Nagendra K.

Azure Data Engineer

Malvern, PA | +1 (484) 716-1024 | nagendrakatigowri76@gmail.com | LinkedIn

SUMMARY

Azure Data Engineer with 3+ years of experience designing and delivering production-grade real-time and batch data pipelines using Azure Data Factory, Databricks, Delta Lake, and Synapse. Expert in integrating AI and ML models to enhance predictive analytics and anomaly detection. Skilled in automation using Terraform and GitHub Actions CI/CD pipelines. Adept with streaming technologies including Kafka, Apache Flink, and Azure Event Hubs to support low latency data processing. Proficient in implementing data governance frameworks compliant with HIPAA, GDPR, NIST, and SOC 2. Collaborative Agile team player focused on scalable cloud-native data lakehouse architectures and data mesh principles.

TECHNICAL SKILLS

- **Languages:** Python, PySpark, T-SQL, SQL, PowerShell
- **Cloud/Data Services:** Azure Data Factory, Azure Synapse, Azure Databricks, ADLS Gen2, Cosmos DB, Azure OpenAI
- **ETL & Streaming:** ADF Pipelines, Delta Lake, Apache Kafka, Azure Event Hubs, Apache Flink, Spark Structured Streaming
- **Big Data & Analytics:** Spark SQL, KQL, Databricks, HDInsight, Power BI
- **Storage & File Formats:** Blob Storage, Parquet, JSON, Avro, CSV
- **DevOps & IaC:** Azure DevOps, GitHub Actions, Terraform, ARM Templates, Key Vault
- **Security & Governance:** RBAC, Private Endpoints, Azure Purview, Data Masking, HIPAA/GDPR/NIST/SOC 2 Compliance
- **Monitoring & Optimization:** ADF Monitoring, Azure Monitor, Log Analytics, Spark/SQL/Flink Tuning
- **Containerization & Orchestration:** Docker, Azure Kubernetes Service (AKS)
- **Methodologies:** Agile/Scrum, Documentation, SDLC, Stakeholder Collaboration, Data Mesh

PROFESSIONAL EXPERIENCE

Elevance Health, PA, USA **Jan 2025 - Current**

Data Engineer

- Automated claims ingestion pipelines using Azure Data Factory, Databricks, and Apache Flink, cutting manual refreshes by 70% and accelerating regulatory reporting.
- Designed real-time streaming solutions with Apache Kafka, Delta Lake, Azure Event Hubs, and Flink that reduced data latency from hours to minutes.
- Embedded AI and ML models into pipelines to deliver predictive analytics, anomaly detection, and operational forecasting across healthcare programs.
- Automated infrastructure deployment with Terraform and GitHub Actions CI/CD pipelines, increasing deployment speed and reducing errors by 85%.
- Enhanced Azure Synapse Analytics and Databricks workloads using AI-powered query optimizations, improving performance and reducing costs.
- Developed and enforced data observability, metadata management, and quality frameworks using Azure Purview and Great Expectations.
- Architected disaster recovery and backup strategies ensuring critical data availability and compliance with data security standards.
- Deployed containerized scalable data pipelines in Docker and AKS to support growing workloads.
- Collaborated in agile cross-functional teams to promote data mesh architectures and decentralized data ownership for enterprise innovation.

CitiusTech, India **Jan 2022 - Jul 2023**

Azure Data Engineer

- Architected scalable ETL and real-time streaming pipelines using Azure Data Factory, Apache Flink, and Kafka for ingesting and processing EHR data.
- Integrated ML models in Databricks to improve clinical risk prediction and analytics accuracy.
- Led Azure infrastructure automation with Terraform, improving reliability and deployment times.
- Spearheaded migration from on-prem Hadoop to Azure cloud-native data lakehouse architectures, optimizing cost and performance.
- Tuned Spark and Flink batch jobs, reducing runtimes by 30%.
- Secured healthcare data using Azure RBAC, Private Endpoints, and maintained compliance with industry regulations.
- Mentored junior engineers and contributed to knowledge-sharing initiatives within global teams.

Mphasis, India **Mar 2021 - Dec 2021**

Data Engineer

- Migrated and automated real-time retail data pipelines using Kafka, Flink, and Azure Synapse, achieving significant latency improvements.
- Embedded AI-driven anomaly detection in streaming pipelines, enhancing data quality and reporting accuracy.
- Implemented Terraform-driven CI/CD pipelines for consistent infrastructure deployment.
- Optimized Spark and Flink jobs for larger batch scalability and improved performance.
- Established monitoring and alerting platforms via Azure Monitor and Log Analytics, reducing production downtime by 50%.
- Created operational documentation, runbooks, and SLAs to streamline incident management and support.

CERTIFICATIONS

Microsoft Certified: Fabric Data Engineer Associate

EDUCATION

Masters in information systems (Concentration: Project Management) **May2025**
Wilmington University, Delaware, USA
Bachelor of Technology in Computer Science and Engineering **Jun 2022**
Guntur Engineering College, India

PROJECTS

Cloud-Based ETL Pipeline Automation using Azure Data Factory

- Architected and deployed automated ETL pipelines processing over 1 million monthly sales and operational records into Azure SQL Database, accelerating data availability by 60% and eliminating manual errors.
- Designed and implemented robust data validation and transformation logic, improving data accuracy and compliance with enterprise governance policies.
- Developed interactive Power BI dashboards with 10+ KPIs, enabling business stakeholders to derive real-time actionable insights and reduce decision-making time by 40%.

Real-Time Event Streaming and Analytics with Apache Kafka & Azure Event Hubs

- Engineered scalable, fault-tolerant data streaming pipelines processing 100,000+ events per hour, supporting real-time analytics and AI-backed anomaly detection.
- Leveraged Spark Structured Streaming and Apache Flink for seamless integration and low-latency fault recovery, reducing data pipeline downtime by 50%.
- Enabled near real-time monitoring dashboards, enhancing operational responsiveness and enabling proactive incident resolution.

AI-Powered Knowledge Extraction Chatbot with Azure OpenAI and LangChain

- Led the development of a secure enterprise chatbot integrating LangChain framework, FAISS vector database, and Azure OpenAI, achieving 90% response accuracy for internal document queries.
- Implemented advanced natural language processing (NLP) techniques to streamline employee access to unstructured data, improving research efficiency by 35%.
- Integrated RBAC and Azure security best practices to ensure compliance with corporate data protection standards.

Retail Sales Dashboard with Power BI and DAX

- Developed interactive retail sales dashboards using Power BI and advanced DAX calculations to enable real-time tracking and comprehensive sales analysis across 100+ stores.
- Integrated multiple data sources including Azure SQL Database and Excel, creating unified visualizations that improved inventory and promotion decision-making.
- Reduced report generation time by 30% and enhanced stakeholder insights with drill-down capabilities and custom visual filters.