Chaoran Yu

(510)-542-7749 yuchaoran2011@gmail.com
5230 Winkworth Pkwy, Syracuse, NY 13215

EDUCATION

University of California, Berkeley 06/2011 - 05/2014
B.S. in Computer Science & Engineering
GPA: 3.843
Graduated with High Honors (Top 10%)
Member of Tau Beta Pi (TBP) & Eta Kappa Nu (HKN) Engineering Honor Societies

EXPERIENCE

Lightbend Inc. 11/2016 - 07/2020
Senior Software Engineer - Fast Data Platform/Cloudflow Engineering *Remote*

- Worked on Lightbend's streaming data platform remotely in a distributed team. Voted a Lightbend Champion for being an outstanding team player.
- Developed billing infrastructure on Google Cloud Platform Marketplace and AWS Marketplace for the commercial version of Lightbend's open source Cloudflow project.
- Focused on running Apache Spark and Flink on Kubernetes (GKE and OpenShift). Contributed various features to and created Helm charts for Kubernetes Operators for Spark and Flink.
- Built a suite of testing and verification tools to enable diagnosing problems with services running in a Mesos cluster. Created and maintained a Jenkins cluster on AWS for internal CI purposes.
- Built customized containerized applications (e.g. Spark Job Server and Apache Hive) to run in a Mesos and Kubernetes environment.
- Wrote and maintained scripts using cloud automation tools e.g. Ansible and Terraform extensively to automate cluster deployment and management in Azure and AWS.

Bloomberg L.P. 01/2016 - 09/2016
Software Engineer - Bloomberg Big Data Services *Princeton, NJ*

- Built REST APIs for Java web services that allow clients to effectively retrieve data records.
- Built a data processing platform that can run batch and stream processing jobs utilizing Apache Spark and Kafka Streams.
- Wrote Python scripts to crawl websites and ingest data into distributed data stores. Worked on Python interface with .NET that makes data in Bloomberg Terminal accessible in Excel.

MediaFirst TV Platform, Ericsson Inc. 05/2014 - 11/2015
Software Development Engineer II - Cloud Analytics *Mountain View, CA*

- Designed and built an end-to-end data pipeline that stores real-time client and service logs in Cassandra and enables cloud services and web clients to fetch logs efficiently.
- Wrote Spark jobs to process user client logs and TV program catalog to generate reports on live TV usage.
- Rebuilt user interface of service monitoring web portal with JQuery, Knockout.js and Bootstrap.js to better visualize service health indicators.
- Served as scrum master for the team for 6 months.

Video and Image Processing Lab, UC Berkeley 05/2013 - 05/2014
Undergraduate Research Assistant (Advisor: Prof. Avideh Zakhor) *Berkeley, CA*

- Developed a particle filtering system that tracks a person's indoor position in real time.
- Set up a Django server that runs particle filter to fuse two modalities of indoor localization: WiFi- and image-based.
- Implemented a human step detection algorithm on Android as motion model of the particle filter.
- Wrote an Android app that collects WiFi and image data for use by pre-existing indoor localization approaches.
- Operated the backpack system that collects building geometry, WiFi RSSI's, and images in an indoor settings. Processed large amount of data collected by backpack to generate floor plans and 3D building models.

TECHNICAL STRENGTHS

Languages	Java, Python, Go, Javascript, Scala, C#, C, Bash, MATLAB
Distributed Systems	Kubernetes, Spark, Flink, Kafka, Mesos
Storage	Cassandra, HBase, MySQL
Operations	Ansible, Terraform, Jenkins, Grafana, Splunk

TALKS

Running Spark and Flink on Kubernetes, A Case Study of Kubernetes Operators 11/2019
Athens Big Data Meetup *Athens, Greece*

· Covered topics such as Custom Resource Definitions, the Operator pattern, designs and implementations of open source Kubernetes operators for Spark and Flink, from Google and Lyft, respectively.

Hands-on Machine Learning with Kafka-based Streaming Pipelines 09/2019
O'Reilly Artificial Intelligence Conference *San Jose, CA*

· Co-taught a three-hour tutorial on serving machine learning models in production. Topics covered include architectures and production considerations for model serving as a service and models as data. Several sample architectures of model serving implemented using TensorFlow Serving, Apache Flink and Apache Spark Structured Streaming are presented.

Just Enough Scala for Spark 06/2017
Spark Summit 2017 & Cognitive Frameworks Festival *San Francisco, CA*

· Gave two full-day tutorials at the Spark Summit and IBM-hosted Cognitive Frameworks Festival on effectively using features of Scala to write idiomatic Spark programs. Topics covered include type inference, pattern matching, case classes, Scala collections etc.

BOOK REVIEW

Apache Cassandra Essentials
Author: Nitin Padalia *ISBN 978-1-78398-910-2 Packt Publishing*

· Reviewed chapters of the book under specified deadlines. Provided feedback to the author regarding code examples, phrasing, and organization of topics. The book was published in October 2015.

PUBLICATION

Simultaneous Fingerprinting and Mapping for Multimodal Image and WiFi Indoor Positioning
Plamen Levchev, Michael N. Krishnan, Chaoran Yu*, Joseph Menke and Avideh Zakhor*

· Proceedings of the 2014 International Conference on Indoor Positioning and Indoor Navigation (IPIN), Busan, Korea, Oct. 2014

CERTIFICATION

Apache Cassandra Certified Developer

· Certification No. V2.1-0003. Certified at Cassandra Summit 2015.

Scalable Machine Learning

· EdX MOOC certificate. Issued in Aug, 2015.

Scrum Alliance Certified ScrumMaster

· Member: 000360473. Certified in Oct, 2014.