KUSH KUMAR
Pune, India
4x Databricks | 2x Azure | 1x Fabric

• • LinkedIn • GitHub • Certifications

Summary

Data Platform Architect/Senior Data Engineer with 5+ years of experience designing and building scalable data platforms using **Azure**, **Databricks**, and **Microsoft Fabric**. Specialized in **Architecting High-Performance** lakehouse pipelines processing **25TB+** datasets and enabling analytics, BI, and ML workloads. Proven track record of improving platform reliability, reducing compute costs, and delivering enterprise-grade data solutions across **Finance**, **Insurance**, **Telecom**, **Healthcare**, and **AI/ML** domains.

Skills

Programming: Python, PySpark, Scala, SQL

Cloud & Data Platforms: Microsoft Azure, Microsoft Fabric, Databricks

Processing & Big Data: Apache Spark, Hadoop, Kafka

Storage & Table Formats: Azure Data Lake Storage (Gen1/Gen2), OneLake, Delta Lake

Data Integration & Orchestration: Azure Data Factory, Fabric Data Pipelines, Dataflow Gen1/Gen2, Apache Airflow

Databases: SQL Server, MySQL, PostgreSQL, MongoDB

Data Architecture & Modeling: Lakehouse Architecture, Medallion Architecture, Data Modeling, Data Marts, ETL/ELT

DevOps & Infrastructure: Git, GitHub, Azure DevOps, Terraform, Docker, Jenkins

Monitoring & Observability: Azure Table Storage Logging, Pipeline Monitoring, Circuit Breaker Frameworks, Grafana

Security & Governance: RBAC, Managed Identity, Azure Key Vault, SHA-2 Encryption

Analytics & Visualization: Power BI, Metabase

Key Achievements

- Reduced pipeline cost from **$10 → $0.02 per run**
- Led migration from **Azure Eco-system → Microsoft Fabric**
- Improved pipeline reliability from **~70% success rate → 99.99% uptime**
- Reduced dashboards refresh **40 min → 4 min**
- Built pipelines processing **25TB+ datasets**

Work Experience

[redacted] **CALIFORNIA, USA**

Software Engineer – Data Engineer

Designed and scaled enterprise lakehouse's data platforms supporting financial analytics and reporting workloads across multiple products.

- **Led architecture standardization across multiple data platforms**, collaborating with data engineering and analytics teams to establish reusable lakehouse patterns and pipeline design practices.
- **Architected multi-hop lakehouse pipelines using Azure Data Factory, Databricks (PySpark), Delta Lake, and Microsoft Fabric**, processing **25TB+ historical datasets and ~100GB daily financial data ingestion** for scalable analytics.
- **Improved pipeline reliability from ~70% success rate to 99.99% uptime** by designing circuit-breaker retry frameworks and implementing proactive monitoring with automated Microsoft Teams alerts.
- **Reduced pipeline execution cost from $57 to $23 per day and from $10 to ~$0.02 per run** through compute optimization, orchestration tuning, and workload scheduling improvements.

- **Enhanced query performance by 95%+** through Delta Lake partitioning strategies and optimized data access patterns across enterprise datasets.
- **Built organization-wide pipeline observability dashboards using Azure Table metrics**, enabling leadership visibility into pipeline health, performance, and operational insights.
- **Designed secure ingestion pipelines for confidential financial datasets using SHA-2 encryption**, enabling compliant processing of sensitive information.
- **Led lakehouse modernization to Microsoft Fabric**, implementing medallion architecture with Dataflow Gen2 and OneLake, reducing pipeline costs by ~30% and improving dashboard load time **from ~1 hour to under 3 minutes**.

[redacted] **NEW JERSEY, USA**

Data Engineer

- **Optimized NLP-based document classification pipelines using SpaCy Phrase Matcher and LLM prompt engineering**, reducing manual policy classification from **51% to 0.3%**.
- **Improved data processing performance by 90%** by implementing parallelized Spark transformations, reducing execution time from **5–7 minutes to under 60 seconds**.
- **Integrated LLM-driven data workflows with Azure services**, improving debugging efficiency and operational visibility by **~50%**.

[redacted] **PUNE, IN**

Data Engineer

- **Optimized ETL workflows and resource utilization**, reducing operational costs by **~20%** across enterprise data pipelines.
- **Designed and implemented configuration-driven data pipelines**, improving reliability and maintainability of large-scale analytics workflows.

[redacted] **BENGALURU, IN**

Data Analyst

- **Executed regression and non-functional testing across multiple enterprise releases**, reducing bug resolution time by **~67%**.
- **Optimized SQL queries and database workflows**, improving response times for healthcare appointment management systems.

[redacted] **BRUSSELS, BE**

Test Engineer

- **Recognized as Employee of the Year after identifying 160 critical defects in Android 10/11 systems**, significantly improving product stability and extending project lifecycle.

[redacted] **PUNE, IN**

Junior Web Developer

- **Developed responsive client websites and improved application reliability**, reducing functional errors by **~5%**.

Education

[redacted] [redacted]

[redacted]

Master's Program (*Big Data Major*) 2022

Simplilearn

Bachelor of Engineering (*IoT Major*) 2018

University of Pune

Personal/Other

Languages: English (Fluent/C2), French(A2/B1).

Community: Non–Profit Volunteer/Educator: Robin Hood Army

Additional Interest: Passionate about history, travel, and creating positive impacts in people's lives.