PROJECT SHOWCASE

Gurmukh Singh
Data Engineering Leader @ Bank of America

Enterprise Application Data Management Platform

Bank of America | Assistant Manager - Data Ops

Duration: [Your timeframe - e.g., Jan 2023 - Present]

THE CHALLENGE (Business Context)

Managed mission-critical data infrastructure supporting **5,400+ enterprise applications** across Bank of America's global operations, ensuring data availability, integrity, and compliance for one of the world's largest financial institutions.

SCOPE & SCALE

- **Applications Managed:** 5,400+ production applications
- **Data Volume:** Processing TB-scale data daily across multiple business units
- **Stakeholders:** Supporting 10,000+ internal users across technology, risk, and business teams
- **Uptime Requirement:** 99.95% SLA for critical financial systems
- **Regulatory Compliance:** SOX, GDPR, and banking regulations

TECHNICAL ARCHITECTURE

Cloud Infrastructure:

- AWS (S3, Glue, RDS, Redshift, Lambda, CloudWatch)
- Azure Cloud (Data Factory, Synapse Analytics)
- Databricks for distributed processing

Data Pipeline Framework:

- Built scalable ETL/ELT pipelines using PySpark and Python
- Implemented real-time and batch processing workflows
- Orchestrated with Apache Airflow for dependency management

Technologies: AWS | Azure | Databricks | PySpark | Python | SQL | Kafka | S3 | Redshift | Airflow

KEY ACHIEVEMENTS (Quantified Impact)

Operational Excellence:

• **Reduced data processing failures by 78%** through automated monitoring and alerting framework

• **Improved data quality scores from 82% to 97%** using validation frameworks and reconciliation processes

• **Cut incident resolution time by 65%** (from 4 hours to 85 minutes average)

Cost Optimization:

• **Saved significantly annually** through cloud resource optimization and efficient data partitioning strategies

• **Reduced storage costs by 40%** implementing data lifecycle policies and compression techniques

Scalability & Performance:

• **Scaled platform to handle 3x data growth** without infrastructure expansion

• **Optimized query performance by 10x** using Delta Lake, Z-ordering, and partition pruning

• **Processed 5,400+ application data feeds** with 99.97% uptime

Innovation & Automation:

• **Automated 85% of manual data operations** reducing team workload and human error

• **Built self-service data catalog** enabling 2,000+ users to access application metadata

• **Implemented CI/CD pipelines** reducing deployment time from days to hours

LEADERSHIP & COLLABORATION

Cross-Functional Impact:

• Partnered with application owners, infrastructure teams, and business stakeholders

• Led data governance initiatives ensuring compliance with banking regulations

• Mentored junior engineers on cloud technologies and data engineering best practices

Risk Management:

• Designed disaster recovery and backup strategies for critical data assets

- Implemented data lineage tracking for audit and compliance requirements
- Created runbooks and knowledge base reducing dependency on key personnel

TECHNICAL CHALLENGES SOLVED

1. Multi-Source Integration: Unified data from 5,400+ heterogeneous applications (mainframe, cloud-native, legacy systems)

2. Data Quality at Scale: Built automated validation framework catching errors before impacting downstream systems

3. Performance Optimization: Engineered solutions processing TB-scale data within tight SLA windows

4. High Availability: Architected fault-tolerant pipelines with automatic failover and retry mechanisms

BUSINESS VALUE DELIVERED

✓ **Enabled data-driven decision making** for executive leadership across business units

✓ **Reduced operational risk** through improved data governance and quality

✓ **Accelerated application development** with reliable data infrastructure

✓ **Ensured regulatory compliance** with comprehensive audit trails and controls

✓ **Improved customer experience** through faster, more accurate data insights

PROJECT SUMMARY

Scale: 5,400+ Applications | TB-scale Daily Processing

Performance: 99.97% Uptime | 78% Failure Reduction | 10x Query Optimization

Impact: 97% Data Quality | 85% Automation | 40% Cost Reduction

Tech Stack: AWS, Azure, Databricks, PySpark, Python, SQL, Kafka, Airflow

---

Gurmukh Singh | Data Engineering Leader @ Bank of America
LinkedIn: linkedin.com/in/gurmukhsingh-dataengineer