HashiCorp Consul on the AWS Cloud

Quick Start Reference Deployment

November 2016
Last update: January 2020 (revisions)

Daniel Callao, HashiCorp, Inc.
Tony Vattathil and Andrew Gargan, Amazon Web Services

Contents

This Quick Start deployment guide was created by Amazon Web Services (AWS) in partnership with HashiCorp, Inc.

Overview

This Quick Start reference deployment guide provides step-by-step instructions for deploying HashiCorp Consul on the Amazon Web Services (AWS) Cloud. Quick Starts are automated reference deployments that use AWS CloudFormation templates to launch, configure, and run the AWS compute, network, storage, and other services required to deploy a specific workload on AWS.

HashiCorp Consul is a tool that provides a foundation for cloud networking automation by using a central registry for service-based networking. Consul's core use cases include the following:

- Service registry and health monitoring, to provide a real-time directory of all services with their health status
- Network middleware automation, with service discovery for dynamic reconfiguration as services scale up, scale down, or move
- Zero Trust network with service mesh, to secure service-to-service traffic with identity-based security policies and encrypted traffic with Mutual-Transport Layer Security (TLS)

Consul is designed to be friendly to both the DevOps community and application developers, making it a good fit for modern, elastic infrastructures.

This Quick Start is for users who are looking for a solution for service discovery, monitoring, or a key/value store. The Quick Start is built using the open-source version of Consul, but is also compatible with Consul Enterprise.

An expanded version of this deployment guide with detailed instructions and screen illustrations is available on the HashiCorp Consul and Consul Enterprise websites.

For other solutions from HashiCorp and AWS, see the AWS Quick Start for HashiCorp Vault.

Please know that we may share who uses AWS Quick Starts with the AWS partner that collaborated with AWS on the content of the Quick Start.

Costs and licenses

You are responsible for the cost of the AWS services used while running this Quick Start reference deployment. There is no additional cost for using the Quick Start.

The AWS CloudFormation template for this Quick Start includes configuration parameters that you can customize. Some of these settings, such as instance type, will affect the cost of deployment. See the pricing pages for each AWS service you will be using for cost estimates.

This Quick Start uses the open-source version of HashiCorp Consul, which doesn't require a license.

Architecture

Deploying this Quick Start with the **default parameters** builds the following Consul environment in its own virtual private cloud (VPC) in the AWS Cloud. For details about the VPC architecture, see the Amazon VPC Quick Start Guide.)

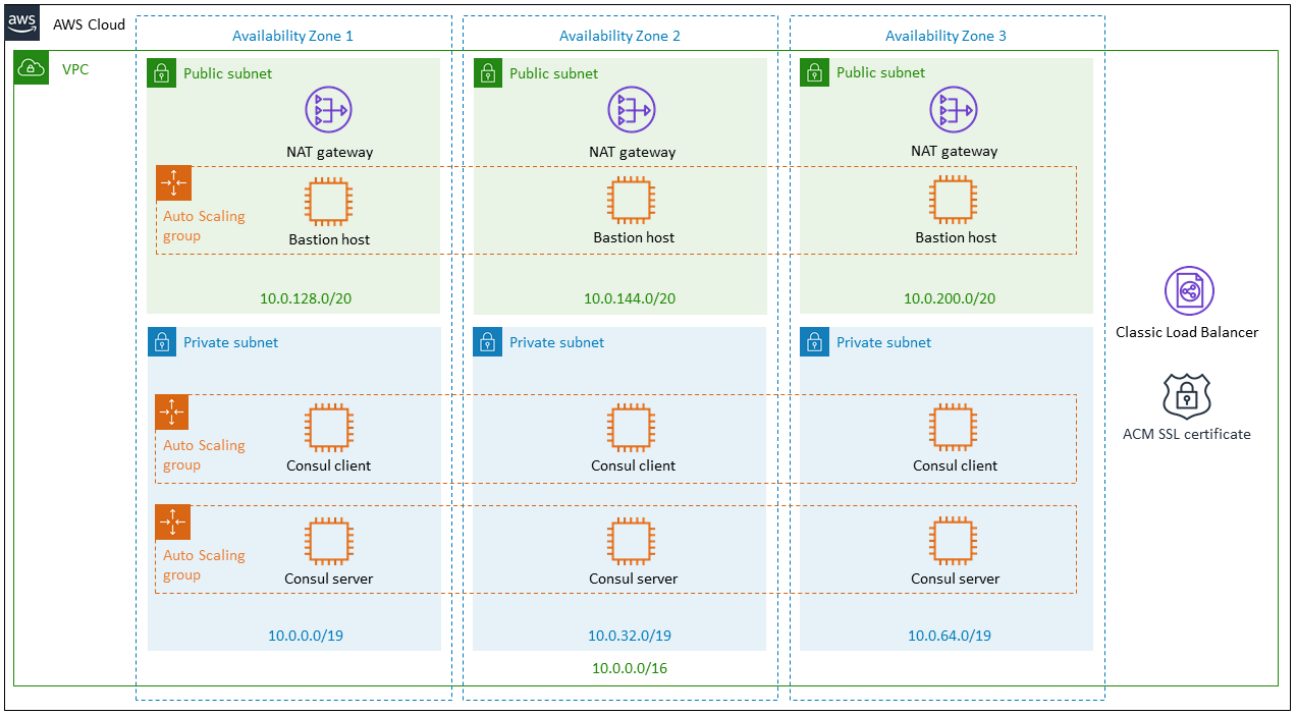


Figure 1: Quick Start architecture for HashiCorp Consul on AWS

The Quick Start provides two deployment options:

- **Deployment of HashiCorp Consul into a new VPC** (end-to-end deployment) builds a new VPC with public and private subnets, and then deploys HashiCorp Consul into that infrastructure.
- **Deployment of HashiCorp Consul into an existing VPC** provisions HashiCorp Consul into your existing infrastructure.

If you use the deployment option to create a new VPC, the AWS CloudFormation template included with the Quick Start will create the following components:

- A VPC with public and private subnets across three Availability Zones.*
- In the public subnets:
  - Linux bastion hosts to allow inbound Secure Shell (SSH) access to EC2 instances in the private subnets.*
  - Managed network address translation (NAT) gateways to allow outbound internet access for resources in the private subnets.*
  - A Classic Load Balancer with AWS Certificate Manager (ACM) attached to the Consul server cluster Auto Scaling group. You can choose to add your own Secure Sockets Layer (SSL) certificate to it.
- In the private subnets:
  - An Auto Scaling group for Consul clients. The number of clients is set to 0 by default, but is user-configurable.
  - An Auto Scaling group for a Consul server cluster in the private subnets. You can choose to create 3, 5, or 7 servers.
  - Consul Template (the `consul-template` daemon) installed on all nodes for integrating applications with Consul's service catalog and key/value store.
  - **Dnsmasq** installed on all nodes for integrating applications with Consul's Domain Name System (DNS) interface for service discovery.

* The template that deploys the Quick Start into an existing VPC skips the components marked by asterisks and prompts you for your existing VPC configuration.

Prerequisites

Specialized knowledge

Before you deploy this Quick Start, we recommend that you become familiar with the following AWS services. (If you are new to AWS, see Getting Started with AWS.)

- Amazon VPC
- Amazon EC2

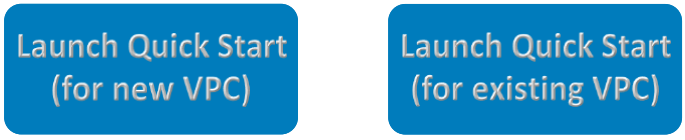
Deployment steps

Step 1. Prepare an AWS account

1. If you don't already have an AWS account, create one at http://aws.amazon.com by following the on-screen instructions.
2. Use the region selector in the navigation bar to choose the AWS Region where you want to deploy HashiCorp Consul on AWS.
3. Create a key pair in your preferred region.
4. If necessary, request a service quota increase for the Amazon EC2 **t2.medium** instance type. You might need to do this if you already have an existing deployment that uses this instance type, and you think you might exceed the default quota with this reference deployment.

Step 2. Launch the Quick Start

1. Choose one of the following options to deploy the AWS CloudFormation template into your AWS account.



The templates are launched in the US West (Oregon) region by default. You can change the region by using the region selector in the navigation bar.

Each stack takes approximately 10 minutes to create.

Note You are responsible for the cost of the AWS services used while running this Quick Start reference deployment. There is no additional cost for using this Quick Start. See the pricing pages for each AWS service you will be using for full details.

2. On the **Select Template** page, keep the default setting for the template URL, and then choose **Next**.
3. On the **Specify Details** page, review the parameters for the template. Enter values for the parameters that require your input. For all other parameters, you can customize the default settings provided by the template.

In the following tables, parameters are listed and described separately for deploying HashiCorp Consul into a new VPC or an existing VPC.

Note The templates for the two scenarios share most, but not all, of the same parameters. For example, the template for an existing VPC prompts you for the VPC and private subnet IDs in your existing VPC environment. You can also download the templates and edit them to create your own parameters based on your specific deployment scenario.

- **Parameters for deployment into a new VPC**

View template

VPC Network configuration:

Parameter label (name)	Default	Description
Availability Zones (`AvailabilityZones`)	*Requires input*	List of Availability Zones to use for the subnets in the VPC. Note: The logical order is preserved. (Three Availability Zones are used for this deployment.)
VPC CIDR (`VPCCIDR`)	10.0.0.0/16	CIDR block for the VPC.
Private Subnet 1 CIDR (`PrivateSubnet1CIDR`)	10.0.0.0/19	CIDR block for private subnet 1 located in Availability Zone 1.
Private Subnet 2 CIDR (`PrivateSubnet2CIDR`)	10.0.32.0/19	CIDR block for private subnet 2 located in Availability Zone 2.
Private Subnet 3 CIDR (`PrivateSubnet3CIDR`)	10.0.64.0/19	CIDR block for private subnet 3 located in Availability Zone 3.
Public Subnet 1 CIDR (`PublicSubnet1CIDR`)	10.0.128.0/20	CIDR block for the public (DMZ) subnet 1 located in Availability Zone 1
Public Subnet 2 CIDR (`PublicSubnet2CIDR`)	10.0.144.0/20	CIDR block for the public (DMZ) subnet 2 located in Availability Zone 2.
Public Subnet 3 CIDR (`PublicSubnet3CIDR`)	10.0.160.0/20	CIDR block for the public (DMZ) subnet 3 located in Availability Zone 3.
Permitted IP range (`AccessCIDR`)	*Requires input*	The CIDR IP range that is permitted to access Consul. Note: A value of 0.0.0.0/0 will allow access from ANY iP address.

Consul setup:

Parameter label (name)	Default	Description
Consul cluster node instance type (`ConsulInstanceType`)	m5.large	The EC2 instance type for the Consul instance.

Parameter label (name)	Default	Description
Number of Consul server nodes (`ConsulServerNodes`)	3	The number of Consul server nodes that will be created. You can choose 3, 5, or 7 nodes.
Number of Consul client nodes (`ConsulClientNodes`)	0	The number of Consul client nodes that will be created.
Key name (`KeyPairName`)	*Requires input*	Public/private key pair, which allows you to connect securely to your instance after it launches. When you created an AWS account, this is the key pair you created in your preferred region.

DNS and SSL configuration:

Parameter label (name(	Default	Description
Load Balancer FQDN (`LoadBalancerFQDN`)	*Optional*	The fully qualified domain name for the load balancer. Use with `HostedZoneID` if you are not using SSL.
Route 53 hosted zone ID (`HostedZoneID`)	*Optional*	Route 53 Hosted Zone ID of the domain name. Used in conjunction with `LoadBalancerFQDN`.
SSL certificate ARN (`SSLCertificateArn`)	*Optional*	The Amazon Resource Name (ARN) of the SSL certificate to use for the load balancer. Use SSLCertificateArn if you are NOT using `LoadBalancerFQDN` and `HostedZoneID`.

AWS Quick Start configuration:

Parameter label (name)	Default	Description
Quick Start S3 Bucket Name (`QSS3BucketName`)	aws-quickstart	S3 bucket name for the Quick Start assets. This bucket name can include numbers, lowercase letters, uppercase letters, and hyphens (-), but should not start or end with a hyphen. You can specify your own bucket if you copy all of the assets and submodules into it, if you want to override the Quick Start behavior for your specific implementation.
Quick Start S3 Key Prefix (`QSS3KeyPrefix`)	quickstart-hashicorp-consul/	S3 key prefix for the Quick Start assets. This prefix can include numbers, lowercase letters, uppercase letters, hyphens (-), and forward slashes (/), but should not start or end with a forward slash (which is automatically added). This parameter enables you to override the Quick Start behavior for your specific implementation.

- **Parameters for deployment into an existing VPC**

  View template

VPC Network configuration:

Parameter label (name)	Default	Description
VPC ID (`VPCID`)	*Requires input*	List of Availability Zones to use for the subnets in the VPC. Note: The logical order is preserved. (Three Availability Zones are used for this deployment.)
VPC CIDR (`VPCCIDR`)	*Requires input*	CIDR block for the VPC.
Private Subnet 1 ID (`PrivateSubnet1ID`)	*Requires input*	ID of private subnet 1 located in Availability Zone 1.
Private Subnet 2 ID (`PrivateSubnet2ID`)	*Requires input*	ID of private subnet 2 located in Availability Zone 2.
Private Subnet 3 ID (`PrivateSubnet3ID`)	*Requires input*	ID of private subnet 3 located in Availability Zone 3.
Public Subnet 1 ID (`PublicSubnet1ID`)	*Requires input*	ID of the public (DMZ) subnet 1 located in Availability Zone 1.
Public Subnet 2 ID (`PublicSubnet2ID`)	*Requires input*	ID of the public (DMZ) subnet 2 located in Availability Zone 2.
Public Subnet 3 ID (`PublicSubnet3ID`)	*Requires input*	ID of the public (DMZ) subnet 3 located in Availability Zone 3.

Bastion host configuration:

Parameter label (name(	Default	Description
Bastion host security group ID (`BastionSecurityGroupID`)	*Requires input*	ID of the bastion host security group to enable SSH connections (e.g., sg-7f16e910).

Consul setup:

Parameter label (name)	Default	Description
Consul cluster node instance type (`ConsulInstanceType`)	m5.large	The EC2 instance type for the Consul instance.
Number of Consul server nodes (`ConsulServerNodes`)	3	The number of Consul server nodes that will be created. You can choose 3, 5, or 7 nodes.

Parameter label (name)	Default	Description
Number of Consul client nodes (`ConsulClientNodes`)	0	The number of Consul client nodes that will be created.
Tag key for Consul cluster nodes (`ConsulEc2RetryTagKey`)	quickstart-consul-cluster	The EC2 instance tag key to filter on when joining to other Consul nodes.
Tag value for Consul cluster nodes (`ConsulEc2RetryTagValue`)	consul-member-node	The EC2 instance tag value to filter on when joining to other Consul nodes.
Key name (`KeyPairName`)	*Requires input*	Public/private key pair, which allows you to connect securely to your instance after it launches. When you created an AWS account, this is the key pair you created in your preferred region.

DNS and SSL configuration:

Parameter label (name(	Default	Description
Load Balancer FQDN (`LoadBalancerFQDN`)	*Optional*	The fully qualified domain name for the load balancer. Use with `HostedZoneID` if you are not using SSL.
Route 53 hosted zone ID (`HostedZoneID`)	*Optional*	Route 53 Hosted Zone ID of the domain name. Used in conjunction with `LoadBalancerFQDN`.
SSL certificate ARN (`SSLCertificateArn`)	*Optional*	The Amazon Resource Name (ARN) of the SSL certificate to use for the load balancer. Use SSLCertificateArn if you are NOT using `LoadBalancerFQDN` and `HostedZoneID`.

AWS Quick Start configuration:

Parameter label (name)	Default	Description
Quick Start S3 Bucket Name (`QSS3BucketName`)	aws-quickstart	S3 bucket name for the Quick Start assets. This bucket name can include numbers, lowercase letters, uppercase letters, and hyphens (-), but should not start or end with a hyphen. You can specify your own bucket if you copy all of the assets and submodules into it, if you want to override the Quick Start behavior for your specific implementation.
Quick Start S3 Key Prefix (`QSS3KeyPrefix`)	quickstart-hashicorp-consul/	S3 key prefix for the Quick Start assets. This prefix can include numbers, lowercase letters, uppercase letters, hyphens (-), and forward slashes (/), but should not start or end with a forward slash (which is automatically

Parameter label (name)	Default	Description
		added). This parameter enables you to override the Quick Start behavior for your specific implementation.

When you finish reviewing and customizing the parameters, choose **Next**.

4. On the **Options** page, you can specify tags (key-value pairs) for resources in your stack and set advanced options. When you're done, choose **Next**.
5. On the **Review** page, review and confirm the template settings. Under **Capabilities**, select the check box to acknowledge that the template will create IAM resources.
6. Choose **Create** to deploy the stack.
7. Monitor the status of the stack. When the status is **CREATE_COMPLETE**, the deployment is complete.
8. You can use the URL displayed in the **Outputs** tab for the stack to view the resources that were created.

Step 3. Access Consul via SSH

To access the Consul environment, first connect to one of the bastion hosts. Use an SSH agent to forward your private key on connection.

> **Important** Do not copy your private key to the bastion host.

For more information on SSH agents, see the GitHub documentation.

To use an SSH agent to access the Consul environment on Mac or Linux:

1. Use the command:

```
ssh-add ~/.ssh/id_rsa
```

2. At the prompt, type your passphrase or press **Enter** for no passphrase.

```
Enter passphrase (empty for no passphrase): [Hit Enter Again or
Enter passphrase]
Enter same passphrase again: [Hit Enter Again or Enter passphrase]
```

3. In the Amazon EC2 console, select one of the two bastion hosts and note its Elastic IP address.

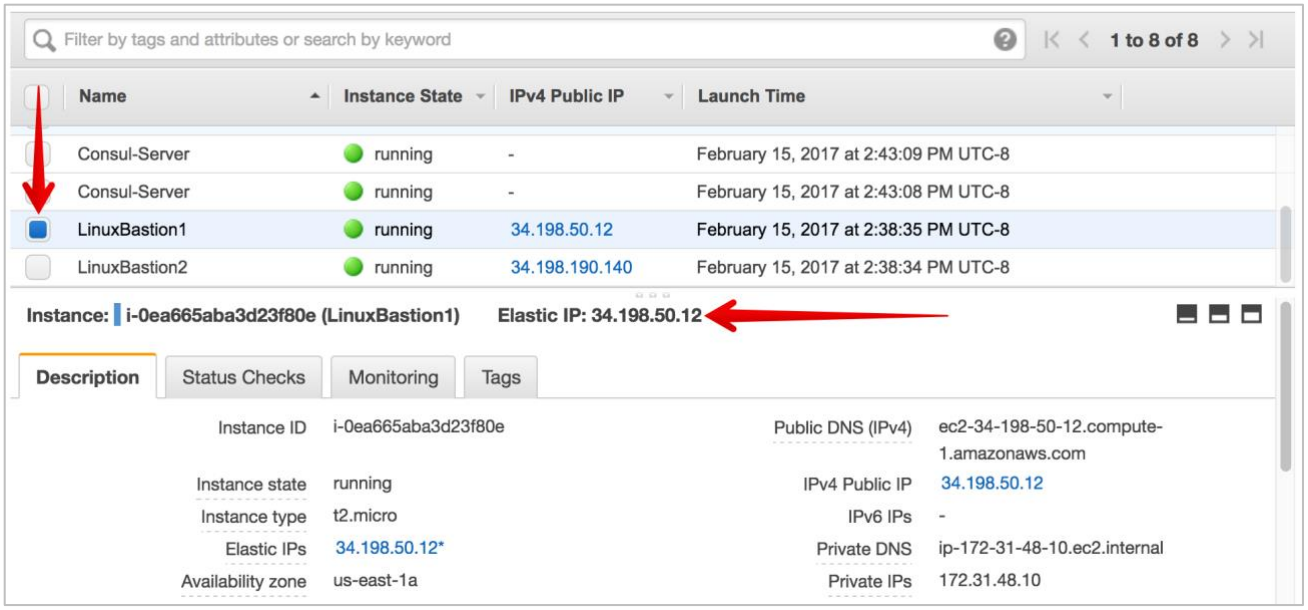
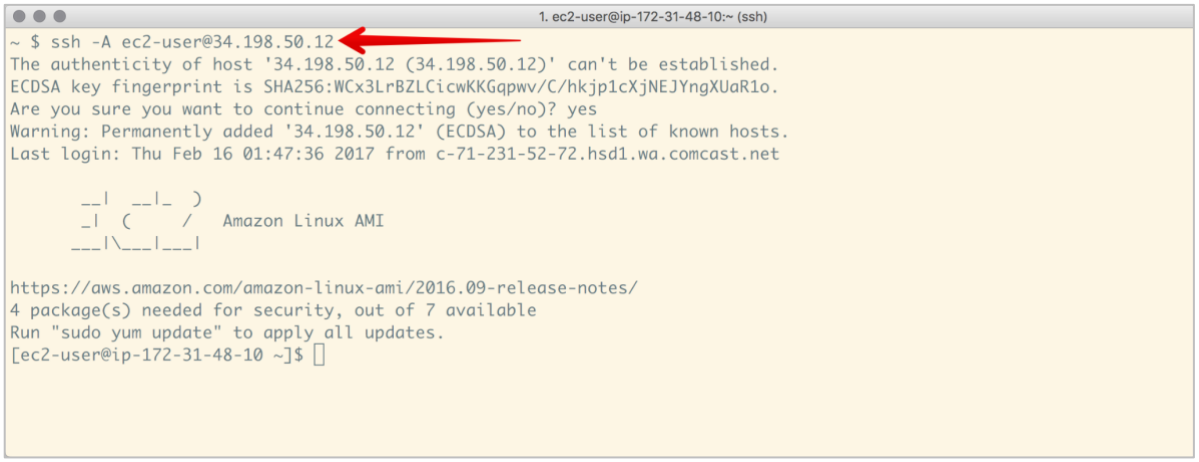


Figure 2: Finding the Elastic IP address for the bastion host instance

In the example in Figure 2, the Elastic IP address for `LinuxBastion1` is **34.198.50.12.**

4. Log in, and type **yes** when prompted to continue connecting:

```
ssh -A ubuntu@34.198.50.12
```



5. In the Amazon EC2 console, select one of the `Consul-Server` or `Consul-Client` hosts and note its private IP address.

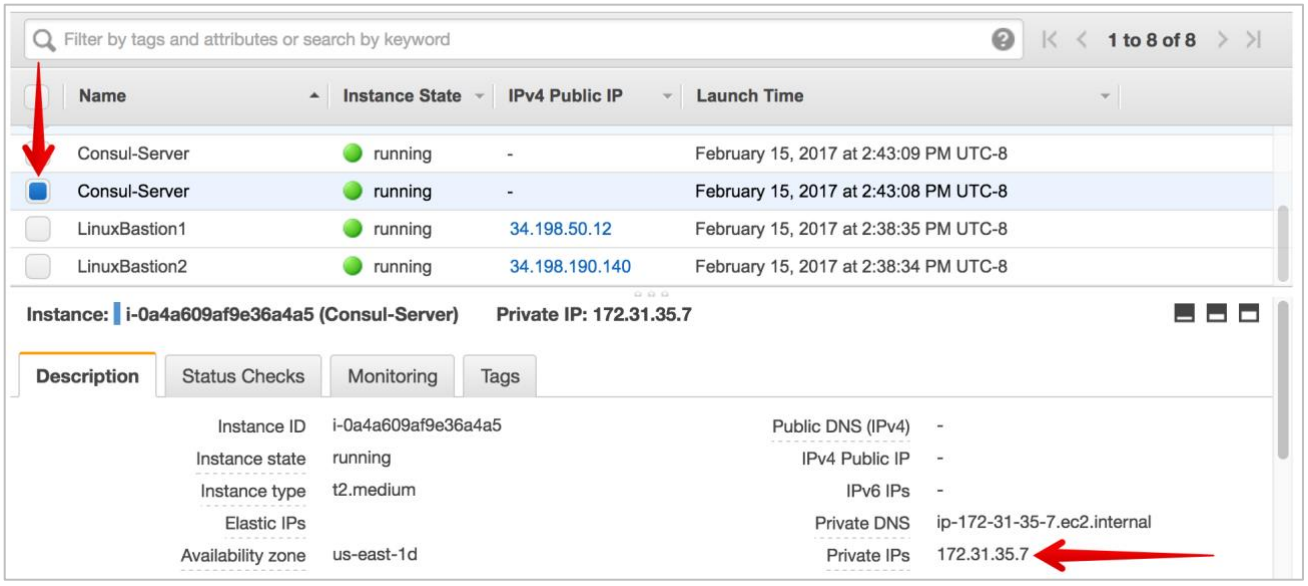


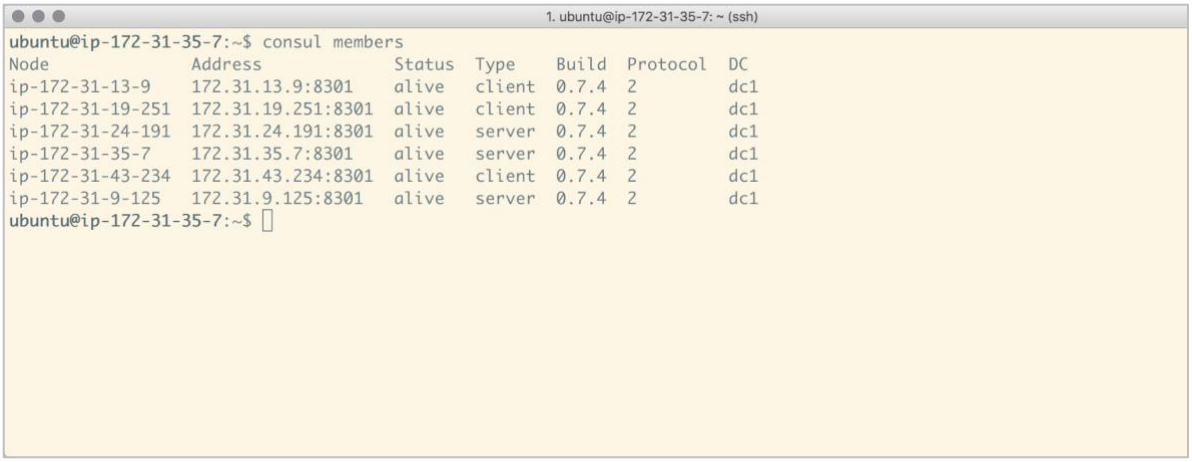
Figure 3: Finding the private IP address for Consul-Server

In the example in Figure 2, the private IP for `Consul-Server` is **172.31.35.7**.

6. From the bastion host, connect to the `Consul-Server` or `Consul-Client` host, using Ubuntu as the user:

7. View Consul members:

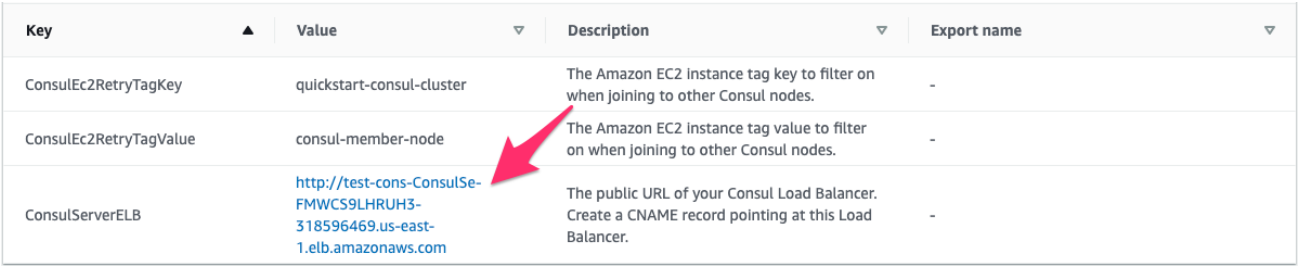
```
consul members
```



Step 4. Test the deployment

To access the Consul server cluster environment, access the Elastic Load Balancing (ELB) endpoint that was created during the deployment.

1. Locate the ELB endpoint address from the **Outputs** tab of the AWS CloudFormation console.



Key	Value	Description	Export name
ConsulEc2RetryTagKey	quickstart-consul-cluster	The Amazon EC2 instance tag key to filter on when joining to other Consul nodes.	-
ConsulEc2RetryTagValue	consul-member-node	The Amazon EC2 instance tag value to filter on when joining to other Consul nodes.	-
ConsulServerELB	http://test-cons-ConsulSe-FMWCS9LHRUH3-318596469.us-east-1.elb.amazonaws.com	The public URL of your Consul Load Balancer. Create a CNAME record pointing at this Load Balancer.	-

2. Use your preferred web browser to open the URL. You will see the Consul server cluster dashboard.

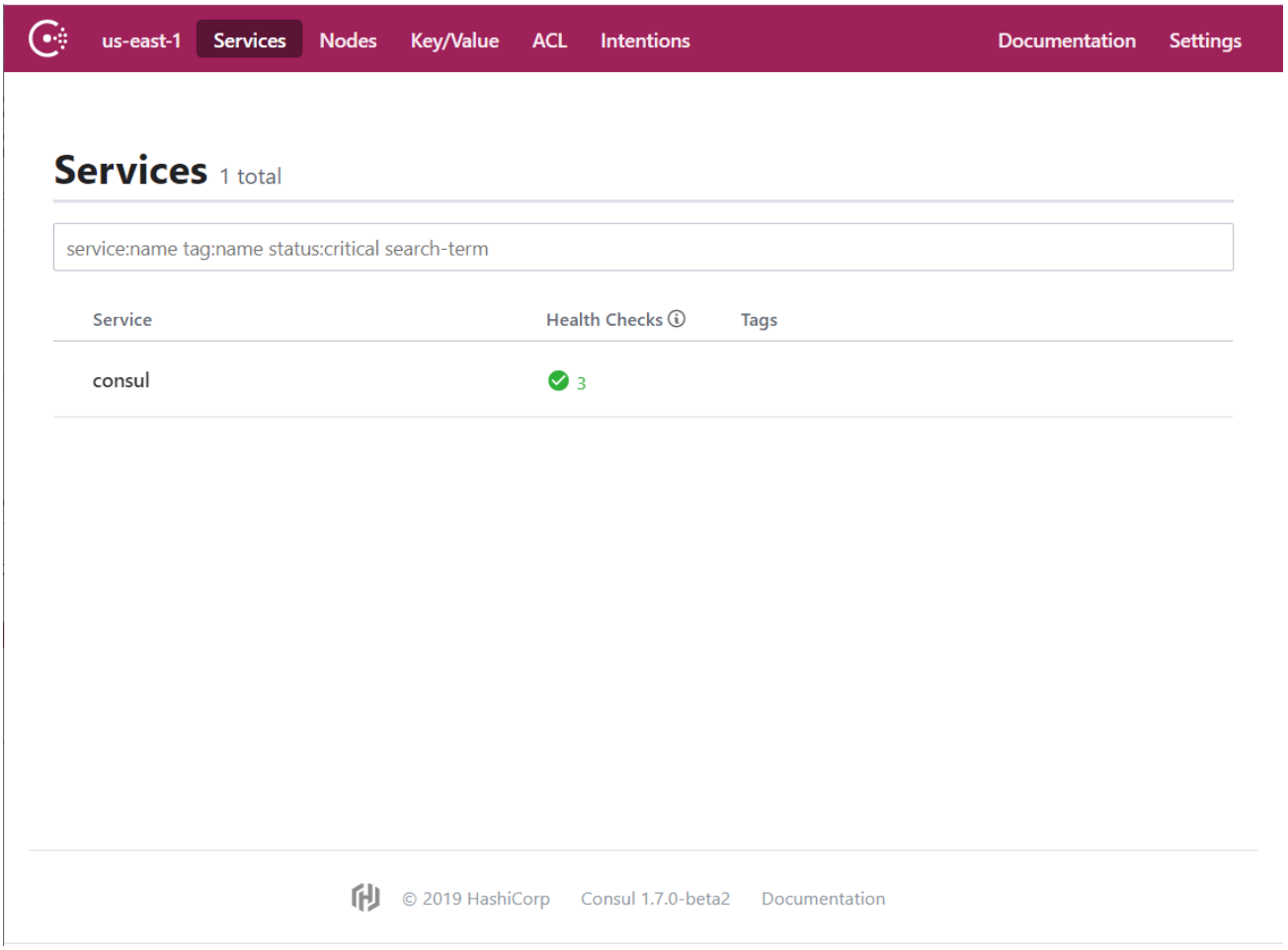


Figure 4: Consul web UI

Step 5. Get started with Consul

To integrate Consul with your environment and get started with Consul services, see the Getting Started section of the HashiCorp Consul website.

Consul Connect and Autopilot are enabled by default.

How to set up a service with Consul Connect - service mesh

Consul Connect is enabled by default. To set up a service on the Consul client nodes, you will need to register the service and proxy with Consul. For more information, please visit the following HashiCorp Learn pages:

- Register the Service and Proxy with Consul
- Register a Dependent Service and Proxy

- Control Communication with Intentions

How to manage Consul Autopilot

Consul Autopilot is enabled by default with the following settings:

```
"autopilot": {
  "cleanup_dead_servers": true,
  "last_contact_threshold": "200ms",
  "max_trailing_logs": 250,
  "server_stabilization_time": "10s",
  "redundancy_zone_tag": "az",
  "disable_upgrade_migration": false,
  "upgrade_version_tag": ""
}
```

Troubleshooting

Q. I encountered a CREATE_FAILED error when I launched the Quick Start. What should I do?

A. If AWS CloudFormation fails to create the stack, we recommend that you relaunch the template with **Rollback on failure** set to **No**. (This setting is under **Advanced** in the AWS CloudFormation console, **Options** page.) With this setting, the stack's state will be retained and the instance will be left running, so you can troubleshoot the issue. (You'll want to look at the log files in `%ProgramFiles%\Amazon\EC2ConfigService` and `C:\cfn\log`.)

> **Important** When you set **Rollback on failure** to **No**, you'll continue to incur AWS charges for this stack. Please make sure to delete the stack when you've finished troubleshooting.

For additional information, see Troubleshooting AWS CloudFormation on the AWS website or contact us on the AWS Quick Start Discussion Forum.

Q. I encountered a size limitation error when I deployed the AWS CloudFormation templates.

A. We recommend that you launch the Quick Start templates from the location we've provided or from another S3 bucket. If you deploy the templates from a local copy on your computer or from a non-S3 location, you might encounter template size limitations when you create the stack. For more information about AWS CloudFormation limits, see the AWS documentation.

Additional resources

AWS services

- Amazon EC2
  http://docs.aws.amazon.com/AWSEC2/latest/WindowsGuide/
- Amazon VPC
  http://aws.amazon.com/documentation/vpc/

HashiCorp Consul

- Consul
  https://www.consul.io
- Consul Enterprise
  https://www.hashicorp.com/consul.html

Quick Start reference deployments

- AWS Quick Start home page
  https://aws.amazon.com/quickstart/
- AWS Quick Start for HashiCorp Vault
  https://s3.amazonaws.com/quickstart-reference/hashicorp/vault/latest/doc/hashicorp-vault-on-the-aws-cloud.pdf

Send us feedback

To post feedback, submit feature ideas, or report bugs, use the **Issues** section of the GitHub repository for this Quick Start. If you'd like to submit code, please review the Quick Start Contributor's Guide.

Document revisions

Date	Change	In sections
January 2020	Updated use cases; added an Auto Scaling group for the bastion hosts; added a Classic Load Balancer with ACM support	Changes in templates and throughout guide
April 2017	Added Linux bastion hosts; updated Consul to version 0.8.0; removed Seed server; added Amazon EC2 retry functionality	Changes in templates and throughout guide
November 2016	Initial publication	—

Notices

This document is provided for informational purposes only. It represents AWS's current product offerings and practices as of the date of issue of this document, which are subject to change without notice. Customers are responsible for making their own independent assessment of the information in this document and any use of AWS's products or services, each of which is provided "as is" without warranty of any kind, whether express or implied. This document does not create any warranties, representations, contractual commitments, conditions or assurances from AWS, its affiliates, suppliers or licensors. The responsibilities and liabilities of AWS to its customers are controlled by AWS agreements, and this document is not part of, nor does it modify, any agreement between AWS and its customers.

The software included with this paper is licensed under the Apache License, Version 2.0 (the "License"). You may not use this file except in compliance with the License. A copy of the License is located at http://aws.amazon.com/apache2.0/ or in the "license" file accompanying this file. This code is distributed on an "AS IS" BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.