Atharva Deshmukh

ad2546@rit.edu | 5859106910 | Rochester, NY | LinkedIn | Portfolio

Summary

Data Engineer with 2+ years of experience in developing, deploying, and managing large-scale data solutions. Proven ability to create efficient and reliable data pipelines. Expertise in data engineering, ETL, and data warehousing.

Technical Skills

Programing Languages: Python, SQL, Bash, Java, JavaScript, C, Git, NoSQL, Scala
Cloud & Infra: AWS (Redshift, Glue, Lambda, S3, IAM), GCP, Azure, Terraform, Docker, Kubernetes, CloudWatch, Schema Versioning.
Data Engineering: Apache Airflow, dbt, Spark, Kafka, ETL/ELT, Data Lakes, Data Marts, Hadoop, Databricks, Snowflake.
Machine Learning & AI: GPT-4, NumPy, SciPy, Pandas, Scikit-learn, Keras, XGBoost, TensorFlow, PyTorch, LangChain, RAG, FAISS.
Visualization & Tools: Power BI, Tableau, Looker, Excel, Matplotlib, Seaborn, Pivot Tables.

Education

Rochester Institute of Technology **May 2025**
Master of Science in Information Technology and Analytics Rochester, NY

University of Mumbai **May 2021**
Bachelor of Engineering in Computer Engineering Mumbai, India

Professional Experience

Data Engineer Consultant **Dark Horse Digital Solutions**
Mumbai, Maharashtra, India **Aug 2022 - Jul 2023**

- Spearheaded the migration of an on-premise Cordy's BPM based CRM system to AWS by re-integrating external services and deploying Docker containers, and monitoring system on CloudWatch resulting in reduction of operational overhead by 25%.
- Designed and maintained a centralized data warehouse in Oracle SQL focused on advanced data modeling and SQL query optimization, enhancing data accessibility for analytics by 30%.
- Collaborated with cross-functional teams to automate data extraction and reporting through complex SQL queries and containerized stored procedures, extracting over 100,000 records daily to support data-driven decision-making.

Data Engineer Consultant Jr **Dark Horse Digital Solutions**
Mumbai, Maharashtra, India **Jun 2021 – Jul 2022**

- Led cross-functional teams to resolve database performance bottlenecks in an IBM BAW (ERP) environment for a health insurance platform by utilizing Docker containers and optimizing SQL queries, ensuring 80% of claims processed in under 10 seconds.
- Orchestrated scalable ETL pipelines using Apache Airflow which integrated into IBM BAW, processing over 10,000 insurance claims daily, and improving claims processing times by 20% within the quarter.
- Developed a scalable data warehouse to centralize insurance data by designing robust ETL pipelines and containerized deployments, improving reporting efficiency and analytics by 30%.

Software Development Intern **VESIT Renaissance**
Mumbai, Maharashtra, India **May 2020 – Jul 2020**

- Defined a performance tracking application for new college students by collecting stakeholder requirements and academic data, leveraging Python and PostgreSQL to build an interactive backend and dashboard that improved academic feedback by 30%.
- Engineered an automated ETL pipeline post data cleansing, validation and preparation using Apache Spark to consolidate diverse student performance data, processing over 10,000 records daily and boosting data accuracy by 20%.
- Orchestrated containerized deployments with Docker to standardize development environments, reducing configuration time by 50% and accelerating feature rollouts.

Projects

Formula 1 Driver Performance Prediction (*AWS Arora, R, Machine Learning)* **Jan 2025 – May 2025**

- Built a scalable data ingestion pipeline to analyze driver telemetry by processing 1M+ records from Fast F1 API into a PostgreSQL data warehouse on AWS Aurora, enabling predictive modeling and live performance dashboards.
- Implemented robust data cleaning and preprocessing techniques including normalization, outlier detection, and feature engineering on 1 million data points structured datasets to ensure data integrity for advanced predictive modeling.

Stars Yapp LLM based astrology *(LLM, Azure DevOps, RAG, Flask)* **May 2020 – Jul 2020**

- Developed and deployed an end-to-end Recruitment Alignment Guide (RAG) application on Microsoft Azure, by integrating OpenAI's GPT-4 with Vedic astrology content, enhancing job-matching processes.
- Leveraged Facebook AI Similarity Search (FAISS) to build a semantic search feature for astrological data, decreasing query response time to under 0.5 seconds and enhancing user experience.