

Rahul Ravichandran *Data Engineer*

✉ rahulravi2u@gmail.com

☎ 9566666688

🌐 [linkedin.com/in/rahulravichandra-n-dataengineer](https://www.linkedin.com/in/rahulravichandra-n-dataengineer)



PROFILE

Data Engineer with 3+ years of experience specializing in building scalable and reliable data infrastructure for healthcare, web3, and population health management. Expertise in designing, implementing, and optimizing ETL/ELT pipelines using a modern data stack including Apache Airflow, AWS, Azure, PySpark, and Delta Lake. Focused on delivering high-quality data through rigorous testing and validation, ensuring seamless data integration and accessibility for data-driven decision-making.

PROFESSIONAL EXPERIENCE

Ideas2It Technologies

2021 - Present

HealthEC - Population Health management

- Spearheaded the development and maintenance of scalable data pipelines for a population health management solution using Apache Airflow, AWS (S3, EventBridge, SQS), and PySpark, processing large volumes of healthcare data (claims, clinical, SDoH).
- Converted various file formats to JSON using Omniparser, loading into AWS Delta Lake.
- Optimized PySpark jobs for data extraction and transformation, ensuring FHIR standard compliance and data integrity, leading to a 15% increase in data processing efficiency.
- Implemented data quality checks using Great Expectations and ensured data security with AES encryption, resulting in a 99.9% data accuracy rate.

Technology Stack: PySpark, Python, Airflow, Kafka, Delta Lake, Trino, Great Expectations, PostgreSQL, AWS, SQS, S3 and EventBridge.

Protocol Labs - Web3

- Built data pipelines to extract, transform, and load data from diverse sources (Postgres, MongoDB, Excel, Airtable, Posthog) into a centralized PostgreSQL data warehouse for Protocol Labs, a web3 company.
- Automated data ingestion and processing using AWS Lambda, Step Functions, and EventBridge, enabling real-time data synchronization and analysis.
- Created interactive dashboards and visualizations using Streamlit and Metabase, providing real-time data insights and performance tracking for key stakeholders.

Technology Stack: AWS Lambda, Step Functions, EventBridge, PostgreSQL, Airtable DataPipeline, Posthog DataPipeline, Streamlit, Metabase and Python.

Sandstone Care - Behavioral Health and Substance Use Treatment

- Developed an automated data extraction system using Azure Functions to integrate data from three distinct platforms into a unified Azure-based environment for a behavioral health and substance use treatment provider.
- Built and maintained an Azure Data Factory pipeline to transfer data to staging tables in Azure Synapse, ensuring efficient data flow and transformation.
- Designed and executed stored procedures for transforming and loading data into different schemas within Azure Synapse, ensuring data integrity and consistency.
- Created comprehensive dashboards in Power BI, utilizing DAX and data transformation techniques for in-depth data visualization and reporting.

Technology Stack: Azure Functions, Azure Storage, Azure Data Factory, Azure Synapse, Power BI and DAX.

Ulabsystems - HealthCare

- Designed and implemented an automated pipeline using AWS services to handle data ingestion, transformation and loading.
- Configured S3 buckets and lambda functions to automate the triggering of AWS Glue Crawlers and ETL jobs.
- Developed Glue ETL jobs for data transformation and loading into Redshift.
- Monitored Glue job status with CloudWatch and set up SNS for email notifications, enhancing process transparency.

Technology Stack: Lambda, Glue, Redshift, SNS and Cloudwatch

SKILLS

Python	Pyspark	SQL	Apache Airflow
Delta Lake	Kafka	Data Modelling	Data Governance
Azure Data Factory	Azure Synapse	Azure Function	PowerBI
Docker	Duck DB	AWS Lambda	AWS EventBridge
Streamlit	AWS Step Function	Metabase	SNS
AWS Glue	AWS Redshift	AWS Cloudwatch	

CERTIFICATES

Microsoft Azure Data Engineer Associate exam DP203

EDUCATION

Bachelor Of Electrical Engineering / St. Joseph's College of Engineering

2017 – 2021