

# Mounika Boggavarapu

📍 Florida | ✉ mounikaboggavarapu123@gmail.com | 📞 813-568-7138 | [🐙 GitHub](#) | [🌐 LinkedIn](#)

## SUMMARY

Results-driven Data Engineer with 3+ years of experience designing scalable data pipelines, ETL workflows, and data models using Apache Spark, PySpark, AWS, and Azure. Skilled in data transformation, real-time streaming, batch processing, and automation, with a proven track record of reducing data processing times by 50% and improving query performance by 40%. Certified AWS Solutions Architect with expertise in data warehousing, cloud infrastructure, and analytics solutions to drive business growth and data-driven decision-making.

## WORK EXPERIENCE

### Tata Consultancy Services (TCS)

Data Engineer

Hyderabad, India  
Aug 2022 - Jul 2023

- **Architected and Automated ETL Pipelines:** Designed and implemented robust **ETL workflows** using **AWS Glue** and **Apache Spark**, automating healthcare data integration and processing to streamline operations.
- **Devised Scalable Data Architectures:** Constructed secure and efficient data storage solutions on **Amazon S3** and developed serverless processing frameworks using **AWS Lambda**, reducing operational costs and enhancing scalability.
- **Optimized Data Analysis Processes:** Automated intricate workflows with Python (Boto3) to implement data transformation processes, enhancing both real-time streaming and batch reporting for accurate insights.
- **Implemented Advanced Warehousing Solutions:** Leveraged **Snowflake** and **Amazon Redshift** for high-performance data warehousing, enabling seamless querying and transforming patient and operational datasets.

### Hexaware Technologies

Cloud Engineer Intern

Chennai, India  
May 2020 - Jul 2022

- **Executed Seamless Cloud Migrations:** Enabled the migration of 15+ customer applications to AWS using Redis, Memcached, Postgres, CloudFront, WAF, Lambda, and CloudFormation, achieving 99.9% system uptime and enhanced scalability.
- **Accelerated Application Performance:** Optimized website load times by 95% through advanced Angular build strategies and AWS CloudFront caching, boosting user retention and engagement by 25%.
- **Engineered Cost-Effective Architectures:** Designed multi-tier applications leveraging private/public subnet models with AWS CloudFormation, reducing annual infrastructure costs by \$30,000 while ensuring security and scalability.
- **Built Serverless Backend Systems:** Contributed to developing a serverless backend for a financial app, leveraging AWS Cognito Hosted UI to reduce development time by 40% and enhance authentication workflows.

## CERTIFICATIONS

[Data Science Professional Certification](#)

[AWS Certified Cloud Practitioner certificate](#)

## SKILLS

**Programming & Development:** Python, Java, SQL, R, PHP, C, JavaScript, Bash, Flask, Django, PLSQL

**Data Science & Machine Learning:** TensorFlow, PyTorch, Keras, Scikit-learn, Tableau, Data Analysis, Data Collection, Time-Series Forecasting, Data Preprocessing

**Cloud & Databases:** AWS (S3, Lambda, Glue, Redshift, EMR, Athena), Azure, Snowflake, PostgreSQL, MySQL, MongoDB

**Development Tools:** Git, Docker, Kubernetes, Terraform, Jenkins, Apache Airflow, Data Governance, and Pipeline Automation

**Specialized Knowledge:** ETL Pipelines, Data Warehousing, Big Data (Apache Spark, PySpark), Generative AI, Deep Learning, NLP, CUDA, Parallel Processing, Data Transformation, Real-Time Processing, Batch Processing, and Data Modeling.

## EDUCATION

**Master of Science in Computer Science:** The University of South Florida (USF)

Aug 2023 - Present

**Bachelor of Engineering in Computer Science:** IIIT Nuzvid (GPA: 3.69)

Aug 2018 - May 2022

## PROJECTS

### PySpark Project: End-to-End Data Analytics and Visualization

- **Tech Stack:** PySpark, Databricks, Python, Big Data Analytics, KPI Calculation, Dashboard Development
- **Engineered Scalable Analytics Solutions and Dashboards:** Designed and implemented a PySpark-based pipeline to process over 2 million records, reducing data preparation time by 40%, doubling data processing speed, and efficiently handling datasets exceeding 10GB. Automated the computation of 10+ KPIs, enabling 25% faster decision-making, and developed interactive Databricks dashboards with dynamic visualizations, improving reporting efficiency by 30%.

### Azure Databricks Project: End-to-End Data Pipeline with Delta Lake Integration

- **Tech Stack:** Azure Databricks, Delta Lake, PySpark, Data Lake Storage, Power BI, Big Data Analytics, API Integration
- **Developed Scalable Data Pipelines and Optimized Analytics:** Designed and implemented a PySpark-based ETL pipeline with advanced data pipeline automation to process over 5 million records, supporting robust real-time streaming and batch workflows, reducing data preparation time by 50% and ensuring 99.9% data accuracy through advanced Delta Lake management. Automated the computation of 15+ KPIs, accelerating decision-making by 40%, created Power BI dashboards that improved reporting efficiency by 30%, and enhanced query performance with Delta Table optimizations, cutting execution time by 60%.