

Gautham Kolluru

GitHub | LinkedIn | Stack Overflow | HackerRank

+1 240-639-9934 | gautham.kolluru@gmail.com | thegauthams.com

Professional Summary

Senior Machine Learning & Data Engineer with 10+ years building production-scale AI/ML systems and data platforms serving 100M+ users. Expertise in designing end-to-end ML infrastructure, deploying LLM-based applications, and architecting distributed data systems processing 100T+ records daily. Proven track record delivering measurable business impact through ML-driven solutions at Fortune 100 companies. Deep experience with cloud-native architectures, MLOps, and real-time streaming systems.

Technical Expertise

- **Machine Learning/AI:** Deep Learning (TensorFlow, PyTorch), LLMs & Generative AI (LangChain, LangGraph, RAG, Prompt Engineering), NLP, Computer Vision, ML Model Optimization, A/B Testing, Feature Engineering
- **MLOps & Infrastructure:** Kubernetes, Docker, Kubeflow, MLflow, Airflow, Model Serving (TensorFlow Serving, TorchServe), CI/CD for ML, Model Monitoring & Observability
- **Big Data & Cloud:** GCP (BigQuery, Dataflow, Vertex AI, Cloud Storage), AWS (SageMaker, EMR, S3, Lambda), Apache Spark, Hadoop, Kafka, Flink
- **Programming & Frameworks:** Python (NumPy, Pandas, Scikit-learn, PySpark), Java, Rust, SQL, Scala
- **Data Engineering:** Real-time Streaming, ETL/ELT Pipelines, Data Lakes, Data Warehousing (Snowflake, Redshift), OLAP/OLTP, Data Quality Frameworks

Professional Experience

Senior Data Engineer, Cisco; Raleigh, NC — 2021-Present

- **Contributed to ML infrastructure development** for network intelligence platform serving Fortune 100 telecom provider (120M+ customers), enabling predictive network analytics and anomaly detection at scale
- **Architected and deployed production LLM-powered application** using LangChain and RAG, reducing network troubleshooting time by 60% and improving customer satisfaction scores by 25%
- **Engineered distributed ML inference pipeline** processing 100T+ daily network events using Kubernetes, achieving 70% latency reduction through intelligent caching and model optimization
- **Built near real-time anomaly detection system** analyzing 10,000+ network devices using parallel processing, preventing \$10M+ in potential network outages
- **Designed fault-tolerant data pipeline** with 99.99% uptime protecting \$10B+ network infrastructure through real-time threat detection

Senior Data Engineer, 84.51°; Cincinnati, OH — 2020-2021

- **Architected enterprise-scale ML feature store** processing 20B rows (5TB+) with 40+ transformation steps, reducing feature engineering time by 80% for data science teams
- **Led performance optimization initiative** achieving 40% latency improvement through distributed computing optimization (Spark tuning, data partitioning strategies) for customer analytics pipelines
- **Implemented real-time recommendation system data pipeline** supporting personalized marketing campaigns, contributing to \$50M+ incremental revenue

Data / Machine Learning Engineer, Cisco; Bengaluru, India — 2019-2020

- **Designed multi-cloud data synchronization platform** enabling real-time data replication between GCP and Snowflake with zero data loss, supporting ML model training workflows
- **Implemented incremental ETL pipelines** using Airflow and Dataflow, reducing data processing costs by 35% through intelligent incremental loading strategies

Software Engineer, Global Data; Hyderabad, India — 2018

- **Engineered scalable web scraping platform** using distributed architecture (Scrapy, Celery, Redis), processing 1K+ pages daily from diverse sources for ML training data
- **Optimized data extraction pipeline** implementing intelligent proxy rotation and multi-processing, achieving 3x throughput improvement and 99.9% reliability

Software Developer, Invincible Software Solutions; Visakhapatnam, India — 2017-2018

- **Led team of 5 engineers** developing 12 full-stack applications for US federal clients using microservices architecture, delivering projects on-time with 98% quality metrics
- **Designed real-time analytics dashboards** using Tableau and SSRS, processing 10M+ daily records and enabling data-driven decision making for stakeholders

Project / Systems Engineer, Royal Oman Police; Muscat, OM — 2013-2015

- **Architected high-performance OLTP/OLAP database system** for nationwide speed camera network, handling 500K+ daily transactions with sub-100ms query latency

Automation Engineer, Mott MacDonald; Abu Dhabi, UAE — 2009-2012

- **Automated industrial control systems** for water treatment facilities using PLC and SCADA, improving operational efficiency by 40%
- **Led installation of automated monitoring systems** at major infrastructure projects (Yas Island development, housing communities)

Education

B. Tech in Computer Science Engineering — Andhra University, 2009

Additional Achievements

- **Scale:** Built ML/data systems processing 100T+ transactions daily in near real-time.
- **Impact:** Delivered solutions contributing to \$20M+ measurable business value