

Lahari Karrotu

Email: laharikarrothu@gmail.com | Phone: 321-367-9669 | Melbourne, Florida 32901

Professional Summary

- Over 3 years of expertise in Data Engineering and Data Warehousing, delivering robust and scalable data solutions.
 - Proficient in designing and implementing ETL pipelines using PySpark, Databricks, and AWS services.
 - Experienced in managing data lakes, warehouses, and SQL databases on AWS cloud platforms.
 - Skilled in building scalable data models, optimizing SQL queries, and resolving star and snowflake schema issues.
 - Developed RESTful APIs with Python and FastAPI for secure and efficient data access.
 - Reduced data processing time by 30% through the development of optimized data pipelines and frameworks.
 - Experience with big data technologies, including Apache Kafka, Hadoop, and Spark.
 - Proficient in CI/CD pipeline automation using Git, Jenkins, and Kubernetes.
 - Expertise in data visualization using Power BI and Tableau, delivering actionable insights through dashboards and reports.
 - Knowledgeable in machine learning and predictive analytics, with experience in implementing ML models on AWS EC2.
 - Skilled in automating workflows using Apache Airflow and Shell scripting for operational efficiency.
-

Education

Master of Science in Computer Science
Florida Institute of Technology, Florida | 2024
Bachelor of Science in Computer Science
KL University, India | 2022

Specialization: Big Data Analytics and Data Science

Technical Skills

- **Cloud Platforms:** AWS (Redshift, Lambda, S3, Glue, Kinesis, Athena, SNS/SQS), Azure
- **Databases & Data Warehousing:** Redshift, PostgreSQL, MySQL, DynamoDB, Athena

- **Programming & Scripting:** Python, Node.js, SQL, JavaScript (ES6), Scala
 - **Big Data Technologies:** Apache Spark, Kafka
 - **ETL & Data Processing:** AWS Glue, Lambda, S3, Redshift
 - **BI Tools:** Tableau, AWS Quicksight, Power BI
 - **DevOps & Automation:** AWS CloudFormation, Docker, CI/CD pipelines
 - **AI Technologies:** Knowledge of integrating AI and ML models in data workflows
 - **Version Control:** Git, GitHub
-

Professional Experience

Data Engineer Intern

Anguliyam, Atlanta | August 2024 – Present

- Developed and deployed AI-powered voice assistants using GPT-based NLP models for medical SOP form filling, streamlining automation and enhancing accuracy.
- Engineered scalable data pipelines integrating AWS S3 and Redshift, automating data flow and enabling real-time reporting.
- Designed and deployed serverless microservices using AWS Lambda, optimizing backend workflows and reducing response time by 30%.
- Built custom APIs to support secure communication between data processing systems and application layers, ensuring compliance with privacy regulations.
- Leveraged data visualization tools like Tableau to deliver dashboards, offering actionable insights to stakeholders.
- Conducted ETL processes, ensuring data integrity and consistent transformation across AWS pipelines.
- Implemented cost optimization strategies in AWS environments, reducing operational expenses while maintaining high system availability.
- Developed automation scripts for data validation, improving data quality and reducing manual intervention by 40%.
- Optimized query performance in Redshift, reducing data retrieval times through advanced SQL techniques.
- Collaborated with business teams to understand analytical requirements, delivering customized reports and insights aligned with key metrics.
- Provided documentation and hands-on training for team adoption of AI solutions, boosting operational efficiency.

Data Engineering Intern

Cognizant, Hyderabad, India | 2022–2023

- Designed and implemented ETL pipelines using AWS services such as S3, Glue, and Redshift for efficient data processing and reporting.

- Leveraged AWS Lambda and S3 for building serverless data processing pipelines, automating data transformation, and improving system performance by 20%.
- Integrated BI tools like Tableau to visualize key business metrics and real-time analytics, enhancing stakeholder decision-making capabilities.
- Optimized Redshift performance with complex SQL queries, improving query speed and reducing data retrieval times.
- Generated, wrote and ran SQL script to implement the DB changes including table update, addition or update of indexes, creation of views and store procedures.
- Responsible for designing new FACT or Dimension Tables to existing Models.
- Implemented stored procedures, functions, views, triggers, packages in PL/SQL.
- Documented the data requirements and system changes into detailed functional specifications.

Data Analyst Intern

EPAM Solutions, Remote, India | 2021–2022

- Reduced data processing time by 30% by developing Data Ingestion framework functionalities in Python.
 - Configured AWS Data Pipeline to load data from S3 to Redshift.
 - Utilized AWS EMR for large data transformations between Amazon S3 and DynamoDB.
 - Imported and exported data using SQL Server Integration Services (SSIS) and Data Transformation Services (DTS).
 - Optimized and tuned Teradata views and SQL to enhance batch performance and data response times.
 - Supported CI/CD and continuous testing activities to streamline deployment processes.
 - Automated scripts and workflows using Apache Airflow and shell scripting for daily production execution.
 - Populated HDFS and PostgreSQL with large datasets using Apache Kafka.
 - Implemented and tuned models on AWS EC2, optimizing for best algorithm performance.
 - Created and maintained reports in Tableau to track model performance.
 - Developed Spark Python modules for machine learning and predictive analytics in Hadoop.
 - Conducted data blending and preparation using Alteryx and SQL for Tableau data sources.
-