

HASWANTH RAJESH

+1 9452411532 | haswanthrajeshn@gmail.com | [LinkedIn](#) | [GitHub](#)

PROFESSIONAL SUMMARY:

Azure Data Engineer with over 3 years of hands-on experience in data engineering and Databricks-based development. Skilled in designing, building, and optimizing scalable data pipelines using PySpark, Delta Lake, SQL, and Azure Data Factory. Experienced in implementing Unity Catalog, CI/CD with Azure DevOps, and data governance frameworks for secure and compliant data solutions. Adept at collaborating in agile environments, troubleshooting performance issues, and delivering production-ready data pipelines for enterprise use cases. Accomplished Azure Data Engineer with 3+ years of experience in data engineering and Databricks development. Proven ability to design, build, and optimize scalable data pipelines using PySpark, Delta Lake, SQL, and Azure Data Factory. Expertise includes implementing Unity Catalog, CI/CD with Azure DevOps, and robust data governance frameworks to ensure secure and compliant data solutions. Collaborative and agile, adept at troubleshooting performance issues and delivering production-ready data pipelines for enterprise applications.

SKILLS:

Databricks & PySpark:	Delta Lake, Databricks Workflows, Auto Loader, Structured Streaming, Unity Catalog
Azure Services:	Azure Data Factory (ADF), Azure Data Lake Storage (ADLS), Synapse, Azure DevOps
Programming & Querying:	Python (PySpark), SQL, APIs (REST/SOAP), JSON, JAVA
Data Engineering:	ETL/ELT, Data Modeling, Schema Evolution, Data Governance, Orchestration Pipelines
DevOps & CI/CD:	Git, Azure DevOps, GitHub Actions, Jenkins (exposure)
Performance Optimization:	Partitioning, Caching, Delta Optimization, Job Scheduling
Tools & Frameworks:	Power BI, JIRA, Airflow (exposure)
Soft Skills:	Analytical Thinking, Agile Collaboration, Problem Solving, Cross-functional Communication

GEN AI SKILL SET:(Learning)

Skill Category	Skills & Technologies	Proficiency	Applications
Foundation Models	LLMs, Midjourney, fine-tuning	Beginner	Custom assistants, enterprise solution design
Vector Databases	Pinecone, Chroma, FAISS, Azure AI Search	Beginner	Semantic search, similarity matching, vector store integration
RAG Methods	Chunking, Embedding Selection, Context Window Optimization, Hybrid Search	Beginner	Document Q&A, knowledge retrieval, intelligent assistants

Prompt Engineering	CoT reasoning, Few-shot learning, System prompt design, Output control	Beginner	Specialized domain adaptation, consistent LLM output, reasoning tasks
GenAI Architectures	Transformers, Attention Mechanisms, Embedding models, Token processing	Beginner	Custom AI architecture, performance optimization
LLM	LangChain, Azure OpenAI, AWS Bedrock, MLflow	Beginner	Production-grade systems
Responsible AI	Hallucination detection, Output safety, Privacy techniques, Responsible deployment frameworks	Beginner	Compliance with AI ethics, enterprise-grade safety policies

PROFESSIONAL EXPERIENCE:

Azure Data Engineer – DatafactZ, Northville, MI | Jul 2023 – Present

- Designed and developed **end-to-end data pipelines** using Databricks, PySpark, and Delta Lake for enterprise data integration.
- Created reusable, parameterized **ETL workflows** in Azure Data Factory (ADF) and orchestrated them using Databricks Workflows.
- Optimized **PySpark transformations and SQL queries**, reducing data processing time by over 40%.
- Implemented **Unity Catalog and data governance frameworks** to ensure access control, lineage tracking, and metadata management.
- Collaborated with cross-functional teams in **agile sprints** to deliver new data features and support production pipelines.
- Developed CI/CD workflows in **Azure DevOps** for automated pipeline deployment and version control.
- Performed **root cause analysis and performance tuning** for slow-running jobs and optimized cluster configurations.

Data Engineer (Databricks) – DatafactZ, Hyderabad, India | Jul 2021 – Aug 2022

- Architected **Medallion Lakehouse architecture** to enable seamless batch and streaming data processing using Databricks Auto Loader.
- Built scalable ingestion and transformation pipelines integrating **Kafka, ADF, and Delta Lake** for multi-domain data.
- Defined and enforced **data quality, encryption, and compliance policies** across multiple projects.
- Automated schema inference and incremental load logic for evolving datasets using PySpark and Delta Lake.
- Contributed to **architecture and design reviews**, providing recommendations on data modeling and partitioning strategies.
- Mentored junior engineers on PySpark best practices, debugging, and code optimization.

PROJECT HIGHLIGHTS

Enterprise Data Pipeline Modernization – Azure Databricks

- Migrated legacy ETL workflows to Databricks, achieving improved reliability and cost efficiency.
- Implemented **Delta Lake-based storage** with time travel and ACID transactions for audit-ready pipelines.
- Set up **ADF orchestration** with conditional triggers and monitoring for data refresh automation.

Metadata-driven Pipeline Framework

- Built configurable ingestion templates supporting multiple source systems with dynamic schema handling.
- Reduced manual coding efforts by 60% and improved maintainability for multi-client environments.

Data Governance Enablement – Unity Catalog Integration

- Integrated Unity Catalog for centralized governance, access management, and auditing across workspaces.
- Documented policies and automated lineage reports for compliance and security audits.

EDUCATION:

- 05/2024 **Master of Science: Computer Science**, The University of Texas at Arlington - Arlington, TX, USA

KEY STRENGTHS

- Strong technical foundation in **Databricks, PySpark, and Azure ecosystem**
- Proven ability to **optimize and troubleshoot large-scale ETL pipelines**
- Skilled in **data modeling, governance, and CI/CD automation**
- Effective communicator and **team collaborator in agile development environments**