Arif Rahman

⦿ Fairfield, IA (willing to relocate) ✉ rahman.mdarif32@gmail.com 📞 (641) 233-9937 in Linkedin ⭘ GitHub

Summary

Results-driven Data Engineer with extensive experience designing, building, and optimizing scalable data pipelines and solutions within the healthcare sector. Skilled in Azure, Snowflake, SQL, and Python, with a strong foundation in data structures, algorithms, and cloud platforms. Adept at monitoring data products, troubleshooting anomalies, and collaborating with cross-functional teams to deliver actionable analytics and maintain robust data processes. Proven ability to support internal and external stakeholders, develop clear documentation, and consistently deliver high-quality, reliable data solutions in fast-paced environments.

Skills

Core Competencies:	Data Engineering, Data Pipelines, Data Structures & Algorithms, ETL/ELT, Data Warehousing, Data Quality, Data Monitoring, Data Process Optimization, Documentation, CI/CD, Problem Solving, Communication
Cloud Platforms:	Azure (Data Factory, Synapse, Logic Apps, DevOps, Functions), AWS (Redshift, S3, Glue, Lambda), GCP
Databases:	Snowflake, MS SQL, PostgreSQL, Redshift, Teradata
Languages:	SQL, Python, PySpark, T-SQL, PL/SQL
Big Data/Tools:	Databricks, Spark, Airflow, Dataiku, DBT, Linux, Docker, Kubernetes
Data Modeling:	Kimball, Star/Snowflake Schema, Data Vault
Healthcare Data:	IMS, IQVIA(MIDAS), HCP, Patient, Payer, Provider Data
Data Visualization:	Power BI, Tableau, Qlik Sense, Excel
DevOps/Version Control:	GitHub, Bitbucket, GitLab
Additional Skills:	Statistical Inference, Training & Mentoring, Stakeholder Collaboration, Customer Service Orientation

Experience

Lead Data Analyst *Remote, United States*
IQVIA *Oct 2021 – Oct 2023*

- Developed and optimized 20+ ETL pipelines using Azure Data Factory, Databricks, and PySpark, ingesting data from multiple sources, including streaming data via Kafka, and building a scalable data lake solution with ADLS. Enhanced query performance and improved operational efficiency by 20% while processing ˜1 TB data per day.
- Orchestrated the development of 14 omni-channel analytical products for pharma client to design ETL using Dataiku, DBT, SQL, Python, and Bitbucket for CI/CD processes. Effectively handled data lakes on AWS S3 and Redshift to support large-scale data operations.
- Developed multiple data models in Power BI for analyzing omni-channel campaign impact, customer engagement, feature store, and CRM Insights. Designed interactive dashboards with advanced DAX to track key performance metrics.
- Developed and implemented data monitoring systems using SQL and Python, creating custom metrics in Dataiku that significantly improved data quality and reduced issues by 20% and prevented unwanted pipeline failures.
- Led the design and development of a SaaS platform for pharma clients, automating data pipelines and integrating APIs with Python, Java, SQL, and Snowflake, enhancing sales team performance by 15%.
- Migrated on-premises data warehouse to Snowflake and streamlined data ingestion processes by utilizing SnowPipe with Azure Blob Storage, enhancing scalability and ensuring efficient, automated pipeline operations.

  Technologies Used: Azure, Snowflake, Redshift, Dataiku, Databricks, Kafka, SQL, Python, PySpark, Power BI.

Senior Data Analyst *Remote, Norway*
Cefalo *Feb 2020 – Sep 2021*

- Developed a data mash and automated social media analytics dashboards in Power BI for an oil shipping

company, by integrating data in AWS Glue from Google Analytics, Facebook, Instagram, and LinkedIn. This initiative resulted in an estimated savings of ~$600K per year in branding budgets.

○ Developed Azure data pipeline using Databricks, Snowflake and Synapse Analytics for reporting solutions, reducing latency and improving decision-making.

○ Engineered Azure-based data pipelines using Ariflow for reports like Subscription KPI, Sales Retention, and Revenue tracking, resulting in a 20% improvement in customer retention strategies.

○ Automated Qlik Sense reporting for Revenue, Subscription, and Corporate Overview analysis, increasing financial planning accuracy by 25%.

○ Collaborated on scalable reporting solutions in Qlik Sense for customer acquisition and churn analysis using Azure Synapse, improving customer behavior insights.

Technologies Used: Azure, AWS, Databricks, Snowflake, SSIS, Airflow, Qlik Sense, Power BI, SQL, Python.

Senior Data Analyst *Remote, United States*
IQVIA *Feb 2015 – Jan 2020*

○ Migrated an on-premises ETL pipeline to Azure Data Factory for a pharma client, integrating data from the Orchestrated Customer Engagement (OCE) in the Salesforce system using SQL, reducing data processing time from 9 hours to 1.5 hours and saving the client ~$155K annually by enhancing performance.

○ Designed tabular data model in SSIS, SSAS with RLS for Power BI CRM reporting, integrating data from Salesforce CRM API and orchestrating a data pipeline in MS SQL Server, supporting 7 different comprehensive report development.

○ Created a comprehensive sales dashboard with 10 years history in QlikView with 3-tier architecture and Teradata for multiple Europe region, improving sales performance tracking by 20%.

○ Automated 80% of Pricing Insight processes using SSIS, SSRS, and SQL, improving data quality and generating monthly reports, while managing data production across 40 European countries for 10+ clients.

○ Delivered SSAS cubes weekly and monthly, processing data in MS SQL Server using SQL, Python, and R to calculate pharmaceutical prices and sales.

Technologies Used: Python, SQL, R, Power BI, QlikView, Qlik Sense, SSAS, SSIS, SSRS, SSMS, Azure Data Factory and blob storage.

Data Analyst *Dhaka, Bangladesh*
Bangladesh Centre for Advanced Studies *Nov 2013 – Jan 2015*

○ Played a pivotal role in identifying research topics and crafting proposals that secured two major projects, driving research initiatives forward and delivering impactful scientific insights for stakeholders.

○ Leveraged data analysis tools like R, SPSS, Excel, and MS Access to process and analyze complex datasets, providing actionable insights that shaped project strategies and facilitated successful donor reporting.

Publications

Fitting Time to First Birth Using Extended Cox Regression Model in Presence of Non-proportional Hazard Jan 2014

Md. Rashedul Hoque, *Md. Arif Rahman*

Dhaka Univ. J. Sci. 63(1): 25-30, 2015 (January)

Conferences

Extended Cox Regression Model in Case of Violation of the Proportionality Assumption: Heaviside Function Approach Nov 2014

Md. Arif Rahman, Md. Rashedul Hoque

ISI Regional Statistics Conference 2014, Kuala Lumpur, Malaysia, 16-19 November 2014

Projects

Iowa Liquor Sales on Azure *github* ↗

○ Developed an end-to-end data pipeline integrating Iowa Gov API and historical data using Azure Functions. Implemented Databricks' Medallion Architecture for data transformation and created Power BI dashboards for actionable sales insights. Enhanced decision-making and operational efficiency. Tools: Azure, Databricks,

Power BI, SQL, Python.

E-Commerce Website *github* ↗

- Built a secure e-commerce platform using Spring Boot, supporting admin, seller, and buyer roles. Implemented role-based access, email notifications, and form validation, enhancing security and user satisfaction. Facilitated seamless e-commerce operations for a scalable online shopping experience. Tools: Spring Boot, Java, PostgreSQL.

Banking System *github* ↗

- Engineered a robust banking application supporting multiple account types and credit card operations (withdrawals, deposits, interest calculations). Implemented design patterns (strategy, command, factory) to improve scalability and maintainability, ensuring high system performance. Tools: Java.

News Portal *github* ↗

- Designed and developed a fully functional news portal with a front-end news page, detailed article view, and admin panel for content approval and analytics. Improved user experience and streamlined content management. Utilized Material UI, Node.js, Next.js, and PostgreSQL.

Education

Maharishi International University	*Dec 2023 – Nov 2024*
Master of Science in Computer Science	*GPA: 3.81/4.00*
University of Dhaka	*Jan 2013 – Dec 2013*
Master of Science in Applied Statistics	*GPA: 3.34/4.00*
University of Dhaka	*Jan 2009 – Dec 2012*
Bachelor of Science in Applied Statistics	*GPA: 3.17/4.00*

github ↗