Proof of Work

A Cloud Migration to Microsoft Fabric

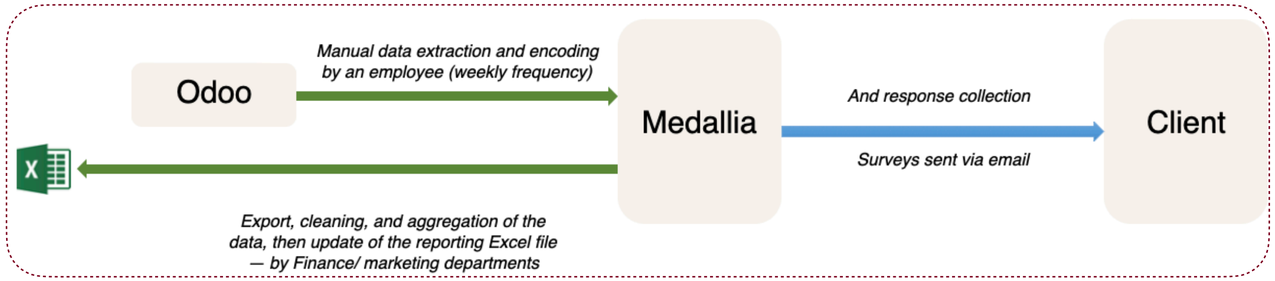
Auteur : **EL KAOUNI Yasser**

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)

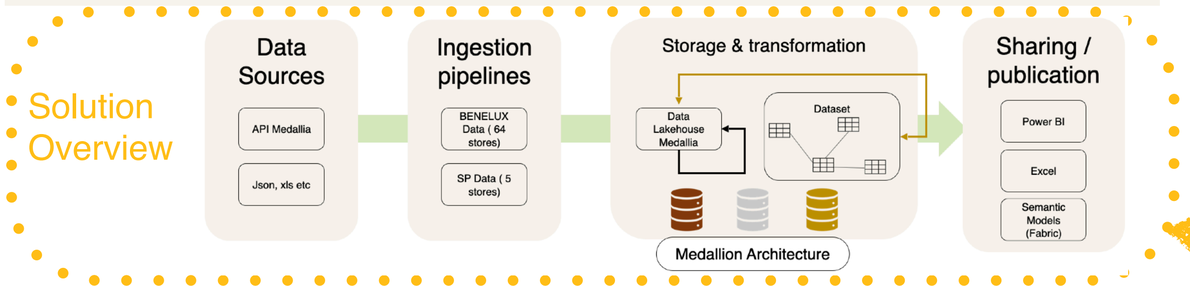


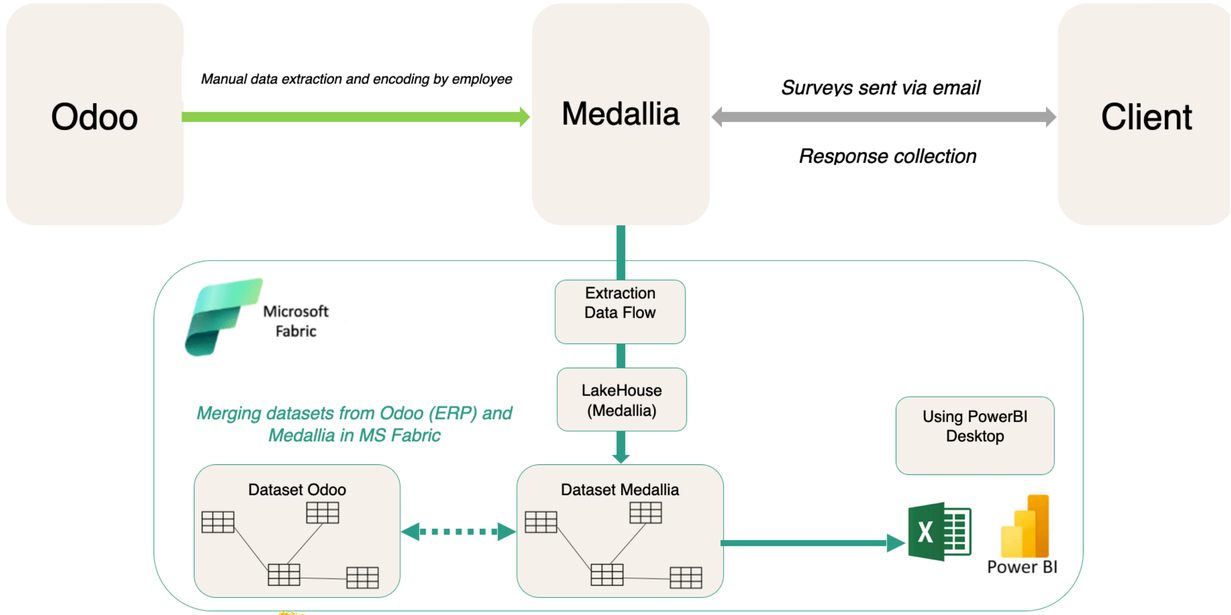
II. Business Problem

- 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

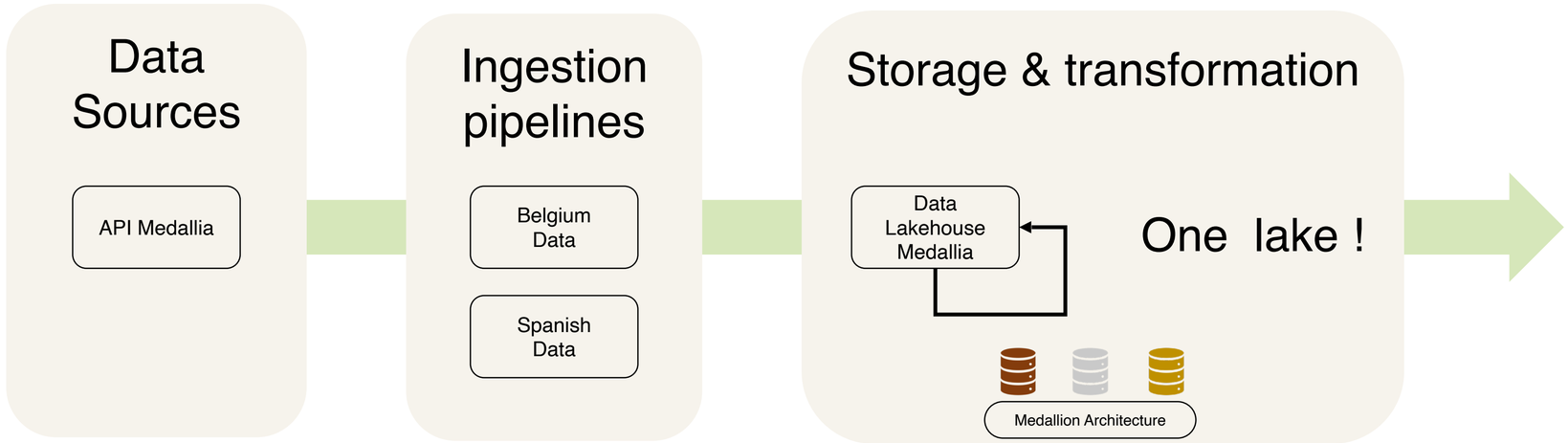
III. My Role (Data Engineer)

- 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**



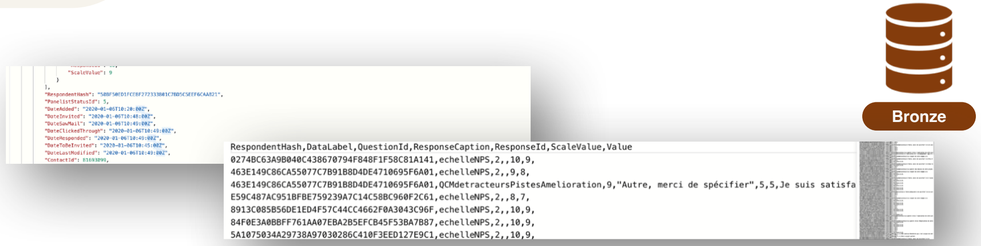


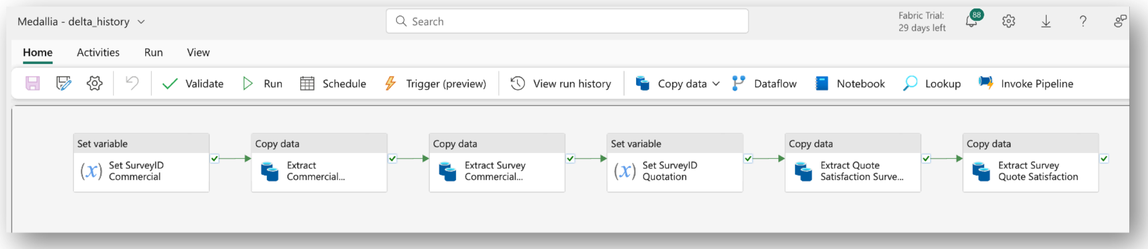
Medallion Architecture: Ingestion & Bronze Layer



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



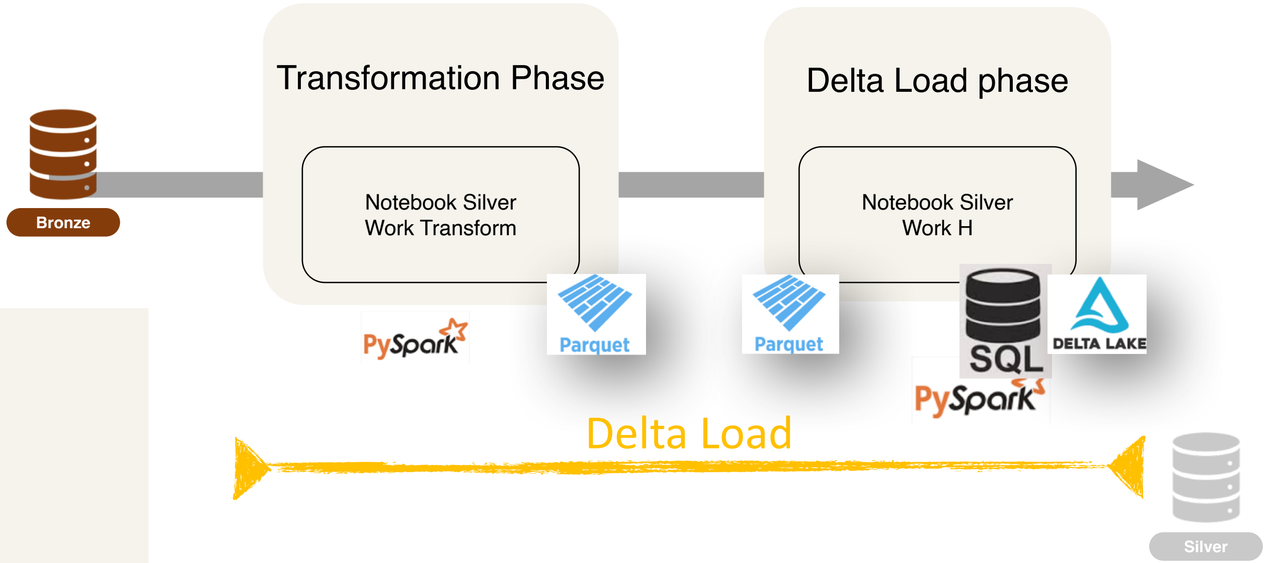


An example ingestion pipeline to the Bronze layer -with Fabric Data Factory

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)



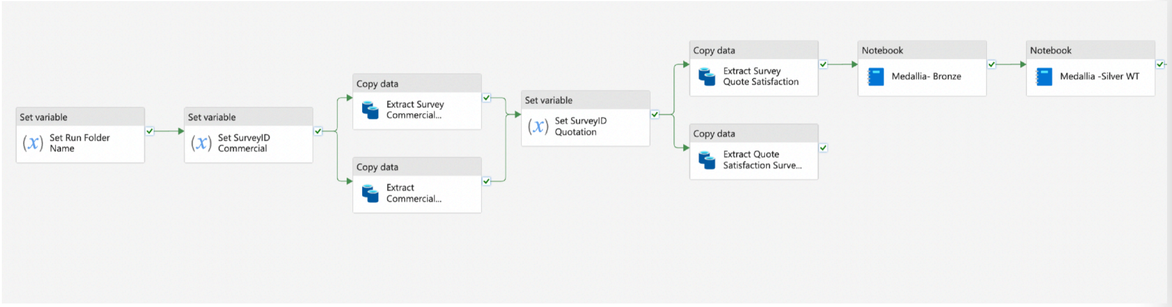
Optimization Phase

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

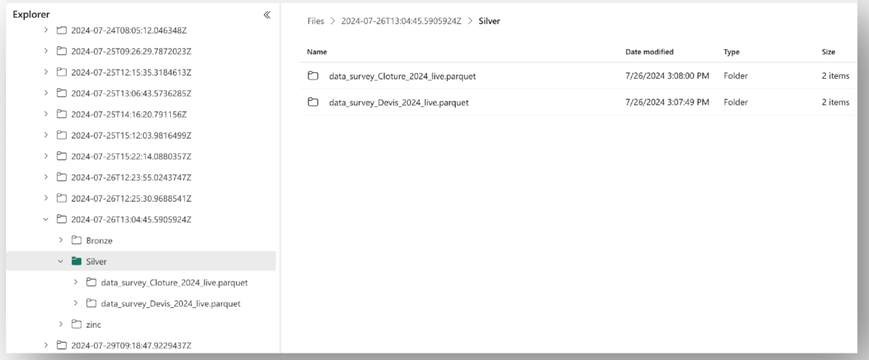
Delta Load technical details

- 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

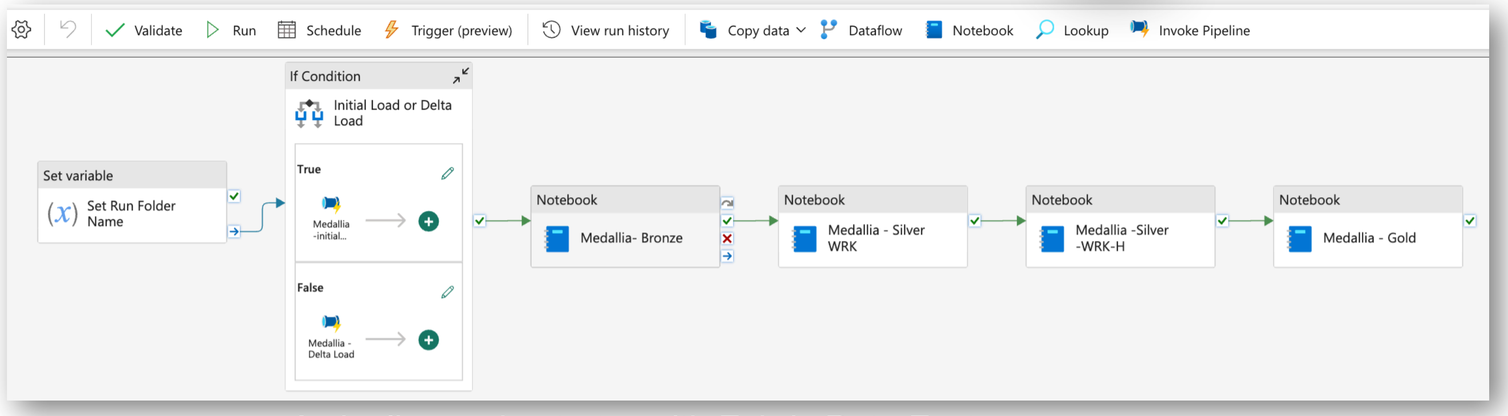


Silver layer — in same Fabric lakehouse

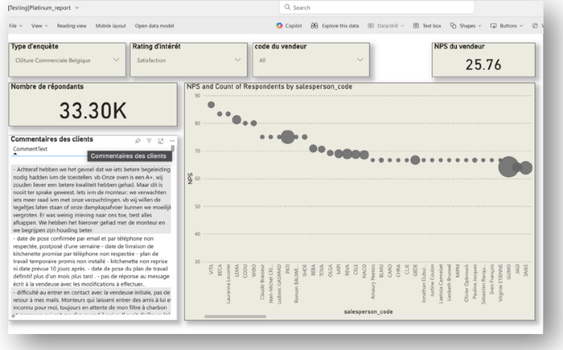
Medallion Gold: Semantic Model & Orchestration

Data Modeling

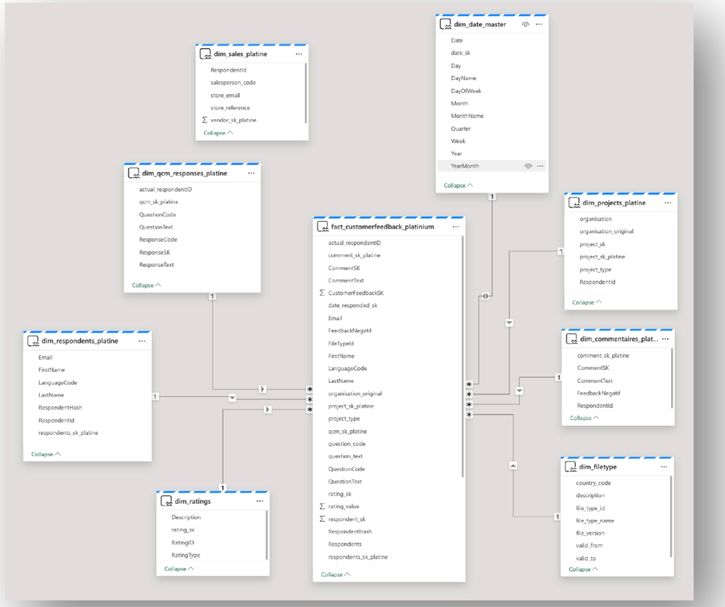
- - Denormalization
- - Model choice → Star Schema (in Gold layer pyspark Notebooks)
- - Data quality validated → test notebooks
- - DAX & measures → KPIs (NPS, churn)
- - Direct usability → Gold notebooks produced for BI layer
- - Data validation with business stakeholders



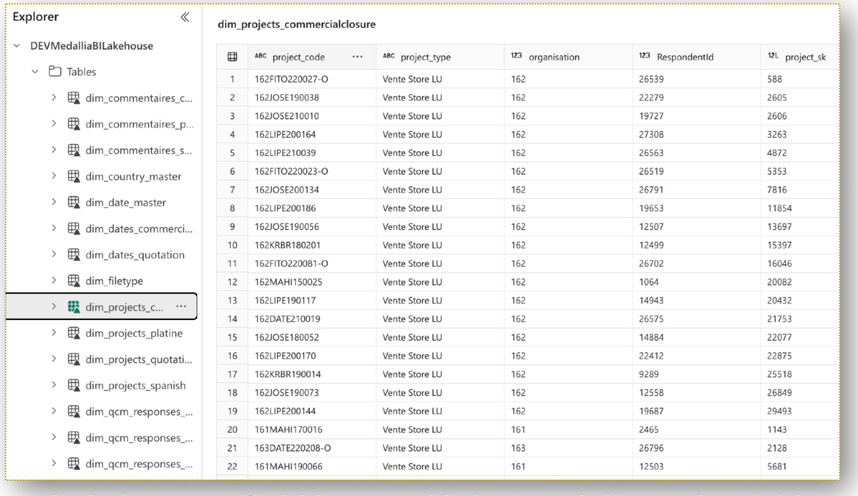
A pipeline orchestrator with Fabric Data Factory



One of Power BI dashboards made



A Star-Schema modeled *Gold layer Semantic Model*



	project_code	project_type	organisation	RespondentId	project_sk
1	162FITO220027-O	Vente Store LU	162	26539	588
2	162JOSE190038	Vente Store LU	162	22279	2605
3	162JOSE210010	Vente Store LU	162	19727	2606
4	162LIPE200164	Vente Store LU	162	27308	3263
5	162LIPE210039	Vente Store LU	162	26563	4872
6	162FITO220023-O	Vente Store LU	162	26519	5353
7	162JOSE200134	Vente Store LU	162	26791	7816
8	162LIPE200186	Vente Store LU	162	19653	11854
9	162JOSE190056	Vente Store LU	162	12507	13697
10	162KRBR180201	Vente Store LU	162	12499	15397
11	162FITO220081-O	Vente Store LU	162	26702	16046
12	162MAHI150025	Vente Store LU	162	1064	20082
13	162LIPE190117	Vente Store LU	162	14943	20432
14	162DATE210019	Vente Store LU	162	26575	21753
15	162JOSE180052	Vente Store LU	162	14884	22077
16	162LIPE200170	Vente Store LU	162	22412	22875
17	162KRBR190014	Vente Store LU	162	9289	25518
18	162JOSE190073	Vente Store LU	162	12558	26849
19	162LIPE200144	Vente Store LU	162	19687	29493
20	161MAHI170016	Vente Store LU	161	2465	1143
21	163DATE220208-O	Vente Store LU	163	26796	2128
22	161MAHI190066	Vente Store LU	161	12503	5681

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 😊