HASWANTH RAJESH

haswanthrajeshn@gmail.com | +1 9452411532 | Farmington Hills, MI | LinkedIn | GitHub

PROFESSIONAL SUMMARY

Data Engineer with **4+ years of experience** specializing in **Azure** data engineering and **Databricks** 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 robust data governance frameworks. Collaborative and agile, adept at troubleshooting performance issues and delivering production-ready data pipelines for enterprise applications.

TECHNICAL SKILLS

Cloud Platforms	Azure Data Factory, Azure Data Lake Storage, Azure Synapse
Big Data & Processing	Delta Lake, Databricks Workflows, Auto Loader, Structured Streaming, Unity Catalog, PySpark, Kafka, Partitioning, Caching, Delta Optimization
Databases	SQL, Data Modeling, Schema Evolution, Data Governance
Programming Languages	Python, Java, JSON
DevOps & CI/CD	Azure DevOps, Git, GitHub Actions, Jenkins, Job Scheduling
Analytics & BI	Power BI
Other Tools & Frameworks	APIs (REST/SOAP), ETL/ELT, Orchestration Pipelines, JIRA, Airflow, Soft Skills, Cross-functional Communication

GENAI & MACHINE LEARNING SKILLS

Foundation LLMs, Midjourney, Vector Pinecone, Chroma, Databases **FAISS**, Azure **AI** Search matching, RAG Methods Chunking, Context Window retrieval, Prompt CoT reasoning, System prompt adaptation, Output control output, **GenAI** Transformers, Attention Custom AI architecture, Embedding models, Token processing, LLM LangChain, Azure **OpenAI**, AWS Bedrock, MLflow, Responsible AI Hallucination detection, Output safety, LLM, LLMs, fine-tuning, Pinecone, vector store, **RAG**, chunking

PROFESSIONAL EXPERIENCE

Azure Data Engineer

DatafactZ, Northville, MI | Jul 2023 – Present

- Architected end-to-end **data pipelines** using **Databricks**, **PySpark**, and Delta Lake, reducing processing time by 40%.
- Spearheaded the development of reusable, parameterized ETL workflows in **Azure Data Factory** (ADF) and orchestrated them with **Databricks Workflows**.

- Implemented **Unity Catalog** and data governance frameworks for access control 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, optimizing cluster configurations for enhanced efficiency.
- Utilized **machine learning concepts** to enhance data processing and analysis capabilities at scale.
- Built scaled data science enablement tools to support engineering processes and empower data scientists.
- Implemented and optimized data warehousing solutions to support business intelligence and analytics.

Data Engineer (Databricks)

DatafactZ, Hyderabad, India | Jul 2021 – Aug 2022

- Spearheaded the development of the Medallion Lakehouse architecture leveraging **Databricks** Auto Loader for seamless batch and streaming data processing, reducing processing time by 30%.
- Designed and implemented scalable ingestion and transformation pipelines integrating **Kafka**, ADF, and Delta Lake, enhancing data processing efficiency by 40%.
- Enforced data quality, encryption, and compliance policies across projects, ensuring 100% data integrity and regulatory compliance.
- Optimized schema inference and incremental load logic using **PySpark** and Delta Lake, increasing data processing speed by 50%.
- Provided expertise in data modeling and partitioning strategies, enhancing architecture efficiency and reducing storage costs by 20%.
- Mentored junior engineers on **PySpark** best practices, resulting in a 25% improvement in code optimization and debugging skills.
- Led the development of scalable and reusable **data pipeline**s to support business intelligence and analytics, streamlining data processing workflows by 30%.
- Collaborated with cross-functional teams to gather requirements and deliver data-driven solutions, ensuring alignment with business goals and enhancing data-driven decision-making processes.
- Implemented and optimized data warehousing solutions to support business intelligence and analytics, improving data accessibility and reliability by 40%.
- 0. Applied **machine learning** concepts to enhance data processing and analysis capabilities at scale, improving data insights and predictive analytics accuracy by 25%.
- Developed scaled data science enablement tools to support engineering processes and empower data scientists in their analyses, increasing efficiency and collaboration among teams.

EDUCATION

Master of Science in Computer Science - University of Texas at Arlington

Graduated: 05/2024