MULTI-CLOUD STRATEGY REVIEW

Multi-Cloud Adoption Strategy

VMware Recommendations and Considerations

Andrew Richardson Senior Client Solutions Architect, Australia & NZ
Dominic Boldeman Senior Client Solutions Architect, Australia & NZ
Sanjay Yadave Vice President, Cloud, Asia Pacific & Japan

April 2023
V3.0

Table of contents

1 Executive Summary

Over the past decade, challenger brands have shaken up the status quo in the business world by pioneering a digital-smart strategy, centered on accelerating innovation via software in the cloud. There are unique qualities that distinguish a digital-smart business: an ability to harness machine learning to turbocharge business insights; a focus on taking automation to the next level; and an obsession with driving the next leap forward in employee engagement and productivity. [1]

Enterprises and established market incumbents are heavily investing across every industry sector to take advantage of the benefits provided by cloud computing, to maintain pace, increase agility and deliver business advantage. VMware's customers report that 76% of application initiatives are being focussed on modernization, and VMware foresees a 26% growth in cloud workloads by 2024[2]. To lead the modernization of the enterprise, CIOs today must be able to:

1. Deliver modern applications at the speed the business demands,
2. Operate across any cloud with the flexibility to run applications in the data centre, at the edge or in any cloud,
3. Drive rapid business transformation while delivering enterprise level resiliency, security, and operations.

However, enterprises face a key challenge as they seek to transform and modernize of large parts of their organization's IT environment: the misalignment between the time needed to modernize applications and the short/medium-term deadlines for migration. Events such as data centre contract renewals create hard deadlines for the exit of an on-premises infrastructure with a migration to public cloud, yet application modernization is a timely exercise that will deliver long-term success in driving down IT costs, improving IT agility and creating new digital revenue streams. Balancing application modernization and transformation with migration velocity is the key to success.

The recommendations in this whitepaper are formulated from observations made through assisting numerous customers across the globe with this challenge. These recommendations will help organisations to successfully accelerate a migration from an on-premises infrastructure to one or more public cloud platforms; by strategically and pragmatically balancing application modernisation with migration velocity, and highlighting the supporting IT services and capabilities which require evolution or transformation to ensure success.

Through VMware's perspective on a successful multi-cloud strategy, twelve detailed recommendations have been formulated which can be grouped into five key recommendation areas:

1. Understand the availability requirements of your applications
2. Respect application boundaries
3. Minimise transformation during migration
4. Develop treatment plans for supporting IT services and capabilities to support migration and transition to multi-cloud
5. Develop an operational model for multi-cloud

[1] *From Cloud Chaos to Cloud Smart, Raghu Raghuram November 4, 2022*
[2] *VMware Market Insights Study, March 2021, based on research of 1,200 organizations globally*

Key to VMware's approach and recommendations is the "Relocate" cloud migration strategy which is unique to workloads hosted on VMware vSphere on-premises. The Relocate strategy extends upon the well-known "five R's" strategies for cloud migration outlined by Gartner in 2010[3] and modified by others in the years since[4]. Where other strategies (including "Rehost", "Replatform", "Repurchase", and "Re-factor") trade simplicity (and therefore velocity) against transformation to differing extents, "Relocate" is by far the simplest as it involves the migration of workloads to public cloud without changing its form-factor or network identity (IP address). This is achieved through migrating to a VMware Cloud on Public Cloud destination. By requiring the least transformation and change of all migration strategies, the Relocate strategy allows organisations to accelerate application modernisation by getting workloads to the cloud and leveraging cloud native services without intensive up-front transformation.

1.1 Customer Migration Journey

Drawing on learnings and best practice across the globe, VMware sees a common path that customers take on their cloud journey. It is possible to define the journey into 3 stages: (1) Discovery (2) Planning (3) Execution. This provides a useful frame of reference but acknowledges that each customer is different in their own way, and experience shows that some customers deviate from this model based on their specific requirements, the pace at which they are adopting cloud, the internal obstacles to be overcome, and the complexities of their organizations. However, this approach captures the key activities on the critical path to success, that seek to avoid project stalls and pre-empts blockers.

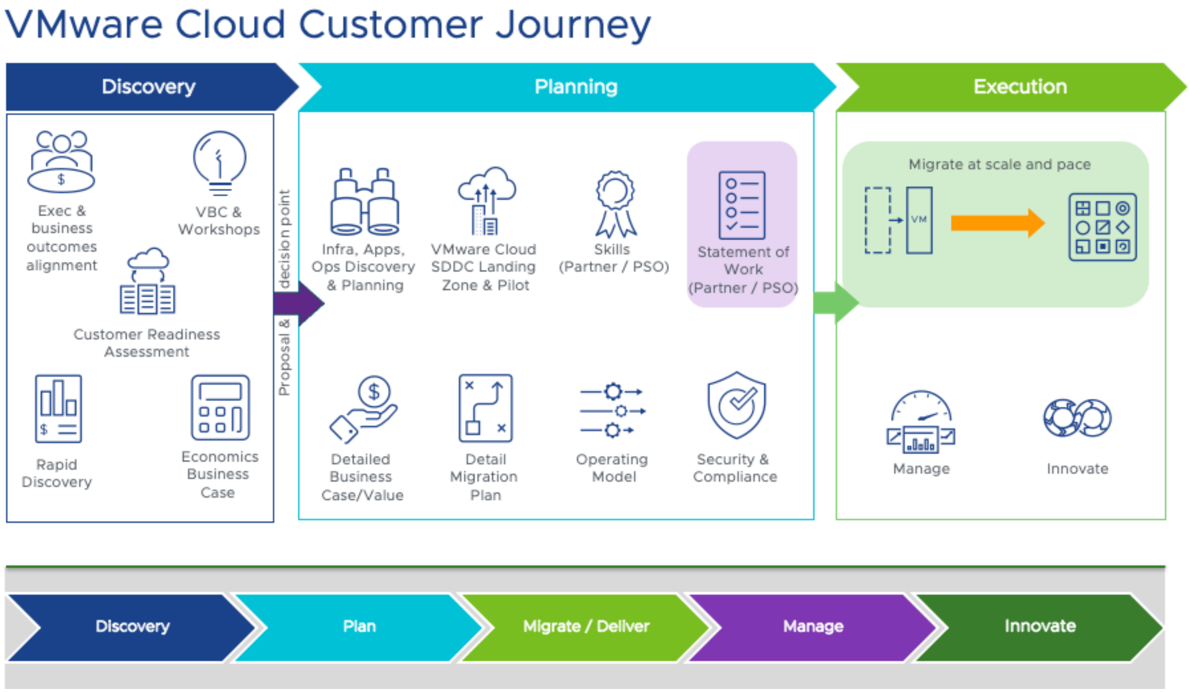
VMware Cloud Customer Journey
Discovery
Planning
Execution
Exec & business outcomes alignment
VBC & Workshops
Customer Readiness Assessment
Rapid Discovery
Economics Business Case
Proposal & decision point
Infra, Apps, Ops Discovery & Planning
VMware Cloud SDDC Landing Zone & Pilot
Skills (Partner / PSO)
Statement of Work (Partner / PSO)
Detailed Business Case/Value
Detail Migration Plan
Operating Model
Security & Compliance
Migrate at scale and pace
VM
Manage
Innovate
Discovery
Plan
Migrate / Deliver
Manage
Innovate

Figure 1 – VMware Cloud Customer Journey

[3] Richard Watson, Migrating Applications to the Cloud: Rehost, Refactor, Revise, Rebuild, or Replace, Gartner.com, 2010, https://www.gartner.com/en/documents/1485116

[4] Stephen Orban, 6 Strategies for Migrating Applications to the Cloud, AWS Cloud Enterprise Strategy Blog, 2016, https://aws.amazon.com/blogs/enterprise-strategy/6-strategies-for-migrating-applications-to-the-cloud/

1.2 Discovery

It is crucial to emphasise that customers who make a successful transformation to cloud, first obtain senior leadership alignment and conviction on the move to cloud. Whilst this paper outlines numerous technical considerations for customers, in the Discovery stage, it is vital that key business outcomes are aligned to the cloud transformation project, with clearly defined observable and measurable metrics. Throughout the Discovery stage VMware works with customers to create a case for change and build executive support through the creation of a business case. This forms a critical asset in demonstrating business value and achieving broad-based support and typically forms the basis of project approval submitted to board or executive approvers.

In the Discovery stage VMware tools and workshops are used to obtain insights into seven business and technical dimensions which will inform the work necessary in the Planning stage to close identified gaps and build the foundations necessary for successful execution:

1. Cloud Vision and Strategy
2. Business Outcomes and Goals
3. Leadership, Governance and Processes
4. People, Tools and Enablement
5. Applications and Development
6. Infrastructure, Data and Platforms
7. Security Controls, Operations & Compliance

It is also important to identify the resources that will deliver the Planning and Execution stages in conjunction with the customers own resources. This maybe a partner, or VMware Professional Services (VMware PS).

1.3 Planning

In Planning, a deep dive is undertaken into the six business and technical dimensions that form the foundation of a successful cloud transformation or migration:

1. Application Assessment & Infrastructure planning - including SDDC and network (HCX) setup.
2. Skills – leveraging existing skills in conjunction with VMware Professional Services and Partners.
3. Detailed business case to secure Executive Support and Sponsorship.
4. Detailed Migration Planning & Governance.
5. Cloud Operating Model.
6. Security and Compliance.

Each of these streams may start and end at different times, but in VMware's experience, every cloud transformation needs to address these at some stage.

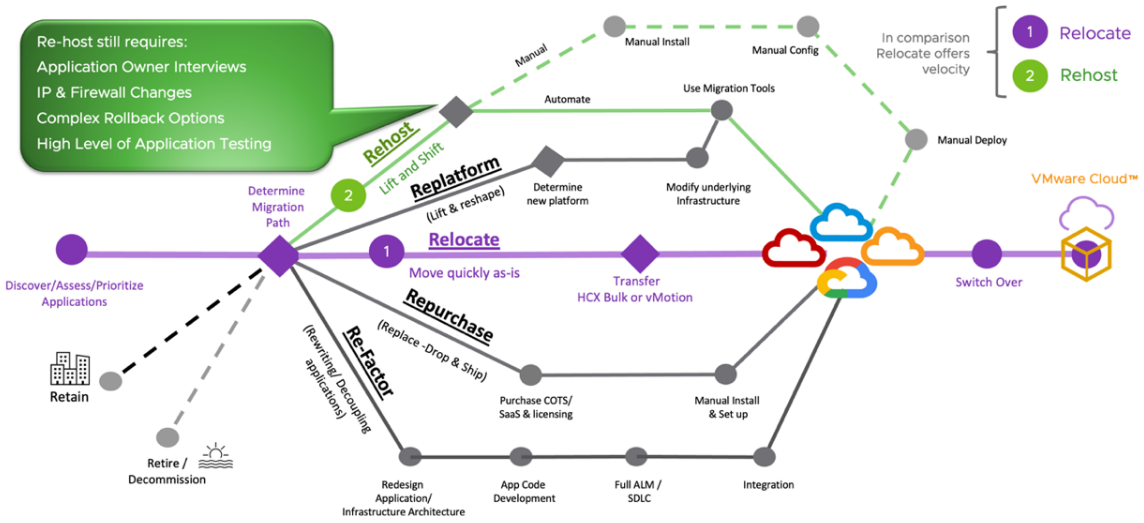

Figure 2 – Application Assessment: "Relocate" cloud migration strategy overview

1.4 Execution

At the end of the Planning stage, customers often execute a pilot migration of an initial subset of workloads. This is a common pattern of success seen with customers who achieve a high velocity in their cloud transformation or migration projects; often with their pilot transitioning to become their production environment. This exercise helps provide a live hands-on experience while helps inform the subsequent waves of migration; enabling customers execute at scale and pace.

1.5 Deep Dive workshops

To accelerate progress and remediate gaps identified in Discovery, it is recommended to undertake deep dive workshops with VMware Professional Services on key topics:

- Developing a multi-cloud strategy & roadmap
- Developing a comprehensive Cloud Architecture based on proven Reference Architectures (for Public, Private, and Hybrid cloud)
- Building a Cloud Operating Model for operations teams and undertaking an operations assessment to track alignment with the Cloud Operating Model
- Developing a migration strategy and creating and executing detailed migration plans

2 Multi-Cloud Migration: Detailed Recommendations

2.1 Understand the availability requirements of your applications.

Item	Comments
Statement	While planning a multi-cloud migration strategy, undertake an assessment of the availability requirements for all applications if this is not already known. Understanding applications' availability requirements ensures that availability is not inadvertently degraded during a migration.
Key Principles	Application availability and resiliency
Justification	The availability of applications is of key importance to a business, and one of the primary design requirements of any infrastructure platform is the ability to cater to the variety of availability requirements which exist within that organisation. Ultimately, application availability is provided by the combination of the components and capabilities of the underlying infrastructure platform and the application itself. From redundant hardware components to infrastructure capabilities like vSphere HA, vSphere Fault Tolerance, and load balancers, to application-level clustering, these components and capabilities work together in specific combinations defined by an architecture to provide required levels of availability to applications, at a trade-off of cost and complexity. However, over time it is not uncommon for the availability requirements of an application, and the architectural and infrastructure components and capabilities required to provide that availability, to become less-well documented and understood. As part of any migration, it is crucial that the availability requirements of applications, and the required architectural and infrastructure components and capabilities, are known and documented. This provides a starting point from which you can ensure that the overall application availability is not degraded inadvertently during the migration. As a simplistic example, migrating a virtual machine from a vSAN datastore with a Failures to Tolerate configuration of 2, to a vSAN datastore with a Failures to Tolerate configuration of 1, will result in an overall reduction of the availability of the application running on that virtual machine, assuming all other things (such as physical disk hardware, etc.) are equal. These kinds of oversights are usually discovered in the most painful way: through an unexpected outage to one or more applications. Being aware of the components and capabilities which contribute to an application's required availability will help to prevent these oversights from occurring during migration. **References:** *About vSphere Availability* *https://docs.vmware.com/en/VMware-vSphere/7.0/com.vmware.vsphere.avail.doc/GUID-63F459B7-8884-4818-8872-C9753B2E0215.html* *Resiliency Design Considerations and Best Practices for VMware Cloud on AWS* *https://aws.amazon.com/blogs/apn/resiliency-design-considerations-and-best-practices-for-vmware-cloud-on-aws/*

2.2 Migrate workloads based on application boundaries.

Item	Comments
Statement	While planning a multi-cloud migration strategy, undertake a detailed analysis of your application portfolio prior to determining the preferred locality of individual application components. Ensure a thorough understanding of application component interactions and dependencies is developed. Ensure migration plans consider the specific performance, availability, and support criteria requirements for each application.
Key Principles	Application Performance, Cost (Data Egress) & Vendor Support
Justification	Leveraging an application centric methodology in developing a multi-cloud migration strategy is critical to the success of the migration. While it may seem attractive to prioritise commercial considerations, such as license discounts applicable to specific cloud providers, when determining workload locality, it is equally important to understand application component interactions and dependencies to ensure that application performance and availability as well as operational support practices are not negatively impacted by decisions on locality. **Performance implications** The largest impacts to application performance are typically related to delays in retrieving data. Processors and RAM are many orders of magnitude faster that the data communication networks and data storage solutions that provide access to external data. Optimising data access latency and throughput is a critical element to achieving suitable application performance. Obviously in a scenario where application components span cloud locations, it is likely overarching application performance will suffer as time for data to traverse the networks connecting the cloud locations is likely to be significantly longer than communicating within a single cloud provider and availability zone. In response to cloud migration activities some application vendors are beginning to specify maximum network latency figures between application components to remain in a supported configuration **Cost implications** Data egress charges can represent a significant cost for organisations adopting a multi-cloud strategy. Within the context of developing an application migration strategy, any decision to locate application components between locations that incur data egress fees needs careful cost analysis. **Availability implications** There are many approaches to application availability, while some applications take responsibility for synchronising data and stateful information between application component instances themselves, others will leverage database or infrastructure services such as storage replication to provide resiliency against component failures.

It is important to understand the availability architecture of an application and its sub-components when determining locality. Inadvertently locating application components across multiple cloud providers may compromise the resiliency of the application.

Where possible avoid solutions that extend failure domains across cloud providers. For example, solutions where management or control plane instances span cloud providers can fail due to an outage occurring in a single provider.

Support implications

Incident response and troubleshooting practices require access to accurate and timely information regarding the application topology and the state of all application components and dependent services responsible for delivering the solution.

Placing application components across cloud providers expands the scope of both the data that needs to be collected and the knowledge that support staff need to address a degraded or failed application.

References:

Cloud Migration Planning

https://www.vmware.com/content/dam/digitalmarketing/vmware/en/pdf/docs/vmw-cloud-aws-cloud-migration-planning.pdf

2.3 Relocate rather than Rehost to retain IP addressing during migration.

Item	Comments
Statement	Avoid re-IP addressing workloads during migration where possible by using the “Relocate” migration strategy rather than “Rehost”. Although there are some valid justifications for re-IP addressing during migration, this changes the migration from the true “lift-and-shift” Relocate strategy to a Rehost strategy: a much more complex transformational approach which introduces countless discovery, assessment, and validation requirements into the migration activity.
Key Principles	Migration Timeframe & Cost vs Workload Placement Flexibility
Justification	**Why Relocate vs Rehost?** One of the primary elements that differentiates VMware Cloud offerings is the ability to enable customers to migrate workloads with high velocity to the cloud. This is achieved through the Relocate migration strategy depicted in figure 1 earlier in this document. This is distinct from the traditional Rehost strategy, whereby the virtual machine is converted to the hyperscaler’s native proprietary format and undergoes a change in network identity (re-IP addressing). VMware leads with this Relocate migration approach as customers often underestimate the time and complexity that goes into rehosting workloads to a cloud provider and the necessary change in network identity (IP address) as part of the migration. If a customer wants or needs to re-IP workloads they are forced to analyse each application in-depth to determine elements throughout the environment which are dependent on each workload’s IP address. These include but are not limited to: • Operating System IPv4/6 addresses • DNS Updates • Load Balancers • Firewalls • Hard-coded Application IP address references Although changing IP addresses might sound simple and, in some instances is necessary, the amount of discovery, change effort, outage length and extensive application testing often heavily impacts customer migration timeframes and causes cost blowout. As such VMware’s approach is to minimize the amount of re-addressing required to the absolute minimum. Otherwise, this often turns what should be a simple “lift & shift” migration project into a transformation project with an unknown number of dependencies and effort that is difficult to estimate upfront. VMware and International Data Corporation (IDC) in the whitepaper *The Business Value of Hybrid Cloud with VMware* estimate that customers who relocate vs rehost decrease costs by 69% and migration effort by 71%. **Caveat with network segment migrations** Although relocating workloads and retaining IP addressing dramatically simplifies the level of effort required for successful migration, there is a trade-off to note. For IP addresses to be retained during a large-scale migration, this implies that all workloads on an entire layer-2 network segment (and the network gateway) must be migrated to the same cloud location, or network segments must remain permanently extended using

VMware HCX Network Extension. The HCX infrastructure required to support network extension will in this case become a permanent and critical component of the IT infrastructure, although the recent addition of HA for Network Extension in VMware HCX dramatically reduces the operational risk of maintaining permanent network extensions,

If migrating to a single public cloud provider this is less likely to be an issue, but migrating entire segments to an individual public cloud provider as part of a multi-cloud strategy could limit migration flexibility and choice on initial migration candidates. This doesn't mean that for this approach to be successful that all workloads should be migrated to a single hyperscaler, however by the same token it relies on network segment being the primary migration criterion rather than application, workload, or software licensing. These factors can still be considered, but for migration velocity it is recommended that these factors are used as a secondary or tertiary criterion for migration bundling and wave planning. If cross-cloud portability is an ongoing requirement for workloads, consider whether a group of workloads could reside together on smaller layer-2 segments with a micro-segmentation network security posture provided by provided by NSX-T, and leveraging the abstraction provided by the VMware vSphere VM form-factor, to achieve cross-cloud portability.

References

VMC on AWS IHS Markit Case Study

https://www.vmware.com/content/dam/digitalmarketing/vmware/en/pdf/customers/vmw-ihs-markit-customer-success-summary-case-study.pdf

The Business Value of Hybrid Cloud with VMware

https://www.vmware.com/content/dam/digitalmarketing/vmware/en/pdf/docs/vmware-idc-paper-the-business-value-of-hybrid-cloud-with-vmware.pdf

Migrating and Retiring Applications AWS Overview

https://docs.aws.amazon.com/prescriptive-guidance/latest/migration-retiring-applications/overview.html

Understanding Network Extension High Availability

https://docs.vmware.com/en/VMware-HCX/4.5/hcx-user-guide/GUID-E1353511-697A-44B0-82A0-852DB55F97D7.html

2.4 Minimize transformation to workloads using infrastructure-level DR.

Item	Comments
Statement	Develop treatment plans for workloads which currently rely on a DR strategy that is incompatible with migration to public cloud. Determine whether SRM or placement on a stretched cluster is suitable for a workload given its specific DR requirements. Backup and restore of workloads for situations where DR fails should be considered as a crucial part of the overall strategy.
Key Principles	Application Availability, Cost (Data Egress) & Vendor Support
Justification	Developing a cloud-ready Disaster Recovery (DR) strategy for workloads is crucial to enable a migration to cloud. Today, there are multiple availability and DR strategies used to protect workloads within the VMware vSphere SDDC on-premises. These range from application-level strategies, where workloads are clustered at the application layer to provide an active-active or active-passive pairing, to infrastructure-level strategies, (such as replication at the storage layer, i.e. NetApp snapshots, or at the hypervisor layer, i.e. vSphere Replication) for workloads which cannot meet their availability/DR requirements through application-level configuration. Not all strategies in use today are compatible with a migration to cloud, specifically where the strategy relies on infrastructure-level capabilities (such as SAN-level replication) which do not exist in the targeted cloud platforms. For workloads relying on these incompatible availability/DR strategies, the availability and DR requirements of the workload must be re-examined, and a treatment plan must be developed to transition workloads to a supported DR strategy in the cloud. To enable selection of the right DR approach for each workload, the DR requirements of the workload must be defined. Key requirements to define for the disaster recovery treatment plan of each workload are: • Recovery Point Objective (RPO): the maximum allowable time between when a disaster occurs and the most recent recovery point that can be restored. • Recovery Time Objective (RTO): the maximum allowable time between the loss of service caused by a disaster and the return of service after recovery has been initiated. Disaster recovery scenarios: What components of the application will be protected against failures by the disaster recovery treatment plan for the workload, and at what point will the disaster recovery process be abandoned in favour of a backup/restore process. Once these requirements are identified, the disaster recovery options can be assessed for each workload. Two primary approaches exist for infrastructure-level disaster recovery of workloads running on VMware Cloud on public cloud: VMware Site Recovery Manager (SRM), and vSphere Stretched Clusters. These approaches are not exactly equivalent and offer pros and cons which are discussed below. VMware Site Recovery Manager (SRM) uses vSphere Replication to provide replication and failover of virtual machines between on-premises and a VMware Cloud SDDC, or between VMware Cloud SDDCs within a region (i.e., across Availability Zones) or across

regions. vSphere Replication supports an RPO as low as 5 minutes, up to as high as 24 hours.

Failover of workloads from one SDDC to another with SRM is initiated manually, at which point pre-defined runbooks are executed to provide asemi-automated failover process. This means the RTO can be very low (depending on the tiered priorities configured within the recovery plans), however since the recovery needs to be initiated manually this can affect the overall RTO. SRM is deployed and managed by the customer, however SaaS equivalents exist in the form of Site Recovery Manager for AVS, and VMware Site Recovery for VMC on AWS.

Benefits of SRM compared with stretched clusters are that SRM supports failover between two logically separate instances of VMware Cloud on public cloud within a public cloud provider, and SRM also supports failover cross-region.

A vSphere stretched cluster is a configuration that stretches a compute cluster with vSAN storage between two Availability Zones within a region of a public cloud provider. Stretched clusters are unique in their ability to provide an RPO of 0, as vSAN synchronously writes all data in both availability zones. Leveraging vSphere HA to restart workloads in the event of a host or site-wide outage, stretched clusters can therefore provide automatic failover of workloads within a region with an RPO of 0 and an RTO of almost 0 (in practice, the RPO is however long it takes for the virtual machines to be restarted by vSphere HA and for the OS to boot).

Benefits of Stretched Clusters compared to SRM is that stretched clusters support an RPO of 0 and near-0 RTO through automatic failover initiation via vSphere HA. Stretched clusters also enable workloads to live migrate cross-site with vMotion for disaster avoidance purposes. However, stretched clusters can only be configured within an SDDC across two availability zones within a single region (there is no cross-region support). A stretched cluster configuration is a single vSphere cluster contained within a single vCenter Server within VMware Cloud on public cloud, so a stretched cluster can be considered a single failure domain for certain kinds of failures (e.g., failures to the management plane, vCenter Server, or failures with a cluster-level scope). This may make stretched clusters unsuitable for some workloads while being highly desirable for others.

A final consideration for either disaster recovery strategy is how network connectivity is provided to virtual machines after failover. While SRM can support re-IP addressing workloads during failover, with either strategy it is highly desirable to avoid re-IP addressing by providing the networks in the recovery site required for workload connectivity after failover. In the SRM approach, orchestration should be added (either in the SRM recovery runbooks or elsewhere) to ensure that the network segments are brought up in the recovery site and routes are advertised upstream. In the stretched cluster approach, the networks will be stretched in the NSX-T overlay, and Tier-0 router failover and route advertisement will be automatic as part of the NSX-T architecture for the stretched cluster.

There are other more specific DR implementations available in certain public cloud providers which leverage SRM or stretched clusters to provide different combinations of RPO, RTO, and cost, and these can be considered once a specific cloud provider has been selected.

References

Design Considerations for Disaster Recovery with VMware Cloud on AWS | AWS Partner Network (APN) Blog

https://aws.amazon.com/blogs/apn/design-considerations-for-disaster-recovery-with-vmware-cloud-on-aws/

Deploy disaster recovery with VMware Site Recovery Manager - Azure VMware Solution | Microsoft Docs

https://docs.microsoft.com/en-us/azure/azure-vmware/disaster-recovery-using-vmware-site-recovery-manager

Disaster Recovery Solutions Brief

https://www.vmware.com/content/dam/learn/en/amer/fy21/pdf/544050_WW_21Q2_VMware_Site_Recovery_DR_as_a_Service_with_VMware_Cloud_on_AWS_Solution_Brief.pdf

2.5 Balance migration and transformation of network security policy.

Item	Comments
Statement	The desire to uplift the network security posture during a migration event should be balanced against the resulting reduction in migration velocity caused by performing transformational activities as part of the migration. The use of tools such as vRealize Network Insight can provide a middle-ground by discovering application flows and formulating firewall policy for NSX-T at different levels of granularity to enable organisations to uplift the network security posture without requiring complex and manual discovery and analysis. For policy defined on-premises in physical firewalls, consider whether an appliance-based option provided by the firewall vendor can reduce the need for policy transformation. Lastly, if required, consider how network security policy will be managed across a disparate range of firewall endpoints across multiple cloud endoints.
Key Principles	Application Security Posture versus Migration Timeframe & Migration Costs
Justification	Network Security Policy or workload firewall rules are a common area that customers need to define a detailed strategy for prior to migrating workloads to the cloud. There are three common areas that are generally considered: 1. Does migrating these workloads to the cloud require an uplift or improvement in firewall security policy? 2. If security policy is sufficient for cloud migration how can these rules be moved to cloud native controls? 3. How do I govern and orchestrate firewall policy across different platforms and providers in a centralized manner? **Uplifting Network Security Policy for Migration** A common inclusion in cloud migration projects is to assess uplifting or improving the network security posture of applications that are migrated into the cloud. This is common for applications or workloads that exist on an existing internal zone-based flat network and don’t necessarily match the network security policy approach that is used in the organization for newly created workloads in the cloud. Although the migration compelling event might be seen as a perfect opportunity to introduce modern security concepts like micro-segmentation, this is an area of migration versus transformation and more specifically in this case migration velocity versus security posture. As a general principal, applying more fine-grained firewall policy to workloads years after they have been put into production is a complex and time-consuming task that often leads to failed migrations and application outages due to incomplete rule sets. This is because documentation often doesn’t exist, or application owners are simply unaware of what rules and network flows are required to run their applications once they have already been implemented in a network with a relatively flat and open security posture. Although the easiest approach for workloads on a flat network would be to simply migrate them into a cloud without any new security policy, there is a middle ground between completely open network policies and more modern object based micro-segmentation.

vRealize Network Insight (or tools with similar functionality) can be used to understand and map network flows and application boundaries with the existing on-premises environment. vRealize Network insight provides a query language that can recommend NSX-T firewall policy at different levels of granularity based on existing network flows between applications and workloads. This can then be used to provide broad coarse-grained network policy e.g., C class to C class or B class to C class -type rules that can be applied either in distributed firewalls or boarder type firewalls.

Migrating Workloads with existing Point to Point Rules

Workloads that communicate inter-zone would commonly have a firewall as the default gateway to allow point to point security rules for an increased security posture. Depending on the nature of these firewalls (physical or virtual appliances) a migration strategy needs to be developed for how these workloads can be migrated into the cloud without a reduction in the security posture that has previously been defined. There are several options on how to achieve this, however some of the common options include:

- Migrate physical firewall appliances to virtual appliances by the same vendor and include them in the migration plan.
- Convert and replicate current firewall policy from physical firewalls to cloud platform rules prior to migration e.g., convert Checkpoint rules to NSX-T Policy.
- Analyze existing network flows and create new policy on cloud platform e.g., Use vRealize Network Insight to assess and recommend policy for NSX-T.

Governing and Orchestrating firewall policy across different providers and platforms

One of the common areas of complexity in a multi-cloud operating model is providing end-to-end security policy across different cloud providers and platforms. As an example, a new application might be provisioned within AWS EC2 and need to communicate to an existing virtual machine running in Azure VMware Solution. Such a network path could include AWS Security Groups, Physical firewalls on-premises, Azure Security Policy and VMware NSX Network Policy.

To ensure the required security posture is maintained and to reduce the operational burden associated with managing this, it is recommended than an in-depth multi-cloud framework, policy, and orchestration tooling be developed to address these use cases going forward to ensure rules can be applied programmatically regardless of the platform and within provider scaling limits.

2.6 Plan Load Balancing for Multi-Cloud.

Item	Comments
Statement	Develop a plan for providing load balancing services in multiple clouds. This will likely require a different approach to the way load balancing services are provided on-premises. In the short to medium term, during workload migrations, be aware of the impact that load balancing transformation will have on migration velocity and implement a tactical plan to get workloads migrated quickly before implementing the desired strategic solution. Post-migration, consider the operational overhead of managing different load balancing solutions in each cloud provider, and assess the benefits provided by a cloud-agnostic load balancing solution such as NSX Advanced Load Balancer (AVI).
Key Principles	Migration Timeframe vs Operational Complexity
Justification	Multiple vendors provide physical and/or virtual implementations for L4-L7 load balancing services for workloads hosted on-premises. Public Cloud providers offer their own load balancing services for workloads on their public cloud platforms, such as AWS ELB, Azure Load Balancer, and the Google Cloud Load Balancer. For customers embarking on a multi-cloud migration strategy, there are a few primary options: 1. Load Balancing infrastructure remains hosted on-premises. 2. Transition physical load balancers to virtual appliances hosted within each cloud provider. 3. Transform load balancing services to the native services offered by each cloud provider. 4. Transform load balancing services to a cloud-agnostic load balancing solution to provide consistent application delivery across multi-cloud environments. From a strategic perspective, aiming to minimise service duplication is a sound architectural goal. However, there is significant technical detail that must be considered to determine an appropriate technical solution. As such VMware will not seek to recommend a specific option from among those three within this paper. We will however offer some observations on migrating to multi-cloud. For customers with a requirement to migrate within a fixed time schedule (such as when exiting an on-premises data centre), VMware will always recommend minimising transformation of associated services. Our experience demonstrates that transformation activities represent the most significant risk to migration velocity. We have found that there are always gaps in the knowledge of application architectures and availability requirements, and the discovery, analysis, and planning activities required to provide full context for specific system configurations always results in extended delays to project timelines, particularly when services integrate with many applications. Post-migration, when transformational activities no longer present a risk to migration velocity, we recommend assessing the benefits provided by cloud-agnostic load balancing solutions such as NSX Advanced Load Balancer (AVI).

	A consistent solution for load balancing across on-premises datacentres, VMware Cloud on Public Cloud solutions, and cloud native Public Clouds, provides huge benefits to the provisioning, operating, and lifecycling of load balancing services across a multi-cloud organisation, and will simplify the activities required to undertake future migration events. Therefore, it is important to develop a strategic plan for providing and operating load balancing services in a multi-cloud context, but adopting a tactical approach to address load balancing within the context of the workload migration is recommended, minimising transformational change to maximise migration velocity.

2.7 Optimize Backup for Multi-Cloud.

Item	Comments
Statement	Assess whether backups can remain within a public cloud provider to avoid data egress costs involved with shipping backup data out of the cloud. If some backup data does need to be shipped outside the cloud provider for archival purposes, investigate whether a tiered model can be implemented to minimise data egress by keeping some backup data within the public cloud provider. If external backup storage is required, determine whether egress bandwidth will be adversely impacted by backup traffic, and whether ingress bandwidth for restore traffic will affect DR SLAs. Ensure the backup tool in use is supported by the public cloud provider, and that all operational and procedural artifacts are updated for any new tooling or processes.
Key Principles	Cost (Data Egress), Application Availability, Disaster Recovery
Justification	The design of the backup service which protects VMs running in the cloud has important implications to the disaster recovery strategy for workloads in the cloud, to cost optimisation of running and maintaining the workloads, and potentially to the migration process itself. The first consideration is where the backups should be archived. A common approach taken by customers is to store backups on native storage within a provider's public cloud platform. The advantage of this approach is avoiding data egress costs involved with shipping backup data off-site, and the high-performance networking within the public cloud platform available to restore backups if required. The downside of this approach is that all backups are ultimately stored within the provider's public cloud platform. The second consideration is what backup tool should be used, and whether this requires a change to the tooling used on-premises today. Many backup tools provide support for VMware workloads on public cloud, but not all tools that traditionally can be used in the datacentre are supported. Changing tools may also require touching each VM during the migration process if an in-guest agent is required, as the migration process must cater for the uninstallation of existing agents used on-premises and the installation and configuration of agents used in the destination VMware Cloud on public cloud environment. Regardless of the approach taken with backup archival and tooling, existing policies and procedures will need to be comprehensively reviewed to ensure they make sense for a workload which has migrated to the cloud, and that the backup/restore process complements the disaster recovery approach as part of the overall DR strategy for each workload.

2.8 Optimize Logging for Multi-Cloud.

Item	Comments
Statement	Assess the magnitude of log data being ingested into any SIEM solution and calculate likely data egress costs specific to logging as part of developing your cloud migration plans. If these costs are determined to be prohibitive then develop a federated logging architecture.
Key Principles	Cost Optimisation
Justification	Aggregation and analysis of logging data is a foundational element of IT Service Management practices. Most IT solution components log large amounts of data to support troubleshooting and analysis of a systems operational state and to support security audit activities. Traditionally, many organisations establish centralised log repositories which are then used to support operational requirements, typically including security incident and event management and incident and problem troubleshooting and analysis processes. The challenge that a multi-cloud strategy presents to centralised log aggregation is the data egress charges associated with transporting log data to the centralised log aggregator solution. VMware's experience has demonstrated that many of our customers haven't adequately considered the impact of data egress charges specifically within the context of log aggregation and are often surprised at the cost impact. This typically results in tactical decision making aimed at reducing cost with the risk that operational practices and security incident and event management practices are compromised. A typical solution VMware has helped customers to implement is the establishment of a federated logging architecture whereby the default position is that log data is aggregated and stored within each cloud provider to support operational troubleshooting activities with specific security related data forwarded to a dedicated SIEM solution.

2.9 Decouple database modernisation from Cloud migration.

Item	Comments
Statement	Adopting DBaaS services provided by public cloud providers is often a key motivator driving migration to cloud. Although this transition may be the desired strategic end-state for database consumption, it is another example of a transformational activity and it is recommended that re-platforming to DBaaS offerings (or database modernization) be decoupled from migration activities such as rehost/relocate when migration deadlines are an important factor.
Key Principles	Migration Timeframe & Cost vs Day-2 Operational Benefits
Justification	A common driver and benefit that IT organizations target when migrating to the cloud is the consumption of Database as a Service (DBaaS) offerings vs in-house managed database platforms used on-premises. DBaaS offerings like AWS RDS, Azure SQL Managed or Google Cloud SQL simplify day-2 lifecycle management and operations of a critical component of the ICT landscape. A key value proposition of VMware Cloud is enabling customers to decouple application modernization from infrastructure modernization, and then leveraging a cloud hosted model to later accelerate application modernization. This simply means re-platforming to DBaaS offerings could be run in parallel to migration activities, however it should be done pre- or post-migration and therefore it should not be a dependency or on the critical path for the migration activity. This is because re-platforming databases to managed cloud offerings is more of an unknown in terms of application compatibility and effort than a simple vMotion of an entire VM to a VMware cloud, and this unpredictability introduces timeframe and cost risk. Re-platforming a database to a managed cloud offering varies in level of effort based on how compatible that application is with a cloud database. For example, moving an Oracle DBMS 12c database to a more modern Oracle or PostgreSQL platform in AWS may introduce application compatibility issues that require changes to the application. Although these issues need to be resolve eventually, placing these transformational re-platforming activities on the critical path can cause the entire program to be delayed, especially if a network segment migration approach is also being used. Although it may be seen as “double-handling” VMware’s recommended approach is to focus on the critical path and the compelling migration deadlines as the main priority and Database modernization as secondary. This means running DB re-platforming as a separate stream to relocate (migrations) and if a database can’t be re-platformed in time for a migration wave the existing database VM is first moved to the desired cloud and then re-platformed later. It is also beneficial to adopt this approach as if all the elements of an application have already been migrated to the cloud, database latency and vendor support become less of an issue for certain applications.

2.10 Undertake Operational Readiness Assessment prior to Production migrations.

Item	Comments
Statement	Undertake an assessment of your operational readiness to adopt and operate a VMware Cloud on public cloud platform prior production migrations. Although a VMware Cloud platform is vSphere-based and largely similar to an on-premises vSphere SDDC, there are some key differences from an ITSM and operational perspective that are important to be plan for. VMware has established a Cloud Operational Readiness service to help customers understand and resolve these differences.
Key Principles	Ensuring migration success beyond the initial transition
Justification	Understanding the impact to existing IT Service Management (ITSM) processes due to the transition to cloud service consumption often lags technical adoption. Specifically, within the context of a vSphere fleet migration to a public cloud provider, it has been a common experience that customers focus on the technical integration and workload migration elements without sufficiently understanding the changes to operational support processes that will be necessary to ensure successful ongoing operations. Areas of potential impact are found across many ITSM practice areas. For example, incident management processes need to consider how information regarding the status of the cloud service can be integrated with existing monitoring capabilities, how notifications regarding cloud service status can be received, interpreted, and actioned and then routed to the right resources troubleshoot and resolve. Change management often needs to adapt to SaaS service upgrade events being driven by the vendor. How does change management understand the schedule of changes, potential service impacts associated with the change and likelihood of impact to dependant systems and users? Capacity management processes provide an additional example. The VMware public cloud services don't limit customers consumption of the cloud resources. Establishing processes to understand the level of resource demand versus available capacity across compute, storage and network resources is the responsibility of the customer. Having a reactive approach to this issue may result in service degradation or disruption. Developing a plan to understand and remediate these types of elements associated with a multi-cloud strategy should be a priority. VMware has established a Cloud Operational Readiness service to help customers identify gaps specifically related to their vSphere fleet migration to public cloud. In a structured engagement we work with the customer's ITSM stakeholders and operations teams to identify any challenges or gaps in their existing ITSM and operational practices and propose treatments and process improvements to ensure they customer is prepared to successfully transition their vSphere fleet to a public cloud provider. **References:** *Organizing for the Cloud - White Paper* *https://www.vmware.com/content/dam/digitalmarketing/vmware/en/pdf/whitepaper/services/vmware-organizing-for-the-cloud-white-paper.pdf*

2.11 Manage efficiency and cost through right-sizing, capacity management, and Transparent Page Sharing

Item	Comments
Statement	While it can be simple for customers to maintain their existing operational practices and VM sizing when moving to VMware Cloud platforms, migration can also provide a good opportunity to consider and assess efficiencies which can reduce host footprint requirements and therefore cost. Right-sizing workloads, implementing best-practices operational processes for capacity management, and a pragmatic use of Transparent Page Sharing (TPS) can provide significant efficiency and cost benefits for the VMware Cloud platform.
Key Principles	Cost Efficiency vs Security Posture and Performance
Justification	**Right-sizing during migration** A migration of virtual machine to VMware Cloud on public cloud provides organisations with an opportunity to right size virtual machine CPU, Memory and Disk allocation. As most vSphere environments on-premises are memory-bound (unless using transparent page sharing), right-sizing allocated memory often makes a significant difference to the total hardware capacity required to support the Virtual Machine footprint. VMware vRealize Operations Manager can provide recommendations and reports for how much of each resource should be allocated to a workload without it experiencing contention. These recommendations can be adopting during or post migration. One of the key benefits of VMware Cloud offerings is the ability to scale clusters both up and down rapidly, meaning that any significant savings from right-sizing can then be used to push down the size of the underlying vSphere platform. **Move to Demand-based capacity management** Capacity management in a VMware vSphere environment has traditionally been performed via two different models: allocation-based and demand-based. Traditionally, the allocation-based model has been widely used, as it is a conservative model of capacity management which aligns to how workload resource sizing was performed prior to the widespread adoption of virtualisation. In the allocation-based model, capacity management is measured by summing the total amount of CPU, Memory, or Disk resources that have been assigned to workloads and comparing against the total amount of resources present in the underlying infrastructure, without consideration for the amount of resource each virtual machine is actually trying to use. Typically, the allocation-based model defines an overcommitment ratio for a resource – for example, a 4:1 CPU overcommitment ratio, meaning four vCPUs for every physical CPU – and free/unused capacity in an environment is calculated using these metrics. The demand-based model is a more modern approach which considers the resource demands each workload guest OS is making, with less regard to the amount they have simply been allocated (and are potentially not using). Take the 4:1 CPU ratio discussed above: using this allocation-based model, an environment with 100 physical CPUs would be at maximum capacity while running 400 vCPUs of workload.

However, if workloads over a period of time are only demanding 15% of their allocation, a demand-based model would allow those workloads to be run on less than the 100 physical CPUs dictated by the allocation-based model with a 4:1 overcommitment ratio. As the demand-based model is a more pragmatic reflection of the amount of resources workloads are demanding, this typically results in higher consolidation ratios of workloads to physical hosts, with a reduction in the number of physical hosts required to run the workloads and a lower cost overall for the infrastructure platform.

Transparent Page Sharing and Large Pages

The vSphere default behaviour of the use of intra-VM only Transparent Page Sharing (TPS) and Large Pages significantly reduces the viability of memory over-commitment for the purposes of capacity management. This in combination with oversized virtual machines can result in a large amount of waste and inefficiency in the environment if workloads are not effectively right-sized, raising the TCO.

The use of vSphere large pages is considered a performance best practice and is well described in the Performance Best Practices for VMware vSphere guide and Large Page Performance white paper. It is important to note that a performance best practice is not simply a general best practice, and performance best practices often are contrary to cost or efficiency best practices. The performance impact of disabling large pages is heavily workload-dependent and ranges from 5% - 15% with noticeable impact generally only observed where the operating systems themselves are requesting the use of large pages and are performing memory intensive activities such as code compilation.

Enabling TPS between VMs (Inter-VM TPS)

Although disabling large pages can by itself provide benefits to memory overcommitment and memory reclamation, larger benefits from an efficiency perspective come when inter-VM TPS is also enabled. Enabling Inter-VM TPS in conjunction with disabling large pages allows vSphere to share identical memory pages between virtual machines which provides a primary benefit of saving empty memory pages or memory whitespace within VMs with the same guest OS. This can be a potentially significant saving. The use and benefit of TPS is more fully described in the Understanding Memory Management in VMware vSphere white paper. Note that Inter-VM TPS was disabled by default in vSphere 5.x for security concerns, however the *Security considerations and disallowing inter-Virtual Machine Transparent Page Sharing* knowledge base article makes the following comment:

"Published academic papers have demonstrated that by forcing a flush and reload of cache memory, it is possible to measure memory timings to try and determine an AES encryption key in use on another virtual machine running on the same physical processor of the host server if Transparent Page Sharing is enabled between the two virtual machines. This technique works only in a highly controlled system configured in a non-standard way that VMware believes would not be recreated in a production environment.

Even though VMware believes information being disclosed in real world conditions is unrealistic, out of an abundance of caution upcoming ESXi Update releases will no longer enable TPS between Virtual Machines by default."

With above considerations in mind, it is a customer's choice of security and performance versus efficiency and TCO. Often, customers choose the option of efficiency for non-production or general-purpose workloads, and performance/security for mission critical or production applications.

References

Performance Best Practices for VMware vSphere® 7.0 - Pg 35 - 2MB Large Memory Pages

https://www.vmware.com/content/dam/digitalmarketing/vmware/en/pdf/techpaper/performance/vsphere-esxi-vcenter-server-70-performance-best-practices.pdf

Large Page Performance

https://www.vmware.com/techpapers/2008/large-page-performance-1039.html

Understanding Memory Resource Management in VMware vSphere 5.0

https://www.vmware.com/content/dam/digitalmarketing/vmware/en/pdf/techpaper/vmware-understanding-memory-resource-management-in-vsphere5.pdf

Security considerations and disallowing inter-Virtual Machine Transparent Page Sharing (2080735)

https://kb.vmware.com/s/article/2080735

Right Sizing VMs with vRealize Operations

https://blogs.vmware.com/management/2020/01/rightsizing-vms-with-vrealize-operations.html

Allocation and Demand Models in vRealize Operations

https://blogs.vmware.com/management/2019/05/allocation-model-for-capacity-management-in-vrealize-operations-7-5.html

2.12 Application Licensing and Support Considerations.

Item	Comments
Statement	Where possible, keep applications together during a migration to cloud, and where a decision is made to fragment an application, ensure that the impact to performance, availability, and troubleshooting is quantified and accepted. Although significant cost savings can be realised through workload placement with licencing costs as the primary driver, the impact to application availability and performance should not be underestimated. The cost savings for fragmenting an application across clouds can be redundant when this fragmentation does not provide an application with the performance or availability it needs to function.
Key Principles	Total Cost of Ownership, Workload Placement Flexibility, Vendor Support
Justification	Licencing and vendor support are crucial considerations for defining workload placement across public clouds. Licencing constraints placed by software vendors on where you can run your workloads without re-purchasing licences can be a significant contributor to the cost of operations in the cloud, and careful consideration of workload placement is an important part of cloud cost optimisation. Oracle and Microsoft both offer enticing licencing cost benefits for placing those vendor's workloads on their own public cloud platforms. Oracle software licencing agreements make it twice as expensive to run a workload in AWS or Azure compared to on-premises or in Oracle Cloud. Microsoft requires customers to repurchase any licencing owned as part of an ELA post-2019 when workloads are moved to a non-Azure cloud, and provide free extended security updates for certain workloads (e.g. Windows Server 2012) when running in Azure. The cost optimisations possible by selectively placing workloads on certain clouds must be carefully balanced against the availability and performance requirements of your applications, however. Each node of an application has a web of dependencies to other endpoints both internally (such as other application nodes) and externally, such as other applications and shared infrastructure services. Applications which have grown up in the on-premises datacentre can be sensitive to latency as they have always enjoyed low latency and high-throughput connectivity. The performance and availability implications to the functionality of applications when latency is suddenly introduced is often unknown and hard to quantify. If an application is fragmented across multiple clouds following a rigid cost-optimisation placement policy, the introduction of latency that will affect the availability and performance of the app is a significant concern and troubleshooting a performance problem across multiple clouds can be an extremely complex task. Vendor support or lack thereof for certain clouds can influence or prohibit certain workload placement options. Examples include Oracle RAC support on VMC on AWS, or conversely, SAP workloads being unsupported on VMC on AWS. Where vendor support is not currently provided, support statements should be clarified through discussion with the vendor.

References

Licensing; Oracle Software; Cloud Computing Environment

https://www.oracle.com/us/corporate/pricing/cloud-licensing-070579.pdf

Cost savings – SQL Server & Windows Server | Microsoft Azure

https://azure.microsoft.com/en-au/overview/azure-vs-aws/cost-savings/

SAP on VMware Cloud on AWS - Migration Options

https://blogs.vmware.com/apps/2019/07/sap-on-vmware-cloud-on-aws-migration-options.html

Oracle RAC on VMware Cloud on Amazon AWS - Virtualize Applications

https://blogs.vmware.com/apps/2017/11/oracle-rac-vmware-cloud-amazon-aws.html

VMware Cloud on AWS for Oracle - AWS Prescriptive Guidance

https://docs.aws.amazon.com/prescriptive-guidance/latest/migration-oracle-database/vmware-oracle.html

3 Glossary of Terms.

Term	Description
Private Cloud	A cloud computing model where the infrastructure is dedicated to a single user organisation. A Private Cloud can be hosted either on-premises, a third-party colocation facility, or via a private cloud provider.
Hybrid Cloud	A cloud computing model that uses a combination of at least one private cloud and at least one public cloud.
Multi-cloud	The operation and consumption of applications running across public clouds, private clouds, and the edge.
Public Cloud Provider / Hyperscaler	Public cloud providers such as Amazon Web Services (AWS), Google Cloud, Microsoft Azure, and Oracle Cloud Infrastructure.
VMware Cloud on public cloud / VMware Cloud	Any VMware Cloud within a public cloud provider such as Azure VMware Solution (AVS), Google Cloud VMware Engine (GCVE), Oracle Cloud VMware Solution (OCVS), and VMware Cloud on AWS (VMC on AWS).
VMware vSphere SDDC	A software-defined data Centre consisting of VMware vSphere, VMware vSAN, and VMware NSX providing virtualisation of compute, storage, and networking respectively. Also used more broadly to refer to VMware vSphere hypervisor platforms located on-premises.
VMC on AWS	VMware Cloud on AWS
AVS	Azure VMware Solution
GCVE	Google Cloud VMware Engine
OCVS	Oracle Cloud VMware Solution
ITSM	IT Service Management
VMware PS	VMware Professional Services
Micro-segmentation	Micro-segmentation is a network security technique that enables security architects to logically divide the data centre into distinct security segments down to the individual workload level (i.e. each individual VM or container), and then define security controls and deliver services for each unique segment.
Right-sizing	Ensuring the resources allocated to a workload are correctly sized (i.e. not over- or under-sized) to provide the most efficient use of physical resources (particularly CPU and RAM) while ensuring application SLAs are met. Typically referring to workloads which are over-sized, which can cause a detrimental impact to the performance of neighbouring workloads, and results in an inefficient utilisation of physical resources.
TPS	Transparent Page Sharing is an optional feature of the VMware vSphere hypervisor that allows identical memory pages across multiple workloads to be shared and deduplicated at the physical memory layer.

VMware, Inc. 3401 Hillview Avenue Palo Alto CA 94304 USA Tel 877-486-9273 Fax 650-427-5001 vmware.com