

Feng Xue

Redondo Beach, CA | 213-249-1040 | xfgavin@gmail.com | [linkedin.com/in/xfgavin/](https://www.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 Engineering Fellow, *Insight*, Los Angeles, CA <http://bit.ly/xfapp> 09/2020 - Present

- Built the “Smart Rider” app to suggest rideshare rates based on historical data and user’s input.
- 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.

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

- Implemented a shell script tool to load and run SAS code, saving about 20% of time per data pull.
- 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 software in MATLAB, increasing data conversion performance by 500%.
- Developed scripts in Shell/MATLAB to streamline QA data collection and preparation for visualization.

System Architect / DevOps (Remote), *Beijing Normal University*, Beijing, China 11/2013 - 10/2020

- Designed and administered a computational cluster using SGE, Kubernetes, Docker, and Singularity, simplified research workflow.
- Launched a set of system administration scripts, reducing ~80% of administrators’ work load.

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

- Guided the team on data collection, storage, processing and modelling for a 200-subject neuroimaging research project that collected ~2TB of neuroimaging and behavioral data.
- Modelled neuroimaging data using general linear model and R, published 7 scientific papers.
- Created a shell script library to automate neuroimaging data management and processing.

EDUCATION

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

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