Ritchie Saul Daniel R

Data Engineer,
Chennai, Tamil Nadu, India

Phone: +91 7639643397 | **Email:** ritchiesauldaniel@gmail.com | **LinkedIn:** In/Ritchie | **Portfolio:** Github

EXECUTIVE SUMMARY

Experienced Data Engineer with over a year of expertise in **ETL processes and cloud-based data solutions**, focusing on data migration and architecture optimization across Manufacturing and B2B2C client platforms. Proficient in working with databases such **as MS SQL Server and MongoDB**.

Skilled in designing and implementing data transformation workflows using ETL tools such as **Azure Synapse Pipelines, Synapse Notebooks, and Databricks**, with proficiency in utilizing **PySpark and Spark SQL** for complex data transformations to enhance integration and performance.

Experienced in orchestrating data workflows using **Apache Airflow** to automate and manage complex data pipelines efficiently. Adept in data governance with **Immuta** and proficient in analyzing valuable insights using **PowerBI** to support data-driven decision-making. Competent in Python and SQL programming, consistently transforming raw transactional data into actionable insights that drive strategic decisions.

FUNCTIONAL EXPERTISE

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

PROFESSIONAL EXPERIENCE

Avasoft – Data Engineer – July 2023 – July 2024

A products and IT solutions company with 1000-1500 employees

Retail Technology (CT, US) - Data Engineer:

- Managed extensive **ETL operations** for a **B2B2C platform**, ensuring seamless data connectivity between partners and customers.
- Designed and integrated multiple work platforms using **GitHub** for version control, **CircleCI & Terraform for continuous integration and deployment.**
- Processed large volumes of transactional **CSV and JSON data from Rest API's, AWS S3 bucket and MongoDB**, transforming them into structured formats.
- Orchestrated **Python-based transformations using Airflow** by creating jobs with appropriate schedule intervals, transforming raw data into valuable insights.
- Utilized Databricks clusters for optimized and scalable data processing, ensuring high performance across the platform.
- Maintained a **Medallion Architecture within Azure Blob Storage** and a data lake environment, ensuring efficient data organization and accessibility.
- Loaded processed data into Snowflake, maintaining databases with various tables and views for reporting. Implemented automated tasks for regular data syncing, ensuring timely updates for dashboards.
- Used **Immuta** to apply policies on PII data, masking customer information for security and compliance.
- Conducted rigorous integration testing, **data validation, and performance testing**, and supported user acceptance testing (UAT) phases to ensure data accuracy and reliability, while actively participating in deployment processes to ensure smooth transitions and minimal disruptions.

- **Analyzed dynamic Power BI** reports to analyze customer and partner activities, product performance, and revenue generation.
- Collaborated closely with stakeholders to gather requirements and incorporate feedback, ensuring reports met business needs. Further adjusted column names and data structures based on evolving business requirements, maintaining flexibility and relevance.
- Resolved bug fixes through **JIRA ticket management** and documented insights on the ETL process in **Confluence** to enhance understanding and facilitate future improvements.

Manufacturing Technology (CT, US) - Data Engineer:
- 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.
- Optimized complex SQL queries and performed **performance tuning** using advanced techniques such as **Common Table Expressions (CTEs),** enhancing query readability and execution efficiency with SQL Server Management Studio (SSMS).
- Recognizing limitations with synapse pipelines in readability and performance due to complex business requirements involving numerous interconnected tables, transitioned the workflow to **Synapse Notebooks** leveraging **PySpark** and **Spark SQL** for enhanced data manipulation, scalability, and overall efficiency.
- Implemented reusable code modules and optimized functions to reduce redundancy and improve maintainability across all ETL processes, resulting in significant time savings and streamlined development efforts.
- Orchestrated ETL workflows within Azure Synapse, maintaining meticulous attention to 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.
- **Engaged in detailed analysis using Power BI dashboards** to derive actionable insights, leveraging the validated golden data to support strategic decision-making.
- Collaborated closely with cross-functional teams, including data analysts and business stakeholders, to gather requirements and adapt the ETL processes to meet evolving business needs.
- Continuously **monitored and optimized the 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, Tamil Nadu – CGPA 8.75 | 2023

- **Higher Secondary School**

  Bharathi Vidhya Bhavan, Erode - 79% | 2019