

Ashik Roshan I | Data Engineer



ashikroshan261@gmail.com

PROFESSIONAL SUMMARY

Dynamic and results-driven technology professional with expertise in programming, cloud computing, data engineering, and analytics.

- Skilled in designing and implementing scalable data pipelines to optimize ETL/ELT processes, enabling efficient ingestion, transformation, and loading of structured and unstructured data.
- Experienced in building and managing data integration workflows and orchestrating tasks to ensure seamless connectivity across diverse systems and applications.
- Proficient in data modeling, schema design, and managing complex data ecosystems to ensure accessibility, reliability, and optimal performance.
- Adept at leveraging Generative AI (GENAI) and Large Language Models (LLMs) to automate workflows, enhance data-driven decision-making, and deliver innovative real-time solutions.
- Advanced expertise in web scraping and web crawling, focusing on automated data extraction, real-time monitoring, and deep analysis to support critical business operations.
- Passionate about applying artificial intelligence and machine learning to solve real-world problems, drive automation, and foster innovation in complex environments.
- Recognized for strong leadership, exceptional organizational skills, and a commitment to adopting cutting-edge technologies to deliver impactful, data-driven solutions.

EXPERIENCE

| | |
|----------------------------|---------------------|
| BlueCloud | Jun 2023 - Mar 2024 |
| Associate Data Engineer | |
| OptiSol Business Solutions | Mar 2024 - Present |
| Data Engineer | |

EDUCATIONAL QUALIFICATION

- **Bachelor of Engineering (B.E. Computer Science):** K.L.N. College of Engineering, Madurai, Tamil Nadu, India.

TECHNICAL SKILLS

- **Programming Languages:** Python, SQL, JavaScript, HTML, CSS
- **Cloud Platforms & Services:** Azure Blob Storage, Azure Data Lake, Azure SQL Database, Azure Open AI, Azure Databricks, AWS S3, AWS Lambda, AWS Glue, AWS Step Functions
- **Data Engineering & Analytics:** ADF (Azure Data Factory), PySpark, DBT (Data Build Tool), Informatica, Snowflake, Power BI, ThoughtSpot
- **Automation & Web Development:** Selenium, Web Scraping, Web Crawling, FastAPI
- **Development Tools & Version Control:** GitHub, Putty
- **Database Management & Administration:** SSMS (SQL Server Management Studio), pgAdmin, MySQL
- **AI & Machine Learning:** Azure Document Intelligence, GenAI

CERTIFICATIONS

- Microsoft: Azure Data Fundamentals
- Snowflake: Snowpro Core Certification
- Databricks: Databricks Lakehouse Fundamentals
- dbt Labs: DBT Learn Fundamental
- Udemy: 100 Days of Code - Python Bootcamp
- Udemy: Snowflake – The Complete Masterclass (2023 Edition)
- HackerRank: SQL (Basic) Certificate

PROJECTS

JIFFY-HANOVER DATA MIGRATION

Technologies Utilized: Azure Data Lake, Azure Data Factory (ADF), SQL Server Management Studio (SSMS), GitHub, Python

Project Description:

The Jiffy-Hanover Data Migration project aimed at transitioning data from a legacy database to an updated data system, leveraging Azure-based technologies and automation to ensure efficient migration and transformation processes.

Key Phases of the Project:

Legacy System Data Mapping:

- Thoroughly analyzed the legacy database schema to understand its structure and domain-specific requirements.
- Mapped data from the legacy system to the new database schema, ensuring compatibility and alignment with business rules.

Data Migration with Azure Data Factory:

- Migrated data stored in CSV format within Azure Data Lake to the legacy database (SSMS) using Azure Data Factory pipelines.
- Configured ADF pipelines, linked services, and datasets to manage the entire data migration workflow efficiently.

Data Transformation & Processing:

- Leveraged Python scripts to perform advanced data transformation tasks that went beyond ADF's native capabilities.
- After transformation, the processed data was stored in Azure Data Lake before being transferred to the legacy system.

Automation and Version Control:

- Used Azure Data Factory Data Flows for most of the transformations to ensure automated, scalable data processing.
- Implemented version control for pipeline configurations, scripts, and resources through GitHub to maintain proper change management and facilitate collaboration among the development team.

Testing & Data Validation:

- Conducted unit testing of the migration process using Python to validate the accuracy and integrity of the data before it was handed over for formal QA testing.
 - Ensured that data transformations were accurate, complete, and met business requirements.
-

REPUBLIC SERVICES (RS) DATA MIGRATION

Technologies Used: Informatica, SQL Server Management Studio (SSMS), Snowflake, DBT, AWS Step Functions, Putty, GitHub, Oracle

Project Overview:

The Republic Services Data Migration project is focused on migrating data from the US Ecology systems to Republic Services (RS), integrating data from multiple sources including Azure SQL Server, Oracle, and MS SQL into a Snowflake staging schema. Informatica was employed as the ETL tool to load the data into Snowflake, and DBT was utilized for data transformation and movement from the staging schema to the Operational Data Store (ODS) schema. This end-to-end migration was carried out across four environments: DEV, STG, QA, and PROD, ensuring a reliable, scalable, and automated data pipeline for ongoing data processing.

Key Contributions and Steps:

ETL Pipeline Development and Maintenance:

- Designed, developed, and maintained ETL pipelines in Informatica to handle both initial full loads and incremental data loads, ensuring the smooth migration of large volumes of data into Snowflake.
- Configured daily pipeline runs to automate data loading, monitoring for potential issues, and implementing resolutions to ensure continuous operation.

Data Transformation Using DBT:

- Utilized DBT Core integrated with Visual Studio Code for efficient data transformations within Snowflake.
- Automated the transition of data from the staging schema to the ODS schema, orchestrated by AWS Step Functions, ensuring optimal data flow across environments.

Quality Assurance and Data Integrity:

- Developed custom DBT macros for data validation, ensuring that data transformations were accurate, met business requirements, and maintained data consistency throughout the migration process.
- Conducted extensive data quality checks across DEV, STG, QA, and PROD environments to ensure that the data met integrity standards before moving to production.

Testing and Automation:

- Performed both manual and automated unit testing using Python scripts to validate the correctness of data loads and transformations.
- Worked closely with the testing team to address and resolve any discrepancies, ensuring that the data migration process adhered to strict business rules.

Version Control and Deployment:

- Utilized GitHub for version control, ensuring that all code, configurations, and scripts were properly maintained and easily accessible for collaborative development efforts.
-

CYCLECOMMERCE INSIGHT

Technologies Used: Python, SQL Server, Snowflake, Azure Data Factory (ADF), Azure Storage Service, Power BI

Project Overview:

The CycleCommerce Performance Analysis project was initiated to enhance business insights and inform decision-making by analyzing transactional data from the online store. The project encompassed a comprehensive approach, involving data migration, modeling, transformation, visualization, and the generation of actionable insights to drive business growth.

Key Contributions and Steps:

Data Extraction and Transformation:

- Developed Python scripts to extract and transform raw transactional data from various sources associated with the online store.
- Loaded the transformed data into SQL Server, preparing it for advanced data processing and analysis.

Data Orchestration with Azure Data Factory (ADF):

- Utilized Azure Data Factory to orchestrate the seamless movement of data from SQL Server to Azure Blob Storage, ensuring smooth and efficient data transfer between systems.
- Implemented data synchronization strategies, enabling real-time data updates and ensuring consistency across platforms.

Data Loading and Staging in Snowflake:

- Managed the transfer of data from Azure Blob Storage into Snowflake using external staging, maintaining strict data integrity and security protocols throughout the process.
- Ensured that the data was readily available and optimized for use within the Snowflake environment for advanced analytics.

Data Modeling and Optimization in Snowflake:

- Created and implemented data models within Snowflake, structuring the e-commerce transaction data to improve accessibility and performance for analytical tasks.
- Optimized data schemas to enhance query performance and support reporting and business intelligence processes.

Visualization and Insights with Power BI:

- Designed and developed Power BI dashboards to visualize key performance indicators (KPIs) from the transaction data.
 - Provided actionable insights to the business team, supporting data-driven decision-making and identifying opportunities for optimization within the online store.
-