

Sahil Shingne

📞 9075767854 — ✉️ sahil45shingne@gmail.com — 🌐 <https://www.linkedin.com/in/sahil-shingne-2185a5185/> — 🏠 <https://github.com/Sahil883>

Summary — Cloud Data Engineer with expertise in Python, SQL, Azure and Generative AI dedicated to designing and implementing scalable data solutions that enhance business insights and decision-making. Skilled in leveraging cloud technologies to drive innovation, efficiency, and data-driven outcomes.

Skills

- **Programming:** SQL (Oracle, MySQL), Pyspark, Python(Data structures and Algorithm), Hadoop, GITLAB.
- **Cloud and Framework:** Azure (ADF, Synapse Analytics, ADLS), AWS(S3), Databricks, Atlan, ETLs, Data Mesh.
- **Artificial Intelligence:** Generative AI, Advanced RAG, Machine Learning, Neural Networks and NLP.

Certifications: DP-203 Azure Data Engineer Associate, Databricks Data Engineer Associate, DP-100 Azure Data Scientist Associate, Az-900 Azure Fundamentals, AI-900 Azure AI Fundamentals.

Experience

Analyst

Jan 2023 – Present

Deloitte USI

Data Migration Engineer

- **Cloud Data Migration:** Collaborated to develop a cloud data migration project, transferring extensive datasets from on-premises to **Azure** infrastructure.
- **ETL Pipeline Development:** Engineered over 10 ETL pipelines in **Azure Data Factory** to migrate 50+ tables, each with 1M+ records, storing outputs in **ADLS Gen2**.
- **ETL Process Optimization:** Built robust ETL processes by implementing SCD Type 2, automated failure notifications, and scheduled triggers.
- **Performance Improvement:** Optimized data flows by converting complex data processes to Synapse Analytics notebooks using **PySpark**, reducing execution times from 9 hours to 30 minutes in SIT (i.e cutting execution time by **90%**).

Generative AI Engineer

- **AI-Driven Framework Development:** Built an AI-driven framework for intelligent business glossary and anomaly detection, leveraging **Claude AI 3.5** and **Collibra** integration.
- **Machine Learning & Unsupervised Learning:** Used clustering algorithms and Stella 400M v5 model for generating high-quality embeddings and analyzing data correlations.
- **Optimization:** Improved job processing efficiency by 15% through **FastAPI**, also achieved **concurrency** with AsyncIO reducing execution time by **60%**.
- **Cloud Data Management:** Managed and stored processed data in **AWS S3**, streamlined deployment workflows with Jenkins.

Data Mesh Engineer

- **Data Mesh Architecture:** Designed and implemented Data Mesh Architecture using Atlan and Python SDK for seamless data management.
- **ETL Pipelines & Data Integration:** Built and managed ETL pipelines to ingest data(20+ tables) from Azure Databricks and GCP into Atlan.
- **Metadata Automation:** Automated metadata updates in Atlan based on user interactions, improving the metadata management process reducing manual efforts by **40%**.
- **Streamlined product onboarding** by contributing foundational code for JIRA and Git integrations in Datamesh.

Projects

Automated Fire detection and Sprinkler System

- **Technologies:** Computer Vision, Arduino
- Developed an automated fire detection and response system, using computer vision to detect fire and activating a sprinkler system with Arduino also sends real-time email alerts upon detection.

Blood Sample Component Analysis

- **Technologies:** Convolutional Neural Network (CNN), Transfer Learning
- Built a CNN model(78% accuracy) to classify white blood cell components (eosinophils, neutrophils, monocytes) in blood samples, enhancing accuracy in medical diagnostics.

Education

Walchand College of Engineering

Aug 2018 – July 2022

Bachelor of Technology in Electrical Engineering

CGPA - 8.19