

Data Engineering Challenge Presentation

This presentation outlines a data engineering solution for the banking sector. We address key challenges in data management and processing. Our goal is to create a scalable, efficient, and secure data architecture. This will meet integration, data quality, and real-time processing needs.



by Johanna Rangel



Data Source Constraints



Annual Tax Declarations

Third-party XML API with limited access.



Credit Card Transactions

JSON API and internal PostgreSQL databases.



Bank Statements

XML API and unstructured files in Amazon S3.

Stakeholder Needs



Data Analysts

Self-service access to consolidated data.



Decision Science

Processed data for credit risk models.



Fraud Team

Real-time transactional data for fraud detection.



Current State (AS-IS)

Disconnected Systems

Data silos hinder access and collaboration.

Batch Processing

Manual processes limit real-time capabilities.

Limited Real-Time

Delays affect analysis and reporting.

Key Challenges



Integration Complexity

Connecting diverse sources presents significant hurdles.



Data Quality Issues

Inconsistencies undermine data reliability.



Latency Problems

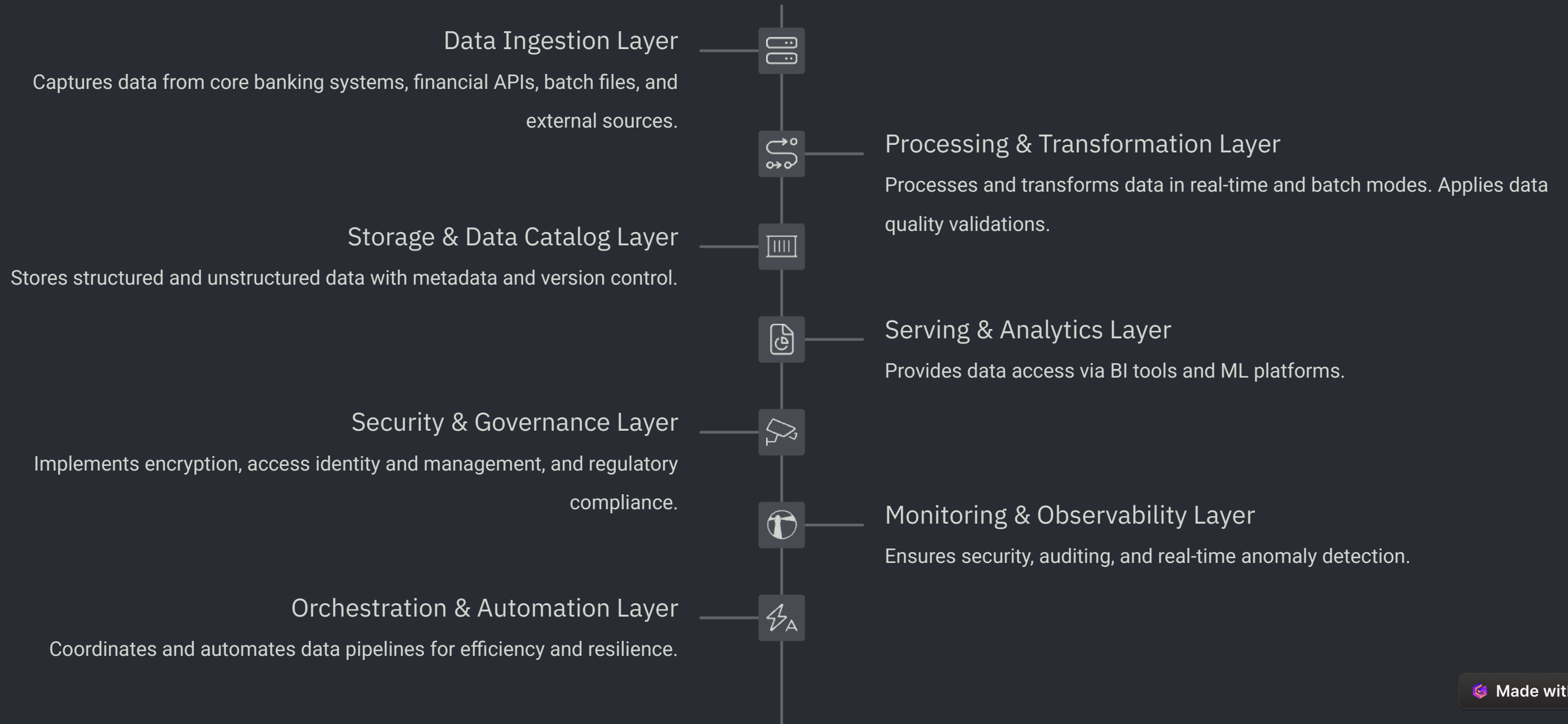
Real-time availability is often lacking.



Security Risks

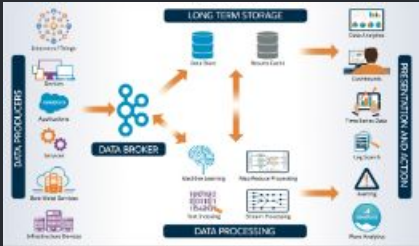
Controlled access and data handling are crucial.

Proposed Solution (TO-BE)



Technical Implementation - Stack

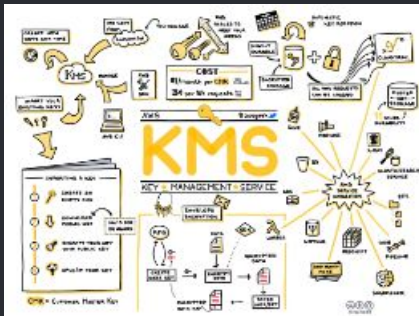
Data Ingestion Layer



Storage & Data Catalog Layer



Security & Governance Layer



Orchestration & Automation Layer



Processing & Transformation Layer




Serving & Analytics Layer



Monitoring & Observability Layer





Expected Benefits

30%

Efficiency

Reduced processing times.

99%

Real-Time

Faster decision-making.

scalable

Scalability

Handles increased data load.

compliant

Compliance

Strict data security.



Conclusion & Next Steps

1 **Phased Implementation**
Begin with a pilot project and expand iteratively.

2 **Testing & Optimization**
Ensure performance and data quality.

3 **Continuous Monitoring**
Maintain system health and compliance.