

Mahaboob Sheik

Data Engineer

+91- 9182912647 | Mahaboobsheik26@gmail.com | [linkedin.com/in/mahaboob-sheik](https://www.linkedin.com/in/mahaboob-sheik) | Hyderabad,Telangana

SUMMARY

Data Engineer with over **3 years** of experience in transforming complex data into actionable insights. Skilled in utilizing Azure cloud services, including **Databricks, Data Factory**, and SQL to streamline data collection, transformation, and analysis. Proficient in **PySpark, SQL, Apache Spark**, and **Python**. Proven success in leading data engineering projects, managing cross-functional teams of up to 10 members, and delivering solutions that improved processing efficiency by up to **20%**.

WORK EXPERIENCE

Data Engineer

April 2021 - Present

Modak Analytics

Hyderabad, Telangana

- Managed and directed a team of 10 individuals to optimize product delivery timelines, resulting in a **20%** increase in on-time project completions.
- Designed Scalable big data pipelines using **Hadoop, Spark, Scala** and other distributed computing technologies to process **50+** Terabytes Monthly.
- Improved query performance by **20%** through analysis and optimization, enhancing data efficiency.
- Developed GLAPI certified **CI/CD** pipelines for automated deployment using SonarQube, SEC API on **Azure DevOps**, leading to **30%** faster deployment.
- Orchestrated seamless data ingestion and storage across multiple sources into Azure Data Lake Storage (**ADLS Gen 2**) amplifying data accessibility by **30%**.
- Optimized **GCP** resources and implemented Storage Lifecycle Management, cutting GCP expenditures by over **10%** annually, enhancing cost-efficiency.
- Developed efficient data pipelines using **Scala** and **Spark**, cutting run time by **20%**.
- Performed root cause analysis on **ETL** process failures and introduced over 10 changes that decreased recurrence by **38%**.
- Created and deployed a fully automated CI/CD pipeline for StreamSets, slashing deployment time by **50%** and facilitating the smooth migration of **40+** pipelines from Dev to QA with minimal downtime.

SKILLS

- **Programming Languages:** PySpark, Python, SQL, Scala, Data Structures and Algorithm.
- **Cloud Services and Technologies:** Azure Databricks, Azure Data Factory, Azure Synapse, ADLS, Azure DevOps and Delta Lake.
- **Data Processing:** Hadoop, Hive, Apache Spark, Kafka, StreamSets.

EDUCATION

Bachelor of Technology in Electronics and Communication Engineering

2017 – 2021

SRKR Engineering college

CERTIFICATIONS

- Microsoft Certified Azure Fundamentals (AZ-900)
- StreamSets WhiteBelt Certification