Banco Wild West – Cloud-Native Data Platform Architecture

Overview:

Banco Wild West, a leading regional bank, has traditionally relied on distributed RDBMS systems to manage its operations, including customer information, transaction details, and loan processing. However, the current N-tier architecture, built on outdated C++ and Java middleware, struggles with high maintenance costs, frequent system failures, and limited scalability. This legacy infrastructure not only hampers innovation but also impacts customer experience, which lags behind that of its competitors. The newly appointed CIDO, Ms. Data First, has initiated a digital transformation aimed at modernizing the bank's data architecture on Microsoft Azure to meet the evolving needs of its customers and regulatory bodies.

The decision to adopt Azure Cloud is driven by its robust set of data services such as Azure Data Factory, Azure SQL Database, and Synapse Analytics, which enable seamless data integration, real-time processing, and advanced analytics. Azure's compliance with global financial regulations and its strong data security measures ensure that the bank remains compliant while scaling its operations. The new architecture is expected to support structured, semi-structured, and unstructured data, providing a Single Source of Truth (SSOT) for all customer profiles and financial transactions. Additionally, it aims to enhance customer interactions through predictive analytics, hyper-personalization, and AI-driven support.

By leveraging Azure's cloud-native capabilities, Banco Wild West aims to overcome its current limitations, achieve operational efficiency, and deliver a superior customer experience. This project outlines the strategic steps to design and implement this next-generation data platform, focusing on scalability, security, and compliance to position Banco Wild West as a leader in digital banking innovation.

Competitor Analysis

Why Azure is the Preferred Choice for Financial Services: In the financial services sector, cloud adoption is critical for scalability, compliance, and real-time analytics. Among major financial institutions, different cloud platforms are chosen based on their strategic needs. Here is how top banks leverage cloud technologies and why Azure stands out as the ideal choice for Banco Wild West.

Data Platform Adoption by Leading Banks:

- **JP Morgan Chase**: JP Morgan has largely adopted AWS for its cloud infrastructure, focusing on real-time risk analysis and data analytics. While AWS offers scalability and big data processing, its pricing model is often higher for financial-grade workloads due to data egress costs and multi-region replication charges.
- **Bank of America (BOA)**: Bank of America strategically selected Microsoft Azure for its digital transformation, primarily due to Azure's strong compliance standards, seamless integration

with Microsoft enterprise tools, and cost-effective storage options. Azure's hybrid capabilities also allow BOA to maintain on-premise control while expanding into cloud-native solutions.

- **Wells Fargo**: Wells Fargo employs a hybrid cloud strategy, utilizing both Google Cloud Platform (GCP) and Azure. GCP handles machine learning and advanced analytics, whereas Azure is leveraged for its compliance capabilities and powerful data lake architecture, supporting high-volume transaction processing.

This competitive analysis shows that while AWS and GCP are strong contenders, Azure emerges as the optimal choice for Banco Wild West due to the following factors:

1. **Cost Efficiency** - Azure's Reserved Instances and tiered storage models (Hot, Cool, Archive) significantly reduce costs compared to AWS's data egress and multi-region charges.
2. **Seamless Integration** - Azure's integration with Microsoft tools like Power BI, SQL Server, and Active Directory streamlines data management and security.
3. **Regulatory Compliance** - Azure meets over 90 compliance certifications globally, including financial regulations like PCI-DSS and ISO 27001, surpassing GCP in financial-grade compliance.
4. **Hybrid Capabilities** - Azure Arc allows Banco Wild West to extend cloud management to on-premises environments seamlessly, which is less mature in AWS and GCP.
5. **Advanced Security Features** - Services like Microsoft Defender, Key Vault, and Azure Sentinel enhance data protection and real-time threat monitoring.

These strengths make Azure not just a platform for cloud storage but a complete data ecosystem for financial institutions, providing scalability, security, and compliance at a lower cost than its competitors.

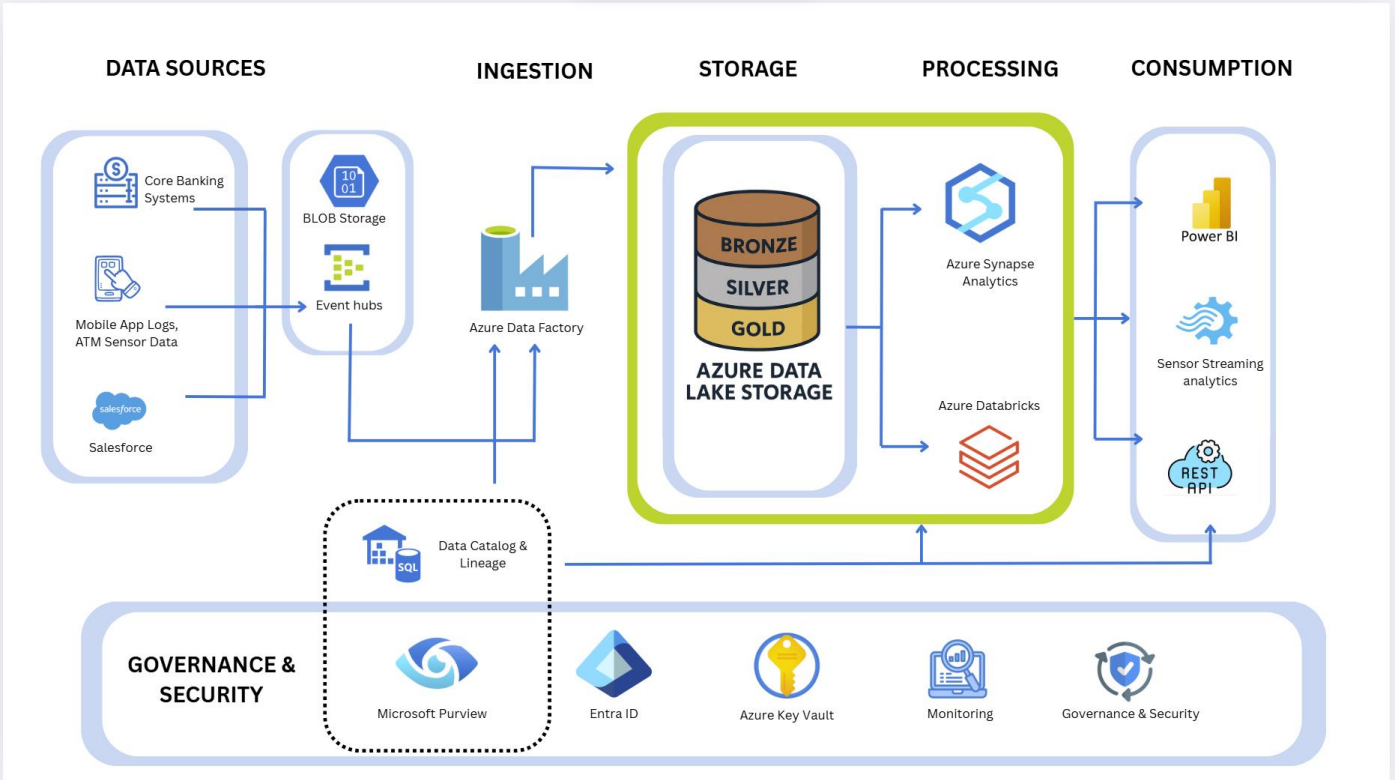
Our Approach to provide solution for Banco Wild West:

To modernize Banco Wild West's data infrastructure, we implemented a **cloud-native Medallion architecture** on **Azure Data Lake Storage Gen2** with **Delta Lake**, structured into Bronze (raw), Silver (cleansed), and Gold (curated) layers. This provides centralized, versioned, and scalable storage across all data types.

We adopted a **hybrid ingestion framework** using **Azure Data Factory** for batch data from systems like Oracle, SQL Server, and **Salesforce**, and **Azure Event Hubs with Databricks Structured Streaming** for real-time data from ATMs, mobile apps, and **Salesforce CDC events**. All data lands in the Bronze Layer for traceability and downstream processing.
In the **Silver Layer**, data is transformed and standardized using **Databricks** and **Delta Live Tables**, applying schema enforcement, deduplication, null handling, and PII masking. Access is managed with **Unity Catalog**, and lineage is tracked via **Microsoft Purview**.

The **Gold Layer** supports business intelligence, compliance, and AI. Curated data feeds **Power BI**, **Azure Synapse**, and **Azure ML** for real-time dashboards and predictive models. Integration with **Salesforce and Agentforce** enables AI-powered customer 360 profiles, fraud alerts, and next-best-action recommendations—streamed back via **Azure Functions** for seamless agent support.

Security and compliance are enforced using **Azure AD**, **RBAC**, **Key Vault**, and encryption protocols (TLS 1.2 / AES-256), aligned with BSA/AML, SOX, and IRS regulations. Lastly, **auto-scaling compute**, **tiered storage**, and **CI/CD pipelines** ensure cost-efficiency, flexibility, and scalability as the platform evolves.



Architectural Design Establishment for the Data Platform

1. Unified Data Access and Storage

1.1 Objective

Banco Wild West aims to eliminate data silos by implementing a unified, cloud-native data platform. This centralized system will integrate data from core banking, CRM, mobile apps, and social media into a scalable, secure foundation for seamless access and cross-departmental analytics.

1.2 Challenges in the Legacy Environment

The bank's legacy on-premises databases (e.g., Oracle, SQL Server) operated in silos, leading t Inconsistent formats and duplicate records, no single source of truth, high maintenance costs and slow insights, inability to handle unstructured or streaming data.:

1.3 Solution Design: Azure-Based Unified Storage Layer

To address these gaps, the proposed design implements a **Medallion architecture** built on **Azure Data Lake Storage Gen2** (ADLS Gen2) and **Delta Lake** format, delivering a consolidated, version-controlled, and query-optimized storage solution.

Key Layers (Medallion Architecture):

- **Bronze Layer – Raw Landing Zone**
  This layer captures **raw, unprocessed data** from multiple ingestion pipelines. It preserves the original fidelity of the data for traceability and auditing.
    - **Tools Used**:
        - *Azure Data Factory* (batch ingestion from SQL Server, Oracle, Salesforce, etc.)
        - *Azure Event Hubs* (real-time ingestion from ATMs, mobile apps, Salesforce CDC)
        - *Azure Databricks (Structured Streaming)* to land real-time streams into Delta tables
        - *ADLS Gen2 with Delta Lake* for scalable, immutable storage
    - **Tasks**: Metadata tagging, partitioning, append-only storage, schema-on-read
- **Silver Layer – Cleansed and Conformed Zone**
  In this layer, raw data is transformed, cleaned, and standardized into consistent, business-aligned schemas.
    - **Tools Used**:
        - *Azure Databricks* (Spark jobs, Delta Live Tables for transformation)
        - *Delta Lake on ADLS Gen2* (supporting upserts, schema enforcement, CDC)
        - *Unity Catalog* (fine-grained access control)
        - *Microsoft Purview* (data cataloging and lineage tracking)
    - **Tasks**: Type casting, MDM deduplication, referential joins, null handling, PII masking

- **Gold Layer – Business-Curated Zone**
  This is the presentation-ready layer for business analytics, dashboards, compliance, and ML pipelines.
    - **Tools Used**:
        - *Azure Databricks* (for aggregations, ML feature prep, model scoring)
        - *Azure Synapse Analytics (Serverless SQL pools)* for BI and ad-hoc queries
        - *Power BI* (connected directly to curated Delta tables)
        - *Azure App Services / Azure Functions* (for APIs exposing Gold tables)
        - *Azure Cosmos DB* (optional caching for low-latency access)
    - **Tasks**: Aggregation, enrichment, segmentation, metric derivation, ML scoring

Technologies Summary:

Component	Purpose
ADLS Gen2	High-throughput, scalable storage for structured and unstructured data
Delta Lake	Provides ACID transactions, schema enforcement, and time-travel capabilities
Azure Data Factory	Batch ingestion from core systems, CRM, and flat files
Azure Event Hubs	Real-time ingestion from mobile apps, ATMs, and Salesforce CDC events
Azure Databricks	Unified processing for batch and stream data into Delta format

1.4 Benefits of the Unified Storage Architecture

- **Single Source of Truth**: All data—structured, semi-structured, and unstructured—is housed in a centralized Delta Lake with full version control.
- **Scalable and Cost-Efficient**: Supports hot, cool, and archive tiers, allowing cost-optimized data retention for 5+ years.
- **Compliance Ready**: Supports lineage tracking, audit logs, and encrypted storage to meet BSA/AML, IRS, and NIST CSF requirements.
- **Flexibility for Diverse Use Cases**: Enables both OLTP (via APIs and Cosmos DB) and OLAP (via Power BI and Synapse) use cases from a unified foundation.
- **Governed Access**: Integrates with Microsoft Purview and Unity Catalog for metadata management and fine-grained access control.

1.5 Strategic Alignment

- This unified data storage modernizes Banco Wild West's infrastructure, enabling cross-functional analytics, faster go-to-market strategies, improved personalization, and reduced total cost of ownership.

Why ADLS Gen2 + Delta Lake?

Banco Wild West selected **Azure Data Lake Storage Gen2 (ADLS Gen2)** combined with **Delta Lake** as the core storage foundation to build a unified, scalable, and secure data platform. This combination brings the flexibility of a data lake with the reliability of a data warehouse—making it ideal for modern banking needs.

Key Benefits:

Feature	Why It Matters
Scalability & Performance	ADLS Gen2 handles petabyte-scale data with high throughput—suitable for handling logs from ATMs, mobile apps, and core banking systems.
Support for All Data Types	Can store structured (tables), semi-structured (JSON), and unstructured (PDFs, logs) data in one place.
Delta Lake = Reliability on Top	Delta Lake adds **ACID transactions**, **schema enforcement**, and **time travel**, turning the data lake into a trusted, versioned data source.
Batch + Streaming Support	Delta Lake supports **real-time streaming (Event Hubs)** and batch processing (Data Factory), simplifying pipelines.
Cost-Efficient	ADLS offers hot, cool, and archive storage tiers—allowing low-cost long-term storage (e.g., for IRS or audit retention).
Compliance and Audit Readiness	Delta Lake's versioning and metadata make it easier to support BSA/AML, IRS reporting, and internal audits.
Open Format	Built on Apache Parquet—ensures long-term compatibility, no vendor lock-in, and easier integration with Spark, ML tools, and BI platforms.

---

2. Prepping Data for Downstream Processes and Removing Irregularities

2.1 Objective

To support reliable analytics and compliance, Banco Wild West established a robust data cleansing strategy to transform raw, inconsistent data into accurate, structured datasets for trusted decision-making.

2.2 Key Challenges in Raw Data

Data from sources like core banking, ATMs, mobile apps, and Salesforce often arrives with issues such as missing values, duplicates, inconsistent formats, schema mismatches, and invalid references. Sensitive fields like PII may also require masking. If unaddressed, these irregularities can compromise report accuracy, model performance, and regulatory compliance.

2.3 Data Cleansing & Standardization Strategy Across Layers

Banco Wild West prepares clean and reliable data through a layered Medallion architecture—**Bronze**, **Silver**, and **Gold**—each playing a key role in handling data irregularities and ensuring downstream usability.

Bronze Layer – Raw Data Capture
Captures unprocessed data from various sources for full traceability.
It stores raw data with minimal transformation, preserving data fidelity for audit and historical

analysis. This layer sets the foundation for schema evolution and error tracking.
Tools: Azure Data Factory, Azure Event Hubs, Azure Databricks (Structured Streaming), Delta Lake on ADLS Gen2

Silver Layer – Cleansed and Conformed Data
Cleans and standardizes raw data for analytical readiness.
Handles type casting, deduplication, null handling, referential joins, and PII masking. This is the primary layer where data irregularities are fixed and trusted datasets are prepared.
Tools: Azure Databricks, Delta Live Tables, Unity Catalog, Delta Lake on ADLS Gen2, Microsoft Purview

Gold Layer – Business-Curated Data
Creates ready-to-consume datasets for BI, ML, APIs, and compliance.
Performs aggregations, calculations, and enrichments to deliver fast, accurate insights to business users and applications.
Tools: Azure Databricks, Azure Synapse Serverless SQL, Power BI, Azure Functions, Delta Lake on ADLS Gen2

Common Cleansing Steps and How They Help:

Step	**Purpose**	**Example**
Type Casting	Ensures data types are consistent	Convert transaction amounts to decimal(18,2)
Deduplication	Removes exact or fuzzy duplicate records	Drop repeated customer records using business key
Null Handling	Replaces or filters missing values	Fill nulls in address with “Unknown” or flag for review
Standardization	Unifies formats across systems	Format dates as yyyy-mm-dd; standardize phone numbers
Referential Integrity Checks	Validates relationships across datasets	Join transactions to customer and branch tables
Master Data Matching (MDM)	Merges duplicate entities across systems	Merge customer profiles from CRM and core banking
Quarantining Bad Records	Isolates corrupt/incomplete data for review	Move failed records to an error table with error reason

Step	Purpose	Example
PII Masking	Protects sensitive data from unauthorized access	Mask SSN or email before sharing data with analysts

2.4 Tools Used and How They Are Applied

Tool	Role in the Process
Azure Databricks	Used to write and run **data transformation logic** in PySpark and SQL. Handles schema standardization, deduplication, joins, and validation rules across millions of rows efficiently.
Delta Live Tables (DLT)	Automates ETL pipeline creation with **built-in quality checks**, schema enforcement, and monitoring. DLT applies transformations on Bronze-to-Silver data and flags failures with audit logs.
Delta Lake on ADLS Gen2	Stores cleaned and versioned **Silver Layer** data. Supports **upserts, schema evolution, and time travel**, allowing incremental improvements without losing data history.
Unity Catalog (Databricks)	Provides **row-level and column-level access control**, ensuring that only authorized users can see sensitive fields (e.g., masking PII for analytics users).
Microsoft Purview	Captures **metadata, lineage, and data quality scores**. Helps track where each field originated, what transformations were applied, and who accessed it—crucial for audits and governance.

3. Scalable Data Ingestion and Extraction

3.1 Objective

To meet real-time, omnichannel, and compliance needs, Banco Wild West requires a scalable ingestion framework to reliably handle high-volume, mixed-speed data from diverse legacy and modern sources.

3.2 Key Challenges in Legacy Ingestion

Banco Wild West's legacy ingestion was limited by siloed batch pipelines, manual ETL, and poor support for streaming data, hindering real-time use cases. The new Azure-based hybrid strategy enables scalable, secure, and efficient batch and streaming data ingestion.

3.3 Ingestion Architecture Design

Banco Wild West's ingestion design uses a **dual-path approach** to support both **bulk ingestion (batch)** and **event-driven ingestion (streaming)**. Both converge into the **Bronze Layer** of the Medallion architecture for traceable, versioned storage.

A. Batch Ingestion – Periodic High-Volume Data Loading

Use Case: Nightly loads from Oracle, SQL Server, Salesforce CRM exports, historical transactions, scanned documents, and third-party data vendors.

- **Toolset**: Azure Data Factory, Azure Databricks, Delta Lake on ADLS Gen2
- **Workflow**:
    1. ADF pipelines connect to RDBMS, FTP, REST APIs, and SaaS platforms.
    2. Data is loaded into Bronze tables in Delta Lake format with minimal transformation.
    3. Metadata (e.g., ingestion time, file status) is captured for auditing.
    4. Data can be replayed or retried in case of failures.
- **Features**:
    - Supports incremental loads with watermarks
    - Offers native connectors for 100+ enterprise data sources
    - Integration with Git for CI/CD deployment of pipelines

B. Streaming Ingestion – Real-Time Event Handling

Use Case: Ingesting ATM sensor logs, mobile app interactions, customer clicks, login events, and Salesforce CDC (Change Data Capture).

- **Toolset**: Azure Event Hubs, Azure Databricks (Structured Streaming), Delta Lake
- **Workflow**:
    1. Data from real-time sources is pushed into Azure Event Hubs (Kafka-compatible).
    2. Databricks Structured Streaming reads and processes this data in near real-time.
    3. Events are appended to Bronze Delta tables, triggering near-instant updates.
    4. Optional windowing, watermarking, and joins with historical Silver tables are applied.
- **Features**:
    - Sub-second latency for streaming analytics
    - Supports millions of events per second with horizontal scaling
    - Integration with fraud detection models and ATM alert systems

C. Data Extraction – Serving Cleaned Data to Consumers

Use Case: Providing cleaned and curated data to BI dashboards, ML models, compliance reports, and customer APIs.

- **Toolset**: Azure Synapse Serverless SQL, Power BI, Azure Databricks SQL, Azure Cosmos DB, Azure Functions, REST APIs
- **Workflow**:
    1. Curated Gold Layer tables are served directly to Power BI and dashboards via Databricks SQL or Synapse.
    2. APIs built with Azure Functions allow real-time access to customer 360 profiles or risk scores.
    3. Cosmos DB is optionally used for caching high-traffic lookup data (e.g., ATM status, last login).
    4. Synapse supports regulatory queries and ad-hoc analytics over curated datasets.
- **Features**:
    - Low-latency data delivery to apps and dashboards
    - Separation of storage and compute for cost efficiency
    - Integration with alert systems for proactive monitoring

3.4 Tools Used in Ingestion and Extraction

Ingestion Tools:
Azure Data Factory, Azure Event Hubs, Azure Databricks (Structured Streaming), Delta Lake on ADLS Gen2
Extraction Tools:
Azure Synapse Serverless SQL, Azure Databricks SQL, Power BI, Azure Functions, REST APIs, Azure Cosmos DB (optional)

3.5 Scalability and Resilience Features

Capability	**How It Helps**
Event Hubs Partitioning	Enables parallel ingestion of high-volume event streams
Databricks Auto-scaling Clusters	Dynamically scale compute based on load during peak ingestion windows
ADF Parallelism and Triggers	Supports concurrency and time-based or event-based execution
Delta Lake Time Travel	Replays and reprocesses ingestion without data duplication

Capability	How It Helps
Blob Storage Tiering (ADLS)	Cost optimization by moving older ingestion files to cool/archive tiers
CI/CD Deployment	Automates changes to ingestion pipelines via GitHub Actions or Azure DevOps

4. Data Security and Data Compliance

4.1 Objective

Banco Wild West aims to enforce end-to-end data protection and regulatory compliance (BSA/AML, IRS, SOX, NIST CSF) by embedding security, governance, and auditability across all data layers—ingestion, storage, processing, and consumption.

4.2 Key Legacy Challenges

Legacy systems lacked centralized access control, had manual policy enforcement, poor lineage visibility, and no unified auditing or encryption—posing compliance risks.

4.3 Security and Compliance Design

The modern architecture applies layered security using Azure-native tools for encryption, RBAC, auditing, data masking, and cataloging, ensuring compliance by design.

A. Data Access and Identity Management

- **Tools**: Azure Active Directory (Entra ID), Unity Catalog, Azure Role-Based Access Control (RBAC)
- **How It Works**:
    - Centralized identity management using Azure AD enables **single sign-on (SSO)**, multi-factor authentication (MFA), and conditional access.
    - Unity Catalog enforces **row-level and column-level access policies** across Delta tables.
    - RBAC ensures fine-grained access control across resources and data layers.

B. Data Encryption and Secure Storage

- **Tools**: Azure Key Vault, ADLS Gen2, Delta Lake
- **How It Works**:
    - Data is encrypted **at rest and in transit** using AES-256 and TLS 1.2 protocols.
    - Azure Key Vault securely stores secrets, connection strings, and encryption keys.

- Delta Lake ensures **transactional integrity and immutability**, critical for audit trails and forensic investigations.

C. Data Governance and Lineage Tracking

- **Tools**: Microsoft Purview, Unity Catalog
- **How It Works**:
  - Purview scans datasets, classifies PII/financial data, and builds **end-to-end lineage graphs**.
  - Tracks data flow from source (e.g., core banking system) to final usage (e.g., IRS report), enabling **traceable compliance**.
  - Enables regulatory and internal audits with versioning, schema history, and usage logs.

D. Regulatory Compliance Features

Regulation	How It's Addressed
BSA/AML	Real-time transaction monitoring and curated Gold tables for suspicious activity reports
IRS 1099/1042	Cleaned and aggregated data in Gold layer mapped to report schema
SOX / NIST CSF	Encryption, RBAC, audit logging, and version control across all data assets
GLBA / GDPR	PII masking, access auditing, and data retention policies via Unity Catalog and Purview

4.4 Tools Used for Data Security & Compliance

Access & Identity:
Azure Active Directory (Entra ID), Unity Catalog, Azure RBAC
Handles authentication, authorization, and access control across all services and datasets.

Encryption & Secrets Management:
Azure Key Vault, ADLS Gen2, Delta Lake
Ensures data is encrypted, securely stored, and protected from unauthorized access.

Governance & Lineage:
Microsoft Purview, Unity Catalog
Provides data classification, sensitivity labeling, and full data lineage for regulatory reporting and auditability.

5. Ease of Hooks for Integration

Banco Wild West has built its modern data platform to be **easy to connect with other systems**—whether it's internal apps, mobile banking, AI tools, or external services like Salesforce. This design makes it faster to build new features, automate workflows, and improve customer experiences.

- **Easy Access Through APIs**
  Key business data like customer profiles, agent performance, and fraud alerts is shared through secure, real-time APIs using **Azure Functions** and **API Management**. These APIs are used by web apps, mobile apps, and internal tools to retrieve the latest insights instantly.
- **Real-Time Event Triggers**
  Actions like ATM alerts, login events, or payment issues are captured using **Azure Event Hubs** and **Event Grid**. These trigger automated processes—like alerting fraud teams or updating CRM cases—without delay.
- **Salesforce CRM and Agentforce Integration**
  The platform connects directly to **Salesforce CRM** and **Agentforce (Service Cloud)** to support customer service and case management.
    - **Batch Sync**: Azure Data Factory pulls daily customer, agent, and case data into the platform.
    - **Real-Time Updates**: Salesforce **Platform Events** and **CDC** stream into **Event Hubs**, keeping the platform updated instantly.
    - **Agent Assist Use Case**: ML models process customer interaction data and send next-best-action suggestions back into Agentforce to help agents serve customers faster and smarter.
- **Self-Service for Business Users**
  Teams can explore and use trusted data using **Power BI**, **Databricks SQL**, and **Unity Catalog**, with access controls to ensure data is securely shared only with the right users.
- **Model Access and Integration**
  The platform integrates with **Azure Machine Learning** and **MLflow**, so data scientists can easily manage and deploy models. These models can be used across systems—like APIs or Agentforce—to support fraud detection, loan risk scoring, and personalization.
- **Secure External Sharing**
  The platform is future-ready for **Delta Sharing**, allowing safe, controlled data sharing with partners and regulators—while keeping full control and audit logs.
- **Data Organization and Tracking**
  Microsoft Purview catalogs all APIs and datasets, providing clear visibility, data lineage, and documentation for easy discovery and governance.

6. Optimal Storage Usage for Different Use Cases

6.1 Objective

Banco Wild West aims to manage data efficiently by using the **right type of storage for the right purpose**. The goal is to reduce costs, maintain performance, and ensure that data is stored in a way that fits its usage—whether it's for real-time apps, historical analysis, compliance reporting, or machine learning.

6.2 Storage Challenges in Legacy Systems

In the past, the bank used one-size-fits-all storage solutions, which caused:

- High costs from storing rarely used data in expensive systems
- Slow performance when querying large historical records
- Manual data archiving without access rules or automation
- Inability to scale with increasing data from digital banking, CRM, and ATMs

6.3 How Banco Wild West Uses Storage Tiers Optimally

To balance **performance and cost**, the bank uses **three ADLS storage tiers**:

- **Hot Tier** – Stores frequently used data (e.g., current KPIs, customer profiles, fraud alerts).
- **Cool Tier** – Stores semi-active data (e.g., weekly performance logs or past month's CRM cases).
- **Archive Tier** – Stores historical or compliance data (e.g., ATM logs from last year, audit records).

With **automated lifecycle policies**, data is moved between tiers based on usage. This ensures fast access to what's needed now while saving costs on older or infrequent data—without manual work.

6.4 Real-World Use Cases and How Storage Is Optimized

Banco Wild West's storage architecture is designed not just for cost and performance, but also to support high-impact business use cases like fraud detection, ATM operations, and personalized customer service. Here's how the right storage tier and layer is chosen for each scenario:

A. Real-Time Fraud Detection (Gold + Hot Tier)

Credit card transactions stream in via Azure Event Hubs and are processed in real time using Databricks Structured Streaming and fraud detection models.

- Bronze Layer captures raw swipe data for traceability.
- Silver Layer enriches it with customer and location data.
- Gold Layer stores curated fraud risk scores and alerts in the Hot tier for instant API/dashboard access.
  Why it matters: Enables immediate fraud response and prevents financial loss.

B. ATM Health Monitoring (Bronze + Archive/Cool Tier)

ATM sensor logs ingested via Event Hubs are stored in the Bronze Layer and used in Databricks for predictive maintenance.

- Recent data stays in Hot/Cool tier for active analysis; older logs move to Archive for compliance.
- Real-time alerts are sent to operations dashboards.
  Why it matters: Improves ATM uptime, reduces field visits, and enhances customer trust.

C. Customer Personalization with CRM and Agentforce (Gold + Hot Tier)

Behavioral and CRM data is streamed and processed to drive personalized recommendations.

- Silver Layer combines behavior and demographics.
- Gold Layer produces churn scores and product suggestions, pushed via real-time APIs to Salesforce CRM and Agentforce.
- Agentforce's AI guides support agents with next-best actions, stored in Hot tier for instant use.
  Why it matters: Boosts loyalty through timely, tailored customer experiences during key interactions.

7. Cost Efficiency Without Reducing the Effectiveness of Use Cases

Banco Wild West uses the cloud-native, modular design of Azure to implement the following cost-saving strategies:

A. Pay-As-You-Go Compute

Tools: Azure Databricks (auto-scaling clusters), Azure Synapse Serverless SQL
How It Helps:
Compute resources scale with workload needs. Fraud models run on-demand and auto-terminate after use. Serverless SQL allows querying without always-on infrastructure.

B. Tiered Storage Strategy

Tools: ADLS Gen2 (Hot, Cool, Archive tiers), Delta Lake
How It Helps:
Frequently accessed data is kept in Hot tier, while older data like ATM logs or audit records moves to lower-cost Cool or Archive tiers automatically, saving storage costs.

C. Unified Storage and Processing (Delta Lake)

Tools: Delta Lake on ADLS Gen2
How It Helps:
Supports both real-time and batch processing with one format. Avoids duplicating data, simplifies architecture, and ensures reliability through ACID compliance.

D. Reusability of Pipelines and Data Layers

Tools: Databricks, Delta Live Tables (DLT)
How It Helps:
Silver and Gold tables are reused for multiple needs (dashboards, models, reports), so teams don't have to build separate data flows—saving development and compute effort.

E. Self-Service and API-Driven Access

Tools: Power BI, REST APIs, Databricks SQL, Unity Catalog
How It Helps:
Teams access data directly without IT involvement. Reduces delays, duplication, and support workload.

F. Salesforce + Agentforce Integration

Tools: Salesforce CRM, Agentforce, Event Hubs, Azure Data Factory
How It Helps:
Centralizes customer data and support processes on a single platform. Agentforce improves productivity by guiding agents with AI-powered next-best actions.
Automated case routing, faster resolution, and reduced call times lead to **lower staffing and training costs**, while also improving customer satisfaction and retention—making Salesforce a **cost-efficient investment**.

Maintaining Use Case Effectiveness: Despite cost optimization, the platform fully supports high-impact use cases like real-time fraud detection, ATM health monitoring, customer personalization (Salesforce + Agentforce), compliance reporting, and AI-driven support—powered by shared Gold-layer data and efficient compute/storage.

8. Scalability in Overall Design to Accommodate Future Business Growth

Proposed Solution:

The project's architecture has been intentionally designed with horizontal scalability, cloud-native flexibility, and modular service composition to support the bank's future growth in customer volume, data size, and feature expansion.

Execution Using Our Data Architecture:

1. Cloud-Native Infrastructure (AWS/Azure):
    - Core components like databases, case management, logging, and analytics are deployed using managed services (e.g., RDS, Cosmos DB, Lambda, Azure Functions) that scale automatically with demand.
    - Use of infrastructure-as-code (Terraform, Bicep) ensures environments can be cloned or extended across regions easily.
2. Microservices & Event-Driven Design:

    - Each functional unit (e.g., customer case processing, notification service, audit logging) is decoupled via queues or event buses.
    - This allows independent scaling—so surges in customer onboarding don't overwhelm fraud-checking or reporting.
3. Scalable Data Pipeline:
    - The data ingestion and processing layer (e.g., using Kafka, Kinesis, or Azure Event Hub) supports high-throughput workloads.
    - Data lakes and warehousing use elastic compute (Redshift, Synapse) so analytics workloads can grow without infrastructure limits.
4. Multi-Tiered Storage:
    - Historical data is offloaded to cold storage (e.g., S3 Glacier, Azure Archive), keeping hot storage lean and fast.
    - This design reduces cost while keeping growth sustainable.
5. API & App Layer Scalability:
    - APIs are hosted on scalable container platforms (e.g., ECS/Fargate, AKS) or serverless platforms.
    - Frontend or admin dashboards can be replicated globally using CDNs and multi-region deployment.

Bank Benefit:

- Future-Proofing: Easily onboard new services, products, or regions without overhauling existing systems.
- Cost Efficiency: Pay-as-you-grow model ensures the bank only pays for resources used.
- Operational Agility: Supports growth in both user base and operational complexity without sacrificing performance or security.
- Rapid Innovation: Modular architecture allows faster integration of AI/ML and fintech features as business evolves.

9. Execution Roadmap:

Phase 1: Infrastructure & Governance Setup *(Weeks 1–3)*
Goal: Lay down secure, scalable cloud foundations and governance for future modules.
Key Tasks:

- Provision core resources: Azure subscriptions, VNet, ADLS Gen2, Key Vault, Azure AD.
- Set up DevOps CI/CD pipelines using Terraform/Bicep + GitHub Actions/Azure DevOps.
- Define RBAC/ABAC roles, encryption protocols (TLS 1.2, AES-256).

- Deploy Microsoft Purview for cataloging, lineage, and PII classification.
- Enforce compliance baseline (BSA/AML, SOX, NIST CSF).

Outcomes:
Secure and governed foundation ready for scalable data onboarding and processing.

Phase 2: Data Ingestion & Bronze Layer Setup *(Weeks 4–6)*
Goal: Build batch and streaming pipelines to land raw data in a traceable format.
Key Tasks:

- Set up Azure Data Factory for batch loads (Oracle, SQL Server, Salesforce).
- Configure Azure Event Hubs + Databricks Structured Streaming for real-time ingest.
- Land data into Bronze Layer in Delta Lake format on ADLS Gen2.
- Apply metadata tagging, partitioning, and schema-on-read.
- Monitor with DLT pipelines for error handling & retry logic.

Outcomes:
All raw data (structured, semi-structured, unstructured) lands into a centralized, versioned store.

Phase 3: Silver Layer - Data Cleansing & Transformation *(Weeks 7–9)*
Goal: Standardize and conform raw data for analytical readiness.
Key Tasks:

- Use Azure Databricks (PySpark/SQL) and Delta Live Tables for transformations.
- Clean, deduplicate, handle nulls, enforce referential integrity, and mask PII.
- Catalog schemas and lineage in Microsoft Purview.
- Apply fine-grained access policies with Unity Catalog.

Outcomes:
Validated, trusted Silver datasets prepared for consumption and downstream modeling.

Phase 4: Gold Layer + BI/ML Integration *(Weeks 10–12)*
Goal: Deliver business-ready curated data to analytics, compliance, and AI services.
Key Tasks:

- Aggregate and enrich Silver data into Gold layer using Databricks.
- Serve Gold data to Power BI, Azure Synapse Serverless SQL, and APIs via Azure Functions.
- Integrate ML pipelines (fraud detection, CLV, document OCR) using Azure ML & MLflow.
- Sync curated data with Agentforce via real-time APIs.

Outcomes:
Business, compliance, and AI use cases powered with clean, curated data.

Phase 5: Optimization, MLOps & Monitoring *(Weeks 13–15+)
Goal: Stabilize, optimize, and enable continuous improvement and monitoring.
Key Tasks:

- Establish MLOps for model deployment, monitoring, and retraining.
- Apply Azure Monitor and Cost Management for performance + budget tracking.
- Implement Delta Sharing for controlled external data exchange.
- Conduct stakeholder training (BI, ML, compliance teams) on self-service access.
- Enable dynamic lifecycle policies (hot/cool/archive) for cost efficiency.

Outcomes:
Production-grade, cost-optimized, continuously monitored platform with reusable assets and scalable governance.

10. ML Use cases:

1. Fraud detection

Banco Wild West's fraud detection system leverages real-time streaming data ingested via Azure Event Hubs and processed using Databricks Structured Streaming, with fraud scoring models deployed through Azure Machine Learning. Events from ATMs, mobile apps, and card swipes are evaluated against trained models in near real-time, enabling immediate fraud alerts. Data flows through the Medallion architecture (Bronze → Silver → Gold), ensuring auditability and trusted outputs. The system is cost-efficient through auto-scaling compute and hot-tier storage, while also being governed with Unity Catalog and Purview for compliance.

2. Agent for customer service - Salesforce ML

Using real-time integration with Salesforce CRM and Agentforce, Banco Wild West delivers AI-driven "next-best-action" suggestions to support agents during customer interactions. Streaming data from digital channels and Salesforce events is processed in Databricks, and recommendations are pushed back via Azure Functions APIs. This seamless flow enhances personalization, boosts productivity, and reduces call times. Data is securely managed via Delta Lake and Unity Catalog, while Power BI and Salesforce dashboards ensure business teams can monitor outcomes—maximizing service quality at minimal cost.

3. Document Checking automatic authentication

For high-volume document intake such as loan applications or ID proofs, Banco Wild West applies OCR and NLP models using Azure ML and Form Recognizer to automate validation. Uploaded files land in the Bronze Layer (ADLS Gen2) and are processed using Databricks notebooks. Extracted text is analyzed for compliance, identity matching, or anomalies. Failed validations are quarantined for

manual review, ensuring audit readiness. This pipeline reduces manual workload, increases speed, and lowers operational cost with serverless compute and cool-tier storage for archived files.

4. Predictive Maintenance of Core Banking Services

By analyzing logs from ATMs, servers, and backend applications streamed via Event Hubs, the platform uses ML models in Azure Databricks to predict failures in core banking systems. These predictions trigger alerts through Azure Functions, allowing proactive maintenance. The solution is built on a scalable Medallion framework with Delta Lake time-travel and Purview lineage ensuring traceability. Storage tiering (hot for current alerts, archive for old logs) and Databricks auto-scaling clusters make it both effective and cost-conscious.

5. Customer Lifetime Value (CLV) Prediction

CLV prediction models in Azure ML analyze cleansed Silver-layer data (CRM, transaction history, product usage) to forecast revenue potential across customer segments. Outputs are stored in Gold-layer Delta tables and surfaced via Power BI and Salesforce dashboards for marketing and product teams. Real-time behavioral data is integrated through Event Hubs and Databricks, ensuring models are continuously updated. The approach supports precision targeting and reduces churn while optimizing compute via on-demand resources and shared pipelines across use cases.

6. Smart Document Processing (OCR + NLP)

Banco Wild West processes large volumes of unstructured documents (e.g., contracts, complaints, feedback) using OCR and NLP pipelines in Azure Databricks and Form Recognizer. Documents are ingested into the Bronze Layer, text is extracted and analyzed for sentiment, topics, or risk flags, and enriched data is stored in the Gold Layer for business use. Integration with Microsoft Purview tracks document lineage and access, ensuring compliance. This automated, serverless architecture reduces manual review cost while enabling insights from previously untapped data sources.

Conclusion:

The adoption of Microsoft Azure as the primary cloud platform for Banco Wild West represents a strategic leap towards modernizing its data architecture. By leveraging Azure's robust cloud services, Banco Wild West is positioned to achieve scalable, secure, and compliant data operations that align with both regulatory demands and evolving customer expectations. The transition addresses the limitations of legacy infrastructure, enabling real-time data processing, predictive analytics, and enhanced digital banking experiences. As the platform matures, Banco Wild West is set to not only improve its operational efficiency but also establish a competitive edge in the rapidly evolving financial sector.

References:

Understanding Azure:

Introduction to Azure Data Lake Storage Gen2
Delta Lake Documentation
Azure Event Hubs Documentation
Azure Data Factory Documentation
Delta Live Tables - Azure Databricks
Microsoft Purview Documentation
Azure Synapse Analytics Documentation
E-learning Course PPT

Pricing Comparison: https://cast.ai/blog/cloud-pricing-comparison/

Salesforce Azure integration:
https://help.salesforce.com/s/articleView?id=ind.sf_contracts_integrate_microsoft_365_azure_with_salesforce.htm&type=5

- Ingest Data from Salesforce using Azure Data Factory
- Salesforce Platform Events Documentation
- Salesforce + Azure Event Hubs CDC Integration Example

Agent force for Finance: https://www.salesforce.com/financial-services/artificial-intelligence/#:~:text=Agentforce%20for%20Financial%20Services%20is,wealth%20management%2C%20and%20insurance%20firms.

Microsoft Azure in Financial Services

Why Bank of America Chose Azure

Wells Fargo's Cloud Strategy : https://www.zdnet.com/article/wells-fargo-to-use-multiple-public-clouds/