

# Ritchie Saul Daniel R

Data Engineer / Analyst,  
Chennai, Tamil Nadu, India



Phone: +91 7639643397 | Email: [ritchiesauldaniel@gmail.com](mailto:ritchiesauldaniel@gmail.com) | LinkedIn: [In/Ritchie](https://www.linkedin.com/in/Ritchie) | Portfolio: [Github](#)

---

## EXECUTIVE SUMMARY

**Data Engineer** with over a year of experience specializing in **ETL processes, cloud-based data solutions, and data architecture optimization** for Manufacturing and B2B2C platforms. Proficient in database management (**MS SQL Server, MongoDB**) and data transformation workflows using tools like **Azure Synapse Pipelines, Synapse Notebooks, Databricks, PySpark, and Spark SQL**. Skilled in automating and managing complex data pipelines with **Apache Airflow** and ensuring data governance using **Immuta**. Adept at extracting actionable insights with **Power BI** to support data-driven decision-making. Strong in Python and SQL programming, **delivering efficient and scalable data solutions to drive strategic decisions**.

## FUNCTIONAL EXPERTISE

- **Language:** SQL, Python
- **Database:** MS SQL server, MongoDB
- **Tools and Platforms:** Azure Synapse, Databricks, Snowflake, Apache Airflow, PowerBI
- **Data Governance :** Immuta
- **Cloud Technologies:** Git Platforms, CircleCI, Terraform
- **Domain Knowledge:** Manufacturing technology, Retail technology

## PROFESSIONAL EXPERIENCE

**Avasoft – Data Engineer – July 2023 – July 2024**

**Retail Technology (CT, US) - Data Engineer:**

### Development:

- Managed extensive **ETL operations** for a **B2B2C platform**, ensuring seamless data connectivity between partners and customers.
- Designed and integrated CI/CD pipelines using **GitHub** for version control, **CircleCI**, and **Terraform** for continuous integration and deployment.
- Processed large volumes of transactional CSV and JSON data from **REST APIs, AWS S3, and MongoDB**, transforming them into structured formats.
- Orchestrated **Python-based transformations** using **Apache Airflow** by creating jobs with appropriate scheduling intervals, converting raw data into valuable insights.
- Utilized **Databricks clusters** for optimized and scalable data processing, maintaining high performance across the platform.
- Maintained a **Medallion Architecture** within **Azure Blob Storage** and a data lake environment for efficient data organization and accessibility.
- Automated data syncing tasks and loaded processed data into **Snowflake**, ensuring timely updates for dashboards and comprehensive reporting.

### Data Governance & Security:

- Used **Immuta** to apply policies on PII data, ensuring masking of customer information for security and regulatory compliance.

### Testing & Deployment:

- Conducted rigorous **integration testing, data validation, and performance testing**.
- Supported user acceptance testing (UAT) phases to **ensure data accuracy and reliability**, actively participating in deployment processes for **smooth transitions and minimal disruptions**.

### Analysis & Reporting:

- Analyzed **dynamic Power BI** reports to analyze customer and partner activities, product performance, and revenue generation.
- Collaborated closely with stakeholders to gather requirements, incorporate feedback, and adjust column names and data structures based on evolving business needs.

### Bug Resolution & Documentation:

- Resolved bug fixes through **JIRA ticket management** and documented insights on the ETL process in **Confluence** to facilitate future improvements and enhance team understanding.

## Manufacturing Technology (CT, US) - Data Engineer:

### Development & Optimization:

- Analyzed and optimized ETL processes utilizing **Azure Synapse Pipelines** to extract, transform, and load data from **MS SQL servers** into the **final data warehouse**, ensuring **efficient data flow and integrity**.
- Transitioned workflows to **Synapse Notebooks** leveraging **PySpark and Spark SQL** for **enhanced data manipulation, scalability, and efficiency** due to the limitations in Synapse Pipelines.
- **Implemented reusable code modules and optimized functions** to reduce redundancy and improve maintainability across all ETL processes, resulting in significant time savings.

### SQL Performance Tuning:

- Optimized complex SQL queries and performed performance tuning using advanced techniques like Common Table Expressions (CTEs), **utilized broadcast joins, and various strategies** to enhance query readability and execution efficiency within SQL Server Management Studio (SSMS).

### Data Pipeline Orchestration:

- Orchestrated ETL workflows within Azure Synapse, focusing on **data lineage and transformation logic** to ensure seamless data flow from source to destination.
- Conducted **rigorous data validation checks** at the final stage to ensure data accuracy and integrity before analysis.

### Analysis & Reporting:

- Engaged in detailed analysis using **Power BI dashboards** to derive actionable insights, leveraging validated data to support strategic decision-making.
- Collaborated closely with cross-functional teams, including data analysts and business stakeholders, to gather requirements and adapt ETL processes to **meet evolving business needs**.

### Monitoring & Optimization:

- Continuously monitored and optimized ETL workflows to enhance performance metrics, such as execution time and resource utilization, ensuring high throughput and efficiency.

## EDUCATION

- **Bachelor of Electrical and Electronics Engineering**  
Coimbatore Institute of Technology, Coimbatore – CGPA 8.75 | 2023
- **Higher Secondary School**  
Bharathi Vidhya Bhavan, Erode – 79% | 2019