

AMOL GULAB NAVATE

Aurangabad, Maharashtra, India / amolnavate11@gmail.com / +91-9021231130
LinkedIn: linkedin.com/in/amol-navate

PROFESSIONAL SUMMARY

Data Engineer with 1+ years of experience in building scalable data pipelines using PySpark, AWS (Glue, EMR, S3, Redshift), Apache Kafka, and Airflow. Expertise in Medallion Architecture, Delta Lake, Apache Hudi, and data quality frameworks. Proven track record in optimizing Spark jobs, implementing CDC pipelines, and delivering production-grade ETL solutions.

TECHNICAL SKILLS

Big Data Technologies:	PySpark, Spark SQL, Databricks, Delta Lake, Apache Hudi, Hadoop
Cloud Platforms:	AWS (Glue, EMR, S3, Redshift, Athena, DynamoDB, IAM, CloudWatch)
Streaming & Messaging:	Apache Kafka, Real-time Data Processing
Orchestration Tools:	Apache Airflow, Apache NiFi
Programming Languages:	Python, SQL, Linux Shell Scripting
Databases:	MySQL, DynamoDB, Redshift, PostgreSQL
Version Control & CI/CD:	Git, GitHub, GitHub Actions
Data Modeling:	Star Schema, Snowflake Schema, SCD Type 1 & 2, Medallion Architecture
Data Quality:	Data Validation, Schema Evolution, Referential Integrity, Anomaly Detection

PROFESSIONAL EXPERIENCE

Data Engineer	2024 – Present
PST IT Solutions, Maharashtra, India	
E-Commerce Data Pipeline Project	
<ul style="list-style-type: none">Designed and implemented end-to-end Medallion Architecture (Bronze-Silver-Gold) data pipeline using PySpark on AWS EMR, processing large-scale e-commerce datasets with automated data quality checksBuilt ingestion pipelines from multiple sources (RDBMS via AWS Glue JDBC, FTP via Apache NiFi, Kafka streams) into Amazon S3, implementing CDC logic and incremental data processing with schema validationDeveloped comprehensive Data Quality Rules framework including completeness validation, uniqueness checks, referential integrity constraints, and anomaly detection to ensure analytics-ready datasetsOptimized Spark jobs using partitioning strategies, bucketing, broadcast joins, and caching techniques, improving processing performance and reducing compute costs while maintaining SLA complianceImplemented Apache Hudi for ACID-compliant upsert operations on S3 data lake, enabling efficient incremental processing and reducing downstream data merge workloadOrchestrated complex ETL workflows using Apache Airflow DAGs with dependency management, retry logic, SLA monitoring, and automated alerting, achieving 99% pipeline reliabilityCreated aggregated summary tables and derived business KPIs to support analytics and reporting requirementsEvaluated and optimized resource allocation for Spark and Glue jobs to balance performance and cost efficiency	

EDUCATION

Bachelor's Degree in Computer Science	Expected: August 2025
Kohinoor College, Aurangabad, Maharashtra / CGPA: 6.29/10	

ACHIEVEMENTS

- Awarded "Rising Star Award" by PST IT Solutions for exceptional performance, timely project delivery, ownership, and strong teamwork
- Successfully reduced data processing costs through implementation of Spark optimization techniques and efficient resource utilization