

# VARSHITH GUDUR

[varshith.gudur17@gmail.com](mailto:varshith.gudur17@gmail.com) | +91 9381924190 | [LinkedIn](#) | [GitHub](#)

## Professional Summary

---

Data & AI Engineer with 3+ years of expertise in **Python, SQL, PySpark, Azure Data Factory (ADF), Databricks, and LLM integration**. Proven success in architecting end-to-end data/AI solutions, including **DocGen.dev**, an AI-powered document generation platform. Skilled in designing scalable cloud architectures, optimizing ETL pipelines, and deploying generative AI models with frameworks like **Hugging Face**.

## WORK EXPERIENCE

---

**Data & AI Engineer NTT DATA | Bangalore, India**

Jan, 2022 – Present

### Data Migration from IBM DataStage to Azure for an Automobile Company:

- Led the successful migration of ETL pipelines from IBM DataStage to Azure Data Factory, ensuring a smooth transition with minimal disruption to ongoing operations and achieving a 40% reduction in processing time.
- Re-engineered data workflows using Kimball methodologies to leverage Azure's cloud capabilities, resulting in enhanced data processing efficiency and a 30% decrease in operational costs.
- Developed dynamic data pipelines to execute multiple data sources in a single workflow, reducing rework and simplifying pipeline management for stakeholders.
- Improved data accessibility and query performance through optimized data partitioning and storage strategies on Azure Data Lake Storage, leading to a 35% improvement in data retrieval times.
- Collaborated with data scientists to integrate pretrained ML models into Azure Databricks for predictive analytics

### Data Migration from AWS S3 to Azure Databricks:

- Created linked services and pipelines on ADF to facilitate the incremental migration of data from AWS S3 to Azure.
- Utilized Azure Databricks for enhanced data processing and analytics, leveraging its advanced capabilities to perform complex transformations and analyses on the migrated data.
- Implemented Unity Catalog in Azure Databricks to centralize data governance and management, enabling consistent access controls, metadata management across the organization.
- Imported and integrated pretrained ML models into Azure Databricks for advanced predictive analytics.

### DocGen.dev – AI Document Generation Platform (*Personal Project | Architect*): [docgen.dev](#)

- **Architected and led development** of an AI-driven document automation tool using Gemini LLM
- Designed fine-tuned LLM pipelines to generate context-aware templates, improving output accuracy by 45%.
- Built a serverless backend on Azure Functions with Azure ADF for data orchestration, reducing manual effort by 50%.
- Integrated React.js for a user-friendly frontend and deployed via Azure Kubernetes Service (AKS).

## Core Skills

---

- **Tools:** Azure Data Factory, Data Bricks, Snowflake, Gen AI, LLMs (GPT, BERT), RAG , Hugging Face.
- **Languages:** Python, SQL, HTML, CSS, JavaScript, Transformers, LangChain
- **Frameworks/Libraries/Databases:** Pandas, PySpark, MySQL, SQL Server, CI/CD pipelines, Azure AI studio

## Certifications

---

- Microsoft Certified- Fundamentals of Az-900
- Databricks Certified Data Engineer Associate
- Databricks Lakehouse Platform Accreditation (v2)
- Microsoft Certified: Azure AI Fundamentals (AI-900)
- Microsoft Certified: Azure AI Engineer Associate (AI-102)

## EDUCATION

---

**Bachelor of Engineering in Computer Science & Engineering**

- Anna University - 8.8 CGPA

Chennai, India  
July, 2018 – June, 2022