

Saurabh Patil

Mumbai, Maharashtra | saurabhpatil.mca@gmail.com | +918652070500 | linkedin.com/in/saurabh-patil
github.com/Saurabhpatil-1997

Summary

Results-oriented Data Engineer with 4+ years of experience building scalable, cloud-native data platforms on AWS and Snowflake. Specialized in designing distributed ETL/ELT pipelines using PySpark and Apache Airflow to process large-scale datasets efficiently. Demonstrated success in reducing processing latency by 30% and eliminating hundreds of manual hours through workflow automation. Proficient in dbt-based transformations, data modeling, and end-to-end system migrations that deliver reliable, analytics-ready data for Business Intelligence reporting and decision-making.

Key Skills

- Programming Languages: **Python, SQL**
- Data Engineering: **ETL/ELT, dbt, Data Pipelines, Data Modeling, Data Quality, Data Validation, Data Storage and Lakehouse**
- Big Data: **PySpark**
- Workflow Orchestration: **Apache Airflow, AWS Step Functions**
- Cloud and Data Warehousing: **Snowflake, AWS (S3, Glue, Lambda, Redshift)**
- Analytics and BI: **Power BI**
- Version Control: **Git, GitHub**

Experience

Software Engineer, Inspiroz IT Services Pvt. Ltd. | Navi Mumbai, Maharashtra Feb 2024 – Present

- Architected distributed ETL frameworks using PySpark and AWS Glue, reducing end-to-end processing latency by 30% for high-volume datasets.
- Established a cloud-native data platform on AWS (S3, Redshift), leveraging Lambda and Step Functions for event-driven ingestion and orchestration, significantly improving data reliability and system availability.
- Streamlined reporting workflows by developing Python and SQL automation solutions, eliminating over 10 hours of manual effort per week and accelerating delivery to business stakeholders.
- Led end-to-end migration of SQL Server (RDS) workloads to Snowflake, building scalable dbt transformation models and maintaining code versioning with Git.
- Awarded “**Highly Recommended by Client**” for consistent delivery of high-quality data solutions.

Data Processing Analyst, Ipsos Pvt. Ltd. | Thane, Maharashtra Jun 2022 – Oct 2023

- Modernized recurring tracker studies (monthly, quarterly, yearly) by implementing Python-based systematized transformations, reducing delivery timelines by 15% and minimizing manual intervention.
- Collaborated with Data Scientists to map complex tabulation requirements into scalable schema for Business Intelligence reporting and Machine Learning modeling.
- Designed Power BI dashboards for weekly/monthly reporting, enabling trend analysis and improved operational visibility.
- Streamlined ETL pipelines using Python and SQL to ingest, clean, and transform large survey datasets from multiple sources, enhancing data reliability and reducing turnaround time for analytics.

Data Analyst, MSDA LLP | Mumbai, Maharashtra Jul 2021 – May 2022

- Implemented scalable web data extraction solutions using Selenium and Requests to collect and structure unstructured data, increasing data availability by 40% for downstream analysis.
- Refactored and optimized complex SQL queries, improving data retrieval performance and reducing dashboard refresh times by 25%.
- Developed interactive Power BI dashboards to monitor key performance indicators, enabling data-driven decision-making for senior leadership.

Projects

1. Inspiroz Insights - Multi-Institution Data Platform

- Delivered end-to-end ETL pipelines for three educational institutions, consolidating operational data into centralized SQL databases for unified reporting.
- Enabled event-driven data ingestion using Azure Functions, ensuring near real-time synchronization across systems.
- Maintained high data integrity and availability, supporting timely decision-making through reliable analytics datasets.

2. KIPP Baltimore - Cloud Data Warehouse Modernization

- Migrated analytical workloads from SQL Server (RDS) to Snowflake, improving scalability, query performance, and cost efficiency.
- Architected an event-driven ingestion framework using AWS Lambda, Step Functions, and EventBridge to enable resilient and scalable pipeline orchestration.
- Developed modular transformation models in dbt with automated testing and documentation, delivering maintainable and analytics-ready datasets.
- Implemented secure and cost-efficient cloud architecture using IAM roles, VPC segmentation, bastion host access, and operational monitoring tools to enhance reliability, governance, and resource utilization.

3. DPPS - Scalable Data Processing Platform

- Engineered a distributed ETL system using PySpark, Python, and Apache Airflow to process large volumes of structured and semi-structured data from APIs and databases.
- Orchestrated end-to-end workflows via Airflow DAGs covering ingestion, transformation, validation, and warehouse loading.
- Introduced data quality controls and pipeline monitoring to meet reliability targets and service-level expectations.

4. Zoho Automation - HR Data Integration Pipeline

- Built a Python-based ingestion pipeline to extract employee records from Zoho REST APIs, replacing manual data exports.
- Transformed raw JSON data into payroll-ready datasets using Pandas, ensuring consistency and usability for downstream processes.
- Delivered accurate monthly datasets to HR teams, eliminating manual effort and saving approximately 5 hours per month.

Education

University of Mumbai, Master of Computer Applications (MCA)

Sep 2018 – Jul 2021

- NCRD's Sterling Institute of Management Studies

University of Mumbai, Bachelor of Science in Computer Science (B.Sc.)

Sep 2015 – Jul 2018

- The SIA College of Higher Education