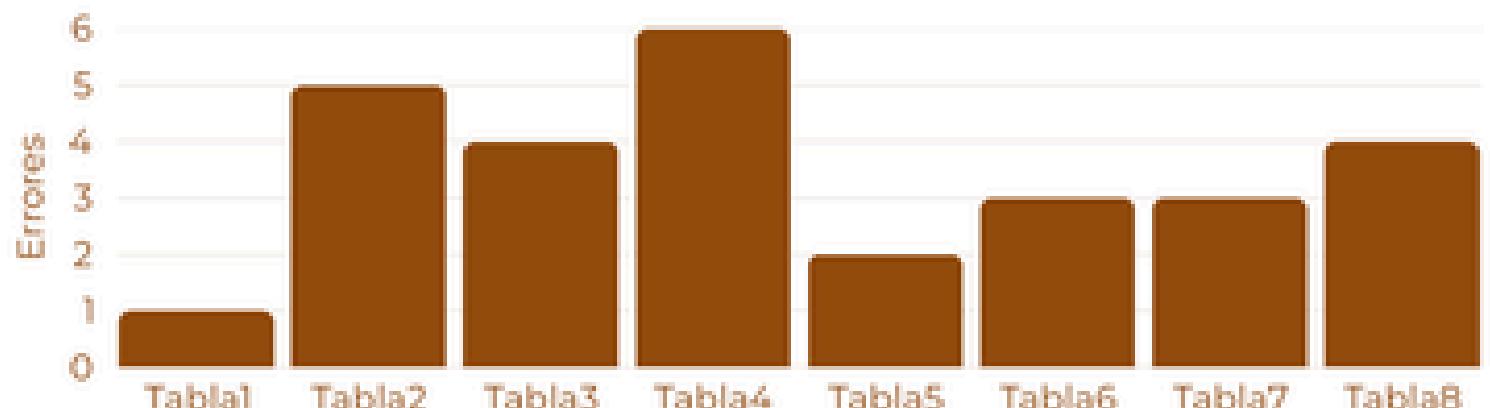


# Data Health

Tabla

Fecha

Errores por tabla



Score global de calidad del dataset

10%

Columnas rotas  
5

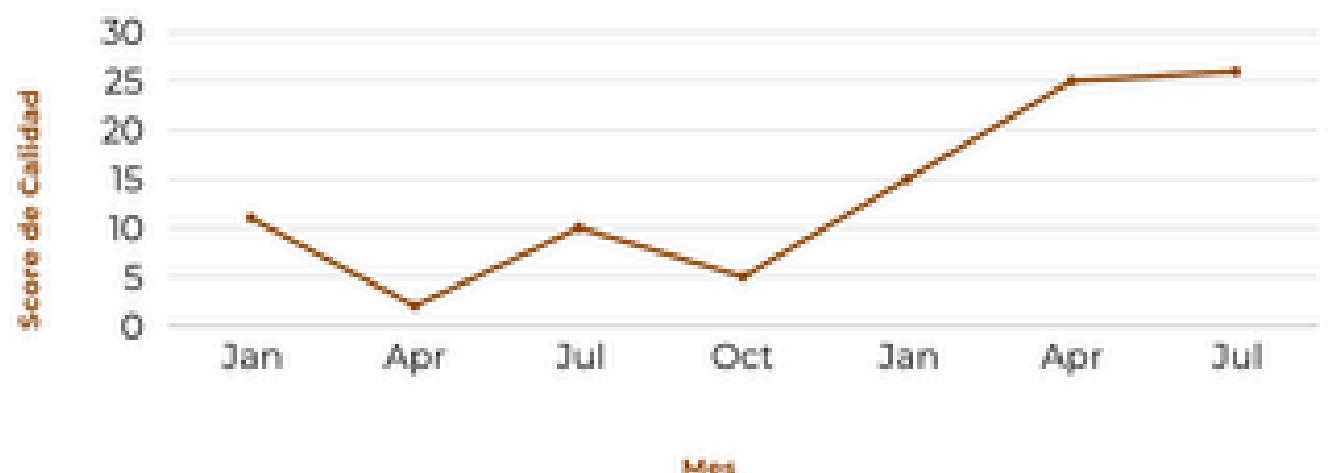
Valores Nulos

3567

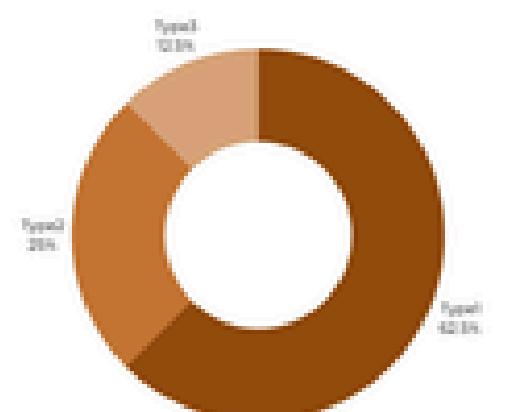
Valores Duplicados

1234

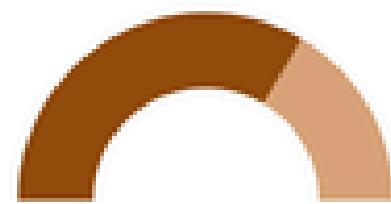
Evolución de la calidad en el tiempo



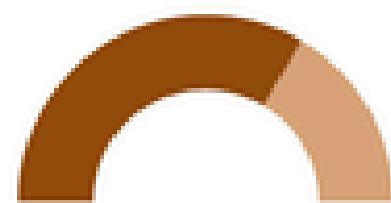
Valores Fuera de Rango/Lista



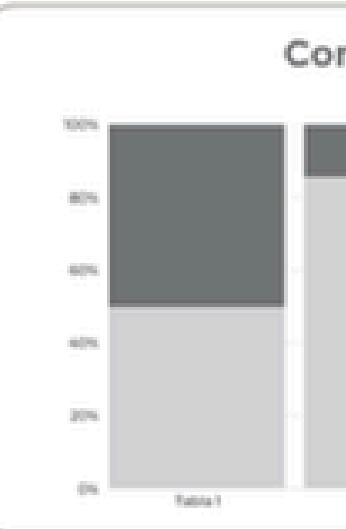
Errores vs. Fuente Externa



Porcentaje de Duplicidad



Dat





# Data Health

Tabla

Fecha

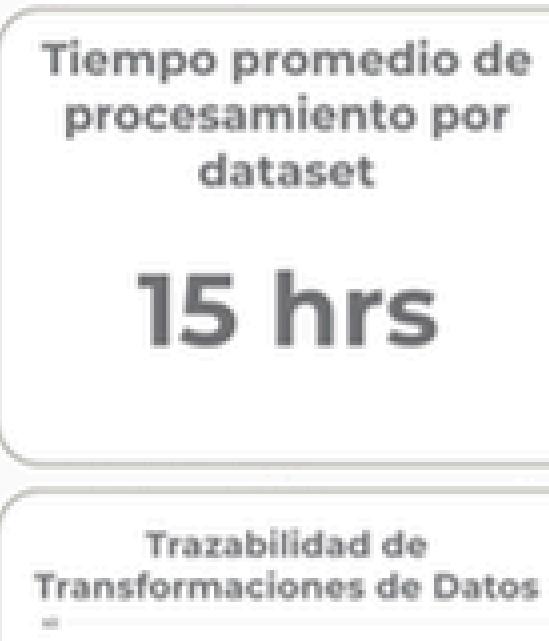
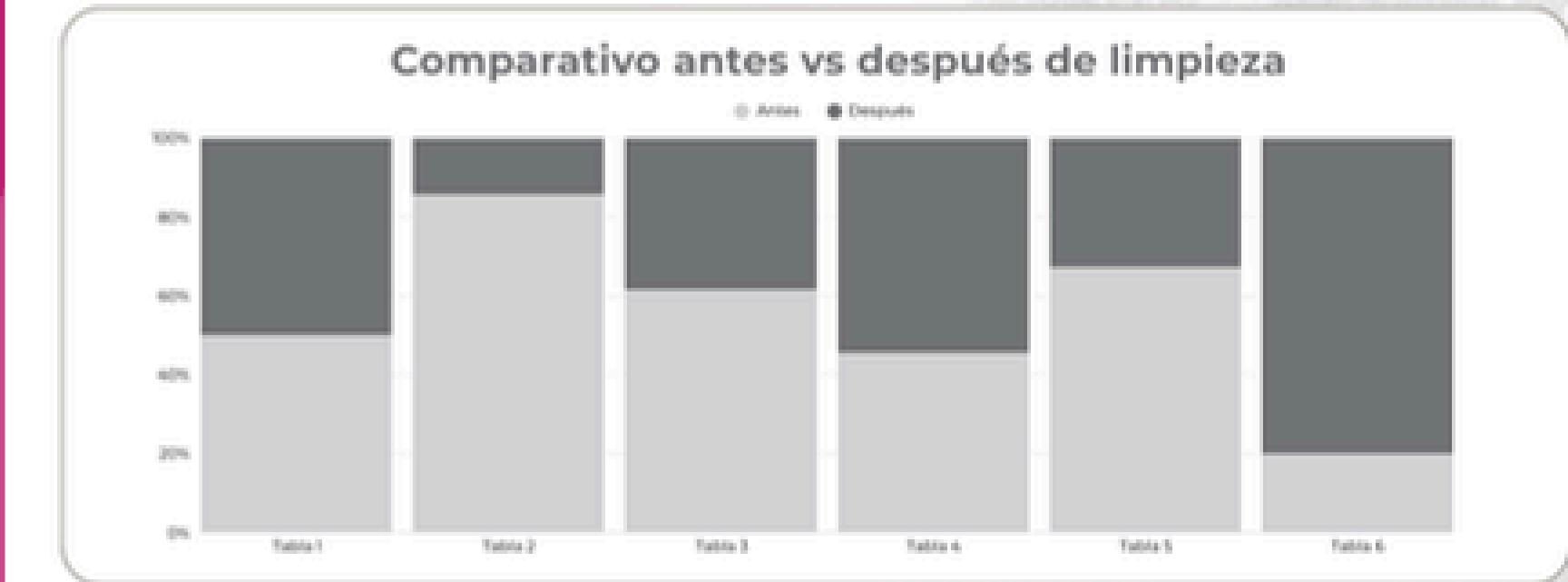


Tabla 1  
Tabla 2  
Tabla 3  
Tabla 4  
Tabla 5  
Tabla 6  
0%

Rentabil  
0



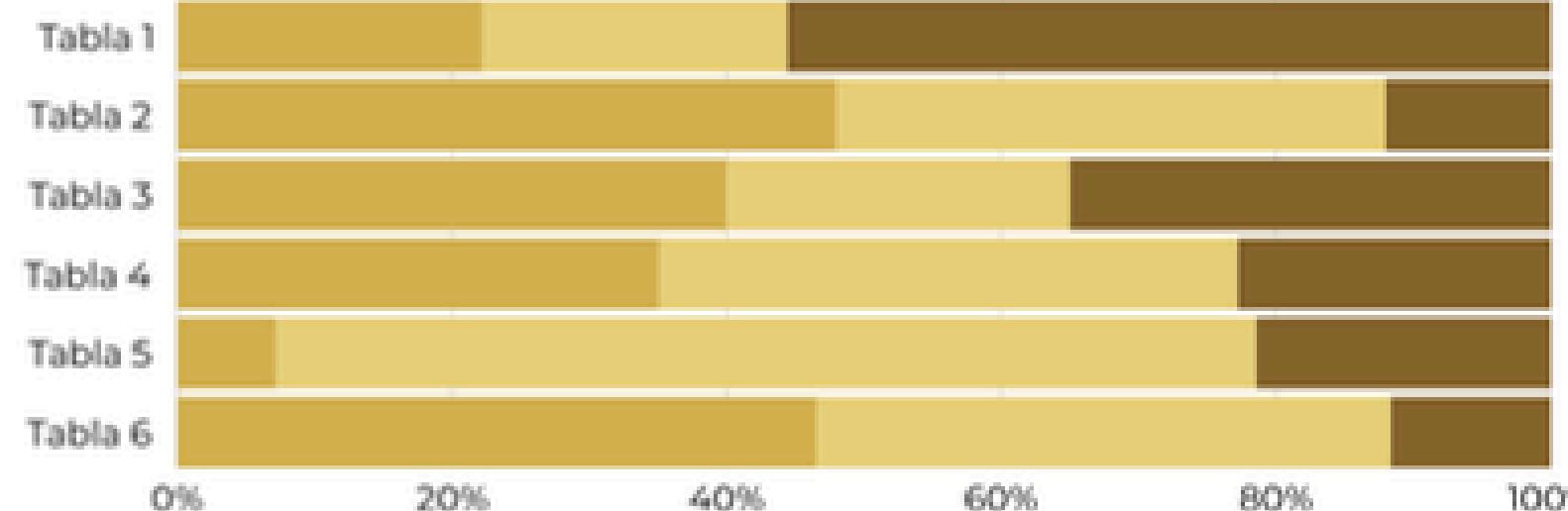
# Data Health

Tabla

Fecha

## Volumen de transacciones por canal

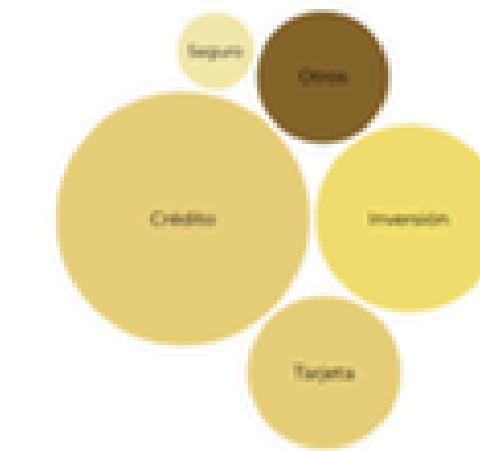
● app móvil ● banca web ● plataforma empresarial



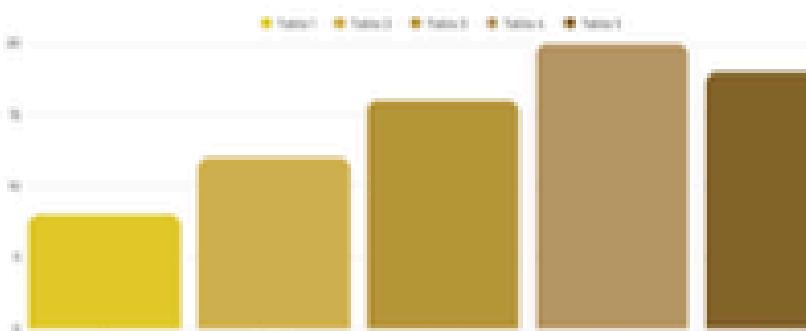
## Usuarios activos mensuales

**1,280**

Distribución del valor de transacciones por tipo de producto y segmento de cliente



## Rentabilidad por segmento o producto

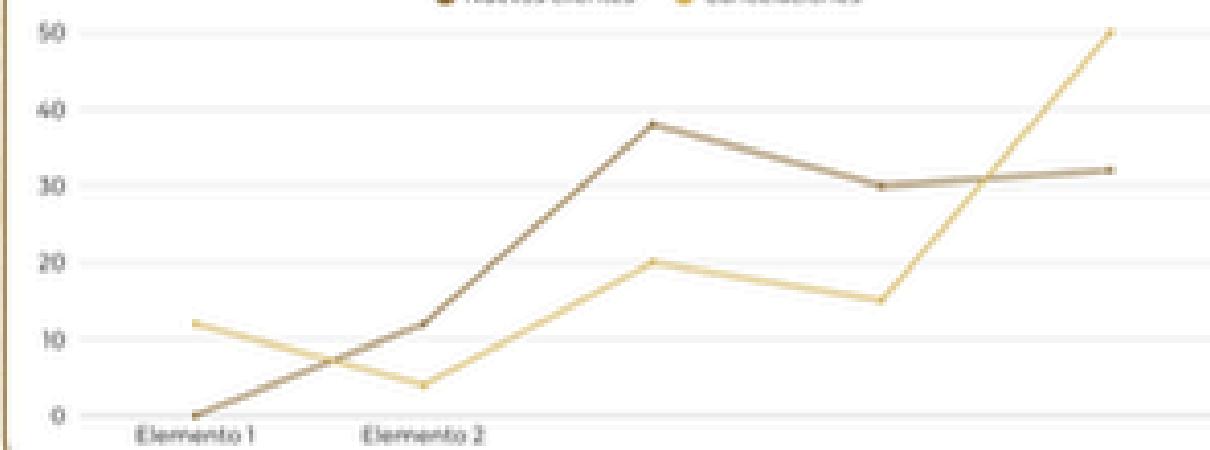


## Tasa de conversión digital

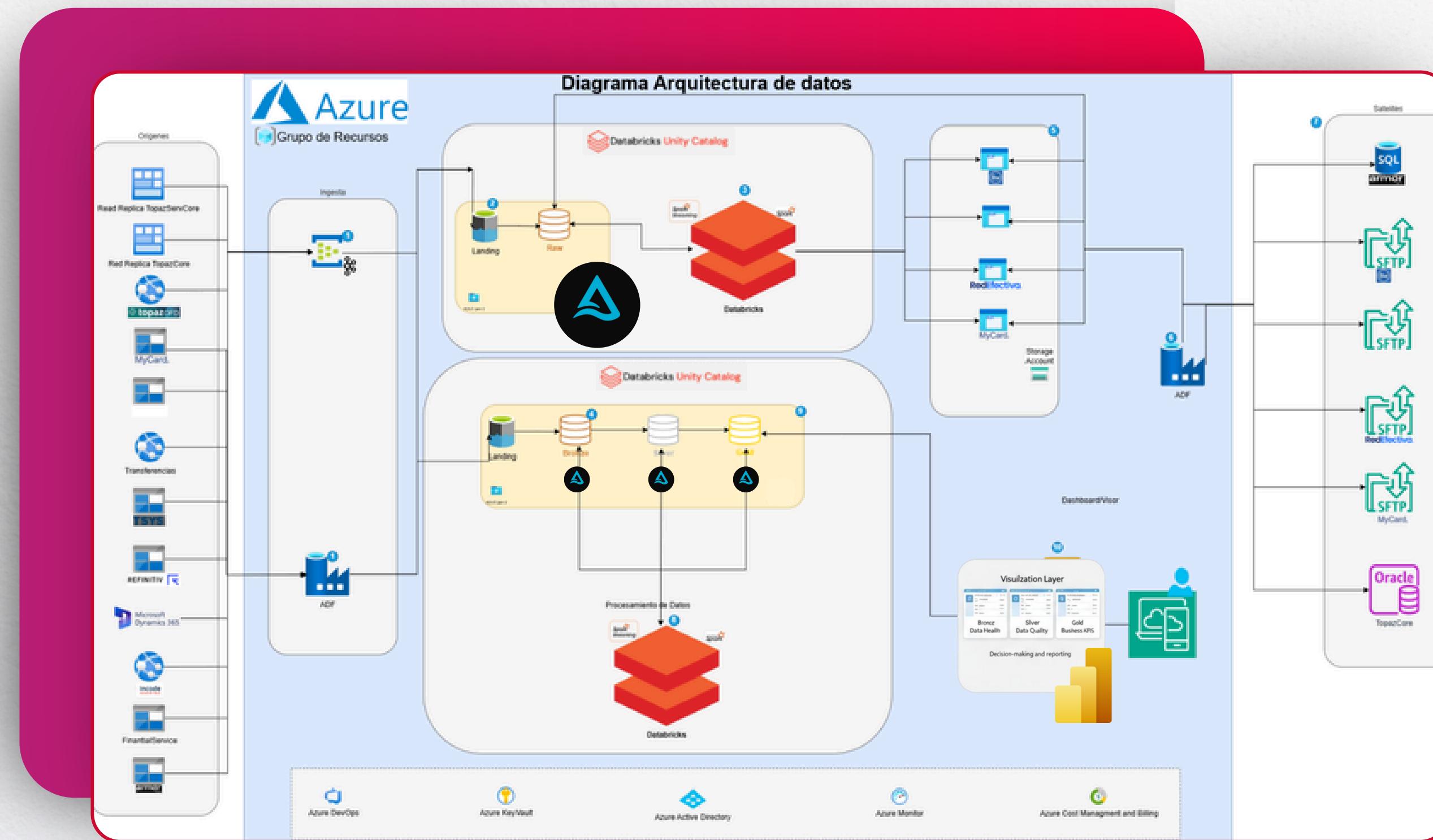
**15%**

## Tendencia de nuevos clientes y cancelaciones

● Nuevos clientes ● Cancelaciones



# Arquitectura



# Herramientas complementarias

## Timeline para implementaciòn

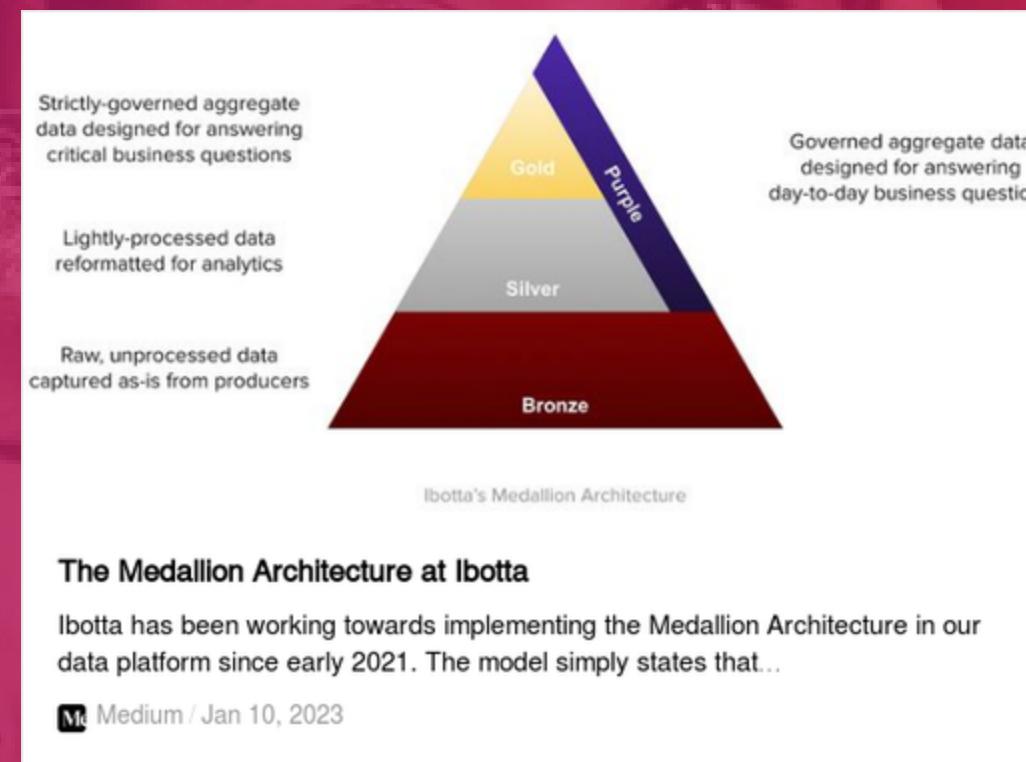
Tasks	Week 1	Week 2	Week 3	Week 4	Week 5
<b>1. Planning &amp; Analysis</b>					
1.1-Identify critical Bronze columns feeding reports.					
1.2- Define DLT quality rules and alert criteria.					
<b>2. Design &amp; Setup</b>					
2.1- Set up Databricks DLT environment in Azure.					
2.2-Configure Medallion layer connections (Raw + Bronze + Silver).					
<b>3. Development</b>					
3.1-Build DLT pipelines for ingestion and validation.					
3.2-Integrate automatic alert system (email/Teams).					
<b>4. Testing &amp; Pilot</b>					
4.1-Run pilot with subset of critical tables.					
4.2-Validate detection of column breaks and data inconsistencies.					
<b>5. Deployment</b>					
5.1-Scale DLT pipelines across all tables.					
5.2-Train BI and Governance teams on monitoring dashboards.					
<b>6. Monitoring &amp; Optimization</b>					
6.1-Ongoing review of alerts and performance.					

## Risk Management Plan

Risk	Impact	Level of Risk	Mitigation
Integration issues between Delta Live Tables and existing Medallion pipelines.	Medium	High	Conduct a pilot in a sandbox environment before full rollout; involve data engineering team in testing.
DLT validation rules may not detect all anomaly types initially.	High	High	Start with critical columns; then update DLT expectations; create feedback loop from BI team.
Alerts may not reach the BI team promptly due to email/Teams configuration issues.	Medium	Medium	Configure multiple alert channels (Teams + email); test notification workflows regularly.
Improper permissions in Databricks workspace might expose sensitive data.	Low	High	Implement role-based access control (RBAC) aligned with Bradesco's security policies.
Increased overhead on Bronze layer due to DLT monitoring tasks.	Medium	Medium	Optimize cluster configuration; schedule monitoring during low-traffic periods.
BI team resistance to adopting new monitoring tools.	Medium	Low	Provide training sessions and demonstrate reduced troubleshooting time.
Unexpected costs due to Databricks compute usage or Delta storage.	Low	Medium	Implement cost monitoring in Azure; use autoscaling clusters.

## Estimaciòn de costos

Category	Details	Cost in USD
Compute (DLT)	<a href="#">500 DBU-h en plan Pro (\$0.38/DBU-h)</a>	~\$190
Almacenamiento	Delta Lake tables (volumen medio de datos del piloto)	~\$200–\$400
Red / Egress	Transferencia de datos fuera de Azure (Delta Sharing, si aplica)	~\$50–\$100
Recursos Humanos	1 Project Lead, 2 Data Engineers, 1 BI/QA Specialist (12–16 semanas)	~\$17,000–\$23,000
Documentaciòn y Capacitaciòn	Manuales: Metadatos, Arquitectura, Implementaciòn; entrenamiento del equipo	~\$2,000–\$3,000
Total Estimado	Piloto	<b>~\$19,440–\$26,700</b>



**Detrás de cada transacción hay una historia. Con datos confiables, cada decisión se convierte en oportunidad.**



bradescard