Graduate Teaching Assistant:

● Facilitate weekly labs to 30-40 undergraduate students  
● Notified students about project due date using Google Mail Merge  
● Enhanced the labs - Linux, Cloud Computing, and Security to reflect the concepts taught in the lecture  
● Other responsibilities included: holding weekly office hours, grading assignments & projects, exam development, exam proctoring, etc.,

Data Science Intern:

Team - Audience Analytics

• Created automated workflow using Databricks and Apache Spark to feed historical data into an XGBoost ML model, resulting in improved accuracy of future audience prediction

• Designed and implemented a Python module for entity resolution using NLP pre-processing techniques, TF-IDF word vectorization, and K-Nearest Neighbor ML model, resulting in more accurate matching of Viacom's program inventory with Gracenote data, and IMDb with TMDB

• Built a data pipeline to ingest 700 million metadata records from IMDb website into AWS S3 bucket, which underwent various data transformations and validations before being loaded into Databricks Delta Tables

• Generated forecast report by aggregating data from several Delta tables using SQL and PySpark, allowing for comparison of predictions made by various models using MAPE and RMSE error metrics

• Converted SQL-based ETL pipelines source code to PySpark, enabling automation and improving efficiency of data-processing

**Skills:** Pandas (Software) · NumPy · Data Cleaning · Data Pipelines · Machine Learning Algorithms · Natural Language Processing (NLP) · Amazon S3 · Databricks · Apache Spark · SQL · Data Analysis · Python (Programming Language)

Senior Software Engineer

Team - IAE Data Reconciliation

• Developed multi-threaded data quality tool in Java to validate holding and transaction data veracity and extrapolate missing financial details based on patterns observed with other users

• Implemented POC for determining suitable database for the project among OLTP (Oracle DB), Mongo DB, and Apache Kudu based on query performance

• Designed Mongo DB Collections, documents with suitable indexing, and configured Spring repositories for application querying

• Implemented Netflix-Ribbon load balancer to achieve fault tolerance and distribution of application traffic

• Created REST API using JAX-RS for sharing reconciliation results with other teams

**Skills:** Core Java · Oracle SQL Developer · Apache Kudu · SQL · MySQL · Java · MongoDB · Spring Boot · PL/SQL · Apache Kafka

Software Engineer  
Team - IAE

• Built web crawlers using Java Selenium APIs to aggregate users' financial data and personal information from banking websites

• Created PDF parser for retrieving data from bank statements using Python PDFMiner text extraction tool

• Implemented regex validator using Java regular expressions for validating PII

• Reviewed the code submitted by peers and mentored couple of new hires

**Skills:** Core Java · PDF Miner · Oracle SQL Developer · SQL · MySQL · PL/SQL · Selenium WebDriver

Intern

I was part of a web-based license application development.

● Worked on Spring MVC framework while building the license application

● Developed web pages using JSP, JQuery, and Ajax

● Performed integration testing of license applications