Shreya Jaiswal

Boston, MA | (617)-238-4538 | shreya.jais1604@gmail.com | [LinkedIn](#) | [Github](#)

EDUCATION

Northeastern University, Boston, MA   **Sep 2022 - Dec 2024**
Master of Science in Information Systems; GPA 3.8/4
Relevant coursework: Program Structures & Algorithms, Data Management and Database Design, Designing Data Architecture for Business Intelligence, Advances in Data Science & Architecture, Network Structures & Cloud Computing

TECHNICAL SKILLS

Programming:	Python, Java, JavaScript, Shell, R, Bash, Scala, PL/SQL, MATLAB
Frameworks:	Hadoop, Spark, Kafka, Scikit-Learn, TensorFlow, PyTorch, LangChain
Tools:	Databricks, Airflow, Talend, Informatica, Alteryx, Tableau, PowerBI, Terraform, Docker, Kubernetes
Database:	Oracle, SQL Server, PostgreSQL, MySQL, Snowflake, Teradata, BigQuery, Redis, MongoDB
Cloud:	Google Cloud Platform, Amazon Web Services (AWS), Microsoft Azure
ML Libraries:	Pandas, Numpy, Scikit-Learn, Plotly, Seaborn, Keras, Matplotlib, TensorFlow, EconML

WORK EXPERIENCE

Northeastern University, Boston, MA   **Jan 2024 - Apr 2024**
Teaching Assistant
- Facilitated learning by reviewing and delivering **10+** coursework modules and exploring technologies in Generative Artificial Intelligence, including Retrieval Augmented Generation (RAG), Large Language Models (LLM), and Vector Databases like Pinecone

Elevance Health, Boston, MA   **Jun 2023 - Dec 2023**
Data Analyst Intern
- Leveraged **Power BI**, **Tableau**, and **SQL** for data analytics to analyze electronic enrollment data from **Oracle** database reducing manual analysis by **60%** and facilitating informed business objectives through effective cross-functional team collaboration
- Designed and executed over **15+ data visualization** solutions using Power BI and Microsoft Excel, examining existing data models to drive a **50%** increase in management review efficiency for daily, weekly, monthly, and yearly reports
- Utilized Oracle tables and Power BI to identify the top **10** vendors with error rates in file inconsistencies of up to **70%**
- executed advanced **SQL** queries to access and analyze data from the relational database for **data modelin**g and reporting

Accenture, Bangalore, India   **Jan 2021 - Jul 2022**
Data Engineer
- Developed distributed **PySpark pipeline** on **Azure Databricks**, enabling data ingestion of historical and incremental **10,000+** XMLs into the data lake for enhanced accessibility and scalability in advanced analytics and business intelligence
- Accelerated batch processing using Apache **Spark Structured Streaming** by over **2 hours** through strategic analysis and efficient job run planning for daily incremental batches of **500 XML** files, resulting in a **25%** boost in processing efficiency
- Orchestrated the design, development, and production support of **ETL data pipelines** handling over **10** terabytes of data using Python, **SparkSQL**, S3, and **Snowflake**, resulting in data-driven decisions through **Big Data** technologies (**Apache Spark, Hadoop**)
- Reduced manual effort by **80%**, improving operational efficiency, by architecting, scripting and automating data integration solutions (Medallion Architecture) using Databricks **PySpark**, **Scala** and **AWS** cloud services (EMR, EC2, S3, DynamoDB SNS/SQS)
- Championed the development of intricate **SQL** queries for data analysis and data extraction, resulting in the identification of critical Claims trends and improved policy retention rates by **10%**
- Collaborated in an **agile** environment with key business stakeholders, gathering requirements and steering innovative ideation sessions through iterative development to drive strategic outcomes

PROJECTS

[Data Warehouse and Analytics Solution for NYC Vehicle Collisions] *(ER Studio, Talend, SQL Server, Power BI)*
- Designed high-performance data model and ETL workflow using Talend, migrating 7+ million records from flat files and BigQuery into SQL Server, while creating 20+ KPI dashboards in PowerBI and Tableau, leveraging data migration and data integration

[NYC Citibike Data Pipeline] *(GCP, BigQuery, Google Looker Studio, DBT, Terraform, Prefect)*
- Built an end-to-end batch data engineering pipeline for NYC's Citibike data, including ETL processes, data extraction, data transformation with DBT, data orchestration with Prefect, and storage in GCP

[PDF Chatbot Application Developer] *(Python, Streamlit, Pinecone, OpenAI, Generative AI)*
- Developed a data science chatbot using Python, Streamlit, Pinecone Vector DB, OpenAI embeddings, and Retrieval-Augmented Generation (RAG) for natural language processing to interact with YouTube links and PDFs

[GCP Cloud Deployment and Automation Project] *(Fast API, Packer, Terraform, GCP, Postgres)*
- Engineered a backend Python application leveraging SQLAlchemy and FastAPI, created custom images with Packer on GCP, automated CI/CD pipelines using GitHub Actions and containerization, reducing deployment time by 40%