



# Bradesco

## Manual de

## Políticas y Prácticas

Maria Dávalos Castro | A01639016  
Lizbeth Vargas Corona | A01752595  
Andrea Ramos Mc Gregor | A045561  
Mariana Nieto Moreno | A0639071



INDEX

# Table of Content

03	04
INTRODUCTION TO BRADESCO	SOFIPO
06	08
METADATOS	DATA AND METADATA CATALOG
09	13
MEDALLION ARCHITECTURE	POLISHING BRONZE INTO DIAMONDS: AUTOMATED MEDALLION MONITORING

# INTRODUCTION

# Bradesco



Banco Bradesco is a major Brazilian bank headquartered in Osasco, São Paulo. It is the second largest private bank in Brazil and Latin America by total assets. The bank has grown significantly over decades, mostly through mergers and acquisitions, including acquiring smaller banks and companies such as Banco Bilbao Vizcaya (BBVA operations in Brazil), Grupo Zogbi, Banco do Estado do Maranhão (BEM), and Banco Morada.

Bradesco offers a comprehensive range of financial services including deposits, various credit types, real estate financing, insurance brokerage, investment banking, credit cards, funds management, leasing, and consumer credit. The bank serves clients across different income levels and has a strong presence in capital markets underwriting. It also operates internationally, including in the United States, under Bradesco Bank.

# Reports Required by Regulatory Authorities in Mexico for SOFIPOS

Regulatory reports requested from entities such as SOFIPOS include financial information, credit operations, compliance with tax and anti-money laundering regulations, as well as reports related to resource management and utilization.

These reports ensure transparency, financial soundness, and regulatory compliance in accordance with the Ley de Ahorro y Crédito Popular (Article 47) and specific provisions issued by the Comisión Nacional Bancaria y de Valores (CNBV).

## Report Generation Sources

Reports are generated from the organization's transactional and analytical systems.

Data sources come from multiple providers, and information is ingested through APIs, SFTP connections, and databases such as Oracle, SQL, and Teradata.

The data is orchestrated into a Data Lake structured under the Medallion architecture (landing, raw, bronze, silver, and gold layers) using tools such as Azure Event Hub, Azure Data Factory, and Databricks.

In the transactional environment, data is processed, while in the analytical environment it is consolidated for the creation of dashboards, regulatory reports, and analytical models.

## Entities Receiving the Reports

Regulatory reports are primarily submitted to:

- Comisión Nacional Bancaria y de Valores (CNBV): Main regulatory and supervisory authority.
- Secretaría de Hacienda y Crédito Público (SHCP): Responsible for fiscal regulations and financial oversight.
- Fondo de Protección de Sociedades Financieras Populares y de Protección a sus Ahorradores: Supervises deposit guarantees.



### Mexican regulatory framework

- SOFIPOs: Ley de Ahorro y Crédito Popular + supervisión CNBV + fondo de protección.
- Bradescard (S. de R.L.): Ley General de Sociedades Mercantiles + regulación financiera aplicable a emisores de crédito.
- En ambos casos: cumplimiento en PLD/FT, protección de datos personales, normativas fiscales y regulatorias.

### Regulatory Requirements for SOFIPOs in México

SOFIPOs in Mexico are regulated by the National Banking and Securities Commission (CNBV) under the Popular Savings and Credit Law and its provisions. Key requirements include:

- Authorization and Operation

SOFIPOs must obtain prior CNBV authorization with detailed documentation about directors and officers, while meeting specific capitalization requirements based on their size and operational level.

- Operational Levels and Prudential Regulation

Operations are categorized by asset levels with differentiated requirements for minimum capital, risk management, credit portfolio provisions, liquidity ratios, and credit manuals for various asset ranges.

- Reporting and Supervision

Mandatory periodic regulatory reporting to CNBV and supervising federations. The CNBV regulates auxiliary supervision and implements corrective measures to ensure financial stability.

- Security and Data Protection

Specific regulations governing information handling, preservation, and security, including electronic authentication and confidentiality, third-party database management contracts, and operational continuity protocols.

- Legal Framework for Roles and Governance

Explicit regulations regarding board composition and responsibilities, technical committees (credit, supervision, compensation), internal and external auditors, and compliance roles.

# Data Governance at Bradescard Mexico (SOFIPO) with Azure Architecture

**Metadata is data that describes other data. It includes information about origin, format, owner, update frequency, quality, business rules, transformations, and relationships between data sets.**

**Proper metadata management is essential for enabling a governed, interoperable, and reliable data ecosystem.**

### Key Principles

- **Centralization:** All metadata must be stored and managed in a central catalog accessible through tools compatible with Azure Data Factory and Databricks, integrated with the Data Lake.
- **Standardization:** The names of tables, columns, metrics, and entities must follow a uniform convention defined in the Data Dictionary.
- **Automation:** The use of automatic metadata discovery tools (e.g., Azure Purview or Databricks Unity Catalog) is prioritized to reduce manual errors and maintain continuous updating.
- **Traceability:** Each data element must have a documented lineage from its source (CRM, Core, Dynamics) to its destination (dashboards, reports, or ML models).
- **Governance and Security:** Metadata must record who the Data Owner and Data Steward are, as well as access permissions, in compliance with the privacy and retention policies defined by CNBV, SHCP, and INAI.

### Metadata Lifecycle Stages

- **Creation:** Metadata is automatically generated during data ingestion (e.g., via Azure Data Factory or Databricks) or manually registered when new sources, tables, or reports are added. It includes technical, business, and quality information.
- **Validation:** Formal review to confirm the accuracy, consistency, and completeness of metadata. Business rules, naming conventions, and dependencies are verified.
- **Publication:** Once validated, metadata is incorporated into the Central Catalog (e.g., Azure Purview or Unity Catalog) and made available to business and analytics teams.
- **Maintenance / Update:** Metadata must be reviewed periodically (at least quarterly) to reflect changes in structures, sources, or definitions.
- **Audit and Monitoring:** Ongoing evaluation of policy compliance, metadata standards, and lineage accuracy.
- **Purging / Retirement:** Outdated, duplicate, or irrelevant metadata is deleted or archived following the organization's retention policies and CNBV-INAI regulations.

## MANUAL'S KEY TOPICS: METADATA

Type	Description
Technical	System-level information: table names, columns, data types, file paths in the Data Lake.
Operational	Details about data ingestion and processing: update date, responsible person, frequency, volume.
Business	Functional and semantic definitions: field descriptions and business meaning.
Quality	Metrics on completeness, accuracy, and validity of the data.
Security	Sensitivity classification and authorized access levels.



---

## MANUAL'S KEY TOPICS: DATA AND METADATA CATALOG

---

# Data and Metadata

The data and metadata catalog is the pillar of data governance. It functions as an inventory and marketplace where users can discover, understand, and trust the data available in the ecosystem.

## Main Objectives

- Accelerate discovery: Help users find relevant and useful data, reducing dependency on technical teams.
- Common language: Help as a main source of business definitions, metrics, entities, etc.
- Foster trust: Supply a complete context about the data, its origin, lineage, quality, providing informed decision making.
- Automatization: Add instruments that help automatization of document sources, tables and relationships, to keep information updated.

## Tools

The main tool implemented is Databricks Unity Catalog. Databricks environment helps all the transformation processes, allowing us:

- Centralized governance: Manage metadata, lineage, access permissions all in a single place.
- Automatic lineage: Record the data and directly feed it to the Data Lineage.
- Integration with Dictionaries: Allows the implementation of descriptions, tags, and comments to current tables and columns, helping standardization to business and technology departments.
- Additional tools: Some additional tools will be Azure Purview to complete visibility, before ingestion to Azure Data Factory.

## Data Validation

The data owner first needs to review the data and validate the information is correctly documented. Then, when the data has been approved, the data is marked as “verified” in the catalog. Lastly, the data is published to those users indicated as trustworthy.



## MANUAL'S KEY TOPICS: MEDALLION ARCHITECTURE

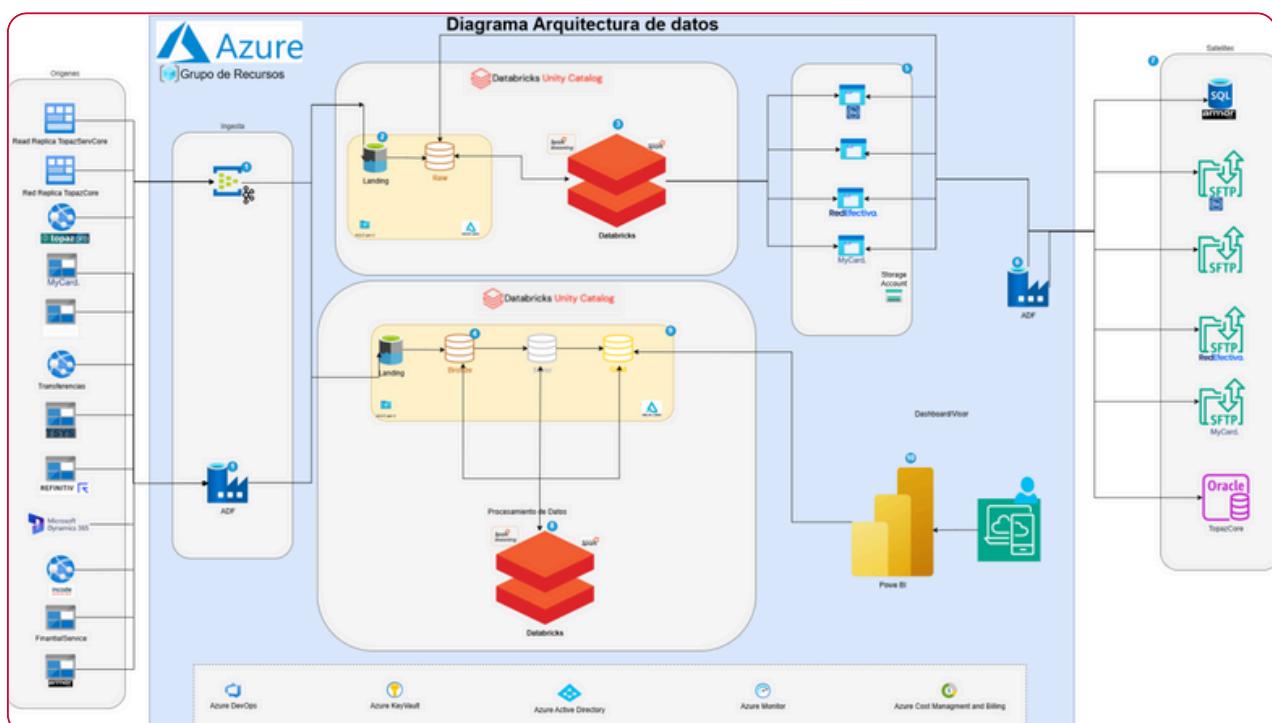
# Medallion Architecture

The medallion architecture is the design that structures and organizes the data. This data is taken from the Data Lake, so that we can process it from raw numbers to high quality, business information. At Bradescard, this model is designed to help comprehend information management, focusing in traceability, quality, and efficiency in data handling.

The objective of this architecture is to smooth the process of ingestion, processing and consumption of information. This scalable structure transforms data into an asset for decision making.

# Current Architecture

This is the current architecture designed for Bradescard. This structure starts from the data ingestion until the business intelligence analysis platforms such as Power BI:



# MANUAL'S KEY TOPICS: MEDALLION ARCHITECTURE

## SOURCES



On the left of the architecture we can see where the data is originated from:

- **Read Replica TopazServCore & Red Replica TopazCore:** Is one of the most popular and well-known core banking systems. It focuses mainly on transactional operations such as savings accounts, checks, loans, deposits, etc.
- **topazOFD:** It's a specialized module from Topaz that mainly focuses on fraud prevention
- **MyCard:** MyCard is a system that manages the lifecycle of their credit cards. This means transactions, credit limits, loyalty programs, etc.
- **Transferencias:** This is a system that only focuses on collecting information from transfers.
- **REFINITIV:** Refinitiv is a globally known financial market data and news. This helps provide real-time foreign data such as exchange rates.
- **Microsoft Dynamics 365:** Dynamics 365 is a customer relationship management that manages non-transactional interactions.
- **incode:** Is an identity verification platform that uses artificial intelligence to validate information such as ID photos and customers selfies.
- **FinancialService:** This system focuses on more specific transactions that aren't covered on the Read Replica TopazServCore & Red Replica TopazCore.
- **armor:** Amor is a globally known cybersecurity company. This system brings data about the current security of the current infrastructure.

## ingests

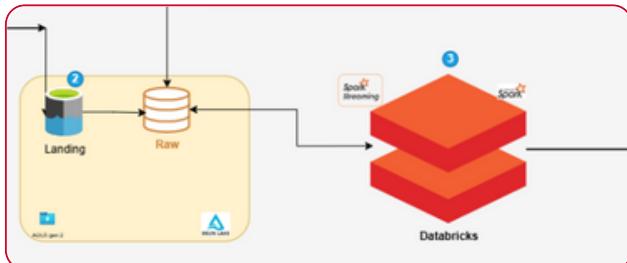
In the ingest is the process between the data lake in Azure in data brick. The main goal is to connect the information in a reliable and secure way. This phase acts as the bridge between all the scattered data sources and the data lake.

- **Azure Data Factory (ADF):** Azure Data Factory or ADF is a data processing system that ingests the data automatically every set period of time, and upload it to the data landing. Additionally, it creates copies of the data
- **Azure Event Hubs:** Azure Event Hubs is a data processing system designed to capture real-time information. Azure Event Hubs is not designed to ingest or copy large capacities of information. Thus, it works with real-time data.



# MANUAL'S KEY TOPICS: MEDALLION ARCHITECTURE

## processing



The processing area is divided into two parts or sections, each focusing on different processes and objectives.

- **Landing (ADLS gen 2):** This is where the data from Azure Data Factory is loaded. The ADLS gen 2 means were the data is physically located.

- **Raw (DELTA LAKE):** From the landing, the data is moved to a structure called "Raw". This is the first layer were the data is structure. This is fundamental as it transforms the data to the final Delta format. This layer provides reliability, a stored data history and improves performance.
- **Databricks:** This is the main processing system were the data is transformed. It uses Apache Spark to process all the data.
  - **Apache Spark:** This process the information from the Raw layer and executes the cleaning, validation, and data combination.
  - **Spark Streaming:** This process the data from Azure Event Hub.

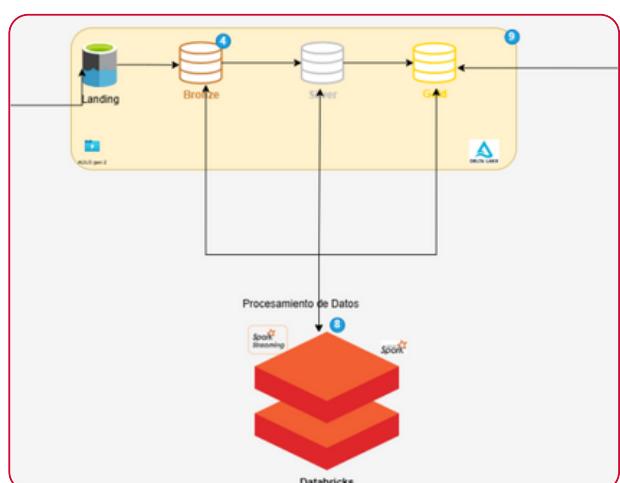
### Importance of Databricks

**Databricks Unity Catalog:** This is not part of the process of dataflow, but a process related to data governance and data security.

## storage

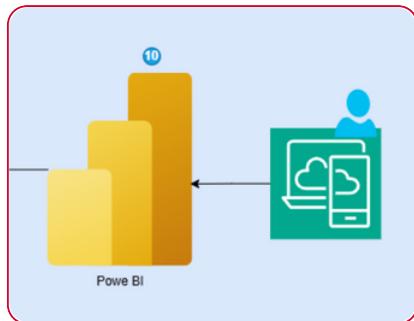
The data is layer store in 3 more layers depending on different factors:

- **Bronze:** This is the first layer after the "Raw" layer. It contains the data from "Raw" without any alterations. It helps as a reference or "historical source". This helps in case any error occurs in the nexts layers, we can have the data with any alterations. It ensures reliability and versioning.
- **Silver:** This is the second layer, and it has the first transformation of the data. In this layer the data is group together into a single source. This version has the data that has been cleaned, validated, standardized and enriched.
- **Gold:** This is the next and final layer, and has the highest quality of data. This data is highly refined and its used for the final consumption. This data is feed to dashboards and reports that are designed to answer specific questions.



## MANUAL'S KEY TOPICS: MEDALLION ARCHITECTURE

### consumption



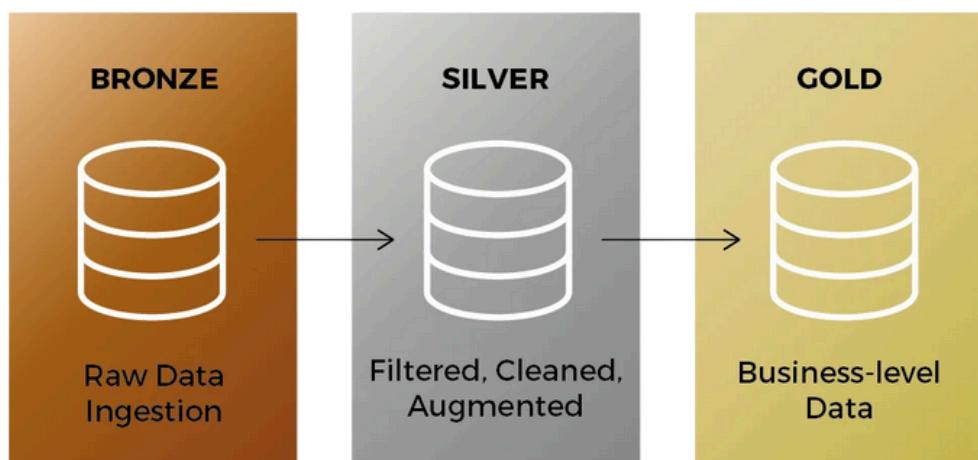
Lastly, on the consumption phase we can use the data from the golden layer for the construction from business intelligence. On this layer we will find the following platforms:

- **Power BI:** This is Microsoft's main Business Intelligence and visualization tool, and lets us create interactive reports and dashboards.
- **End-Users and Accessibility:** This is where we can see the reports created in Power BI, in just one place. In here, user can personalize the reports based on a number of filters previously selected, depending on their necessities. Its main goal is to have democratize data, making it available everywhere, anytime.

## Benefits

The medallion architecture offers a lot of benefits and advantages, for example:

- **Traceability:** It allows us to track data from the ingestion to its final consumption and detect any errors on time.
- **Data Quality:** It assures the ensures the data quality control on each layer guaranteeing reliability.
- **Scalability:** This type of architecture separates the processes of ingestion, transformation and consumption. This helps optimization processes and scalability of the data in the ecosystem.
- **Analysis:** The different layers on the architecture enables different teams to access different levels of detail depending on the necessities.



# Polishing Bronze into Diamonds: Automated Medallion Monitoring

## CHALLENGE

When a column is broken, renamed, or contains inconsistent data, Power BI dashboards fail to reflect accurate information, impacting decision-making.

Currently, monitoring and validation in the Bronze layer are limited, allowing errors to propagate to Silver and Gold layers before detection.

## Turning Raw Data into Diamond-Quality Insights

Our proposal is to enhance Bradescard's existing Azure Databricks architecture by implementing Delta Live Tables (DLT) on top of the current Parquet layer, transforming static Delta Tables into live, automated data pipelines. This solution will automate ingestion and transformation across the Medallion layers while continuously monitoring and validating data quality in the Bronze stage. By detecting missing or renamed columns, null values, and structural inconsistencies in real time, the system will generate immediate alerts to the BI team and preserve a versioned history of data for rollback, auditing, and traceability. This ensures a seamless flow of reliable data from Bronze to Gold, turning raw inputs into diamond-quality insights ready for accurate and trustworthy Power BI dashboards.

## Benefits

- ★ Early detection of errors
- ★ Automated monitoring
- ★ Reliable reporting
- ★ Efficiency and scalability
- ★ Auditability and traceability

## Implementation

- 1 Identify critical tables and columns in Bronze feeding key dashboards.
- 2 Define DLT data quality rules (expectations) for each critical column.
- 3 Build DLT pipelines for incremental ingestion, validation, and transformation.
- 4 Integrate automated alerts via email, Teams, or Slack for immediate notification.
- 5 Develop Power BI dashboards for each layer



# DASHBOARD PROTOTYPES

BRONZE

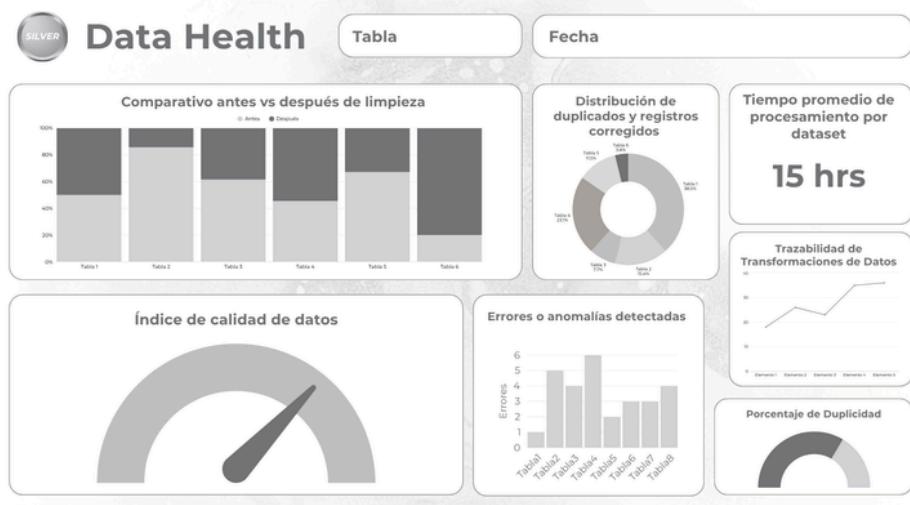
## Data Health



BRONZE LAYER

SILVER

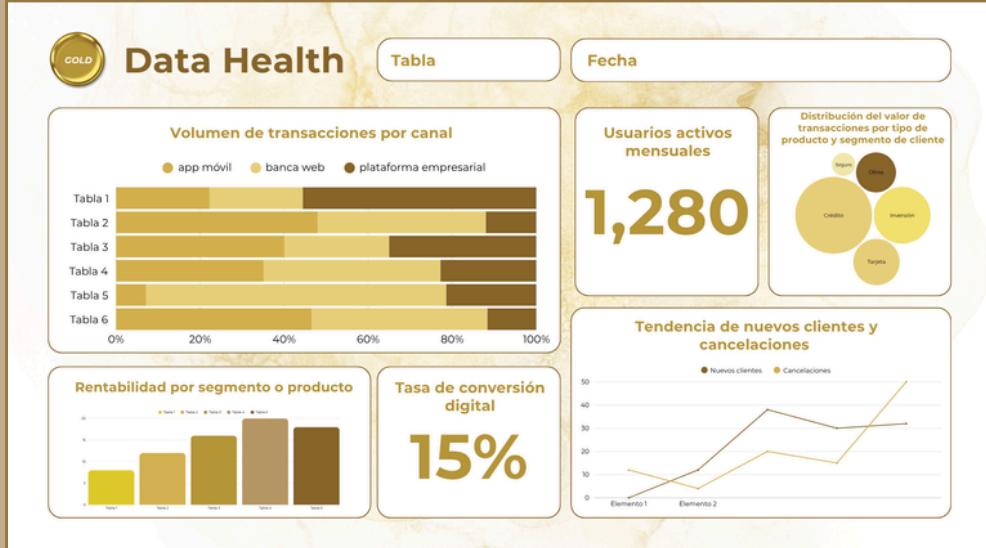
## Data Health



SILVER LAYER

GOLD

## Data Health



GOLD LAYER

# 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



# GANTT CHART

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.					



**THANK YOU!**