Hewlett Packard Enterprise

HPE Nimble Storage Deployment Guide for Amazon Web Services

Contents

Notices

The information contained herein is subject to change without notice. The only warranties for Hewlett Packard Enterprise products and services are set forth in the express warranty statements accompanying such products and services. Nothing herein should be construed as constituting an additional warranty. Hewlett Packard Enterprise shall not be liable for technical or editorial errors or omissions contained herein.

Confidential computer software. Valid license from Hewlett Packard Enterprise required for possession, use, or copying. Consistent with FAR 12.211 and 12.212, Commercial Computer Software, Computer Software Documentation, and Technical Data for Commercial Items are licensed to the U.S. Government under vendor's standard commercial license.

Links to third-party websites take you outside the Hewlett Packard Enterprise website. Hewlett Packard Enterprise has no control over and is not responsible for information outside the Hewlett Packard Enterprise website.

Acknowledgments

Intel®, Itanium®, Pentium®, Intel Inside®, and the Intel Inside logo are trademarks of Intel Corporation in the United States and other countries.

Microsoft® and Windows® are either registered trademarks or trademarks of Microsoft Corporation in the United States and/or other countries.

Adobe® and Acrobat® are trademarks of Adobe Systems Incorporated. Java® and Oracle® are registered trademarks of Oracle and/or its affiliates.

UNIX® is a registered trademark of The Open Group.

Publication Date

Thursday April 26, 2018 13:55:06

Document ID

gse1470689224282

Support

All documentation and knowledge base articles are available on HPE InfoSight at *https://infosight.hpe.com*.
To register for HPE InfoSight, click the *Create Account* link on the main page.

Email: *support@nimblestorage.com*

For all other general support contact information, go to *https://www.nimblestorage.com/customer-support/*.

Overview

The topic of cloud can be a complex discussion. There are widespread misconceptions and assumptions about what the cloud is, what it can and cannot do, and how best to adopt it. This paper provides an overview of key cloud concepts, a brief introduction to Amazon Web Services (AWS), and specific guidance for deploying HPE Nimble Storage predictive flash arrays in a hybrid cloud architecture that is connected to AWS.

In addition to discussing specific types of clouds, this document includes a generic high-level description of the concept of cloud. For many years now, shared infrastructure has been popular, and it is a well-known architecture for IT. The arrival of mainstream virtualization enabled IT departments to realize the ability to share integrated resources that traditionally had been overprovisioned and underutilized. However, the problem with shared infrastructure was that resources were still manually provisioned. Cloud, at its most fundamental level, is the application of a highly automated provisioning and management paradigm to a shared infrastructure.

Cloud Characteristics

What makes a cloud, a cloud? In several ways, the popular joke that "a cloud is just someone else's computer" is very much accurate. Realistically, however, it is a bit more complex than that.

The National Institute of Standards and Technology (NIST) has done a great job of defining several aspects of cloud (characteristics, deployment models, and service models). *NIST special publication 800-145* identifies the essential cloud characteristics as "on-demand self-service, broad network access, resource pooling, rapid elasticity, and measured service."

On-Demand Self-Service

On-demand self-service is the ability to automatically provision resources in the cloud without the need for human interaction. This aspect is perhaps the most differentiating characteristic for cloud as compared to a traditional shared infrastructure architecture.

Broad Network Access

Broad network access describes the ability to access cloud resources from anywhere over standard network connections, whether private or public. Most companies access public cloud resources over a secure connection. For example, AWS Virtual Private Cloud (VPC) is used to provide private cloud resources in a public cloud architecture.

Resource Pooling

Resource pooling is the traditional concept of shared infrastructure and secure multi-tenancy. This arrangement allows cloud service providers to deploy large amounts of physical resources that can be partitioned logically and securely across multiple customers. The ability to get every bit of usage out of physical resources is one of the keys to providing low-cost cloud services.

Rapid Elasticity

Rapid elasticity describes the ability to provision and release resources immediately, without limitations based on the number of physical resources available. Alternatively, this could be described as rapid scaling of the logical and physical infrastructure. Physically, it means deploying enough hardware to meet, or more realistically to exceed, expected customer demand.

Measured Service

Measured service is the ability to easily manage and monitor cloud resources. Manageability and simplicity are qualities that have become increasingly popular in recent years. Clearly, they have always been important, but they are even more so today. Their resurgence in popularity is due not only to millennials' general expectations of immediate delivery, but also to the pace of growth in demand for data, compute resources, bandwidth, and more.

In addition, customers must be able to plan and predict costs associated with operating workloads in the cloud. An often overlooked aspect of a measured service is that "you get what you pay for." If you pay for X IOPS, you get X IOPS, no more and no less.

Cloud Types and Service Models

Four types of clouds exist today:

- Public
- Private
- Hybrid
- Community

In addition, there are three common service models:

- Software as a service (SaaS)
- Platform as a service (PaaS)
- Infrastructure as a service (IaaS)

The National Institute of Standards and Technology (NIST) provides concise definitions for each cloud type, known as deployment models, and cloud service models.

Deployment Models

NIST special publication 800-145 contains the following definitions for cloud deployment models:

- **Public Cloud:** "*The cloud infrastructure is provisioned for open use by the general public. It may be owned, managed, and operated by a business, academic, or government organization, or some combination of them. It exists on the premises of the cloud provider.*"
- **Private Cloud:** "*The cloud infrastructure is provisioned for exclusive use by a single organization comprising multiple consumers (e.g., business units). It may be owned, managed, and operated by the organization, a third party, or some combination of them, and it may exist on or off premises.*"
- **Hybrid Cloud:** "*The cloud infrastructure is a composition of two or more distinct cloud infrastructures (private, community, or public) that remain unique entities, but are bound together by standardized or proprietary technology that enables data and application portability (e.g., cloud bursting for load balancing between clouds).*"
- **Community Cloud:** "*The cloud infrastructure is provisioned for exclusive use by a specific community of consumers from organizations that have shared concerns (e.g., mission, security requirements, policy, and compliance considerations). It may be owned, managed, and operated by one or more of the organizations in the community, a third party, or some combination of them, and it may exist on or off premises.*"

Service Models

NIST special publication 800-145 contains the following definitions for cloud service models:

- **Software as a Service:** "*The capability provided to the consumer is to use the provider's applications running on a cloud infrastructure. The applications are accessible from various client devices through either a thin client interface, such as a web browser (e.g., web-based email), or a program interface. The consumer does not manage or control the underlying cloud infrastructure including network, servers, operating systems, storage, or even individual application capabilities, with the possible exception of limited user specific application configuration settings.*"
- **Platform as a Service:** "*The capability provided to the consumer is to deploy onto the cloud infrastructure consumer-created or acquired applications created using programming languages, libraries, services, and tools supported by the provider. The consumer does not manage or control the underlying cloud infrastructure including network, servers, operating systems, or storage, but has control over the deployed applications and possibly configuration settings for the application-hosting environment.*"

- **Infrastructure as a Service:** "*The capability provided to the consumer is to provision processing, storage, networks, and other fundamental computing resources where the consumer is able to deploy and run arbitrary software, which can include operating systems and applications. The consumer does not manage or control the underlying cloud infrastructure but has control over operating systems, storage, and deployed applications; and possibly limited control of select networking components (e.g., host firewalls).*"

Storage as a Service

The terminology in common use is not limited to what is defined by NIST. Storage as a service (STaaS) is another service model that is of particular interest to Hewlett Packard Enterprise (HPE) and other storage vendors. HPE currently offers storage on demand (SoD), which is a pay-as-you-go service for customers' storage needs. SoD customers pay only for storage or application use versus the amount of capacity that has been provisioned (unused space). By its very nature, SoD aligns well with the resource pooling characteristic of cloud.

STaaS is a cloud service model that can easily extend existing SoD offerings into a cloud service. If storage arrays are deployed into a collocation data center to gain broad network access, then an existing SoD service can be extended to STaaS by addressing on-demand self-service (building a portal to provision storage), rapid elasticity (deploying enough storage to meet demand), and measurable service (enabling the storage to be easily managed and monitored).

Amazon Web Services

AWS has the largest market share of the public cloud, and it offers an ever-changing and staggering number of services. The following figure provides a high-level overview of the services that are accessible through the AWS console. Even beyond the initial list, however, each service normally contains several subcomponents. For example, EC2 is displayed under the Compute section. When you drill down to the next level of EC2, several subcomponents are displayed, such as instances, volumes, key pairs, elastic IPs, snapshots, security groups, load balancers, placement groups, and dedicated hosts. It is easy to become overwhelmed when trying to take it all in at once.

Fortunately, for the purposes of this document and for connecting an HPE Nimble Storage array to AWS, it is necessary to understand only a small subset of AWS services. The *AWS website* contains more information about the wide range of cloud services available from AWS.

Figure 1: AWS services overview

Compute Services

There are two key compute services:

- *Elastic Compute Cloud (EC2)*
- *Elastic Block Storage (EBS)*

EBS is included because it is used strictly with EC2 compute instances and is not provisioned as a separate storage service as, for example, the Amazon Simple Storage Service (S3) is.

Elastic Compute Cloud

Amazon EC2 is the scalable compute cloud service. The core product of the EC2 service is a compute instance. Many different kinds of compute instances can be run, from general purpose instances (T2, M4, and M3), compute optimized (C4 and C3), memory optimized (X1 and R3), GPU (G2), to storage optimized (I2 for high I/O instance and D2 for dense storage instances). In most cases, within the different kinds of instances, several different instance types are available.

Instance types offer a range of options in number of CPUs, amount of memory, type and amount of storage (EBS only or SSD), and networking performance. For more information about instance types, see *Amazon EC2 Instance Types*.

Elastic Block Storage

EBS volumes are block storage objects attached to EC2 instances that can be used just like any regular block storage volume or LUN. EBS volumes are put in a specific availability zone and replicated automatically for data protection. EBS volumes can be backed by either HDD or SSD, depending on your storage performance requirements.

Initially, EBS was its own separate entity in the AWS console, but that is no longer the case. EBS volumes are automatically procured during the process of deploying a new EC2 instance. Depending on the EC2 instance type you are deploying, EBS HDD (specified as "EBS only") might be the only option available, or you might have the choice between using HDD or SSD.

EBS volumes have specific drawbacks:

- Although EBS volumes are replicated, they are replicated within the localized availability zones (for example, `us-west-1a` to `us-west-1b`).
- Depending on the EC2 instance you choose, you are restricted to a specific type of storage (HDD or SSD).
- SSD storage in EBS is limited in capacity (nowhere near the amount of SSD capacity that an HPE Nimble Storage array can provide).
- Customers pay for all provisioned EBS storage, versus paying only for the capacity that they use.

Networking Services

There are two key networking services:

- *Direct Connect*
- *Virtual Private Cloud (VPC)*

Both Direct Connect and VPC are used when connecting on-premises storage arrays to cloud-based compute resources.

Direct Connect

AWS Direct Connect (DX) allows a dedicated network connection to be established between your on-premises data center and AWS. Using a dedicated network connection to AWS provides more consistent network connectivity, increased network security, and increased bandwidth.

Before requesting a direct connection from AWS, it is important to understand what localized direct connect location is nearest your data center. Direct connections can be public, private, or both. IPsec is used to provide secure tunnels between your on-premises network and the Direct Connect location. Multiple virtual interfaces can be created to partition the connectivity.

EC2 instances that are running within a VPC can be accessed over a direct connection. For more information on Amazon VPC, see the section *Virtual Private Cloud* on page 11.

Cost for AWS Direct Connect breaks down into four core items:

- ISP connectivity costs
- Billable port hours
- Data transfer
- AWS specific resource costs

ISP costs are related to bandwidth and setup charges. Billable port hours are measured by AWS and depend on how many ports on their network are being used (partial hours are billed as full hours). Data transfer is also billed. Data transfer into the cloud is no cost, whereas data transfer out of the cloud is billed (data into the cloud is free, data out is charged). AWS specific resources are related to the specific services you are accessing, for example an EC2 instance or S3 object store.

For more information about requirements and limitations for AWS Direct Connect, see the latest *User Guide for AWS Direct Connect*.

Virtual Private Cloud

A VPC allows the creation of a private network within AWS. Users have control over the IP address range used, gateways, subnets, and routing tables. When combined with Direct Connect, a VPN connection can be deployed to connect the VPC to an on-premises data center network. VPN connections can be further secured by being encrypted if that is desired.

A VPC can be peered to another VPC in AWS to bridge the networks together. Elastic IP addresses (public IPs) can be attached to a network interface to make an instance publicly accessible over the internet. Although it is possible to attach public IPs to a specific instance, this is a security risk as it relates to VPC, which is meant to provide a private network in AWS. Exercise caution when assigning public IP addresses to your AWS resources.

A great native feature of VPC is network flow logs, which log inbound and outbound network traffic on the VPC. If network connectivity issues are encountered, the VPC flow logs are helpful in understanding whether traffic is making it into or out of the VPC.

Deploying EC2 Direct Connected Storage

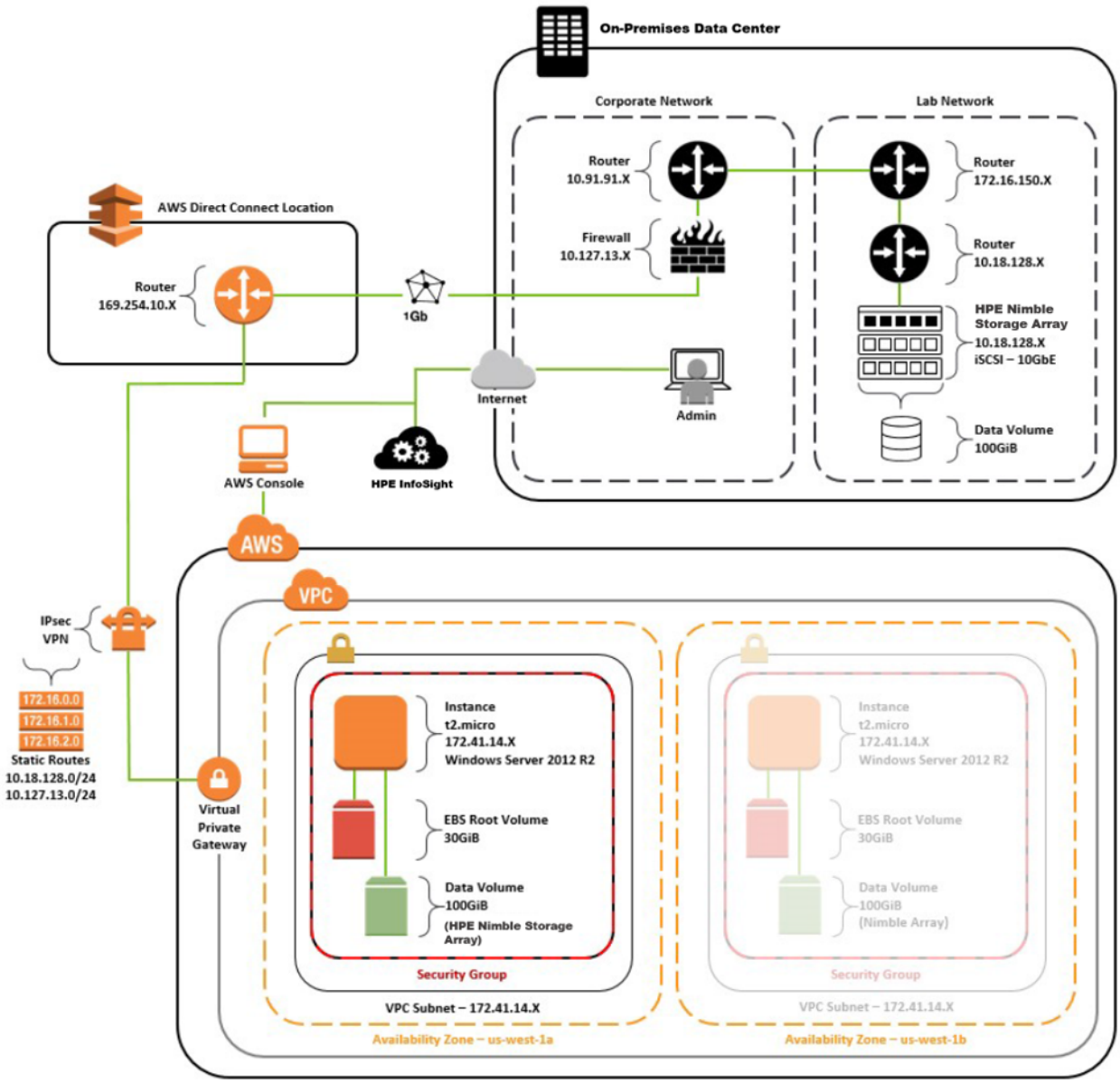
HPE Nimble Storage cloud-connected storage for AWS enables customers to leverage on-premises storage arrays into the cloud, which provides several key benefits:

- Storage can stay on-premises to provide the best performance and latency for critical workloads (not in the cloud), while allowing access to on-premises data by cloud compute resources.
- Cloud compute resources have access to data without having to first migrate the data into the cloud.
- HPE InfoSight provides visibility into data and predictive analytics.
- Remote offices or users can access cloud compute resources that access on-premises data directly.
- Volumes can be moved quickly and nondisruptively between different tiers of storage (for example, from hybrid to all flash or vice versa).
- You gain enterprise-class storage features such as deduplication and compression for your data sets while still leveraging the benefits of cloud compute infrastructure.

Architecture Overview

The following diagram provides an overview of the solution architecture.

Figure 2: AWS EC2 Direct Connect storage architecture overview



The on-premises data center in this architecture uses multiple layers of switching and routing to demonstrate more complex networking. What is essential to the architecture is that the storage array discovery IP address and data interfaces are routable to the AWS direct connection.

Although they are not explicitly shown in this diagram, two IPsec tunnels to the AWS Virtual Private Gateway (VGW) provide fault tolerance and failover capabilities (active-passive). It is possible to configure the IPsec VPN to use an active-active configuration.

A 1 Gbps connection was provisioned for this reference architecture. Although sub-1 Gbps connections can be provisioned (through AWS partners), two speeds are available when a direct connect is initiated from the AWS console: either 1 Gbps or 10 Gbps. One or more connections can be provisioned at whatever bandwidth rates are necessary to support the data access requirements.

Reference this architecture while planning and designing connectivity between on-premises HPE Nimble Storage arrays and AWS compute resources through AWS DX. Each customer's requirements and network

connectivity are expected to vary slightly, but the basic architectural foundation shown in this document will be common. Nevertheless, AWS DX supports different connectivity designs.

Private AWS DX Connectivity Options

There are three common private AWS DX connectivity options to consider: static VPN, dynamic VPN, and resilient dynamic VPN. This section focuses on the specific components present between the AWS VGW (the entry point into the VPC) and the customer's on-premises or collocated networks when using one of the connectivity options.

Establishing VPN connectivity to AWS requires using Border Gateway Protocol (BGP) if the VPN connection will be dynamically routed. BGP allows routing information (prefixes) to be exchanged between neighbors on a network. The following key BGP-related terms and definitions are related to AWS connectivity:

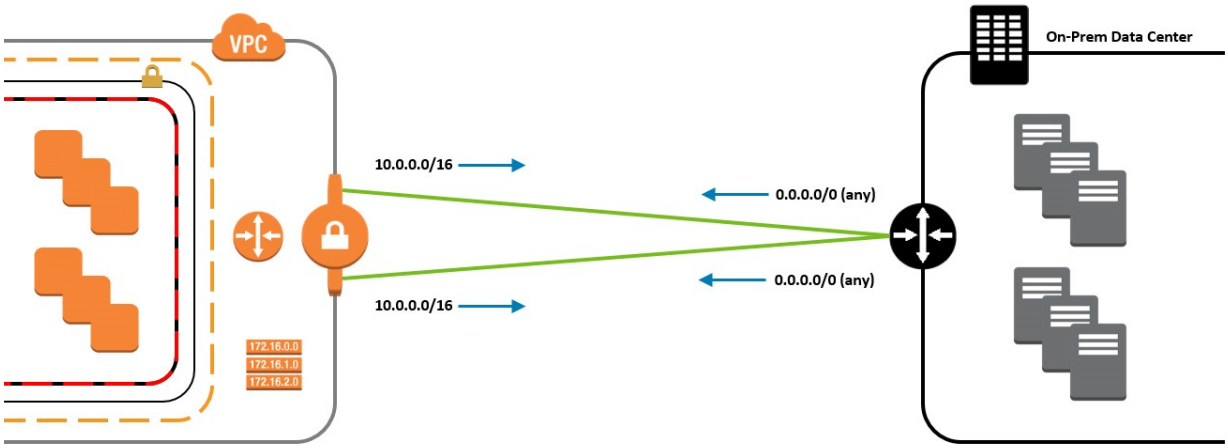
- **Autonomous systems (AS) number.** Allows a network to identify itself to BGP neighbors on a network. AWS has a single AS number (7224) that is used for the entire AWS network.
- **AS path.** The measure of the distance of a network. Distance is measured by appending the AS number to traffic as it passes through network hops.
- **Prefixes.** Consists of a Classless Inter-Domain Routing (CIDR) block and an AS path.
- **iBGP.** The protocol used when a peering relationship is established between two neighbors in the same AS number.
- **eBGP.** The protocol used when a peering relationship is established between two neighbors in different AS numbers.
- **Local preference**. Allows specific connections over BGP to be weighed by artificially appending AS numbers to the `AS_PATH`. This is useful if, for example, a higher bandwidth path with a longer `AS_PATH` should be used over a lower bandwidth path that has a shorter `AS_PATH`.

Static VPN

In the static VPN model, two IPsec tunnels are used to connect to two end points in AWS, one for each availability zone (for example, `us-west-1a` and `us-west-1b`). AWS supports only two security associations (SAs) per tunnel, one inbound and one outbound, for a total of four SAs. Some routers use a different SA for each network being routed, but that arrangement does not work for AWS.

Access control lists (ACLs) must be consolidated to cover all traffic, from all networks, that is routed over the AWS DX connection. Specific protocols can be filtered by using security groups within the VPC configuration. VPN routing uses static routes that are configured from the AWS console. The following diagram provides an overview of a static VPN configuration.

Figure 3: Example of AWS Direct Connect through a static VPN

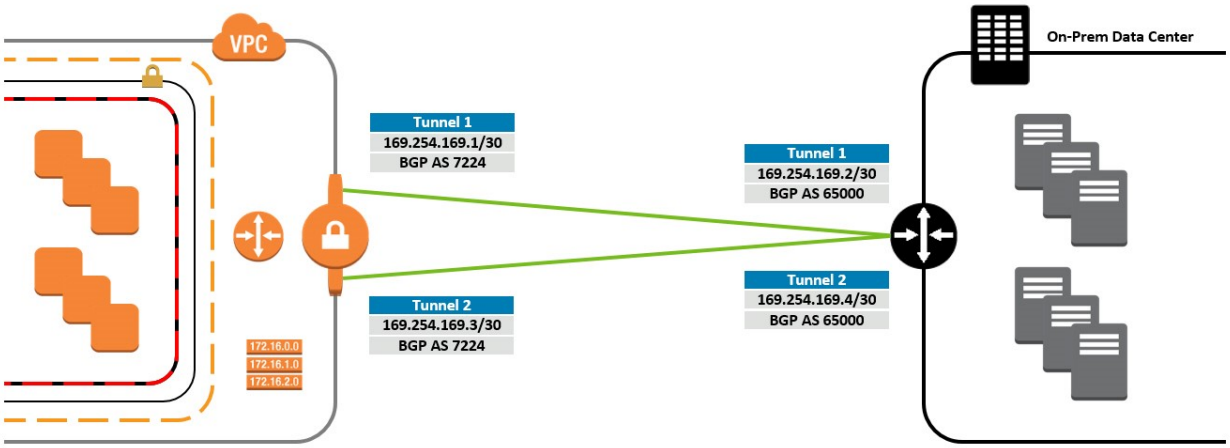


Dynamic VPN

Similar to the static VPN connectivity option, two IPsec tunnels are used to connect to two end points in AWS. IP addresses must be specified at each end of the tunnels; AWS automatically generates them during the configuration process. As routes are announced over BGP to AWS (to the VGW), the route table in the VPC must be updated. It can be updated manually, or if route propagation is enabled, the route table is automatically updated by AWS.

AWS uses a single AS number for the entire AWS network. The AWS AS number is generally 7224, but it can vary by region. (The number is shown in the downloaded configuration from AWS.) The customer configuration can use any valid AS number, either a public AS number that is already in use, or an AS number from the reserved block of AS numbers (for example, 65000). The following diagram shows an example of a dynamic VPN configuration.

Figure 4: Example of AWS Direct Connect through a dynamic VPN

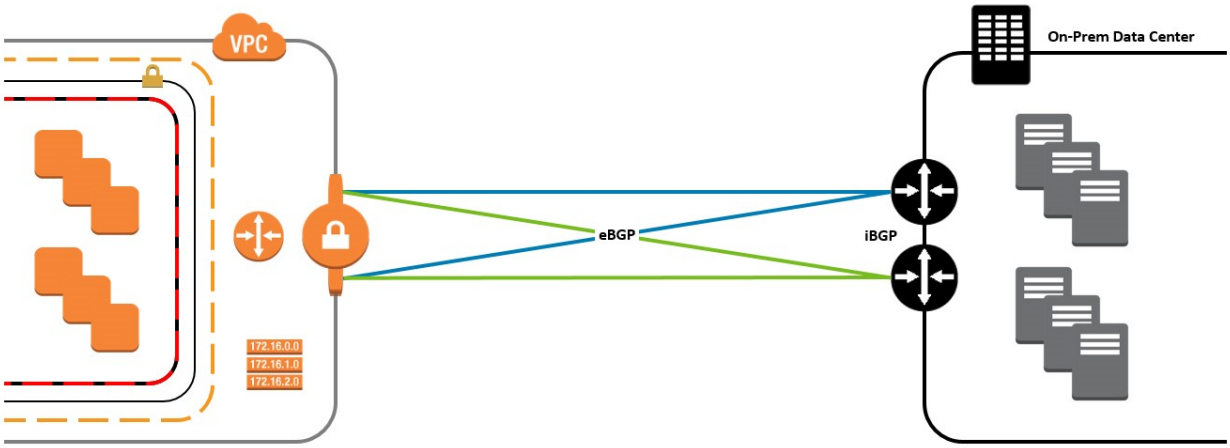


Resilient Dynamic VPN

The resilient dynamic VPN connectivity option builds on the dynamic VPN option. Instead of a single VPN connection, two VPN connections are used (with two customer gateway devices). Each VPN terminates in a separate availability zone in the VPC. Each customer router establishes two IPsec tunnels each, for a total of four IPsec tunnels going to AWS.

With two gateway devices, an internal routing protocol (such as iBGP or OSPF) must be used to determine how internal traffic should be routed to AWS. The most obvious benefit of using this connectivity option is the elimination of the single point of failure (SPOF) on the customer side. The following diagram shows an overview of a resilient dynamic VPN configuration.

Figure 5: Example of AWS Direct Connect through a resilient dynamic VPN



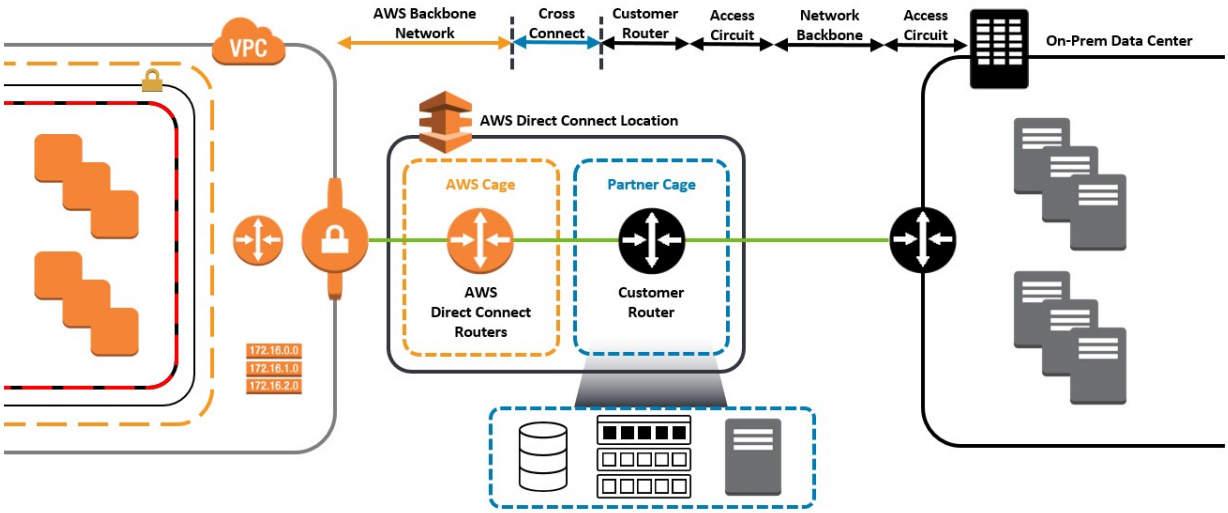
On-Premises and Collocation

Any connectivity option can be used in an architecture that combines on-premises and collocated, on-premises-only, or collocated-only resources. The collocated-only option is not covered in this section because it is the same as the combination of on-premises and collocated, without the on-premises components.

On-Premises and Collocated

This section assumes that the collocation facility being used is the AWS partner in the region. AWS partner collocation gives the collocated resources the best connectivity into AWS. The following diagram shows an overview of the components and the demarcation of physical connectivity. The AWS partner is responsible only for the cross-connection between the collocated customer gateway device and the AWS routers.

Figure 6: Overview of on-premises and collocated AWS Direct Connect

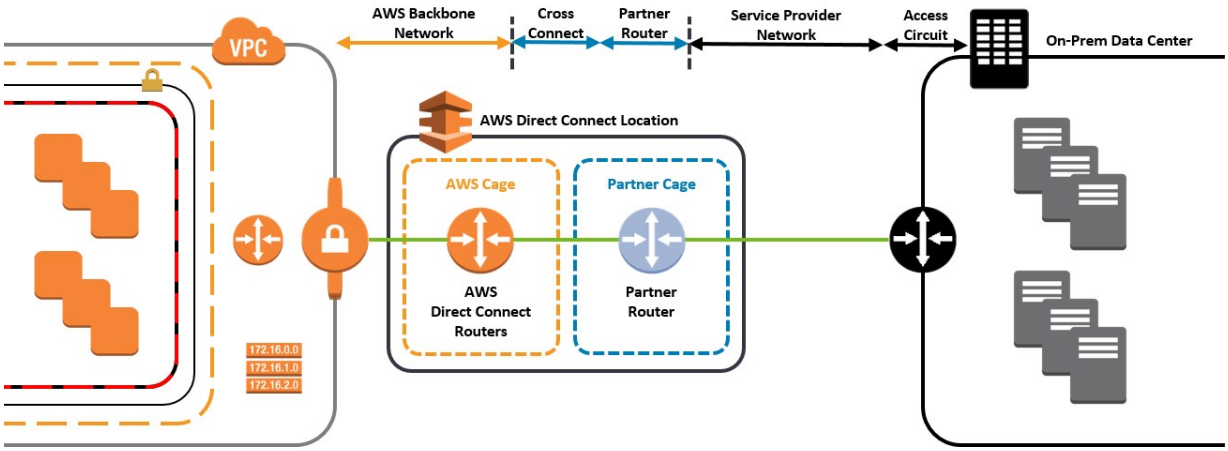


On-Premises Only

In this configuration, the customer gateway device connects to the AWS partner gateway device. The AWS partner is then responsible for the cross-connect between the partner router and the AWS DX routers. The

following diagram shows an overview of the components and the demarcation of physical connectivity. This configuration is used for the example architecture in this document.

Figure 7: Overview of on-premises-only AWS Direct Connect



Deploying AWS Virtual Private Cloud

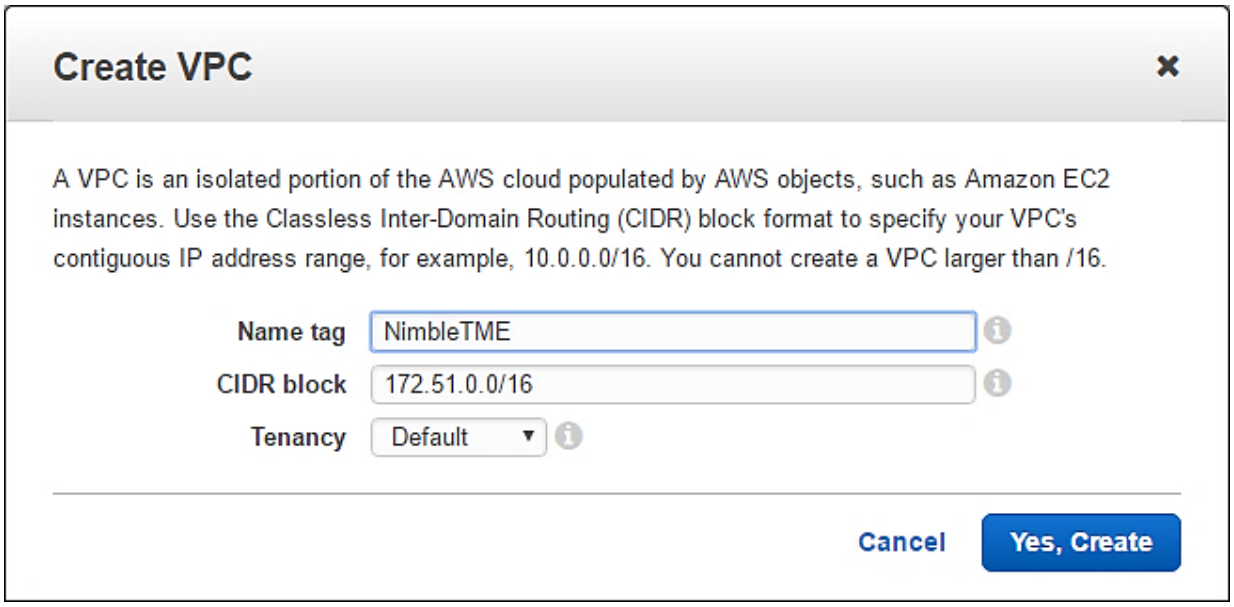
Deploying an EC2 instance into a VPC provides multiple benefits. For example, an EC2 instance in a VPC can use a persistent private IP address, enable the filtering of inbound and outbound traffic (security groups), and enable the use of network ACLs (filter traffic for the entire VPC).

Create and Configure a New VPC

There are many ways to set up a VPC. For more information about available options, see the *AWS VPC User Guide*. The following procedure provides a walk-through of an example VPC setup.

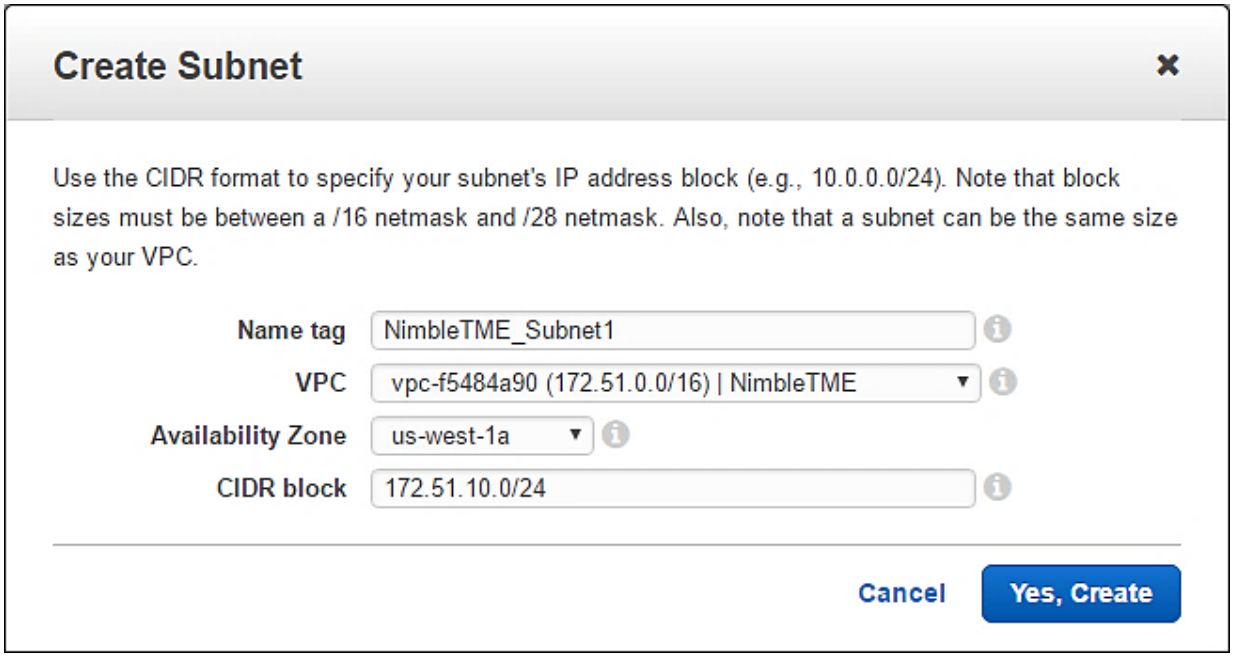
Procedure

1. Log in to the AWS console.
2. Click the **VPC** link in the **Networking** section of the main AWS service dashboard.
3. In the **Virtual Private Cloud** pane on the left side of the page, click the **Your VPCs** link.
4. Click **Create VPC**.
5. In the **Create VPC** window, verify or set the name tag, CIDR block, and tenancy options.



Note Use the appropriate naming conventions and CIDR blocks based on the specific configuration. The values shown throughout these steps are used only for the purposes of this example.

6 Click **Yes, Create**.
7 In the **Virtual Private Cloud** pane on the left side of the page, click the **Subnets** link.
8 Click **Create Subnet**.
9 In the **Create Subnet** window, verify or set the name tag, VPC, availability zone, and CIDR block options.



10 Click **Yes, Create**.

Verify Current Network ACL Settings

At this point in the process, the VPC is created and has a single subnet. Three additional objects are automatically created and associated with the VPC:

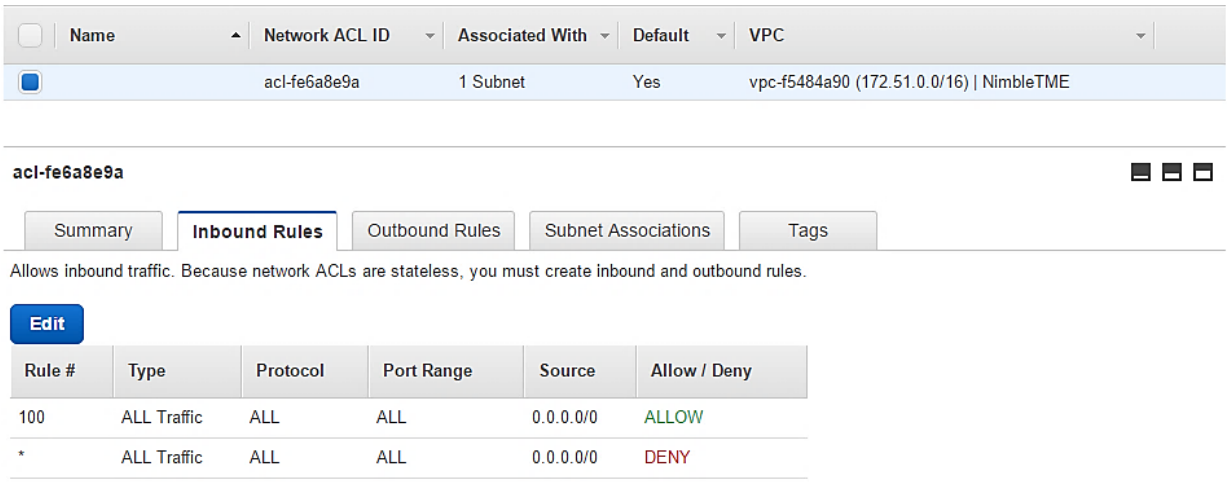
- A DHCP options set
- A route table
- A network ACL

Of the three, the only one that should be modified at this point is the network ACL (which controls what inbound and outbound traffic is allowed for the VPC).

Procedure

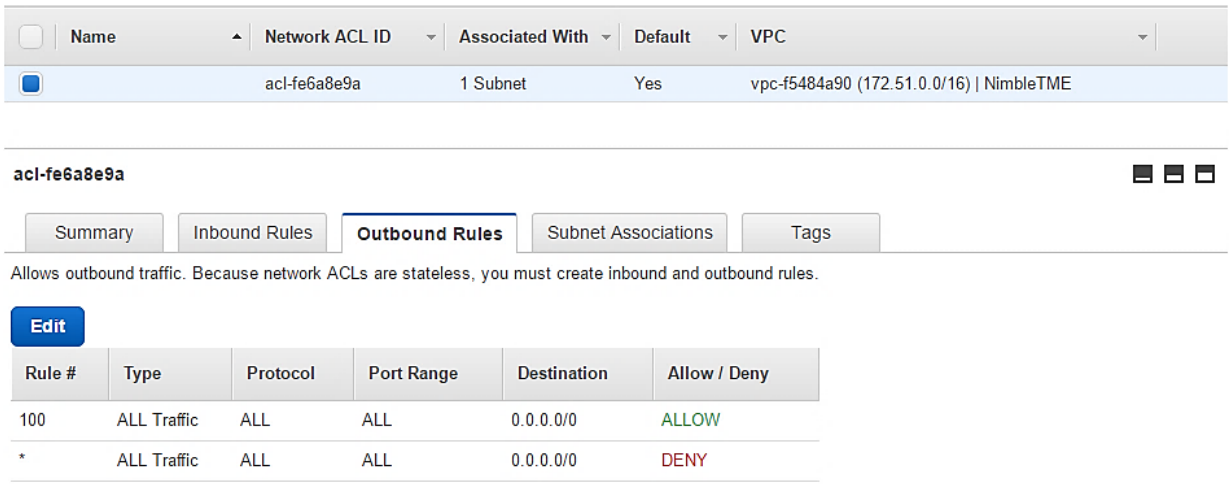
1. In the **Security** section on the left side of the page, click the **Network ACLs** link.
2. From the list of network ACLs, locate the appropriate entry by verifying it in the **VPC** column of the list, and select the checkbox to the left of the network ACL entry.
3. In the properties area at the bottom of the page, click the **Inbound Rules** tab.

   For this example, the default rule that allows all traffic is sufficient.



4. In the properties area at the bottom of the page, click the **Outbound Rules** tab.

   For this example, the default rule that allows all traffic is sufficient.



Name	Network ACL ID	Associated With	Default	VPC
	acl-fe6a8e9a	1 Subnet	Yes	vpc-f5484a90 (172.51.0.0/16) | NimbleTME

acl-fe6a8e9a

Summary | Inbound Rules | Outbound Rules | Subnet Associations | Tags

Allows outbound traffic. Because network ACLs are stateless, you must create inbound and outbound rules.

Edit

Rule #	Type	Protocol	Port Range	Destination	Allow / Deny
100	ALL Traffic	ALL	ALL	0.0.0.0/0	ALLOW
*	ALL Traffic	ALL	ALL	0.0.0.0/0	DENY

Note Recommendations for network ACLs are included in the VPC user guide for AWS, specifically in the *Recommended Network ACL Rules for Your VPC* section.

Deploying AWS Direct Connect

The process for deploying AWS DX is covered in detail in the *AWS Direct Connect User Guide*. Before requesting a new AWS DX connection, review the requirements that are listed in the user guide. They key requirements are as follows:

- The customer network must meet the following conditions:
  - Connections to AWS DX require single-mode fiber, 1000BASE-LX (1310nm) for 1 Gb Ethernet, or 10GBASE-LR (1310nm) for 10 GbE. Auto-negotiation for the port must be disabled. You must support 802.1Q VLANs across these connections.
  - Your network must support Border Gateway Protocol (BGP) and BGP MD5 authentication. Optionally, you may configure Bidirectional Forwarding Detection (BFD).
- Before you can connect to a VPC, you must complete the following tasks:
  - Provide a private autonomous system number (ASN). Amazon allocates a private IP address to you in the 169.x.x.x range.
  - Create a VGW and attach it to your VPC. For more information about creating a VGW, see *Adding a Hardware Virtual Private Gateway to Your VPC* in the Amazon VPC User Guide.

AWS DX setup (including VPN deployment) has four parts that must be completed:

- The AWS DX process
- Virtual private gateway (VGW) setup
- Customer gateway (CGW) setup
- VPN setup

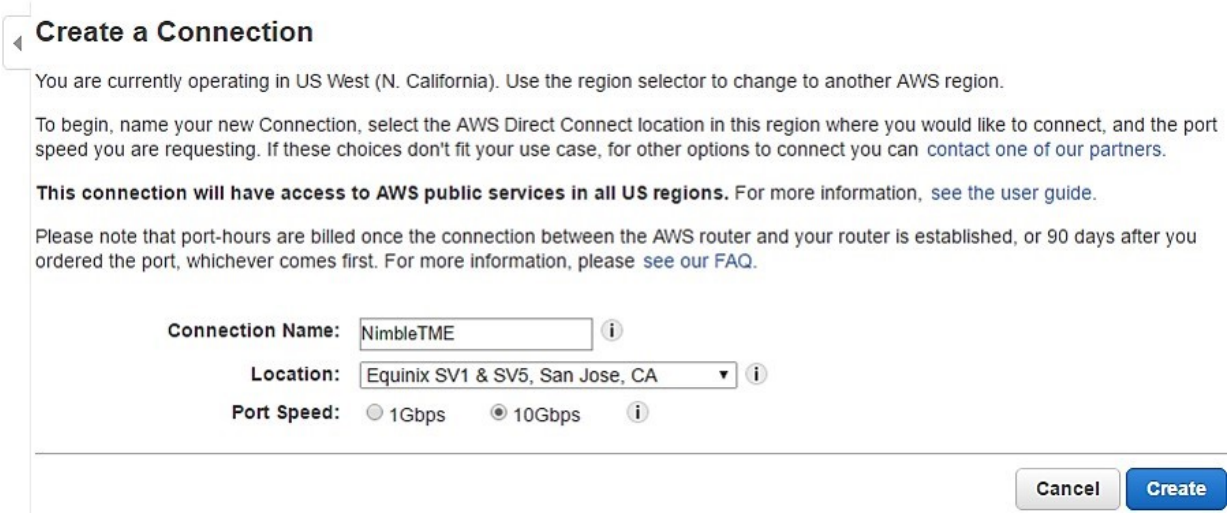
Completing the AWS Direct Connect Process

Create an AWS Direct Connect Connection

The first task in the AWS Direct Connect process is to create a new AWS Direct Connect connection.

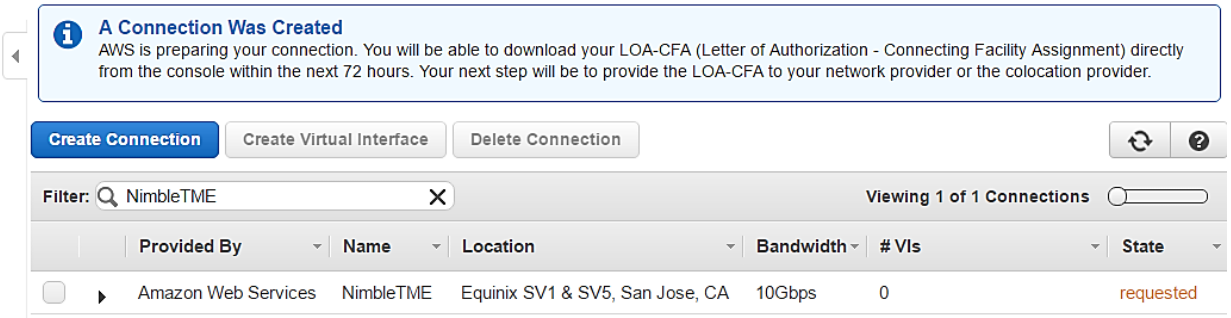
Procedure

1. Log in to the AWS console.
2. Click the **Direct Connect** link in the **Networking** section of the main AWS service dashboard.
3. Click **Create Connection**.
4. In the **Create a Connection** dialog box, verify or set the connection name, the location, and the port speed.

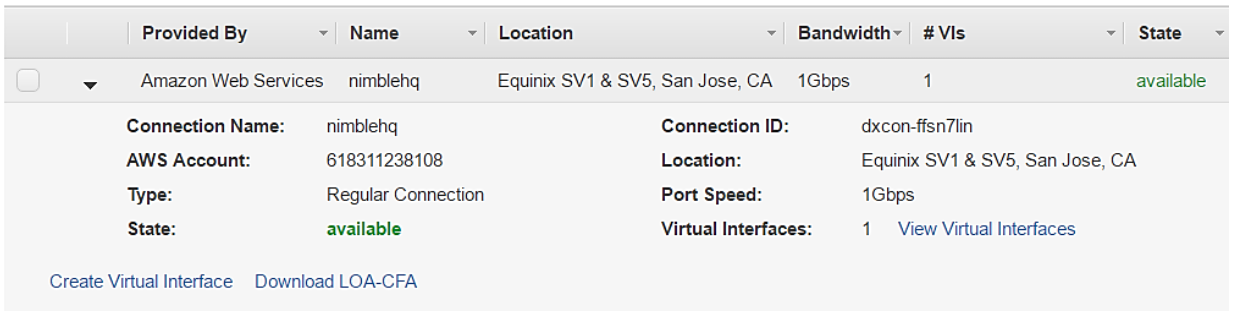


5. Click **Create**.

   In the list of connections, the **State** column for the newly created connection displays **Requested** in orange. In order to proceed, the Letter of Authorization – Connecting Facility Assignment (LOA-CFA) must be available. The LOA-CFA is made available from download within 72 hours of requesting the connection.



6. Verify that the LOA-CFA link is active for the requested connection by expanding the connection properties (click the black arrow icon). When the LOA-CFA link is active, click the **Download LOA-CFA** link.



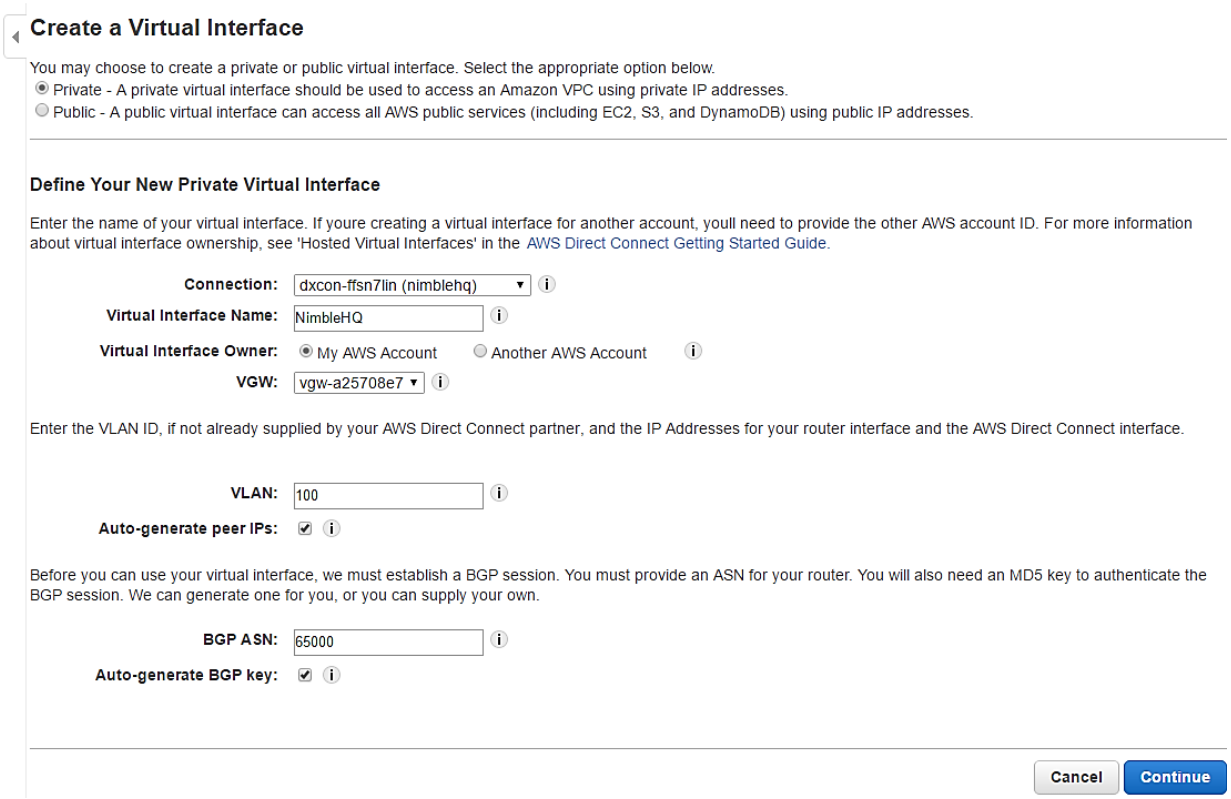
7 In the **Download LOA-CFA** dialog box, enter an appropriate provider name, if applicable, and click **Download**.
8 After the LOA-CFA letter is downloaded, contact the facility provider and complete the cross-connect request. (For instructions on how to complete the request, see the *AWS User Guide*).

Create a Virtual Interface

After the cross-connect is complete, create a virtual interface.

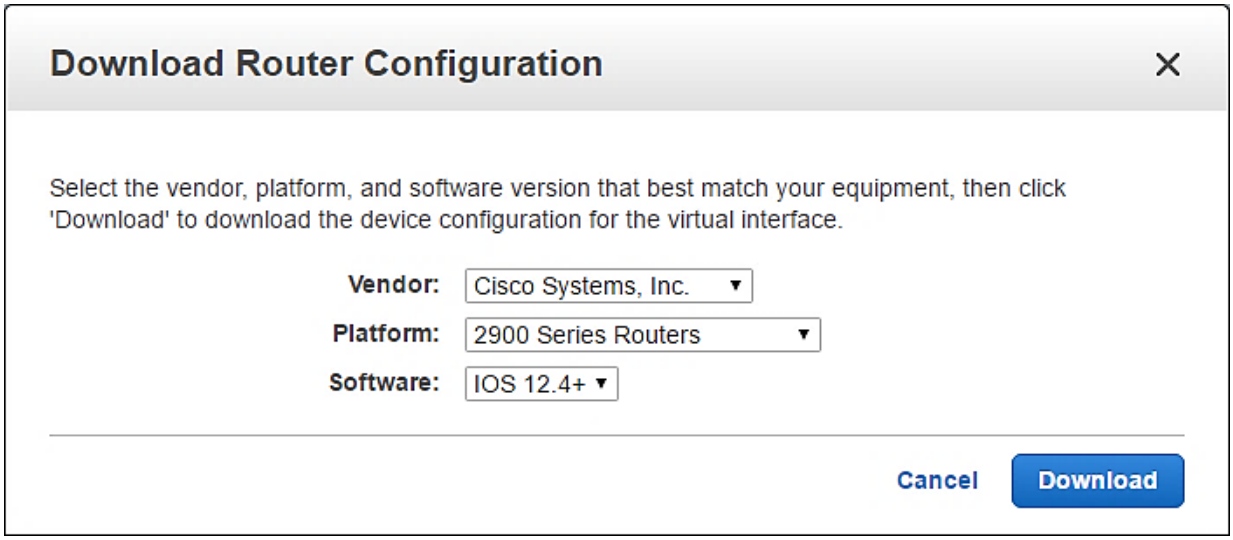
Procedure

1 Log in to the AWS console.
2 Click the **Diret Connect** link in the **Networking** section of the main AWS service dashboard.
3 Expand the connection properties for the connection and click the **Create Virtual Interface** link.
4 In the **Create a Virtual Interface** dialog box, verify or set the connection, the virtual interface name and owner, the VGW, the VLAN ID, and the BGB ASN.



Note The **VGW** field displays the ID of the VGW to use. To verify the correct ID to use, go to the **VPC Management** section in the AWS console and click **Virtual Private Gateways** (in the **VPN Connections** section on the left side of the dashboard). The **ID** column in the VGW list contains the ID string for the VGW.

5 Click **Continue** to return to the virtual interface dashboard and verify that the newly created virtual interface is listed.
6 From the list, select the checkbox for the appropriate virtual interface (for example, **NimbleHQ**).
7 Click the **Download Router Configuration** link located in the virtual interface properties.
8 In the **Download Router Configuration** dialog box, select the vendor, the platform, and the softwware that match the collocated or partner router.



Note If the proper device is not listed, it will be necessary to translate the configuration steps into steps that work for the specific device. (That task should not be difficult because the settings are very basic).

9 Click **Download**.
10 Apply the downloaded configuration to your CGW device.

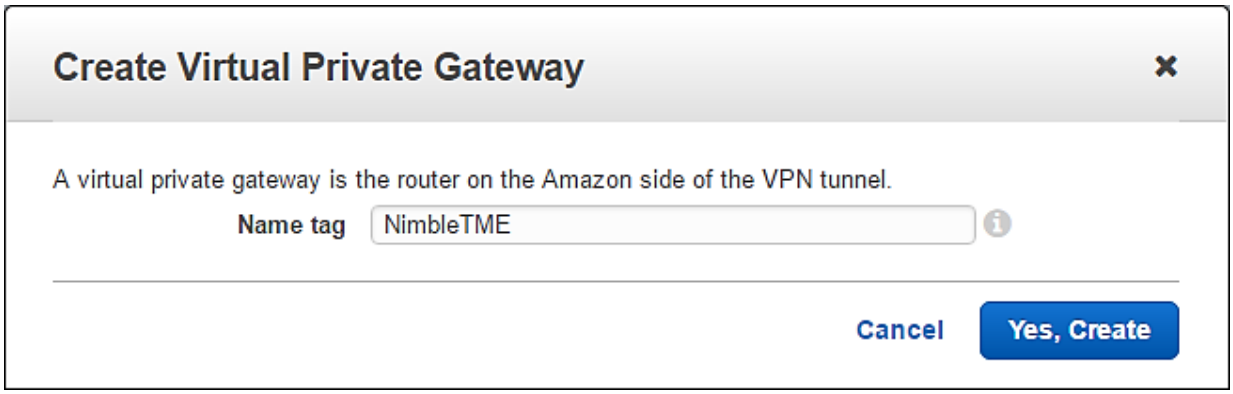
This configuration is what is necessary to apply to the router that will be connected directly to the AWS DX routers. If customer routers are collocated within the AWS DX partner facility, the configuration is applied to the customer routers. If on-premises-only connectivity is used, the configuration is applied to the partner router that routes the customer connection into AWS.

Set Up a Virtual Private Gateway

Create a VGW and attach it to the VPC.

Procedure

1 Log in to the AWS console.
2 Click the **VPC** link in the **Networking** section of the main AWS service dashboard.
3 In the **VPN Connections** section on the left side of the page, click the **Virtual Private Gateways** link.
4 Click **Create Virtual Private Gateway**.
5 In the **Create Virtual Private Gateway** dialog box, enter an appropriate name tag.



6 Click **Yes, Create**.
7 From the list of VGWs, locate the newly created **NimbleTME** entry. The **State** column for the entry displays **unattached** in red. Select the checkbox for the VGW entry.
8 Click **Attach to VPC**.
9 In the **Attach to VPC** dialog box, select the appropriate VPC from the drop-down list.

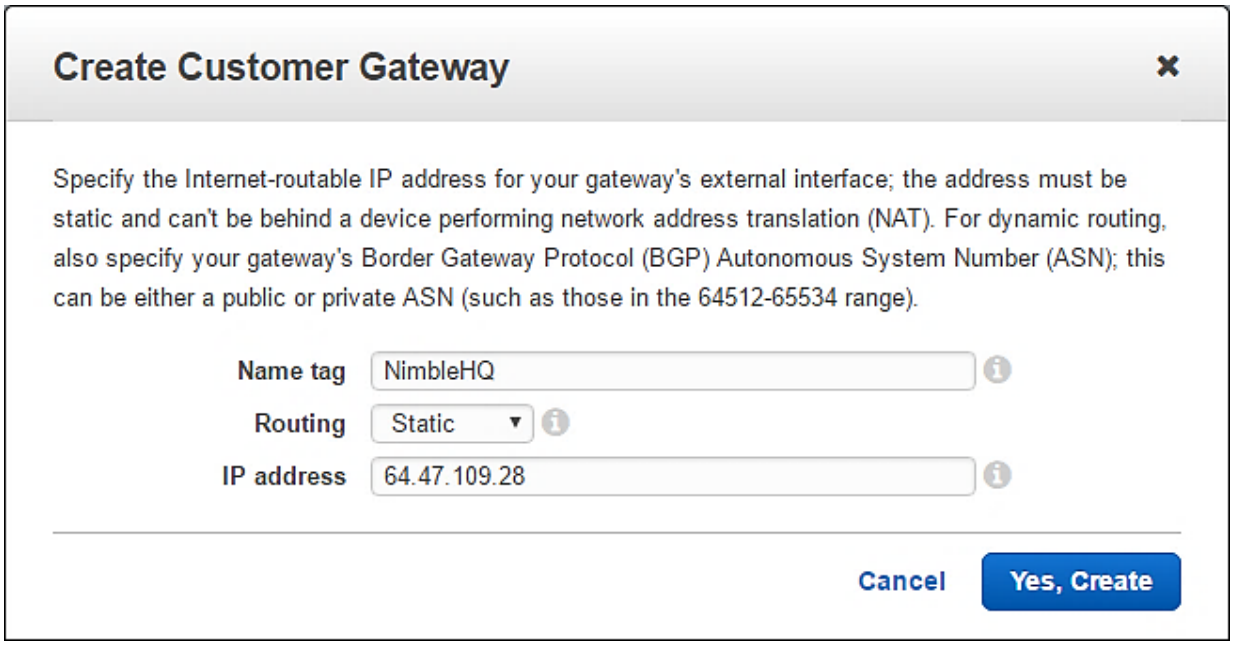
10 Click **Yes, Attach**.
11 Verify that the **State** column for the VGW displays **attached** in green.

Set Up a Customer Gateway

Depending on your connectivity, it might be necessary to create more than one CGW (for example, for a resilient dynamic VPN).

Procedure

1 Log in to the AWS console.
2 Click the **VPC** link in the **Networking** section of the main AWS service dashboard.
3 In the **VPN Connections** section on the left side of the page, click the **Customer Gateways** link.
4 Click **Create Customer Gateway**.
5 In the **Create Customer Gateway** dialog box, verify or set the name tag, the routing option, and the IP address.



6 Click **Yes, Create**.
7 Verify that the newly created CGW is listed in the customer gateway dashboard.

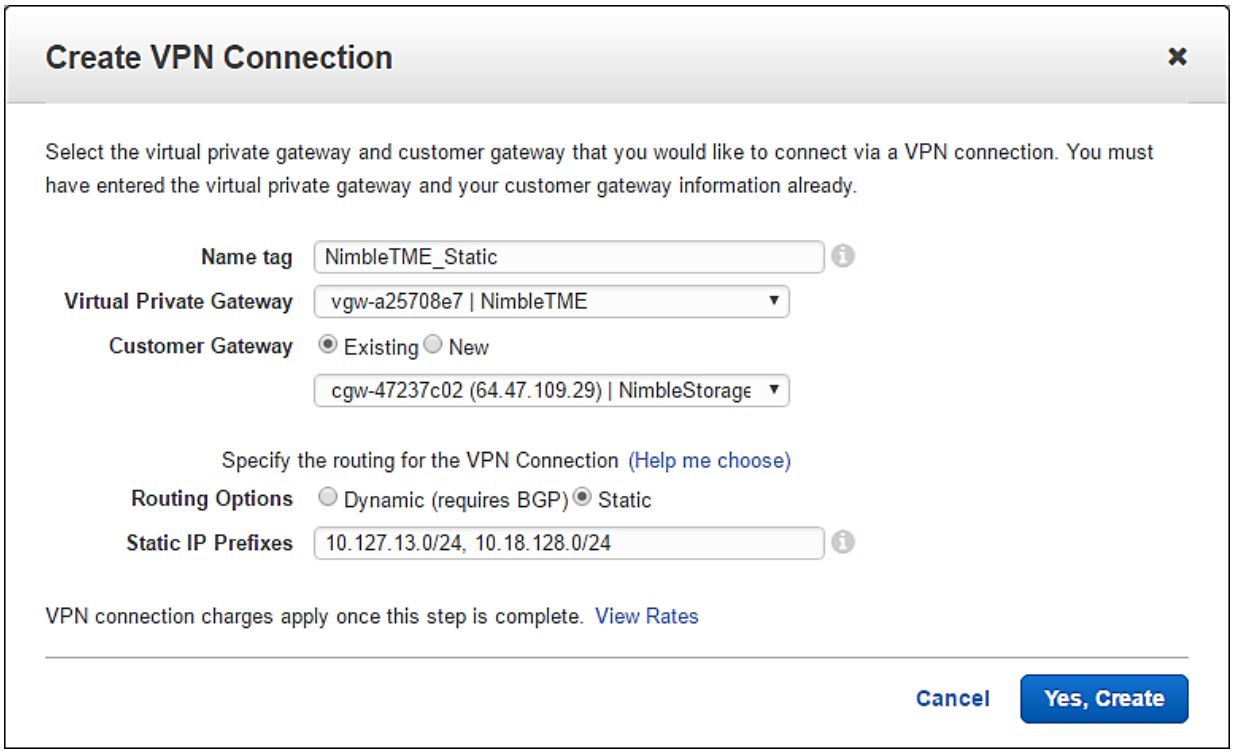
Set Up a Virtual Private Network

To complete the setup of secure and private connectivity into the VPC, set up a VPN connection over AWS Direct Connect.

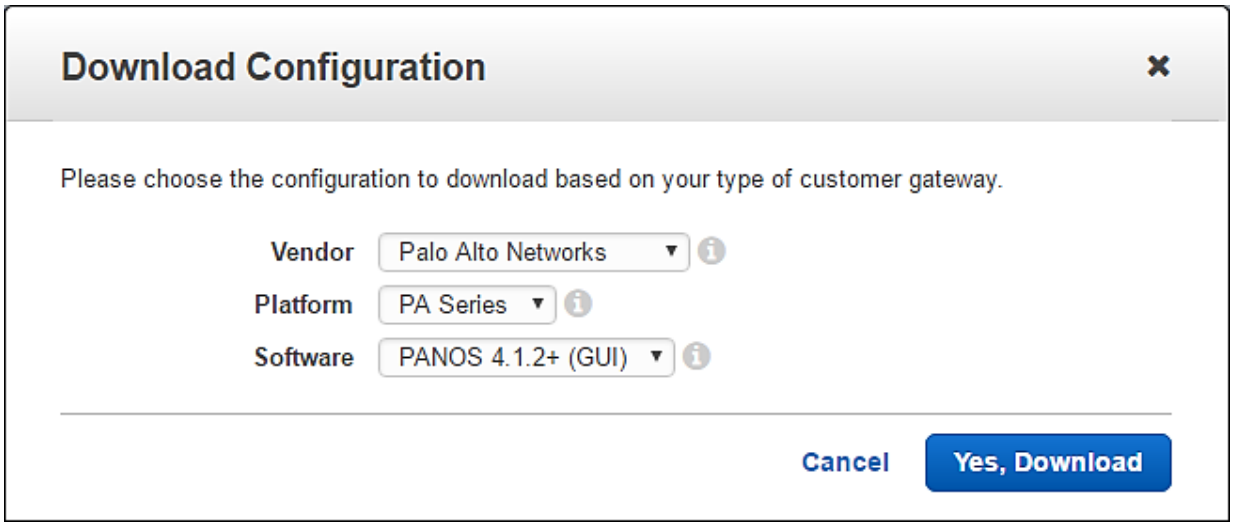
Note Some steps might differ slightly for each customer because of varying specific router models and OS versions. General web searches can offer helpful resources for specific devices. For example, the HPE Nimble Storage example used throughout this document uses a Palo Alto Networks (PAN) firewall, and *Rich's IT Blog* contains a very helpful post that is specific to PAN firewalls. Nevertheless, these types of resources should be used only if necessary because there is no guarantee that they are 100% correct.

Procedure

1 Log in to the AWS console.
2 Click the **VPC** link in the **Networking** section of the main AWS service dashboard.
3 In the **VPN Connections** section on the left side of the page, click the **VPN Connections** link.
4 Click **Create VPN Connection**.
5 In the **Create VPN Connection** dialog box, verify or set the name tag, the virtual private gateway, the customer gateway, the routing option, and the static IP prefixes.



6 Click **Yes, Create**.
7 From the list, select the checkbox for the VPN connection.
8 Click **Download Configuration**.
9 In the **Download Configuration** dialog box, select the vendor, platform, and software options that match the CGW device.



10 Click **Yes, Download**.
11 Apply the downloaded configuration to the CGW device.
12 After configuration is complete, deploy a LINUX or Windows EC2 instance with the ability to ping and trace routes to verify private IP connectivity from AWS to the on-premises or collocated hosts, storage arrays, or other pingable device.

Deploying an EC2 Instance

The process of deploying and configuring an EC2 instance includes two tasks:

- Launching a new EC2 compute instance
- Connecting to and then configuring the new instance

Launch a New EC2 Instance

Procedure

1. Log in to the AWS console.
2. Click the **EC2** link in the **Compute** section of the main AWS service dashboard.
3. Click **Launch Instance** in the **Create Instance** section of the EC2 dashboard.
4. Scroll to locate the **Microsoft Windows Server 2012 R2 Base** Amazon Machine Image (AMI). Click **Select** at the right of the AMI entry.

Note In this example, Windows Server 2012 R2 is used. However, the same general steps can be used to deploy any supported AMI.

5. Locate the instance type that matches your requirements and click the box located in the left column for the appropriate row (a blue square indicates which instance has been selected).

   For this example, the following instance type was selected:

   - Family = **General purpose**
   - Type = **m4.xlarge**
   - vCPUs = **4**
   - Memory (GiB) = **16**
   - Instance Storage (GB) = **EBS only**
   - EBS-Optimized Available = **Yes**
   - Network Performance = **High**

6. Click **Next: Configure Instance Details** at the bottom right of the page.
7. On the **Configure Instance Details** page, verify or set the following options:

   - Number of Instances = **1**
   - Network = **vpc-23577f46 (172.41.0.0/16) | NimbleAWSTunnel**
   - Subnet = **subnet-a75061fe (172.41.0.0/16) | us-west-1a**
   - Auto-assign Public IP = **Enable**

   Note Retain the default eth0 network interface that is listed. The network and subnet settings will be different because they are specific to each AWS customer. The network selected must be the VPC with the direct connect to the on-premises data center or collocation. Other options on the page can be configured as desired to meet operational requirements.

8. Click **Next: Add Storage**.

9 On the **Add Storage** page, verify that the root volume is listed.

Additional AWS storage volumes can be added if desired.

10 Click **Next: Tag Instance**.

11 On the **Tag Instance** page, click **Next: Configure Security Group**.

12 On the **Configure Security Group** page, click **Create a new security group** and enter a name and description for the new security group.

Note An existing security group can be used if desired.

13 Set or verify the security group rule.

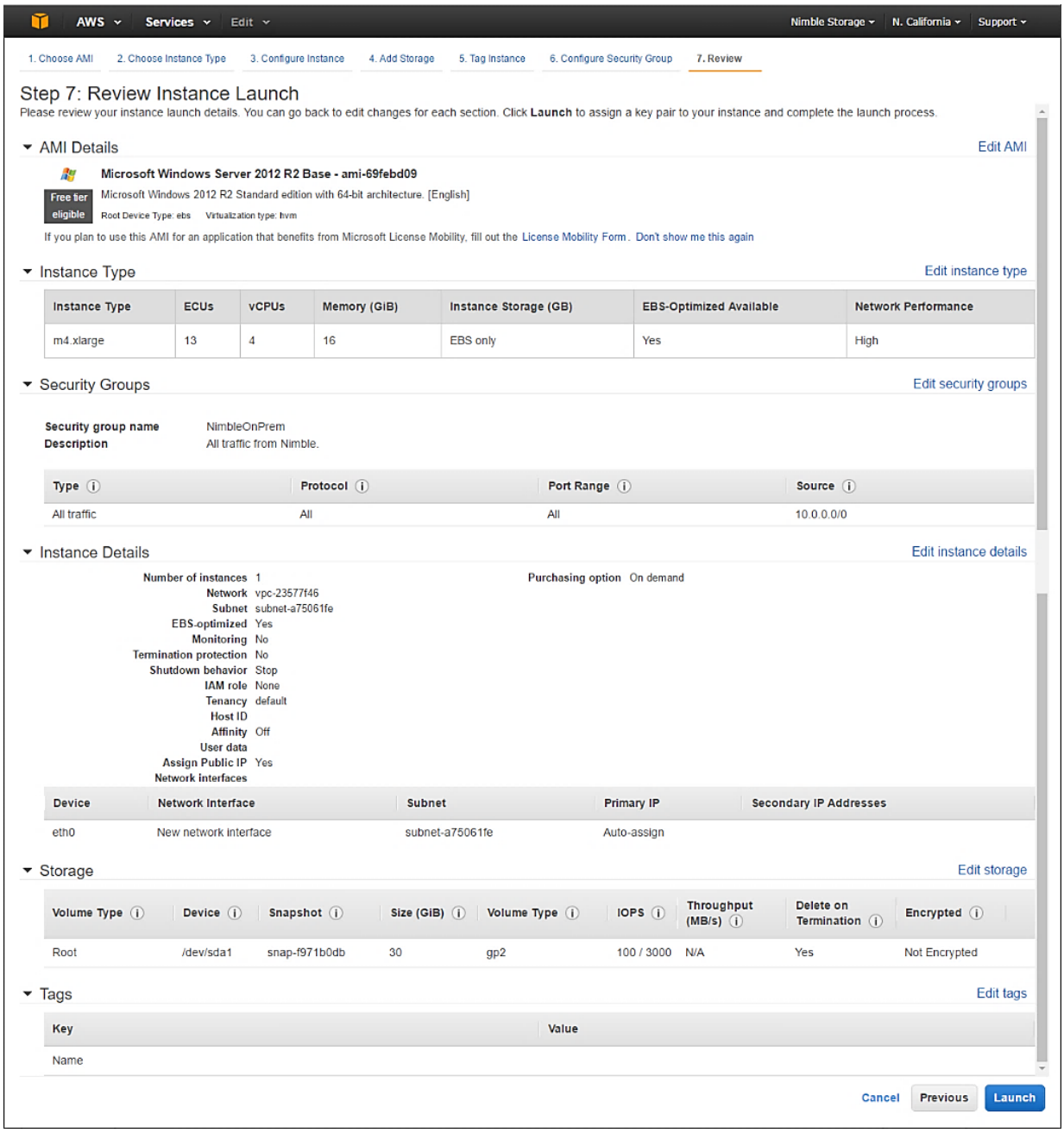
The default security group rule is for Remote Desktop Protocol (RDP). For this example, the following rule was used:

- Type = **All traffic**
- Source = **Anywhere | 10.0.0.0/0**

Note To keep the configuration simple and open to connectivity, only a single rule was used. Depending on your security requirements, it might be necessary to add multiple rules for specific protocols. If more restrictive rules are added, verify that at minimum the protocol used to connect to the instance is allowed (for example, RDP for Windows based instances). If necessary, changes can be made to the security group configuration after the instance is launched.

14 Click **Review and Launch**.

15 On the **Review and Launch** page, verify the instance configuration and then click **Launch**.



16 In the key pair dialog box, either create a new key pair or select an existing key pair.

In this example, an existing key pair named **TMEInstance** is selected. Select the checkbox for the appropriate key pair to acknowledge that the user has access to the `<key_pair_name>.pem` file, and click **Launch Instances**.

Note The key pair file (`.pem`) is critical to being able to decrypt the password to access the instance. Always back up the key pair file and secure it from improper use. For shared instances, HPE recommends that the administrator generate the password for the user, rather than allowing multiple users access to the key pair file.

17 When the **Launch Status** page displays, note the statement on that page that it takes a few minutes to launch the instance before you can connect.

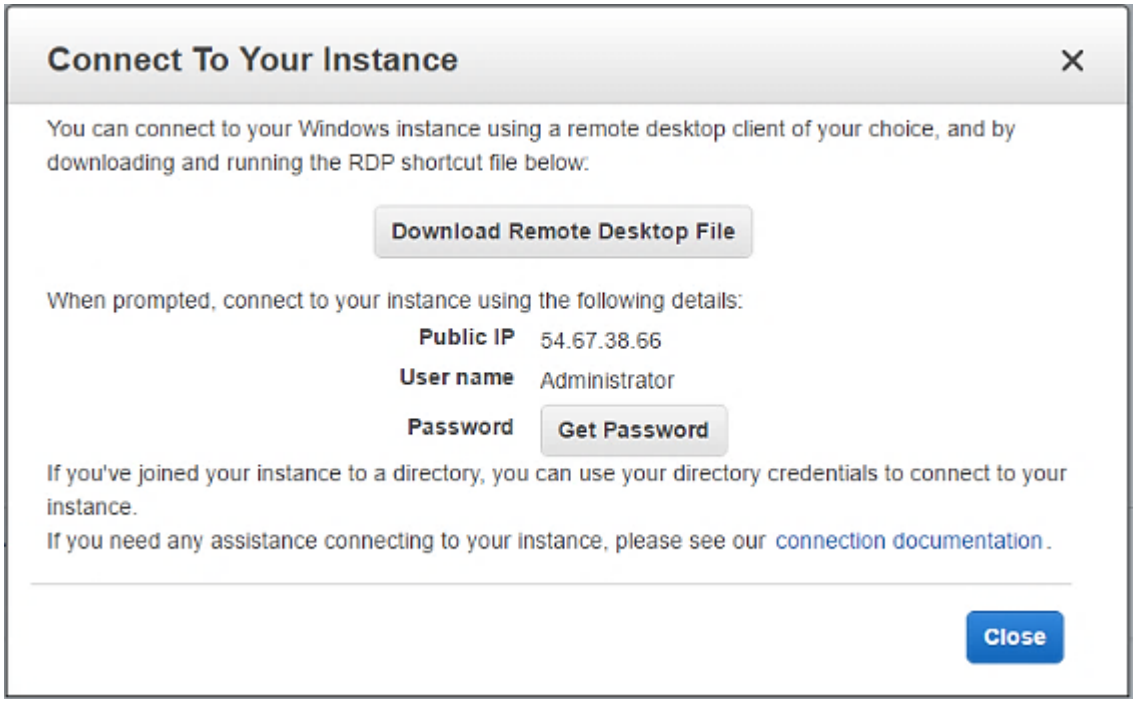
Click **View Instances** to see when the instance is ready. The **Status Checks** column on the instances dashboard indicates **Initializing** while the instance is being launched. When the column displays **2/2 checks passed**, you can connect to the instance.

Connect to and Configure the Instance

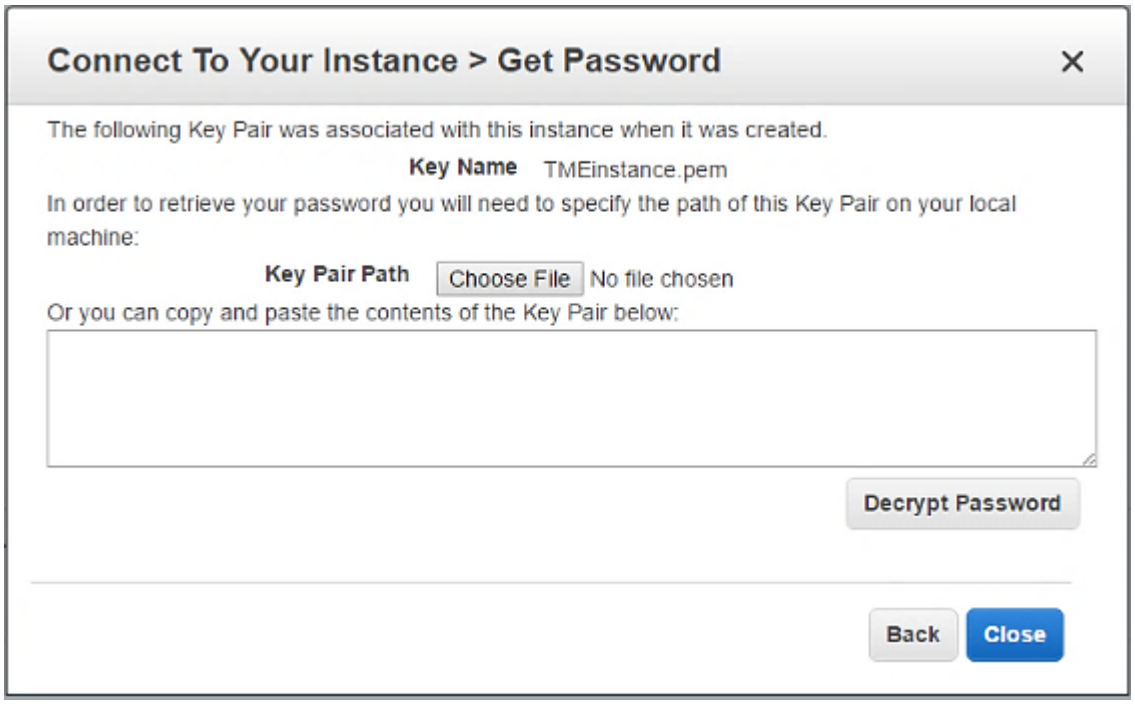
Procedure

1. Open the EC2 instances dashboard from the AWS console.
2. Locate the instance in the list and click the grey box to the left of the row to select the instance.

   When selected, the grey box becomes blue.
3. Above the instances dashboard, click **Connect**.
4. In the **Connect to Your Instance** dialog box, click **Get Password**.

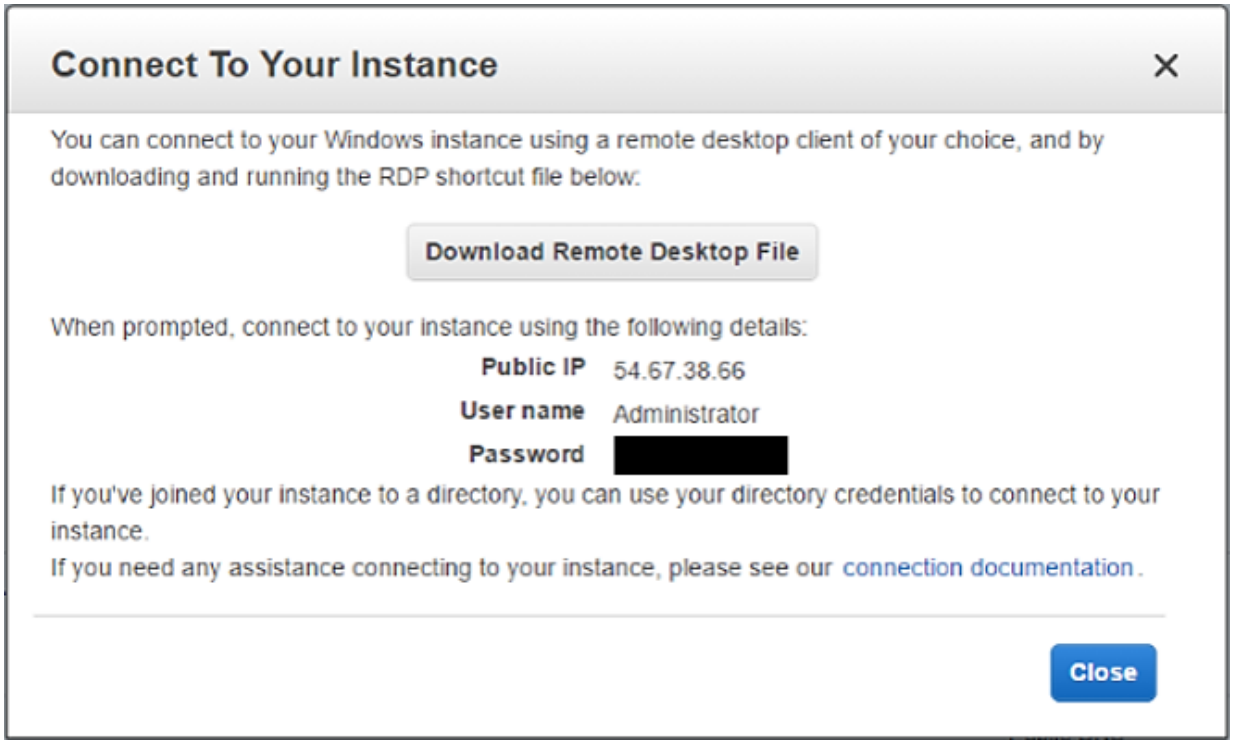


5. When the **Key Name** field displays the expected key pair file that must be used (for example, **TMEinstance.pem**), click **Choose File**, and then navigate to the designated key pair file and open it.



6 Click **Decrypt Password** to display the password in the **Password** field.

Record the password and the IP address because they will be used to open a remote desktop connection to the instance.



7 Click **Close** and exit the AWS console.

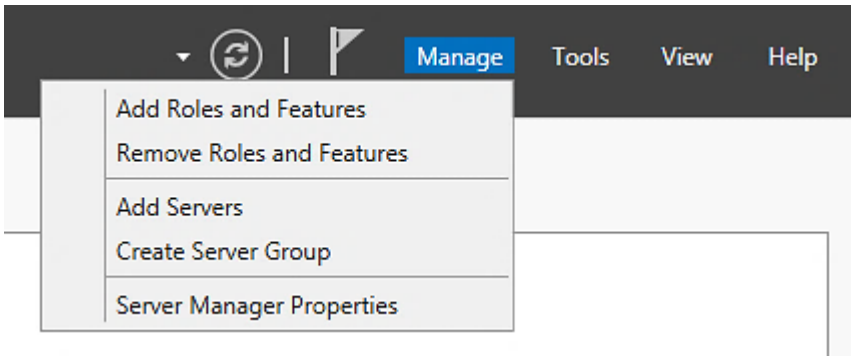
Note The next steps are related to connecting to the instance from a Windows based system through a remote desktop connection. With the credentials determined at this point, MAC and LINUX users should understand how to connect to the instance by using the appropriate local OS tools.

8 Open the **Remote Desktop Connection** application on your local host.
9 Enter the appropriate IP address in the **Computer** field and click **Connect**.
10 In the **Windows Security** window, enter **Administrator** for the user name and use the previously generated password.

The system might prompt you to accept a certificate before connecting to the remote system. If this window appears, click the **Don't ask me again for connections to this computer** checkbox and then click **Yes** to display the remote desktop.

Note Because the AMI used in this example is Windows Server 2012 R2, the next steps apply to that operating system specifically.

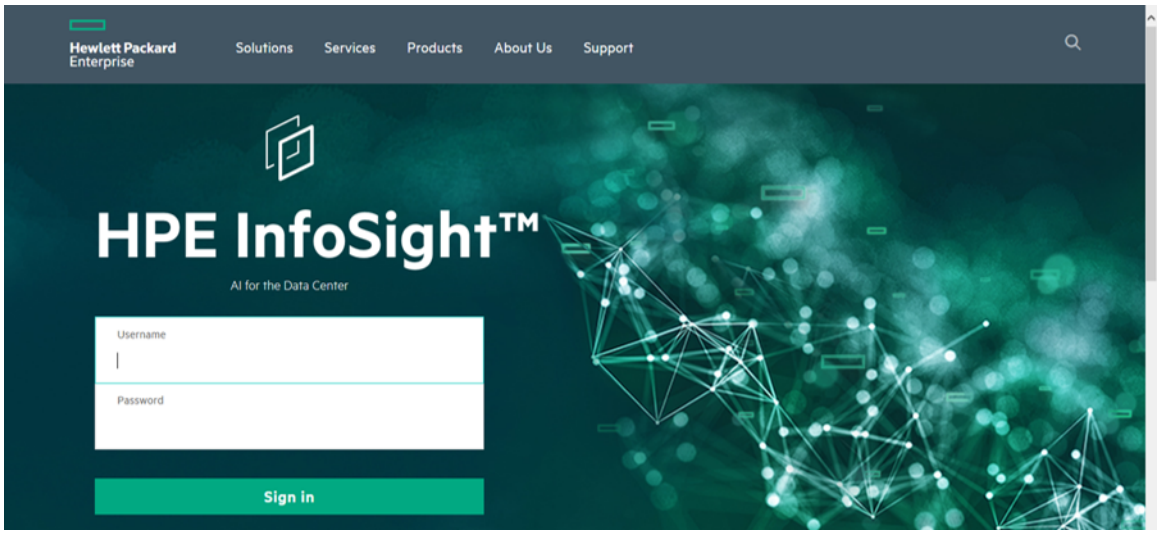
11 Click **Windows** and select **Server Manager** from the start menu to begin installing multipath I/O on the system.
12 From the menu options located at the top right of the window, select **Manage** and then click **Add Roles and Features**.



13 On the **Before you Begin** page of the **Add Roles and Features** wizard, click **Next**.
14 On the **Installation Type** page, verify that **Role-based or feature-based installation** is selected and click **Next**.
15 On the **Server Selection** page, verify that **Select a server from the server pool** is selected and that the Windows Server instance is highlighted in the **Server Pool** list. Click **Next**.
16 On the **Server Roles** page, click **Next**.
17 On the **Features** page, locate **Multipath I/O** in the Features list and select the checkbox for the feature. Click **Next**.
18 On the **Confirmation** page, click **Install** and wait for the installation to be completed.
19 On the **Results** page, click **Close**.
20 Close the **Server Manager** window.
21 Open **Internet Explorer** and navigate to *HPE InfoSight*.

Note Internet security is high by default. It will be necessary to either add blocked sites as you navigate or lower the internet security settings in order navigate to HPE InfoSight.

22 Enter the appropriate user name and password to log in to HPE InfoSight. If you do not have existing credentials, click the **Create Account** link and register for an account (then return to this step).



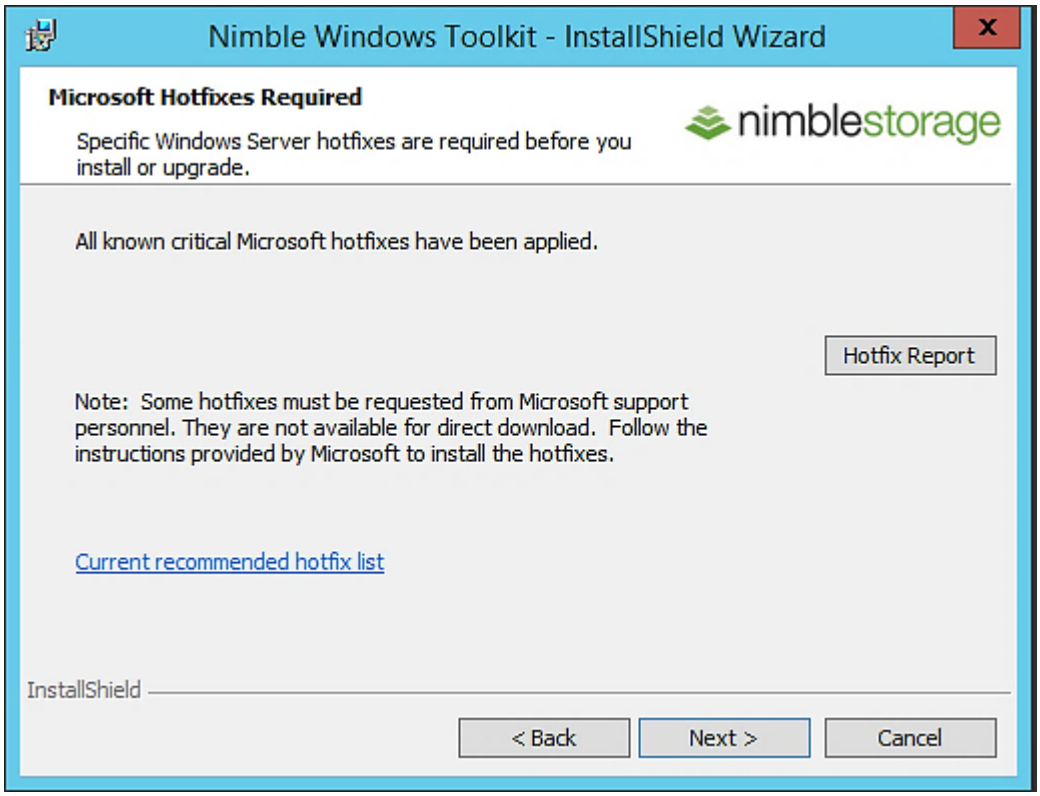
23 From the main HPE InfoSight dashboard, select the **Resources** menu and select **Software Downloads** from the list.
24 Click the **HPE Nimble Storage Windows Toolkit (NWT)** link from the list on the left side of the page under **Integration Kits**.
25 Click the **Show other versions** link located next to the **Current Version** section.
26 Locate the entry for **3.2.0.410 (RC)** from the **Other Versions** list and click the **Software (64-bit)** link to download the NWT.

   After it is downloaded, launch the installation file.

   Note The version of the NWT used should be the same as or higher than the version of NimbleOS installed on the HPE Nimble Storage array. Although this example uses 3.2.0.410 (RC), any supported version that is appropriate for the configuration can be used.

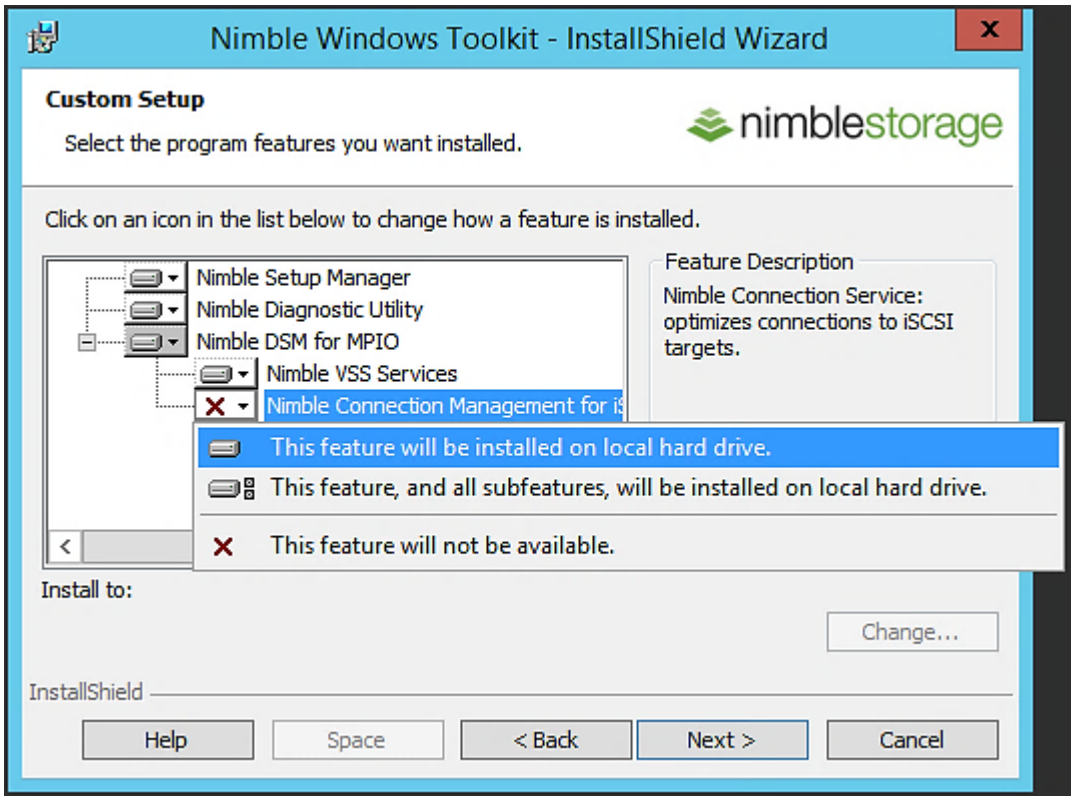
27 On the **Welcome** page of the installation wizard, click **Next**.
28 Accept the terms of the license agreement and click **Next**.
29 Verify that all critical Microsoft hotfixes have been applied to the system and click **Next**.

   To generate a report that identifies any gaps, click **Hotfix Report**. Resolve gaps by installing the appropriate hotfixes. The topic of installing hotfixes is outside the scope of this document, but you must resolve any problems before continuing the installation.



30 On the **Nimble Logs Directory** page, click **Next**.

31 On the **Custom Setup** page, click **Nimble Connection Management for iSCSI** in the features list, and select the option **This feature will be installed on local hard drive**. Click **Next**.



32 On the **Ready to Install the Program** page, click **Install**.
33 During installation, when a dialog box appears, indicating that HPE will adjust registry key values, click **OK**.
34 When installation is complete, click **Finish**.
35 Click **Yes** to restart the system and finalize the installation process.
36 When the system restart is complete, click **Windows**.
37 Locate the **HPE Nimble Storage** applications in the start menu and select the **HPE Nimble Storage Connection Manager** application.
38 Record the IQN string located in the **Initiator Name** field because it will be needed for configuring the HPE Nimble Storage array.

Deploying an HPE Nimble Storage Array

The topic of basic installation and configuration of an HPE Nimble Storage array is outside the scope of this document. For those instructions, see the appropriate installation and administration guide *documentation* available from HPE InfoSight.

After the basic setup and configuration of the array is complete, you can proceed to complete the array configuration steps that are specific to deploying a volume for the EC2 instance.

NimbleOS 3.4 and later includes the next-generation array management interface. NimbleOS versions earlier than 3.4 run the classic array management interface. The steps for configuring the array differ slightly between the two interfaces.

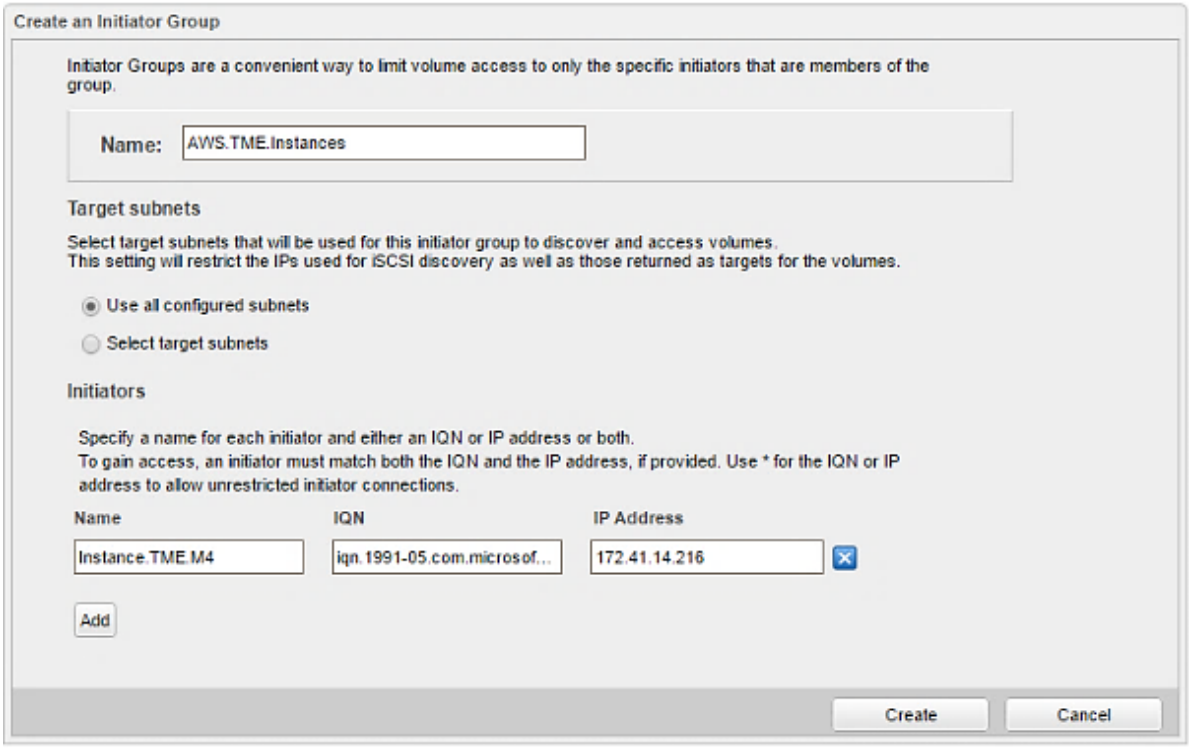
Configure the Array in the Classic Array Management Interface

Procedure

1. Open a web browser and navigate to the array management interface.
2. Enter the appropriate user name and password and click **Log In**.
3. On the array dashboard page, select the **Manage** menu and select **Initiator Groups** from the list.
4. On the **Initiator Groups** page, click **Create**.
5. In the **Create an Initiator Group** window, enter an appropriate name for the initiator group in the **Name** field (for example, **AWS.TME.Instances**), and verify that **Use all configured subnets** radio button is selected.

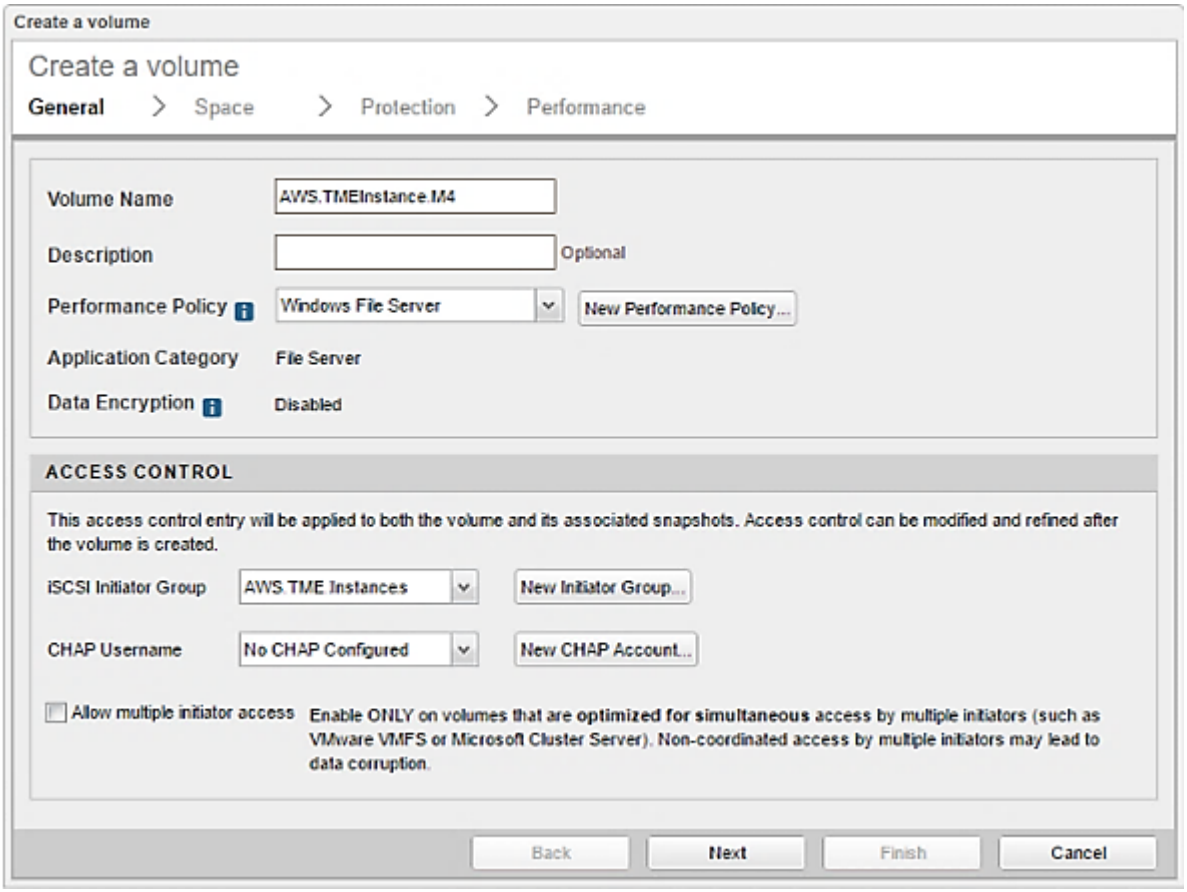
   Note A single initiator group can be used for all IQNs that are associated with AWS EC2 instances.

6. In the **Initiators** section, provide a name for each initiator and either the IQN or the IP address, or both:
   - **a** In the **Name** field for the IQN entry, specify a name that describes the specific EC2 instance (for example, **Instance.TME.M4**).
   - **b** In the **IQN** field, specify the IQN string that was recorded from the EC2 instance (for example, **iqn.1991-05.com.microsoft.win-9jtu9s7f4v2**).
   - **c** In the **IP Address** field, enter the private IP address of the EC2 instance (for example, **172.41.14.218**).



7. Click **Create**.
8. Select the Manage menu and select **Volumes** from the list.
9. On the volumes dashboard, click **New Volume** to open the **Create a volume** wizard.
10. On the **General** page of the **Create a volume** wizard, provide the volume name, the performance policy, and the iSCSI initiator group, and then click **Next**:
    - **a** In the **Volume Name** field, specify **AWS.TMEInstance.M4**.
    - **b** In the **Performance Policy** field, select **Windows File Server**.
    - **c** In the **iSCSI Initiator Group** field, select **AWS.TME.Instances**.

Note These values are presented for the example in this document. Use the values that are appropriate for each specific configuration.



11 On the **Space** page, enter a numeric value (for example, **1**) in the **Size** field and set the appropriate capacity (for example **TiB**), and then click **Next**.

12 On the **Protection** page, select the appropriate radio button for the protection policy.

In this example, the volume is not in a replication relationship, so **No volume collection** is selected.

13 Click **Finish**.

Configure the Array in the Next-Generation Array Management Interface

Procedure

1 Open a web browser and navigate to the array management interface.

2 Enter the appropriate user name and password and click **Log In**.

3 On the array dashboard, select the **Manage** menu and select **Data Access** from the list.

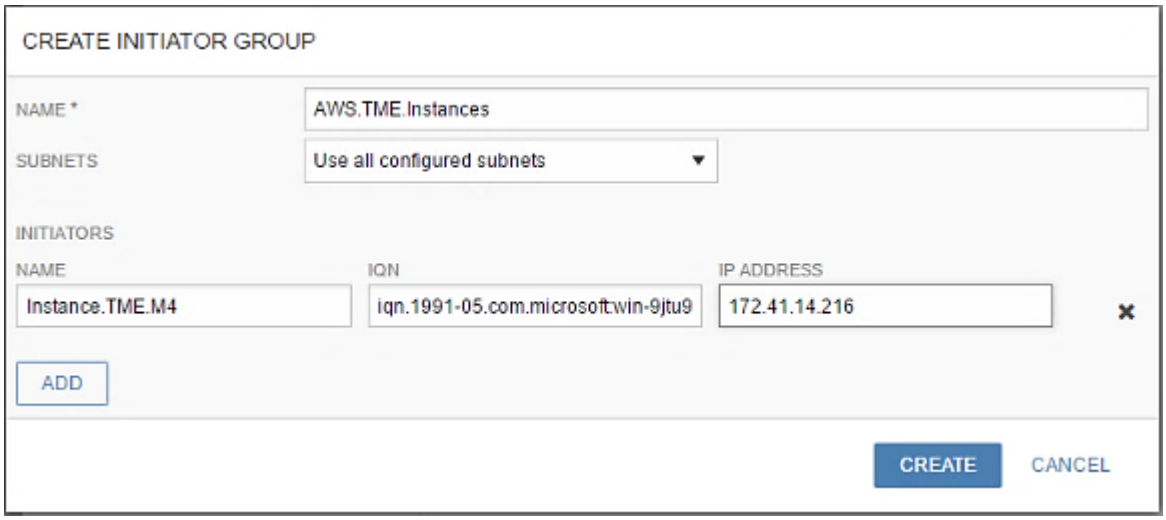
4 On the **Data Access** page, click the **+** button to add an initiator group.

5 In the **Create Initiator Group** window, enter an appropriate name for the initiator group in the **Name** field (for example, **AWS.TME.Instances**) and verify that **Use all configured subnets** is selected in the **Subnets** field.

Note A single initiator group can be used for all IQNs that are associated with AWS EC2 instances.

6 In the **Initiators** section, click **Add**, and provide a name for each initiator, as well as the IQN and the IP address:

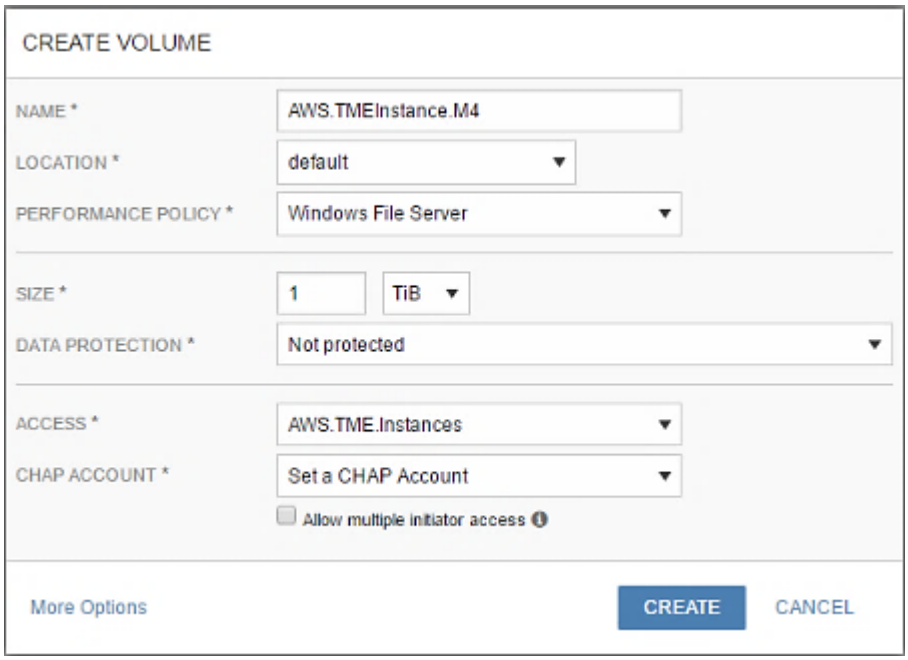
a In the initiator **Name** field, specify a name for the IQN entry that describes the specific EC2 instance (for example, **Instance.TME.M4**).
b In the **IQN** field, enter the IQN string that was recorded from the EC2 instance (for example, **iqn.1991-05.com.microsoft:win-9jtu9s7f4v2**).
c In the **IP Address** field, provide the private IP address of the EC2 instance (for example, **172.41.14.218**).



7 Click **Create**.
8 Select the **Manage** menu and select **Data Storage** from the list.
9 On the **Data Storage** page, click the **+** button to add a new volume.
10 In the **Create Volume** window, provide the information to define the volume:

a In the **Name** field, enter **AWS.TMEInstance.M4**.
b In the **Performance Policy** field, select **Windows File Server**.
c In the **Size** field, specify **1 TiB**.
d In the **Data Protection** field, select **Not protected**.
e In the **Access** field, select **AWS.TME.Instances**.

Note These values are presented for the example in this document. Use the values that are appropriate for each specific configuration.



11 Click **Create**.

Deploying an HPE Nimble Storage Volume on an EC2 Host

With the Windows Server base prepared and the HPE Nimble Storage array configuration complete, the next task is to complete the host volume configuration, which requires connecting the storage volume and allocating the volume.

Connect the Storage Volume

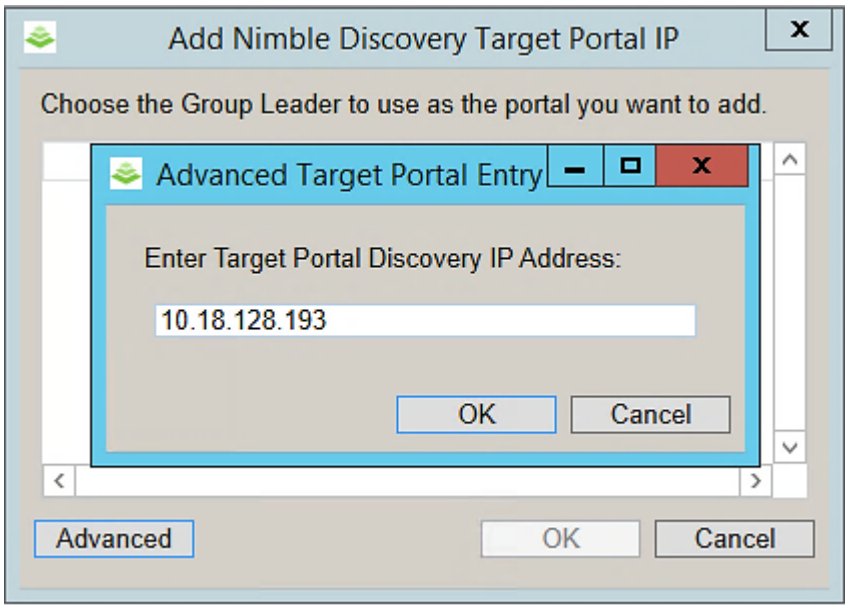
Procedure

1. Open the **Remote Desktop Connection** application on your local host.
2. Enter the appropriate IP address in the **Computer** field and click **Connect**.
3. In the **Windows Security** window, enter **Administrator** for the user name and use the previously generated password.

   The system might prompt you to accept a certificate before connecting to the remote system. If this window appears, select the **Don't ask me again for connections to this computer** checkbox and then click **Yes** to display the remote desktop.

   Note Because the AMI used is Windows Server 2012 R2, the next steps apply to that operating system specifically.

4. If the HPE Nimble Storage Connection Manager application is not already open, click **Windows** to open it and then locate **Nimble Storage** applications in the start menu and click the **Nimble Connection Manager** application.
5. In the **Nimble Discovery IP (Port 3260)** section of the **Nimble Connection Manager** main window, click **Add**.
6. In the **Add Nimble Discovery Target Portal IP** window, click **Advanced**.
7. In the **Advanced Target Portal Entry** window, enter the iSCSI discovery IP address of the HPE Nimble Storage array (for example, **10.18.128.193**), and click **OK**.



8 In the **Add Nimble Discovery Target Portal IP** window, click **OK**.

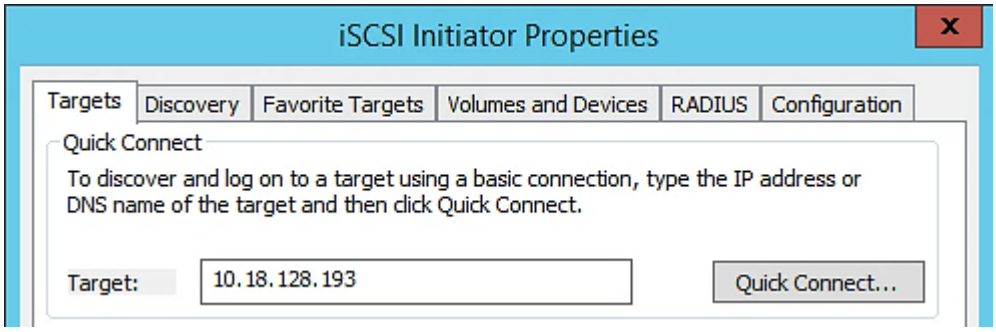
9 In the **Nimble Connection Manager** main window, verify that the discovery IP address has been added to the list in the **Nimble Discovery IP (Port 3260)** section.

Leave the application open.

10 Click **Windows** and click the search icon (the magnifying glass) on the **Start** page.

11 Type **iscsi initiator** in the search field and select the **iSCSI Initiator** application entry from the search list.

12 On the **Targets** tab of the **iSCSI Initiator Properties** window, enter the iSCSI discovery IP address of the HPE Nimble Storage array in the **Target** field (for example, **10.18.128.193**), and click **Quick Connect**.

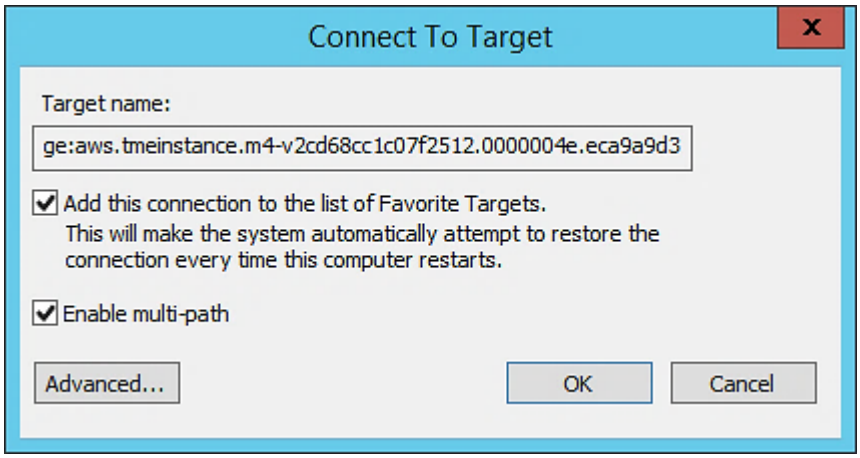


13 In the **Quick Connect** window, verify that the proper IQN is listed and that the **Status** column indicates **Connected**, and click **Done**.

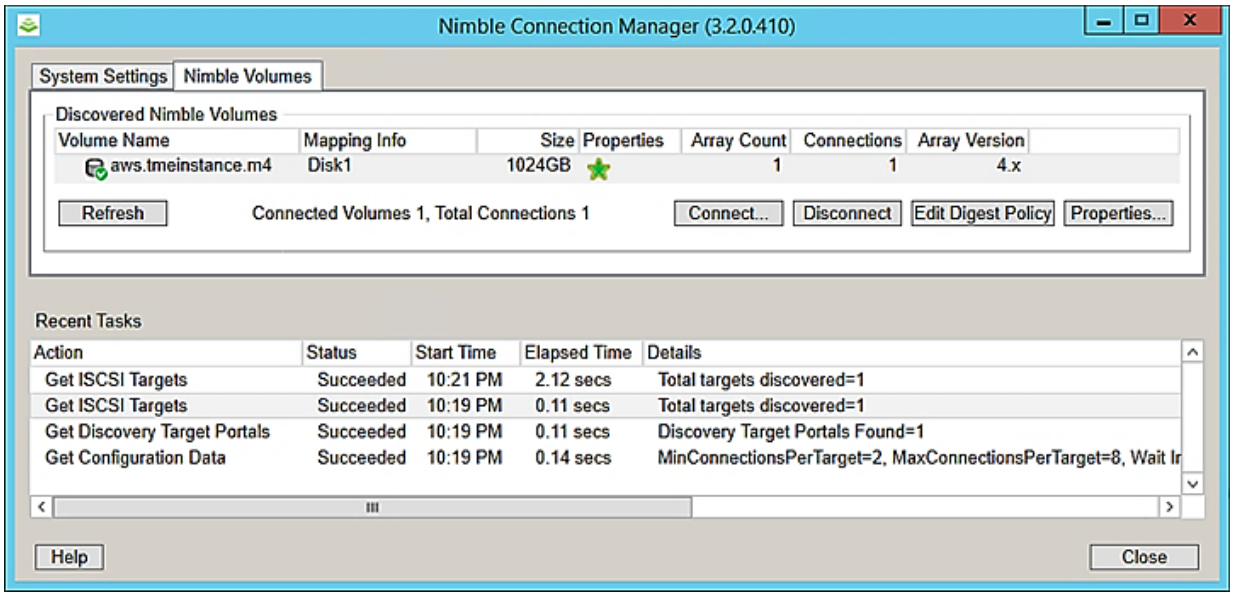
14 In the **Discovered Targets** section of the **iSCSI Initiator Properties** window, click the IQN that is associated with the HPE Nimble Storage array.

With the IQN highlighted, click **Connect**.

15 In the **Connect To Target** window, select the **Enable multi-path** checkbox and click **OK**.



16 Close the **iSCSI Initiator Properties** window.
17 Open the **Nimble Connection Manager** application (which should have been left open from step 9).
18 Click the **Nimble Volumes** tab.
19 Verify that the volume is listed as connected and healthy in the **Discovered Nimble Volumes** section.



Allocate the Windows Volume

Procedure

1 On the **Start** page, click **Windows** and click the search icon (the magnifying glass).
2 Type **disk management** in the search field and select **Create and format hard disk partitions** from the search list.

   In the main window of the **Disk Management** application, an unknown disk entry should display that matches the size of the volume created on the HPE Nimble Storage array (for example, **Disk 1, Unknown, 1024.00 GB, Offline, Unallocated**).

3 Right-click the unknown disk icon and select **Online** from the menu.
4 Right-click the unknown disk and select **Initialize Disk** from the menu.

5 In the **Initialize Disk** window, click **OK**.

The disk entry should show as online and basic, but still unallocated.

6 Right-click the **Unallocated** section of the disk entry and select **New Simple Volume** from the menu.

7 On the **Welcome** page of the **New Simple Volume** wizard, click **Next**.

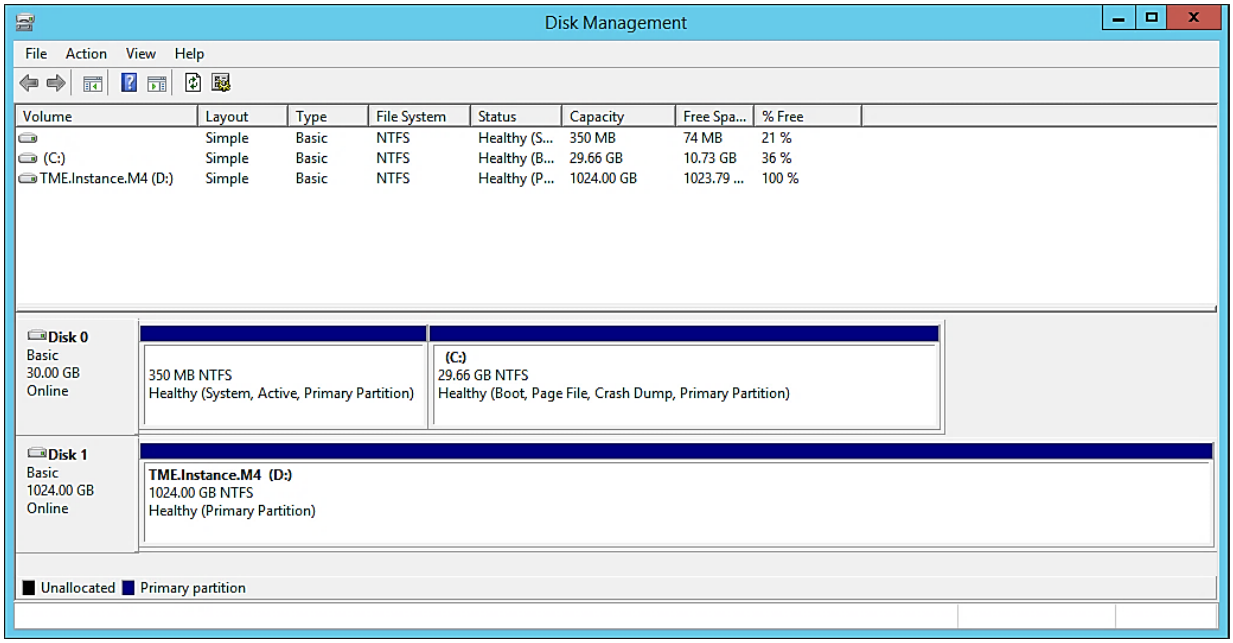
8 On the **Specify Volume Size** page, verify that the maximum size is selected and click **Next**.

9 On the **Assign Drive Letter or Path** page, verify that the **Assign the following drive letter** radio button is selected and the drop-down menu lists an appropriate drive letter (for example, **D**), and then click **Next**.

10 On the **Format Partition** page, enter an appropriate name in the **Volume Label** field (for example, **TME.Instance.M4**) and click **Next**.

11 On the **Complete** page, click **Finish**.

The disk entry should show the status of the disk as **Formatting**. When complete, the disk status changes to **Healthy (Primary Partition)**.



12 Open Windows File Explorer and verify that the new volume is shown and is accessible.

Note To further verify that the volume is working correctly, try copying or creating files in the new volume.

Conclusion

The deployment of on-premises or collocated storage arrays in support of cloud compute resources offers multiple benefits. You can take advantage of the enterprise features, availability, accessibility, and scalability of best-in-class storage solutions. Most important, you can retain control over what is perhaps a company's most critical asset, its data.

Cloud isn't the answer to everything (at least not yet), but it is clear that the new reality of IT revolves around leveraging both on-premises and cloud architectures in a hybrid model. Hybrid clouds and hybrid on-premises/cloud architectures deliver the most realistic and robust technology capabilities for businesses today.

This document will continue to be updated to reflect the latest deployment guidance and recommendations. For more information about AWS, consult the extensive online documentation available from the *AWS website*. Additional information about HPE Nimble Storage arrays can be found in the documentation section of *HPE InfoSight*.

Version History

Version	Release Date	Description
1.0	October 2017	Initial release