Managing cloud HPC with infrastructure-as-code

Matt Vaughn
Principal Developer Advocate
HPC Engineering
Amazon Web Services
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?
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"}
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

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

This Photo by Unknown Author is licensed under CC BY-ND
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?

**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

<table>
<thead>
<tr>
<th>Stack name</th>
<th>Status</th>
<th>Created time</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>lugdemo-02</td>
<td>UPDATE_COMPLETE</td>
<td>2023-04-24 10:19:32 UTC-0700</td>
<td>-</td>
</tr>
<tr>
<td>lugdemo-01</td>
<td>CREATE_COMPLETE</td>
<td>2023-04-24 06:06:11 UTC-0700</td>
<td>-</td>
</tr>
<tr>
<td>fsxlustredemo</td>
<td>CREATE_COMPLETE</td>
<td>2023-04-24 05:53:41 UTC-0700</td>
<td>Creates an FSxL filesystem of PERSISTENT_2 type plus the Security Group needed for use with ParallelCluster</td>
</tr>
<tr>
<td>CDKToolkit</td>
<td>CREATE_COMPLETE</td>
<td>2023-04-21 11:20:45 UTC-0700</td>
<td>This stack includes resources needed to deploy AWS CDK apps into this environment</td>
</tr>
<tr>
<td>techshorts042023</td>
<td>CREATE_COMPLETE</td>
<td>2023-04-18 16:35:23 UTC-0700</td>
<td>-</td>
</tr>
<tr>
<td>pcui-042023-pc351-ParallelClusterAPI-FBFJPZPZB8Z8</td>
<td>CREATE_COMPLETE</td>
<td>2023-04-18 10:00:44 UTC-0700</td>
<td>Template for the ParallelCluster API</td>
</tr>
<tr>
<td>pcui-042023-pc351</td>
<td>CREATE_COMPLETE</td>
<td>2023-04-18 10:00:38 UTC-0700</td>
<td>-</td>
</tr>
<tr>
<td>parallelcluster-ui-cognito</td>
<td>CREATE_COMPLETE</td>
<td>2023-02-20 08:52:32 UTC-0800</td>
<td>ParallelCluster UI Cognito User Pool</td>
</tr>
</tbody>
</table>
Cluster definition in a single file

Cluster Configuration File

OS
- Image: os: centos7
- HeadNode:
  - InstanceType: c5.4xlarge
  - Networking:
  - Ssh:
    - KeyName: My_PC3_KeyPair
    - AllowedIps: 0.0.0.0/0
- Scheduling:
  - Scheduler: slurm
  - SlurmSettings:
    - ScaleDownIdleTime: 10
    - DisableManagedDns: true
  - SlurmQueues:
    - Name: q1_ondemand
  - ComputeSettings:
    - ComputeResources:
      - Name: compute-resource-1
      - InstanceType: c5.n18xlarge
      - Efa:
        - Enabled: true
      - MinCount: 0
      - MaxCount: 64
      - Networking:
        - SubnetIds:
          - subnet-a12321bc
      - PlacementGroup: Enabled:true
- SharedStorage:
  - Name: myebs
  - StorageType: Ebs
  - EbsSettings:
    - VolumeType: gp3
    - Size: 100
  - Name: myfsx
  - StorageType: FsxLustre
  - FsxLustreSettings:
    - StorageCapacity: 1200
    - DeploymentType: SCRATCH_2
    - ImportPath: s3://myhpcbucket

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

(Physical) cluster design

Cluster deployed on AWS

AWS Cloud region
- VPC
  - Public subnet
    - Head instance
    - EBS Volume (NFS export)
    - NAT gateway
  - Private subnet
    - Compute instances
    - Visualization instance
- EC2 Fleet
- Amazon FSx for Lustre
- Amazon S3 Object storage

Cluster configuration file content:

Image:
- os: centos7

HeadNode:
- InstanceType: c5.4xlarge
- Networking:
- Ssh:
  - KeyName: My_PC3_KeyPair
  - AllowedIps: 0.0.0.0/0

Scheduling:
- Scheduler: slurm
- SlurmSettings:
  - ScaleDownIdleTime: 10
  - DisableManagedDns: true
- SlurmQueues:
  - Name: q1_ondemand
- ComputeSettings:
  - ComputeResources:
    - Name: compute-resource-1
    - InstanceType: c5.n18xlarge
    - Efa:
      - Enabled: true
    - MinCount: 0
    - MaxCount: 64
    - Networking:
      - SubnetIds:
        - subnet-a12321bc
    - PlacementGroup: Enabled:true
- SharedStorage:
  - Name: myebs
  - StorageType: Ebs
  - EbsSettings:
    - VolumeType: gp3
    - Size: 100
  - Name: myfsx
  - StorageType: FsxLustre
  - FsxLustreSettings:
    - StorageCapacity: 1200
    - DeploymentType: SCRATCH_2
    - ImportPath: s3://myhpcbucket
Dynamic compute resource scaling

Allocate Instance(s) & Run jobs

Cluster created
No Compute nodes allocated
Waiting for jobs...

Scale up when jobs are waiting

Scale down when the cluster is idle

(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
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

- Learn
- Design
- Build
- Test
- IoT - ETL - Exploratory Analysis - Simulation - Model Training - Inference - Visualization

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

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
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/