Machine Learning Infrastructure Engineer

Ø jonathanj.in **@** me@jonathanj.in **in** linkedin.com/in/jinnovation **♀** github.com/jinnovation

EXPERIENCE

Present 03/2021

Spotify, Senior Machine Learning Software Engineer, New York

Member of ML Platform.

- > Working on: centralized, multi-tenant ML orchestration infrastructure with Kubernetes and Kubeflow; and userfacing pipeline-authoring SDK based on TFX
- Implemented standardized, on-by-default resource quotas to mitigate risk of noisy neighbor effect in centralized **Kubeflow clusters**
- Oversaw comprehensive formalization of SLO-tracking strategy, using Terraform to formalize SLOs in GCP for all clusters in our multi-cluster in reproducible fashion
- Spearheaded development of a custom metrics exporter, transforming Kubernetes events into actionable Prometheus metrics to address gaps in our observability/reliability strategy

TensorFlow TFX Kubernetes Kubeflow GCP Terraform Prometheus gRPC

01/2021 12/2019

NVIDIA, SENIOR SYSTEMS SOFTWARE ENGINEER, AI INFRASTRUCTURE, New York

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

- > Developing 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"
- Authored self-service, reproducible, and traceable workflows to generate "miniaturized" production datasets, enabling rapid iteration/prototyping of training infrastructure refinements

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.

- > Core contributor to ML Workflows, Twitter's Airflow-based platform for productionizing ML pipelines
- > Spearheaded initial integration and cross-compatibility of TensorFlow Extended (TFX) with ML Workflows to 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

07/2018 07/2016

Uber, SOFTWARE ENGINEER, New York

Member of Observability Applications. Worked on forecasting and anomaly detection for time series metrics.

- Re-architected time-series metric forecasting pipeline to support concurrent batch backfilling; reduced asymptotic burden on underlying data store by 90%
- Extended M3-based anomaly detection platform to support multiple forecasting models; carried out migration to intercommunicating services with zero downtime and full backwards compatibility

Go Java M3 Apache Thrift Cassandra

07/2016 07/2015

OkCupid, Software Engineer, New York

Contributed to backend service development as part of a 10-person backend engineering team.

Implemented prototype collaborative filtering functionality for matching between prospectively compatible users.

C++

SKILLS

Programming Languages Python, Go, Bash, C++, Java

> Machine Learning Kubeflow, TensorFlow Extended (TFX), TensorFlow

Distributed Systems Kubernetes, gRPC

Infrastructure Tooling Bazel, Prometheus, Grafana, M3, Cassandra, Apache Airflow

Cloud Infrastructure Google Cloud Platform (GCP), Terraform

EDUCATION

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

JONATHAN JIN NOVEMBER 4, 2021