

Contact

sonya_ling1947@yahoo.com

www.linkedin.com/in/sonya-ling-699bb7b (LinkedIn)

github.com/threecuptea (Personal)

Top Skills

AI Agents

Python (Programming Language)

PyTorch

Certifications

Cloudera Certified Developer for Apache Hadoop (CCDH)

Redhat Certified System Administrator (RHCSA/ RHCT),

Sun Certified Programmer for Java 2

Learning Kubernetes

Data Science Foundations: Python Scientific Stack

Honors-Awards

Community TA of 'Functional Programming Principles in Scala' Coursera course

Sonya Ling

Staff Software Engineer@OpenX (OpenAudience, Big Data, ML, GCP, Vertex AI)

Hacienda Heights, California, United States

Summary

A sharp-minded Staff Software Engineer with over 15-year experiences and a tenacious and proven-record tech lead of building data pipelines and high-volume distributed system end-to-end

- I. Very quick learner of both emerging technologies, (GenAI & LLMs) and business knowledge, with extensive experience on Ads audiences and targeting. Familiar with Ads regulations like GDPR and CTV digital marketing
 - II. Great cross-discipline breath: Cloud Platform GCP & AWS, Linux, Docker, CI/CD, Kubernetes, and Vertex AI
 - III. Excellent working with cross-function teams: Product Managers, SREs, Data Scientists and UX team
 - IV. Proficiency in Python, Java, Go, TypeScript, BigQuery, GraphQL, PostgreSQL, and NoSQL MongoDB
 - V. Professional knowledge of Machine Learning and Deep Learning.
- Languages: Python (6 years), JavaScript (2 years), Go (3 years), Bash script (2 year) & Java (10+ years) and Scala (1+ year), C++ (1-year)
 - Cloud Computing: GCP (6 years), AWS (1+ year)
 - Linux and Shell commands (15 years)
 - SQL Databases: Google BigQuery (6 years), Postgres (4 years), MySQL (4 years), SQL (15 years)
 - Hasura: GraphQL engine (6 years)
 - NoSQL: MongoDB (4+ years), Administrative & advanced level
 - BigData: Google BigTable (2 years), Hadoop (3 years) and Apache Spark (1+ year)
 - Machine/ Deep Learning: Panda/ Numpy (2+ years) , Scikit-learn(1 year), Vertex AI (1- year), PyTorch (1- year), GenAI tools & LLMs (1- year)
 - Message systems: Google Pubsub (5 years), Kafka (2+ years)
 - Misc. : Airflow (6 years), Kubenetes (3 year), CI-CD (3 year), Grafana/ Prometheus (3 years)

Experience

OpenX

6 years 7 months

Staff Software Engineer

March 2023 - October 2025 (2 years 8 months)

Pasadena, California, United States

Led building a new Lookalike middle-tier pipeline using TypeScript API and WTFs, customized workflow processing modules written in Python, that listens on Google PubSub and being triggered on messages. This effectively connected the UI with backend Lookalike ML model-trained and inference jobs orchestrated by Google Vertex AI pipeline. Worked closely with ML engineers on troubleshooting.

LAL's ML model is using a BigQuery table as its features store. Performance tuning the BQ query to eliminate frequent job failures due to BQ 6-hour execution limit by breaking it down into small manageable queries being persisted as temp tables and redirect the BQ job to a project with higher time-slot cap.

Built Hadron Service to integrate with an external partner using OpenRTB protocol in Delivery User Data Service (UDS); facilitating the partner's contextual segments on deal targeting. Tested throttling and scaled up traffic to the full.

Spearheaded onboarding IP, and ID5 Advertiser IDs. Built the pipeline by enhancing multiple components written in Python, Java, or Go. UDS is a high-throughput, low-latency distributed system written in Go. Refactored UDS by adopting the best practices of object-oriented programming and dependency injection for easy on-boarding of more Ads Identity IDs. All OpenX systems are hosted in GCP.

We failed to on board IPs badly due to the fact that IPs are not good hash keys by nature and led to CPU spikes constantly in one BigTable cluster. Solved the problem by prepending 4 bytes hash of parsed IPs. Onboarding ID5s involves decryption algorithm and access to its key pairs for decryption. Wrote decryption algorithm in Go and designed a memory store for fast lookup with atomic concurrent access in mind. Also optimize loading with versioning.

Senior Software Engineer

April 2019 - March 2023 (4 years)

Pasadena

OpenAudience's well known Identity-Graph is a BigQuery table built by creating presumable Ads user nodes (OAIDs) centrally around shared edges (Ads IDs like cookies, devices, and Ramp IDs). Onboarded LL IDs as stratum 3 of the graph by building hashed emails edges between LL's and EXP's. The clean version of Identity-Graph takes GDPR and opt-out into consideration. CTV Graph is for the reverse lookup of CTV devices to OAIDs.

Built OpenAudience's audience activation and identity IDs export pipeline by leveraging the Identity-Graph to bring in connected identity IDs for exports. Automated the weekly batch process by AADF. AADF is a group of Python modules attempting to balance OpenX resources and the effectiveness of IDs based upon predefined business rules and is triggered by Airflow.

Responsible for integrating segment, audience, deal, and deal spending data from GraphQL and BQ to feed reporting needs of data warehousing

Maintained LegacyAudience system implemented with Hadoop. Designed an early version of Lookalike implemented using Apache Spark.

TiVo

Staff Engineer, Software Deveopement

February 2018 - March 2019 (1 year 2 months)

Santa Clara, California, United States

Developed a new Danube-UDF (Unified Data Foundry) Pipeline which integrated the TiVo batch data processing with the Rovi Danube near real-time pipeline. Applied A/B testing using Spark DataFrame to ensure data integrity.

Coded and migrated the Danube interface from RabbitMQ to Apache Kafka 0.10.0. Added Avro Schema for serialization and de-serialization of Kafka messages, which utilized a compact data format and improved versioning management.

Implemented a single thread-safe KafkaProducer to publish messages with well-distributed keys, which optimized the load distribution across partitions. Multiple KafkaConsumers joins ConsumerGroup across multiple Kafka partitions to enhance processing speed. Reduced the costly re-balance process during the initial group joining stage.

Rovi Corporation (now TiVo)

4 years

Principal Software Engineer

July 2017 - February 2018 (8 months)

Burbank, CA

Diagnosed using VisualVM profiling and resolved OOM issues in the Danube Resolver. Inspected memory snapshots following heap dumps to identify and dereference objects that were incorrectly held in memory, which reduced memory consumption.

Analyzed the advantages and disadvantages of different Garbage Collection(GC) algorithms, specifically finding that while G1GC offers short pause times suitable for web applications, its aggressive memory acquisition was excessive for the backend-focused Danube application, helping guide future architecture decisions.

Senior Software Engineer

2014 - July 2017 (3 years)

Burbank, CA

Maintained the Danube Data Pipeline as the principal data engineer, successfully ingesting terabytes of data at a processing speed of over 2 million records per hour. Managed the production release process.

Administered a MongoDB cluster consisting of master, slave, and arbiter replica sets, ensuring high availability and reliability of the database and demonstrated expertise in MongoDB shell commands and the aggregation framework.

Payvia

Senior Software Engineer

October 2012 - December 2013 (1 year 3 months)

Built Mobile Messenger empowering consumers to donate charity or campaign securely

Zumbox

Senior Software Engineer

November 2011 - October 2012 (1 year)

Westlake Village, CA

Built online mail delivery service by redirecting mails when users chose 'Go Paperless'

Education

University of California, Irvine - The Paul Merage School of Business
MBA, Management Information System · (September 1990 - July 1991)

Case Western Reserve University

Business Administration and Management, General · (September 1989 - June 1990)

National Taiwan University

BA, English Language and Literature · (September 1981 - June 1986)