Ganesh Makkina

New Brunswick, NJ | (732)558-8640 | gm832@scarletmail.rutgers.edu | linkedin.com/in/ganesh-makkina/

EDUCATION

Rutgers University — **New Brunswick, NJ**
Masters in Statistics - Data Science — **September 2023 - May 2025**
Relevant Coursework: Reg. & Time Series, Prob. & Statistics, Adv. DB Management, Fin. Data Mining, Data Wrangling.

TECHNICAL SKILLS

- **Programming Skills:** Python, Scala, Java, SQL, NoSQL, C++
- **Cloud Platforms:** AWS (S3, Glue, Redshift, Lambda, Kinesis, EMR, IAM, Lake Formation, Step Functions), GCP (BigQuery, Dataflow, Data Fusion, Airflow)
- **Big Data & ETL Tools:** Databricks, Apache Spark, Apache Kafka, AWS Glue, Hadoop, Informatica, ELT Frameworks
- **Databases & Data Warehousing:** SQL Server, MongoDB, MySQL, Redshift, Neo4j, ArangoDB, Snowflake
- **Data Stream Processing:** AWS Kinesis, Apache Kafka
- **API and Data Frameworks:** RESTful APIs, Spring Boot, Microservices Architecture, Domain-Driven Design
- **Data Visualization & Other Tools:** Tableau, AWS QuickSight, Power BI, Looker, Git, Docker, Kubernetes, Jenkins, Azure DevOps, Unix/Linux Shell Scripting

PROFESSIONAL WORK EXPERIENCE

Quantiphi (AWS Partner) — **Mumbai, India**
Data Engineer — **August 2021 – August 2023**

Project 1: Enterprise Data Lake Migration and Modernization

- Collaborated on migration of a global technology client's legacy data infrastructure (e.g., Elasticsearch) to an AWS-based data lake using S3, Databricks for processing and Redshift for warehousing, improving performance.
- Validated and transformed historical datasets containing millions of records using Scala Spark on Databricks, ensuring data integrity for real-time data streaming via Kinesis and Lambda for dynamic front-end updates.
- Engineered SQL queries in Redshift (and Spark SQL on Databricks) with Common Table Expressions (CTEs) and stored procedures, improving query speeds by 20% and enabling faster data access for critical business applications.
- Built scalable AWS data pipelines (using S3, Glue for ELT, Lambda, Databricks, Redshift) for seamless data processing, using Managed Airflow (MWAA) to automate workflows and ensure real-time dashboard updates.
- Spearheaded the design and implementation of a cost-optimized AWS data architecture, achieving a 30% infrastructure cost reduction while significantly boosting query and front-end performance.

Project 2: Scalable COVID-19 Data Ingestion and Analytics Platform

- Collaborated with a leading healthcare client to design and implement end-to-end AWS data pipelines for processing 10+ million records daily from APIs of healthcare agencies and public datasets.
- Leveraged AWS Lambda for real-time API extraction, S3 for ingestion, AWS Glue, and Python Spark on Databricks for large-scale transformations, ensuring quality data for analytics/ML.
- Built scalable data architectures using S3 data lake, Hive metastore, and Databricks SQL to query and transform millions of records daily, achieving a 30% reduction in preparation time and enabling rapid insights for public health.
- Orchestrated workflows with Managed Airflow (MWAA), integrated Kinesis for real-time updates, and CI/CD (AWS CodePipeline & CodeBuild) for deployment/monitoring, ensuring reliable delivery of cleansed datasets.
- Developed interactive dashboards in AWS QuickSight, visualizing COVID-19 trends and actionable insights, empowering stakeholders to make critical data-driven decisions during the pandemic.

PROJECTS

Project - Rutgers CSO Operations & Scheduling Automation Platform

- Developed a full-stack automation platform (React, Node.js) for Rutgers, featuring a scalable MySQL DB (AWS RDS) optimized for complex scheduling & compliance, reducing manual workload by ~30 hrs/week.
- Designed intricate data models for diverse operations and built robust Node.js RESTful APIs to handle complex logic, CRUD, and transformations, ensuring data integrity and system reliability.
- Architected data integrations (e.g., Google Calendar API) and pipelines with scalable S3 storage (concept); engineered for AWS deployment utilizing EC2/ECS and RDS, applying cloud data management best practices.

Project - Twitter Data Analytics and Search Engine for High-Performance Querying

- Engineered a high-performance Twitter search engine storing millions of records in MySQL (users) and MongoDB (tweets/retweets), improving query efficiency through optimized indexing and caching.
- Built a Streamlit-based analytics dashboard with predefined search filters and real-time query tracking, reducing response times by 80% to under 200ms and enhancing search performance.

CERTIFICATION

- **AWS Certified Associate Data Engineer:** Proficient in designing and building scalable data pipelines and data warehouses on the AWS platform.
- **GCP Certified Associate Cloud Engineer:** Skilled in managing and optimizing cloud infrastructure and solutions on Google Cloud Platform.