Shaharukha Pathan

*Azure Data Engineer*

Highly skilled and experienced Azure Data Engineer with a solid background in designing, developing, and implementing data solutions on the Microsoft Azure platform. Possessing four years of professional expertise, I have a strong understanding of data engineering concepts, cloud technologies, and data management best practices.

**Email:** [shrkhpthn@gmail.com](mailto:shrkhpthn@gmail.com)

**Mobile: +91 8668759518**

**PROFESSIONAL SUMMARY**

* **ACA Global:** Senior data engineer Nov-2024 – Till date
* **ACA Global:** Data engineer Jun-2023 – Nov 2024
* **Datametica:** Associate Engineer III Apr-2023 – Jun 2023
* **Datametica:** Associate Engineer I Dec-2021 – Mar 2023
* **Datametica:** Associate Analyst Jul-2021 – Dec 2021

PROGRAMMING LANGUAGE: PYTHON, SQL

CLOUD: Azure data factory, Azure data bricks, GCP, Airflow, DAGS.

Company: ACA Global  
Duration**:** June 2023 – Present

**Job Description:**

Currently working as a **Senior Data Engineer** at **ACA Global**, where I contribute to **Market Abuse Surveillance (MAS)**. My role involves onboarding clients and building end-to-end **data pipelines** using **Azure Data Factory** to support surveillance and regulatory compliance efforts.

**Key Responsibilities:**

* Design and implement scalable **ETL pipelines** in **Azure Data Factory** to ingest, transform, and store data from various sources.
* Onboard new clients by setting up their data integration workflows and ensuring data quality and completeness.
* Create and manage **datasets** and **linked services** within ADF to support dynamic and secure data flows.
* Ingest data from multiple sources such as **REST APIs, SFTP servers, SQL Databases, CSV files, and JSON files**.
* Utilize **stored procedures** to load and manage data in SQL-based systems and Azure data tables.
* Store and organize structured and semi-structured data in **Azure Storage** for downstream processing and compliance reporting.
* Develop **transformation scripts in Python** to preprocess and enrich data for generating **Indications of Interest (IOIs)** and other analytical outputs.

**Skills & Technologies:**

* **Cloud Platforms:** Microsoft Azure
* **Data Tools:** Azure Data Factory, Azure Storage, SQL Server
* **Scripting:** Python (data transformation and automation)
* **Data Sources:** REST APIs, SFTP, SQL Databases, CSV, JSON
* **Database Processing:** Stored Procedures, ETL/ELT Pipelines
* **Domain Expertise:** Market Abuse Surveillance (MAS), Regulatory Data Processing

Company**:** Datametica  
**Duration:** Jul 2021 – Jun 2023

**Job Description:**

At Datametica, I worked as a **Data Engineer**, primarily focused on **cloud migration projects**, helping clients transition their on-premises data infrastructure to **Google Cloud Platform (GCP)**.

**Key Responsibilities:**

* Led the end-to-end **data migration process** from on-premises systems to GCP, ensuring minimal disruption and optimized performance.
* Designed and developed **robust data pipelines** using tools like **Apache Airflow**, **Big Query**, and **Cloud Storage**.
* Built scalable and maintainable **DAGs in Apache Airflow** to automate data ingestion from multiple sources such as **SFTP, FTP, Oracle, and Netezza**.
* Implemented **incremental and batch pipelines** to efficiently handle both historical and real-time data loads.
* Ensured data integrity and consistency throughout the migration process and optimized queries for performance in Big Query.
* Collaborated closely with clients to understand their data landscape and provide tailored solutions for seamless cloud integration.

**Skills & Technologies:**

* **Cloud Platforms:** Google Cloud Platform (GCP)
* **Data Tools:** Big Query, Apache Airflow and **Cloud Storage**.
* **Databases & Sources:** Oracle, Netezza, Big Query
* **ETL/ELT Pipelines:** Batch processing, Incremental loads, Historical data migration
* **Scripting & Automation:** Python (for DAGs and data transformation logic)

EDUCATION

MSC (Computer Science) (2018-2020)

BSC (Computer Science) (2015-2018)

**Certifications**: Google Cloud Certified Associate Cloud Engineer, **Databricks for Data Engineering, Python for Data Science**

REWARDS AND RECOGNITION

* WALL OF FAME I got Recognition for my outstanding performance in ACA

October 2024

* SPOT AWARD In Recognition of an outstanding performance Success at Datametica

March 2022

DECLARATION

I hereby declare that all the statements above are true to the best of my knowledge and belief.