

ROHAN REDDY B

TAMPA, FL | +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

University of South Florida, Tampa FL USA

August 2022 - May 2024

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, Trifacta, 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 | Truist | USA | March 2023 – April 2024

- Maintained global data pipelines, integration, ETL processes, and data management in data lakes and warehouses using **AWS Glue, Azure Synapse, S3, and Redshift**, ensuring smooth data flow and governance across environments.
- Built complex distributed systems for data handling, using **Kafka, Hadoop, MapReduce, Hive, AWS, and Snowflake** to handle

large-scale data, collecting metrics, and building analytical pipelines.

- Led ETL pipeline creation, migration of ETL code from **Talend** to **Informatica**, and wrote complex **SQL/PLSQL** queries while automating ingestion using **Python** and **Scala** for seamless data transformations and migration across platforms.
- Installed **Kafka** on **Hadoop** clusters, integrated with **HDFS**, and optimized NoSQL solutions using **DynamoDB** and **HBase** for real-time data streaming and unstructured data management.
- Designed **Star** and **Snowflake** schemas for **AWS Redshift**, created data models, implemented data quality checks with **Great Expectations**, and managed metadata using **AWS Glue Data Catalog** to maintain data integrity and structure.
- Employed **CircleCI**, **GitLab**, and **Jenkins** for automating continuous integration, deployment of ETL workflows, and data pipeline management, improving efficiency and reducing deployment times.
- Implemented **Prometheus** and **Grafana** for monitoring ETL pipelines and optimized pipeline performance using **Airflow**, **AWS Lambda**, and **Azure Kubernetes Services**, leveraging **Docker** for microservice isolation.
- Leveraged **AWS Lambda** and **Kinesis Analytics** for serverless processing and real-time insights and implemented **AWS Step Functions** to automate machine learning workflows in **Amazon Sage Maker**.
- Collaborated with business analysts to develop **Tableau** dashboards and reports pulling from data lakes, providing deep insights and aligning data analytics with business requirements.

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

- Worked with analysis and management teams to support their requirements, involved in the entire project lifecycle from design discussions to production deployment, and conducted ETL and data validation using **SQL Server Integration Services (SSIS)**.
- Managed and monitored **Hadoop clusters** using **Cloudera Manage**, investigated data quality issues, and generated presentable narratives based on data biases due to incompleteness.
- Developed and optimized **ETL pipelines** using **Python**, **PySpark**, and **Azure Data Factory**, ingesting data from **MongoDB**, **Blob Storage**, and on-premise systems like **MySQL** and **Cassandra** into **Azure Data Lake** for analysis in **Azure Synapse Analytics**.
- Implemented **Azure Purview** for data governance, **Azure Key Vault** for secure credential storage, and integrated **Azure IAM roles** for access control, enhancing overall data security and management.
- Used **dbt** (data build tool) for transforming data in the warehouse, ensuring data integrity and consistency across datasets, and orchestrated ETL workflows using **Apache Airflow**.
- Collaborated with data scientists and product owners to automate machine learning model training and deployment using **Azure ML Studio**, and optimized data models in **Azure SQL Data Warehouse** using the **Star Schema**.
- Created visualizations and performance reports using **Power BI**, **Tableau**, and **Excel**, pulling data from **Azure Analysis Services** and presenting findings to business users and stakeholders.
- Managed and monitored **Azure Kubernetes Services (AKS)** for efficient resource scheduling and scaling of big data clusters, and optimized SQL queries to improve data infrastructure performance and reliability.
- Conducted unit testing, UAT, and End-to-End automation design reviews in an Agile environment, communicating project progress and insights to leadership and stakeholders during sprint demos.

ACADEMIC PROJECTS

- **Real time Stock Market Analysis:** Designed and implemented a real-time data processing pipeline using Python, Apache Kafka, and AWS services. Developed a Python-based stock market simulation app to generate data from CSV files, which was sent to Kafka on Amazon EC2 using Boto3. The data was then consumed into Amazon S3, cataloged by AWS Glue, and queried with Amazon Athena for analysis. ([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:** Accomplished data analyst with expertise in leveraging IMDb data to analyze and visualize trends in Netflix's vast content library. Proficient in using datasets to uncover insights on content popularity, directorial influence, and cross-platform performance. Strong skills in data visualization, trend analysis. ([Link](#))

CERTIFICATION

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