Spot On!

A tale of how Lyft saved ⅔ of their cloud computing costs through AWS EC2 Spot Instances.

Deepshika Dhanasekar

About Me

Senior at U.C. Berkeley

EECS & Business Administration

Currently: Working with Bill Allison (CTO of U.C. Berkeley)

Prev: EECS Course Staff, Intern @ Lyft Level 5, BlackRock, Apple

Case Study

Level 5

Lyft Level 5

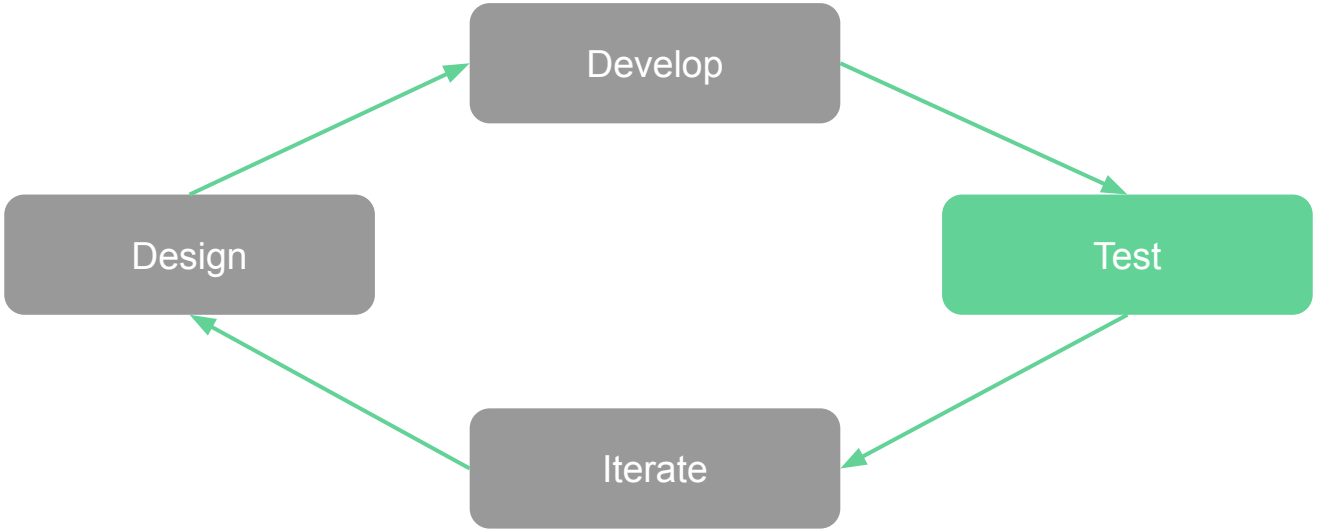
Autonomous Vehicles

Along with its focus on shared rides, Lyft launched its autonomous vehicles division in 2017.

Value Proposition: Lyft's access to large amounts of ride data will give it's self-driving technology a competitive advantage.

AV Self Driving Technology

...is expensive to develop and test!



AV Simulations

> Using petabytes of data gathered from its AV & rideshare fleet, Lyft's engineers run millions of simulations to improve the performance and safety of its self-driving technology before testing in the real world.

Current Architecture

These simulations are run on an **internal web application** that ingests data collected from test drives and recreates the test drive conditions. The web application runs as a **microservice using AWS.**

Problem

All of these simulations are computationally intensive and could **cost thousands of dollars.**

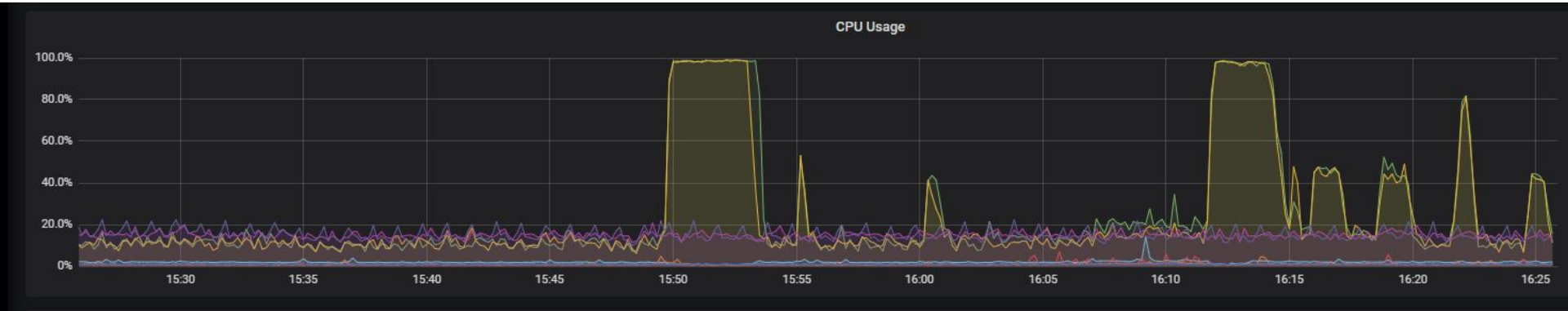
Solution

A combination of Amazon **Elastic Compute Cloud** Spot Instances and Amazon Elastic **Kubernetes** Service.

Unlike Lyft's rideshare business, which has consistent and uniform demand throughout the day, Level 5 has different compute constraints...

1. Need to service large, batch-style workloads with very spiky profiles

2. Need the ability to burst up to high peak loads and then turn everything down when its not being used



Enabling Simulations to Run Efficiently

Problems

1. Spot Instances may not be available because of high demand.
2. Available spot instances may not be the instance typically used for the simulation engine.

Engineering Changes

1. EKS

Used EKS to prioritize and scale resource pools so jobs were efficiently using instances, taking into account regional zone usage.

2. Fleet Diversity

Ensure the simulation stack works on whichever type of instance is currently available.

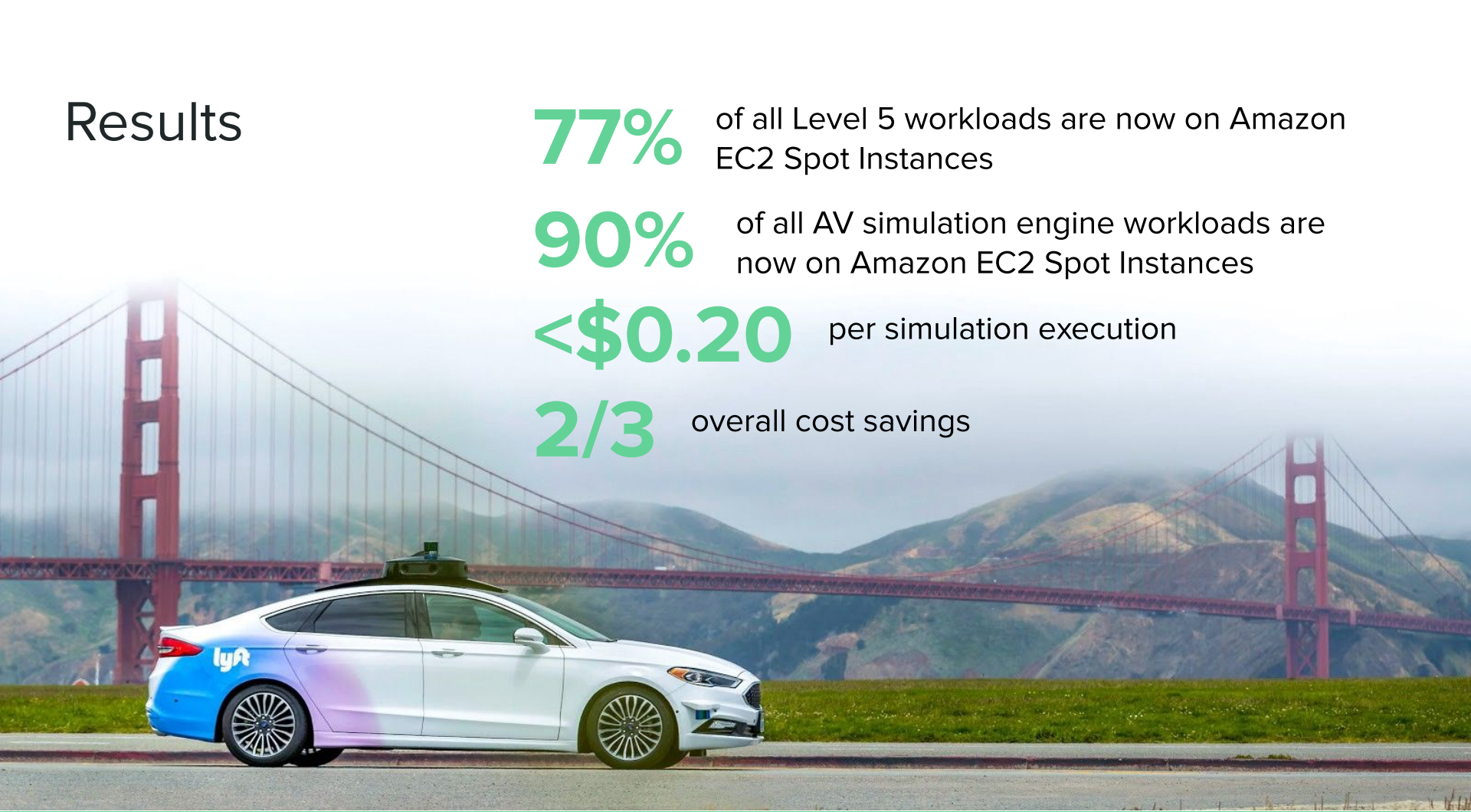
Results

77% of all Level 5 workloads are now on Amazon EC2 Spot Instances

90% of all AV simulation engine workloads are now on Amazon EC2 Spot Instances

<$0.20 per simulation execution

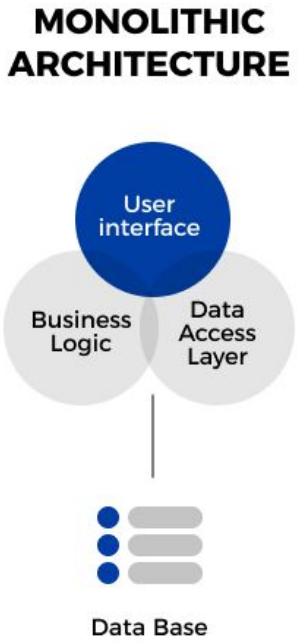
2/3 overall cost savings

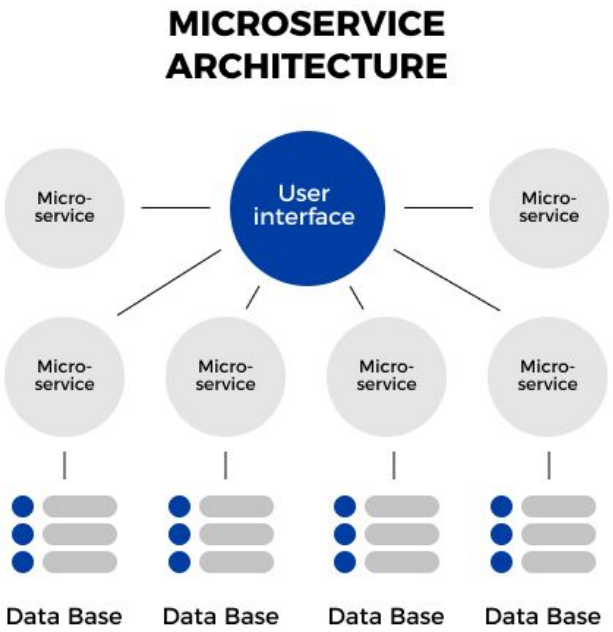


Questions?

My contact info:
Deepshika Dhanasekar
ddhanasekar@berkeley.edu

My first foray into cloud computing was understanding ...





...monolithic vs microservice architectures.

From Virtualization to Containerization

Virtualization

Use Virtual Machines (VM) to isolate microservices from each other. The host operating system (Hypervisor) provides virtual CPU, memory, & other resources.

Disadvantages

- VMs work as if they are running on physical hardware
- Provisioning of resources takes time
- System resources are used inefficiently

Containerization

Virtualize the operating system, rather than the hardware. Provide a virtual OS to each individual microservice, i.e. in its own container.

Advantages

- Run multiple microservices alongside each other
- Impose dynamic limitations on their resource utilization

Commonly used tool in industry: **Docker**

Managing Containers

Kubernetes: *"...an open-source system for automating deployment, scaling, and management of containerized applications."*

Cluster: the entire Kubernetes instance

Pod: the basic unit of deployment that contains all the containers that need to coexist together

Node: the individual VMs that the cluster uses to run the pods.

- **Master Node:** controls the scheduling of pods across the worker nodes. Ensures that the desired state of the cluster is maintained.
- **Worker Node:** where the application actually runs. Has individual network proxies to communicate within the cluster and expose the application to the public.

Amazon Elastic Compute Cloud (EC2)

Provides scalable computing capacity in the Amazon Web Services (AWS) cloud, eliminating the need to invest in hardware. Use EC2 to scale up or down your compute instances to handle changes in requirements or traffic.

Amazon EC2 Instances