Data Platform Modernisation & Migration for the Department of Human Settlements

Prepared by: Rifumo Empowerment Holdings

Date: 29 September 2025

Table of Contents

1. Executive Summary

This proposal presents a comprehensive, secure and scalable data platform modernisation and migration programme tailored to the mandate of the Department of Human Settlements (DHS). The programme consolidates fragmented housing, beneficiary, project, land, and settlement data into a governed Microsoft Fabric platform backed by OneLake. It enables real-time analytics for programmes such as Breaking New Ground (BNG), Informal Settlements Upgrading, Social Housing, Community Residential Units (CRU), and First Home Finance (FLISP). The result is a single source of truth for planning, allocation, delivery tracking, compliance reporting, and spatial decision-making.

Key Outcomes:

- A unified Human Settlements Data Lakehouse (OneLake) with governed domains (Beneficiaries, Projects, Finance, Land & GIS, Contractors, Compliance).
- End-to-end data ingestion and quality pipelines from core systems (e.g., HSS, SHRA MIS, HDA Land Bank, municipal billing/ERP, provincial systems).
- High-performance semantic models and dashboards in Power BI (Direct Lake) for delivery tracking, subsidy management, budget burn, and backlog eradication KPIs.
- Automated statutory and management reporting with data lineage and auditing.
- POPIA-aligned security, role-based access, and data protection controls.
- A pragmatic migration roadmap using a domain-by-domain (Strangler) approach to minimise risk and ensure continuity.

1.1 Scope of Work

- Discovery & Roadmap: system inventory, data profiling, dependency mapping, migration backlog and business case.

- Fabric Landing Zone: workspaces (Dev/Test/Prod), OneLake, RBAC, Managed Identities, Key Vault, CI/CD.

- Data Integration: batch + streaming ingestion, data quality rules, metadata, observability.

- Semantic Models & Reporting: Direct Lake datasets, RLS/OLS, curated dashboards & reports.

- Governance & Security: Purview catalog + lineage, sensitivity labels, retention and audit.

- Cutover & Hypercare: parallel run, reconciliation, go-live, decommission legacy, KT & runbooks.

2. Solution Overview

2.1 Architecture at a Glance

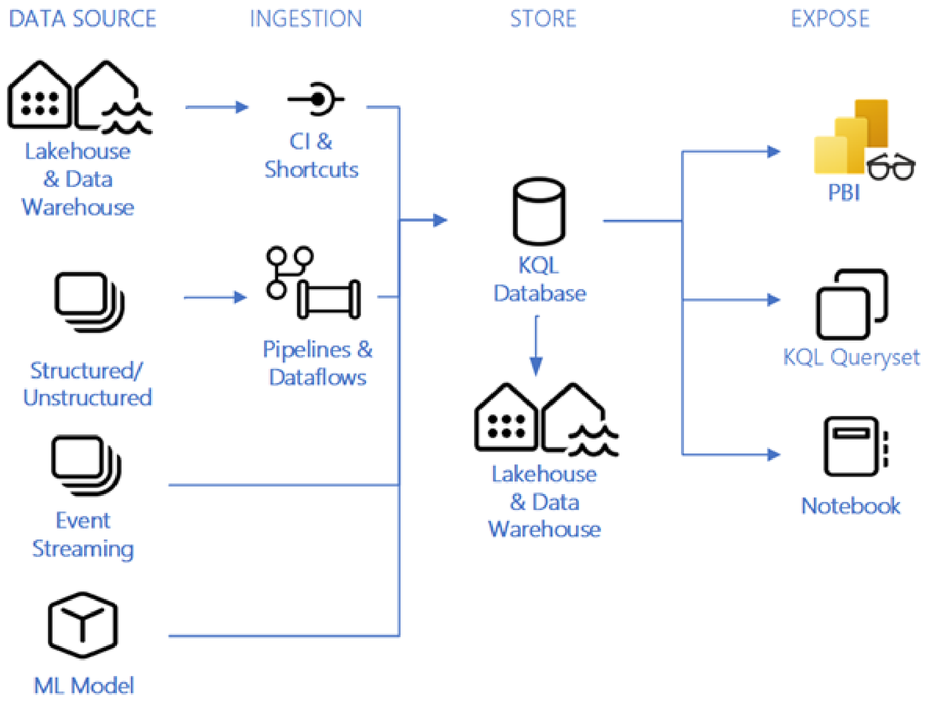
The solution uses Microsoft Fabric as the unified analytics platform with OneLake as the single logical data lake. Data is ingested via Fabric Data Factory into a medallion lakehouse (Bronze/Silver/Gold). Curated Gold datasets are served to Power BI in Direct Lake mode for sub-second interactivity without heavy refresh cycles. Orchestration, monitoring and CI/CD ensure reliability and repeatability across environments.

Data Integration and Dashboards

The architecture leverages Microsoft Fabric as the foundational platform, integrated with Databricks for advanced data engineering, machine learning, and real-time analytics, along with dedicated workspaces to optimize collaboration, management, and data governance.

Layered Architecture Overview

The architecture will be divided into several layers to ensure scalability, security, and seamless integration:



Data Ingestion Layer

- **Sources:** Diverse data sources, including on-premise databases (e.g., Oracle, Informix, SQL Server), cloud-based services, streaming platforms, and APIs.
- **Tools:** Microsoft Fabric (Data Factory) for automated data pipelines and Databricks for real-time streaming ingestion (Kafka or Event Hubs).
- **Capabilities:** Supports both batch and streaming ingestion to accommodate diverse use cases.

Data Lakehouse Storage Layer

- **Microsoft Fabric OneLake:** Acts as the unified storage layer for structured, semi-structured, and unstructured data.
- **Databricks Delta Lake:** Integrated with OneLake for efficient storage management with features like ACID transactions and schema enforcement.
- **Partitioning and Indexing:** Optimized data storage using partitioning and indexing techniques to enhance query performance.

Data Processing and Transformation Layer

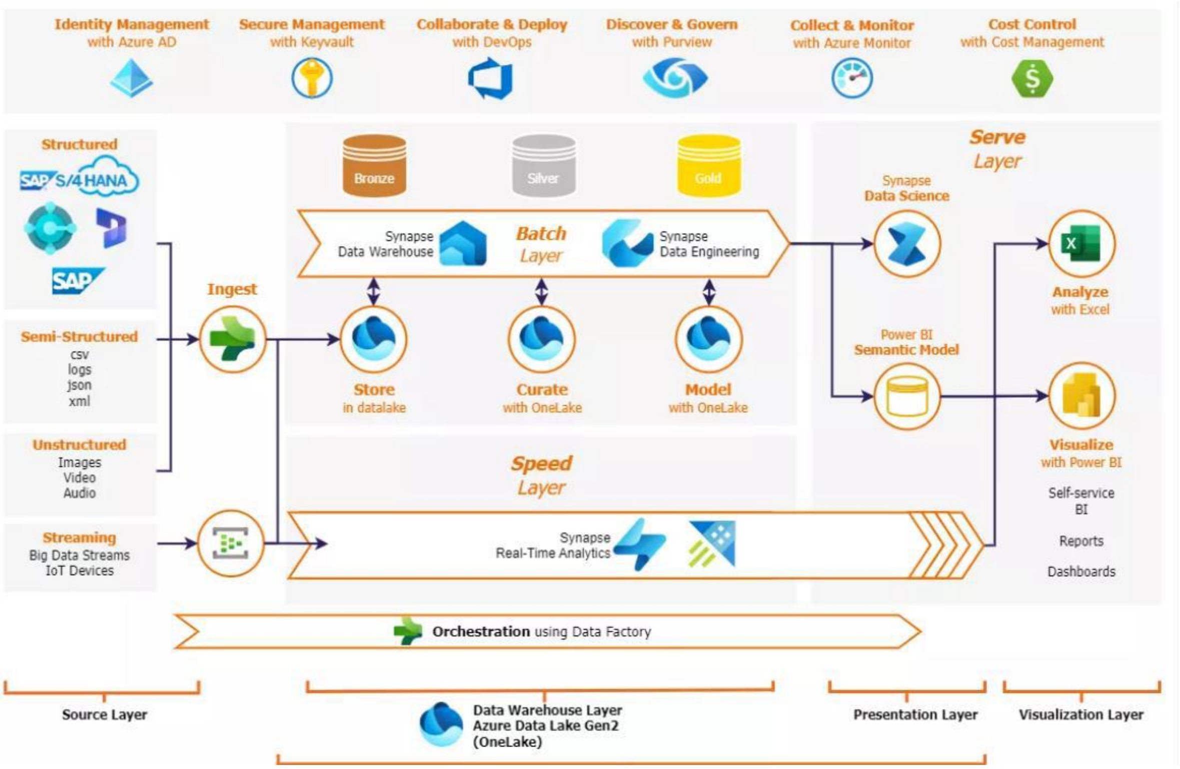
- **Databricks Workspaces:** Dedicated environments for ETL development, data engineering, and machine learning model training.
- **Microsoft Fabric Synapse Analytics:** For big data processing, SQL-based analytics, and advanced data transformations.
- **Orchestration Tools:** Integration of Databricks workflows and Fabric pipelines for automation and coordination of complex data processes.

Data Governance Layer

- **Microsoft Fabric Compliance Center:** Provides centralized governance with role-based access control (RBAC), encryption, and regulatory compliance (e.g., POPIA).
- **Databricks Unity Catalog:** Ensures consistent metadata management and fine-grained security policies across Databricks environments.
- **Auditing and Monitoring:** Tools like Microsoft Purview and Databricks monitoring features for enhanced governance and observability.

Analytics and Visualization Layer

- **Microsoft Power BI:** Integrated with OneLake and Synapse Analytics for real-time data visualization, reporting, and dashboard creation.
- **Databricks SQL:** Enables direct querying of Delta Lake for analytics use cases requiring high performance and scalability.
- **Real-Time Analytics:** Databricks facilitates streaming analytics by processing data in near real-time and rendering insights via Power BI.



Key Components

The following components will play a pivotal role in the architecture:

Microsoft Fabric

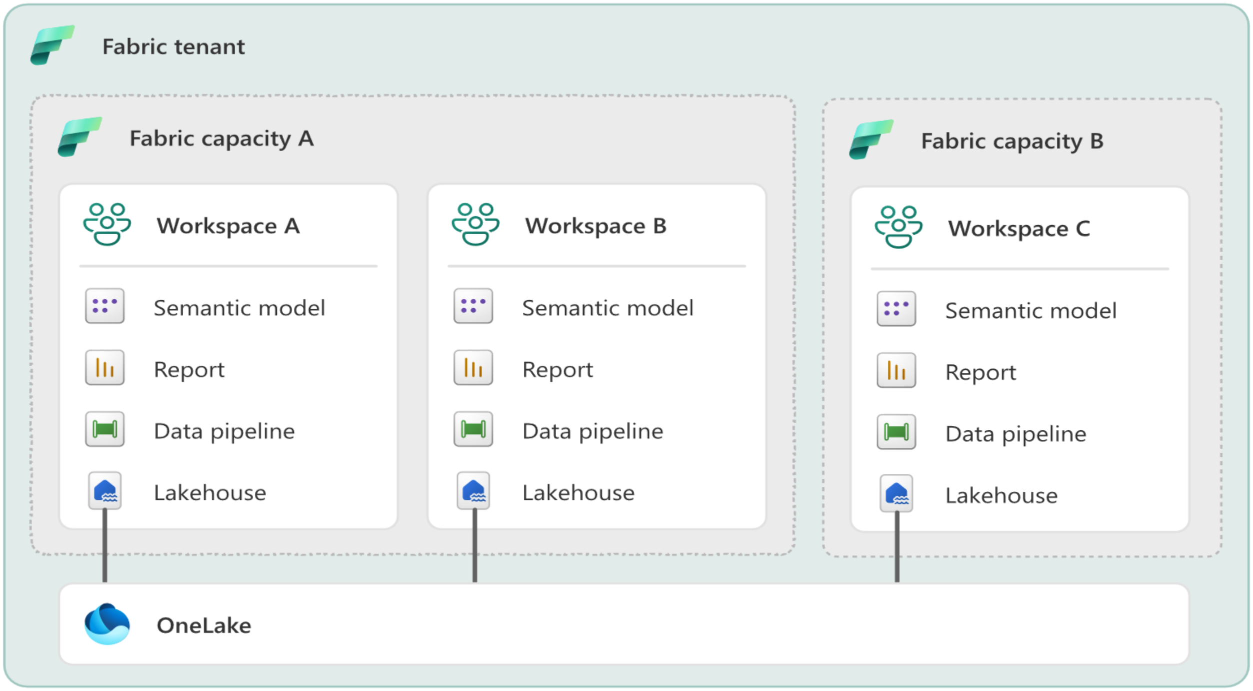
- **Data Factory:** Automates data ingestion and ETL tasks across diverse sources.
- **OneLake:** Serves as the central storage hub, enabling seamless data management and universal access.
- **Synapse Analytics:** Supports complex data processing and analytical workloads.
- **Power BI:** Provides rich visualization and reporting tools tailored for end users.

Databricks

- **Delta Lake:** Ensures efficient, reliable, and scalable storage of data.
- **Workspaces:** Offers isolated environments for data engineers and scientists to collaborate on data pipelines, models, and experiments.
- **Streaming Frameworks:** Real-time data ingestion using Kafka, Event Hubs, or Delta Live Tables.

Workspaces

- **Dedicated Databricks Workspaces:** Create specialized environments for machine learning, ETL processes, and data analytics.
- **Microsoft Fabric Workspaces:** Allow teams to manage datasets, analytics reports, and shared dashboards effectively.



Workflow Example

Step 1: Data Ingestion

- Raw data is ingested into the architecture via Microsoft Fabric's Data Factory pipelines (batch data) or Databricks streaming mechanisms (real-time data).

Step 2: Storage

- Ingested data is stored in Fabric OneLake and Databricks Delta Lake. Metadata is captured in Databricks Unity Catalog for governance.

Step 3: Processing

- Data engineers use Databricks Workspaces for ETL operations and transformation, with workflows orchestrated through Microsoft Fabric pipelines.

Step 4: Analytics

- Transformed data is made available to Power BI for visualization and reporting. Analysts access data directly via Databricks SQL for ad-hoc queries.

Step 5: Governance and Security

- Role-based access control (RBAC), auditing, and encryption are enforced by Microsoft Fabric Compliance Center and Databricks Unity Catalog.
- 9.Benefits of This Architecture
- **Unified Platform:** Integrates Microsoft Fabric and Databricks seamlessly for streamlined data management.
- **Scalability:** Accommodates large datasets and diverse analytics requirements.
- **Flexibility:** Offers dedicated workspaces for specialized tasks, improving collaboration and productivity.
- **Real-Time Capabilities:** Enables near real-time data processing and analytics.
- **Security and Governance:** Ensures robust protection and compliance with industry standards.

Methodology and Project Approach: Data Platform Modernization

Phase-Wise Implementation Plan

Each phase lays the foundation for subsequent stages. The roadmap ensures technical maturity and adoption readiness.

Phase 1: Discovery & Assessment

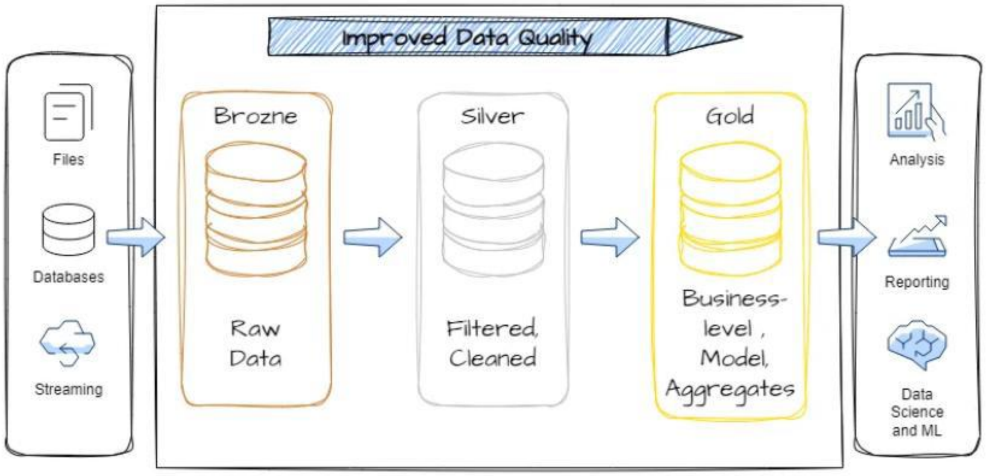
- Inventory of legacy and current systems.
- Data profiling, quality checks, and lineage mapping.
- Catalog metadata and identify integration touchpoints.

Phase 2: Platform Design and Architecture

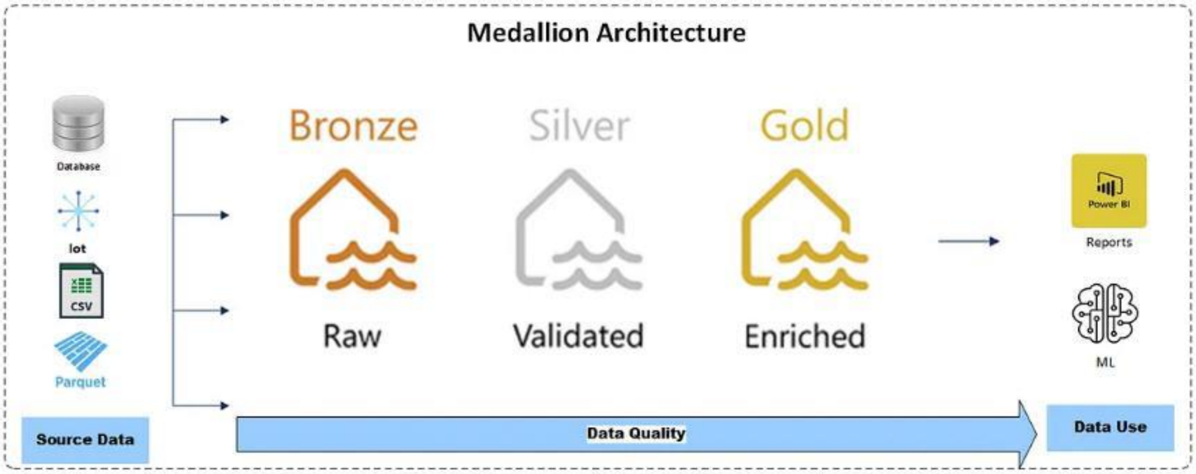
- Deploy Microsoft Fabric with OneLake, Dataflows Gen2, and Spark notebooks.
- Implement Medallion architecture:
    - Bronze: Raw data ingestion.
    - Silver: Cleansed, conformed data.
    - Gold: Business-ready models for reporting.

Phase 3: Data Integration and Pipeline Engineering

- Use Fabric Pipelines to ingest legacy data and automate ETL.
- Define pipeline schedules, error handling, and dependency logic.
- Integrate batch and real-time ingestion patterns.



Medallion Architecture



Medallion Architecture Overview (Fabric Lakehouse)

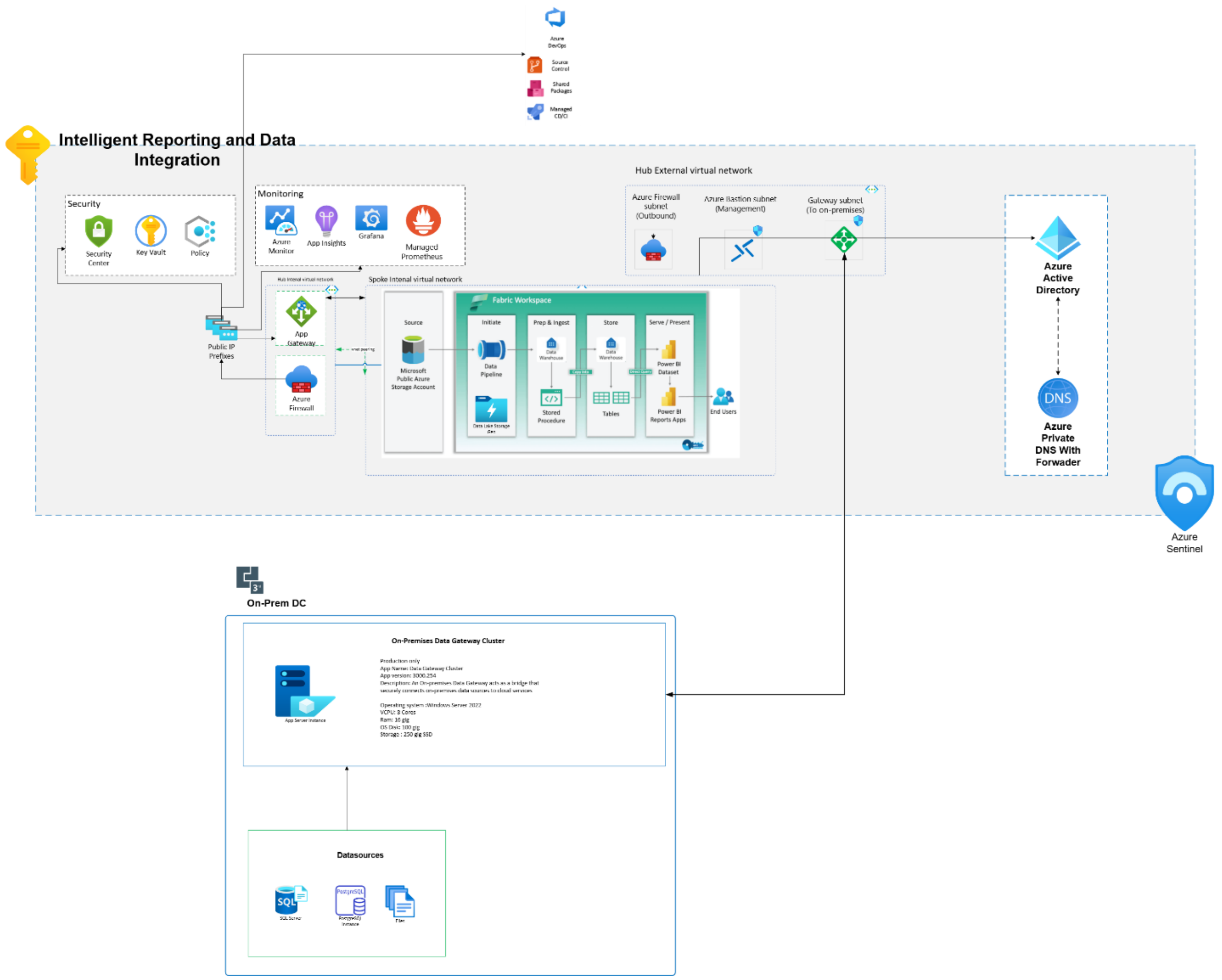
Layer	Purpose	Key Tools in Fabric
Bronze	Raw data ingestion and storage	Dataflows Gen2, Data Pipelines, OneLake
Silver	Cleaned and transformed business data	Lakehouse SQL, Notebooks, Spark
Gold	Business-ready, aggregated/curated datasets	Lakehouse SQL, Power BI, Warehouse

Data Gateway Design

Secure and scalable gateway infrastructure is critical to hybrid integration.

Design Principles:

- Implement Microsoft On-premises Data Gateway for SAP
- Isolate gateway configuration per environment (Dev/Test/Prod).
- Monitor availability and performance using gateway metrics.



Data Cleansing, Standardization, and Verification

Maintaining data quality is essential to usability and trust.

Techniques Used:

- Referential Integrity Checks to validate relationships between entities.
- Rule-based Validation against business logic
- Duplicate resolution and enrichment routines.
- Anomaly detection using AI modelling within Fabric notebooks.

Data Governance and Security Strategy

Compliance and control form the backbone of enterprise-grade modernization.

Governance Tactics:

- Implement Microsoft Purview for metadata management and lineage.
- Apply sensitivity labels and RBAC across workspaces.
- Utilize Azure Key Vault for secure credential and key storage.
- Enable audit logging for traceability and monitoring.

Workspace Design and Semantic Modelling

Modular workspace design allows structured development and controlled access.

Design Elements:

- Create separate workspaces by data domain.
- Enforce naming conventions and role-based access.
- Define semantic models with calculated measures, KPIs, and hierarchies.
- Use deployment pipelines for lifecycle promotion from Dev to Prod.

Real-Time Reporting Architecture Using Power BI

Responsive, self-service analytics empower users across the organization.

Architecture Overview:

- Direct Lake Mode ensures high-speed dashboard rendering.
- Build composite models for learner, financial, and operational domains.
- Embed dashboards into existing portals for seamless user experience.
- Configure filters, drill-through, and navigation for decision support.

- Training, Support, and Change Enablement

Sustainable change requires education, documentation, and ongoing engagement.

Approach to Enablement:

- Design targeted training tracks for analysts, managers, and IT staff.
- Provide embedded help features and glossary terms within dashboards.
- Collect feedback for iterative improvement.
- Schedule quarterly adoption review sessions.

- Performance Optimization and Monitoring

Ongoing measurement ensures platform health and data reliability.

Monitoring Strategies:

- Use Fabric’s built-in metrics for pipeline performance.
- Track dashboard load times and user engagement.
- Implement alerting for data failures and ingestion anomalies.
- Define thresholds for data freshness and reporting lag.

Migration Approach & Methodology

We deliver in two-week sprints using a Strangler-Fig approach: wrap legacy endpoints, build new pipelines and models in Fabric, run in parallel for reconciliation, then switch by domain. Continuous testing, security, and observability are applied throughout.

SSIS → Fabric Patterns

Control flows → Fabric Data Factory pipelines; data flows → Dataflows Gen2 or Notebooks.

Externalize configs and secrets to Key Vault; parameterize per environment (Dev/Test/Prod).

Observability via Fabric monitoring, Application Insights custom events, and alerts; robust retries and compensation.

SSAS → Fabric Semantic Models

Inventory measures, partitions, and roles; convert model metadata (TMDL/Tabular Editor).

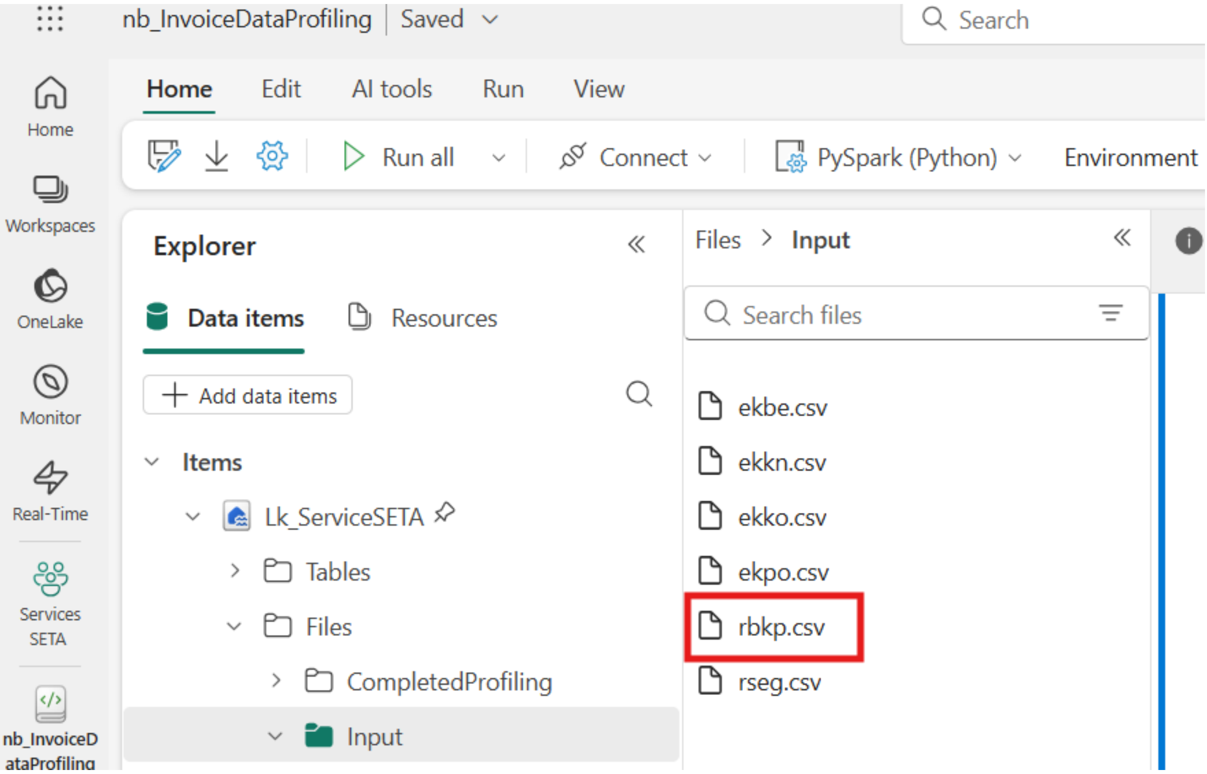
Optimize for Direct Lake (star schemas, aggregations, partition pruning).

Implement RLS/OLS; validate DAX parity and calculation accuracy with test packs.

Demo

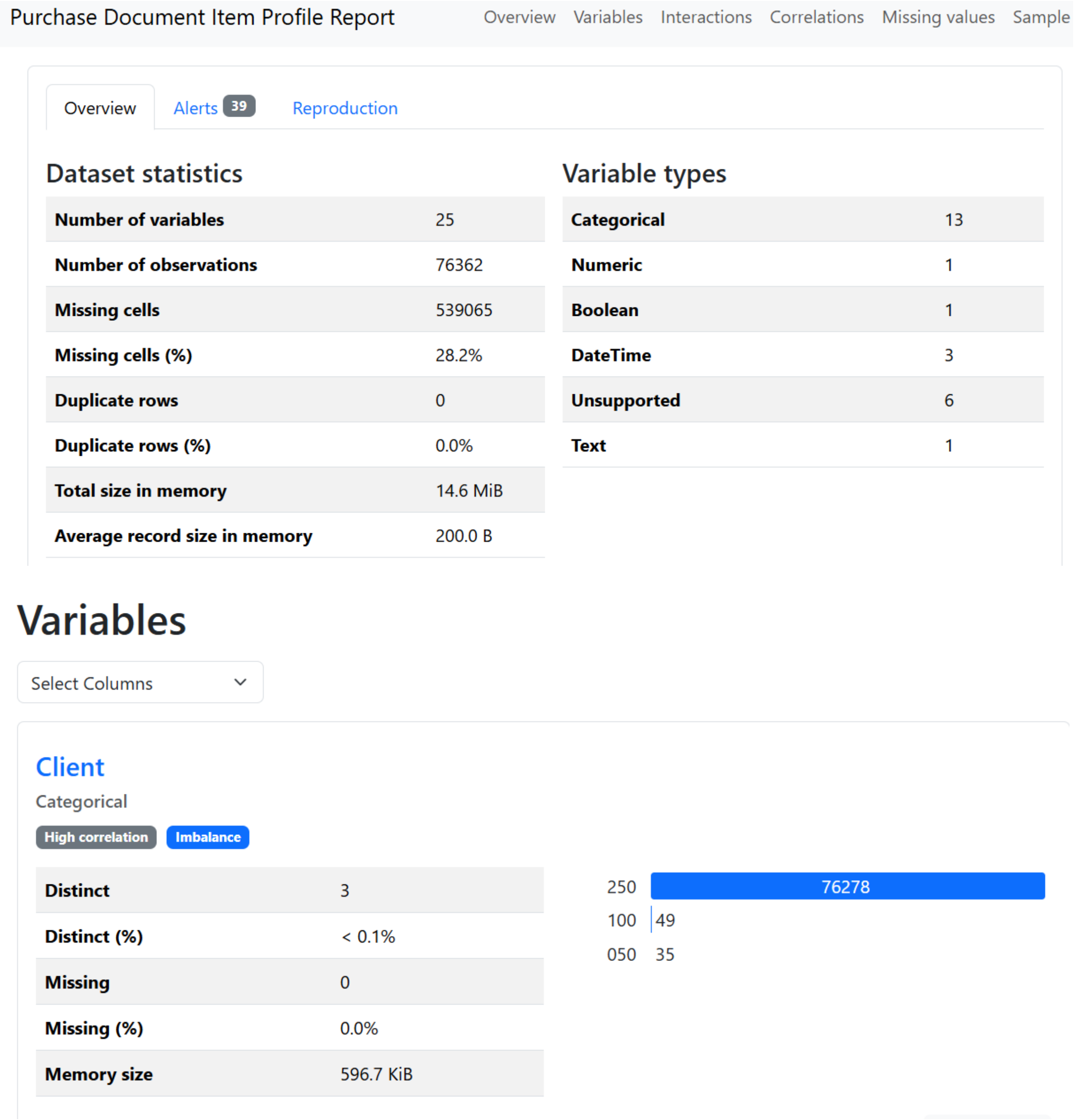
Bronze Layer

This layer is taking source data from legacy system (e.g Oracle) into the datalake (first layer of medallion). See below the RBKP file dropped into the datalake.



Data Profiling

Using Python Spark tool we create data profiling results for different source files. The results are stored in the OneLake folder as html files. See below examples:



Purchase Document Item Profile Report

Dataset statistics

Number of variables	25
Number of observations	76362
Missing cells	539065
Missing cells (%)	28.2%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	14.6 MiB
Average record size in memory	200.0 B

Variable types

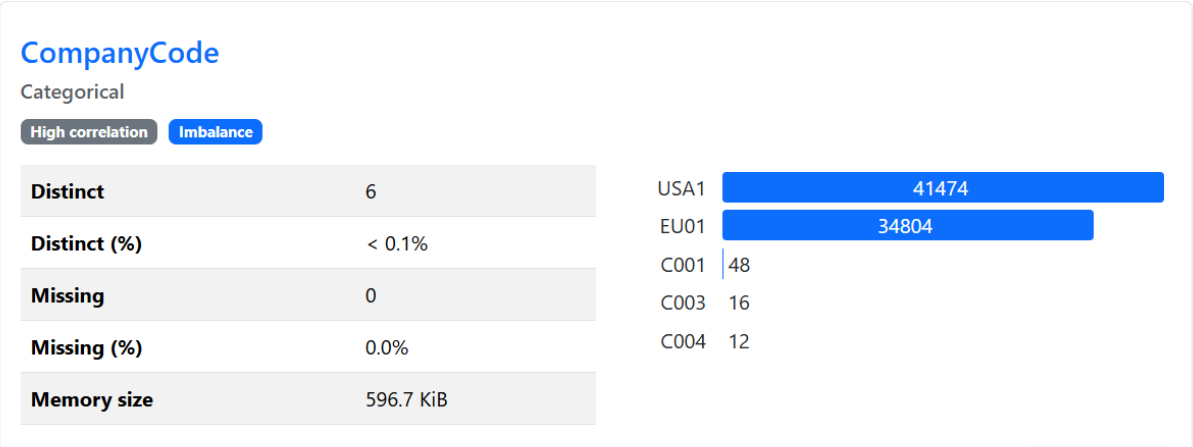
Categorical	13
Numeric	1
Boolean	1
DateTime	3
Unsupported	6
Text	1

Variables

Client

Categorical

Distinct	3
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	596.7 KiB



CompanyCode

Categorical

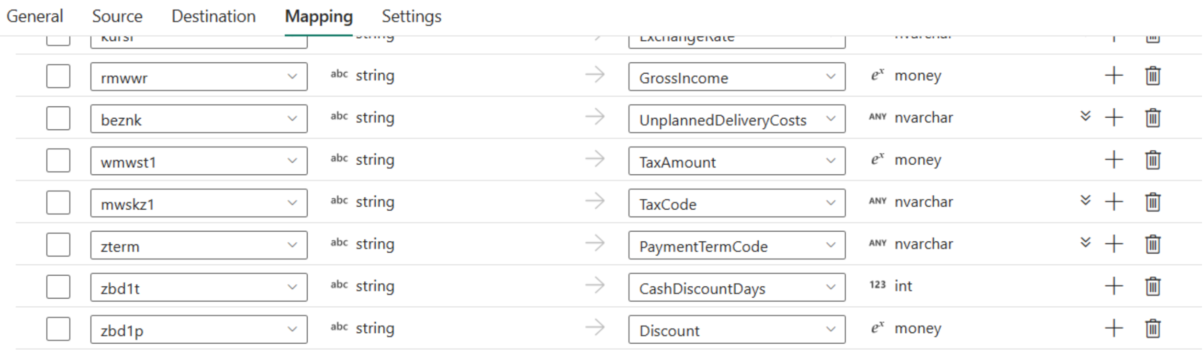
High correlation Imbalance

Distinct	6
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	596.7 KiB

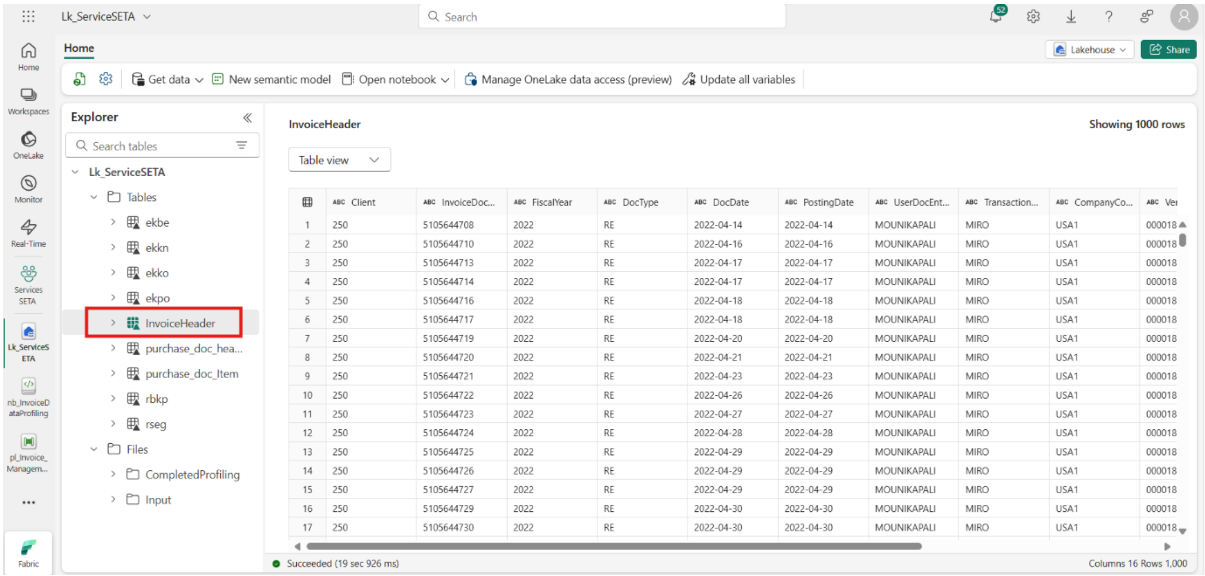
USA1	41474
EU01	34804
C001	48
C003	16
C004	12

Silver Medallion Layer

After applying any business rules, mapping, and data cleanup using stored procedures based on data profiling and data dictionary, the files are stored as tables in the database, see below:



□	kursf	abc string	→	ExchangeRate	ANY nvarchar	
□	rmwwr	abc string	→	GrossIncome	money	
□	beznk	abc string	→	UnplannedDeliveryCosts	ANY nvarchar	
□	wmwst1	abc string	→	TaxAmount	money	
□	mwskz1	abc string	→	TaxCode	ANY nvarchar	
□	zterm	abc string	→	PaymentTermCode	ANY nvarchar	
□	zbd1t	abc string	→	CashDiscountDays	123 int	
□	zbd1p	abc string	→	Discount	money	



InvoiceHeader

Showing 1000 rows

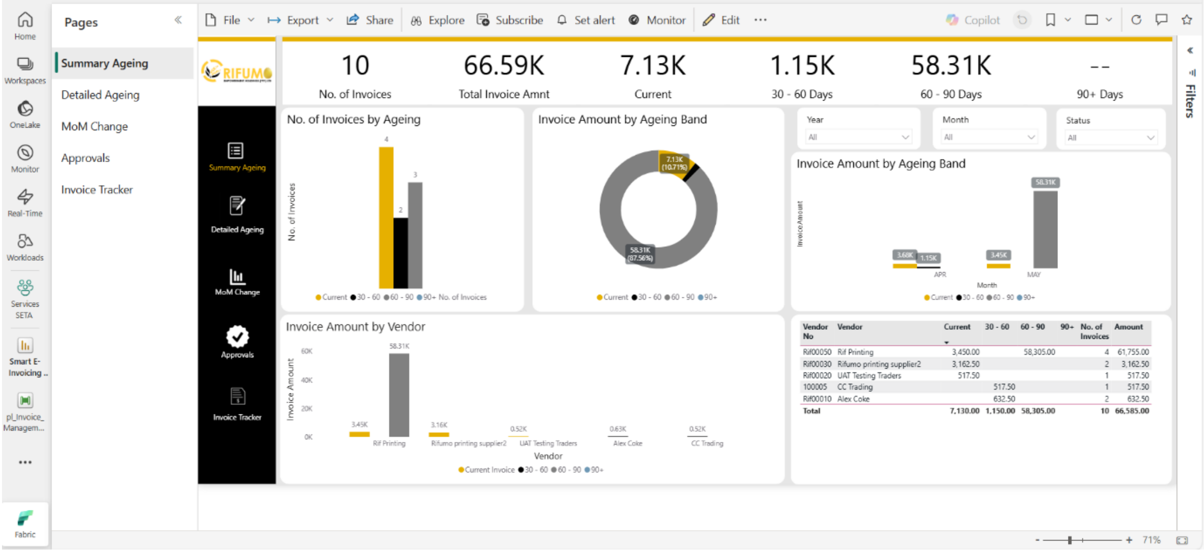
	Client	InvoiceDoc...	FiscalYear	DocType	DocDate	PostingDate	UserDocEnt...	Transaction...	CompanyCo...	Ver
1	250	5105644708	2022	RE	2022-04-14	2022-04-14	MOUNIKAPALI	MIRO	USA1	000018
2	250	5105644710	2022	RE	2022-04-16	2022-04-16	MOUNIKAPALI	MIRO	USA1	000018
3	250	5105644713	2022	RE	2022-04-17	2022-04-17	MOUNIKAPALI	MIRO	USA1	000018
4	250	5105644714	2022	RE	2022-04-17	2022-04-17	MOUNIKAPALI	MIRO	USA1	000018
5	250	5105644716	2022	RE	2022-04-18	2022-04-18	MOUNIKAPALI	MIRO	USA1	000018
6	250	5105644717	2022	RE	2022-04-18	2022-04-18	MOUNIKAPALI	MIRO	USA1	000018
7	250	5105644719	2022	RE	2022-04-20	2022-04-20	MOUNIKAPALI	MIRO	USA1	000018
8	250	5105644720	2022	RE	2022-04-21	2022-04-21	MOUNIKAPALI	MIRO	USA1	000018
9	250	5105644721	2022	RE	2022-04-23	2022-04-23	MOUNIKAPALI	MIRO	USA1	000018
10	250	5105644722	2022	RE	2022-04-26	2022-04-26	MOUNIKAPALI	MIRO	USA1	000018
11	250	5105644723	2022	RE	2022-04-27	2022-04-27	MOUNIKAPALI	MIRO	USA1	000018
12	250	5105644724	2022	RE	2022-04-28	2022-04-28	MOUNIKAPALI	MIRO	USA1	000018
13	250	5105644725	2022	RE	2022-04-29	2022-04-29	MOUNIKAPALI	MIRO	USA1	000018
14	250	5105644726	2022	RE	2022-04-29	2022-04-29	MOUNIKAPALI	MIRO	USA1	000018
15	250	5105644727	2022	RE	2022-04-29	2022-04-29	MOUNIKAPALI	MIRO	USA1	000018
16	250	5105644729	2022	RE	2022-04-30	2022-04-30	MOUNIKAPALI	MIRO	USA1	000018
17	250	5105644730	2022	RE	2022-04-30	2022-04-30	MOUNIKAPALI	MIRO	USA1	000018

Succeeded (19 sec 926 ms)

Columns 16 Rows 1,000

Gold Medallion Layer

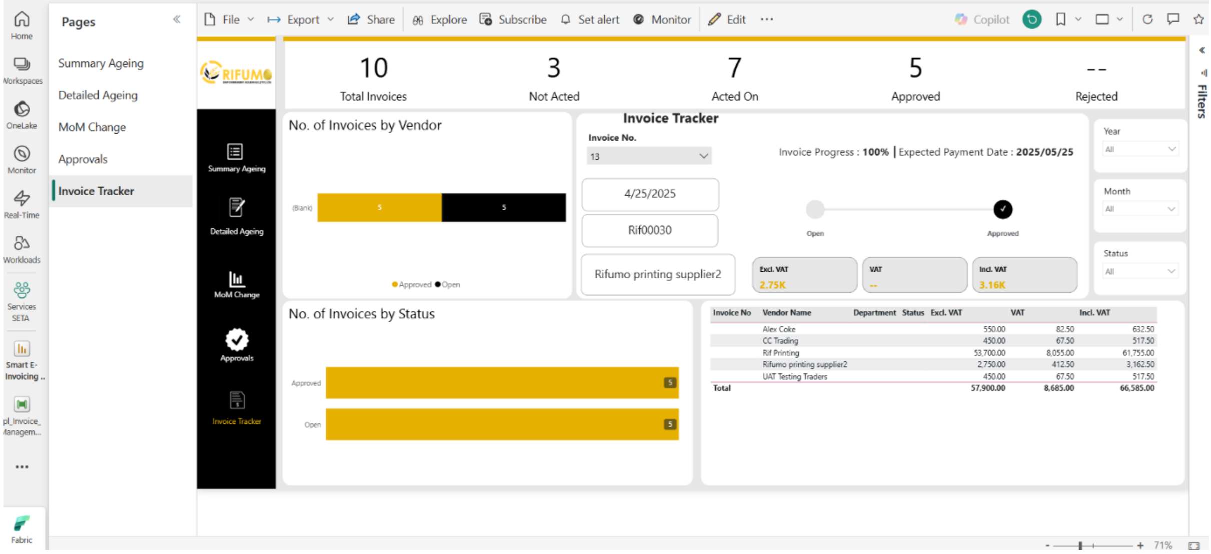
In the 3rd layer of the medallion architecture, the data is then reported using power BI which is integrated in the fabric workspace.



10	66.59K	7.13K	1.15K	58.31K	--
No. of Invoices	Total Invoice Amnt	Current	30 - 60 Days	60 - 90 Days	90+ Days

No. of Invoices by Ageing
4
2
3
No. of Invoices
● Current ● 30 - 60 ● 60 - 90 ● 90+ No. of Invoices
Invoice Amount by Ageing Band
7.13K (10.71%)
58.31K (87.56%)
● Current ● 30 - 60 ● 60 - 90 ● 90+
Year
All
Month
All
Status
All
Invoice Amount by Ageing Band
58.31K
3.68K
1.15K
3.45K
APR
MAY
Month
InvoiceAmount
● Current ● 30 - 60 ● 60 - 90 ● 90+
Invoice Amount by Vendor
60K
40K
20K
0K
Invoice Amount
58.31K
3.45K
3.16K
0.52K
0.63K
0.52K
Rif Printing
Rifumo printing supplier2
UAT Testing Traders
Alex Coke
CC Trading
Vendor
● Current Invoice ● 30 - 60 ● 60 - 90 ● 90+
71%

Vendor No	Vendor	Current	30 - 60	60 - 90	90+	No. of Invoices	Amount
Rif00050	Rif Printing	3,450.00		58,305.00		4	61,755.00
Rif00030	Rifumo printing supplier2	3,162.50				2	3,162.50
Rif00020	UAT Testing Traders	517.50				1	517.50
100005	CC Trading		517.50			1	517.50
Rif00010	Alex Coke		632.50			2	632.50
Total		**7,130.00**	**1,150.00**	**58,305.00**		**10**	**66,585.00**



10 Total Invoices | 3 Not Acted | 7 Acted On | 5 Approved | -- Rejected

Invoice Tracker

Invoice No.: 13

4/25/2025

Rif00030

Rifumo printing supplier2

Invoice Progress : **100%** | Expected Payment Date : **2025/05/25**

Excl. VAT: 2.75K

VAT: --

Incl. VAT: 3.16K

Year: All

Month: All

Status: All

Invoice No	Vendor Name	Department	Status	Excl. VAT	VAT	Incl. VAT
	Alex Coke			550.00	82.50	632.50
	CC Trading			450.00	67.50	517.50
	Rif Printing			53,700.00	8,055.00	61,755.00
	Rifumo printing supplier2			2,750.00	412.50	3,162.50
	UAT Testing Traders			450.00	67.50	517.50
Total				**57,900.00**	**8,685.00**	**66,585.00**

Data Handling and Protection

Data Obfuscation

Data obfuscation is the process of replacing sensitive information with data that looks like real production information, making it useless to malicious actors. It is primarily used in test or development environments—developers and testers need realistic data to build and test software, but they do not need to see the real data.

During the Customer Review Sessions of the Source-to-Target and Report Mappings recommendations for Sensitive and Private Data will be reviewed for Data Obfuscation Decisions.

There are three primary data obfuscation techniques:

I. **Masking-Out**

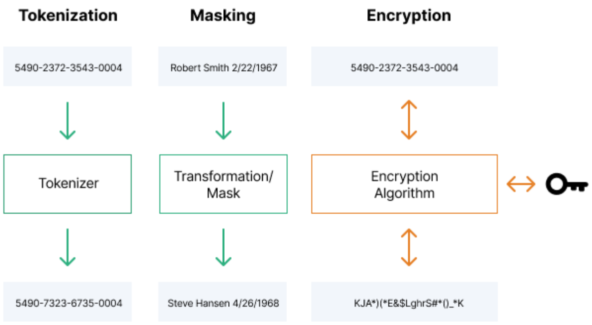
Is a way to create different versions of the data with a similar structure. The data type does not change, only the value change. Data can be modified in a number of ways, for example shifting numbers or letters, replacing words, and switching partial data between records.

II. Data encryption

Uses cryptographic methods, usually symmetric or private/pub key systems to codify the data, making it completely unusable until decrypted. Encryption is very secure, but when you encrypt your data, you cannot manipulate or analyze it.

III. Data tokenization

Replaces certain data with meaningless values. However, authorized users can connect the token to the original data. Token data can be used in production environments, for example, to execute financial transactions without the need to transmit a credit card number to an external processor.



Key reasons organizations rely on data obfuscation methods:

- **Third parties can't be trusted**—sending personal data, payment card information or health information to any third party is dangerous. There is a dual risk—it increases the number of people who have access to the data beyond the organization's control, and it exposes the organization to violations of regulations and standards.
- **Business operations may not need real data**—any use of customer, employee, or user data is risky because it exposes the data to employees, contractors, and others. Many business processes, such as development, testing, analytics, and reporting, do not necessarily need to process real personal data. By obfuscating the data, the organization can maintain the business process but eliminate the risk.
- **Compliance—**many compliance standards require data to be obfuscated under certain conditions.

Data Security

As the data being stored and presented is extremely sensitive role-based security has been requested for the Power BI data model.

- Role based security will be implemented on an organizational hierarchy perspective.
- Where if a user has access to the Head Office Layer, he/she would be able to see everything at head office and below,.
- Only users with Active Directory Accounts will be able to access the Power BI reports, and the organisation will need to provide a list of which user groups (Not specific individual users) need to have access at which levels.

Architecture Layers:

Layer	Purpose & Key Capabilities
Data Ingestion	Fabric Data Factory pipelines (batch/stream), on-premises data gateway, APIs; scheduler, retries, alerting.
Data Lakehouse (OneLake)	Single logical lake with Delta tables, domain-aligned folders, schema management, partitions and retention.
Processing & Quality	Spark notebooks/SQL for transformations, rule-based data validation, SCD handling, reconciliation.
Governance & Security	Purview catalog & lineage, Entra ID RBAC, sensitivity labels, Key Vault secrets, audit trails.
Analytics	Power BI semantic models (Direct Lake), RLS/OLS, KPIs and dashboards embedded in portals.

2.2 Methodology & Project Approach

Delivery follows a two-week sprint cadence and a Strangler pattern: wrap legacy interfaces, build new domain pipelines and semantic models in Fabric, run parallel for reconciliation, then cut over by domain. Continuous testing, security checks, and observability are applied from Sprint 2 onwards.

2.3 Human Settlements Priority Use Cases

- Beneficiary & Subsidy Management: end-to-end tracking across HSS/First Home Finance (FLISP), duplicate detection, and verification.

- Informal Settlements Upgrading (UISP): settlement profiling, bulk services progress, relocations and permanency register.

- Project & Delivery Management: pipeline status, budget burn (MTEF), contractor performance, remedial actions.

- Social/CRU Rental: stock register, arrears, occupancy, yield, SHRA compliance indicators.

- Land & Spatial (SPLUMA): land availability, zoning, SDF alignment; GIS overlays for planning and reporting.

- Reporting & Compliance: standard packs for EXCO, MINMEC, Portfolio Committee, and Open Data excerpts.

3. Data Governance, Security & Compliance

- POPIA alignment: lawful processing, purpose limitation, minimality, and data subject rights baked into design.

- Security: Entra ID RBAC/ABAC, Managed Identities, encryption at rest and in transit, secrets in Key Vault.

- Catalog & Lineage: Purview collections per domain, business glossary, end-to-end lineage for audits.

- Operations: SRE runbooks, alerting, SLOs, error budgets, backup/DR and health dashboards.

4. Project Plan

Phase	Timeline	Key Activities
Phase 1: Discovery & Assessment	Weeks 1–6	• Mobilisation, stakeholder mapping, system & data inventory, profiling, and current-state review. • Target architecture & migration roadmap; value cases & initial KPI catalogue. • Prioritised domain backlog and delivery plan.
Phase 2: Foundations & Landing Zone	Weeks 7–10	• Fabric tenant configuration, OneLake, workspaces (Dev/Test/Prod), RBAC & Key Vault. • CI/CD pipelines, standards, naming conventions, and observability. • On-premises data gateway and network patterns.
Phase 3: Wave 1 Delivery (Priority Domains)	Weeks 11–20	• Build ingestion & quality pipelines for Beneficiaries & Projects. • Gold semantic models and executive dashboards (Direct Lake). • UAT, training, and parallel run.
Phase 4: Wave 2 Delivery (Additional Domains)	Weeks 21–26	• Extend to Land & GIS, Finance, and Rental (Social/CRU). • Performance tuning, data products and APIs. • UAT and readiness assessments.
Phase 5: Go-Live, Cutover & Hypercare	Weeks 27–30	• Controlled cutover by domain with freeze window. • Decommission legacy,

handover, and skills transfer.
• 4 weeks hypercare with defect SLAs.

5. Budget & Commercials

The professional services budget is capped at R20,000,000 (ex VAT). USD figures are indicative and based on the exchange rate noted below.

Item	Allocation %	Amount (ZAR)	Amount (USD)
Phase 1 – Discovery & Assessment	15%	R 3,000,000.00	$ 173,711.64
Phase 2 – Foundations & Landing Zone	20%	R 4,000,000.00	$ 231,615.52
Phase 3 – Wave 1 Delivery	35%	R 7,000,000.00	$ 405,327.16
Phase 4 – Wave 2 Delivery	20%	R 4,000,000.00	$ 231,615.52
Phase 5 – Go-Live & Hypercare	10%	R 2,000,000.00	$ 115,807.76

Exchange rate used for USD conversion: 1 USD = R17.27 (as at proposal date). Final billing occurs in ZAR; USD figures are for comparative purposes only.

6. Assumptions, Dependencies & Exclusions

- Azure consumption and Microsoft Fabric capacity are excluded from this services budget and will be billed to the client subscription.
- Timely access to source systems, test data, and subject matter experts (national, provincial and municipal).
- Delivery primarily remote; on-site travel & expenses billed at cost when pre-approved.
- Security reviews and penetration testing beyond standard hardening can be scoped separately.
- All reporting templates and KPIs will be agreed during Discovery and iteratively refined.

7. Risks & Mitigations

Risk	Likelihood	Impact	Mitigation
Legacy logic undocumented in source systems (e.g., HSS)	Medium	High	Reverse-engineer with SMEs; create test harness; parallel run for two sprints.
Model performance at scale	Medium	High	Star schema; aggregations; Direct Lake; incremental patterns.
Under-sized capacity causing throttling	Low	High	Monitor Fabric metrics; tune pipelines; scale capacity as needed.
UAT stakeholder availability	Medium	Medium	Early calendar blocks; UAT playbooks; BA-led daily triage.
Late security audit findings	Low	High	Security checkpoints from Sprint 2; Purview/Defender scans continuously.

8. Team & RACI (Summary)

Role	Responsibilities
Programme Manager	Governance, schedule, risk & stakeholder management.
Solution/Technical Lead	Architecture, standards, performance & security patterns.
Lead Data Engineer & Data Engineers	Ingestion, transformation, orchestration, testing.
Lead BI Developer & BI Developers	Semantic models, DAX, dashboards, performance tuning.
Data/Business Analysts	Requirements, mappings, testing, change & training.