Santa Clara, CA, USA
jiahuawang7@gmail.com

Jiahua Wang (Alvin)

Technical Skills

Likes: distributed-systems database web-services hadoop hdfs hive mapreduce apache-spark apache-kafka akka akka-streams scala java sql python ruby javascript node.js kubernetes tensorflow machine-learning deep-learning

Experience

Software Engineer – VMware   *Dec 2018 → Current*
java, azure, mongodb, javascript, reactjs, backend, kubernetes, docker

Worked on backend and frontend for Hydra, the service for managing virtual desktops in Microsoft Azure.

Software Engineer – Tesla   *Jun 2017 → Aug 2018*
hadoop, hbase, apache-spark, apache-kafka, apache-kafka-connect, hdfs, mapreduce, akka, akka-streams, java, scala, python, backend, kubernetes, reactjs, docker, javascript

Built core analytics infrastructure for Model S/X/3, energy, supercharger, and manufacturing data. This data was used to improve our products, debug issues, and support court cases.

Migrated streaming ETL pipelines to use Akka Streams, improving reliability and peak throughput by 10x.

Designed and implemented OAuth2-based authentication and authorization for our primary service for providing data on demand to our users. This security was instrumental in getting our service to meet GDPR requirements. The service was used by >1000 people and >10 applications.

Software Engineer, Tools and Infrastructure – Google   *Oct 2016 → Jun 2017*
guice, protocol-buffers, testing, unit-testing, integration-testing, productivity, backend, machine-learning, scalability, bazel, java, python, c++

Worked on testability and productivity tools for an anti-payment fraud system at Google which used AI/ML to determine fraud probability.

Learned about how TensorFlow is used at Google and applied it to a toy problem.

Software Engineer – Cask Data   *Feb 2014 → Mar 2016*
hadoop, mapreduce, hdfs, hive, prestodb, java, soa, high-availability, scalability, distributed-system, command-line-interface, mysql

Simplified big data application development via the Cask Data Application Platform (CDAP), a platform that simplifies developing applications that interact with Hadoop and its ecosystem. Worked on realtime and batch data ingestion, security, and various applications.

Developed an extensible and configurable ETL framework. This project helped Cask acquire its largest early customer by proving immediate value wrt to ETL.

Developed a command-line interface, featuring context-sensitive autocompletion. Reduced the friction around using CDAP significantly, which helped Cask land several customers at a time when the UI wasn't fully usable.

Software Engineer – Electronic Arts   *May 2013 → Feb 2014*
mysql, android, ios, amazon-ec2, distributed-system, jvm, hibernate, rabbitmq, amazon-elb, rest, angularjs, jquery, java, javascript

Developed an ad event tracking service that was instrumental in driving advertisement revenue for most of EA's mobile games. The service processed 1.4k QPS on average.

Developed a low-overhead monitoring solution for our backend infrastructure supporting most of EA's major mobile games. The system handled approximately 20k QPS, uncovered numerous previously-unknown issues, and was able to depict the health of the system in near-realtime so that production issues could be spotted quickly.

Software Engineer Intern – Facebook   *May 2012 → Aug 2012*
memcached, php, c++, c, python

Developed a low-overhead feature for reducing the size of metadata being stored in our memcached cluster.

Reduced photo metadata size by ~50% and provided a way to reduce the size of other types of metadata by 20-50%.

Education

Computer Science – University of Southern California   *2010 → 2013*
c#, c++, c, java, unity3d, unrealscript, unreal-development-kit

Developed status effect system and lobby client/server for Core Overload, a multiplayer 3d space combat game.
https://youtu.be/Ka9GG7ZD9Q4

Certifications

Structuring Machine Learning Projects by deeplearning.ai on Coursera – https://www.coursera.org/account/accomplishments/verify/W4FJ3EFH2WTT

Apr 2019 → Apr 2019

Improving Deep Neural Networks: Hyperparameter tuning, Regularization and Optimization by deeplearning.ai on Coursera – https://www.coursera.org/account/accomplishments/verify/9X77X29V3VEH

Mar 2019 → Mar 2019

Neural Networks and Deep Learning by deeplearning.ai on Coursera – https://www.coursera.org/account/accomplishments/verify/JU2U9CECJJL2

Dec 2018 → Dec 2018

Machine Learning by Stanford University on Coursera – https://www.coursera.org/learn/machine-learning

Sep 2015 → Sep 2015

Projects & Interests

Cask Data Application Platform (CDAP) – https://github.com/caskdata/cdap
hadoop, mapreduce, hdfs, apache-spark, hive, sql, java, scala

Aug 2014 → Mar 2016

Open source framework for building and running data applications on Apache Hadoop.