

# HASWANTH RAJESH

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

## PROFESSIONAL SUMMARY

Senior Business Analyst with 4+ years of experience in **Azure**-based data engineering, **Databricks** development, and **PySpark** pipelines. Skilled in designing scalable data solutions with **Delta Lake**, **SQL**, and **Azure Data Factory**. Proficient in implementing **Unity Catalog**, CI/CD pipelines, and enterprise data governance for secure, reliable platforms. Specializing in Generative AI solutions, including LLM-powered applications, RAG pipelines, and vector database integration for intelligent automation and decision support. Currently leading GenAI projects focused on prompt engineering, embeddings, and API-based model orchestration to deliver scalable, business-driven AI solutions. A collaborative problem-solver dedicated to driving measurable impact through high-quality data and AI platforms.

## TECHNICAL SKILLS

<b>Cloud Platforms</b>	Azure, Azure Data Factory, Azure Data Lake Storage, Azure Databricks, Microsoft Fabric, Synapse
<b>Big Data &amp; Processing</b>	Delta Lake, Databricks Workflows, PySpark, Kafka
<b>Databases</b>	SQL, Data Modeling, Schema Evolution, SQL Server, MySQL Server, Azure Data Studio
<b>Programming Languages</b>	Python, Java, JSON, Scala
<b>DevOps &amp; CI/CD</b>	Azure DevOps, Git, GitHub Actions, Jenkins
<b>Analytics &amp; BI</b>	Power BI, Tableau
<b>Tableau Other Tools &amp; Frameworks</b>	APIs (REST/SOAP), ETL/ELT, Orchestration Pipelines, JIRA, FAISS, Azure AI Search matching, RAG, Methods Chunking, Context, Window retrieval, Prompt CoT reasoning, System prompt adaptation, Output control, GenAI Transformers, Attention, Custom AI architecture, Embedding models, Token processing, LLM, LangChain, Azure OpenAI, AWS Bedrock, MLflow, Responsible AI Hallucination detection, Output safety, LLMs, fine-tuning, Pinecone, SSMS, SSIS

## PROFESSIONAL EXPERIENCE

### Azure Data Engineer

DatafactZ, Northville, MI | Jul 2023 – Present

- Architected end-to-end **data pipelines** in **Databricks** and **Delta Lake** using **PySpark** to reduce processing time by 40%.
- Spearheaded the development of reusable, parameterized **ETL workflows** in **Azure Data Factory (ADF)**, orchestrated with **Databricks Workflows**.
- Implemented **Unity Catalog** and **data governance frameworks** for access control and metadata management in cloud-native solutions.

- Developed **CI/CD workflows** in **Azure DevOps** for automated pipeline deployment and version control, utilizing **Python scripts**.
- Collaborated in agile sprints to deliver new **data features** and support production pipelines, showcasing strong **data engineering** experience.
- Utilized **machine learning concepts** to enhance data processing and analysis capabilities at scale, empowering data scientists.
- Conducted root cause analysis, tuning **cluster configurations** for enhanced efficiency, showcasing strong **SQL skills**.
- Built scaled **data science enablement tools** to support engineering processes, integrating machine learning models.
- Implemented and optimized **data warehousing solutions** for **business intelligence** and analytics, showcasing expertise in **data engineering**.
- Optimized performance by writing efficient **Python scripts** and **SQL queries**, ensuring seamless data processing and analytics.

## **Data Engineer (Databricks)**

*DatafactZ, Hyderabad, India | Jul 2021 – Aug 2022*

- Architected **Medallion Lakehouse** using **Databricks** Auto Loader for batch and streaming data processing, reducing time by 30%.
- Spearheaded design of scalable pipelines integrating **Kafka**, ADF, and **Delta Lake**, enhancing efficiency by 40%.
- Enforced data quality, encryption, and compliance policies, ensuring 100% data integrity.
- Optimized architecture efficiency and reduced storage costs by 20% through data modeling and partitioning strategies.
- Mentored junior engineers on PySpark best practices, resulting in a 25% improvement in code optimization.
- Led development of reusable data pipelines, streamlining workflows by 30%.
- Collaborated with cross-functional teams to deliver data-driven solutions aligned with business goals.
- Implemented and optimized data warehousing solutions, improving accessibility and reliability by 40%.
- Applied machine learning concepts to enhance data processing capabilities, improving insights and accuracy by 25%.
- Developed data science tools to support engineering processes and empower data scientists, increasing efficiency and collaboration.

---

## **EDUCATION**

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

*Graduated: May 2024*