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

SUMMARY

Experienced Data Engineer with 4+ years of success supporting clinical research through advanced data management, statistical modeling, and analytics.

- Specialized in curating longitudinal registries, harmonizing real-world data, and automating quality control processes to drive accurate reporting and health surveillance.
- Proficient in SQL, R, Python, Stata, and Git, with hands-on experience in EHR data extraction, clinical systems integration, and developing secure, scalable pipelines across research environments.
- Experienced in configuring server tasks, managing data dictionaries, validating datasets, and executing analyses aligned with IRB, HIPAA, and NIH compliance standards.
- Adept at working with interdisciplinary teams, mentoring junior scientists, and delivering visualizations, documentation, and statistical outputs that inform research publications and funding proposals.

EDUCATION

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

SKILLS

- **Programming & Scripting**: Python, R, SQL, C++, Java, Bash, Unix Shell, SAS
- **Data Science & Statistical Tools**: Scikit-learn, TensorFlow, Bioconductor, MATLAB, NumPy, pandas, SciPy, Statsmodels
- **Data Engineering & Processing**: Apache Airflow, Apache Spark, HL7 Integration, REST APIs, JSON, Parquet, ETL Design
- **Databases & Warehousing:** PostgreSQL, SQL Server, MySQL, Snowflake, Oracle, REDCap API
- **Biomedical & Research Systems:** EHR Systems (Epic, Cerner), Clinical Registries, NIH/UCSF Standards
- **Cloud & HPC:** Azure (Synapse, Data Factory, Key Vault, Databricks), AWS (S3, Lambda, RedShift, EC2, Glue, Athena), GCP (BigQuery, DataFlow, Pub/Sub), Linux HPC Clusters, Terraform, Kubernetes, Docker
- **Machine Learning & AI:** Keras, PyTorch, XGBoost, Hugging Face Transformers, LangChain, NLTK, SpaCy, Reinforcement Learning, NLP, Time Series Forecasting, MLOps (CI/CD, Model Registry)
- **Visualization & Reporting**: Tableau, Power BI, Streamlit, Dash, Matplotlib, Seaborn, Plotly
- **Project & Agile Tools:** JIRA, Confluence, Trello, Scrum/Kanban, Reproducibility & Documentation Standards
- **Version Control & DevOps:** Git, GitHub, GitLab, Azure DevOps, CI/CD Pipelines
- **Professional Competencies:** Data Governance, Research Collaboration, Data Quality, Cross-Disciplinary Communication

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 **Azure Synapse** 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**, **Databricks**, 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 in **Azure Data Factory** 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 cloud 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, Databricks, Azure Synapse, Azure Data Factory, Snowflake, 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

- Business Analysis & Process Management – (Link)
- Working with BigQuery – (Link)
- Database Operations in MariaDB Using Python – (Link)

ACADEMIC PROJECTS

- **Real time Stock Market Analysis:** Designed and implemented a real-time data pipeline using Python, Apache Kafka, and AWS services. Built a stock market simulation app to generate CSV data, sent to Kafka on Amazon EC2 via Boto3, and stored in Amazon S3. Cataloged data with AWS Glue and analyzed it using Amazon Athena. (Link)
- **Made Easy Pharmacy Store Website:** We developed and implemented a user-friendly and visually appealing e-commerce website for a pharmacy store, providing seamless online shopping experience for customers. We used technologies to develop the website HTML, CSS, JavaScript, Flask. (Link)
- **Airbnb project:** With SQL Developer, created a database like Airbnb's business workflow using the data-dimensional modeling approach. It helps capture business data and can handle operations smoothly. From the data, we can infer targeted customers and the footfall of the customers. (Link)
- **Data visualization Project:** Developed interactive Tableau dashboards to analyze the EPL 2018-19 season, assess the growth of Indian startups, explore Netflix content trends using IMDb data, and evaluate the health impacts of wildfires.(Link)