

Chaitanya Ghadling

McKinney, TX | +1(862) 622-9631 | chaitanyaghadling@gmail.com | [Github](#) | [LinkedIn](#) |

PORTFOLIO

EDUCATION

Master of Science in Data Science

May 2023

New Jersey Institute of Technology • Newark, NJ • 3.9 GPA

Newark, New Jersey, USA

Bachelor of Engineering in Computer Engineering

Jun 2021

Savitribai Phule University • Pune, MH • 8.0 GPA

Pune, India

TECHNICAL SKILLS

- **Programming Languages:** Python, R, Java, C++, C#, Javascript
- **Databases:** MySQL, PostgreSQL, MongoDB, T-SQL
- **Data Science Stack:** Apache Spark, Apache Kafka, Apache Hadoop, Informatica, Snowflake, DataBricks, Tableau, PowerBI, Docker, Machine Learning, Deep Learning, Statistical Analysis.
- **Cloud Tools:** AWS: S3, EC2, RDS, Redshift, Glue, Lambda, EMR, Azure, GCP.

WORK EXPERIENCE

Data Engineer • *Frontier Communications*

July 2023 - Nov 2024

- Designed, implemented, and maintained an ETL pipeline integrating **NetSuite** ERP data with **Apache Kafka, Airflow, Spark, Docker, and AWS** services for seamless data flow and processing.
- Configured **Kafka** topics and consumers for real-time ingestion of transactional data from **NetSuite**, reducing data latency by **40%** and ensuring robust and low-latency data pipelines.
- Leveraged **AWS Glue** for cataloging NetSuite metadata transforming ERP data into analytics-ready formats.
- Optimized data models for NetSuite-generated structured data, stored in **AWS RDS** and **Amazon Redshift** to support advanced reporting and BI use cases.
- Incorporated both unstructured data (logs, documents) and structured data into the pipeline by leveraging **AWS S3** as a storage layer and processing it with **Apache Spark**, reducing processing time by **50%**.
- Implemented monitoring and logging solutions using **AWS CloudWatch** and **AWS CloudTrail** to track pipeline performance metrics, identify bottlenecks, resulting in a **25% improvement in pipeline efficiency**.
- Containerized the entire ETL pipeline environment using **Docker**, facilitating seamless deployment and scalability across different environments.
- Collaborated closely with cross-functional teams including data analysts, developers, and stakeholders to gather requirements, implement solutions, and ensure alignment with business objectives.

Research Assistant • *New Jersey Institute of Technology*

Feb 2023 - July 2024

- Developed a novel image rendering module for a 3D video conferencing tool, utilizing various Deep Learning tools like **Pytorch** and **Tensorflow** for efficient and scalable model training in a distributed environment.
- Implemented advanced Deep Learning techniques like **GANs, U-Net, Encoder/Decoder** framework to achieve high-quality results in the inpainting task.
- Created a custom dataset and dataLoader from scratch tailored to model requirements for better results.
- Collaborated closely with PhD students to research state of art algorithms and seamlessly integrate novel methodologies into existing software architecture, ensuring smooth functionality and enhanced performance.

Data Engineer • *GCB Services*

May 2022 - Dec 2022

- Automated daily tasks of the company by developing multiple standalone applications, deployed on **AWS S3**, using **Selenium** and **Xlsxwriter**, reducing routine workload by 70%.
- Pioneered the company's transition of data storage and analysis from **Excel/VBA** to **SQL** by designing a centralized database from scratch and deploying it on **AWS RDS** Server.
- Extracted valuable insights by performing extensive data analysis from diverse sources, for performance

tracking and enabling more informed client billing decisions, resulting in 23% increase in company revenue.

- Enhanced project readability by thoroughly documenting existing code, providing clear explanations, and improving overall code maintainability.
- Streamlined the visualization of critical KPIs by developing a web application using the **Streamlit** framework, deployed on **Heroku** facilitating informed decision-making. [\[Preview\]](#)

Data Analyst Intern • PYSIST Pvt. Ltd.

April 2020 - Sep 2020

- Established ETL workflows using Informatica to extract, transform, and load data from various sources into the data warehouse, resulting in a 30% reduction in data processing time.
- Devised ad-hoc analysis using **SQL**, **Python** and executed **A/B tests** resulting in improved business decision making.
- Exhibited proficiency in executing complex **SQL** queries and effectively utilizing existing tables to create views based on specific requirements.
- Designed custom visualizations in **PowerBI** to display complex data in an interactive manner.

PROJECTS

Background Geolocation Tracking

Oct – Nov 2023

- Created a deployment ready cross platform (Android, IOS) mobile application using **Flutter** that tracks location of 30+ users periodically and stores it in a database hosted on **AWS RDS** server.
- Hosted a backend system using **node.js** that handles application interaction with databases.
- Developed multiple **APIs** to transfer information between backend and front-end systems.

Exploring Data Linkages

Oct – Dec 2022

- Worked in a team of 5 to develop a project aimed at exploring data linkages and hidden biases in datasets.
- Established linkages between 4 NYC open datasets, totaling over 5.5M rows, based on common attributes such as age, gender, and race using **Python**.
- Conducted data pre-processing and initial exploratory data analysis (EDA) using **pandas** and **matplotlib**.
- Generated multiple visualizations and interactive dashboards on **Tableau** drawing insights from the data.

Twitter Sentiment Analysis

Feb – March 2022

- Engineered a robust pipeline for sentiment analysis, performing binary classification on a dataset of 10,000 tweets to accurately determine the nature of each tweet.
- Leveraged **Spark** streaming and Tweepy to dynamically retrieve tweets and utilized the Py-Spark MLlib Library for their classification, achieving an impressive accuracy rate of approximately 85%.
- Employed Ensemble methods such as bagging and weighted **Random Forest** Algorithm for effective classification, and **Natural Language Processing** techniques for text pre-processing and feature extraction.

Library Monitoring System

Jan – March 2021

- Collaborated in designing the frontend using **JavaFX**, creating a user friendly interface for library staff.
- Developed an entire database management system capable of efficiently handling the data of over 10,000+ college students.
- Executed a diverse range of complex **SQL** queries, including those involving intricate date-time formats, to extract crucial information and optimize database operations.
- Played a pivotal role in the project lifecycle, actively contributing to the design, development, and successful deployment of the system in the college library.

CERTIFICATIONS AND TRAINING

- | | |
|--|---|
| • Machine Learning A-Z: Hands-on Python and R in Data Science | Udemy Aug 2021 |
| • Neural Networks and Deep Learning | deeplearning.ai, Coursera Nov 2020 |
| • Python for Everybody (P4E) 5 course specialization | University of Michigan, Coursera Aug 2020 |
| • Microsoft Technology Associate for Intro to Programming using JavaScript | Microsoft Jan 2020 |
| • AWS Certified Solutions Architect Associate certificate training | SOFTECH Data Securities Oct 2019 |