

AWS re:Invent

NOV. 28 – DEC. 2, 2022 | LAS VEGAS, NV

Best practices for high performance computing in the cloud

Maxime Hugues

Principal HPC Specialist Solutions Architect
AWS

Lowell Wofford

Principal Solutions Architect
AWS

Agenda

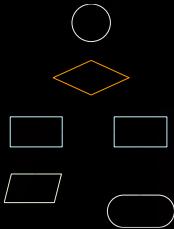
Presentation: Fundamentals of infrastructure for HPC and AWS ParallelCluster

Hands-on: Create a HPC cluster using AWS ParallelCluster

Presentation: Amazon FSx for Lustre and application management using Spack

Hands-on: HPC cluster update to dynamically mount Amazon FSx for Lustre

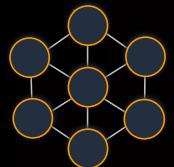
Compute and orchestration building blocks



Workflow, notifications, queues

- Workflow management
- Notification and message queues

Workflow management and communication

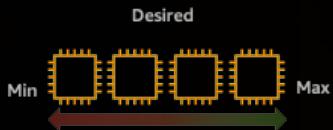


Serverless & containers

- Event-driven functions (AWS Lambda)
- Batch schedulers; containers orchestrators

Abstraction

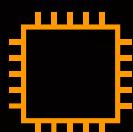
Focus on the workload and not infrastructure



Provisioning

- Amazon EC2 Auto Scaling groups: scale up & down
- Instance fleets: capacity at scale across AZs

Compute on events or requests



Instances

- Virtualized
- Bare metal

Different capabilities (CPU, RAM, SSD, network, and accelerators)

Base compute layer

Base infrastructure

VPC and subnets

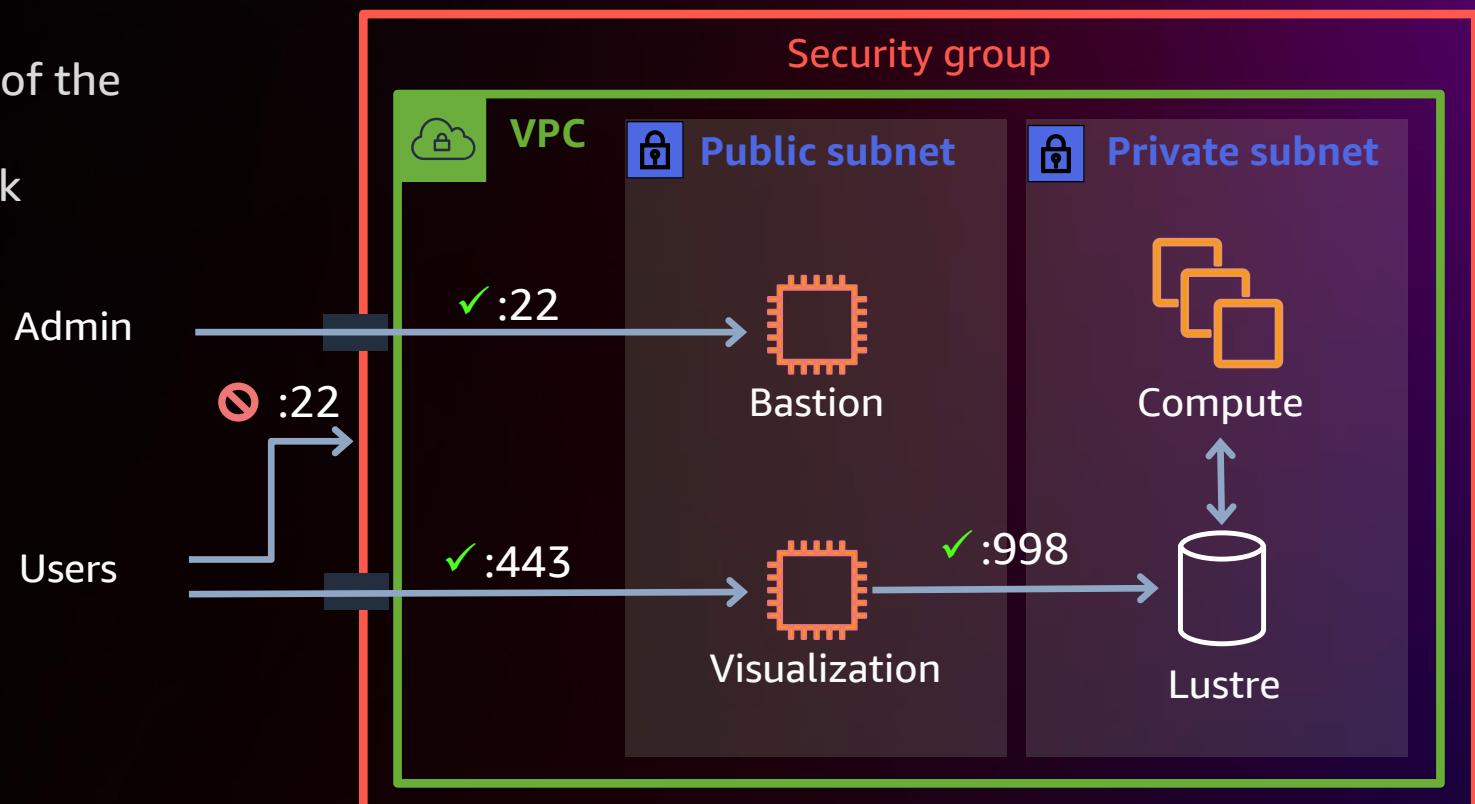
- Virtual private cloud: logical section of the cloud provider infrastructure
- Subnet: logical partition of a network

Security groups

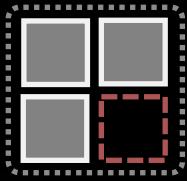
- Virtual firewalls
- VPC & instances

Instances and services

- Instances
- Managed services
- ...

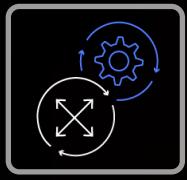


Amazon EC2 Auto Scaling in more detail



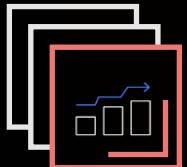
Logical unit

Purpose of scaling or management



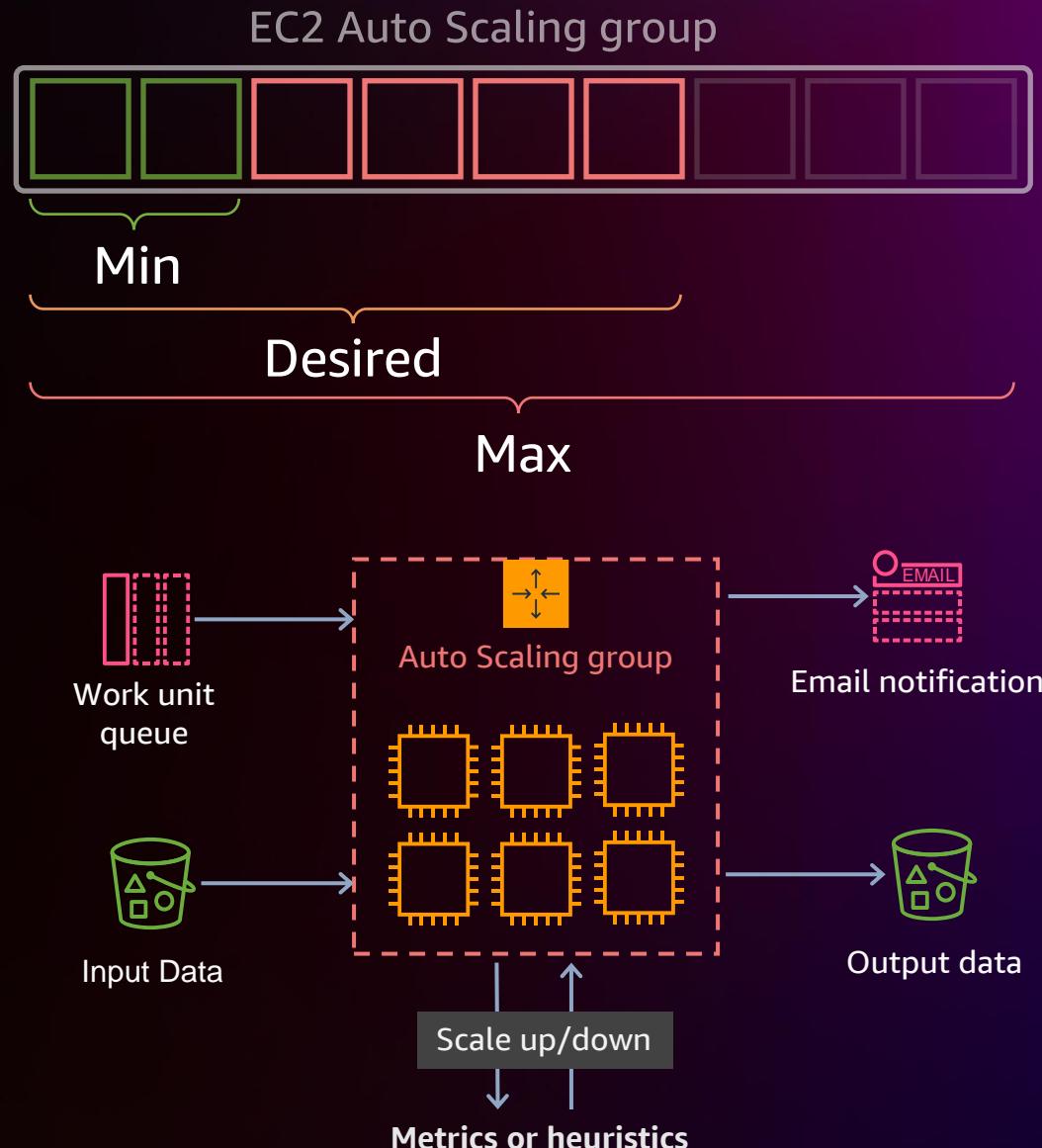
Launch templates

Kind, size, storage, SSH key pair, user data, and security groups



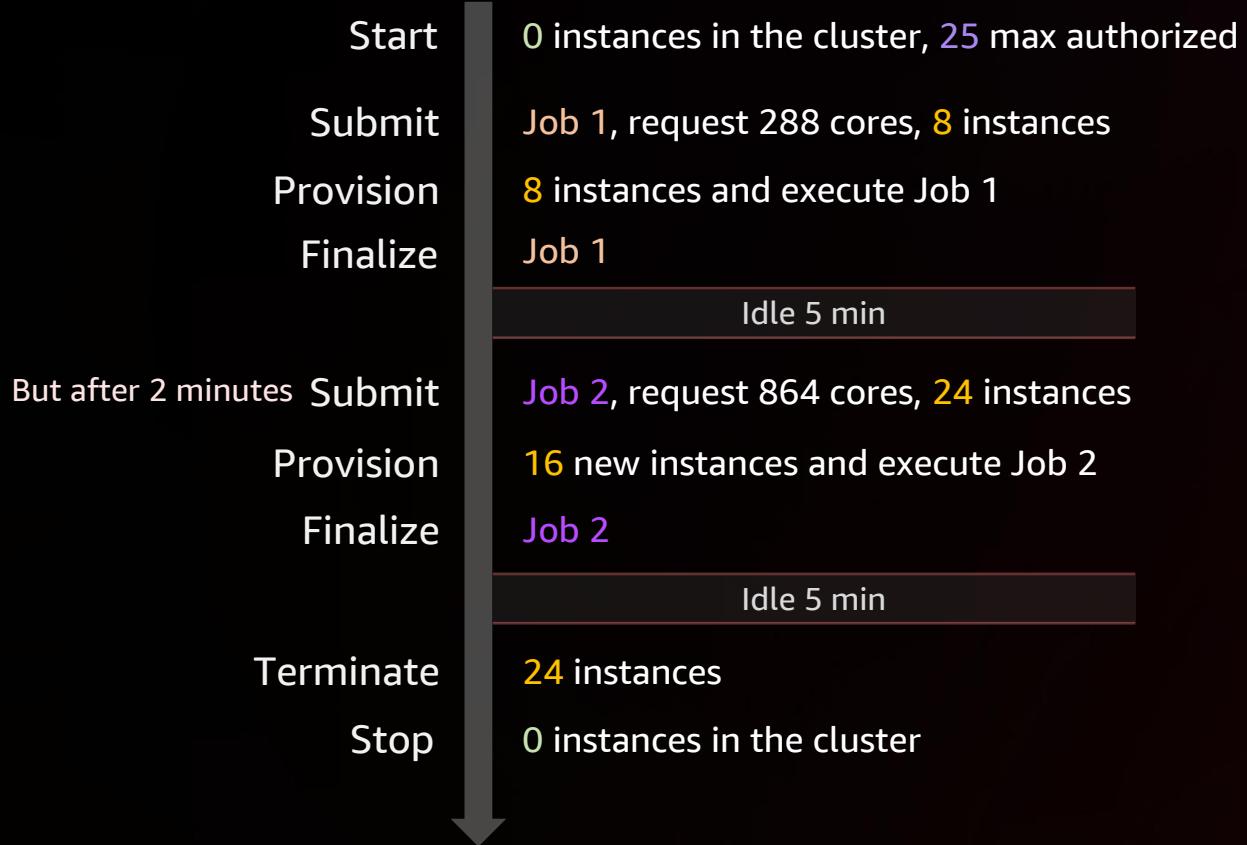
Scaling options

- Manual, schedule, predictive
- Notify on start, stop, terminate . . .

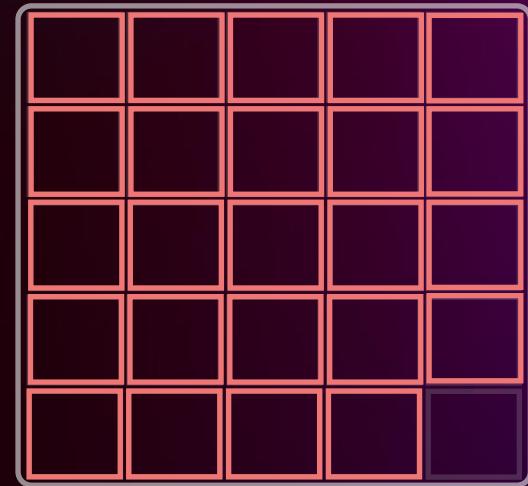


EC2 Auto Scaling compute system?

Imagine that nodes are added when jobs are submitted and removed when they finish



Auto Scaling compute cluster



Min: 0 Desired: 24 Max: 25
cores / node: 36

```
aws autoscaling deleteScalingGroup \
  --auto-scaling-group-name my-auto-scaling-group \
  --desiredCapacity 0 \
  --region us-east-1 \
  --maxSize 4096 \
  --minSize 0 \
  --vpcZoneType=bySubnet \
  --subnet-ids="subnet-5ea0c127"
```

Putting it all together

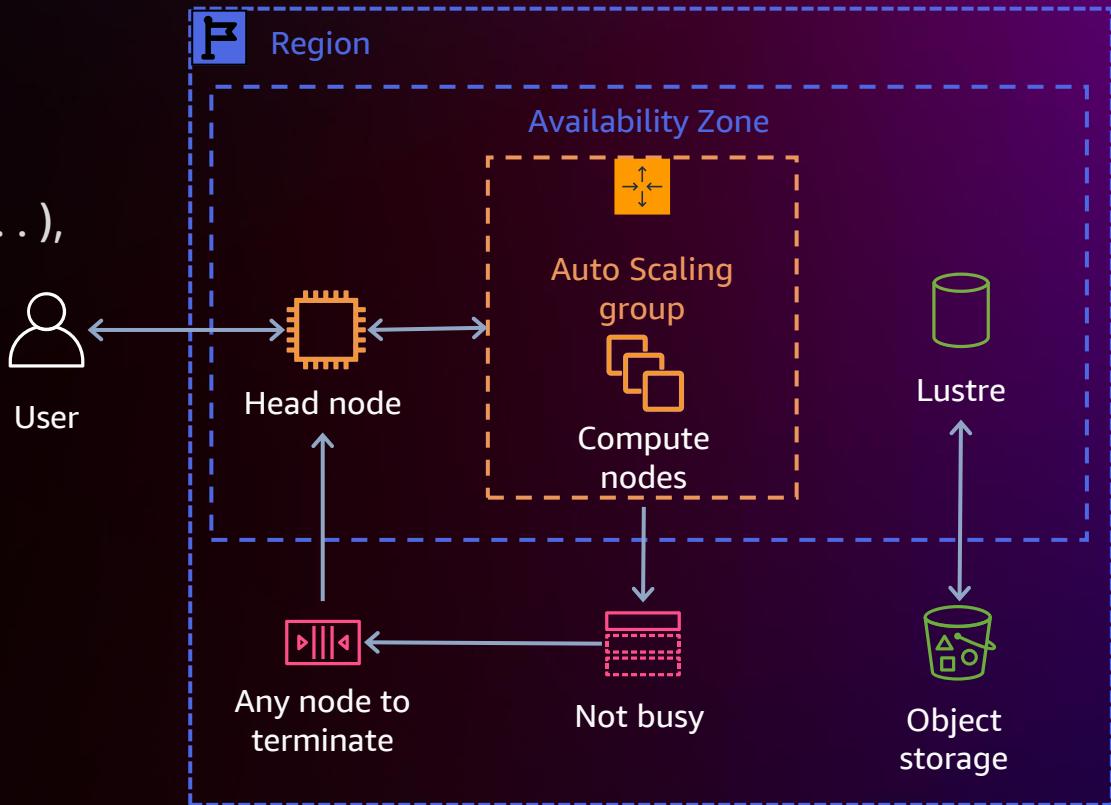
Building an auto-scaling HPC cluster

- Similar to on premises but with auto scaling
- Still a classical HPC system with a scheduler (SLURM, SGE, . . .), Lustre, and placement groups (tightly coupled)
- The same familiar interface with an elastic capacity

Additional technical considerations

When ready, instances send notifications to a message queuing service – the scheduler watches this queue and adds the compute nodes as they appear

When not busy, they will lock themselves up, check the scheduler queue, send a notification, and terminate



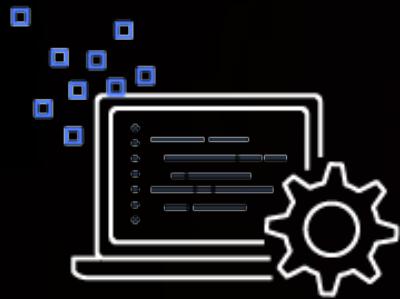
Part of the next hands-on



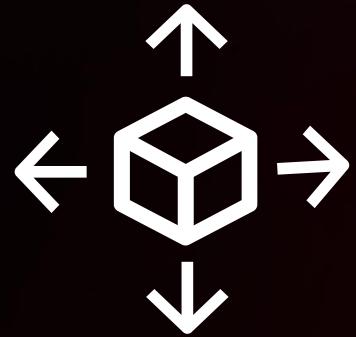
AWS ParallelCluster

Open-source cluster management for HPC on AWS

Why use AWS ParallelCluster?



Easy cluster management



Automatic resource scaling



Seamless migration to the cloud

Easy cluster management

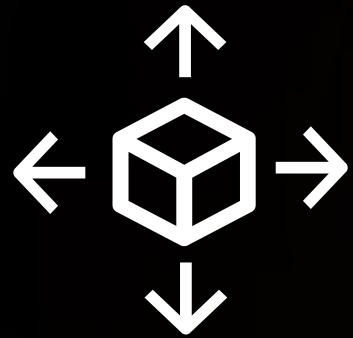


“pcluster configure” to set up a cluster in minutes

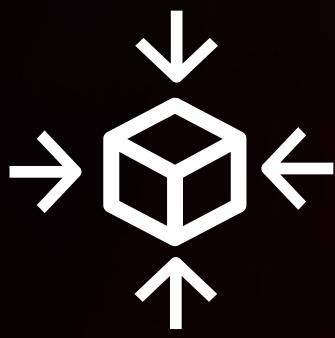
Use config files to define details of replicable clusters

Launch, stop, and restart clusters on demand

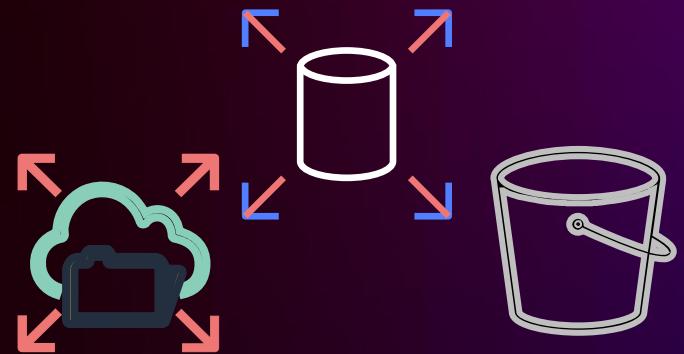
Automatic resource scaling



Scale up when jobs
are waiting

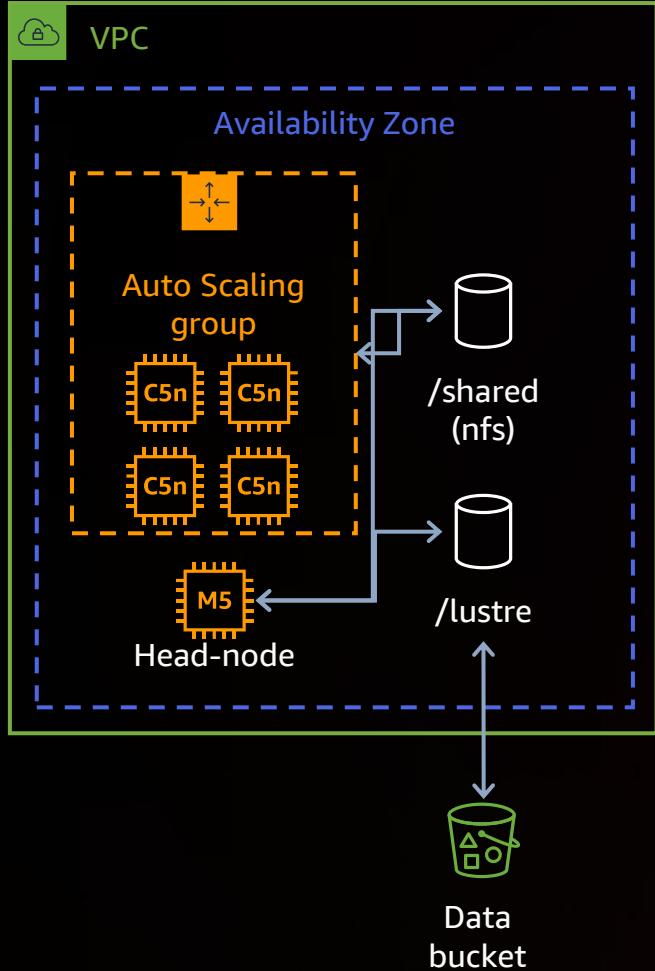


Scale down when the
cluster is idle



Your data storage and file system
scale to match your compute

Simple architecture



- Post install configuration
 - Install applications scripts
- Amazon EBS snapshot bootstrap
 - Application installations or static configurations
- Other details
 - Amazon EFS can be shared across clusters
 - Lustre partition can be mounted but per AZ
 - Public/private subnets for head/compute
 - Link to AD for user mapping if required

Example of configuration file

```
Region: [AWS_REGION]
Image:
  Os: alinux2
SharedStorage:
  - Name: Ebs0
    StorageType: Ebs
    MountDir: /shared
    EbsSettings:
      VolumeType: gp3
      DeletionPolicy: Delete
      Size: '100'

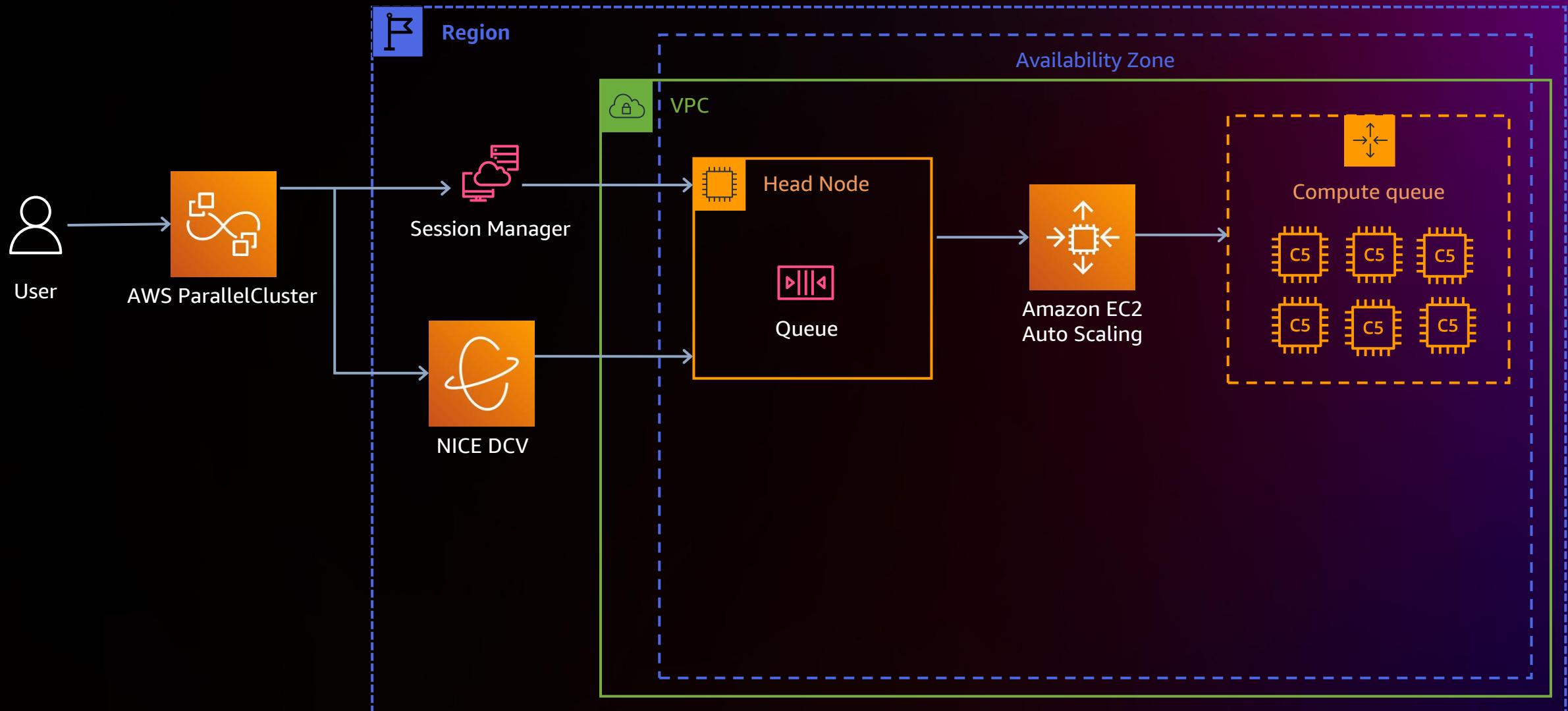
HeadNode:
  InstanceType: c5.xlarge
  Ssh:
    KeyName: [KEY_NAME]
  Networking:
    SubnetId: [subnet-12345689]
  Dcv:
    Enabled: true
  LocalStorage:
    RootVolume:
      VolumeType: gp2

Scheduling:
  Scheduler: slurm
  slurmQueues:
    - Name: compute-queue
      ComputeResources:
        - Name: hpc6a
          MinCount: 0
          MaxCount: 4
  InstanceType: hpc6a.48xlarge
  Efa:
    Enabled: true
  DisableSimultaneousMultithreading: true
  Networking:
    SubnetIds:
      - [subnet-12345689]
  PlacementGroup:
    Enabled: true
```

Hands-on: Create a HPC cluster using AWS ParallelCluster



Architecture for this workshop



Getting started with this workshop

- As a participant, you will have access to an AWS account with any optional pre-provisioned infrastructure and IAM policies needed to complete this workshop
- The AWS account will only be available for the duration of this workshop; you will lose access to the account thereafter
- The optional pre-provisioned infrastructure will be deployed to a specific region; check your workshop content to determine whether other regions will be used
- Be sure to review the terms and conditions of the event; do not upload any personal or confidential information in the account

Step 1: Sign-in via your preferred method

<https://catalog.workshops.aws/join>

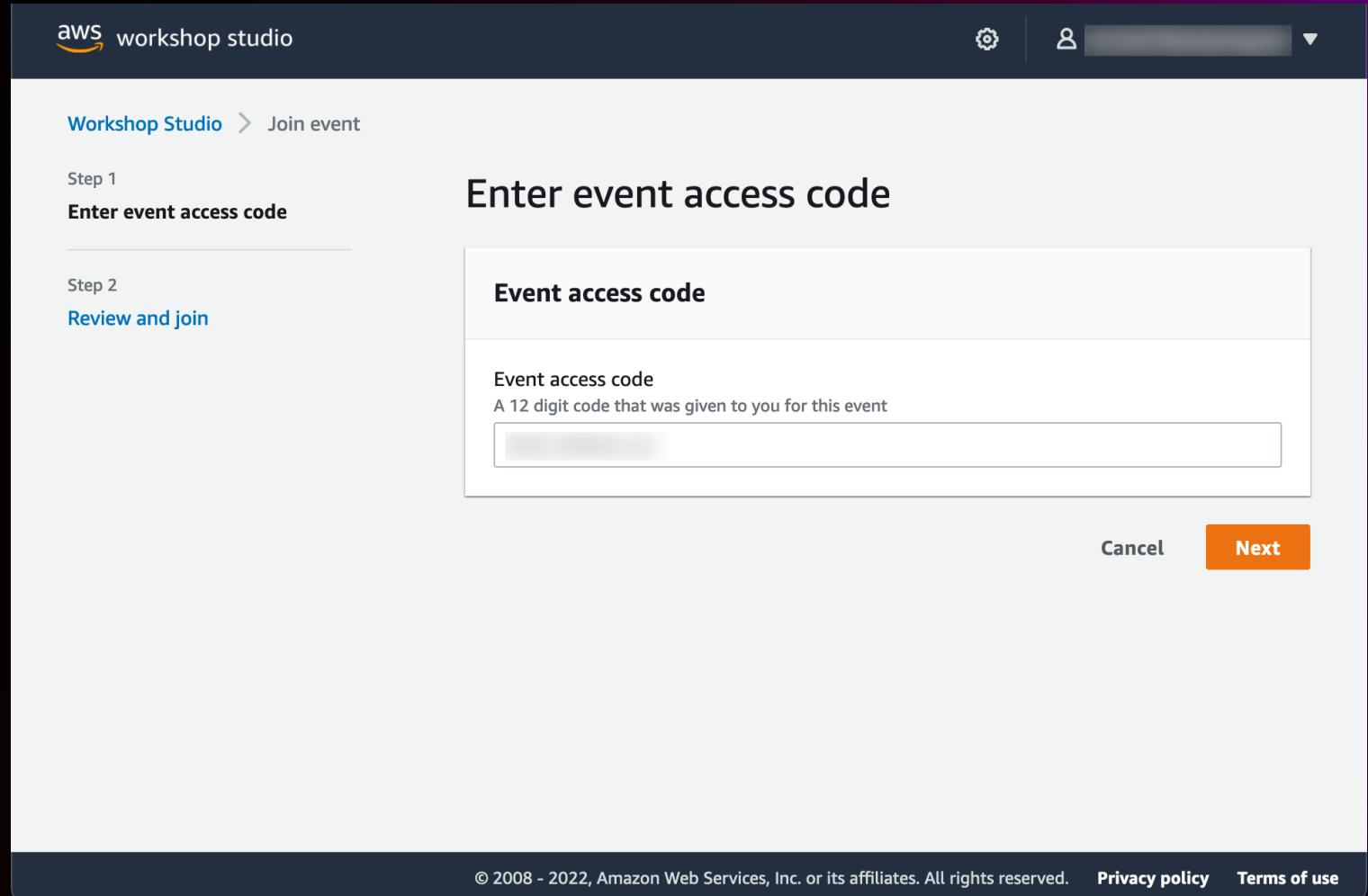


A screenshot of the AWS Workshop Studio sign-in interface. The top navigation bar includes the AWS logo and the text "workshop studio". Below the navigation, the text "Workshop Studio > Sign in" is displayed. A central box is titled "Sign in" with the sub-instruction "Choose a preferred sign-in method". It features three options: "Email one-time password (OTP)" (highlighted with an orange background), "Login with Amazon" (with a light gray background), and "Amazon employee" (with a light gray background). Each option has a corresponding descriptive text below it. At the bottom of the page, a dark footer bar contains the text "© 2008 - 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved." and links for "Privacy policy" and "Terms of use".

Step 2: Enter event access code

Enter 12-digit event access code

If you were given a one-click join link, you can skip this step



The screenshot shows the AWS Workshop Studio interface. The top navigation bar includes the AWS logo, 'workshop studio', a gear icon, a user icon, and a dropdown menu. The main navigation on the left shows 'Workshop Studio' and 'Join event'. The main content area is titled 'Step 1 Enter event access code' and 'Step 2 Review and join'. To the right, a large callout box is titled 'Enter event access code' and contains a section for 'Event access code' with the sub-instruction 'A 12 digit code that was given to you for this event'. At the bottom right of the callout are 'Cancel' and 'Next' buttons. The footer of the page includes the copyright notice '© 2008 - 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.' and links for 'Privacy policy' and 'Terms of use'.

Step 3: Review terms and join event

The screenshot shows the 'Review and join' step of the AWS Workshop Studio 'Join event' process. The page is titled 'Review and join' and displays 'Event details' and 'Terms and Conditions'.

Event details

Name	Start time	Duration	Level
AWS General Immersion Day	9/23/2022 01:13 AM	12 hours	-

Terms and Conditions
Read and accept before joining the event

1. By using AWS Workshop Studio for the relevant event, you agree to the AWS Event Terms and Conditions and the AWS Acceptable Use Policy. You acknowledge and agree that are using an AWS-owned account that you can only access for the duration of the relevant event. If you find residual resources or materials in the AWS-owned account, you will make us aware and cease use of the account. AWS reserves the right to terminate the account and delete the contents at any time.
2. You will not: (a) process or run any operation on any data other than test data sets or lab-approved materials by AWS, and (b) copy, import, export or otherwise create derivative works of materials provided by AWS, including but not limited to, data sets.
3. AWS is under no obligation to enable the transmission of your materials through Event Engine and may, in its discretion, edit, block, refuse to post, or remove your materials at any time.
4. Your use of AWS Workshop Studio will comply with these terms and all applicable laws, and your access to AWS Workshop Studio will immediately and automatically terminate if you do not comply with any of these terms or conditions.

I agree with the Terms and Conditions

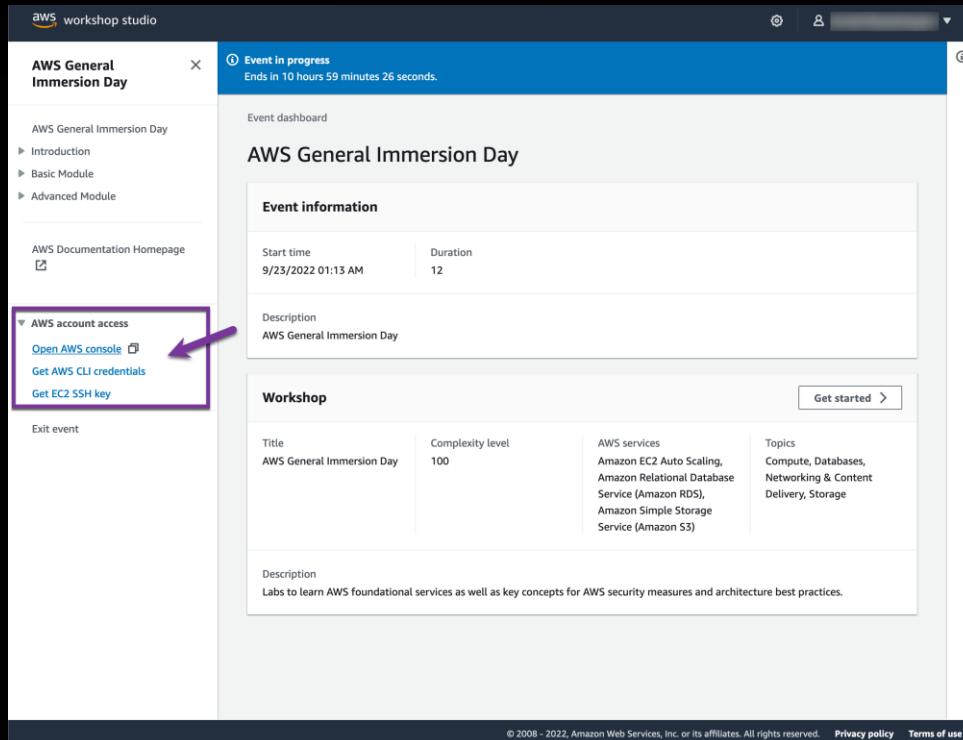
Cancel Previous Join event

© 2008 - 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved. [Privacy policy](#) [Terms of use](#)



Step 4: Access AWS account

Access the AWS Management Console or generate AWS CLI credentials as needed



AWS General Immersion Day

Event in progress
Ends in 10 hours 59 minutes 26 seconds.

Event dashboard

AWS General Immersion Day

Event information

Start time	9/23/2022 01:13 AM	Duration	12
------------	--------------------	----------	----

Description

AWS General Immersion Day

Workshop

Title	AWS General Immersion Day	Complexity level	100
AWS services	Amazon EC2 Auto Scaling, Amazon Relational Database Service (Amazon RDS), Amazon Simple Storage Service (Amazon S3)	Topics	Compute, Databases, Networking & Content Delivery, Storage

Workshop

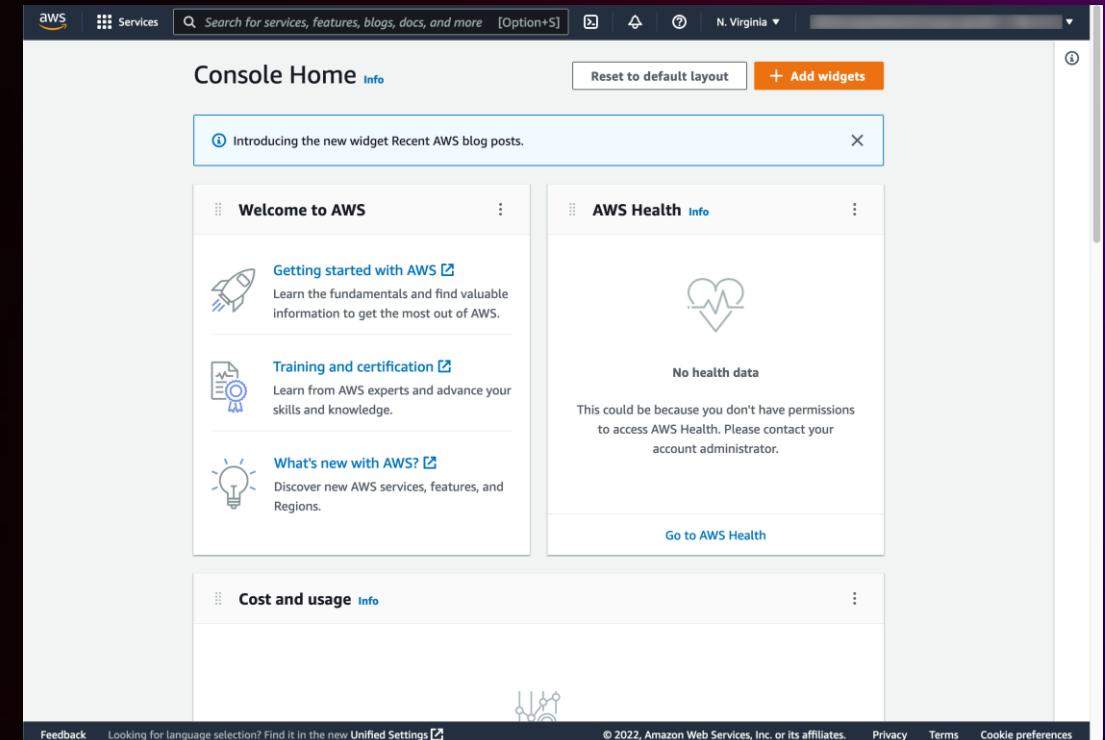
Get started >

AWS account access

- Open AWS console
- Get AWS CLI credentials
- Get EC2 SSH key

Exit event

© 2008 - 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved. [Privacy policy](#) [Terms of use](#)



Console Home

Search for services, features, blogs, docs, and more [Option+S]

Reset to default layout + Add widgets

Welcome to AWS

- Getting started with AWS
- Training and certification
- What's new with AWS?

AWS Health

No health data

This could be because you don't have permissions to access AWS Health. Please contact your account administrator.

Go to AWS Health

Cost and usage

Feedback Looking for language selection? Find it in the new Unified Settings

© 2022, Amazon Web Services, Inc. or its affiliates. [Privacy](#) [Terms](#) [Cookie preferences](#)



Step 5: Get started with the workshop

AWS workshop studio

AWS General Immersion Day

Event in progress
Ends in 10 hours 59 minutes 26 seconds.

Event dashboard

AWS General Immersion Day

Event information

Start time	Duration
9/23/2022 01:13 AM	12

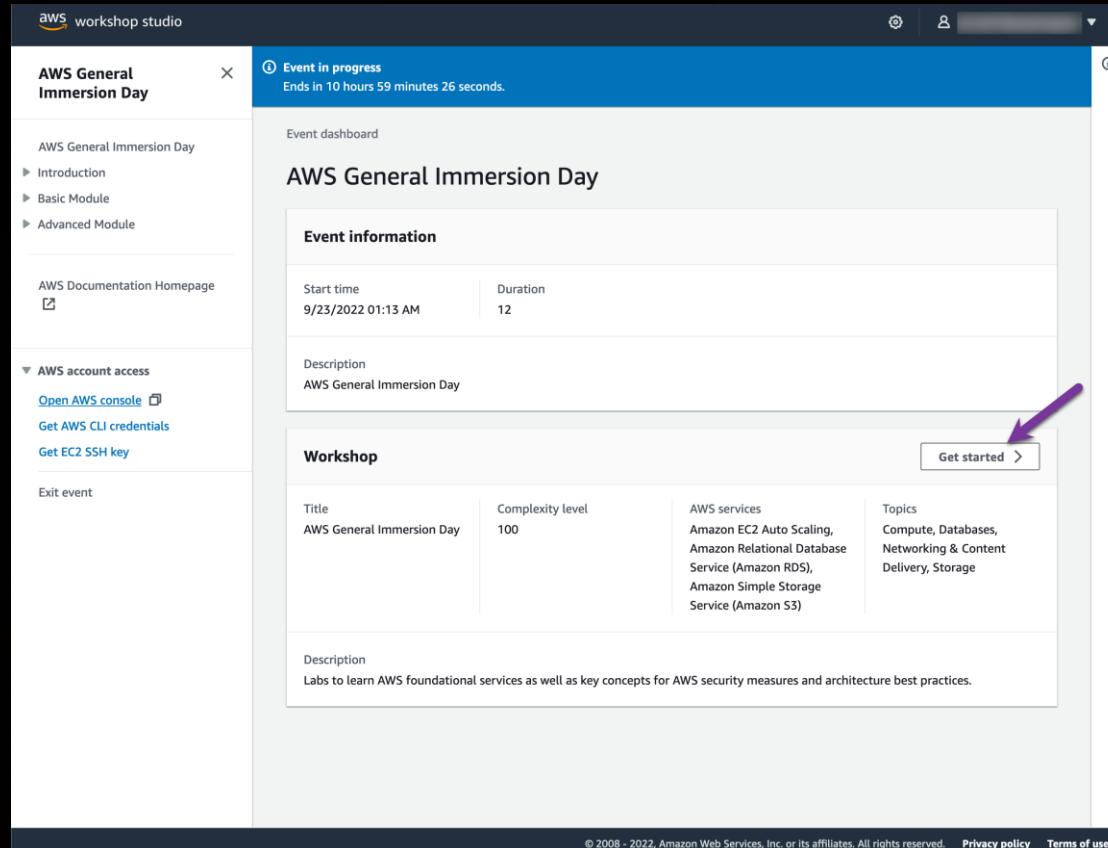
Description
AWS General Immersion Day

Workshop

Title	Complexity level	AWS services	Topics
AWS General Immersion Day	100	Amazon EC2 Auto Scaling, Amazon Relational Database Service (Amazon RDS), Amazon Simple Storage Service (Amazon S3)	Compute, Databases, Networking & Content Delivery, Storage

Description
Labs to learn AWS foundational services as well as key concepts for AWS security measures and architecture best practices.

Get started >



© 2008 - 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved. [Privacy policy](#) [Terms of use](#)



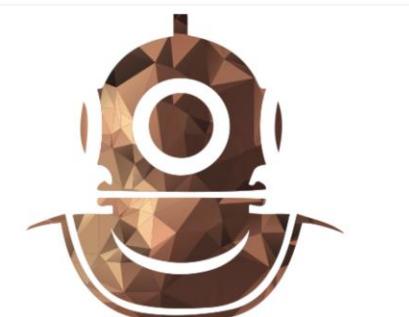
AWS workshop studio

AWS General Immersion Day

Event in progress
Ends in 10 hours 57 minutes 12 seconds.

Event dashboard > AWS General Immersion Day

AWS General Immersion Day



IMMERSION DAYS

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](#)

© 2008 - 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved. [Privacy policy](#) [Terms of use](#)



Amazon FSx for Lustre



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Storage should not be the bottleneck for compute-intensive workloads

Faster storage reduces time to results and overall TCO by saving compute costs

Vast datasets

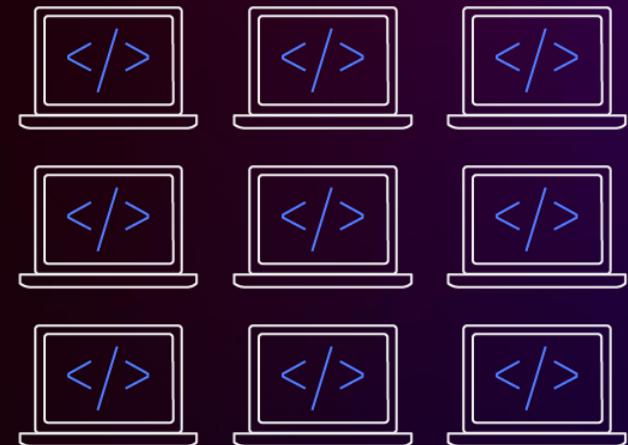


Ability to serve data quickly:

Fast storage
High-speed network
Low latency



Scale out compute capacity
(hundreds to millions of cores)



Amazon FSx for Lustre



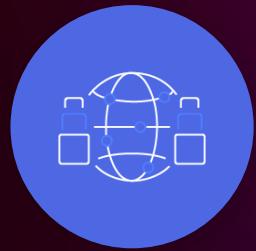
Fully managed Lustre file system for high-performance workloads



Massively scalable
performance



Seamless access to
your data repositories



Simple and
fully managed



Native file
system interface



Cost optimized for
compute-intensive workloads



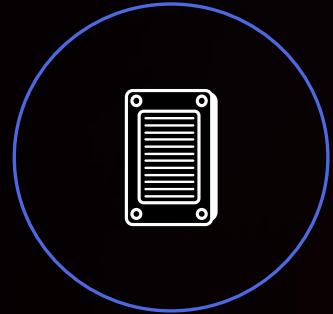
Secure
and compliant

High, scalable performance



High, scalable performance

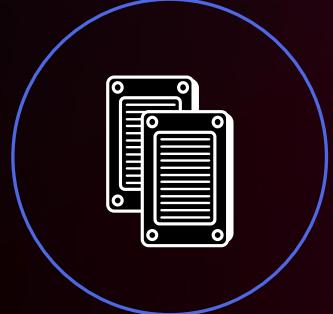
SSD



Scratch

Short-term processing
Spin up → process → spin down
Single copy of data

SSD or HDD

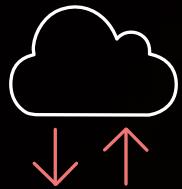


Persistent

Longer-term processing
Highly available (HA) file servers
Replicated copies of data

Amazon FSx for Lustre control plane (API, management layer, file system control) is designed to be HA for both options

What FSx for Lustre brings to your workloads



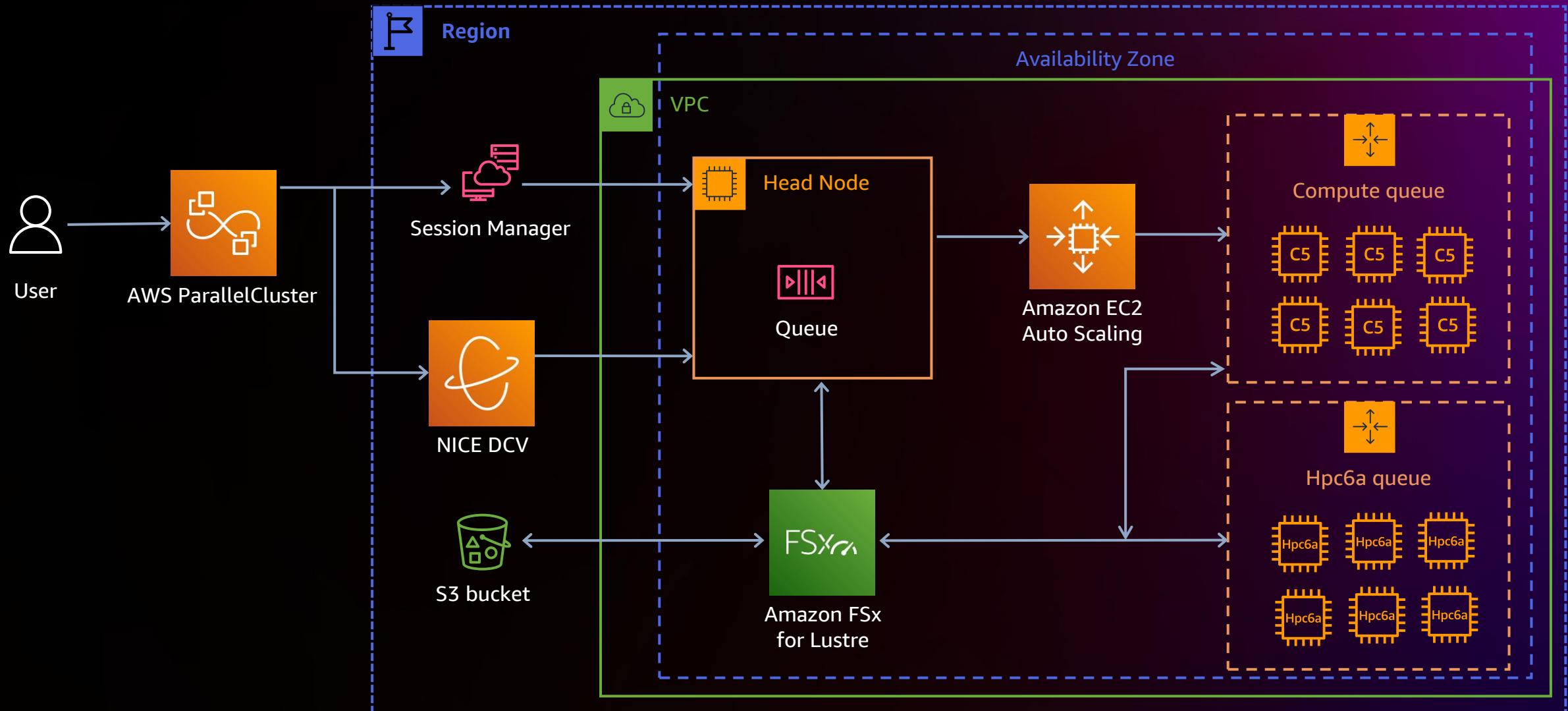
Access to Amazon S3 data through a fast file interface



Hands-on: HPC cluster update to use Amazon FSx for Lustre



Architecture for this workshop



Thank you!

Maxime Hugues

maxhaws@amazon.com

Lowell Wofford



Please complete the session
survey in the **mobile app**