

TRUXTEN COOK

San Francisco, California · truxtencook@gmail.com · 6028108711

EDUCATION

Arizona State University

MS Computer Science & Big Data Systems, BS Computer Science *GPA: 4.0*

Tempe, AZ

Aug 2018 - Dec 2022

WORK EXPERIENCE

Tatari

Software Engineer - Data Platform Team

San Francisco, California

May 2022 - Present

- Designed and authored a framework to automatically cut and deploy infrastructure for MLFlow machine learning models on Databricks. Collaborated with multiple Engineering and Data Science teams to ensure that stakeholders needs and ML-OPS best practices were followed. Final pipeline included automatic Terraform plans and applies through the CI, a bespoke infrastructure-as-code solution to dynamically shift the model deployed to best suite business needs, and a Python framework to accelerate deployment of machine learning models.
- Spearheaded a large-scale optimization of S3 storage across hundreds of buckets, reducing reducing monthly costs by over \$30,000 across hundreds of buckets and a multi-petabyte Data Lake. Designed and implemented data lifecycle policies with Terraform, reducing the company's cost for some highly used buckets by over %60.
- Ported multiple terabyte-scale, business-critical ETL workflows from Redshift to Databricks using PySpark. Led the development of new ETLs, managed the backfilling of historical data, and ensured data integrity during the transition from the old Redshift pipeline to the new Databricks platform. This overhaul resulted in a 40% increase in speed for essential data ingestion jobs and yielded significant cost savings by reducing Redshift compute expenses.
- Architected and implemented a bespoke local Airflow deployment to optimize end-to-end (E2E) testing, significantly accelerating new feature development cycles. This versatile solution enabled non-technical stakeholders to conduct basic job tests autonomously, while providing technical users with the tools to manage a comprehensive Kubernetes (K8s) and Databricks development environment with ease.
- Served On-Call for the Data Platform team, rapidly responding to incidents and helping stake-holder developers with questions regarding Databricks, PySpark, Redshift, and more.

Sandia National Laboratories

Machine learning R&D Intern

Gilbert, AZ

May 2021 - Aug 2021

- Used ML-OPS best practices in PyTorch to create an end-to-end automated machine learning pipeline for various tasks such as real-time object detection and machine translation. This let developers create a PoC for new projects in a matter of hours rather than days.
- Implemented automatic logging and visualization of relevant hyperparameters and metrics using MLFLow.

TECHNICAL SKILLS

Programming Languages: Python, Terraform, C/C++, SQL, JavaScript

Technologies: Databricks, Spark, MLFlow, Airflow, Kubernetes, PostgreSQL, PyTorch, S3, ECR, EC2

Developer Tools: JIRA, Git, Confluence

PROJECTS

Distributed Database Hotspot Analysis using Apache Sedona Scala, Apache Hadoop, Apache Sedona

Used Scala, Hadoop, and Apache Sedona to create distributed map reduce system to solve the ACM SIGSPATIAL CUP 2016 problem.

Resource Description Framework (RDF) Database Java

Converted a prebuilt Relational Database written in Java to use the Resource Description Framework data model. Changes made were vast and involved directly modifying and creating low level database systems, such as custom B+ Tree index schemes, a custom data page structure & record design, and task specific query optimization and evaluation.

AWARDS

Moore Award

Given for graduating with a 4.0 G.P.A in under 8 semesters

Arizona State University

December 2022