

Group 02



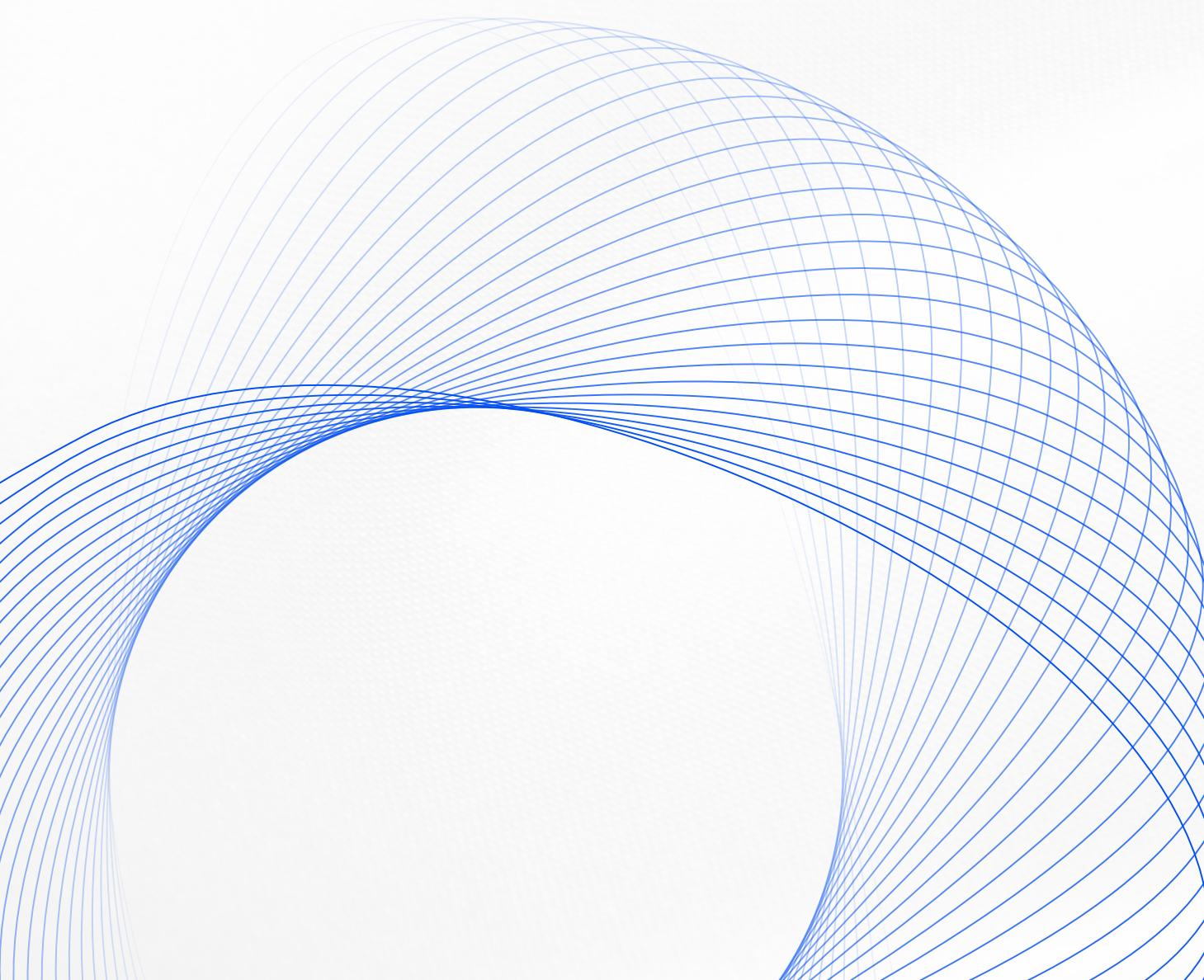
BUAN 6335.502

# Banco Wild West Cloud-Native Data Platform Architecture

Analyzing Challenges, Strategies, and Results

Sahil · Sri Krishna · Bhumika · Ananya · Saurabh · Utkarsh · Ambuj

# Agenda

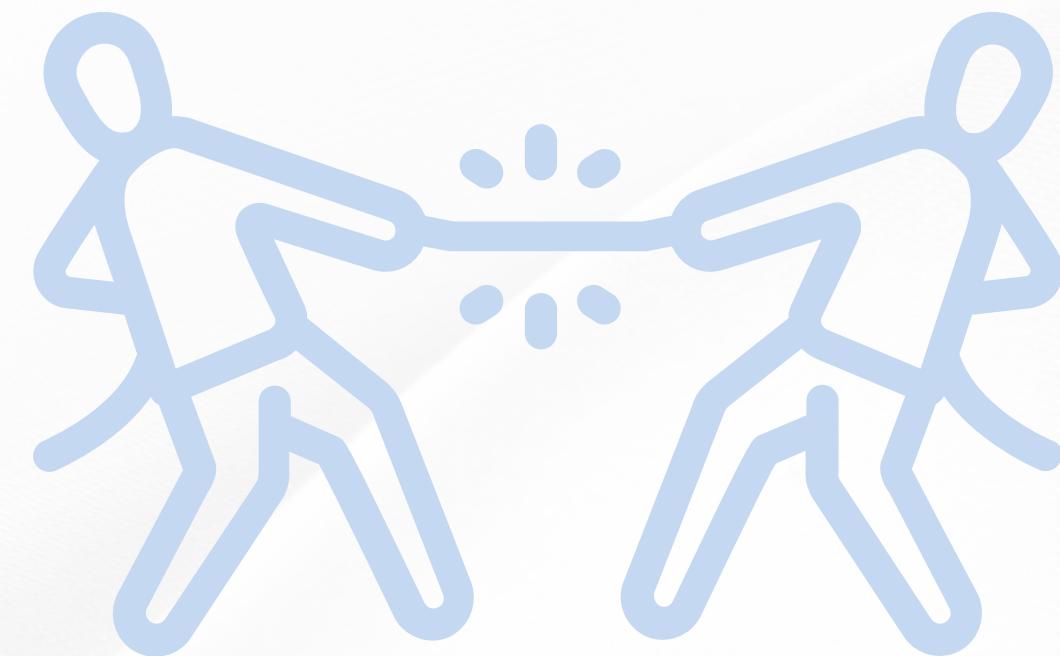
- 
- 
- 1 Project Overview & Competitor Landscape
  - 2 Components Walkthrough
    - Data Sources
    - Ingestion
    - Storage
    - Processing
    - Consumption & AI
  - 3 Governance & Compliance
  - 4 Data Flow Architecture
  - 5 Why Azure? Final Thoughts

# Context

- Legacy systems using C++/Java middleware and distributed RDBMSs
- High maintenance costs, limited scalability, poor customer experience
- New CIDO Ms. Data First pushes cloud transformation
- Microsoft Azure chosen for modernization due to compliance, scalability, and real-time analytics support

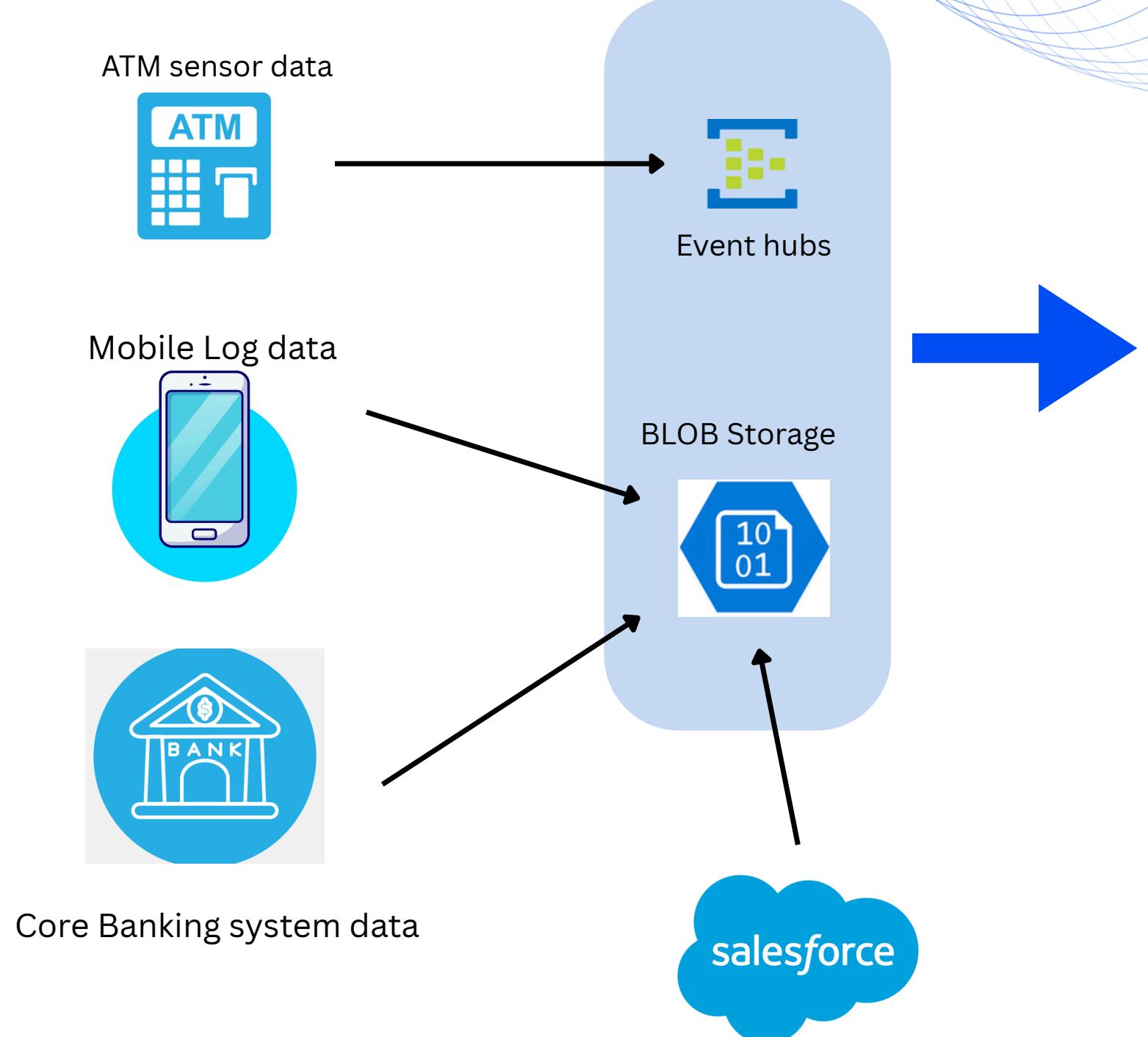
# Competitor Analysis

- JP Morgan uses AWS: good for big data, but costly
- Bank of America uses Azure: compliance, hybrid flexibility, lower cost
- Wells Fargo uses Azure + GCP: Azure for data lake, GCP for ML
- Why Azure: Best fit for Banco Wild West's compliance, hybrid, and integration needs



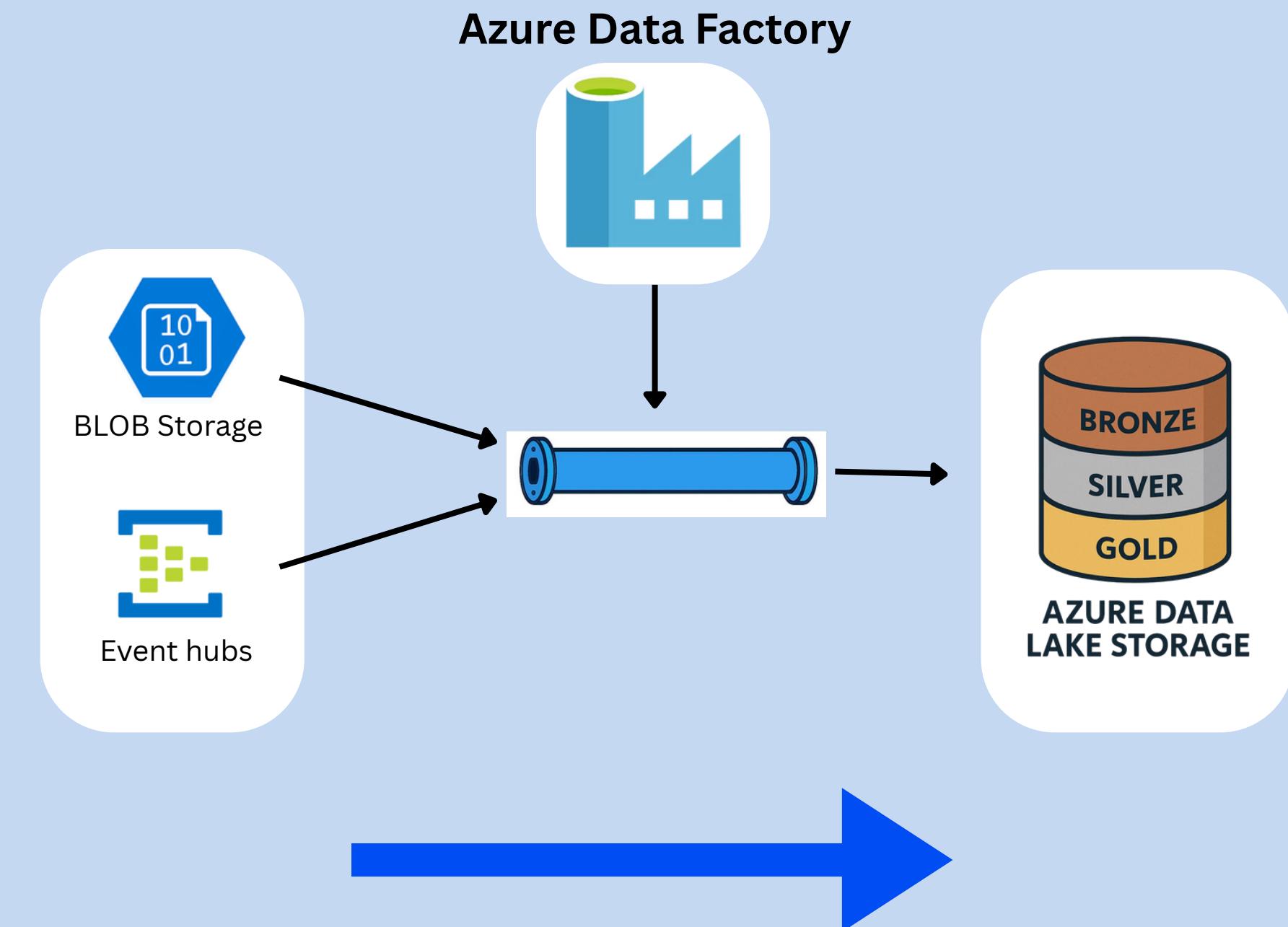
Category	Examples	Data Type
Customer Information	PII, KYC details, account profiles	Structured
Transactions	Payments, deposits, transfers	Structured
Loans & Mortgages	Application forms, scanned documents	Semi-structured / Unstructured
Credit Scores & Risk	Scores from bureaus, internal risk models	Structured
Channel Interactions	CRM entries, ATM logs, mobile/web clickstreams	Semi-structured
Communication Data	Emails, voice calls, chatbot logs	Unstructured
Social & External Feeds	Social media, regulator APIs, bureau feeds	Semi-structured

# Data Sources



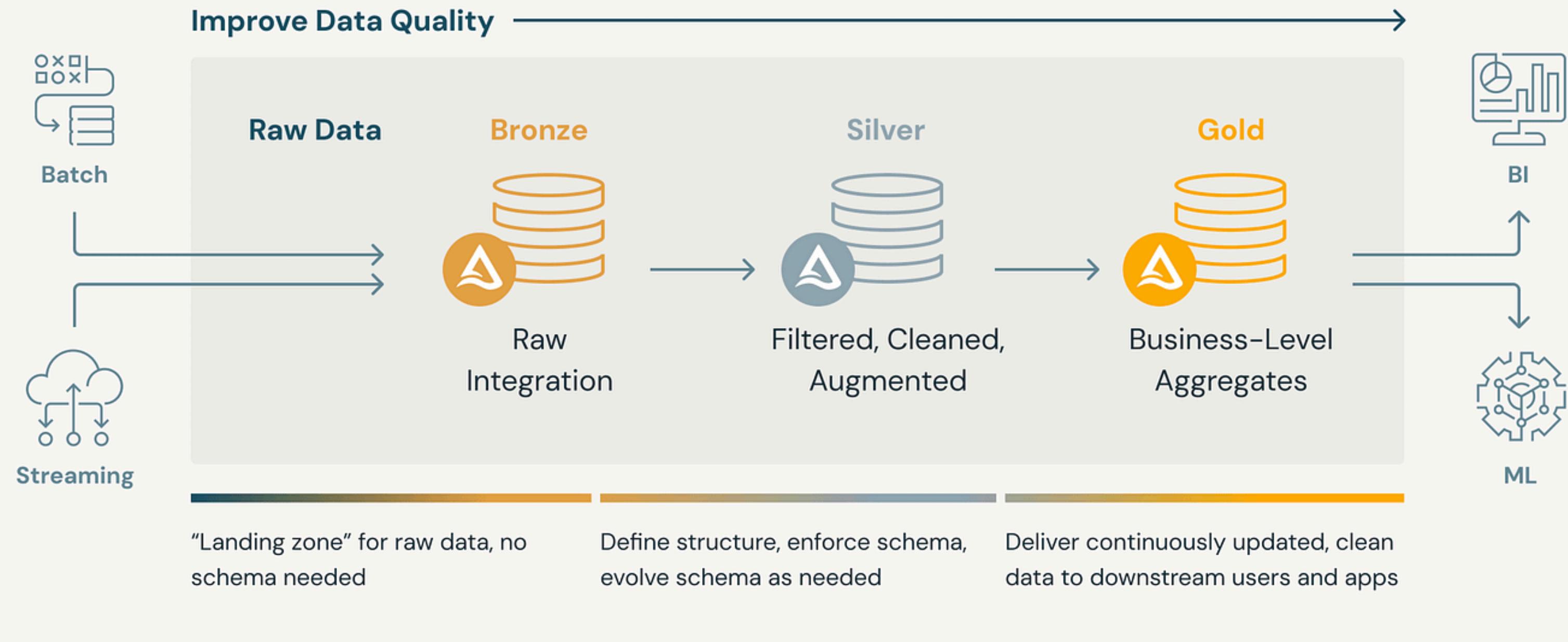
- **Ingests Data from Diverse Sources**
  - Connects to databases, files (BLOB), real-time streams (Event Hubs), SaaS (Salesforce), and more.
- **Orchestrates ETL/ELT Processes**
  - Manages the extraction, transformation, and loading of data.
- **Loads Data into Azure Data Lake Storage**
  - Serves as the primary destination for organized data storage.
- **Enables Analytics with Azure Synapse & Databricks**
  - Prepares data for downstream processing and analysis.
- **Provides Automated and Monitored Data Pipelines**

# Ingesting the Data

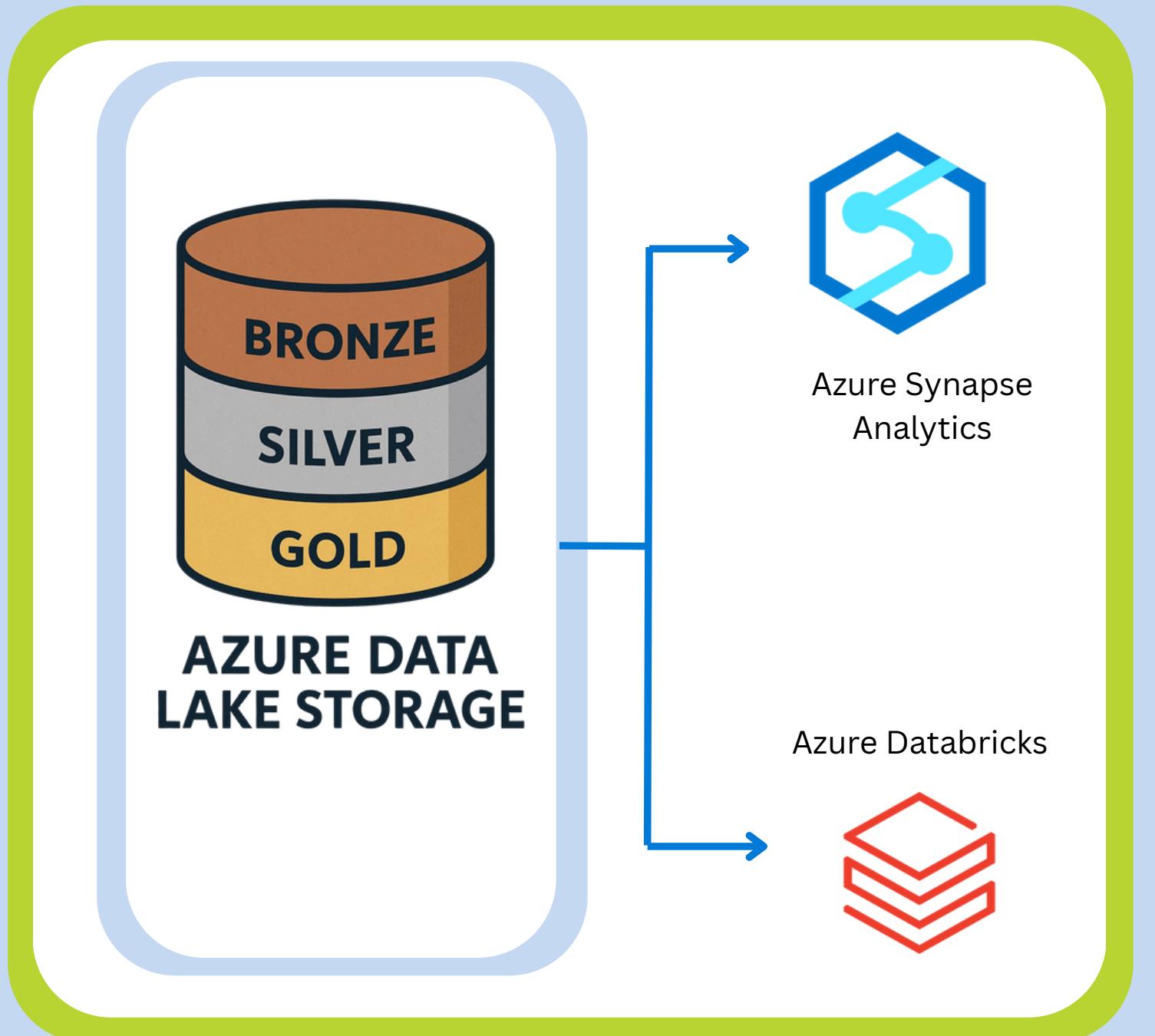


# The Storage

Building reliable, performant data pipelines with  **DELTA LAKE**



# Analytical Tool



Component	Azure Synapse Analytics	Azure Databricks
<b>Role</b>	Data warehouse + big data analytics platform	Unified analytics for big data + machine learning
<b>Data Flow</b>	Reads/Writes data from/to Azure Data Lake Storage	Reads/Writes data from/to Azure Data Lake Storage
<b>Use Cases</b>	Structured data analysis, large-scale SQL, BI workloads	Real-time + batch data processing, ETL/ELT, ML
<b>Integration</b>	Works with ADF, Power BI, ADLS, Purview	Works with ADF, MLflow, ADLS, Purview
<b>Processing Engine</b>	SQL-based (T-SQL, Spark SQL), Pipelines	Apache Spark-based notebooks (PySpark, Scala, SQL)
<b>Data Governance</b>	Integrated with Microsoft Purview for lineage/catalog	Integrated with Microsoft Purview for lineage/catalog
<b>Security</b>	Controlled via Entra ID, Key Vault	Controlled via Entra ID, Key Vault

# Agentic AI Integration

High volumes of repetitive customer queries (e.g., balance checks, loan status) overwhelm agents, slowing down resolution time and impacting service quality.

solutions

- Agent Assist via Salesforce + Databricks ML
- REST APIs
- Power BI Dashboards on Gold Layer
- ML Model Deployment via Azure ML & MLflow
- Delta Sharing for External Stakeholders

# Governance, Security & Licensing

## Access Control

- Entra ID (Azure AD) + MFA
- RBAC for service-level permissions

## Data Governance

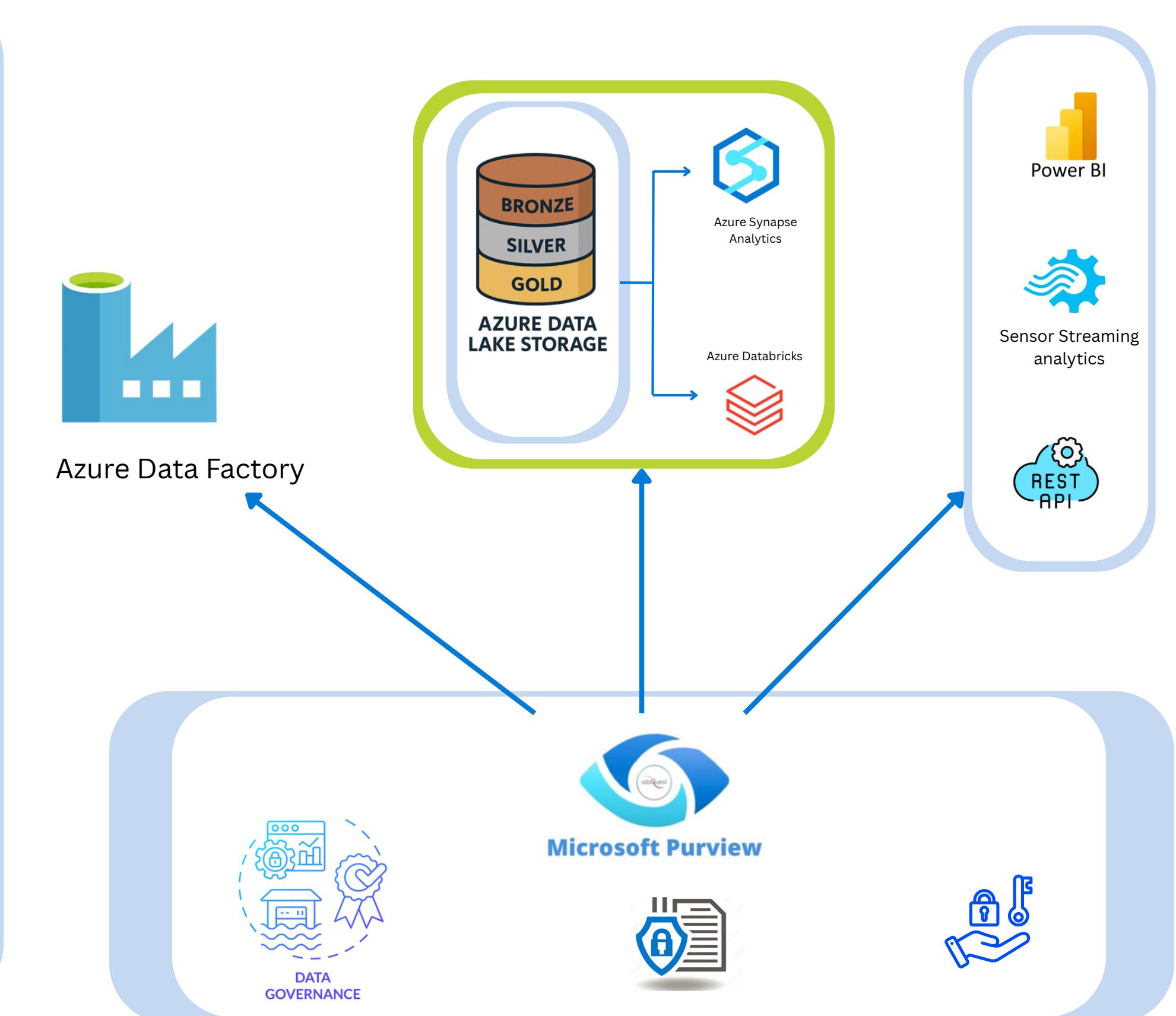
- Microsoft Purview: PII tagging, lineage, auditing
- Unity Catalog: Row/column-level access

## Security

- AES-256 (at rest), TLS 1.2 (in transit)
- Azure Key Vault for secrets & keys

## Licensing & Storage

- Usage-based Azure licensing
- Optional: OneDrive/SharePoint for sharing



# Why Azure?

## Trusted & Compliant

- 90+ compliance certifications (e.g., PCI-DSS, ISO 27001)
- Built-in governance tools: Purview, Key Vault, RBAC

## Seamless Integration

- Natively connects with Power BI, SQL Server, Azure AD
- Smooth fit within the Microsoft ecosystem

## Cost Efficiency

- Lower storage & data egress costs vs. AWS/GCP
- Tiered storage (Hot, Cool, Archive) for optimized spending

## Hybrid & Multi-Cloud Ready

- Azure Arc enables on-prem, hybrid, and multi-cloud deployments

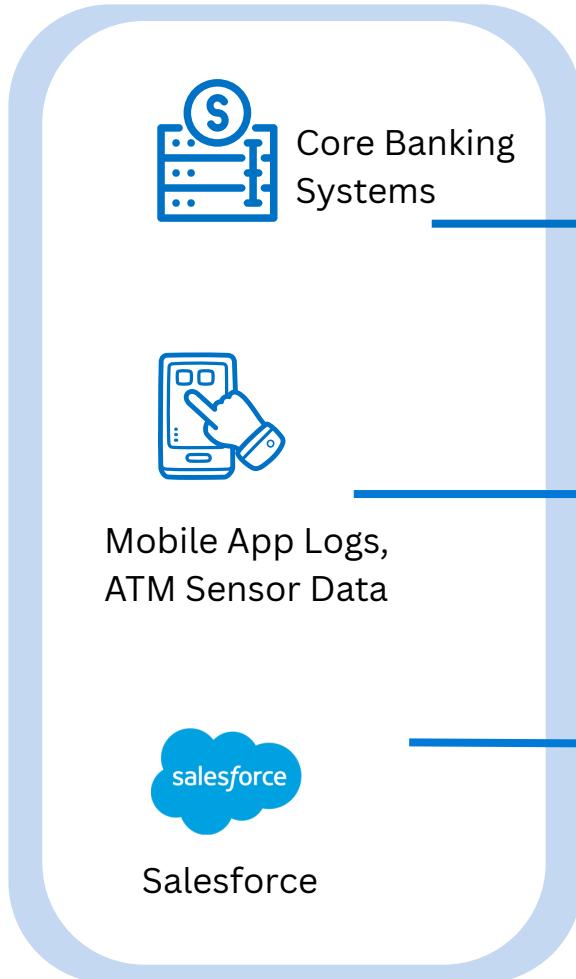
## Strategic Fit

- Supports compliance, reduces cost, and accelerates modernization

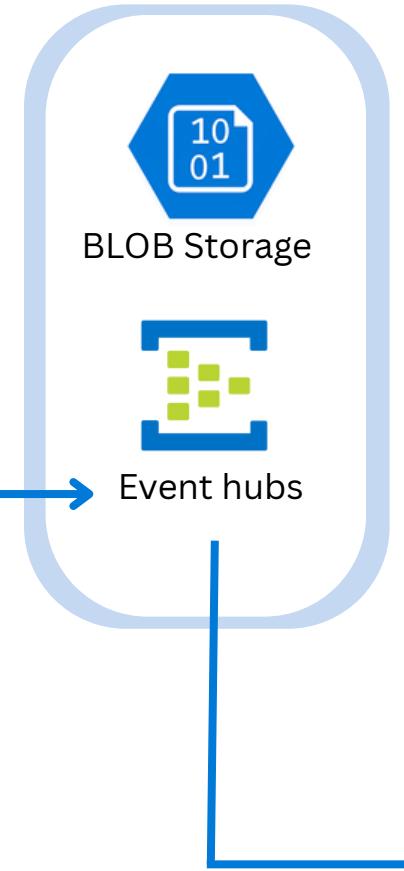


# Data Flow Architecture

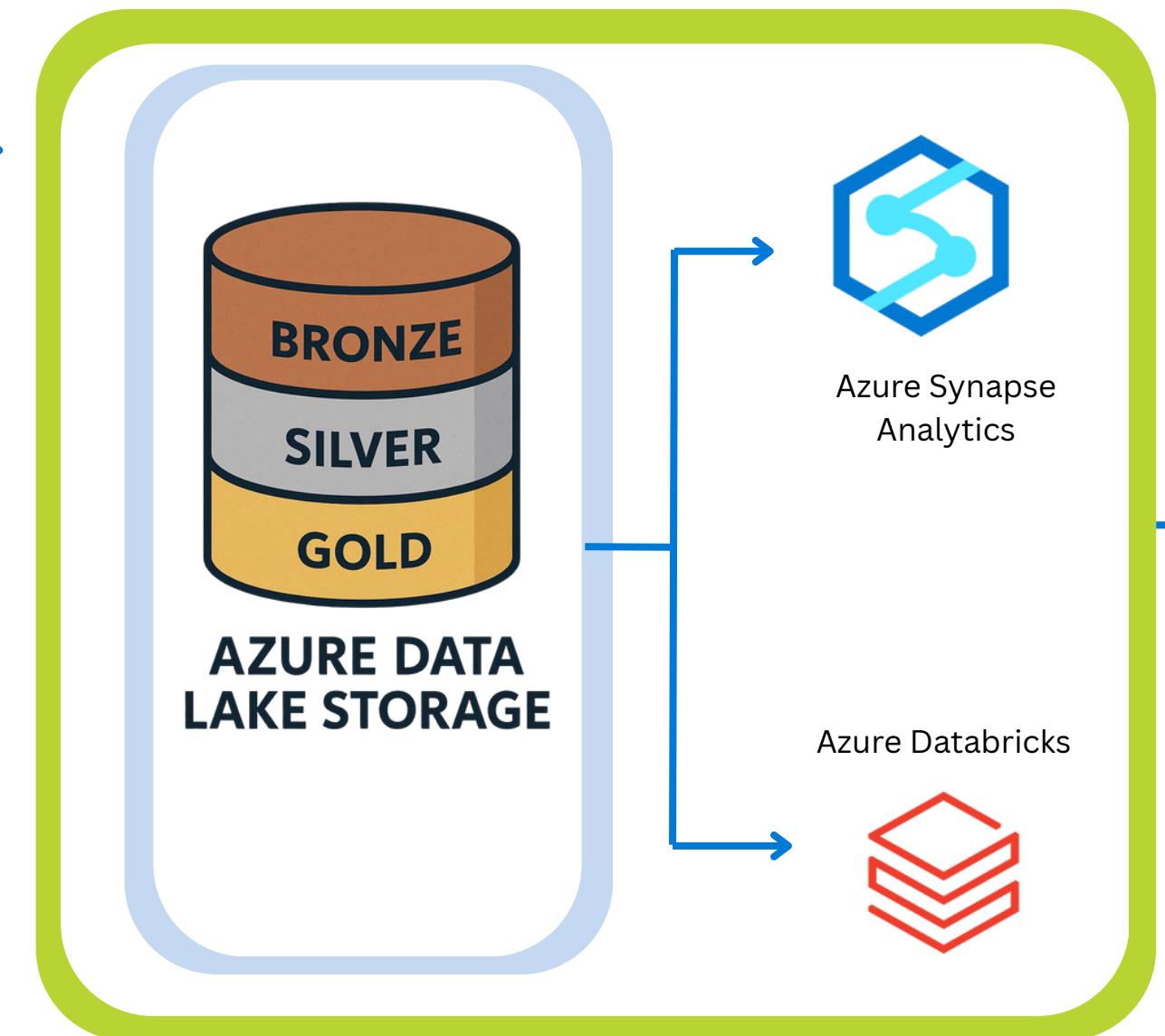
## DATA SOURCES



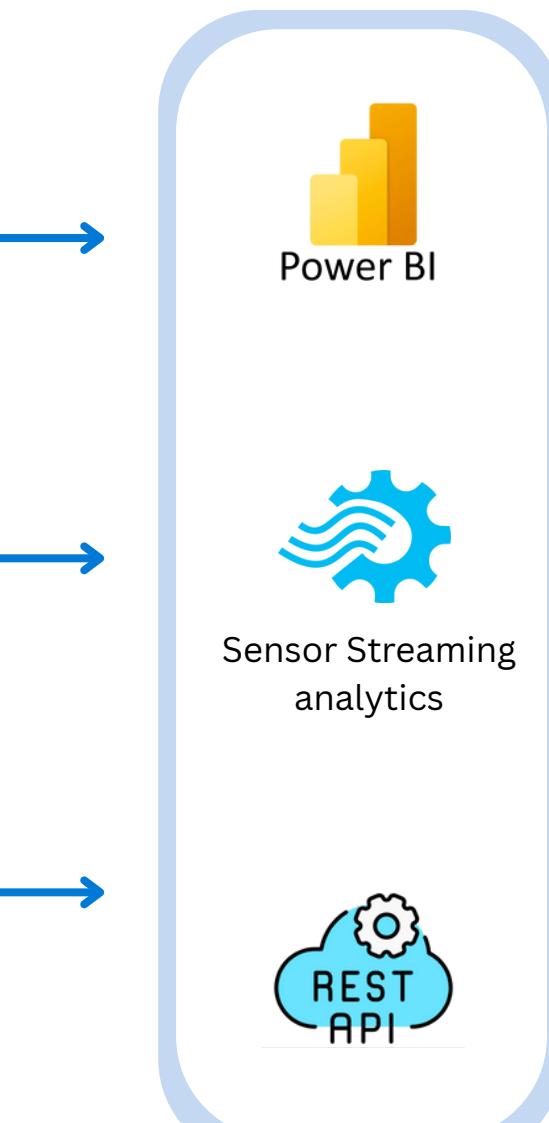
## INGESTION



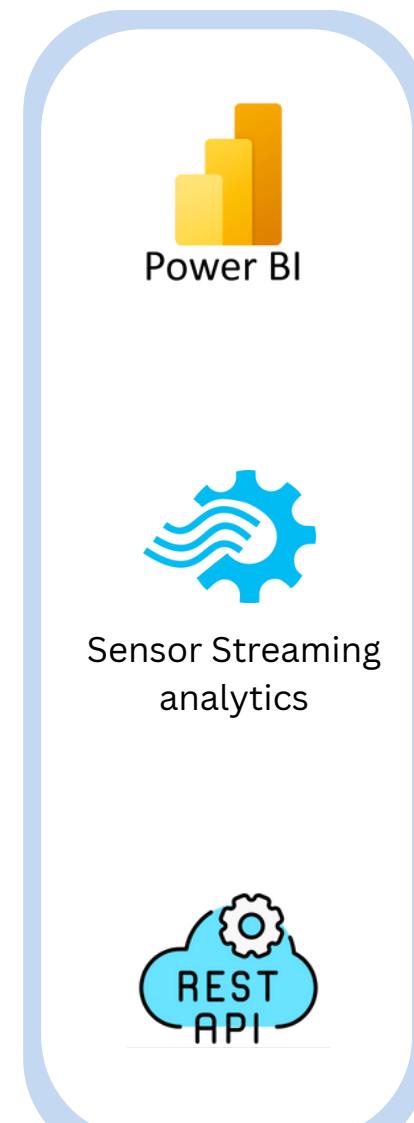
## STORAGE



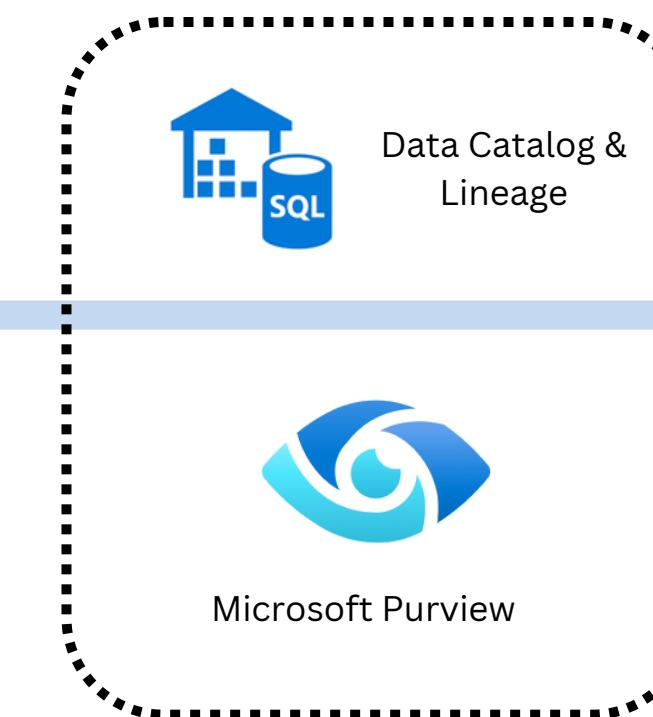
## PROCESSING



## CONSUMPTION



## GOVERNANCE & SECURITY



# Conclusion and CTA

- **Cloud-native & AI-ready platform**
  - Built for real-time data, machine learning, and automation
- **Secure & compliant foundation**
  - Encryption, access control, and audit tracking are built-in
- **Delivers real-time value**
  - Enables fraud detection, customer personalization, and smarter decisions
- **Modular, scalable, and future-proof**
  - Easy to expand as the bank grows or adds new services
- **Next Step: Start phased implementation**
  - Begin with infrastructure, then move to ingestion, processing, and ML integration

