Jason Xu

669-214-9107 | jasonxu20000505@gmail.com | jason-xu-b7b70322b | https://sonaj000.github.io/

Professional Summary

Junior Data Engineer with experience designing automated data pipelines, orchestrating scalable data workflows, and integrating databases using Python, SQL, PySpark, and AWS. Background in computational science and machine learning with practical applications in real-time data systems and analytics.

Education

University of California Santa Cruz — Santa Clara,CA
Master of Science in Computational Science — *September 2023 – June 2025*

Northwestern University — Evanston, IL
Bachelor of Arts in Neuroscience, Computations and Systems Modeling — *September 2018 – June 2022*

Experience

Data Engineer Intern — April 2025 – June 2025
ASE US Inc — *Sunnyvale, CA*

- Built an automated data pipeline in **Python** using **Pandas** and **PySpark** to map **hundreds** of new quotation pricing records per cycle into the internal pricing database, leveraging fuzzy matching and rule-based transformations.
- Achieved accurate mapping of over **85%** of new quotation data, with approximately **10–15% flagged for manual review**, reducing manual processing time from **4–8 hours to under 1 hour** while improving **data consistency** and **audit traceability**.
- Collaborated with the **Sales Operations team** to develop **data validation** and **error-flagging logic**, surfacing **10–15%** of records per cycle for review to ensure pricing accuracy and maintain data quality standards.
- Designed a Python-based archival system, writing processed quotation data to a secure **PostgreSQL** database; improved data retrieval speeds by **60%** for reporting and downstream analytical tasks

Graduate Student Researcher — September 2023 - March 2025
University of California Santa Cruz — *Santa Cruz, CA*

- Designed and implemented an **automated testing** framework in C++ leveraging the Go-Explore Algorithm to detect reachability bugs in 3D levels within **Unreal Engine 5**. Our framework was able to get more than **80% coverage** relative to a human play tester.
- Engineered and exported a **supervised hybrid neural network** using **PyTorch** and **Python** to **Unreal Engine 5**, enabling model-driven prediction of optimal game states for exploration and driving a **55% increase** in total test coverage across evaluated levels.
- Conducted statistical analysis and data modeling in **Python** across **3 test levels** and **15 test runs**, comparing human vs. algorithmic performance, evaluating coverage efficiency improvements, and visualizing key insights in **Tableau**.

Research Intern — July 2022 – January 2023
GUII Lab — *Santa Clara, CA*

- Developed a data pipeline to process multimodal data (chat logs, facial recordings, biometrics, gameplay points) from **20+ players**, performing **Python (Pandas)** preprocessing, feature engineering, and **time-series analysis** across **thousands of datapoints** to detect player stress patterns in an ARG game.

Projects

FFXIV Market Data Pipeline | *Python, AWS, Airflow, PostgreSQL*

- Built an end-to-end data pipeline to ingest, process, and store external market data from the Universalis API, leveraging **AWS S3** for **data lake storage** and **PostgreSQL** for structured reporting and analytics.
- Orchestrated scheduled batch **ETL** workflows using **Apache Airflow**, with integrated data validation and operational monitoring to ensure data accuracy and pipeline reliability.
- Designed relational data models to support the Faerie server market **trend analysis**, **item demand forecasting**, and **cluster detection** to help users identify profitable items to sell or craft.

Skills

Programming Languages:Python, SQL (PostgreSQL, MySQL), C++, JavaScript, HTML/CSS, Bash
Frameworks& Tools: PySpark, AWS, Airflow, Pandas, NumPy, Tableau, Selenium, Unreal Engine 5
Developer Tools: Git, Docker, AWS
Certifications: AWS Certified Cloud Practitioner, Sophos Central Endpoint and Server v4.0 – Engineer/Architect
Languages: English (Native), Mandarin Chinese (Professional), Spanish (Basic)