Intelligent Reporting and Integration

Table of Contents

1 Executive Summary

1.1 Executive Summary and Key Introduction

This programme delivers 35 production-grade dashboards and advanced analytics across five waves on Microsoft Fabric. It establishes platform foundations (OneLake, Lakehouse/Warehouse, Data Factory, Purview), curated semantic models with Direct Lake, and iterative dashboard and ML delivery with adoption and scale-out operations.

1.2 Business Objectives & Outcomes

- Deliver 35 governed, high-performing dashboards with measurable adoption.
- Stand up shared data & semantic assets to accelerate future use cases and self-service BI.
- Embed analytics patterns (feature store, experimentation, MLOps baseline) tied to business KPIs.
- Institutionalize governance: catalog, lineage, sensitivity, and quality rules across domains.
- ≥ 30% reduction in manual reporting effort across addressed domains.
- P95 dashboard load time < 6 seconds (Direct Lake).
- Adoption KPIs tracked (MAU, query success) per domain.
- Purview lineage coverage for 90%+ curated datasets; data dictionary delivered.

1.3 Scope of Work

- Discovery & Design: Wave-based discovery, KPI definition, data mapping, and UX prototyping; analytics charters.
- Data Integration Factory: Shared ingestion & transformations to OneLake (Delta) via Data Factory/Notebooks; parameterised & observable.
- Semantic Modelling: Warehouse/Lakehouse star schemas and Fabric semantic models (RLS/OLS, certified datasets).
- Dashboards Delivery: Power BI dashboards per UC (visual design system, accessibility, performance, deployment pipelines).
- Advanced Analytics: Feature engineering & experiments; model registry and CI/CD for selected UCs each wave.
- Adoption & Training: UAT playbooks, enablement, release notes, CoE patterns; telemetry for usage and satisfaction.
- Operations & Optimisation: Capacity monitoring, cost/perf optimisation, governance expansion, and benefits realisation.

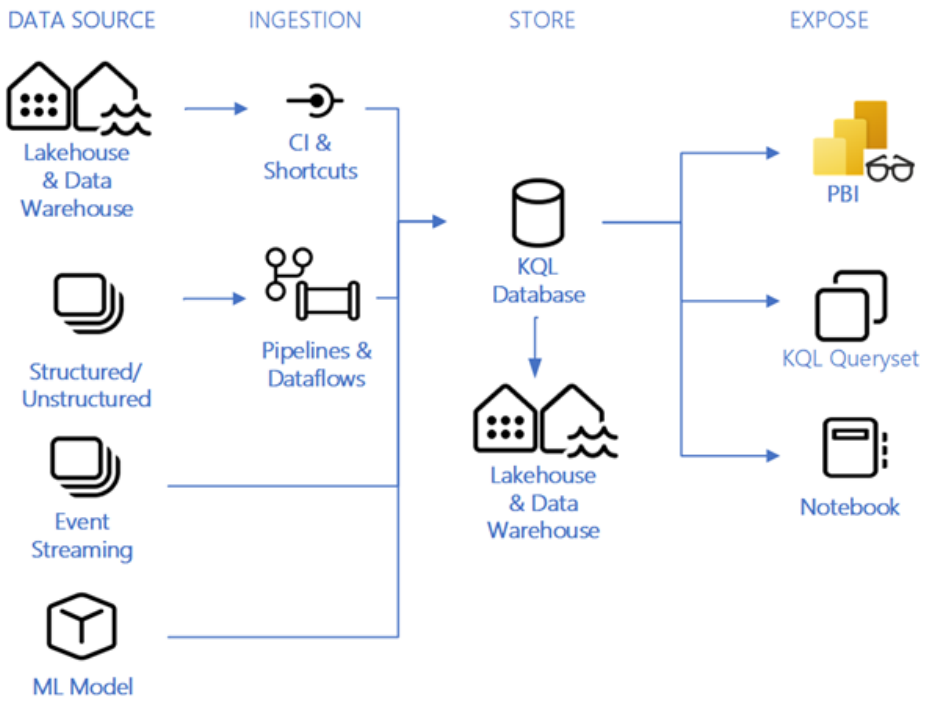
2 Solution Overview

2.1 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.

2.1.1 Layered Architecture Overview

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



2.1.1.1 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.

2.1.1.2 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.

2.1.1.3 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.

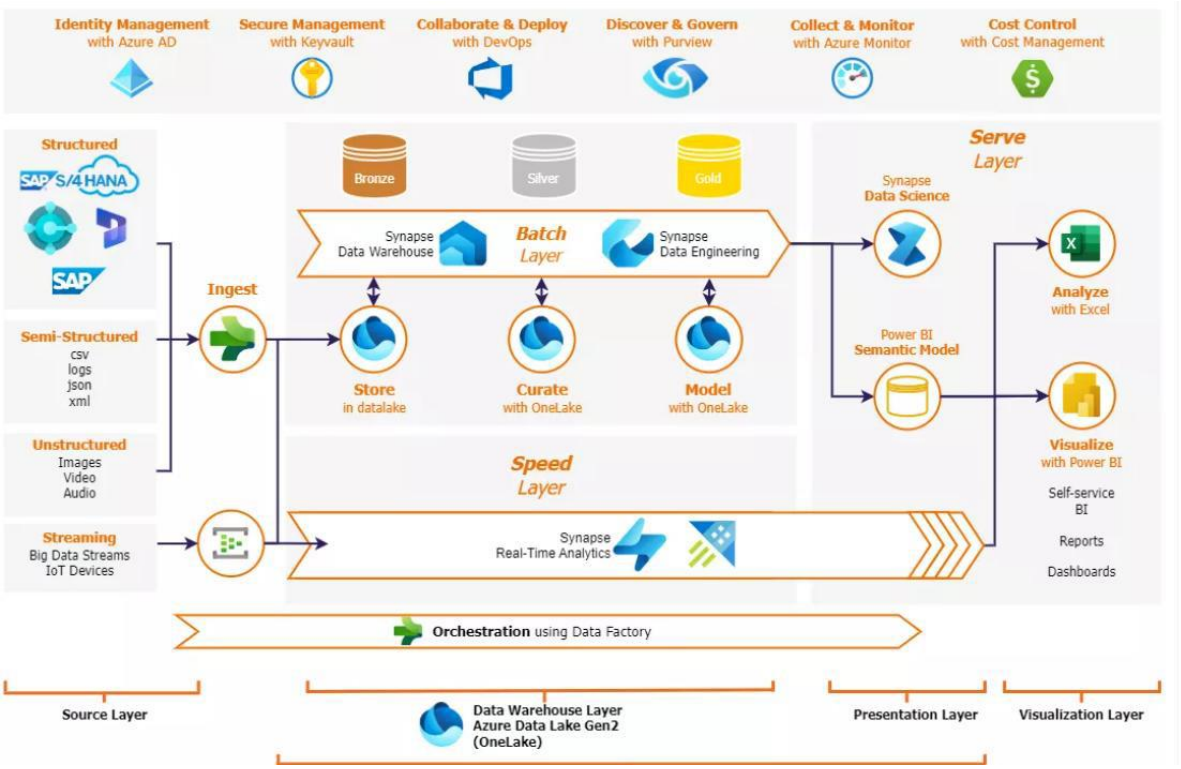
2.1.1.4 Data Governance Layer

- **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.

2.1.1.5 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.



2.1.1.6 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.

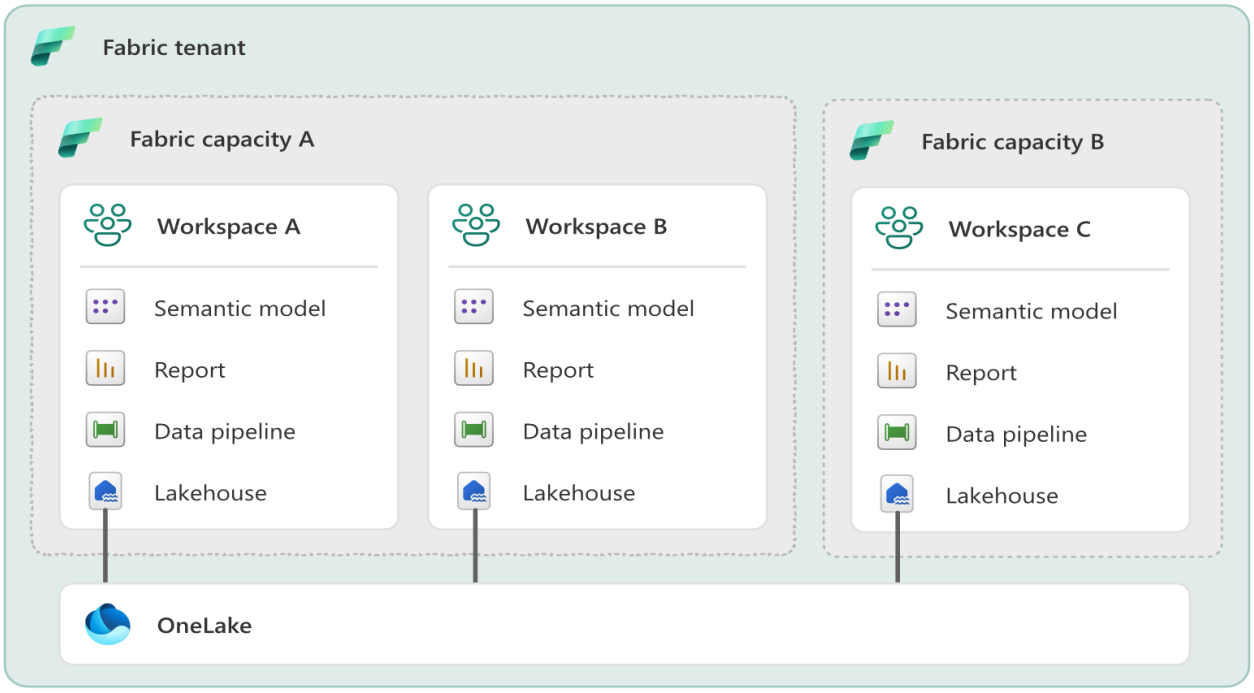
Databricks

- **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.



2.1.1.7 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.

2.1.2 Methodology and Project Approach: Data Platform Modernization

2.1.2.1 Phase-Wise Implementation Plan

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

Phase 1: Discovery & Assessment

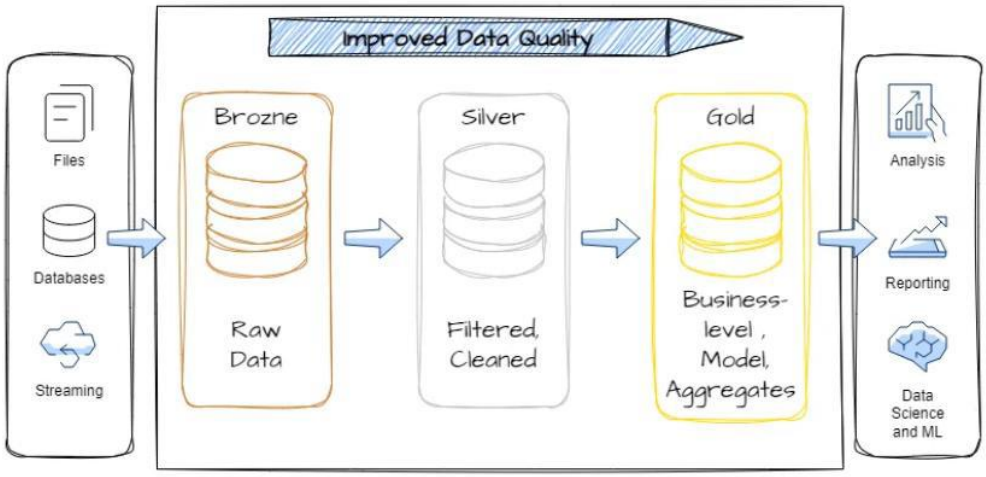
- Inventory of legacy and current systems.
- Data profiling, quality checks, and lineage mapping.
- Catalog metadata and identify integration touchpoints.

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

2.1.2.2 Medallion Architecture Overview (Fabric Lakehouse)

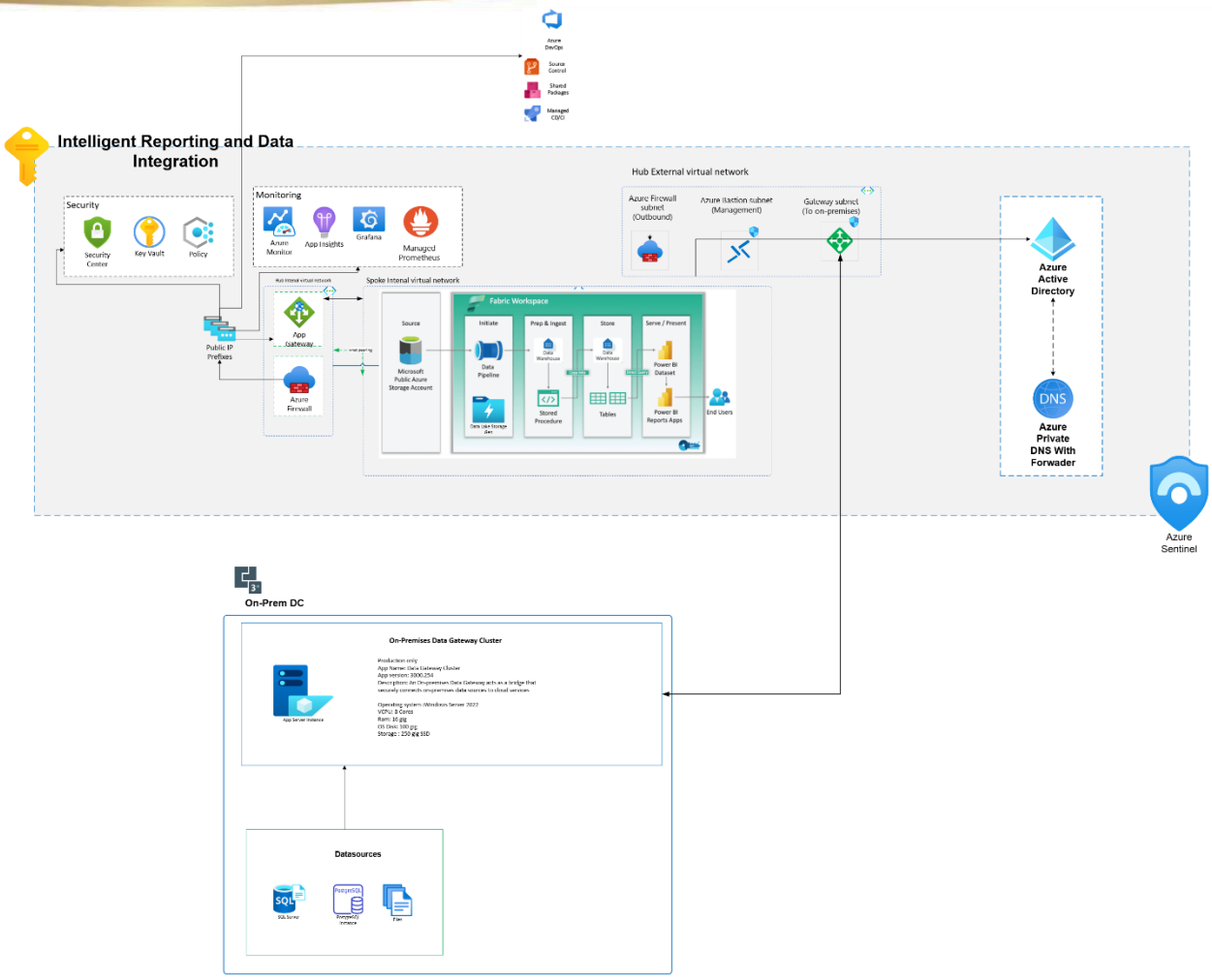
Layer	Purpose	Key Tools in Fabric
Bronze	Raw data ingestion and storage	Dataflows Gen2, Data Pipelines, OneLake
Silver	Cleaned and transformed business data	Lakehouse SQL, Notebooks, Spark
Gold	Business-ready, aggregated/curated datasets	Lakehouse SQL, Power BI, Warehouse

2.1.2.3 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.



2.1.2.4 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.

2.1.2.5 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.

2.1.2.6 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.

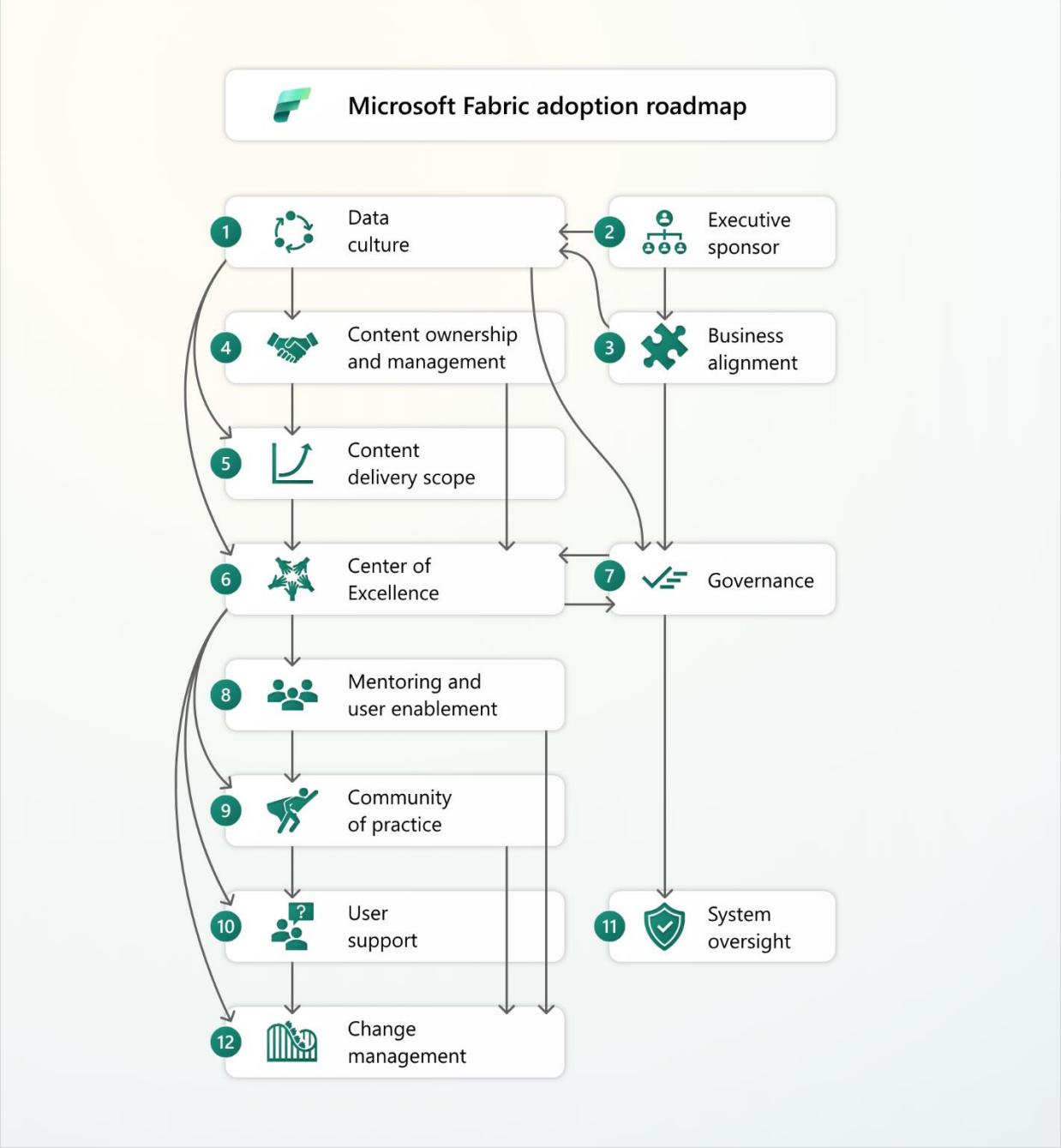
- Performance Optimization and Monitoring

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.

3 Approach & Methodology



3.1.1 Ingestion and Storage with Microsoft Fabric

- **Unified Data Platform:**

Microsoft Fabric provides a single, integrated platform to ingest data from various sources, including raw data, streaming data, and data from third-party APIs.

- **Data Lakehouse/Data Warehouse:**

Data is typically stored in a Fabric Lakehouse, Data Warehouse, or KQL database, centralizing all data and making it more accessible.

- **OneLake:**

Fabric's OneLake centralizes data storage, eliminating data silos and streamlining the data pipeline.

3.1.2 Data Transformation and Preparation

- **Data Transformation:**

Use Fabric's tools to transform and prepare raw data, ensuring data quality and consistency for reporting.

- **Real-time Processing:**

Utilize Fabric Event Streams to process live data and feed it into real-time dashboards in Power BI.

3. Advanced Analytics and AI Integration

- **Machine Learning:**

Integrate predictions and data from Machine Learning models developed within Fabric into Power BI dashboards for advanced insights.

- **AI-driven Insights:**

Apply AI-driven insights directly from Fabric to Power BI, allowing for the discovery of trends and anomalies within the data.

4. Dashboard Development and Visualization with Power BI

- **Data Consumption:**

Power BI connects to the data stored in Fabric's data lakehouse, data warehouse, or other Fabric data sources to build reports and dashboards.

- **Real-time Dashboards:**

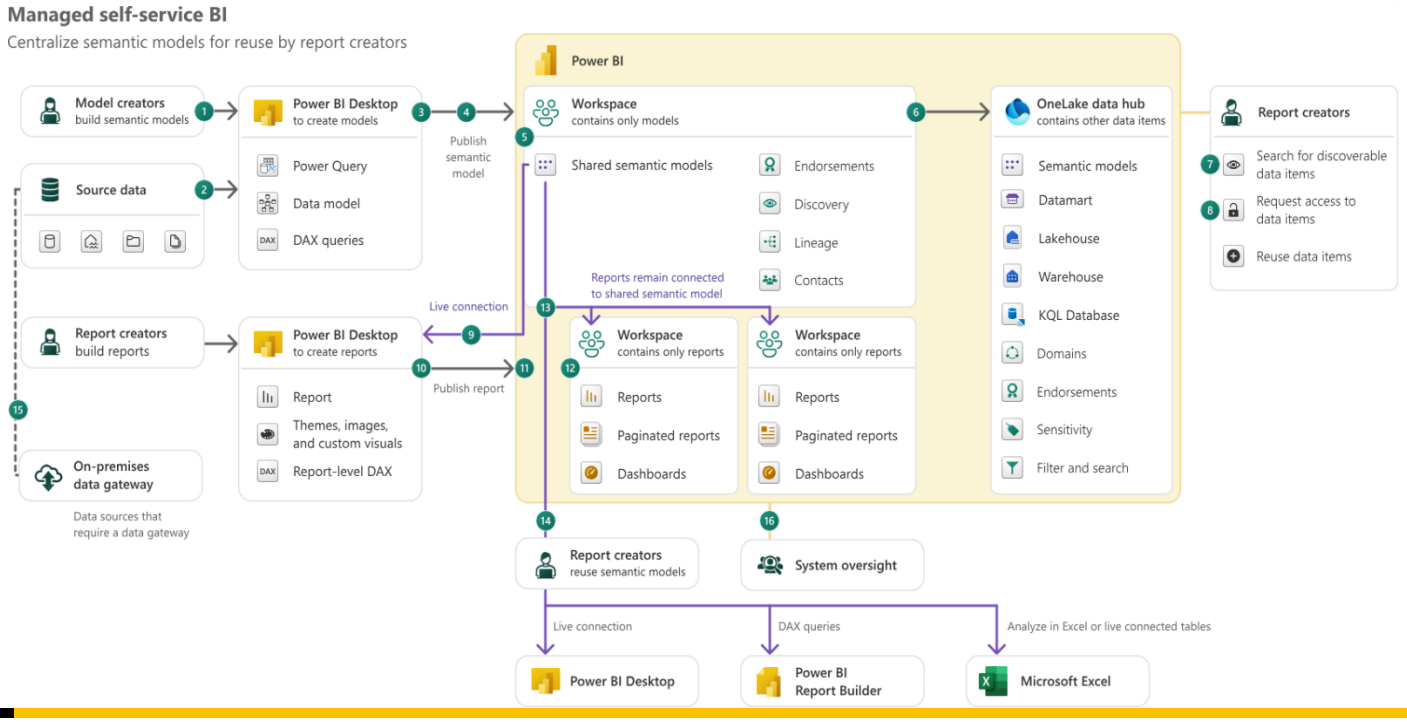
Create real-time dashboards by connecting live data streams processed by Fabric.

- **Publish and Share:**

Publish the completed reports to the Power BI Service for sharing with stakeholders, collaboration, and to maintain updated reports through scheduled data refreshes.

Benefits of this Methodology

- **Simplified Data Governance:** A unified data platform in Fabric simplifies data governance.
- **Improved Data Quality:** Robust systems for managing data quality can be built within the Fabric architecture.
- **Enhanced Self-Service Analytics:** Makes it easier for users to implement and utilize self-service analytics.
- **Seamless Integration:** Integrates data, analytics, and AI into a cohesive platform, eliminating the need for manual integrations.
- **Centralized Data:** Reduces data silos by centralizing data storage in OneLake.



4 Data Governance, Security & Compliance

- **Catalog & Lineage**: Register all data products in Microsoft Purview; define domains/collections; enable end-to-end lineage.
- **Security**: Entra ID RBAC/ABAC, Managed Identities; encryption at rest and in transit; secrets in Key Vault.
- **Compliance**: Sensitivity labels; retention policies; immutable audit logs for critical events; POPIA-aligned controls.
- **Operations**: SRE runbooks, alerting, and dashboards; measurable SLOs and error budgets.

5 Milestones & Cost

Milestone	Proposed % (ex VAT)	Amount (USD)
Phase 0 – Foundations complete	10%	$ 220,248.92
Wave 1 – Use Cases 1–7 complete	15%	$ 330,373.38
Wave 2 – Use Cases 8–14 complete	15%	$ 330,373.38
Wave 3 – Use Cases 15–21 complete	15%	$ 330,373.38
Wave 4 – Use Cases 22–28 complete	15%	$ 330,373.38
Wave 5 – Use Cases 29–35 complete	15%	$ 330,373.38
Phase 3 – Scale, Ops & Optimisation complete	15%	$ 330,373.38

5.1 Pricing & Commercials

Item	Amount (USD)
Labour (ex VAT)	$ 1,655,975.98
Non-labour (ex VAT)	$ 430,594.90
Delivery contingency 7% (ex VAT)	$ 115,918.32
Total (ex VAT)	$ 2,202,489.20
VAT 15% (if applicable)	$ 330,373.38
Total (incl. VAT)	$ 2,532,862.58

5.2 Non-Labour Items (Detail)

Item	Ex VAT (USD)	Incl VAT (USD)
Fabric capacity (F-sku) & Power BI Premium (24 months combined allowance)	$ 260,623.23	$ 299,716.71
Azure DevOps & GitHub Enterprise seats (24 months)	$ 33,994.33	$ 39,093.48
Training, Adoption & UAT enablement (venues, labs, materials)	$ 45,325.78	$ 52,124.65
Travel & on-site waves (Justice cluster)	$ 22,662.89	$ 26,062.32
Data migration/analytics accelerators hardware & tools	$ 45,325.78	$ 52,124.65
Data governance & quality tooling (Purview add-ons, scanners)	$ 22,662.89	$ 26,062.32

6 Risks & Mitigations

Risk	Likelihood	Impact	Mitigation
Wave scope creep across 35 use cases	Medium	High	Strict wave scoping, sprint demos, MoSCoW, change control
Fabric capacity contention during peak loads	Medium	High	Capacity metrics monitoring, Dev/Test isolation, autoscale, reservations
Data quality issues delay analytics	High	High	Early profiling, DQ rules, remediation backlog, data contracts with owners
Model performance & cost in Direct Lake	Medium	Medium	Agg tables, incremental refresh, storage/partition strategy, usage review
Key SME availability for UAT	Medium	Medium	UAT calendar blocks, BA-led triage, executive escalation path