Ninaad Sawant

Data Engineer | Python, Pyspark, Scala, SQL | ETL, ELT, Airflow | AWS, Azure
Worcester, MA | 774-420-0444 | ninaadsawant23@gmail.com | Linkedin | Portfolio | GitHub

SUMMARY

Data Engineer with 2.5 years of experience building scalable data pipelines, APIs and cloud data platforms. Skilled in Python, PySpark, AWS, Azure with strong focus on performance optimization and data quality. Proven ability to enable data-driven insights.

EDUCATION

Master of Science in Computer Science (STEM) January 2024 - May 2025
Clark University, Worcester, MA

Bachelor of Engineering in Electronics August 2016 - November 2020
Vidyalankar Institute of Technology, Mumbai, India

TECHNICAL SKILLS

- **Languages/ Frameworks:** Python, Java, Scala, SQL, Spark, PySpark, Hive, dbt
- **Big Data Engineering:** ETL/ELT, Kafka, HDFS, Apache Airflow, Hadoop, Snowflake
- **Machine Learning/ Analytics:** Regression, Classification, Clustering, NLP, TensorFlow, PowerBI, Tableau, Excel
- **Cloud Computing:** AWS (S3, Redshift, RDS, Glue, IAM, QuickSight), Azure (Data Factory, Databricks), GCP (BigQuery)
- **Databases:** MySQL, PostgreSQL, MongoDB, Elasticsearch
- **Additional Tools:** GitHub, Jenkins, Postman, Unix, Control-M, Jira
- **Certifications:** AWS Certified Cloud Practitioner, AWS Certified Data Engineer - Associate

PROFESSIONAL EXPERIENCE

Infosys Limited, Pune, India May 2021 - November 2023
Data Engineer

- Improved data throughput by 33% through optimization of over 25 ETL pipelines using PySpark and Scala, each processing more than 10 million records daily across HDFS, Amazon S3, and Elasticsearch.
- Reduced data latency by 23% by building and maintaining real-time ingestion pipelines with Apache Kafka, enabling near-instant event processing and supporting time-sensitive data processing.
- Refined data reporting by 40% by writing and optimizing approximately 150 advanced SQL queries using joins, window functions, and CTEs, accelerating analytics and business decision-making.
- Optimized pipeline troubleshooting by 35% and improved data pipeline up-time, by leveraging advanced debugging techniques in HiveQL and Spark SQL to quickly identify root causes across distributed datasets.
- Served over 50,000 daily data requests by developing secure, high-performance REST APIs using Python and Java Spring Boot.
- Improved API response times by 83% through iterative code-level optimization, validating API responses in Postman and finalizing performance statistics based on client feedback.
- Enhanced CI/CD pipelines using Jenkins and Git, optimizing workflows, managing version control to improve release reliability.
- Automated over 25 ETL pipelines in Apache Airflow as per client SLAs, eliminating manual ETL runs by 99%, improving reliability through access validations, source data freshness checks, and automated failure alerts to relevant teams.
- Collaborated with cross-functional teams and stakeholders in weekly Agile meetings to define data pipelines, data modeling and API functionality, enhancing SLA compliance by 90%.
- Monitored over 25 critical ETL pipelines in production, performing root cause analysis on failures and coordinating with relevant teams for resolution, ensuring timely support and reducing data issues by 97%.
- Maintained data quality and security with regular code reviews focused on test coverage, coding standards, and compliance with security best practices.
- Documented workflows and project progress using Confluence to ensure transparency with stakeholders and engineering teams.

TECHNICAL PROJECTS

Tableau Sales Analytics Dashboard *(Tech: Tableau, Excel)* June 2025

- Enabled data-driven insights through a Tableau dashboard visualizing over 10,000 transactions with sales/profit KPIs and trends.

Snowflake Revenue Analytics Pipeline *(Tech: Python, SQL, SnowFlake, Data modeling)* May 2025

- Delivered 20 business KPIs by building a Snowflake data warehouse using a star schema model to track revenue and performance.

Azure Synapse Medallion Pipeline *(Tech: PySpark, SQL, Azure Data Factory, Databricks, Synapse, PowerBI)* April 2025

- Improved Power BI reporting by 25% through developing a scalable optimized data pipeline, processing 75,000 API records.