PRAMEEL KOMMINENI

SENIOR DATA ENGINEER

KPK111298@GMAIL.COM
+1 (309)-612-5411
linkedin.com/in/pkomm

Professional Summary:

Senior Data Engineer with 5+ years of experience building and modernizing cloud data platforms across AWS and Azure. Experienced in designing scalable data architectures and enterprise data platforms. Specialized in migrating legacy ETL systems to scalable Snowflake and Databricks architectures, and designing batch and real time pipelines using Spark, dbt, and Airflow. Experienced in building feature pipelines and governed data models to support analytics, fraud detection, and machine learning use cases. Worked in multi cloud environments, with AWS as the core platform and selective analytics workloads running on Google BigQuery. Known for improving performance, reducing costs, and delivering reliable data platforms across logistics, insurance, healthcare, and enterprise systems.

Technical Skills:

- **Cloud Platforms:** AWS (S3, Glue, Lambda, Redshift), Azure (ADF, Data Lake, Fabric), Google Cloud (BigQuery)
- **Data Engineering:** Snowflake, Databricks, Apache Spark, dbt, Apache Airflow, Delta Lake, Data Architecture
- **Data Integration:** Fivetran, HVR, Informatica, SSIS
- **Databases & Warehouses:** Snowflake, Redshift, BigQuery, PostgreSQL, SQL Server, MySQL
- **Programming:** Python, SQL, Scala (basic), Shell scripting
- **BI Tools:** Power BI, Tableau, SSRS
- **DevOps & CI/CD:** Git, Jenkins, Azure DevOps
- **Big Data:** Hadoop, Hive, Kafka

Professional Experience:

Client: U.S. Xpress 02/2025 – Present
Role: Data Engineer II Remote Chattanooga, TN

Project Overview: Led the migration of Microsoft Dynamics 365 CRM and legacy systems to a modern Snowflake data platform on AWS. Built batch and real-time pipelines, implemented governed data models with dbt, and enabled self-service analytics, reducing data latency by 60% and infrastructure costs by 40%.
Environment: Snowflake and AWS based data platform processing 3–5 TB of logistics and CRM data daily across 60+ batch and real time pipelines, supporting analytics for operations, finance, and customer experience teams.
Responsibilities:

- Built ingestion pipelines to move data from Dynamics 365 and legacy systems into AWS S3 using Fivetran and HVR.
- Standardized high volume replication and created automated data validation scripts in Python.
- Designed ETL and ELT pipelines to load raw data from S3 into Snowflake using Snowpipe and AWS Glue.
- Built real time ingestion pipelines with Kafka and AWS Kinesis, reducing data latency by 60 percent.
- Built reusable dbt models to transform raw data into curated analytics layers.
- Built governed data models and feature datasets to support fraud detection and operational forecasting.
- Built feature pipelines and curated training datasets to support machine learning models for route optimization and demand forecasting.
- Supported cross platform analytics by integrating select datasets into Google BigQuery for partner reporting and data science use cases.
- Implemented data governance controls such as masking, retention rules, and identity based access in Snowflake.
- Enabled self service analytics by building reusable data marts and semantic layers for Power BI and Tableau.
- Migrated and decommissioned legacy infrastructure, reducing infrastructure costs by 40 percent.
- Set up CI and CD pipelines, automated monitoring, and peer code reviews to maintain engineering standards.
- Mentored junior engineers on dbt modeling, CI and CD practices, and data governance.
- **Tools:** Python, SQL, dbt, Fivetran, HVR, Snowflake, AWS S3, Glue, Lambda, Kafka, Kinesis, Google BigQuery, Power BI, Tableau, CI/CD

Client: American Family Insurance 12/2023 – 10/2024
Role: Data Engineer Remote - Tampa, FL

Project Overview: Migrated legacy SSIS analytics pipelines to a cloud native lakehouse using Databricks and Microsoft Fabric. Built scalable Spark pipelines and governed data layers, improving transformation speed by 50 percent and reducing reporting cycle times by over 60 percent.
Environment: Azure and Databricks lakehouse platform processing insurance and financial data across 40+ pipelines, supporting reporting, risk modeling, and regulatory analytics for business and finance teams.

- Migrated legacy SSIS workflows into PySpark based Databricks notebooks, improving scalability and maintainability.
- Built ingestion and transformation pipelines using Azure Data Factory and Databricks to move data into Azure Data Lake.
- Applied medallion architecture principles across bronze, silver, and gold layers to standardize ingestion, transformation, and consumption.
- Built Databricks and Spark ETL workflows that improved transformation speed by 50 percent and made dashboards three times faster.
- Integrated Delta Lake with Azure Data Lake to provide ACID compliant, version controlled data storage.
- Designed analytics workflows using Microsoft Fabric components such as OneLake and Lakehouse to centralize insurance and financial data.
- Built feature tables and curated training datasets to support fraud detection and risk scoring models.
- Worked with data science teams to deliver versioned datasets and pipeline orchestration for model training and batch scoring workflows.
- Collaborated with analysts and business stakeholders to design semantic models and reporting layers aligned with insurance KPIs.
- Built reusable dbt models with strong documentation, testing, and data quality checks.
- Enhanced SSRS and Power BI dashboards by 60 percent through optimized semantic models and DAX queries.
- Implemented data masking, retention, and compliance controls for regulatory reporting aligned with security standards.
- Built CI and CD pipelines in Azure DevOps to automate deployment and testing across analytics and reporting environments.
- Managed Azure resources using RBAC, cost optimization, and autoscaling strategies to maintain performance and budget efficiency.
- Reduced overall reporting cycle times by more than 60 percent.
- **Tools:** Databricks, Spark, Azure Data Factory, Azure Data Lake, Delta Lake, Microsoft Fabric, dbt, Python, Power BI, SSRS, Azure DevOps, Git

Client: Harman International 09/2020 – 06/2023
Role: Data Engineer Onsite -Bangalore, India

Project Overview: Modernized legacy ETL pipelines by migrating Informatica workflows to cloud native Spark pipelines on AWS and Snowflake. Built scalable ingestion, transformation, and dimensional data models, reduced processing time by 50 percent, and lowered Snowflake warehouse costs by 20 percent while supporting analytics and machine learning workloads.
Environment: AWS based data platform processing 2–3 TB of data daily across 30+ batch pipelines, supporting analytics for product, operations, and executive reporting teams.

- Modernized legacy ETL pipelines by migrating them to cloud native Spark workflows on AWS to support analytics and machine learning workloads.
- Rebuilt legacy Informatica workflows using PySpark, Scala, and Airflow, reducing processing time by 50 percent.
- Built Spark based ingestion and transformation pipelines on Databricks to process structured and semi structured data stored in AWS S3.
- Built and operated data pipelines on AWS using S3, Glue, EMR, Lambda, and Redshift.
- Designed and migrated dimensional data models using Kimball methodology to Snowflake for scalable reporting and analytics.
- Partnered with analytics teams to define KPI driven data models for executive dashboards and operational reporting.
- Reduced Snowflake warehouse consumption by 20 percent through query optimization and performance tuning.
- Implemented performance tuning and cost optimization strategies across Snowflake and Amazon Redshift environments.
- Deployed containerized microservices to deliver reusable REST APIs for analytics and machine learning integrations.
- Built and deployed interactive dashboards in Power BI, Tableau, and SSRS to provide real time insights for executive teams.
- Established CI and CD pipelines using Jenkins and Git, conducted code reviews, and enforced engineering best practices.
- Built scalable data pipelines that supported downstream analytics and partner integrations.
- **Tools:** Python, PySpark, SQL, Scala, Databricks, AWS S3, Glue, EMR, Lambda, Redshift, RDS, Snowflake, Airflow, Jenkins, Git, Power BI, Tableau

Client: Zeta Biotechnology Limited, 06/2019 – 09/2020
Role: Hadoop Developer Hybrid - Hyderabad, India

Project Overview: Developed big data ingestion and transformation frameworks on Hadoop and Spark for healthcare analytics. Built Spark and Hive pipelines, automated data quality checks, and improved query performance by 25 percent while reducing manual QA effort.
Environment: Hadoop cluster processing healthcare datasets across 20+ ingestion pipelines, supporting compliance reporting, data quality monitoring, and analytics teams.

- Built big data ingestion and transformation pipelines on Hadoop and Spark to support healthcare analytics and compliance reporting.
- Deployed and managed Hadoop clusters across development, staging, and production environments to ensure stable and high performance processing.
- Designed a data quality framework on Hadoop to automate validation and process control across multiple ingestion pipelines.
- Partnered with engineering teams to define standardized ETL frameworks aligned with healthcare data governance and compliance needs.
- Built Spark Scala applications and Hive scripts to transform large scale datasets into fact and dimension tables using complex business rules.
- Optimized Hive performance through partitioning, bucketing, and efficient storage formats such as ORC and Parquet.
- Integrated MongoDB to support unstructured data ingestion pipelines.
- Automated ingestion and quality assurance workflows using Oozie, Bash, and scheduled jobs to improve delivery speed.
- Built auditing pipelines to monitor ingestion performance and support compliance reporting.
- Improved Hive query performance by 25 percent through storage and query optimizations.

- Reduced manual QA effort by automating validation workflows and improving delivery timelines.
- **Tools:** Apache Spark, Scala, Apache Hadoop, Apache Hive, Apache Oozie, Apache Sqoop, Bash, MongoDB, CRON, ORC, Parquet

Certifications:

- MICROSOFT CERTIFIED: FABRIC DATA ENGINEER ASSOCIATE
- SNOWFLAKE CERTIFIED: SNOWPRO CORE

Education:

- Master of Science in Information Systems – 2024 – University of Illinois Springfield, IL, USA.
- Bachelor's in Business administration – 2019 – Osmania University, Hyd, India.