Microsoft

Microsoft Cloud Platform System

powered by Dell

Contents

Introduction

Service providers and enterprises that wish to deploy workloads using a service provider-style solution are looking for cost-effective, flexible, and highly scalable designs. However, building and operating a cloud is a complex undertaking. Integrating the hardware, installing and configuring the software, and optimizing the overall infrastructure for performance, scale, and reliability all can make many cloud deployments fall short of their goals.

From start to finish, it can take several months to purchase and deploy the hardware and install and configure the software. This process typically requires a large, specialized team of IT professionals, ranging from networking experts to storage experts and, of course, virtualization and operating system experts. Once the infrastructure is configured, necessary templates for services and applications also need to be created, requiring additional time and investment.

Customer Expectations

With the growing adoption of public cloud offerings, and the always on and elastic characteristics they provide, customer expectations are continually evolving as to what they expect from a private cloud provider delivering Infrastructure or any service. The capability of public clouds to provide instantaneous service, reliability, predictability and availability are top of mind for end users.

Those expectations transfer to private cloud operators as well. End users expect to be able to connect to a self-service portal and do anything they want to their resources. Many cloud deployments fail while trying to meet these expectations. Ultimately, the question is "Why is it so difficult to deliver a cloud?" In order to answer this question, Microsoft is rethinking the way that private clouds are designed, built and deployed.

Microsoft Cloud Platform System – powered by Dell

The Microsoft Cloud Platform System (CPS) is designed specifically to reduce the complexity and risk of implementing a self-service cloud. CPS includes all the needed software and hardware, service providers and enterprises can give customers the self-service offerings they are demanding. As a result, these providers can respond quickly to business opportunities—without dealing with the complexities associated with deploying and operating a cloud.

CPS is integrated hardware, powered by Dell, that is an Azure-consistent cloud in a box which runs in your datacenter. That means that there is a consistent experience between what end users interact with in Azure as there is in CPS.

CPS is specifically built to maximize the economic benefits of the software-designed datacenter when operating cloud services. By creating a layer of software abstraction across physical storage, network and compute, the agility of the system passes the benefits directly to the operator by decoupling the physical fabric from the tenant services that run on top of it.

Finally, CPS is deployed with a Microsoft led support and orchestrated update process. With CPS, the operator does not need to call individual component vendors for support – a single call to Microsoft is all that's needed. Microsoft support personnel will lead the support process no matter what component in the system needs attention.

Components and Structure

CPS is composed of standard Microsoft software and Dell hardware. For software, Windows Server 2012 R2, System Center 2012 R2 and Windows Azure Pack (WAP) make up the core of the system. The design and implementation of the software is based on learnings derived while operating the Microsoft Azure public cloud. An orchestrated patching and updating system keeps the both the core Microsoft software as well as firmware up to date so that there is no impact on tenant services, and optimized run-books for multiple Microsoft applications are provided as guidance to enable customers to deploy these applications on CPS.

Just as all the elements of the Microsoft software are standard, every effort has been made to ensure that, from a Dell perspective, all of the hardware is "off-the-shelf" and that there are no special elements that require additional support.

Logical Architecture

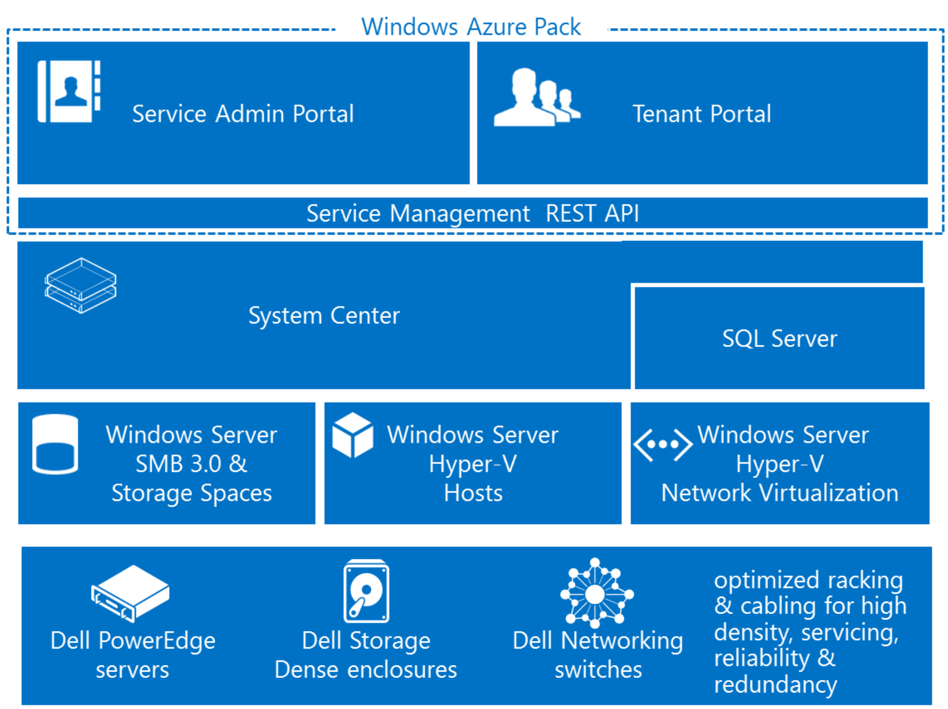

Figure: Logical view of CPS Components

When an end user connects to CPS they are given the option to use a URL that takes them to a **Tenant Portal** which offers the same user experience that they would see in Microsoft Azure. The end user is provided access to the similar user experience in order to manage their resources. Similar to Microsoft Azure, users also can leverage a REST API to access and manage resources.

Analogous to the tenant portal is the **Service Admin** portal. In order to provide a truly self service experience for tenants from signup through resource management, services and capacity need to be bundled and managed in a way that is consistent with public cloud services. Through the Service Administrator portal, providers are able to create offerings to which tenants can subscribe. From these offerings and subcriptions, providers can either charge-back or show-back usage reports for their customers.

Both these portals are built on a REST based API known as the **Service Management API**. This API enables providers who do not wish to use the provided portals to customize and create their own branded user experience while maintaining the rich features of CPS. Additionally, if advanced users want to develop programmatically against the resources they can use this API to integrate their applications and services with CPS resources. Because this is a REST based API, end users can use a variety of tools to interact with the system, from Windows and Linux through Java, Python, or many other framworks and languages.

CPS requires management capabilities such as deployment, configuration, monitoring and automation All of these management functions are deployed in Hyper-V based highly available virtual machines and the core management workloads are kept on a dedicated management cluster. **System Center** is the familiar toolset used to deliver those services. The over all system configuration data is kept in **SQL Server** 2012 R2.

From a tenant networking perspective, all virtual machines that are provisioned through infrastructure-as-a-service (IaaS) are created on isolated virtual networks using **Windows Server Hyper-V Network Virtualization**. This allows tenants to create networks created on demand without providers having to reprogram the physical networks. This isolation provides data channel isolation at the network layer and creates flexibility for tenants and administrators who don't have to worry about such things as overlapping IP addresses, conflicting machines, or configuration errors when working with physical networking.

CPS implements the best practices of Microsoft software-defined storage virtualization technology - that is delivered through **Windows Scale-out File Server, Windows Server SMB 3.0 & Storage Spaces**. All of the **Windows Server Hyper-V** hosts place their VMs on an SMB-backed share which is built on top of storage pools that are created from the underlying physical storage.

Dell PowerEdge Servers provide the physical compute. **Dell PowerVault Dense Enclosures** house the just-a-bunch-of-disks (JBOD) configuration for system storage. And **Dell Force10 Networking switches** compose the physical networking layer.

CPS Design Principles

When rethinking the way that private clouds are designed, built and deployed, many techniques that are essential to the efficient operation of Azure became principles of CPS design:

Comprehensive System Architecture: The overall design of the system must be specific from the PDU's, through the cabling and all the way through to the end user portals. All of the physical and software systems in CPS are designed to be resilient and redundant to failure. By designing the system as a whole, the right tradeoffs between multiple components for redundancy and price have been made.

Predictable Supply Chain: The larger the deployment, the higher the chances of component failure. The system needs to be resilient to such failures, preventing interruption to tenant services, while at the same time being optimized to replace hardware in a predictable and efficient manner. By working with standard hardware from Dell, a predicatable cadence of maintenance can be established - allowing technicians to respond to failures in an orderly and predictable way.

Validated Deployment: Customization is one of the driving factors of choosing to deploy a private cloud. However, that customization can be the very thing that causes problems. CPS provides a standardized and validated hardware and fabric deployment that can be easily maintained in an automated fashion. By decoupling the fabric from the services, the customization can take place at the service level where impact to the overall system is minimal.

Orchestrated Updates: Just as it is assumed that all hardware will fail, it's also assumed that software will need to be updated. CPS has an automated and orchestrated update system that is analagous to the predictable supply chain for hardware. When the core CPS components need software or firmware updates, the orchestration process performs the operation with impact to tenant resources in mind.

Validated Workloads: End users expect that their cloud experience have robust and easy to use patterns that they are able to deploy in a self servicve fashion. Microsoft has validated workloads, such as Microsoft Sharepoint and Microsoft SQL which are provided as guidance to enable customers to deploy and operate these workloads on CPS in a repeatable way to reduce errors.

Unified Support: Finally, a single defined point of contact for support is vital to operating a cloud. The complexity of cloud computing infrastructure would be overwhelming if each vendor needed to be contacted for every problem and the operator be burdened with finding the responsible party. Microsoft leads that support experience and facilitates the contact with each CPS infrastructure component vendor.

Resiliency and Redundancy

One of the fundamental elements of cloud computing is the notion that all systems need to be designed for resiliency and redundancy. One way to accomplish this is to ensure that for every component in the system is repeated many times. However, this approach introduces additional cost into the system. With CPS, considerable investment was made to determine the optimal balance between what is resilient and redundant in the system and what the overall cost of the system should be.

Cloud Operations Simulation

In order to prove that CPS is more than just a design, Microsoft engineers calculate the amount of failures expected for a physical system over the course of one year, and then actually perform those failures while tenant activity is continuously taking place over the course of one week. Some examples of the operations and failures are listed in the table below.

Example Operational Activites	**Example Forced Failures**
Live Migration of Virtual Machines	Node Failures
Live Storage Migration	JBOD Failures
VM Creation	Disk Drive Failures per day
VM Deletion	NIC/Cable Failures and Switch Failures
IO Bursts for VMs	SAS Cable failures and HBA Failures
Many others…	

Table: Example Operator Activities and Forced failures performed against the system by Microsoft.

By validating CPS in such a manner, the customer is assured of the ability of system to remain operational in the event of inevitable failures in various parts of the system.

Tenant Cloud Services

When choosing cloud providers, end users expect to have choices, not only in pricing and service levels, but also in the service models by which they consume and manage resources. CPS delivers multiple service models through a single portal that is consistant with Microsoft Azure. With CPS service providers can create plans and offerings that combine Infrastructure-as-a-Service (IaaS), Platform-as-a-Service (PaaS) and Database-as-a-Service (DBaaS).

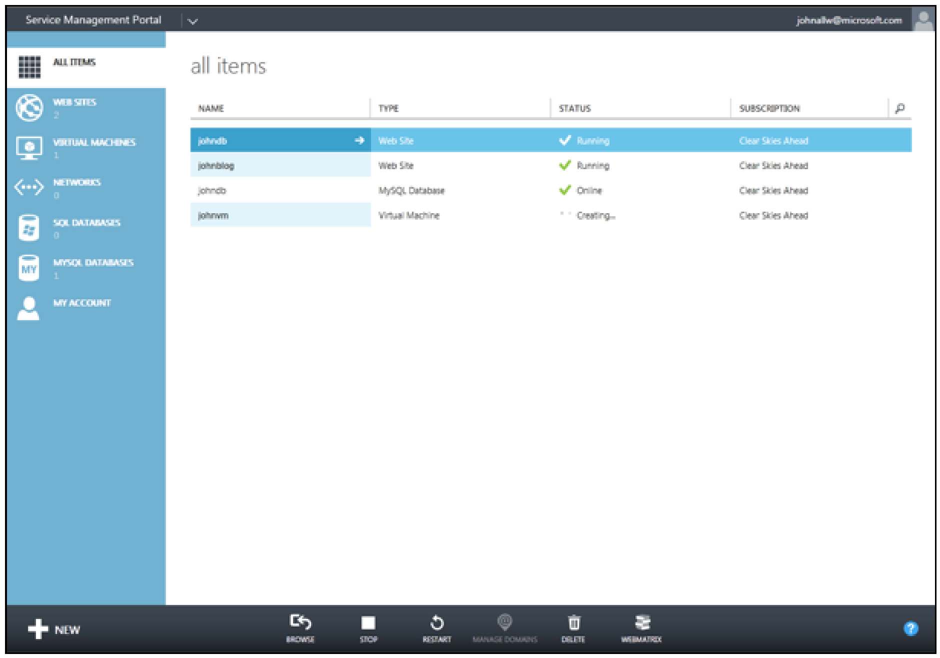

Figure: Management portal for tenants (note the strong consistency with the Microsoft Azure portal)

Virtual Machines (IaaS)

CPS provides IaaS services for provisioning virtual machines running either Windows or Linux. Service administrators can offer various pre-configured images through a gallery. This provides a very simple and standardized way to not only provide virtual machines with a base OS image, but also pre-configured application stacks within those images.

Virtual Networks

Virtual networks enable the self-service creation of isolated, tenant-specific IP address schemes independent of the actual IP infrastructure on which they are layered on. This means that tenant VMs can be migrated between hosts or racks without the need to reconfigure IP address schemes or the risk of colliding with other tenants' IP address spaces. The tenant virtual networks are isolated from the underlying infrastructure as well as from other tenant virtual networks to provide a multi-tenant capable architecture.

Web Sites (PaaS)

CPS can be extended to deliver web sites as a service. The Web Sites Service enables high-density, multi-tenant, and highly secure web application offerings for service providers and enterprises. Delivering this service to customers and users allows for deployment of modern web applications. Many programming languages and template web applications are supported, as well as integration with developer tools and popular source control repositories. Web sites is an optional component that can be added to CPS.

Databases (DBaaS)

CPS supports Microsoft SQL Server and MySQL in a Database-as-a-service (DBaaS) model for tenants. This capability is available to web applications in the Web Sites Service or in multi-tiered applications. DBaaS is an optional capability that can be added to CPS.

Workloads

A major part of the appeal of the Cloud Services model is the variety of pre-built templates that can accelerate productivity for the end user. With Cloud Platform System, service providers and enterprises can build a robust portfolio of potential workloads they can make available to customers and end users.

Pre-defined runbooks for Microsoft workloads

To help providers quickly start delivering highly scalable workloads, Microsoft has authored workload patterns for CPS such as Microsoft SharePoint and Microsoft SQL Server. These patterns (known as *VMRoles*) are co-designed with each of the Microsoft product teams and serve as samples that providers can customize for their tenants' specific needs.

Additionally, non-Microsoft workloads will be available as well, such as CentOS and Oracle DB.

Custom Workloads

Service providers and enterprises can create their own custom workload templates to meet unique or specific customer requirements—adding more value to their services. Specifically, providers can configure applications, operating systems, deployment rules, and the UI Wizard that the tenant uses in the self-service portal.

Enterprises can make business applications available to users in various departments to speed up deployment of those critical in-house applications.

System Architecture

CPS is a combination of forward-looking design, testing and automation that ensures that the system maintains the highest service availability levels with as little human intervention as possible. Understanding the system architecture helps explain how IT Staff can get out of the day-to-day operations of the system and focus on innovation on top of the actual infrastructure.

Understanding Stamps

CPS ships in units referred to as a Stamp which entails the management domain of a CPS instance. A single stamp ranges from a minimum of one rack to a maximum of four racks. As the cloud capacity needs grow, a stamp can be scaled out, effectively expanding the aggregate pool of compute, storage and network resources.

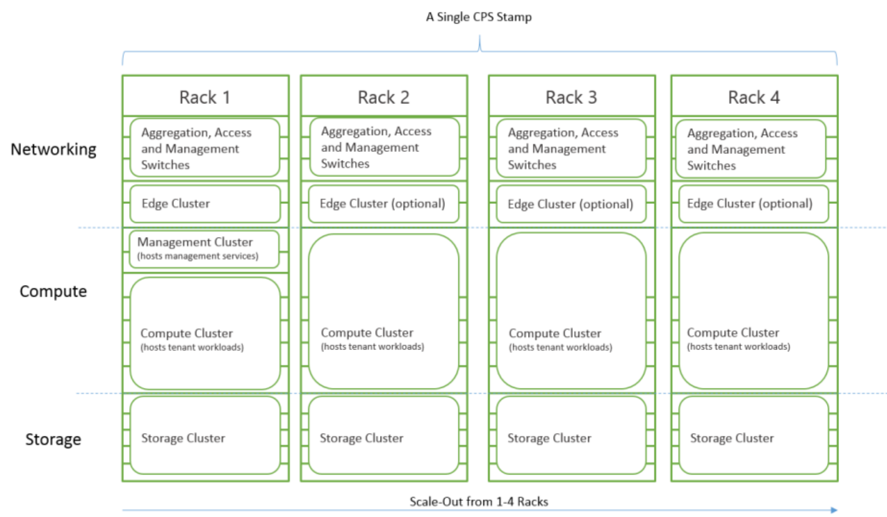

Figure: Diagram of a single CPS Stamp scaling from one to four racks.

All of this arrives with CPS predeployed and preconfigured with a capacity of up to 8000 VMs** in a full stamp.

In cloud computing, it is essential to transition thinking away from servers as the unit of measure to pooled units of resources. In this sense, the following table illustrates the capacity available as CPS stamp is scaled out.

	Rack 1	Per Additional Rack	Max per Stamp
Compute Nodes available for workloads	24*	32	120
Physical CPU Cores available for workloads	384	512	1920
Memory for workloads	6 TB	8 TB	30 TB
Storage Used for Backup	126 TB	126 TB	504 TB
Storage Available for Workloads	136 TB	156 TB	604 TB

Table: Detailed Pooled Capacity for a single stamp – at one rack, through a full rack. *The first rack has 32 compute nodes, but 8 are dedicated to the management and edge clusters.

** This number is based on VMs which are configured as follows: 2 vCPU's, 1.75 GB RAM, 50 GB Disk

Software-Defined Datacenter Technologies

The CPS software-defined infrastructure provides flexibility to customers. The key goal is to deliver better economics through the use of industry-standard hardware at scale, and then using software to make that hardware infrastructure reliable, available, and better serviceable.

Storage

Storage in CPS starts with a **JBOD Chassis** (as illustrated below). Each chassis contains 60 disks which are a mix of 48 HDDs and 12 SSDs. This combination of disks is used to create tiers of storage – keeping the most commonly accessed data on the most performant tier. **Storage Spaces**, a feature of Windows Server 2012 R2, aggregates all of the physical storage into pools, and does the work of moving the data between tiers. CPS is deployed with three pools, of which two of the pools are used for Tenant Workloads and the third is used for Backup.

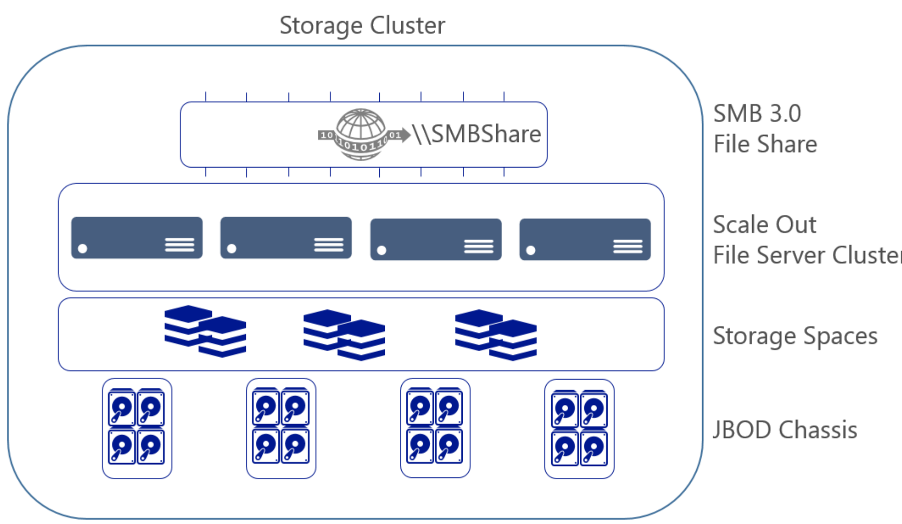

Figure: Logical structure of the Storage Cluster

The pooled storage is exposed to the fabric via a four node **Scale Out File Server cluster**. Ultimately, each rack has 282TB of usable storage. The two tenant workload pools are triple mirrored across all of the disks to provide redundancy and performant write semantics while the backup pools are configured with dual parity and de-duplication for optimal capacity.

Compute

Hyper-V provides the virtualization platform for CPS, the same technology that powers Microsoft Azure, running hundreds of thousands of VM's at a global scale. All of the Hyper-V nodes connect to the shared storage through an **SMB 3.0 File Share** to place the storage for each virtual machine. CPS enables virtual machine management using workload templates, automated deployment and provisioning, virtualized networking that moves with the VM, and the live migration features of Hyper-V.

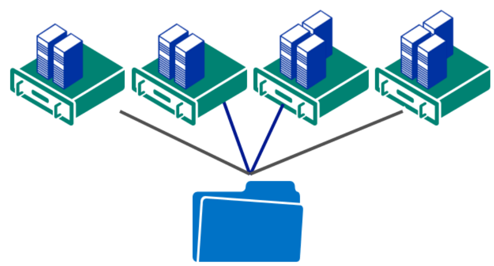

Figure: Compute nodes connecting to the SMB 3.0 Files Share on the storage cluster.

Each compute node is attached to four 10GB Ethernet connections. Two Ethernet connections with Remote Direct Memory Access (RDMA) enabled are used to access storage while the other two (with Network Virtualization using Generic Routing Encapsulation (NVGRE) offload) provide redundant (Load Balancing with Fail Over (LBFO) connecticvity for tenant network traffic.

Networking

Based on learnings from building and operating Microsoft Azure, CPS is designed with a flat Layer 3 physical network with network virtualization enabled for tenant networks. Each compute node has two network ports for tenant traffic. Each card is connected to each of the access switches in the rack – providing redundancy in the network path. The network interfaces themselves are then configured for teaming, providing both load balancing and failover for tenant network traffic. If both physical interfaces are operational, they are used to provide 20 Gbps from the physical host, but if one interface experiences downtime, the system automatically fails over to the remaining interface.

Each Node is connected to the storage that houses the Virtual Hard Disks (VHD) for the VMs. Similar to network connectivity, two cards are used, each providing 10 Gbps to the storage. Over this connection, SMB over RDMA is utilized to deliver high throughput, high resiliency, and low latency traffic from the storage subsystem to the Hyper-V node.

The physical switches in CPS are cabled and configured in redundancy mode to enable operational availability in case of failures.

Management and Operations

Management Cluster

The Management Cluster is effectively at the center of the system - providing all the services needed to operate the cloud. Physically, the management cluster consists of six servers, configured as a Hyper-V failover cluster, separated from the other compute nodes. The storage for the cluster comes from the shared storage from the Scale Out File Servers (as discussed in the storage section on page 13). All of the core management functions are deployed as virtual machines on the cluster. CPS arrives with all of the infrastructure and management functions, such as directory services, portals, monitoring, backup etc., pre-deployed and pre-configured in a highly available or resilient manner. Typical management operations such as resetting passwords are automated to minimize manual involvement in the operations of the system.

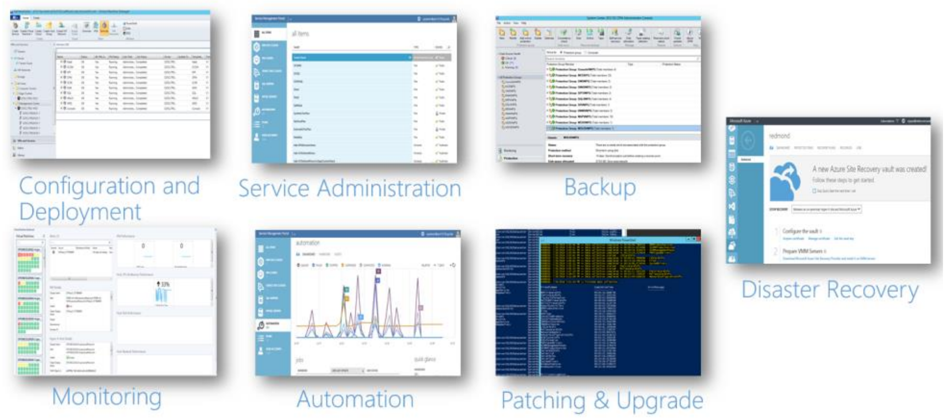

Figure: Core management functions delivered through CPS.

Configuration and Deployment

System Center Virtual Machine Manager (VMM) is used to deploy and configure additional racks when the stamp is scaled. VMM Service templates are also the basis of all other management services. When CPS is pre-configured in the factory, VMM is the first management service that is deployed and all other management systems are deployed from it. Through this standardized approach, patching and updates to the management systems can be automated and integrated into the Orchestrated Update and Patching subsystem.

Service Administration

For service administration, the Windows Azure Pack Service Administration portal is available to operators to be able to create and configure plans which then can be made available to tenants. It is also the main portal for interacting with Service Management Automation (SMA), which serves as the automation engine that drives CPS. SMA is the on-premises instantiation of Azure automation.

Backup

Backup services for both tenant workloads and management cluster are provided through System Center Data Protection Manager (DPM) and are included with CPS. For tenant workloads, automation enables the protection for newly created virtual machines during a protection window in which they get added to the backup schedule. An integrated alerting mechanism is in place to ensure that all tenant workloads conform to backup SLAs.

For the Management services, CPS uses a pre-deployed backup service using DPM of the management cluster - helping to ensure continued uptime of the system and services. All of the components in the management cluster are automatically backed up, and automated consistency checks are performed post-recovery should it be required.

Monitoring

Using System Center Operations Manager (OM) capabilities, administrators can monitor the fabric compute, storage, and network components using centralized dashboards purpose-built for CPS. From the dashboards, they can see component health at a glance and drill into granular health, performance, and capacity. Since the exact hardware and software configurations are known, and have been operated first hand by Microsoft, CPS designed management packs are optimized to eliminate alerting noise and help administrators focus on what is important.

Automation

Service Management Automation (SMA) drives much of the automated processes for CPS and has been discussed in the context of the management systems themselves. One CPS automation that is not explicitly part of any management system, but is universal to all is the Password Management automation. Ensuring consistent passwords according to password rotation policies can be challenging for many customers. The Password Management functionality enables administrators to change passwords with no impact to tenant services. This system is integrated with the monitoring system for notifications of password rotation times. As customers operate CPS, they can add to the library of automated tasks relevant to their environments and processes.

Orchestrated Patching and Updates

One of the key learnings from operating global clouds at scale is that manual tasks are one of the main sources of errors and service interruption. Extreme levels of automation are required to successfully deliver cloud services at scale. Automating such tasks frees up vital IT resources to focus on business critical issues. One of the most important points where outages can occur is patching the system. CPS ships with all of the administrative tasks automated in a way that is consistent with the overall system design.

Figure: The orchestrated patching and update process is designed to not disrupt tenant workloads.

The overall flow of the patching process reflects the level of validation done to ensure customer uptime. First, the relevant patches, from firmware through portals are deployed on internal Microsoft development CPS stamps. Once the automated upgrades are successfully tested, they are rolled out to an internal Microsoft production environment for further validation. Only then are they made available to customers for deployment in their own environments. When deployed, the automation finger-prints the current state of the system to determine dependencies and then orchestrates the patching across CPS. While updating, tenant workloads are automatically managed (such as being live migrated from the Hyper-V host being patched) in conjunction with the process to optimize for service availability and impact to tenant services. This process is performed against the *entire* infrastructure.

Disaster Recovery with Azure Site Recovery

For many organizations, disaster recovery is a complicated process, difficult to test and troublesome to manage. If a service provider or enterprise owns multiple datacenters, they can use Azure Site Recovery

(ASR) to create a disaster recovery plan between two CPS stamps. This service is provided with CPS at no additional cost. With ASR, disaster recovery is simplified and incorporated into the overall design of the system that keeps both tenant and management systems highly available. Only management metadata is used in the Azure service itself to structure the recovery. Tenant and management data is transferred directly from the main to the DR site. ASR plans can orchestrate the recovery of resources at a designated site. ASR further simplifies the disaster recovery process by enabling testing of failovers and restorations of systems.

Conclusion

The world of IT is undergoing constant change. With the ever-growing complexity and massive speed of innovation, service providers and enterprises must keep up in order to stay competitive. The Microsoft Cloud Platform System offers a complete end-to-end, Azure-consistent on-premises cloud solution in a pre-integrated, fully validated, and supported package. This enables IT staff to deliver value rapidly and focus on innovation of services to benefit business outcomes. Through converged systems, the software-defined infrastructure, and unified (and thus simplified) IT operations, customers can achieve greater flexibility while reducing cost. Developing new business offerings can now happen at cloud speed in the datacenter.