



LUG 2023 | MAY 1-4, 2023

Managing cloud HPC with infrastructure-as-code

Matt Vaughn

Principal Developer Advocate
HPC Engineering
Amazon Web Services

© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.



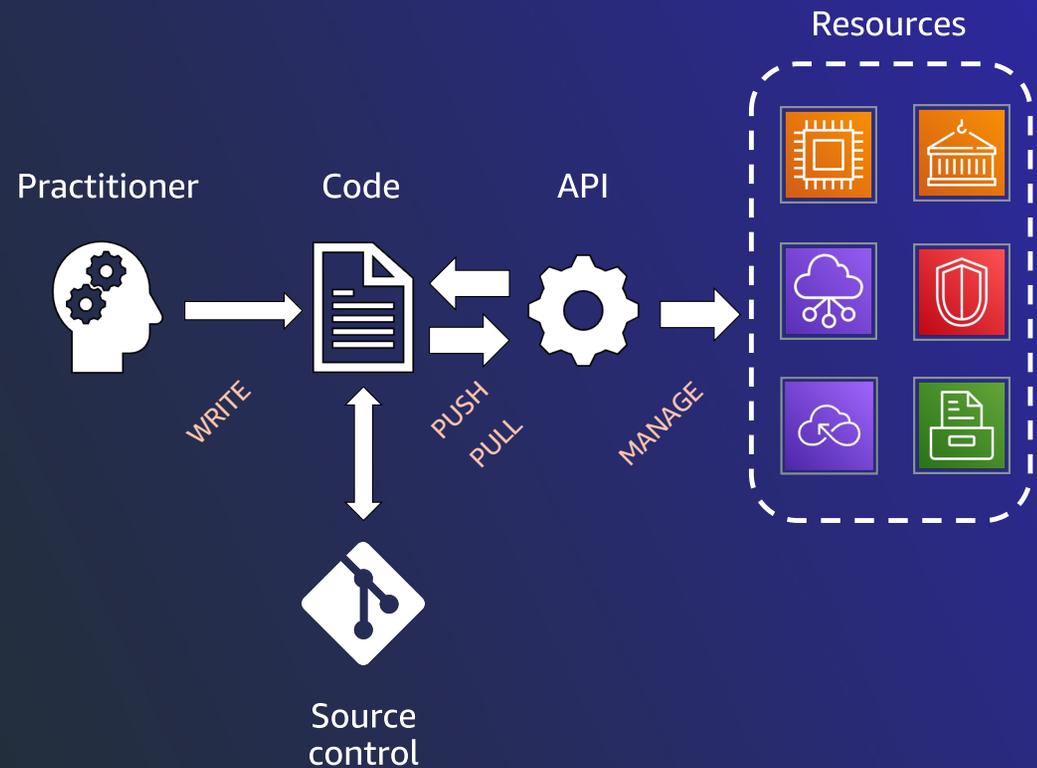
1. What is infrastructure as code?
2. Key IaC technologies
3. Interactive vs managed deployments
4. Infrastructure by composition
5. Exemplar HPC infrastructure as code
6. Use case 1: Complex compute environment
7. Use case 2: HPC-Ops
8. Summary and conclusion



What is Infrastructure as Code (IaC)?

Managing and provisioning of infrastructure through code instead of manual processes.

- Practitioner writes code
- Code managed under source control
- Push/pull to automation server
- Resources managed declaratively



Some key IaC technologies

AWS CloudFormation

AWS CDK

Terraform

CDKTF

Pulumi

Ansible/Chef/Puppet/Salt*



How does this apply to HPC?



© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.



AWS FSx Console

Create file system

File system details

File system name - optional [Info](#)

Maximum of 256 Unicode letters, whitespace, and numbers, plus + - = . _ : /

Deployment and storage type [Info](#)

Select a deployment type and storage type to fit your workload requirements

Persistent, SSD

Persistent, HDD

with SSD cache

Scratch, SSD

Throughput per unit of storage [Info](#)

Throughput (MB/s) per unit of storage (TiB)

125 MB/s/TiB

250 MB/s/TiB

500 MB/s/TiB

1000 MB/s/TiB

Storage capacity [Info](#)

TiB

Supported sizes: 1.2 TiB or increments of 2.4 TiB

Throughput capacity [Info](#)

Throughput capacity = Storage capacity (TiB) * Per unit storage throughput (MB/s)

240000 MB/s

Data compression type [Info](#)

Data compression reduces the physical disk space needed to store file data. Select LZ4 to enable data compression

LZ4

AWS CLI v2

```
% aws fsx create-file-system \  
--file-system-type LUSTRE \  
--storage-capacity 240000 \  
--storage-type SSD \  
--subnet-ids subnet-0123456789 \  
--security-group-ids sg-0123456a \  
--lustre-configuration \  
{  
  "DeploymentType": "PERSISTENT_2",  
  "PerUnitStorageThroughput": 1000,  
  "DataCompressionType": "LZ4"  
}
```



[This Photo](#) by Unknown Author is licensed under [CC BY-SA](#)

```
resource "aws_fsx_lustre_file_system" "lugdemo" {  
  deployment_type = "PERSISTENT_2"  
  per_unit_storage_throughput = 1000  
  storage_type = "SSD"  
  data_compression_type = "LZ4"  
  storage_capacity = 240000  
  subnet_ids = ["subnet-0123456789"]  
  security_group_ids = ["sg-0123456a"]  
}
```

Hashicorp Terraform



[This Photo](#) by Unknown Author is licensed under [CC BY-ND](#)

FSxLFilesystem:

Type: **AWS::FSx::FileSystem**

Properties:

FileSystemType: **LUSTRE**

FileSystemTypeVersion: **"2.12"**

StorageType: **SSD**

StorageCapacity: **240000**

SecurityGroupIds:

- **sg-0123456a**

SubnetIds:

- **subnet-0123456789**

LustreConfiguration:

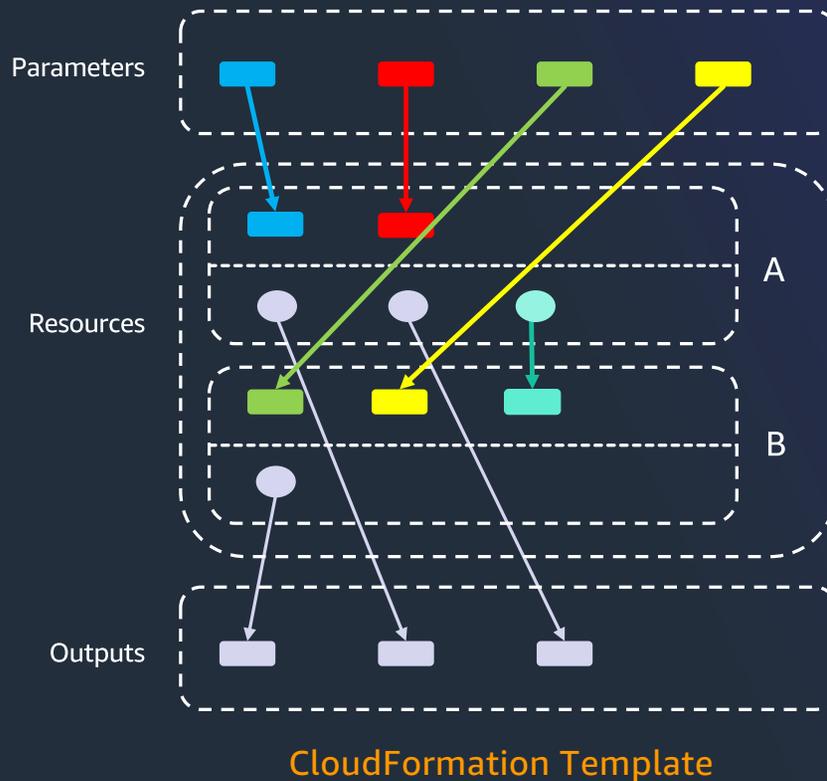
DataCompressionType: **LZ4**

DeploymentType: **PERSISTENT_2**

PerUnitStorageThroughput: **1000**

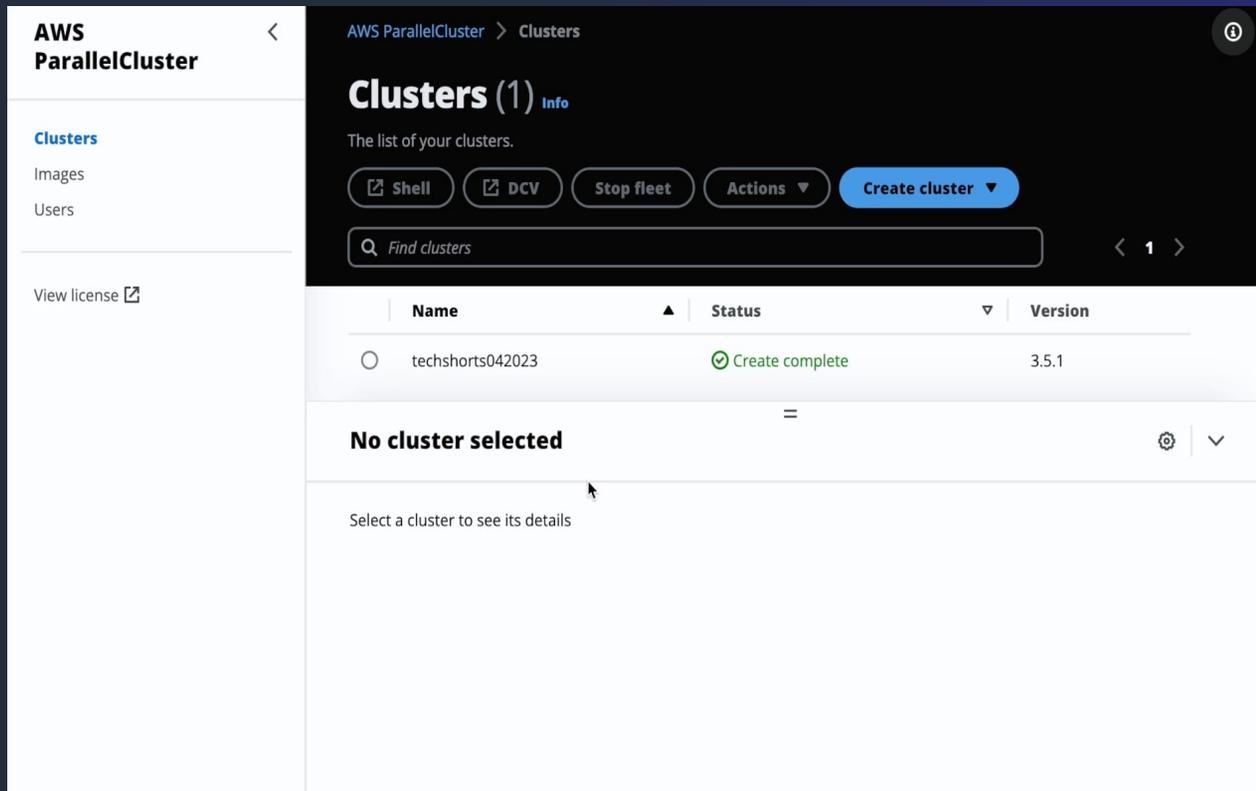
AWS CloudFormation

Infrastructure via composition



- A CloudFormation template deploys a *stack*
- Each template has Parameters, *Resources*, Outputs, Mappings, & Conditionals
- Resources are AWS (and other) types, including other CloudFormation stacks
- Resources have Properties which can be defined by parameters or by properties of other resources
- Outputs from nested stacks can be as parameters and property values for dependent resources
- CloudFormation service handles order of execution, drift detection, etc.
- **CloudFormation templates are DAGs**

But... how does this apply to HPC?



The screenshot shows the AWS ParallelCluster console. The left sidebar contains navigation links for Clusters, Images, and Users. The main content area displays the 'Clusters (1)' page with a search bar and a table of clusters. The table has columns for Name, Status, and Version. One cluster is listed: 'techshorts042023' with a status of 'Create complete' and version '3.5.1'. Below the table, there is a message 'No cluster selected' and a prompt 'Select a cluster to see its details'.

Name	Status	Version
techshorts042023	Create complete	3.5.1

ParallelCluster is a first-class example of Infrastructure as Code

1. Cluster architecture defined with IaC
2. Under the hood, ParallelCluster uses CloudFormation to implement cluster components
 1. Networking/Security/IAM
 2. Storage (EFS, FSx, EBS, S3)
 3. Compute/Head Node resources
 4. (A lot of other stuff)

```
pcluster create-cluster -n lugdemo-2 -c demo-config.yml
```



ParallelCluster clusters as CloudFormation stacks

CloudFormation > Stacks

Stacks (21)

Active 

View nested

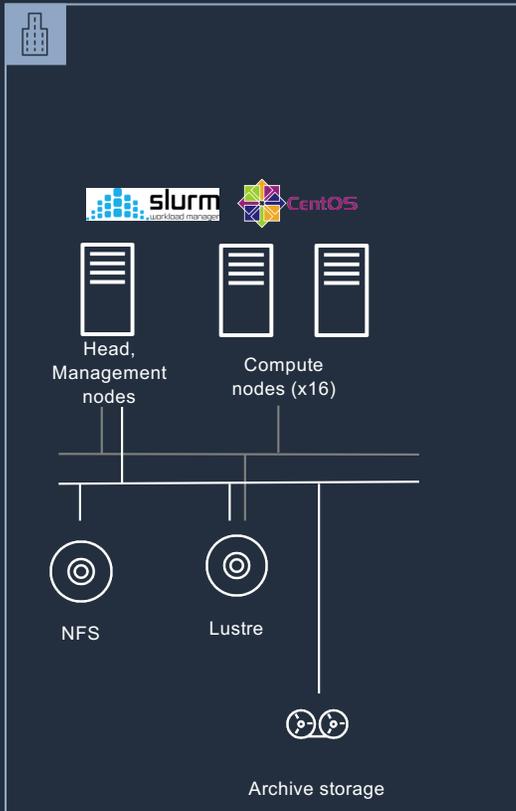
< 1 > 

Stack name	Status	Created time	Description
<input checked="" type="radio"/> lugdemo-02	 UPDATE_COMPLETE	2023-04-24 10:19:32 UTC-0700	-
<input type="radio"/> lugdemo-01	 CREATE_COMPLETE	2023-04-24 06:06:11 UTC-0700	-
<input type="radio"/> fsxlustredemo	 CREATE_COMPLETE	2023-04-24 05:53:41 UTC-0700	Creates an FSxL filesystem of PERSISTENT_2 type plus the Security Group needed for use with ParallelCluster
<input type="radio"/> CDKToolkit	 CREATE_COMPLETE	2023-04-21 11:20:45 UTC-0700	This stack includes resources needed to deploy AWS CDK apps into this environment
<input type="radio"/> techshorts042023	 CREATE_COMPLETE	2023-04-18 16:55:23 UTC-0700	-
<input type="radio"/> pcui-042023-pc351-ParallelClusterApi-FBFJSPZB0BZ8 NESTED	 CREATE_COMPLETE	2023-04-18 10:00:44 UTC-0700	Template for the ParallelCluster API
<input type="radio"/> pcui-042023-pc351	 CREATE_COMPLETE	2023-04-18 10:00:38 UTC-0700	-
<input type="radio"/> parallelcluster-ui-cognito	 CREATE_COMPLETE	2023-02-20 08:52:32 UTC-0800	ParallelCluster UI Cognito User Pool



Cluster definition in a single file

(Physical) cluster design

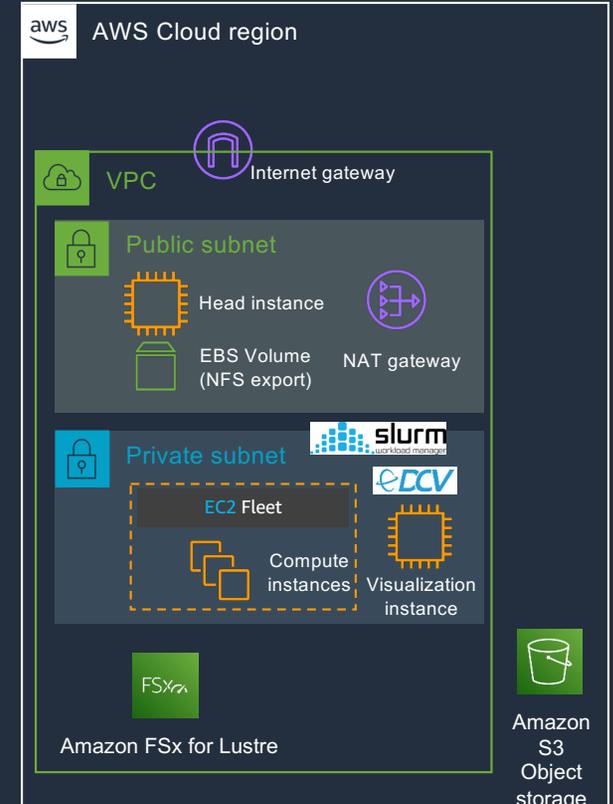


Cluster Configuration File

```
Image:
  Os: centos7
HeadNode:
  InstanceType: c5.4xlarge
  Networking:
    SubnetId: subnet-b46032ec
  Ssh: KeyName: My_PC3_KeyPair
  AllowedIps: 0.0.0.0/0
Scheduling:
  Scheduler: slurm
  SlurmSettings:
    ScaledownIdleTime: 10
  Dns:
    DisableManagedDns: true
SlurmQueues:
  - Name: q1_ondemand
ComputeSettings:
  LocalStorage:
    RootVolume:
      Size: 100
  CapacityType: ONDEMAND
  ComputeResources:
    - Name: compute-resource-1
      InstanceType: c5.n18xlarge
      Efa:
        Enabled: true
        MinCount: 0
        MaxCount: 64
      Networking:
        SubnetIds:
          - subnet-a12321bc
        PlacementGroup:
          Enabled: true
SharedStorage:
  - MountDir: /shared
    Name: myefs
    StorageType: Ebs
    EbsSettings:
      VolumeType: gp3
      Size: 100
  - MountDir: /lustre
    Name: myfsx
    StorageType: FsxLustre
  FsxLustreSettings:
    StorageCapacity: 1200
    DeploymentType: SCRATCH_2
    ImportPath: s3://myhpcbucket
```

OS
Head node
Scheduler Settings
Compute nodes
Network settings
NFS Storage
Lustre storage

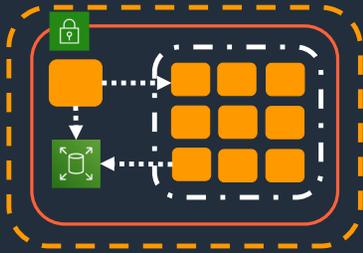
Cluster deployed on AWS



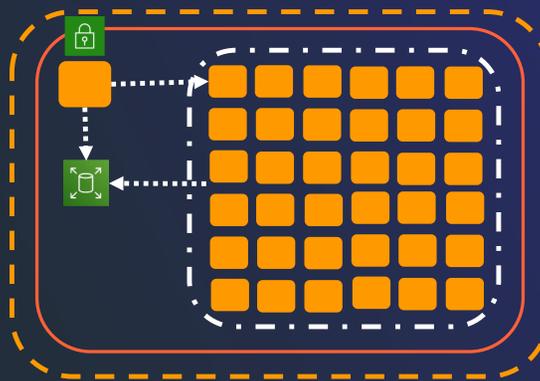
aws AWS Cloud region
VPC Internet gateway
Public subnet Head instance EBS Volume (NFS export) NAT gateway
Private subnet slurm ECV Compute instances Visualization instance
Amazon FSx for Lustre Amazon S3 Object storage

Dynamic compute resource scaling

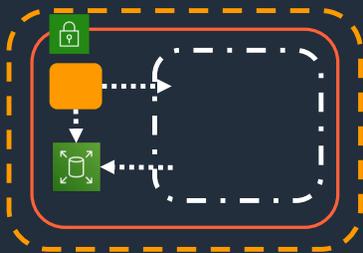
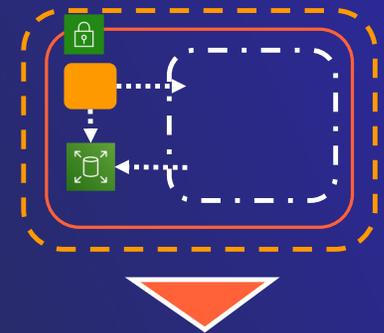
Allocate Instance(s)
& Run jobs



Scale up when
jobs are waiting



Scale down when
the cluster is idle

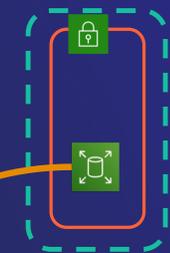


Cluster created

No Compute nodes allocated

Waiting for jobs...

(Optional) Delete the cluster
after data sync to Object
storage



CloudFormation for cluster deployments

New in ParallelCluster 3.6

Define your cluster as a **CloudFormation template**

No user-installed CLI or web UI needed

Should also work with AWS CDK

Embed HPC in complex IT systems

- Interoperability between compute environments
- *Ops-models for HPC workloads



```

AWSTemplateFormatVersion: '2010-09-09'
Description: AWS ParallelCluster CloudFormation Template

Parameters:
  AvailabilityZone:
    Description: Availability zone where instances will be launched
    Type: AWS::EC2::AvailabilityZone::Name
    Default: us-east-2b
  KeyName:
    Description: KeyPair to login to the head node
    Type: AWS::EC2::KeyPair::KeyName
  ComputeInstanceMax:
    Description: Maximum number of compute instances
    Type: Number
    Default: 10

Resources:
  PclusterVpc:
    Type: AWS::CloudFormation::Stack
    DeletionPolicy : Delete
    UpdateReplacePolicy: Delete
    Properties:
      Parameters:
        PublicCIDR: 10.0.0.0/24
        PrivateCIDR: 10.0.16.0/20
        AvailabilityZone: !Ref AvailabilityZone
      TemplateURL: !Sub
        - https://{{Region}}-aws-parallelcluster.s3.{{Region}}.amazonaws.com/
          parallelcluster/{{Version}}/templates/networking/public-private.cfn.
          json
        - { Version: 3.6.0, Region: !Ref AWS::Region }

  PclusterCluster:
    Type: Custom::PclusterCluster
    Properties:
      ServiceToken: !GetAtt [ PclusterClusterProvider , Outputs.
        ServiceToken ]
      DeletionPolicy: Retain
      ClusterName: !Sub 'c-${AWS::StackName}'
      ClusterConfiguration:
        Image:

```

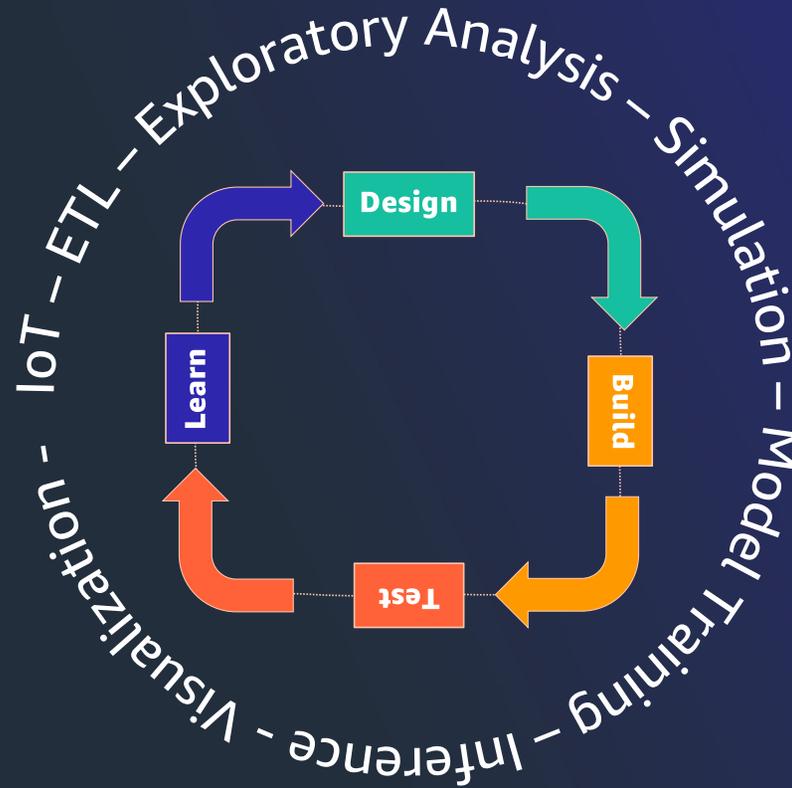
- Choose Availability Zone for instances
- Select SSH key name
- Specify max # instances
- Provision VPC + subnets
- Provision cluster in VPC & subnets

```

aws cloudformation create-stack \
  --stack-name lugdemo-3 \
  --template-body file://demo.yml

```

A real-world complex computing environment



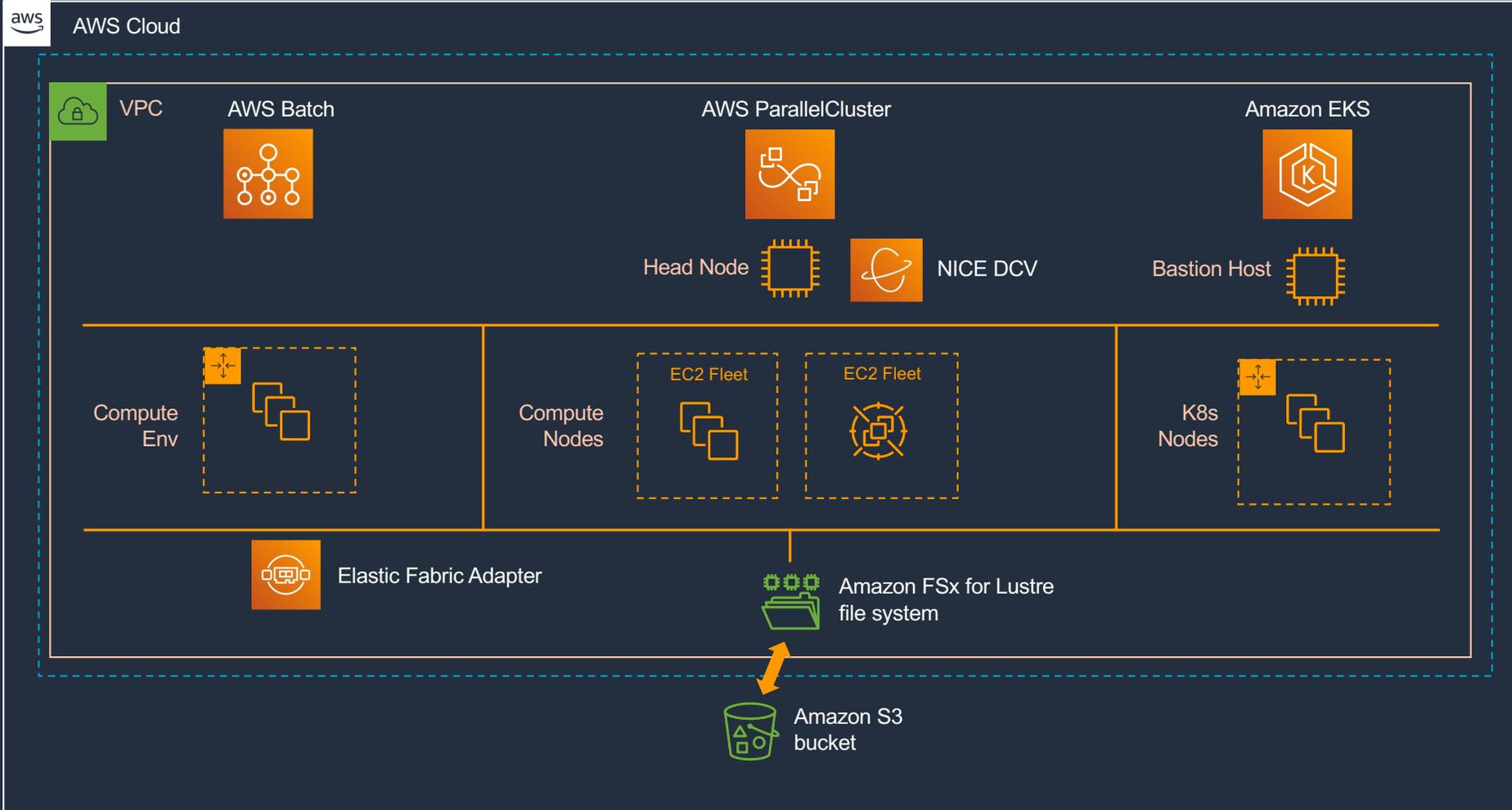
Option 1 - Fit everything in one computing paradigm

- HPC?
- Batch?
- K8s?

Option 2 - Integrate computing paradigms

- HPC
- Batch
- K8s

Unifying multiple compute systems with FSx for Lustre



HPC-Ops – an emerging urgent computing paradigm made easier with IaC

MAXAR

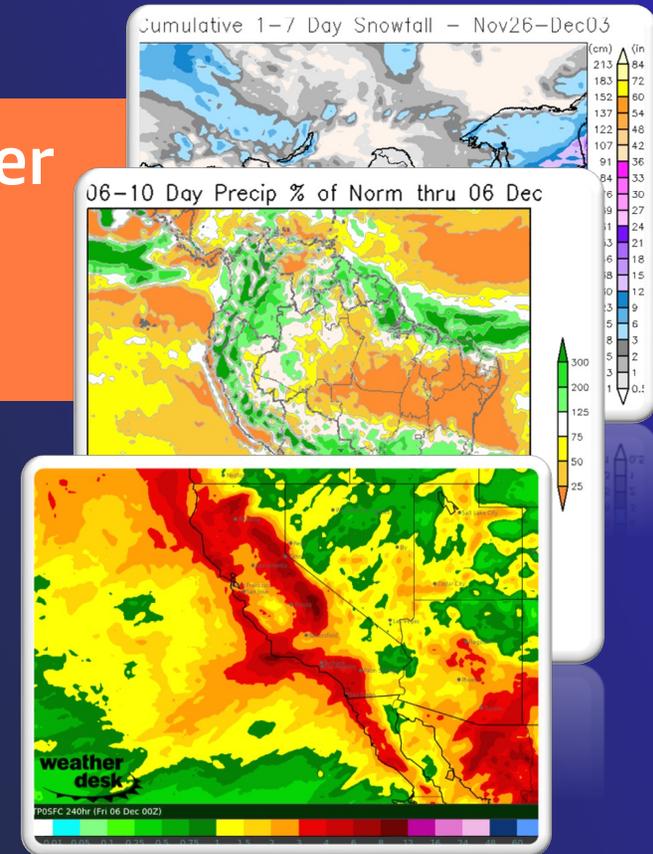
58% Faster than NOAA Supercomputer
45% Lower Compute Cost
~ 5600 Cores

Each forecasting job is run on an ephemeral HPC cluster

Charges incurred for ~45min



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



Conclusions

- Infrastructure as code is a powerful approach for modeling and managing complex resource deployments
- AWS ParallelCluster uses IaC to deploy and manage dynamic, autoscaling HPC
- AWS ParallelCluster 3.6 supports cluster deployment directly with CloudFormation
- This allows more sophisticated IT integrations with HPC
- It also makes it easier to implement DevOps, MLOps, DataOps, etc.



Questions

Supplementary Resources

- HPC Workshops
 - <https://www.hpcworkshops.com/>
 - <https://workshops.aws/categories/HPC>
- Media:
 - AWS HPC Blog : <https://aws.amazon.com/blogs/hpc/>
 - HPC Tech Shorts YouTube: <https://www.youtube.com/c/hpctechshorts>
 - Community Site: <https://day1hpc.com/>

