

Ramachandra Nalam

nrcvamsi@gmail.com | +1 5086140301 | www.linkedin.com/in/ramachandra-nalam/ | Seattle, WA – 98104

PROFESSIONAL SUMMARY

Result-oriented professional with 7+ years in building large-scale data pipelines, real-time analytics, and data warehousing across technology, retail, and financial domains. Expertise in Apache Spark, Kafka, Airflow, and Snowflake, designing pipelines, and reducing data lag by 35%. Proficient in data modeling, automation, dashboard development, and achieving cost savings and workload reduction. Proven success in partnering with cross-functional teams, delivering high-impact data solutions, and driving measurable business outcomes.

EXPERIENCE

Data Engineer: Meta

Sep 2024 – Present

- Created a real-time messaging analytics pipeline handling 50B+ daily events using Apache Kafka and Spark, applying dbt transformations, simplified data modeling by 60% and supported self-serve metrics
- Developed multi-dimensional data models for user engagement tracking, creating reusable dbt semantic layers powering 15+ executive dashboards in Looker and reducing report generation time
- Implemented data quality framework with Great Expectations and Python validators, maintaining 99.9% data accuracy and preventing 12+ business-critical incidents through anomaly detection automation
- Coordinated cross-functional analytics projects with Product, Growth, and Safety teams, building Airflow ETL pipelines processing 2TB+ daily, contributing to an 8% improvement in message delivery success rate
- Designed Tableau dashboards for marketing and sales teams, tracking key performance indicators, achieving a 40% increase in data accessibility for business stakeholders
- Designed data warehouse solutions on Snowflake for Messenger analytics, applying query optimizations and partitioning strategies, lowering costs by 40% while supporting 500+ monthly active users

Data Engineer: Amazon

Jul 2022 – Aug 2024

- Led ETL pipeline development and optimization in AWS using AWS Glue, processing 750GB daily from S3 to Aurora DB with 98% data accuracy
- Managed Apache Kafka streaming pipelines handling 500,000 events per second, shortening data lag by 35% and enabling real-time analytics for business reporting
- Developed PySpark and Spark SQL applications, transforming 1TB daily from CSV, Parquet, and JSON files, increasing data aggregation precision
- Presented weekly data visualizations and actionable insights to stakeholders, integrating SQL, Python, and BI tools, automating reports, improving strategic planning timeliness, increasing campaign reach by 20%
- Optimized Redshift queries for reporting pipelines, reducing CPU utilization by 52% and saving 9 hours weekly in monitoring, troubleshooting, and on-call tasks for analysts

Data Engineer: UB

Jan 2021 – Jun 2022

- Consolidated 20,000+ student records, identifying graduation timelines and patterns, improving predictive accuracy for academic interventions by 28%
- Orchestrated different machine learning models, including logistic regression, decision tree, and random forest, enhancing risk prediction for 3,500 at-risk students
- Achieved 88.74% accuracy with Random Forest, contributing to a strategic initiative, increasing 4-year graduation rates by 23% for targeted student groups
- Deployed Snowflake and Redshift data warehouses, partitioning tables and optimizing queries, leading to a 40% boost in resource consumption and supporting internal users
- Managed cloud storage on AWS S3 and Azure Data Lake for 50TB+ data, enabling standardized access, governance, and data reliability across departments

Data Engineer: Nike India

Aug 2018 – Dec 2020

- Orchestrated ETL pipelines for 5 queues, ingesting over 6 million records daily from different sources, achieving 98% data availability and consistency for business intelligence teams
- Automated PySpark jobs processing 20TB+ daily datasets, transforming all records per BI requirements, minimizing data processing errors while supporting cross-functional analytics needs

- Executed Apache NiFi workflows transferring data across shared folders, triggering automated notifications, and ensuring 95% of Hive table loads completed each business day
- Engineered Python scripts aggregating login activities for 150,000 users, transferring processed data to Snowflake, and achieving a 75% reduction in potential security incidents for the organization

Analytics Engineer: Freelancer

Jul 2017 – Dec 2020

- Deployed a TensorFlow-based load forecasting system for households, leading to a 10% reduction in energy costs and minimizing customer disruptions by 8%
- Led a cross-functional team to prototype system deployment, monitored outcomes, and implemented CI/CD with GitHub Actions and AWS, achieving a 30% acceleration in deployment speed

SKILLS

- | | | | |
|-------------------|----------------------|-----------------------|----------------|
| • Python/R | • Apache Airflow | • OLAP/OLTP | • Git |
| • Scala | • AWS Glue | • Jenkins | • Compliance |
| • SQL | • Azure Data Factory | • NLP | • Azure DevOps |
| • JavaScript | • Jira | • Canonical Datasets | • Big Data |
| • TypeScript | • Tableau | • Data Vault Modeling | • Databricks |
| • Apache Spark | • Power BI | • KPI Dashboards | • OpenAI |
| • Apache Hadoop | • Google Data Studio | • Azure | • Enterprise |
| • Apache Kafka | • Matplotlib | • CI/CD | • Data Lakes |
| • Google BigQuery | • Seaborn | • Cataloging | |

PROJECTS

Competitive Marketing Intelligence using Gen-AI

Feb 2022 – Jul 2022

- Directed development of a GenAI chatbot with LLMs, NLP, and vector search, extracting intelligence from earnings calls, contributing to a 30% improvement in data accessibility
- Engineered automation synthesizing financial pages weekly, lowering manual research workload by 20%, accelerating insight extraction and decision readiness for cross-functional executive teams

EDUCATION

University at Buffalo

Jan 2021 – Aug 2022

Master of Science in Data Science

K L University

Jun 2014 – Aug 2018

Bachelor of Technology in Computer Science