PRAMOD KUMAR

Lead Data Engineer

Salisbury, NC | Pramodsk369@gmail.com | pramodsola.com | github.com/pramodsola

PROFESSIONAL SUMMARY

Lead Data Engineer with 10+ years of experience across enterprise ETL, data engineering, and Azure-based data platforms. Designs and builds batch, continuous streaming, and event-driven solutions using Azure Data Factory, Databricks, PySpark, Spark SQL, Delta Lake, ADLS Gen2, Event Hubs/Kafka, Service Bus, and Azure Functions. Provides technical direction across solution design, code reviews, production support, platform modernization, and cross-team delivery. Experienced in unified data governance with Unity Catalog (access controls, lineage), serverless modernization, and applied GenAI/LLMOps.

CORE TECHNICAL SKILLS

Azure & Data Platform: Azure Data Factory, ADLS Gen2, Azure Functions, Event Hubs, Service Bus, Logic Apps, Event Grid, Key Vault, Managed Identity, RBAC, Azure SQL, Synapse Analytics, Self-hosted Integration Runtime

Databricks & Spark: PySpark, Spark SQL, Structured Streaming, Auto Loader, Delta Lake / MERGE, Delta Live Tables (DLT) / Lakeflow Declarative Pipelines, Lakeflow Jobs, Databricks Workflows, Serverless Compute, Photon, cluster/runtime modernization, performance tuning

Databricks Governance & Data Intelligence: Unity Catalog, unified data & AI governance, access controls, lineage, catalogs/schemas/volumes, Delta Sharing, AI/BI Dashboards, Genie One, Genie Agents, Genie Code

Databricks GenAI / Agentic AI: AI Search (Vector Search), RAG, embeddings, Model Serving, Foundation Model APIs, Custom Agents, MCP Services, Databricks Apps, Lakebase, Unity Gateway / AI Gateway, MLflow 3 GenAI tracing/evaluation/monitoring, LLM judges, prompt & agent observability

DevOps & Platform Engineering: Databricks Asset Bundles, Azure DevOps, Git/GitHub, CI/CD, REST APIs / SDKs, environment-aware DEV/QA/PROD deployments, orchestration and operational recovery

Streaming & Integration: Apache Kafka, Event Hubs, Azure Service Bus topics/subscriptions/filters, Google Cloud Pub/Sub topics/subscriptions, checkpointing, offset management, schema evolution, idempotent / at-least-once processing, REST APIs

Data Formats & Interfaces: Delta / Delta Lake, Parquet, JSON, XML, CSV, delimited/text files, relational and semi-structured data, REST APIs

Engineering, Observability & Reliability: Python, SQL/T-SQL, shell scripting, Azure Monitor, Application Insights, Log Analytics, KQL, ADF/Databricks monitoring, audit/control tables, reconciliation, data quality, restartability, alerting and RCA

PROFESSIONAL EXPERIENCE

Large U.S. Grocery Retailer | **Salisbury, NC** **Mar 2020 – Present**

Senior Azure Data Engineer | Retail Supply-Chain Data Platform

- **Data platform design:** Designed and delivered Azure and Databricks data solutions for retail supply-chain workloads, translating business and non-functional requirements into practical ingestion, transformation, orchestration, security and operational patterns.
- **Event-driven processing:** Designed facility-level batch-completion flows using Event Hubs, Azure Functions, ADF, Databricks and Control-M to trigger downstream processing from business events while retaining scheduled operational controls and graceful handling of incomplete batches.
- **Streaming engineering:** Built Event Hubs/Kafka ingestion with Spark Structured Streaming and Auto Loader across DELTA, JSON, Parquet, CSV and other structured/semi-structured data, including checkpoint/offset management, partitioning, schema evolution, large-message handling and reliable restart/recovery.
- **Batch + streaming delivery:** Selected scheduled batch, micro-batch or event-triggered processing per workload based on latency, reliability and operational requirements, supporting up to 14 intraday execution windows.
- **Messaging integration:** Designed messaging patterns across Azure Service Bus and Google Cloud Pub/Sub using topics/subscriptions, routing and filtering strategies, independent retry/monitoring boundaries, and managed-identity based integration where applicable.
- **Databricks modernization & governance:** Contributed to the migration of dozens of WMS/DB2 data assets, Databricks runtime upgrades and Kafka-native ingestion; assessed legacy mount/Hive Metastore dependencies and applied Unity Catalog patterns for unified data governance, access controls, lineage and governed catalog/schema practices across DEV/QA/PROD.
- **Reusable engineering frameworks:** Built metadata-driven ETL/ELT and Delta Live Tables (DLT) workflows using ADF, Databricks, PySpark, SQL, Delta Lake, and ADLS Gen2. Automated pipelines using event-based, scheduled, and tumbling-window triggers, with dependency handling and environment-specific execution across all environments.
- **Technical leadership:** Mentor 5+ engineers and provide technical direction; review code and solution designs, troubleshoot complex delivery issues and work with architects, platform, security, QA, MuleSoft and application teams on integration decisions.

- **Security & delivery:** Secured pipelines with Key Vault, managed identities, RBAC, private-network integration and environment parameterization; coordinated Git/Azure DevOps promotion, QA/E2E testing, release validation, hypercare and business sign-off.
- **Production observability & performance:** Investigate pipeline and Azure Function failures using Azure Monitor, Application Insights, Log Analytics/KQL and ADF/Databricks telemetry, tracing dependencies, reconciling incomplete runs, recovering failed jobs and driving RCA; tuned Spark/SQL workloads, partitioning, file layout and Databricks compute/runtime settings for reliable execution.
- **Serverless modernization:** Modernized Databricks workloads for serverless execution by replacing legacy DBFS/mount assumptions with ABFSS and Unity Catalog-compatible access patterns, parameterizing environments, resolving dependency/configuration incompatibilities and aligning catalog permissions for governed execution. Supported cluster/runtime modernization, Serverless Compute adoption, and production performance tuning.
- **Outbound API integration:** Staged and batched curated data for downstream consumers, including JSON generation, payload sizing, grouping, business-key normalization and validation, with managed-identity delivery coordinated with MuleSoft and application teams.
- **Supply-chain domain depth:** Delivered pipelines for warehouse management, yard management, inbound purchase orders, ASN validation and distribution-center space and inventory across retail distribution-center operations.

Environment: Azure Data Factory, Databricks, Structured Streaming, Auto Loader, Delta Live Tables, Unity Catalog, Event Hubs, Kafka, Service Bus, Google Cloud Pub/Sub, Azure Functions, Azure Monitor, KQL, Control-M, Azure DevOps, DB2

SELECTED APPLIED AI / GENAI & LLMOPS PROJECTS

Independent Engineering Portfolio **2025 – Present**

Data + AI Engineering Projects

- **Databricks RAG / LLMOps:** Built production-style RAG and research-assistant workflows using governed Unity Catalog data, Databricks AI Search (Vector Search), embeddings, Model Serving / Foundation Model APIs, and MLflow 3 tracing, evaluation, monitoring and LLM-judge patterns.
- **Data Intelligence & Agentic AI:** Built and integrated governed AI workflows using AI/BI Dashboards, Genie One, Genie Agents, Genie Code, Custom Agents, MCP Services, Databricks Apps, Lakebase and Unity Gateway / AI Gateway patterns for tool-using agents, persistent state and controlled model/tool access.
- **Modern Databricks platform engineering:** Implemented Lakeflow Jobs, Spark Declarative Pipelines, Databricks Workflows, Serverless Compute, Photon and Databricks Asset Bundles with environment-aware CI/CD, REST APIs / SDKs, monitoring and repeatable DEV/QA/PROD deployment patterns.

EARLIER PROFESSIONAL EXPERIENCE

Manufacturing Enterprise | **Indiana** **Feb 2019 – Feb 2020**

Azure Data Engineer

- **Azure migration & orchestration:** Designed on-premises SQL Server-to-ADLS data movement with ADF, self-hosted Integration Runtime and Databricks; built parameterized pipelines, incremental loads, event/schedule triggers, PySpark/SQL transformations, logging, error handling and recovery patterns.
- **ETL engineering:** Built complex ETL packages from staging to partitioned tables with incremental loading, error handling, logging, configuration and deployment best practices; designed normalized SQL objects (tables, views, stored procedures, indexes) and used Python notebooks to generate fact datasets.

Industrial Manufacturing Company | **Illinois** **Mar 2016 – Jan 2019**

Data Engineer

- **Azure data platform modernization:** Designed storage, integration, processing and analytics solutions while migrating on-premises workloads to ADLS, ADF, Databricks, Azure SQL and Azure SQL Data Warehouse / PolyBase; built Spark Streaming/PySpark processing and tuned SQL for downstream analytics.
- **Legacy modernization:** Analyzed legacy SQL scripts and redesigned them as PySpark solutions; created linked services to Azure Storage and on-premises SQL Server; delivered tabular models with row-level security for downstream reporting.

Financial Services Company | **Hyderabad, India** **Jun 2010 – Jun 2014**

ETL Developer

- **Enterprise ETL & performance:** Delivered enterprise ETL using Teradata, Informatica PowerCenter, Oracle, BTEQ, FastLoad/MultiLoad/TPT and UNIX shell; built historical/incremental loads, tuned Teradata SQL and supported requirements, testing, deployment and production workflows.
- **Workflow engineering:** Built data flows with execute-SQL tasks, sequence containers and for-each loop, and Informatica workflows with event-wait, event-raise, command and email tasks; supported the full SDLC from business analysis through deployment and documentation.

EDUCATION & TRAINING

M.S., Texas A&M University–Kingsville, 2015 | **B. Tech,** JNTU Hyderabad, India, 2010.
M.B.A. (IT Management), Campbellsville University, 2020.
Postgraduate Program in AI / Machine Learning, The University of Texas, Austin, TX, 2026.