

NEW YORK | JULY, 10, 2024



© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

# Upcoming Local Events



Amazon JFK27 – Hank

## Karpenter: Amazon EKS best practices and cloud cost optimization

Tuesday 27<sup>th</sup> August 09:00 - 14:00

Come and join this hands-on workshop to learn about optimization on Amazon EKS with Karpenter and EC2 Spot. Discover implementing Karpenter best-practices in a dedicated AWS environment.

## DevOps for Efficient Compute

Thursday 19<sup>th</sup> September 09:00 - 16:00

Languages covered: Python, Go or Java  
Deployment options covered: EKS with Karpenter or Autoscaling Groups on AWS Graviton



© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

KUB301

# Harness the power of Karpenter to scale, optimize & upgrade Kubernetes



© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.



Carlos

ruecarlo@

## Carlos Rueda

(he/him)  
WW Tech Leader  
Core Compute  
AWS



Rachel

rleekin@

## Rachel Leekin

(she/her)  
Sr. Specialist Solutions  
Architect, Containers  
AWS

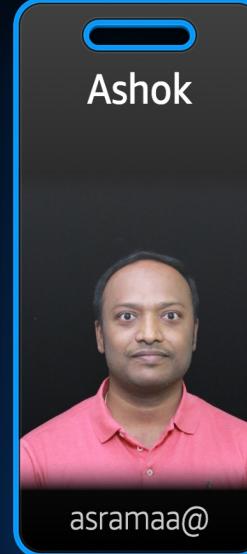


Vinod

singwvin@

## Vinod Singh

(he/him)  
Sr. Solutions Architect,  
ISV  
AWS



Ashok

asrama@

## Ashok Srirama

(he/him)  
Sr. Specialist Solutions  
Architect, Containers  
AWS

# Tips for the day!

- Collaborate and share (be mindful there are multiple customers here)
- Ask for help!
- We want to hear feedback on services/usage !
- Talk with our specialists for any real-world challenges



Join at [menti.com](https://menti.com) | use code **28 51 66 0**



# Instructions

Go to

**www.menti.com**

Enter the code

**28 51 66 0**



Or use QR code



© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.



# What's your role title?



32 responses

DevOps

cloud engineer

CTO

Cloud architect

Vice President

Software Engineer

SWE

Cloud Architect

Platform Engineer

DevOps Architect

Software Engineer

Platform Engineer

Engineering Manager

Product Owner

MLOps Engineer

Cloud Architect

Software Engineer

Principal cloud  
architect

DevOps Engineer

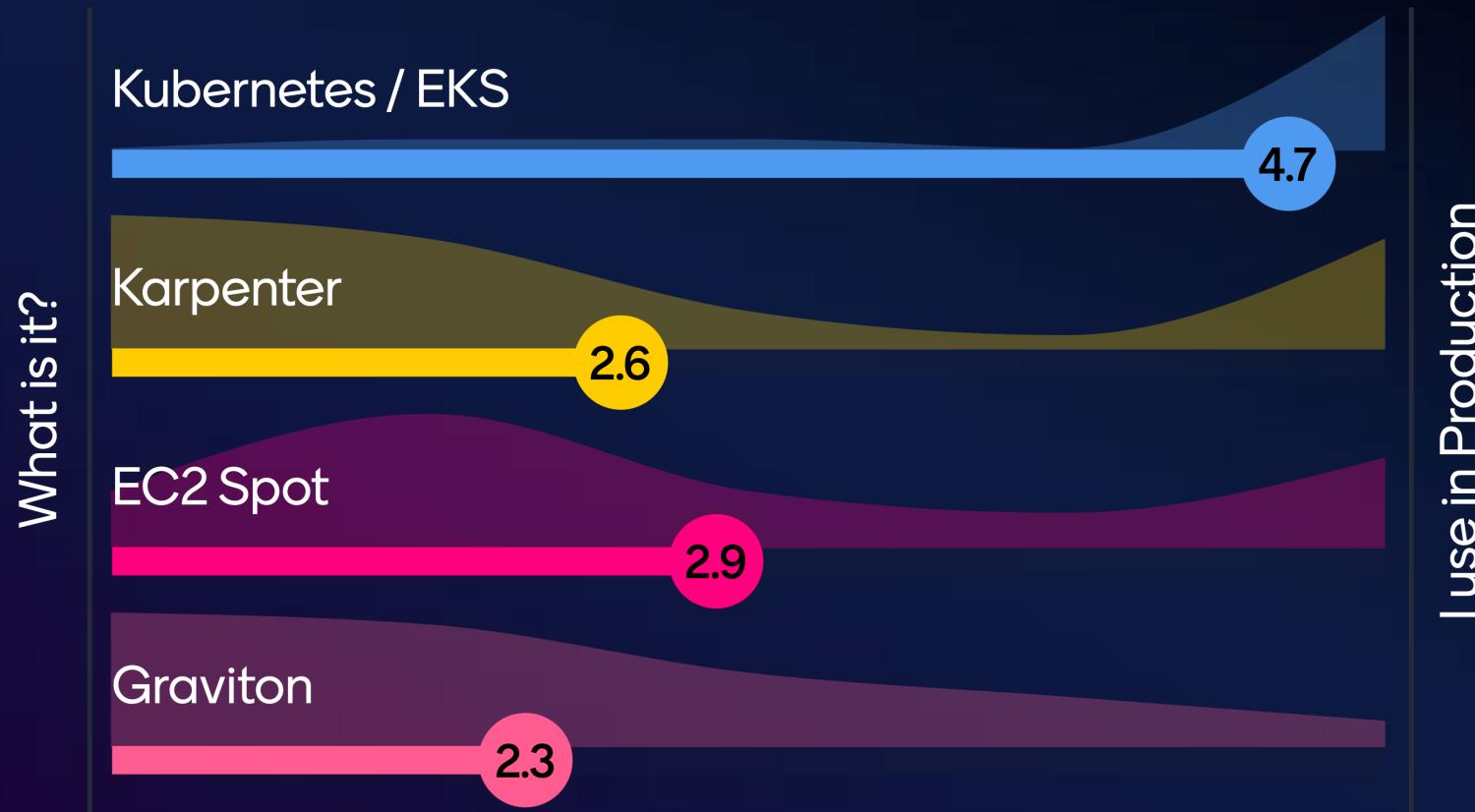
ML Lead



© 2024 Amazon Web Services, Inc. or its affiliates. All rights reserved.



# What's your current adoption?



# What do you want to get of today?



36 responses

yaml:)

Hands on skills of  
optimizing the use of  
EKS

How to leverage  
Karpenter to minimize  
costs.

understand karpenter

Learn about  
karpenter

Knowledge

kube

hands on karpenter

Understand about  
Karpenter, it's

Learning Stuff

Scaling

Learn about  
Carpenter

Poc or use case

Code :)

scaling

roadmap, nodepool  
info, selector best  
practices

A better  
understanding of  
karpenter

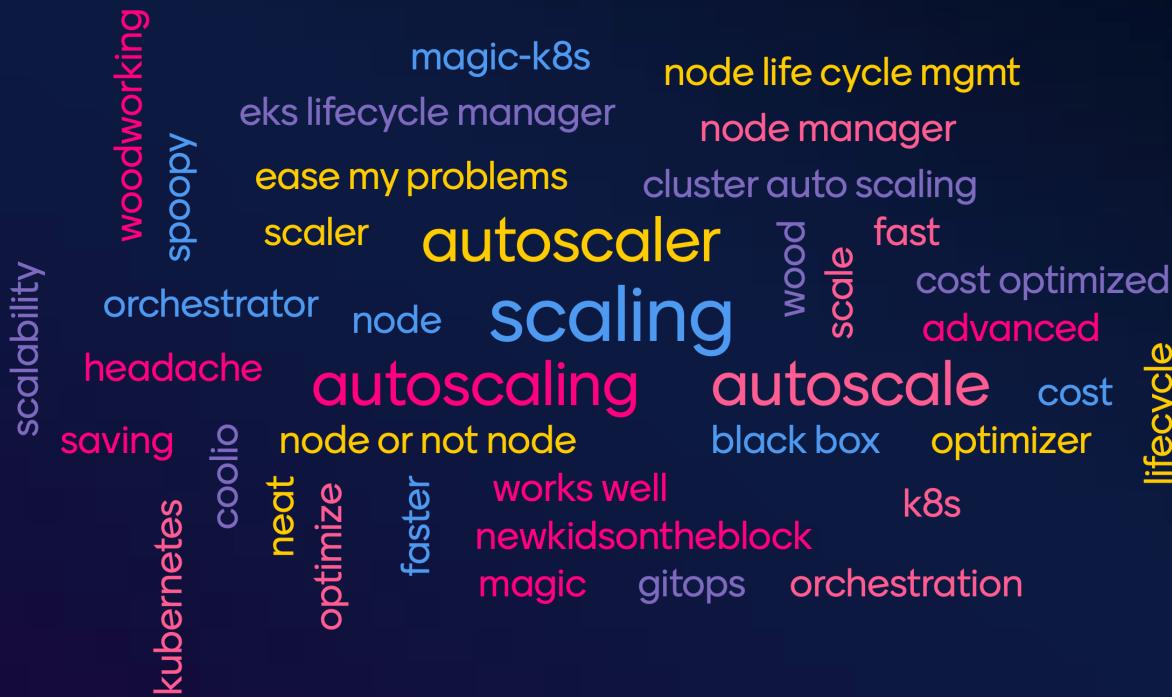
upgrading eks using  
karpenter



© 2024, Amazon Web Services, Inc. or its affiliates. All ri

# One-word description of Karpenter

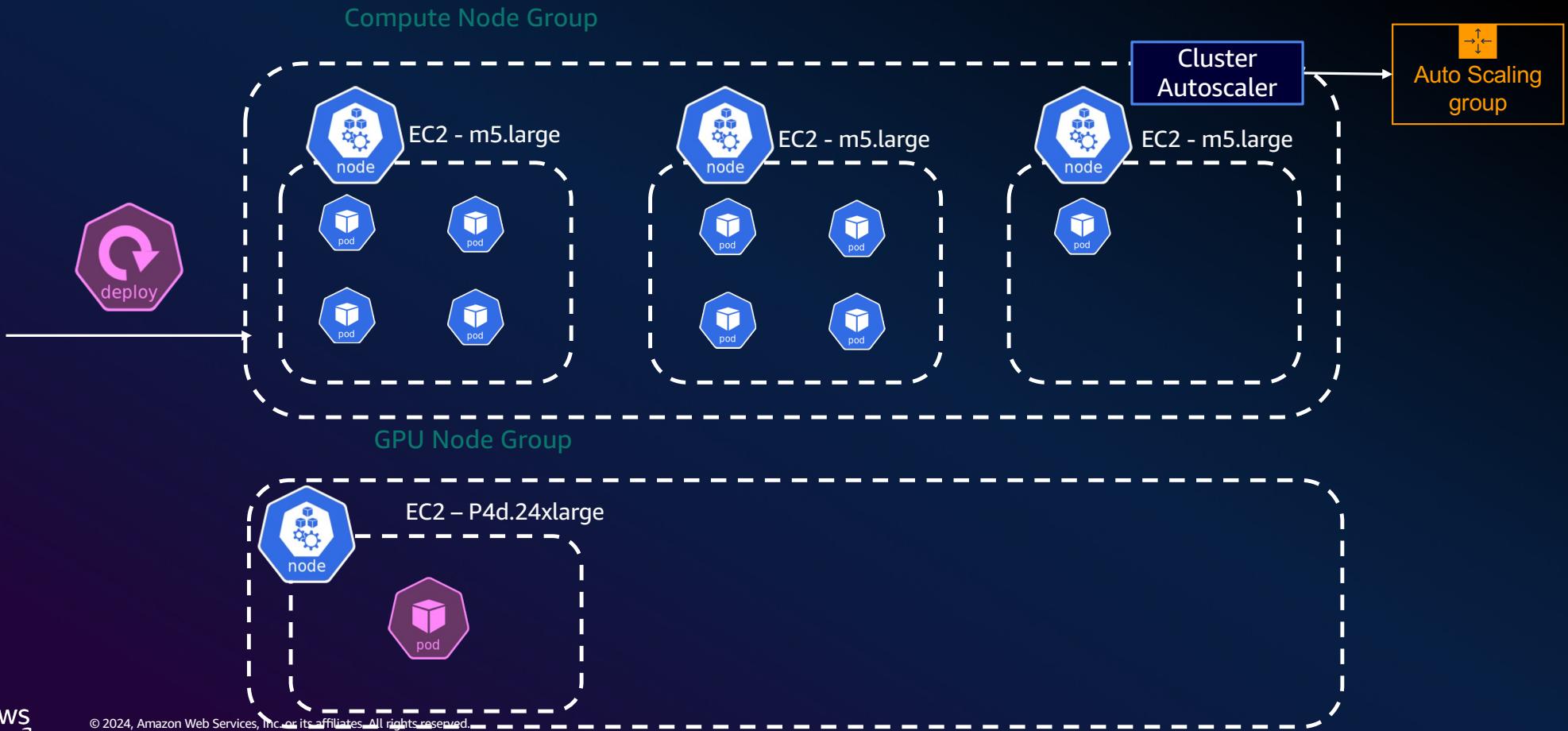
44 responses



# Cluster Autoscaler + ASG Challenges



Amazon EKS



# Challenges managing node groups and ASGs

- Overhead of multiple ASGs across AZs for high availability
- Workloads need different compute resources
- Added overhead increases cost of operations



# Karpenter



Amazon EKS



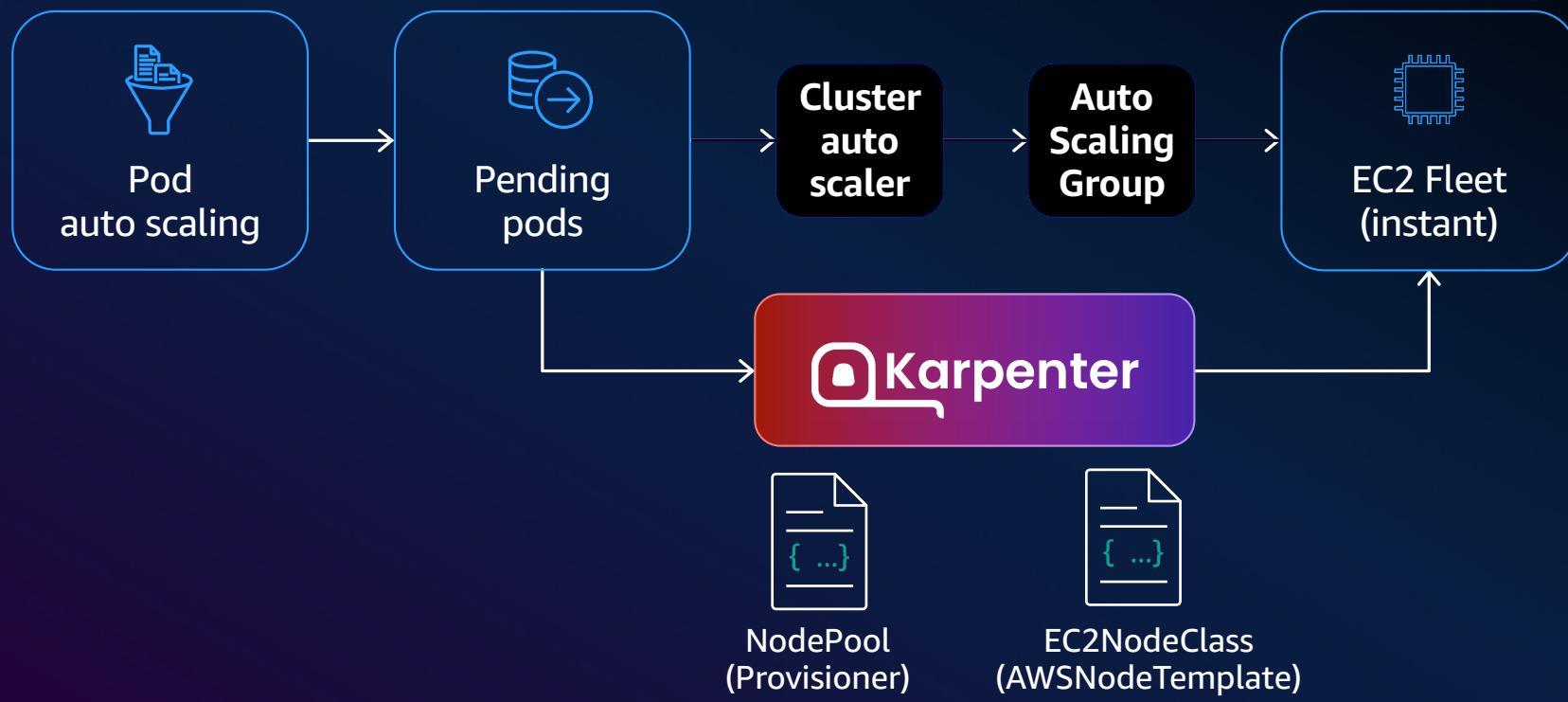
## What if we remove the concept of node groups?

- Simplified configuration
- Cost efficient to run diverse workloads
- Kubernetes native, Open-source
- Flexible compute built-in



© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

# How Karpenter works



# NodePool

## Instance type flexibility

- Attribute-based requirements → sizes, families, generations, CPU architectures
- Setup limits on total vCPU or memory of nodes

## AZ flexibility

- Provision in specified AZs

## CPU Arch flexibility

- x86-64 or Arm64

## Purchase options flexibility

- On-demand, if nothing specified
- Prioritizes Spot if both present



© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

```
apiVersion: kapenter.sh/v1beta1
kind: NodePool
metadata:
  name: default
spec:
  template:
    spec:
      requirements:
        - key: kapenter.k8s.aws/instance-family
          operator: In
          values: ["c5","m5","r5"]
        - key: kapenter.k8s.aws/instance-size
          operator: NotIn
          values: ["nano","micro","small"]
        - key: topology.kubernetes.io/zone
          operator: In
          values: ["us-west-2a","us-west-2b"]
        - key: kubernetes.io/arch
          operator: In
          values: ["amd64","arm64"]
        - key: kapenter.sh/capacity-type
          operator: In
          values: ["spot","on-demand"]
      limits:
        cpu: 100
```

# Selecting AMI and more with EC2NodeClass

```
apiVersion: karpenter.k8s.aws/v1beta1
kind: EC2NodeClass
metadata:
  name: default
spec:
  subnetSelectorTerms:
    - tags:
        karpenter.sh/discovery: "${CLUSTER_NAME}"
  securityGroupSelectorTerms:
    - tags:
        karpenter.sh/discovery: "${CLUSTER_NAME}"
  amiFamily: AL2
  amiSelectorTerms:
    - tags:
        karpenter.sh/discovery: "${CLUSTER_NAME}"
  userData: |
    echo "Hello world"
  role: "KarpenterNodeRole-${CLUSTER_NAME}"
  tags:
    team: team-a
  blockDeviceMappings: {...}
  metadataOptions: {...}
```

- Select which subnets are eligible to launch EC2
- Attach security groups to EC2
- Select AMI for EC2
- By default, AL2 is prepopulated
- Supported AMIs: AL2, AL2023, BR, Ubuntu, Win 19/22, custom
- Run userData, define EBS, attach tags, and more

# Benefits of leveraging Karpenter

## Operational efficiency

### Node auto-provisioning

- ✓ Optimal Compute Infrastructure E.g. Intel, Graviton, GPUs
- ✓ Flexibility E.g. No need to configure every node group)

### Simplified Resource management

- ✓ Easy upgrades
- ✓ Efficient management E.g. many clusters and node groups

## Cost Optimization

### Node consolidation

- ✓ Increases utilization E.g. Consolidates and bin packs pods to fewer node groups
- ✓ Lowers costs E.g. consolidation from On-Demand->On-Demand, On-Demand->EC2 Spot

## Application Availability

### Rapid scaling

- ✓ Faster E.g. responds quickly and automatically to changes in application load, scheduling, and resource requirements

### Compute availability

- ✓ Mixes purchasing options as needed E.g. Prioritizes Spot, pulls in On-Demand if Spot is not available



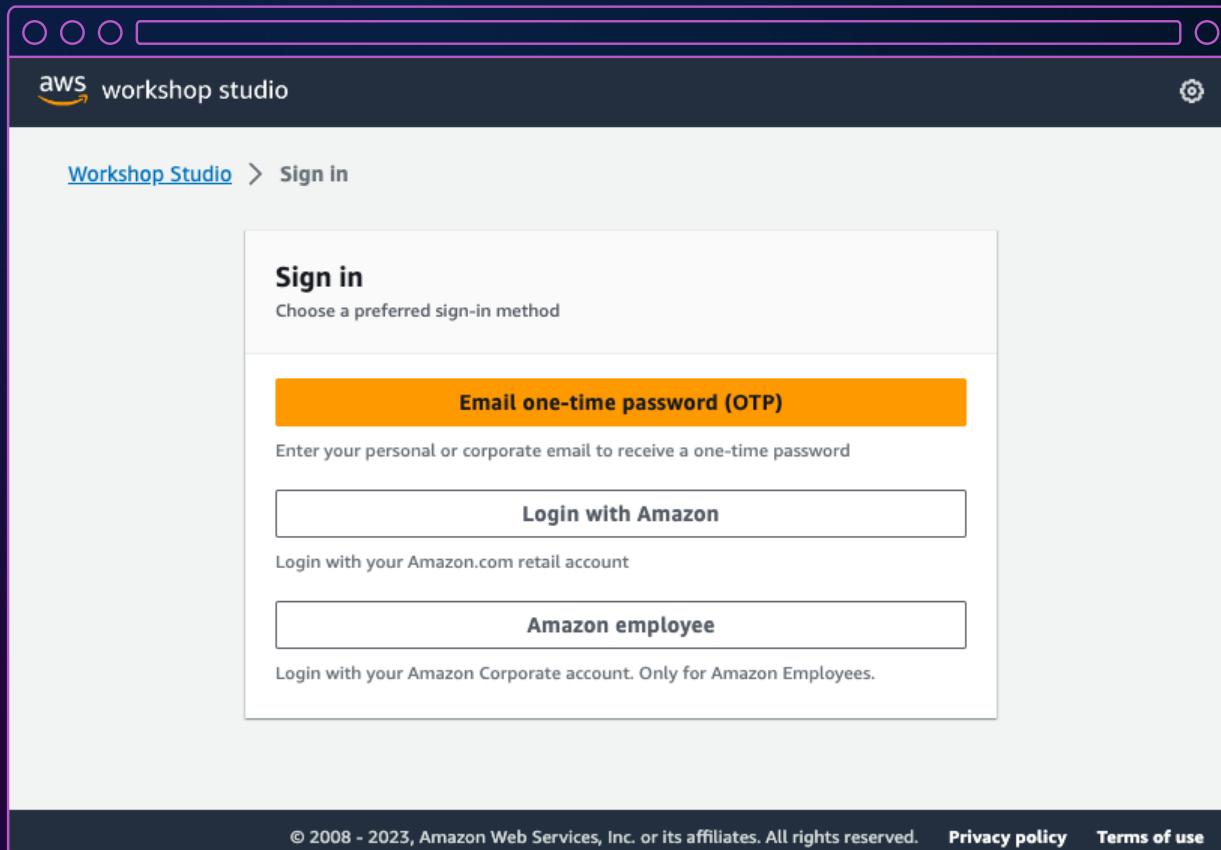
**Event code:**

**<https://s12d.com/ny-summit-kub301>**



© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

# Step 1: Sign in using your preferred method



The screenshot shows the sign-in page for the AWS Workshop Studio. At the top, there's a header bar with the AWS logo and the text "aws workshop studio". Below the header, the URL "Workshop Studio > Sign in" is visible. The main area is titled "Sign in" and contains the instruction "Choose a preferred sign-in method". There are four options listed: "Email one-time password (OTP)" (which is highlighted in orange), "Login with Amazon", "Amazon employee", and "Login with your Amazon.com retail account". At the bottom of the page, a dark footer bar contains the copyright notice "© 2008 - 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.", and links for "Privacy policy" and "Terms of use".



© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

# Step 2: Enter the event access code

The screenshot shows a web browser window for 'aws workshop studio'. The title bar says 'aws workshop studio'. The main content area is titled 'Workshop Studio > Join event'. On the left, there's a sidebar with 'Step 1: Enter event access code' and 'Step 2: Review and join'. The main content area has a heading 'Enter event access code' and a sub-section 'Event access code' with a text input field containing 'abcd-012345-ef'. At the bottom right are 'Cancel' and 'Next' buttons.

Access code is  
automatically added



© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

© 2008 - 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved. [Privacy policy](#) [Terms of use](#)

# Step 3: Review terms and join event

The screenshot shows the 'Review and join' step of the AWS Workshop Studio process. At the top, it displays the event details: Name - re:Invent 2023 Workshop, Start time - 8/28/2023 10:59 AM, Duration - 12 hours, Level - 100. Below this is a section titled 'Terms and Conditions' with a link to 'Read and accept before joining the event'. It lists 11 numbered terms of use. A checkbox at the bottom left of the terms section is checked, with the label 'I agree with the Terms and Conditions'. At the bottom right, there are 'Cancel', 'Previous', and 'Join event' buttons.

aws workshop studio

Workshop Studio > Join event

Step 1  
Enter event access code

Step 2  
Review and join

**Review and join**

**Event details**

Name	Start time	Duration	Level
re:Invent 2023 Workshop	8/28/2023 10:59 AM	12 hours	100

Description  
Workshop at re:Invent 2023

**Terms and Conditions**  
Read and accept before joining the event

Read and accept before joining the event:

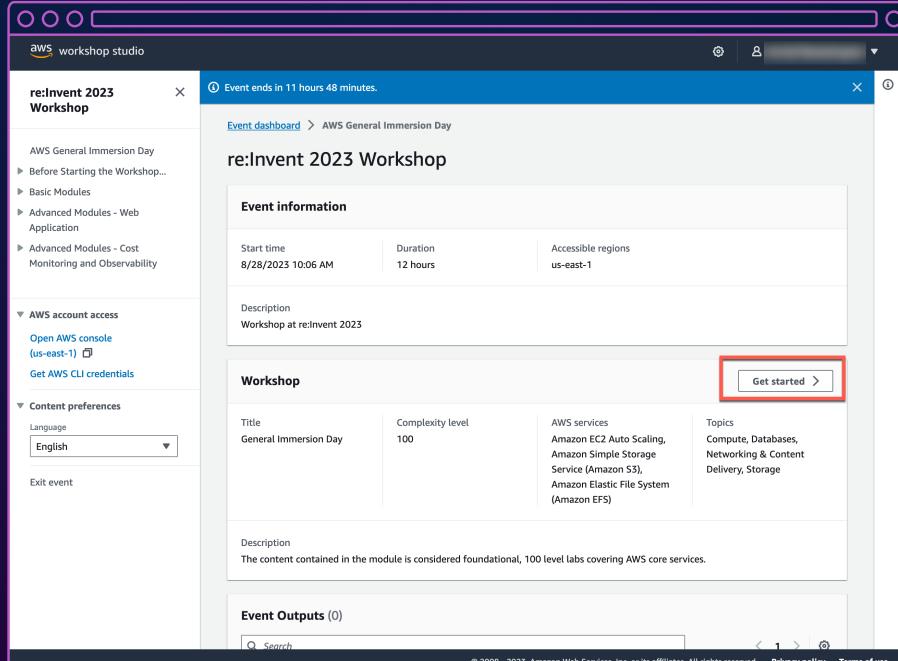
- By using AWS Workshop Studio for the relevant event, You agree to the [AWS Event Terms and Conditions](#) and the [AWS Acceptable Use Policy](#).
- If You are under 18 years old, You may participate in the relevant event, using AWS Workshop Studio: (a) if You are at least the minimum age below based on the country or region in which You reside, and (b) with the involvement of a parent, guardian or educator.
- You acknowledge and agree that You are using an AWS-owned account that You will be only be able to access during the relevant event. You have no ownership rights over this AWS-owned account.
- During the relevant event, while using this AWS-owned account, You will not use, import, input, or introduce any data, dataset, or other material that contains personal data, financial information, or any other data or materials that may be subject to laws and regulations such as the General Data Protection Regulation or The Health Insurance Portability and Accountability Act of 1996.
- If You find residual resources or materials in this AWS-owned account, You will notify Your Event Operator immediately.
- AWS, its affiliates and any entities or persons acting on AWS's behalf, reserves the right to terminate this AWS-owned account and to delete its contents at any time, without any notice to You.
- During the relevant event, while using this AWS-owned account, You will not: process or run any operation on any data other than test datasets or lab materials that have been approved by AWS.
- You will not copy, import, export or otherwise create derivative works of materials provided by AWS for use outside of the event.
- AWS, its affiliates and any entities or persons acting on AWS's behalf, has no obligation to enable the transmission of Your materials through AWS Workshop Studio, and may, in its discretion, edit, block, refuse to post, or remove Your materials at any time, without notice to You.
- If You are an AWS Partner, using AWS Workshop Studio as part of Your participation in the AWS Partner Network Program, Your use of AWS Workshop Studio is governed by these terms, the AWS Partner Network Terms and Conditions, and Your applicable customer agreement with AWS.
- Your use of AWS Workshop Studio will comply with these terms and all applicable laws. If You fail to comply with any of these terms, Your access to AWS Workshop Studio may be immediately terminated, without notice to You.

I agree with the Terms and Conditions

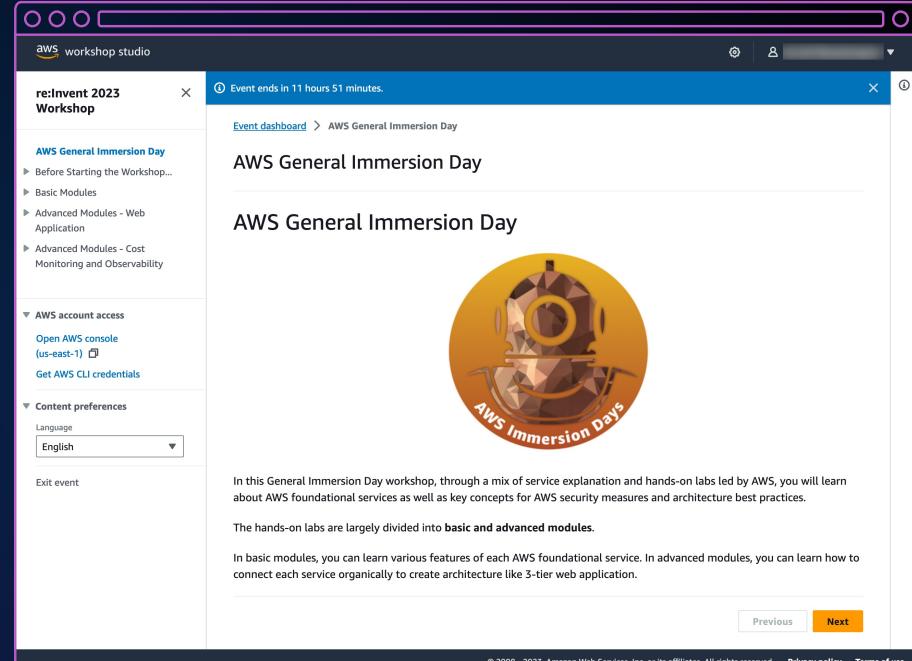
Cancel Previous Join event



# Step 4: Get started with the workshop



The screenshot shows the AWS Workshop Studio interface for the re:Invent 2023 Workshop. The 'Event information' section displays the start time (8/28/2023 10:06 AM), duration (12 hours), and accessible regions (us-east-1). The 'Workshop' section shows the title 'General Immersion Day', complexity level (100), AWS services (Amazon EC2 Auto Scaling, Amazon Simple Storage Service (Amazon S3), Amazon Elastic File System (Amazon EFS)), and topics (Compute, Databases, Networking & Content Delivery, Storage). A large orange arrow points from the 'Get started' button in the workshop details to the expanded description below.



The screenshot shows the expanded 'Get started' description for the 'General Immersion Day' workshop. It states: 'In this General Immersion Day workshop, through a mix of service explanation and hands-on labs led by AWS, you will learn about AWS foundational services as well as key concepts for AWS security measures and architecture best practices.' Below this, it says: 'The hands-on labs are largely divided into basic and advanced modules.' At the bottom, there are 'Previous' and 'Next' navigation buttons.



# Step 5: Access AWS account

AWS General Immersion Day

In this General Immersion Day workshop, through a mix of service explanation and hands-on labs led by AWS, you will learn about AWS foundational services as well as key concepts for AWS security measures and architecture best practices.

The hands-on labs are largely divided into **basic** and **advanced** modules.

In basic modules, you can learn various features of each AWS foundational service. In advanced modules, you can learn how to connect each service organically to create architecture like 3-tier web application.

Previous Next

Console Home

Recently visited

- CloudWatch
- Service Quotas
- AWS Organizations
- Athena
- CloudFormation
- AWS Health Dashboard
- Resource Groups & Tag Editor
- Amazon EventBridge
- AWS Cost Explorer
- S3
- EC2
- Lambda
- CodeCommit
- Amazon Redshift

Welcome to AWS

Getting started with AWS

Training and certification

What's new with AWS?

Cost and usage

Current month costs  
\$0.00

Forecasted month end costs  
-

Last month costs  
\$0.00

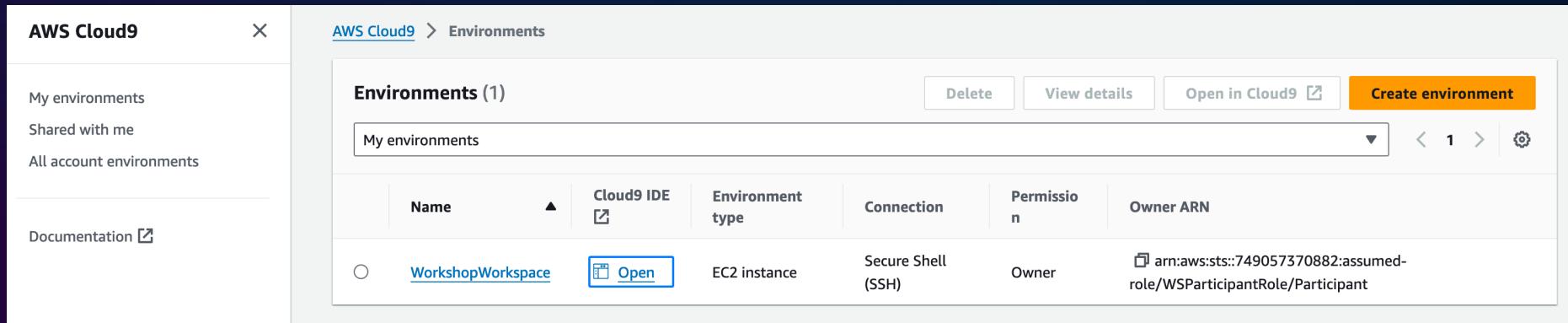
Select Open AWS console



# Step 6: Navigate to Cloud9 and open IDE



The screenshot shows the AWS search interface with the query 'cloud9' entered. The search results page displays various AWS services and features. The 'Cloud9' service card is highlighted with a red box. The card includes the text 'A Cloud IDE for Writing, Running, and Debugging Code'.



The screenshot shows the AWS Cloud9 environments page. The left sidebar lists 'My environments', 'Shared with me', 'All account environments', and a 'Documentation' link. The main content area shows a table titled 'Environments (1)'. The table has columns for Name, Cloud9 IDE, Environment type, Connection, Permission, and Owner ARN. One environment is listed: 'WorkshopWorkspace' (Cloud9 IDE, EC2 instance, Secure Shell (SSH), Owner, arn:aws:sts::749057370882:assumed-role/WSParticipantRole/Participant).

Name	Cloud9 IDE	Environment type	Connection	Permission	Owner ARN
<a href="#">WorkshopWorkspace</a>	<a href="#">Open</a>	EC2 instance	Secure Shell (SSH)	Owner	arn:aws:sts::749057370882:assumed-role/WSParticipantRole/Participant



# Hands-on lab sections

1. Install Karpenter
2. Basic NodePool (skip drift sub-section)
3. Cost optimization (skip on-demand & spot ratio split sub-section)
4. Scheduling constraints (challenges!)

The rest of modules are optional



# References

- Karpenter best practices  
<https://aws.github.io/aws-eks-best-practices/karpenter/>
- Karpenter workshop  
<https://catalog.workshops.aws/karpenter>
- Karpenter blueprints  
<https://github.com/aws-samples/karpenter-blueprints>
- AWS re:Invent 2023: Harness the power of Karpenter  
[https://www.youtube.com/watch?v=lkq\\_9ETHeks&t=15s](https://www.youtube.com/watch?v=lkq_9ETHeks&t=15s)



© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

# Upcoming Local Events



Amazon JFK27 – Hank

## Karpenter: Amazon EKS best practices and cloud cost optimization

**Tuesday 27<sup>th</sup> August 09:00 - 14:00**

Come and join this hands-on workshop to learn about optimization on Amazon EKS with Karpenter and EC2 Spot. Discover implementing Karpenter best-practices in a dedicated AWS environment.

## DevOps for Efficient Compute

**Thursday 19<sup>th</sup> September 09:00 - 16:00**

Languages covered: Python, Go or Java

Deployment options covered: EKS with Karpenter or Autoscaling Groups on AWS Graviton



© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

# Thank you!



Please complete the session  
survey in the mobile app

**Carlos Manzanedo Rueda**  
(he/him)  
WW SA Leader, Flexible  
Compute  
AWS

**Rachel Leekin**  
(she/her)  
Sr. Specialist Solutions  
Architect, Containers  
AWS

**Vinod Singh**  
(he/him)  
Sr. Solutions Architect,  
ISV  
AWS

**Pankaj Walke**  
(he/him)  
Sr. Cloud Support  
Engineer  
AWS

**Ashok Srirama**  
(he/him)  
Sr. Specialist Solutions  
Architect, Containers  
AWS



© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.