

# Banco Wild West: Modern Data Platform

Investing in Tomorrow, Today.

Group 2 | BUAN 6335 – Spring 2025

## Meet Our Team:

Thriksha Giriraju, Prasanna Warad,  
Srijananie Nagasubburajan, Wenting Lu,  
Hrushikesh Khandare, Sahil Salvi,  
Shivapreya Durgaprasad

# AGENDA:

- Banco Wild West relies on legacy RDBMS with Java/C++ middleware high costs , low agility.
- System lacks support for real-time insights, AI/ML, and advanced analytics.
- CIDO mandate: Build scalable, secure Azure-based cloud data platform.

## Key Objectives:

- Support unified OLTP and OLAP workloads.
- Enable AI/ML analytics
- Drive real-time insights, compliance, and cost efficiency

# KEY REQUIREMENTS

## Technical:

- Ingest structured, semi-structured, and unstructured data
- Enable batch, micro-batch, streaming pipelines
- Centralized metadata and SSOT
- Regulatory security (NIST, GDPR, IRS)



## Business:

- Personalized banking experiences
- Fraud detection, credit scoring
- Audit-ready infrastructure
- Scalable and cost-efficient design

# Cloud Platform Evaluation & Decision

| Criteria                          | AWS (Market Leader)   | GCP (AI/ML Innovator)                                 | Azure (Selected)  |
|-----------------------------------|---|---|---|
| Enterprise Integration            | ✗ Limited Microsoft support (AAD, Office, Power BI)               | ✗ External to Microsoft stack                         | ✓ Full native integration (AD, 365, Power BI, Teams)        |
| Data Governance & Compliance      | ✗ Fragmented tools (IAM, IAMS, Macie, Lake Formation)             | ⚠ Improving via Dataplex; still maturing              | ✓ Unified toolset (Purview, Key Vault, Sentinel)            |
| Lakehouse Architecture            | ✗ Delta Lake not native; setup with S3, Glue, Redshift is complex | ⚠ BigLake & Iceberg evolving                          | ✓ Native Delta Lake with ADLS, Synapse, Databricks          |
| AI/ML Enablement                  | ⚠ SageMaker powerful, but DevOps-heavy                            | ✓ Vertex AI strong but lacks prebuilt NLP/OCR         | ✓ Cognitive Services + MLOps + Power BI                     |
| BI & Visualization                | ✗ No native BI; external tools needed                             | ⚠ Looker available, limited enterprise use            | ✓ Power BI deeply integrated & widely adopted               |
| Cost Optimization                 | ⚠ Complex pricing (Glue, Redshift, egress)                        | ✓ Tiered storage; serverless Synapse                  | ✓ Efficient tiered storage + integration                    |
| Regulated Industry Fit            | ⚠ Manual controls needed; compliance burden                       | ✗ Limited use cases in enterprise finance             | ✓ Trusted in finance, health, public sector                 |
| Tool Ecosystem & Interoperability | ⚠ Requires custom integration                                     | ✓ Strong GCP ecosystem, but less external flexibility | ✓ Interoperable with many tools & Microsoft-native stack    |
| Adoption & Maturity               | ✓ Market dominant with widespread adoption                        | ⚠ Strong in AI/ML, weaker enterprise penetration      | ✓ Enterprise-grade, strong government & enterprise presence |

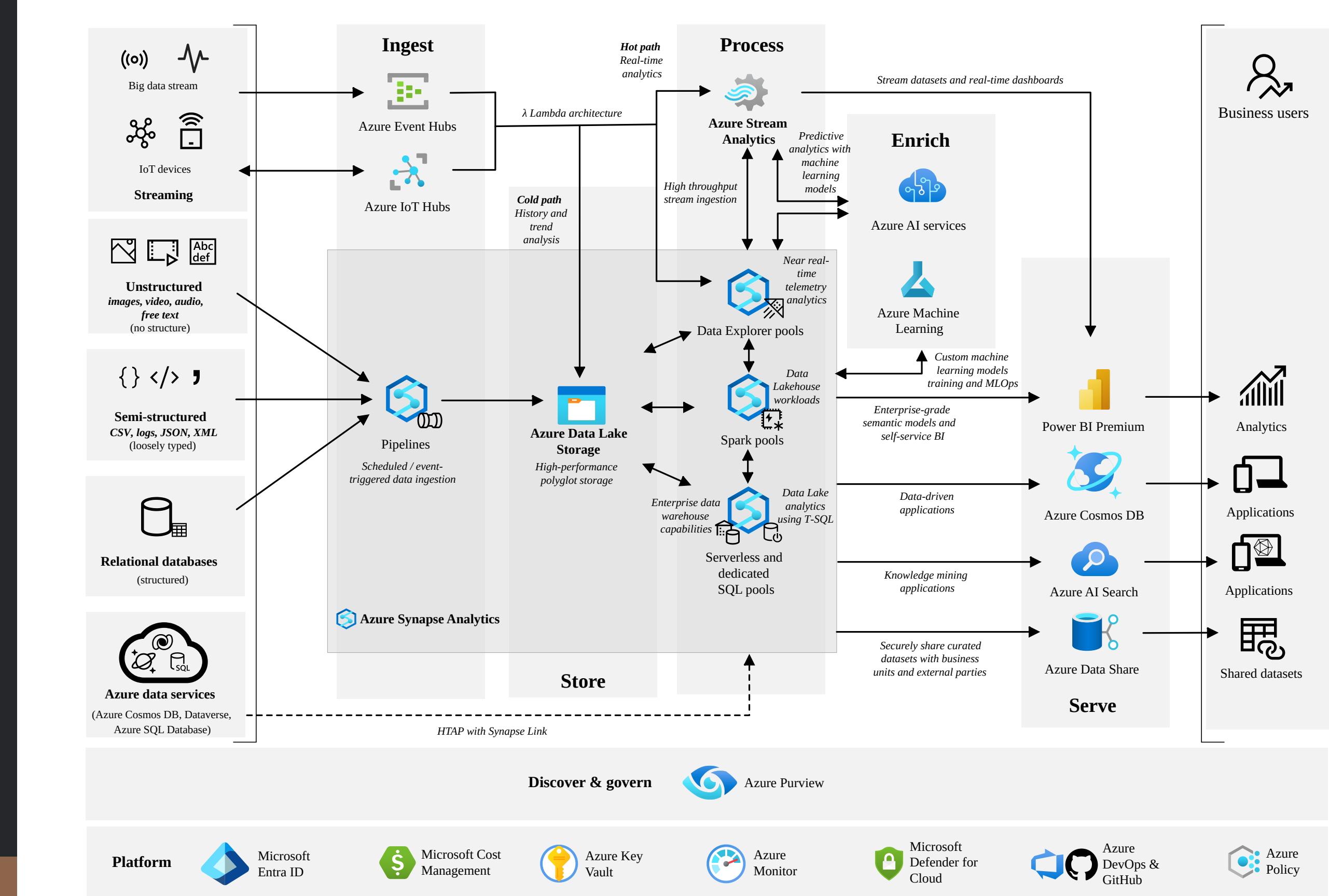


# Selected Azure Components

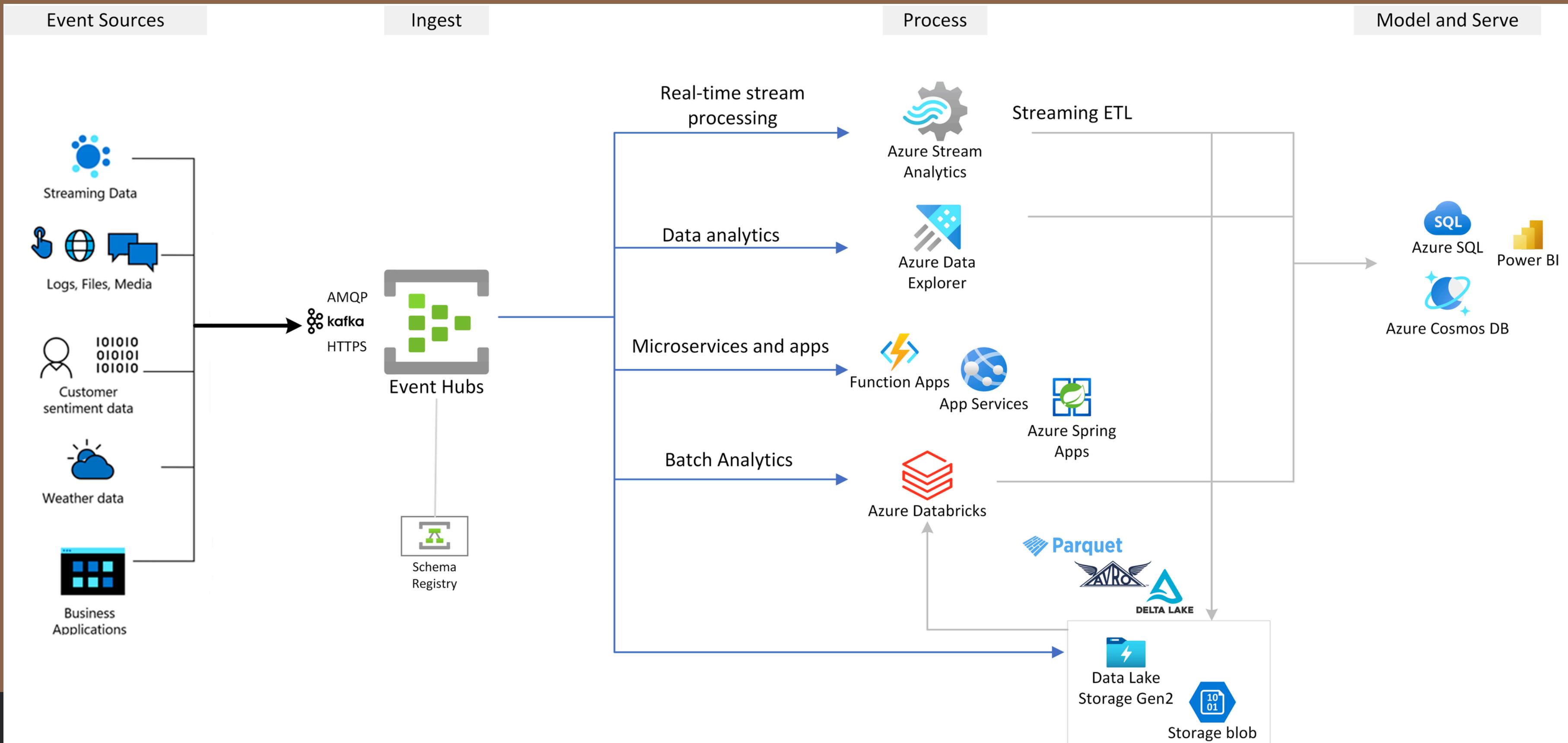


# Lakehouse Architecture Overview

- Combines the strengths of a data lake (scale) + warehouse (structure)
- Bronze: Raw → Silver: Cleaned → Gold: Analytics
- Supports OLTP + OLAP, ML pipelines, API and dashboard integration
- Ensures versioning, rollback, and schema enforcement via Delta Lake
- Seamless integration with reporting tools, APIs + ML workflows



# Data Ingestion Architecture



# Data Ingestion Architecture

## Source Systems



### AdventureWorksLT

- SalesLT.Address
- SalesLT.CustomerAddress
- SalesLT.Customer
- SalesLT.SalesOrderDetail
- SalesLT.SalesOrderHeader
- SalesLT.Product



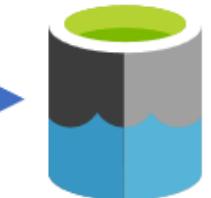
### WideWorldImporters

- Application.Cities
- Application.StateProvinces
- Application.Countries
- Sales.Customers
- Sales.Orders
- Sales.Orderlines
- Warehouse.StockItems



## Bronze Data Lake

### Raw Zone



### adventureworksLT Container

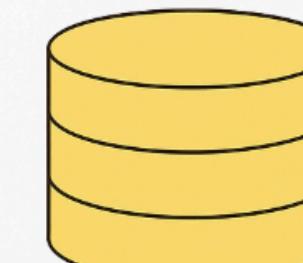
- SalesLT/Address/new (folder structure)
- SalesLT/CustomerAddress/new
- SalesLT/Customer/new
- SalesLT/SalesOrderDetail/new
- SalesLT/SalesOrderHeader/new
- SalesLT/Product/new



### WideWorldImporters Container

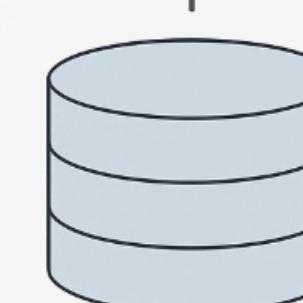
- Application/Cities/new (folder structure)
- Application/StateProvinces/new
- Application/Countries/new
- Sales/Customers/new
- Sales/Orders/new
- Sales/Orderlines/new
- Warehouse/StockItems/new

## Data Storage and Security



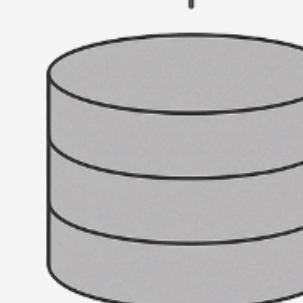
### Gold Layer

Aggregated, curated data



### Silver Layer

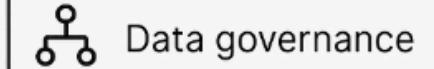
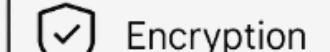
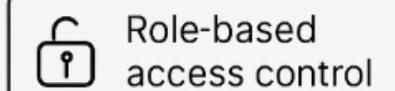
Cleansed, consolidated data



### Bronze Layer

Raw data storage

- Structured, semi-structured, unstructured
- Delta Lake format
- ACID compliance
- Data partitioning



# CHALLENGES

## Legacy System Complexity

Tightly coupled monolithic systems (e.g., old Java/C++ middleware, Teradata) are difficult to integrate with modern cloud-native solutions.

## Real-Time Data Latency

Difficulty in streaming structured + semi-structured data concurrently.

## Metadata Overload

Difficulty tracking data lineage, versioning, and data owner accountability.

## Governance Complexity

Regulations (GDPR, AML, IRS) demand transparency and restricted access.

## Data Silos Across Departments

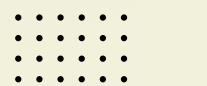
Different teams maintain separate data stores with no consistent structure or sharing mechanism.

## Tool Proliferation

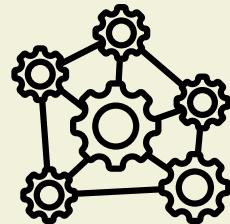
No standardization leads to redundancy and siloed development.



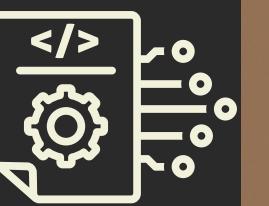
# MITIGATION PLAN



## Legacy System Complexity

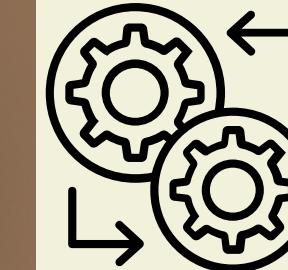


Use Azure Data Factory or Synapse Pipelines to bridge legacy RDBMS via connectors.



## Metadata Overload

Use Azure Purview for centralized metadata cataloging and lineage tracking.



## Data Silos Across Departments

Use Azure Data Lake Gen2 as a central repository (lakehouse).



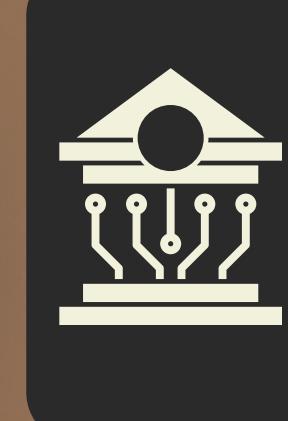
## Real-Time Data Latency

Introduce Lambda Architecture: combine batch (cold path) + real-time (hot path) analytics.



## Tool Proliferation

Consolidate analytics to Power BI Premium for standard dashboards.

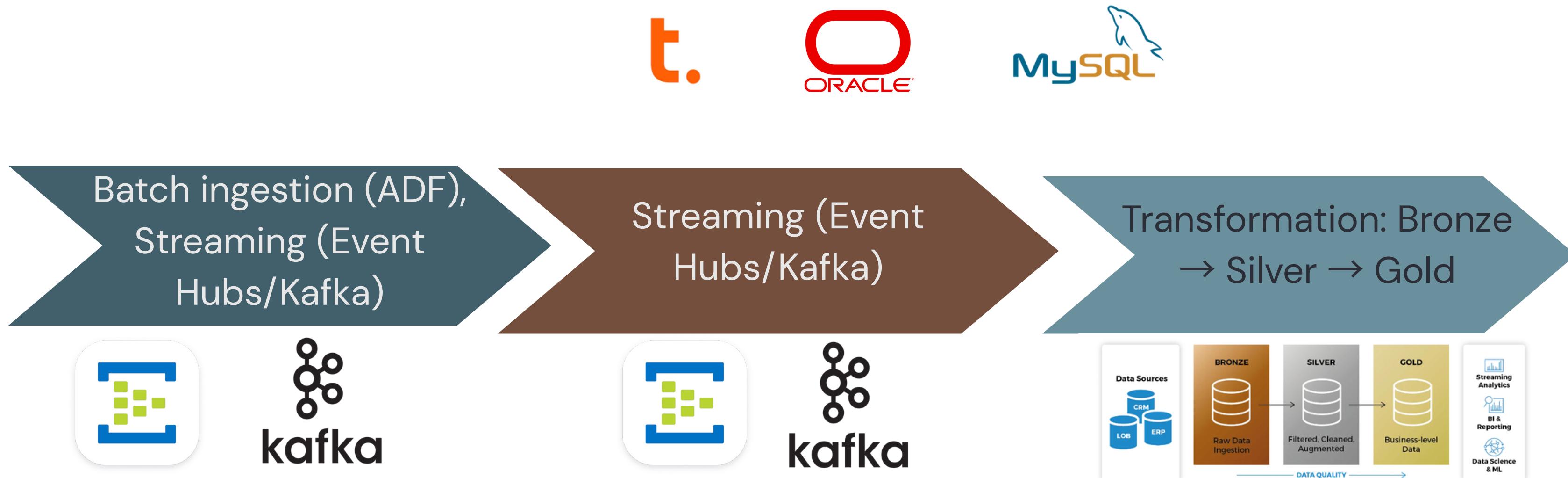


## Governance Complexity

Implement Azure AD + RBAC for fine-grained role-based access.

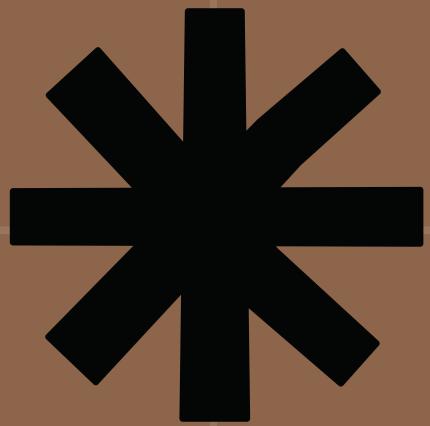
# Data Migration Strategy

Legacy Sources: Teradata, Oracle, MySQL, SQL Server



- Delta Lake Advantage: Validation, schema enforcement via Databricks
- Parallel run → final cutover → decommissioning
- OCR pipelines for unstructured documents

# ML / AI Applications



Fraud detection using real-time analytics

Detects suspicious transaction patterns by applying real-time anomaly detection and supervised models, reducing financial risk and boosting customer trust.

Credit risk scoring with Azure ML

Uses machine learning models to assess loan default risk by analyzing customer demographics, credit history, and income, streamlining approvals and minimizing bad debt.

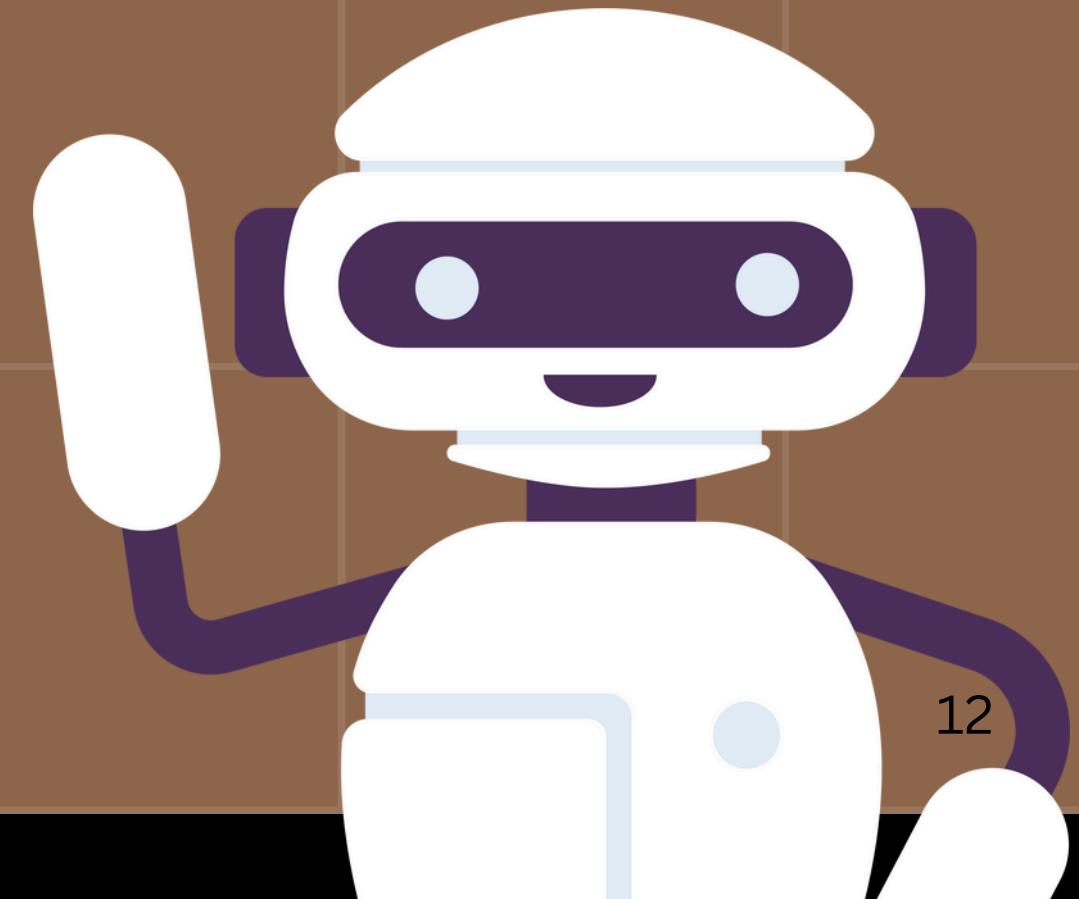
Predictive ATM maintenance and staff optimization

Forecasts potential ATM failures using telemetry data and time-series models, enabling proactive maintenance and optimizing branch workforce planning.

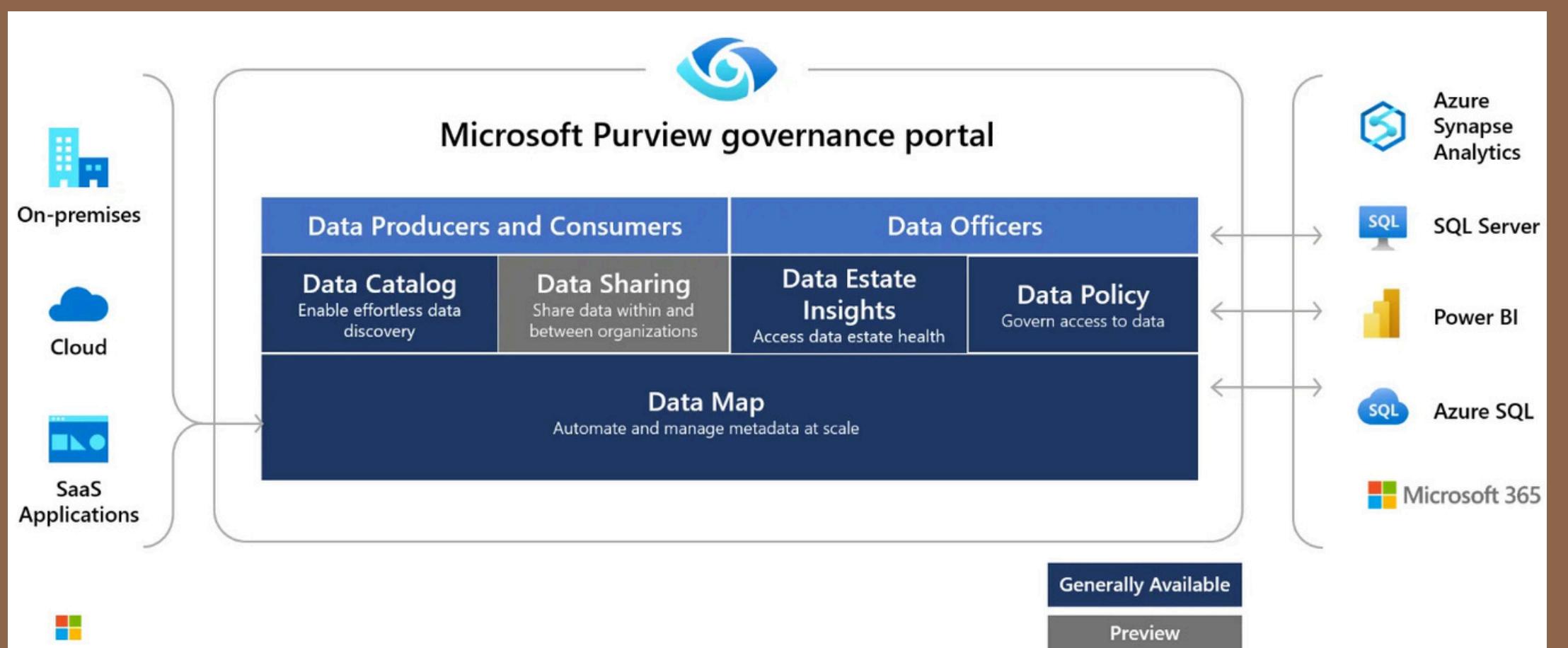
Hyper-personalized banking experiences

Recommends relevant products or services by analyzing customer behavior and preferences using collaborative filtering and predictive analytics.

Chatbots for enhanced customer support  
Leverages NLP from Azure Cognitive Services to automate routine queries, improve customer service availability, and reduce call center workload.



# COMPLIANCE AND DATA GOVERNANCE



| Governance Area          | Azure Tool                        |
|--------------------------|-----------------------------------|
| Metadata Management      | Azure Purview (Data Map, Catalog) |
| Secure Access Control    | Azure AD + RBAC                   |
| Encryption Management    | Key Vault (BYOK)                  |
| Audit & Monitoring       | Log Analytics, Azure Monitor      |
| Threat Detection         | Defender for Cloud, Sentinel      |
| Compliance Standards Met | NIST, GDPR, IRS                   |

Microsoft Purview acts as a central governance hub, unifying metadata, access policies, and lineage tracking across diverse environments, forming the foundation for secure SSOT enablement.



# EXECUTION ROADMAP

## Quarterly implementation plan:

Q1: Infra setup,  
RBAC,  
cataloging

Q2: ADF pipelines,  
Bronze/Silver  
setup

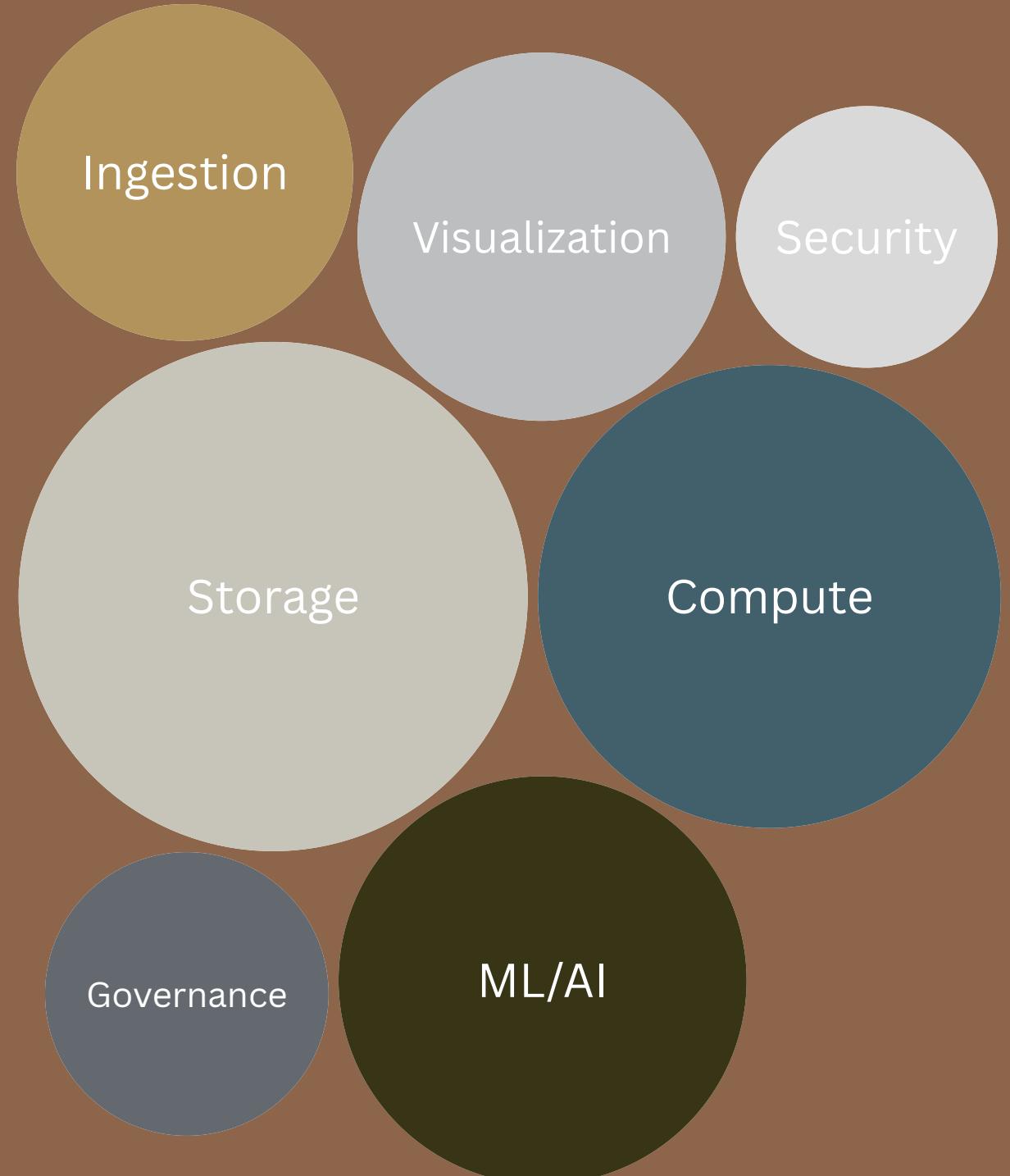
Q3: Event Hubs,  
ML prototypes,  
Synapse dashboards

Q4: BI launch,  
API Gateway,  
monitoring setup

Post-Q4: ML retraining, cost optimization,  
onboarding new use cases



# Cost and ROI Analysis

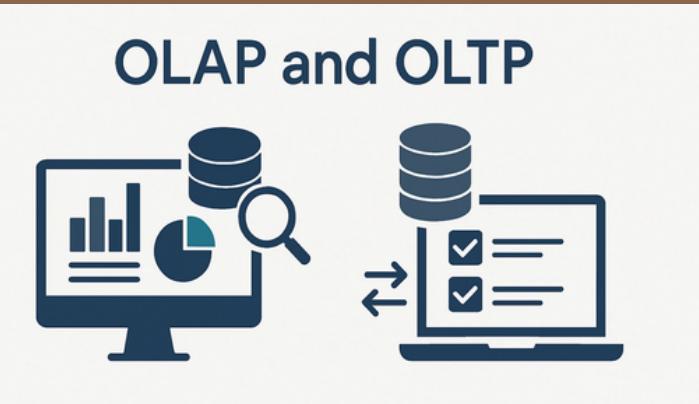


Monthly Cost Estimation~\$30500, saving 35-40%

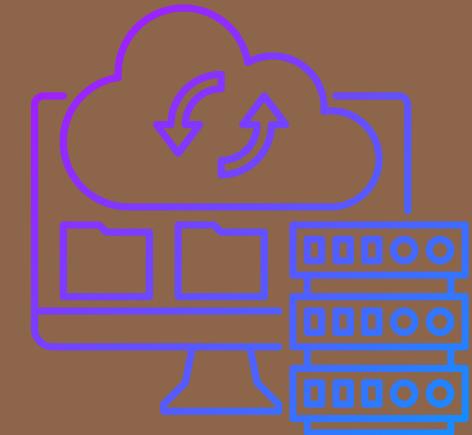


Return On Investment

# CONCLUSION & BUSINESS IMPACT



Unified & Scalable



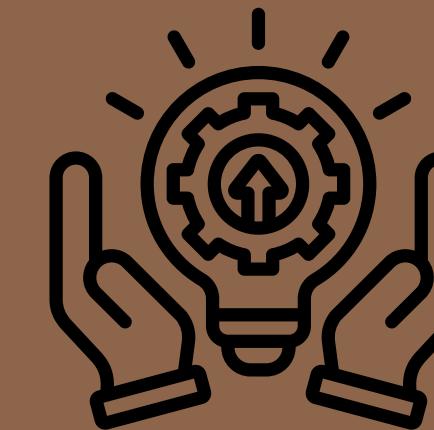
AI-Ready



Secure &  
Compliant



Insight-Driven text



Future-Proof

# REFERENCES

- <https://learn.microsoft.com/en-us/azure/event-hubs/event-hubs-about>
- <https://learn.microsoft.com/en-us/azure/architecture/solution-ideas/articles/iot-azure-data-explorer>
- <https://learn.microsoft.com/en-us/azure/architecture/ai-ml/idea/many-models-machine-learning-azure-machine-learning>
- <https://learn.microsoft.com/en-us/azure/dms/migration-using-azure-data-studio?tabs=azure-sql-mi>



# Thank You