Nagendra K.
AI/ML & Azure Data Engineer
Malvern, PA | +1 (484) 716-1024 | nagendrakatigowri76@gmail.com | [LinkedIn](#)

SUMMARY

AI/ML & Azure Data Engineer with 3+ years of experience designing and optimizing enterprise-scale data platforms using Azure, Databricks, and Synapse. Proven record of building high-performance ETL and streaming pipelines processing 5+ TB of data daily, reducing latency by 70%, and integrating AI/LLM models using LangChain, FAISS, and Mistral for real-time semantic search and anomaly detection. Skilled in cloud automation using Terraform, GitHub Actions, and container orchestration with AKS. Adept in implementing Unity Catalog–based governance, ensuring 100% regulatory compliance with HIPAA, GDPR, and SOC 2. Passionate about leveraging ML, embeddings, and vector databases to drive predictive analytics and business intelligence in scalable cloud environments.

TECHNICAL SKILLS

- **Languages:** Python, PySpark, SQL, T-SQL, PowerShell
- **Cloud & Data Platforms:** Azure Data Factory, Databricks, Synapse, ADLS Gen2, Snowflake, Azure Fabric, Azure OpenAI
- **ETL & Streaming:** Delta Lake, Apache Kafka, Flink, Event Hubs, Delta Live Tables, Spark Structured Streaming
- **AI & ML:** LangChain, FAISS Vector DB, Mistral, Sentence Transformers, Embeddings & Chunking, MLflow, Feature Store, Hugging Face, LLMOps
- **Data Governance:** Unity Catalog, Azure Purview, Data Lineage, RBAC, Masking, Classification, Policy Enforcement
- **DevOps & IaC:** Terraform, GitHub Actions, Azure DevOps, Docker, AKS, ARM Templates, Key Vault
- **Visualization & Analytics:** Power BI, DAX, Synapse Analytics, KQL, HDInsight
- **Methodologies:** Agile/Scrum, SDLC, Documentation, Stakeholder Collaboration, Data Mesh

PROFESSIONAL EXPERIENCE

Elevance Health | Malvern, PA
Data Engineer | Jan 2025 – Present

- Engineered 20+ ingestion pipelines using Azure Data Factory and Databricks, processing 5 TB+ healthcare claims data daily, cutting manual refreshes by 70%.
- Developed real-time AI-driven analytics pipelines integrating LangChain, FAISS, and Azure OpenAI, enabling sub-5-second response times for intelligent data retrieval.
- Deployed Mistral-based LLM models in Databricks for operational forecasting, boosting anomaly detection accuracy by 35%.
- Streamlined infrastructure deployment with Terraform and GitHub Actions CI/CD, reducing release times from 2 hours to 15 minutes and improving reliability by 85%.
- Orchestrated Unity Catalog and Azure Purview integration, ensuring 100% lineage tracking and secure access control for 50+ datasets.
- Optimized Synapse and Databricks workloads, improving query performance by 40% and cutting compute costs by 30%.
- Containerized pipelines with Docker and AKS to handle 10K+ concurrent event streams for real-time processing.
- Led data governance and compliance initiatives maintaining HIPAA and GDPR audit readiness with zero violations.

CitiusTech | Mumbai, India
Azure Data Engineer | Jan 2022 – Jul 2023

- Designed scalable ETL pipelines via ADF and Flink handling 2 TB+ EHR data daily, improving processing efficiency by 60%.
- Integrated Sentence Transformer embeddings into Databricks workflows for similarity search, reducing ML feature retrieval latency by 40%.
- Built Azure infrastructure using Terraform, cutting deployment time by 75% and minimizing configuration drift.
- Migrated on-prem Hadoop clusters to Azure Delta Lakehouse, lowering storage costs by 30% and boosting performance by 25%.
- Implemented Unity Catalog for unified metadata tracking and data classification across 15+ domains.
- Tuned Spark and Flink jobs for large-scale healthcare workloads, reducing job runtimes by 35%.
- Enhanced data security via RBAC and private endpoints, improving compliance audit scores by 20%.

Mphasis | Bengaluru, India
Data Engineer | Mar 2021 – Dec 2021

- Developed real-time data pipelines using Kafka, Flink, and Synapse, reducing data latency from 2 hours to under 5 minutes.
- Implemented LLM-based anomaly detection in streaming pipelines, increasing operational accuracy by 30%.
- Created Terraform CI/CD pipelines, enabling zero-downtime infrastructure rollouts and consistent multi-region deployments.
- Configured Azure Monitor and Log Analytics dashboards, reducing mean time to recovery (MTTR) by 50%.
- Documented detailed runbooks, SLAs, and process workflows, improving cross-team support efficiency by 25%.

CERTIFICATIONS

- **Microsoft Certified: Fabric Data Engineer Associate**

EDUCATION

Master's in Information Systems (Project Management Concentration) – Wilmington University, DE   **May 2025**

Bachelor of Technology in Computer Science and Engineering – Guntur Engineering College, India   **Jun 2022**

PROJECTS

- **AI-Powered Knowledge Extraction Chatbot (LangChain + Azure OpenAI + FAISS + Mistral)**

Built a chatbot integrating LangChain, FAISS, and Mistral, achieving 90% response accuracy and cutting document lookup times by 80%. Used Sentence Transformers and embedding chunking for contextual retrieval across 50K+ enterprise documents. Applied Unity Catalog-based RBAC for secure access and lineage control.

- **Cloud ETL Pipeline Optimization (ADF + Databricks + Synapse)**

Developed ETL workflows processing 1M+ monthly records, increasing data freshness by 60%. Implemented Delta Live Tables and validation rules to maintain 99.9% data accuracy. Created Power BI dashboards with 10+ KPIs, reducing executive reporting time by 40%.

- **Real-Time Event Streaming and Analytics (Kafka + Flink + Event Hubs)**

Processed 100K+ events/hour in low-latency pipelines with Spark Structured Streaming. Enabled predictive monitoring and ML-driven alerts, improving incident response times by 45%.

Work Authorization: Authorized to work in the U.S. without sponsorship | Open to Relocation and Hybrid Roles