

Kattamuri Kowshiq

+91 7036160563 | kattamurikowshiq@gmail.com | LinkedIn | GitHub | Portfolio | LeetCode

SUMMARY

Previously worked as AI Data Engineer specialized in Data Warehousing (DWH), scalable data processing platforms, and automated ingestion frameworks with built-in change detection. Proven expertise in building robust, secure AI-powered data infrastructure, executing Enterprise AI integrations (RAG/MCP servers via Python/FastAPI), and deploying LLM agent orchestration pipelines.

EDUCATION

Indian Institute of Technology Roorkee <i>Master of Technology AI</i>	Roorkee, India 2026 – 2028
National Institute of Technology Calicut <i>Bachelor of Technology CGPA: 7.86 / 10</i>	Kozhikode, India 2020 – 2024

TECHNICAL SKILLS

Languages: Python, SQL, C++, Unix Shell Scripting
Data Engineering & DWH: Apache Spark, PySpark, Kafka, dbt, Snowflake, Databricks, Delta Lake, Hadoop
AI / ML: Enterprise AI Integrations, RAG Pipelines, Agentic AI, MCP Server, LangChain, Vector Databases
Cloud & Infra: AWS (EKS, RDS, S3, IAM, Secrets Manager), Terraform, Kubernetes, Docker
DevOps & Observability: CI/CD (GitHub Actions), Git, Jira, Airflow
Databases & Visualization: PostgreSQL, MySQL, MongoDB, Streamlit, Data Visualization Concepts (PowerBI/Tableau)

EXPERIENCE

Data Engineering, Management and Governance Analyst <i>Accenture</i>	April 2026 – July 2026 Hyderabad, India
- Designed and built data ingestion pipelines that connect to 9+ observability platforms (CloudWatch, Dynatrace, Jira, ServiceNow) and normalize their heterogeneous JSON/JSONB payloads into standardized schemas within a multi-tenant PostgreSQL database, ensuring proper tenant isolation and maintaining 99%+ data accuracy across concurrent writes. - Implemented a rules-driven data correlation engine that intelligently links cross-source events (incidents, metrics, logs, tickets) by matching on entity keys (service name, environment, team) and time proximity windows, reducing manual investigation time for SRE teams by eliminating data silos across platforms. - Built and integrated a knowledge retrieval system using Amazon Titan Embeddings to index 50K+ SRE runbooks and incident postmortems into OpenSearch, enabling semantic search queries that surface relevant context automatically—reducing LLM API costs through intelligent caching with 90%+ hit rate optimization. - Developed comprehensive audit logging for all data operations (ingestion, transformation, access) across the multi-tenant system; implemented role-based access control (RBAC) with tenant-level permissions and data lineage tracking, ensuring compliance and accountability for sensitive infrastructure data. - Integrated with multiple LLM providers (Azure, OpenAI, Bedrock) to automatically extract intent from user queries and generate intelligent analysis recommendations; built a semantic caching layer that stores analysis results by query intent and question type, reducing API costs while improving query response times.	
Advanced Application Engineering Analyst	November 2024 – April 2026
- Engineered and maintained scalable data processing platforms and ELT pipelines utilizing Python, PySpark, and dbt, implementing incremental loads (SCD Type 2) to ensure high data availability and architecture acceleration for enterprise analytics. - Architected real-time, event-driven data streaming workflows using Kafka and AWS Lambda, enabling faster data insights and highly reliable asynchronous processing with fault-tolerant execution. - Optimized distributed data workflows and complex SQL queries by tuning partition strategies, minimizing shuffle operations, and implementing advanced indexing, reducing overall processing time and query latency by up to 35%. - Drove the full software development lifecycle (SDLC) for data applications, utilizing Unix shell scripting, GitHub Actions (CI/CD), and stringent unit testing to ensure the on-time delivery of secure, production-grade pipelines.	

PROJECTS

Enterprise RAG Analytics Platform Python, LangGraph, LangChain, Qdrant, OpenAI, APIs	GitHub
- Engineered an end-to-end Retrieval-Augmented Generation (RAG) pipeline using Python, LangGraph, and Qdrant, automating the ingestion, chunking, and vectorization of complex unstructured documents to ensure high data availability. - Architected an intelligent agentic routing workflow integrating OpenAI LLMs and Tavily web search APIs to orchestrate context retrieval, real-time data lookups, and automated claim verification, directly aligning with enterprise AI integration patterns. - Deployed a scalable multi-session UI with streaming capabilities using Streamlit, accelerating the delivery of actionable data insights and optimizing the end-user analytics experience.	

CERTIFICATIONS

Professional Data Engineer in Python — DataCamp, 2026
--