

GUDUR VARSHITH

Bangalore, India | varshith.gudur17@gmail.com | +91 9381924190 | [LinkedIn](#) | [GitHub](#)

SKILLS

- **Tools:** Microsoft Azure , Power BI, DataBricks, Airflow, PySpark, Postman, Hadoop, Jupyter Notebooks, Gen AI, LLM, RAG.
- **Languages:** Python, SQL, HTML, CSS, JavaScript .
- **Frameworks/Libraries/Databases:** Pandas, PySpark, Numpy, Matplotlib, MySQL, Linux

WORK EXPERIENCE

NTT DATA | Bangalore, India

Data Engineer

Jan, 2022 – Present

- Designed and deployed robust data ingestion pipelines, integrating various sources such as **SQL databases, APIs, Excel spreadsheets, CSV files, MySQL, SFTP and SharePoint**.
- Engineered streamlined ETL processes using DBT and snowflake for data extraction, transformation, and loading, ensuring data quality and consistency. Achieved a remarkable **95% reduction in operational time**.
- Developed Python-based Databricks notebooks for data transformation and enrichment, implementing Parquet, Apache Avro, and Apache ORC techniques for optimized storage, resulting in a **60% cost reduction**.
- Leveraged **PySpark** for advanced data manipulations, aggregations, and joins, enhancing data usability for reporting and analysis purposes.
- Spearheaded data migration projects to **Azure Data Lake Storage (ADLS) Gen 2**, ensuring smooth and secure transfers from diverse sources.
- Implemented efficient data partitioning and optimization strategies in ADLS Gen 2, enhancing **data accessibility and query performance**.
- Collaborated with offshore clients to align data solutions with their business objectives, understanding their requirements and use cases.
- Monitored data pipelines for consistent flow, promptly identifying and resolving issues to maintain high availability and reliability.
- Developed **automation scripts** to schedule data tasks, **minimizing manual intervention** and improving overall efficiency.
- Designed scalable data solutions to accommodate expanding volumes and evolving business needs.

PROJECTS

Nucleus Command Center - Core Insights :

Oct, 2022 - Present

- Extracting insights from diverse source systems across API, SharePoint, MySQL, and SQL Server platforms, employing tools like Postman for API response assessment.
- Developing Python functions for function apps and deploying them efficiently, resulting in an 80% efficiency improvement.
- Building multi-layered code structures, executing data transformations, and managing code repositories on Azure DevOps.
- Utilizing Azure DevOps for monitoring project progress and facilitating smooth data migration to higher environments.
- Creating and executing ETL pipelines within Azure Data Factory for seamless progression to production.

Unity-Catalog & Gen AI :

Jan, 2022-Nov

- Led the successful implementation of a Unity Catalog in Databricks to create a centralized data catalog, enabling streamlined data discovery and management.
- Seamlessly integrated and provided access to various environments from the existing ecosystem, resulting in a notable cost reduction of 15% and a significant reduction in the time required to change different environments.
- Developed over 190+ lines of code to build and customize the Unity Catalog, ensuring it met specific organizational requirements and enhancing its functionality, data accessibility, enabling faster decision-making and data-driven insights.
- Developed a document processing pipeline with integrated relevance-aware grouping (RAG) functionality, efficiently generating document embeddings and prioritizing document chunks based on relevance to user queries.
- Managed the creation of a vector database to store embeddings and conducted extensive testing to validate system accuracy, collaborating with domain experts to optimize algorithms and enhance user experience over time.

EDUCATION

Bachelor of Engineering in Computer Science & Engineering

Chennai, India

- Anna University - 8.8 CGPA

July, 2018 – June, 2022

Certifications

- **Microsoft Certified- Fundamentals of Az-900**
- **Databricks Certified Data Engineer Associate**
- **Databricks lakehouse Platform Accreditation (v2)**