Ranga Rohit Nallamolu

nallamolu.rangarohit@gmail.com | linkedin.com/in/ranga-rohit-nallamolu | https://github.com/rangarohitnallamolu | +1 978 569 7460

SUMMARY

Experienced data engineer skilled in building scalable data pipelines and real-time analytics using AWS, Azure, Apache Spark, Kafka, and Airflow. Proficient in integrating machine learning with Python to enhance business operations and leveraging visualization tools like Tableau and Power BI for data-driven decision-making. Seeking a challenging role to apply technical expertise and solve business challenges.

EDUCATION

University of Massachusetts — Lowell, MA
Master of Science in Computer Science — **Sep 2022 - May 2024**

Christ University — Bangalore, India
Bachelor of Technology in Computer Science — **Aug 2017 - May 2021**

SKILLS

Programming Languages: Python, Java, R, Java Script, SQL, NoSQL, XML, Unix / Linux Scripting

Cloud Platforms and Databases: AWS, Azure, GCP, Snowflake, DBT, MySQL, DynamoDB, Cassandra, PostgreSQL

Data Engineering: ETL, Spark, Kafka, Hadoop, Airflow, Docker, Redshift, Databricks, Tableau, PowerBI, Data Analytics, Pandas, Numpy

EXPERIENCE

Data Engineer | *University of Massachusetts, Lowell, MA* — **Dec 2022 - May 2024**

- Designed and implemented efficient AWS data pipelines for Graduate, Online & Continuing Education, and Finance departments to ETL data into a warehouse using Cassandra and python-Spark, reducing manual efforts by 60%
- Built a real-time data monitoring system with AWS CloudWatch and Lambda, enabling proactive detection and resolution of data pipeline issues, minimizing downtime and ensuring continuous data flow
- Automated data validation processes in the ETL pipeline, ensuring data integrity and accuracy, which led to a 25% reduction in error rates and improved overall data quality
- Optimized complex data warehouse queries in AWS Redshift using partitioning and indexing, improving performance by 30%
- **Capabilities**: Python, EC2, ECR, ECS, S3, SNS, VPC, DynamoDB, Scala, Cassandra, Tableau, Terraform, Apache Spark

Research Assistant | *University of Massachusetts, Lowell, MA* — **Dec 2022 - Nov 2023**

- Engineered robust data pipelines in Azure Data Lake and Azure Databricks to process and analyze 12 years of microbial shotgun metagenomics data, identifying a 70% correlation between Coprococcus bacteria and diabetes
- Developed and validated machine learning models using Random Forest and XGBoost in python for diabetes risk prediction, achieving an accuracy of 85%, which deepened the understanding of the gut microbiome's influence on metabolic health
- Created interactive Power BI dashboards to visualize key findings, including the significant role of Coprococcus in diabetes, facilitating effective communication of complex data analyses and supporting data-driven decision-making in healthcare research
- Collaborated with cross-functional research teams to integrate findings from multiple projects, enhancing data representation and outcomes
- **Capabilities**: Azure Data Lake Storage, Azure Databricks, Python, R, Machine Learning, Statistical Analysis, Power BI

Data Engineer | *Anakin, Bangalore, India* — **Mar 2021 - Jul 2022**

- Engineered and implemented robust data pipelines using AWS, MongoDB, and Spark to process over 10 million records daily, achieving a 20% increase in operational efficiency
- Optimized data ingestion and transformation processes with Apache Spark and AWS, leading to a 30% boost in processing speed and ensuring data integrity across diverse sources
- Developed and managed CI/CD pipelines using tools like Jenkins, AWS CodePipeline, and Kubernetes to automate deployment processes, increasing deployment frequency and reducing downtime
- Designed and constructed a scalable Hive data model for a large data lake, handling over 10 million daily transactions with 99.9% accuracy
- **Capabilities**: AWS, EC2, S3, Lambda, RDS, Redshift, Glue, EMR, CloudWatch, Apache Spark, Apache Kafka, Python, Scala, CI/CD

PROJECTS

Real-Time Stock Market Data Analysis

- Acquired real-time stock data for Google, Amazon, Nvidia, and Apple via Yahoo Finance API. Set up a data processing service on AWS EC2, implementing Apache Kafka for seamless data transfer to AWS S3, and used AWS Glue for transformation
- Applied an LSTM model with 80% accuracy and visualized trends in Tableau, enhancing real-time analytics

Data Analysis for Supply Chain & Sales

- Automated the ingestion of supply chain and sales data through an ETL pipeline, utilizing dbt for data transformation, Snowflake for centralized warehousing, and Apache Airflow for efficient workflow management
- Analyzed key metrics, delivering insights that improved decision-making by 40%, and visualized results through interactive dashboards