
aws
INNOVATE
ONLINE CONFERENCE

aws INNOVATE
ONLINE CONFERENCE

Cost-optimize Amazon EC2 workloads at scale

Boyd McGeachie
Senior Product Manager – Technical,
EC2 Pricing & Launch Services, AWS

Chad Schmutzer
Principal Developer Advocate, Spot,
EC2 Pricing & Launch Services, AWS

Learning objectives

- 300 level

- Answer questions such as:
  - “How can I scale my compute cost effectively?”
  - “How can I reduce my compute spend?”
  - “What are Amazon EC2 Spot Instances?”
  - ”How can I mix On-Demand and Spot Instances?”
  - “How can I mix multiple Amazon EC2 instance types in a single EC2 Auto Scaling group?”

Agenda

- Intro
- Amazon EC2 launch templates / Demo
- Amazon EC2 Fleet / Demo
- Amazon EC2 Auto Scaling groups / Demo
- Q&A

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

m1.small

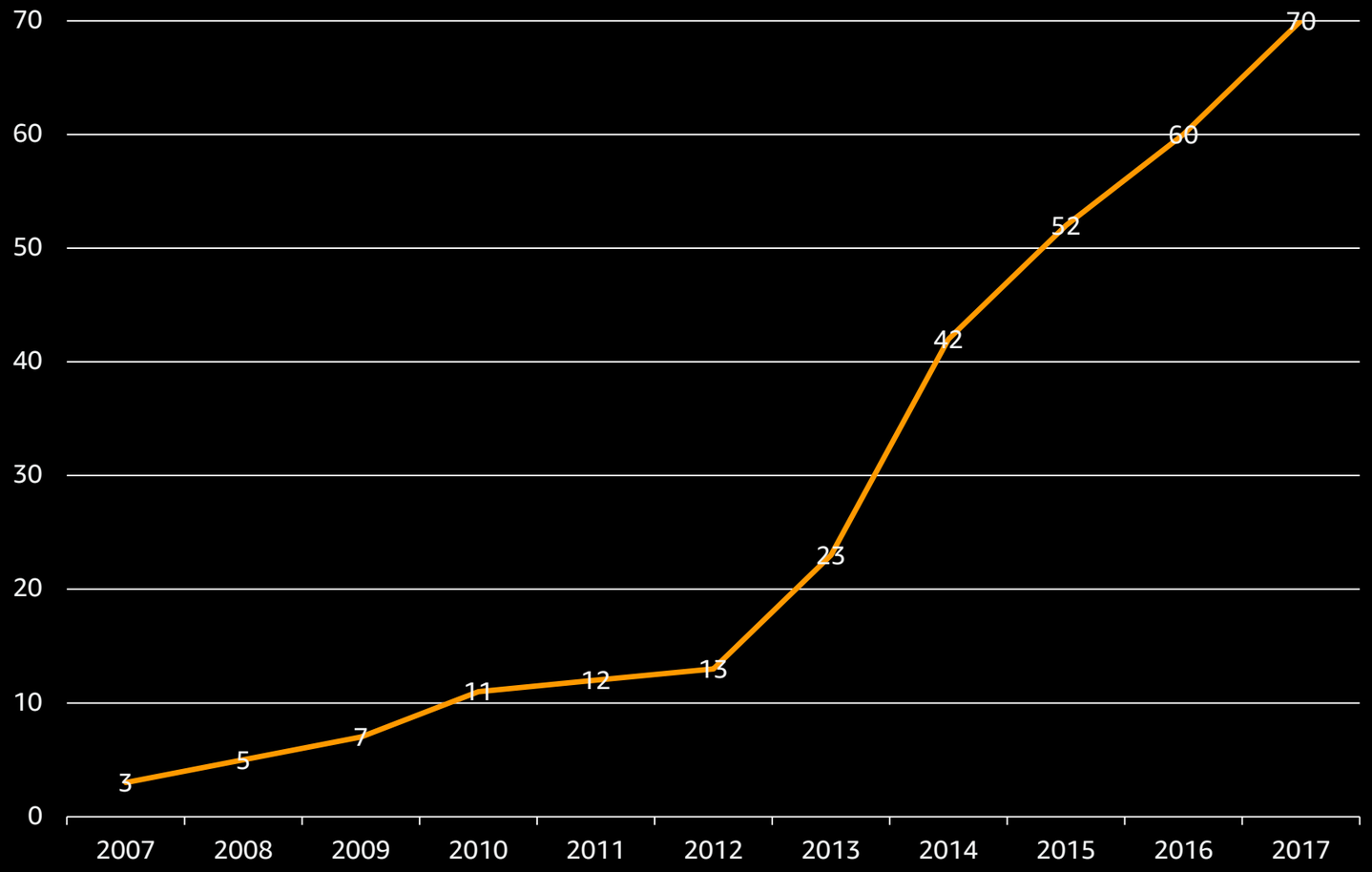
Then we added some new instance types

m1.small

m1.medium

m1.large

Then we added a lot more instance types



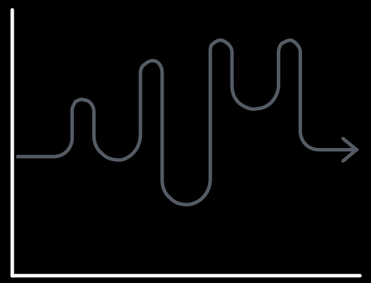
Recent launches:

Compute-optimized: c5{d,n}
General-purpose: m5{d,a}, t3, a1
Accelerated computing: g3s, g4
Memory-optimized: z1d, r5{d,a}

Amazon EC2 purchasing options

On-Demand

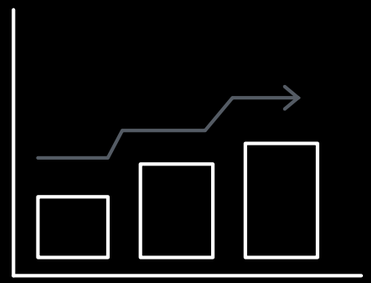
Pay for compute capacity by **the second** with no long-term commitments



Spiky workloads, to define needs

Reserved Instances

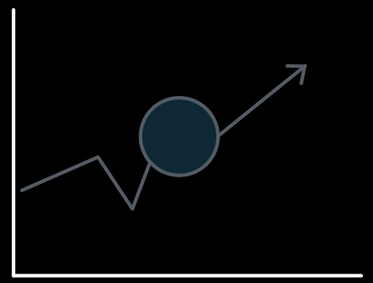
Make a 1- or 3-year commitment and receive a **significant discount** off On-Demand prices



Committed & steady-state usage

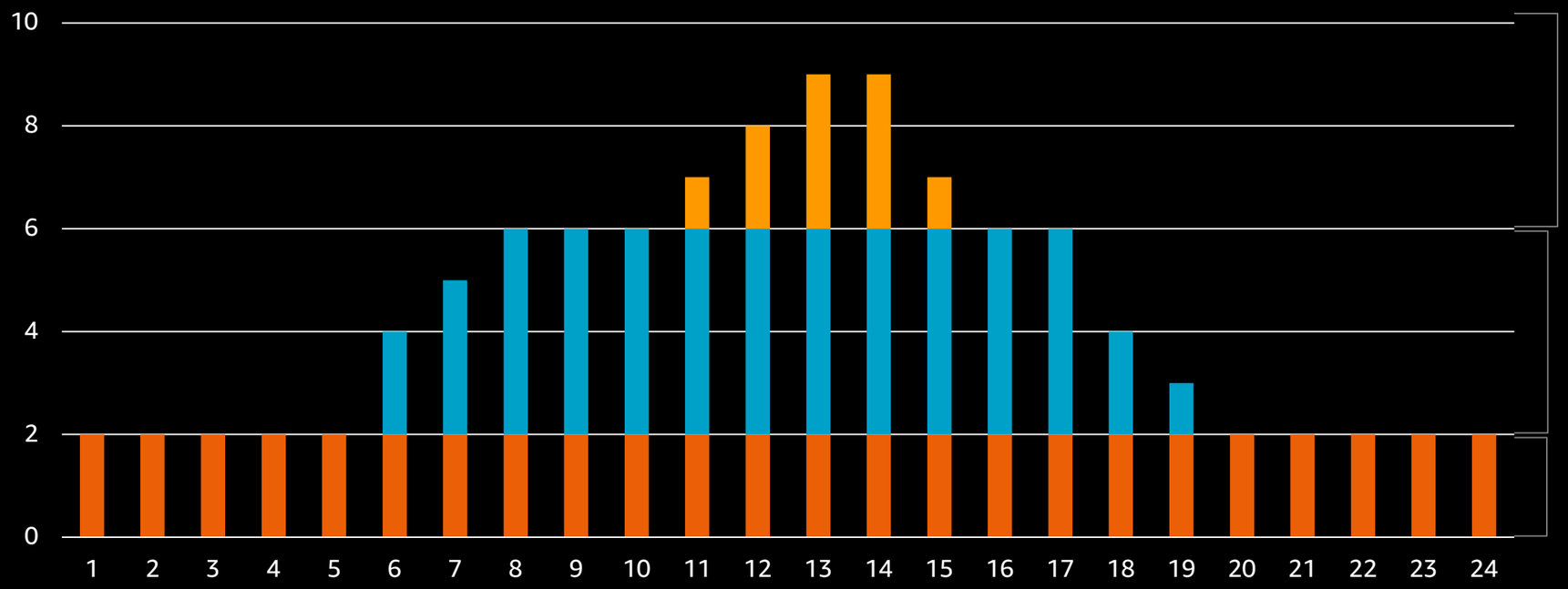
Spot Instances

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



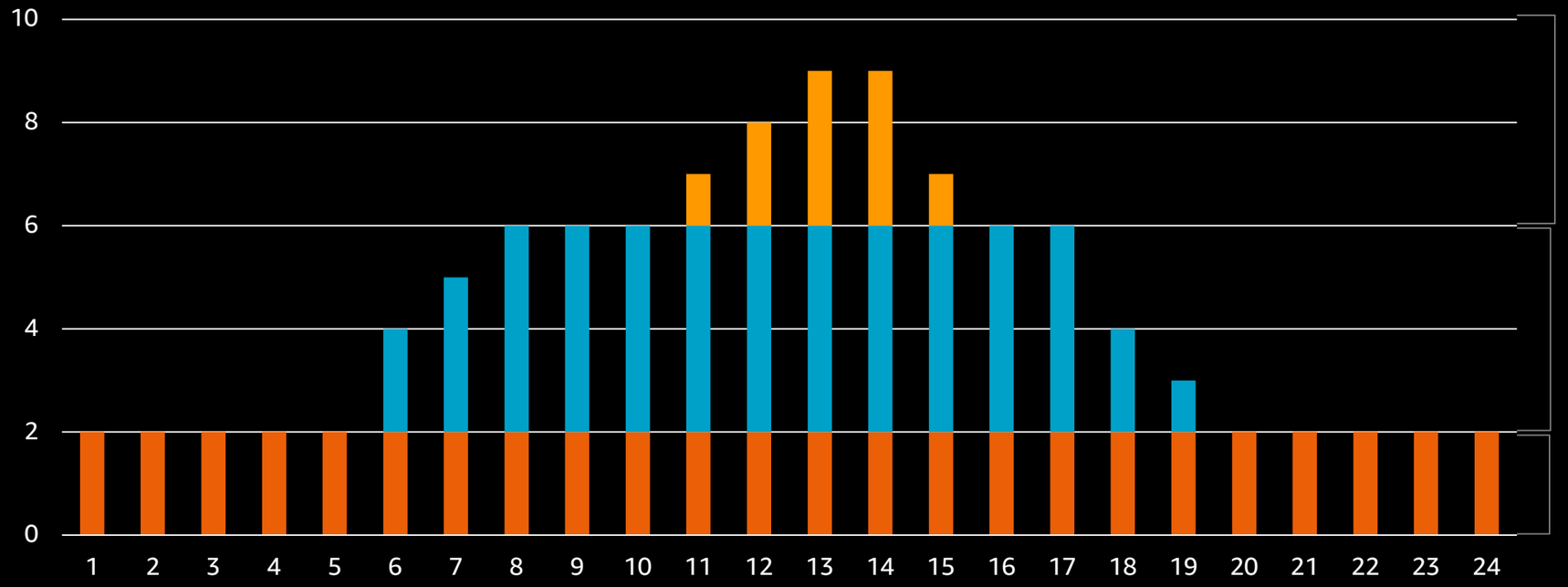
Fault-tolerant, flexible, stateless workloads

Combine purchase options to optimize at scale



Use RIs for known/steady-state workloads

Combine purchase options to optimize at scale

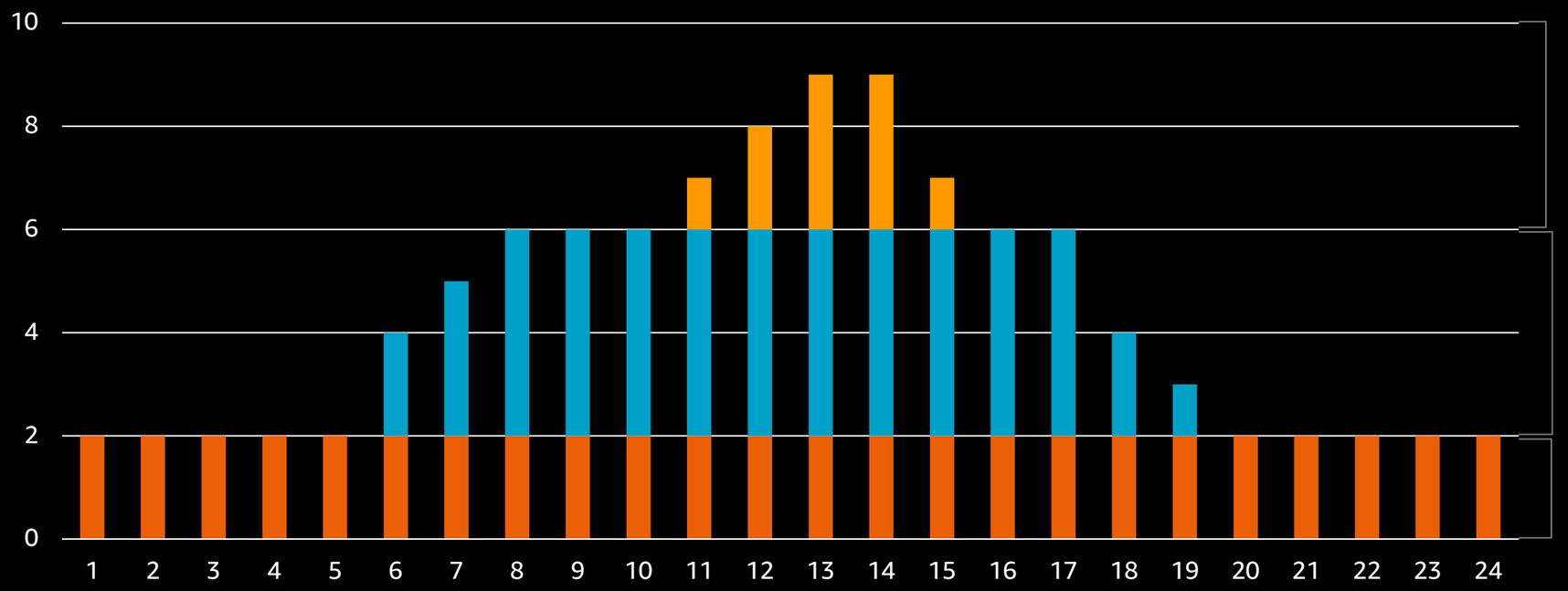


Scale using Spot

On-Demand, or both

Use RIs for known/steady-state workloads

Combine purchase options to optimize at scale



Scale using Spot

On-Demand, or both

Use RIs for known/steady-state workloads

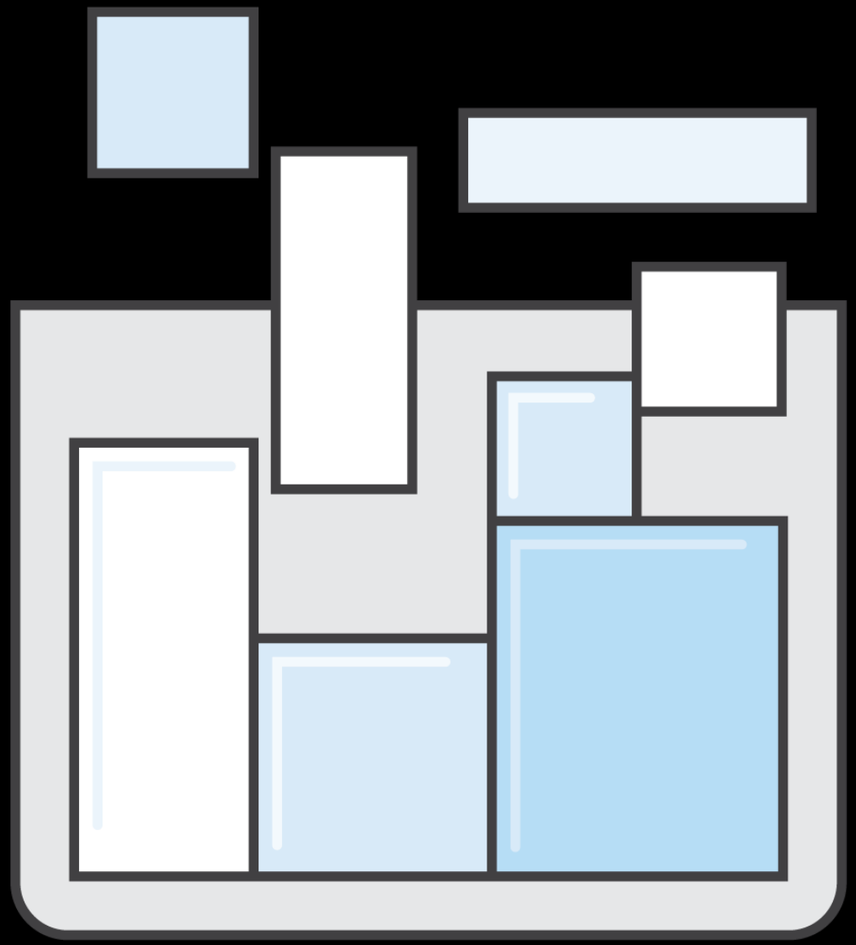
On-Demand capacity reservations for your reservation needs

Why combine instances and purchase models?

To turbo-boost an application using Spot Instances

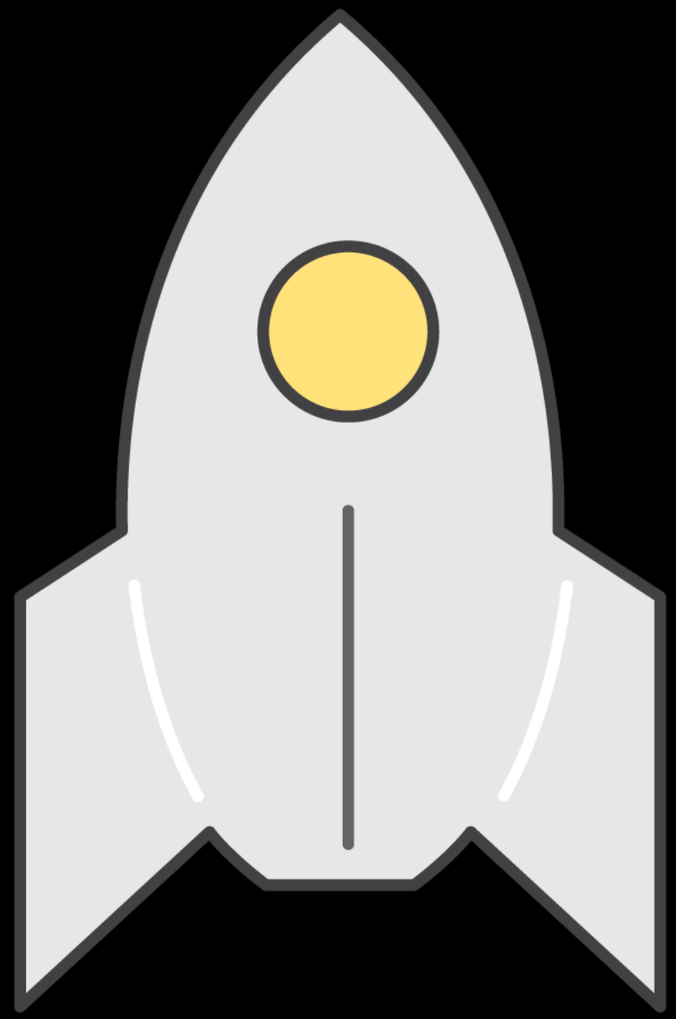
Why combine instances and purchase models?

To scale on vCPUs, memory, or containers



Why combine instances and purchase models?

To scale 1,000x



Which customers are combining instances and purchase models, and with what workloads?

aws INNOVATE
ONLINE CONFERENCE

Containers + Spot = Match made in heaven

- ✓ Containers are often stateless, fault-tolerant, and a great fit for Spot Instances
- ✓ Deploy containerized workloads and easily manage clusters at any scale at a fraction of the cost with Spot Instances
- ✓ Spot Instances can be used with Amazon Elastic Container Service (Amazon ECS) or Kubernetes to run any containerized workload

skyscanner

Skyscanner is a travel fare aggregator website and travel metasearch engine based in Edinburgh, Scotland

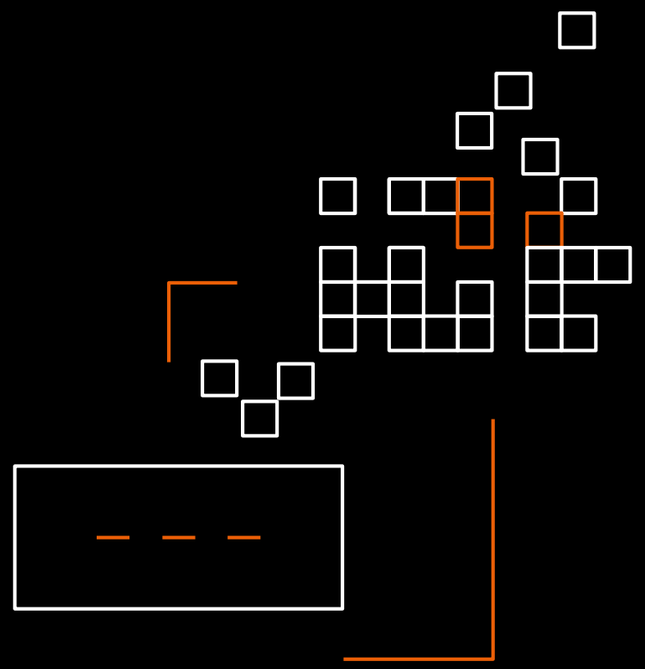
"We are currently tracking **74% savings** over all regions."

Paul Gillespie
Principal Architect/Tribe Lead

lyft yelp MapBox Expedia Caltech

aws | intel

Workload example: Big data



Aol. Zillow FINRA

Amazon EMR hadoop APACHE Spark

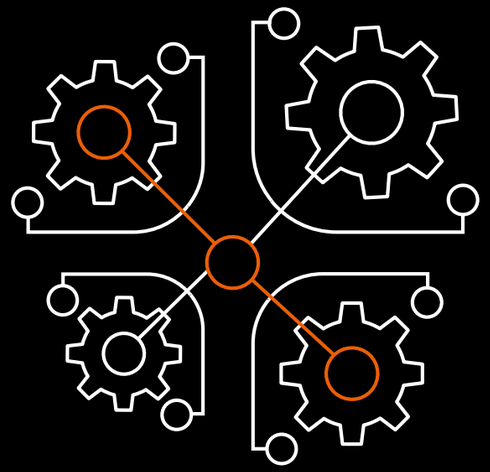
- ✓ Spot Instances provide acceleration, scale, and cost savings to run hyperscale workloads for data analysis
- ✓ Scale to large numbers of parallel nodes via Amazon EC2 Fleet
- ✓ Use Spot Instances with Amazon EMR, Hadoop, or Spark to process massive amounts of data

"A job that took weeks in our data center, due to limited resources, took hours on Spot thanks to the great parallelism, at a very cost-efficient price."

Shay Asoolin
Sr. Director, Development Infrastructure, Mobileye

aws | intel

Workload example: CI/CD



sprinklr

zuora

revcontent

Jenkins

Bamboo

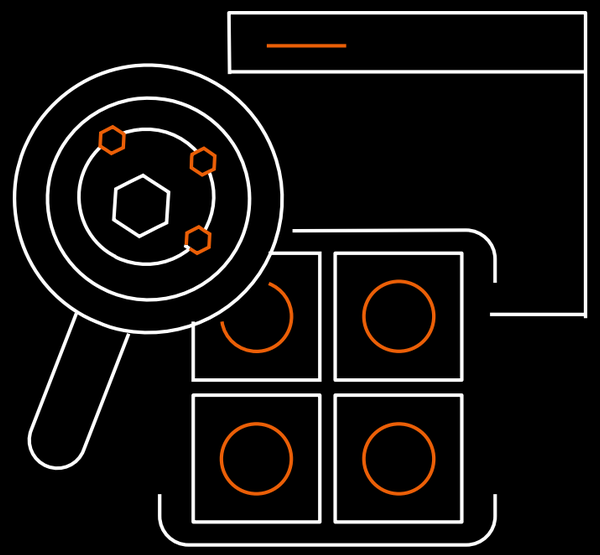
- ✓ Configure Jenkins with the EC2 Spot plugin to automatically scale a fleet of Spot Instances based on the number of CI/CD jobs
- ✓ Increase cost savings by leveraging older generation instances for CI, as these processes do not require a lot of power for testing

"By using [Amazon EC2] Spot instances, we've been able to save up to 75 percent a month simply by changing four lines of code. It makes perfect sense for saving money when you're running continuous integration workloads or pipeline processing."

Matthew Leventi
Lead Engineer, Lyft

aws | intel

Workload example: Web services



NetSeer

codewise

MirrorWeb

Amazon EC2 Auto Scaling

Amazon EC2 Fleet

- ✓ Scale, throughput, and deep cost savings for large-scale web operations
- ✓ Launch and manage a collection of diversified Spot Instances across pools via EC2 Fleet and Auto Scaling group
- ✓ NEW! Include Spot with RIs and On-Demand in a single Auto Scaling group

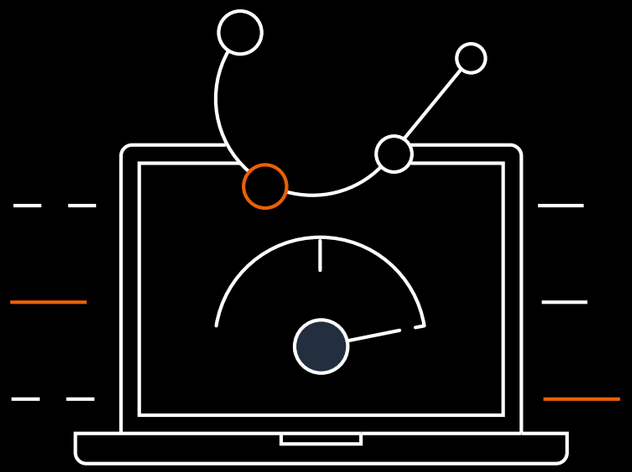
Quantcast scales ad services, saves 60% using Amazon EC2 Spot Instances

"As we roll out more infrastructure to AWS, Amazon EC2 Spot Instances are helping us control costs and scale our systems to meet demand."

Leah Blank
Senior Systems Engineer, Quantcast

aws | intel

Workload example: HPC



AWS Batch

AWS CloudFormation

Amazon EC2 Fleet

- ✓ Accelerate HPC workloads such as genomic sequencing, CFD, and algorithmic trading by running massively parallel jobs
- ✓ Run multiple projects simultaneously; launch & decommission 1,000s of nodes
- ✓ Spot Auto Scaling groups; F1(FPGA), eg1 (Elastic GPUs), cluster GPU instances to accelerate processing

Illumina saves nearly $400,000 monthly, speeds Genomics Analysis using Spot Instances

"We are able to offer our customers a lower-cost, high-performance genomic-analysis platform, which can help them speed their time to answers."

Andy Nelson
Associate Director, Informatics & Cloud Operations, Illumina

The tools

aws INNOVATE
ONLINE CONFERENCE

Automate cost optimization & capacity management

Amazon EC2 launch templates

Amazon EC2 Fleet

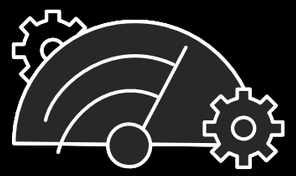
Amazon EC2 Auto Scaling

… Let’s see how this all works together to automatically optimize scale, performance, and cost behind the scenes

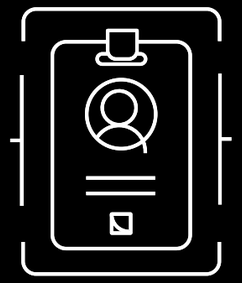
Amazon EC2 launch templates

aws INNOVATE
ONLINE CONFERENCE

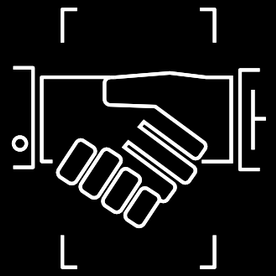
Use launch templates to achieve …



Increased productivity



Simplified permissions

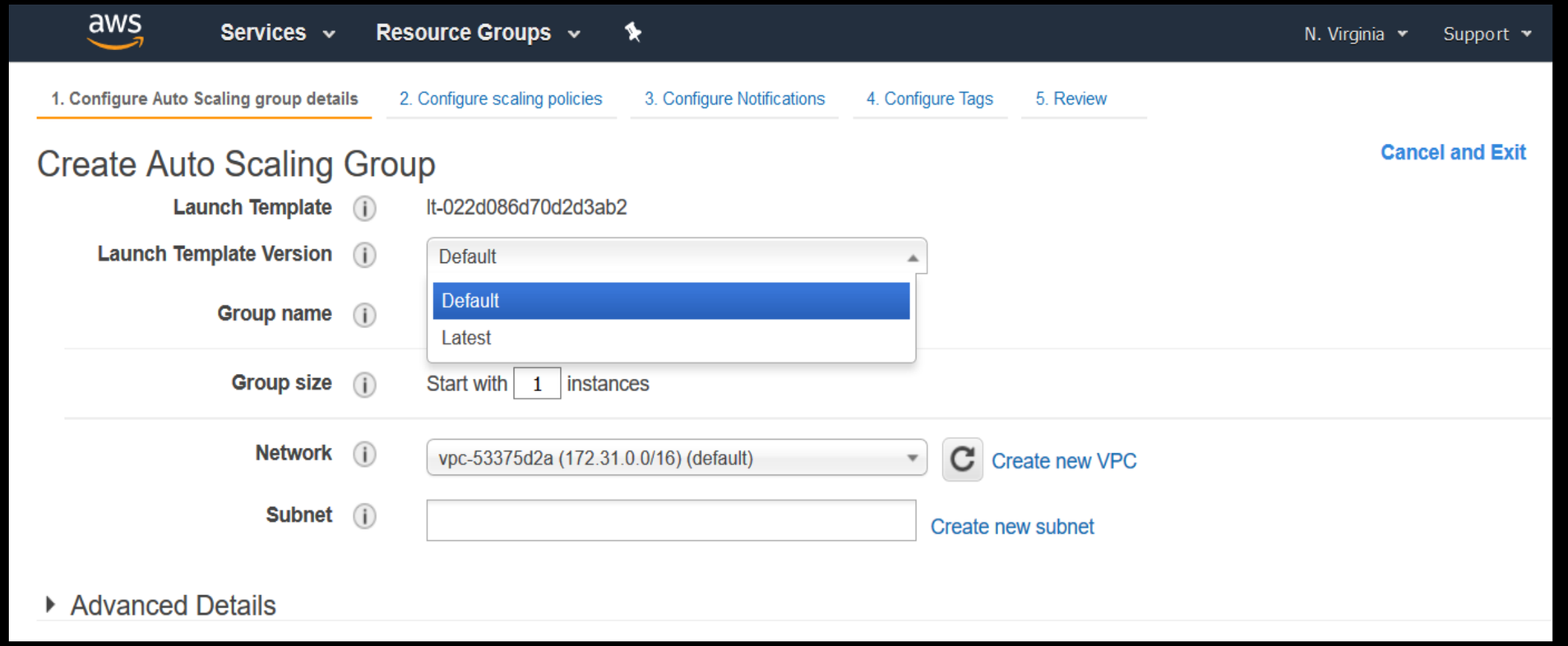


Governance & best practices

Consistent experience

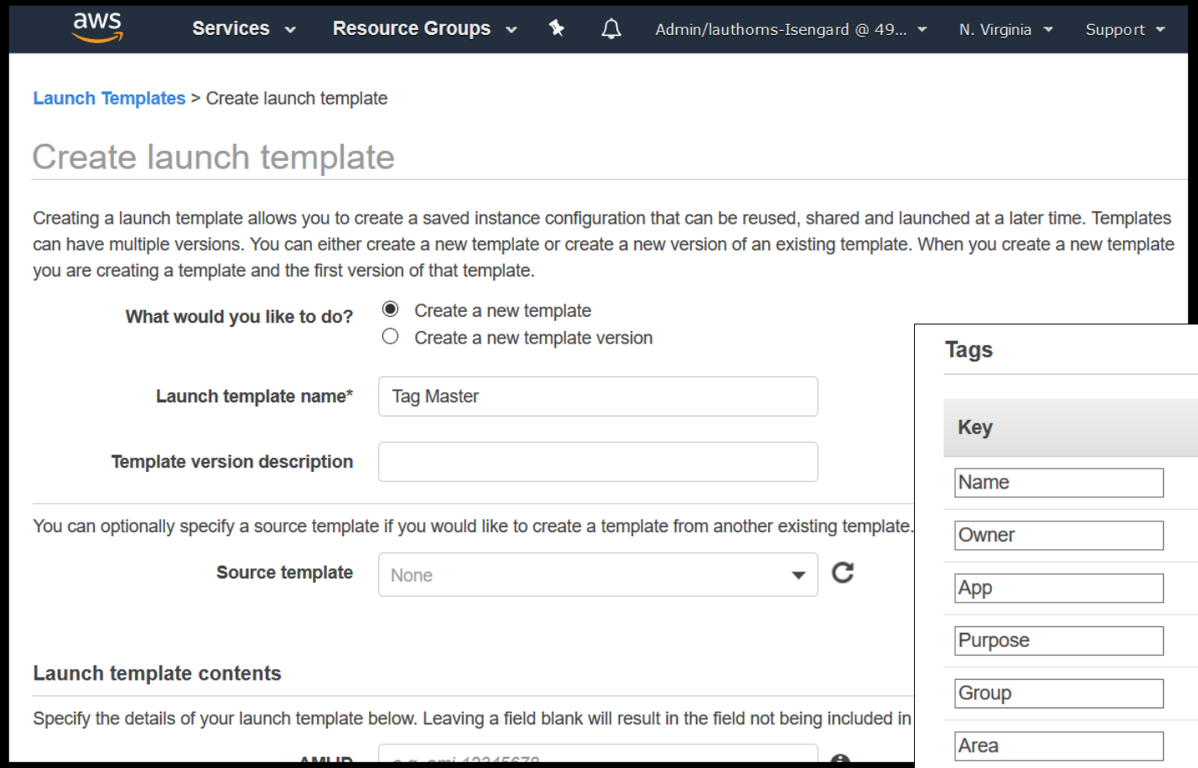
Increased productivity: Automated updates

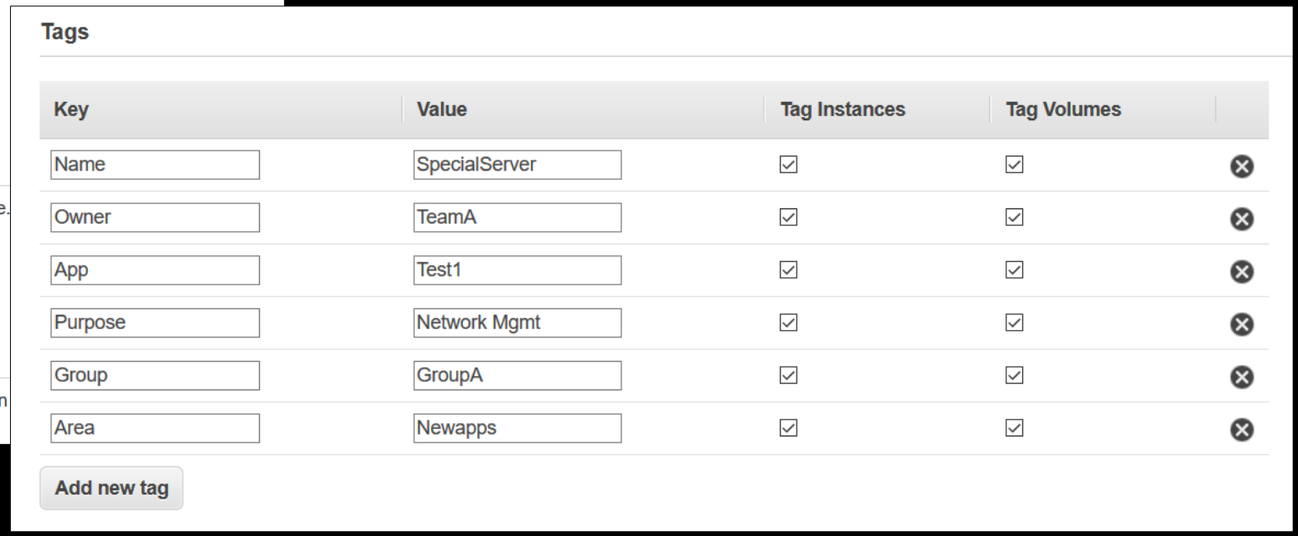
For example, push a patched AMI to EC2 Auto Scaling groups



Increased productivity: Eliminate repetitive tasks

For example, save tags in a launch template





Use launch templates as an Auth vehicle

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
         "ec2:RunInstances"
      ],
            "Resource": [
         "arn:aws:ec2:us-east-1::image/*",
         "arn:aws:ec2:us-east-1:1234567890:subnet/*",
         "arn:aws:ec2:us-east-1:1234567890:network-interface/*",
         "arn:aws:ec2:us-east-1:1234567890:security-group/*",
         "arn:aws:ec2:us-east-1:1234567890:key-pair/*",
         "arn:aws:ec2:us-east-1:1234567890:instance/*",
         "arn:ec2:ec2:us-east-1:1234567890:snapshot",
         "arn:ec2:ec2:us-east-1:1234567890:elastic-gpu/*"
      ]
    },
    {
      "Effect": "Allow",
      "Action": [
         "ec2:RunInstances"
      ],
      "Resource": [
         "arn:aws:ec2:us-east-1:1234567890:volume/*",
      ],
      "Condition": {
         "NumericLessThan": {
            "ec2:VolumeSize" : "X"
          }
       }
    },
    {
      "Effect": "Allow",
      "Action": [
         "ec2:CreateTags"
      ],
      "Resource": "arn:aws:ec2:us-east-1:1234567890:*/*",
      "Condition": {
         "StringEquals": {
            "ec2:CreateAction" : "RunInstances"
          }
       }
    }
  ]
}
```

```
{
   "Version":"2012-10-17",
   "Statement":[
      {
         "Effect":"Allow",
         "Action":"ec2:RunInstances",
         "Resource":"*",
         "Condition":{
            "ArnLike":{
"ec2:LaunchTemplate":"arn:aws:ec2:region:acc
ount:launch-template/(* or actual template
id)"
            }
         }
      }
   ]
}
```

Amazon EC2 launch templates: Demo

aws INNOVATE
ONLINE CONFERENCE

About me

Chad Schmutzer / schmutze@amazon.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 – Worldwide Public Sector (~3 years)
  - Associate Director for Academic Research Computing at the California Institute of Technology in Pasadena, CA (~15 years)
- Little League coach

Amazon EC2 launch templates: Demo

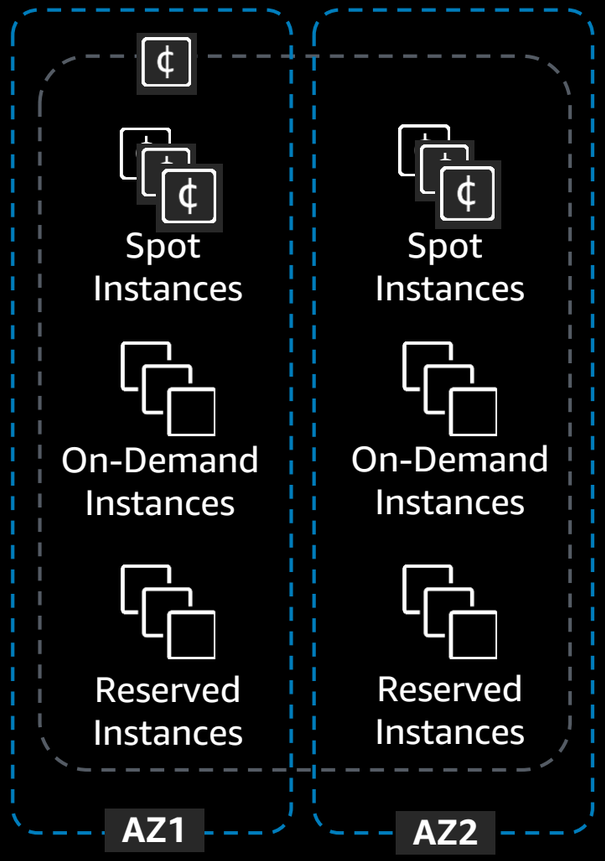
aws INNOVATE
ONLINE CONFERENCE

Amazon EC2 Fleet

aws INNOVATE
ONLINE CONFERENCE

Amazon EC2 Fleet

Simplifies provisioning of EC2 capacity across different instance types, AZs, and purchase models with a single API



Use all three purchase models to optimize costs

Automatic optimization behind the scenes with software

Benefits

Reduce costs

Increase operational efficiency

Key features

Flexible capacity allocation

Massive scale

Simplified provisioning

Amazon EC2 Fleet and allocation strategies

Amazon EC2 Fleet

Provisions capacity across multiple instance types according to allocation strategies

Allocation strategies

On-Demand prioritized list of instance types

Spot Instances across the N lowest-priced instance pools

Amazon EC2 Fleet types

request

- Places an asynchronous one-time request without maintaining capacity or submitting requests in alternative capacity pools if capacity is unavailable

maintain (default)

- Places an asynchronous request for your desired capacity, and maintains it by replenishing interrupted Spot Instances

instant (new-ish!)

- Places a synchronous one-time request, and returns errors for any instances that could not be launched

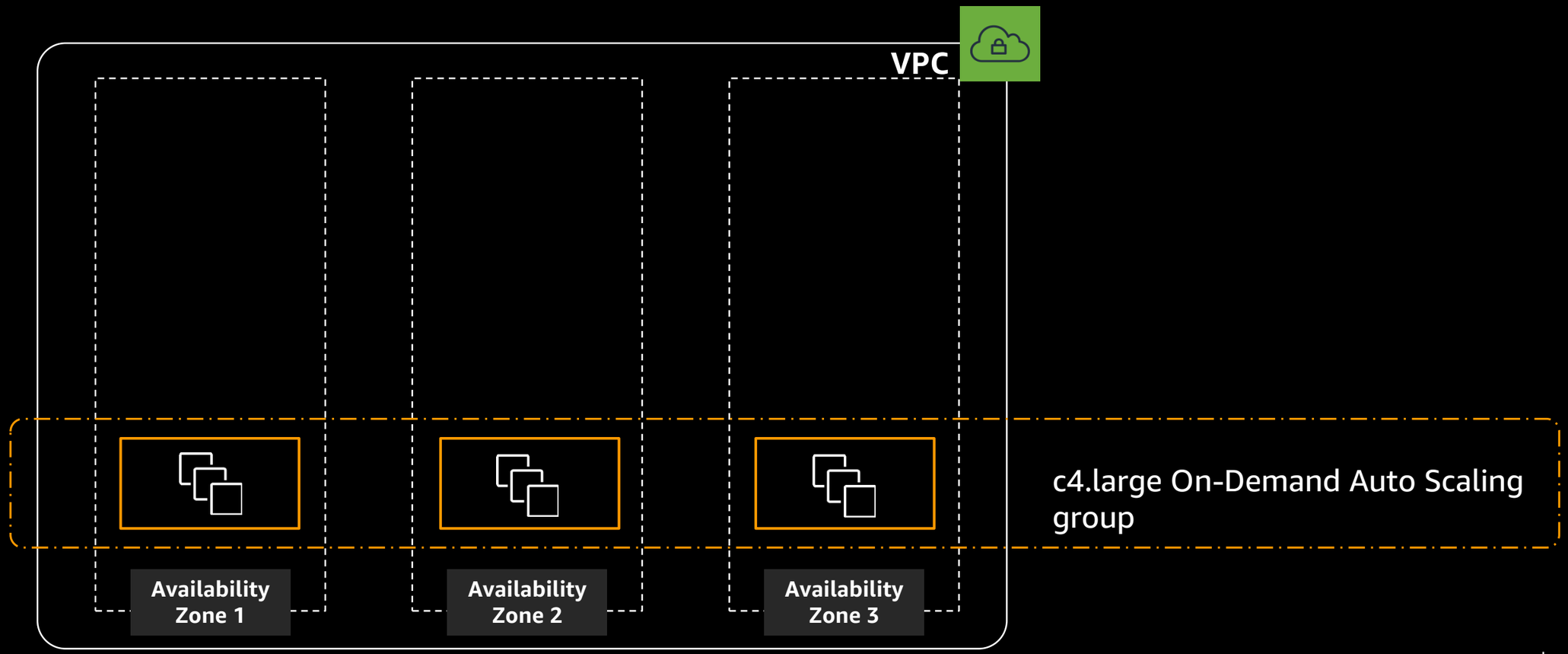
Amazon EC2 Fleet: Demo

aws INNOVATE
ONLINE CONFERENCE

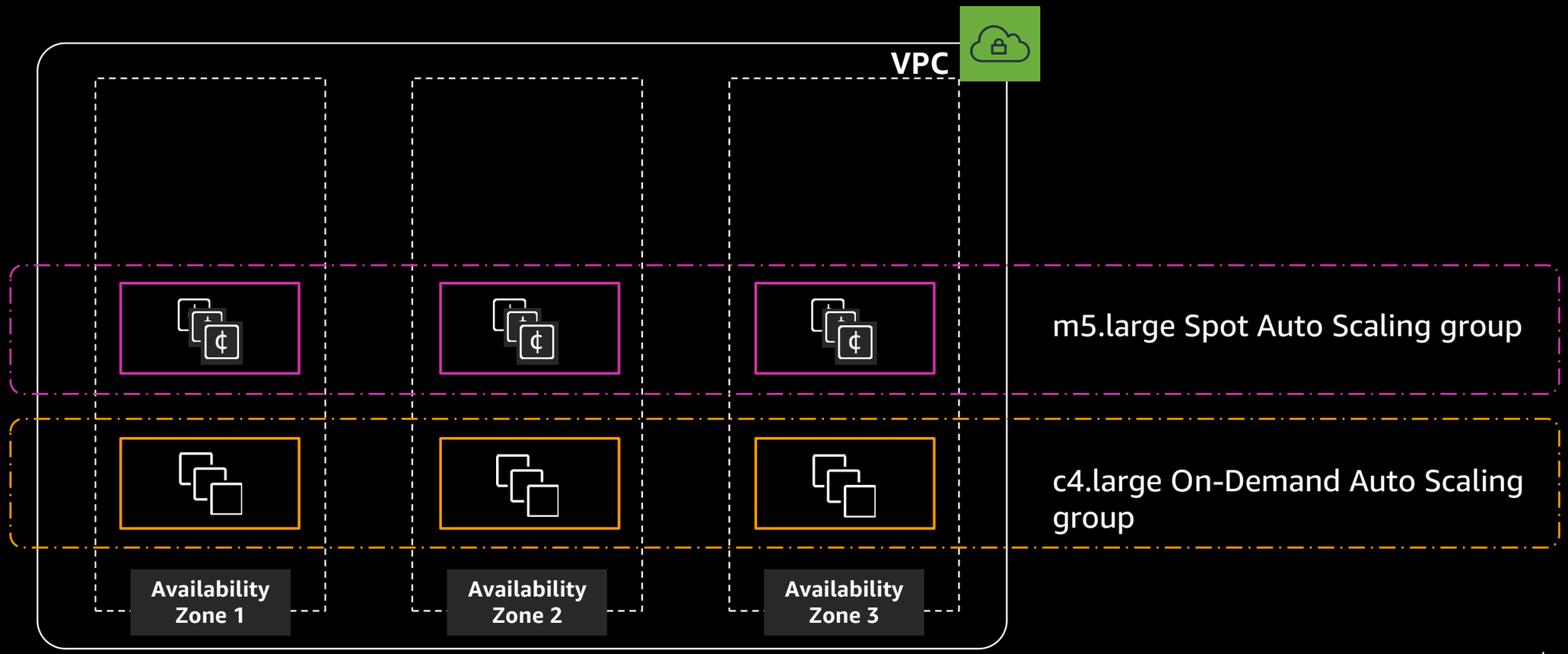
Amazon EC2 Auto Scaling (with multiple purchase options and instance types)

aws INNOVATE
ONLINE CONFERENCE

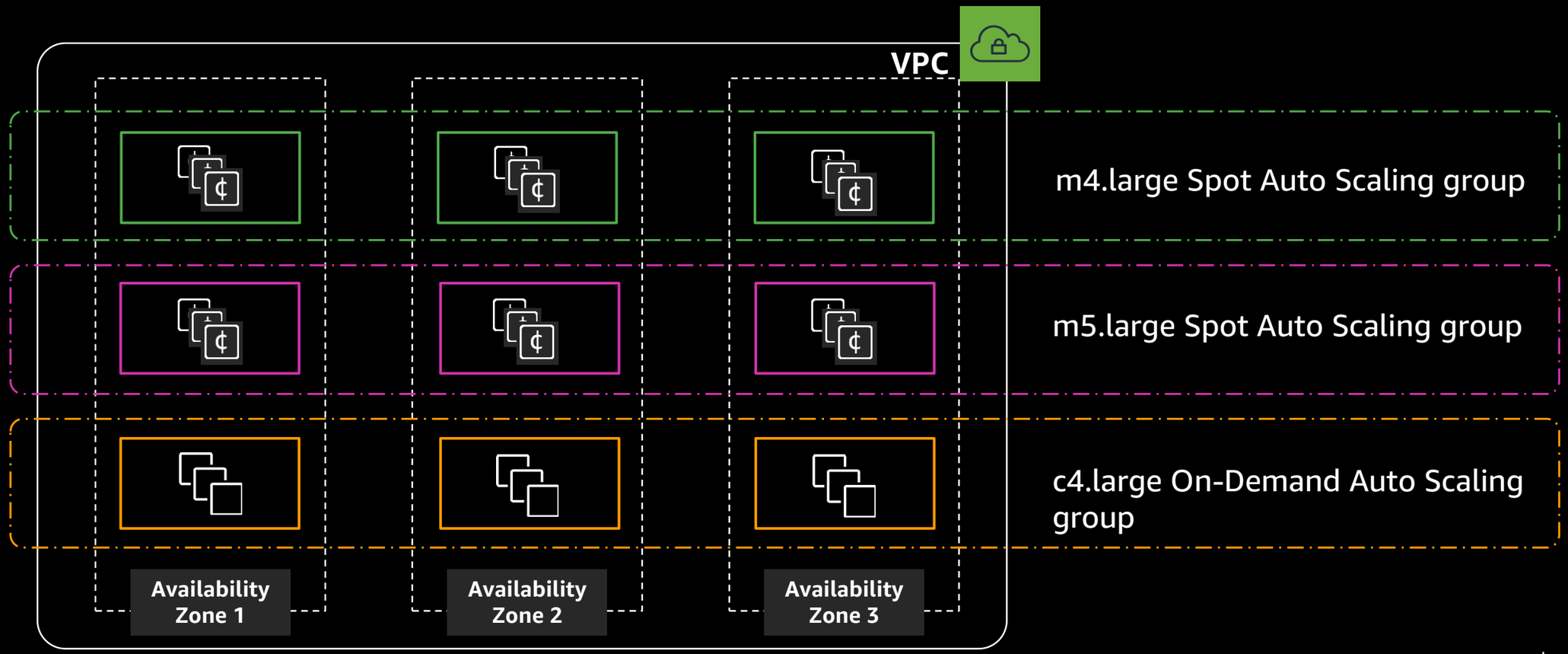
Before: Multiple Auto Scaling groups to use Spot, On-Demand, and RIs together



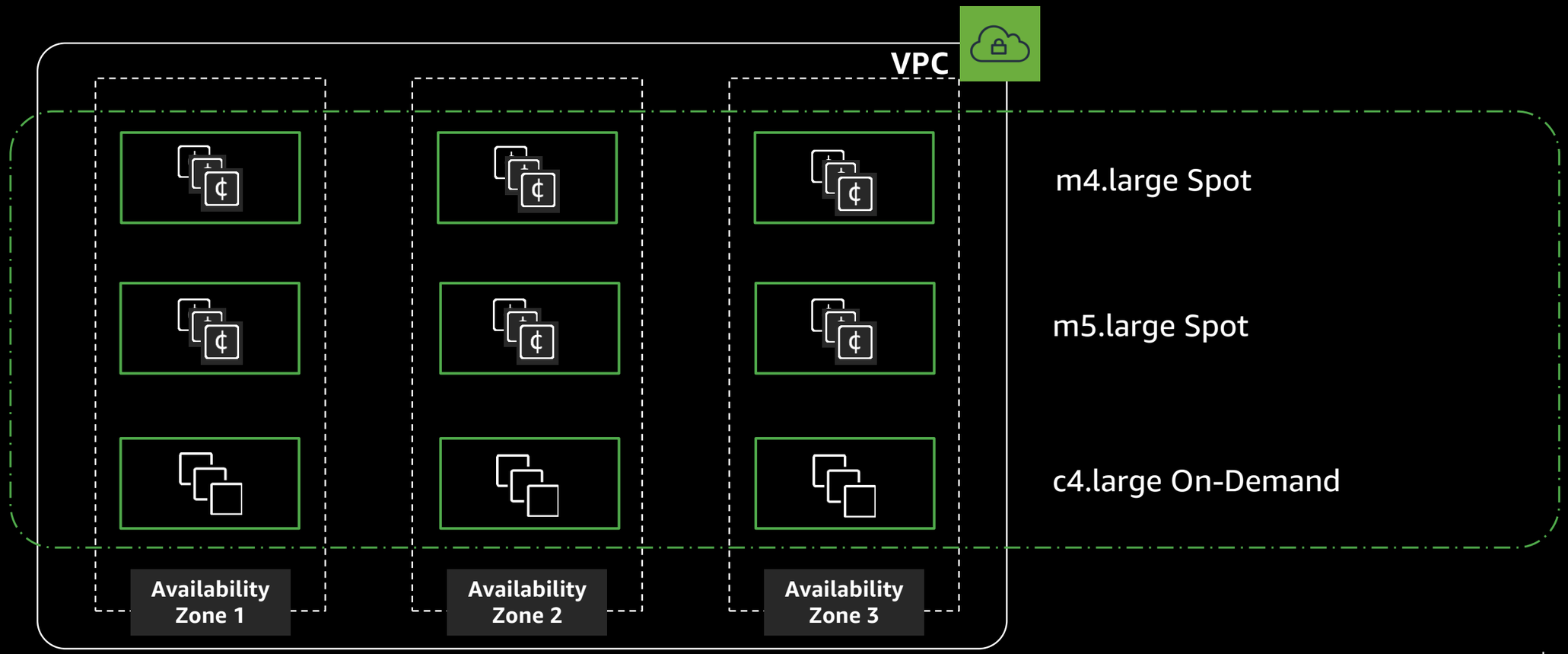
Before: Multiple Auto Scaling groups to use Spot, On-Demand, and RIs together



Before: Multiple Auto Scaling groups to use Spot, On-Demand, and RIs together



After: Include Spot, On-Demand, and RIs in a single Auto Scaling group



Save up to 90% using EC2 Auto Scaling and EC2 Fleet

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

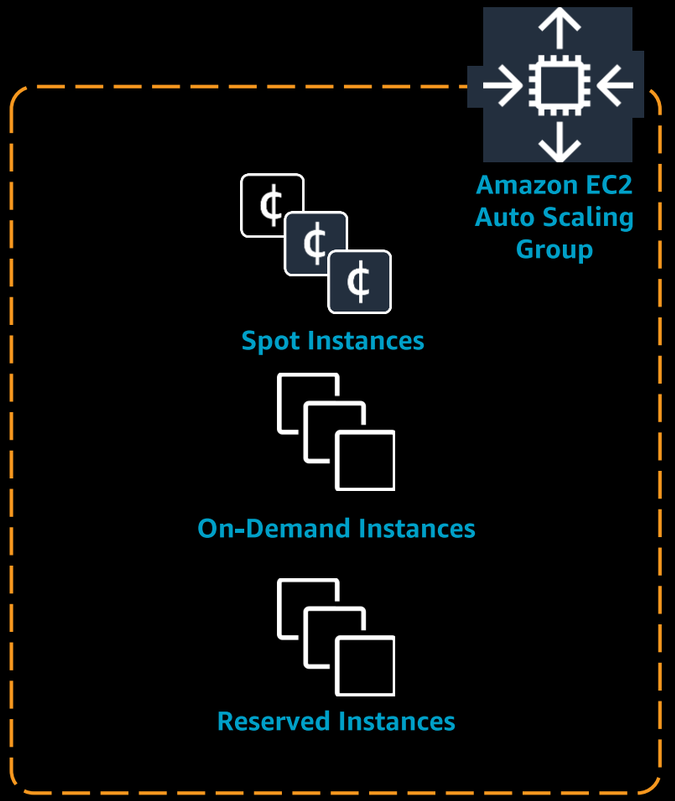
Lowest cost

Specify what percentage of your Auto Scaling group capacity should be fulfilled by On-Demand Instances and Spot Instances to optimize cost

Prioritized list

Use a prioritized list for On-Demand Instance types to scale capacity during an urgent, unpredictable event to optimize performance

Reduce operational overhead



Reduce cost | Optimize performance | Reduce operational overhead

Demo: EC2 Auto Scaling with multiple purchase options and instance types

aws INNOVATE
ONLINE CONFERENCE

In conclusion

- Be aggressive with your cost optimization strategies for compute—Spot is very easy to use and can help you save at scale
- Be flexible and diversify across instance types and purchase models
- Adopt Amazon EC2 launch templates
- Use Amazon EC2 Auto Scaling groups

In conclusion

- All examples and demos are published in GitHub:
  - https://github.com/awslabs/ec2-spot-labs/tree/master/ec2-fleet
- Try out some self-paced workshops yourself:
  - https://ec2spotworkshops.com/
- Follow us on Twitter:
  - @BoydMcgeachie
  - @schmutze

Learn from AWS experts. Advance your skills and knowledge. Build your future in the AWS Cloud.

Digital Training

Free, self-paced online courses built by AWS experts

Classroom Training

Classes taught by accredited AWS instructors

AWS Certification

Exams to validate expertise with an industry-recognized credential

Ready to begin building your cloud skills?
Get started at: https://www.aws.training/

Why work with an APN Partner?

APN Partners are uniquely positioned to help your organization at any stage of your cloud adoption journey, and they:

- Share your goals—focused on your success
- Help you take full advantage of all the business benefits that AWS has to offer
- Provide services and solutions to support any AWS use case across your full customer life cycle

APN Partners with deep expertise in AWS services:

AWS Managed Service Provider (MSP) Partners

APN Partners with cloud infrastructure and application migration expertise

AWS Competency Partners

APN Partners with verified, vetted, and validated specialized offerings

AWS Service Delivery Partners

APN Partners with a track record of delivering specific AWS services to customers

Find the right APN Partner for your needs: https://aws.amazon.com/partners/find/

Thank you for attending AWS Innovate

We hope you found it interesting! A kind reminder to **complete the survey.**
Let us know what you thought of today's event and how we can improve the event experience for you in the future.


aws-apac-marketing@amazon.com


twitter.com/AWSCloud


facebook.com/AmazonWebServices


youtube.com/user/AmazonWebServices


slideshare.net/AmazonWebServices


twitch.tv/aws