

Ken Wu

Brooklyn, New York

Senior Data Engineer at iHeartRadio

Summary

A passionate and talented senior big data architect/full stack engineer with 12+ years of professional software development experience, various open-source projects contributions with specialties in data-warehousing, system architecture, scripting skills, data-analysis/engineering, multi-threaded robust applications

Highlighted Core Skills - As of Fall, 2018

- Cloud computing/Big Data: 4+ years (with 3+ yrs in Spark/Hive/RDS/Redshift A.W.S., 2+ yr in AirFlow/Luigi/Jenkins Pipeline)
 - Data structure/Algorithms: 7+ years (with 7 yrs in Java enterprise/5 yrs in python; 6 yrs in complex data algorithms processing, 1 yrs in M.L./Recommender algorithms)
 - System architecture/design: 3+ years (Built and maintained the Hadoop/Spark cluster, implemented the Data Lake storage system, data-modeled the consumer aggregated report systems)
-

Experience

Senior Data Engineer at iHeartRadio

January 2019 - Present (9 months)

- Leading the effort to maintain and improve and expand the analytics and machine learning infrastructure and data pipelines
- Collaborating with teams of data scientists here implementing some M.L. models with the data scientists here using Stochastic Gradient Descent (SGD) and Neural Networks/Deep Learning.
- Upgrading the AirFlow 1.10.5 with python 3 on the whole infrastructure (the legacy system was on 1.9.0 and python 2.7)
- Won 2 hackathon programming competitions with both best-engineering award

Big Data/Cloud Engineer Lead at Unified

December 2015 - December 2018 (3 years 1 month)

- 1) Built and implemented the first-generation ETL data pipeline from raw s3 data received from first data party provider and then injected to the AWS redshift using Hive and Spark (PySpark DataFrame API); And for second generation, re-architected it in a data-lake on RDS version to further save company's storage costs with most jobs hooked up with data quality checks to ensure data correctness. Examples of the ETL spark jobs are to i) calculate the markov_chain problems,

identify all audiences's user numbers that are increasing each quarter by 10%, and ii) writing customized pyspark udf function for aggregations.

2) Hooked up the above second generation ETL data pipeline using Jenkins DSL groovy and Apache AirFlow for full automation and task schedulings/executions.

3) Contributed to lot to their python code base: Communique and Audience Generation tools.

4) Help data-science team to automate their existing manual long running jobs by chaining them together using Jenkins Pipeline in the Jenkins environment.

Technical Skills: Hive, Spark, Redshift, RDS, Apache AirFlow, Python, Pandas, Jenkins Pipeline, React, redux.

Achievements:

- Won the company's 2017 Hackathon programming contest on creating a smart writing a chatbot in 48 hours.
- Implemented features to support Facebook Audience Generation and developed code to support iHeart 1st party data ingestion.
- Coached and provided lot of technical guidelines and assistance to the junior members in the teams who did not have computer science background.

Senior Big Data Architect at Qualia Media

August 2014 - December 2015 (1 year 5 months)

- 1) Built the ETL data pipeline along with the consumer intent signal systems on a huge hadoop/spark cluster using luigi as the work-flow management tool for job scheduling/executions. Some highlights of the achievements are implementing the automatic parallelization on different data topics and the pre-downloaders of remote assets off the HDFS from S3 which speeds up the daily work flow by 30+ % at least.
- 2) Lot of Dev ops responsibilities such as managing all the AWS EC2 nodes and making sure they are mostly up and be responsive, developing puppet configuration scripts to create and provision the cluster.
- 3) Contributed to the Apache Pig and ElasticSearch open source community for some of the bugs I found and fixed and created the pull request – such as PIG-4623.
- 4) Led the entire software life-cycle including hands-on development on platform code base such as Kafka-client, scala on Spark, pig scripts . . . etc, code reviews, unit-testings and integration testings and deployments on Jenkins environment.

Technical Skills: Big data, Spark, Hadoop, HDFS, HIVE, MapReduce, python, mesos, react and circle-CI.

Achievements:

- The per-down-loaders I envisioned and developed saved lot of network overhead to s3 on the cluster and increased the data-locality since they are now available in the HDFS for processing.
- I used puppet to build a Hadoop/Spark cluster on top of Mesos and it was successfully deployed to production and has currently been used. The direct consequence is to save money for my company for additional cluster cost for Spark since right now Spark can co-sit with Hadoop on the same cluster on top of Mesos.
- I designed an efficiently smart partitioning scheme on storing and processing the scoring-data to optimize the work-flows resulting dramatic performance improvements.

Senior Software Engineer / Team Lead at TravelClick

January 2013 - July 2014 (1 year 7 months)

Technical Skills: JAVA 6 and 7, JBoss, Hadoop, HDFS, Kafka, ElasticSearch, Oracle coherent cache, JMS and Fiorano MQ, Agile software development, Scrum.

Achievements:

- The optimization engine that I built reduces lot of database overhead and resulted in a huge performance improvement (at least 35% faster on average).
- The live monitoring system has been widely used across the business and received a huge recognition from different product teams.
- Continuously worked on resolving the hard production issues that troubled the products and reduced the JIRA tickets by 20-30% monthly on average.
- Provided the mentor to the other new and junior developers so as to speed them up in the development; Being a scrum master in a rotation basic.
- Envisioned the needs of transforming from SVN to GIT and took a lead in the transformation and being a Git master to other teammates.

Markit On-demand Sr. Core Java developer, A.V.P. at Markit

October 2009 - January 2013 (3 years 4 months)

distributed Java EJB system, multi-threaded systems, server side programming, java GWT and Web developments

Senior Software Developer at Western Union

March 2007 - October 2009 (2 years 8 months)

Primary project was 1) an online payment service applications in C#, which was a high profile payments processing

application: maintained and enhanced the server side and the middle layer library codes that were shared across different systems; developed business logic implementation handle a mix of change of business requirement. 2) An event-driven GUI web applications using ASP.NET for the world-wide external users: developed original payment website in highly MVC architecture, encompassing design, coding, testing, and maintenance; worked with project managers to specify, design and execute business application projects; reviewed code produced by other teammates and ensures high code quality; mentored peers on the team who were less familiar with the .net language and web developments. Provided architecture level of planning on upgrading of their .NET legacy systems from 2.0 to 3.5

Software developer at CCSI - Contemporary Computer Services, Inc.

May 2006 - March 2007 (11 months)

ASP.net, database programming

Education

State University of New York at Stony Brook

M.S, Computer Science, 2005 - 2006

Activities and Societies: Member of Stony Brook Association for Computing Machinery (A.C.M) programming team, Upsilon Pi Epsilon C.S. Honor Society

State University of New York at Stony Brook

B.S, Computer Science, Applied Mathematics and Statistics, 2002 - 2005

Activities and Societies: Member of Stony Brook Association for Computing Machinery (A.C.M) programming team, Upsilon Pi Epsilon C.S. Honor Society, Cum Laude Honor

Ken Wu

Brooklyn, New York

Senior Data Engineer at iHeartRadio

LinkedIn® Recruiter

2 people have recommended Ken

"I was fortunate to work with Ken on the DataLake project at Unified and he had a consistently positive impact on the successful finish of the project. While I was just freshly graduating from Carnegie Mellon University with master degree in Software Engineering few years ago but not much real world enterprise data pipeline experience, Ken was an accomplished technical lead with exceptional technical and leadership skills to guide me to implement the first phrase DataLake project from scratch to production. I had worked closely with him in the past 1 year and had admired his dedication for perfection and best practices."

—Tianbing Leng, Software Engineer, Unified, worked indirectly for Ken at Unified

"Few people have the opportunity to with strong team players like Ken. I was fortunate to hire Ken three years ago and it was one of the best decisions I made. Not only is he a sharp technologist, he's also willing to share that knowledge and help others out. He's quick to learn new languages and stops at nothing less then total understanding. As a team member, Ken earns my highest recommendation."

—Kevin Ferguson, Head Of Engineering, Vetterly, managed Ken at Unified