Rohan Reddy Bandi

+1 (813) 893 5627 | brohanreddy24@gmail.com | Tableau | LinkedIn | Portfolio

SUMMARY

Results-driven **Data Engineer** with expertise in building scalable **ETL/ELT pipelines**, cloud data platforms, and real-time analytics solutions across **AWS**, **Azure**, and **Snowflake** environments. Proficient in automating workflows and optimizing performance using **Databricks**, **Kubeflow**, **Argo Workflows**, and **Apache Airflow**, while ensuring efficient resource utilization. Skilled in leveraging **Snowpark**, **AWS Glue**, **Terraform**, and advanced **SQL** for data transformation, performance tuning, and operational reporting. Experienced in enabling enterprise-wide analytics and machine learning initiatives through clean, secure, and auditable data pipelines. Adept at collaborating with cross-functional teams to support data-driven decision-making while ensuring compliance with **HIPAA**, **IRB**, and **HL7** standards.

EDUCATION

Master of Science in Business Analytics and Information Systems	**CGPA - 3.94/4**
University of South Florida, Tampa FL USA	August 2022 - May 2024

SKILLS

- **Programming & Scripting:** Python, R, SQL, C++, Java, JavaScript, Bash, Unix Shell, SAS, Scala, Go
- **Data Engineering & Processing:** ETL/ELT Pipeline Design, Apache Airflow, dbt, Apache Kafka, HL7 Integration, REST APIs, GraphQL APIs, JSON, Parquet, Avro, Data Integration, Data Modeling, Data Lake Architecture, DataOps
- **Databases & Data Warehousing:** PostgreSQL, SQL Server, MySQL, Snowflake, Redshift, Oracle, REDCap API, MongoDB, Cassandra, Azure Cosmos DB, Amazon DynamoDB, Kubernetes, Docker, Terraform, IaC
- **Cloud & HPC:** Azure (Synapse, Data Factory, Key Vault, Databricks, Cosmos DB), AWS (S3, Lambda, Redshift, EC2, Glue, Athena, EMR, Kinesis), GCP (BigQuery, DataFlow/Pub-Sub, DataProc), Linux HPC Clusters,
- **Data Preparation & Automation Tools:** Alteryx, Salesforce Integration, Apache NiFi, Talend, Informatica
- **Data Science & Statistical Tools:** Scikit-learn, TensorFlow, MATLAB, NumPy, pandas, SciPy
- **Machine Learning & AI**: Keras, PyTorch, XGBoost, Hugging Face Transformers, LangChain, NLTK, SpaCy, Reinforcement Learning, NLP, Caret (R ML Toolkit), Predictive Lead Scoring, APIs for ML/AI Features
- **Visualization & Reporting:** Tableau, Power BI, Streamlit, Dash, Matplotlib, Seaborn, Plotly, Looker, QlikView
- **Project & Collaboration Tools:** JIRA, Confluence, Trello, Scrum/Kanban, Asana, Reproducibility & Documentation
- **Version Control & DevOps:** Git, GitHub, GitLab, Azure DevOps, CI/CD Pipelines, Jenkins, GitOps

PROFESSIONAL EXPERIENCE

Data Engineer | CVS Health | USA
January 2024 - Present

- Developed multiple scalable ETL pipelines using **Databricks**, **Kubeflow**, and **Snowflake**, seamlessly integrating with **Azure Cloud**, **Azure Data Lake Storage (ADLS)**, **AKS**, and **SFTP** to/from JDA servers.
- Designed and implemented an end-to-end pipeline using **Snowpark**, leveraging **Snowflake warehouse compute** exclusively to run and score ML models, eliminating the need for tools like Airflow and Databricks, and reducing compute costs by **20%**.
- Built reusable **Argo workflow templates** and developed node-specific ETL jobs in **Kubeflow**, optimizing resource allocation and cutting infrastructure costs by **50%**.
- Contributed to a **$10 million business impact**, including **$500K** in value from enhanced precision tracking enabled by Kubeflow over Databricks.
- Improved workload efficiency in **Azure Kubernetes Service (AKS)** through advanced workflow automation and strategic resource provisioning.
- Reduced unnecessary node computations by optimizing model input features, enhancing overall ML pipeline performance.
- Integrated **Snowflake stored procedures** and **vectorized UDFs** for complex data transformations, improving analytical capabilities.
- Engineered robust ETL pipelines in **Databricks** to migrate data from **Oracle to Snowflake**, addressing schema mismatches and preserving data integrity.
- Automated and orchestrated workflows using **Apache Airflow** and fine-tuned processing with **PySpark**, improving runtime performance.

- Diagnosed and resolved critical issues in data workflows, ensuring accurate event processing and stabilizing key business operations.
- Improved data pipeline runtimes by **20%** through **Apache Spark** performance tuning and efficient debugging practices.

Tools & Tech: Python, SQL, Snowflake, Databricks, Snowpark, Kubeflow, Azure (ADLS, AKS), Oracle, Apache Airflow, PySpark, Argo Workflows, SFTP, Git, Bash, Linux, Vectorized UDFs, Stored Procedures

Data Engineer | Accenture | Hyderabad, India
August 2020 - July 2022
- Designed and deployed scalable ETL/ELT pipelines using *Python* and *SQL* to extract, transform, and harmonize EHR and clinical data from multiple systems into **AWS Redshift** and **Snowflake**, enabling integrated HIV registry analytics.
- Built research-grade data models and schemas to support CFAR clinical studies, optimizing performance for real-time and batch processing while ensuring compliance with **HIPAA**, **IRB**, and **HL7** standards.
- Leveraged **Apache Airflow**, **AWS Glue**, and **Terraform** to automate data workflows, manage cloud infrastructure as code, and execute statistical modeling using *Python* (**Pandas**, **NumPy**) and *R* (**tidyverse**, **dplyr**) for multi-site collaborations.
- Developed complex **SQL** views, stored procedures, and monitoring scripts to support automated quality assurance metrics, data validation, and operational reporting within **AWS Glue** and **AWS Redshift** environments.
- Maintained a centralized data dictionary and metadata repository for multi-source datasets, collaborating cross-functionally to deliver clean, secure, and auditable data assets for **machine learning** and research publications.
- Optimized **data warehouse** performance and **AWS** resource usage by tuning **SQL** queries, applying partitioning strategies, and provisioning infrastructure with **Terraform**, achieving a 30% reduction in processing time and cloud costs.

Tools & Tech: *Python*, *SQL*, *R*, **Apache Airflow**, **AWS Glue**, **AWS Redshift**, **Snowflake**, **S3**, *PostgreSQL*, *Microsoft SQL Server*, **Terraform**, **GitLab**, *Confluence*, **HIPAA**, **IRB**, **HL7**, *Bash*

Associate Data Analyst | Tech Mahindra| Hyderabad, India
May 2019 - July 2020
- Led end-to-end **data analysis initiatives on AWS**, supporting insights for over **100 enterprise applications** with scalable and cost-efficient analytics solutions.
- Built and maintained **data pipelines** using **AWS Glue**, **S3**, and **Redshift**, transforming and processing **5+ TB of data daily** for real-time reporting.
- Designed and delivered **interactive dashboards** in **AWS QuickSight**, **Power BI**, and **Tableau**, supporting strategic decisions across sales, operations, and leadership.
- Wrote and optimized complex **SQL queries** in **Amazon Redshift** to extract, clean, and analyze large datasets, identifying trends that improved efficiency by **15%**.
- Conducted advanced analysis using **Python** and **Pandas**, including **anomaly detection**, **A/B testing**, and **predictive modeling** to support data-driven marketing and operational strategies.
- Automated **ETL workflows** and data validation using **Python** and **AWS services**, improving reliability and reducing manual intervention by **30%**.
- Partnered with business stakeholders to gather reporting requirements and deliver **custom analytics solutions**, while creating **data dictionaries** to support self-service BI.

Tools & Tech: SQL, Python, AWS, Power BI, Tableau, Amazon Redshift, Git, Cloud Monitoring, RBAC, A/B Testing, Anomaly Detection, Predictive Modeling

CERTIFICATIONS
- **AWS Data Engineer Certification**
- **AWS Machine Learning Associate**
- **Business Analysis & Process Management**
- **Database Operations in MariaDB Using Python**