Microsoft Exchange Server 2010 in the AWS Cloud: Planning & Implementation Guide

Shankar Sivadasan, Ulf Schoo

July 2013

(Please consult **http://aws.amazon.com/whitepapers/** for the latest version of this whitepaper.)

Table of Contents

ABSTRACT 3
ABOUT THE GUIDE 3
BEFORE YOU GET STARTED 3
PLANNING YOUR EXCHANGE SERVER DEPLOYMENT IN THE AWS CLOUD 4
Stage 1: Produce a Candidate Design 4
Processor Sizing 4
Memory Sizing 8
Storage Sizing 8
Log Replication Sizing 11
Network Traffic Sizing 11
Stage 2: Test the Storage Components 12
Stage 3: Verify Server Performance 14
DESIGNING FOR SITE RESILIENCY AND HIGH AVAILABILITY 15
Regions 15
Availability Zones 15
Active Directory Site Design 15
Cross-Region Deployment 16
Client Access Load Balancing 17
CONFIGURING FOR HIGH AVAILABILITY 18
ADDITIONAL DEPLOYMENT BEST PRACTICES 21
Instance Configuration 21
Monitoring 22
Patch Management 22
Message Hygiene 22
Network Security 23
Backup Options 23
SAMPLE DEPLOYMENT SCENARIO 24
Small Business Deployment Scenario (250 mailboxes) 24
Scenario Details 24
Solution Overview 24
Architecture 27
CONCLUSION 28
FURTHER READING 28
APPENDIX 30

Abstract

Amazon Web Services (AWS) provides a comprehensive set of services and tools for deploying Microsoft Windows-based workloads, including Microsoft Exchange Server 2010, on a reliable and secure cloud infrastructure.

Exchange Server is one of the most mission-critical messaging platforms in the enterprise today. It provides email, scheduling, and tools for custom collaboration and messaging-service applications across the entire organization. AWS cloud provides a suite of infrastructure services and features that enable you to deploy Exchange Server in a highly available, affordable and fault-tolerant manner. By deploying a high-availability and site-resilient Exchange Server architecture in the AWS cloud, customers can leverage the powerful functionality of Exchange Server along with the flexibility and security of AWS.

This guide targets Exchange Server IT administrators and deployment engineers. It discusses the planning and deployment tools that you are already familiar with and discusses how to use these tools in the context of AWS environment. After reading it, you should have a good understanding of the architectural considerations and steps involved to plan and deploy an Exchange Server 2010–based messaging service in the AWS cloud.

About the Guide

This planning and implementation guide discusses planning topics, architectural considerations, and configuration steps relevant before and after launching the necessary AWS services such as Amazon Elastic Compute Cloud (Amazon EC2) and Amazon Virtual Private Cloud (Amazon VPC) to run a high-availability and site-resilient Exchange architecture.

This guide discusses the planning and architectural considerations in the context of a sample deployment scenario. The sample scenario highlights a small/medium business (SMB) deployment with up to 250 mailboxes, each mailbox providing up to 5 GiB of mailbox storage and up to 5 GiB of personal archive mailbox storage.

With this guide, we also provide you with sample AWS CloudFormation templates that are designed to help you deploy the necessary and correctly configured infrastructure for the sample scenario in a repeatable and reliable manner.

Note: You can deploy Exchange Server and many other Windows server application licenses in the AWS cloud using the Microsoft License Mobility through Software Assurance program.

Before You Get Started

Implementing Exchange Server in the AWS cloud is an advanced topic. If you are new to AWS, see the Getting Started section of the AWS documentation at *http://docs.aws.amazon.com/gettingstarted/latest/awsgsg-intro/intro.html*. In addition, you should already be familiar with the following topics:

- Amazon EC2 and Amazon VPC
- Windows Server 2008 R2
- Windows Server Active Directory and DNS
- Windows Server Failover Clustering (WSFC)

- Microsoft Exchange Server

This document does not discuss general Exchange Server installation and software configuration tasks, but focuses on infrastructure configuration topics that require careful consideration when you are planning and deploying Exchange Server in the AWS cloud. For general Exchange Server software configuration guidance and best practices, consult the Microsoft product documentation.

Planning your Exchange Server Deployment in the AWS Cloud

The methodology and tools you use to design an Exchange Server deployment are the same whether the target platform is physical, virtual, or a cloud-based infrastructure. Always size Exchange Server deployments as though you plan to install them on dedicated physical hardware. Make sure, however, that the design takes into consideration the performance characteristics of a shared computing environment.

Planning for a production Exchange Server deployment includes the following stages:

1. Produce one or more candidate designs using the **Exchange Server Role Requirements Calculator**.
2. Test the storage component of each candidate design to verify that you've attained the required input/output operations per second (IOPS) within the required thresholds for latency using the Exchange Server version of **Microsoft Jetstress**.
3. Verify that each candidate design passes the **Microsoft Exchange Load Generator 2010** storage-performance validation test. Compare client-access performance data against established baselines for acceptable latency.

Stage 1: Produce a Candidate Design

Much of the capacity planning for a new Exchange Server deployment design centers on the Exchange Server Mailbox role. The other Exchange Server roles (for example, Client Access Server or Hub Transport Server) are generally an output of the quantity and size of Exchange Server Mailbox servers required to satisfy the target usage scenario. With that in mind, the first planning tool we usually consult in any new Exchange Server deployment is the Exchange 2010 Server Role Requirements Calculator. At this writing, the current version of the calculator is V20.8.

To download the calculator, go to v20.8 of the Exchange 2010 Server Role Requirements Calculator at *http://gallery.technet.microsoft.com/office/Exchange-2010-Mailbox-Server-Role*.

The following sections address specific areas of the calculator as they pertain to the AWS cloud infrastructure. Configure any areas that we do not mention as though your Exchange Server environment were to be deployed using dedicated hardware.

Processor Sizing

The guest operating system gives us some visibility into the processor specifications of the virtualization host's physical hardware. The actual performance of the virtual CPU type and cores presented to the virtual machine will vary from the

dedicated physical equivalent, however, because the underlying hypervisor controls the scheduling of its tasks. To produce a more quantifiable performance metric, AWS provides the Elastic Compute Unit (ECU) baseline. Each instance type has one or more virtual cores, each generally performing at one or more ECUs.

For information, see the definition of an ECU in the Amazon EC2 General FAQ at *http://aws.amazon.com/ec2/faqs/*.

In the Mailbox Server Role Requirements Calculator spreadsheet, three fields on the **Input** tab pertain to processor performance. These inputs determine the following:

- Whether you should deduct an arbitrary overhead percentage from the resulting processor-performance calculation based on the cost of virtualization
- The number of processor cores that will be made available to the virtual machine
- The optional SPECint2006 performance rating of the target platform onto which the Mailbox role will be deployed

Because you might deploy cloud-based virtual machines onto various hardware platforms with various specifications, you can't always reliably apply a platform-specific baseline such as SPECint2006. For this reason, and because the target physical platform might not have been evaluated by SPEC, the Microsoft Exchange Server team provides the Exchange Processor Query Tool. At this writing, the current version of the tool is v1.1. (To download the tool, go to v1.1 of the Exchange Processor Query Tool at *http://gallery.technet.microsoft.com/Exchange-Processor-Query-b06748a5*.) The tool takes a particular processor number as its input. Then it searches the SPECint2006 baseline for systems that include that processor number, and provides as an output an average performance rating for the specified processor.

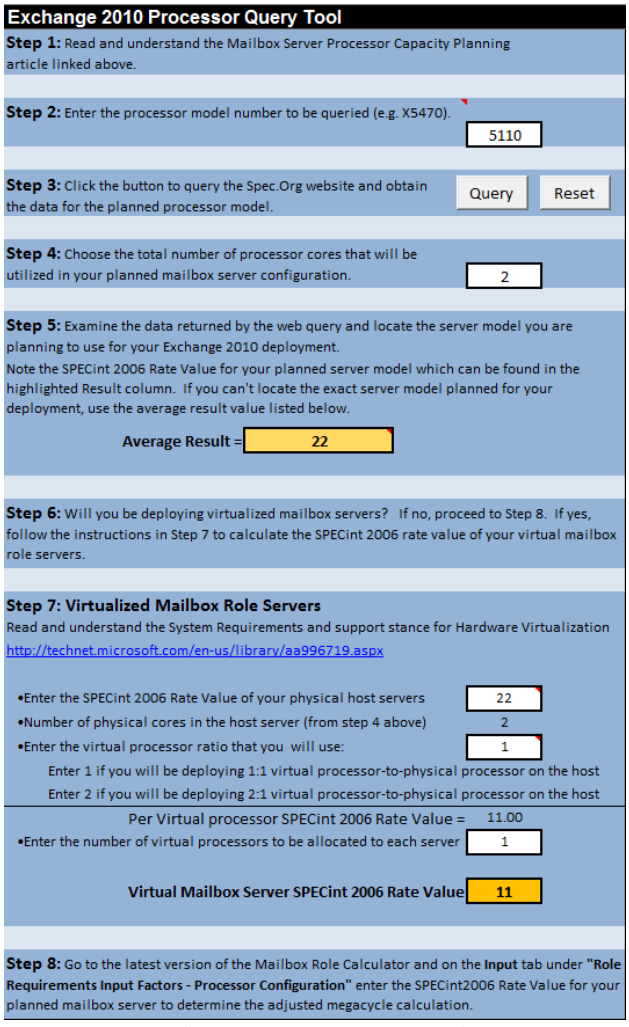


Figure 1: Exchange Processor Query Tool

To come up with **SpecInt2006 Rate Value**, we recommend that you use an approximation for your closest matching processor. For example, you could input the processor number 5110, which will output the average rating of 22 for dual-core configuration, with a per-core rating of 11, as shown in the Figure 1.

To work with the output from the Exchange Processor Query Tool:

1. Set the **SPECint2006 Rate** value to the number of virtual cores (vCPU) multiplied by the number of ECUs per virtual core (ECU/vCPU). Multiply the resulting product by the generic SPECint2006 rate for an ECU (e.g., 11 in our preceding example). Expressed as a formula this looks like this:

   (vCPU*(ECU/vCPU))*SPECint2006 Rate Value

   Example 1:

   A memory-optimized extra-large instance (m2.xlarge),

like the non EBS-optimized instance we use in our deployment scenario later in this paper, is defined as having two virtual cores (vCPUs) and 6.5 ECUs. The result of this calculation is a SPECint2006 rate of **71.5** ((2*(6.5/2))*11).

Example 2:

A general-purpose extra-large instance (m1.xlarge), like the EBS-optimized instance we use in our deployment scenario, is defined as having four virtual cores (vCPUs) and 8 ECUs. The result of this calculation is a SPECInt2006 rate of **88** ((4*(8/4))*11).

2. Set the **Processor Cores/Server** value to the number of virtual cores provided with the target AWS instance type.

Server Configuration	Processor Cores / Server	SPECint2006 Rate Value
Mailbox Servers	2	71.5

Figure 2: Mailbox Server Values for an m2.xlarge Instance Type

3. The resulting ECU SPECint2006 equivalent rating factors in virtualization overhead, so you set the Server Role **Virtualization** value on the **Input** tab of the Server Role Requirements Calculator to **No**, as shown in Figure 2:

Exchange Environment Configuration	Value
Global Catalog Server Architecture	64-bit
Server Multi-Role Configuration (MBX+CAS+HT)	Yes
Server Role Virtualization	No

Figure 3: Server Role Virtualization Value Set to No

Based on general guidelines:

- In a standalone deployment, the Mailbox-role processor usage should not exceed 75 percent.
- In a multi-role server configuration, processor usage should not exceed 35 percent.
- For Database Availability Group (DAG) members, standalone Mailbox-role processor usage should not exceed 80 percent.
- For DAG members, in a multi-role server configuration, processor usage should not exceed 40 percent after a single or double member failure.

You might need to deploy more Exchange Server Mailbox role instances, or choose a more suitable instance type, to reduce the estimated CPU utilization to within the recommended thresholds.

Server Configuration	/ Primary Datacenter Server (Single Failure)
Recommended RAM Configuration	16 GB
Server Total Available Adjusted Megacycles	12710
Mailbox Role CPU Megacycle Requirements	2750
Mailbox Role CPU Utilization	22%
Possible Storage Architecture	RAID

Figure 4: Mailbox Role CPU Utilization on the Role Requirements Tab of the Exchange 2010 Mailbox Server Role Requirements Calculator

Role-Specific Processor Sizing Considerations

Initial sizing for all other Exchange Server roles is typically based on a ratio of role processor cores to Mailbox role processor cores:

- Mailbox : Hub Transport =
  - 7:1 (with no antivirus scanning on Hub Transport server)
  - 5:1 (with antivirus scanning on Hub Transport server)
- Mailbox : Client Access = 4:3
- Mailbox : Client Access and Hub Transport Combined Role = 1:1
- Mailbox : Active Directory Global Catalog =
  - 4:1 (32-bit Global Catalog servers)
  - 8:1 (64-bit Global Catalog servers)

The Server Role Requirements Calculator determines the required megacycles for the Mailbox role based on your inputs and divides that by the megacycles provided by the processor configuration you specified on the **Input** tab. The ratio results appear in the **Processor Core Ratio Requirements** field on the **Role Requirements** tab. If you chose a multi-role configuration, only the **Number of Mailbox Cores Required to Support Activated Databases** and **Recommended Minimum Number of Global Catalog Cores** values are populated, because the Hub Transport and Client Access roles share the same cores as the Mailbox role.

Processor Core Ratio Requirements	/ Datacenter 1	/ Datacenter 2
Number of Mailbox Cores Required to Support Activated Databases	6	6
Recommended Minimum Number of Hub Transport Cores	2	2
Recommended Minimum Number of Client Access Cores	5	5
Recommended Minimum Number of Global Catalog Cores	1	1

Figure 5: Processor Core Requirements

The Hub Transport and Edge Transport roles are supported with a single core. All other Exchange Server roles require a minimum of two cores. Always defer to the minimum system requirements if your calculations produce lower results. Amazon Web Services currently does not offer any single core AWS instance types that include the Exchange Server minimum of 4 GB of RAM, so all instance types used for Exchange Server have at least two virtual cores.

Choose an AWS instance type for each Exchange Server based on the roles it will provide and according to the initial sizing ratios above. Consider redundancy and availability, too: While a single Global Catalog server might meet the processor core requirement, according to Microsoft recommendations, you should have at least two Global Catalog servers per Active Directory site.

Do not base processor sizing for the Edge Transport role on Mailbox core ratios, but rather on estimated peak workload. For more information, see the Edge Transport Server topic on Microsoft Technet at *http://technet.microsoft.com/en-us/library/dd346701%28v=exchg.141%29.aspx#Edge*.

Unified Messaging role servers should generally use a minimum of four cores. If you will be using Voice Mail Preview, increase this minimum to eight cores. Microsoft does not recommend combining the Unified Messaging role with other Exchange Server roles.

To avoid excessive context switching, Microsoft recommends that individual role servers not exceed 12 cores, and that multi-role servers not exceed 24 cores. AWS recommends that you scale your Exchange Server deployment outward, rather than upward.

Memory Sizing

Exchange Server memory requirements start with a minimum of 4 GB for all roles. The minimum memory requirement for multi-role Exchange servers is 8 GB (the recommended maximum is 4GB plus 3-30 MB per mailbox). The **Recommended RAM Configuration** output value on the **Role Requirements** tab of the Server Role Requirements Calculator continues upward from 8 GB based on the target quantity and size of mailboxes that you specify on the **Input** tab.

AWS offers a number of instances types with different memory configuration options. The general best practice is to pick the instance type with exact or slightly lower memory configuration. For example, pick m1.large if you need 8GB of memory. You may want to switch to a higher memory configuration instance type (For example, m1.xlarge) to help reduce the risk of reduced database cache size which might result in higher IOPS required for a given target mailbox-usage scenario.

Server Configuration	/ Primary Datacenter Server (Single Failure)
Recommended RAM Configuration	8 GB

Figure 6: Recommended RAM Configuration on the Role Requirements Tab

Role-Specific Memory Sizing Considerations

All Exchange Server roles require a minimum of 4 GB of RAM. You should configure multi-role servers with a minimum of 8 GB of RAM.

While Microsoft documentation indicates support for the Client Access role with a minimum of 4 GB of RAM, you should use the larger of 8 GB of RAM or 2 GB RAM per core for optimal performance. These guidelines also apply to servers that combine the Client Access and Hub Transport roles.

Configure Hub Transport, Edge Transport, and Unified Messaging role servers with 4 GB to 8 GB of RAM.

Configure Active Directory Global Catalog servers with enough memory to cache the entire Active Directory database (NTDS.DIT) in RAM. This configuration is required to support the 8:1 Mailbox to 64-bit Global Catalog server processor-to-core ratio.

Storage Sizing

The total storage capacity required to support the target deployment scenario is the result of a variety of parameters, including primary and archive mailbox quantity and size, other mailbox usage profile parameters, database storage overhead, logical unit number (LUN) free space percentage, the number of copies for each database (DAGs), and others.

The AWS equivalent of a LUN is an Amazon Elastic Block Storage (Amazon EBS) volume. You can very quickly provision or decommission Amazon EBS volumes to adjust to increasing or decreasing storage demand. Amazon EBS volumes appear to the guest operating system as SCSI targets attached to a virtual SCSI controller. At this writing, you can provision up to 1 TB per Amazon EBS volume.

You might need to enter a custom maximum database size on the **Input** tab of the Server Role Requirements Calculator to avoid having the resulting configuration include LUNs that exceed the maximum size of an Amazon EBS volume, as shown in Figure 7.

Database Configuration	Value
Maximum Database Size Configuration	Custom
Maximum Database Size (GB)	900

Figure 7: Custom Maximum Database Size

Each Amazon EBS volume attaches to one of 26 possible mount points on the instance: */dev/sda1* and *xvdb* through *xvdz*. The operating system root volume is mounted at */dev/sda1*. You can attach additional Amazon EBS volumes or host-based *instance storage* to the 25 remaining mount points.

You might need to add Mailbox role instances to avoid having the resulting configuration exceed the maximum number of EBS volume mount points per instance.

Exchange Environment Configuration	Value
Global Catalog Server Architecture	64-bit
Server Multi-Role Configuration (MBX+CAS+HT)	Yes
Server Role Virtualization	No
High Availability Deployment	Yes
Number of Mailbox Servers Hosting Active Mailboxes / DAG (Primary Datacenter)	4
Number of Database Availability Groups	1

Figure 8: Number of Mailbox Servers Hosting Active Mailboxes

While you might be able to attach as many Amazon EBS volumes as supported by the operating system virtualization drivers, the instance might not be able to access each of these volumes at the required IOPS or within acceptable thresholds for latency. In a shared compute and storage environment, a variety of factors affects storage performance for a given instance.

Each instance shares storage and general network traffic across all elastic network interfaces (ENI) attached to the instance. Each instance also shares an underlying host physical network interface.

For our calculations, we can assume that standard Amazon EBS volumes perform at an average of approximately 100 IOPS. As an alternative, you can provision Amazon EBS volumes with a specific IOPS target of up to 4000 IOPS per volume. In addition, you can launch some Amazon EBS-optimized instances, providing a dedicated throughput between your Amazon EC2 instance and Amazon EBS storage between 500 Mbps and 1000 Mbps.

You might need to use the Server Role Requirements Calculator to produce several possible configurations and validate the performance of each using the Exchange Server version of the Microsoft Jetstress tool.

Instance storage is ephemeral; the configured instance storage appears as a new disk each time the instance is stopped and started again. All data located on instance storage is lost after the instance is stopped. Instance storage is included in the runtime cost of the instance. Instance storage is not included with all instance types and, for those that do include it, the capacity and number of instance storage disks is preset based on the instance type. For information, see Amazon EC2 Instance Types at *http://aws.amazon.com/ec2/instance-types/*.

Instance store volumes are ideal for temporary storage of information that changes frequently, such as operating system paging and temporary file storage. Instance storage generally performs faster than Amazon EBS because it is local to the

instance. To help provide a reliable operating environment, however, you require an automated mechanism to initialize, partition, format, and mount the instance storage as well as reconfigure the operating system to use the new storage on each instance start. The next section includes an example of how to provision an instance store ephemeral volume for use as the OS dedicated paging volume in the AWS CloudFormation templates.

Here are some specific recommendations to keep in mind when you plan the quantity and size of volumes that your Exchange Server instances will require for the volumes that will *not* host Exchange Server database files:

- Choose a root volume size that will provide sufficient capacity for patching and diagnostic logging. Microsoft recommends a minimum OS partition for several products, including Exchange Server, of 80 GiB. You can change the default root-volume size when launching the instance either through the AWS Management Console or by using AWS CloudFormation templates.
- To avoid disk contention between the operating system and Exchange Server, install Exchange Server to a path on an Amazon EBS volume other than the partition holding the OS. This will also help protect the operating system if Exchange Server data (or other application data stored on the same volume) overruns the available capacity of the volume.
- Microsoft recommends that the operating system paging file and temporary files path for Exchange Server be located on volumes separate from the operating system files for optimal performance. Instance storage is an excellent candidate for both of these types of data.

Here are some recommendations to keep in mind when you plan the quantity and size of volumes to store Exchange Server database files:

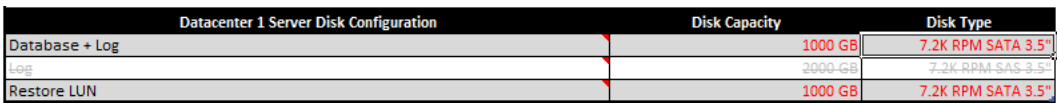
- There is an indirect relationship between mailbox size quota and required storage IOPS. If you use a smaller mailbox-size quota, more mailboxes are co-located on the same volume and more IOPS are required for each volume. If you choose a larger mailbox size quota, fewer mailboxes are co-located on the same volume, and fewer IOPS are required per volume. For example, an Amazon EBS volume with 100 500-MB mailboxes has a higher IOPS requirement than one with 25 2-GB mailboxes.

The Server Role Requirements Calculator only allows JBOD configurations where there are three or more copies of a database in a DAG configuration. For all other configurations, the tool suggests using RAID. All Amazon EBS volumes are mirrored on the backend. This, combined with regular backups, meets the data protection requirement.

Storage Options	Value
Consider Storage Designs Utilizing JBOD (if applicable)	Yes

Figure 9: Storage Options

For **Datacenter # Server Disk Type**, select 7.2K RPM SATA 3.5" for the **Disk Type** value and set the **Disk Capacity** value to the maximum size Amazon EBS volume, 1000 GB. This disk type selection most closely resembles the performance characteristics of a standard Amazon EBS volume.



Datacenter 1 Server Disk Configuration	Disk Capacity	Disk Type
Database + Log	1000 GB	7.2K RPM SATA 3.5"
Log	2000 GB	7.2K RPM SAS 3.5"
Restore LUN	1000 GB	7.2K RPM SATA 3.5"

Figure 10: Disk Configuration

The **Backup Methodology** value must be set to either Software VSS Backup/Restore or Exchange Native Data Protection.

Note: Microsoft does not recommend that you use only Exchange Native Data Protection (no backups/replication only) unless you have configured three or more copies of each database within a Database Availability Group.

Backup Configuration	Value
Backup Methodology	Software VSS Backup/Restore

Figure 11: Backup Configuration

Role-Specific Storage Sizing Considerations

You generally base your storage and monitoring plan Exchange Server roles other than the Mailbox role on your core volume layout (that is, root, paging, temp, and Exchange Server installation). After you have established your baseline volume sizes, use continuous monitoring of available disk capacity to determine if you need to adjust the volume sizes.

Exchange Server requires at least 1.2 GB on the installation volume, an additional 500 MB for each Unified Messaging language pack that you plan to install, 200 MB on the root volume, and 500 MB on the volume where the queue database for the Hub Transport or Edge Transport roles is stored.

Note: If Exchange Server runs low on disk space, it will begin refusing connections due to *back pressure*.

Log Replication Sizing

Amazon Availability Zones within the same region connect through high-speed links. Start to plan your Exchange Server DAG transaction log replication by choosing **Fast Ethernet** for the **Network Link Type** on the **Input** tab of the Server Role Requirements Calculator. Leave **Network Link Latency** at the default of 50.00 or run your own tests between temporary instances to establish a more precise value.

Network Configuration	Value
Network Link Type	Fast Ethernet
Network Link Latency (ms)	50.00

Figure 12: Network Configuration

Network Traffic Sizing

You can use the Exchange Client Network Bandwidth Calculator to predict the network bandwidth requirements for a specific set of clients. The prediction algorithms used within this calculator are entirely new and are derived from significant testing and observation. At this writing, the current version of this tool is Public.0.48BETA4.

To download the tool, go to the Exchange Client Network Bandwidth Calculator page at *http://gallery.technet.microsoft.com/office/Exchange-Client-Network-8af1bf00*.

Stage 2: Test the Storage Components

Before you place any Exchange Server Mailbox role design into production, you *must* test the storage subsystem to ensure that the design supports the required IOPS within acceptable thresholds for latency. Storage subsystem performance is critical to an acceptable Exchange Server client experience.

Microsoft Exchange Server Jetstress 2010 is a free tool provided by Microsoft's Exchange Server team to simulate realistic Exchange Server I/O patterns against one or more test databases. The tool creates the specified test databases on the target volumes and performs simulated transactions for client access, background maintenance, and transaction log replication. You can customize the behavior of the tool through the GUI and the configuration XML file. Throughout the simulation, the tool collects values for a variety of Exchange Server performance counters and, at the conclusion of the simulation, compares them to acceptable thresholds for latency for database and transaction log operations. The current version of the tool is 14.01.0225.017.

To download the tool, go to the Microsoft Exchange Server Jetstress page at *http://www.microsoft.com/en-us/download/details.aspx?id=4167*. You can also download the tool from the **Support** tab of the EC2ConfigService Setting app.

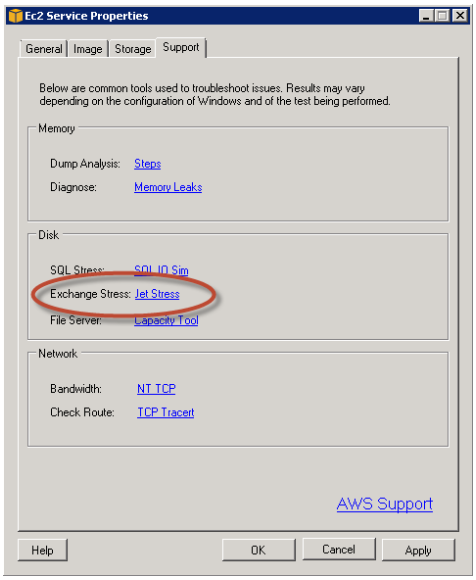


Figure 13: Download Jetstress

Before you install Jetstress, read the Jetstress Field Guide at *http://gallery.technet.microsoft.com/Jetstress-Field-Guide-1602d64c*. This guide includes important information how to configure and use the tool, and how to interpret the results. The current version of the guide is v1.0.0.16.

Important: Do not install Exchange Server before you install Jetstress. Jetstress will interfere with database performance counters if you install it after Exchange Server.

Consider the following recommendations, when you install and run Jetstress in AWS:

- Deploy each Exchange Server Mailbox role instance and its associated Amazon EBS volumes for your intended design. Tag associated volumes for each instance to help provide easy identification through the AWS Management Console.
- After you install Jetstress, ensure that the version of the performance counter and ESE database engine files you installed match exactly the versions (including patch level) to be used in your production Exchange Server deployment.
- Prior to database initialization, partition and format each Amazon EBS volume that will contain Exchange Server databases, using NTFS and a 64 K allocation unit size. If a volume is already formatted, you can check the current allocation unit size using the following command: `fsutil fsinfo ntfsinfo <letter:>`. Note the value for **Bytes Per Cluster** in the output; this value should be **65536**. If the value for this parameter is any other number, the best option is to reformat the volume with the appropriate allocation unit size.

```
C:\>fsutil fsinfo ntfsinfo f:
NTFS Volume Serial Number :       0x6a8ef03a8eeffbff
Version :                         3.1
Number Sectors :                  0x00000000027fe7ff
Total Clusters :                  0x0000000000004ffcf
Free Clusters  :                  0x0000000000027a3d
Total Reserved :                  0x0000000000000000
Bytes Per Sector  :               512
Bytes Per Physical Sector :       <Not Supported>
Bytes Per Cluster :               65536
Bytes Per FileRecord Segment    : 1024
Clusters Per FileRecord Segment : 0
Mft Valid Data Length :           0x0000000000200000
Mft Start Lcn  :                  0x000000000000c000
Mft2 Start Lcn :                  0x0000000000000001
Mft Zone Start :                  0x000000000000c020
Mft Zone End   :                  0x000000000000cca0
RM Identifier:        47A7F156-041E-11E1-AB44-FE285F0720CF
```

Figure 14: Output From the fsutil fsinfo ntfsinfo <drive letter:> Command

- When you deploy Exchange Server into a shared compute and storage environment, consider taking the following steps:
    - Provide dedicated throughput between your Amazon EC2 instance and Amazon EBS storage (500 Mbps or 1000 Mbps) by selecting an EBS-optimized instance type that meets your requirements for processor performance and memory configuration.

Instance Type	Dedicated Throughput
m1.large	500 Mbps
m1.xlarge	1000 Mbps
m2.2xlarge	500 Mbps
m2.4xlarge	1000 Mbps
m3.xlarge	500 Mbps
m3.2xlarge	1000 Mbps
c1.xlarge	1000 Mbps

Figure 15: EBS-optimized instance types

  - Provision Amazon EBS volumes with a specific IOPS target. Provisioned IOPS (PIOPS) supports up to 4000 IOPS per volume. You can stripe (RAID0) multiple volumes to deliver thousands of IOPS to your application. A provisioned IOPS volume must be at least 10 GB in size and follow the recommended IOPS:GB ratio of 10:1. For example, a volume with 4000 IOPS must be at least 400 GB.
  - As with standard Amazon EBS volumes, there is first-read penalty when you access the data. Performance is restored after you access the data once. To help obtain maximum performance consistency with new database volumes, we generally recommend that you pre-initialize the volume using a *full* format, which performs a surface scan to access every block of the target volume. Ensure that the **Quick Format** option is deselected when you format a volume using the Disk Management snap-in in Windows. **Note:** Depending on the instance type, a full-format of a large Amazon EBS volume can take many hours.
- Run Jetstress on all instances that will host the Exchange Server Mailbox role.
  Note: The selected instance type has a direct effect on the number of Amazon EBS volumes an instance can support at the required IOPS and within acceptable thresholds for latency. You might find that you can achieve or exceed the required IOPS but fail to do so within acceptable thresholds for latency. If so, you must either reduce the number of volumes per instance or change the instance type.
- Temporarily upgrade the instance type to reduce the amount of time it takes for full Amazon EBS volume pre-initialization and Jetstress test database initialization. After you format the target volumes and Jetstress test databases initialization, shut down the instance and change the instance type back to match your target design.

Stage 3: Verify Server Performance

Microsoft Exchange Load Generator 2010 simulates the server workload that is generated by user interaction with various messaging client software. It is a useful tool for server administrators or messaging deployment engineers who are sizing servers and validating deployment plans. Exchange Load Generator helps you determine whether each of your servers can handle the load that they are intended to carry. You can also use Exchange Load Generator to help validate the overall solution. Because Exchange Load Generator simulates client requests, you can also validate the effect of server-side solutions such as archiving, antivirus, or anti-spam products.

We recommend that you run Exchange Load Generator against your candidate Exchange Server deployment to validate its ability to handle the anticipated client load. Exchange Load Generator affects the performance of all systems involved, so you should run the tool after you have fully configured Exchange Server for your target deployment and before you introduce production user mailboxes or production data.

Exchange Load Generator uses many simulated user mailboxes to create the server workload. Because mailboxes must be part of a domain user's account, the Exchange Load Generator tool creates many domain user accounts to support these user mailboxes. Exchange Load Generator requires that the password associated with these domain accounts be the same.

You cannot use Exchange Load Generator for storage design validation, because it does not replicate the correct IOPS per user. For storage design validation, use Jetstress.

Exchange Load Generator does not provide a complete picture of the user experience, and you should not interpret its results as if it did. Exchange Load Generator is a community-supported tool and is provided by Microsoft as-is.

Designing for Site Resiliency and High Availability

Amazon EC2 is available in multiple geographic locations. One of the key elements to achieving greater fault tolerance and site resiliency is to distribute your application geographically. If a single AWS datacenter fails for any reason, you can protect your application by running it simultaneously in a geographically distant datacenter. This section describes how to design your highly available and site-resilient Exchange Server environment so that it replicates your resources across locations to take advantage of the AWS cloud.

Regions

You can provision instances in multiple geographic locations called regions. There are 9 regions available around the world today. You can launch Amazon EC2 instances in these regions so your instances are closer to your customers. For example, you might want to launch instances in Europe to be closer to your European customers or to help you meet your legal requirements.

Availability Zones

Within each Region are Availability Zones (AZs). Availability Zones are distinct locations that are engineered to be insulated from failures in other Availability Zones and provide inexpensive, low latency network connectivity to other Availability Zones in the same Region. By launching instances in separate Availability Zones, you can protect your applications from a failure (unlikely as it might be) that affects an entire zone. From an Exchange Server design perspective, think of each Availability Zone as a separate data center. To help achieve high availability, design your Exchange Server deployment to span two or more Availability Zones.

If your business needs require it, you can design your Exchange Server deployment to span multiple regions, as well. This, however, is more complex and requires additional networking and security, as well as more thorough testing and continuous monitoring.

Active Directory Site Design

The design of your Active Directory site topology in Amazon Web Services determines the high-availability service failover behavior, the distribution of load-balanced traffic, and the number of servers required for your design. You can stretch an Active Directory site across multiple Availability Zones, but we do *not* recommend that you design Active Directory sites to span multiple AWS regions. The following section discusses the benefits and drawbacks of different Active Directory site designs.

Spanning Active Directory Sites across Availability Zones

The benefits of spanning Active Directory sites across Availability Zones are as follows:

- Active Directory replication latency is minimal for configuration changes and failovers.
- If an individual role service fails in a particular Availability Zone, Exchange Server roles in that Availability Zone can use services provided by Exchange Server roles in another Availability Zone within the same logical Active Directory site.
- You can configure Exchange EdgeSync so that internally, Exchange Edge Transport and Hub Transport roles located in the same Active Directory site balance the load and failover between one another for message

transport. Externally, the SMTP protocol includes a weight-based mechanism to balance the load and failover between those hostnames configured in DNS as MX records.

The drawbacks of spanning Active Directory sites across Availability Zones are as follows:

- You can only configure one Client Access Array object per Active Directory site. All Client Access servers will belong to the same Client Access Array and consequently share the same FQDN (fully qualified domain name). This might result in a client using a Client Access role server in one Availability Zone to access their mailbox, which is located in a database that is active on a Mailbox role server in a different Availability Zone.
- Active Directory and Exchange Server use high-speed protocols for communication between all servers, which results in more inter-Availability Zone network traffic. This is subject to the Regional Data Transfer monthly charge rate. In some cases, you can manually configure servers in each Availability Zone to localize service access.

Confining Each Active Directory Site to a Single Availability Zone

The benefits of confining each Active Directory site to a single Availability Zones are as follows:

- Each Availability Zone is configured with its own Client Access Array object and associated site-specific FQDN. Client Access servers are members of the Client Access Array for their local Active Directory site. (For extensive information regarding Exchange Server cross-site proxy and redirection behavior, see Understanding Proxying and Redirection at *http://technet.microsoft.com/en-us/library/bb310763%28v=exchg.141%29.aspx*.
- You can stretch a DAG across Active Directory Sites; Microsoft only requires that the latency between DAG members does not exceed 500 ms. If you stretch a DAG, Active Directory will use inter-site transport optimizations to reduce the bandwidth required for replication.

The drawbacks of confining each Active Directory site to a single Availability Zones are as follows:

- You can only configure inter-site Active Directory replication to as low as 15 minutes, so you might have to manually force replication across the Active Directory site topology after you make critical configuration changes, or target specific Domain Controllers when you use Exchange Server cmdlets.
- To manage the single FQDN used for initial service access to Outlook Web Access (OWA) or Exchange Autodiscover, you must either use a Global Traffic Manager or an intelligent DNS service, or plan to manually modify records during a failure if you are using round-robin DNS.
- The Exchange Server Mailbox role requires at least one of each of the following roles in each Active Directory site where you install Exchange Server: Global Catalog, Client Access, and Hub Transport. A Mailbox server in one Availability Zone cannot use services provided by Exchange Server roles in another Active Directory site. For intra-site service failover, you must configure at least two instances of each role.
- Disabling a send connector in one Active Directory site causes the Hub Transport servers in that site to route outbound messages to Hub Transport servers in another site for delivery through that site's enabled send connector. This affects your Regional Data Transfer charges.

Cross-Region Deployment

All communications between regions is across the Internet, so you should use the appropriate encryption methods to protect your data.

If you plan to span a DAG across regions, follow the same recommendations that apply to spanning a DAG across your physical data centers without a dedicated interconnection. Test the configuration to confirm that network latency does not exceed 500 ms, unnecessary failovers do not occur, and that failovers, switchovers, and client access all work as expected.

Data transfer between regions is charged at the Internet data transfer rate for both the sending and the receiving instance. For more information, see Amazon EC2 Pricing - Data Transfer at *http://aws.amazon.com/ec2/pricing.*

Client Access Load Balancing

Using load-balanced client access FQDNs can help provide a near-seamless user experience during failovers or switchovers, but you must meet specific requirements for Exchange Server client access protocols to ensure reliable service.

For an overview of load balancing with Exchange Server, see Understanding Load Balancing in Exchange 2010 at *http://technet.microsoft.com/en-us/library/ff625247%28v=exchg.141%29.aspx*.

For detailed information about protocol-specific requirements for load balancing, see Load Balancing Requirements of Exchange Protocols at *http://technet.microsoft.com/library/ff625248%28exchg.141%29*.

The Exchange Server team has also produced a list of hardware and software load balancers that have been tested and verified to work with Exchange Server client access protocols. For more information, see Exchange Server 2010 load balancer deployment at *http://technet.microsoft.com/en-us/exchange/gg176682.aspx/.*

At this writing, Amazon elastic load balancers cannot balance the load of Exchange Server client access protocols. For information about Amazon elastic load balancing, see the Elastic Load Balancing Developer Guide at *http://docs.aws.amazon.com/ElasticLoadBalancing/latest/DeveloperGuide/Welcome.html.*

At this writing, the list of third-party load balancers that are compatible with AWS and Exchange Server client access protocols include the Riverbed Stingray Traffic Manager, Citrix Netscaler VPX, Loadbalancer.org Enterprise EC2 and F5 Big-IP Virtual Edition for AWS Local Traffic Manger (LTM) and Global Traffic Manager (GTM).

Configuring for High Availability

After you install Exchange Server and perform basic configuration steps (which include configuring send connectors, receive connectors, and changing the default database paths), configure the DAG for high availability of your mailbox databases. Here are some tips on how to configure the DAG to get the most out of deploying Exchange Server in the AWS cloud.

Use two elastic network interfaces on DAG member instances. To isolate replication traffic, assign the second elastic network interface to a different AWS security group, as shown in Figure 16.

(Note that this configuration provides logical separation, but does not provide any additional instance bandwidth, because the instance network quota is shared across all ENIs attached to the instance).

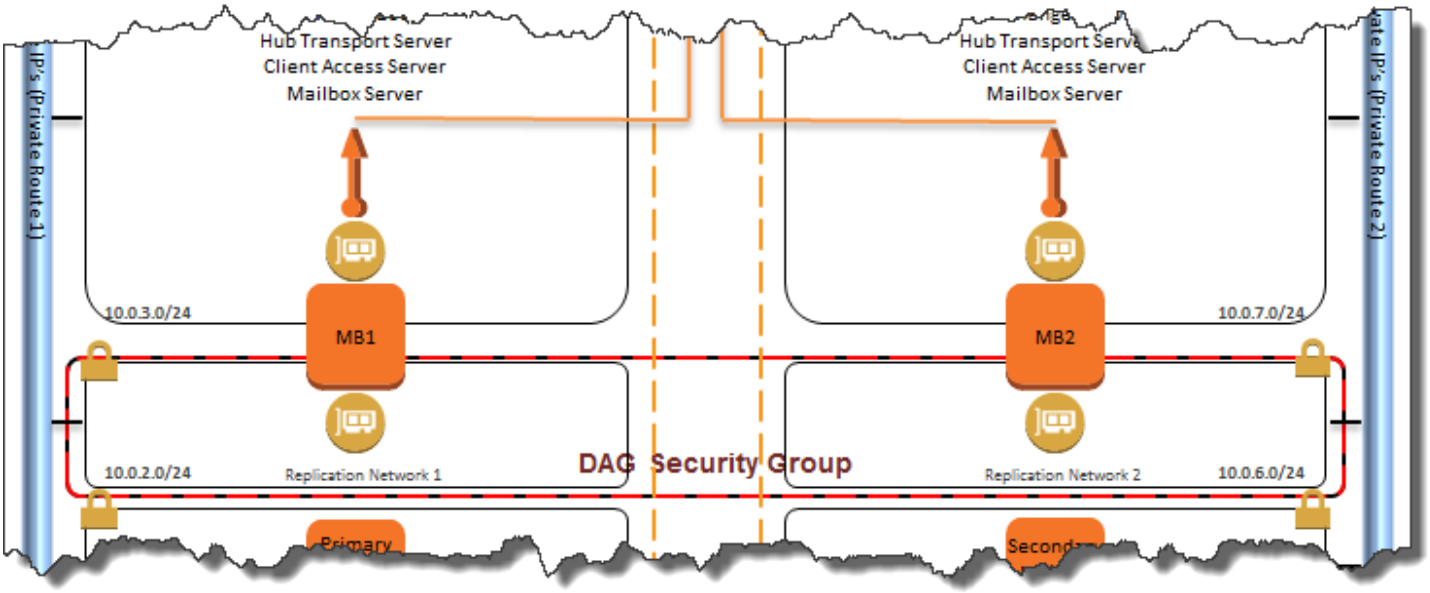


Figure 16: Mailbox Server with Two ENIs and Separate DAG Security Group

Configure the ENIs on the instances hosting the Mailbox servers as follows:

	Primary Private IP	Secondary Private IP	EIP
MB1			
eth0	10.0.3.10	10.0.3.100	<your EIP>
eth1	10.0.2.10		
MB2			
eth0	10.0.7.10	10.0.7.100	<your EIP>
eth1	10.0.6.10		

Figure 17: IP Configuration of the ENIs

Note: Because we configured the instance with two ENIs, you must configure Windows to disable the default route on the second ENI, and ensure that the interface is not configured to register itself with DNS.

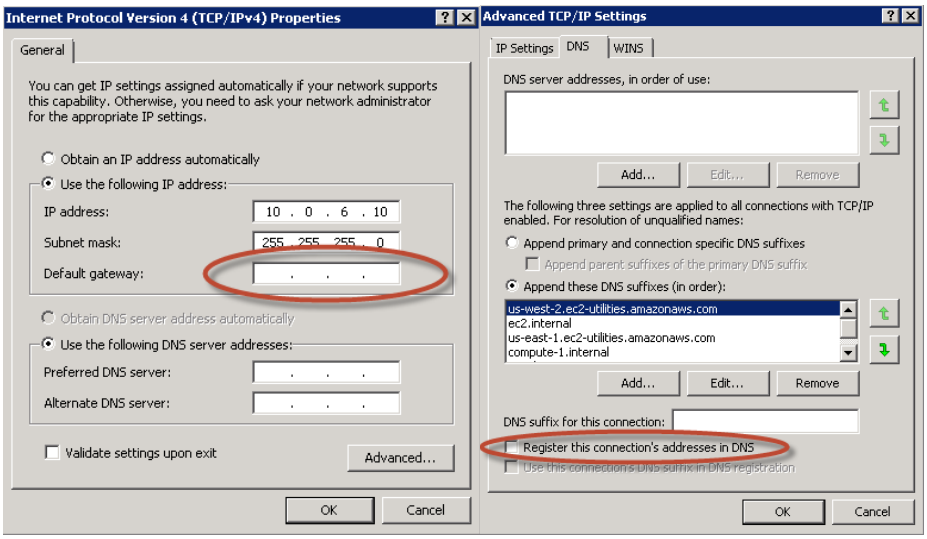


Figure 18: Windows Configuration of the Secondary ENI

Because our sample replication networks are isolated in separate Availability Zones, add static routes for the DAG replication network subnets in each Availability Zone. The following adds a route to the replication connection for gateway 10.0.2.1 using the `netsh` command:

```
netsh interface ipv4 add route 10.0.6.0/24 "Connection Name" 10.0.2.1
```

Note: If you are not using an Exchange Role Server for your file share witness server, you must add the Active Directory Exchange Trusted Subsystem group to the local Administrators group on the target server before you create the DAG.

Before you configure DAG networking, add a secondary private IP address (for example, 10.0.3.100) to the ENIs associated with the MAPI networks in both Availability Zones, as shown in Figure 17.

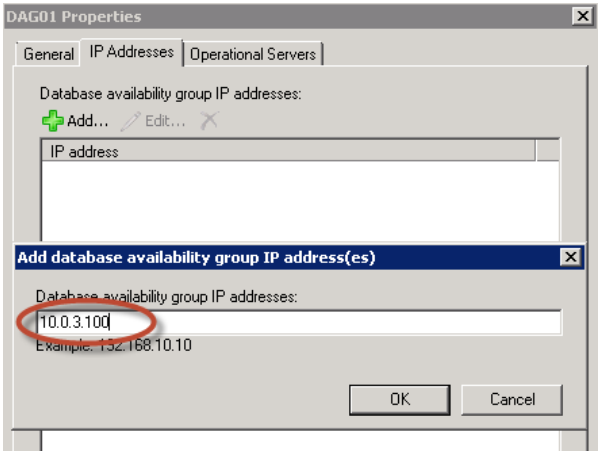


Figure 19: Use a Secondary Private IP Address to Configure the DAG Network.

Disable replication on the MAPI network and enable it on the replication network.

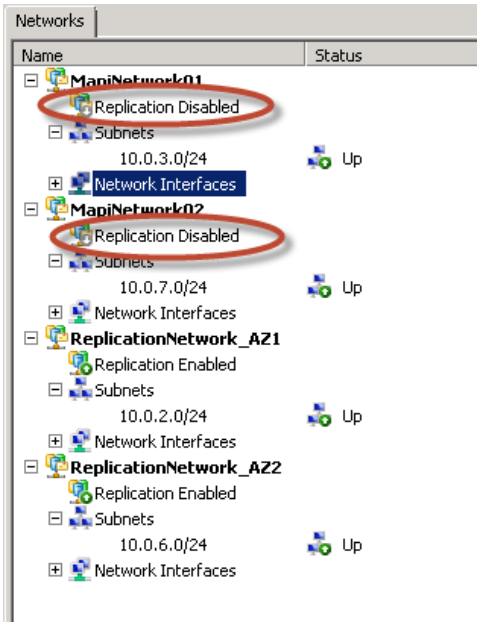


Figure 20: DAG Configuration for Small Business Deployment

After the server is in production, monitor the DAG for unexpected failovers and review the event logs for causes of failovers. If an individual database is failing over, check the health of the volume and check for disk errors in the event logs. If the entire cluster is failing over, you might need to adjust the cluster timeout thresholds using either the `cluster.exe` command or Windows PowerShell.

Note: At this writing, an issue in Windows Failover Cluster causes the cluster to failover when the witness server is unavailable. This does not usually affect any databases, but the server that owns the cluster and the Primary Active Manager will change. You can use the following command to move the cluster back to your preferred node:

```
cluster group "Cluster Group" /moveto:nodename
```

To simplify failover between Availability Zones if all instances in one Availability Zone become unavailable, ensure that your DAG is set to Datacenter Activation Coordination (DAC) mode.

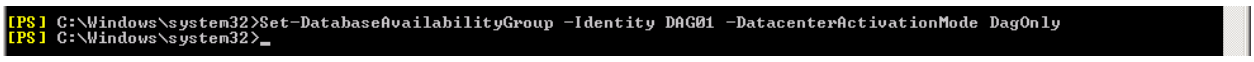


Figure 21: Set DAC Mode.

Configure cross-site silent redirection for Outlook Web App to help provide a seamless client experience after a failover or switchover of databases. Exchange Autodiscover and MAPI will handle redirection of other clients.

For information, see OWA Cross-Site Silent Redirection in Exchange 2010 SP2 at *http://blogs.technet.com/b/exchange/archive/2011/12/12/owa-cross-site-silent-redirection-in-exchange-2010-sp2.aspx*.

In a multiple Availability Zone deployment, consider setting the *PreferredGlobalCatalog* parameter if a Mailbox database operation, such as mounting a new database, fails because of the replication latency that occurs between the configured domain controllers and the preferred global catalog. For information, see the *You cannot create a new Exchange Server 2010 Mailbox database in a multiple domain environment* Microsoft Knowledgebase article at *http://support.microsoft.com/kb/977960.*

```
Set-ADServerSettings –PreferredServer <DC FQDN>
```

Test your failover strategy at all levels to ensure all components behave as expected, both after initial configuration and periodically.

Additional Deployment Best Practices

You might encounter additional deployment best practices in your specific deployment. This section discusses the ones you may be most likely to encounter.

Instance Configuration

You can deploy your Exchange solution stack repeatedly and reliably using AWS CloudFormation in combination with Windows Powershell. The sample AWS CloudFormation templates use standard Windows Server 2008 R2 Amazon Machine Images (AMIs) and then perform, in a scripted fashion, all the necessary configuration tasks. These tasks include the following:

- Renaming the instance
- Creating a forest and domain
- Joining Mailbox Servers to the domain
- Performing network configuration tasks
- Downloading all necessary Exchange installation files

Alternatively, you can pre-configure some of items that are common to all your Exchange servers, like the Exchange installation files, by creating a custom AMI. Creating a custom AMI helps ensure a consistent system baseline and can save you a significant amount of time when you are deploying a larger number of Exchange servers.

For more information, see Creating Your Own Windows AMI at *http://docs.aws.amazon.com/AWSEC2/latest/WindowsGuide/Creating_WinAMI.html.*

In addition, keep the following in mind when you create a custom Windows AMI:

- Be sure any software that you install is compatible with Microsoft Sysprep.
- Do not install Exchange Server in your custom AMI. Exchange Server does not support Microsoft Sysprep.
- Do not join your custom AMI to the Active Directory Domain.

When you launch the instance from your custom AMI, if your chosen instance type supports it, relocate the operating system paging file and temporary files path to instance storage using custom initialization scripts to provide optimal performance for these types of data.

Monitoring

Amazon CloudWatch provides useful statistical data, graphing, and alarms regarding instance performance. You can configure the level of monitoring and you can monitor more than just instances. For example, you can also monitor elastic load balancers, Amazon EBS volumes, and more.

We suggest that you also use an operational monitoring tool with built-in intelligence for monitoring Exchange Server and related products and services. For example, you can combine Microsoft System Center Operations Manager with Amazon CloudWatch, or use a product that can integrate with Amazon CloudWatch, such as the AWS Management Pack for Microsoft System Center, available at *http://aws.typepad.com/aws/2013/05/aws-management-pack-for-microsoft-system-center.html*. These tools can provide invaluable, intelligent alerting and historical performance data with reporting.

Finally, consider tagging instance volumes to make it easy to identify them in your monitoring tools.

For more information, see Tagging Your Amazon EC2 Resources in the AWS documentation at *http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/Using_Tags.html.*

Use the AWS CloudFormation templates provided with this guide to start your proof of concept or production deployment and to have all necessary resources tagged via the template. You can edit the tags in the template to match your naming conventions.

Patch Management

We recommend that you use a patch management tool such as Microsoft Windows Server Update Services or Microsoft System Center Configuration Manager to help provide a consistent operating baseline for your systems, and help reduce your monthly Internet data transfer charge by centralizing patch download and distribution.

If you don't use a patch management tool, configure the Windows Update service through Group Policy to meet your company's patch management policy.

Message Hygiene

Exchange Server provides basic anti-spam filtering using the Edge Transport role or the Hub Transport role with the Edge Transport agents installed. We advise that you also invest in a product at the edge for message hygiene; either a locally installed product or a cloud-based service such as Microsoft Exchange Online Protection.

If you must manage internal message hygiene, make sure that your instance design includes sufficient computing resources to handle the additional load on your Exchange servers (as you would for any operating system monitoring or security agent).

In order to maintain the quality of Amazon EC2 addresses for sending email, we enforce default limits on the amount of email sent from Amazon EC2 accounts. If you want to send larger amounts of email from Amazon EC2, you can apply to have these limits removed from your account by filling out this form at *https://portal.aws.amazon.com/gp/aws/html-forms-controller/contactus/ec2-email-limit-rdns-request.*

If you intend to send email to third parties from Amazon EC2 instances, we also suggest that you provision one or more Elastic IP addresses (EIPs) and provide them to Amazon Web Services through the same form. Amazon Web Services

works with ISPs and Internet anti-SPAM organizations such as Spamhaus to help reduce the chance that email sent from your addresses will be flagged as SPAM.

You can also help avoid having your email flagged as SPAM by assigning a static reverse DNS record to the EIP used to send email. You have the option to provide Amazon Web Services with a reverse DNS record (such as foo.yourcompany.com) to associate with your EIP(s). Note that a corresponding forward DNS record (A Record) pointing to your EIP must exist before we can create your reverse DNS record. **Note:** It may take up to a week before the anti-SPAM organizations approves your EIP(s).

After AWS has confirmed that the elastic IP addresses are approved by the anti-SPAM organizations, go to the Spamhaus.org Blocklist Removal Center at *http://www.spamhaus.org/lookup/*. Look up your elastic IP addresses and request that they be removed from the PBL list. You *must* do this before you send any outbound email, because otherwise your elastic IP addresses might be flagged by other block list services, who you will then need to contact to remove your addresses from their lists as well.

Network Security

Never open RDP up to the entire Internet—not even for testing purposes or temporarily. Always restrict ports and source traffic to the minimum necessary to support the functionality of the application. For information about securing Remote Desktop Gateway, see the *Securing the Microsoft Platform on Amazon Web Services* whitepaper at *http://media.amazonwebservices.com/AWS_Microsoft_Platform_Security.pdf.*

We recommend against providing direct access to Exchange Server Client Access servers for Exchange client access. Instead, consider using a security product such as Microsoft Unified Access Gateway 2010 to provide pre-authentication, filtering, reverse proxying, or HTTP-based load balancing of your client access services.

If your deployment goals and scenarios permit, avoid directly publishing inbound SMTP access to Hub Transport servers that are installed on a multi-role Exchange Server. Instead, use Edge Transport servers, and an edge-based inbound filtering service. At a minimum, install the Edge Transport agents on the Hub Transport servers that will receive email directly from the Internet.

Backup Options

Amazon Machine Images (AMIs) are the basic building blocks of Amazon EC2. An AMI is a template that contains a software configuration (operating system, application server, and applications). You can create an AMI of an Exchange Server as a backup strategy if you also use application-level backup software that uses VSS to provide consistency of Exchange Server data. If you experience data corruption, you can use the application-level backup software to restore the affected data. If you have a catastrophic failure, you can first restore the Exchange Server from the latest AMI, and then restore the latest application-level backup. At this time, we do not recommend using the Amazon EBS Snapshot functionality for your Exchange Server Backup.

Note: Some third-party backup products (such as CommVault Simpana) store backup data on Amazon Simple Storage Service (Amazon S3) for greater durability and at a potentially lower cost than storing backup data on EBS volumes.

For a list of certified Exchange Server backup solutions, consult the Windows Server Catalog at *http://windowsservercatalog.com*.

Sample Deployment Scenario

The following use case scenario provides one example of possible Exchange Server deployments in the AWS cloud. You can easily deploy this scenario by using the AWS CloudFormation templates provided. We designed the templates to provision the necessary networking, compute, and storage resources for the scenario, so that you can focus on configuring Exchange Server to meet your business needs. The source of the sample templates are provided so you can edit them for your specific deployment using the AWS Toolkit for Visual Studio or any other JSON editor.

Note: AWS does not provide installation media for Microsoft software. If you use this guide to set up a test or evaluation environment, you may be able to download a trial version at *http://www.microsoft.com/en-us/download/details.aspx?id=36768*. For a production deployment, you may be able to use your volume licensing software and mobilize the license as described in the Microsoft License Mobility through Software Assurance program.

Small Business Deployment Scenario (250 mailboxes)

This scenario's sample solution is designed to provide cost-optimized high availability and high performance in a site design that stretches across two Amazon EC2 Availability Zones.

Scenario Details

- 250 Mailboxes
- 5 GB Primary Mailbox Quota per User
- 5 GB Archive Mailbox Quota per User
- 100 Messages Sent/Received Per User Per Day
- Average Message Size of 75 KB
- 60% Outlook Anywhere, 30% Exchange ActiveSync, 10% Outlook Web App

Solution Overview

We deploy the sample solution for this scenario in three parts, two of which are scripted via the included AWS CloudFormation templates:

Part 1: Scripted Launch and Configuration of the Virtual Network and Active Directory Infrastructure

This template performs the following tasks:

1. Sets up the virtual network for the Exchange Server deployment in the AWS cloud, including subnets in two Availability Zones
2. Configures private and public routes
3. Launches Windows Server 2008 R2 AMIs and sets up and configures Active Directory and DNS
4. Creates a user who is a member of the Domain Admins group, and an Exchange admin user who is a member of the Enterprise Admins and Schema Admins group
5. Enables administrative ingress and egress into your Amazon VPC via Remote Desktop Gateway and NAT instances
6. Configures Amazon EC2 security groups to control network traffic between the Exchange Servers and Active Directory

To launch the AWS CloudFormation template in the US-West-2 (Oregon) region, click **Launch**. This will load the sample template in the AWS Management Console Wizard.

The sample deployment consists of a single Amazon VPC with ten subnets. The subnets are split up evenly across two Availability Zones, AZ1 and AZ2. Four subnets, two in each AZ, will be *public*, enabling administrative ingress and egress into the environment and inbound SMTP Access. Amazon VPC security groups isolate the network traffic. An Amazon VPC Internet Gateway provides external network access for the four public subnets.

Two Active Directory Domain Controllers are configured as Global Catalog servers (one per Availability Zone). They also serve as internal DNS servers. A single site Active Directory site topology is configured. Additionally, one of the Global Catalog servers will host the file share for the File Share Witness role required for Exchange Server Database Availability Group quorum.

Part 2: Scripted Launch and Configuration of the Exchange Server Multi-Role Servers

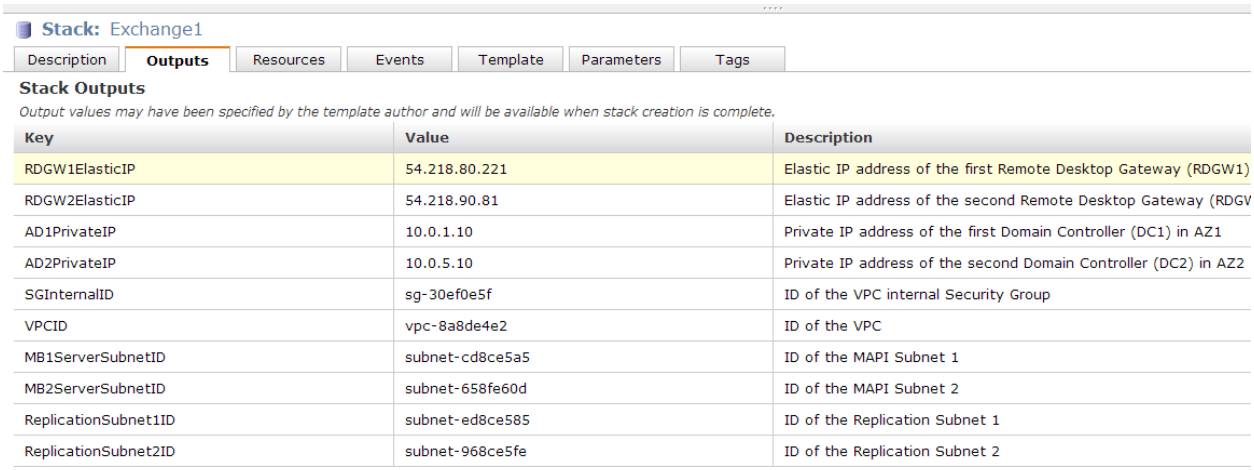
This template performs the following tasks:

- Launches standard Windows Server 2008 R2 AMIs
    - Joins the instances to the domain
    - Formats the disks holding the MB database volumes
    - Configures the page file on the instance
    - Downloads a trail version from Microsoft and unpacks the Exchange Server installation files
    - Installs the required Exchange Server installation prerequisites
    - Creates an unattended installation script and places it on the desktop
- Attaches two ENIs per instance, one dedicated to the mailbox database replication and one dedicated to client and inbound SMTP access
- Attaches EIPs to the ENIs dedicated to client and inbound SMTP access
- Configures a static route between the dedicated DAG networks in each Availability Zone

To launch the AWS CloudFormation template in the US-West-2 (Oregon) region using standard EBS volumes for the Mailbox Database Volumes, click **Launch**.

To launch the AWS CloudFormation template in the US-West-2 (Oregon) region using EBS-optimized instances and provisioned IOPS EBS volumes for the Mailbox Database Volumes, click **Launch**.

Note: When you launch the second AWS CloudFormation sample template, use the output values from the first template.


Stack: Exchange1

Description | Outputs | Resources | Events | Template | Parameters | Tags

Stack Outputs

Output values may have been specified by the template author and will be available when stack creation is complete.

Key	Value	Description
RDGW1ElasticIP	54.218.80.221	Elastic IP address of the first Remote Desktop Gateway (RDGW1)
RDGW2ElasticIP	54.218.90.81	Elastic IP address of the second Remote Desktop Gateway (RDGV
AD1PrivateIP	10.0.1.10	Private IP address of the first Domain Controller (DC1) in AZ1
AD2PrivateIP	10.0.5.10	Private IP address of the second Domain Controller (DC2) in AZ2
SGInternalID	sg-30ef0e5f	ID of the VPC internal Security Group
VPCID	vpc-8a8de4e2	ID of the VPC
MB1ServerSubnetID	subnet-cd8ce5a5	ID of the MAPI Subnet 1
MB2ServerSubnetID	subnet-658fe60d	ID of the MAPI Subnet 2
ReplicationSubnet1ID	subnet-ed8ce585	ID of the Replication Subnet 1
ReplicationSubnet2ID	subnet-968ce5fe	ID of the Replication Subnet 2

Figure 22: Outputs from Template 1

In this sample deployment, two Exchange Server multi-role servers host four mailbox databases and participate in a Database Availability Group. Mailboxes are evenly distributed among the two mailbox servers. Each mailbox database has two copies (including the active copy), and under normal operating conditions, each Exchange Server server hosts two active database copies. Each Exchange Server server hosts the Mailbox, Client Access, and Hub Transport roles and includes two elastic network interfaces; one dedicated to mailbox database replication, and one dedicated to client and inbound SMTP access.

Part 3: Install and Configure Exchange

Follow these steps to install Exchange:

1. Open the Remote Desktop Connection application (mstsc.exe) and connect to the Remote Desktop Gateway (RDGW1) in AZ1 using its Elastic IP address (e.g., 107.23.221.99).
2. Your Remote Desktop Gateway is domain-joined. Log in using the credentials of the Domain Admin user and Domain Admin Password (e.g., UID: Contoso\StackAdmin and Password: Password123).
3. After successfully logging into the Remote Desktop Gateway, open the Remote Desktop Connection application, and connect to the Exchange mailbox server (EX01) in AZ1 using its NetBIOS name (for example, EX01).
4. Use the credentials of the Exchange Admin User and Exchange Admin Password (for example, UID: contoso\exadmin and Password: Password123) to log into the instance.

The template we launched in Part 2 placed a pre-configured unattended install file on the desktop.

1. Launch the unattended Exchange installation by right-clicking the **InstallExchange2010.bat** file.
2. Select **Run as administrator**.
3. After Exchange is installed but before you configure the DAG, add two secondary Private IP addresses to the instances hosting the Mailbox Servers (as described in the preceding section, Configuring for High Availability).

If your business and deployment needs require it, replace the Remote Desktop Gateway instances deployed in Step 1 with two Microsoft Forefront Unified Access Gateway 2010 servers. These UAG servers provide secure, pre-authenticated client access to Exchange Server and serve as a Remote Desktop Gateway for secure administrative Remote Desktop Protocol access to your Active Directory servers or any other server deployed inside your Amazon VPC.

You also might want to help provide a near-seamless user experience during failovers or switchovers by providing load-balanced client access FQDNs through a third-party load balancer.

Architecture

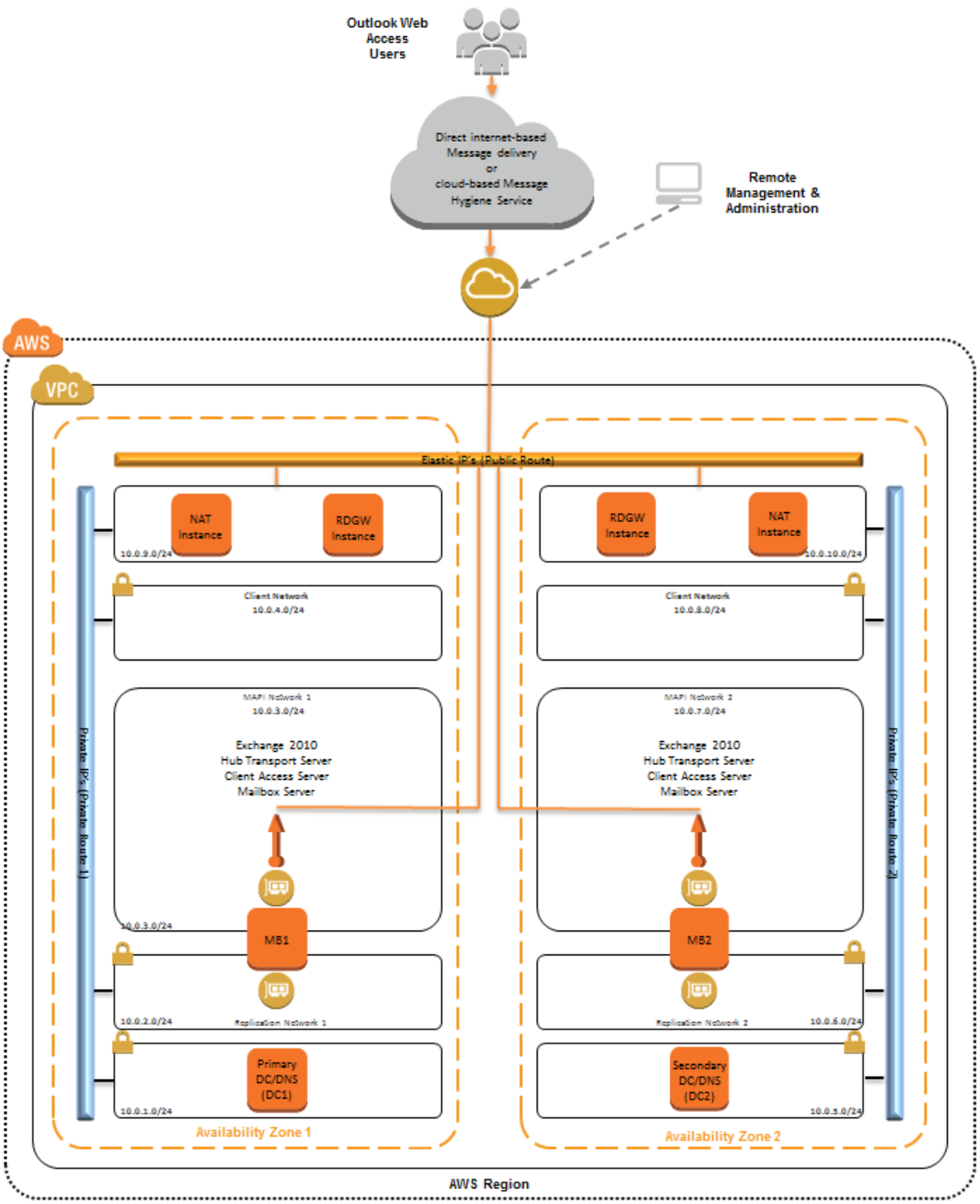


Figure 23: Small Business Sample Architecture

Conclusion

The reliable and secure AWS cloud infrastructure enables highly available and site-resilient Exchange Server deployments.

In this whitepaper, we walked you through the steps that an Exchange Server deployment engineer or Exchange Server administrator would perform to deploy Exchange Server in AWS environment. Using the tools that you are already familiar with, we provided recommendations and planning tips so you can configure your Exchange Server environment based on your individual requirements and business needs. We also provided a sample deployment scenario with AWS CloudFormation templates that helps you deploy the environment with few clicks. We further highlighted some advanced topics that require careful consideration so you can successfully deploy Exchange Server in the AWS cloud with confidence.

Further Reading

- AWS Case Study:
  - Choice Logistics
    http://aws.amazon.com/solutions/case-studies/choicelogistics/
- Microsoft on AWS:
  - http://aws.amazon.com/microsoft/
- Amazon EC2 Windows Guide:
  - http://docs.amazonwebservices.com/AWSEC2/latest/WindowsGuide/Welcome.html?r=7870
- Microsoft AMIs for Windows and SQL Server:
  - http://aws.amazon.com/windows
  - https://aws.amazon.com/amis?ami_provider_id=1&platform=Windows&selection=ami_provider_id%2Bplatform
- Microsoft License Mobility:
  - http://aws.amazon.com/windows/mslicensemobility
- Whitepapers/Articles:
  - Deploy a Microsoft SharePoint 2010 Server Farm in the AWS Cloud in 6 Simple Steps
    http://aws.amazon.com/articles/9982940049271604
  - Microsoft SharePoint 2010 on AWS: Advanced Implementation Guide
    http://media.amazonwebservices.com/AWS_SharePoint_Reference_Implementation_Guide.pdf

- Microsoft SharePoint Server on AWS: Reference Architecture
  http://awsmedia.s3.amazonaws.com/SharePoint_on_AWS_Reference_Architecture_White_Paper.pdf

- Secure Microsoft Applications on AWS
  http://media.amazonwebservices.com/AWS_Microsoft_Platform_Security.pdf

- Implementing Microsoft Windows Server Failover Clustering (WSFC) and SQL Server 2012 AlwaysOn Availability Groups in the AWS Cloud
  http://media.amazonwebservices.com/AWS_WSFC_SQL_Server_AlwaysOn.pdf

Additional Resources

- Exchange Server Resources:
  - V20.8 of the Exchange 2010 Server Role Requirements Calculator:
    http://gallery.technet.microsoft.com/office/Exchange-2010-Mailbox-Server-Role-
  - Exchange Processor Query Tool:
    http://gallery.technet.microsoft.com/Exchange-Processor-Query-b06748a5
  - Jetstress:
    http://www.microsoft.com/en-us/download/details.aspx?id=4167
  - Jetstress Field Guide:
    http://gallery.technet.microsoft.com/Jetstress-Field-Guide-1602d64c
  - Exchange Server Load Generator 2010:
    http://technet.microsoft.com/en-us/library/dd335108(v=exchg.141).aspx
  - Exchange Client Network Calculator
    http://gallery.technet.microsoft.com/office/Exchange-Client-Network-8af1bf00
  - Exchange Backup solutions listed in the Windows Server Catalog:
    http://windowsservercatalog.com/results.aspx?text=exchange+backup&=Go&bCatID=1282&avc=10&ava=0&OR=5&chtext=&cstext=&csttext=&chbtext
  - Exchange, Firewalls, and Support… Oh, my!:
    http://blogs.technet.com/b/exchange/archive/2013/02/18/exchange-firewalls-and-support-oh-my.aspx
- AWS Resources:
  - AWS Toolkit for Visual Studio:
    http://aws.amazon.com/visualstudio/
  - AWS Windows and .NET Developer Center:
    http://aws.amazon.com/net
  - AWS Management Pack for Microsoft Systems Center:
    http://aws.typepad.com/aws/2013/05/aws-management-pack-for-microsoft-system-center.html
- Other Resources:
  - Spamhouse.org block list removal center:
    http://www.spamhaus.org/lookup/

Appendix

Amazon EC2 Security Group Configuration

AWS provides a set of building blocks, such as Amazon EC2 and Amazon VPC, that customers can use to provision infrastructure for their applications. In this model, some security capabilities, such as physical security, are the responsibility of AWS and are highlighted in the *AWS Security Whitepaper* at *http://media.amazonwebservices.com/pdf/AWS_Security_Whitepaper.pdf.* Other areas, such as controlling access to applications, are the responsibility of the application developer and the tools provided in the Microsoft platform.

If you followed the scripted deployment option described in this paper, AWS CloudFormation templates configured the necessary security groups for you. The security groups are listed here for your reference:

Subsystem Port Mappings

Subsystem	Associated With	Inbound Interface	Port(s)
SGInternal	DC1, DC2, EX01 (eth0), EX02 (eth0)	SGInternal	TCP 1-3388, TCP 3390-65535, UDP 1-65535, (ICMP -1)
		PublicSubnet1	TCP 3389, (ICMP -1)
		PublicSubnet2	TCP 3389, (ICMP -1)
		DomainMemberSG	UDP123, TCP135, UDP138, TCP445, UDP445, TCP464, UDP464, TCP49152-65535, UDP49152-65535, TCP389, UDP389, TCP636, TCP3268, TCP3269, TCP54, UDP53, TCP88, UDP67, UDP2535
DomainMemeberSG	RDGW1, RDGW2	PrivateSubnet1 (subnet where DC1 is deployed)	TCP53, UDP53, TCP49152-65535, UDP49152-65535
		PrivateSubnet5 (subnet where DC2 is deployed)	TCP53, UDP53, TCP49152-65535, UDP49152-65535
NAT1SecurityGroup	NAT1	0.0.0.0/0	TCP 22
		PrivateSubnet1CIDR (subnet where DC1 is deployed)	ALL 1-65535
		PrivateSubnet2CIDR (subnet where the Replication network in AZ1 is deployed)	ALL 1-65535
		PrivateSubnet3CIDR (subnet where the mail server in AZ1 is deployed)	ALL 1-65535
		PrivateSubnet4CIDR (subnet where the mail clients in AZ1 are deployed)	ALL 1-65535
NAT2SecurityGroup	NAT2	0.0.0.0/0	TCP 22
		PrivateSubnet5CIDR	ALL 1-65535

		(subnet where DC2 is deployed)	
		PrivateSubnet6CIDR (subnet where the Replication network in AZ2 is deployed)	ALL 1-65535
		PrivateSubnet7CIDR (subnet where the mail server in AZ2 is deployed)	ALL 1-65535
		PrivateSubnet8CIDR (subnet where the mail clients in AZ2 are deployed)	ALL 1-65535
RDGWSecurityGroup	RDGW1, RDGW2	0.0.0.0/0 *	TCP3389
SGDAG	EX01 (eth1), EX02 (eth1)	SGDAG	ALL 1-65535
SMTPSecurityGroup	EX01 (eth01), EX02 (eth0)	0.0.0.0/0	TCP 25, TCP 465
WebCasSecurityGroup	EX01 (eth01), EX02 (eth0)	0.0.0.0/0	TCP 80, TCP 443

Note: It is important that RDP port **never be** opened up to the entire Internet—not even for testing purposes or temporarily. Always restrict ports and source traffic to the minimum necessary to support the functionality of the application. For a further discussion of securing Remote Desktop Gateway, see the *Securing the Microsoft Platform on Amazon Web Services* whitepaper at *http://media.amazonwebservices.com/AWS_Microsoft_Platform_Security.pdf*.