

Data Engineer with over 3 years of experience in designing and building reliable data pipelines that transform raw data into meaningful business insights. Proficient in leveraging a diverse set of tools and technologies, including Apache Airflow, PySpark, Databricks, Python, SQL, and cloud platforms such as AWS (EC2, S3, MWSAA). Dedicated to continuous learning and staying current with emerging technologies to drive both professional and personal growth. A strong advocate for teamwork and collaboration, committed to delivering efficient and scalable data engineering solutions.

SKILLS

Languages - Python, SQL, Scala, Shell Scripting, Pyspark

Cloud - AWS, EC2, S3, EKS, MWSAA (Airflow), Databricks, Snowflake

Tools & Technologies - Apache Spark, Teradata, Data Analysis, Docker, Git, Agile, CI/CD, Kubernetes, Data Warehouse, Vault, Informatica, Spark SQL, Trino

WORK EXPERIENCE

DEVELOPMENT ENGINEER 2 | COMCAST | Apr'2024 - Present

- Migrated Databricks pipelines to an on-premises Kubernetes environment as part of a cost-saving initiative, achieving a 40% reduction in infrastructure costs while maintaining performance and scalability.
- Built and led the Oracle-to-Snowflake Migration Project using the python-batch-runner library, managing enterprise-level migration activities and reducing operational support effort by 50%.
- Designed and implemented comprehensive Data Quality (DQ) checks using Informatica, Shell Script, Python, and Scala Spark for on-premises critical pipelines and common frameworks, ensuring 100% data integrity validation across 400+ datasets and improving reliability for end-users.
- Ingested and peer-reviewed 300+ data pipelines to enable seamless data transfer from Snowflake to the Data Lake (S3) as part of an enterprise-wide Data Consolidation initiative.
- Deployed and configured an MWSAA (Managed Workflows for Apache Airflow) environment in AWS, and built 100+ pipelines to streamline migration from an in-house orchestration system, boosting reliability and enhancing operational efficiency by 30%.
- Developed subject matter expertise in Spark, Streaming, and Pyrunner, mastering key functional areas essential to my role.

DEVELOPMENT ENGINEER 1 | COMCAST | Jun'2022 - Mar'2024

- Delivered data model changes across 100+ datasets/tables in Databricks and Spark, including column additions and data type updates, enabling 50+ business use cases and ensuring alignment with evolving data requirements.
- Optimized Databricks clusters used in data processing pipelines, enhancing performance while maintaining run-time efficiency and productivity, resulting in a 10% cost reduction.
- Created a reusable Databricks Shard Migration Utility (Python + Databricks API) to automate environment migrations, enabling seamless scaling across multiple teams and improving migration turnaround time by 80%.
- Collaborated with cross-functional teams on POCs using Airflow and EMR, improving the stability and reliability of ETL pipelines.
- Trained and mentored new team members on both technical and functional aspects, accelerating their onboarding process and enabling them to contribute to deliverables within weeks.

INTERN | COMCAST | Jan'2022 - Jun'2022

- Presented a proof of concept (POC) for an ELK (Elasticsearch, Logstash, Kibana)-based analytics solution that unified usage logs from Trino and Teradata, enabling data-driven performance optimization and achieving a 10% reduction in customer query runtime.
- Strengthened foundational understanding of SDLC and data warehousing concepts through continuous learning initiatives, enhancing overall design and implementation quality in data projects.

EDUCATION

SASTRA University
B.Tech. Computer Science & Engineering

2018 – 2022
8.79 CGPA

ACHIEVEMENTS

- High-Five Champ Award for outstanding performance and excellence in achieving project milestones in Q4 2022, Q2 2023, Q1 2024, and Q4 2024.
- Pinnacle Award for consistently demonstrating exceptional skills, and contributing towards organizational goals in Q1 2023.
- Spotlight Award for being a reliable, accountable, and consistent top performer in the team in Q1 2025.