

How to build scalable web based applications for less with Amazon EC2 Spot Instances

Chad Schmutzer, Principal Developer Advocate, EC2 Spot

April 2020

About me

Chad Schmutzer / schmutze@amazon.com / [@schmutze](https://twitter.com/schmutze)

Principal Developer Advocate – EC2 Spot
Pasadena, CA

Maintain [Jenkins EC2 Spot Fleet plugin](#) and [EC2 Spot labs on GitHub](#)

Previously

- Specialist Solutions Architect – EC2 Spot (~2.5 years)
- Solutions Architect – World Wide Public Sector (~3 years)
- Associate Director for Academic Research Computing at the California Institute of Technology in Pasadena, CA (~15 years)

Little league coach

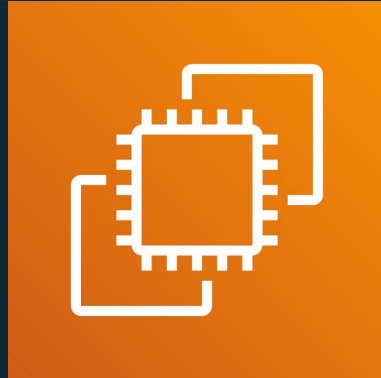


Agenda

- EC2 Spot Instances - Key concepts to use EC2 Spot successfully
- Demo - Let's build a web app using EC2 Spot
- Q&A

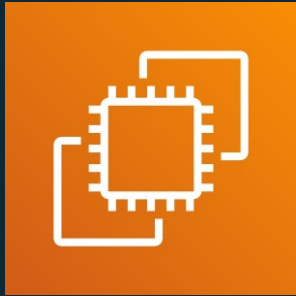
EC2 Spot Instances - Key concepts to use EC2 Spot successfully

At first there was Amazon Elastic Compute Cloud (Amazon EC2)

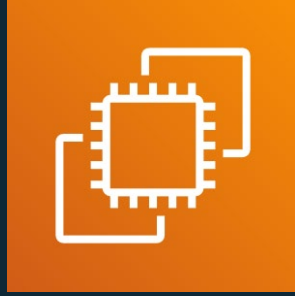


m1.small

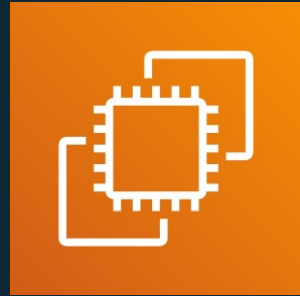
Then we added some new instance types



m1.small

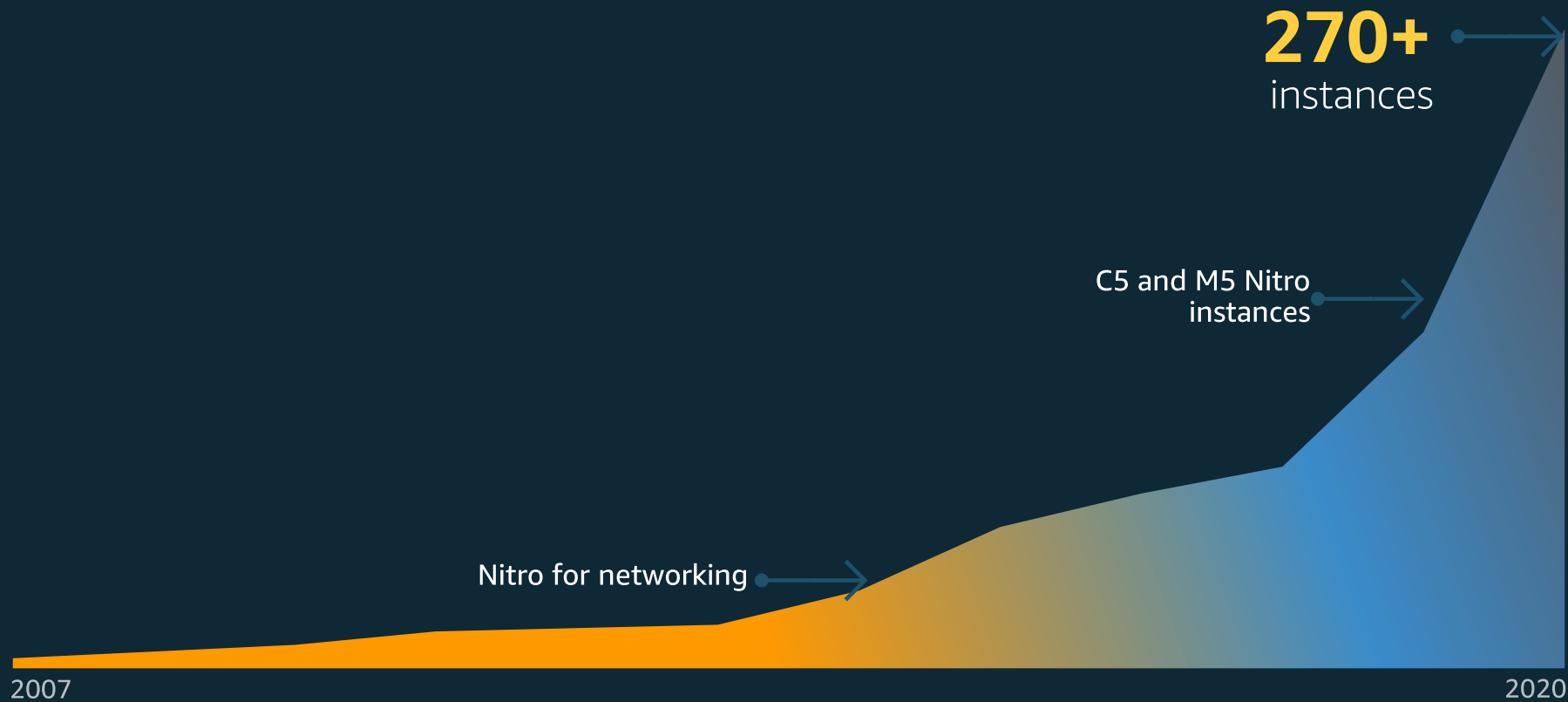


m1.large



m1.xlarge

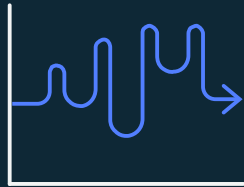
Then we added a lot more instance types



Amazon EC2 purchase options

On-Demand

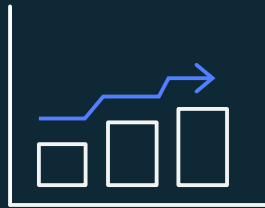
Pay-for-compute capacity
by the second with no
long-term commitments



Spiky workloads,
to define needs

Savings Plans & Reserved Instances

Make a commitment and receive a
significant discount off compute



Committed &
steady-state usage

Spot Instances

Spare EC2 capacity at
savings of up to 90%
off On-Demand prices



Fault-tolerant, flexible,
stateless workloads

What do I need to know about Spot Instances?



Spot is spare capacity

Same infrastructure as
On-Demand



Spot pricing

Smooth, infrequent
changes, no spikes,
more predictable



Interruptions

Spot is an interruptible
product. Interruptions happen
as there are shifts in
supply and demand

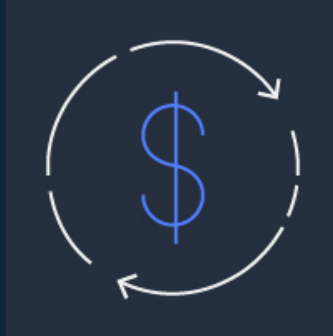


Be flexible

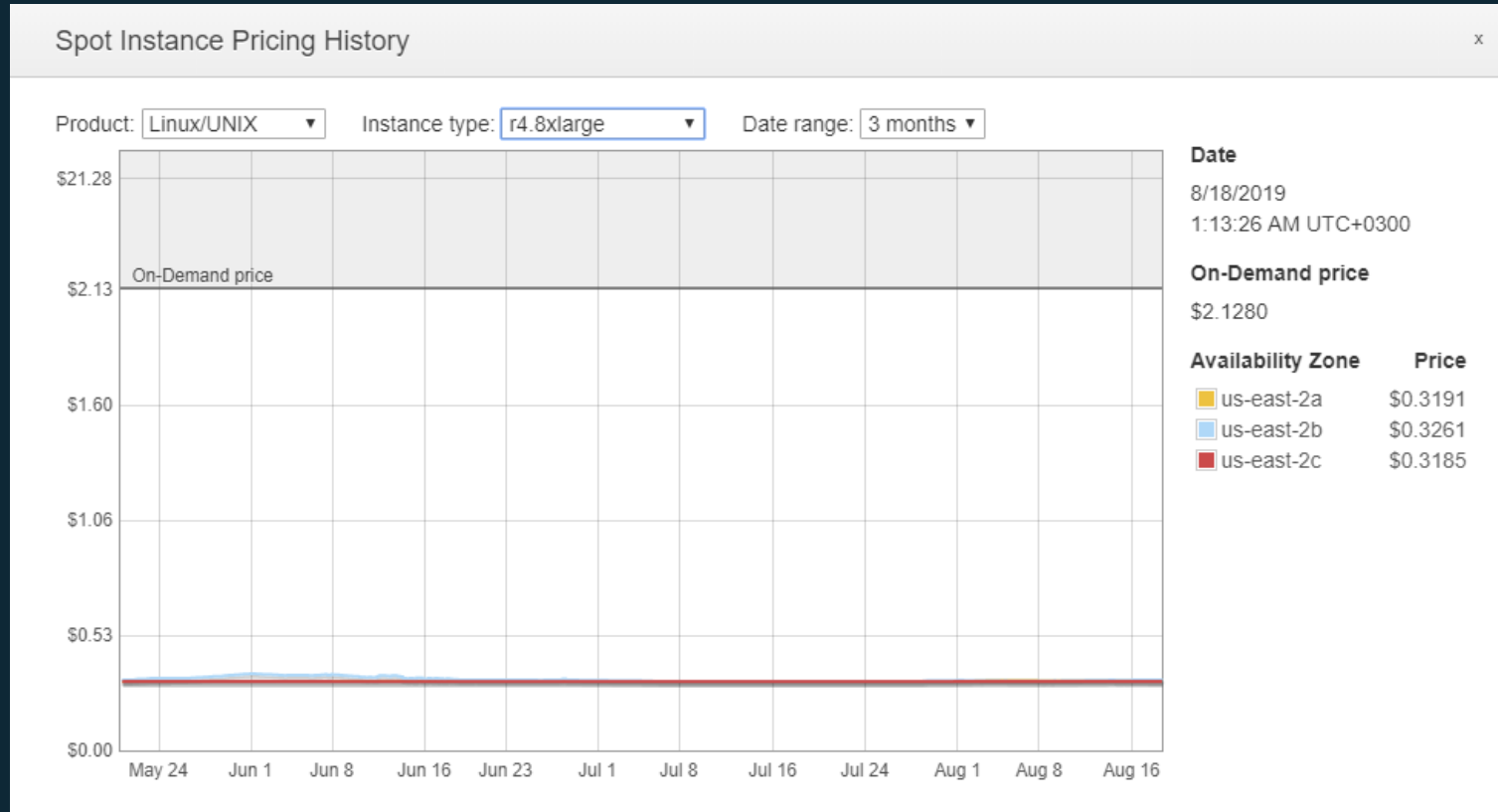
Choose multiple instance
types, sizes, and
Availability Zones

How much spare capacity is there?

"On average, every week, AWS customers are using more compute capacity on Amazon EC2 Spot Instances than customers were running across all of Amazon EC2 in 2014."



Pricing model



EC2 capacity pools - instance type flexibility

C4	1a	1b	1c	On-Demand
8XL	\$0.28	\$0.27	\$0.29	\$1.76
4XL	\$0.21	\$0.19	\$0.16	\$0.88
2XL	\$0.08	\$0.07	\$0.08	\$0.44
XL	\$0.04	\$0.05	\$0.04	\$0.22
L	\$0.01	\$0.01	\$0.02	\$0.11

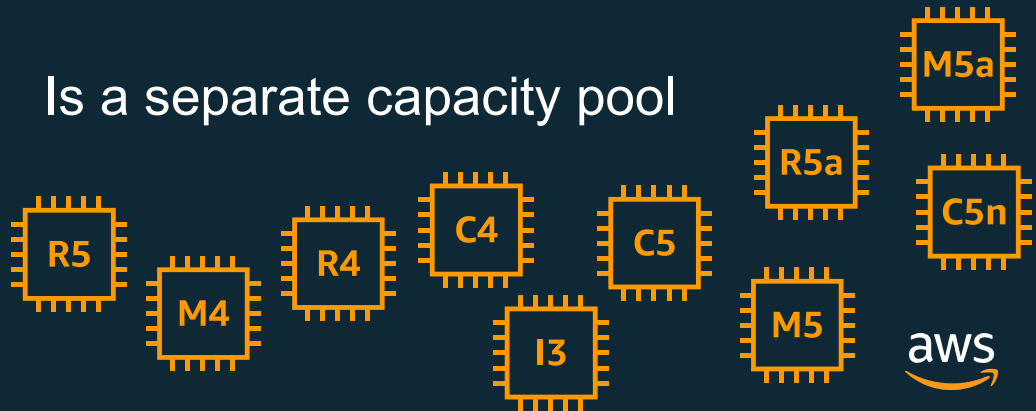
Each instance family

Each instance size

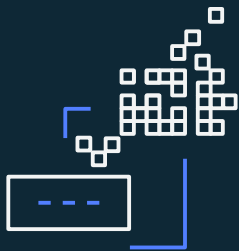
Each Availability Zone

In every Region

Is a separate capacity pool



Getting started with Spot



Big data



CI/CD



Web services



High performance computing



Or any containerized workload

Spot is ideal for:

- Fault-tolerant
- Flexible
- Loosely coupled
- Stateless workloads

High level modernization requirements

Design your application or workload to:

1. Be completely instance type flexible
2. Become completely interruption tolerant
3. Utilize dynamic scaling policies

Low level technical modernization requirements

1. Adopt EC2 Launch Templates
2. Use EC2 Auto Scaling
 1. Use mixed instances group policy
 2. Set a baseline of On-Demand capacity and tie that to Savings Plans
 3. Use as many instance types as possible
 4. Use the capacity optimized allocation strategy
 5. Enable dynamic scaling policies
 1. Target tracking
 2. Predictive scaling
3. Use an Application Load Balancer

Save up to 90% using EC2 Auto Scaling

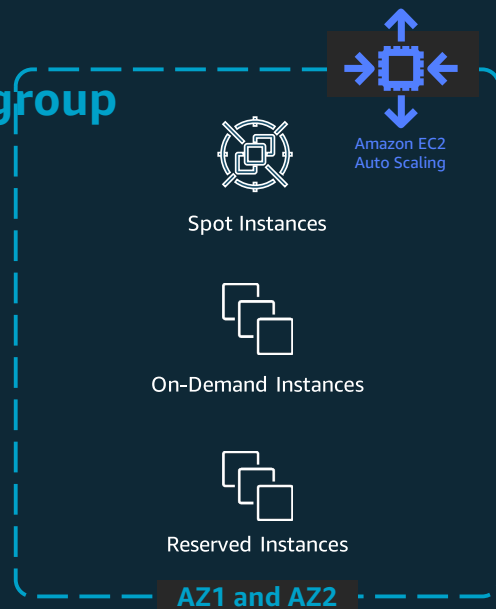
Automatically scale instances across instance families and purchase models in a single Auto Scaling group

- Capacity optimized



- Lowest cost

- Prioritized list

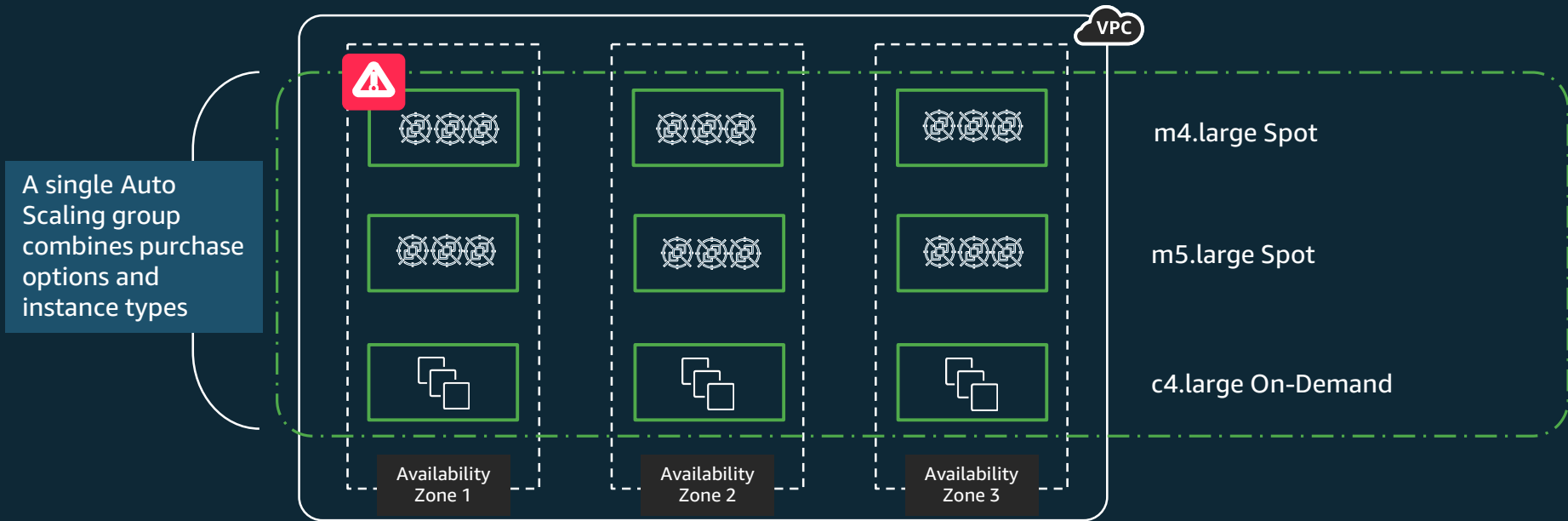


Reduce cost

Optimize performance

Eliminate operational overhead

EC2 Auto Scaling – Mixed Instances Policy



Demo

Q&A

Thank you!

schmutze@amazon.com / [@schmutze](#)

- <https://aws.amazon.com/ec2/spot/>
- <https://aws.amazon.com/ec2/autoscaling/>
- <https://github.com/aws-labs/ec2-spot-labs>
- <https://ec2spotworkshops.com/>