Technical White Paper: 2022

Comparing VMware Cloud to Traditional Public Cloud by Total Cost of Ownership

Table of contents

Introduction

When weighing a move to “The Cloud”, there is a lot to consider. Clouds are complicated, and cloud providers offer a great degree of similar capabilities implemented in proprietary ways, while trying to add their own unique value.

We took a different approach with VMware Cloud™. VMware Cloud was built specifically for multi-cloud, providing a consistent platform across a wide array of hyperscaler cloud providers. This platform offers a broad set of integrated infrastructure capabilities that support the requirements of any modern enterprise application in the data center. It enables rapid migration to any cloud and unifies all new and old environments with consistent operations and security.

When all this is taken together, VMware delivers a cloud platform that provides more functionality at a lower cost. The following sections offer insight to the costs associated with optimal cloud functionality, as well as the differences that distinguish VMware Cloud.

This has been proven over many thousands of customer engagements in the last three years by the VMware Cloud Economics team. See an overview of what the VMware Cloud Economics team does.

VMware Cloud is not the same as traditional public cloud

One of the most frequent misconceptions we hear from customers is that “VMware Cloud MUST be more expensive!” Some customers assume that VMware must be deploying VMware Cloud on top of the full software stack of the public cloud in question, most frequently Amazon Web Services. But actually, VMware deploys the full Software-Defined Data Center stack, VMware Cloud Foundation™, on bare-metal hosts.

The second, and more pernicious, of the misconceptions has to do with how the comparison is set up and an assumption that a basic compute instances, like Amazon EC2 (the compute instances analog), is functionally equivalent to VMware Cloud Foundation running on a bare-metal host, which contains VMware’s market-leading compute virtualization platforms: vSphere®; NSX®, for network virtualization; vSAN™ for storage virtualization; and all the hardware resources of the host. It is this last misconception that this paper seeks to address.

How this paper is structured

For the sake of comparison, we will focus on Amazon Web Services to demonstrate that VMware Cloud costs less in total than any functionally equivalent traditional public cloud. We will cover our cloud economics model and the structure behind it, as well as underlying assumptions. We will give you an “apples-to-apples” comparison between VMware Cloud and the traditional public cloud from a functional perspective, so as to build a comparison where each cloud environment is as close to being functionally equivalent as possible. Then, we will discuss some of the intrinsic VMware technology savings realized in this comparison. Finally, we will offer a few suggestions on how you can get your own cloud economic analysis.

VMware Cloud Economics model

Let us examine how we think about the total cost of a decision to move to the cloud. It is important to be specific when discussing what costs should be considered. An example of the format of a VMware Cloud Economics model is shown in Figure 1. Here, we quantify all aspects of the purchase, installation and operation of a set of workloads (commonly thought of as Virtual Machines, or VMs), which are migrated onto either VMware Cloud or to a traditional public cloud.

Output of a VMware Cloud Economics model			
3-year migration occupation to Public Cloud	VMC on AWS	AWS public cloud	Difference
Compute	$1,394,420	$1,094,566	($299,854)
Storage	$0	$577,462	$577,462
Network	$139,200	$129,600	($9,600)
Functional equivalence software	$521,583	$908,007	$386,423
Operating system licenses	$989,314	$1,673,254	$683,940
Facilities	$0	$0	$0
Labor	$122,402	$122,402	$0
DR	$0	$0	$0
Professional services	$0	$0	$0
Support	$326,195	$540,000	$213,805
Other	$0	$0	$0
Total 3-year operations costs	$3,493,115	$5,045,291	$1,552,176
3-year TCO cost with value-added savings	$3,509,453	$8,907,696	$5,398,243

VMware Cloud added value savings	VMC on AWS	AWS public cloud	Difference
Compute			
CPU oversubscription*	$0	$732,516	($732,516)
Memory overcommit*	$0	$183,129	($183,129)
Storage			
Storage deduplication	$0	$289,077	($289,077)
Storage compression	$0	$144,250	($144,250)
Availability			
App mobility/migration downtime	$0	$83,222	($83,222)
Automated host failure remediation	$0	$1,760,760	($1,760,760)
Switching costs			
Migration (rehosting)	$14,351	$685,417	($671,066)
Re-skilling	$1,987	$167,163	($165,176)
Total added value savings	$16,338	$3,862,405	($3,846,067)

Functional equivalence software detail	VMC on AWS	AWS public cloud	Difference
Networking			
Load balancer	$10,092	$9,836	$255
IPSEC VPN & NAT gateway	$0	$160,999	($160,999)
Distributed network firewall	$0	$127,017	($127,017)
Transit gateway	$99,279	$99,279	$0
Security			
Traceflow	$0	$2,077	($2,077)
Packet monitoring	$0	$6,824	($6,824)
Micro-segmentation	$0	$255,834	($255,834)
Advanced firewall (L7-APP ID)	$160,560	$0	($160,560)
Management			
Observability	$0	$83,222	($83,222)
Logging	$0	$1,760,760	($1,760,760)
vRealize Log Insight	$3,427	$0	$3,427
vRealize Operations Cloud	$248,225	$0	$248,225
Total functional equivalence costs	$521,583	$908,007	($386,423)

Support	VMC on AWS	
Transit gateway	$191,600	
Security	$134,595	
Total support costs	$326,195	

** Note: We count the feature that provides the greatest savings and assume that includes the savings from the other feature.*

These calculations include acquisition costs of the compute, storage and network resources. Additionally, this paper also includes a discussion of labor, support, maintenance and facilities costs. Only then are we are able to analyze the sustainability savings related to moving workloads to the cloud. Our model illustrates carbon metrics savings related to sustainability for those customers who are interested in the environment, social and governance metrics being requested by many investors and boards of directors.

The theory behind the model

The fundamental notion behind VMware's Cloud Economics model is that the software, which makes up our offering in VMware Cloud, gives customers more capabilities for lower total cost over the life of a cloud project. Figure 2 calls into relief the mistake some customers make by focusing on acquisition of compute resources alone. By this measure, traditional public cloud IaaS is indeed less expensive. However, this is not an apples-to-apples comparison because its starting from the wrong baseline. In order do to this correctly, we must look at the total cost of ownership, including advanced networking for things such as micro-segmentation, shared storage, support and of course "switching costs" in the form of cloud migration.

The migration from an on-premises datacenter to the cloud, or from cloud to cloud, is not to be overlooked since it can account for up to 30% of the cost of the project[2] and will determine the time that the project takes. A discussion of the full cost model of migration cloud resources is not the scope of this paper and will be addressed in future work.

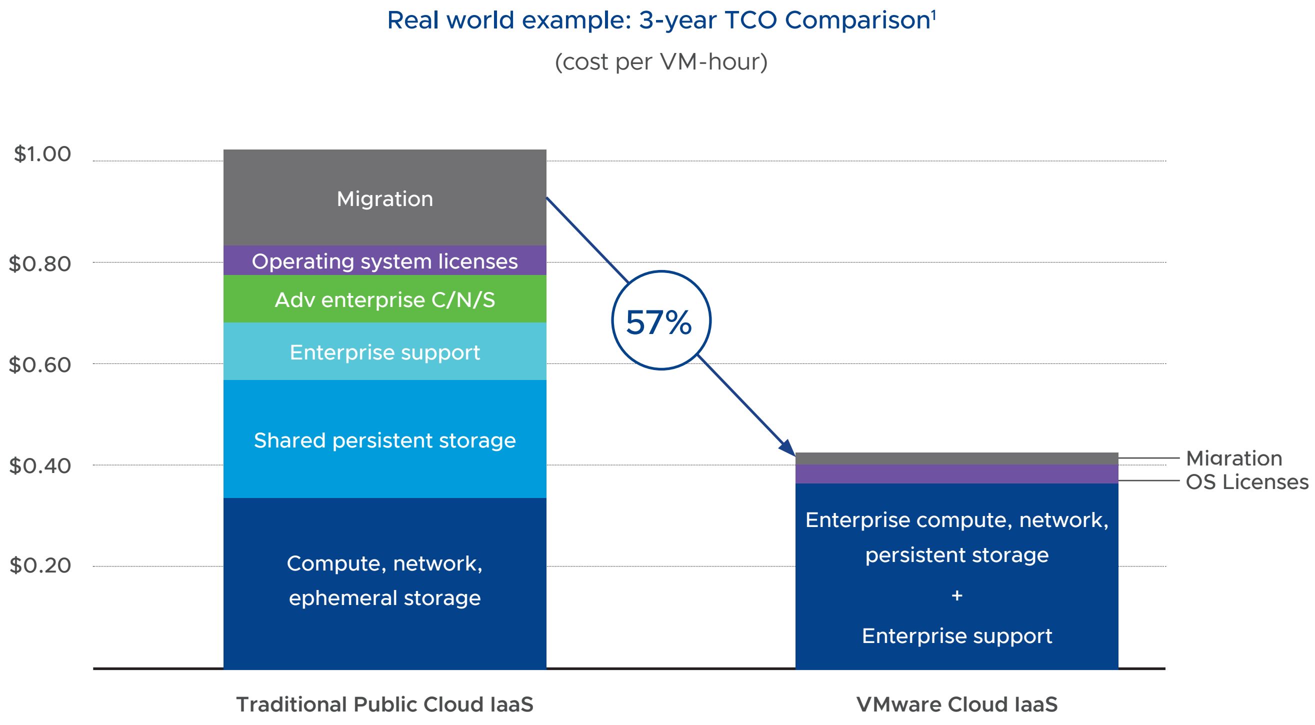


Figure 2: The structure of VMware Cloud savings

The model

Let us look at the overall model used to make the comparison. The VMware Cloud Economics methodology is designed to compare VMware Cloud on AWS and public cloud traditional implementations on as close to an apples-to-apples basis as possible. Our model design methodology delivers defendable results to provide confidence in your decision.

While the two environments appear to be similar on the surface, evaluating the functionality and services of each will help you make a better-informed decision on how well the value of the two alternatives compare. We designed VMware Cloud on AWS to be a complete customer solution. Creating the same operational functionality on the public cloud, however, requires the subscription of additional services over and above the published compute and storage costs needed to deliver the same, and in some cases, less enterprise-ready services. Our exploration of functional equivalence identifies those additional services and aggregates the additional cost on a feature-by-feature basis.

The following assumptions were used as the baseline for this comparison:

- Three-year prepaid reserved instance pricing
- 1,000 VMs running on twenty-eight (28) i3 nodes, with about 300TB of NVMe block storage
- Average VM utilizes 4 vCPUs with 16 GB vRAM
- AWS m5.xlarge EC2 instance-type
- Elastic Block Storage (EBS)
- No discounting applied to either solution

A base of one thousand VM instances was used as a default workload for the model to reflect an average “Enterprise” workload comprised of multiple production applications. In addition, the model was based on data derived from the VMware Cloud on AWS full workload set. The m5.xlarge EC2 instance type was chosen to closely match the average VM size provisioned across the VMware Cloud on AWS cloud infrastructure. Elastic Block Storage was the AWS storage service that matched the performant NVMe block storage used on the VMC on AWS hosts.

Migration and the migration bubble

As suggested by Figure 2, migrating to a cloud is a substantial part of any cloud project, in terms of both time and money. While VMware makes this easy due of its HCX[2] technology, a discussion of the savings VMware provides because of its ease of migration to the cloud is outside the scope of this paper and will be covered in future work.

Suffice it to say that the time difference in moving VMs to the different platforms, whether to VMware Cloud or the public cloud, is substantial. The core VMware Cloud Economics models 11 hours of staff time to migrate a VM from vSphere to a hyperscaler platform. Experience has shown that the average time to move a vSphere VM to VMware Cloud is about 15-20 minutes, due to the vMotion feature and the HCX software package that comes with VMware Cloud at no cost. Clearly, as the number of VMs increases, the difference between the two different types of platforms becomes substantial.

Because migrating to the cloud cannot be done all at once, other factors arise that can unnecessarily raise the total cost of a project. A migration of this sort must have both environments running at the same: the current and the future platform. This means that organizations will incur increased infrastructure costs because they will have to run two platforms at a time, with two sets of storage costs, two sets of management costs and so on, leading to a cost bubble during migration. So, the longer the migration, the longer the time spent paying for two platforms. Organizations must seriously consider the length of time it takes to migrate workloads to the cloud, as this will have a direct impact on the cost of the project.

Like-for-like comparison

To fairly compare two platforms, the platforms must be functionally equivalent. It is grossly unequal to compare a set of cloud virtual machines to the same workload running on the VMware Cloud platform. VMware Cloud's platform contains a wide array of essential technologies for building, running and maintaining enterprise-grade applications. In order to make a fair comparison, we have included comparable technology from AWS to match the functionality included in VMware Cloud's Platform. This paper will first cover the direct costs, that is, the services that an organization would have to purchase from AWS to achieve functional parity, as far as is possible. This will be discussed in the section entitled Direct Costs. Next, this paper will cover the intrinsic value the VMware Cloud platform delivers, namely the costs and savings due to efficient resource usage and cost avoidance. This will be discussed in the section entitled Intrinsic VMware technology value.

Direct costs

Now that the model and basis of comparison have been established, let us discuss the direct costs incurred when building out a like-for-like comparison.

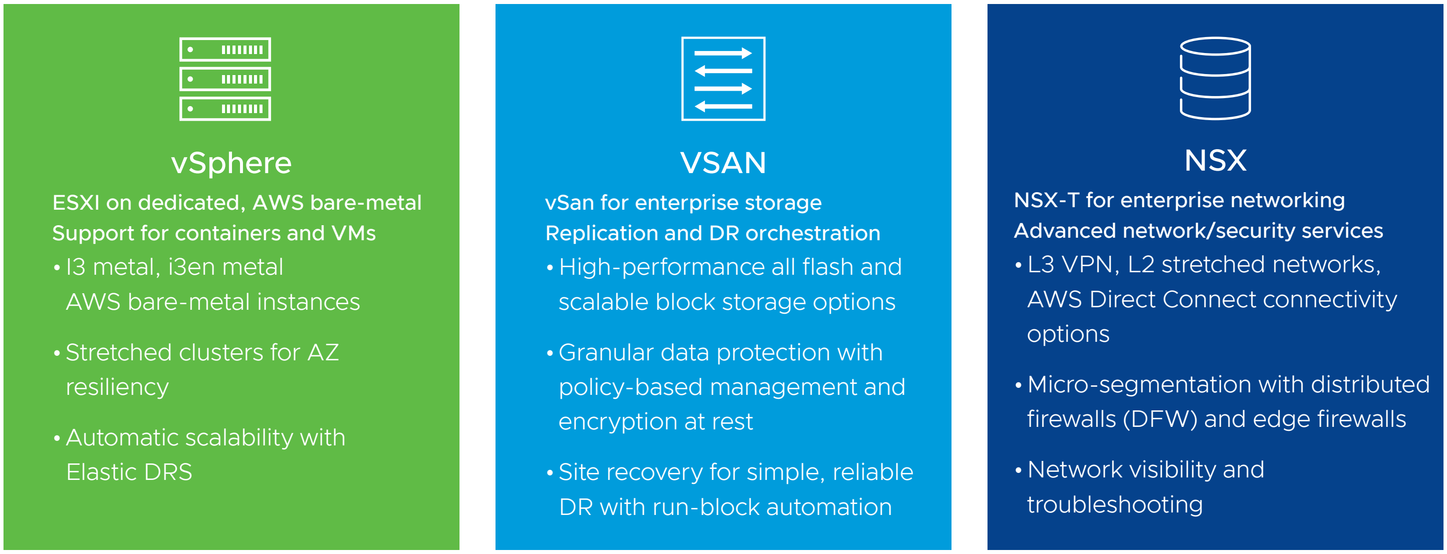
Direct costs refer to the additional native cloud services that need to be added to have feature parity with the out-of-the box features of the VMware Cloud on AWS solution. The VMware Cloud on AWS solution has been jointly engineered to be complimentary to the adjacent services offered natively by AWS. However, for this study we are comparing features as if one of the solutions was being leveraged over another. This highlights the embedded enterprise features and capabilities that comprise the VMC on AWS solution. While this analysis is not an exhaustive comparison of all native cloud services, all efforts have been made to ensure feature parity to the extent possible. This includes adding additional VMware services and costs where applicable.

Baseline

To compare the distinct solutions, we will lay out the baseline architecture used for each. This baseline would be the foundation for a cloud infrastructure implementation running production workloads.

The VMware Cloud on AWS includes the following:

- VMware Cloud Console
- VMware software-defined data center (SDDC): vSphere, vSAN, NSX, vCenter Server®
- VMware HCX
- VMware Tanzu services: Tanzu Kubernetes Grid service + Tanzu Mission Control essentials
- Dedicated Amazon EC2 bare metal instances (i3– 36 Cores/512GB RAM/~10.7TB RAW)
- VMware global support
- Lifecycle management by VMware (updates, patches and upgrades)
- Support for high availability (HA) and stretched clusters
- Service-level agreement (SLA)
- Transit gateway connect



vSphere

ESXI on dedicated, AWS bare-metal
Support for containers and VMs

- I3 metal, i3en metal AWS bare-metal instances
- Stretched clusters for AZ resiliency
- Automatic scalability with Elastic DRS

VSAN

vSan for enterprise storage
Replication and DR orchestration

- High-performance all flash and scalable block storage options
- Granular data protection with policy-based management and encryption at rest
- Site recovery for simple, reliable DR with run-block automation

NSX

NSX-T for enterprise networking
Advanced network/security services

- L3 VPN, L2 stretched networks, AWS Direct Connect connectivity options
- Micro-segmentation with distributed firewalls (DFW) and edge firewalls
- Network visibility and troubleshooting

Figure 3: VMware Cloud on AWS SDDC baseline architecture

The inclusion of the NSX software suite is notable, as it provides a centralized network configuration and management across all clusters. It provides a scale-out architecture for network elasticity with absolute security coverage. NSX is application-aware and provides granular analytics and firewall policy formulation, as well.

The following architecture provides compute, storage and basic connectivity for AWS native cloud services for production workloads. The additional services will be built on top of this baseline in the following sections.

These services provide a baseline AWS native services for our comparison:

- AWS console
- Single region
- Virtual private clouds (VPC) – public and private subnets
- Elastic Block Storage (EBS)
- Elastic compute containers (EC2)
- Multiple availability zones
- Direct connect

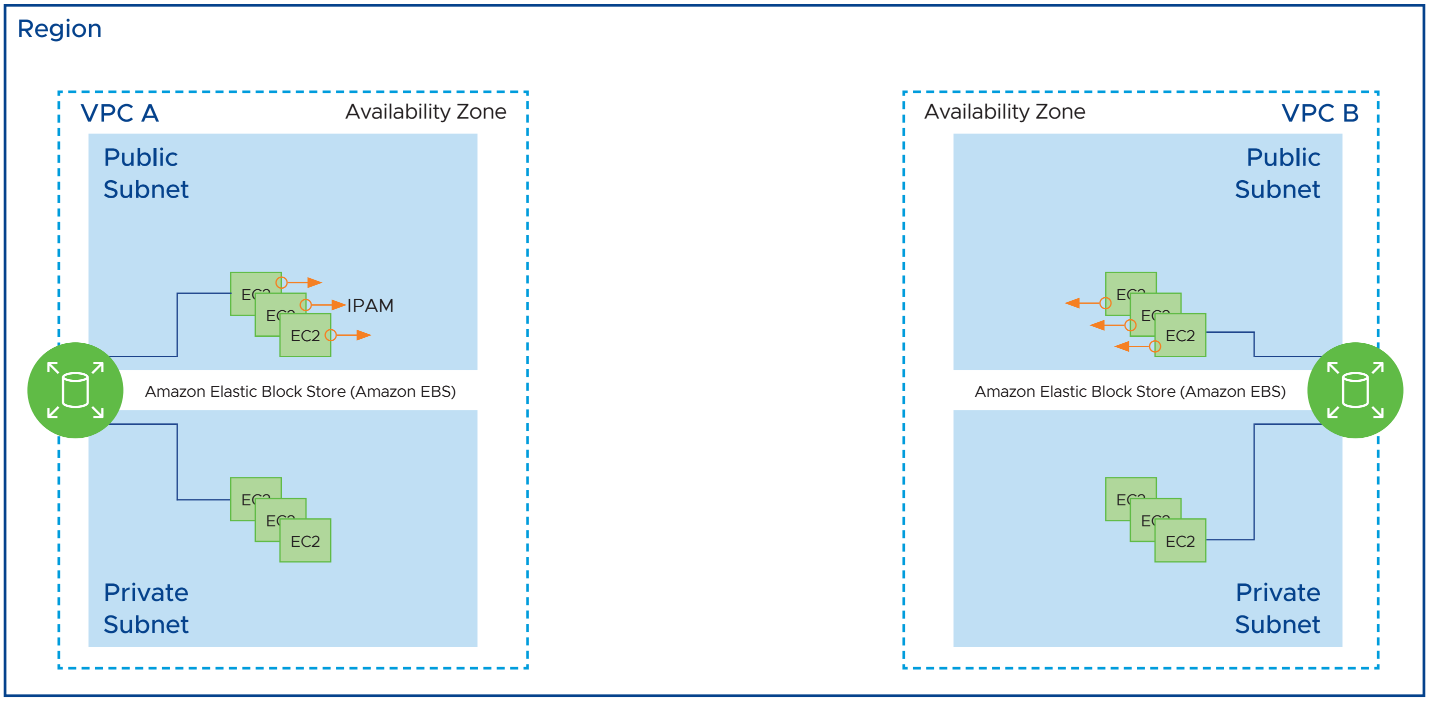


Figure 4: AWS baseline architecture

A medium enterprise application workload was selected, like the one described above, to represent the baseline assumption for the comparison. This reflects a realistic starting point for analysis that may represent the entire production stack or subset for a larger enterprise. While there are several variables that affect the cost for egress/ingress network traffic, we have assumed a percentage of the EBS storage costs as an anchor for data transfer. The default percentages were chosen based on the AWS position within the flow of data and are shown in the table below. This allows the costs to scale accordingly and can be changed easily. If you would like to see what the model looks like with different assumptions, email cloud-economics@vmware.com to start the process of building your own model.

However, more functionality is needed and will be described in the following subsections

AWS overall assumptions	
VM total	1000
Storage per VM (GB)	300
Total VM storage (GB)	300,000
Data transfer outside AWS for transit gateway %	15%
Data transfer outside AWS for VPN %	10%
Data transfer outside AWS for NAT gateway %	5%
Transit gateway data processing %	15%
ELB LCU processed %	20%
NAT gateway data processing %	10%
Network firewall data processing %	20%
CloudWatch log data ingested %	5%
CloudTrail lake log data ingested %	1%
CloudTrail lake log data ingested %	1%

Network services: VMC on AWS SDDC networking

Let us take a look into the network services provided by VMware Cloud, and the services needed in the public cloud to match it.

There is a huge breadth of services offered via VMware Cloud Foundation and NSX[4]. The following diagram highlights the default network services included with a VMC on AWS SDDC subscription.

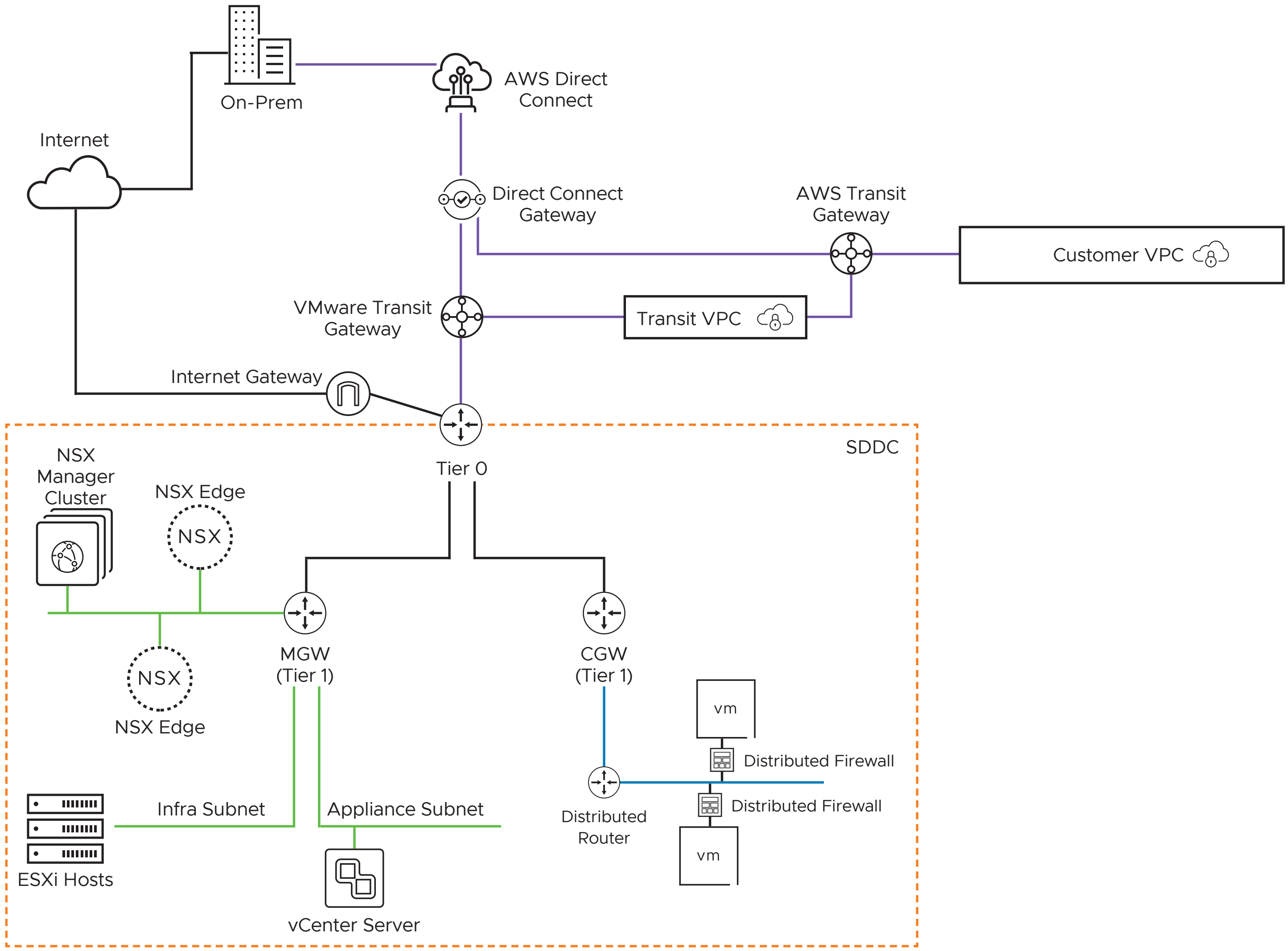


Figure 5: VMC on AWS network topology

NSX provides strong, multi-cloud, easy-to-operationalize network defenses that secure application traffic within and across clouds. NSX goes a step further in making it easy to enable Zero Trust application access, which secures traffic across applications and individual workloads with security controls that are consistent, automated, attached to the workload and elastic in scale.

SDDC network has two notional tiers:

- Tier 0 handles north-south traffic (traffic leaving or entering the SDDC, or between the management and compute gateways). In the default configuration, each SDDC has a single Tier-0 router.
- Tier 1 handles east-west traffic (traffic between routed network segments within the SDDC). In the default configuration, each SDDC has a single Tier-1 router. You can create and configure additional Tier-1 gateways if you need them.

NSX® Edge™ appliance

- The default NSX Edge Appliance is implemented as a pair of VMs that run in active/standby mode. This appliance provides the platform on which the default Tier 0 and Tier 1 routers run, along with IPsec VPN connections and their BGP routing machinery. All north-south traffic goes through the default Tier 0 router.

Management gateway (MGW)

- The MGW is a Tier 1 router that handles routing and firewalling for the vCenter Server and other management appliances running in the SDDC. Management gateway firewall rules run on the MGW and control access to management VMs.

Compute gateway (CGW)

- The CGW is a Tier 1 router that handles network traffic for workload VMs connected to routed compute network segments. Compute gateway firewall rules, along with Network Address Translation (NAT) rules, run on the Tier 0 router.

Networking functions included with VMware Cloud on AWS[5]:

1. **Site-to-site VPN** — NSX supports IPSec Virtual Private Network (IPSec VPN) and Layer 2 VPN (L2 VPN) on an NSX Edge node. IPSec VPN offers site-to-site connectivity between an NSX Edge node and remote sites.

2. **NAT gateway** — NSX supports Source NAT (SNAT), Destination NAT (DNAT) and Reflexive NAT leveraging the tier0/tier-1 gateways running in active-standby mode.

3. **Load balancing** — The NSX Advanced Load Balancer architecture provides a centrally managed, dynamic pool of load balancing resources on commodity x86 servers, VMs, or containers to deliver granular services close to individual applications. This allows network services to scale up without the added complexity of managing hundreds of disparate appliances.

4. **Network firewall** — Distributed Firewall (DFW) is a stateful firewall that runs on all SDDC hosts. It provides protection for traffic within the SDDC and enables micro-segmentation to allow fine-grained control over traffic between workloads. Distributed Firewall rules apply at the VM (vNIC) level and control east-west traffic within the SDDC.

5. **Transit gateway** — VMware Transit Connect Gateway (vTGW) provides connectivity between SDDCs in a single region on VMware Cloud on AWS. Each region has its own vTGW, which can be attached to a vTGW in another region. vTGW provides high bandwidth and low latency connectivity. It also enables connectivity between the SDDC Group and multiple AWS native Virtual Private Clouds (VPCs), as well as customers' on-premises environments connected via an AWS Direct Connect Gateway.

Native AWS networking

Having established the included VMC network services, we can look at the functionality needed to build something comparable in the public cloud. The following diagram highlights the additional native services that would need to be added to equal the default services in the VMC on AWS SDDC.

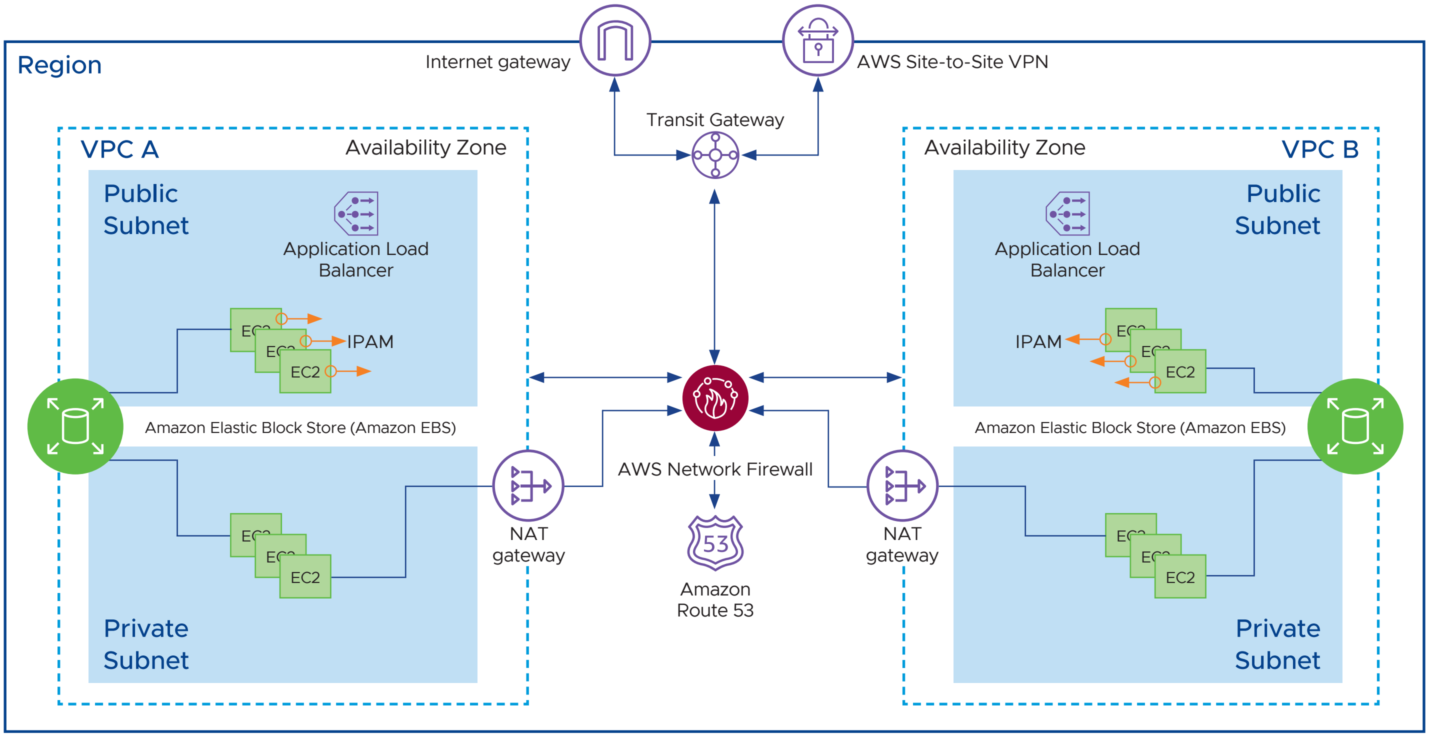


Figure 6: Baseline network configuration for AWS

The additional networking services are comprised of five core components:

1. **Site-to-site VPN** — AWS Site-to-Site VPN is a service that creates a secure connection between an on-premises data center or branch office and the AWS resources using IP Security (IPSec) tunnels. When using Site-to-Site VPN, you can connect to both your Amazon Virtual Private Clouds (VPC) as well as AWS Transit Gateway and two tunnels per connection are used for increased redundancy. The model assumes (2) IPSEC tunnel connections for redundancy, fully provisioned for the month. It also assumes that 10% of instance storage capacity (EBS) will be transferred out of AWS to the Internet.
2. **NAT gateway** — A NAT gateway is a Network Address Translation (NAT) service. You can use a NAT gateway so that instances in a private subnet can connect to services outside the VPC, but external services cannot initiate a connection with those instances. The model assumes (2) full-provisioned NAT gateways for redundancy attached to separate VPCs. The model assumes that the NAT gateways will process 10% of instance storage capacity and that 5% of instance storage capacity (EBS) will be transferred out of AWS to the Internet.
3. **Elastic load balancer** — The Application Load Balancer operates at the request level (layer 7), routing traffic to targets (EC2 instances, containers, IP addresses and Lambda functions) based on the content of the request. This is ideal for advanced load balancing of HTTP and HTTPS traffic. The model assumes leveraging the Application Load Balancer across (2) VPCs and processes 10% of instance storage (EBS) capacity and assumes the processes bytes are the maximum load-balancer capacity unit pricing metric.
4. **Network firewall** — AWS Network Firewall is a managed service that makes it easy to deploy essential network protections for Virtual Private Clouds (VPCs). The model assumes that the Network Firewall is deployed in a centralized manner as a VPC attachment to the Transit gateway. This design maintains symmetric routing to the same zonal firewall, where traffic can be filtered inbound or outbound, to or from Internet gateways, Direct Connect gateways, VPN Site-to-Site and client gateways, NAT gateways and even between other attached VPCs and subnets. The model assumes the Network Firewall will be fully provisioned and process 20% of instance storage capacity (EBS) each month.
5. **AWS transit gateway** — The AWS Transit gateway routes all traffic to and from each VPC or VPN. The model assumes (1) Transit gateway per region with (5) attachments: (2) VPCs, (2) IPSEC VPN tunnels and (1) Network Firewall. The model also assumes traffic processed is 15% of EBS capacity, as well as 15% of EBS instance storage capacity for data transferred out of AWS.

Table outlines the mapping of network functions onto both the VMC on AWS and the native AWS. Note the differences.

Network services pricing						
Function	VMC on AWS service	Native AWS service	VMC on AWS cost metric	VMC pricing	Native cost metric	Native pricing
Load balancer	Advanced load balancer	Elastic load balancer	$ per ALB active per hour	$0.38	$ per ELB active/hour	$0.0225
					$ per load balancer capacity unit in processed bytes	$0.0080
Remote connectivity	NSX Edge router	AWS VPN	NA	Included	$ per hour/connection	$0.05
					$ per GB of data processed	$0.09
Network address translation	NSX Edge router	NAT gateway	NA		$ per hour/NAT Gateway is available	$0.045
					$ per GB data processed	$0.045
					$ per GB data transferred out of AWS	$0.09
Firewall	NSX distributed network firewall	Network firewall	NA		$ per Firewall endpoint/hour	$0.0395
					$ per GB data processed/month	$0.065
VPC routing	Transit connect gateway	Transit gateway	Same as AWS	Passed on	$ per attachment	$0.05
					$ per GB data processed	$0.02
					$ per GB data transferred out of AWS	$0.09

Based on our default workload, following are the networking services assumptions.

AWS networking assumptions	
Transit gateway hourly usage per month	730
Transit GW data processed (GB) per month	44985
Transit GW attachments	3
Transit GW data transferred out per month	44985
Site-to-site VPN connection hours per month	730
Site-to-site VPN connections	2
Site-to-site VPN data transferred out	29990
NAT gateway provisioned	2
NAT gateway hourly usage per month	730
NAT gateway data processed	29990
NAT gateway data transferred out per month	14995
Network firewall endpoints	5
Network firewall endpoints hourly usage	730
Network firewall data processed	59980
Elastic load balancers provisioned	1
Elastic load balancer usage per month	730
Elastic load balancer usage load balancer capacity units (LCUs) per month	59980

If you would like to see what the model looks like with different assumptions, email cloud-economics@vmware.com to start the process of building your own model.

Security

Security is a fundamental element for VMware Cloud, with NSX and the concept of micro-segmentation taking center stage. In this section, we look at the additional security components of NSX and then discuss their analogues in the public cloud.

Distributed firewall design topology

Establishing a security baseline

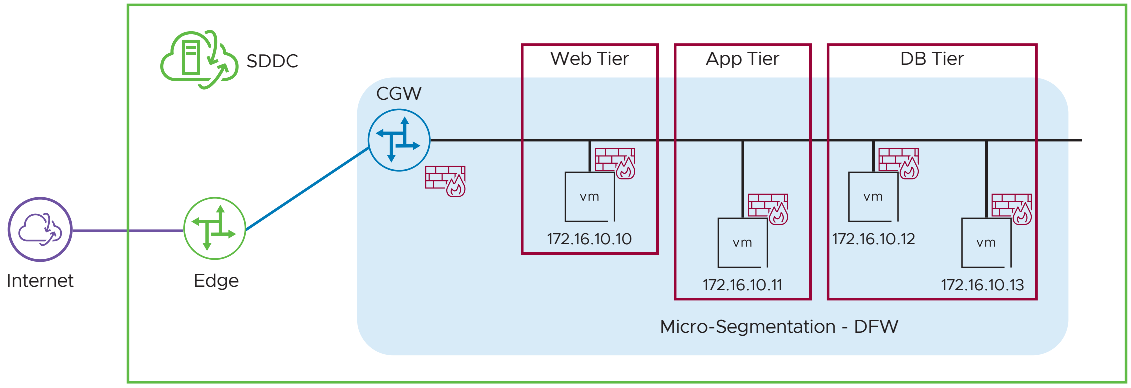


Figure 7: Overview of NSX security

NSX security features include the following advanced capabilities:

Reachability analysis — Traceflow allows you to inject a packet into the network and monitor its flow across the network. This allows you to monitor your network and identify issues, such as bottlenecks or disruptions. With Traceflow, you can identify the path (or paths) a packet takes to reach its destination or, conversely, where a packet is dropped along the way. Each entity reports the packet handling on input and output, so you can determine whether issues occur when receiving a packet or when forwarding the packet.

Traffic mirroring — NSX port mirroring can send mirrored traffic from a source to a destination appliance in the SDDC or to an on-premises network.

Micro-segmentation — This feature is a native capability of NSX. Edge firewalls run on the management and compute gateways. These stateful firewalls examine all traffic into and out of the SDDC. In addition, these distributed firewalls allow fine-grained control over traffic between workloads.

Native AWS security architecture

The following diagram highlights the additional native AWS security services that would need to be added to equal features and functionality provided by NSX.

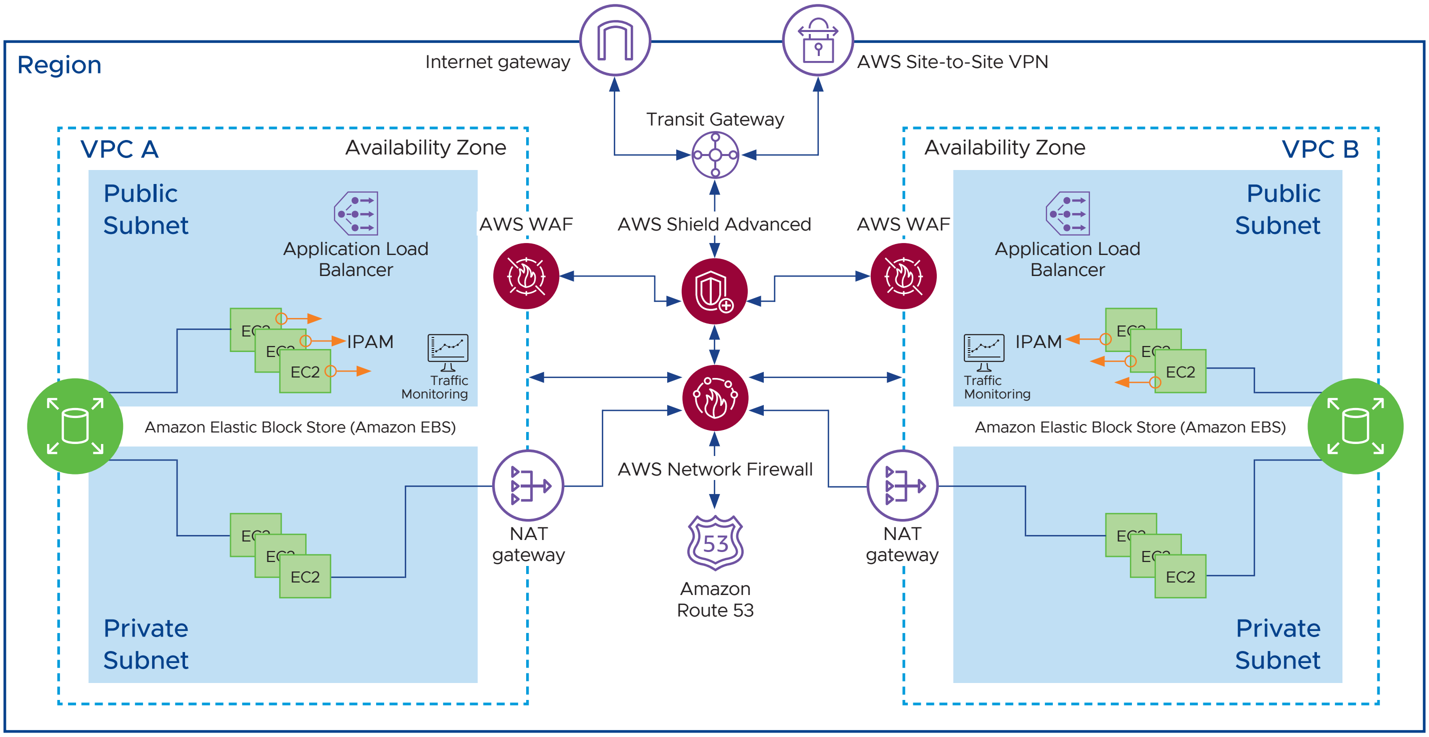


Figure 8: Security services

The additional AWS security services are listed below:

Reachability analysis — This static configuration analysis tool enables analysis and debugging of network reachability between two resources in a VPC. By specifying the source and destination resources, Reachability Analyzer produces hop-by-hop details of the virtual path between them when they are reachable and identifies the blocking component when they are unreachable. The model assumes 1x the number of ENIs assigned to the EC2 instances.

Traffic mirroring — Traffic mirroring gives direct access to the network packets flowing through the VPC. The model assumes that 3% of the Elastic Network Interface (ENI) assigned to EC2 instances will be actively monitored per month.

Micro-segmentation — Often provided by leveraging existing third-party tools that align with existing security and compliancy requirements. However, it also can be provided by layering multiple AWS security services:

- Web application firewalls (WAF)
- Route tables and security groups per VPC
- AWS shield advanced

For example, inbound traffic requests are first screened by AWS Shield. The request then is analyzed by the AWS Web Application Firewall (WAF) to restrict things, such as SQL insertion. The inbound traffic then is sent to the Application Load Balancer for the public subnets. The load balancer redirects to one of the WAFs. These firewalls apply IDS/IPS, malware, sandboxing and sometimes Secure Socket Layer (SSL) decryption for packet-level inspection by security information and event management. Next, the request is sent to the VPC Routing Table. The Routing Table applies Security Group policies, which restrict the source, destination, ports and routes for the traffic to ensure that only specific services can communicate within and between VPCs (public vs private). All the processing done by the VPC is captured in VPC flow logs and routed to CloudTrail.

AWS micro-segmentation implementation default assumptions:

- 20 Web application firewall Web ACLs each with 3500 rules
- 4 Web ACL groups each with 3500 rules
- 5 managed Web ACL groups
- 5 million WAF requests per month
- AWS shield advanced subscription

The pricing for the components mentioned above is outlined in the table below.

Security function pricing

Function	VMC on AWS service	Native AWS service	VMC on AWS cost metric	VMC pricing	Native cost metric	Native pricing
Packet flow analysis	Traceflow	Reachability analyzer	NA	Included	$ per analysis processed	$0.10
Packet inspection	Port mirroring	Traffic mirroring	NA		$ per ENI assessed per hour	$0.015
Micro-segmentation	NSX distributed edge firewall	Web application firewall	NA		$ per WAF WebACL/month	$5.00
					$ per rule per WAF WebACL/month	$1.00
					$ per rule group per WebACL/month	$1.00
					$ per rule in each rule group	$1.00
					$ per rmanaged rule group per WebACL/month	$1.00
					$ per million WAF requests	$0.0000006
Advanced firewall	NSX advanced firewall Add-on	AWS shield advanced	$ per host per year	$10,704	$ per month (1-year subscription)	$36,000

For network security, our model default assumptions are:

Network security defaults	
Reachability analyses per month	1000
Traffic mirroring hourly usage per month	730
Elastic network interfaces (ENIs)	3
Using traffic mirroring	30
WAF web ACL utilized per month	20
WAF web ACL rules per ACL per month	3500
WAF web ACL rule groups per month	4
WAF web ACL rules per rule group per month	3500
WAF managed web ACL rule groups	10
WAF requests per month (million)	5
AWS shield advanced monthly subscription	1

If you would like to see what the model looks like with different assumptions, email cloud-economics@vmware.com to start the process of building your own model.

Observability & management

Observability and management are key to understanding the full cost of moving to the cloud, since the largest portion of time in the lifecycle of an application is spent in running in production. Figure 9 highlights the vRealize® Operations™ Cloud features added to the analysis to provide feature parity with AWS CloudWatch and CloudTrail services.

vRealize Operations Cloud delivers a unified management platform to optimize, plan and scale hybrid cloud deployments from applications to infrastructure as a service, powered by AI/ML.

VMware vRealize Operations Cloud delivers self-driving operations from apps to infrastructure to help optimize, plan and scale VMware Cloud and multi-cloud environments. vRealize Operations Cloud delivers continuous performance optimization, efficient capacity and cost management, proactive planning, intelligent remediation and integrated compliance as a VMware Cloud service.

Continuous performance optimization

Assure hybrid cloud performance at a minimal cost. Based on operational and business intent, real-time predictive analytics drive actions to automatically balance workloads and proactively avoid contention, continuously optimizing VMware Cloud on AWS SDDC and hybrid cloud environments. Automate workload balancing and placement to reduce software license costs, optimize based on performance tiers, consolidate clusters, or enforce compliance.

Efficient capacity and cost management

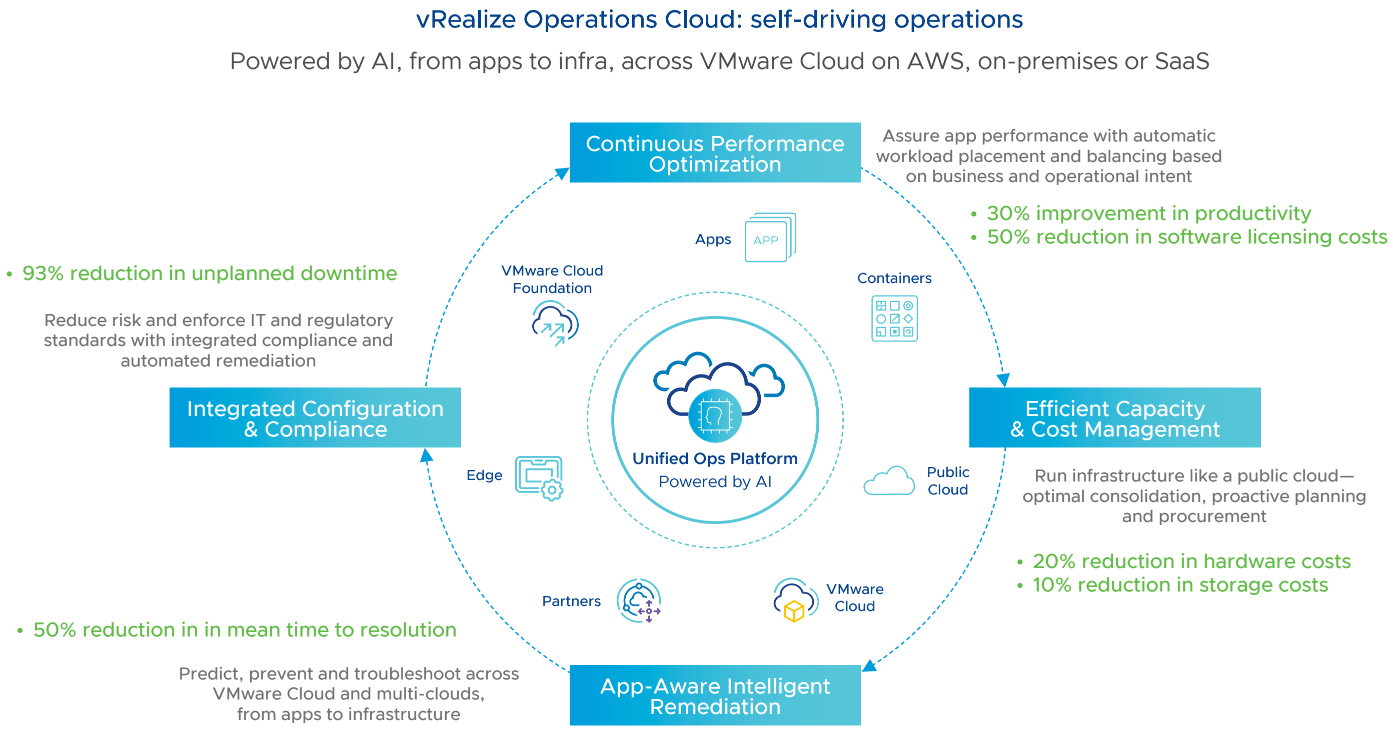
Reduce cost and improve efficiency with real-time, predictive capacity and cost analytics, delivering optimal consolidation and proactive planning. Predict future demand, get actionable recommendations and automate reclamation and right-sizing. Integrate costs and capacity analytics to optimize utilization and reduce costs. Advanced what-if scenarios help plan capacity and model the best fit for new workloads, hardware procurement, HCI planning, cost comparison across data centers and migration planning to public clouds.

Intelligent remediation

Predict, prevent and troubleshoot faster with actionable insights correlating metrics, events, logs and configuration data to deliver AI-based anomaly detection across hybrid clouds. Extend monitoring visibility to multiple public clouds. It is possible to centralize IT Operations management using native SDDC and VMware Cloud on AWS integrations, with management packs for scalability and extensibility.

Integrated compliance

Reduce risk and enforce IT and regulatory standards for SDDC and VMware Cloud on AWS with integrated compliance and automated drift remediation. Ensure the environment's adherence to requirements.



vRealize Operations Cloud: self-driving operations

Powered by AI, from apps to infra, across VMware Cloud on AWS, on-premises or SaaS

Continuous Performance Optimization

Assure app performance with automatic workload placement and balancing based on business and operational intent

- 30% improvement in productivity
- 50% reduction in software licensing costs

Efficient Capacity & Cost Management

Run infrastructure like a public cloud—optimal consolidation, proactive planning and procurement

- 20% reduction in hardware costs
- 10% reduction in storage costs

App-Aware Intelligent Remediation

- 50% reduction in in mean time to resolution

Predict, prevent and troubleshoot across VMware Cloud and multi-clouds, from apps to infrastructure

Integrated Configuration & Compliance

- 93% reduction in unplanned downtime

Reduce risk and enforce IT and regulatory standards with integrated compliance and automated remediation

Figure 9: vRealize Operations Cloud features

VMware vRealize® Log Insight™ Cloud is a VMware cloud service that collects and analyzes log data generated by all resources in your VMware Cloud on AWS environment to centralize log management, accelerate IT troubleshooting and provide deep, operational visibility across VMware Cloud on AWS and private cloud environments. Every VMware Cloud on AWS subscription includes a trial of VMware vRealize Log Insight Cloud focused on audit and diagnostic capabilities. The full subscription cost has been added to the model for feature parity with native cloud. Key features include:

- Real-time log search
- Analysis and visualization
- Dashboard creation and sharing
- Usage statistics; 1 gigabyte per day per org log ingestion; real-time and query-based alerting; log forwarding to on-premises and SaaS destinations like Splunk
- Log export for offline analysis
- Metric extraction from log messages; public query APIs; log filtering, tagging and masking; and audit dashboards

Native management and observability

Here we look at the management and observability features that align to the VMware Cloud Architecture, using Amazon Web Services as our default public cloud.

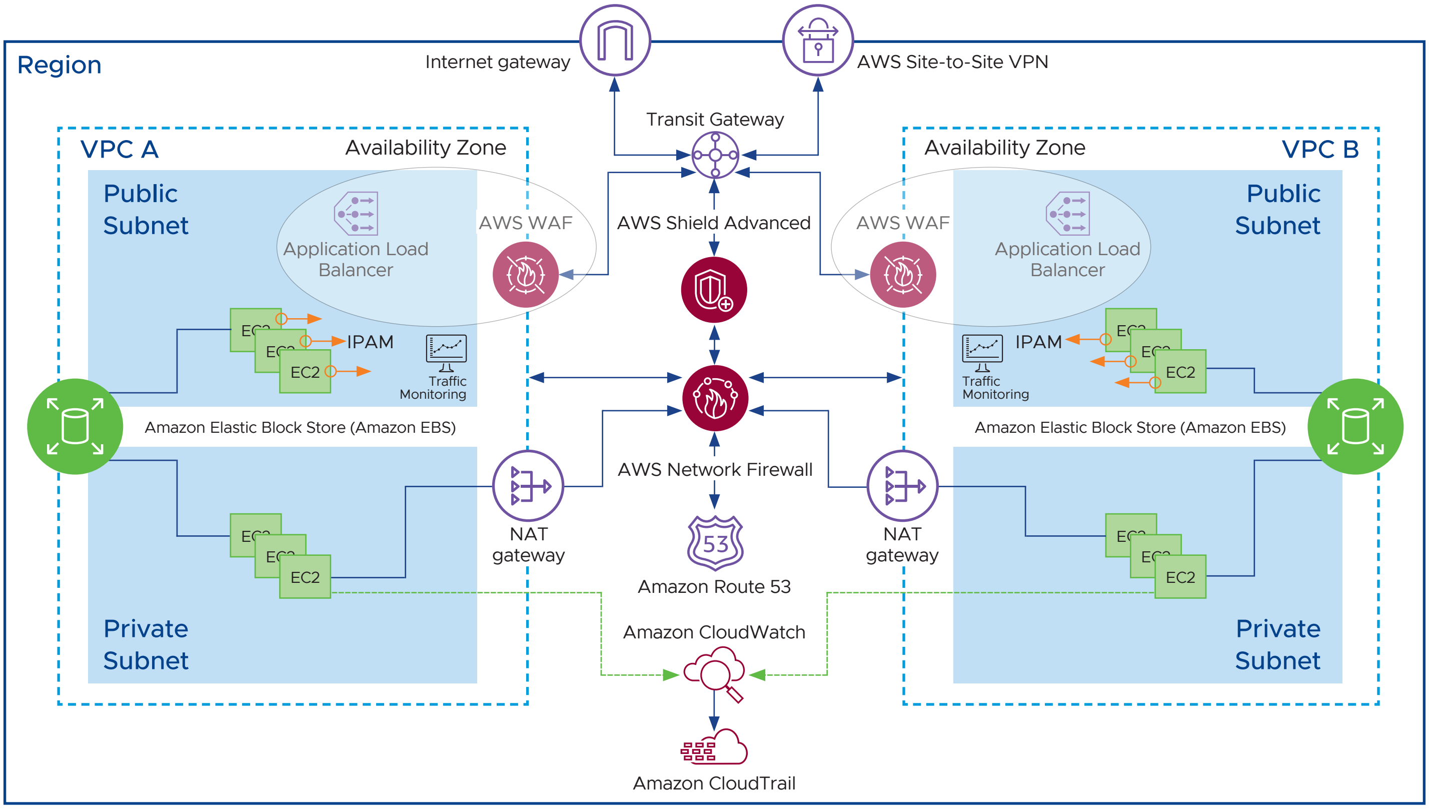


Figure 10: Management services

The AWS management services included to match the VMware Cloud features are listed below:

Amazon CloudWatch — CloudWatch provides data and actionable insights to monitor applications, respond to system-wide performance changes and optimize resource utilization. CloudWatch collects monitoring and operational data in the form of logs, metrics and events. The model assumes (7) metrics per instance captured and 5% of EBS storage capacity processed each month.

AWS CloudTrail — CloudTrail enables auditing, security monitoring and operational troubleshooting by tracking user activity and API usage. CloudTrail records two types of events: management events capturing control plane and data events capturing high-volume data plane actions, such as reading or writing an Amazon S3 object or CloudWatch Logs. The model defaults include use of CloudTrail Lake with 1% of EBS capacity ingested and stored per month, along with 1% of EBS capacity analyzed per month. Pricing for these functions is listed in the table below.

Management function pricing

Function	VMC on AWS service	Native AWS service	VMC on AWS cost metric	VMC pricing	Native cost metric	Native pricing
Observability	vRealize Operations Cloud	CloudWatch	Per OSI/VM	$11.95 per VM/month	$ per metric monitored	$0.30
					$ per GB flow log data ingested	$0.50
Logging	vRealize Log Insight Cloud	CloudTrail	Per GB indexed (defaulte is 1GB per VM)	$1.65 per 10GB indexed/month	$ per GB data ingested and stored	$1.00
					$ per GB data processed	$0.01

Our management and observability model default assumptions are referenced below.

Management and logging defaults

CloudWatch metrics per EC2 instance	7
CloudWatch log data ingested (GB)	14995
CloudTrail lake data ingested and stored	2999
CloudTrail lake data processed	2999

Intrinsic VMware technology value

As we have shown, there are direct costs associated with public cloud beyond what VMware Cloud offers out-of-the-box at no extra charge. In addition, there are several areas where the VMware Cloud platform avoids costs by efficient use of resources, which obviate the need for additional computing/network/storage purchases. In this section, we will cover areas of cost avoidance, which we are calling Intrinsic VMware Technology Value.

Compute savings and CPU oversubscription

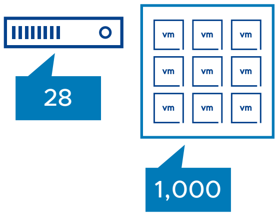
In this section, we will discuss how the VMware Cloud platform allows for optimization of computing resources via a feature called CPU oversubscription.

The value of CPU oversubscription is based on the ability to oversubscribe CPUs on VMC hosts, which increases the number of workloads that can be sustained effectively on the physical hosts. Just as a juggler can keep multiple balls in the air while using only two hands, oversubscription accomplishes the same thing for virtualized workloads with many workloads in the air (running) on just a few hands (CPUs).

Static size example with i3 host

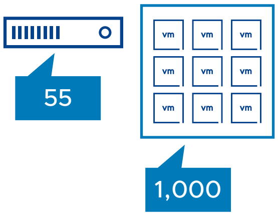
VMware Cloud on AWS i3 Host with Oversubscription

- 1000 VMs
- 4 VCPUs per VM
- 36 cores
- 72 vCPUs per host
- 28 I3 Hosts



EC2 i3.metal Host without CPU Oversubscription

- 1000 VMs
- 4 VCPUs per VM
- 36 cores
- 72 vCPUs per host
- 28 I3 Hosts



Hosts = Number of VMs / (72 vCPUs / vCPUs per VM)
= 1000 / (72/4)
= 55
VMs per Host = 1000 / 55 = 18

Additional i3.metal hosts = 17
Oversubscription Value = 27 i3.metal hosts * $61,043 (3yr price per host)
= $1,648,161

- The oversubscription value represents the VMware Cloud host commitment savings based on CPU Oversubscription and increased VM density by the reduction of physical hosts.
- The ability to oversubscribe CPUs effectively reduces the number of physical hosts by 27 with a 3-year cost savings of $1,648,161

Figure 11: CPU oversubscription

With CPU oversubscription, fewer physical hosts are required on VMC to support the same workloads as compared with hosts that are not configured for oversubscription. In order to value this ability, we estimate the number of hosts that would be required to support the same number of VMs based on the CPU allocation to each VM. For simplicity, we use the average vCPUs per VM.

We compute the additional hosts that would be required as the difference between the non-oversubscribed hosts and the VMC hosts. The total value of CPU consolidation is the product of the number of hosts saved and the VMC cost per host.

This value represents the total savings recognized by having the ability to oversubscribe the CPUs on the physical costs, thereby effectively allowing you to deliver the same work with approximately 30% fewer hosts.

Storage and compression

VMware Cloud on AWS Storage with deduplication

Public Cloud EBS Storage without deduplication

300 TB usable
75 TB uncompressed
300 TB usable

Uncompressed space = Number of TBs * 1.25 (avg compression rate)
= 300 * 1.25
= 375 TB

- Requires less storage for VMware Cloud on AWS (using 1.25 storage compression factor)
- Requires additional monthly storage costs for Public Cloud
- Provides the ability to share storage not available in the Public Cloud

Cost of uncompressed storage = Additional storage (GB) * EBs cost per GB per month
= (75 * 1024) * ($0.10 * 36)
= $276,480

- The uncompressed storage value represents the potential VMware Cloud on AWS storage savings based on a 1.25 rate of compression
- The ability to compress storage effectively reduces the amount of storage needed by 20% with a 3-year cost savings of $276,480

Figure 12: Storage compression

The value of storage compression is relatively straight forward. With VMware Cloud, we can average a 1.25 data compression rate, which means 500GB of information could be stored in 400GB of storage space, for example. This would occur by using standard data compression algorithms. This capability can reduce the amount of storage you would need to purchase to retain the same information in public cloud storage, resulting in higher monthly operational costs for the public cloud.

We compute this benefit by taking the total GB of storage needed in the VMC hosts and multiplying by the 1.25 compression factor to arrive at a public cloud non-compressed storage total. The compressions factor is an indication of how much more storage you would need without the compression algorithm. We then calculate the difference in the VMC storage and apply the monthly public cloud cost per GB.

This produces a cost for the additional public cloud storage expenses that would occur without the ability to compress storage.

Storage and deduplication

VMware Cloud on AWS Storage with deduplication

Public Cloud EBS Storage without deduplication

300 TB usable
150 TB duplicated
300 TB usable

Cost of duplicated storage = Additional storage (GB) * EBs cost per GB per month
= (150 * 1024) * ($0.10 * 36)
= $552,960

Duplicated time = Number of TBs * 150 (avg duplication rate)
= 300 * 150
= 450 TB

- Requires less storage for VMware Cloud on AWS (using a 1.5 deduplication factor)
- Requires additional monthly storage costs for Public Cloud
- Provides the ability to share storage not available in the Public Cloud
- The duplicated storage value represents the potential VMware Cloud on AWS storage savings based on a 150 rate of deduplication
- The ability to deduplication storage effectively reduces the amount of storage needed by 33% with a 3-year cost savings of $552,960

Figure 14: Duplication

The value of storage deduplication is relatively straightforward. With VMware Cloud, we can average a 1.5 data deduplication rate, which means 500GB of information could be stored in 333GB of storage space, for example. It does this by storing each unique file only once. This is especially useful when you have may copies of the same file, as you would have with multiple instances of an operating systems. This capability can reduce the amount of storage you would need to purchase to retain the same information in public cloud storage resulting in higher monthly operational costs for public cloud.

We compute this benefit by taking the total GB of storage needed in the VMC hosts and multiplying by the 1.5 deduplication factor to arrive at a public cloud non-deduplicated storage total. The deduplication factor is the amount of storage the virtual machine would need if deduplication was not being used. We then calculate the difference in the VMC storage and this value and apply the public cloud monthly cost per GB. In the above example, 300TB of deduplicated storage can contain the same amount of usable information as 450TB of non-deduplicated storage, resulting in a cost avoidance of approximately $553K over three years. This produces a cost for the additional public cloud storage expenses that would occur without the ability to deduplicate storage.

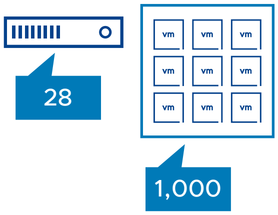
Memory and memory overcommit

Like CPU oversubscription, memory overcommit also is leveraging the computer's ability to "juggle" the RAM utilized by workloads.

Static size example with i3 host

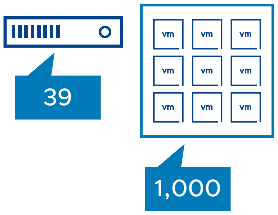
VMware Cloud on AWS i3 host with overcommit

- 1000 VMs
- 16 GB RAM per VM
- 512 GB vRAM per host
- 28 I3 Hosts



EC2 i3.metal host without overcommit

- 1000 VMs
- 16 GB RAM per VM
- 512 GB vRAM per host
- 28 I3 Hosts



Hosts = Number of VMs / (512 GB RAM per host / vRAM per VM) * 1.25
= 1000 / (512/16) * 1.25
= 39
VMs per Host = 1000 / 39 = 26

Additional EC2 i3.metal hosts = 11
Oversubscription Value = 11 i3.metal hosts * $61,043 (3yr price per host)
= $671,473

- The overcommit value represents the VMware Cloud host commitment savings based on RAM overcommit and increased VM density enabled by the reduction of physical hosts.
- The ability to overcommit RAM effectively reduces the number of physical hosts by 11 with a 3-year cost savings of $671,473

Figure 13: RAM overcommit

The value of RAM overcommit is based on the ability to overcommit RAM on VMC hosts, which increases the number of workloads that can be effectively sustained on the physical hosts. With RAM overcommitment, fewer physical hosts are required on VMC to support the same workloads as compared with hosts that are not configured for memory overcommit. In order to value this ability, we estimate the number of VMC hosts that would be required to support the same number of VMs based on the RAM allocation to each VM. For simplicity, we use the average vRAM per VM.

We compute the additional hosts that would be required as the difference between the non-overcommitted hosts and the VMC hosts. The total value of memory overcommit is the product of the number of hosts saved and the VMC cost per host.

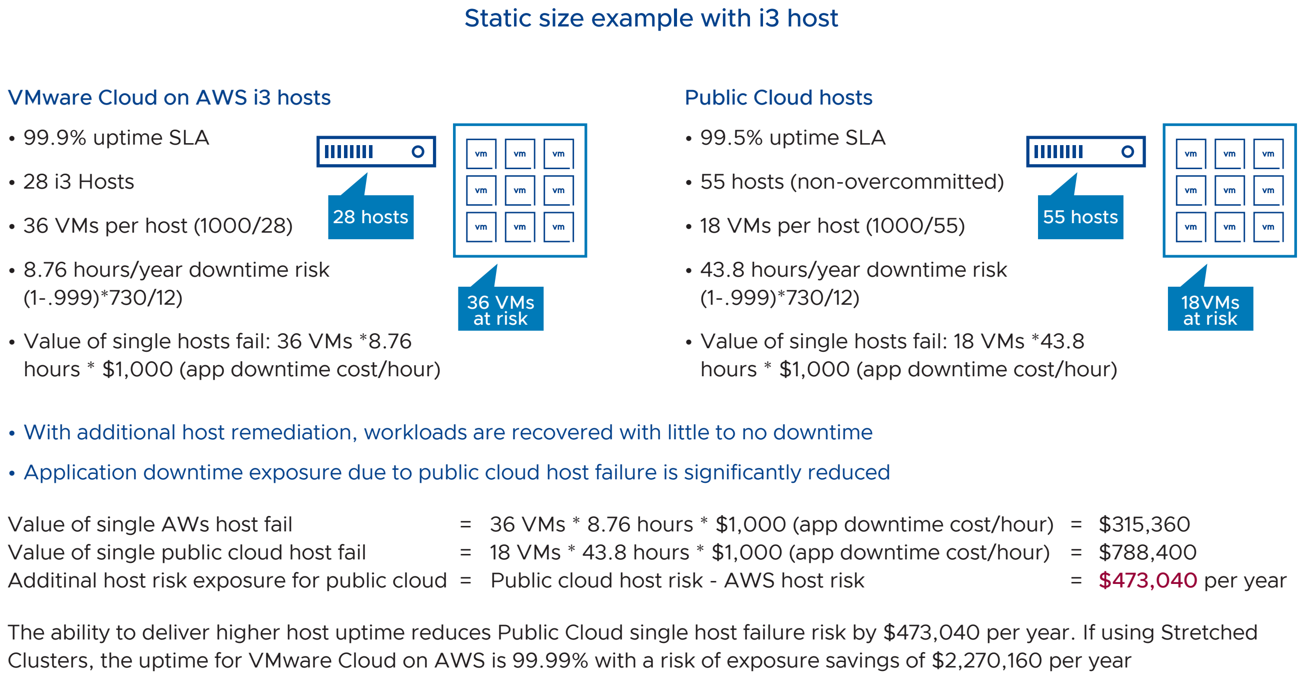
This value represents the total savings recognized by having the ability to overcommit the memory on the physical costs effectively, thereby allowing you to deliver the same work with approximately 27% fewer hosts.

Availability

Another vector of indirect savings in the VMware Cloud platform is that several features allow customers to prepare for, and avoid, system failures using several failure mode features in the VMware Cloud platform. This section will discuss three availability features that can lower total overall costs by limiting costs related to outages. These features include Automated Host Remediation, Application Mobility and Stretch Cluster.

Automated host remediation

One such availability feature is Automated Host Remediation, described in Figure 15, below.



Static size example with i3 host

VMware Cloud on AWS i3 hosts

- 99.9% uptime SLA
- 28 i3 Hosts
- 36 VMs per host (1000/28)
- 8.76 hours/year downtime risk (1-.999)*730/12)
- Value of single hosts fail: 36 VMs *8.76 hours * $1,000 (app downtime cost/hour)

Public Cloud hosts

- 99.5% uptime SLA
- 55 hosts (non-overcommitted)
- 18 VMs per host (1000/55)
- 43.8 hours/year downtime risk (1-.999)*730/12)
- Value of single hosts fail: 18 VMs *43.8 hours * $1,000 (app downtime cost/hour)

- With additional host remediation, workloads are recovered with little to no downtime
- Application downtime exposure due to public cloud host failure is significantly reduced

Value of single AWs host fail	=	36 VMs * 8.76 hours * $1,000 (app downtime cost/hour)	=	$315,360
Value of single public cloud host fail	=	18 VMs * 43.8 hours * $1,000 (app downtime cost/hour)	=	$788,400
Additinal host risk exposure for public cloud	=	Public cloud host risk - AWS host risk	=	$473,040 per year

The ability to deliver higher host uptime reduces Public Cloud single host failure risk by $473,040 per year. If using Stretched Clusters, the uptime for VMware Cloud on AWS is 99.99% with a risk of exposure savings of $2,270,160 per year

Figure 15: Automated host remediation

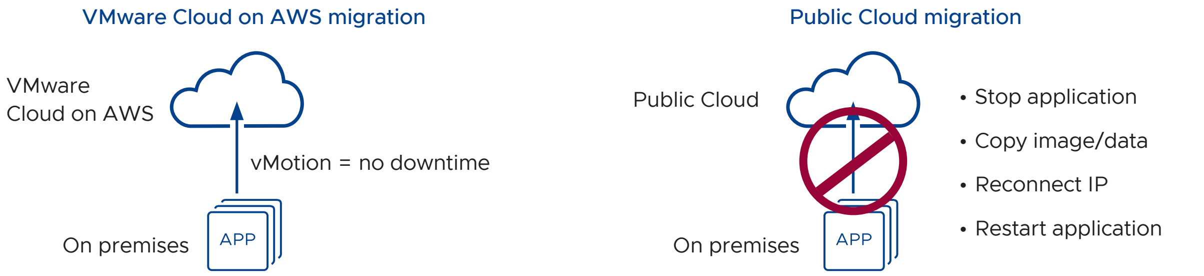
The value of the VMC automated host remediation is based on the uptime SLA differences between VMC and standard public cloud traditional instances. In this benefit analysis, we compare the standard VMC 99.9% uptime SLA with the standard public cloud uptime SLA of 99.5%. Based on 12 months, with 730 hours per month, we compute the number of hours of expected downtime for each — 43.8 hours per year for public cloud and 8.76 hours per year for VMC.

We then compute the total outage exposure as the difference in these two values multiplied by the number of VMC hosts required for a conservative estimate. The total exposure benefit then is determined by multiplying those total outage hours by the downtime cost per hour ($1,000 per hour by default). In the above example, the potential risk mitigated by the included VMC automated host remediation ability equates to approximately $473K per year.

A key benefit of VMC is that if a host fails, all the workloads on that node are recovered within minutes with no service interruption. Should the public cloud node fail, all instances on that host also would fail without notice or uninterruptible failover in many cases. Furthermore, because the instance placement on hosts is abstracted away from the customer, you may have no idea what workloads would be impacted by a node failure, potentially cascading into other connected instances not on the failing node.

Application mobility

Another availability feature in the VMware Cloud Platform centers around the application mobility functionality.



Outage time = 1000 wokloads * Avg size (300GB) / transfer speed GB.hour (1024/1024*60*60)
= 1000 * 300 / 3600
= 83 Hours

- vMotion migrates the workload with no downtime
- Time to move is based on the average amount of application data and the data transmission speed

Value of application mobility = Migration outage time * $1,000 (app downtime cost/hour) = $83,000

The ability to move workloads without interruption potentially saves $83,000 in downtime over the migration period

Figure 16: Application mobility

The ability of vSphere to move operating workloads without service interruption is a key benefit of VMC. We determine the value of this benefit by computing the amount of time it would take to move a workload to the cloud based on transmission speed, the average cost of downtime per hour ($1,000 per hour by default) and the number of instances to move.

Stretched cluster

In the example above, we estimate the downtime savings avoided by having vMotion ability at 83 hours of migration transfer time for 1,000 rehosted workloads, or approximately an $83K potential downtime cost.

The ability to stretch clusters is unique to VMC on AWS and allows nodes to be replicated across physically separate data centers. This is done so that if one data center goes off line, those workloads will be automatically shifted to the remaining data center. The benefit for VMC is that this capability only requires the subscription of stretched nodes and inter-region data transmission costs, while dramatically lowering the time and cost for outages.

For a fair comparison, we assume that the same number of public cloud nodes will be required to match the VMC node count.

Delivering a similar, but likely inferior, capability on public cloud would require the purchase of additional instances by the node, transmission costs and complex software to provide the connectivity fabric.

We compute the number of additional instances needed by multiplying the number of nodes that are stretched by the number of instances estimated per public cloud node. The value of this capability is then computed as the product of the additional instances and the TCO per public cloud instance cost.

This benefit is shown for illustration as duplicating stretched cluster operational functionality may be impractical and unreliable.

Conclusion

The only real way to understand the cost of any Cloud project is to calculate the total cost of ownership across the life of a project. To fairly compare the VMware Cloud platform against a public cloud, we need to make an apples-to-apples comparison. In the paper, we have shown that comparison and proven that public cloud imposes many costs on the user beyond the costs of the compute portion of a cloud project. Consequently, a cloud offering that provides a bundle of services in a manner that provides a consistent platform, like VMware Cloud, clearly shows a lower total cost of ownership.

VMware Cloud Economics: get your own TCO model

If you have detailed questions about the math behind our model, or how to get a model for your business, please contact us at cloud-economics@vmware.com, or talk to your VMware Account Manager and have them ask for a Cloud Economics analysis.

Acknowledgements

The authors of this paper, from VMware's Cloud Economics team, are Brandon Da Costa, Craig Stanley and Bill Roth. The authors wish to thank the following people at VMware for their advice and advance reviews: Adam Osterholt, Drew Mazeitis, Sandeep Sharma, Ron Fuller, Jake Bloom and Mark Fleischman.

[1] n=74. Based on last 74 Econ Analysis with reliablePublic Cloud data with 3o outliers filtered out. Pulled 6/16/2022, Craig Stanley.

[2] VMware Cloud Economics Model Database, Pulled May 26, 2022, n=152. Modeled off of most recent final models with valid data, since Feb 2021.

[3] HCX web page: https://www.vmware.com/products/hcx.html, references May 2022.

[4] When we refer to NSX, we are referring to something called NSX-T, which is described at https://docs.vmware.com/en/VMware-NSX-T-Data-Center/index.html

[5] Iv VMware Cloud on AWS networking and security product documentation: https://docs.vmware.com/en/VMware-Cloud-on-AWS/services/com.vmware.vmc-aws-networking-security/GUID-658253DB-F384-4040-94B2-DF2AC3C9D396.html

[6] V Blog Post by Brett Valentine: https://securityintelligence.com/posts/making-the-case-for-network-segmentation-in-aws/

Item No: Comparing VMware Cloud to Traditional Public Cloud 11/22