AWS re:Invent

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

CMP305-R1

Getting started with HPC on AWS

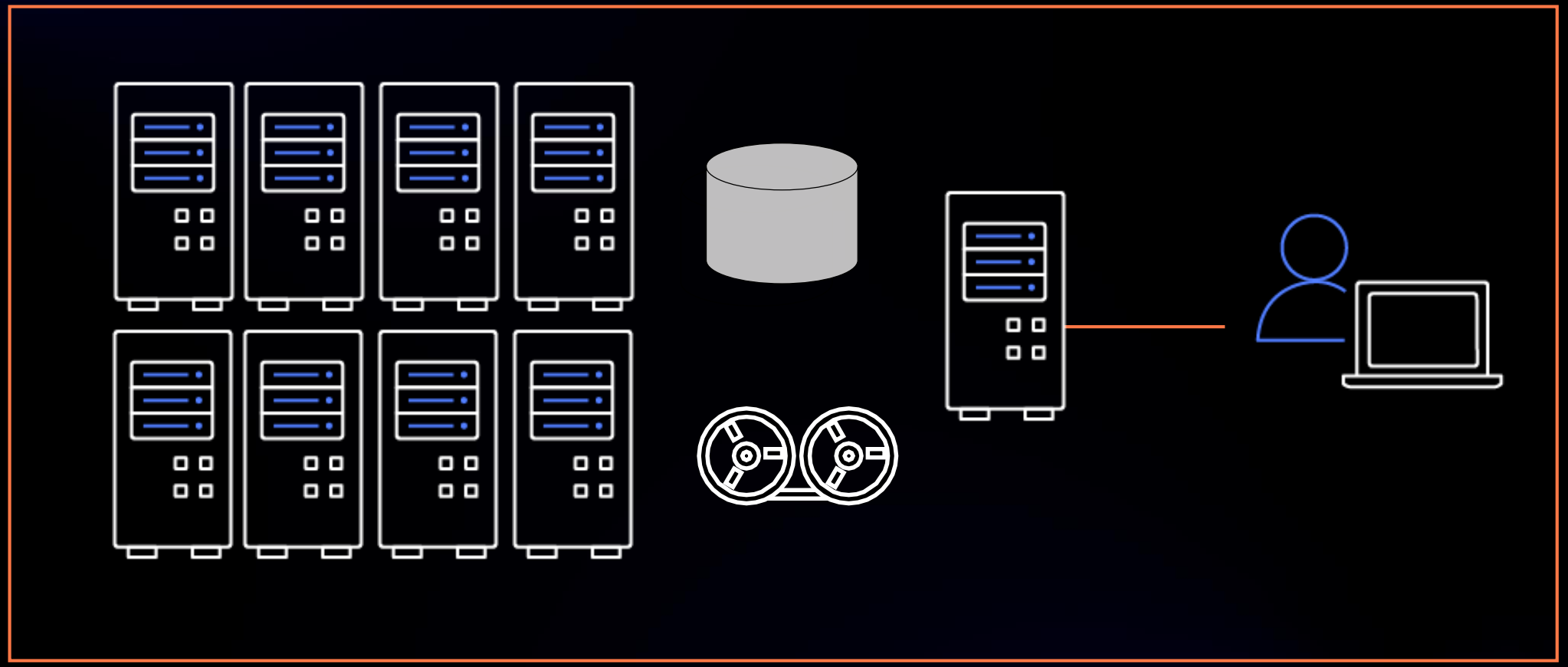
Pierre-Yves Aquilanti, PhD
Principal Specialist Solutions Architect, HPC
Amazon Web Services

Anh Tran
Senior Manager, Specialist Solutions Architect, HPC
Amazon Web Services

Charlie Gruenwald, PhD
Senior Software Development Engineer, HPC
Amazon Web Services

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

Familiar view of an HPC system



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

Base infrastructure

Amazon VPC and subnets

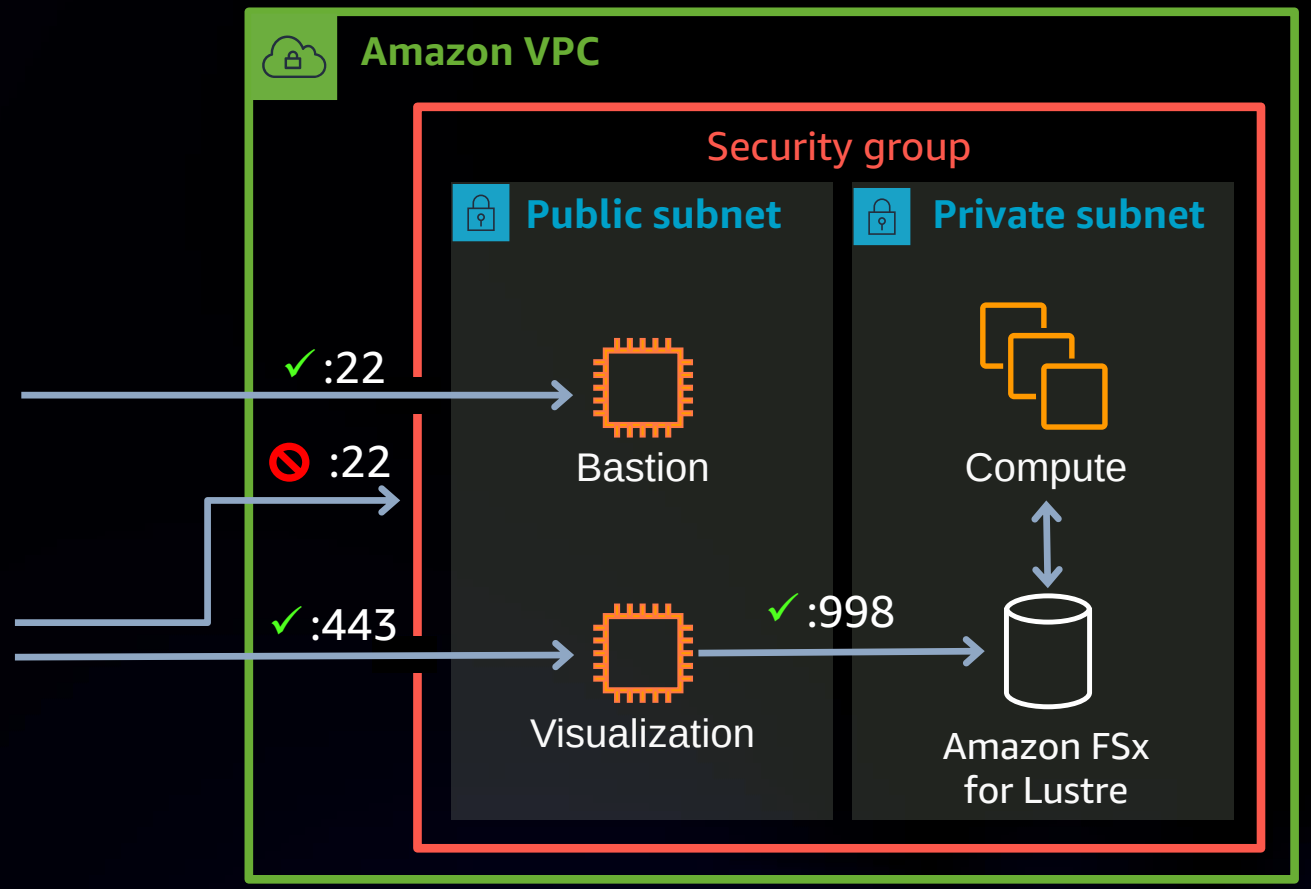
- Amazon Virtual Private Cloud: logical section of the cloud provider infrastructure
- Subnet: logical partition of a network

Security groups

- Virtual firewalls
- Amazon VPC and instances

Instances and services

- Instances
- Managed services
- . . .



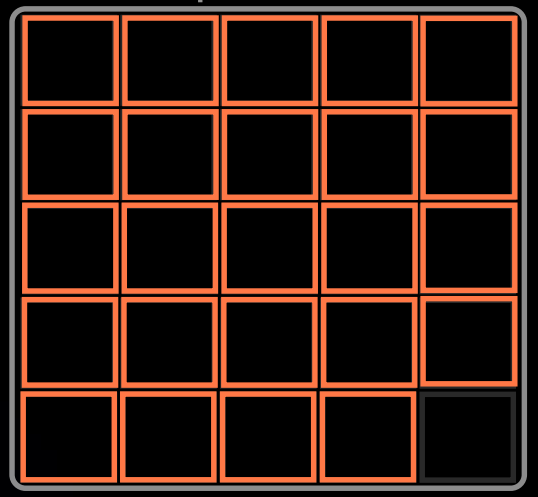
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon EC2 Auto Scaling compute system

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

	Start	0 instances in the cluster, 25 max authorized
	Submit	Job 1, request 288 cores, 8 instances
	Provision	8 instances and execute Job 1
	Finalize	Job 1
		Idle 5 min
But after 2 minutes	Submit	Job 2, request 864 cores, 24 instances
	Provision	16 new instances and execute Job 2
	Finalize	Job 2
		Idle 5 min
	Terminate	24 instances
	Stop	0 instances in the cluster

Amazon EC2 Auto Scaling compute cluster



Min: 0 **Desired: 24** Max: 25

cores / node: 36

```
aws autoscaling set-desired-capacity \
    --auto-scaling-group-name my-asg-scaling-group \
    --desired-capacity 24
```

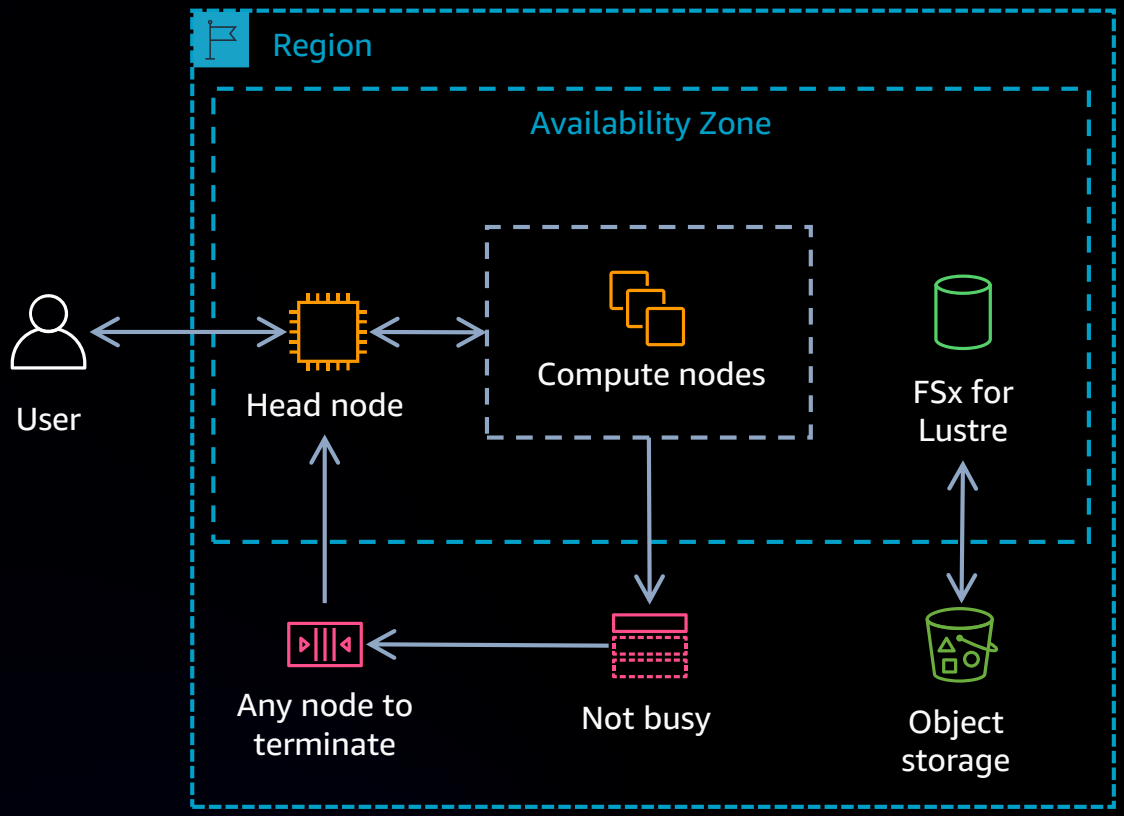
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Putting it all together

Building an auto scaling HPC cluster

- Still a classical HPC system with a scheduler (SLURM), Lustre, and placement groups (tightly coupled)
- The same familiar interface with an elastic capacity

AWS CloudFormation does the heavy lifting on your behalf to build and configure your clusters



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

Example of configuration file

```
Region: us-east-2

Image:
  Os: alinux2

HeadNode:
  InstanceType: c5n.18xlarge
  Networking:
    SubnetId: subnet-xxxxxxx
  Ssh:
    KeyName: hpc

SharedStorage:

  - FsxLustreSettings:
      StorageCapacity: 1200
    MountDir: /fsx
    Name: fsx
    StorageType: FsxLustre
```

Operating system

Head node configuration

Shared file system

```
Scheduling:
  Scheduler: slurm
  SlurmQueues:
    - Name: c5n-queue
      Networking:
        SubnetIds:
          - subnet-xxxxxxx
        PlacementGroup:
          Enabled: True
      CapacityType: ONDEMAND
      ComputeResources:
        - Name: c5nresources
          InstanceType: c5n.18xlarge
          DisableSimultaneousMultithreading: true
          MinCount: 0
          MaxCount: 5
          Efa:
            Enabled: true
```

Compute queue

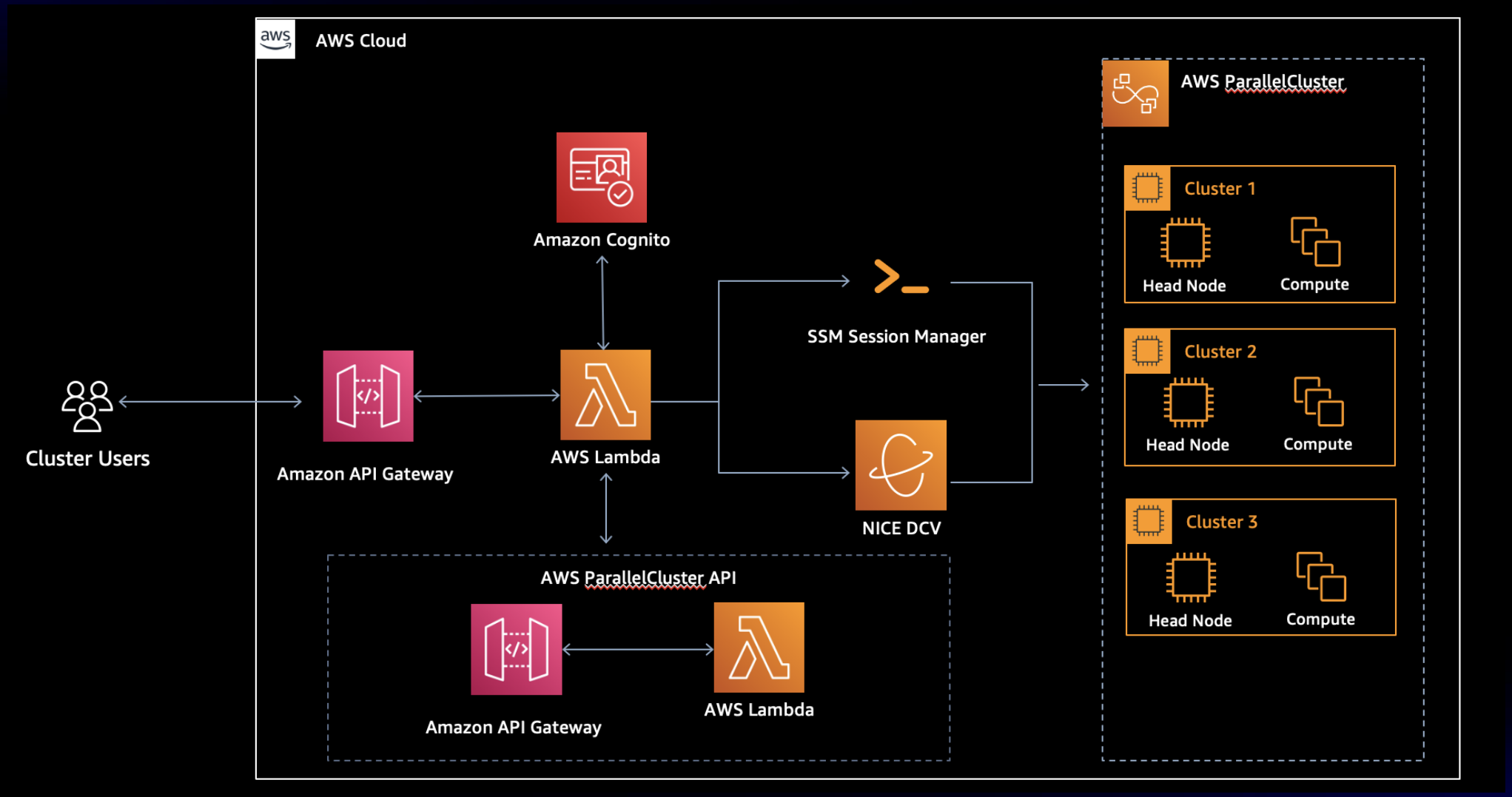
But today we are using a Web UI

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

Hands on

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

What you will build today

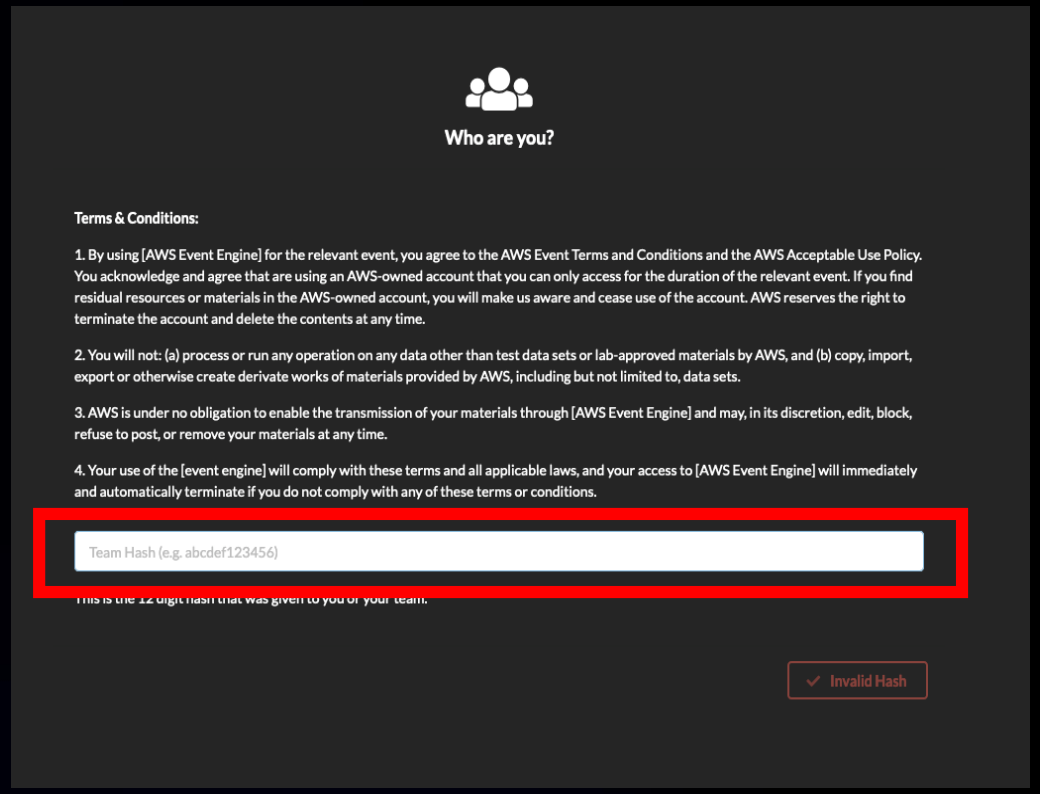


© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Log into your account

Use the 12-character hash to log into your account

a364-16a07736d4-1b

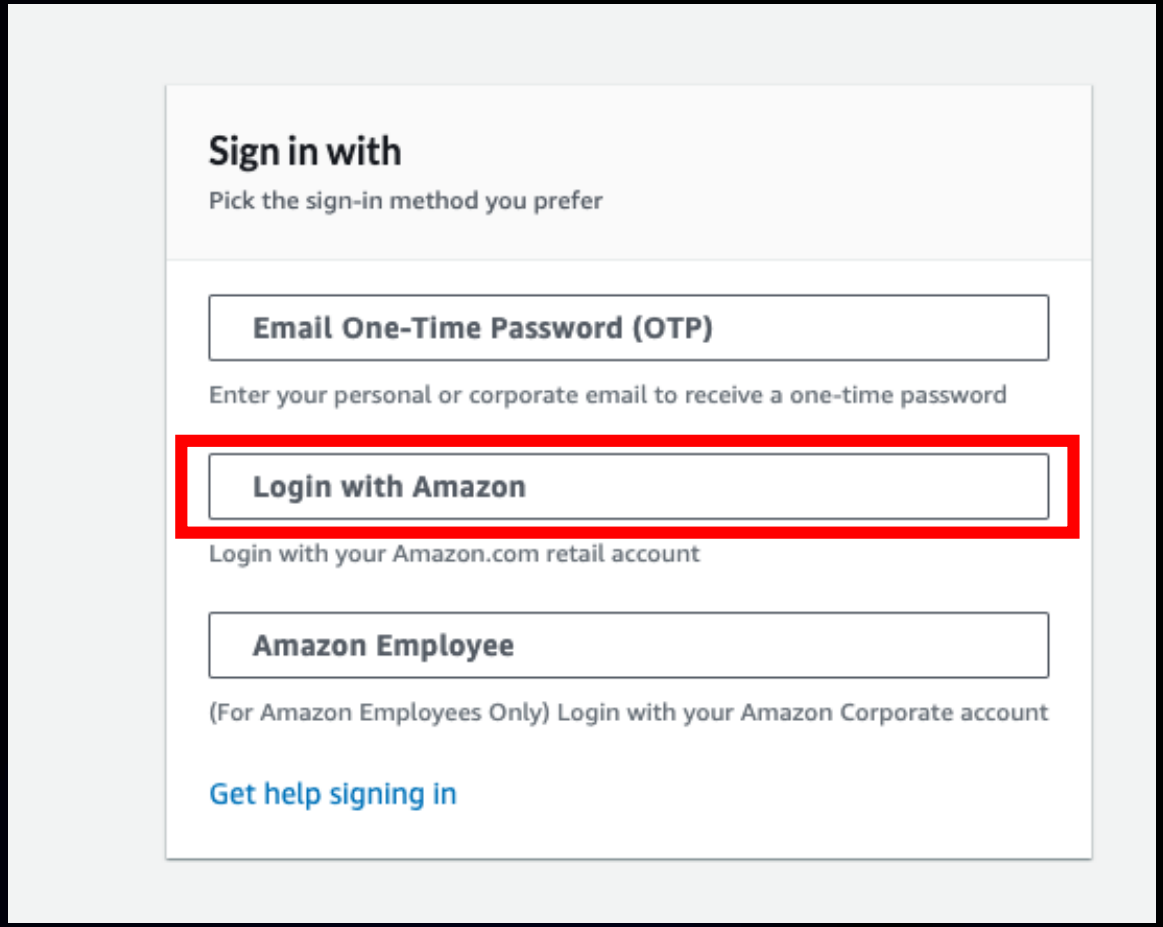


https://dashboard.eventengine.run

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

Log in with Amazon

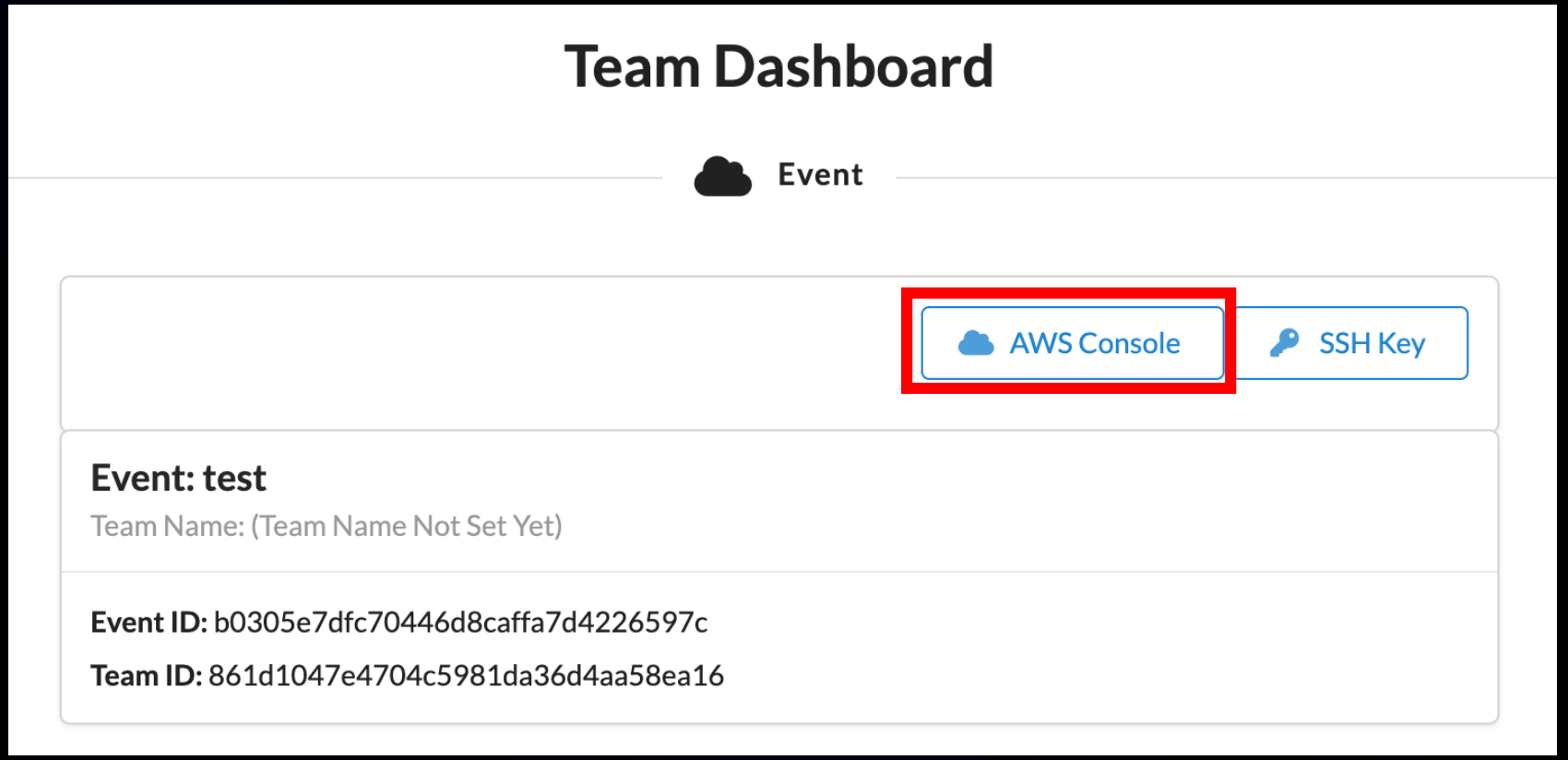
Choose login with Amazon



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

Access the console 1/2

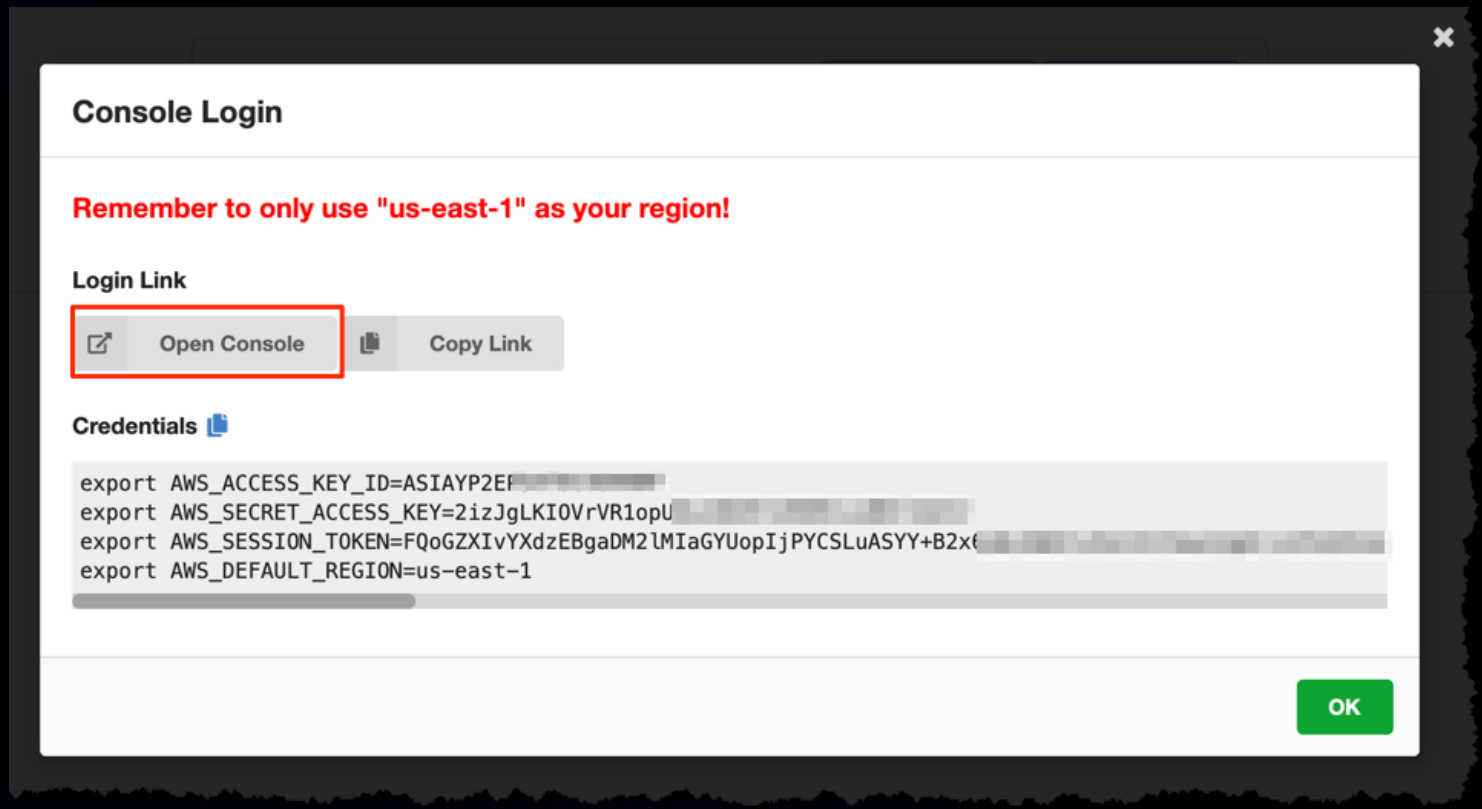
Click AWS Console



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

Access the console 2/2

Click on Open Console to access your AWS account

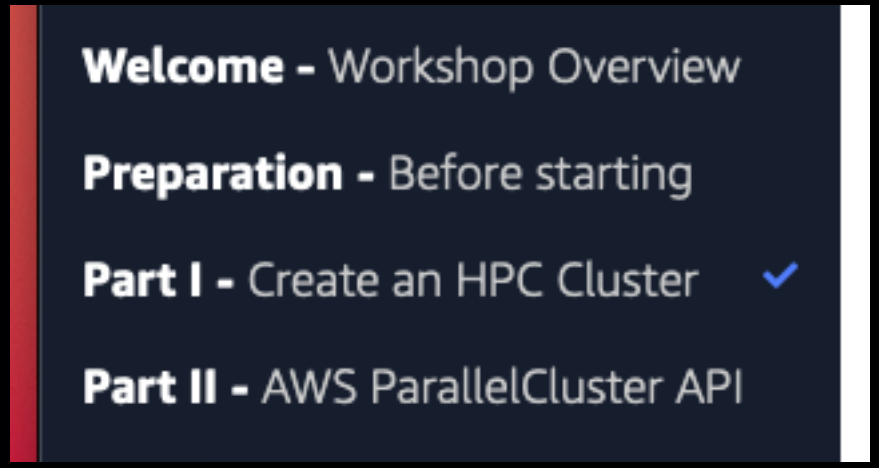


© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Starting the hands-on part

Objectives for this lab

1. Create a cluster with PCluster Manager
2. Run an application, observe the scaling
3. Interact with the AWS ParallelCluster API



Login: Use your 12-character hash **a364-16a07736d4-1b**

https://dashboard.eventengine.run

Lab: Connect to http://ri21.hpcworkshops.com

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

Thank you!

Pierre-Yves Aquilanti, PhD
pierreya@amazon.com

Anh Tran
trnh@amazon.com

Charlie Gruenwald, PhD
cgruenwa@amazon.com

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