ROHAN REDDY BANDI

TAMPA, FL | +1 813-893-5627 | brohanreddy24@gmail.com|Portfolio|Tableau

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 | Truist | USA | March 2023 – April 2024

- Performed extensive data analysis and processed over 50 petabytes of streaming data using **Apache Spark, Apache Flink, Apache Kafka, Hadoop**, and **Python.**

- Loaded and transformed around 100 terabytes of data into **Amazon Redshift** and **Google Big Query,** ensuring data consistency and integrity through **ETL** development with **Talend, Apache NiFi, Hadoop,** and **Python.**
- Optimized query performance and maintained high availability for 5 different Database Management Systems (DBMS), handling a combined workload of 500 queries per hour, including **MySQL, PostgreSQL, MongoDB,** and **Amazon Redshift.** Fine-tuned SQL queries implemented indexing strategies, and partitioning techniques to achieve a 30% increase in query performance.
- Designed and deployed cloud-based solutions across **AWS, GCP,** and **Microsoft Azure,** handling over 1,000 data workflows using services such as **Amazon S3, Google Cloud Pub/Sub, Azure Event Hubs,** and **Azure Synapse Analytics**. Leveraged Python for scripting and automation in cloud environments.
- Implemented data quality checks using tools like **Talend, Apache NiFi, Great Expectations,** and **Tableau** to ensure data accuracy and completeness. Utilized Python for scripting data quality checks.
- Implemented data security measures using **Apache Ranger, AWS Key Management Service, Azure Key Vault**, and **Hadoop** to encrypt sensitive data and ensure compliance with security standards. Used Java for implementing security measures.
- Designed and implemented **CI/CD** workflows for data processing pipelines, ensuring automated integration, testing, and deployment of data transformations. **Utilized Jenkins, Git, and Python for CI/CD implementation.**
- Managed orchestration, achieving a 30% increase in scalability for data processing by containerizing 10 data applications using **Apache Flume, Docker, Hadoop, and Java.**
- Enabled real-time visibility into system performance by implementing monitoring and logging solutions, reducing troubleshooting time by 25% for data processing bottlenecks using **ELK Stack, Prometheus, Grafana, Hadoop, and Python.**

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

- Leveraged **Databricks platform** to optimize big data processing, resulting in a **25%** increase in processing speed and efficiency. Utilized Scala and Python for developing scalable data processing applications.
- Spearheaded **A/B testing procedures**, designing experiments, and scrutinizing data to furnish crucial insights for informed decision-making, showcasing adeptness in experiment design and hypothesis testing using **Tableau**, **Python**, and **SQL**.
- Developed scalable data processing applications in **Scala** leveraging the **Spark** framework for efficient and high-performance data analytics solutions. Utilized **Python** for additional scripting and automation.
- Implemented and optimized complex **ETL pipelines** using **Spark** to process large volumes of data stored in **Google BigQuery** and **Hadoop**, ensuring data integrity, performance, and reliability. Utilized **Python** for scripting within **ETL processes**.
- Engaged with **15 cross-functional** teams to gather requirements and create **data models**, resulting in a 25% improvement in retrieval speed and a **20%** reduction in storage costs. Used **SQL** for data modeling and querying.
- Achieved a **30%** increase in query performance by fine-tuning **SQL** queries, implementing indexing strategies, and partitioning techniques.
- Implemented validation processes leading to a **15%** decrease in data errors or inconsistencies, adhering to data governance best practices. Utilized **Python** for scripting validation processes.
- Automated 50 data workflows, resulting in a **30%** reduction in manual effort and improved efficiency in monitoring data pipelines' health. Employed **Python** for scripting automation.
- Improved database system availability by 25%, resolving 100 critical issues within a year, ensuring reliability for critical operations. Utilized Java for troubleshooting and resolving critical issues.
- Handled **20 data streams** with **Apache Kafka, AWS Kinesis, Google Cloud Pub/Sub**, and **Hadoop**, enabling real-time
- processing, resulting in a **40%** reduction in processing time. Utilized **Java** for handling data streams.

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)