Deploying Qumulo Clusters Using the AWS-SA-WAF-CLUSTER CloudFormation Template

Dack Busch

Aug 31, 2021

Table of Contents

Intended Audience

This document is intended for anyone using the AWS Qumulo Solution Architect (SA) Well Architected Framework (WAF) Cluster CloudFormation templates. These templates and associated files are referred to simply as the aws-sa-waf-cluster. This document is comprehensive and experienced AWS users may find it repetitive or obvious. Less experienced users will benefit from the detailed documentation. Leverage the Table of Contents to jump to the area of interest.

Overview

This guide provides instructions for deploying the aws-sa-waf-cluster set of templates in AWS to fully provision the infrastructure for a Qumulo cluster. Any subsequent configuration for the file data platform is outside the scope of this document. Refer to the Qumulo Knowledge Base for further information.

Software Licences

Prior to executing any of the aws-sa-waf-cluster templates a Qumulo AMI subscription must be accepted in the AWS Marketplace for a chosen AWS account. The templates documented here are much more comprehensive than those available by default in the AWS Marketplace, so don't deploy from the AWS Marketplace. Simply accept the offer for the AMI of interest and close the AWS Marketplace session.

Architecture

Deploying the aws-sa-waf-cluster templates will instantiate the following infrastructure in AWS.

Qumulo WAF Deployment in an Existing VPC - Limited by Cloud Formation Template to 1TB-6PB of Capacity

As shown in Figure 1, the CloudFormation template expects the following:
- A VPC configured with a private subnet, public subnet, NAT gateway, and Internet Gateway
  - This basic infrastructure and the associated route tables are required to provide the automation infrastructure and cluster Internet connectivity.

As shown in Figure 1, the CloudFormation template sets up the following:
- A cluster of Qumulo EC2 instances
  - 4 to 20 nodes
  - 1TB to 6PB of usable capacity

- Multiple AWS Elastic Block Store (EBS) volumes connected to the Qumulo cluster
  - 28 unique EBS Volume configurations (ie Disk Configs) are available
  - All flash (SSD) and Hybrid (SSD+HDD) Disk Configs are available
- The Qumulo Sidecar which consists of two Lambda functions
  - One Lambda function sends Qumulo cluster metrics to Cloudwatch
  - The other Lambda function monitors the health of EBS volumes and automatically replaces any unhealthy volumes
- (Optional) AWS Route 53 private hosted zone
  - If R53 is resolving all DNS for the VPC then a R53 private hosted zone may be optionally provisioned to configure DNS A-records for the cluster.
- (Optional) Public IP Management
  - If desired a Network Load Balancer with an Elastic IP (public static) may also be configured
  - Note, this NLB listens on port 443 and has some inherent DoS capabilities.
  - As a best practice the cluster should be managed long-term via a secure host in the public or private subnet.
  - This functionality is only intended for initial configuration inspection and validation in the event no other EC2 machines exist in the VPC to manage the cluster
- Provisioner
  - An EC2 instance configures multiple parameters on the Qumulo cluster and within the AWS infrastructure
  - This instance automatically shuts down after successful provisioning
  - It is also restarted during stack updates to provision modifications to the infrastructure

Planning the Deployment

Specialized Knowledge

Deploying this template requires only a moderate level of familiarity with AWS services. If you're new to AWS, see https://aws.amazon.com/getting-started/.

AWS Account

Deployment of this template requires an AWS account. If you don't have an account visit https://aws.amazon.com.

Technical Requirements

Before you launch the template, review the following information and requirements and ensure that your account and privileges are properly configured. Otherwise the deployment may fail.

Resource Quotas

Review current quota utilization and ensure that resources are available for the resources required for this deployment.

Resource	This Deployment Uses
Elastic IP (optional for Public Management)	1
Security Groups	2
AWS IAM roles	4
m5.2xlarge, m5.4xlarge, m5.8xlarge, m5.16xlarge, m5.24xlarge, c5n.4xlarge, c5n.9xlarge, or c5n.18xlarge	4 to 20 of the chosen instance type and associated vCPUs
m5.large	1
Lambda functions	2
EBS (gp2, st1, or sc1) config dependent	2.34TiB to 6.25PiB

Supported Regions

Review current quota utilization and ensure that resources are available for the resources required for this deployment.

Americas

- us-east-1, US East (N. Virginia)
- us-east-2, US East (Ohio)
- us-west-1, US West (N. California)
- us-west-2, US West (Oregon)
- us-gov-west-1
- us-gov-east-1
- ca-central-1, Canada (Central)
- sa-east-1, South America (São Paulo)

Asia Pacific

- ap-northeast-1, Asia Pacific (Tokyo)
- ap-northeast-2, Asia Pacific (Seoul)
- ap-south-1, Asia Pacific (Mumbai)
- ap-southeast-1, Asia Pacific (Singapore)
- ap-southeast-2, Asia Pacific (Sydney)
- ap-east-1, Asia Pacific (Hong Kong)

Europe/Middle East/Africa

- eu-central-1, Europe (Frankfurt)
- eu-west-1, Europe (Ireland)
- eu-west-2, Europe (London)
- eu-west-3, Europe (Paris)
- eu-north-1, Europe (Stockholm)
- eu-south-1, Europe (Milan)
- me-south-1, Middle East (Bahrain)

IAM Permissions

Before launching the template, you must sign in to the AWS Management Console with IAM permissions for the resources that the template deploys and the services it leverages. The *AdministratorAccess* managed policy within IAM provides sufficient permissions, although your organization may choose to use a custom policy with more restrictions. The following AWS services are required in a custom IAM role or IAM user to deploy this template:

application-autoscaling:*	**elasticloadbalancing:***	**route53:***
applicationinsights:*	**events:***	**s3:***
autoscaling:*	**health:***	**secretsmanager:***
cloudformation:*	**iam:***	**sns:***
cloudtrail:*	**kms:***	**ssm:***
cloudwatch:*	**lambda:***	**sqs:***
compute-optimizer:*	**logs:***	**tag:***
ec2:*	**resource-groups:***	

Deployment Options

This template deploys into an existing VPC. There are optional parameters which are denoted in the template with OPTIONAL:
1. AWS Local Zone or Outpost deployment
2. Public IP Management
3. R53 Private Hosted Zone configuration
4. Customer Managed Key configuration
5. SNS Notifications
6. CloudWatch audit log configuration

Deploying in a VPC without Internet access

The architecture referenced previously expects a NAT Gateway in the existing VPC. It is possible to use the aws-sa-waf-cluster templates to deploy without internet access.
To support deployments without public Internet connectivity the VPC must be configured with:
- VPC Interface Endpoints to support AWS services
- A security group applied to each VPC Interface Endpoint
- Qumulo Core software in the S3 bucket prefix aws-sa-waf-cluster/upgrade folder

VPC Interface Endpoints

It is recommended to leave the policy for each endpoint as Full Access. The following VPC Interface Endpoints must be configured when deploying without Internet Access:

cloudformation	**lambda**	**secretsmanager**
cloudtrail	**logs**	**sns**
ec2	**monitoring**	**ssm**
events	**profile**	**sts**
kms	**S3**	**sqs**

VPC Interface Endpoints Security Group rules

To allow traffic from the provisioning node and the Qumulo cluster the security group applied to the VPC interface endpoints must have the following Security Groups added as Inbound Rule Sources:
- [stack name]-PROVISIONINGSTACK-[aws 12 digit code]-**ProvisionerSG**-[aws 12 digit code]
- [stack name]-QSTACK-[aws 12 digit code]-**QumuloSecurityGroup**-[aws 12 digit code]

Automated Qumulo Core upgrades during deployment

If you enter a newer software version than the Qumulo Core AMI was released with, the provisioning instance will expect to find the code via the Internet, or in this case from the S3 bucket, since the provisioning instance will not be able to reach the public Internet. It will look for code in the template's S3 bucket at the following prefix: [your prefix]/upgrade. Typically this is aws-sa-waf-cluster/upgrade. All quarterly released images between the AMI ID release version to, and including, the requested version must be present in the bucket. For example if the AMI was released with 4.0.6 and the desired version is 4.2.3, then 4.1.0.1, 4.2.0, and 4.2.3 versions must be in the bucket. If the software images are not present in the S3 bucket the Provisioning instance will not shutdown, nor will it progress with provisioning activities. To see the quarterly release cadence just cat the file [your prefix/cfn-init/upgrade-order.txt. Place the image(s) in the bucket and restart the provisioning instance if this occurs. The image(s) must be in the Qumulo Trends format: **qumulo_upgrade_cloud_x.y.z_num.qimg**. If you don't have access to Qumulo Trends, contact Qumulo for support.

Deployment Steps

Download the aws-sa-waf-cluster CloudFormation Folder

1. Go to https://github.com/Qumulo/aws-sa-waf-cluster and select the green **Code** button and then click on **Download Zip**.
2. Unzip the file on your local machine
3. Copy the top level folder and all contents, unmodified, to an AWS S3 bucket that your account has access to. Numerous tools exist to place this content in your S3 bucket like the AWS S3 Console, AWS CLI commands, or custom S3 browser utilities. All that is critical is that all files are copied and the directory hierarchy is maintained.
4. If deploying without Internet connectivity ensure the VPC Interface Endpoints, their Security Groups, and the desired Qumulo Core software version are setup per the previous *Deploying Without Internet Access* section.

Sign in to your AWS account

1. Sign in to your AWS account at https://aws.amazon.com with an IAM user role that has the necessary permissions. For details, see Planning the Deployment earlier in this document.
2. Make sure that your AWS account is configured correctly, as discussed in the Technical Requirements section earlier in this document.

Subscribe to a Qumulo Marketplace Offer

A Qumulo subscription for the chosen configuration and deployment region is required.
1. Go to https://aws.amazon.com/marketplace and type **Qumulo** in the search bar. Select one of the available offerings with a designated usable capacity. NOTE: If you have a private offer you will receive an email with a link to accept the private offer. The **Customizable File Storage Node** configuration in the Marketplace requires a private offer unless you choose to accept it with 320TiB per EC2 instance.
2. Click **Continue to Subscribe** in the upper right corner. The subscription will take a couple of minutes to process.
3. At this point the Marketplace process is complete when using the aws-sa-waf-cluster CloudFormation template. Close the Marketplace tab. Do not deploy from Marketplace.

Launch the CloudFormation Template

1. Go to the **CloudFormation** view in the AWS Console
2. Select the AWS Region to deploy the Qumulo Cluster in from the upper-right corner of the AWS Console.
3. Duplicate the browser tab and switch to the **S3** view in the AWS Console
4. Navigate to the bucket you placed the aws-sa-waf-cluster folder in
5. Select the **aws-sa-waf-cluster** folder
6. Click on the template **qcluster-existingVPC.cft.yaml**
7. Copy the **Object URL** (in blue text) for the template (NOT the S3 URL)
8. Switch back to the CloudFormation browser tab
9. Click **Create Stack** and select **With new resources**
10. Keep the defaults and paste the template URL into the **Amazon S3 URL** field
11. Click **Next**
12. The template is now launched and you will see the list of parameters on the **Specify Stack Details** page

Entering Parameters in the Template

The template parameters are largely self documenting with intuitive names and detailed descriptions. However, this template does a lot of work, so some additional explanation is warranted. All sections of the template will be captured below and additional information provided for the parameters.

Stack Name Parameter

Specify stack details

Stack name

Stack name

QCluster1

Stack name can include letters (A-Z and a-z), numbers (0-9), and dashes (-).

While stack names can be fairly complex, this template extensively tags resources with the stack name, nested stack name, and unique stack numbers generated by AWS. Thus, less is more for the stack name. In this example **QCluster1** will suffice.

AWS Template Configuration Parameters

This section consists of basic AWS parameters. Since this is the beginning of the parameters for the template, the aws-sa-waf-cluster template version number is also referenced here.

S3 Bucket Name

This is simply the name of the S3 Bucket the aws-sa-waf-cluster folder was placed in. Note, the account you use to provision the cluster must have access to this bucket. Also, you may be wondering why does the template need to know the bucket and prefix for this template when it was just launched with the URL? This template uses nested stacks. The top-level stack uses the bucket and prefix supplied to generate URLs for the nested stacks.

S3 Key Prefix Name

For those new to AWS, this parameter is simply the path in the bucket inclusive of the aws-sa-waf-cluster folder. It must NOT start with a "/" and it MUST end with a "/". This is the AWS S3 Key Prefix Name convention.

S3 Bucket Region

This is the region the S3 bucket is hosted in that contains the aws-sa-waf-cluster folder. In many cases this will NOT be the region the template is being executed in. S3 buckets are hosted in a region and available globally throughout AWS. To identify the region for your S3 bucket simply go to the AWS Console **S3** page, select **Buckets**, and the region will be listed just to the right of the bucket name.

AWS Key-Pair Name

This field is a dynamic drop-down populated with all the key-pairs in the account for the region. Here a simple name of OR was chosen for the Oregon (us-west-2) region. Any parameter field with an arrow on the far right is dynamically populated with options to select. Further, these fields must all be populated with a selection or the template will fail upon execution.

OPTIONAL: Environment Type

This field enables some options for development environments that should not be used in production environments.

AWS Network Configuration Parameters

The next section consists of multiple parameters to configure the networking for the cluster and supporting elements.

AWS Network Configuration

AWS VPC ID
AWS VPC ID.

[▼]

Qumulo Security Group
An IPv4 CIDR block for specifying the generated security group's allowed addresses for inbound traffic. Set to the VPC CIDR.

[172.31.0.0/16   ]

AWS Private Subnet ID
AWS Private Subnet in the VPC.

[▼]

Is the Qumulo Cluster being deployed in a Local Zone or Outpost?
AWS Lambda services are not supported in Local Zones or Outposts and will be placed in the subnet specified below.

[NO   ▼]

Qumulo Sidecar Lambdas Private Subnet ID
An AWS subnet ID MUST be selected. Select a subnet in one of the region's AZs (not in a Local Zone, not an Outpost Subnet). It is only used if YES for Local Zone or Outpost was selected above.

[▼]

OPTIONAL: Provision Public IP for Qumulo Management
Select YES to provision an Elastic IP Address (public static) attached to a Network Loadbalancer listening only on port 443 for Qumulo Managment. Not supported in Local Zones or Outpost.

[NO   ▼]

OPTIONAL: Enable Replication Port for Qumulo Public IP
Select YES to enable port 3712 for replication from on-prem Qumulo systems using the Elastic IP (public static) for Qumulo Managment. Requires YES to Public Management above.

[NO   ▼]

AWS Public Subnet ID
An AWS subnet ID MUST be selected. It is only used if YES for Public Management was selected above.

[▼]

OPTIONAL: FQDN for R53 Private Hosted Zone
IF NONE.local, R53 config will be skipped. Private R53 DNS Fully Qualified Domain Name. The .local domain is one way to provide public query resolution for overlapping names: qumulo.com vs qumulo.local

[NONE.local   ]

OPTIONAL: R53 Record Name for Qumulo RR DNS
ONLY APPLICABLE if a domain name was provided above. Record Name for R53 Private Hosted Zone Qumulo Cluster floating IPs.

[qumulo   ]

AWS VPC ID

This field is a dynamic drop-down populated with all the VPCs in the account for the region. Choose the VPC to deploy in.

Qumulo Security Group

The Qumulo Security Group CIDR generally should match the VPC CIDR. This allows the cluster to communicate with any subnet within the VPC.

AWS Private Subnet ID

The cluster is deployed in a private subnet within the VPC. This field is a dynamic drop-down populated with all the subnets available within the Region. Select the subnet, and hence the Availability Zone (AZ), the cluster resides in. Make sure the subnet chosen is relevant for the VPC chosen. This can sometimes be confusing if the tags don't clarify the VPC the subnet resides in.

Is The Qumulo Cluster Being Deployed in a Local Zone or Outpost?

Since AWS Lambda functions are not yet supported in Local Zones or Outposts, change this field to **YES** if your cluster is being deployed in a Local Zone or on an Outpost. An example is the LA Local Zone that is part of the Oregon region with four AZs.

Qumulo Sidecar Lambda Private Subnet ID

In the event the cluster is being deployed in a Local Zone or on an Outpost, choose a subnet to deploy the AWS Lambdas in that is in an AZ for the AWS region. Just like any other CloudFormation dynamic drop down, an option must be chosen. In this example the same private subnet ID was selected to make the parser happy.

OPTIONAL: Provision Public IP for Qumulo Management

If this option is selected, an Elastic IP (public static IP) is assigned to an AWS Network Load Balancer that is connected to the cluster and listens only on port 443. This is not recommended for production clusters with sensitive data. To minimize the attack surface area, manage the cluster from an EC2 instance within your VPC. Note, AWS Local Zones and Outposts do not support NLB target groups, so this option is not supported if deploying the cluster in a Local Zone or on an Outpost.

OPTIONAL: Enable Replication Port for Qumulo Public IP

The replication port can be enabled on the public interface for customers that want to replicate from one Qumulo cluster to another via the public Internet. This is fine for test purposes but should not be used on production clusters. Either a VPN or AWS Direct Connect should be used for production replication.

AWS Public Subnet ID

This parameter must be populated, even if the public management option is left at the default of **NO**. This is a basic template requirement for the CloudFormation parser even though the field is only used for provisioning in the event public management was selected.

OPTIONAL: FQDN for R53 Private Hosted Zone

The template will optionally create a Route 53 Private Hosted Zone with equally weighted records, TTL=0, for the floating IPs on the cluster. If left at the default of **NONE.local** this option will be ignored. Here an FQDN of **test.local** was chosen since this VPC has no Active Directory server with DNS. [DNS options in AWS to enable IP Failover and Client Distribution](#)

R53 Record Name for Qumulo RR DNS

This field is the DNS A-Record name for the Route 53 records that are created. This is only relevant if an FQDN was entered above other than **NONE.local**.

Qumulo File Data Platform Configuration Parameters

This section configures all the parameters that are specific to the Qumulo Cluster.

Qumulo File Data Platform Configuration

Is the Qumulo Cluster being deployed for Disaster Recovery?
Select YES if the cluster is being deployed for a Disaster Recovery workload.

NO	▼

Qumulo AWS Marketplace Offering Accepted
Select the Qumulo Cluster usable capacity per the accpeted AWS Marketplace offering. Customizable offerings typically leverage a Private Offer via AWS Marketplace and can be used to deploy 1TB to 6PB with this template.

1TB-Usable-All-Flash	▼

Qumulo AWS AMI ID
Only used for the Specified AMI-ID option above. Amazon Machine Image ID supplied by Qumulo.

Only for Specified AMI-ID Offering

Qumulo EC2 Instance Type
EC2 instance type for Qumulo nodes.

m5.2xlarge	▼

Total Number of Qumulo EC2 Instances
Total number of EC2 instances, or Qumulo Nodes, in the Qumulo Cluster: (4-10). NOTE: This field may be used to add nodes with a CloudFormation Stack Update after initial provisioning.

Select for Custom Offering OR Expanding Cluster	▼

EBS Volume Configuration per EC2 Instance
Choose the EBS Volume configuration and type for the Qumulo EC2 instances: AF= SSD, Hybrid st1= SSD+HDD st1, Hybrid sc1= SSD+HDD sc1. NOTE: This must match the EBS capacity and type of the Customizable Private Offer.

Select for Custom Offering	▼

Floating IP for IP Failover
Number of EC2 Secondary IPs to be configured for each instance in the cluster, 1-4.

3	▼

Qumulo Core Software Version
Software version to install on the cluster. NOTE: This field CAN NOT be used to upgrade the cluster with a CloudFormation Stack Update. All Updates after initial creation must follow the quarterly release cadence using the Web UI or REST API.

4.2.0

Qumulo Cluster Name
Name must be an alpha-numeric string between 2 and 15 characters. Dash (-) is allowed if not the first or last character.

Cloud-Q

Qumulo Cluster Admin Password
Minumum 8 characters and must include one each of: uppercase, lowercase, and a special character.

OPTIONAL: AWS EBS Volumes Encryption Key
Leave Blank and AWS will generate a key. To specify a Customer Managed Key provide the KMS CMK ID: 12345678-1234-1234-1234-1234567890ab

OPTIONAL: AWS Permissions Boundary Policy Name
Apply an IAM Permissions Boundary Policy to the Qumulo IAM roles that are created for the Qumulo cluster and provisioning instance. This is an account based policy and is optional. Qumulo's IAM roles conform to the least privilege model.

OPTIONAL: Qumulo Instance Recovery Topic
Optionally enter the ARN of an SNS topic that receives messages when an instance alarm is triggered.

OPTIONAL: Send Qumulo Audit Log messages to CloudWatch Logs?
Select YES to create a CloudWatch Logs Group for the Qumulo Cluster that captures all Qumulo Audit Log Activity.

NO	▼

Enable Termination Protection
Enable Termination Protection for EC2 instances and the CloudFormation stack

YES	▼

Is the Qumulo Cluster being deployed for Disaster Recovery?

If the cluster is for Disaster Recovery from another AZ, Region, or an on-prem cluster then select YES. Template rules will then guide the appropriate choice of EC2 instance types and EBS volume configurations.

Qumulo AWS Marketplace Offering Accepted

This field is a dynamic drop-down populated with the Qumulo Marketplace offers noted with usable capacity or the customizable option. Select the Marketplace offer that you accepted directly or via a private offer.. In some cases Qumulo may specify the AMI-ID to be used, which is why the Specified-AMI-ID option is available.

Qumulo AWS AMI ID

This parameter is only used if Specified-AMI-ID was selected as an option. Qumulo will provide the appropriate AMI ID.

Qumulo EC2 Instance Type

This field is a dynamic drop-down populated with the supported EC2 instance types. For low performance clusters an m5.2xlarge instance type may be selected. For very high performance clusters a c5n.18xlarge instance type may be used. If you don't know what instance type will support your workload contact Qumulo for recommendations.

Total Number of Qumulo EC2 Instances

This field is only used for Custom-1TB-6PB or the Specified-AMI-ID during initial provisioning. This is another dynamic drop-down providing options from 4 to 20 instances. Capacity and performance scale with instance count. Most cloud based workloads can be facilitated with 20 or less nodes. If you need more than 20 nodes, contact Qumulo. Note, this field can be used to increase the node count in a subsequent Stack Update.

EBS Volume Configuration per EC2 Instance

This field is only used for Custom-1TB-6PB or the Specified-AMI-ID. This is the disk configuration per EC2 Instance. All supported Qumulo disk configurations are available in this dynamic drop-down. Qumulo will provide the disk configuration that is appropriate for Customizable offerings or Specified-AMI-ID offerings.

Floating IP for IP Failover

EC2 Instances running Qumulo Core support EC2 Secondary IP configuration. Within the Qumulo cluster these are referred to as Floating IPs. The cluster takes over ownership and management of these IPs to protect client connections in the event of an EC2 instance failure. By default, three floating IPs per instance are recommended. Should any instance fail its load will be distributed to three adjacent instances in the cluster. Options for 1-4 floating IPs per instance are available. These IP addresses are configured in a Route 53 Private Hosted Zone if that option was selected in the previous section of the template. Otherwise, A-records must be set up in the chosen DNS server.

Qumulo Core Software Version

Qumulo releases AWS AMI IDs quarterly. As of the date of this document Qumulo Core 4.0.6 is the latest version. As Qumulo software continues to evolve you may enter any newer software version when provisioning new clusters. The template will update all nodes to the desired software version before forming the first quorum for the cluster. Note, this field can't be used in a Stack Update to change the software version. Nor can it be used to install a version of software that is older than the version the AMI ID shipped with. For post deployment software upgrades refer to Qumulo KB [Qumulo Core Upgrades via UI](#).

Qumulo Cluster Name

This parameter is simply the name given to the cluster. If you have multiple clusters in an AWS region, give them unique names.

Qumulo Cluster Admin Password

This parameter is the Admin Password assigned to the cluster. This password is also stored in AWS Secrets Manager for subsequent reference.

OPTIONAL: AWS EBS Volumes Encryption Key

All data at rest is encrypted. IF this field is left blank an AWS generated key is used. You may optionally specify a Customer Managed Key that has been setup in the AWS Key Management System as shown in the above example. Note, the Key Management System CMK policy for the key must not contain any leftover SIDs from previous provisioning. Assuming there are no other legitimate SIDs in the policy, a clean CMK policy should follow the following JSON structure below.

OPTIONAL: AWS Permissions Boundary Policy Name

Specify the name of the IAM Permissions Boundary Policy in this field to restrict the IAM roles created for the Qumulo cluster EC2 instances and the Provisioning instance to the desired boundary. This is not necessary and is optional. Care must be taken that the boundary policy is not overly restrictive or features and fuctions of Cloud Q on AWS may be impacted. The IAM roles created for these resources conform to a least privilege model.

OPTIONAL: Qumulo Instance Recovery Topic

In the event of an EC2 Instance failure, Amazon Simple Notification Service (SNS) text or email messages will be sent via the configured SNS topic. Paste in your SNS topic ARN here.

OPTIONAL: Send Qumulo Audit Log Messages to CloudWatch Logs?

This parameter defaults to **NO**. To enable Qumulo Audit logs to be stored in a CloudWatch Log Group select **YES**.

Enable Termination Protection

It is highly recommended this parameter be left at the default of YES. Both EC2 termination protection and CloudFormation stack termination protection will be enabled. This guards against any accidental deletions of the cluster or an instance that may result in data loss.

Qumulo EBS Monitoring/Replacement & CloudWatch Metrics Configuration

This section configures the Qumulo Sidecar.

Qumulo EBS Monitoring/Replacement & CloudWatch Metrics Configuration

Provision Qumulo Sidecar Lambdas
By default the Qumulo Sidecar Lambdas are deployed to monitor and replace failed EBS volumes plus send metrics to CloudWatch. Leave this at default for production environments with WAF compliance. The ability to disable this provisioning is offerred just for test environments.

> YES ▼

Qumulo Sidecar Username

> SideCarUser

Qumulo Sidecar Password
Minumum 8 characters and must include one each of: uppercase, lowercase, and a special character. This field must be filled in regardless of the Sidecar provisioning option above.

> ••••••••

Qumulo Sidecar Software Version
Software Version should match the desired cluster version at creation. NOTE: This field may be used to upgrade the SideCar software version with a CloudFormation Stack Update after upgrading the cluster via the Web UI or REST API.

> 4.1.5

OPTIONAL: Qumulo Sidecar SNS Topic
Optionally enter an SNS topic ARN that lambda errors and successful disk replacements will be published to.

>

Provision Qumulo Sidecar Lambdas

By default the Sidecar Lambdas are provisioned. For test environments the option to disable Sidecar provisioning is provided. Qumulo recommends using the Sidecar in all production networks that conform to AWS Well-Architected standards.

Qumulo Sidecar Username

A username is configured on the cluster for Sidecar Lambda access. You may change this username if desired.

Qumulo Sidecar Password

This field must be filled in, even if the Sidecar is not being provisioned. A password is created for the Sidecar username and these credentials are stored in Secrets Manager for subsequent reference.

Qumulo Sidecar Software Version

The Sidecar software version should match the cluster software version previously specified. Note, this field may be used with a Stack Update to install a newer version of the Sidecar Lambda software. This is especially helpful for production clusters that are upgraded post deployment via the Qumulo UI or API, so the Sidecar software version is synchronized with the cluster software version.

OPTIONAL: Qumulo Sidecar SNS Topic

In the event of an EBS Volume failure, Amazon Simple Notification Service (SNS) text or email messages will be sent via the configured SNS topic. Paste in your SNS topic ARN here.

Deploy the CloudFormation Stack

1. After completing all parameter entries choose **Next**.
2. On the **Configure stack options** page you can specify additional tags. When finished choose **Next**.
3. On the **Review** page confirm the template settings. Under **Capabilities**, at the bottom of the page, select the two check boxes to acknowledge that the template may create IAM resources and may require the ability to automatically expand macros. *The template will fail if these check boxes are left unchecked.*
4. Choose **Create stack** to deploy the stack.

CloudFormation Nested Stack Deployment

The stack will take 10 minutes, or less, to deploy depending on the options chosen in the template. You can monitor the progress of the stack by choosing **Events** on the top-level stack named **QCluster1** in this example. Below is a description of each nested stack's function in the order they are deployed.

QLOOKUP Nested Stack

The first three stacks are launched in parallel. The QLOOKUP Stack is a hierarchical region map that finds the AMI-ID for the marketplace offer by region. Note, this stack is not executed when selecting the Specified-AMI-ID option.

SECRETSSTACK Nested Stack

The first two stacks are launched in parallel. The SECRETS Stack stores usernames and passwords in Secrets Manager for the cluster, sidecar, and downloading software from Qumulo Trends. The purpose of leveraging Secrets Manager is to provide a secure reference for subsequent stacks and users who may forget the passwords assigned.

QIAMSTACK Nested Stack

The QIAM Stack creates an IAM profile for the Qumulo Cluster to enable the cluster to manage EC2 Secondary IP addresses (Floating IPs), decrypt data, send CloudWatch alarms, and send audit logs to CloudWatch Logs. Note, creating IAM roles in AWS takes some time so don't be alarmed if the QIAM stack takes 3 or 4 minutes.

QSTACK Nested Stack

The Q Stack spins up all the EC2 instances and EBS volumes for the cluster. It also creates a placement group for the cluster and tags all the EC2 instances with the appropriate stack name and node number. In addition, it creates CloudWatch alarms for EC2 instance failure and a security group for the cluster with the CIDR specified in the template.

MGMTNLBSTACK Nested Stack

All stacks subsequent to QSTACK execute in parallel. If public management of the cluster was chosen in the template this nested stack is executed as long as the cluster is NOT being deployed in an AWS Local Zone. It spins up a Network Load Balancer with a public Elastic IP. The load balancer listens only on port 443 and optionally on port 3712 if the replication port was selected. This load balancer connects to the primary EC2 IP address on each node. These are known as the persistent IPs in the Qumulo UI.

DNSSTACK Nested Stack

If the Route 53 Private Hosted Zone FQDN was configured then the DNS stack is executed. It creates the private hosted zone and all the A-records with the name assigned in the template. All records are given a TTL=0. While round-robin behavior is the goal, Route 53 doesn't

provide perfect round-robin. Instead the records are given an equal probability of resolution. Clients are well distributed, but not perfectly symmetric.

PROVISIONINGSTACK Nested Stack

This stack spins up an EC2 instance with custom user data. It configures the Qumulo Cluster and some AWS environment requirements.

Qumulo Configuration
- Software upgrades of QSTACK created nodes
- Forms the first quorum for the cluster
- Assigns Floating IP addresses to the cluster
- Configures Sidecar username, password, and custom RBAC role
- Configures Audit Logging for CloudWatch Logs
- Changes the admin password

AWS Configuration
- Checks for Public Internet reachability with a CURL test to trends.qumulo.com
- Assigns a QSTACK Policy to protect the cluster in subsequent Stack Updates
- Edits the Customer Managed Key Policy so Sidecar can create CMK encrypted volumes
- Tags EBS volumes with the stack name and volume type
- Tracks software versions, cluster IPs, instance IDs, & UUID in AWS Parameter Store
- Tracks the provisioning instance 'last-run-status' in Parameter Store
- Configures Termination Protection for the Stack and the EC2 instances

This instance automatically shuts down upon completion of its provisioning tasks. DO NOT DELETE THIS EC2 INSTANCE. IT WILL BE USED FOR STACK UPDATES.

CLOUDWATCHSTACK Nested Stack

This stack creates resource groups, a CloudWatch dashboard, and a CloudWatch log group (optional) for the cluster. First, it creates a resource group for the EC2 instances and then it creates one or more resource groups for the EBS volumes. The resource groups created for the EBS volumes depend on the EBS volume configuration of the cluster. All Flash clusters will have just one resource group with the stack name and -SSD. Hybrid clusters will have two resource groups for EBS: one with -SSD and one with -HDD. The purpose of these resource groups is to provide a simple means to create a filtered view in CloudWatch for the EC2 and EBS metrics native to AWS.

A CloudWatch Dashboard is also created that presents key metrics sent by the Sidecar Metrics Lambda function. These are Qumulo specific metrics.

Finally, if Audit Logging was enabled a CloudWatch log group is created for the cluster. All administrative activity, Lambda access, and file/directory create/modify write activity is captured in this log.

QSIDECARSTACK Nested Stack

Assuming the provisioning option was left as YES, the SIDECAR stack is deployed. It creates two Lambda functions with the specified Sidecar software version. The first is the Metrics Lambda that sends Qumulo metrics to CloudWatch. The second is the Disk Recovery Lambda that monitors EBS volumes and automatically replaces any failed EBS volumes. IAM roles, permissions, and events are created for each Lambda function.

Below is the event view for the top-level **QCluster1** stack. Note the timestamps to manage expectations on the duration for each nested stack.

Timestamp	Logical ID	Status
2021-08-15 17:40:03 UTC-0700	QCluster1	⊘ CREATE_COMPLETE
2021-08-15 17:39:59 UTC-0700	QSIDECARSTACK	⊘ CREATE_COMPLETE
2021-08-15 17:37:48 UTC-0700	PROVISIONINGSTACK	⊘ CREATE_COMPLETE
2021-08-15 17:37:48 UTC-0700	DNSSTACK	⊘ CREATE_COMPLETE
2021-08-15 17:37:23 UTC-0700	MGMTNLBSTACK	⊘ CREATE_COMPLETE
2021-08-15 17:34:55 UTC-0700	CLOUDWATCHSTACK	⊘ CREATE_COMPLETE
2021-08-15 17:34:45 UTC-0700	QSIDECARSTACK	ⓘ CREATE_IN_PROGRESS
2021-08-15 17:34:45 UTC-0700	PROVISIONINGSTACK	ⓘ CREATE_IN_PROGRESS
2021-08-15 17:34:44 UTC-0700	DNSSTACK	ⓘ CREATE_IN_PROGRESS
2021-08-15 17:34:44 UTC-0700	CLOUDWATCHSTACK	ⓘ CREATE_IN_PROGRESS
2021-08-15 17:34:44 UTC-0700	MGMTNLBSTACK	ⓘ CREATE_IN_PROGRESS
2021-08-15 17:34:44 UTC-0700	PROVISIONINGSTACK	ⓘ CREATE_IN_PROGRESS
2021-08-15 17:34:44 UTC-0700	QSIDECARSTACK	ⓘ CREATE_IN_PROGRESS
2021-08-15 17:34:44 UTC-0700	CLOUDWATCHSTACK	ⓘ CREATE_IN_PROGRESS
2021-08-15 17:34:43 UTC-0700	MGMTNLBSTACK	ⓘ CREATE_IN_PROGRESS
2021-08-15 17:34:43 UTC-0700	DNSSTACK	ⓘ CREATE_IN_PROGRESS
2021-08-15 17:34:41 UTC-0700	QSTACK	⊘ CREATE_COMPLETE
2021-08-15 17:33:39 UTC-0700	QSTACK	ⓘ CREATE_IN_PROGRESS
2021-08-15 17:33:39 UTC-0700	QSTACK	ⓘ CREATE_IN_PROGRESS
2021-08-15 17:33:35 UTC-0700	QIAMSTACK	⊘ CREATE_COMPLETE
2021-08-15 17:31:08 UTC-0700	SECRETSSTACK	⊘ CREATE_COMPLETE
2021-08-15 17:31:08 UTC-0700	QLOOKUPSTACK	⊘ CREATE_COMPLETE
2021-08-15 17:30:58 UTC-0700	SECRETSSTACK	ⓘ CREATE_IN_PROGRESS
2021-08-15 17:30:57 UTC-0700	QLOOKUPSTACK	ⓘ CREATE_IN_PROGRESS
2021-08-15 17:30:57 UTC-0700	QIAMSTACK	ⓘ CREATE_IN_PROGRESS
2021-08-15 17:30:57 UTC-0700	SECRETSSTACK	ⓘ CREATE_IN_PROGRESS
2021-08-15 17:30:56 UTC-0700	QIAMSTACK	ⓘ CREATE_IN_PROGRESS
2021-08-15 17:30:56 UTC-0700	QLOOKUPSTACK	ⓘ CREATE_IN_PROGRESS
2021-08-15 17:30:50 UTC-0700	QCluster1	ⓘ CREATE_IN_PROGRESS

Post Deployment Steps

Once the top-level stack event log shows **CREATE_COMPLETE,** CloudFormation has completed instantiation of all stack resources. Below are the steps to validate the deployment.

Review & Verify the AWS Infrastructure

Verify the Cluster Instances are Running

In the **AWS EC2 Console** filter on the stack name, clear the running instance filter, and verify the number of instances for the cluster is as expected. Four in this example.

Verify the Provisioning Instance has Stopped
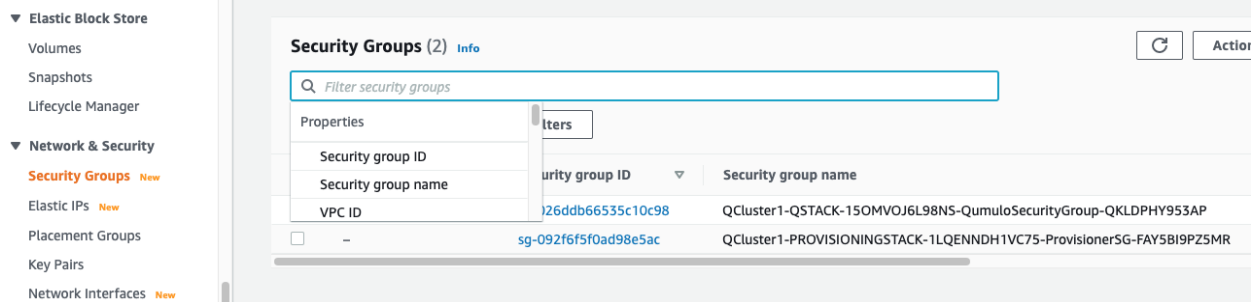
CloudFormation has completed the instantiation of all resources, but this does not mean all resources are fully initialized and running. Specifically, the Provisioning instance will still be initializing. Given all the tasks the Provisioning instance has to accomplish it will require at least 4 minutes AFTER stack completion to finish all tasks. For a large cluster (ie 20 nodes) it may require up to 10 minutes. This variability is due to software upgrades of the instances and tagging EBS volumes. When it is finished it will automatically shutdown. If the provisioning instance has not stopped after 10 or more minutes, jump to the troubleshooting section.

Verify the EC2 Security Groups
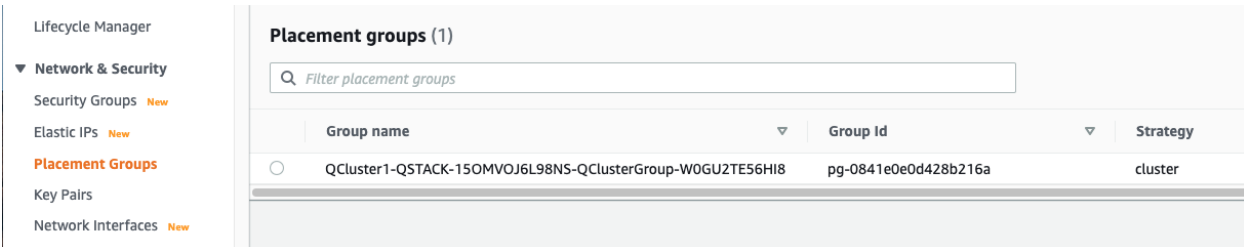
In the AWS Console go to the EC2 **Security Groups** page and filter on the top-level stack name. There will be two Security Groups that have been created. Select either to inspect the ports and CIDRs configured.

Verify the EC2 Placement Group for the Cluster

In the AWS Console go to **Placement Groups**. A placement group with the stack name has been created.
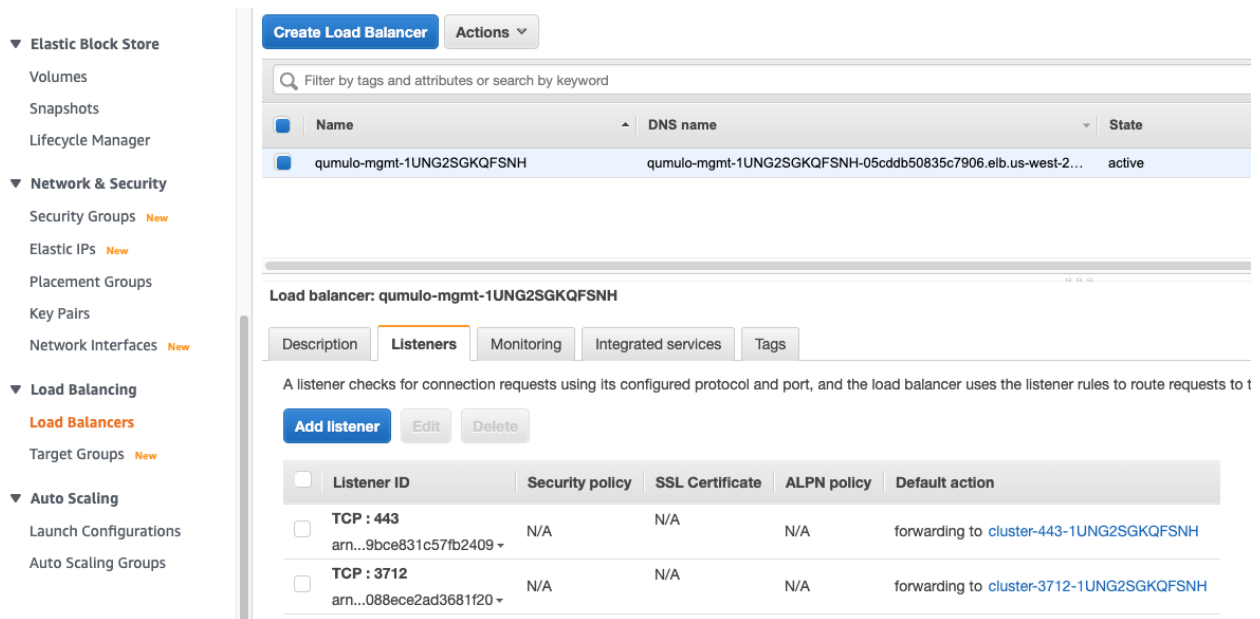
Verify the Load Balancer for Public Management (Optional)

In the AWS Console go to **Load Balancers**. If Public Management was selected in the template a load balancer has been created. It will be listening on 443, and if selected in the template, 3712 for replication.

Verify EBS Volume Tags

If the Provisioning instance has stopped the EBS volumes will be tagged accordingly for the cluster and EBS volume configuration. Go to the AWS Console **Elastic Block Store Volumes** page to verify. The type and number of EBS volumes will vary depending on EBS volume configuration chosen in the template and the number of EC2 instances.

	Name	Volume ID	Size	Volume Type	IOPS	Tl
	QCluster1-QSTACK-15OMVOJ6L98NS-boot	vol-04cf38e8…	60 GiB	gp2	180	-
	QCluster1-QSTACK-15OMVOJ6L98NS-boot	vol-09761f15…	60 GiB	gp2	180	-
	QCluster1-QSTACK-15OMVOJ6L98NS-boot	vol-0acc328…	60 GiB	gp2	180	-
	QCluster1-QSTACK-15OMVOJ6L98NS-boot	vol-0be5ffc7…	60 GiB	gp2	180	-
	QCluster1-QSTACK-15OMVOJ6L98NS-gp2	vol-00ae521…	100 GiB	gp2	300	-
	QCluster1-QSTACK-15OMVOJ6L98NS-gp2	vol-028b746…	100 GiB	gp2	300	-
	QCluster1-QSTACK-15OMVOJ6L98NS-gp2	vol-029c1d0…	100 GiB	gp2	300	-
	QCluster1-QSTACK-15OMVOJ6L98NS-gp2	vol-03e732a…	100 GiB	gp2	300	-
	QCluster1-QSTACK-15OMVOJ6L98NS-gp2	vol-045989f2…	100 GiB	gp2	300	-
	QCluster1-QSTACK-15OMVOJ6L98NS-gp2	vol-0528401…	100 GiB	gp2	300	-
	QCluster1-QSTACK-15OMVOJ6L98NS-gp2	vol-05e5767…	100 GiB	gp2	300	-
	QCluster1-QSTACK-15OMVOJ6L98NS-gp2	vol-06efcc73…	100 GiB	gp2	300	-
	QCluster1-QSTACK-15OMVOJ6L98NS-gp2	vol-081444c…	100 GiB	gp2	300	-
	QCluster1-QSTACK-15OMVOJ6L98NS-gp2	vol-08a574a…	100 GiB	gp2	300	-
	QCluster1-QSTACK-15OMVOJ6L98NS-gp2	vol-0945092…	100 GiB	gp2	300	-
	QCluster1-QSTACK-15OMVOJ6L98NS-gp2	vol-09f9a714…	100 GiB	gp2	300	-
	QCluster1-QSTACK-15OMVOJ6L98NS-gp2	vol-0a2564e…	100 GiB	gp2	300	-
	QCluster1-QSTACK-15OMVOJ6L98NS-gp2	vol-0b464cb…	100 GiB	gp2	300	-
	QCluster1-QSTACK-15OMVOJ6L98NS-gp2	vol-0ba9cc1…	100 GiB	gp2	300	-
	QCluster1-QSTACK-15OMVOJ6L98NS-gp2	vol-0cb1258…	100 GiB	gp2	300	-
	QCluster1-QSTACK-15OMVOJ6L98NS-gp2	vol-0d183f78…	100 GiB	gp2	300	-

Create Volume   Actions ∨

search : QCluster1 ⊗   Add filter

Verify EBS Encryption with a CMK (Optional)
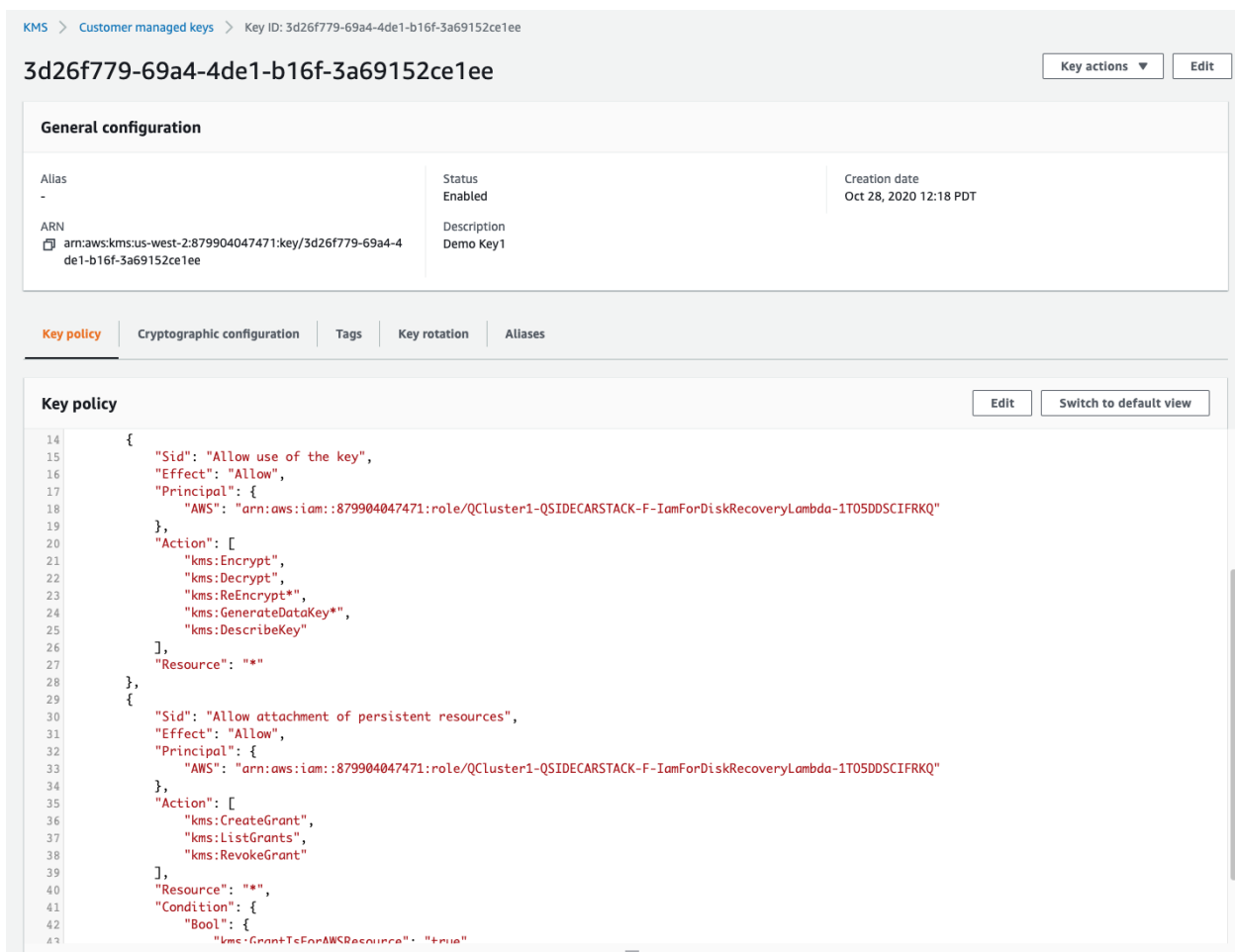
On the same page scroll to the right to verify that the volumes are encrypted with the Customer Managed Key assigned in the template. This is only relevant if a CMK was specified. If the field was left blank in the template, AWS will generate a key to encrypt the data at rest.

Volume Status	Encryption	KMS Key ID
✅ Okay	Encrypted	3d26f779-69a4-4de1-b16f-3a69152ce1ee
✅ Okay	Encrypted	3d26f779-69a4-4de1-b16f-3a69152ce1ee
✅ Okay	Encrypted	3d26f779-69a4-4de1-b16f-3a69152ce1ee
✅ Okay	Encrypted	3d26f779-69a4-4de1-b16f-3a69152ce1ee
✅ Okay	Encrypted	3d26f779-69a4-4de1-b16f-3a69152ce1ee
✅ Okay	Encrypted	3d26f779-69a4-4de1-b16f-3a69152ce1ee
✅ Okay	Encrypted	3d26f779-69a4-4de1-b16f-3a69152ce1ee
✅ Okay	Encrypted	3d26f779-69a4-4de1-b16f-3a69152ce1ee
✅ Okay	Encrypted	3d26f779-69a4-4de1-b16f-3a69152ce1ee
✅ Okay	Encrypted	3d26f779-69a4-4de1-b16f-3a69152ce1ee
✅ Okay	Encrypted	3d26f779-69a4-4de1-b16f-3a69152ce1ee
✅ Okay	Encrypted	3d26f779-69a4-4de1-b16f-3a69152ce1ee
✅ Okay	Encrypted	3d26f779-69a4-4de1-b16f-3a69152ce1ee
✅ Okay	Encrypted	3d26f779-69a4-4de1-b16f-3a69152ce1ee
✅ Okay	Encrypted	3d26f779-69a4-4de1-b16f-3a69152ce1ee
✅ Okay	Encrypted	3d26f779-69a4-4de1-b16f-3a69152ce1ee
✅ Okay	Encrypted	3d26f779-69a4-4de1-b16f-3a69152ce1ee
✅ Okay	Encrypted	3d26f779-69a4-4de1-b16f-3a69152ce1ee
✅ Okay	Encrypted	3d26f779-69a4-4de1-b16f-3a69152ce1ee
✅ Okay	Encrypted	3d26f779-69a4-4de1-b16f-3a69152ce1ee
✅ Okay	Encrypted	3d26f779-69a4-4de1-b16f-3a69152ce1ee

Verify the KMS CMK Policy (Optional)

In the AWS Console go to the **Key Management Service** page and select the CMK that was chosen in the template. Verify that the policy has been updated with two SIDs, one for the Metrics Lambda and one for the Disk Recovery Lambda. If the policy is not updated it is likely the Provisioning node will not have shutdown because the policy was not cleaned up prior to launching the template. Without this policy modification in place the Sidecar will not be able to create a new EBS volume to replace a failed EBS volume.

Verify Secrets Manager Secrets

In the AWS Console go to the **Secrets Manager** page and filter on the top-level stack name. There will be three secrets that have been created to store username/password pairs. Select any of them to see the credentials.

Verify the IAM Roles

In the AWS Console go to the IAM page and filter on the top-level stack name. There will be four IAM roles that have been created: two for the Sidecar, one for the cluster, and one for the provisioning instance.

Verify Sidecar Lambdas

In the AWS Console go to the **Lambda** page and filter on the top-level stack name. There will be two Lambda functions. Select the **Disk Recovery Lambda** and then choose **Monitor**. In the populated graphs check that the Error Count and Success Rate shows 100% green and 0% red. This confirms the Disk Recovery Lambda is communicating with the cluster. Review the Metrics Lambda in the same manner.
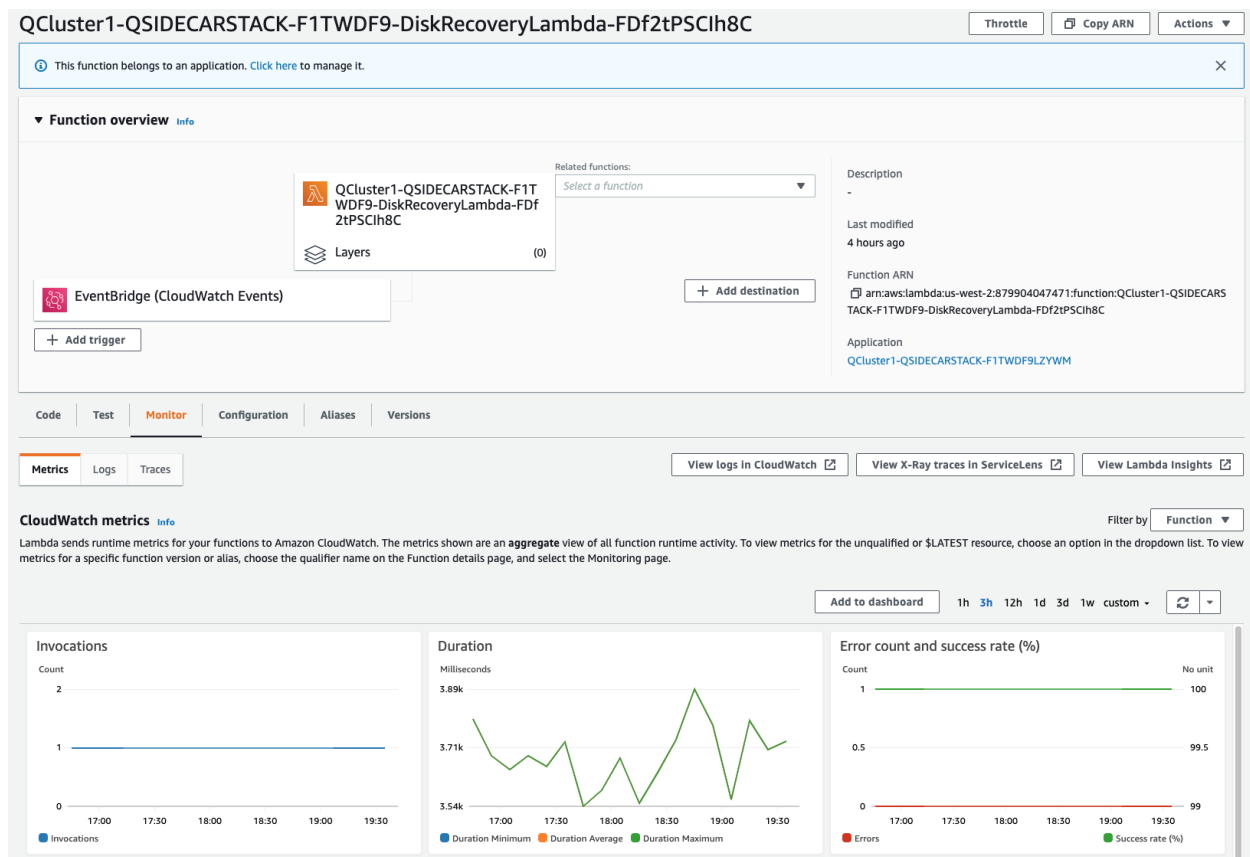
Verify Route 53 Private Hosted Zone for DNS (Optional)

In the AWS Console go to **Route 53**. Select the Private Hosted Zone that was created. In this case it is **test.local**. Verify the A-records were created with the A-record name specified in the template. This is only relevant if an FQDN was specified, otherwise Route 53 configuration is skipped. Note, 12 A-records were created, one for each floating IP, since 4 EC2 instances with 3 floating IPs were chosen in the template.

Route 53 > Hosted zones > test.local

test.local Info

▶ **Hosted zone details**

Records (14) | Hosted zone tags (1)

Records (14) Info
Automatic mode is the current search behavior optimized for best filter results. To change modes go to settings.

|	Record name ▽	Type ▽	Routin... ▽	Differ... ▽	Value/Route traffic to
☐	test.local	NS	Simple	-	ns-1536.awsdns-00.co.uk. ns-0.awsdns-00.com. ns-1024.awsdns-00.org. ns-512.awsdns-00.net.
☐	test.local	SOA	Simple	-	ns-1536.awsdns-00.co.uk. awsdns-hostmaster.amazon.com. 1 7200 900 1209600 86400
☐	qumulo.test.local	A	Weighted	0	172.31.76.253
☐	qumulo.test.local	A	Weighted	0	172.31.79.91
☐	qumulo.test.local	A	Weighted	0	172.31.79.122
☐	qumulo.test.local	A	Weighted	0	172.31.64.25
☐	qumulo.test.local	A	Weighted	0	172.31.77.135
☐	qumulo.test.local	A	Weighted	0	172.31.76.22
☐	qumulo.test.local	A	Weighted	0	172.31.77.155
☐	qumulo.test.local	A	Weighted	0	172.31.70.23
☐	qumulo.test.local	A	Weighted	0	172.31.78.16
☐	qumulo.test.local	A	Weighted	0	172.31.65.196
☐	qumulo.test.local	A	Weighted	0	172.31.69.146
☐	qumulo.test.local	A	Weighted	0	172.31.72.160

Verify Resource Groups

In the AWS Console go to **CloudWatch**. In the first filter box choose **EC2** and then in the **Filter by resource group box** select the cluster with **Qumulo-Cluster-EC2-[Stack Name]**. This provides a CloudWatch filtered view of the EC2 instances for the cluster. CPU Utilization, network stats, boot volume stats, and alarm events are available.
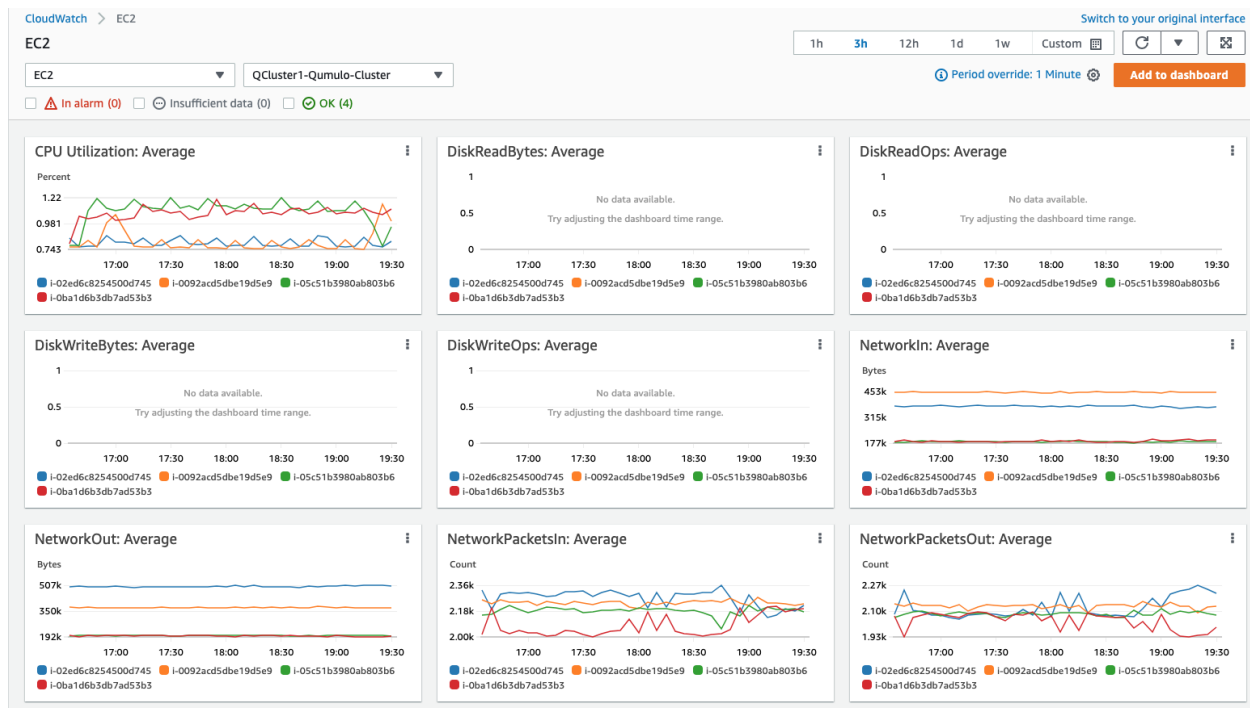
Now clear the **Filter by resource group field** and select **EBS** in the first filter box. Now in the **Filter by resource group field** choose the cluster with **Qumulo-Cluster-[SSD or HDD]-[Stack Name]**. This is a CloudWatch view of the EBS volumes for the cluster. Note, boot volumes are not included in this view.
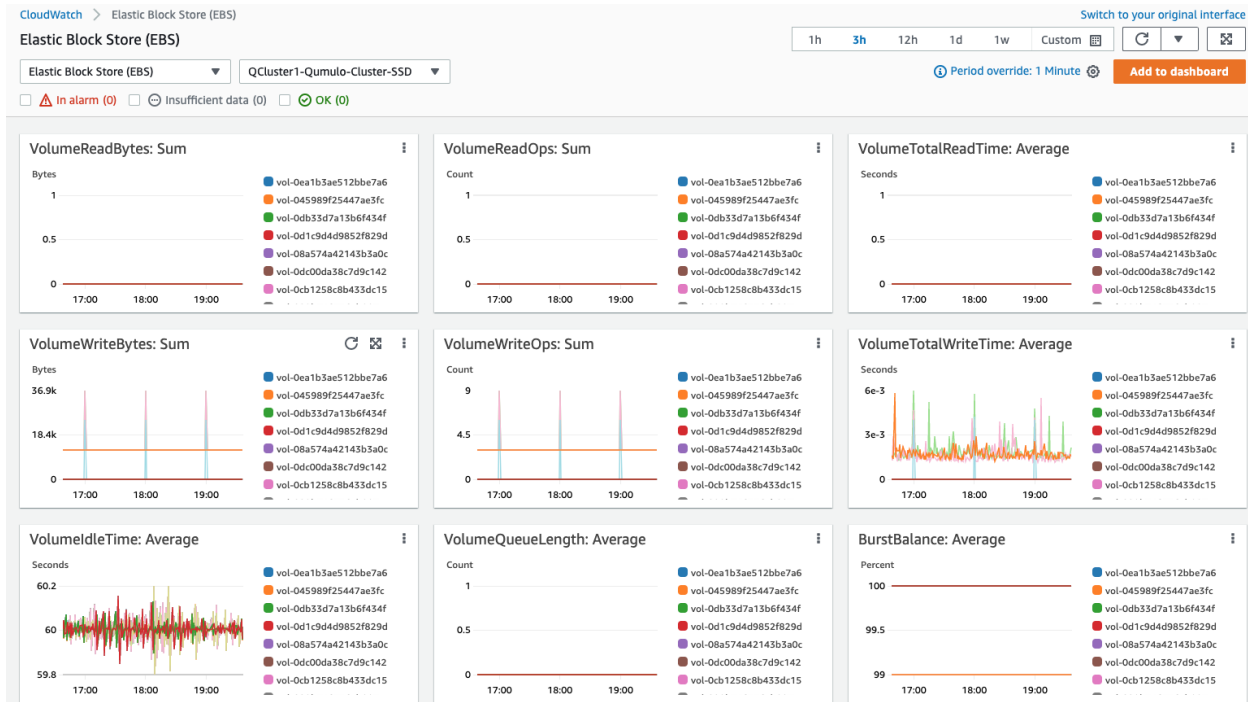
Verify CloudWatch Dashboard
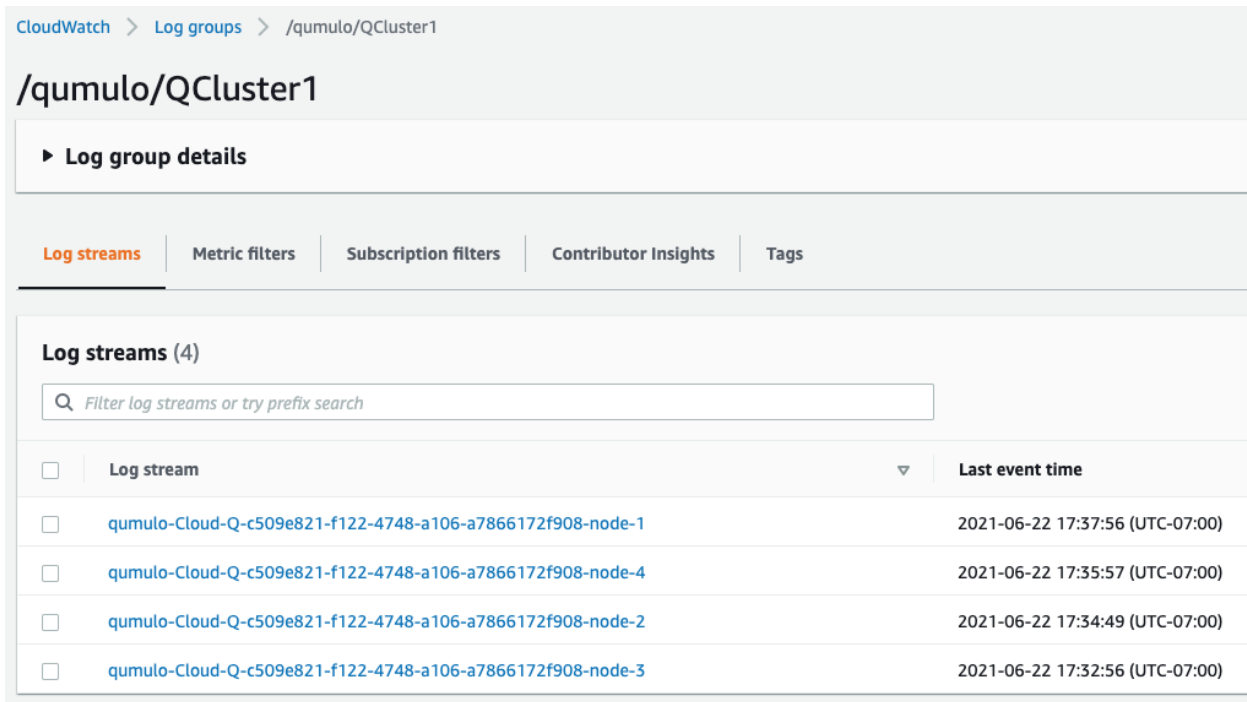
In the AWS Console go to **CloudWatch > Dashboard > Qumulo-Cluster-[Stack Name]-QSTACK-[123456789ABCD]**. This is a dashboard that has been built to display the metrics sent by the Qumulo Sidecar Metrics Lambda function. Instance health, EBS health, Available Capacity, and Performance data are all available. This dashboard is very useful for historical data that is over 72 hours old. For real-time data visit the Qumulo cluster's UI. Note: If you are deploying multiple clusters in an AWS region give them unique Qumulo Cluster Names. Metrics are filtered based on the Qumulo Cluster Name.
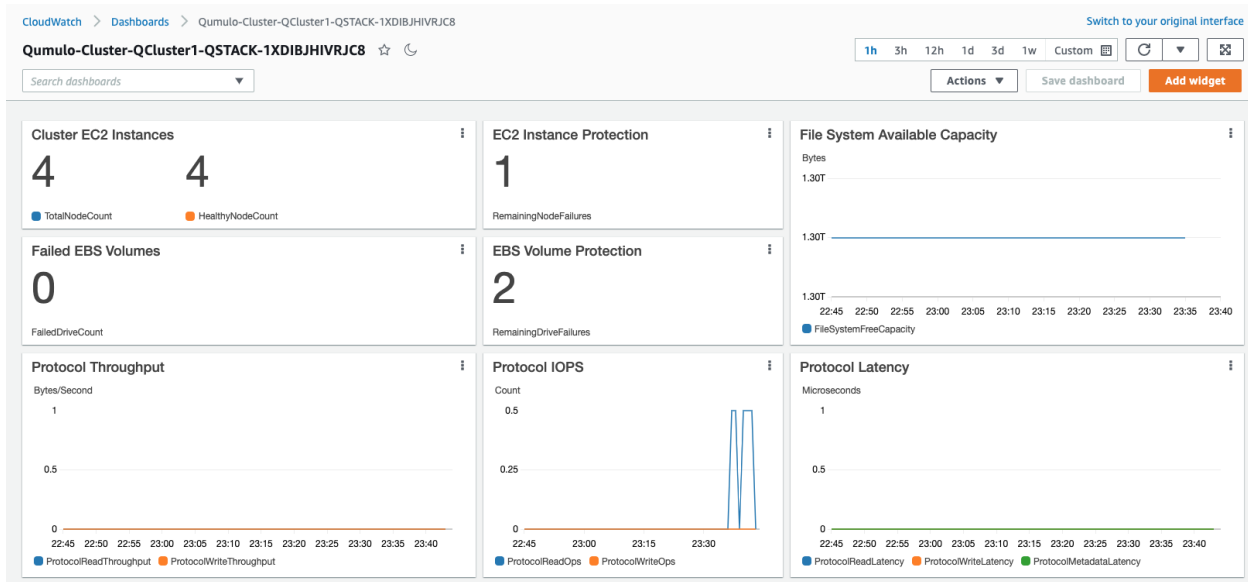
Verify CloudWatch Logs (Audit Logging)

In the AWS Console go to **CloudWatch > Log Groups > /qumulo/[Stack Name]**. This log group is configured if Audit Logging was enabled in the CloudFormation template. Log files will immediately be available for each instance in the cluster.

Provided by Qumulo
© 2021 Qumulo, Inc.
Contact Us

Review & Verify the Qumulo Cluster Configuration

Review the Outputs of the CloudFormation Stack

Go to the **CloudFormation** page and select the top-level stack name, **Qcluster1**. Choose **Outputs**. If Route 53 was configured a URL to the private addresses, resolved by Route 53, will be shown. If Route 53 was skipped, a URL to the first node's primary IP address will be displayed. Likewise, if Public Management was chosen a URL to the Elastic IP (public static) address will be shown. If connecting via the public Internet, open a page from your local machine using the **QumuloPublicIP** URL. If connecting from within your VPC, paste the **QumuloPrivateIP** URL into the browser of an EC2 instance running Chrome.
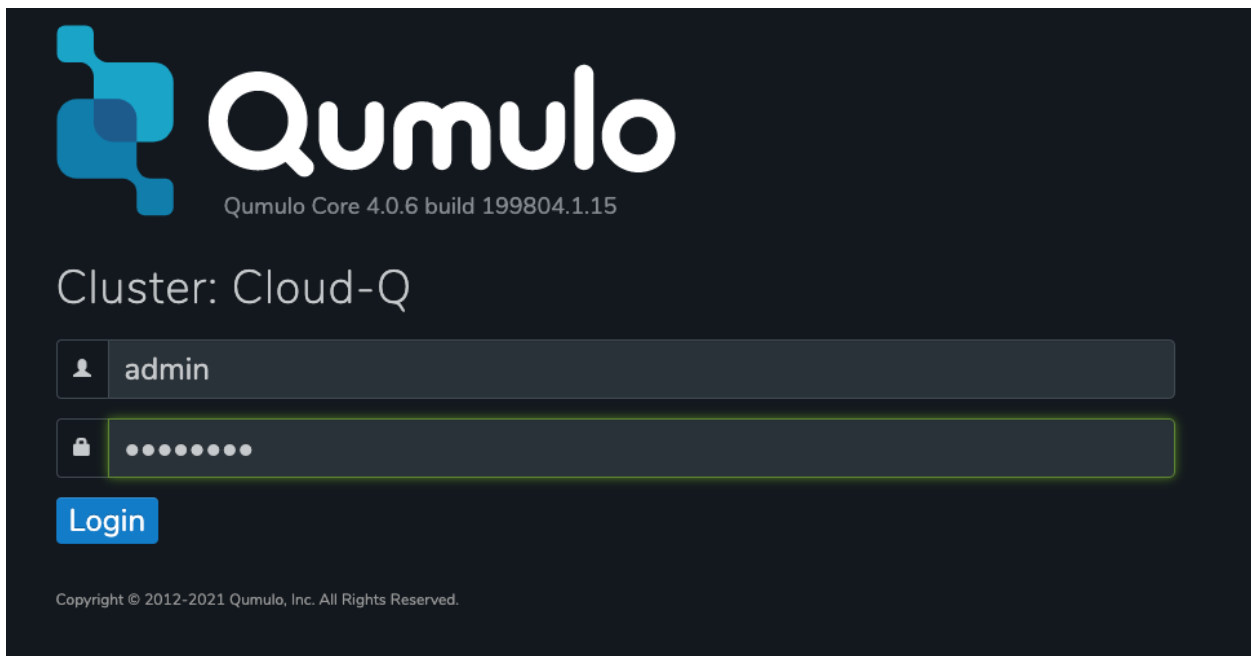
Verify Admin Password

The login page should authenticate with the credentials:
Username: **admin**
Password: '***your chosen Admin password***'
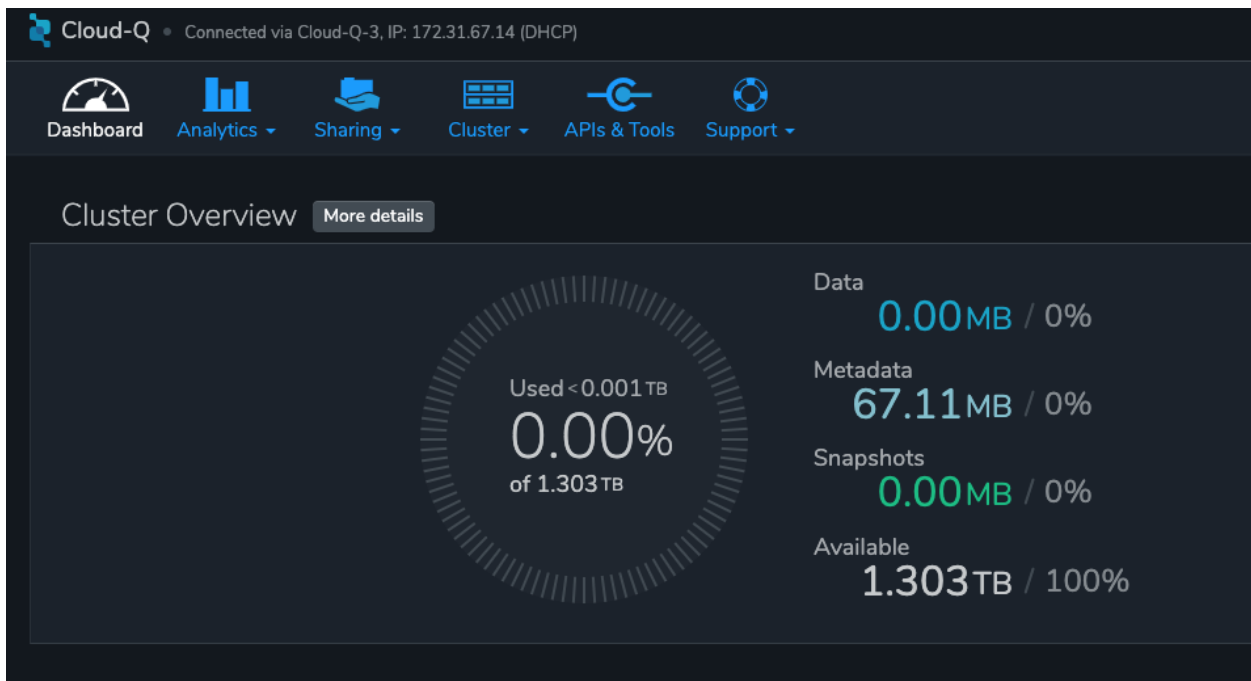If you've forgotten the admin password entered in the template go to Secrets Manager and retrieve it.

Verify Quorum and Protection

After logging in, the cluster dashboard should be displayed. IF it isn't the cluster failed to form quorum. Jump to troubleshooting.


Choose **More details**. The number of nodes in the cluster should match what was provisioned in the template. Further, to the right is the protection status showing protection for 1 node failure or 2 disk failures.
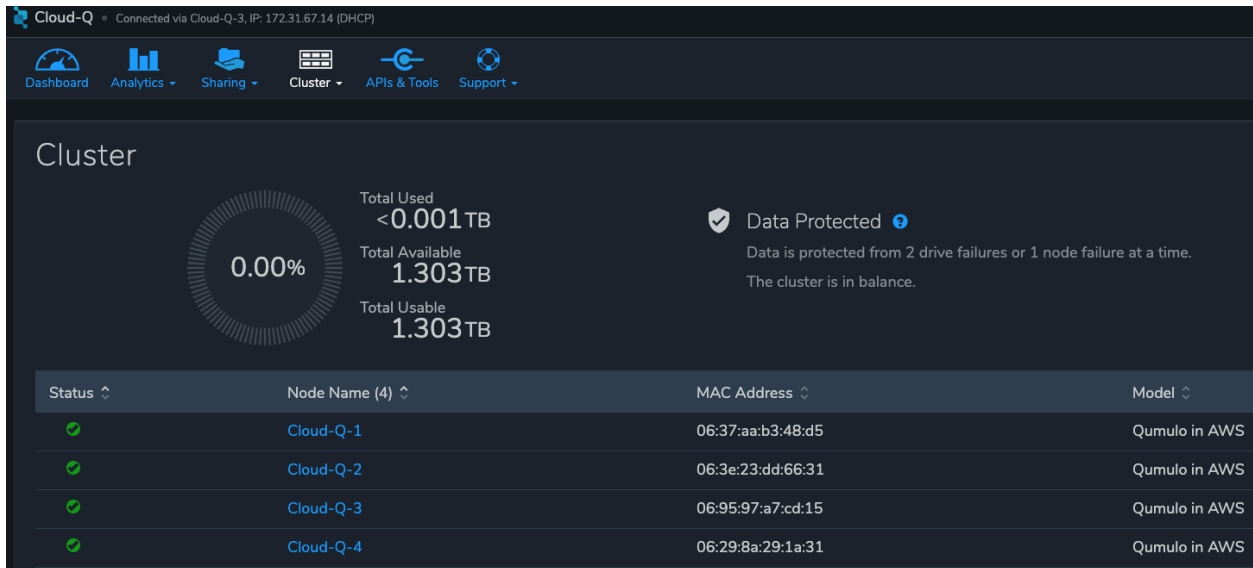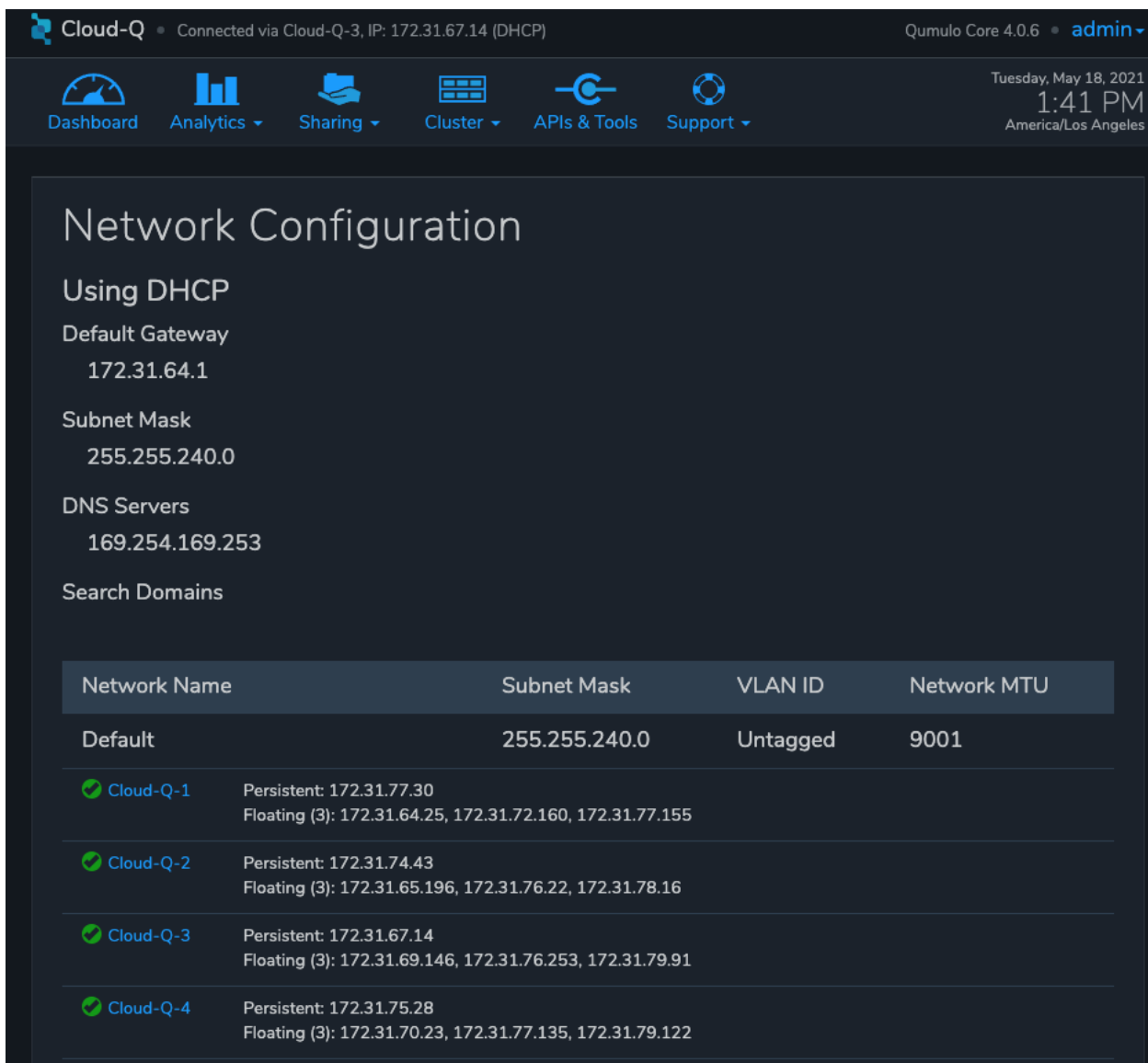
Verify Software Version

In the top right of the Qumulo UI the software version is displayed. This should match the software version requested when the template was filled in. Here it shows Qumulo Core version 4.0.6 as expected.
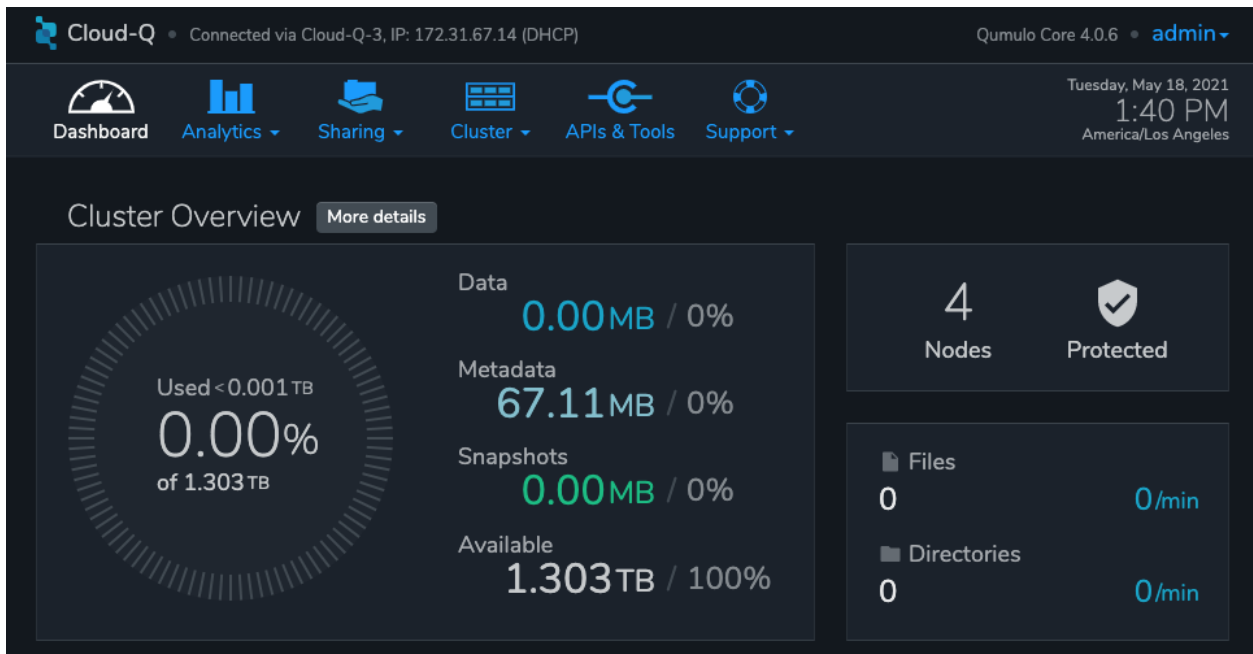
Verify Floating IPs

Go to the **Cluster** tab and select **Network Configuration**. Each node will have a persistent IP. This is the EC2 primary IP address that is provided via DHCP at creation and subsequently never changes unless the instance is destroyed (i.e. terminated). Also, each node will have floating IPs associated with it. In this case the default of 3 floating IPs per instance was chosen. These IPs are EC2 secondary IPs that the cluster now manages as floating IPs. The AWS EC2 console will only display what EC2 secondary IPs were assigned to an instance at creation. For real-time status always refer to the Qumulo UI.

Verify Sidecar User and Custom RBAC Configuration

Previously the Sidecar Lambda function connectivity to the cluster was verified. There's no need to review the Sidecar User and RBAC configuration. If you desire to review these they are under **Cluster** -> **Local Users & Groups** and **Cluster** -> **Role Management**, respectively.

Stack Update Options

Supported Stack Update Parameters

	Add	Del	Change
Total Number of Qumulo EC2 Instances			**increase**
Provision Qumulo Sidecar Lambdas	✅		
Qumulo Sidecar Software Version			✅
Enable Termination Protection	✅	✅	✅
OPTIONAL: Provision Public IP for Qumulo Management	✅	✅	✅
OPTIONAL: Enable Replication Port for Qumulo Public IP	✅	✅	✅
OPTIONAL: FQDN for R53 Private Hosted Zone	✅	✅	✅
OPTIONAL: R53 Record Name for Qumulo RR DNS	✅	✅	✅
OPTIONAL: Send Qumulo Audit Log messages to CloudWatch Logs?	✅	✅	✅
OPTIONAL: AWS Permissions Boundary Policy Name	✅	✅	✅

Adding Node(s) to the Cluster

A Qumulo cluster may be grown in both capacity and performance by adding additional nodes (EC2 instances) to the cluster. This stack supports adding as many as 16 nodes in one stack update for a maximum of 20 nodes total in the cluster. Each node added increases compute, networking, and storage capacity. To add nodes to a cluster follow the procedure below. Note, total instance count may only be increased, not decreased. If total instance count is decreased the stack update will fail and rollback.
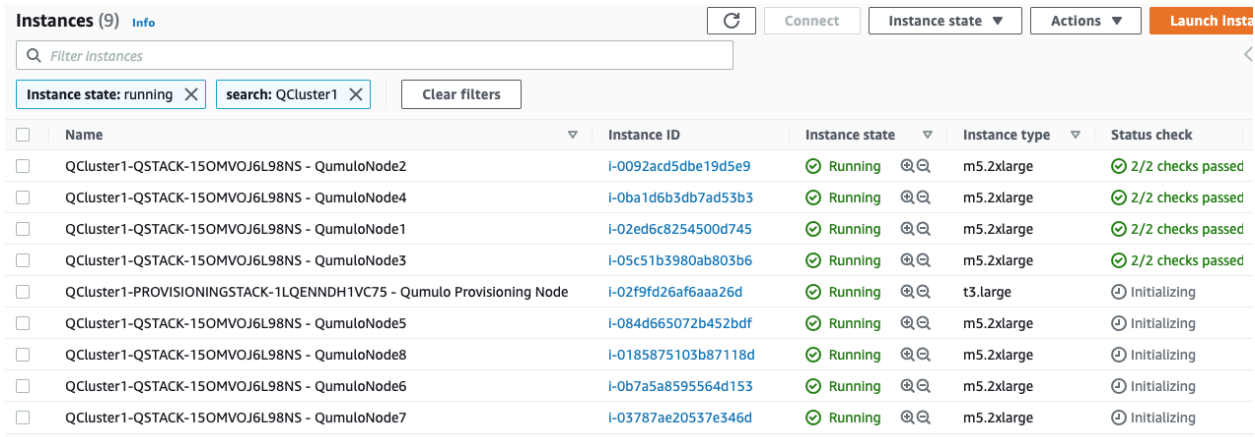
IF you have upgraded the software on the cluster after initial deployment leave the software version for the cluster in the template as it was originally provisioned. The stack is unaware of this update and the software version field for the cluster can not be used for upgrades after initial deployment.

1. Go to the **CloudFormation** view in the AWS Console

2. Select the top-level stack name, **QCluster1** in this example
3. Select **Update** in the upper right corner
4. Keep the default **Use Current Template**
5. Select **Next**
6. The template as last populated will be displayed
7. Scroll down to the **Total Number of Qumulo EC2 Instances**
8. Increase the number of instances to the chosen value, **8** in this example
9. Select **Next**
10. Select **Next** again
11. **Check both boxes** acknowledging that CloudFormation may create IAM roles and that it may leverage CAPABILITY_AUTO_EXPAND.
12. Select **Update stack**

The stack will commence updating. In this case four nodes will be added to the cluster. This is not service impacting as the existing nodes are left untouched. There is a brief quorum bounce to add the four new nodes to the cluster. Below is a view of the AWS EC2 Console showing the new instances initializing.

Notice that the Provisioning instance is also being restarted. This is by design. The Provisioner will query the latest version of software running on the cluster and upgrade all new nodes to this version of software before joining them to the cluster. Further, it tags all the new EBS volumes and updates the floating IPs.

This stack provisioned Public Management and Route 53 originally. With the addition of new nodes, IP addresses need to be added to the Load Balancer and the Route 53 Private Hosted Zone. The stack will automate these updates as well. You may review any nested stack to see what resources were modified or added in the stack **Events** tab. At the completion of node

addition you may review any and all of the AWS infrastructure referencing the former section. As a final check make sure the Provisioning node shutdown which indicates success of all secondary provisioning.

Finally, login to the cluster and verify the node addition.

Updating the Sidecar Software Version

A stack update may be used to update the Sidecar software version. Follow the same steps as a Node Addition, but change the **Sidecar Software Version** field to the desired version instead of changing the number of EC2 instances (steps 7 & 8 above). This is typically done after updating the cluster software via the Qumulo UI.

Adding or Removing Public Management

A stack update may be used to add or remove public management. Since this update is completely separate from the cluster there's no changes required to the cluster infrastructure or infrastructure touched by the Provisioning instance. Hence, it will not restart. Follow the same steps as a Node Addition, but change the **OPTIONAL: Provision Public IP for Qumulo Management** parameter to 'YES/NO' instead of changing the number of EC2 instances (steps 7 & 8 above). Note, the MGMTNLBSTACK will be deleted when removing public management. This is expected. The stack will show as DELETE_FAILED for a period of time while CloudFormation retries the delete of the Elastic IP. Ultimately it will succeed.

Adding or Removing Route53 DNS Private Hosted Zone

It is possible to change the R53 FQDN, but AWS requires the deletion of the current Private Hosted Zone and a new one will be rebuilt if the FQDN is modified in a stack update. To remove the private hosted zone, set the FQDN back to NONE.local. In the stack update pages you can review the changes the update will make. Follow the same steps as a Node Addition, but change the **OPTIONAL: FQDN for R53 Private Hosted Zone** parameter to the desired value instead of changing the number of EC2 instances (steps 7 & 8 above).

Enabling or Disabling Audit Logging

A stack update may be used to enable or disable Qumulo audit logging. These logs are stored in a CloudWatch Logs log group. If a stack update is used to disable audit logging the log group will be deleted. Likewise, if audit logging is enabled in a stack update a log group will be created with the name **/qumulo/[Stack Name]**. Follow the same steps as a Node Addition, but change the **OPTIONAL: Send Qumulo Audit Log messages to CloudWatch Logs?** parameter to 'YES/NO' instead of changing the number of EC2 instances (steps 7 & 8 above).

Adding the Qumulo Sidecar Lambdas

If the Sidecar was not deployed with the Cluster originally, it may be added subsequently to the stack. Follow the same steps as a Node Addition, but change the **Provision Qumulo Sidecar Lambdas** parameter to 'Yes' instead of changing the number of EC2 instances (steps 7 & 8 above). Removing the Sidecar lambdas is not supported.

Enabling or Disabling Termination Protection

A stack update may be used to enable or disable Termination Protection for the EC2 instances and the CloudFormation stack. Termination protection should be enabled in all production environments. Only disable it with a stack update prior to deleting the stack.

Other Stack Updates and the QSTACK Policy

The only restrictions placed on stack updates are for the Qumulo cluster. Specifically this is the QSTACK. The stack policy is applied by the Provisioning instance, and it forbids any modifications, deletions, or recreations of QSTACK EC2 and EBS infrastructure. This is to protect production environments from erroneous stack updates. In the event a stack update is attempted for an unsupported change the update will simply fail and rollback without harm. Many stack updates are possible and not all permutations have been tested. The common examples are documented above that are most productive and well tested.

Changing Instance Types and EBS Volume Types

Qumulo does not support changing the cluster instance types with a stack update. This is prevented with the aforementioned stack policy. While it would be possible if allowed, it would stop all the instances, change the instance type, and restart them. This would be service impacting in a production environment. Instead Qumulo recommends shutting down an instance at a time so the cluster can leverage floating IPs and maintain the production workload.
Due to the permutations of EBS volume configurations the likelihood of user error is high attempting to change EBS volume types with a stack update. Rather than risk data loss this is blocked by the QSTACK policy.
For both instance type changes and EBS volume type changes Qumulo offers simple scripts that are production friendly.

Protecting Production Environments

In production deployments it is wise to enable Termination Protection for the entire stack and the EC2 instances. The template provides this protection by default.

Deleting the Stack

When a cluster is no longer needed ensure all critical data has been removed from the cluster. Qumulo's SHIFT functionality may be used to natively copy data from the cluster to S3. Alternatively, Qumulo supports S3 Snapshots but rehydration will require a cluster with the same EBS volume configuration. Once the data has been archived with the chosen method then use CloudFormation to update the stack to **Disable Termination Protection**. Finally, select the **top-level stack** in CloudFormation and choose **Delete**. All resources will be deleted.

If a Customer Managed Key was used for encryption at rest, the KMS CMK policy must be cleaned up. It's simplest to do this after the stack is completely deleted. AWS CloudFormation does not support CMK policy modifications so it is unable to track these changes that the Provisioning instance applied. Go to the **AWS Key Management Service** and select the **CMK** that was used. Then **Edit** the policy. **Delete** the two SIDs for the Sidecar and select **Save**. If the key policy had no other SIDs applied to it, aside from the Qumulo Sidecar SIDs, it will have the following JSON structure before and after being cleaned up.

After Stack Deletion but before Cleanup

Key policy

```
14        {
15            "Sid": "Allow use of the key",
16            "Effect": "Allow",
17            "Principal": {
18                "AWS": "AROA4ZXS6IFX6B4SWQZI5"
19            },
20            "Action": [
21                "kms:Encrypt",
22                "kms:Decrypt",
23                "kms:ReEncrypt*",
24                "kms:GenerateDataKey*",
25                "kms:DescribeKey"
26            ],
27            "Resource": "*"
28        },
29        {
30            "Sid": "Allow attachment of persistent resources",
31            "Effect": "Allow",
32            "Principal": {
33                "AWS": "AROA4ZXS6IFX6B4SWQZI5"
34            },
35            "Action": [
36                "kms:CreateGrant",
37                "kms:ListGrants",
38                "kms:RevokeGrant"
39            ],
40            "Resource": "*",
41            "Condition": {
42                "Bool": {
43                    "kms:GrantIsForAWSResource": "true"
44                }
```

After Cleanup

> KMS > Customer managed keys > 3d26f779-69a4-4de1-b16f-3a69152ce1ee > Edit policy
>
> # Edit key policy
>
> **Key policy**
>
> ```
>  1 {
>  2     "Id": "key-consolepolicy-3",
>  3     "Version": "2012-10-17",
>  4     "Statement": [
>  5         {
>  6             "Sid": "Enable IAM User Permissions",
>  7             "Effect": "Allow",
>  8             "Principal": {
>  9                 "AWS": "arn:aws:iam::879904047471:root"
> 10             },
> 11             "Action": "kms:*",
> 12             "Resource": "*"
> 13         }
> 14     ]
> 15 }
> ```
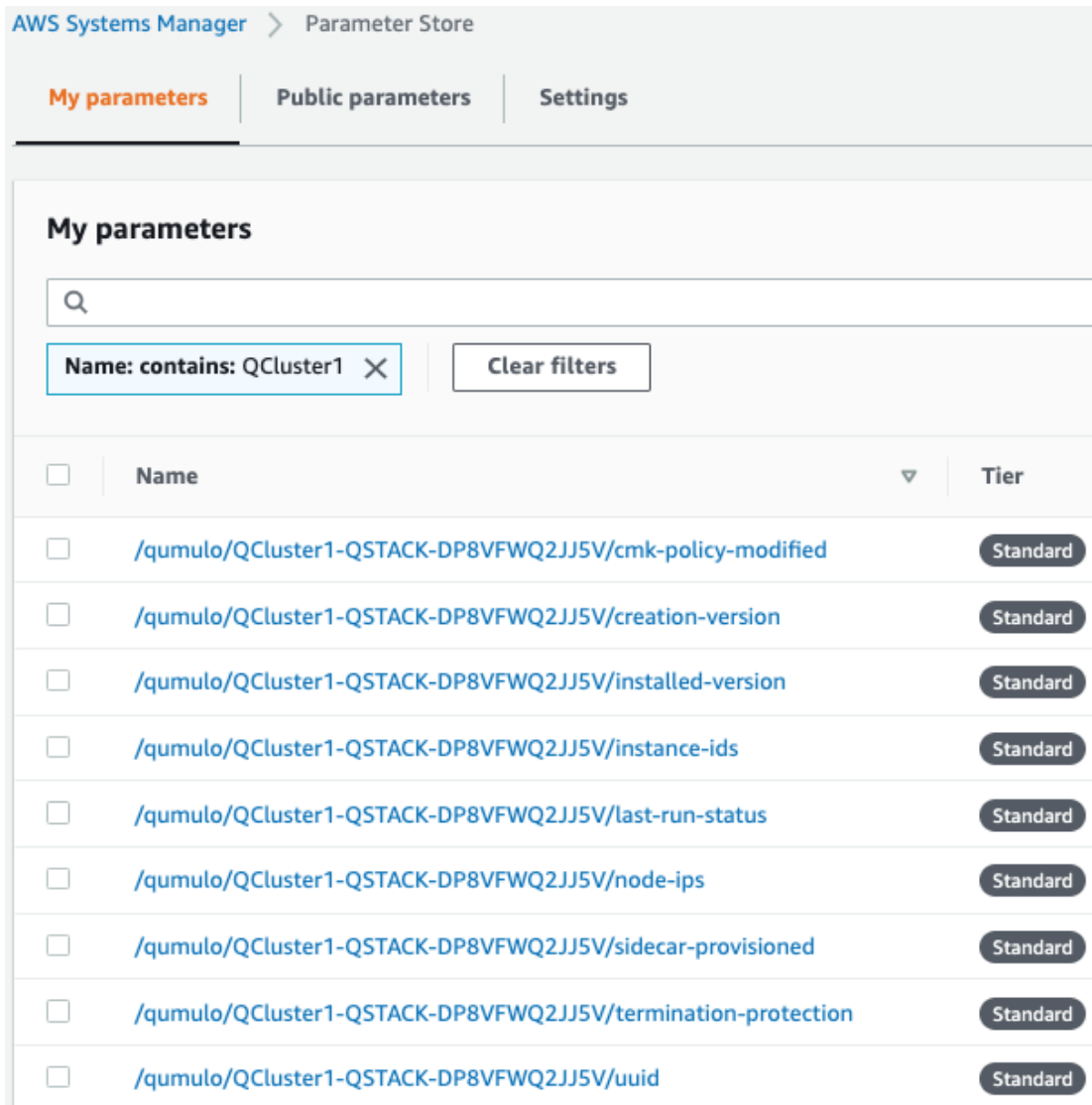
As of the date of this document AWS CloudFormation will fail to delete all of the MGMTNLB stack resources (If Public Management was provisioned). Simply let the deletion finish, reselect the MGMTNLB stack and delete it again, and then delete the top-level stack.

Troubleshooting

Where's the UUID for the cluster?

The Provisioning instance grabs a copy of the UUID for the cluster after the first quorum is formed. Go to **Parameter Store** and filter on the top-level stack name. The following parameters are stored by the Provisioning instance. The UUID is last on the list. Select it to view the UUID.

Forgot the cluster admin password

The admin password entered when the cluster was originally provisioned is stored in AWS Secrets Manager. Go to **Secrets Manager** and filter on the top-level stack name. Then look for **ClusterSecrets** and then **Retrieve secret value**.   Also, if the admin password is changed post-deployment, it must be updated in Secrets Manager for stack updates to function correctly.

The Stack failed on the first nested stack, QLOOKUP or SECRETSSTACK

The S3 Bucket, Key Name Prefix, or Object URL are not correct. Delete the stack and relaunch the template with the correct S3 parameters. Do NOT use the S3 URL, use the Object URL for the template or the stack will fail.

The Stack failed when provisioning the QSTACK

The four most common causes for this are:
1. An AWS Marketplace offer has not been accepted that matches the **Qumulo AWS Marketplace Offering Accepted** parameter entered in the template
2. The EBS volumes configuration doesn't match the requirements for the **Qumulo AMI ID** entered when using the Specified-AMI-ID option
3. The cluster failed to place in the placement group
4. Service Quotas were not pre-planned and the QSTACK failed

Review the AMI ID and marketplace subscriptions. Double check the EBS volume config selected in the template. If the cluster failed to place, choose a different AZ to deploy the cluster in to find more available resources by selecting a different private subnet ID within the VPC. Adjust Service Quotas if necessary. Delete the failed stack and relaunch the template after rectifying the problem.

The Stack Update failed and rolled back

No harm is done. No Qumulo Cluster parameters for the QSTACK, except the Number of EC2 Instances, can be changed. The number of instances can't be decreased.

The Cluster didn't form quorum

The four most common causes for this are:
1. The software version specified in the template doesn't exist
2. The software version specified in the template is older than the AMI software version
3. The S3 Bucket Region specified is incorrect
4. The VPC doesn't have public internet access and the image wasn't placed in the /upgrade folder in the S3 bucket, see **The Provisioning instance didn't shutdown**

Check for typos by reviewing the parameters entered in the template in the CloudFormation console. Double check the software version specified for the cluster and make sure it is equal to or newer than the version the Marketplace offer lists. Finally, double check the S3 Bucket Region entered in the template. Rectify and restart the Provisioning instance or delete the stack and redeploy.

The Provisioning instance didn't shutdown

Common Causes

The four most common causes for this are:
1. The VPC doesn't have access to the public Internet or DNS resolution is not functioning. Without access to public infrastructure the Provisioning instance can't talk to AWS services like Secrets Manager, KMS, Parameter Store, or download the desired version of Qumulo Core software. Review the public and private subnets, their route tables, and the NAT Gateway. Review the AWS Parameter Store **last-run-status** to verify public internet connectivity (see the section below on last-run-status). Also double check that there are no Network ACLs blocking traffic.
2. The VPC doesn't have access to the public Internet, but this was planned. One or more VPC Endpoints may be missing. The VPC Interface Endpoints Security Group is not correct. The desired Qumulo Core software version has not been placed in the S3 Bucket /upgrade folder. **See the section Deploying without Internet Access**.
3. A Customer Managed Key was provisioned and the policy was unable to be modified for the CMK because the policy didn't have valid SIDs before the template was launched.
4. A stack update was executed to add nodes. The stack update succeeded but the provisioning instance didn't shutdown and the nodes aren't added to the cluster. Most likely the Cluster's admin password was changed post deployment. If this is the case go to **Secrets Manager**, filter on the top-level stack name, and look for **ClusterSecrets**. **Retrieve secret value** and **Edit**. Update the admin password and save the secret. Then stop and restart the provisioning instance.

Cleanup the CMK, correct the VPC infrastructure, update the admin password and restart the provisioning instance. See the sections that follow on restarting the provisioning instance, monitoring its status in the Parameter Store, and downloading logs.

AWS Parameter Store last-run-status

If the Provisioning instance doesn't automatically shutdown, the AWS Systems Manager Parameter Store **last-run-status** parameter may be checked to see where it stopped. As shown below, the parameter history shows the major blocks in the code the provisioning instance executes. In this example QCluster1 was built for the first time as noted by the NEW CLUSTER update to the last-run-status parameter

/qumulo/QCluster1-QSTACK-DP8VFWQ2JJ5V/last-stati

Overview   **History**   **Tags**

Versions

	Version	Value	Tier
○	13	Shutting down provisioning instance	Standard
○	12	Tagging untagged EBS volumes	Standard
○	11	Applying CMK policy	Standard
○	10	Provisioning Sidecar info on Cluster	Standard
○	9	Forming first quorum and configuring cluster	Standard
○	8	Upgrading unconfigured nodes to 4.2.0	Standard
○	7	Downloading qumulo_upgrade_cloud_4.2.0.qimg from Trends.qumulo.com	Standard
○	6	All nodes out of quorum, NEW CLUSTER	Standard
○	5	Checking quorum state and boot status	Standard
○	4	Updating Termination Protection. Enabled=YES	Standard
○	3	Installing jq, qq, and nginx if needed and reading secrets	Standard
○	2	BOOTED. Public Internet Reachable	Standard
○	1	null	Standard

Restarting the Provisioning Instance

The Provisioning instance is designed to restart with every Stack Update. Further, it may be manually stopped from the AWS Console, if it doesn't automatically stop within 10 minutes, and then manually restarted. Examples where this may be very helpful are if software wasn't placed in the S3 bucket when deploying without internet access, a CMK policy wasn't cleaned up prior to deployment, or intended internet connectivity wasn't functioning as expected and has been rectified.

Download the Provisioning instance log

In the event none of the troubleshooting steps help to rectify the problems it's likely the Provisioning instance log will be helpful. To retrieve the log follow these steps:
1. Go to the AWS Console **EC2 Instances** page
2. **Check the box** beside the Provisioning instance
3. Select **Actions** in the upper right corner
4. Select **Monitor & troubleshoot**
5. Select **Get system log**
6. Select Download in the upper right corner

Feel free to review the log right in the AWS console or download it to collaborate with Qumulo to resolve the problem. Often the log will show an obvious error pointing you to the resolution.

Provisioning instance flow chart

The provisioning instance executes the code in user data every boot cycle. The abbreviated logic diagram below shows the major branches and AWS SSM Parameter Store values for **last-run-status** throughout the execution of the code.

Provisioning Flowchart

Provisioning Instance Boots

- **Internet reachable?**
  - NO → last-run-status: BOOTED. NO Public Internet Connectivity. VPC Endpoints must be provisioned for AWS Services.
  - YES → last-run-status: BOOTED. Public Internet Reachable

last-run-status: Installing jq , qq , and nginx if needed and reading secrets

- **qq, jq, nginx already installed?**
  - NO → Download and install qq, jq, nginx
  - YES → Get Secrets, usernames & passwords

- **Change Term Protection?**
  - YES → last-run-status: Updating Termination Protection → Modify Term Protection
  - NO → (continue)

- **All nodes out of Quorum?**
  - NO → last-run-status: 4 or more nodes in quorum, checking for node adds
    - **New Nodes?**
      - NO → (continue)
      - YES → (continue)
  - YES → last-run-status: All nodes out of quorum, NEW CLUSTER → Set QSTACK Protection Policy

- **SW Upgrade required?**
  - NO → (continue)
  - YES →
    - **SW Already downloaded?**
      - NO → Download Qumulo Core x.y.z from Internet or S3 Bucket if no Internet access
      - YES → Upgrade Nodes

- **New Cluster?**
  - NO → last-run-status: Quorum already exists, adding nodes to cluster → Add nodes to cluster and check for quorum
  - YES → last-run-status: Forming first quorum and configuring cluster → Form quorum and configure cluster

- **Prov Sidecar?**
  - YES → last-run-status: Provisioning Sidecar Info on Cluster → Config Cluster for Sidecar
  - NO → (continue)

- **CMK for encryption?**
  - YES → last-run-status: Applying CMK Policy → Modify KMS CMK Policy
  - NO → (continue)

last-run-status: Tagging untagged EBS Volumes → Tag Volumes by Type → last-run-status: Shutting down provisioning instance

Provisioning Instance Stopped

Contact Us