Rahul Kengeri

Columbia, SC | (516) 923-1045 | rahulgowdakk@outlook.com | rahulkengeri.net | github.com/ruby4785

SKILLS

Programming Languages: Python (Pandas, NumPy, Scikit-Learn, Matplotlib, Seaborn) | SQL | Java | PySpark
Databases: MySQL | MongoDB | PostgreSQL | Snowflake
Cloud Platforms: AWS (EC2, S3, EMR, Lambda) | Azure (Synapse, Data Factory, Stream Analytics)
Data Visualization: Tableau | Power BI | Excel | SAS
Tools & Technologies: Apache Kafka | Apache Airflow | Git | Azure DevOps | Docker | CI/CD pipeline

WORK EXPERIENCE

Quality Engineer Co-op   *Middletown, PA*
Schneider Electric   August 2024 – December 2024
- Built a data pipeline to calculate DCR KPI to replace manual updates using **Python** and **Databricks**, which pulls data from **Amazon Redshift**, saving 10 hours monthly and reducing reliance on single team members.
- Analyzed NQC (Non-Quality Cost) data to identify and target high-impact SKUs, improving **process Sigma** from **3.62 to 4.91** leading to an **84.4% reduction** in quality-related waste.
- Improved MRB (Material Review Board) data collection rate by **130%** with a redesigned, accessible QR code form and created a **Tableau dashboard** that turned data into actionable insights, which led to better inventory tracking and decision-making.
- Scripted multiple **SQL queries**, including one to extract customer names for 18K delivery numbers, saving **150 hours** of manual lookups and improving the Quality team's efficiency.

Analytics Engineer Intern   *Bengaluru, KA*
Radiant Info Systems Ltd   August 2021 – February 2022
- Collaborated on an automated **ELT data pipeline** ingesting real-time passenger data from 900+ city buses into Snowflake data warehouse, using **Apache Kafka** for streaming data and **Spark** for transformations.
- Enhanced **data ingestion efficiency** by 15% during peak hours by implementing micro-batching in Kafka Connector configuration which reduced the number of individual database connections and enabled managing **8M+ events** daily.
- Overhauled the existing data warehouse in Snowflake by implementing **Snowflake schema** which resulted in a **12% decrease in storage needs**, reducing costs.
- Created 5+ **data-driven** dashboards featuring comprehensive **analytics and visualizations**, providing **management teams** with **actionable insights** that enhanced the experience for 4M+ daily commuters.

PROJECTS

Data Science Job Market Analysis: ETL Pipeline and Visualization| (AWS, PySpark, Airflow, Snowflake, Tableau)
- **Engineered an ETL data pipeline** using **Airflow** for a 10k-row dataset containing jobs and salaries in the data science domain, leveraging Apache PySpark, AWS services (EC2, EMR, S3), and Snowflake.
- Performed **Star Schema** data modeling using SQL to build a **data warehouse for analytics**, visualized processed data in Tableau.
- Built tailored **Tableau dashboards** with 5+ key **insights** for job seekers, enhancing career decision-making.

Supermarket Data Pipeline and Analysis Using Azure Services | (Azure, PySpark, Power BI)
- Orchestrated an end-to-end **data processing pipeline** leveraging **Azure services** to analyze 10,000+ sales data from a nationwide superstore, establishing an automated pipeline for deriving sales insights.
- Utilized **Data Factory** to ingest raw data, performed data cleaning, and transformation using **PySpark** on Azure Synapse, and deployed **Dedicated SQL Pool** as a data warehouse, enabling ad-hoc and advanced querying.
- Designed an intuitive **Power BI dashboard** containing key KPIs (optimal discount range, lucrative product categories, sales figures) for stakeholders to drive more sales and profitability.

EDUCATION

Master Of Data Science   *Rochester, NY*
Rochester Institute of Technology   August 2022 – December 2024

Bachelor of Computer Science and Engineering   *Bengaluru, KA*
Global Academy of Technology   August 2018 – July 2022

CERTIFICATIONS
- Microsoft Certified: Azure Data Engineer Associate (June 2024)
- Turnip Innovations: SQL For Big Data Applications (July 2020)