AWS re:Invent

NOV. 29 – DEC. 3, 2021 | LAS VEGAS, NV

CMP307

Running efficient Kubernetes clusters on Amazon EC2 with Karpenter

Jillian Kuentz
Software Development Manager
AWS

Carlos Manzanedo Rueda
Principal Solutions Architect
AWS

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

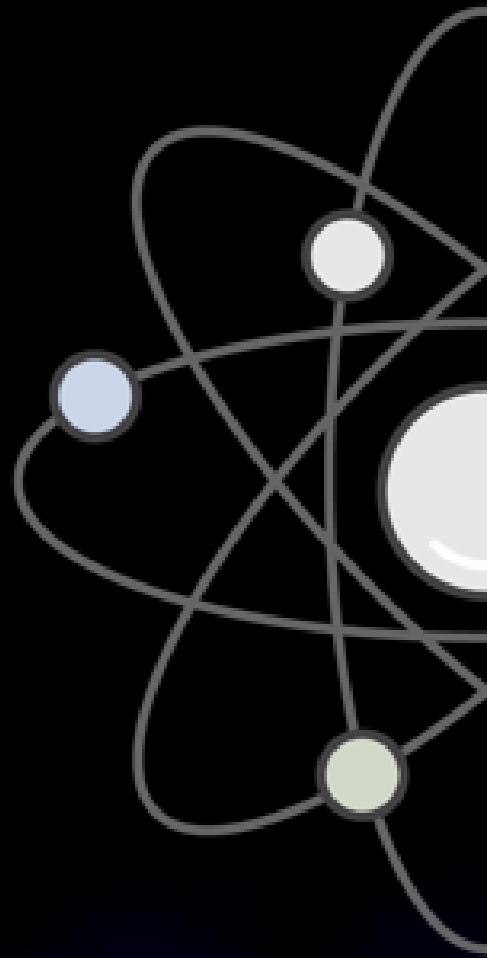
Agenda

Amazon EKS and Amazon EC2 Spot overview

Karpenter and auto scalers

Spot interruption handling

Workshop



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Kubernetes and Spot







© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

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 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% saving** over all regions."

Paul Gillespie
Principal Architect/Tribe Lead

lyft yelp MapBox Expedia Caltech

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

EC2 Spot instance pools explained

C5	1a	1b	1c	On-demand
8XL	$0.28	$0.27	$0.29	$1.76
4XL	$0.16	$0.17	$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	$0.04	$0.11

Instance flexibility and **diversification**

Each instance family

Each instance size

Each availability zone

In every region

Is a separate **Spot pool**



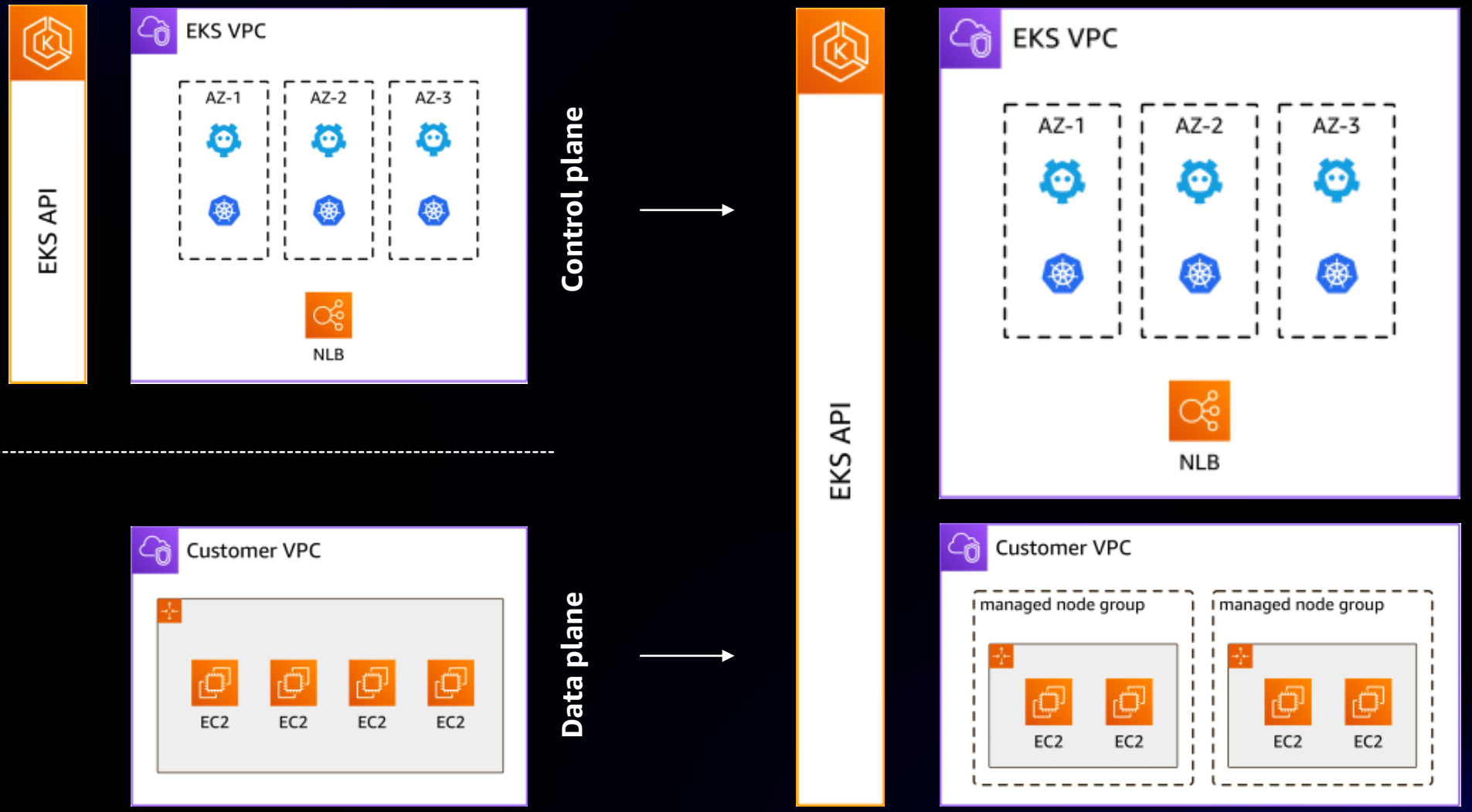
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

A high-level view of an Amazon EKS cluster



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

EKS node groups and managed node groups



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Karpenter and Auto Scaling

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

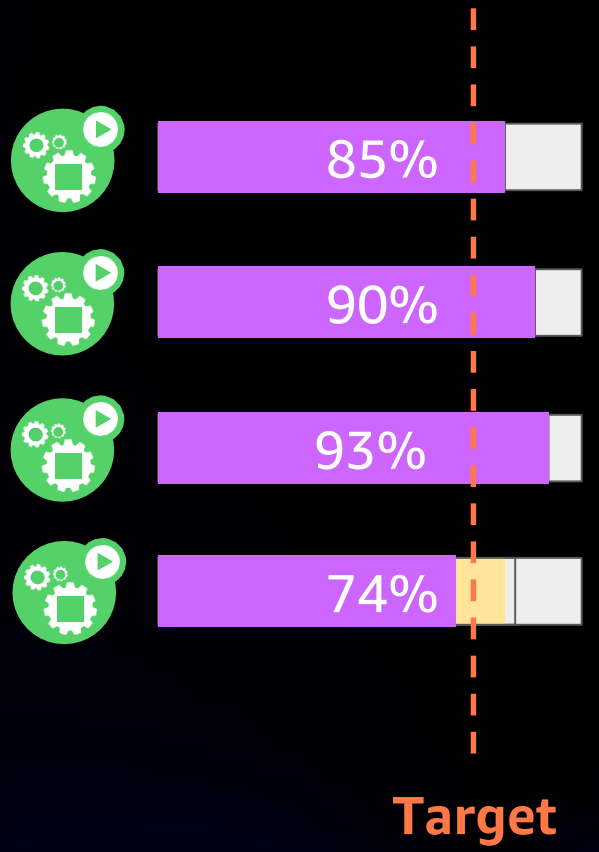
HPA intro

If pods are heavily loaded, then **starting new pods** may bring the average load down

If pods are barely loaded, then **stopping pods** will free resources

Specify the target for the load

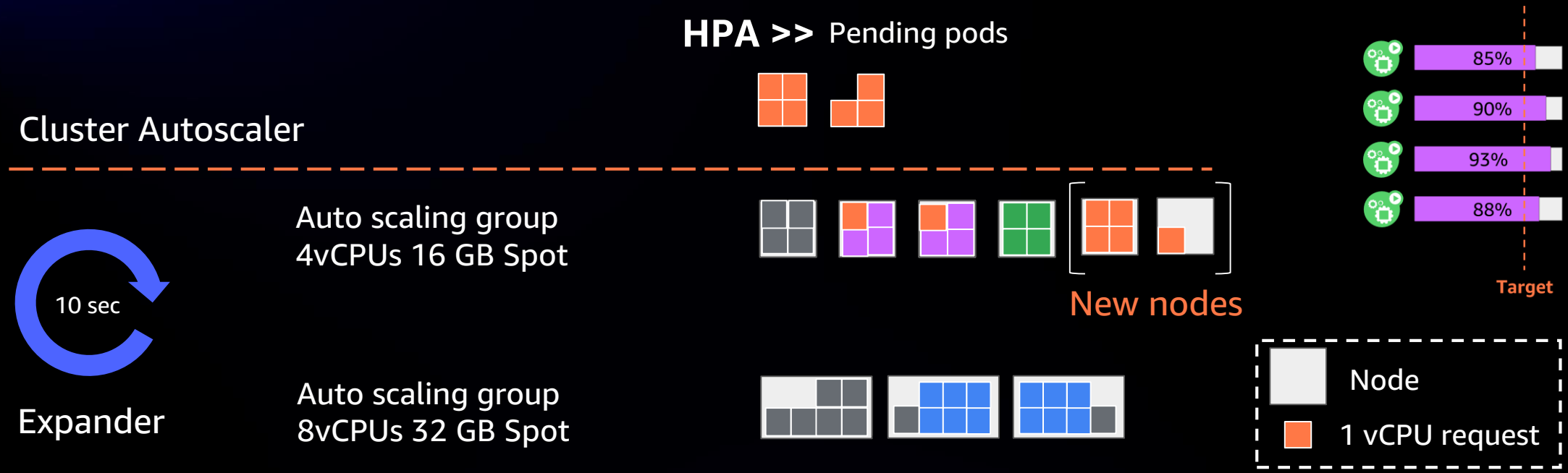
Ex: Target = CPU utilization 70%



aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Cluster Autoscaler scale-up



Cluster AutoScaler RunOnce:

- Reconciliation and filtering
- Scale up (simulation and expander logic)
- Scale down and filtering

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Karpenter: Groupless auto scaling

What if we remove the concept of node groups?

- Simplification of configuration
- Reduce simulation complexity
- Generate node properties from pod constraints

Provisioner CRDs

```yaml
apiVersion: karpenter.sh/v1alpha4
kind: Provisioner
metadata:
  name: team1
spec:
  architectures:
  - amd64
  - arm64
  provider:
    apiVersion: extensions.karpenter.sh/v1alpha1
    kind: AWS
    capacityTypes:
    - on-demand
    - spot
    cluster:
      endpoint: ${CLUSTER_ENDPOINT}
      name: ${CLUSTER_NAME}
    instanceProfile: KarpenterInstPrf-${CLUSTER_NAME}
  taints:
  - effect: NoSchedule
    key: application-intent-1
```

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Karpenter

Karpenter

- Open source / AWSLabs
- Vendor neutral
- Provisioner CRDs (custom resource definition)
- Track nodes using K8s labels
- Workload isolation support

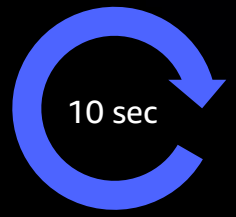
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Karpenter scale-up

HPA >> Pending pods



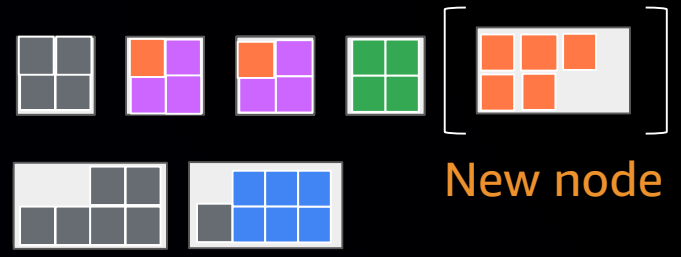
Karpenter



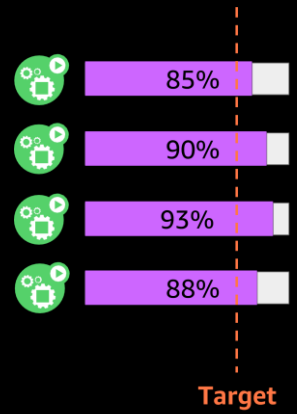
Default: all instance types, excluding metal

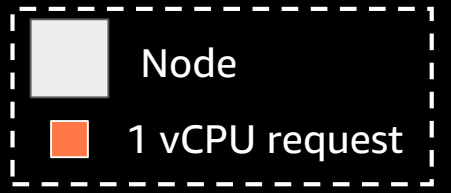
or

instanceTypes:
[m5.large, m5.2xlarge, …]



New node





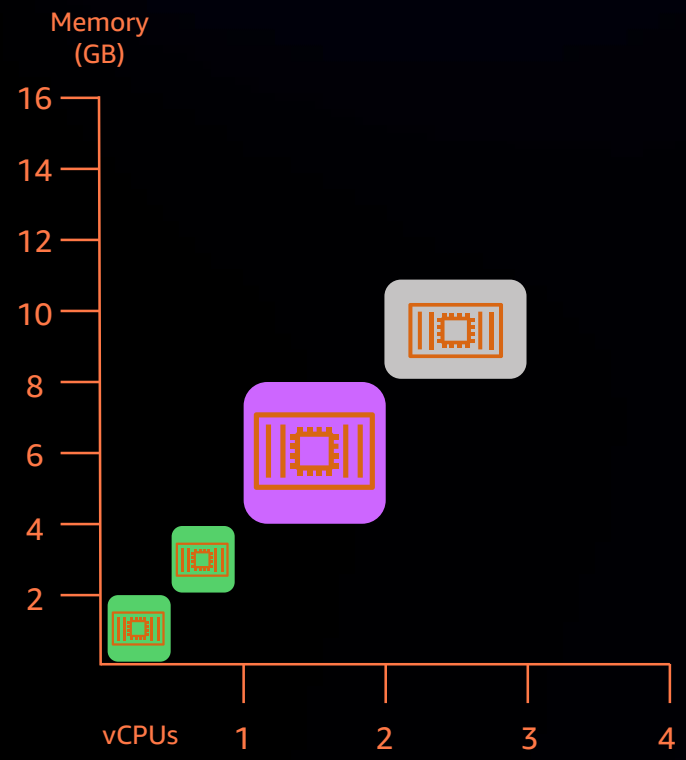
Provisioning and scheduling decisions

- Early binding to provisioned nodes vs. placeholder instances
- Remove scheduler version dependency

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Karpenter: Binpacking



Online binpacking

- First-fit descending
- Multidimensional knapsack
  - CPU, memory, euclidean
  - Other algorithms

Well-known labels

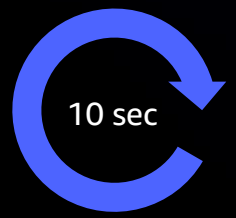
- node.k8.aws/capacity-type=spot
- kubernetes.io/arch=arm64
- topology.Kubernetes.io/zone=us-west-2a
- node.kuberenetes.io/instance-type=m5.large

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

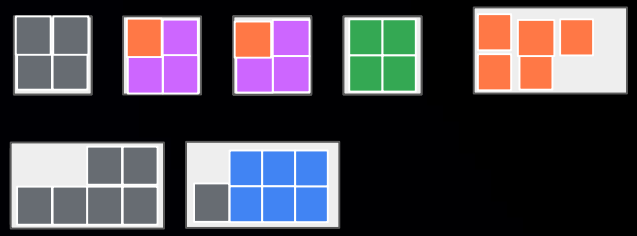
Karpenter scale-in

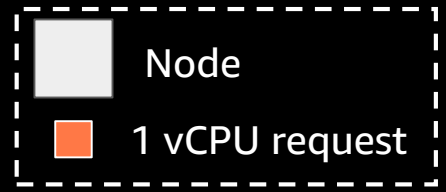
HPA<< Pending pods

Karpenter



ttlSecondsAfterEmpty: seconds the controller will wait before attempting to delete a node, measured from when the node is detected to be empty





Terminations

- Remove underutilized nodes (empty nodes)
- Node TTL
- Future work: defragmentation

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Spot interruption handling

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Handling interruptions



Identify	Spot instance is being reclaimed
2-minute notification	Gracefully prepare the node for termination
Cordon	Cordon to prevent new pods from being placed
Drain	Drain connections on the running pods
Replace	To maintain the desired capacity

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

AWS Node Termination Handler

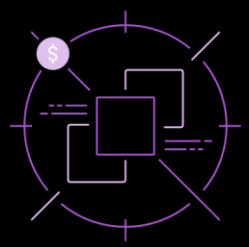
Feature	IMDS processor	Queue processor
K8s DaemonSet	✓	✗
K8s deployment	✗	✓
Spot instance interruptions (ITN)	✓	✓
Scheduled events	✓	✓
EC2 instance rebalance recommendation	✓	✓
ASG lifecycle hooks	✗	✓
EC2 status changes	✗	✓
Setup required	✗	✓



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.







EKS & Spot workshop

Workshop URL: https://cmp307.ec2spotworkshops.com

Event Engine URL: https://dashboard.eventengine.run/login

Event hash: 903c-180bd18ae4-dc

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Thank you!

Jillian Kuentz
@jillmon

Carlos Manzanedo Rueda
@ruecarlo

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.