Somu shekhar Potnuru

Data Engineer

✉ shekharsomu98@gmail.com 📞 7730054406 📍 Hyderabad 📅 17/06/1998 🏴 Indian

in linkedin.com/in/potnurusomushekhar98

PROFILE

Results-driven Data Engineer with over 3 years of experience in building and optimizing **ETL pipelines**, data workflows, and cloud-based data solutions. Proficient in Azure Data Services, including **Azure Data Factory**, **Databricks**, **Azure Data Lake Gen2**, and **Azure Synapse Analytics**, with expertise in **Python**, **Scala** and **PySpark** for scalable data processing. Strong knowledge of **SQL**, **NoSQL** databases, Data **Warehousing**, and Big Data frameworks like **Hadoop**. Experienced in pipeline orchestration, data modeling, performance tuning, and end-to-end deployment of data engineering solutions. Passionate about designing efficient, reliable, and scalable data systems to drive business insights and decision-making.

EDUCATION

Bachelor's of computer science and technology   2016 – 2020 | Jalandhar, India
Lovely professional university
6.1 CGPA

Intermediate education   2014 – 2016 | Visakhapatnam, India
Sri Gayatri college
79%

School   2014 | Visakhapatnam, India
Siva sivani school
8.8 GPA

PROFESSIONAL EXPERIENCE

Associate Consultant | Senior Analyst | Analyst   05/2022 – present | Hyderabad, India
Capgemini Technology Services India Pvt. Ltd.
Role Summary:
Led and contributed to multiple data engineering and analytics initiatives, including the design and implementation of scalable data pipelines, cloud-based ETL solutions, and predictive analytics models using modern **Azure** and **AWS** technologies. Played a key role in automating **data workflows**, and driving actionable insights for business stakeholders.
Key Responsibilities & Achievements:
- Data Engineering & Cloud Technologies:
  - Designed and built robust ETL pipelines using **Azure Data Factory**, **Databricks**, and **PySpark** to process large-scale datasets from multiple sources.
  - Automated **data ingestion**, **transformation**, and **validation** workflows using **CI/CD** pipelines in **Azure DevOps**.
  - Deployed and managed datasets on **Azure Data Lake Storage, Azure Synapse Analytics,** and **Azure SQL** Databases for scalable storage and querying.
- Big Data & Analytics:
  - Hands-on experience working with Apache Spark for **distributed data processing** and implementing optimized transformations for large datasets.
  - Developed and deployed big data solutions using Azure Databricks, contributing to faster **analytics** and **cost-effective** storage.
  - Built dashboards and visualizations using **Tableau** to communicate business insights to stakeholders across domains.
- AI/ML & Predictive Modeling
  - Worked with **AWS SageMaker** to train, evaluate, and deploy machine learning models.
  - Developed Proof of Concepts using Weibull analysis, **regression models**, and clustering for asset performance monitoring and operational insights.
- DevOps & Automation
  - Developed reusable components and CI/CD pipelines to automate model deployment and data refresh cycles using **Azure DevOps** and **GitHub** Actions.
  - Integrated Terraform for infrastructure-as-code and streamlined deployment of cloud resources.
- Collaboration & Agile Delivery

- Worked closely with cross-functional global teams including data scientists, architects, and business stakeholders to gather requirements and deliver scalable solutions.
- Practiced **Agile** methodologies, delivering incremental value through sprint-based planning and continuous integration.
- Security & Compliance
  - Ensured compliance with organizational and client-specific security standards through **role-based access control** and **data encryption** in transit and at rest.
  - Participated in security assessments and audits in adherence to Microsoft Cloud Background Check standards.

SKILLS

Big Data
 Map reduce, Hadoop, distributed computing.

Web Development
Django framework and REST APIs

Tools
Jupiter notebook, VS Code, MySQL, Postgres, SageMaker

Programming
Python, SQl, No SQL, Apache Spark

Cloud
Azure Datafactory, Azure databricks, Azure Synapse Analytics, Azure SQL Databases, Azure Event hub.

CERTIFICATES

- Azure cloud fundamental (AZ-900)
- AWS cloud practitioner (CLF-01)

AWARDS

Rising star award
Capgemini

Shining star award
Unilever

Best team award
Unilever