Proof of Work

A Cloud Migration to Microsoft Fabric

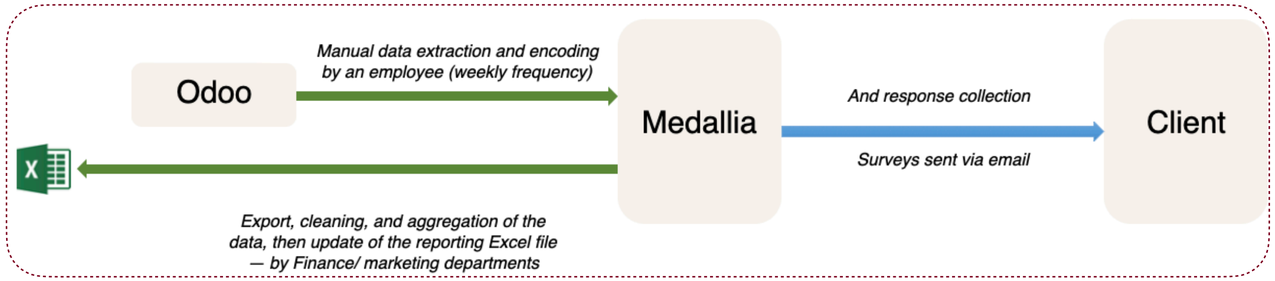
Auteur : **EL KAOUNI Yasser**

Microsoft Fabric

Context , Problem Statement , Proposed technical solution

I. General Context

- Belgian retail group undergoing a **cloud migration → Microsoft Fabric**
- 92 multi-region stores: **64 BE • 23 ES • 5 MENA**
- My **Project scope**: customer satisfaction data collected via **Medallia** —> End to End
  (NPS surveys + open-text comments)

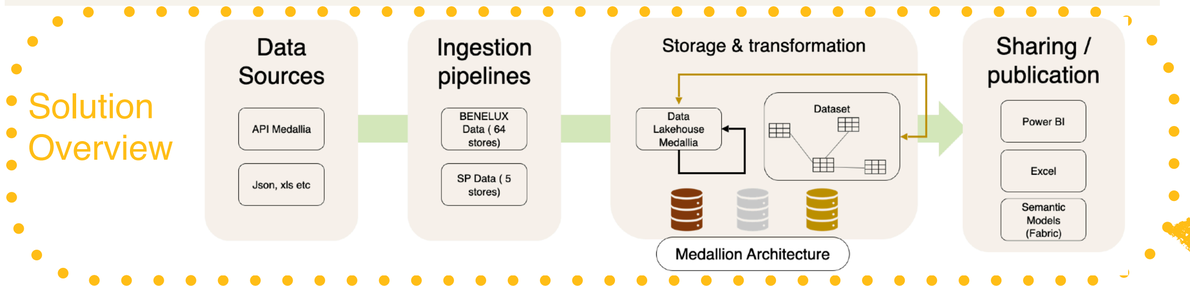


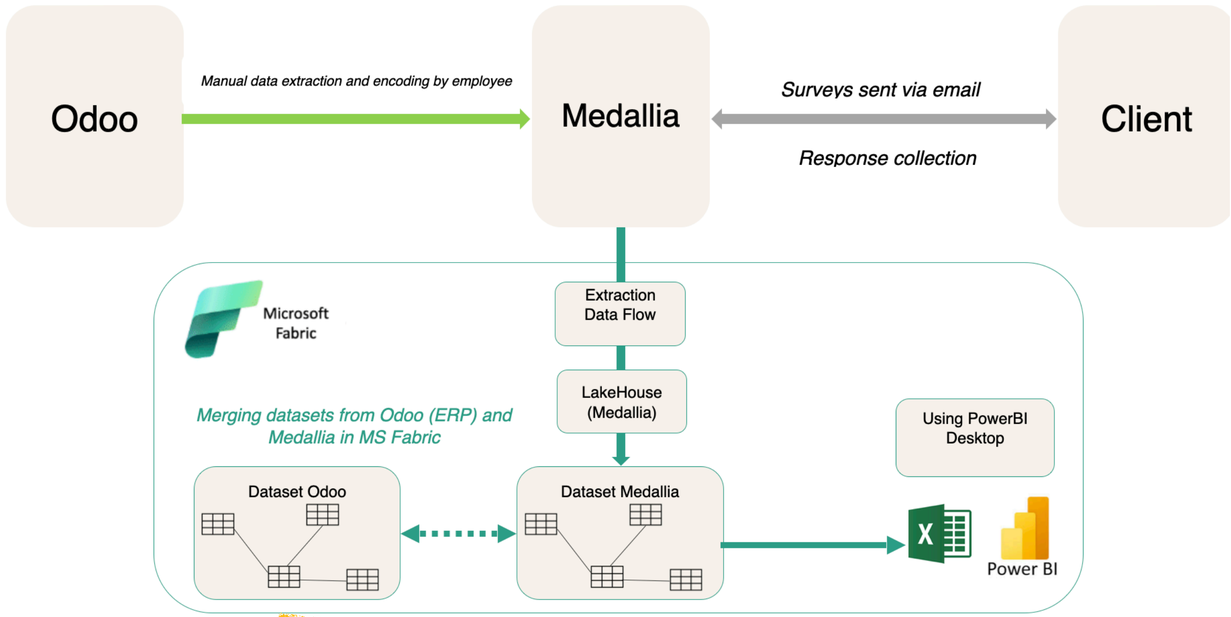
II. Business Problem

- Current process **manual / Excel-based**: weekly export by employee, cleaning, aggregation
- Latency: 7 days → decisions made too late
- Value loss: unstructured textual comments are unusable, copy-paste errors likely
- **Human cost**: ≈ 10 hours/week mobilized on finance & marketing side

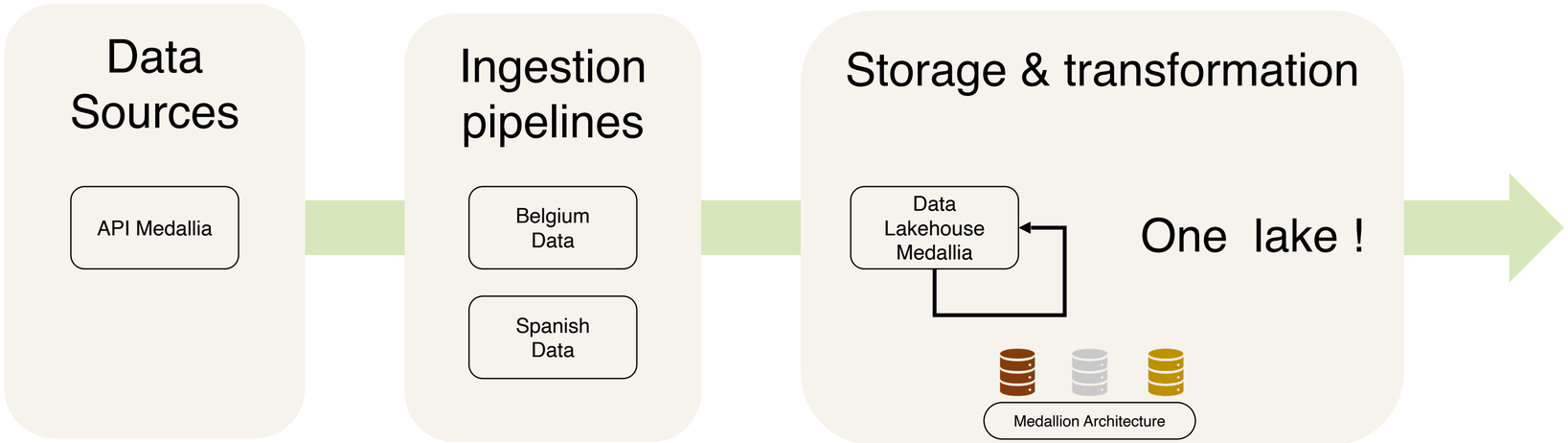
III. My Role (Data Engineer)

- Full end-to-end automation: Medallia API → Bronze/Silver/Gold (Fabric Lakehouse)
- Delta Load pipeline optimized: reconciliation from **5 min → 1 min (with delta load !)**
- Multi-source centralization: Medallia + Excel files (per country)
- Power BI dashboard (**daily** NPS & churn) with region-level RLS
- **Resulting impact: ~10 h/week saved on manual operations → decisions made in < 2 h**



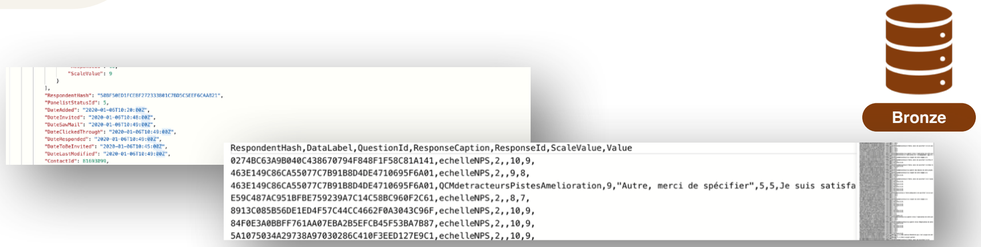


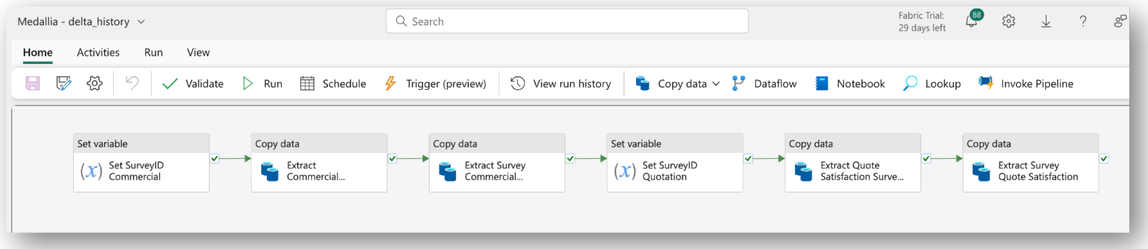
Medallion Architecture: Ingestion & Bronze Layer



Ingestion Phase

- Set up secure data extraction flows using REST API and Data Factory
- Multi-source & multi-country ingestion (Spain, Belgium, Netherlands)
- Data stored in Bronze layer of a Fabric Lakehouse
- Ingestion phase:
  - Load all historical data up to today
  - Load daily deltas from the last full load



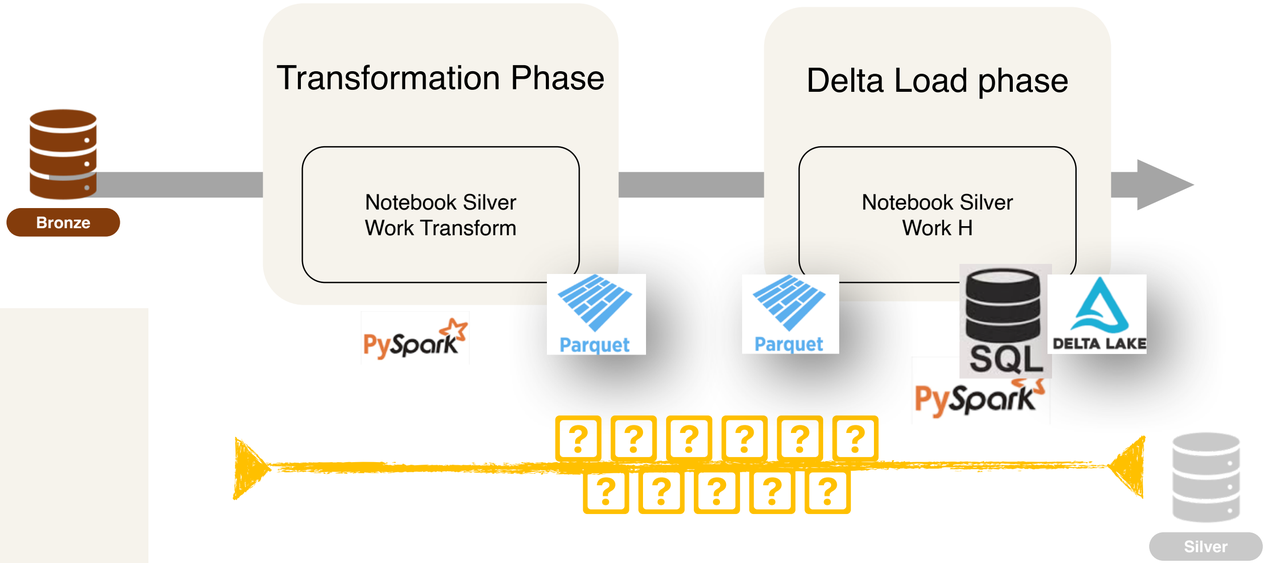


An example ingestion pipeline to the Bronze layer -with Fabric Data Factory

Medallion Architecture: Transformation, Optimization

Transformation Phase (Silver W Transform Layer)

- Data cleaning
- Enforcing schema and strict typing (columns, formats, etc.)
- Ensuring data quality and regulatory compliance
- Exploratory statistics
- Complex transformations (e.g. store mapping)



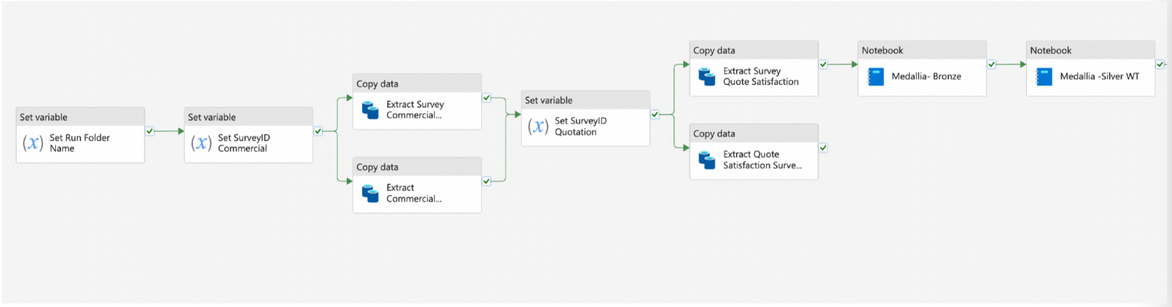
Optimization Phase

- Historical data merging (Delta Lake Upsert)
- Schema reconciliation
- Schema drift handling strategy (Delta)
- Data lineage & logging for auditability
- Orchestrator configuration

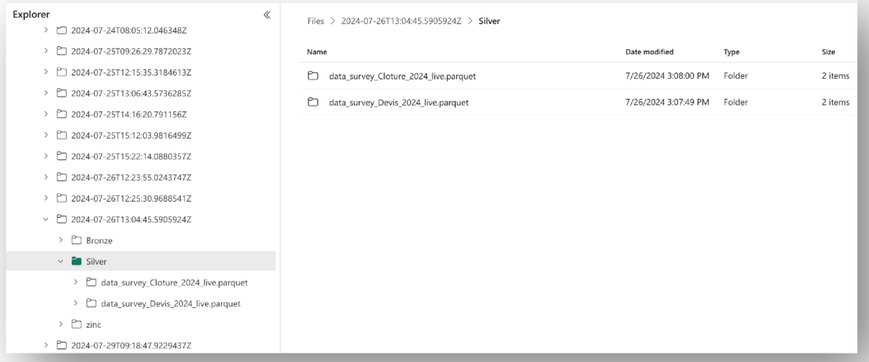
Delta Load technical details

- Silver WRK_T phase: snapshot version
- Reconciliation phase: O.S.S.T
- Performance improvement: 52 seconds (vs. 5 minutes)

Unit tests performed using notebooks:
- Temporal tests for Silver WRK_T
- Join tests via SQL endpoints
- Kolmogorov–Smirnov distribution comparison



An orchestrated transformation pipeline

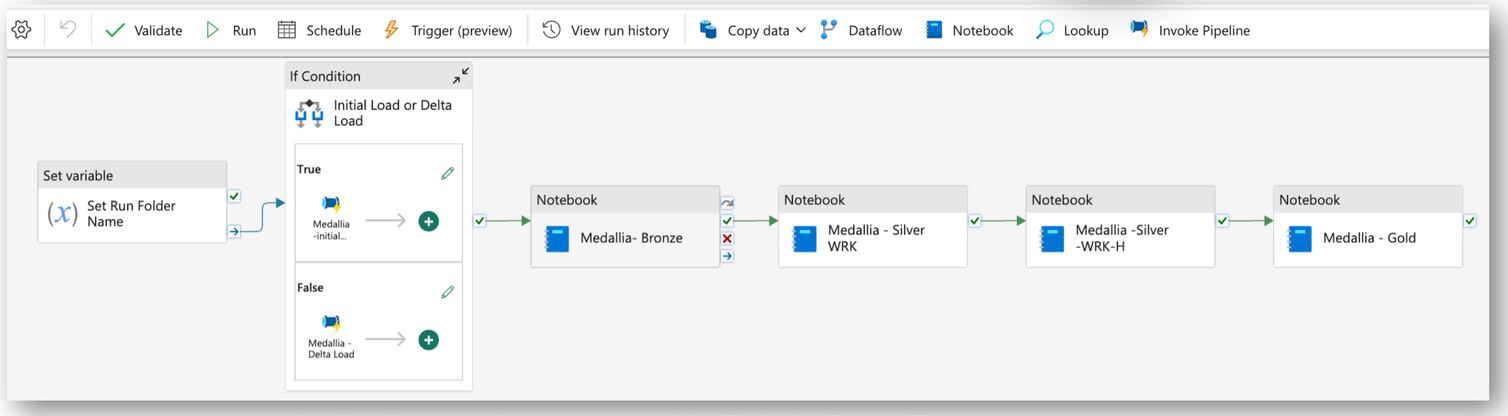


Silver layer — in same Fabric lakehouse

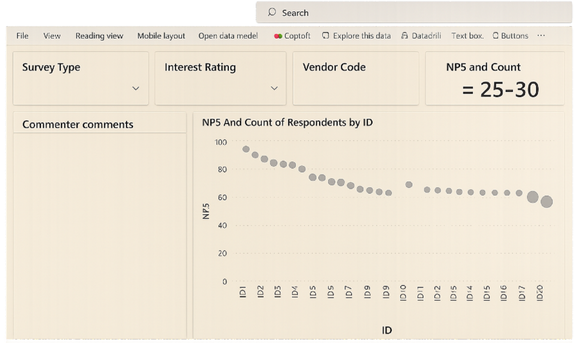
Medallion Gold: Semantic Model & Orchestration

Data Modeling

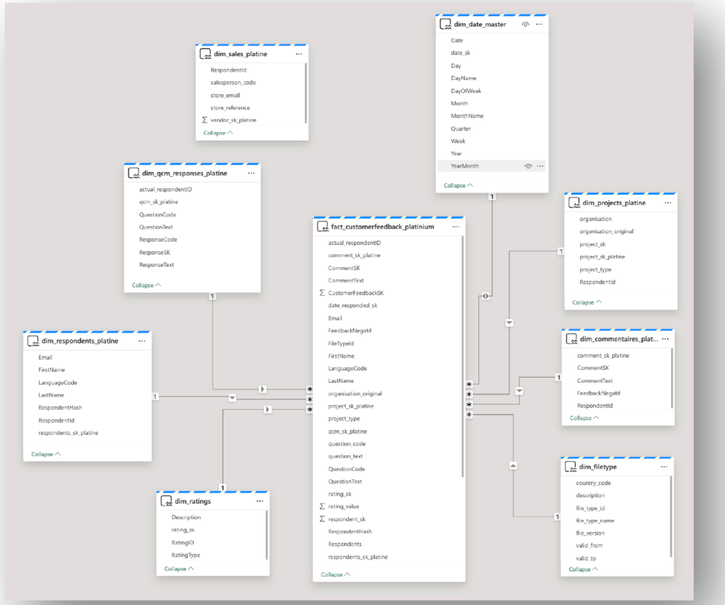
- - Denormalization
- - Model choice → Star Schema (in Gold layer pyspark Notebooks)
- - Data quality validated → test notebooks
- - DAX & measures → KPIs (NPS, churn)
- - Direct usability → Gold notebooks produced for BI layer
- - Data validation with business stakeholders



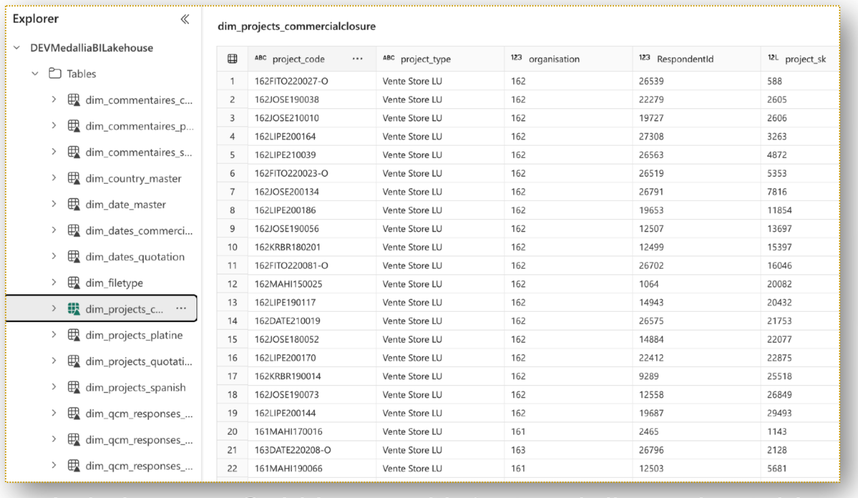
A pipeline orchestrator with Fabric Data Factory



(Redacted) sample of Power BI dashboards



A Star-Schema modeled *Gold layer Semantic Model*



Lakehouse - Gold Layer with fact and dimension tables

Summary of Results

- *- Data cleaning & analysis*
- *- Integration pipelines (Belgium, Spain)*
- *- Implementation of the Medallion architecture*
- *- Optimization: deployment of **Medallion Plus***
- *- Unit testing and SQL Endpoint validation (MS Fabric Lakehouse)*
- *- Gold **data modeling** via star schema*
- *- Implementation of centralized Gold+ layer (Platinum)*
- *- Data validation with business stakeholders (Finance, B.As, etc.)*
- *- Deployment → UAT validated by Business and technical analysts*

Thank you 😊