vmware PRESS

VMware vCloud® Architecture Toolkit (vCAT)

Technical and Operational Guidance for Cloud Success

VMware vCAT Team

Foreword by **Pat Gelsinger**, CEO, VMware

VMware vCloud® Architecture Toolkit (vCAT)

Technical and Operational Guidance for Cloud Success

predictability that enable stakeholders to improve value and align spending with business goals. It also automates the core financial processes needed to easily plan, charge, and optimize the cost and value of IT. The ITBM suite includes these components:

- **IT Costing:** Maps the connections between IT services and their underlying cost drivers using an intuitive graphical approach that enables total cost of ownership (TCO) and unit cost tracking
- **IT Demand Management and Budget Planning:** Facilitates accurate, fact-based IT budgeting, planning, and forecasting
- **IT Showback and Chargeback:** Gives business units visibility into IT costs and alternatives, including fully itemized billing and chargeback
- **IT Cost Optimization:** Automatically identifies potential areas for ongoing cost reduction, such as candidates for virtualization and consolidation, storage tiering, SLA reduction, end of life, deferral of upgrades, and support reduction
- **Vendor Manager:** Provides a control and optimization mechanism for vendor agreements that proactively governs contractual commitments
- **SLA Manager:** Sets, tracks, and reports on SLAs, key performance indicators (KPIs), and key value indicators (KVIs) for services, vendors, and customers, and performs root cause and business impact analysis at all levels

4.6.1.2.1 Relationship Between Chargeback and ITBM Suite

vCenter Chargeback collects virtualization and vCloud cost data by integrating with vSphere and vCloud Director. It then provides cost data to the ITBM suite for inclusion in cost models.

Both products are connected by the vCenter Chargeback Connector, which scans vCenter Chargeback for a specific hierarchy and creates a report schedule to generate cost reports for this hierarchy on a daily basis. The connector also retrieves both generated and archived reports and provides the cost data for each virtual machine in the hierarchy to the IT Business Management Suite.

Based on the cost data collected by the connector, the IT Business Management Suite populates detailed analysis reports in its cost model and CIO dashboard. This integration provides visibility to CIOs across all IT assets and enables them to easily identify cost reduction opportunities by comparing virtualization, vCloud, and physical costs.

4.6.1.2.2 Cost Models

The ITBM Suite provides out-of-the-box (OOTB) cost models. A *cost model* is a multitiered set of allocation rules that map the financial relationships from the general ledger up to the business units within the organization. The relationships reveal which entities drive the cost of other entities.

The *cost browser* provides a simple way for users (typically the IT finance administrator) to create and modify a cost model that defines the cost relationships in their business structure. Cost models can be modified periodically by adding or deleting elements and changing dependencies to reflect the current contributory relationships between cost object types.

The OOTB cost model does not necessarily reflect all financial aspects of a fully mature IT organization. Instead, it provides immediate value to typical IT organizations and introduces design guidance for object types and common allocation rules that allocate cost end to end from the general ledger to the business units. If needed, the model can be enhanced to reflect any organization cost structure and data sources.

4.6.1.2.3 Integration with vCloud

Using the ITBM Suite, the customer gains unprecedented visibility and transparency across all IT components (physical, virtual, and vCloud). The ITBM Suite enables automatic tracking and processing of IT cost and service data across the organization. ITBM dashboards provide a 360-degree view of what IT services cost to deliver and the service levels that are provided. This visibility enables IT to run like a business and helps IT executives make fact-based decisions.

4.7 vCloud Service Control

vCloud Service Control deals with service governance and lifecycle management, and the design and development of vCloud-based IT services.

4.7.1 vCloud Service Governance and Lifecycle Management

The purpose of vCloud Service Governance and Lifecycle Management is to implement a standard methodology and control over the proposal, acceptance or rejection decision, definition, and end-to-end disposition of services and service offerings. It also provides governance and control over the quality of available services. Elements include Service Portfolio Management, Catalog Management, and Service Level Management.

4.7.1.1 Service Portfolio and Catalog Management

The purpose of the *service portfolio* is to accept or reject service proposals and maintain the overall catalog of services, whether rejected, under development, deployed, or retired. A primary responsibility of Service Portfolio Management is to verify that the services accepted for development and deployment align with the strategic and business requirements of the organization and its customers. This includes continuous review to allow adjustment of services due to new requirements and retirement of existing services due to lack of demand or replacement with a newer service.

The purpose of a *service catalog* is to maintain the active set of services. Active services are those under development or currently offered to customers for use in the vCloud environment. In this context, the service catalog is part of the overall service portfolio (as opposed to the consumer service catalog, from which customers deploy service offerings). The tool that supports the service portfolio and contains the service catalog provides a mechanism for automatically populating the consumer self-service catalog from the service offerings defined in the service catalog as part of the service offering release process. Regular reviews of the service catalog should be performed and adjustments made in line with feature changes in future releases of vCloud Director, vSphere, or other supporting products.

4.7.1.1.1 vCloud Service Catalog Components

The service catalog for the vCloud that vCloud Director supports offers service components to the end customer. At a minimum, the service catalog must define the following:

- **Organization container:** The *container* for the customer's IaaS, with attributes that hold basic, default service configuration information. Typically, only one organization container is purchased per customer.
- **Organization virtual datacenters:** The boundaries for running the virtual machines within the IaaS service, configured with sizing information based on the customers' requirements, with an appropriate SLA assigned to them. A minimum of one organization virtual datacenter is required for a customer to offer a service. Additional organization virtual datacenters can be requested, if required.

In addition to these core vCloud components, an organization can establish a standard set of offerings within the vCloud service catalog to provide customers with vApps (standardized groupings of preconfigured virtual machines) and media (installable software packages).

After being accepted into the service portfolio, the service and constituent service offerings should be defined with at least the following components:

- Service description
- Service requirements
- Service-level agreements
- Support terms and conditions
- Service lifecycle considerations
- Projected demand information for capacity planning
- Pricing and chargeback requirements
- Compliance requirements (regulatory and otherwise)
- Security requirements
- Monitoring and other operational requirements

Including pricing and chargeback, compliance, security, and operational requirements in the service definition is critical because these are core considerations during the service design and development process.

4.7.1.1.2 Service Types

Service types include business user services and technology services.

- **Business user services:** Defined as Software as a Service (SaaS) offerings, these services are generally directly consumed by users and are available as part of the organization's enterprise service catalog.

- **Technology services:** Defined as Infrastructure as a Service (IaaS) or Platform as a Service (PaaS), these technology services are not consumed directly by users, but they enable infrastructure automation that enhances an IT organization's capability to provide business user services.

4.7.1.1.3 Service Interrelationships

For optimal vCloud business user services, all types of technology services must be seamlessly integrated, usually with a workflow engine named the *orchestration layer*. Invoking a business user service can automatically trigger one or more technology services. The rules governing these workflows need to be preconfigured and preapproved for control. They are also needed to provide an agreed-to level of service to the business user. This agreed-to level of service is known as a *service-level agreement* (SLA).

4.7.1.1.4 vCloud Service Catalog Evolution

To improve the vCloud service catalog process and help realize vCloud benefits, as many service offerings as possible should be made available to users through automated provisioning.

In the virtualization world, the initial process for procurement of virtual machines generally follows the model that is applied to physical infrastructure. Although it is effective, it is not the most efficient mechanism for providing services, and vCloud benefits cannot be fully realized unless the process is changed. Figure 4.11 gives a logical representation of the evolution of the vCloud service catalog from this current state to the desired end state.

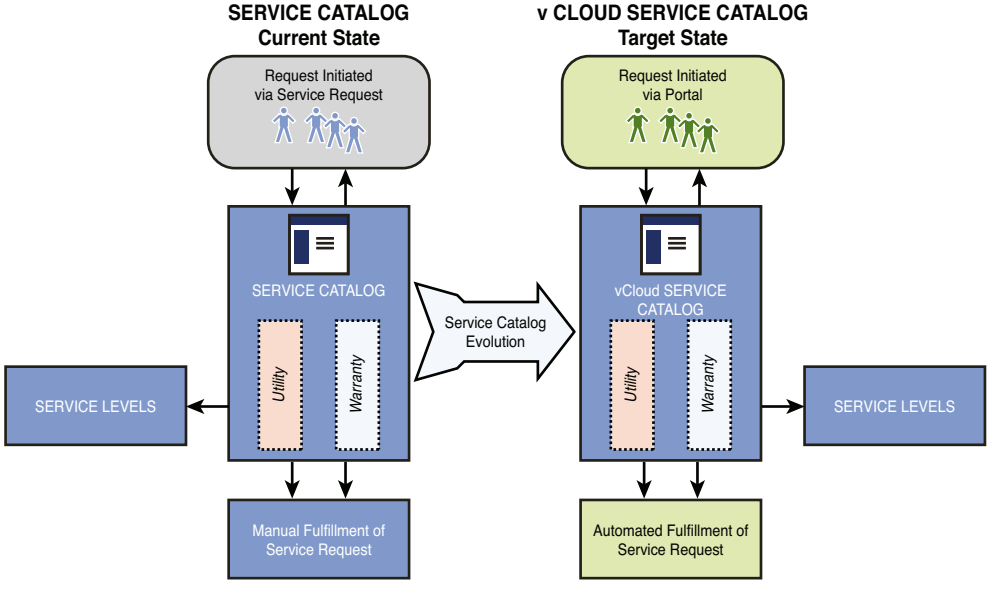

FIGURE 4.11 Service catalog evolution

In the service catalog current state, when a new service is requested, a service request is submitted to select and provision an offering from the service catalog. In addition to the utility (vApp or organization virtual datacenter) to be provided to the customer, the request includes the required service level provided by the virtual datacenter in which the vApp is to be provisioned, as well as any built-in availability features within the vApp itself. After the service is ordered, the end customer must wait for staff to fulfill the service request for the virtual machines to provide the service to be provisioned.

To satisfy the self-service, on-demand attribute of vCloud computing, the customer should be able to connect to a portal, select the required service offering, and have it automatically provisioned. This removes the need for manual selection from the service catalog and also removes delay in the provisioning processes. Figure 4.11 shows this process as the vCloud service catalog target state.

vCloud Director provides the capability to manage these requests from the service catalog. For vApps, an organization administrator can determine who within the organization has rights to request and provision vApps and thus provide end-to-end self-service. With vCloud Director, the user can select and provision the vApp and also specify the organization's virtual datacenter in which it is to be deployed. Because organization virtual datacenters are associated with provider virtual datacenters, the user is essentially selecting the required level of service.

To transition to the target state vCloud service catalog:

1. Continue with the service request process until the vCloud service catalog is available on the portal.
2. Enable IT staff to perform vCloud service catalog requests with automated provisioning on behalf of the user, including required approvals.
3. Add the capability for users to access the vCloud service catalog and request services that result in automated provisioning of the corresponding vApps, including required approvals.

4.7.1.1.5 Standardization of vCloud Offerings into the Service Catalog

Standardization of service offerings is essential to achieving a scalable, cost-efficient vCloud environment. Typically, compute resource-based service offerings (CPU, memory, and storage) provide a baseline for vCloud consumption and should be standardized as much as possible, regardless of whether they apply to organization virtual datacenters or vApps (and their associated virtual machines).

Compute resources for organization virtual datacenters available in the service catalog should be standardized into various sizes. The required compute resource configurations vary depending on the selected vCloud Director allocation model (allocation pool, pay as you go, or reservation pool) because attributes such as CPU speed and CPU or memory guarantee vary. Combining these two components means that the service catalog can offer differently sized organization virtual datacenters for each type of allocation model.

Similarly, to create a vApp catalog item (public or organization), standardization should be used as possible. From a compute resource point of view, standard-sized virtual machines should be created to use in a *pick list* of machines for vApp creation. These standardized virtual machines can vary in resource size for CPU, memory, and storage (for example, Standard, Standard Plus, Advanced, Premium, and Premium Plus). Because a vApp consists of one or more individual virtual machines, the appropriately sized virtual machines can be selected from the pick list during the vApp catalog creation process.

In addition to the basic compute offerings of the virtual machines within the vApps, it is necessary to develop the service catalog to include vApp software configurations. These can be basic groupings of compute resources and can be expanded over time to offer more advanced services. Table 4.2 shows sample vApp offerings.

TABLE 4.2 Sample vApp Offerings

vApp	Configuration
2-Tier Standard Compute	1x Standard RHEL Web virtual machine
	1x Standard Windows Server 2008 Application virtual machine
3-Tier Standard Compute, Advanced Database	1x Standard RHEL Web virtual machine
	1x Standard RHEL application virtual machine
	1x Advanced MySQL Database virtual machine
3-Server Standard Plus Compute (not necessarily tiered)	3x Standard Plus Windows Server 2008 Application virtual machine

4.7.1.1.6 Establish Service Levels for vCloud Services in the Service Catalog

To provide an appropriate level of service for the vCloud customers' requirements, services should be further differentiated by their corresponding service levels. Service levels can be defined with availability and recoverability attributes such as Recovery Time Objective (RTO), Recovery Point Objective (RPO), and incident response times. The attributes can be applied to the different components within the service catalog.

It is possible to design for different service levels for the virtual machines contained in a vApp. For example, a vApp could contain multiple web servers to provide resilience in the event of server failure, and thus a lower RTO for the service.

Virtual datacenters provide abstracted physical and virtual resources. Different service levels can be defined by using (or not using) the underlying hardware technology (such as server capabilities, storage array technologies, storage protocols, and replication) and virtualization technology (HA, DRS, VMware vSphere vMotion®, and others).

Combined, vApps and the capabilities of the virtual datacenters on which they can be deployed offer the capability to create a powerful and extensive vCloud service catalog.