

ROHAN REDDY

+1 (813) 893 5627 | brohanreddy24@gmail.com | [Tableau](#) | [LinkedIn](#)

PROFESSIONAL SUMMARY

Self-Motivated Data Professional with 4 years of experience in data engineering and analysis, proficient in SQL, Python, and Scala. Processed 10+ terabytes of data weekly using Apache Spark and Kafka, and designed 50+ ETL pipelines on AWS, GCP, and Azure. Created 30+ dashboards in Tableau and Power BI, reducing reporting time by 40%, and built 10 predictive models using TensorFlow, improving forecasting by 25%. Optimized database performance by 20%, handling 500+ queries daily, and automated 80% of deployments with Kubernetes.

- **Data Processing & Analysis:** Processed over 10 terabytes of streaming data weekly using Apache Spark and Kafka, driving actionable insights for business decisions.
- **ETL Development:** Designed and implemented 50+ ETL pipelines with tools like Talend and Apache Ni-Fi, transforming and loading data into Amazon Redshift and Google Big Query with 99.9% accuracy.
- **Database Management:** Managed and optimized 500+ queries per day across 5 database systems, improving query performance by 20% through indexing and optimization.
- **Cloud Platforms:** Built and maintained 100+ data workflows on AWS, GCP, and Azure, ensuring 24/7 uptime and scalability for large-scale data operations.
- **Data Quality & Security:** Implemented automated validation checks and security protocols, achieving 95% compliance with internal and external data governance standards.
- **Data Visualization:** Developed 30+ interactive dashboards using Tableau and Power BI, reducing manual reporting time by 40% for stakeholders.
- **Automation & Optimization:** Automated 80% of deployment processes using Kubernetes and Python, resulting in a 15% efficiency improvement in data pipelines.
- **API Development:** Designed and deployed 10+ RESTful APIs, enabling seamless data exchange and real-time insights for cross-functional teams.
- **Machine Learning:** Built 10 predictive models using Python and TensorFlow, improving forecasting accuracy by 25%.
- **Programming Expertise:** Proficient in Python, SQL, and Unix systems, with hands-on experience in scripting and automation.

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 |
| 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 | CVS Health | USA

January 2024 - Present

- Designed and implemented scalable ETL pipelines using **Spark, Informatica, and AWS Glue** to process **patient records, claims data, and pharmacy transactions**, reducing data processing times by **50%** and improving regulatory compliance across **Azure and Google Cloud**.

- Built distributed systems with **Kafka, Apache Pulsar, and AWS Kinesis** to optimize real-time ingestion of electronic health records (EHR), insurance claims, and prescription data into **Snowflake, Amazon Redshift, and Google BigQuery**, enhancing efficiency for **5+ healthcare enterprise systems**.
- Optimized SQL query execution speed by **40%** in multi-terabyte environments, enabling faster processing of healthcare analytics, population health metrics, and predictive modeling using **Hadoop, Cassandra, and PostgreSQL**.
- Designed Star and Snowflake schemas for **AWS Redshift** to manage structured healthcare data, incorporating **100+ data quality validations** for patient demographics, clinical trials, and claims data, ensuring **99.9% accuracy** and compliance with **HIPAA regulations**.
- Automated workflows with **Apache Airflow and Kubernetes**, streamlining ETL processes for patient outcomes, insurance claims, and provider performance reporting, improving system uptime by **25%** in **AWS and Azure Event Hubs** ecosystems.
- Developed high-performance healthcare data lakes and warehouses using **Talend, NiFi, and Databricks**, processing over **10 TB** of medical and pharmacy data monthly with **Flink and Scala** to support advanced analytics.
- Enhanced data security and compliance by reducing security incidents by **60%**, ensuring **HIPAA and HITRUST compliance** through rigorous validations using **Great Expectations** and advanced statistical analysis techniques.
- Delivered **dynamic dashboards** with **Tableau and Power BI**, providing insights into patient care, prescription trends, and hospital performance by integrating data from **Teradata, Oracle, and SQL Server**.
- Collaborated with business analysts to create **Tableau reports**, aligning analytics with healthcare business needs and supporting initiatives for cost optimization, patient engagement, and provider efficiency across **15+ business units**.

Data Engineer | Accenture | Hyderabad, India

August 2020 - July 2022

- Developed and optimized **scalable ETL pipelines** using **Apache Spark, Hadoop, and Python**, processing **5+ terabytes** of data daily for **real-time data integration** and transformation.
- Designed **data architectures** leveraging **Hadoop, Apache Spark, and Azure Synapse Analytics**, ensuring efficient processing of **structured and unstructured data** for **20+ applications**.
- Implemented advanced **query tuning** techniques on **SQL Server**, improving response times by **40%** for complex analytical queries, enhancing performance for **500+ reports**.
- Utilized **geospatial analytics tools** (e.g., **PostGIS, ArcGIS**) to develop data solutions for **mapping and spatial analysis**, addressing **10+ client use cases** with high precision.
- Automated **data ingestion workflows** using **Python** and **Apache Airflow**, reducing manual effort by **30%** and increasing system reliability by **25%**.
- Collaborated with cross-functional teams to create reusable **data pipeline templates** and **ETL patterns**, standardizing practices across **15+ projects**.
- Led efforts to implement **data governance frameworks** ensuring compliance with **GDPR** and regulatory standards, achieving **100% compliance**.
- Supported cloud migration from legacy systems to **AWS** and **Azure**, minimizing downtime to **5 hours**, ensuring seamless transition for global operations.
- Created technical documentation and delivered training on **ETL, data pipelines, and cloud tools** to **15+ junior engineers**, promoting knowledge sharing and team development.

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.
- Supported **data governance** efforts by implementing **role-based access control (RBAC)** and ensuring data integrity through proactive **quality checks** and **cloud-based monitoring**.

CERTIFICATIONS

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