THE vSPHERE ADMIN'S QUICK REFERENCE GUIDE TO AWS VIRTUAL PRIVATE CLOUD

By David M. Davis, vExpert

Partner, ActualTech Media

ActualTech Media
www.ActualTechMedia.com

Background

There are many good reasons to run enterprise workloads in the public cloud. With most enterprises running vSphere and with most public cloud workloads being run in Amazon’s EC2/VPC, vSphere Admins need to get up to speed quickly on how to administer EC2/VPC, and that’s exactly why you need this quick reference guide.

The Companies

Known originally for selling books, Amazon has become the most well-known public cloud player with over 1 million active customers and 80% more than all other cloud providers, combined.

VMware is known for being the first company to first successfully virtualize the x86 architecture and then they innovated with unique features like vMotion, Distributed Resource Scheduler (DRS), and High Availability (HA). With over 500,000 customers, VMware is in use by 100% of the Fortune 100 and 99.5% of the Fortune 500.

The Products

Amazon Web Services (AWS)

AWS is made up of 450+ different services that provide just every type of cloud computing imaginable. The AWS service that is used to “host virtual servers” (as Amazon calls it) is EC2, or elastic compute cloud. EC2 provides infrastructure public cloud services.

PC, or Virtual Private Cloud, runs on top of Amazon's core computing service EC2, just like many VMware offerings run on top of vSphere. VPC provides each administrator their own secure virtual private cloud to administer and control.

vCloud Suite

VMware’s vCloud is a bundle with the core being the vSphere hypervisor and associated advanced features plus vCenter, vRealize management, and Site Recovery Manager (SRM) for disaster recovery.

The vCloud Suite is used to run on-premises workloads in your local datacenter.

Does AWS EC2 run in the enterprise datacenter?

No. AWS EC2 is used to run off-premises workloads, in the public cloud and there is no on-premises hypervisor.

Does VMware’s vCloud Suite run in the public cloud?

No, vCloud Suite runs only in your local datacenter. VMware does, however, have a public cloud offering called VMware’s vCloud Air which competes with EC2.

AWS Product Line

It's worth pointing out that with the 450+ services that AWS offers, they likely have a public cloud service for every possible business need. Here are some of the most common (infrastructure-related)

AWS services:

- EC2 Container Services – run compatible directly in the cloud without having to deploy a new instance (or install and maintain a guest operation system)
- Elastic Beanstalk – run and manage web apps
- Autoscaling – autoscaling of your workload across instances
- Elastic Load Balancing – network load balancing
- S3 – object store
- CloudFront – content delivery network (CDN)
- EBS – elastic block storage
- Glacier – archival
- Workspaces – Virtual Desktop as a Service (DaaS)

AWS Cloud Watch

Amazon CloudWatch is a monitoring service for AWS cloud resources and the applications you run on AWS. You can use Amazon CloudWatch to collect and track metrics, collect and monitor log files, and set alarms.

VMware Product Line

VMware has many products to handle many different IT challenges with the vCloud Suite, vSphere, and vCenter being the products used for enterprise infrastructure.

Other well-known VMware products are:

- vCloud Air – VMware's public cloud that offers infrastructure, disaster recovery, and object storage (similar to AWS S3), and more
- NSX – network virtualization
- Horizon Suite – desktop and application virtualization
- vRealize Suite – management and automation of the unified hybrid cloud

vRealize Operations

To manage and monitor capacity and performance in your vSphere infrastructure, VMware offers vRealize Operations, as part of the vRealize Suite. Optionally VMware offers other vCloud Management products in the vRealize Suite such as vRealize Automation, vRealize Log Insight, and vRealize Business.

Workload Management

If you are a vSphere Administrator and want to run workloads in Amazon's AWS public cloud under their EC2 service, here's what you need to know. While both are designed to run virtual machines, you'll find that the two use different terminology and have their differences.

Virtual Private Cloud (VPC)

AWS VPC is your own private cloud (with seemingly infinite capacity), in the public cloud that you administer, just as you would your vSphere infrastructure (but with more cloud functionality).

Instance

A virtual server in EC2 is called an "instance". EC2 is designed to run your guest operating systems and applications which are called instances.

Instance Sizes

In EC2, instances are pre-configured based on use case, then in sizes like small, medium, and large. For example:

- T2 – general purpose, burstable
- M4 – general purpose with EBS (elastic block store)
- M3 – general purpose
- C4 – compute optimized, with EBS
- C3 – compute optimized
- R3 – memory optimized
- G2 – with GPU
- I2 – storage optimized for high I/O
- D2 – dense storage instances

The full list of EC2 instance types is here.

vSphere Virtual Data Center (VDC)

vSphere is very focused on virtualizing the datacenter. Thus when working with vSphere you are administering a virtual data center, or VDC, that is managed by VMware vCenter. VDCs can map to physical data centers or there can be multiple VDCs in a physical DC.

Virtual Machine (VM)

A virtual server in vSphere is called a virtual machine, or VM. vSphere runs operating systems and applications inside virtual machine that are hardware independent and portable.

VM Sizing

In vSphere, a VM can be sized based on pre-configured sizes based on the guest operating system that you select or, more typically, virtual machines resources are completely customized where the admins give each VM the resources that they feel it needs (vCPU, vRAM, vNICs, and vDisk).

amazon web services

On-Demand and Reserved Instance Tiers

In EC2, on-demand and reserved instances are all about the price that you pay to use the instances. On-demand instances cost more than reserved as you are making no long term commitment to use those resources. With reserved pricing, you are paying a reduced rate but are committed to paying for some time (however if you overcommitted you have the option of selling those resources to someone else).

Free Tier

Below the on-demand tier (in cost and performance) is the AWS EC2 free tier where you can get started running an instance on EC2 at no cost up to a limited amount of resources and usage. Details on the EC2 free tier are available here.

AMI File

An AWS EC2 instance can be stored in a single file on disk called an AMI file, because it has an AMI file extension (which stands for Amazon Machine Instance).

It's important to note that when an AMI is used to create a new instance it is copied. An AMI is a template that many instances can be started from.

vmware®

Reservations

vSphere reservations ensure that resources are available for a particular VM when resources are overcommitted.

Make no mistake, vSphere's resource reservation options are not similar to EC2's reserved instances.

Free vSphere

You can run virtual machines on vSphere for free using the free vSphere hypervisor. The catch is that, unlike EC2 where VMs are run in the public cloud, you'll need to provide a physical machine (server) to run vSphere.

VMDK File

A vSphere virtual machine can be stored in a single file on disk called a virtual machine disk file, which ends in a ".vmdk" extension.

An important differentiation between an AMI and a VMDK is that vSphere virtual machines are made up of more than just VMDKs. vSphere VMs have VMX configuration files that define the configuration of the VM.

It may make more sense to compare an AMI with the OVA / OVF virtual appliance packaging format however OVA files also contain configurations for the virtual appliance (such as CPU, memory, and disk).

Spot Instance

A unique EC2 feature is the ability to bid on resource capacity and run stateless instances that process until someone outbids you (and your instance is gone). Spot instances are very cool but only applicable to stateless applications.

Availability Zones

Within each AWS EC2 region are multiple availability zones and each offers its own redundant infrastructure. Availability zones are connected via multiple low-latency, high speed links. When you create new instances you select which availability zone you want to start the instance in.

You should distribute instances across multiple availability zones to ensure that if one availability zone fails, instances in another AZ continue to function. Examples of EC2 availability zones within the us-east-1 region are:

- us-east-1a
- us-east-1b
- us-east-1c
- us-east-1d

No vSphere Equivalent

vSphere Availability

In enterprise data centers, you could create an availability zone where two, let’s say, separate racks of servers both had independent storage, network, and power, such that if one rack had a catastrophic failure, the other would continue to operate. This could be done with vSphere HA clusters and redundancy in the infrastructure.

Networking

VPC IP Addressing

With EC2 VPC, all instances receive private IP addresses from the subnet assigned to your VPC. If your VPC spans multiple availability zones then you'll have multiple subnets, at least one from each availability zone.

Special VPC IP addresses are:

- 10.0.0.0: Network address
- 10.0.0.1: Reserved by AWS for the VPC router
- 10.0.0.2: Reserved by AWS for mapping to the Amazon-provided DNS
- 10.0.0.3: Reserved by AWS for future use
- 10.0.0.255: Reserved network broadcast address but broadcast is not supported

Internet Gateways

For an instance to access the public Internet, an internet gateway must be enabled on your VPC as VPC networks must use private IP addresses. Internet gateways provide a default route for Internet-bound traffic and perform NAT (network address translation) to translate between the private and public networks. EC2 Internet gateways scale horizontally, are fully redundant, and highly available.

vSphere IP Addressing

Many enterprises use third-party DHCP servers that provide private IP addresses to the virtual machine guest operating systems. IP address schemes are designed based on the size and the needs of the enterprise.

Accessing the Internet

vSphere infrastructures are typically on private networks as well, like a VPC, and access the Internet through Internet gateways / firewalls, owned and configured by the enterprise.

Elastic IP Addresses

In EC2, public IP addresses that you can dynamically map to your private instances, as needed, are called elastic IP addresses. When a new instance is created in a default VPC, it receives an elastic IP address. Instances created in non-default VPCs don’t receive elastic IP addresses by default. You will be charged for elastic IP addresses that are assigned to an instance even if that instance is off. Running instances with an elastic IP address aren’t charged for use unless they need more than one.

VPC Peering

With EC2 VPC Peering allows you to create a virtual private network (VPN) connection between two VPCs with a few clicks and no need to install or configure network equipment.

AWS Direct Connect

Direct Connect provides the option to connect to AWS from specific AWS-partnered co-location facilities using a dedicated WAN link (not the Internet). That dedicated connection can be partitioned into multiple VLANs using 802.1q

Public IP Addresses in vSphere

Typically, a network administrator will create a DMZ where a few virtual machines connect who need to be on the public Internet, full time. Those VMs have static NAT addresses mapped to their private IP addresses or they may have public IP addresses assigned.

Connecting Datacenters

For companies with multiple datacenters, they are usually connected with either a site-to-site VPN connection (over the public Internet) or a dedicated WAN circuit.

Security

Key Pairs

In EC2, you must create and use a key pair (public and private key) to login to your instance.

Amazon EC2 uses public–key cryptography to encrypt and decrypt login information.

In EC2, Linux instances have no password, and you use a key pair to log in using SSH. With Windows instances, you use a key pair to obtain the administrator password and then log in using RDP.

vSphere Login Security

vSphere has login credentials on ESXi hosts, vCenter management servers, and instead each guest virtual machine. With vSphere 6, the PSC, or platform services controller, provides single sign on (or SSO) for a vSphere infrastructure.

VPC Network Security

In EC2, a security groups controls the network traffic between one or more instances in a VPC. When you launch a new instance in a VPC, you associate a security group with it. Like a virtual firewall, security groups have rules that allow traffic to and from other instances. You can add another layer of defense to your VPC by with optional network access control lists (or network ACLs).

vSphere Network Security

There are multiple ways to control vSphere network security. Traditionally, vCloud Networking and Security (vCNS) was used but, with it's discontinuation, VMware recommends NSX and there are numerous third-party virtual firewalls available for controlling virtual network traffic.

Storage

EBS Volumes

Elastic Block Storage, or EBS, volumes provide EC2 instances with persistent block level storage volumes. EBS volumes are highly available as they are automatically replicated within the availability zone and you are only charged for what you provision.

vSphere Shared Storage

vSphere infrastructures typically use shared storage to storage all virtual machines. That shared storage could be a dedicated NFS NAS, iSCSI SAN, or Fibre Channel (FC) SAN. In many cases today, companies are considering software-defined storage (SDS) that is distributed across multiple hosts running a hypervisor and then packaged as hyperconvergence.

Certification

To get up to speed on new technology and to prove knowledge and obtain new employment or compensation, certifications are commonly pursued by IT professionals. Both AWS and VMware offer a number of certifications in EC2/VPC and vSphere.

AWS Certifications

The two most popular AWS certification for those getting started are:

- AWS Certified Solutions Architect Associate
- AWS Certified SysOp Administrator Associate

For great training on Amazon's VPC, consider Nigel Poulton's Pluralsight course on AWS VPC Operations.

VMware Certifications

The two most popular vSphere certifications for those getting started are:

- VMware Certified Associate – Data Center Virtualization (VCA-DCV6)
- VMware Certified Professsional – Data Center Virtualization (VCP-DCV6)

For great training on VMware vSphere 6 Foundations, consider David Davis's Pluralsight course on vSphere 6 Foundations.