

Transforming \$30B+ AUM Data into Actionable Insights with 1Platform by Modernizing Data Architecture

Client Overview

Our client is a leading wealth management firm headquartered in the US with a significant operational footprint across Latin America. The company primarily caters to high-net-worth individuals by offering a comprehensive suite of services, including brokerage, advisory, and platform solutions.

In terms of its market status and operational scale, they manage information for over **\$30 billion in Assets Under Management (AUM)**. Consisting of both in-house and independent professionals—and serves a diverse clientele of over **32,000 individuals**.





Key bottlenecks

The client faced significant operational and technical hurdles that hindered their growth and data-driven decision-making:



Lack of a Centralized Repository:

Data was scattered across various SQL databases, APIs, and siloed platforms, resulting in a non-unified view.



Duplicated Records:

There was no de-duplicated "golden record" for customers, financial accounts, securities, or advisors.



Unstructured Data Overload:

Thousands of legacy **KYC PDF forms** (both scanned and digital) had accumulated without a way to convert them into structured, tabular formats.



Manual Operations:

Pipeline triggering and monitoring were entirely manual, leading to consistent daily operational challenges.



Fragmented Customer Profiles:

Siloed systems meant financial advisors lacked a consolidated view of a customer's status across different custodians.



Ineffective Governance:

Gaps in data accuracy and consistency existed due to a lack of centralized data governance across custodians.

Solution & Methodology

To address these challenges, we utilized a framework centered around 1Platform, a metadata-driven unified data analytics solution

1 Architecture Framework:
We implemented a Medallion Architecture (comprising Bronze, Silver, and Gold layers) on Microsoft Fabric.

2 Integrated MDM Layer:
A crucial Master Data Management (MDM) layer was introduced between the Silver and Gold layers to handle dimensions and ensure data integrity.

3 Hybrid OCR Approach with interactive UI:
To solve accuracy issues with legacy forms, we developed a hybrid approach combining Python scripts with OCR fallback, which significantly increased speed and accuracy compared to standard LLM-only models.

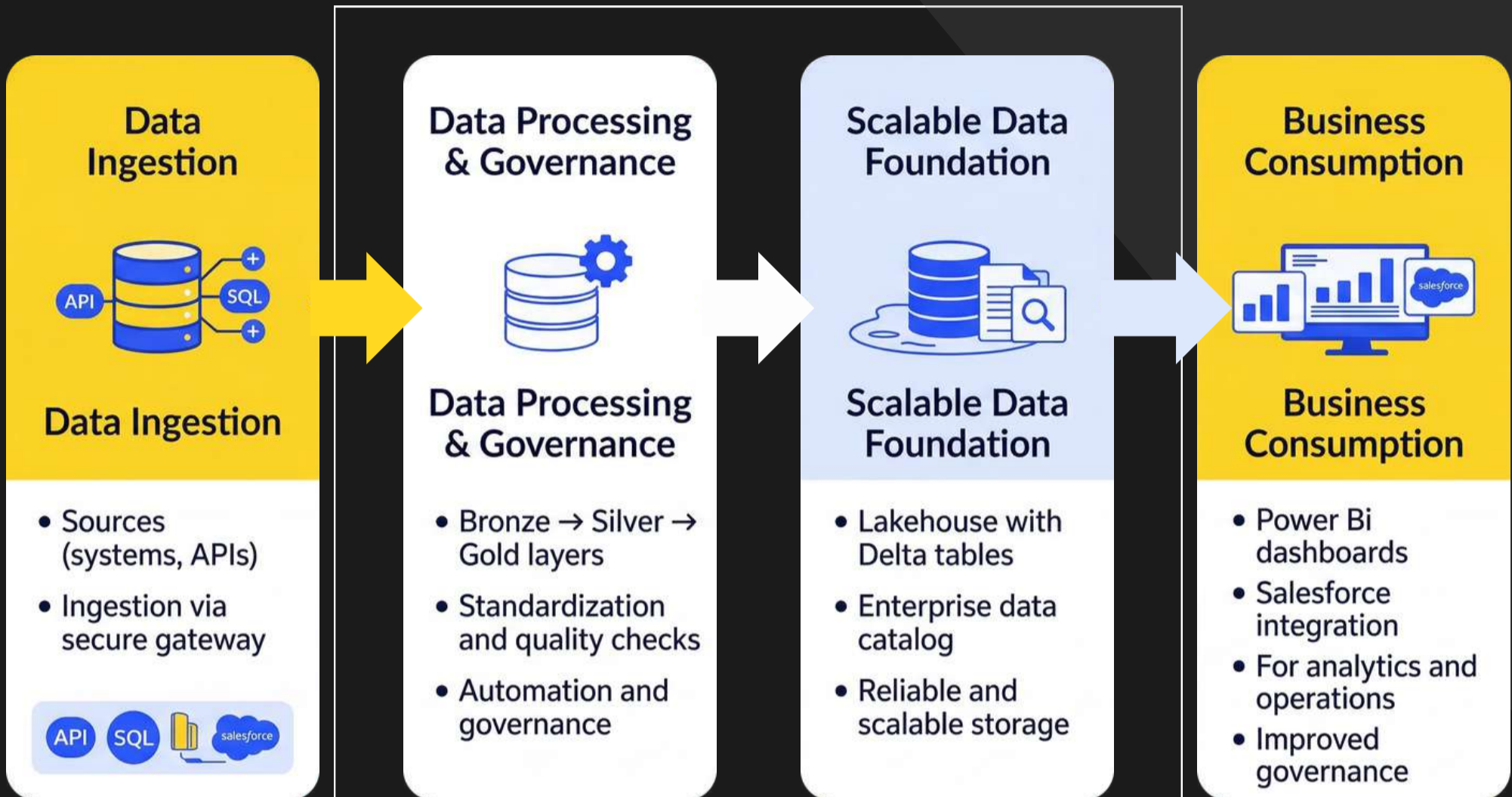
4 Automated Orchestration:
Using the Data Nexus tool within 1Platform, we created an automated environment for ingestion, transformation, and real-time monitoring of data pipelines.

5 Deduplication Strategy:
A robust algorithm was employed, using weighted scoring across phonetic matches (e.g., for names) and exact column-level matches (e.g., for tax IDs).

6 Collaborative Governance:
We established strict standards for nomenclature, data quality rules, and incremental loading strategies to ensure consistency across all artifacts



Platform based architecture for the product



Platform capabilities for

- Ingestion
- Transformation
- Optical Character Recognition (OCR)
- Orchestration
- MDM Deduplication
- Golden Record Creation
- ▶ Data Storage based on Lakehouse
- ▶ Fabric Data Pipeline based orchestration
- ▶ Data movement and execution in Microsoft Fabric

Significant details about the project

Scale of Pipeline

We developed over **220 automated ingestion and transformation pipelines.**

Data Volume

The warehouse successfully manages information for over **\$30 billion in Assets Under Management (AUM).**

Processing Efficiency

The refined OCR process reduced form-processing time from 20-25 minutes down to **2 minutes per form**, with accuracy jumping from 60% to **95%.**

Robustness

Pipelines were designed to be **idempotent**; if a pipeline fails, it is built to automatically fetch previous data during the next run without impacting data integrity



Business Impact



Unified Executive View

A clear, non-duplicate version of records for strategic decision-making



Elimination of Manual Effort

Automated monitoring and email notifications for pipeline failures



Legacy Data Unlocked

Approximately **7,000+ KYC PDFs** were converted into structured data



Reduced Rework

Estimated rework **reduced by 80%** and increased reliability of future AI projects



Real-Time Insights

The solution broke down data silos, powering real-time decisions and providing financial advisors with a holistic view of client portfolios



About

POLESTAR

ANALYTICS



Polestar Analytics is a leader in Data, Analytics, AI, and Enterprise Planning helping organizations to unlock intelligent outcomes through our proprietary products like TPlatform, accelerators, and services. Our expertise spans data engineering, data science, agentic and generative AI, and advanced planning for CPG/Retail, Pharmaceuticals, Manufacturing, IT/ITeS, and Financial Services