|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Rahul Rajaram**   |  |  |  |  | | --- | --- | --- | --- | | [rahul\_rajaram@outlook.com](mailto:rahul_rajaram@outlook.com) | | +1 (602) 552-2914 | | | [Rahul2040](https://github.com/Rahul2040) | [rajaram-rahul](https://www.linkedin.com/in/rajaram-rahul/) | | Irving,TX | | | | |
| PROFESSIONAL SUMMARY | | | |
| Data Engineer with 2.5+ years of experience building cloud-native data pipelines on GCP, AWS, and Azure. Master’s in Computer Science (Big Data Systems), currently suppo-rting real-time analytics and pricing workflows at CVS Health. Skilled in PySpark, SQL, Airflow, and BigQuery, with a focus on pipeline optimization, legacy system migration, and scalable data solutions. | | | |
| EDUCATION | | | |
| **Master of Science Computer Science (Big Data Systems)** | | *Aug 2023 – May 2025* | |
| Arizona State University | *Tempe, AZ* | | ***GPA: 3.80/4*** | |
| Relevant Courses: Database Management System Implementation, Data Mining, Data Visualization, Distributed Data Systems, Cloud Computing, Multimedia and Web Data systems, Artificial Intelligence | | |  |
| **Bachelor of Engineering in Electronics and Communication Engineering** | | *July 2017- May 2021* | |
| Anna University | *Chennai, TN, India* | | ***GPA: 8.21/10*** | |
| Relevant Courses: Data Structures and Algorithms, Cryptography and Network Security, Machine Learning Techniques | | |  |
| SKILLS | | | |
| **Languages** | Python, SQL, Bash | | |
| **Tools & Technologies** | Data Modeling, ETL, Data Warehousing, PySpark, Airflow, BigQuery, Snowflake, AWS S3, AWS DynamoDB, Jenkins, Docker, Git, Linux | | |
| **Soft Skills** | Problem-Solving, Communication, Team Collaboration | | |
| PROFESSIONAL EXPERIENCE | | | |
| **CVS health | Irving, TX, USA** | | | |
| ***Data Engineer*** | | *Jun. 2025 - Present* | |
| * Designed BigQuery data models and ETL pipelines for pricing, improving traceability, audit compliance, and analytical consistency across systems. * Automated pricing workflows on GCP using PySpark, Airflow, and Jenkins, ensuring scalable, reliable deployments with clear SLA enforcement. * Collaborated with data scientists and engineers to align models, metrics, and logging to product analytics and strategic pricing goals. * Created reusable Python query generation engine supporting linear programming models, enhancing performance transparency and pricing logic adaptability. * Implemented validation checks and dashboards to track data quality, SLA adherence, and operational reliability across all pricing pipelines. * Led initiatives to enforce access controls and data lineage tracking, improving compliance and auditability of production systems and datasets. | | | |
| ***Analytics Engineering Intern*** | | *May 2024 – Aug 2024 & Jan. 2025 – May 2025* | |
| * Migrated complex data pipelines from Teradata to **BigQuery** and **PySpark**, improving **scalability, performance,** and **cost-efficiency** across systems. * Converted legacy SQL logic to **PySpark** with testing, ensuring exact match between old and new pipeline results under various conditions. * Built and deployed **Airflow DAGs** with **Jenkins integration**, automating secure, reliable cloud workflows using GCP-native tools. * Enabled **cluster parallelization** on DataProc, reducing execution time and costs by optimizing resource utilization and task scheduling. * Partnered with product teams to expose cleaned datasets for **analytics, enabling timely insights** for pricing and operations. * Enhanced metadata, **logging**, and **exception handling** to simplify monitoring and reduce time to diagnose issues in production workflows | | | |
| **LTIMindtree | Chennai, TN, India** | | | |
| ***Senior Engineer IIOT*** | | *Aug. 2021 – June 2023* | |
| * Developed **ETL pipelines** for **1M+ IoT** messages daily, optimizing transformation and schema for faster, cleaner business insights. * Replaced costly **Databricks** workflows with **Snowpipe** and SQL, achieving **90% cost savings** without sacrificing performance or reliability. * Created **React** and **Angular**-based interfaces with **PostgreSQL**, delivering interactive dashboards and tools for operations and analytics teams. * Designed AWS-hosted web app for L&T Construction to track electricity usage and automate meter billing workflows. * Coordinated across backend, frontend, and cloud teams, ensuring timely project delivery and effective cross-functional communication. * Led internal AWS cloud training and mentored team on best practices for scalable architecture and secure data pipelines. | | | |
| PROJECTS | | | |
| ***Generic Drug Pricing*** | | *May 2024 - Present* | |
| * Built Python query generator engine to support LP pricing logic, enabling **modular**, **traceable**, and **auditable** pricing decision-making. * Constructed **BigQuery** data pipelines orchestrated by **Airflow** and **Composer** to manage pricing updates with **high availability** and **compliance**. * Designed scalable workflows on **DataProc** clusters to handle **complex model logic** and **deliver timely** pricing outputs. * Reduced manual interventions by **automating** execution and validations through **Jenkins**, improving deployment **reliability** and regulatory readiness. * Collaborated with product, legal, and analytics teams to ensure pricing logic aligned with business goals and compliance standards. * Drove implementation of monitoring dashboards and failure alerts, enhancing transparency, debuggability, and operational visibility for stakeholders. | | | |
| ***Columnar Database Design*** | | *Jan. 2024 – May 2024* | |
| * Developed a custom column-store database on Java-based Minibase to optimize analytical queries involving **grouping**, **filtering**, and **projections**. * Extended query engine to support columnar operations with **indexing**, boosting performance in **analytical** workloads versus row-store storage. * Designed and tested storage format to improve access speed for large-scale aggregations and minimize read latency. * Implemented support for selection, projection, and group-by, benchmarking against traditional formats to demonstrate analytical advantages. * Strengthened debugging and Java system-level skills while troubleshooting performance bottlenecks in query execution. * Gained deep insights into database internals and architectural tradeoffs involved in analytical database engine design. | | | |
| ***Connected Batteries*** | | *Nov. 2022- Jun. 2023* | |
| * Integrated IoT sensors on trucks to capture telemetry and GPS data for **real-time fleet and battery health monitoring**. * Built data pipelines using **Azure** **IoTHub**, **Eventhub**, and **Snowflake** to process streaming telemetry at scale. * Created stored procedures for data cleaning and transformation, improving consistency and downstream usability of raw data. * Coordinated with product and cloud teams to deliver reliable, end-to-end pipeline for battery monitoring analytics. * Delivered business dashboards with fleet-level analytics, aiding predictive maintenance and reducing vehicle downtime. * Ensured production readiness with exception handling, retry logic, and system health metrics in Snowflake-based workflows. | | | |