Victor Yap

Email: vicyap24@gmail.com
Github: https://github.com/vicyap
LinkedIn: https://www.linkedin.com/in/vicyap/

EXPERIENCE

Zitara (Battery Management Software), Fully Remote Mar 2024 - Nov 2024
Senior DevOps Engineer (Part-Time, Full-Time)

- Implemented Infrastructure as Code and imported all infrastructure into Terraform and Terragrunt. Infrastructure changes changed from manual to managed through code using CI/CD in preparation for SOC2 Compliance.
- Simplified and migrated the deployment of the app (Flask + Celery + React) from manual standalone, long-lived EC2 instances to automated push-button deployments using Github Actions + ECS + MQ + CloudFront.
- Separated AWS shared, stage and prod environments from a single AWS account to multiple AWS accounts. Increased confidence and safety when deploying new changes or when trying new ideas.
- Reduced CI build times from 10 minutes to 5 minutes by implementing Docker base images and Docker external caching with S3. Managed developer tools to streamline working with Docker locally. Implemented integration tests using Docker.
- Worked with other teams by providing DevOps skills and guidance. Helped migrate to Dagster for our data pipelines which enabled a huge productivity increase for the data science and battery modeling teams. Helped the embedded software team manage ephemeral Windows VMs to replicate customer environments for testing.

Self-Employed, Fully Remote Apr 2022 - Mar 2024
DevOps Consultant/Contractor

- Worked with small companies (< 100 people) that lacked DevOps. Provided guidance regarding best practices, ideal solutions for their company's level of engineering and advice on how to hire a full-time DevOps role to fill their gaps.
- Helped clients get started with Infrastructure as Code (IAC), understand why it is important and where it does and does not apply.
- Implemented custom tools and extended existing tools and providers to achieve solutions.

Rev.com (ASR / Speech Recognition), Fully Remote Dec 2020 - Oct 2022
Senior MLOps Engineer

- Managed an AWS account for the entire team of about 15 data scientists with Terraform. This enabled data scientists to focus on building more accurate models – a key metric for the ASR business.
- Designed and implemented a highly elastic, distributed batch processing cluster with Ray, Kubernetes on EKS and Karpenter for autoscaling nodes. The cluster scales up based on incoming Pod Requests using Spot Instances and scales down as Pods idle. The cluster reduces the friction for a data scientist to run locally and in the cloud.
- Designed and built a model monitoring dashboard with Snowflake and Looker to evaluate our ASR models daily. This provides a daily metric for anyone in the company to view our model's accuracy and performance.
- Kept the research team's DevOps practices aligned with the broader engineering organization. Imported all of the team's AWS infrastructure into Terraform code as the entire engineering organization moved towards infrastructure as code.

- Maintained a relationship with AWS and other vendors to get early access to products and steer their roadmap for Rev's needs. Added the Rev logo to AWS FSx for OpenZFS customer references. https://aws.amazon.com/fsx/openzfs/customers/
- Increased productivity and decreased overall onboarding time and effort by creating and clearly documenting the steps for a self-serve process for new hires. New hires onboard themselves.

Vectra AI (AI/ML Network Security), Austin, TX Aug 2017 - Dec 2020
Platform Software Engineer

- Vectra AI applies artificial intelligence to detect and respond to cyberattacks in real time. My roles included packaging internal software for production upgrades, monitoring customer appliances for failing services or failing hardware, taking ownership of legacy internal services, bootstrapping AWS accounts for data engineering and bootstrapping Kubernetes with AWS EKS.
- Worked on diagnosing customer appliance defects and failures. This involved developing an internal tool for system health monitoring. This tool is widely used by customer support and other engineering teams to help identify system health issues faster.
- Rewrote the end-to-end test framework, in Python with Pytest, which saved developers time and improved overall test coverage. Refactored our PXE boot scripts with Ansible to provide more consistent and reliable configuration during manufacturing.

Uhnder (Automotive Radar), Austin, TX Dec 2016 - Aug 2017
Software Engineer

- Uhnder develops radar chips for automotive applications. My roles included developing the test automation framework, Linux system administration, developing internal tools to enable developer productivity.
- Deployed Jenkins to test software with continuous integration. Developed an internal Jenkins library for creating jobs and composing those jobs into pipelines.
- Designed a network-controlled Raspberry Pi + electric relay in Python in order to automate testing for the product against physical instruments. This enabled developers to automate their instrumentation.

EpochLabs (Distributed Database), Austin, TX June 2016 - August 2016
Software Developer Intern

- Contributed to a distributed database for Linux and written in C++. Scripts and other non-product code were written in Python and Bash.
- Benchmarked latency and throughput of other distributed databases such as Riak, Cassandra and Scylla using Basho Bench and YCSB benchmarking tools. Wrote custom scripts to automatically deploy cluster configurations to multiple machines.

Applied Research Laboratories, Austin, TX Oct 2014 - Mar 2016, August 2016 - Dec 2016
Student Research Assistant

- Lidar point cloud and digital elevation model processing toolbox application written in C++11 and used Qt for the GUI interface. The toolbox features easy-to-use automatic classification options and provides a visualizer for point clouds and digital elevation models.
- Refactored core interfaces with more efficient algorithms. New code was continuously tested with a continuous integration server to ensure high test code coverage. Additionally developed internal tools for other developers.

Indiana University, Bloomington, IN June 2013 - July 2013
Summer Research Intern

- Prototyped an additional approach to 3D point cloud segmentation. The approach uses surrounding neighbor classifications to produce more cohesive segmentations.
- Developed using Python with ROS (robot operating system) and sklearn to improve the overall accuracy of classifying 3D point clouds.
- Experimented by using different probabilistic graphical models, machine learning, classification and optimization techniques.

PUBLIC

See https://vicyap.com/public for links.

- One Cluster to Rule Them All - ML on the Cloud Using Ray on Kubernetes and AWS | MLOps World 2022
- Contributions to hashicorp/terraform-provider-aws

EDUCATION

Bachelor of Science in Computer Science Dec 2016
Bachelor of Science in Electrical Engineering
The University of Texas at Austin, Austin, TX