

SANKET VISHWAKARMA

sameer.vishk@gmail.com | +91-7980956276 | LinkedIn: sanket-vishwakarma

SUMMARY

Data Engineer with 3+ years of experience building scalable data pipelines and AI-powered solutions. Expertise in AWS, Snowflake, Python, PySpark, and modern data orchestration tools. Proven track record of reducing manual effort by 65-75% through automation and improving data quality using LLM-based solutions. Strong background in statistical analysis, machine learning, and cross-functional collaboration.

TECHNICAL SKILLS

Languages & Frameworks: Python, PySpark, SQL, Terraform, DBT, Airflow, Dagster

Cloud & Data Platforms: AWS (Glue, Lambda, Step Functions, Elastic Beanstalk, Appflow), Snowflake, Snowflake Streamlit

Data Engineering: ETL/ELT Pipelines, Data Architecture, Data Quality & Governance, Delta Lake, Batch & Stream Processing

AI/ML: Large Language Models (LLMs), GenAI, Machine Learning, Statistical Analysis

DevOps & Tools: CI/CD Pipelines, Git, Power BI, API Integration, Pentaho, Selenium

Databases: RDS, D365 ERP Systems, OData API

PROFESSIONAL EXPERIENCE

Data Engineer

September 2023 – Present

Kellanova (formerly Kellogg's), Global Data Analytics Team

Mumbai, India

- Architected and deployed a **resource monitoring dashboard** using Python, AWS Step Functions, Lambda, and APIs, reducing manual effort by **70%** and providing real-time insights into ingestion and orchestration jobs across the organization
- Built a **custom Snowflake Streamlit marketplace** for Data Quality & Governance with DBT integration, real-time dashboards, and approval workflows; implemented **LLM-powered SQL generation** from natural language, improving data quality by **65%**
- Designed and implemented **DevOps CI/CD pipelines** with auto-merge, advanced branching strategies, and rule sets, improving deployment reliability and reducing production risks
- Led POC initiatives for Snowflake, Airflow, DBT, Dagster, and AWS Appflow to optimize data architecture; collaborated with business users and regional teams to deliver AI/GenAI-powered solutions
- Created Power BI reports delivering actionable insights for business stakeholders to monitor job performance and failures

Data Engineer

April 2022 – September 2023

Edelweiss Tokio Life Insurance

Mumbai, India

- Developed complex data pipelines for agent commission processing with delta loads, API integrations, and multi-source data handling using AWS Glue and PySpark
- Led POC and production deployment of C-Data to replace decommissioned ingestor tool for D365 ERP integration, achieving **seamless migration with minimal downtime** and zero impact on downstream jobs
- **Saved \$2,000/month** by automating critical Glue jobs using sync database for ERP systems; identified and removed orphan tables and non-relevant jobs for cost optimization
- Improved pipeline performance by **40%** through Spark optimization techniques and delta load implementations
- Supported data science team with model building and customer experience predictions; managed customer data provisioning via OData API using Python

IT Programmer

Cognizant

October 2021 – March 2022

Kolkata, India

- Performed statistical analysis on e-commerce user data for Macy’s shopping website; conducted EDA on multiple data sources to extract actionable insights
- Automated web testing and debugging using Selenium and XPath, reducing manual testing time by **50%**

EDUCATION

Bachelor of Science in Mathematics and Statistics

Jadavpur University

January 2015 – November 2020

Academic Excellence Award (Class XII)

CERTIFICATIONS

Data Analysis with Python | Machine Learning (Coursera) | Google Data Analytics Professional | Apache Spark Big Data Engineering | Hands-On Essentials: Data Application Builders Workshop