Jonathan Jin

Machine Learning Infrastructure Engineer

jonathanj.in | me@jonathanj.in | linkedin.com/in/jinnovation | github.com/jinnovation

Select Experience

Present – 03/2021 | **Spotify** Senior Machine Learning Software Engineer | New York

Member of ML Platform. Working on: multi-cluster, ML-centric cloud infrastructure with Kubernetes and Kubeflow; and ML engineering SDKs, based around TFX, for pipeline orchestration, feature engineering, data processing, model analysis + validation, model deployment, etc.

- Spearheading design and development of a unified user-facing configuration plane for platform's products; authored and drove consensus of architecture and design; guided multi-quarter development roadmap and coordinated cross-team collaboration;
- Designed formal, opinionated ontology of MLOps semantics to enable auditability of ML usage in production and form the basis of Spotify's Response AI initiatives, e.g. model cards;
- Drove team's primary product (managed Kubernetes cluster infrastructure with companion SDK) to general availability; collaborated with engineering and product leadership to identify and establish baseline product excellence, e.g. API design philosophy, documentation tooling, explicit reliability guarantees (SLOs), etc.;
- Implemented formalized SLO tracking for multi-cluster, pipeline execution infrastructure; used Terraform to canonicalize SLO definitions for monitoring and violation reporting in Google Cloud Monitoring; implemented custom Kubernetes listener to implement nuanced and domain-specific SLIs.
- Onboarded four new junior engineers in less than a year; held biweekly one-on-ones with each to address concerns, collaborate to identify growth opportunities, and provide technical and career coaching.

Kubernetes | Ray | GCP | Terraform | TensorFlow | TFX | Kubeflow | Prometheus | gRPC

01/2021 – 12/2019 | **NVIDIA** Senior Systems Software Engineer, AI Infrastructure | New York

Member of AI Infrastructure. Contributor to MagLev, NVIDIA's AI infrastructure for autonomous vehicle development. Also contributed to Modulus, the deep learning SDK for autonomous vehicle R&D.

- Initiated development of solution for "hybrid data/model parallelism" using a Ray-based parameter server design and Horovod to enable horizontally-scalable multi-task training;
- Co-delivered a Kubernetes-based scheduling mechanism to enable priority access to cluster resources for select use cases, e.g. prep for upcoming external demos, via virtual "resource shares";

Kubernetes | TensorFlow | Horovod | Ray | gRPC | Bazel | SwiftStack

12/2019 – 08/2018 | **Twitter** Machine Learning Software Engineer | New York

Member of Cortex, Twitter's central ML platform organization. Worked on: workflow orchestration; experiment management/iteration; and overall ML engineering productivity.

- Spearheaded initial integration of TensorFlow Extended (TFX)with legacy Airflow-based orchestration platformto increase agility of workflow development, iterative execution/experimentation, etc.
- Enabled distributed training of TensorFlow models in Apache Mesos from an Airflow pipeline via Deepbird, Twitter's TensorFlow-based model training/evaluating/serving framework

Apache Airflow | Apache Aurora | TensorFlow

Skills

Programming Languages	Python, Go, Bash, C++, Java
Machine Learning	Kubeflow, TensorFlow Extended (TFX), TensorFlow, Ray
Distributed Systems	Kubernetes, gRPC, Docker
Infrastructure Tooling	Bazel, Prometheus, Grafana, M3, Cassandra, Apache Airflow
Cloud Infrastructure	Google Cloud Platform (GCP), Terraform

Education

2015 **University of Chicago**, B.S. Computer Science, B.A. Economics