Rohan Reddy Bandi

Allen, TX | +1 813-893-5627 | brohanreddy24@gmail.com | Portfolio | Tableau | LinkedIn

PROFESSIONAL SUMMARY

Self-Motivated and Accomplished Data Engineer adept at leveraging **SQL, Python, and Scala** to drive robust data solutions, visualization, and machine learning across diverse platforms. Proficient in

- **Data Processing & Analysis:** Managed 50+ petabytes of streaming data using **Apache Spark, Apache Flink, and Kafka**, ensuring actionable insights and outcomes.
- **ETL Expertise**: Orchestrated transformations of 100+ terabytes into **Amazon Redshift** and **Google BigQuery**, ensuring data integrity via **Talend**, Apache NiFi, and stringent ETL processes.
- **Database Management & Optimization:** Fine-tuned query performance across 5 different DBMS, handling 500+ queries hourly, bolstering efficiency and system availability.
- **Cloud-Based Solutions and Big-data:** Designed and deployed cloud solutions across **AWS, GCP, and Azure**, managing 1,000+ data workflows, ensuring seamless operations.
- **Data Quality & Security:** Implemented robust data quality checks, security measures, and compliance standards using Talend, Apache NiFi, and security protocols.
- **Automation & Optimization:** Automated deployment processes, orchestrated Kubernetes, improving system efficiency by 20% through performance tuning and optimized data pipelines.
- **Data Visualization:** Crafted professional visualizations using **Tableau, Power BI**, enhancing data representation for stakeholders, enabling insightful decision-making.
- **Machine Learning Insights:** Derived actionable insights-built 30+ predictive models using **Python, R, TensorFlow, PyTorch**, contributing to data-driven decision-making and efficiency.
- **API Development and Integration:** Proficient in designing and implementing APIs to facilitate seamless data exchange and drive efficient communication between systems, enhancing overall data accessibility and real-time insights.
- Proficient professional with comprehensive expertise in Unix systems, adept coding skills, and a mastery of Python programming language. Experienced in **RESTful API** development and implementation, coupled with a strong background in Artificial Intelligence (AI) technologies. Excels in leveraging these skills to engineer innovative solutions and deliver robust, data-driven applications in dynamic environments.

EDUCATION

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

SKILLS

Programming	: R, PYTHON, SQL, NoSQL, C, C++, JAVA, JAVASCRIPT, Scala, bash, Unix,
Data Modelling & ETL	: Spark, Flink, Nifi, Kafka, Informatica, Talend, Data Bricks
Database Management Systems	: Hadoop, Cassandra, Hive, Apache HBase, Amazon Redshift, Google Big Query, Microsoft SQL Server, MySQL, PostgreSQL, Oracle, MongoDB,
Data ware Housing	: Amazon Redshift, Google Big Query, Snowflake, Teradata, Microsoft Azure Synapse Analytics
Data Visualization	: Tableau, Power BI
Cloud Platforms	: Amazon Web Services (AWS), Google Cloud Platform (GCP), Microsoft Azure
Data Modelling and SQL	: SQL, Apache Avro, Parquet, Schema Registry
Version Control	: Git, GitHub
Containerization	: Docker, Kubernetes
Monitoring and Logging	: ELK Stack (Elasticsearch, Logstash, Kibana), Prometheus, Grafana, Splunk, New Relic
Data Quality and Validation	: Nifi, Talend, Trifecta, Great Expectations
Data Streaming	: Kafka, Apache Pulsar, AWS Kinesis, Google Cloud Pub/Sub, Azure Event Hubs, Flink
Data Security and Encryption	: Apache Ranger, Apache Knox, Amazon KMS, Azure Key Vault, GCP Key Management Service
Job Scheduling and Orchestration	: Apache Airflow, Kubernetes, Luigi
Data Serialization	: JSON, Avro, Parquet, ORC, XML, Protobuf
Machine learning tools	: TensorFlow, Pandas, NumPy, PyTorch, scikit-learn, PySpark
Soft skills	: Problem solving skills, mentoring, strong critical thinking.

PROFESSIONAL EXPERIENCE

Data Engineer | Tryittech- CVS Health| USA | January 2024 – Present

- **Designed and implemented scalable ETL pipelines** using **Spark**, **Informatica**, and **AWS Glue**, reducing data processing times by **50%** and enhancing governance across **Azure** and **Google Cloud**.
- Built distributed systems with **Kafka**, **Apache Pulsar**, and **AWS Kinesis**, optimizing real-time data ingestion into **Snowflake**, **Amazon**

Redshift, and **Google BigQuery**, improving efficiency for **5+ enterprise systems**.

- Improved SQL query execution speed by **40%** in multi-terabyte environments using **Hadoop**, **Cassandra**, and **PostgreSQL**, addressing complex data modeling requirements.
- Integrated machine learning models with **Scikit-Learn**, **TensorFlow**, and **Pandas**, achieving a **40% boost** in reporting efficiency with **Hadoop**, **MapReduce**, and **Hive**.
- Designed **Star and Snowflake schemas** for **AWS Redshift**, incorporating **100+ data quality validations** with **Great Expectations**, and managing metadata through **AWS Glue Data Catalog**, ensuring **99.9% data accuracy**.
- Automated workflows with **Apache Airflow** and **Kubernetes**, improving system uptime by **25%** and streamlining operations in **AWS** and **Azure Event Hubs** ecosystems.
- Developed high-performance **data lakes** and **warehouses** using **Talend**, **NiFi**, and **Databricks**, processing over **10 TB of data monthly** with **Flink** and **Scala** for seamless transformations.
- Enhanced data security and compliance by reducing security incidents by **60%**, conducting rigorous validations using **Great Expectations**, and applying advanced **Statistical Analysis techniques**.
- Delivered **30+ dynamic dashboards** with **Tableau** and **Power BI**, combining data from **Teradata**, **Oracle**, and **SQL Server** for enterprise-wide insights.
- Collaborated with business analysts to create **20+ Tableau reports**, aligning analytics with business needs and providing insights that impacted **15+ business units**.

Data Engineer | Accenture | Hyderabad, India | May 2019 – July 2022

- **Supported analysis and management teams** in delivering over **15 projects**, actively participating in design, development, and production phases. Improved data pipeline performance by **30%** through ETL and validation processes using **SSIS**.
- Supervised **three Hadoop clusters** with **Cloudera Manager**, addressing **95% of data quality issues** and developing data narratives to resolve biases stemming from incomplete datasets.
- Designed ETL pipelines utilizing **Python**, **PySpark**, and **Azure Data Factory**, handling **10 TB of data monthly** across sources like **MongoDB**, **Blob Storage**, and **on-premise systems**. Integrated data into **Azure Data Lake**, boosting analytics readiness for **Azure Synapse Analytics**.
- Enhanced governance and security for **20+ datasets** by implementing **Azure Purview** and **Azure IAM roles**, achieving **100% compliance** and ensuring seamless access control via **Azure Key Vault**.
- Transformed data using **dbt (data build tool)** to ensure **99.9% consistency** and minimize transformation errors by **40%**, while orchestrating workflows with **Apache Airflow** for improved operational efficiency.
- Partnered with data scientists to streamline machine learning workflows, automating deployment for **10+ models** through **Azure ML Studio**, and optimized storage performance by employing the **Star Schema** in **Azure SQL Data Warehouse**.
- Created **50+ visualizations and reports** using **Power BI**, **Tableau**, and **Excel**, facilitating business insights and supporting strategic decisions with enhanced data reliability.
- Managed **Azure Kubernetes Services (AKS)** to scale **big data clusters**, cutting operational costs by **20%** and boosting infrastructure performance through optimized SQL queries.
- Conducted rigorous testing and automation reviews during **10 Agile sprints**, presenting progress and outcomes effectively to stakeholders in sprint demos.

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)

CERTIFICATION

- Database Operations in MariaDB Using Python from Infosys - (Link)
- Business Analysis & Process Management - (Link)
- Working with Big Query – (Link)