

XINGE FENG (MOMO)

momofeng95@gmail.com | +1 404-955-2104 | linkedin.com/in/xinge-feng

EDUCATION

Washington University in St. Louis (GPA: 4.0/4.0)

Aug 2022 - May 2024

Master of Science, Engineering in Data Analytics & Statistics

Related Coursework: Machine Learning, Cloud Computing, Artificial Intelligence, Data Warehouse, Scalable Analysis (PySpark, Hadoop)

University of North Carolina at Chapel Hill (GPA: 3.7/4.0)

Aug 2018 - Dec 2021

Bachelor of Science, Computer Science & Mathematics

Related Coursework: Data Warehouse, Database management, Data Mining, Big Data, Data Modeling, Data Structure, Airflow

TECHNICAL SKILLS

- **Languages:** SQL, Python (Pandas, Statsmodels, NumPy, SciPy, Matplotlib), Scala, Java, JavaScript, R, HTML/CSS
- **Database:** PostgreSQL, Oracle, SQL Server, DynamoDB, MongoDB
- **Tools:** AWS (Redshift, S3, Kafka), Azure, GCP, Snowflake, Databricks, Spark, Hadoop, Airflow, Power BI, Tableau, Rest APIs, CI/CD
- **Related Certificates:** AWS Certified Cloud Practitioner, Tableau Desktop Specialist

EXPERIENCE

Amazon

May 2023 - Aug 2023

Data Engineer Intern

Nashville, Tennessee, United States

- Transferred structured data from Andes **Data Lake** to Amazon **Data Warehouse** Redshift clusters, saving 3 hours weekly for the business
- Developed and optimized robust **ETL pipelines** with the requests from finance teams, resulting in a decrease in data processing time
- Employed **AWS Glue** to craft an ETL process converting JSON dataset to CSV format, subsequently storing the output file in an **S3 Bucket**, and monitoring the applications using **CloudWatch**
- Executed over 100 complex SQL queries using DataGrip, extracted data from Amazon Redshift to **Power BI**, and constructed AI-driven Power BI dashboards for Amazon Delivery HR, resulting in 80% decrease in workload for the analysis team
- Supported data streams integration using **Kafka**, ensuring real-time data analysis to track associates' attrition aligned with stakeholder objectives, and assisted the project manager in saving 380 hours daily within North America's Amazon factory workers

Wang Lab for Biomedicines at Washington University in St. Louis

Aug 2022 - Dec 2022

Data Researcher

St. Louis, Missouri, United States

- Managed storage of experimental patient data in **SQL server** and wrote 50+ SQL queries to extract the data based on the UPI team's need
- Developed a signal inverse fitting model dealing with thousands of noisy water molecular diffusion data with 98% fitting accuracy in **Python**
- Utilized Monte Carlo Simulation to create randomized diffusion weighted imaging signal data, refined the model to achieve an 84% fitting accuracy, and implemented the enhanced model to contribute to the research efforts of the dMRI team

AVIC Securities

Jul 2021 - Aug 2021

Data Engineer Intern

Beijing, China

- Developed 5 automatic **ETL pipelines** in SSIS and **Azure Data Factory** to integrate data from 10+ data sources and conducted transformation to generate new metrics ensuring consistency
- Utilized Azure Stream Analytics for real-time data processing and insights extraction
- Designed and implemented data validation and unit testing procedures to ensure data quality
- Built data visualization dashboards in **Tableau** to present the weekly changing rate of bond and currency data and gave weekly presentations

iQIYI Inc

Mar 2021 - Jul 2021

Data Engineer Intern

Beijing, China

- Executed nearly 70 **SQL** queries to extract data like user gross data and video view data from **Oracle**, defined the ETL pipelines with PySpark and SQL in **Spark**, and monitored the tasks dependencies and scheduled pipelines weekly execution in **Airflow**
- Created DBT models to define data transformations on the datasets as SQL code and load the processed soil datasets with 200+ features to Google BigQuery in **GCP**
- Migrated data from Excel to **Power BI**, and built visualization dashboard on DAU and page view data that automatically analyzed user activity
- Developed and optimized a retention score prediction model, reducing the Mean Squared Error to 0.001 by adding a monomial expression in the Linear Regression model in **Python**, and gave recommendations based on the result to the BI group and project manager's decision making

PROJECTS

Like & Recommend: Amazon Review Data Project

Spring 2020

- Designed and constructed a weighted model on the helpfulness of reviews using SVD algorithm and Ridge Regression
- Obtained the word features for helpful reviews using the NLP machine learning model and made prediction on the helpful votes
- Predicted the weighted reputation in time series analysis
- Won Honorable Mention Prize in Mathematical Contest in Modeling 2020