NetIQ® Cloud Manager

Product Overview

June 2016

Legal Notice

For information about legal notices, trademarks, disclaimers, warranties, export and other use restrictions, U.S. Government rights, patent policy, and FIPS compliance, see https://www.netiq.com/company/legal/.

Copyright © 2016 NetIQ Corporation. All Rights Reserved.

Contents

About This Book and the Library 7

1 What Cloud Manager Does 9

1.1 Provides Multi-Tenancy . . . 9
1.2 Organizes and Monitors Resources . . . 10
1.3 Provides a Catalog for Building Business Services . . . 10
1.4 Customizes Service Offerings for Organizations . . . 10
1.5 Exposes Business Service Costs . . . 11

2 The Cloud Manager Components 13

3 The Cloud Manager Environment 17

3.1 Zones and Resource Groups . . . 18
3.2 Workload Templates . . . 19
3.3 Service Levels . . . 21
3.4 Organizations and Business Groups . . . 22

4 Cloud Deployment Structures 25

4.1 Private or Hosted Private Cloud . . . 25
4.2 Public Cloud . . . 27
4.3 Hybrid Cloud . . . 28

5 Basic Orchestration Concepts 31

5.1 Understanding Cloud Manager Orchestration Architecture . . . 31
5.1.1 The Orchestration Server . . . 31
5.1.2 The Orchestration Agent . . . 33
5.1.3 The Resource Monitor . . . 34
5.1.4 The Orchestration Console and Command Line Tools . . . 34
5.1.5 Entity Types and Managers . . . 34
5.1.6 Jobs . . . 37
5.1.7 Constraint-Based Job Scheduling . . . 40
5.2 Understanding Orchestration Functionality . . . 41
5.2.1 How Do I Interact with the Orchestration Server? . . . 41
5.2.2 How Orchestration Components Communicate . . . 43
5.2.3 Resource Virtualization . . . 45
5.2.4 Policy-Based Management . . . 45
5.2.5 Grid Object Visualization . . . 46
5.2.6 Understanding Job Semantics . . . 46
5.2.7 Distributed Messaging and Failover . . . 47

6 Server Discovery and Multicasting 49

6.1 Multicast Routes . . . 49
6.2 Multi-homed Hosts . . . 49
6.3 Multiple Subnets . . . 50

6.4 Datagrid and Multicasting . . . 50
6.5 Datagrid Multicast Interface Selection . . . 50

7 Cloud Manager Orchestration and LDAP Authentication 51

7.1 What is LDAP? . . . 51
7.2 Understanding LDAP Structure . . . 52
7.2.1 The Distinguished Name . . . 52
7.2.2 The Relative Distinguished Name . . . 52

8 Cloud Manager Orchestration Security 55

8.1 User and Administrator Password Hashing Methods . . . 55
8.2 User and Agent Password Authentication . . . 55
8.3 Password Protection . . . 56
8.4 TLS Encryption . . . 56
8.4.1 Setting TLS Options . . . 56
8.4.2 Updating the TLS Server Certificate . . . 57
8.5 Security for Administrative Services . . . 58

9 User Concepts 59

9.1 Organization Scope versus System Scope . . . 59
9.2 Cloud Manager Roles . . . 59
9.2.1 Descriptions . . . 60
9.2.2 Rights . . . 61
9.2.3 Rights by Organization Role . . . 67
9.3 Cloud Manager User Groups versus LDAP User Groups . . . 68
9.4 Roles That Can Create User Accounts and User Groups . . . 69

10 Workload Template Concepts 71

10.1 Workload Template Components . . . 71
10.2 Pre-Populated Template Settings . . . 72
10.3 Workload Template Changes . . . 72
10.4 VM Template Changes . . . 72
10.5 Workload Template Retirements . . . 72
10.6 Workload Template Deletions . . . 73
10.7 Catalog Manager Role . . . 73

11 Resource Concepts 75

11.1 Resource Group Concepts . . . 75
11.1.1 VM Host Recommendations . . . 75
11.1.2 Shared and Dedicated Resource Groups . . . 75
11.1.3 Service Levels . . . 75
11.1.4 Resource Group Examples . . . 76
11.2 Advanced Network Configuration Concepts . . . 76
11.2.1 Understanding IP Address Discovery for Native IPAM . . . 76

12 Task Concepts 79

12.1 Types of Tasks . . . 79
12.2 Task Order in the Workflow Process . . . 80
12.3 Task Assignments and Owners . . . 80

A Cloud Manager Terminology **81**

About This Book and the Library

The *Product Overview* provides information about the features, functionality, and operational concepts of Cloud Manager.

Intended Audience

This document is intended for IT staff, such as IT service providers and operators, as well as business and technical users, who use Cloud Manager to build, manage, and consume cloud computing resources. A knowledge of VMware hypervisor technology is assumed.

Information in the Library

The library for this product is available in HTML and PDF formats on the Cloud Manager Documentation (https://www.netiq.com/documentation/cloud-manager) website.

The Cloud Manager library provides the following information resources:

Release Notes

Provides information about new features and enhancements in the release, as well as any known issues.

Product Overview

Provides information about the features, functionality, and operational concepts of Cloud Manager.

Installation and Upgrade Guide

Provides detailed planning and installation information for the software, as well as upgrade information.

Procedures Guide

Provides conceptual information, an overview of the user interface, and step-by-step guidance for common tasks.

Administrator Reference

Provides detailed reference information about tools and interfaces used by this product.

Help

Provides context-sensitive information and step-by-step guidance for common tasks as you work in the interface.

Feedback

If you have suggestions for documentation improvements, click **comment on this topic** at the bottom of any page in the HTML version of the documentation. You can also email Documentation-Feedback@netiq.com. We value your input and look forward to hearing from you.

Additional Resources

We encourage you to use the following additional resources online:

- Cloud Manager Forum (https://forums.netiq.com/forumdisplay.php?13-Cloud-Manager): A web-based community of product users where you can discuss product functionality and advice with other product users.
- Cloud Manager Product page (https://www.netiq.com/products/cloud-manager/): A web-based product brochure that provides information about features, how to buy, technical specifications, frequently asked questions, and a variety of resources such as videos and white papers.
- NetIQ User Community (https://www.netiq.com/communities/): A web-based community with a variety of discussion topics.
- NetIQ Support Knowledgebase (https://www.netiq.com/support/kb/): A collection of in-depth technical articles.
- NetIQ Support Forums (https://forums.netiq.com/forum.php): A web location where product users can discuss NetIQ product functionality and advice with other product users.
- MyNetIQ (https://www.netiq.com/f/mynetiq/): A website offering product information and services, such as access to premium white papers, webcast registrations, and product trial downloads.

Technical Support

For specific product issues, please contact our Technical Support team at www.netiq.com/support.

1 What Cloud Manager Does

Cloud Manager transforms your virtual infrastructure into a true Cloud environment. Built to operate with your existing VMware virtual hosts, Cloud Manager accelerates delivery of services through on-demand requesting of *workloads* and automated provisioning of the workloads.

Whether you are an private or public service provider, your customers demand timely access to the computing resources needed to run their businesses. Even with a virtual infrastructure in place, providing new services to customers can require you to research the requirements, create or customize the appropriate virtual machines, and then deploy the virtual machines accordingly, all of which can exhaust valuable time and effort.

Cloud Manager lets you expose your virtual computing resources in a manner that enables your customers to easily consume them for business services and you to deliver the services efficiently, automatically, and on time.

The following sections explain what Cloud Manager does to transform your virtual infrastructure into flexible, reliable, and secure Cloud environment:

- Section 1.1, “Provides Multi-Tenancy,” on page 9
- Section 1.2, “Organizes and Monitors Resources,” on page 10
- Section 1.3, “Provides a Catalog for Building Business Services,” on page 10
- Section 1.4, “Customizes Service Offerings for Organizations,” on page 10
- Section 1.5, “Exposes Business Service Costs,” on page 11

1.1 Provides Multi-Tenancy

Cloud Manager enables you to provide Cloud services to multiple tenants, referred to as *organizations*, at one time. As the service provider, you assign the resources that an organization can use for its business services. These resources might be dedicated to a single organization or shared among multiple organizations, depending on your business model and customer requirements.

Within an organization, members of the organization are assigned *roles* that let them control and monitor the deployment of business services for their organization. Some members might have rights to request business services, some to approve or deny business service requests, some to control organization membership, and some to create business groups (organization subunits) and allocate organization resources to the business groups.

Depending on the scope of your Cloud services, an organization can represent different units. For example, if you are a public service provider, each organization would most likely represent a company. However, if you are an enterprise IT department, an organization might represent your company or a single department or cost center within your company.

1.2 Organizes and Monitors Resources

One of the more difficult management activities involved with providing Cloud services to multiple organizations is ensuring that each organization has access to only the resources that it should. To alleviate this problem, Cloud Manager enables you to group resources and assign the resource groups to the appropriate organizations. When an organization creates a business service, the business service is deployed to a resource group assigned to the organization.

A *resource group* is a collection of hosts or clusters and their associated resources (CPUs, memory, networks, and storage). In VMware vSphere environments, a resource group can also be a resource pool.

You can have both dedicated and shared resource groups. A dedicated resource group services only one organization, while a shared resource group services more than one organization.

Cloud Manager also monitors resource utilization for the entire system and for individual organizations. By comparing used resources against resource capacity, you can ensure that your overall system and each individual organization has sufficient resources

1.3 Provides a Catalog for Building Business Services

Most customers don't want to concern themselves with the details of your virtual infrastructure. All they really want is to run their business services and to know that they are receiving the level of support needed for those services.

Cloud Manager removes all customer interaction with your virtual infrastructure through the use of a catalog. The catalog consists of *workload templates* and *service levels* that you create and make available to the customer.

- **Workload templates:** A *business service* can have one or more workloads (virtual machines). The workloads are created from workload templates the customer selects from the catalog. The workload template identifies a virtual machine and the amount of resources (virtual CPUs, memory, networks, and disk space) it needs to run.
- **Service levels:** A service level associates a business service workload with 1) the resource group where it will be deployed, 2) the support objectives (such as availability, response time, and quality), and 3) the cost of the resources and support. The customer selects a service level for each workload in the business service.

When a customer needs a new business service, he or she creates the business service workloads from the workload templates you have made available and selects the service level for each of the workloads. They don't need to know anything about the virtual infrastructure to successfully deploy their business service.

1.4 Customizes Service Offerings for Organizations

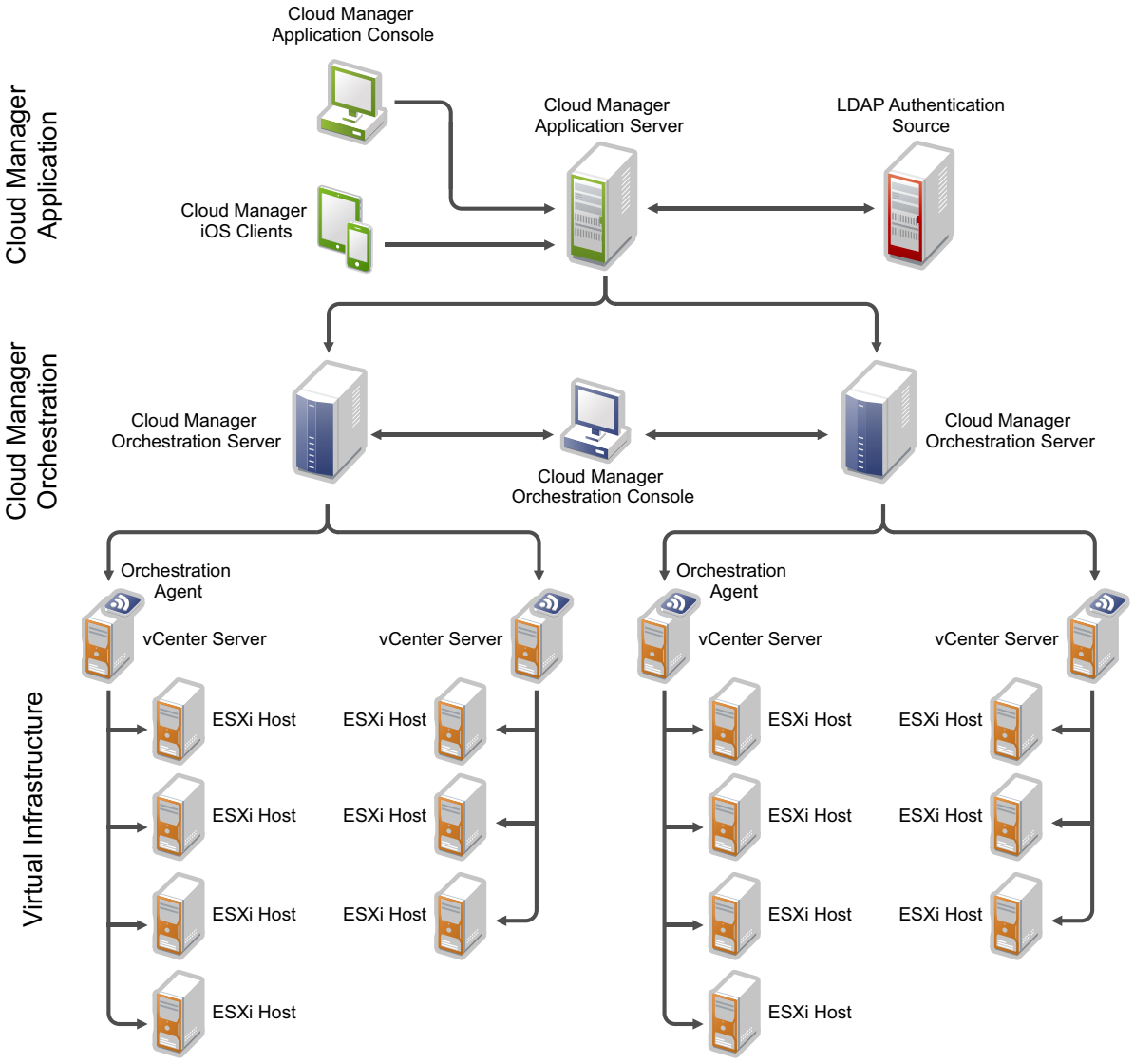
Not all organizations need the same types of business services. Cloud Manager lets you customize your service offerings for each organization by determining which workload templates and service levels are available to an organization. While your catalog might contain hundreds of workload templates and service levels, you can assign only the ones needed by an organization.

1.5 Exposes Business Service Costs

You can assign costs to the various components associated with a business service. This includes a workload's setup and license costs, its resource costs, and its support costs. The total monthly cost for the business service is available to the customer before requesting the service and is also available through cost reports.

2 The Cloud Manager Components

Cloud Manager consists of two main components: the Cloud Manager Application Server and the Cloud Manager Orchestration Server. The Cloud Manager Application Server and the Cloud Manager Orchestration Server sit on top of your virtual infrastructure to automate Cloud services for your customers.



Cloud Manager Application Server

The Cloud Manager Application Server provides the portal for initiating and managing business services. When a customer requests a business service through the Application Console, the Application Server sends instructions that the Orchestration Server uses to provision the service's workloads (virtual machines) through the virtual infrastructure technologies.

- **Application Console:** A web application that can be run on any computer with a supported web browser. The console is for both Cloud Manager administrators and users. Cloud Manager administrators use the console to organize computing resources so that users can consume them as business services. Users access the console to request and manage business services. Login to the console occurs through an LDAP directory designated as the authentication source.
- **Application Server:** Supports the Application Console and communicates with Orchestration Servers to provide instructions for deploying, managing, and removing business service workloads. It also performs user authentication with the LDAP source.

Cloud Manager Orchestration Server

The Cloud Manager Orchestration Server automates the creation and management of business service workloads in the virtual infrastructure. When the Orchestration Server receives a business service request from the Application Server, the Orchestration Server directs the creation of the service's workloads from the appropriate VM template and the deployment of the workloads to the appropriate VM host. In addition, Cloud Manager Orchestration Server discovers and surfaces your virtual infrastructure resources (hypervisor technologies, VM hosts, VM templates, and so forth) in the Cloud Manager Application Console so that you can organize them into the catalog components that customers use to build their business services.

- **Orchestration Server:** Receives workload instructions from the Application Server and directs the creation and management of those workloads by the virtual infrastructure. Depending on the size of your virtual infrastructure, you might have one or many Orchestration Servers.
- **Cloud Manager Orchestration Console:** Monitors and manages the activity of the Orchestration Servers, enabling you to view and troubleshoot jobs associated with workload creation and management.
- **Cloud Manager Orchestration Agent:** Provides communication between the Orchestration Server and the VM hosts managed by the server.

Virtual Infrastructure

The VMware virtual infrastructure forms the foundation of the Cloud Manager physical topology. The hypervisor technology virtualizes the underlying physical resources and enables the creation and management of virtual machines.

The illustration shown above does not represent all components of the virtual infrastructure (such as networks, storage, virtual machines, and so forth). It is intended simply to show how the Cloud Manager components sit on top of your virtual infrastructure and interact with it to provide cloud services. Cloud Manager documentation assumes that the person who will implement Cloud Manager is knowledgeable about your virtual infrastructure components and management. Refer to your VMware documentation for information.

Cloud Marketplace 2.2.1 App for iPad

The Cloud Marketplace mobile app for iPad accesses available Cloud Manager 2.5 resources to provide Cloud Manager sales representatives a fast and easy way to assemble a customized business service quote for interested customers. Using the built-in shopping cart of quotes, customers can compare these costs and build a deployable business services order wherever they are. The app is available for free download at the at the Apple iTunes Store.

You can access several video tutorials to help you become familiar with this product. See Tutorials for Cloud Marketplace (https://www.youtube.com/playlist?list=PLuO-OPQ1FcWjxuFTdQnabShinOwX2AL3p) on YouTube. The links to these videos are also available on the Cloud Marketplace help page.

The Cloud Marketplace app can be re-themed with the look and feel or localization that you want. For information about how to re-theme, see “Rebranding the Cloud Marketplace Mobile App” in the *Cloud Manager Administrator Reference* manual. The templates you need for the rebranded themes you create are available in a downloadable `.zip` file (https://www.netiq.com/documentation/beta/cloud-manager-2-5/resources/marketplace_rebrand.zip).

3 The Cloud Manager Environment

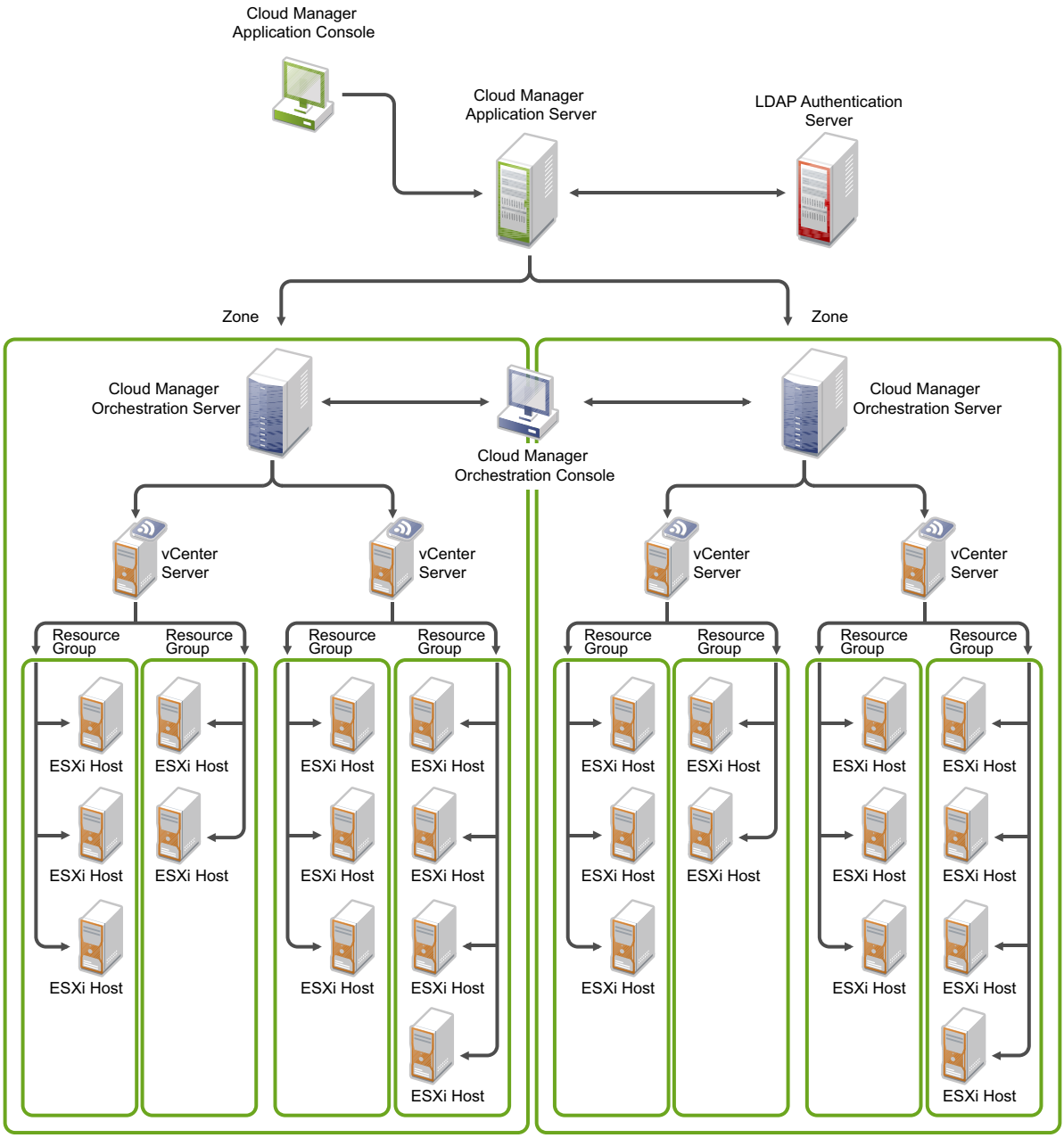
Cloud Manager turns your virtual infrastructure into a Cloud environment that provides automated business services to your customers. The Cloud environment consists of a variety of components. Some of the components, such as *zones* and *resource groups*, provide ways to organize your virtual infrastructure resources so that Cloud Manager knows where to run business services. These components are mostly hidden to users. Other components, such as *service levels* and *workload templates*, form the core of business services and are readily visible to users.

The following sections introduce the key components of the Cloud Manager environment:

- Section 3.1, “Zones and Resource Groups,” on page 18
- Section 3.2, “Workload Templates,” on page 19
- Section 3.3, “Service Levels,” on page 21
- Section 3.4, “Organizations and Business Groups,” on page 22

3.1 Zones and Resource Groups

A Cloud Manager zone is an Orchestration Server and its managed resources (hosts, clusters, resource pools, networks, storage, and so forth). Within a zone, these resources are organized into resource groups, as shown in the following illustration.



A resource group identifies a collection of hosts (and their associated networks and storage). When a workload is deployed, it is assigned to the resource group and provisioned using any of the resources within the group.

A resource group has the following characteristics:

- Can include standalone hosts and clusters. Optionally, a resource group can be a vSphere resource pool. All host or pool resources (CPUs, memory, networks, disks, and so forth) should provide the same performance level so that a workload can run equally well on any of the resources.
- Cannot span zones. All resources in the group must reside in the same zone.
- Cannot share storage repositories with other resource groups.

As an example, you might form a Business Critical resource group that consists of high-performance vSphere hosts intended for critical production workloads. At the same time, you might have a Lab resource group that consists of standard-performance vSphere hosts intended for non-production workloads.

3.2 Workload Templates

Workload templates are used to create business service workloads. A workload template defines the following:

- The VM template used to create the workload.
- Resource customizations to apply to the workload. For example, if the VM template provides 2 CPUs, you can increase that number to 4 CPUs.
- The license and setup costs associated with a workload created from the template.

You create a catalog of workload templates from which users can choose when requesting business services. Depending on the needs of your users, you might have many workload templates.

The examples in Figure 3-1 are based on workload operating system, but you might have workload templates that provide not only the operating system but also applications or other services.

Figure 3-1 *Example Workload Templates*



Windows Server 2008 64-bit

VM Template:	WinServer2008_64
Operating System:	Microsoft Windows Server 2008 64-bit
Hypervisor:	Hyper-V
CPUs:	2
Memory:	4 GB
NICs:	2
Additional Disks:	2 (800 GB)



Windows Server 2003 32-bit

VM Template:	WinServer2003_32
Operating System:	Microsoft Windows Server 2003 32-bit
Hypervisor:	VMWare
CPUs:	4
Memory:	12 GB
NICs:	4
Additional Disks:	2 (2000 GB)



SUSE Linux Enterprise Server 10 64-bit

VM Template:	SLES10_64
Operating System:	SUSE Linux Enterprise Server 10 SP2 64-bit
Hypervisor:	Xen
CPUs:	4
Memory:	8 GB
NICs:	3
Additional Disks:	3 (1500 GB)

3.3 Service Levels

Service levels are associated with resource groups. They determine how much it costs to run business service workloads on the resources (vCPUs, memory, networks, and storage). They can also include performance expectations for those resources as well as the level of IT support provided for workloads running. The performance and support expectations, referred to as service level objectives, can also have costs associated with them.

Consider the example service levels in Figure 3-2 for Platinum, Gold, and Bronze:

Figure 3-2 *Example Service Levels*

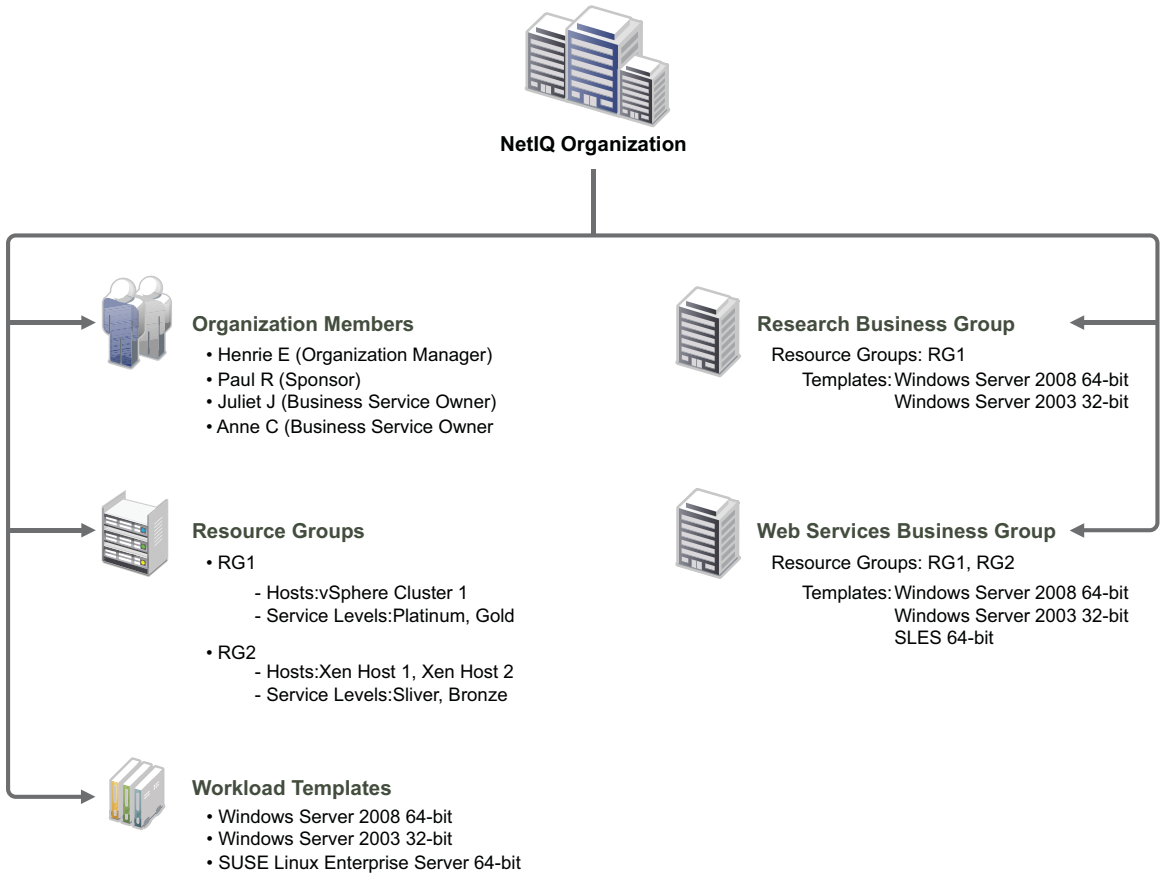


The Platinum service level runs workloads in the Business Critical resource group and sets service objectives of 99% availability and 2 hour response time for support issues. It has the highest resource cost and objectives cost. The other two service levels provide slightly lower resource quality and service objectives for the workloads at a lesser cost.

3.4 Organizations and Business Groups

An organization represents a tenant for which you are providing Cloud services. A business group is a subunit within the organization for which business services are deployed. An organization must have at least one business group. Typically, an organization represents a company and business groups represent the departments or cost centers that need to deploy business services.

Figure 3-3 *Example Organization*



To provide organizations with the resources needed to deploy business services, resource groups and workload templates are assigned to organizations. In the illustration above, two resource groups (RG1 and RG2) and three workload templates (Windows Server 2008 64-bit, Windows Server 2008 32-bit, and SUSE Linux Enterprise Server 64-bit) are assigned to the NetIQ organization.

Because an organization's business groups might not need access to the same resources and workload templates, the organization's resources and workload templates must be assigned to the business units. In the above illustration, the organization has two business groups (Research and Web Services). Both resource groups and all three workload templates are assigned to the Web Services business group to be used for its business services, but only one resource group and two templates are assigned to the Research business group for its business services.

Users are added as organization members and given roles within the organization. A role provides rights to perform specific activities, such as deploying business services or managing the organization's membership. In the above illustration, both Juliet J and Anne C have the Business Service Owner role for the organization. This allows them to create business services for both the Research and Web Services business groups. It is also possible to assign a user a role at the

business group level rather than at the organization level. For example, if Juliet J was assigned the Business Service Owner role for the Research business group rather than for the organization, she could create business services only for the Research business group.

4 Cloud Deployment Structures

Cloud Manager components can be configured for and deployed in a variety of cloud computing structures.

- Section 4.1, “Private or Hosted Private Cloud,” on page 25
- Section 4.2, “Public Cloud,” on page 27
- Section 4.3, “Hybrid Cloud,” on page 28

4.1 Private or Hosted Private Cloud

In a private cloud structure, the networking between the cloud provider and cloud consumers is controlled by one organization. All cloud resources belong to that organization, and are hosted on premise.

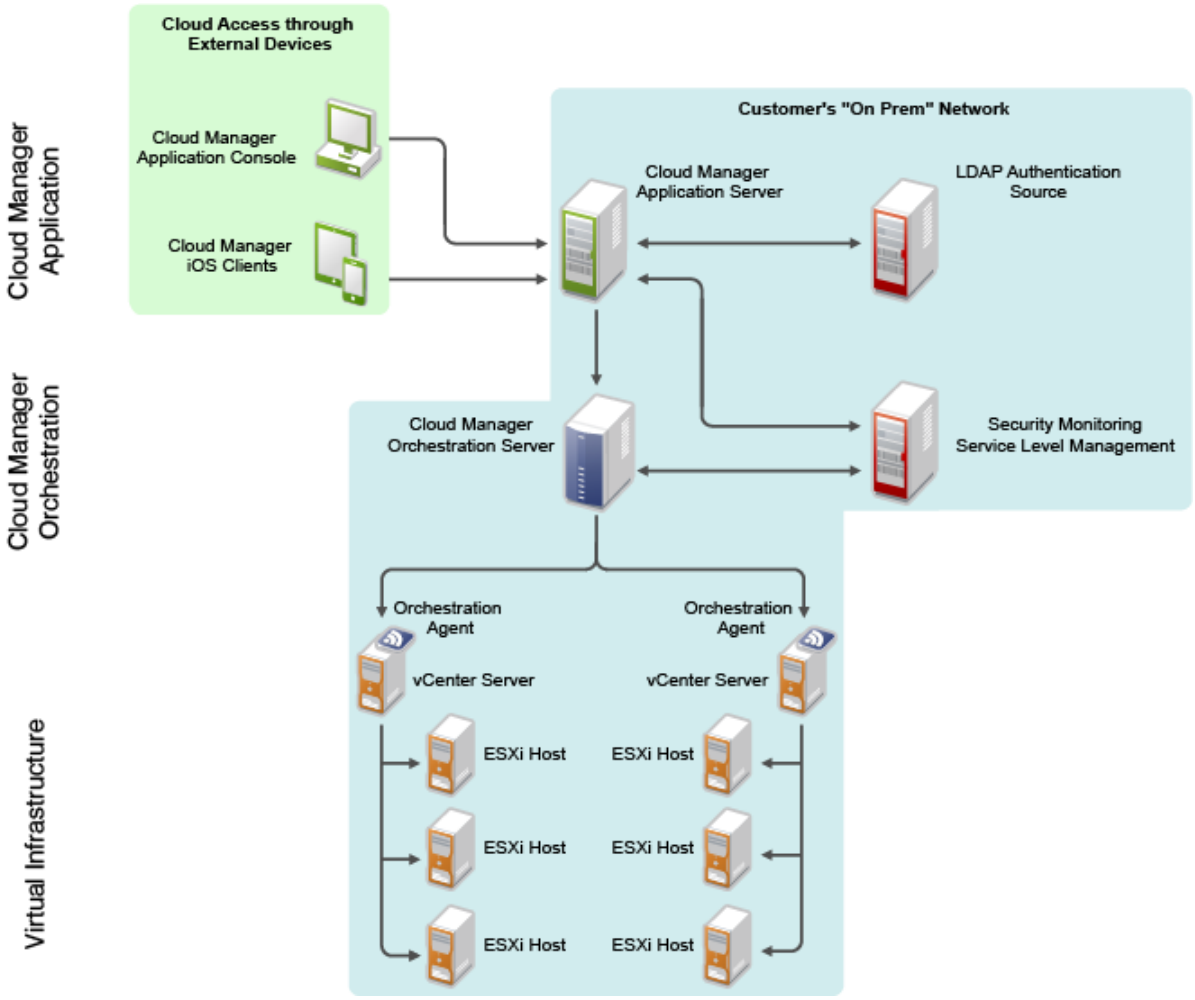
A *hosted* private cloud is hosted in a service provider’s environment, contracted for use by an organization. In this arrangement, the organization has the operational responsibility, while access control, power management and networking are provided by the service provider.

Depending on security requirements, the private cloud is implemented as dedicated components in a rack of servers, or perhaps in a secured room or cage. In most cases, a private cloud is custom-designed, and fulfills a particular set of business requirements. Cloud Manager complies with the security, heterogeneity and availability requirements of such a solution. It can also enable granular service level management and compliance delivery by integrating other NetIQ solutions.

If you use Cloud Manager as the basis of a private cloud, your organization should be committed to virtualizing its business computing resources. This requires a significant decisions regarding your existing resources.

The following diagram illustrates the components you should consider as integral in a private cloud environment where Cloud Manager is the principal management tool of virtual computing resources.

Figure 4-1 *Cloud Manager components in a private cloud structure*



There are several deployment considerations in a Cloud Manager private cloud structure:

- Access to virtual machine consoles is provided by the Cloud Manager Application Server Console (a web-based interface). Alternatively, users can install VNC or RDP services in their cloud workloads and use those services to log into their workloads remotely, over the organization network.
- If you set up a firewall between the Cloud Manager Application Server and cloud consumers, the firewall must allow inbound connections to the server on port 8183 (the default port) and on the remote console port (a high port between 8000 and 65535, as configured in Cloud Manager).
- Cloud Manager must have access to the ports that expose workload consoles on the hypervisors. For VMware 5.*x,* these are ports 50000-50999.
- The organization owns all operations responsibilities for the cloud, including template management, storage, and resource pooling. Cloud Manager administrative requirements vary according to how you structure the cloud Administrative duties can be assigned directly or delegated.

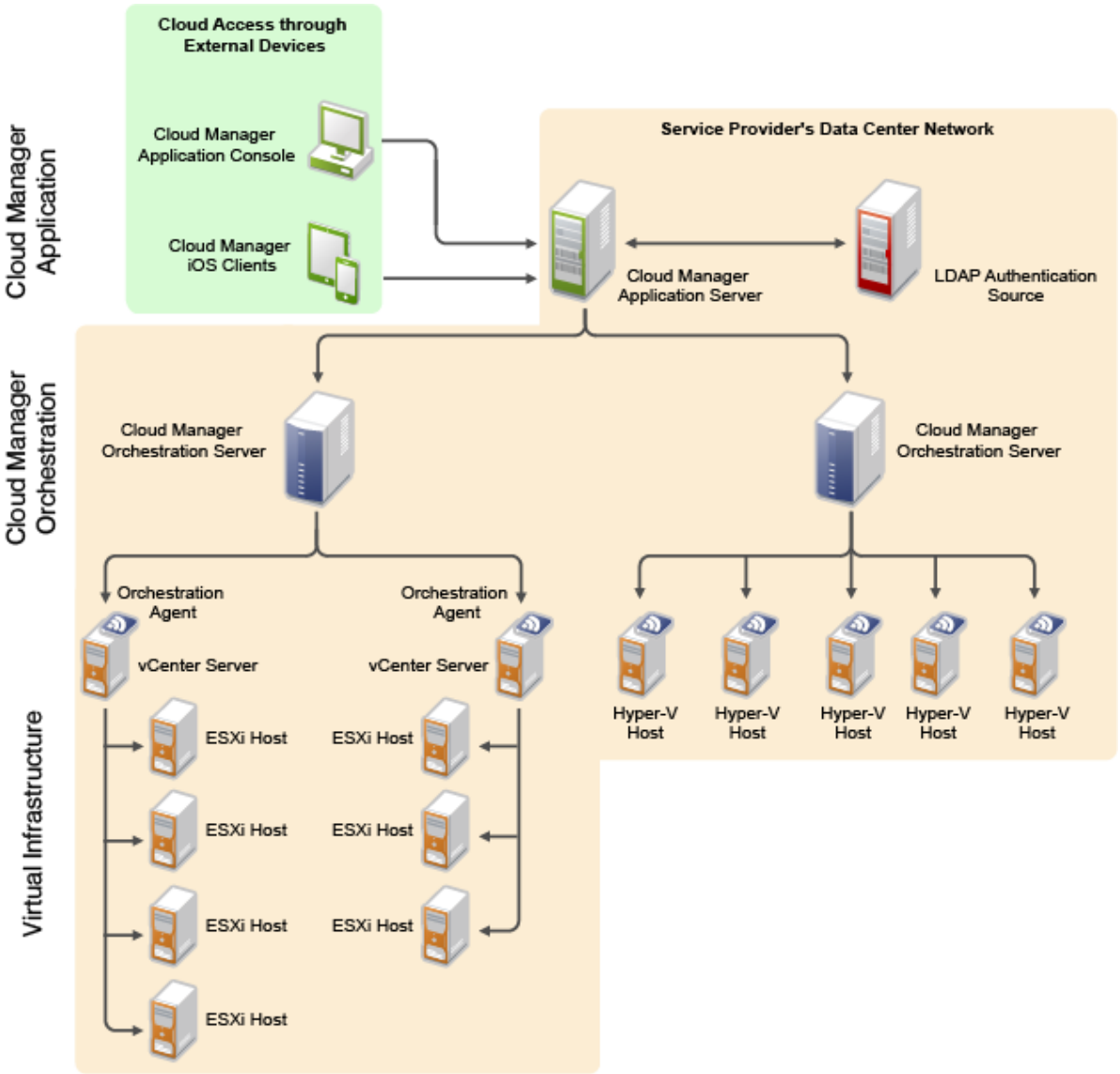
4.2 Public Cloud

In a public cloud environment, a service provider makes computing services (applications and storage, for example) available and accessible to the general public over the Internet. Public cloud services can be free or offered on a pay-per-usage model. The self-service capabilities of Cloud Manager, combined with its embedded approval and change management, delivers the base for a trusted, low-maintenance public cloud solution where all costing elements are exposed and provided prior to consumption.

Although there might be little or no difference between public and private cloud architectures, service providers typically incorporate substantial security measures before they offer services over a non-trusted network for customers in the public domain.

The following diagram illustrates the basic components you should consider as integral if you plan on building a public cloud environment where Cloud Manager is the principal management tool of virtual computing resources.

Figure 4-2 *Cloud Manager components in a public cloud structure*



There are several deployment considerations in a Cloud Manager public cloud structure:

- Access to virtual machine consoles is provided by the Cloud Manager Application Server Console (a web-based interface).
- A reverse proxy (such as NetIQ Access Manager) takes care of the SSL endpoint work and redirects from port 80/443 to port 8183 on the Cloud Manager Application Server.
- The external firewall must allow incoming connections to the reverse proxy on port 80/443 and on the remote console port (a high port between 8000 and 65535, as configured in Cloud Manager). Alternatively the remote console port can be tunneled through the reverse proxy.
- Cloud Manager must have access to the ports that expose workload consoles on the hypervisors. For VMware 5.*x*, these are ports 50000-50999.
- Administrative functions are separated between the service provider and the consuming organizations. The service provider (whose administrative rights can be directly assigned or delegated) maintains and operates the Cloud Manager infrastructure (including such things as template management, storage, and resource pooling) while the organizations paying for services use the Cloud Manager Application Server Console to manage a VM's life cycle (including change requests).

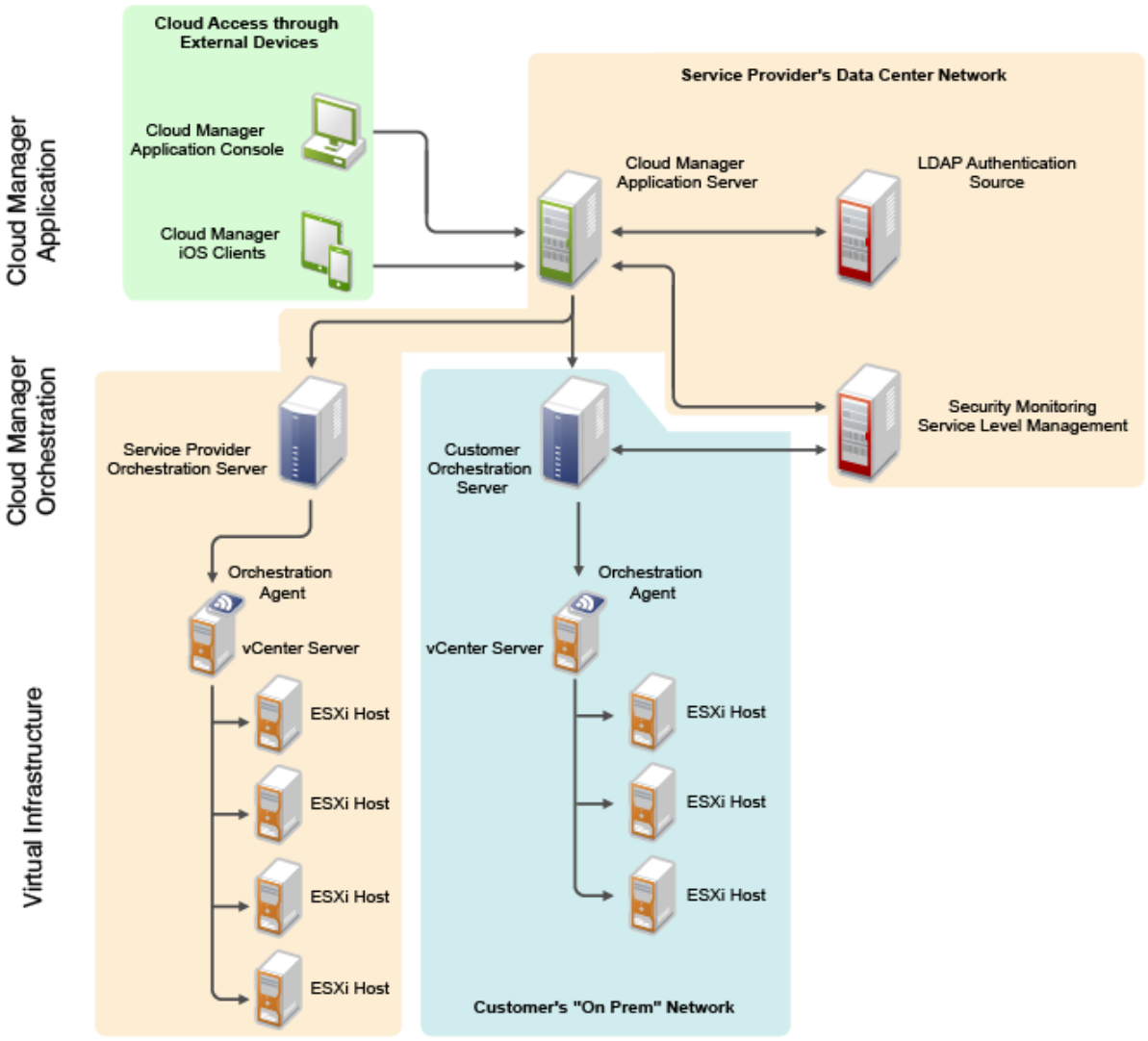
4.3 Hybrid Cloud

A hybrid cloud environment is similar to a public cloud environment because it is operated by a service provider. However, the resources available to a consuming organization (such as applications and storage) are a mix of resources owned by the service provider and resources owned by the organization itself.

In Cloud Manager, a hybrid cloud is implemented by using an on-premise zone. The Cloud Manager Application Server runs on the computing infrastructure of the service provider, while the Cloud Manager Orchestration Server that manages the resources of the consuming organization runs on the computing infrastructure of that organization. Tunneling or a VPN is necessary to establish the communication path between the Cloud Manager Application Server and the on-premise Cloud Manager Orchestration Server and the virtual infrastructure.

The following diagram illustrates the basic components you should consider as integral if you plan to allow an external service provider to use Cloud Manager to administer the on-premise virtual resources of your organization.

Figure 4-3 *Cloud Manager components in a hybrid cloud configuration*



There are several deployment considerations in a Cloud Manager hybrid cloud structure:

- Access to virtual machine consoles is provided by the Cloud Manager Application Server Console (a web-based interface).
- A reverse proxy (e.g. NetIQ Access Manager) takes care of the SSL endpoint work and redirects from port 80/443 to port 8183 on the Cloud Manager Application Server.
- The external firewall must allow incoming connections to the reverse proxy on port 80/443 and on the remote console port (a high port between 8000 and 65535, as configured in Cloud Manager). Alternatively the remote console port can be tunneled through the reverse proxy.
- In the service providers infrastructure and in the on-premise organization infrastructure, Cloud Manager must have access to the ports that expose workload consoles on the hypervisors. For VMware 5.*x*, these are ports 50000-50999.

  You should set up either have server-to-server VPN networking or tunneling that is activated to the on-premise organization infrastructure.
- The minimum bandwidth requirements for efficient remote management (including a remote console session) is 16 MB. Without a remote console session, the requirement is 1 MB.

- Responsibility for template management, storage and resource pooling is owned by the service provider and by the organization, respectively, in their separate infrastructures. All workload templates must reside on the infrastructure where their corresponding virtual machines run: templates can only be consumed only in the infrastructure where they reside (service provider or organization).
- Administrative functions are separated between the service provider and the consuming organizations. The service provider (whose administrative rights can be directly assigned or delegated) maintains and operates the Cloud Manager infrastructure (including such things as template management, storage, and resource pooling) while the organizations paying for services use the Cloud Manager Application Server Console to manage a VM's life cycle (including change requests).
- Operational tasks in the cloud environment as a whole are the responsibility of the service provider. The paying organization consumes that infrastructure, which also encompasses the resources of the customer's own on-premises zone. The organization's activity in the cloud would typically include VM life cycle operations, including change requests.
- Administrative responsibilities such as template management, storage, and resource pooling are separately executed by the service provider or the organization, depending on the infrastructure involved. Workload templates can only be consumed in the infrastructure (that is, the Cloud Manager zone) where they reside. The administrative role can be directly assigned or delegated.

5 Basic Orchestration Concepts

This section describes concepts about the Orchestration architecture and functions.

- Section 5.1, “Understanding Cloud Manager Orchestration Architecture,” on page 31
- Section 5.2, “Understanding Orchestration Functionality,” on page 41

5.1 Understanding Cloud Manager Orchestration Architecture

This section contains information about the following topics:

- Section 5.1.1, “The Orchestration Server,” on page 31
- Section 5.1.2, “The Orchestration Agent,” on page 33
- Section 5.1.3, “The Resource Monitor,” on page 34
- Section 5.1.4, “The Orchestration Console and Command Line Tools,” on page 34
- Section 5.1.5, “Entity Types and Managers,” on page 34
- Section 5.1.6, “Jobs,” on page 37
- Section 5.1.7, “Constraint-Based Job Scheduling,” on page 40

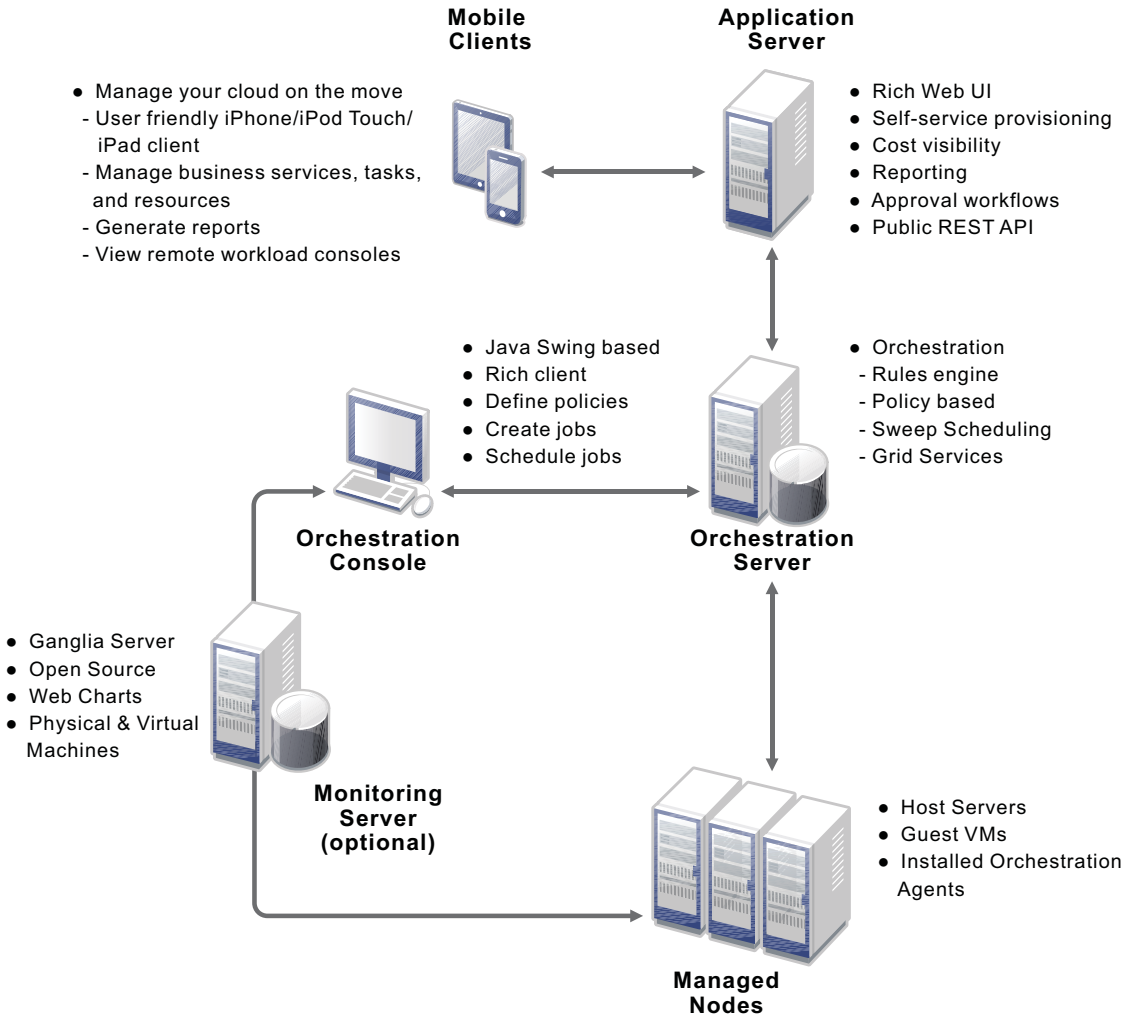
5.1.1 The Orchestration Server

The Cloud Manager Orchestration Server is an advanced data center management solution designed to manage all network resources. It provides the infrastructure that manages group of ten, one hundred, or thousands of physical or virtual resources.

The Orchestration Server can perform a wide range of distributed processing problems including high performance computing, and breaking down work, including VM life cycle management, into jobs that can be processed in parallel through distributed job scheduling.

The following figure shows a high-level perspective of how the Orchestration Server fits into Cloud Manager architecture.

Figure 5-1 *Cloud Manager Orchestration Architecture*



The Orchestration Server is the gateway between enterprise applications and resource servers. The server has two primary functions:

- To manage the resource servers
- To manage jobs to run on the computing resource

In the first function, the server manages the computing resources by collecting, maintaining, and updating their status availability, service cost, and other facts. Changes to the computing resources can be made by the administrator.

The second function of the server is to run remote applications—called jobs—on the computing resources. The Orchestration Server uses a policy-based broker and scheduler to decide when and how a job should run on the computing resources. The decisions are based on many controlled factors, including the number of computing resources, their cost, and a variety of other factors as specified by the policy constraints set up by the server administrator. The Orchestration Server runs the job and provides all the job's output responses back to the user. The server provides failover capabilities to allow jobs to continue if computing resources and network conditions degrade.

The core strength of the Orchestration Server is the capability to automatically, rapidly, and securely create and scale heterogeneous virtual environments by using a specialized VM provisioning adapter jobs that discover VMs already existing in the VMware hypervisor environment. Once the VMs are discovered, they can be cloned and then provisioned as workloads to suit the business service needs of Cloud Manager users.

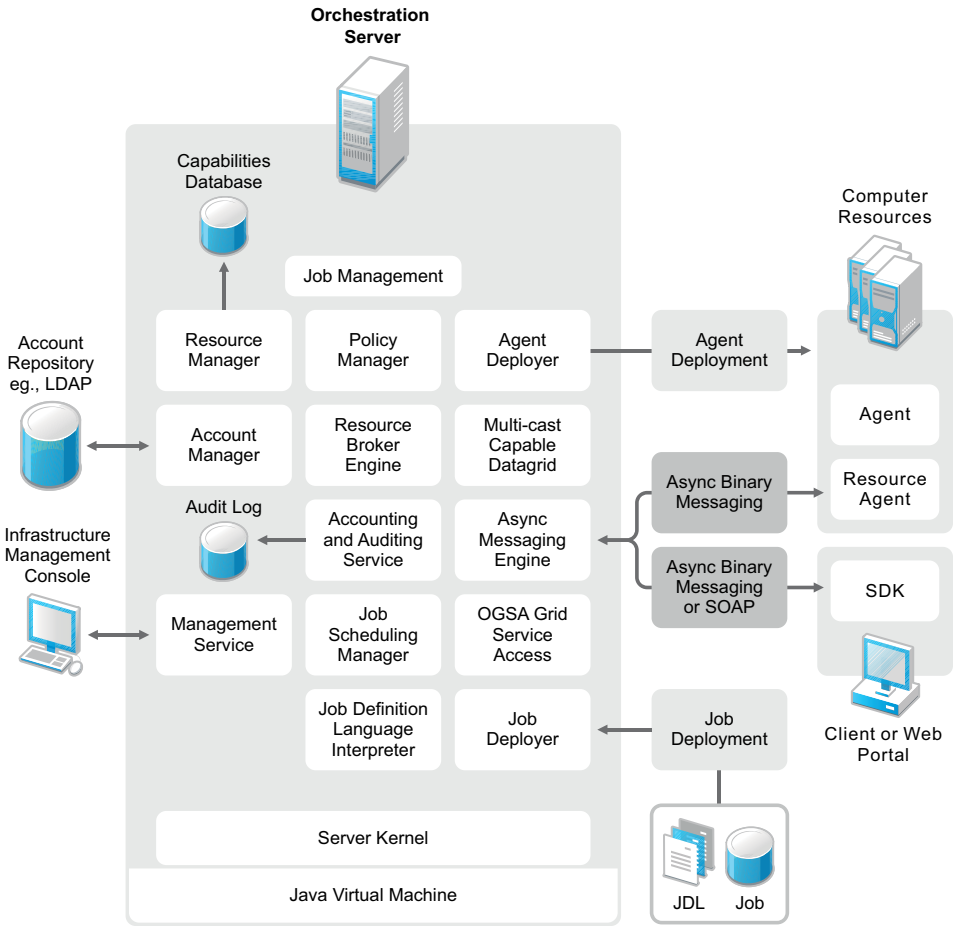
5.1.2 The Orchestration Agent

Agents are installed on all managed resources as part of the product deployment. The agent connects every managed resource to its configured server and advertises to the Orchestration Server that the resource is available for tasks. This persistent and auto-reestablishing connection is important because it provides a message bus for the distribution of work, collection of information about the resource, per-job messaging, health checks, and resource failover control.

After resources are enabled, Cloud Manager Orchestration can discover, access, and store detailed abstracted information—called “facts”—about every resource. Managed resources, referred to as “nodes,” are addressable members of the Orchestration Server “grid” (also sometimes called the “matrix”). When integrated into the grid, nodes can be deployed, monitored, and managed by the Orchestration Server, as discussed in Section 5.2, “Understanding Orchestration Functionality,” on page 41.

An overview of Cloud Manager Orchestration grid architecture is illustrated in the figure below, much of which is explained in this guide:

Figure 5-2 *Cloud Manager Orchestration Grid Architecture*



5.1.3 The Resource Monitor

Cloud Manager Orchestration enables the monitoring of your system computing resources by using its built-in Resource Monitor.

5.1.4 The Orchestration Console and Command Line Tools

The Orchestration Console and the Orchestration command line tools (sometimes called the “Orchestration Clients”) let a computing resource administrator troubleshoot, initiate, change, or shut down server functions for the Orchestration Server and its computing resources. The clients also monitor all managed computing resource job activity and provide facilities to manage application jobs. When you install the Clients on a computing resource, you are installing the following tools:

- zos command line interface
- zosadmin command line interface
- Orchestration Console
- Java SDK (toolkit)

The Orchestration Console is a graphical user interface running on Java. It provides a way for the server administrator to troubleshoot and to initiate, change, or shut down the functioning of the Orchestration Server and its resources. It also functions as a monitor of all Orchestration job activity, and it provides an interface for managing Orchestration Server jobs. For more information about the console, see the *Cloud Manager Administrator Reference*.

5.1.5 Entity Types and Managers

The following entities are some of key components of the Orchestration Server model:

- “Resources” on page 34
- “Users” on page 35
- “Job Definitions” on page 35
- “Job Instances” on page 35
- “Policies” on page 35
- “Facts” on page 35
- “Constraints” on page 36
- “Groups” on page 36
- “VM: Hosts, Images, and Instances” on page 36
- “Templates” on page 36

Resources

All managed resources, which are called nodes, have an agent with a socket connection to the Orchestration Server. All resource use is metered, controlled, and audited by the Orchestration Server. Policies govern the use of resources.

The Orchestration Server allocates resources by reacting as load is increased on a resource. As soon as we go above a threshold that was set in a policy, a new resource is allocated and consequently the load on that resource drops to an acceptable rate.

You can also write and jobs that perform cost accounting to account for the cost of a resource up through the job hierarchy, periodically, about every 20 seconds. For more information, see “Auditing and Accounting Jobs” on page 40.

A collection of jobs, all under the same hierarchy, can cooperate with each other so that when one job offers to give up a resource it is reallocated to another similar priority job. Similarly, when a higher priority job becomes overloaded and is waiting on a resource, the system “steals” a resource from a lower priority job, thus increasing load on the low priority job and allocating it to the higher priority job. This process satisfies the policy, which specifies that a higher priority job must complete at the expense of a low priority job.

Users

An Orchestration user is an individual who authenticates to the Orchestration Server for the purpose of managing (that is, running, monitoring, canceling, pausing, stopping, or starting) a deployed job, or a user who authenticates through the Cloud Manager Web Console or a Cloud Manager mobile client to manage virtual machines. The Orchestration Server administrator can use the Orchestration Console to identify users who are running jobs and to monitor the jobs that are currently running or that have run during the current server session.

Orchestration Server users must authenticate to access the system. Access and use of system resources are governed by policies. For more information, see “The User Object” in the *Cloud Manager Administrator Reference*.

Job Definitions

A job definition is described in the embedded enhanced Python script that you create as a job developer. Each job instance runs a job that is defined by the Job Definition Language (JDL). Job definitions might also contain usage policies.

Job Instances

Jobs are instantiated at runtime from job definitions that inherit policies from the entire context of the job (such as users, job definitions, resources, or groups).

Policies

Policies are XML documents that contain various constraints and static fact assignments that govern how jobs run in the Cloud Manager Orchestration environment.

Policies are used to enforce quotas, job queuing, resource restrictions, permissions, and other job parameters. Policies can be associated with any Orchestration Server object.

Facts

Facts represent the state of any object in the Orchestration Server grid. They can be discovered through a job or they can be explicitly set.

Facts control the behavior a job (or joblet) when it’s executing. Facts also detect and return information about that job in various UIs and server functions. For example, a job description that is set through its policy and has a specified value might do absolutely nothing except return immediately after network latency.

The XML fact element defines a fact to be stored in the Grid object's fact namespace. The name, type and value of the fact are specified as attributes. For list or array fact types, the element tag defines list or array members. For dictionary fact types, the `dict` tag defines dictionary members.

Facts can also be created and modified in JDL and in the Java Client SDK.

There are three basic types of facts:

- **Static:** Facts that require you to set a value. For example, in a policy, you might set a value to be False. Static facts can be modified through policies.
- **Dynamic:** Facts produced by the Orchestration Server system itself. Policies cannot override dynamic facts. They are read only and their value is determined by the Orchestration Server itself.
- **Computed:** Facts derived from a value, like that generated from the cell of a spreadsheet. Computed facts have some kind of logic behind them which derive their values.

See the example, `/opt/novell/zenworks/zos/server/examples/allTypes.policy`. This example policy has an XML representation for all the fact types.

Constraints

The constraint element of a policy can define the selection and allocation of Grid objects (such as resources) in a job. The required type attribute of a constraint defines the selection of the resource type.

For example, in order for the Orchestration Server to choose resources for a job, it uses a "resource" constraint type. A resource constraint consists of Boolean logic that executes against facts in the system. Based upon this evaluation, the Orchestration Server considers only resources that match the criteria that have been defined in constraints.

Groups

Resources, users, job definitions and virtual machines (VM) are managed in groups with group policies that are inherited by members of the group.

VM: Hosts, Images, and Instances

A virtual machine host is a resource that is able to run guest operating systems. Attributes (facts) associated with the VM host control its limitations and functionality within the Orchestration Server. A VM image is a resource image that can be cloned and/or provisioned. A VM instance represents a running copy of a VM image.

Templates

Templates are images that are meant to be cloned (copied) prior to provisioning the new copy.

5.1.6 Jobs

The Orchestration Server manages all nodes by administering jobs (and the functional control of jobs at the resource level by using joblets), which control the properties (facts) associated with every resource. In other words, jobs are units of functionality that dispatch data center tasks to resources on the network such as management, migration, monitoring, load balancing, etc.

The Orchestration Server provides a unique job development, debugging, and deployment environment that expands with the demands of growing data centers.

As a job developer, your task is to develop jobs to perform a wide array of work that can be deployed and managed by the Orchestration Server.

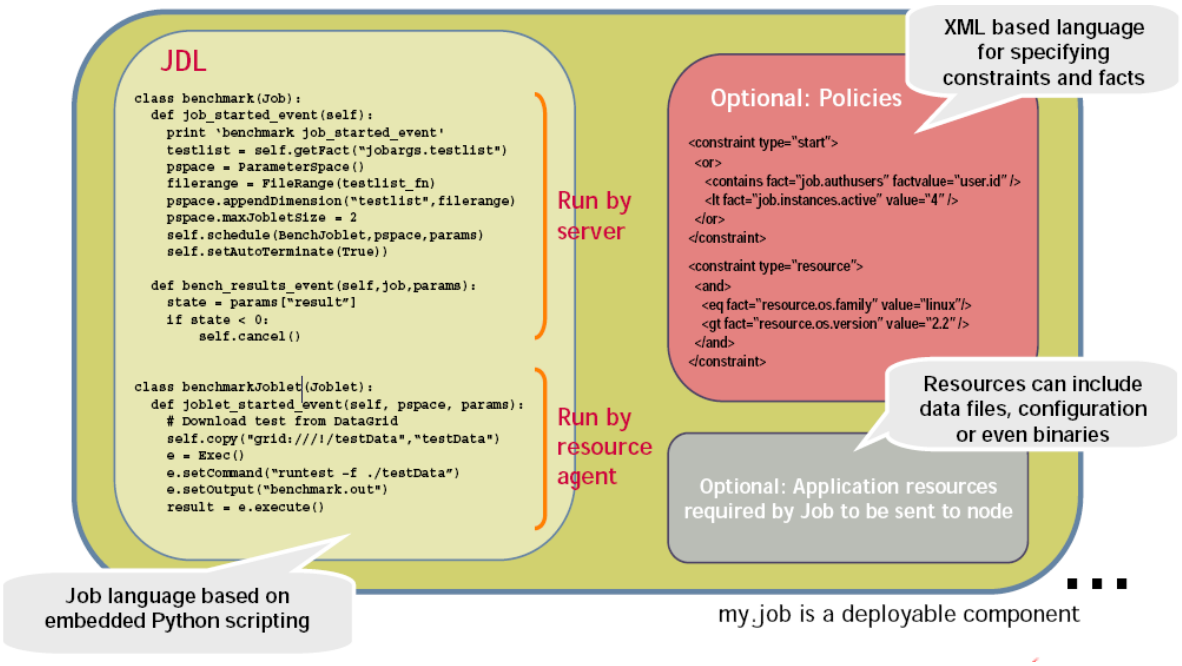
Jobs, which run on the Orchestration Server, can provide functions within the Orchestration environment that might last from seconds to months. Job and joblet code exist in the same script file and are identified by the `.jdl` extension. The `.jdl` script contains only one job definition and zero or more joblet definitions. A `.jdl` script can have only one Job subclass. As for naming conventions, the Job subclass name does not have to match the `.jdl` filename; however, the `.jdl` filename is the defined job name, so the `.jdl` filename must match the `.job` filename that contains the `.jdl` script. For example, the job files (`demoIterator.jdl` and `demoIterator.policy`) included in the `demoIterator` example job are packaged into the archive file named `demoIterator.job`, so in this case, the name of the job is `demoIterator`.

A job file also might have policies associated with it to define and control the job's behavior and to define certain constraints to restrict its execution. A `.jdl` script that is accompanied by a policy file is typically packaged in a job archive file (`.job`). Because a `.job` file is physically equivalent to a Java archive file (`.jar`), you can use the JDK JAR tool to create the job archive.

Multiple job archives can be delivered as a management pack in a service archive file (SAR) identified with the `.sar` extension. Typically, a group of related files are delivered this way.

As shown in the following illustration, jobs include all of the code, policy, and data elements necessary to execute specific, predetermined tasks administered either through the Cloud Manager Orchestration Console, or from the zos command line tool.

Figure 5-3 *Components of a Job (`my.job`,)*



Because each job has specific, predefined elements, jobs can be scripted and delivered to any agent, which ultimately can lead to automating almost any data center task. Jobs provide the following functionality:

- “Controlling Process Flow” on page 38
- “Parallel Processing” on page 38
- “Managing the Cluster Life Cycle” on page 38
- “Discovery Jobs” on page 39
- “System Jobs” on page 39
- “Provisioning Jobs” on page 40
- “Auditing and Accounting Jobs” on page 40

Controlling Process Flow

Jobs can written to control all operations and processes of managed resources. Through jobs, the Orchestration Server manages resources to perform work. Automated jobs (written in JDL), are broken down into joblets, which are distributed among multiple resources.

Parallel Processing

By managing many small joblets, the Orchestration Server can enhance system performance and maximize resource use.

Managing the Cluster Life Cycle

Jobs can detect demand and monitor health of system resources, then modify clusters automatically to maximize system performance and provide failover services.

Discovery Jobs

Some jobs provide inspection of resources to more effectively management assets. These jobs enable all agents to periodically report basic resource facts and performance metrics. In essence, these metrics are stored as facts consisting of a key word and typed-value pairs like the following example:

```
resource.loadaverage=4.563, type=float
```

Jobs can poll resources and automatically trigger other jobs if resource performance values reach certain levels.

The system job scheduler is used to run resource discovery jobs to augment resource facts as demands change on resources. This can be done on a routine, scheduled basis or whenever new resources are provisioned, new software is installed, bandwidth changes occur, OS patches are deployed, or other events occur that might impact the system.

Consequently, resource facts form a capabilities database for the entire system. Jobs can be written that apply constraints to facts in policies, thus providing very granular control of all resources as required. All active resources are searchable and records are retained for all off-line resources.

The following `osInfo.job` example shows how a job sets operating system facts for specific resources:

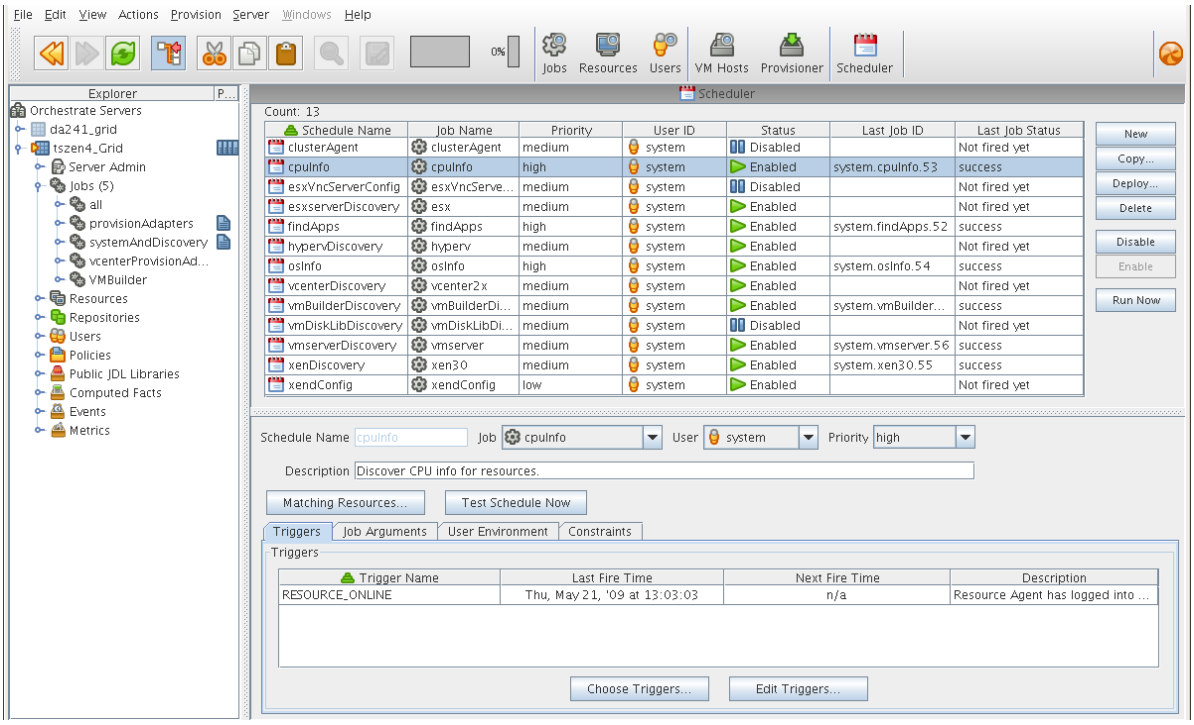
```
resource.cpu.mhz (integer) e.g., "800" (in Mhz)
    resource.cpy.vendor (string) e.g. "GenuineIntel"
    resource.cpu.model (string) e.g. "Pentium III"
    resource.cpu.family (string) e.g. "i686"
```

`osInfo.job` is packaged as a single cross-platform job and includes the Python-based JDL and a policy to set the timeout. It is run each time a new resource appears and once every 24 hours to ensure validity of the resources.

System Jobs

Jobs can be scheduled to periodically trigger specific system resources based on specific time constraints or events. As shown in the following figure, the Orchestration Server provides a built-in job scheduler that enables you or system administrators to flexibly deploy and run jobs.

Figure 5-4 *The Job Scheduler*



Provisioning Jobs

Jobs also drive provisioning for virtual machines (VMs) and physical machines, such as blade servers. Provisioning adapter jobs for various VM hypervisors are deployed and organized into appropriate job groups for management convenience.

The provisioning jobs included in the Orchestration Server are used for interacting with VM hosts and repositories for VM life cycle management and for cloning, moving VMs, and other management tasks. These jobs are called “provisioning adapters” and are members of the job group called “provisionAdapters.”

Auditing and Accounting Jobs

You can create Cloud Manager Orchestration jobs that perform reporting, auditing, and costing functions inside your data center. Your jobs can aggregate cost accounting for assigned resources and perform resource audit trails.

5.1.7 Constraint-Based Job Scheduling

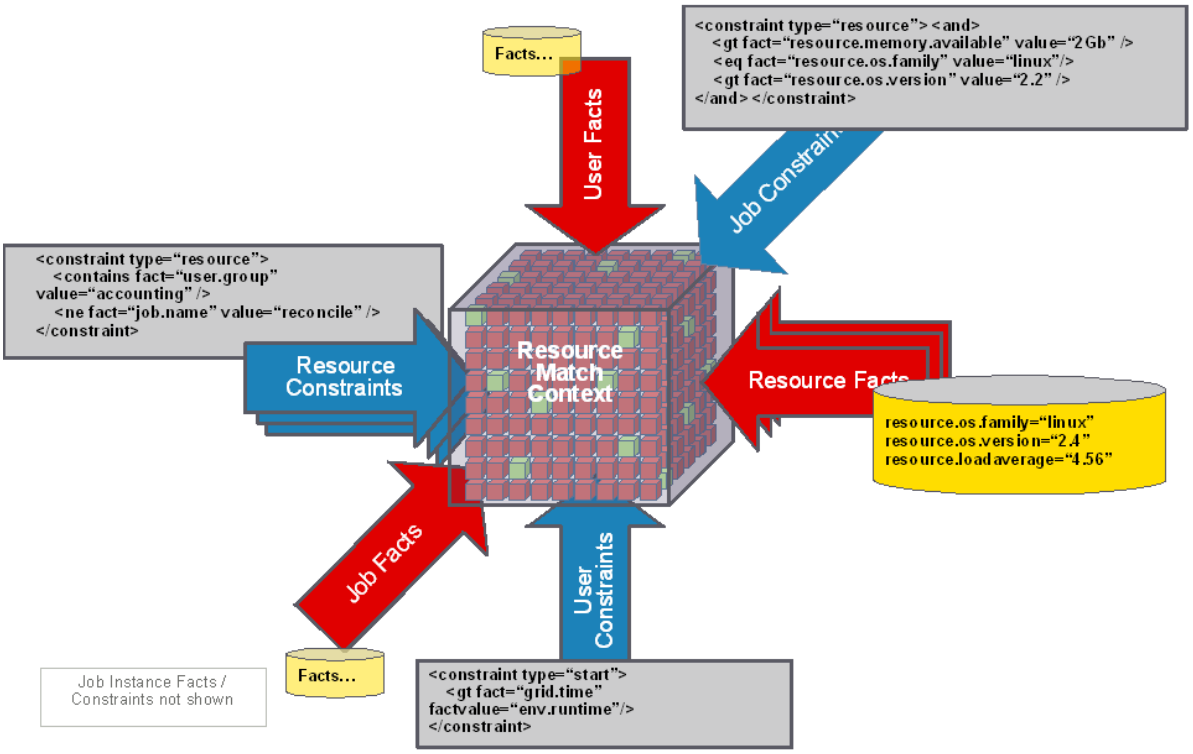
The Orchestration Server is a “broker” that can distribute jobs to every “partner” agent on the grid. Based on assigned policies, jobs have priorities and are executed based on the following contexts:

- User Constraints
- User Facts
- Job Constraints
- Job Facts
- Job Instance

- Resource User Constraints
- Resource Facts
- Groups

Each object in a job context contains the following elements:

Figure 5-5 *Constraint-Based Resource Brokering*



5.2 Understanding Orchestration Functionality

- Section 5.2.1, “How Do I Interact with the Orchestration Server?,” on page 41
- Section 5.2.2, “How Orchestration Components Communicate,” on page 43
- Section 5.2.3, “Resource Virtualization,” on page 45
- Section 5.2.4, “Policy-Based Management,” on page 45
- Section 5.2.5, “Grid Object Visualization,” on page 46
- Section 5.2.6, “Understanding Job Semantics,” on page 46
- Section 5.2.7, “Distributed Messaging and Failover,” on page 47

5.2.1 How Do I Interact with the Orchestration Server?

Orchestration administrators and users perform their activities by using their own graphical tool or command line interface tools. In general, the same functions are available in either the graphical or the command line tools. The toolset is summarized in the chart below.

Table 5-1 *Summary of the Orchestration Toolset*

Role	Tool Type	Description	Common Function
Orchestration Administrator	Graphical Interface	The Orchestration Console	◆ Stops or starts the Orchestration Server(s) ◆ Deploys jobs. ◆ Manages Group and Policy associations. ◆ Monitors jobs. ◆ Helps troubleshoot jobs/ policies. ◆ Monitors computing resource usage. ◆ Creates and manages user accounts.
	Command Line Interface	Sample command line for help: `zosadmin` *`command`* `--help`	
	Graphical Interface	The Cloud Manager Web Console or a Cloud Manager Mobile Client	◆ Discovers host servers in the Orchestration grid ◆ Discovers existing VMs ◆ Creates, edits, installs, and deletes VMs ◆ Manages VM repositories ◆ Stops, starts, pauses, or suspends VMs ◆ Migrates or moves VMs ◆ Installs the Orchestration Agent on VMs ◆ Creates and clones VM templates ◆ Provides group management of VMs, host servers, storage locations, and templates ◆ Resyncs state of VMs and hosts with Orchestration Server ◆ Provides access to VM and Host consoles ◆ Shows details of VM and host configurations ◆ Provides error log and progress views

Role	Tool Type	Description	Common Function
Orchestration User	Command Line Interface	Sample command line for help: `zos command --help`	◆ Displays deployed jobs. ◆ Displays available computing resources. ◆ Runs jobs. ◆ Monitors running jobs. ◆ Manages the user's own jobs. That is, a user can cancel, pause, restart, and change job priority.

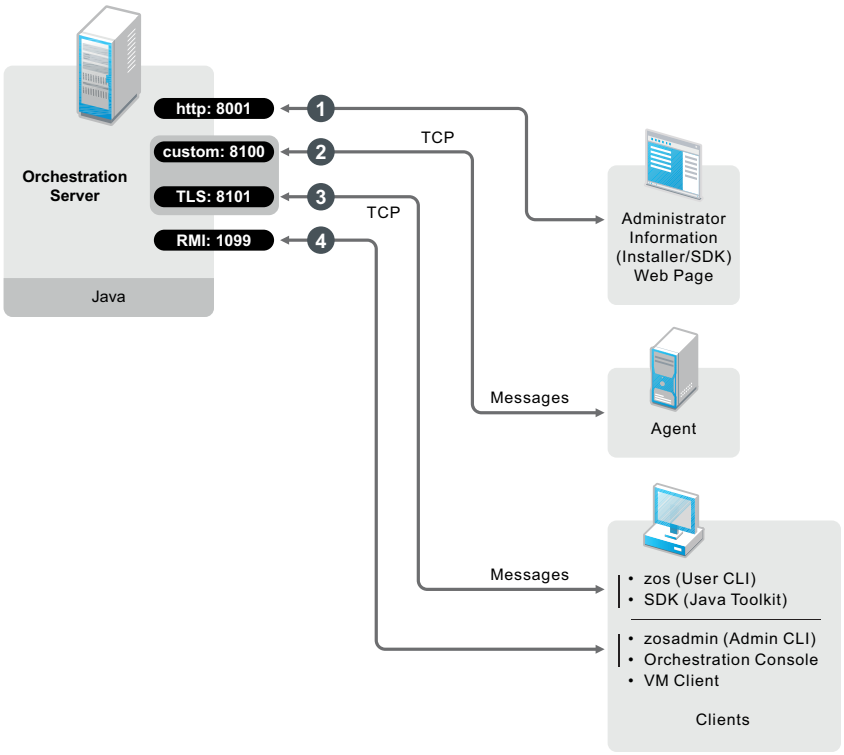
Other functions can also be performed by using either the graphical or command line tools. To help you understand how these tools can be used, you can find more information in the following sections:

- "Orchestration Server zos Command Line Tool" in the *Cloud Manager Administrator Reference*.
- "Orchestration Server zosadmin Command Line Tool"in the *Cloud Manager Administrator Reference*.

5.2.2 How Orchestration Components Communicate

The following diagram illustrates how the various components of Cloud Manager Orchestration communicate with the Orchestration Server. An explanation for each communication link follows the diagram.

Figure 5-6 *Communication Ports Used By the Orchestration Server*



1. Administrators who want more information about Cloud Manager Orchestration and a method to access or install additional clients or agents can access the Administrator Information page. To do so, open a web browser and enter the URL to the Orchestration Server, followed by the port designated for the Web Info page during installation. In a basic installation of Orchestration, the default is port 8001. The URL would therefore be entered as follows:

   `http://`*`DNS_Name_or_IP_Address`*`:8001`

2. The Orchestration Server establishes and maintains contact with an installed Orchestration Agent on a computing resource through port 8100, using a custom protocol.
3. When a user invokes the `zos` command line interface (available after Orchestration clients are installed on a machine), or when using the Java toolkit SDK, those client tools communicate with the Orchestration Server over ports 8100 and 8101.
4. When the administrator invokes the `zosadmin` command line interface (available after Orchestration clients — including the Orchestration Console — are installed on a machine), or when using the Orchestration Console, those client tools communicate with the Orchestration Server over port 1099, which uses a Java RMI (Remote Method Invocation) protocol.

NOTE: The Orchestration Server Console uses TCP port 1099 for its initial connection and then selects a port in the ephemeral port range (ports 32768-65535) for additional communications. If you have problems connecting to the Orchestration Console, ensure that these ports on the server are reachable from the client.

IMPORTANT: Currently, the RMI connection between the Orchestration Server and the JMX server where the Orchestration Console, `zos`, and `zosadmin` tools reside (see items 3 and 4, above) transmits the Orchestration password in plain text, although it is not stored on the JMX server. For this reason, we recommend that you use a secured (such as an SSL link), trusted connection between the computer running the Orchestration Server and an instance of the Orchestration Client tools.

5.2.3 Resource Virtualization

Host machines or test targets managed by the Orchestration Server form nodes on the grid. All resources are virtualized for access by maintaining a capabilities database containing extensive information (facts) for each managed resource.

This information is automatically polled and obtained from each resource periodically or when it first comes online. The extent of the resource information the system can gather is customizable and highly extensible, controlled by the jobs you create and deploy.

5.2.4 Policy-Based Management

Policies are aggregations of facts and constraints that are used to enforce quotas, job queuing, resource restrictions, permissions, and other user and resource functions. Policies can be set on all objects and are inherited, which facilitates implementation within related resources.

Facts, which might be static, dynamic or computed for complex logic, are used when jobs or test scenarios require resources in order to select a resource that exactly matches the requirements of the test, and to control the access and assignment of resources to particular jobs, users, projects, etc. through policies.This abstraction keeps the infrastructure fluid and allows for easy resource substitution.

Of course, direct named access is also possible. An example of a policy that constrains the selection of a resource for a particular job or test is shown in the sample below. Although resource constraints can be applied at the policy level, they can also be described by the job itself or even dynamically composed at runtime.

```
<policy>
  <constraint type="resource">
    <and>
      <eq fact="resource.os.family" value="Linux"/>
      <gt fact="resource.os.version" value="2.2" />
    <and>
  </constraint>
</policy>
```

An example of a policy that constrains the start of a job or test because too many tests are already in progress is shown in the following sample:

```
<policy>
  <!-- Constrains the job to limit the number of running jobs to a
defined value but exempt certain users from this limit. All jobs
that attempt to exceed the limit are queued until the running jobs
count decreases and the constraint passes. -->

<constraint type="start" reason="Too busy">
    <or>
      <lt fact="job.instances.active" value="5"/>
      <eq fact="user.name" value="canary" />
    </or>
  </constraint>
</policy>
```

5.2.5 Grid Object Visualization

One of the greatest strengths of the Cloud Manager Orchestration solution is the ability to manage and visualize the entire grid. This is performed through the Cloud Manager Orchestration Console and the Cloud Manager VM Monitoring System.

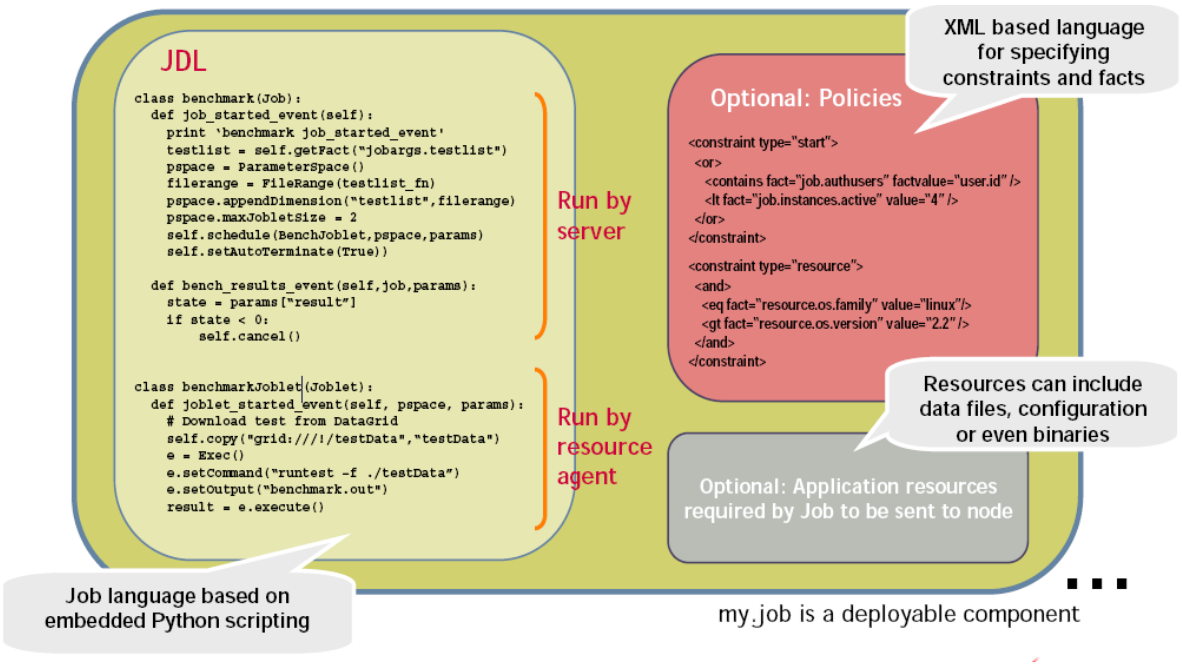
The desktop Orchestration Console is a Java application that has broad platform support and provides job, resource, and user views of activity as well as access to the historical audit database system, cost accounting, and other graphing features.

The Orchestration Console also applies policies that govern the use of shared infrastructure or simply create logical grouping of nodes on the grid. For more information about the console, see the *Cloud Manager Administrator Reference*.

5.2.6 Understanding Job Semantics

As mentioned earlier, the Orchestration Server runs jobs. A job is a container that can encapsulate several components including the Python-based logic for controlling the job life cycle (such as a test) through logic that accompanies any remote activity, task-related resources such as configuration files, binaries and any policies that should be associated with the job, as illustrated below.

Figure 5-7 *Components of a Job*



Workflows

Jobs can also invoke other jobs, creating hierarchies. Because of the communication between the job client (either a user/user client application or another job) it is easy to create complex workflows composed of discrete and separately versioned components.

When a job is executed and an instance is created, the class that extends job is run on the server and as that logic requests resources, the class(es) that extend the joblet are automatically shipped to the requested resource to manage the remote task. The communication mechanism between these distributed components manifests itself as event method calls on the corresponding piece.

5.2.7 Distributed Messaging and Failover

A job has control over all aspects of its failover semantics, which can be specified separately for conditions such as the loss of a resource, failure of an individual joblet, or joblet timeout.

The failover/health check mechanisms leverage the same communications mechanism that is available to job and joblet logic. Specifically, when a job is started and resources are employed, a message interface is established among all the components as shown in Figure 5-8 on page 48.

Optionally, a communication channel can also be kept open to the initiating client. This client communication channel can be closed and reopened later based on jobid. Messages can be sent with the command

```
sendEvent(foo_event, params, ...)
```

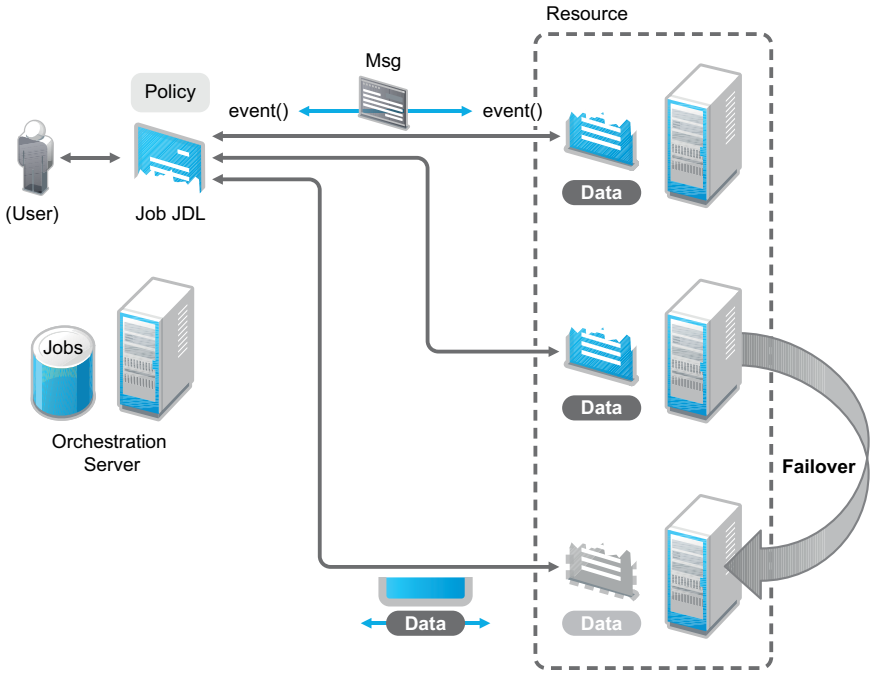
and received at the other end as a method invocation

```
def foo_event(self, params)
```

If a job allows it, a failure in any joblet causes the Orchestration Server to automatically find an alternative resource, copy over the joblet JDL code, and reestablish the communication connection. A job also can listen for such conditions simply by defining a method for one of the internally generated events, such as `def joblet_failure_event(...)`.

Such failover allows, for example, for a large set of regression tests to be run (perhaps in parallel) and for a resource to die in the middle of the tests without the test run being rendered invalid. The figure below shows how job logic is distributed and failover achieved:

Figure 5-8 *A Job in Action*



6 Server Discovery and Multicasting

The Cloud Manager Orchestration Server, the Orchestration Agent, and other Orchestration tools use IP multicast messages to locate servers and to announce when servers are started or shut down. If multicasting is not supported in your existing network environment, all Orchestration components allow a specific machine to be specified instead of using multicast discovery. Multicast support is not required to run the Orchestration Server, the Orchestration Agent, or any Orchestration tools.

This section includes the following multicast information:

- Section 6.1, “Multicast Routes,” on page 49
- Section 6.2, “Multi-homed Hosts,” on page 49
- Section 6.3, “Multiple Subnets,” on page 50
- Section 6.4, “Datagrid and Multicasting,” on page 50
- Section 6.5, “Datagrid Multicast Interface Selection,” on page 50

6.1 Multicast Routes

A common problem with multicasting, particularly on Linux, is the lack of a default route or multicast network route. Most systems are configured to have at least a “default” route, and on such systems, multicast messages use the default route like any other network traffic. Systems do not necessarily require a default route. Multicasting might not function correctly on systems that lack a default route. Attempts to send messages on such systems fail with a `Network Unreachable` message because the operating system is unable to determine the correct network interface on which to send the message.

The quick solution is to add a default route on such systems. In some environments, however, it might not make sense to add a default route. In such cases, another solution is to add a network route for the 224.0.0.0/4 block representing the multicast IP address space. On Linux, for example, issue the following command as the `root` user:

```
route add -net 224.0.0.0 netmask 240.0.0.0 dev eth0
```

This command tells the system to send all multicast datagrams to the `eth0` network card by default. Substitute `eth0` with a different interface name if applicable.

6.2 Multi-homed Hosts

A multi-homed host is a machine with more than one network interface configured. This can be anything from a Linux system being used as a network router to a laptop computer with both an active Ethernet connection and an active wireless connection. If there are two or more network interfaces active at the same time (even if only one is actually being used) the system is “multi-homed.”

Some operating systems like Linux provide only very rudimentary routing as a default part of the operating system. They rely on external routing software like “mrouted” to support full multicast routing. As a result, problems might arise in multi-homed machines because outgoing multicast messages are sent on only one interface in the absence of more sophisticated routing software. It is possible that the interface chosen by the operating system is incorrect. The Orchestration Server and

its associated tools make a best effort to ensure that discovery queries and announcements are sent on all available interfaces. It should not be necessary to run an external routing program with the current Orchestration Server.

6.3 Multiple Subnets

By default, the Orchestration Server and its associated tools are configured to allow multicast messages to pass through up to two gateways. This allows discovery to work in multi-subnet environments, provided that the network routers on your network are properly configured to perform multicast routing. Consult the vendor's documentation for information on configuring multicast routing on your network routers.

6.4 Datagrid and Multicasting

The Cloud Manager Orchestration datagrid facility provides a multicast-based file distribution service that allows large multi-gigabyte files to be simultaneously delivered to a large number of recipient machines while using far less network bandwidth than would be used by copying the file individually to each node. This service is available only in network environments that support IP multicasting. Aside from the file multicast service, all other features of the datagrid use normal unicast network operations and do not require multicast support. The routing and troubleshooting pointers provided above for network discovery also apply to datagrid multicasting. In addition, due to the potentially large bandwidth used by file transfers, you might want to limit the set of interfaces on which files are multicasted.

6.5 Datagrid Multicast Interface Selection

On multi-homed servers, the datagrid multicast service sends outbound control and data packets on all available interfaces on the system. This allows datagrid multicasting to work "out of the box" with multi-homed servers. This behavior might not be optimal if you require multicasting of files only to a subset of the available interfaces. You can instruct the datagrid to multicast only to the desired interfaces by selecting the correct interfaces from the Orchestration Server Console on the Info/Configuration tab under **Data Grid Configuration**. Restricting the set of datagrid multicast interfaces prevents large amounts of file data from being sent to uninterested subnets.

7 Cloud Manager Orchestration and LDAP Authentication

Although the Cloud Manager Orchestration Server has its own user database and authentication mechanism, it also allows integration with an existing Lightweight Directory Access Protocol (LDAP) system for authenticating user credentials.

This section includes the following information:

- Section 7.1, “What is LDAP?,” on page 51
- Section 7.2, “Understanding LDAP Structure,” on page 52

7.1 What is LDAP?

At a high level, LDAP is a protocol designed to allow quick, efficient searches of directory services. Built around Internet technologies, LDAP makes it possible to easily update and query directory services over standard TCP/IP connections, and it includes many powerful features, including security, access control, data replication and support for Unicode.

LDAP is based on the Directory Access Protocol (DAP), which was designed for communication between directory servers and clients compliant to the X.500 standard. However, DAP can be difficult to implement and use, and is not suitable for use with web applications. LDAP is a simpler, faster alternative, offering much of the same basic functionality without the performance overhead and deployment difficulties of DAP.

Because LDAP is built for a networked world, it is based on a client-server model.The system consists of one (or more) LDAP servers, which host the public directory service, and multiple clients, which connect to the server to perform queries and retrieve results. LDAP clients are built into most common address book applications, including email clients like Microsoft Outlook; however, since LDAP-compliant directories can store a diverse range of data (not just names and phone numbers), LDAP clients are also increasingly making an appearance in other applications. LDAP support is included in the Orchestration Server to allow it to integrate with existing user authentication mechanisms that are being used in many data centers.

There are many similarities between the Internet Domain Name System (DNS) model and LDAP: both are global directories that can be split across multiple hosts, both have built-in redundancy and replication features, and both include referral capabilities that make it possible to retrieve data that is not available locally from other hosts in the system.

7.2 Understanding LDAP Structure

An LDAP directory is usually structured hierarchically as a tree of nodes (the LDAP directory tree is sometimes referred to as the Directory Information Tree, or DIT). Each node represents a record, or “entry” in the LDAP database.

This section includes the following information:

- Section 7.2.1, “The Distinguished Name,” on page 52
- Section 7.2.2, “The Relative Distinguished Name,” on page 52

7.2.1 The Distinguished Name

An LDAP entry consists of numerous attribute-value pairs. It is uniquely identified by what is known as a “distinguished name” (DN).

To draw a parallel with a relational database management system (RDBMS), an LDAP entry is analogous to a record, its attributes are the fields of that record, and a DN is a primary key that uniquely identifies each record.

Consider the following example of an LDAP entry:

```
dn: mail=joe@novell.com, dc=novell, dc=com
objectclass: inetOrgPerson
cn: Joe
sn: Somebody
mail: joe@novell.com
telephoneNumber: 1 234 567 8912
```

This is an entry for a single person, Joe Somebody, who works at Novell. The components of the entry – name, email address, telephone number – are split into attribute-value pairs, with the entire record identified by a unique DN (the first line of the entry). Some of these attributes are required and some are optional, depending on the object class being used for the entry; however, the entire set of data constitutes a single entry, or node, on the LDAP directory tree.

7.2.2 The Relative Distinguished Name

Every entry in the directory tree has a “relative distinguished name” (RDN) consisting of one or more attribute-value pairs. An RDN must be unique at that level in the directory hierarchy. In the example above, for instance, the following are all valid RDNs for the entry:

```
cn=Joe
```

or

```
cn=Joe+sn=Somebody
```

or

```
cn=Joe+sn=Somebody+telephoneNumber=12345678912
```

or

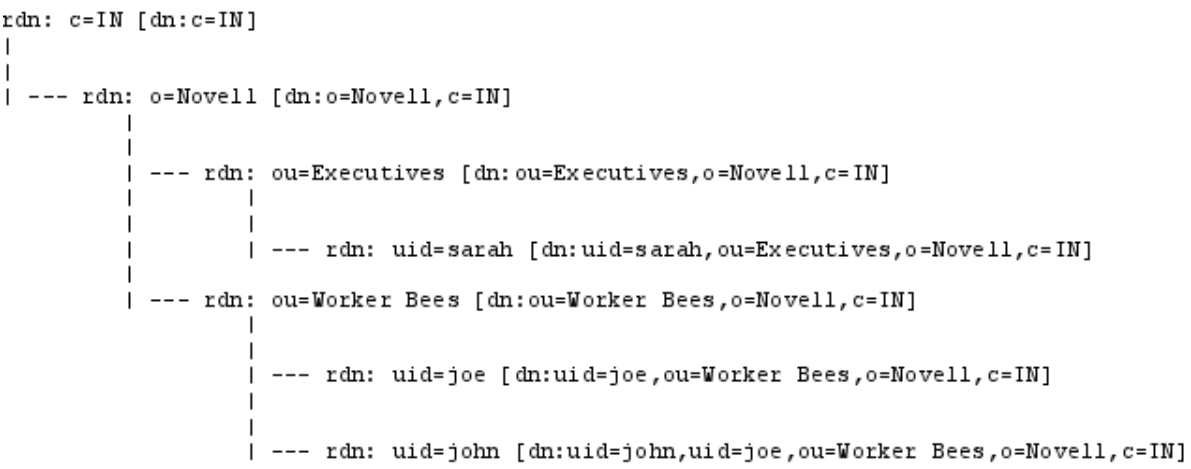
```
mail=joe@novell.com
```

There are no set rules regarding which attributes of a particular entry should be used for the RDN; the LDAP model leaves this decision to the directory designer, specifying only that the RDN of an entry must be such that it can uniquely identify that entry at that level in the DIT.

Because RDNs exist for every entry in the tree, the DN for any entry is formed by sequentially appending the RDNs of all the nodes between that entry and the root entry. In this way, you can use the DN to easily locate any node in the directory tree, regardless of its location or depth in the hierarchy.

For example, consider the following LDAP directory:

Figure 7-1 *Sample LDAP Directory*



To identify the node belonging to Joe Somebody (the DN for Joe Somebody's entry) you would add all the RDNs between that entry and the root of the tree:

```
uid=joe,ou=Worker Bees,o=Novell,c=IN
```

In a similar manner, the DN for the node belonging to Sarah would be

```
uid=sarah,ou=Executives,o=Novell,c=IN
```

while the DN for the Novell node would be

```
o=Novell,c=IN
```

Because LDAP entries are arranged in a hierarchical tree, and because each node on the tree can be uniquely identified by a DN, the LDAP model lends itself to sophisticated queries and powerful search filters.

8 Cloud Manager Orchestration Security

This section explains various security issues related to Cloud Manager Orchestration:

- Section 8.1, “User and Administrator Password Hashing Methods,” on page 55
- Section 8.2, “User and Agent Password Authentication,” on page 55
- Section 8.3, “Password Protection,” on page 56
- Section 8.4, “TLS Encryption,” on page 56
- Section 8.5, “Security for Administrative Services,” on page 58

8.1 User and Administrator Password Hashing Methods

All passwords stored in the Orchestration Server are hashed using Secure Hash Algorithm-1 (SHA-1). However, user passwords are no longer hashed when sent from the client to the server. Instead, the plain text password entered by the user is sent over an encrypted authentication connection to the server to obtain a unique per-session credential issued by the server. This allows the server to “plug in” to alternative user directories such as Active Directory or OpenLDAP. Agent credentials are still stored, singly hashed, on the disk on the agent machine. The first pass hashing prevents “user friendly” passwords entered by administrators from being compromised by storing them on the agent machines. The server’s password database (for agents and for users not using an alternative user directory) stores all passwords in a double-hashed form to prevent a stolen password database from being used to obtain passwords.

WARNING: The `zosadmin` command line and the Orchestration Console do not use SSL encryption, nor do they support TLS/SSL, so they should only be used over a secure network.

All agent and client connections support TLS encryption. This includes the zos command line and the Orchestration Agent.

8.2 User and Agent Password Authentication

The Orchestration Server stores all user and agent passwords in its data store as double-hashed strings. User clients such as the zos command send the plain text password over a TLS encrypted authentication connection to obtain a randomly generated per-session credential issued by the server. This session credential is retained by the client, either in memory or in a temporary disk file for the duration of the session.

It is not possible to obtain the user’s password from the session credential, however. It should be protected to prevent unauthorized users from taking over the session. Agents send a singly hashed password as their login credential, which is in turn hashed once more on the server to authenticate new agent connections. Upon authentication, agents receive the same type of session credential as user clients.

Singly-hashed password strings are used as a special case for agents, because agents typically must store their plain text credentials to disk to allow the agents to start up on host or VM reboot. The use of a once hashed version of the password on the agent prevents administrators from compromising “user friendly” text passwords by storing them unhashed on agents. The use of single hashing on the agents and double hashing on the server database prevents stolen credential data from being used to obtain actual user or administrator-entered passwords

8.3 Password Protection

You should take measures to protect the passwords and credentials on both the Cloud Manager Orchestration Server and the Cloud Manager Orchestration Agents by ensuring that only the user account of the Orchestration Server (currently `root` or `Administrator`, by default) has access to the `/store` and `/tls` directories on the server, so that general users are prevented from obtaining the password. On agents, allow only the agent users (normally `root` or `Administrator`) to have access to the `agent.properties` file, which contains the agent’s authentication credential.

Currently, the Orchestration Server restricts file access on the server, but we recommend that you disallow shell accounts on server machines for general users as a precaution.

For users, none of the NetIQ-provided client utilities stores the user-entered password to disk in either plain text or hashed form. However, temporary once-per-session credentials are stored to the disk in the users `$HOME/.novell/zos/client` directory. Theft of this session credential could allow someone else to take over that user session, but not to steal the user’s password. Users can protect their logged-in session by making sure the permissions either on their home directory or on the `~/.novell/zos/client` directory are set to forbid both read and write access by other users.

Orchestration Agents use the same authentication protocol and password hashing as users (agent passwords are stored to disk in hashed form, not plain text) with the exception that agent passwords are not salted, allowing agents to be renamed by the server. Because agent passwords are not salted, we recommend that you generate and use random non-mnemonic strings for agent passwords.

Administrators can enhance security when configuring new agents by setting the `zos.agent.password` property to the asterisk character (*). This causes the agent to automatically generate a new random credential not based on any easily guessed plain text word. When the new agent is “accepted” by the administrator, the newly generated credential is stored by the server. This is the default behavior when the Orchestration Agent is first installed.

n addition, the `zos.agent.password` property can be set to a plain text password in `agent.properties`. If this is done, the agent automatically replaces the plain text password with the hashed version when it next starts. This allows administrators to more easily set up an initial password for agents.

8.4 TLS Encryption

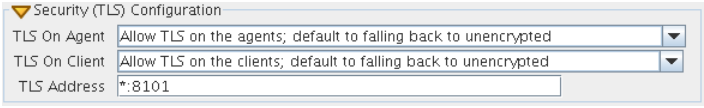
- Section 8.4.1, “Setting TLS Options,” on page 56
- Section 8.4.2, “Updating the TLS Server Certificate,” on page 57

8.4.1 Setting TLS Options

Cloud Manager Orchestration uses Transport Layer Security (TLS) to provide encryption for both user and agent connections. The Orchestration Agent, the Orchestration Console, and other Orchestration clients use TLS to initiate their connections to the Orchestration Server, and then the

server specifies whether to “fall back” to plain text or continue the session fully encrypted. Although you can manually configure the agent and clients to either always require TLS encryption or to fully disable TLS encryption, we recommend that you leave the agents and clients in their default configuration, and then use the Orchestration Console on the server to specify the default behavior. This is the purpose of the TLS Options section on the main server tab of the Orchestration Console.

Figure 8-1 *TLS Options in the Cloud Manager Orchestration Server Console*



Here, there are 4 levels that you can set separately for both agent connections and user/client connections:

- **Forbid TLS for (agents/clients):** This option is to fully disable and prohibit TLS encryption altogether. This is the least secure option and is therefore usually not the desirable choice, but it could be required in countries that restrict encryption or in low security environments where performance is more critical than security.
- **Allow TLS on the (agents/clients); default to falling back to unencrypted:** This option (the factory default for both agents and clients) is to allow TLS encryption if the agent or client explicitly requests it, but to default to falling back to plain text after authentication.

  NOTE: Authentication always occurs over SSL, regardless of settings.

- **Allow TLS on the (agents/clients); default to TLS encrypted if not configured encrypted:**

  This option is similar to the second option. Agents/clients may specify whether or not to use TLS, but if they use the default of “server specified,” the server defaults to using TLS.
- **Make TLS mandatory on the (agents/clients):** This option is the most secure, locked down option. It requires TLS at all times, and fails connections if the agent or the client tries to specify plain text.

8.4.2 Updating the TLS Server Certificate

Understanding Transport Layer Security (TLS) encryption is particularly important if you reinstall the server and have an old server certificate in either your agent or client user profile similar to `ssh` shared keys. If you have an old certificate, you need to either manually replace it or delete it and allow the client or agent to download the new one from the server using one of the following procedures:

- **For the Agent:** The TLS certificate is in `<agentdir>/tls/server.pem`. Deleting this certificate will cause the agent, by default, to log a minor warning message and download a new one the next time it tries to connect to the server. This is technically not secure, since the server could be an impersonator. If security is required for this small window of time, then the real server’s `<serverdir>/<instancedir>/tls/cert.pem` can be copied to the above server.pem file.
- **For the Client:** The easiest way to update the certificate from the command line tools is to simply answer “yes” both times when prompted about the out-of date certificate. This is, again, not 100% secure, but is suitable for most situations. For absolute security, hand copy the server’s cert.pem (see above) to `~/.novell/zos/client/tls/<serverIPAddr:Port>.pem`.
- **For Java SDK clients:** Follow the manual copy technique above to replace the certificate. If the local network is fairly trustworthy, you can also delete the above `~/.novell/.../*.pem` files, which will cause the client to auto-download a new certificate.

8.5 Security for Administrative Services

The Orchestration Console and the zosadmin command line tool are clients to the MBean and RMI servers. Cloud Manager Orchestration does not provide encryption for these administrative services, so you should be careful to use them only in a secure environment.

When the user logs in using either `zosadmin login` or the Orchestration Console, the user's password is sent to the server, and then the server issues a per-session credential to be used for further operations. The user's clear text password is never stored to disk; however, it is currently sent "over the wire" in plain text form. For this reason, the administrative clients should only be used in a secure, trusted environment.

The zosadmin client stores the session credential obtained from a `zosadmin login` request in a temporary file for use by subsequent operations. This credential cannot be used to obtain the user's password, but it could be used to take over the user's current session until it times out or expires. For this reason, the files in the user's `.novell/zoc/` directory should be configured to disallow access by other users.

9 User Concepts

Users must have a Cloud Manager account in order to perform activities within the Cloud Manager system. The following sections provide information you should understand as you create and manage user accounts and user groups:

- Section 9.1, “Organization Scope versus System Scope,” on page 59
- Section 9.2, “Cloud Manager Roles,” on page 59
- Section 9.3, “Cloud Manager User Groups versus LDAP User Groups,” on page 68
- Section 9.4, “Roles That Can Create User Accounts and User Groups,” on page 69

9.1 Organization Scope versus System Scope

When you create a Cloud Manager account for a user, you can give the user an Organization scope or a System scope. The scope determines what roles can be assigned to the user.

A user with the Organization scope (referred to as an *Organization user* or *Organization member*) is assigned membership in a specific organization and can hold Organization roles. These roles provide rights to perform activities within the user’s organization, such as creating business groups, allocating organization resources to the business groups, and deploying business services.

A user with the System scope (referred to as a *System user*) is not assigned membership in an organization and can hold System roles. These roles provide rights to perform system-level activities, such as configuring the Cloud Manager system, creating zones, creating organizations, creating groups of resources for use by organizations, and monitoring zone and organization resource capacity. In addition, System users can be assigned Organization roles for any organization in the system.

Organization and System scopes also apply to user groups, meaning that there are *System user groups* and *Organization user groups*. Users groups are discussed in “Cloud Manager User Groups versus LDAP User Groups” on page 68.

Both System roles and Organization roles are discussed in detail in “Cloud Manager Roles” on page 59.

9.2 Cloud Manager Roles

A user must have one or more Cloud Manager roles in order to do anything in Cloud Manager. There are nine Cloud Manager roles. All nine roles can be given to System users, while only four of the roles can be given to Organization members. Each role carries its own set of rights and responsibilities for the Cloud Manager system or for an organization within the system.

- Section 9.2.1, “Descriptions,” on page 60
- Section 9.2.2, “Rights,” on page 61
- Section 9.2.3, “Rights by Organization Role,” on page 67

9.2.1 Descriptions

Table 9-1 defines the 11 roles that you can hold as a System user. Each role carries its own set of responsibilities for the Cloud Manager system or for an organization within the system.

Table 9-1 *System User Roles*

Key	System User Role	Rights
CA	Cloud Administrator	Has full rights to the Cloud Manager system. Can perform all tasks in the system.
ZA	Zone Administrator	Has rights to manage the resources for one or more assigned zones. Only System users can be Zone Administrators.
CM	Catalog Manager	Has rights to create, modify, and delete workload templates. Workload templates must be assigned to organizations by the Cloud Administrator.
A	Approver	Has rights to approve or deny a business service request based on available zone and organization resource capacity.
BA	Build Administrator	Has rights to complete pre-build and post-build configuration for workloads in requested business services.
BGV	Business Group Viewer	Has rights to view business services for a business group.
BSO	Business Service Owner	Has rights to create, modify, and delete business services for an organization or for specific business groups within an organization.
OM	Organization Manager	Has rights to manage users, roles, resource assignments, and business services within an assigned organization. Both System users and Organization users can be Organization Managers. System users can be assigned as Organization Managers in multiple organizations. Organization users can be assigned as Organization Managers only in their own organization.
SM	Sales Manager	Has rights to view users, roles, resource assignments, and business services within an assigned organization. Both System users and Organization users can be Sales Managers. System users can be assigned as Sales Managers in multiple organizations. Organization users can be assigned as Sales Managers only in their own organization. Can create quotes for an organization in Cloud Marketplace.
S	Sponsor	Has rights to approve or deny a business service request based on financial reasons.
SV	System Viewer	Has rights to view anything in the Cloud Manager system, but cannot perform any tasks. Only System users can be System Viewers.

Table 9-2 defines the 5 roles that you can hold as Organization roles. Both Organization users and System users can be assigned these roles.

Table 9-2 *Organization Roles*

Key	Organization Role	Rights
BGV	Business Group Viewer	Has rights to view business services for a business group.
BSO	Business Service Owner	Has rights to create, modify, and delete business services for an organization or for specific business groups within an organization.
OM	Organization Manager	Has rights to manage users, role assignments, resource assignments, and business services within an assigned organization. System users can be assigned as Organization Managers in multiple organizations. Organization users can be assigned as Organization Managers only in their own organization.
SM	Sales Manager	Has rights to view users, roles, resource assignments, and business services within an assigned organization. Both System users and Organization users can be Sales Managers. System users can be assigned as Sales Managers in multiple organizations. Organization users can be assigned as Sales Managers only in their own organization. Can create quotes for an organization in Cloud Marketplace.
S	Sponsor	Has rights to approve or deny a business service request based on financial reasons.

9.2.2 Rights

Table 9-3 *System Management*

System Management Rights											
Roles (See Table 9-1)	**CA**	**ZA**	**CM**	**A**	**BA**	**BGV**	**BSO**	**OM**	**SM**	**S**	**SV**
USERS											
Create System user accounts and user groups, either manually or by importing from an LDAP directory	✔										
Modify System user and user group properties (email, phone number, and so forth)	✔										
ROLES											
Assign Cloud Administrator role	✔										
Assign Zone Administrator role	✔										

System Management Rights											
Roles (See Table 9-1)	**CA**	**ZA**	**CM**	**A**	**BA**	**BGV**	**BSO**	**OM**	**SM**	**S**	**SV**
Assign Catalog Manager role	✔										
Assign Build Administrator role	✔										
Assign Approver role	✔										
Assign System Viewer role	✔										
CAPACITY AND REPORTS											
View resource capacity for system	✔										✔
Generate resource capacity reports for system	✔										

Table 9-4 *Zone Management*

Zone Management Rights											
Roles (See Table 9-1)	**CA**	**ZA**	**CM**	**A**	**BA**	**BGV**	**BSO**	**OM**	**SM**	**S**	**SV**
Assign Zone Administrator role	✔										
Create, modify, retire, and delete zones	✔										
Create, modify, and delete resource groups for zones	✔	✔									
View resource capacity for zones	✔	✔							✔		✔
Generate resource capacity reports for zones	✔	✔									

Table 9-5 *Organization Management*

Organization Management Rights											
Roles (See Table 9-1)	**CA**	**ZA**	**CM**	**A**	**BA**	**BGV**	**BSO**	**OM**	**SM**	**S**	**SV**
USERS											
Create Organization user accounts and user groups, either manually or by importing from an LDAP directory	✔							✔			
Modify Organization user and user group properties (email, phone number, etc.)	✔							✔			
ROLES											
Assign Organization Manager role	✔							✔			
Assign Sponsor role	✔							✔			
Assign Business Group View role	✔							✔			
Assign Business Service Owner role	✔							✔			
Assign Sales Manager role	✔							✔			
ORGANIZATIONS											
Create modify, and delete organizations	✔										
Assign a cost factor (discount or markup) to organizations*	✔							✔			
Assign resource groups to organizations	✔										
Assign workload templates to organizations	✔										
Assign networks to organizations	✔										
BUSINESS GROUPS											
Create modify, and delete business groups	✔							✔			
Assign workload templates from an organization to its business groups	✔							✔			

Organization Management Rights											
Roles (See Table 9-1)	**CA**	**ZA**	**CM**	**A**	**BA**	**BGV**	**BSO**	**OM**	**SM**	**S**	**SV**
Assign resource groups from an organization to its business groups	✔							✔			
Assign networks from an organization to its business groups	✔							✔			
View information and business services for business groups									✔		✔
CAPACITY AND REPORTS											
View resource capacity for organizations	✔					✔		✔			✔
Generate resource capacity reports for organizations	✔							✔			

***** Applies only to an Organization Manager who is a System user. An Organization Manager who is a member of the organization cannot change the cost factor for the organization.

Table 9-6 *Resource Management*

Resource Management Rights											
Roles (See Table 9-1)	**CA**	**ZA**	**CM**	**A**	**BA**	**BGV**	**BSO**	**OM**	**SM**	**S**	**SV**
RESOURCE GROUPS											
Create, modify, and delete resource groups	✔	✔									
Assign resource groups to organizations	✔										
Assign resource groups to an organization's business groups	✔							✔			
SERVICE LEVELS											
Create modify, retire, and delete service levels	✔										
Create modify, and delete service level objectives	✔										

Resource Management Rights											
Roles (See Table 9-1)	**CA**	**ZA**	**CM**	**A**	**BA**	**BGV**	**BSO**	**OM**	**SM**	**S**	**SV**
Assign resource costs to service levels	✔										
Assign service levels to resource groups	✔										
CAPACITY AND REPORTS											
View resource capacity for organizations	✔							✔	✔		✔
View resource capacity for zones	✔	✔									✔
View resource capacity for the system	✔										✔
Generate resource capacity reports for organizations	✔							✔			
Generate resource capacity reports for zones	✔	✔									
Generate resource capacity reports for the system	✔										

Table 9-7 *Catalog Management*

Catalog Management Rights											
Roles (See Table 9-1)	**CA**	**ZA**	**CM**	**A**	**BA**	**BGV**	**BSO**	**OM**	**SM**	**S**	**SV**
Assign Catalog Managers	✔										
Create, modify, retire, and delete workload templates	✔		✔								
Assign workload templates to organizations	✔										

Table 9-8 *Business Service Management*

Business Service Management Rights											
Roles (See Table 9-1)	**CA**	**ZA**	**CM**	**A**	**BA**	**BGV**	**BSO**	**OM**	**SM**	**S**	**SV**
Import existing VMs as business services	✔										
Request new business services	✔						✔	✔			
Request changes to existing business services	✔						✔	✔			
Provide Administrator approval or rejection of business service requests (new and change)	✔			✔							
Provide Sponsor approval or rejection of business service requests (new and change)	✔									✔	
Complete pre- and post-build workload configuration tasks for business service requests (new and change)	✔				✔						
Delegate business service ownership to other users	✔							✔			
Lifecycle business service workloads	✔						✔	✔			
Remotely access business service workloads	✔						✔	✔			
View business services						✔		✔	✔		✔
Delete business services	✔						✔	✔			

9.2.3 Rights by Organization Role

Table 9-9 *Organization Management*

Organization Management					
Roles (See Table 9-2)	**BGV**	**BSO**	**OM**	**SM**	**S**
USERS					
Create Organization user accounts and user groups, either manually or by importing from an LDAP directory			✔		
Modify Organization user and user group properties (email, phone number, etc.)			✔		
ROLES					
Assign Organization Manager role			✔		
Assign Sponsor role			✔		
Assign Business Group View role			✔		
Assign Business Service Owner role			✔		
BUSINESS GROUPS					
Create modify, and delete business groups			✔		
Assign workload templates from an organization to its business groups			✔		
Assign resource groups from an organization to its business groups			✔		
Assign networks from an organization to its business groups			✔		
View information and business services for a business group	✔			✔	
CAPACITY & REPORTS					
View resource capacity for organization			✔	✔	
Generate resource capacity reports for organization			✔		

Table 9-10 Resource Management

Resource Management					
Roles (See Table 9-2)	**BGV**	**BSO**	**OM**	**SM**	**S**
RESOURCE GROUPS					
Assign resource groups to an organization's business groups			✔		
CAPACITY AND REPORTS					
View resource capacity for organization			✔	✔	
Generate resource capacity reports for organization			✔		

Table 9-11 Business Service Management

Business Service Management					
Roles (See Table 9-2)	**BGV**	**BSO**	**OM**	**SM**	**S**
Request new business services		✔	✔		
Request changes to existing business services		✔	✔		
Provide Sponsor approval or rejection of business service requests (new and change)					✔
Delegate business service ownership to other users			✔		
Cycle (start, suspend, stop) business service workloads		✔	✔		
Remotely access business service workloads		✔	✔		
View business services	✔			✔	
Delete business services		✔	✔		✔
Generate reports for business service costs	✔	✔	✔		✔

9.3 Cloud Manager User Groups versus LDAP User Groups

Rather than assign roles to individual users, you can create user groups and assign roles to the user groups. Users who are added to a group inherit the group's roles.

There are two types of user groups:

- **Cloud Manager:** These groups are created in Cloud Manager. The group's membership is maintained in Cloud Manager. You can add both users and other groups (including LDAP user groups) to the group.

- **LDAP:** These groups are imported from your LDAP authentication source. The group's membership is maintained in the LDAP source. You cannot add users or other groups to the group in Cloud Manager.

9.4 Roles That Can Create User Accounts and User Groups

The following roles have rights to create users and user groups. This includes manually entering information to create users or groups and importing users or groups from the LDAP authentication source.

- **Cloud Administrator:** Can create both System and Organization users and groups.
- **Organization Manager:** Can create Organization users and groups for assigned organizations. For example, a System user who is an Organization Manager for multiple organizations can create users in each of the assigned organizations. An Organization user who is an Organization Manager for his or her organization can create users only for that organization.

10 Workload Template Concepts

Workload templates are used to create the workloads for business services. The following sections provide information you should understand as you create and manage workload templates:

- Section 10.1, “Workload Template Components,” on page 71
- Section 10.2, “Pre-Populated Template Settings,” on page 72
- Section 10.3, “Workload Template Changes,” on page 72
- Section 10.4, “VM Template Changes,” on page 72
- Section 10.5, “Workload Template Retirements,” on page 72
- Section 10.6, “Workload Template Deletions,” on page 73
- Section 10.7, “Catalog Manager Role,” on page 73

10.1 Workload Template Components

A workload template consists of the following:

- **Template name:** This is the display name that Business Service Owners see when selecting from the list of workload templates. You should use a naming scheme that makes sense to the users who will consume the template.
- **Costs:** You can add one-time setup costs and monthly software license costs to the template. The costs are applied to all workloads created from the template.
- **VM template:** This is the VM template that is used to create the workloads. It is located in one of the Cloud Manager Orchestration repositories.
- **Custom VM template settings:** The VM template includes default resource settings for CPUs, memory, networks (NICs), and disk space. You can increase or decrease the settings to customize the workload template. For example, if the VM template is configured with one CPU, you can increase the CPU setting to two CPUs. When a workload is created from the workload template, the resulting VM is configured with two CPUs rather than one.
- **Windows or Linux settings:** If the VM template’s operating system is Windows, the template includes Windows settings. If the operating system is Linux, the template includes Linux settings. You can pre-populate any settings that you want to be the same for all workloads created from the template. For example, if you want all workloads to use the same password for the local Administrator account, you could pre-populate that setting. See the next section, “Pre-Populated Template Settings” on page 72.
- **Associations:** (Optional) After they are created, you can associate the workload template with one or more business groups or organizations.

10.2 Pre-Populated Template Settings

When a Windows-based workload is created from a template, certain Windows settings, such as the computer name, the domain or workgroup, and the Windows product key, must be supplied. Likewise, when a Linux-based workload is created from a template, certain Linux settings, such as the host name, must be supplied.

If there are settings that will be the same for all workloads created from a template, you can pre-populate those settings. For example, if you want all Windows workloads to use the same password for the local Administrator account, you can provide that password in the template.

Any settings that you do not pre-populate must be filled in when requesting a new business service (by the Business Service Owner) or when performing the pre-build configuration (by a Cloud Administrator or a Build Administrator).

10.3 Workload Template Changes

You can change workload template settings at any time, even if the template has been used to create workloads. The only restriction is that you cannot specify a different VM template if the workload template is in use by a requested or deployed workload.

Changing a workload template has no immediate effect on deployed workloads. However, if a change is requested for a deployed workload, the workload settings are validated against the new workload template settings. This might require the Business Service Owner to change settings that he or she did not plan to change. For example, suppose that you create a workload template that allocates 4 CPUs. A Business Service Owner creates a workload (with 4 CPUs) from the workload template. You then change the workload template's CPU allocation from 4 to 2. After the change, the Business Service Owner requests a change to the workload's number of disks. When creating the change request, the Business Service Owner must also change the CPUs from 4 to 2 because 4 CPUs are no longer supported by the new workload template.

10.4 VM Template Changes

VM templates are managed (created, changed, and deleted) through the Cloud Manager Orchestration Server console. If you change a VM template that is referenced by a workload template, the change is not recognized for the workload template unless you open the workload template in Edit mode and then save it again. You do not need to change any workload template settings, you only need to save the template again.

10.5 Workload Template Retirements

You can retire a workload template in order to prevent the template from being available for creating new workloads. You can import retired workload templates or make change requests for them. New workloads based on retired templates will fail to provision. You can select this option when you create a workload template in order to hide it from being used at this time.

10.6 Workload Template Deletions

You can delete a workload template unless it is currently being used to build a workload. As soon as the workload is built and deployed, you can delete the workload template.

Deleting a workload template has no effect on deployed workloads, even if the Business Service Owner of one of the workloads requests a change to it.

10.7 Catalog Manager Role

As a Cloud Administrator, you have full rights to the Cloud Manager system. This includes creating, editing, and deleting workload templates. However, you might decide that you want the person who creates your VM templates on the Cloud Manager Orchestration Server or in the hypervisor tools to also be the person who creates your Cloud Manager workload templates.

To facilitate the delegation of workload template responsibilities, Cloud Manager includes a System role called Catalog Manager. A Catalog Manager has rights only to create, modify, and delete workload templates. The Catalog Manager can’t see anything else (organizations, business services, resources, reports, and so forth) in the Cloud Manager console. Because of this, you (or other Cloud Administrators) must assign workload templates to organizations; the Catalog Manager cannot make template assignments.

11 Resource Concepts

This section includes information about advanced concepts you need to comprehend for Cloud Manager resources:

- Section 11.1, “Resource Group Concepts,” on page 75
- Section 11.2, “Advanced Network Configuration Concepts,” on page 76

11.1 Resource Group Concepts

A resource group defines a set of VM hosts that an organization can use for its business services. In addition to the VM hosts, the resource group includes one or more service levels that define the cost of the host resources (vCPUs, memory, storage, and networks) and the service objectives (availability, support response time, and so forth).

- Section 11.1.1, “VM Host Recommendations,” on page 75
- Section 11.1.2, “Shared and Dedicated Resource Groups,” on page 75
- Section 11.1.3, “Service Levels,” on page 75
- Section 11.1.4, “Resource Group Examples,” on page 76

11.1.1 VM Host Recommendations

All VM hosts that you include in a resource group should be identical in terms of hypervisor technology, operating system version, network configuration, storage repository configuration, and hardware capabilities. This ensures a consistent environment for business services regardless of the host. It also ensures that the resource group’s service levels apply to all hosts.

11.1.2 Shared and Dedicated Resource Groups

A resource group can be shared among multiple organizations, which means that each organization’s business services utilize the same resources, or a resource group can be assigned to only one organization, in which case only that organization’s business services consume the resources.

11.1.3 Service Levels

A resource group identifies a collection of VM hosts to which workloads can be deployed. However, a resource group does not includes any costs associated with running workloads on the hosts. A resource group also does not include any service objectives for the workloads (such as host availability or support response time). The resource costs and service objectives are applied to resource groups through the use of service levels.

A service level defines the monthly cost for each type of host resource (vCPUs, memory, storage, and networks). For example, you might set the cost of one vCPU at $25 per month. If a workload requires two vCPUs, $50 is added to the monthly cost of the workload.

A service level can also include service objectives. Objectives typically define measurable behaviors such as host availability (uptime) or support response time and have a cost associated with them. Any service objective costs are added to the monthly cost of a workload that is deployed in the resource group.

A service level can be assigned to multiple resource groups. For example, two identical resource groups might require the same service level.

Multiple service levels can also be assigned to a single resource group. For example, two service levels might have the same host resource costs but different service support levels - the first with 24x7x365 support and the second with 12x5x365 support. The user, when requesting a business service, could select the service level with the desired support level.

11.1.4 Resource Group Examples

As an example, you might create a Business Critical resource group that consists of high-performance hosts intended for mission critical applications and services. You assign the resource group a Platinum service level with costs that reflect the more expensive hardware and service contract. Any business service that is provisioned to the resource group's hosts automatically inherits the resource and service costs.

Or, you might create a Lab resource group that consists of standard-performance hosts intended for software testing. You assign the resource group a Bronze service level with costs that reflect the less expensive hardware and service contract.

11.2 Advanced Network Configuration Concepts

Although much of network configuration is readily understandable because of the network configuration interface in the Web Console (see "Editing Network Configurations" in the *Cloud Manager Procedures Guide*), this section includes information about more advanced network configuration concepts.

- Section 11.2.1, "Understanding IP Address Discovery for Native IPAM," on page 76

11.2.1 Understanding IP Address Discovery for Native IPAM

To create its own IP address pool, Cloud Manager checks all known or existing workloads for NICs that are on the IP pool's network. It also checks for all NICs that are on the network and within the range of the pool. Cloud Manager then synchronizes the NICs with the IP pool, reserving the addresses in the pool and storing IP Address Manager (IPAM) information on the NIC so that it can be used for change requests and deletes.

In an extended IP Address discovery, Cloud Manager surveys the Orchestration zone that provided the network, looking for unassigned (that is, unimported) VMs on that zone that have NICs on that network and falling within the range of the IP pool. For any VM that matches these criteria, Cloud Manager "reserves" the address in the IP pool as an "External_VM." This designation identifies the VM as unavailable for assignment to another workload.

If you want to continue to add VMs to the zone, synchronizing their discovery with the Cloud Manager Application Server, you can run the following command at the `karaf` console to perform another instance of extended IP address discovery:

```
cm:ipam-native-start-extended-discovery <-n networkId> <-p poolId>
```

Things to know about this command:

- The `networkId` and `poolId` parameters are mutually exclusive; you need to use only one of them.
- The format of the networkID parameter is <*`zoneId`*>`:Network:`<*`networkName`*>. You can find this value in the Cloud Manager Web Console: **Resources** > *Networks* > *network_name* > *Edit* > *General* (see upper left corner under the *Network* label).
- The `poolId` value is not visible in the Web Console interface. If you want to use this parameter in the karaf command, you could find the `poolId` value of a given network by using this command at the karaf console:

  ```
  cm:ipam-native-list-pools -l
  ```

  This command yields a list of all of the IP address pools by ID, along with other information such as start and end address, last address allocated, host name pattern, and Network ID.
- The command starts an Application Server job called `IPPoolAddressExtendenedIPDiscoveryCMOSJob`. To determine if the job has run, see the Application Server activity log located on the server at `/opt/netiq/cloudmanager/logs/cloudmanager_server.log`.

12 Task Concepts

Business services can be created, changed, and deleted. Each of these actions results in a *creation* request, a *change* request, or a *deletion* request. Each request must go through a workflow process in order for the request to be approved (or denied) and the resulting business service workloads to be configured if necessary.

During the workflow process for a request, tasks are generated. These tasks must be completed for the business service to be created, changed, or deleted. The following sections provide information you should understand in order to successfully manage the tasks generated for you and other roles in your system:

- Section 12.1, “Types of Tasks,” on page 79
- Section 12.2, “Task Order in the Workflow Process,” on page 80
- Section 12.3, “Task Assignments and Owners,” on page 80

12.1 Types of Tasks

There are two types of tasks generated during the workflow process:

- **Approval Tasks:** Approval tasks require the task owner to either approve or deny the business service request. There are two approval tasks generated during the workflow process: an Administrator approval task and a Sponsor approval task.

  The Administrator approval task is generated for an administrator (Approver or Cloud Administrator) to provide IT approval. If the administrator approves the request, a Sponsor approval task is then generated for the Sponsor to provide financial approval.
- **Configuration Tasks:** Configuration tasks require the task owner to complete the configuration of business service workloads. There are two configuration tasks generated during the workflow process: a pre-build configuration task and a post-build configuration task.

  The pre-build configuration task occurs before the requested business service’s workloads are built, and a post-build configuration task occurs after the workloads are built. The tasks are generated for Build Administrators and Cloud Administrators.
- **Trigger Tasks:** Trigger tasks require user input before performing an action, for example, a user can choose to immediately reboot a virtual machine as part of a change request, which could potentially lose data. Alternatively, a trigger task can be scheduled, allowing the user to delay a task to complete at a later time. For example, a user could schedule a VM reboot for midnight on a Saturday.

  When a change request is made for a business service, the reboot trigger task occurs after the pre-build configuration task, and before the changes are applied to the workloads in the system build step. The trigger task is generated for the Business Service Owner and for Build Administrators and Cloud Administrators.

12.2 Task Order in the Workflow Process

During the workflow process, tasks are generated sequentially as needed so that the process flows correctly. The following table shows the workflow process and where the approval and configuration tasks occur in the process. The table also indicates the roles that can perform each task.

Workflow Process	Performed by
 Create a request	Business Service Owner
	Organization Manager
	Cloud Administrator
 Administrator approval task	Approver
	Cloud Administrator
 Sponsor approval task	Sponsor
	Cloud Administrator
 Pre-build configuration task	Build Administrator
	Cloud Administrator
 Reboot trigger task (at Change Request)	Business Service Owner
	Organization Manager
	Cloud Administrator
 Build the workloads	System
 Post-build configuration tasks	Build Administrator
	Cloud Administrator
 Deploy the business service	System

The workflow process shown above is for requests for new or changed business services. Requests for deleted business services go through the two approval stage

s only. No configuration tasks are required and the business service is deleted after the Sponsor approval.

12.3 Task Assignments and Owners

Tasks are assigned to roles and not specific individuals. For example, the Sponsor approval task is assigned to all users who have the Sponsor role for the organization or business group that requested the business service. Likewise, the configuration tasks are assigned to all users who are Build Administrators for the organization or the business group that requested the business service. The Reboot trigger task is assigned to users who are Business Service Owners or administrators for the organization or business group that requested a change in the business service.

Individual users can claim a task to become the *task owner*. No other users can work on the task while it is owned by the user. If necessary, owned tasks can be released by the owner or claimed by another user.

A Cloud Manager Terminology

Approver

A Cloud Manager role that provides Application Console rights to approve or deny business service requests based on available resource capacity for an organization or zone.

Build Administrator

A Cloud Manager role that provides Application Console rights to complete pre-build and post-build configuration for workloads in requested business services.

Business Group

A subunit of an organization. A business group can be assigned all or some of the organization's resources (such as its hosts, templates, and networks) to use for deploying business services.

A business group might represent a cost center or a department that needs to deploy business services.

When organization members are associated to one or more business groups, they are assigned rights to use the resources that the business groups provides. For example, an organization member might be assigned as a Business Service Owner or Sponsor for a specific business group.

Business Group Viewer

A Cloud Manager role that provides Application Console rights to view business services for a business group.

Business Service

A collection of workloads that are deployed together.

Business Service Owner

A Cloud Manager role that provides Application Console rights to create, modify, and delete business services for an organization or for specific business groups within an organization.

Catalog Manager

A Cloud Manager role that provides Application Console rights to create, modify, and delete workload templates.

Cloud Administrator

A Cloud Manager role that provides Application Console rights to perform all tasks.

Cloud Manager Application Console

A web application that can be run on any computer with a supported web browser. The console is for both Cloud Manager administrators and users. Cloud Manager administrators use the console to organize computing resources so that users can consume them as business services. Users access the console to request and manage business services. Login to the console occurs through an LDAP directory designated as the authentication source.

Cloud Manager Application Server

The server component that supports the Cloud Manager Application Console and communicates with Cloud Manager Orchestration Servers to provide instructions for deploying, managing, and removing business service workloads. It also performs user authentication with the LDAP source.

Cloud Manager Orchestration Agent

A client component installed on VM hosts to enable them to be managed by a Cloud Manager Orchestration Server.

Cloud Manager Orchestration Console

The administrative interface for the Cloud Manager Orchestration Server. The console monitors and manages the activity of the Orchestration Servers, enabling you to view and troubleshoot jobs associated with workload creation and management.

Cloud Manager Orchestration Server

The server component that receives workload instructions from the Cloud Manger Application Server and directs the creation and management of those workloads by the virtual infrastructure. Depending on the size of your virtual infrastructure, you might have one or many Orchestration Servers.

Host

A computer (that is, a physical machine) that hosts one or more virtual machines (VM).

Organization

A tenant for which you are providing Cloud services. You assign resources to the organization that it can use for deploying business services.

An organization includes users (referred to as “members”) who serve in roles within the organization. These roles facilitate the management of business services and the management of the organization. A user can belong to only one organization.

An organization typically represents a company. In the case of a private service provider or enterprise IT department, an organization could represent small company units such as business units, cost centers, and departments.

Organization Manager

A Cloud Manager role that provides Application Console rights to manage users, role assignments, resource assignments, and business services within an assigned organization.

Resource Group

A collection of hosts or clusters and their associated resources (CPUs, memory, networks, and storage). In VMware vSphere environments, a resource group can also be a resource pool.

You can have both dedicated and shared resource groups. A dedicated resource group services only one organization, while a shared resource group services more than one organization.

Role

A set of rights that allows a user to perform specific activities in the Cloud Manager Application Console.

Service Level

A service level defines resource costs (vCPUs, memory, networks, and storage) for a business service. If desired, it can also define service objectives relating to such items as business service uptime, computing performance, or support availability. Any cost associated with an objective becomes part of the service level cost.

Service Level Objective

A measurable objective such as business service uptime, support availability, or support response time. Each objective can have a cost associated with it. When added to a service level, the objective's cost becomes part of the service level cost.

Sponsor

A Cloud Manager Application Console role that provides rights to approve or deny business service requests based on an organization's financial policies.

Workload

A virtual machine.

Workload Template

A template used to create a workload. The template defines the VM template from which the workload is created and allows for customizing of the VM template settings to increase or decrease the resources (vCPUs, memory, disk storage, and networks) required for the workload.

Zone

A Cloud Manager Orchestration Server and its managed resources (VM hosts, VM templates, and so on). The Cloud Manager Orchestration Console might also refer to this as a "grid."

Zone Administrator

A Cloud Manager Application Console role that provides rights to manage the resources for one or more assigned zones.