Chih-Feng Lin

90 Christopher Columbus Dr, Apt 2411, Jersey City, NJ 07302
(650)-526-8759 | chihfenl@alumni.cmu.edu | github.com/ChihFengLin | www.linkedin.com/in/chihfenl

Education

Carnegie Mellon University M.S. in Electrical and Computer Engineering | 08/01/2014 - 12/23/2015 | Pittsburgh, PA
National Taiwan University M.S. in Applied Mechanics | 09/01/2009 - 07/31/2011 | Taipei, Taiwan
National Taiwan University B.S. in Civil Engineering | 09/01/2005 - 06/30/2009 | Taipei, Taiwan

Professional Experience

VP, Big Data Platform Architect | 01/15/2018 – Present
TD Securities | New York, NY

Architected Hadoop big data infrastructure used by various trading desks and internal engineering teams from the ground up. Headed big data project and improved platform to achieve multi-tenancy, secure and highly availability. As big data platform product owner, my primary job function included:

- Built Big Data Infrastructure on Premises and Cloud
  - Pioneered infrastructure-as-code workflow and developed Ansible script to automatically deploy Hortonworks distribution Hadoop Big Data Platform (HDP) on premises environments.
  - Led HDP SaltSatck script development and successfully improved efficiency to deploy complex Hadoop ecosystem (HDFS, YARN, HBASE, HIVE, SPARK, KAFKA) into diverse cloud like Azure.
- Developed and Orchestrated Scalable ETL Pipelines
  - Established Airflow cluster with Celery and RabbitMQ backend from scratch, which elevated capabilities of handling multiple ETL tasks and enhanced visibility of data management.
  - Implemented Spark based end of day market data ingestion pipeline digested from Kafka, which successfully boosted data processing time over 100%.
- Administered Prod and Dev Big Data Platform
  - Led Hadoop administration group to maintain, upgrade platform and recover from disaster. Advised solutions for clients across New York, Toronto and London with over 50 developers tp onboard big data applications.
  - Devised code based configuration managemnet provisioning methodology integrated with CI/CD pipeline. Improved reliability of managing over 200 configuration parameters in Hadoop ecosystem across all prod and dev environment.

Selected Projects

Predicting Aircraft Fuel Consumption Using Deep Learning - NASA Sponsored Project

- Established Stacked Autoencoder (SAE) model with 182 aircraft feed-in features and constructed the training workflow for multilayer network architecture, which can be effectively applied to other prediction tasks. Succeeded in improving prediction rate by 40% on testing data as compared with pure linear regression model. (https://goo.gl/tcTZdR)

Cloud-Based Twitter Analytics Web Service

- Created a REST API web service (Python Tornado Framework) with MySQL and HBase backend data storage combined with AutoScalingGroup (ASG) and ElasticLoadBlancer (ELB) functions to analyze Twittwr data based on MapReduce structure, which successfully demonstrated with stable high QPS capability under arbitrary query test.

Previous Professional Experience

Data Engineer | 02/15/2016 – 01/01/2018
Bomoda (Acquired by Weber Shandwick) | New York, NY

- Designed ETL Process: Built a message queue system and API to digest raw data from vendor, integrated with cron jobs to automatically ship raw data to S3 data storage and a real-time error tracking system such as Sentry to improve the reliability of the process.
- Developed Python And Scala Based Backend System: Transformed the conventional single-machine data processing system into a distributed architecture that is deployed on Amazon AWS cluster using Spark framework, which significantly reduced the computation time from 10 hours to 1 hour and converted the data into clean and structured format.
- Bridged From Raw Data To Readily Accessible Information: Established the cron jobs that automatically ship cleaned data to the connected BigQuery database and designed all the schemas to normalize data, provided with an easy and queryable interface for data strategy team, which successfully doubled the company's productivity.
- Managed Product And QA Environment: Utilized Docker and Vagrant to build deployment-like server environment on the local machine. Programmed Ansible playbook to automate ETL process locally and deployed code on the product server after passing testing.

Software Engineering Intern | 06/21/2015 – 08/15/2015
Renault Innovation Silicon Valley | Sunnyvale, CA

- Prototyped In-Car Seat Driver's Heartbeat Detection System: Developed system including hardware and software. Embedded sensor into car seat to capture vibration signal from driver and communicated digital signal with computer via Arduino.
- Created Heartbeat Detection Algorithm: Analyzed the extracted siganl and designed a new algorithm based on time-frequency analysis to convert vibration siganl to human readable heartbeat number. Successfully achieved error rate within 10% as compared with ground truth.

Skills

Programming: Python, Scala, Java, SQL, Bash, C
BigData: Hadoop Ecosystem(HDFS, MapReduce, Spark, Kafka, Yarn), AWS, Azure
DataWarehouse: BigQuery, MySQL, MongoDB, HBase, S3, Alluxio, Hive, Redis
DevOPs: Ansible, Salt, CloudFormation, Docker, Vagrant, Packer, Jenkins, Kubernetes
Others: Ambari, Airflow, UnitTest, Git, Maven, Vim, Sentry, LaTeX