NAVEEN DHARAVATH

928-221-8870 | naveendharavath10032000@gmail.com | linkedin.com/in/naveen-dharavathde

PROFESSIONAL SUMMARY

Results-oriented Data Engineer with 3+ years of real-time experience in designing, building, and optimizing data pipelines, ETL workflows, and cloud-based data infrastructure across banking and healthcare domains. Skilled in Python, SQL, PySpark, Airflow, AWS, and Snowflake, with proven expertise in data modeling, pipeline orchestration, and real-time analytics enablement. Adept at collaborating with cross-functional teams to ensure data quality, reliability, and scalability in large-scale environments.

TECHNICAL SKILLS

- **Programming**: Python, SQL, Java, Scala, Shell Scripting
- **Data Engineering**: PySpark, Airflow, Pandas, ETL/ELT, Data Warehousing, Databricks
- **Databases**: MySQL, PostgreSQL, Snowflake, MongoDB, Redshift
- **Cloud Platforms**: AWS (S3, EC2, Glue, Lambda, Redshift), Azure Data Factory
- **Big Data Tools**: Spark, Kafka, Hadoop, Hive
- **Data Modeling**: Star/Snowflake Schema, Dimensional Modeling, ER Diagrams
- **DevOps/Automation**: Docker, Jenkins, Git, CI/CD, Terraform
- **Visualization**: Power BI, Tableau
- **Operating Systems**: Linux, Windows, MacOS

PROFESSIONAL EXPERIENCE

Fifth Third Bank **Oct 2024 - Present**
Data Engineer *Kentwood, MI*
Project: Enterprise Data Platform for Credit Risk & Fraud Analytics

- Developed and automated ETL pipelines using PySpark and Apache Airflow for real-time ingestion and transformation of financial transactions and risk data from multiple systems.
- Designed and implemented data models in Snowflake to support downstream analytical and BI workloads.
- Built AWS-based data lake architecture integrating S3, Glue, and Redshift for scalable storage and querying of structured and semi-structured data.
- Created data validation and quality frameworks to monitor data integrity and latency across ingestion layers.
- Implemented workflow orchestration, dependency management, and DAG scheduling using Airflow, resulting in more reliable and maintainable ETL operations
- Optimized SQL queries and Spark jobs, improving data pipeline performance by 35%.
- Integrated data lineage and cataloging through AWS Glue Data Catalog and maintained schema evolution.
- Collaborated with analytics teams to deliver curated data marts for fraud detection dashboards in Power BI.
- Automated pipeline monitoring and alerts through CloudWatch and custom logging utilities.
- Ensured compliance with data governance and security policies (GDPR, Basel III) across data environments.

Cyient **Jun 2021 - Dec 2023**
Data Engineer *Hyderabad, India*
Project: Clinical Data Integration and Patient Risk Insights Platform

- Designed and deployed data ingestion and transformation pipelines integrating EHR, clinical trial, and patient data using Airflow, Python, and SQL, which enhanced data accessibility and supported timely insights
- Developed ETL frameworks for structured/unstructured data and automated schema mapping for healthcare datasets, improving processing efficiency and data consistency
- Implemented data quality checks, deduplication, and validation logic to ensure accuracy and regulatory compliance (HIPAA).
- Migrated legacy ETL jobs to AWS Glue and PySpark, improving runtime efficiency by streamlining processes and reducing operational costs
- Managed data warehouse design in Snowflake supporting advanced analytics and reporting.
- Automated data partitioning, metadata tracking, and workflow scheduling using Airflow DAGs.
- Built audit tables and change-data-capture logic to track data flow and lineage.
- Collaborated with BI teams to develop interactive Power BI dashboards for patient analytics.
- Implemented CI/CD integration using Jenkins for pipeline versioning and deployment automation.
- Improved overall data processing performance by 40% and reduced manual ETL overhead.

EDUCATION

Northern Arizona University, Flagstaff, USA **May 2025**
Master of Science, Computer Science

CMR Institute of Technology, Hyderabad, India **2023**
Bachelor of Technology, Computer Science and Engineering