

EDUCATION

University of Pittsburgh

Master of Science in Information Science (3.7/4.0)

Pittsburgh, PA

Aug. 2017 – May. 2019

Xiamen University

Bachelor of Management in Accounting (3.6/4.0)

Xiamen, FJ, China

Sept. 2013 – July 2017

Relevant Courses: Cloud Computing (CMU), Algorithm, Advanced Database Management, Web Development, Data Structure, Computer Network, Information Retrieval, Information Visualization, Operating System

WORK EXPERIENCE

Software Engineer Intern

CDAR Center

Pittsburgh, PA

May. 2018 – Dec. 2018

The project involves developing a web application which bridges multiple molecular simulation programs and serves as an integrated online computational platform. (PHP, Laravel, React, MySQL, Apache, Python, HTML, Sass)

- Constructed a task management system based on **Laravel MVC framework** and improved the extensibility of function
- Designed and built responsive interface using **Bootstrap** and **Sass**, and created custom reusable **React** Components Library
- Implemented a set of **RESTful APIs** that encapsulated powerful GPU-Accelerated computing resources in backend with **Facade Pattern** to provide web service for third-party developers and reduced latency by 70% via **Job Queue** and **Redis**
- Optimized the database operations with **Eloquent ORM framework**, which also minimizes the efforts of switch databases
- Developed GUI plugin by **Python** that utilizes the APIs to generate molecular topologies for visualization system

SKILLS

Languages: Java, Scala, Python, JavaScript, SQL, HTML, CSS/Sass, PHP

Frameworks: Laravel, Vert.x, Undertow, Flask, React, jQuery, Spark, Hadoop, Kafka, Samza

Platform and Tools: Node.js, PostgreSQL, MySQL, MongoDB, Neo4j, Redis, Amazon Web Service (AWS), Azure, Google Cloud Platform (GCP), Terraform, Docker, Kubernetes, Git, Ubuntu, JUnit, Apache, Maven

PROJECTS

Twitter Analytics Web Service on the Cloud

Led a team of 3 to build performant and reliable web services to process HTTP requests for queries of Twitter data and to conduct data analytics within a specified budget (Java, Scala, Vert.x, PostgreSQL, Docker, AWS, Azure)

- Developed fully functional web servers in the frontend with **Java** and **Vert.x framework** to process queries using **multi-threaded design** and **in-memory cache** to achieve 18K responses per second and zero error rate
- Optimized the schema of database and used **PostgreSQL** to handle large scale data and a high volume of requests
- Implemented Extract-Transform-Load on a Twitter dataset (~1TB) through **distributed batch processing** by **Scala** and performed cluster computing via **Spark** on **Azure**, achieved 90% less running time compared with Hadoop MapReduce
- Wrapped the web servers in **Docker** and deployed on **Amazon ECS** for **Auto Scaling**

New York City Taxi Application Backend System

Implemented a Uber-like backend system that could process ride requests and provide real-time taxi fare estimate, driver matching and advertising service (Python, Flask, Google Cloud Platform, Java, Apache Kafka, Apache Samza)

- Trained a fare predictor on **Google ML Engine** with **XGBoost** model based on over 10 million NYC taxi records and performed hyperparameter tuning to improve the accuracy and served to support real-time prediction
- Developed **RESTful APIs** with **Python** and **Flask** that could accept text, speech or image queries and extract locations coordinates using **Google Cloud ML API** and responds with text or speech result of price estimated
- Converted ride requests and profile of clients into real-time data streams with **Java** and **Apache Kafka**
- Implemented data processing pipelines to consume and analyze data streams via **Apache Samza** and deployed on YARN cluster to provide live driver-matching and advertising