Implementing a Hybrid Cloud Strategy

Using vCloud Air, VMware NSX and vRealize Automation

TECHNICAL WHITE PAPER

vmware®

Table of Contents

Purpose and Overview

This white paper demonstrates how to successfully implement a hybrid cloud strategy and aims to provide a better understanding of its use case and its potential. The scenario described in this white paper is based on a successful real-world customer implementation by VMware employees who are part of CTO Ambassador group.

The VMware Office of the CTO runs the CTO Ambassador program. The CTO Ambassadors are members of a small group of our most experienced and talented customer-facing, individual contributor technologists. They are pre-sales systems engineers (SEs), technical account managers (TAMs), professional services (PSO) consultants, architects and global support services engineers. The ambassadors help to ensure a tight collaboration between R&D and our customers so that we can address current customer issues and future needs as effectively as possible.

Executive Summary

IT has long debated the merits of public and private cloud. Public clouds allow organizations to gain capacity and scale services on-demand, while private clouds allow companies to maintain control and visibility of business-critical applications. But there is one cloud model that stands apart: hybrid cloud.

Hybrid clouds provide the best of both worlds: secure, on-demand access to IT resources with the flexibility to move workloads onsite or offsite to meet specific needs. It's the security you need in your private cloud with the scalability and reach of your public cloud. Hybrid cloud implementations should be versatile, easy to use, and interoperable with your onsite VMware vSphere® environment. Interoperability allows the same people to manage both onsite and offsite resources while leveraging existing processes and tools and lowering the operational expenditure and complexity.

Customers have acknowledged the following five key use cases for a hybrid cloud implementation.

- **Development and Testing**
  Hybrid cloud provides businesses with the flexibility to gain needed capacity for limited time periods without making capital investments for additional infrastructure.
- **Extending Existing Applications**
  With hybrid cloud, businesses can extend current standard applications to the cloud to meet the needs of rapid growth or free up onsite resources for more business-critical projects.
- **Disaster Recovery**
  Every organization fears an outage, or outright loss, of business-critical information. While onsite disaster recovery solutions can be expensive, preventing businesses from adopting the protection plans they need, a hybrid cloud can offer an affordable disaster recovery solution with flexible commitments, capacity, and cost.
- **Web and Mobile Apps**
  Hybrid cloud is ideal for cloud-native and mobile applications that are data-intensive and tend to need the elasticity to scale with sudden or unpredictable traffic spikes. With hybrid cloud, organizations can keep sensitive data onsite and maintain existing IT policies to meet the application's security and compliance requirements.
- **Development Operations**
  As developers and operations teams work closer together to increase the rate of delivery and quality of deployed software, a hybrid cloud allows them to not only blur the lines between the roles, but between Dev/Test and production, and between onsite and offsite placement of workloads.

VMware vCloud® Air™ is a public cloud platform built on the trusted foundation of vSphere, compatible with your on-premises data center, that includes infrastructure, disaster recovery, and various application service offerings. vCloud Air allows you to extend your workloads into the cloud with ease. You can migrate existing onsite virtual machines (VMs) to the public cloud or develop new applications directly in the cloud. Being built on the same platform, vCloud Air extends to your private data center, branch office or vCloud Air Network partner so you can also easily port VMs and other business-critical workloads to the location of your choice—all with the secure and capable foundation of vSphere.

The extensibility of the VMware hybrid cloud provides flexible choice for organizations to optimize their workload placement based on the needs and demands of the application. By integrating vCloud Air with VMware vRealize™ Automation™, it is possible to provision these workloads on- or off-premises while maintaining consistency in deployment and experience using a familiar interface.

In this white paper the private data center is extended into vCloud Air which is used for the deployment of production and scale-out applications, as well as for a Dev/Test platform for an enterprise organization. The focus of this paper is on the developer and the challenges facing the organization's large development team. This team needs to scale in and out depending on the time of the year. The team must also release new versions of their solution on an almost daily basis. vRealize Automation is the self-service portal for developers and application owners allowing them to deploy new services when required. These services, whether on- or off-premises, are accessible in an identical way without any change in the operational model.

The Conceptual Architecture

The primary drivers for this architecture are flexibility, agility and simplicity. In order to achieve this, the architecture proposed contains an abstraction layer that is capable of orchestrating the deployment of workloads on different types of resources. In the following diagram a simple conceptual workflow is depicted.

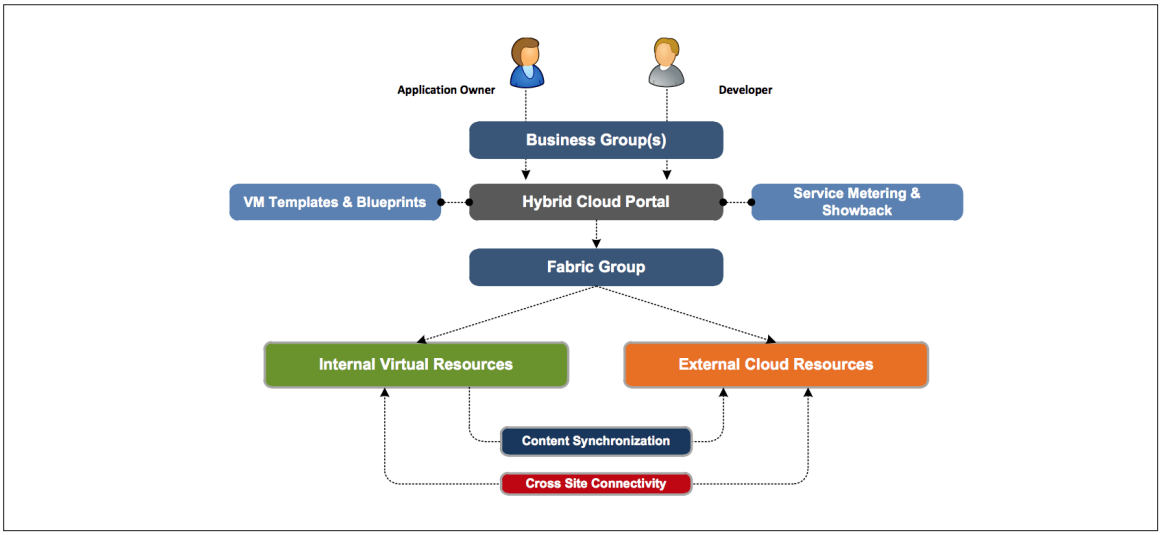

Figure 1.

There are two main types of consumers of the service, namely the developers and application owners. Both will be capable of defining their requirements when they request resources. The outcome in this scenario is one, or multiple, virtual machine(s) deployed in the location determined suitable for running these workloads by the policy engine and automation layer and, more importantly, also configured to the needs of the consumer. Besides the provisioning process itself, the whole life cycle will be managed by vRealize Automation. This includes things like the approval process, management, retirement and archival of requested services.

Creating this seamless hybrid cloud experience requires some fundamental components, including a common management and orchestration platform, unified networking, a common security model, and one place to call for support.

This means an organization can remain in control using all the same procedures and tools they already know, and the key stakeholders can get to the cloud faster without having to re-architect the application. The diagram below gives a high level overview of this integration.

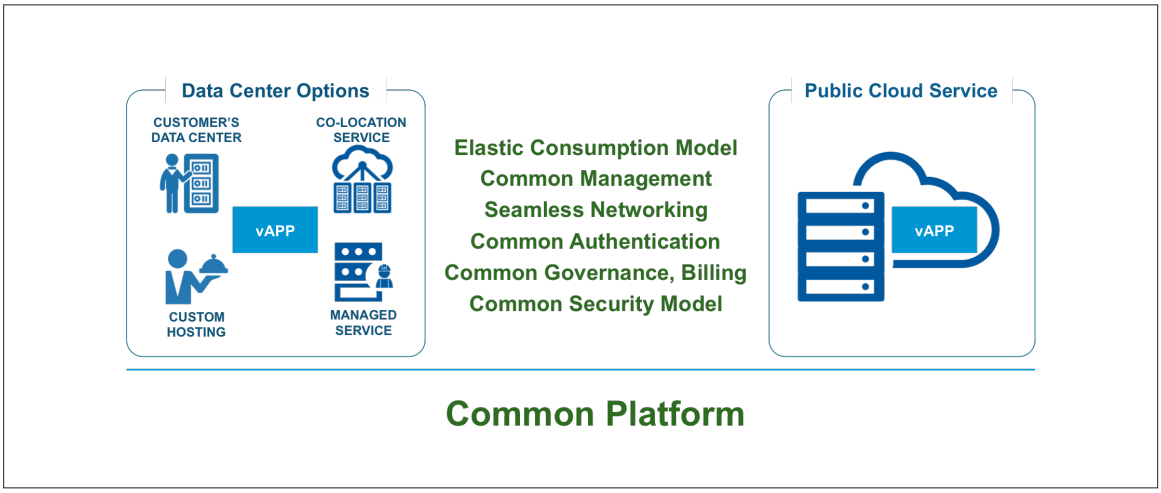

Figure 2.

Building out a hybrid cloud infrastructure is akin to building another data center—one that is virtual, but another data center nonetheless. One of the parallels is how multiple sites and clouds are connected together. Organizations have the ability to connect sites together using either an IPsec VPN or a Direct Connect private line. The decision criteria for which to use is typically based on bandwidth and security requirements and cost. If the applications deployed require high bandwidth consumption, then a 1Gbps or 10Gbps Direct Connect line provides higher throughput than traffic over the Internet. However, if lower bandwidth is acceptable, then a simple and cost effective site-to-site IPsec VPN may be preferable.

The diagram below shows at a conceptual level how a company can extend their data center to the public cloud.

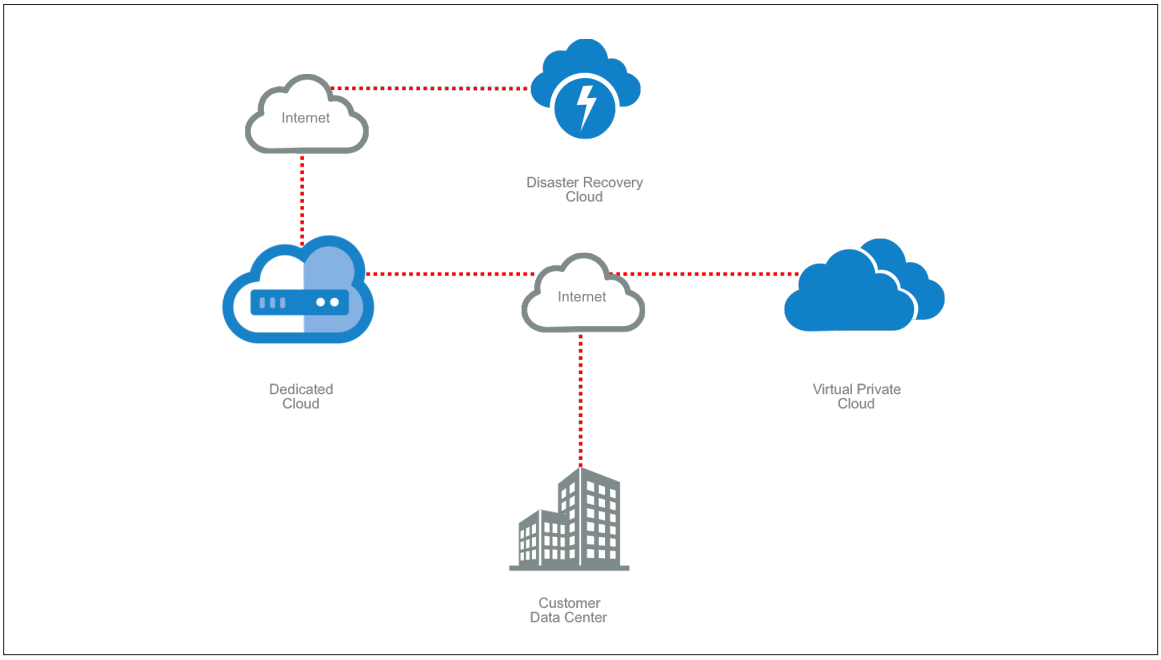

Figure 3.

This diagram shows two environments: an on-premises data center and multiple off-premises virtual data centers (clouds). These are connected via multiple VPN links across the Internet. In this scenario there is a permanent always-on link between the multiple sites. Using this approach allows you to deploy workloads in multiple locations, and communicate across the sites.

Authentication is a critical consideration when deploying workloads in this type of hybrid cloud model. If there are two application servers running off-premises in the cloud, continually authenticating against a directory service running in the on-premises data center, this authentication traffic may saturate a VPN connection. In this scenario it may be beneficial to extend the directory service to the cloud, and create a new authentication site. By doing so, there is a local authentication service for the workloads which reduces the amount of traffic travelling across the VPN connection, but the directory service remains in sync with the original one running on-premises. This technique is commonly used when building out multiple data centers, and can be replicated in a hybrid cloud.

With multiple end points in different locations, businesses offer more choice to their developers as to where they can deploy workloads. For example the diagram below depicts how vRealize Automation can deploy to different endpoints based on particular policies defined for that workload.

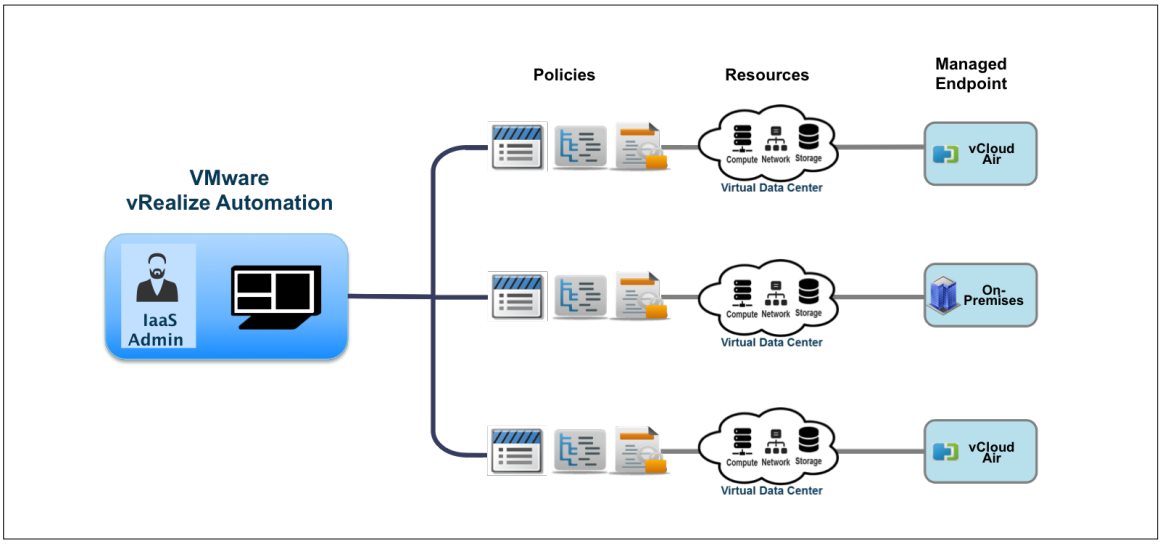

Figure 4.

As the above diagram shows, an organization can define policies that place applications across different geographical locations, local data centers, or even clouds with different performance characteristics.

Consider a business that needs to support teams across geographical locations. This company may have a physical data center in the west coast of the United States, but some of their developers may be based in the United Kingdom. By leveraging vRealize Automation and vCloud Air, this company can have a vCloud Air environment in the UK, and define a policy that only allows particular applications or workloads to be deployed in the UK. This provides greater granularity and control over how and where workloads are deployed.

Policies can also separate endpoints based on different performance characteristics. With vCloud Air, different environments can be attached to different tiers of storage; one virtual data center can be attached to high-performance storage while another virtual data center can be attached to a more economical standard tier of storage. By leveraging policies, a business can direct applications to the best location for that workload based on cost.

Logical and Physical Architecture

There are four main areas of interest in this architecture. The following diagram shows an overview of the full architecture and the four areas in a bit more detail.

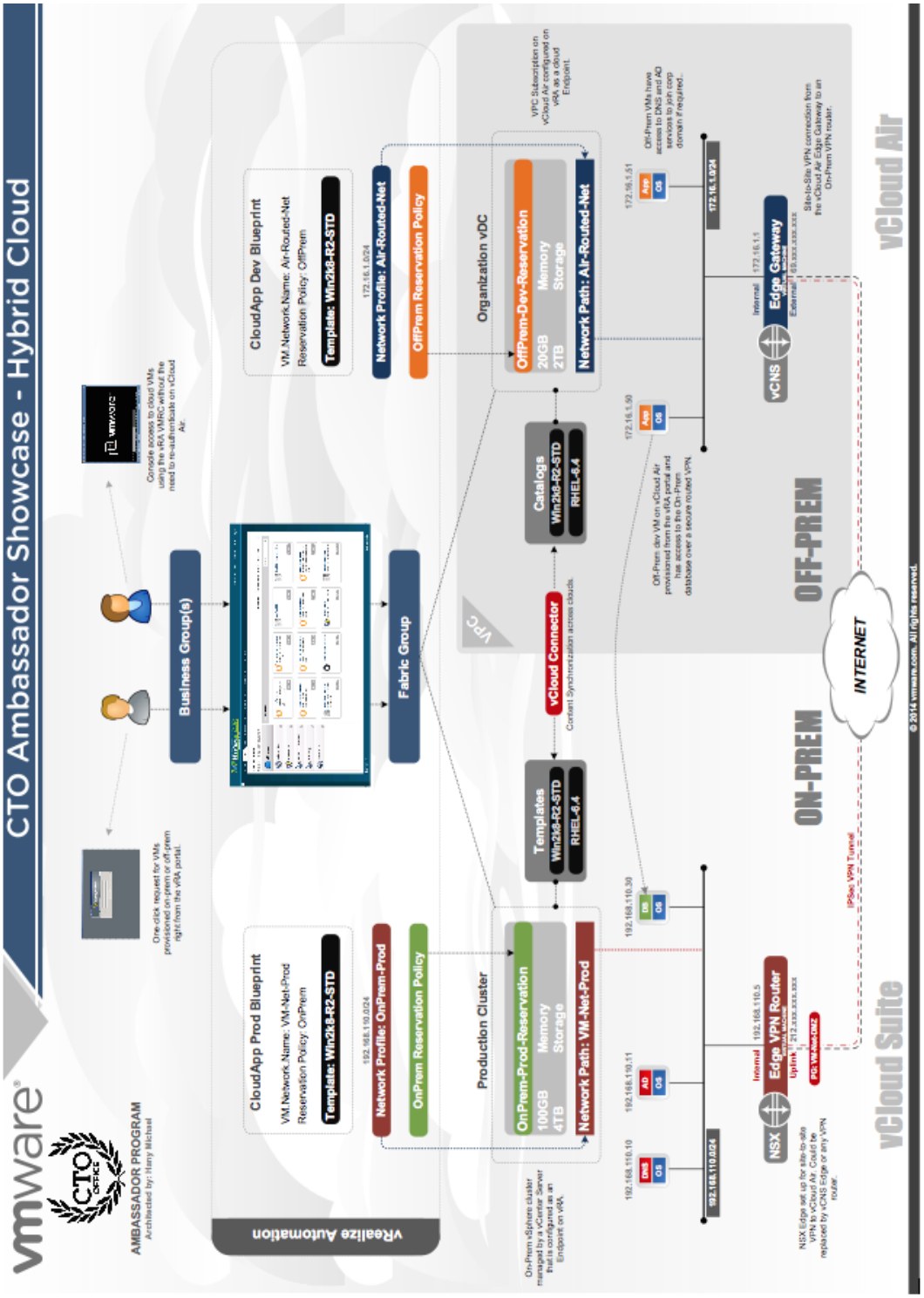

Figure 5.

Starting from the bottom, VMware NSX™ for vSphere is running a site-to-site VPN tunnel to an edge gateway on vCloud Air. This layer is used to bridge the communication between on- and off-premises environments.

On the left is a traditional vSphere infrastructure to which NSX is connected. On the right side is a Virtual Private Cloud (VPC) on vCloud Air. (The latter could be also a Dedicated Cloud on vCloud Air depending on business requirements.) At the top of the architecture is vRealize Automation, which acts as the single entry point to this hybrid cloud infrastructure. vRealize Automation contains policy-based configurations mapped to the business groups. This top-level orchestration and automation layer will enable the consumers of the service to provide their requirements and select the applicable policies and receive fully deployed and configured virtual machines as a result within minutes.

Further detail on each of the four areas follows.

The Networking Infrastructure

The underlying network infrastructure is required for cross-site communications. The workloads that are provisioned off-premises must be able to communicate to the on-premises infrastructure services. To achieve this, a site-to-site VPN tunnel between two gateways is configured. Those gateways shown in the architecture diagram are NSX 6.0 for vSphere (on-premises) and the vCloud Air edge gateway. Please note that customers can utilize VMware vCloud® Networking and Security™ (part of the VMware vCloud Suite®) edge appliances or any hardware-based VPN gateway instead of NSX if they choose.

With NSX, two interfaces are needed: one internal to the corporate network fabric and another one external to the public Internet. The latter needs to be set with a public IP-address or NAT'ed in a DMZ provided that it can receive the VPN connection requests from the internet (in this case, from the edge gateway in vCloud Air). It is recommended to leverage NSX here due to the incredible flexibility that it has when it comes to the design and deployment.

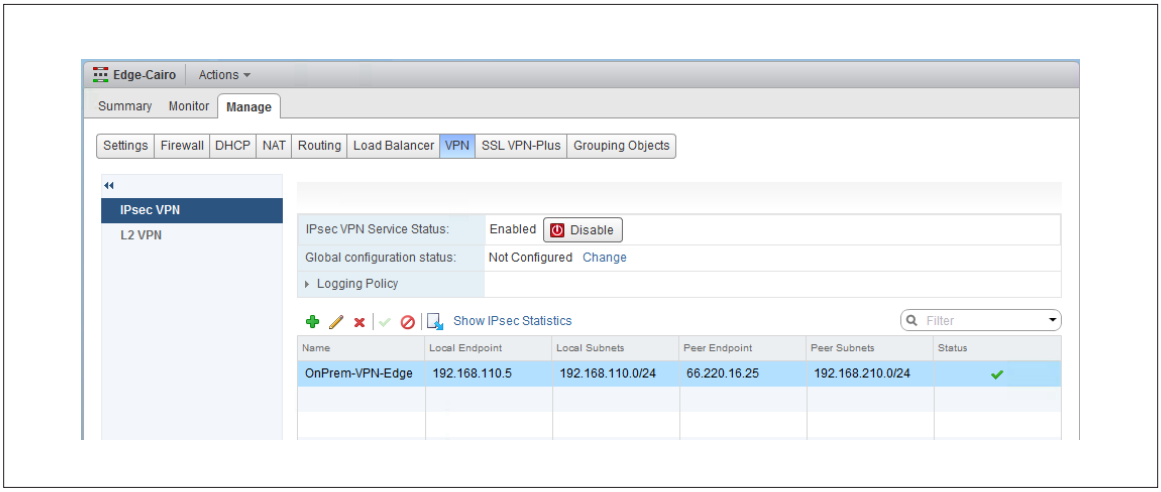

Figure 6.

Within vCloud Air, the edge gateway associated with the cloud environment must be configured with the off-premises VPN information. At the time of this writing, these configuration parameters are not exposed in the vCloud Air user interface (UI) itself, which means it must be managed from the vCloud Director UI. Nevertheless, it is quite straightforward and it is identical to what needs to be configured on-premises.

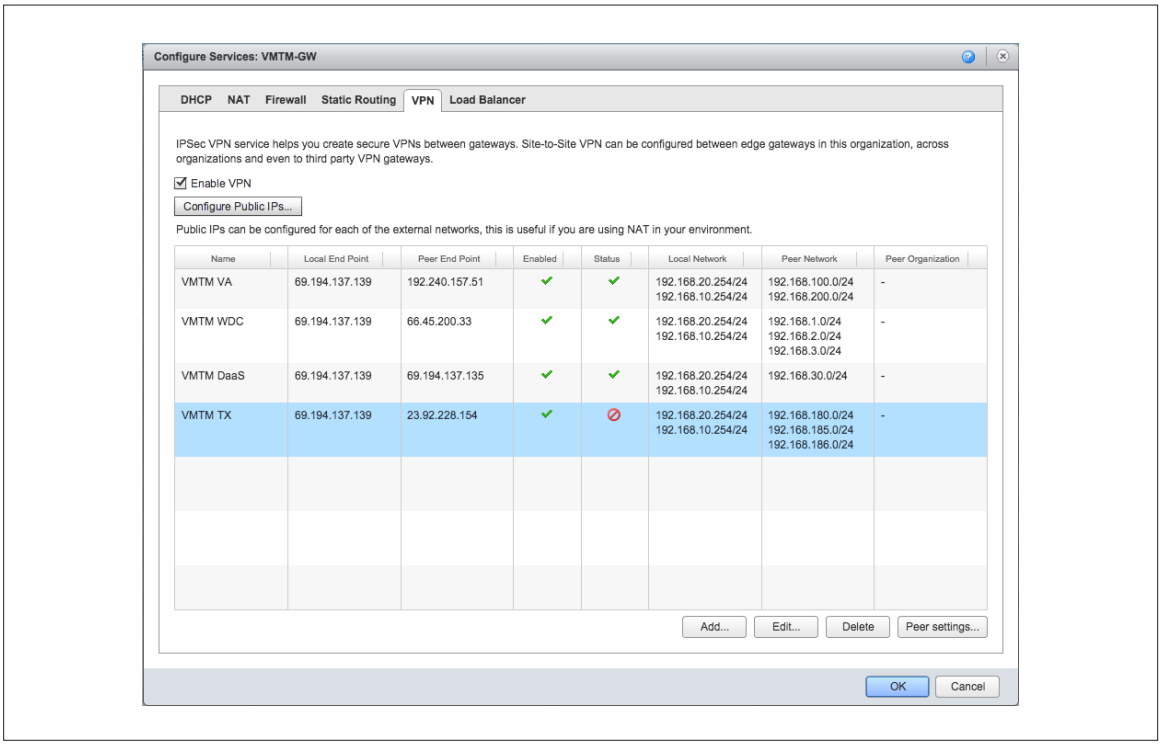

Figure 7.

The Compute Resources

Once the on- and off-premises clouds are connected and traffic is able to flow back and forth between these clouds, it is time to organize the compute resources within. The on-premises environment would typically be a vSphere cluster where the workloads reside, labeled here as "Production". This may also host the UAT workloads or another cluster can be allocated for that. The important part here is that the NSX Edge must have a network interface/route to the VM Networks on that cluster. This is depicted in the architecture diagram as 192.168.110.0/24.

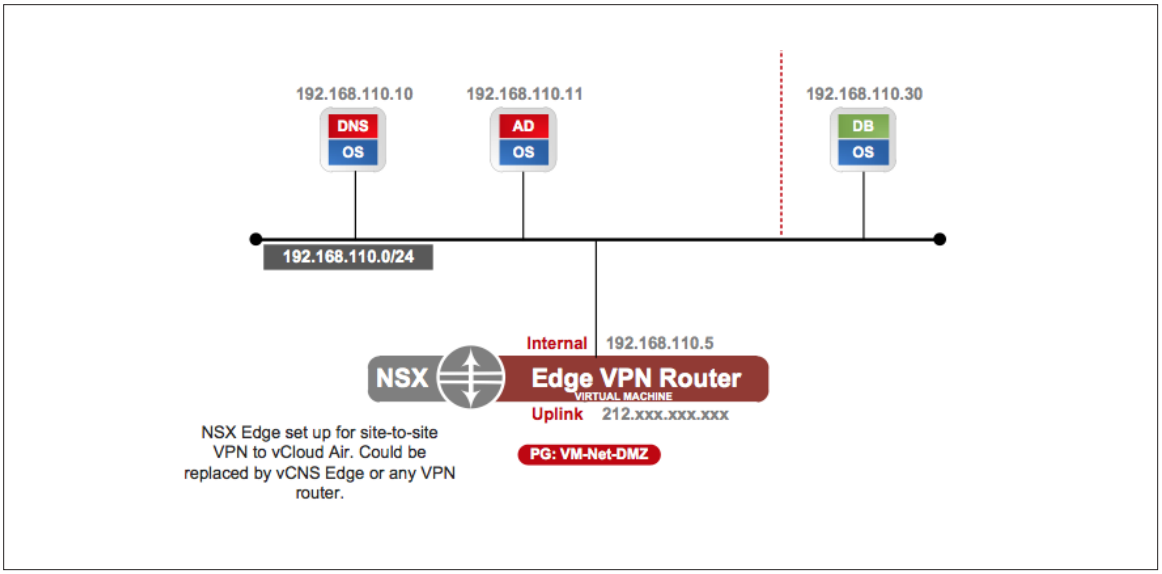

Figure 8.

The cloud environment in this architecture is a Virtual Private Cloud on vCloud Air with 10GHz of compute and 20GB Memory. This capacity can be scaled up per demand. Like with the on-premises vSphere cluster, this VDC must have a Routed Network that is acting as the internal interface for the edge device. This is illustrated in the architecture diagram as 172.16.1.0/24.

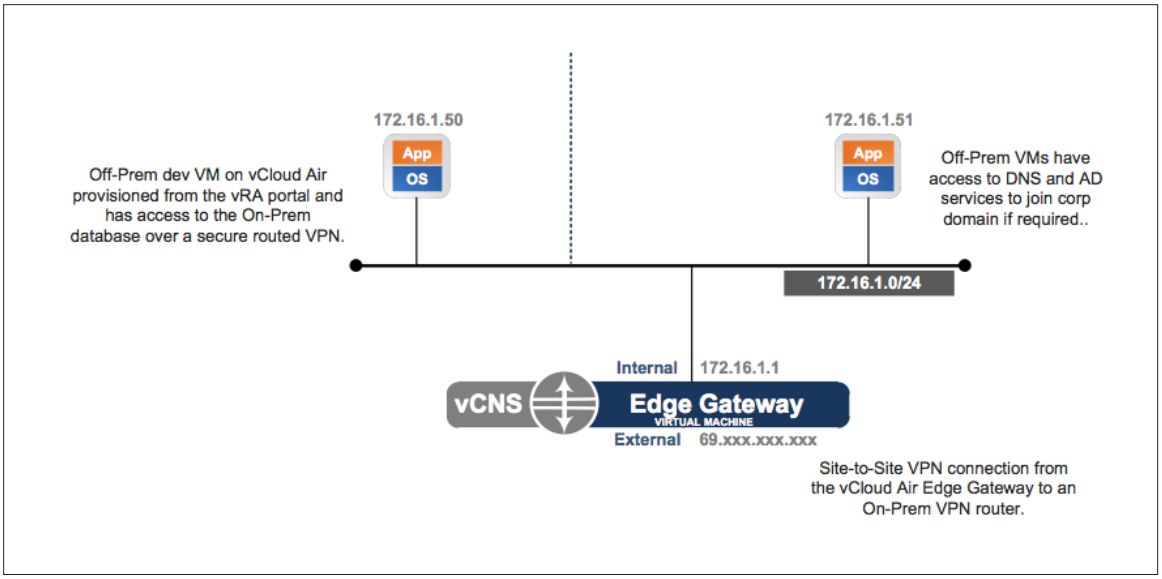

Figure 9.

Within the vCloud Air UI, this network is displayed in the following way:

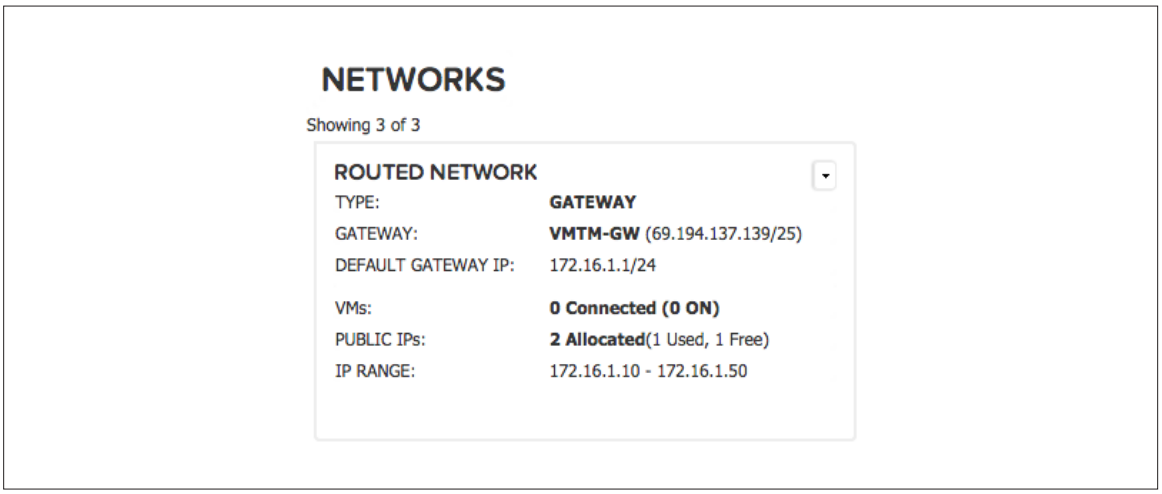

Figure 10.

Content Synchronization

As an integrated hybrid cloud solution, it is essential that the developers and application owners have access to a single, common set of core templates that are used in both on-premises and off-premises environments. The best way to do that is to maintain the core templates on-premises and have these templates synchronized (or pushed out) to the public cloud for consumption through vRealize Automation blueprints. The free vCloud Connector® tool can be leveraged for this purpose. vCloud Connector is more widely known as a tool to manage the movement of virtual machines across vSphere and vCloud Director-based environments. However, it also includes a catalog synchronization capability that is important in a hybrid cloud. vCloud Connector (vCC) is installed and managed from the on-premises environment and requires two components – vCC Server and vCC Node. The vCC Server is a one-time setup that typically sits in the management cluster of the data center. This acts as a management hub that coordinates the workload migrations and template synchronizations across your private and public clouds. The vCC Node is the data endpoint through which all vCC traffic passes. A vCC node must be deployed for every unique endpoint (e.g. vSphere cluster) or public cloud (e.g. vCloud Air). In the case of the latter, a vCC node is already provided to customers. Organizations just need to know the correct endpoint address. More detailed specifics can be found in the vCloud Connector documentation for vCloud Air (http://pubs.vmware.com/vca/index.jsp#com.vmware.vcc.vca.doc/GUID-2645E1B6-031C-4791-92EF-9F790DDCB059.html).

Once the vCC Server and vCC Node components are setup on-premises, "subscribing" to the template location will ensure that they are synchronized to your public cloud endpoint.

VMware vRealize Automation

Leveraging vRealize Automation as a single policy-driven portal delivers a true hybrid cloud. First, the two endpoints need to be added to vRealize Automation: the local VMware vCenter Server™ that is managing the production cluster and the vCloud Director® endpoint of the Virtual Private Cloud on vCloud Air. This white paper will not go through the exact configuration steps for adding these endpoints, but there are some excellent and detailed videos on YouTube by the VMware Technical Marketing team (http://www.youtube.com/watch?v=KrgQMIS5Jmk).

Some important guidelines specific to this architecture that are needed:

- **Reservation Policies:** Two Reservation Policies are needed – one for the local compute cluster running on vSphere, and the second for the remote compute resources running on vCloud Air.
- **Blueprints:** Multiple Blueprints are needed. At a minimum, one for the local/production VMs and one for the remote/development VMs. Each one should be pointing to the relevant Reservation Policy that was created.
- **Templates:** The Templates in the blueprints should be selected respectively from vSphere and vCloud Air. Both will be identical since they are synchronized by vCloud Connector as explained earlier.
- **Customization:** vCenter Customization Specification can be leveraged for the on-premises environment to customize templates while provisioning. This is not necessary for the vCloud Air side as the platform takes care of this.
- **Network Profiles:** Both the external and internal IP addressing through the Network Profile configurations will be maintained. You do not have to work around IP address conflicts even on the public cloud since both vRealize Automation and vCloud Air will keep track of the consumed IPs.

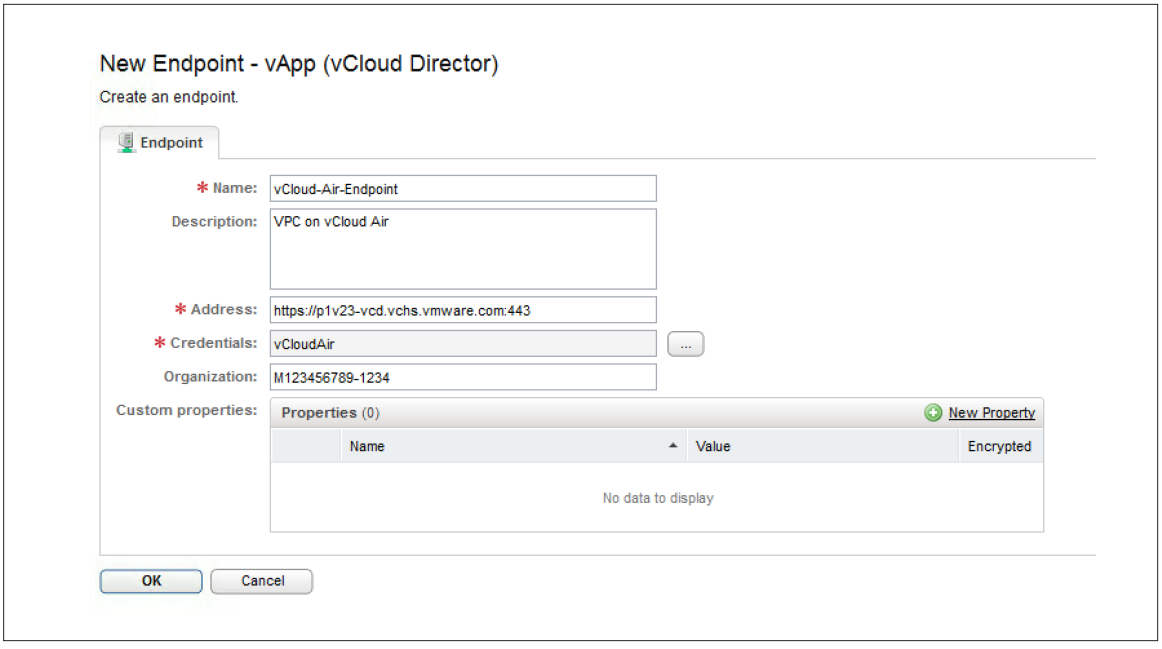

Figure 11.

Putting It All Together

The last piece in this architecture is to create the entitlements for each Blueprint/Service and their relevant approval processes. What is even more interesting is what can be done to delegate specific and precise capabilities to internal users or even external contractors or consultants working on different projects. Note that depending on the role of the consumer, the process can be different as depicted in the diagram below.

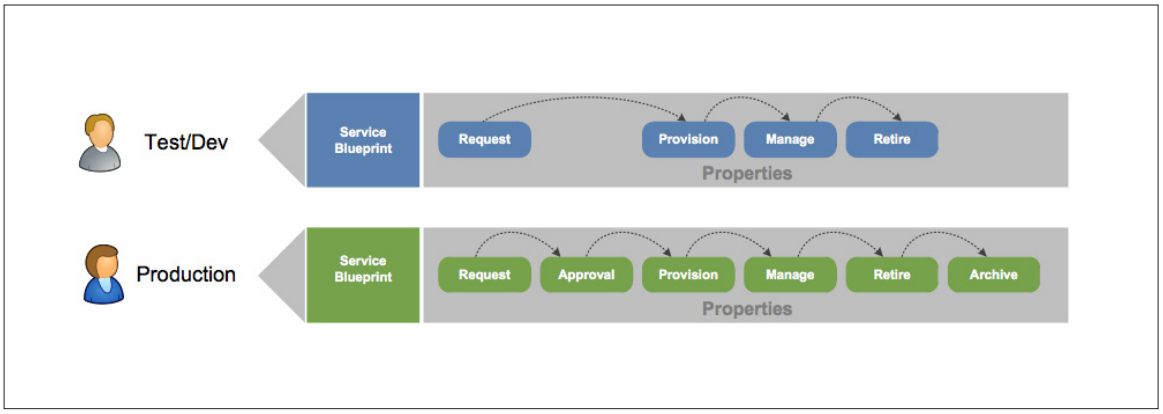

Figure 12.

One example scenario is when a new developer is hired and ready to start the process of developing an application from Test/Dev, to UAT and Production. An administrator can simply assign the developer an account on Active Directory, provide him access to the vRealize Automation portal, and give him the entitlement to provision workloads only to vCloud Air. This is defined by the administrator's approval process in vRealize Automation. Once the virtual machine is deployed, the developer can access the console of his VM(s) on vCloud Air through the VMRC without touching vCloud Director or knowing anything about it. He can also connect to the on-premises infrastructure services per the security policies that were defined within VMware NSX. It is as if everything is one large environment… it truly is hybrid.

During the software development cycle, the developer may want to create a snapshot before he or she upgrades a specific component in the environment to ensure the ability to restore to a previous build. Developers can perform this on their own since they have that permission defined in their vRealize Automation portal; no need to bother the operations team. But what if the developer wants to start fresh? This is still not an issue as the developer can "re-provision" the VM and have it prepared from scratch in a matter of minutes. In this case, the policy can be designed so that it does not require another approval since the developer already got it the first time. Automating these policies can save both developers and administrators a lot of time.

The same applies to the Application Owner. If there is a need to scale out an application then new instances can simply be requested by going to the vRealize Automation portal. New instances will be deployed as soon as they are approved, and can also be retired in a similar fashion when no longer required.

Conclusion

This white paper has shown how to implement a hybrid cloud strategy using vCloud Air, VMware vSphere, VMware NSX and VMware vRealize Automation. With this solution, existing, and/or new application deployments can be horizontally and vertically scaled across on-premises and off-premises environments.

This hybrid approach avoids the need to have idle excessive capacity on-premises as additional capacity can easily be procured when needed. New projects can be supported on-demand while providing a single place for developers and application owners to request resources. This approach lowers capital expenditure but also operational expenditure and provides business agility by allowing organizations to scale out applications at times of need and scale back when applicable.

About the Authors

Hany Michael is a Lead Architect in the SEMEA Professional Services Organization at VMware and a CTO Ambassador. Hany has an extensive experience in architecting and building private and public clouds across EMEA for large enterprises, service providers and Telco's. He has also published many KB Articles and reference architectures for different VMware solutions and technologies. In his spare time, he blogs about various technical subjects and he is well known in the industry for his architecture based diagrams and blueprints.

- Follow Hany's personal blog at http://www.hypervizor.com
- Follow Hany on Twitter: @hany_michael

David Hill is a Senior Technical Marketing Architect at VMware in the Cloud Services Business Unit. At VMware David has produced a number of vCloud based reference architectures, including the vCloud Architecture Toolkit and vCloud Monitoring Architecture. He is currently working on developing cloud solutions for customers moving to the cloud and has been the lead architect on current feature releases including Data Protection and Storage Tiers. He holds multiple VMware certifications and has been awarded vExpert status for the past four years running.

- Follow David's personal blogs at http://www.davidhill.co
- Follow David on Twitter: @davehill99

Duncan Epping is a Chief Technologist at VMware in the Office of CTO. In that role, he serves as a partner and trusted adviser to VMware's customers primarily in EMEA. Main responsibilities are ensuring VMware's future innovations align with essential customer needs and translating customer problems to opportunities. He specializes in Software Defined Storage, hyper-converged infrastructures and business continuity / disaster recovery solutions. Duncan has 3 patents pending on the topic of availability, storage and resource management. He is a VMware Certified Design Expert (VCDX007.)

- Follow Duncan's personal blogs at http://www.yellow-bricks.com
- Follow Duncan on Twitter: @DuncanYB

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

 Item No: VMW7356-TWP-IMPLEMNT-HYBRID-CLOUD-STRATEGY-USLET-103 03/15