

PREMKUMAR CHIMAKURTHI

St. Louis, MO (Open to Relocation) | +1(224) 715-7017 | prem.chimakurthi@gmail.com | [linkedin.com/in/premkch](https://www.linkedin.com/in/premkch) | [Github](https://github.com/premkch)

SUMMARY

Data & AI Engineer with 5+ years building production-grade platforms on Azure + Databricks (Spark, Delta Lake, Unity Catalog). Delivers modern ELT/ETL, streaming, and batch pipelines processing 10M–100M+ records daily. Hands-on with MLflow, Terraform, CI/CD, Azure DevOps, Git, and Agile delivery. **Microsoft Certified: Azure AI Engineer & Data Scientist Associate.** **Strong collaboration and teamwork skills in agile environments** with a proven track record of problem solving across data engineering, machine learning feature pipelines, and data observability. Experienced in building and optimizing **distributed data processing systems at scale** using Spark and cloud-native architectures. **Microsoft Azure–Certified. Open to relocation for onsite roles.**

EDUCATION

WEBSTER UNIVERSITY, St. Louis, MO

Graduated May 2025

Master of Science in Business Analytics. Coursework: Cloud Computing, Data Engineering, Computer Science, Machine Learning, Statistics, Information Systems

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY, Guntur, India

Bachelor of Technology in Mechanical Engineering

EXPERIENCE

ROBERTHALF, INC. – Client: Dent Wizard (Automotive Services)

Nov 2023 – Present

Senior Data & AI Engineer | St. Louis, MO

- Shipped production-grade **Azure + Databricks platform** (ADLS Gen2, Delta Lake Bronze/Silver/Gold, Unity Catalog) processing 50M+ records/day using PySpark and Spark Structured Streaming.
- Built **AI-enabling foundations: MLflow experiment tracking, ML-ready Delta Lake feature tables** for demand forecasting; reduced data latency hours → near real-time via CDC with Delta Live Tables & Auto Loader.
- Automated IaC with **Terraform + CI/CD (Azure DevOps)**; enforced security-by-design via Key Vault and Unity Catalog governance. Cut ETL runtimes 55% via Z-ordering, partitioning, broadcast joins.
- Owned data quality with **Great Expectations + observability dashboards**; reduced pipeline failures 40% across 50M+ daily records.
- Collaborated with data science and ML teams in an agile environment to build machine learning feature engineering and classification pipelines, delivering curated feature tables in Delta Lake for demand forecasting models.

TATA CONSULTANCY SERVICES LTD. – Client: DaVita Healthcare

May 2022 – Oct 2023

Data Engineer | Hyderabad, India

- Designed **GCP lakehouse (BigQuery, Dataflow)** with Kimball star schema and SCD Type 1 & 2 models; built real-time **Kafka streaming pipelines** processing 20M+ healthcare records/day across 50+ HIPAA-compliant centers.
- Orchestrated 30+ DAG workflows via **Apache Airflow (Cloud Composer)**; built GCP Monitoring dashboards for observability — reduced incidents 40% through root cause analysis
- Orchestrated complex DAG workflows using Apache Airflow (Cloud Composer) with dependency management, retry logic, and SLA monitoring; managed 30+ scheduled pipelines processing **20M+ healthcare records daily** across 50+ dialysis centers in a HIPAA-compliant environment.
- Developed data observability dashboards using GCP Monitoring; performed root cause analysis and applied problem-solving algorithms to resolve pipeline failures, reducing data-related incidents by 40% through cross-team collaboration.

HCL TECHNOLOGIES LTD. – Client: ExxonMobil

Dec 2019 – Mar 2022

Data Engineer | Chennai, India

- Built batch & streaming pipelines on **AWS (S3, Glue, Redshift, Kinesis, Lambda)** processing 10M+ IoT records/day; designed Redshift MPP schemas cutting query times 35%.
- Delivered **Python feature engineering datasets for ML predictive maintenance models**; reduced unplanned equipment downtime by 22%.
- Implemented data validation and monitoring using CloudWatch alarms and SNS notifications; built automated alerting for schema drift, null-rate spikes, and freshness SLO violations through teamwork with site reliability engineers.
- Partnered with process engineers to deliver feature engineering datasets using Python (Pandas, NumPy) for predictive maintenance machine learning models; identified equipment degradation patterns that reduced unplanned downtime by 22%.

SKILLS

Azure/Cloud: ADLS Gen2, ADF, Event Hubs, Functions, Key Vault, Synapse, VNets, Databricks (Spark, Delta Lake, Unity Catalog, Workflows, MLflow) | **AWS:** S3, Glue, Redshift, Kinesis | **GCP:** BigQuery, Dataflow

Languages: Python (PySpark, Pandas), SQL (Delta SQL, Spark SQL), Scala (basic), Go | **Data Modeling:** Kimball/Inmon/Data Vault, Star Schema, SCD Type 1 & 2, Lakehouse

Engineering: Terraform/Bicep (IaC), CI/CD (Azure DevOps, GitHub Actions), Git, code review, dbx/pytest, Docker, Great Expectations | **Streaming:** Kafka, Event Hubs, Structured Streaming

ML/AI: MLflow, feature engineering, feature stores, model registry, ML pipelines, Generative AI | **Databases:** Snowflake, Synapse, Cosmos DB (NoSQL), Redshift, PostgreSQL, SQL Server

CERTIFICATIONS

Microsoft Certified: Azure Data Scientist Associate | Microsoft Certified: Azure AI Engineer Associate | AWS ML Foundations