Anoop Musale
anoopmusale27@gmail.com | (607) 761-0926

CAREER OBJECTIVE

Proven and hard-working data engineer with 5+ years of experience. Proficient in Python, Databricks, SQL and Informatica PowerCenter. Track record in building robust software solutions and data engineering systems, with significant achievements in optimizing SQL queries by 30%, increasing informatica and python code performance by 40%, developing efficient data pipelines and leading cross-functional teams to deliver innovative data solutions. Adept at driving data strategy by aligning technical insights with business goals.

SKIILS & TOOLS

Data Engineering Tools: Python, PySpark, Databricks

Cloud Technologies: Amazon Web Service (AWS), AWS S3, AWS RDS, Delta Lake

ETL Tools: Informatica PowerCenter, Informatica Intelligent Cloud Services (IICS)

Databases & Query Optimization: Oracle SQL, MySQL, Advanced SQL query tuning

DevOps & CI/CD: Liquibase, Jenkins, GitHub, Concourse, Harness, Rancher, Kubernetes, Docker

Others: Jira, ServiceNow, Pandas, NumPy, API design

WORK EXPERIENCE

Senior Data Engineer   April 2022 – Current
Nationwide Insurance (Contract)

- Led a team of 5 data engineers and 2 data analysts in designing and implementing a universal data modeling framework using Python, Databricks, PySpark, and AWS, consolidating nationwide business data into a centralized AWS RDS Oracle database to streamline access for underwriting, business, and agent teams.
- Optimized SQL queries by 30%, improving batch performance and ensuring faster, more efficient data processing for critical business operations.
- Onboarded and implemented UAC Stonebranch, configuring workflows, triggers, and scheduling jobs for daily, monthly, and yearly runs, while creating real-time dashboards to monitor and report job statuses.
- Developed a fully automated database change management pipeline using Liquibase and GitHub, supporting the company's transition toward continuous integration and deployment (CI/CD) for streamlined, error-free database updates.
- Designed and deployed data ingestion pipelines using Python, Databricks, AWS S3, and Delta Lake, enabling real-time ETL processing for risk assessment and portfolio insights for underwriting decisions.
- Created automated reporting jobs delivering daily, monthly, quarterly, and yearly insights into claims, product performance, and regional trends, enabling leadership to make informed, data-driven strategic decisions.
- Demonstrated strong people management and collaboration skills by leading cross-functional teams, aligning diverse stakeholders, and fostering a collaborative environment that emphasizes knowledge sharing and teamwork.
- Maintained a strong client focus by understanding stakeholder needs, translating business objectives into technical solutions, and ensuring timely delivery of data insights that supported strategic decision-making at all organizational levels.

Data Engineer   August 2019 – March 2022
Nationwide Insurance (Contract)

- Led the successful upgrade of Informatica PowerCenter and Guidewire Policy Center versions while conducting Proof of Concept (POC) for migrating Guidewire Policy Center, Informatica PowerCenter, and Oracle databases to Amazon Web Services (AWS) cloud.
- Improved the performance of Informatica PowerCenter (ETL) jobs and SQL queries, achieving a 40% reduction in batch runtime by implementing process optimization techniques and efficient database strategies.
- Designed and implemented a comprehensive system for data preprocessing and cleaning, ensuring improved data quality while enforcing access control on sensitive customer information to maintain compliance.
- Developed unit and system testing frameworks in Python and TOSCA DI, enhancing code validation processes and achieving a 30% improvement in overall code quality.
- Engineered a robust continuous integration and continuous deployment (CI/CD) pipeline leveraging tools such as GitHub, Jenkins, Concourse, Docker, and Kubernetes to streamline and automate code deployment workflows.
- Performed advanced SQL query optimization by implementing index strategies, leveraging explain plans, collecting DBSTATS, and refining join logic to enhance database performance.
- Handled over 100+ change requests using Jira and ServiceNow, actively participating in requirement gathering, request analysis, development changes, and end-to-end testing, thereby contributing to continuous improvement and adaptability to evolving business needs.
- Collaborated with development and QA teams to resolve 250+ code and technical issues within a 5–6-week span. This proactive approach unblocked critical production issues and prevented the accumulation of technical debt.
- Played a pivotal role in designing scalable, efficient cloud migration strategies with AWS to ensure streamlined operations for enterprise-level data processing.
- Utilized advanced performance tuning techniques to identify bottlenecks and implemented solutions that drastically reduced batch processing runtimes, ensuring faster data delivery.
- Automated deployment workflows using cutting-edge CI/CD tools and containerization strategies, improving developer productivity and system reliability while ensuring a seamless, error-free deployment pipeline.

EDUCATION

State University of New York, Binghamton
Master of Science (M.S.), Computer Science   Aug 2017 - Jun 2019

Mumbai University
Bachelor of Engineering (B.E.), Information Technology   Aug 2013 - Jun 2017