KOMALI GOSHTU

www.linkedin.com/in/komaligoshtu

+91 9030370333 | Komali.gosthu55@gmail.com

Azure Data Engineer with 2 years of experience in designing, developing, and optimizing data solutions on cloud platforms. Proficient in utilizing PySpark for large-scale data processing and analytics, alongside expertise in Azure services such as Azure Data Lake, Azure Databricks, and Azure SQL Database. Strong background in ETL processes, data warehousing, and data visualization, committed to leveraging data to drive business insights.

TECHNICAL SKILLS

- **Azure Services:** Azure Data Factory, Azure Databricks, Azure Blob Storage, Azure SQL Database, Azure Logic Apps
- **Programming Languages:** Python, PySpark, SQL
- **Data Storage and Processing:** Delta Lake, Data Lakes, Data Warehousing, Relational Databases
- **Visualization Tools:** Power BI
- **ETL/ELT Tools:** Azure Data Factory, Logic Apps

EXPERIENCE

DATA ENGINEER

INFOSYS
MAR 2022 – FEB 2024

- Built and maintained data pipelines using tools like Azure Data Factory, ensuring seamless ETL processes to support data flow and integration.
- Developed and optimized data storage solutions in Azure Synapse Analytics and Azure SQL Database to enable efficient data retrieval and analysis.
- Utilized Azure Databricks and Spark for large-scale data processing and analysis to extract actionable insights.
- Implemented data quality checks, data validation, and cleansing processes to maintain data accuracy and reliability.
- Designed real-time data streaming solutions using Event Hubs and Stream Analytics for real-time business insights.
- Worked with cross-functional teams to align data solutions with business objectives, ensuring data infrastructure meets the evolving needs of the organization.

EDUCATION

MASTER OF BUSINESS ADMINISTRATION IN HUMAN RESOURCES

Jawaharlal Nehru Technological University, Kakinada

June 2021 – June 2023 | GPA: 7.5

BACHELOR OF SCIENCE IN STATISTICS

Krishna University, Machilipatnam

Jan 2018 – Apr 2021 | GPA: 8.0

PROJECTS

RETAIL SALES DATA PIPELINE WITH ADF

- Developed and managed scalable data pipelines using Azure Data Factory and Azure Databricks, integrating data from various sources for analytics.

- Designed optimized data storage solutions (Azure Blob Storage, Azure SQL Database) and implemented efficient data models (star and snowflake schemas) to support business intelligence.

- Enhanced pipeline performance through PySpark optimizations and integrated Power BI to create interactive dashboards, enabling data-driven decision-making.

- Ensured data quality with robust validation rules, maintained governance standards, and collaborated with stakeholders to deliver tailored solutions. Comprehensive documentation was provided for scalability and onboarding.

CERTIFICATES

- Microsoft Certified PowerBI Data Analyst Associate
- Microsoft Certified Azure Data Engineer Associate