

TEJDEEP MADDIRALA

Potsdam, NY | +1 (315)-742-0527 | tej.d50@gmail.com | www.linkedin.com/in/tejdeep-tezz | www.github.com/Tezz1999

Data Engineer with 2+ years of experience in designing, developing, and optimizing data pipelines and big data solutions. Proficient in deploying scalable systems, predictive modeling, and delivering actionable insights. Adept at adapting to fast changing technologies and collaborating with cross-functional teams to drive data-driven decision-making.

EXPERIENCE

DATA ENGINEER (INTERN) | CLARKSON UNIVERSITY | MAY 2023 - MAR 2024

- Spearheaded data management for St. Lawrence Public Transit right from data acquisition, profiling and storage.
- Designed, developed, and automated workflows for GTFS data pipelines to create a platform for analytics, ultimately optimizing various factors including schedules, routes, fuel efficiency, and coverage.
- Implemented comprehensive data quality checks with immediate notification alerts to ensure high data integrity and reliability throughout the pipeline.
- Worked on designing Snowflake warehousing schemas to store historical data, focusing on efficient and performant analytics querying.
- Developed visualizations using Tableau to showcase key metrics and insights for county heads, enabling data-driven decision-making and strategic planning.

ASSISTANT PROFESSOR |BIG DATA AND CLOUD ARCHITECTURE | AUG 2022 – MAY 2024

- Initiated and developed an entire course for students on building data pipelines using AWS and industry-leading technologies, tackling problems in real-life data migration, ETL, analysis, and machine learning workflows in collaboration with a professor.
- Lectured students weekly and assigned projects on discussed technologies. Conducted lab sessions to address their doubts and ensure hands-on learning.
- Mentored nearly 30 major projects till now, based on development and monitoring of efficient batch and real-time data streaming pipelines for both analytics and machine learning workflows.

EDUCATION

MASTER OF SCIENCE IN DATA SCIENCE (3.68/4) |AUG 2022 - MAY 2024 | CLARKSON UNIVERSITY, POTSDAM, NY

TECHNICAL SKILLS

- **Core:** Python, R, SQL, Java, C, Airflow, AWS, Apache Kafka, Apache Hadoop, Git, Linux
- **Database Management:** MySQL, PostgreSQL, NoSQL, Cassandra, Snowflake, Redshift
- **Tools & Technologies:** Informatica, Hadoop, Spark, Hive, Docker, Kubernetes, Tableau, Terraform, .
- **Certifications:** IBM data Engineering, GCP Data engineering, Understanding and Visualizing Data With Python(University of Michigan).
- **Frameworks & Packages:** Scikit-learn, Pandas, Numpy, Pyspark, Boto3, Seaborn, NLTK, TensorFlow, PyTorch, Keras

PROJECTS

- Built a real-time data streaming pipeline, migrating API data to Kafka and Spark, then to Cassandra, orchestrated by Apache Airflow and containerized in Docker.
- Created a change data capture (CDC) for real-time data streaming to capture changes made in database.
- Employed feed forward ANN to analyze mammogram image data, to detect breast cancer attaining dice coefficient of 0.84, sensitivity of 0.82 and specificity of 0.9.