

# Rubeena Shaik

• +91 7013419285 • shaikrubeena.s23@gmail.com • <https://www.linkedin.com/in/shaik-rubeena-37188b1a3/>

## PROFESSIONAL SUMMARY

---

Proficient Azure Data Engineer with 4+ years of experience in building scalable solutions for analytics projects using Azure Data Factory and Azure Databricks. Proven track record as a results-driven professional, streamlining complex ETL workflows and enhancing end-to-end data visibility. Skilled in deploying cloud-based data solutions aligned with best practices.

## SKILLS

---

- |                    |                |                           |                           |
|--------------------|----------------|---------------------------|---------------------------|
| • ADF              | • Python       | • SQL                     | • Spark                   |
| • Spark SQL        | • PySpark      | • Azure Data Lake storage | • Azure Synapse Analytics |
| • Azure Logic Apps | • Databricks   | • Azure Key Vault         | • Azure SQL Database      |
| • GIT              | • Azure DevOps | • Data Warehouse          |                           |

## PROFESSIONAL EXPERIENCE

---

### Mass Mutual Global Services India LLP, Hyderabad Associate

Jan'2024 – Present

- **Designed and developed a Master Data Ingestion Framework**—an end-to-end, scalable pipeline within a single notebook that supports ingestion of any master data source, significantly reducing the need for multiple notebooks and manual intervention.
- **Enhanced scalability and maintainability** by automating the onboarding of new master data sources. This framework improved operational efficiency and ensured consistency across data ingestion processes.
- **Implemented a unified and dynamic merge logic** to handle inserts and updates across all dimension tables using an active indicator flag. This design allows seamless integration of new master files with consistent schema but different data values.
- **Built a robust and high-performance Fact Data Ingestion pipeline** capable of processing large volumes of data efficiently. Utilized techniques such as checkpointing, broadcasting, and data explosion to optimize performance and reduce processing time. Collaborated with stakeholders to gather transformation rules.
- **Developed a dependency-aware fact pipeline retriggering mechanism**, enabling automatic parallel invocation of dependent fact pipelines in Databricks whenever upstream master data is refreshed. This ensured data consistency across reporting layers without manual reprocessing.
- **Used Git-based version control** (Azure Repos) to manage ADF pipeline versions and collaborated with the team through pull requests and code reviews.
- **Worked with Azure Key Vault** to manage credentials and securely configure linked services in ADF, following enterprise security best practices.
- **Set up alerts in ADF** to monitor pipeline failures and performance issues, enabling timely notifications and faster resolution in the production environment.

**Programmer Analyst**

- Assisted in developing and maintaining data pipelines using Azure Data Factory for automating data ingestion from SQL Server and cloud storage.
- Participated in writing basic data transformation logic using Mapping Data Flows to cleanse and standardize incoming data.
- Followed version control practices using Azure Repos (Git) for source code management and collaborated through pull requests.
- Helped configure schedule-based triggers for automating pipeline execution and monitored pipeline performance using built-in ADF tools.
- Worked closely with the DevOps team to understand CI/CD deployment processes using Azure DevOps pipelines for ADF artifacts.
- Created basic documentation for pipelines and participated in team meetings to discuss ongoing data integration tasks.

**Programmer Analyst Trainee**

- Extracting, transforming and loading data from source system to targets using Informatica Intelligent Cloud Services (IICS) and Informatica PowerCenter.
- Re-Platforming the Database from Teradata to Vertica in existing ETL flow.
- Developed mappings using various transformations like update strategy, lookup, stored procedure, Router, Filter, sequence generator, joiner, aggregate transformation and expression.
- Developing Application views to fetch data from various source tables.
- Worked with different sources such as Relational and flat files.
- Involved into performance tuning of mapping and automating the ETL process using Tivoli workload scheduler.

**EDUCATION**

---

**Bachelor of Engineering, Sona college of Technology, Salem**  
Electronics and Communication Engineering: (CGPA: 9.27/10.0)

May-2021

**CERTIFICATION**

---

AZ-900: Microsoft Azure Fundamentals Exam

Mar-2024

**APPRECIATIONS/ACHIVEMENTS**

---

- Successfully completed a critical, time-sensitive project data load, earning appreciation from senior management for the timely delivery and impact.
- Received an Outstanding Performance Rating for two consecutive years: 2022–2023 and 2023–2024.
- Honored with the "**Pat on the Back**" Award in 2024 for excellence in project delivery.
- Recognized by the business team for proactively identifying a critical production bug before go-live, demonstrating strong commitment and attention to detail.
- Consistently appreciated by the business team with multiple kudos for identifying and resolving critical production issues effectively.