Feng Xue

Redondo Beach, CA | 213-249-1040 | xfgavin@gmail.com | linkedin.com/in/xfgavin/ | github.com/xfgavin

SKILLS

Programming Languages: Python, Matlab, Bash, Excel VBA, SQL, HTML, SAS, R

Data Engineering Tools: Spark, Airflow, PostGIS, AWS, Tableau

Frameworks and Tools: Git, Kubernetes, Docker, Singularity, Dash, Ansible, Sun Grid Engine (SGE)

EXPERIENCE

Data Engineer, *Insight Data Science,* Los Angeles, CA   09/2020 - Present

- Built the "Smart Rider" app (http://bit.ly/xfapp) to suggest rideshare rates based on historical data and user's input, including geo location, month, traffic situation, and social restriction e.g. COVID.
- Developed data processing pipeline using Python/Spark/Airflow/PostGIS/Dash/Docker to process a ~240GB historical geolocation based dataset on Amazon AWS.
- Optimized the pipeline by adding postgres stored procedures, gained >10 times of performance boost.
- Developed Ansible playbooks and Docker images for quick deployment.

Senior Medical Informatics Analyst, *Blue Shield California,* El Segundo, CA   10/2019 - Present

- Automated monthly data reporting using a set of Shell scripts, saved ~200% of report generation time.
- Wrote an Excel VBA pre-negotiation checklist tool to help facilitate negotiating contracts with hospitals.

Computational & Data Science Research Specialist III, *UCSD*, San Diego, CA   02/2017 - 10/2019

- Optimized a neuroimaging data processing pipeline, increased data conversion performance by 500%.
- Developed scripts in Shell/MATLAB to streamline QA data collection and preparation for visualization.

DevOps Engineer (Remote), *Beijing Normal University,* Beijing, China   11/2013 - 09/2020

- Designed and administered a computational cluster using SGE, Kubernetes, Docker, and Singularity to support 30+ users to process neuroimaging data in an optimized workflow.
- Launched a set of Linux system administration scripts, reducing ~80% of administrators' work load.

Research Scientist, *University of Southern California,* Los Angeles, *CA*   01/2011 - 02/2017

- Lead the team on data collection, storage, processing and modelling for a 200-subject neuroimaging research project that collected ~2TB of neuroimaging and behavioral data.
- Created a shell script library to automate neuroimaging data management and processing.
- Published 7 scientific papers.

EDUCATION

Beijing Normal University,   *Ph.D. in Cognitive Neuroscience, 2010*   Beijing, China

Beijing Normal University,   *B.S. in Psychology, 2002*   Beijing, China