CHAPTER
2

Cloud Planning
and Architecture

A s an IT organization, you may have recently decided to move to the cloud and now need to plan how to do so. Or perhaps your business has already begun its cloud migration by moving to public clouds, but now you are looking to avoid the risk of noncompliance by providing comparable services in an internal cloud. In both cases, adequately planning and designing the appropriate cloud architecture is key to providing a successful implementation that meets the demands of your internal customers while also reducing costs.

Cloud Planning

There could be multiple reasons why your organization is considering moving to a private cloud. Typical motivations include the following:

- Consolidating infrastructure and streamlining operations

- Reducing capital expenditure (capex) in lieu of increasing operational expenditure (opex)

- Faster time in delivering key business services

Having well-defined goals and articulating them in a cloud plan ensures that the technology decisions address the right problems, stakeholders have the right expectations, and the cloud implementation goes ahead smoothly. Various people need to get involved in the cloud project planning from both IT and business organizations to make the cloud project successful.

Determining Stakeholders and Responsibilities

Deploying a cloud requires stakeholders to agree on an architecture that transforms the IT environment to take advantage of the IT as a service paradigm. This may require not only repurposing existing software and hardware resources into well-defined services, but also making shifts in organizational processes. For example, a cloud architecture that supports Database as a Service (DBaaS) may require the database administrators and

storage administrators to agree on standard software configurations and common automation workflow for provisioning the services.

Cloud Architects

Cloud architects understand and know how to leverage public and private cloud computing technologies to form business solutions. Cloud architects understand how to configure IT assets to enable services that meet the needs of the business. Generally, cloud architects should be familiar with the concepts, architecture, and other aspects of the cloud and how to map them to the business needs.

Cloud Infrastructure Administrators

Cloud infrastructure administrators are responsible for putting together the initial physical cloud infrastructure and determining the physical and logical partitioning of cloud resources based on IT and business constraints. These individuals are also responsible for the ongoing cloud operations for managing the elastic cloud capacity, managing performance and configuration, administering cloud infrastructure and deprovisioning the physical infrastructure, including server, storage, and networking. These individuals may come from multiple disciplines, such as

- System administrators

- Network administrators

- Storage administrators

Tasks performed by the cloud infrastructure administrators include

- Setting up and pooling the overall physical infrastructure (servers, storage, and networks).

- Gauging cloud infrastructure requirements and making physical infrastructure available for those requirements. The cloud infrastructure administrator may or may not be aware of the services that will consume the infrastructure.

- Ongoing monitoring and management of the cloud infrastructure.

Cloud Self-Service Administrators

One of the tenets of cloud computing is to allow cloud consumers to request cloud resources in a self-service manner. This allows for a pay-as-you-go model where consumers view the cloud as having unlimited resources from which they can request what they need, and pay for what they consume. Organizations must establish clear operating agreements between the providers and consumers of cloud services. Both sides need to understand the details and consequences of areas such as change management, financial chargeback, and problem resolution. Cloud self-service administrators define these governance policies for the cloud self service users.

Tasks performed by the cloud self-service administrators include

- Defining access boundaries for cloud self-service users/consumers. For example, self-service users may be restricted to provision resources in a specific network, from a specific location, or from a predefined resource pool, or they may be restricted to select among only certain images during deployment.

- Setting up quotas for self-service users in terms of the aggregate amount of infrastructure (server, storage, network) resources they are allowed to request.

- Publishing cloud services in a self-service catalog for the users and controlling their access.

- Setting up unit charges and, optionally, payment terms for the users.

- Establishing maintenance windows, SLAs, and support services.

A single individual may act as the self-service administrator for an organization and perform all the preceding tasks, or a group of administrators may take up the various responsibilities. For example, a separate chargeback administrator may be responsible for setting up charge plans.

Cloud Users/Application Teams

The application teams are generally the de facto "users" or "consumers" of the cloud. In many cases, the cloud infrastructure administrators will belong to central IT, whereas the cloud users will be testers and developers in the

application teams from the line of business side of the house. These users have access to a self-service interface or portal application. They can browse a service catalog to select a service for deployment. The users have no visibility into the underlying cloud infrastructure that hosts that service.

Tasks performed by the cloud users include

■ Requesting resources (virtual machines, databases, middleware, and so on) from the cloud

■ Monitoring and administering requested resources

■ Scaling out and scaling back existing deployments in the cloud

■ Reviewing charges accrued for the cloud services consumed

Discovering and Assessing Existing Assets and Current Workloads

As discussed in Chapter 1, a cloud project, especially a private cloud project, may be considered either *greenfield* or *brownfield*. As a reminder, a greenfield project is one that involves new infrastructure, including, in some cases, a new data center. A brownfield project is one where existing assets are consolidated and transformed into a cloud. For the brownfield projects, it is important to discover existing assets (technology portfolio), make them manageable by Oracle Enterprise Manager 12*c*, and then baseline their performance for consolidation.

As part of planning for the cloud, the process of discovery and assessment looks for the software and hardware assets in the environment along with the relationships among them. Many companies do not have an accurate, current inventory of both their physical and virtual environments. As the first step to developing a plan for the new cloud environment, it is important to take a correct inventory of the current technology portfolio. Companies may choose to rationalize all the assets in their IT environment in one project, but this "boil the ocean" approach can introduce significant risks. A more practical approach is to develop a strategy that involves multiple projects, each focusing on rationalizing only the assets that support a horizontal or vertical "slice" of business functionality, and is aligned with a business goal or objective.

TIP
For easier transition to the cloud, select the part of the IT environment that is not organizationally or technically complex and where there is opportunity for short-term savings, via license or support savings. Near-term success and payback will help justify continued investment.

Rationalizing the technology portfolio involves two steps:

1. Capturing the existing technology portfolio. This can be done either manually, with spreadsheets, or with automated tools such as Oracle Enterprise Manager 12*c*. Typical information captured includes: asset name and description, owner, location, department, number of users, packaged/custom, hardware, operating system, and database.

2. Mapping the technology portfolio to the business capabilities. Once the inventory is complete, the organization must map the assets to either capabilities or business processes within the business architecture. This makes it easy to identify redundancies and gaps. Completing such an inventory and mapping exercise will usually reveal many overlapping and duplicate assets that are candidates for consolidation. When a recommendation for an asset is not obvious, a more detailed evaluation may be required.

Oracle Enterprise Manager can be used to discover and collect information about datacenter assets either one time or on an ongoing basis. To do automatic host discovery, Oracle Enterprise Manager uses an agentless Nmap (Network Mapper) protocol. Nmap can discover hosts only for a subnet, so the process has to be repeated for every subnet. For each subnet, Nmap is executed as a job from one host where the agent is already deployed. Figure 2-1 shows the process of automatic discovery.

In automatic host discovery, a single Oracle Enterprise Manager Agent is tasked to scan the entire network based on IP address ranges that the administrator specifies. It then returns a list of "unmanaged" host machines—that is, host machines that do not yet have the Oracle Enterprise Manager Agent installed—with a list of ports in use that match
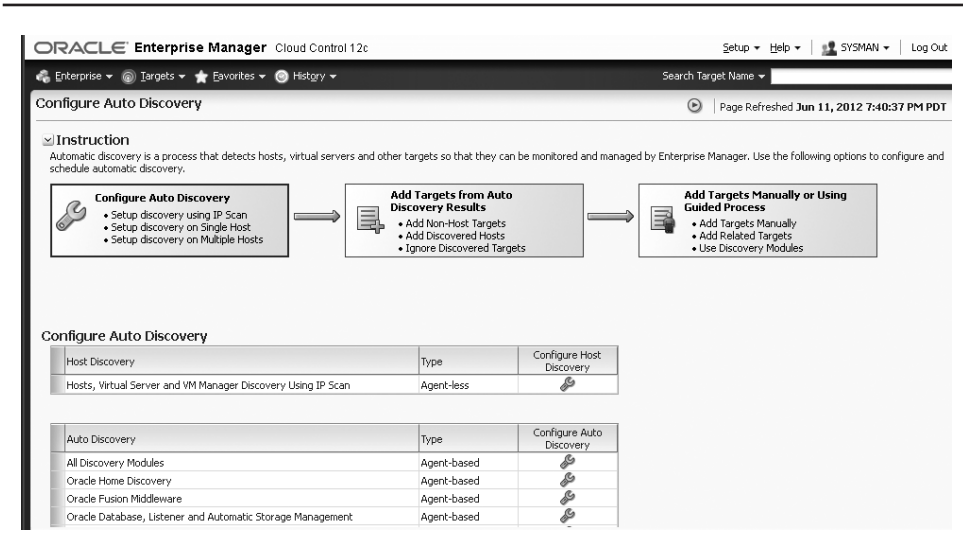
FIGURE 2-1. *Automatic discovery in Oracle Enterprise Manager*

the ranges specified. The name of the service using each port is also returned.

By looking at the list of services and ports, an administrator should be able to determine what types of Oracle components have been discovered. For example, if a host is returned with port 7001 in use, the administrator can reasonably assume that this port is associated with an Oracle WebLogic Server domain that can be promoted to managed target status. Similarly, a host returned with port 1521 in use would indicate a database server.

The next step is to deploy Oracle Enterprise Manager Agents to the hosts if the administrator wants to promote the host to a managed status. The Agent Push application is integrated with the discovery flow, so that the agent is pushed into the hosts. Once an Oracle Enterprise Manager Agent is deployed to the host, any Oracle components running on the host will be discovered and reported as potential targets. These components can then be promoted to managed target status, enabling them to be managed and monitored by Oracle Enterprise Manager Cloud Control 12*c*.

To be most effective, automatic discovery should ideally be run by a network administrator with an overall understanding of which Oracle components are running on which ports.

Let's now look at the deep discovery process that takes place once the Oracle Enterprise Manager Agent is placed on the host. The discovery process is automated, cataloging all the elements in your data center and the dependencies that exist between systems. The dependencies are particularly important when planning for the cloud, as failures can occur when interdependencies are not properly configured, both in the cloud and across cloud/physical lines. Supported asset types include

- Servers, storage, and network (hosts and virtual machines, for example)

- Databases (a database instance, for example)

- Groups, systems, and services (a database system, for example)

- Middleware (application deployments, WebLogic Servers, and WebLogic Domains, for example)

- Business applications (such as Oracle Fusion Applications, Oracle Siebel, Oracle E-Business Suite, etc.)

The discovery process is thorough in detecting software and targets through multiple means. However, if for some reason a target is not discovered, Oracle Enterprise Manager provides a guided, manual flow to add the target.

Oracle Enterprise Manager 12*c* supports a group of metrics and compliance rules to be applied to the newly discovered or created targets in the form of Monitoring Templates. For example, there could be different Monitoring Templates for monitoring production and test databases. When the targets get promoted, the Monitoring Templates for subsequent management can be automatically applied, thereby making the targets manageable from the onset.

Figure 2-2 shows the Inventory and Usage Details page of Oracle Enterprise Manager Cloud Control 12*c*, which provides a summary view of the discovered targets and allows you to drill down to view details about each target. Configuration information that is periodically collected is stored in the Oracle Enterprise Manager's Repository, allowing you to view, save, track, compare, search, and customize up-to-date configuration information for your entire enterprise.
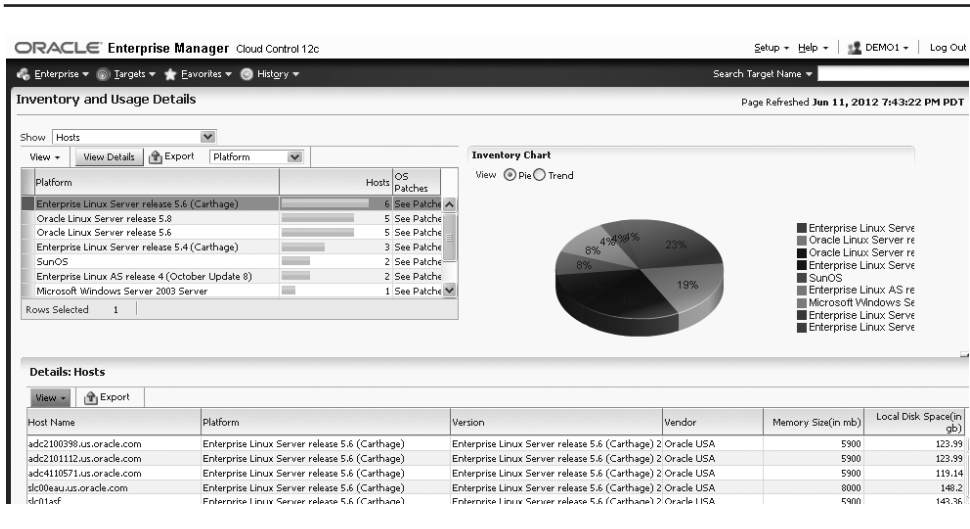
FIGURE 2-2. *Oracle Enterprise Manager asset inventory*

Administrators can segregate the targets into Administration Groups, which greatly simplify the process of setting up targets for management in Oracle Enterprise Manager by automating the application of management settings such as monitoring settings or compliance standards. Typically, these settings are manually applied to individual target, or perhaps semi-automatically using custom scripts. However, by defining Administration Groups, Oracle Enterprise Manager uses specific target properties to direct the target to the appropriate Administration Group and then automatically apply the requisite monitoring and management settings. This level of automation simplifies the target setup process and also enables a data center to easily scale as new targets are added to Oracle Enterprise Manager for management. Administrators can also store with each target additional target attributes, like Cost Center, Location, Lifecycle Status, and so forth, for better reporting and decision making. Figure 2-3 demonstrates the typical Administration Group workflow.

The first step involves setting a target's Lifecycle Status property when a target is first added to Oracle Enterprise Manager for monitoring. At that time, you determine where in the prioritization hierarchy that target belongs, with
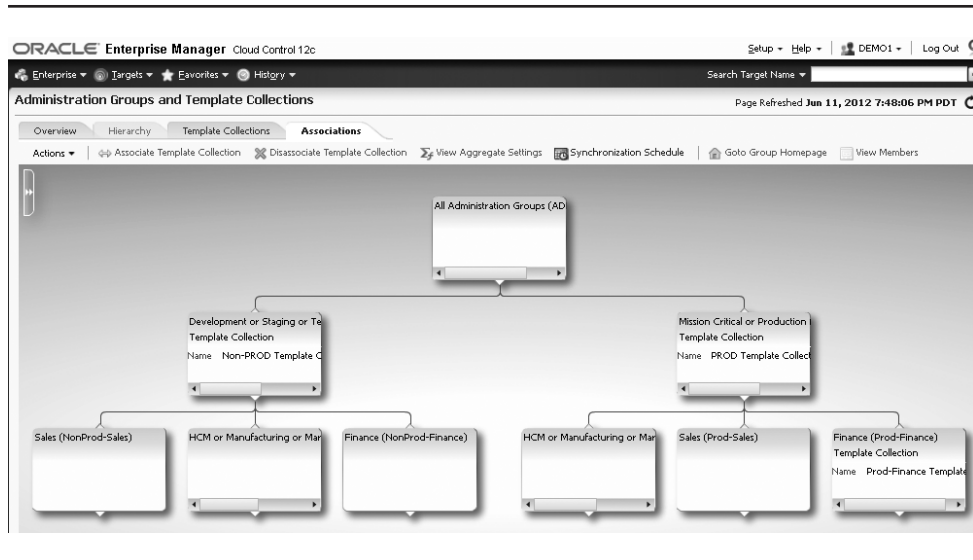
FIGURE 2-3. *A typical Administration Group*

the highest level being "mission critical" and the lowest being "development." Target Lifecycle Status prioritization consists of the following levels:

■ Mission Critical (highest priority)

■ Production

■ Staging

■ Test

■ Development (lowest priority)

As shown in the second tier of Figure 2-3, once Lifecycle Status is set, Oracle Enterprise Manger uses it to determine which Administration Group the target belongs to.

To prevent different monitoring settings from being applied to the same target, Administration Groups were designed to be mutually exclusive with

other Administration Groups in terms of group membership. Administration Groups can also be used for hierarchically classifying targets in an organization. For example, Figure 2-3 shows an Administration Group hierarchy consisting of two subgroups, Production targets and Test targets, with each subgroup having its own Template Collection. In this example, targets in the Production subgroup inherit monitoring settings from Monitoring Template A while targets in the Test subgroup inherit monitoring settings from Monitoring Template B.

Consolidation Planning

The discovery and assessment process results in the identification of targets, after which their workloads are monitored to determine if they are suitable for moving to the cloud. Depending on the procurement cycle or the specific hardware vendor agreement in effect, enterprises may acquire different types of server hardware and operating systems for building the cloud, inadvertently creating a confusing array of systems that administrators must manage, administer, patch, and upgrade. This, in turn, increases labor costs and ongoing maintenance and support costs against IT budgets. Hence, it is important to find a way to migrate existing, often disparate hardware and workloads to standardized operating systems and hardware in the cloud, with the ultimate goal of reducing costs.

Before delving deeper into the Oracle Enterprise Manager 12*c* Consolidation Planner, let us understand how Oracle Engineered Systems and server virtualization technologies can function as consolidated platform for the cloud.

Oracle Engineered Systems and Server Virtualization Technologies

Engineered systems are hardware and software that are pre-integrated to reduce the cost and complexity of IT infrastructures while increasing productivity and performance. Oracle's engineered systems contain software components such as the operating system, middleware, database, and so forth that are optimized for performance, reliability and scalability. Typically, customers can house tens and thousands of application components in a relatively dense configuration to save on data center floor space, power and cooling costs. That is why engineered systems are often the preferred building blocks for a consolidated cloud platform. Examples of engineered

systems include Oracle Exadata Database Machine, Oracle Exalogic Elastic Cloud, Oracle SPARC SuperCluster, and Oracle Database Appliance.

Enterprises are also increasingly investigating server virtualization technologies, such as Oracle VM, as a foundation for clouds. Server virtualization masks physical server resources such as processors, memory and storage from applications running on the server. The server administrator uses a virtualization software called the hypervisor to divide one physical server into multiple isolated virtual environments. The virtual environments are also known as guests, instances, containers, or emulations, depending on the platform. This makes it possible to use the shared hardware infrastructure while getting the benefits of isolation that virtualization provides. Guest virtual machines running on Oracle VM Server can be configured to use one of the following virtualization modes:

- **Paravirtualized mode (PVM)** In this mode, the kernel of the guest operating system is modified to distinguish that it is running on a hypervisor instead of on the bare-metal hardware. As a result, I/O actions and system clock timers in particular are handled more efficiently, as compared with nonparavirtualized systems where I/O hardware and timers have to be emulated in the operating system. Oracle VM supports PV kernels for Oracle Linux and Red Hat Enterprise Linux, offering better performance and scalability.

- **Hardware virtualized mode (HVM) or fully virtualized mode** When support for hardware virtualization (Intel VT-x or AMD-V) is available in the host server hardware, the guest operating system may run unmodified and depend on the hardware features to ensure proper operation in a virtualized environment. However, HVM can introduce additional overhead since it relies on binary translation and device emulation. But it brings the benefits of compatibility: largely, any "unmodified" operating system image such as Linux, Solaris, or Windows that is not virtualization-aware can run as an HVM guest without any changes, making it easy to consolidate existing physical server images, leaving them essentially untouched. With the recent development in Intel and AMD processors, in some scenarios, hardware-assisted virtualization will also provide performance benefits.

■ **Hardware virtualized mode using paravirtualized drivers (PVHVM)**   This mode is identical to HVM, but with additional paravirtualized I/O drivers installed in the guest's operating system to improve virtual machine performance. This allows the operating system to essentially run unmodified, but uses PV drivers that have been modified specifically to execute I/O commands efficiently in a virtualized environment to minimize overhead. Oracle provides PV drivers for Microsoft Windows. PV drivers are provided as part of the operating system for Oracle Linux and Oracle Solaris. HVM works with all CPUs featuring the AMD-V (SVM) extension. Intel Virtualization Technology requires a computer system with an enabled Intel processor, BIOS, virtual machine monitor (VMM), and, for some uses, certain platform software enabled for it.

Depending on whether the cloud infrastructure will be physical or virtualized, the workloads are potential P2V (physical to virtual) or P2P (physical to physical) migration candidates. P2E (physical to engineered systems) can be considered to be another variant of consolidation planning. Consolidating applications, systems, or data centers reduces costs and simplifies operations.

Consolidation, however, could result in a case of putting more eggs in a single basket. To consolidate successfully, you need to ensure that you have a "stronger basket." The principles of virtualization, standardized processes, and management automation help to ensure that IT operations and systems are more robust. There are many other aspects of resiliency, such as security and disaster recovery, that also need to be enhanced. By consolidating, however, you can afford the investments needed to implement more robust technologies and processes, and thus the consolidated systems can indeed be made stronger.

Oracle Enterprise Manager 12*c* Consolidation Planner

Oracle Enterprise Manager 12*c* Consolidation Planner enables you to map servers you want to consolidate to the generic physical machines, Oracle Exadata or Exalogic engineered systems, or Oracle VM servers on which they can be consolidated. By leveraging metric and configuration data collected from servers managed by Oracle Enterprise Manager, Consolidation Planner helps you to determine the optimum consolidation scenarios that will produce the maximum cost benefits for the risk you are willing to take.

The goal of consolidation is to identify underutilized servers and find a way to consolidate them, enabling the enterprise to free up as many servers as possible while continuing to maintain service levels. Since servers have different levels of CPU capacity, Consolidation Planner uses computer benchmark data to normalize the CPU usage for a given hardware in the consolidation process. Specifically, Consolidation Planner uses the SPECint2006 benchmark, published by Standard Performance Evaluation Corporation (SPEC), for the different classes of hardware. More information on this benchmark is available on the SPEC website (www.spec.org).

Let us look at how workload characteristics can be taken into account when consolidating into a cloud environment.

Identifying the Right Workloads   Source servers are existing servers in your data center that are candidates for consolidation based on existing workload. Effective consolidation must also take into account the nature of the workload and the supporting technologies. For example, you may want to consolidate the databases of two different applications. If the applications currently run on different versions of the database, you may decide to keep each software stack pristine, and use CPU virtualization to simply consolidate the applications onto a pool of hardware servers. In this case, the historical data stored in the Oracle Enterprise Manager repository for the CPU resource consumption by the database workloads is used for consolidation (see Figure 2-4).

Destination servers are either existing servers or yet-to-be-purchased servers that source servers will be consolidated to. These may be individual machines or an integrated system such as Oracle Exadata or Oracle Exalogic.

Consolidation Project   A consolidation project is created for each consolidation effort, and then individual consolidation scenarios are created and executed within the scope of the project. A consolidation project defines the scope of consolidation in terms of the following:

- The type of consolidation. In Oracle Enterprise Manager 12*c*, two types of consolidation schemes are supported:

   - **P2V**   From physical source servers to Oracle VM destination servers

   - **P2P**   From physical source servers to physical destination servers
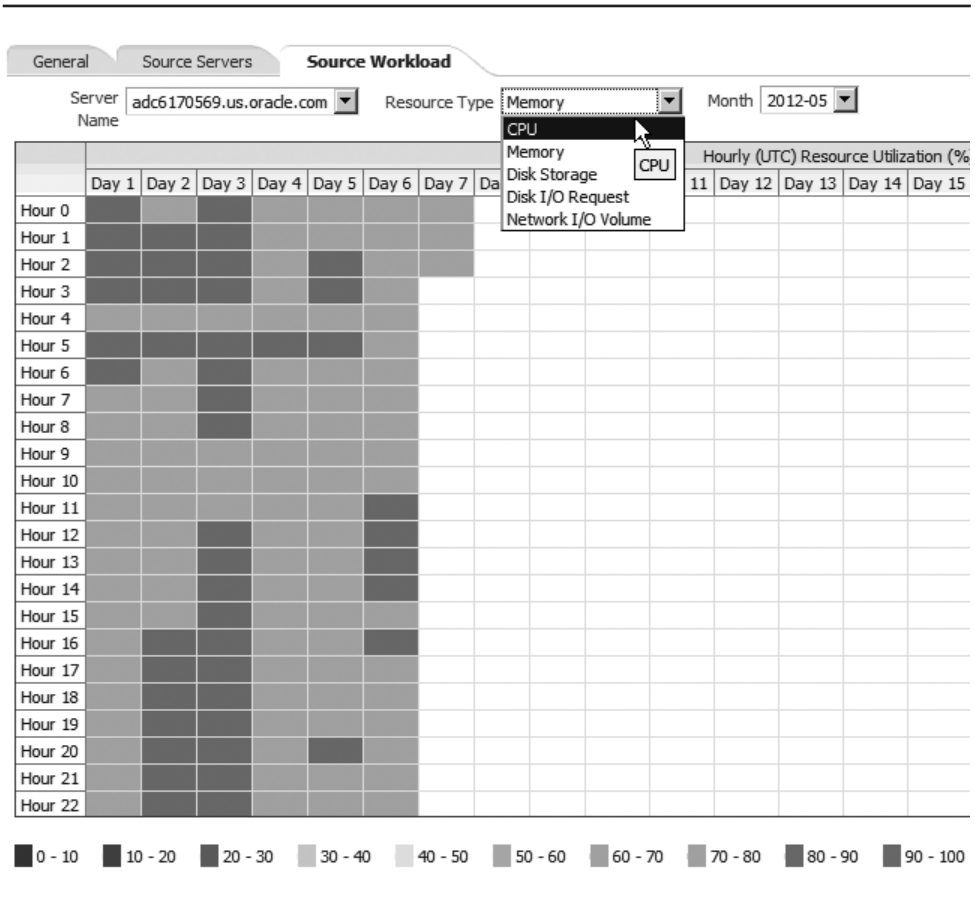
FIGURE 2-4. *Oracle Enterprise Manager's Consolidation Planner*

- The preliminary set of candidate source servers to consider consolidating from.

- The preliminary set of candidate destination servers to consider consolidating to.

- The duration over which resource usage data will be collected for the workloads on the source servers. This data is used to create consolidation scenarios.

After the project is defined, an Oracle Enterprise Manager job is executed to bring the historical performance data for the source servers from the Oracle

Enterprise Manager repository into the Consolidation Planner application. Once the job finishes, the project becomes an active project, and data collection happens on regular intervals—usually once a day.

Consolidation Scenario Each consolidation project contains one or more consolidation scenarios. You can create and execute scenarios and then compare them to determine which consolidation strategy makes the most sense. Consolidation scenarios can be created by providing the following information:

- The metrics to take into account for consolidation. These could be one or more of CPU, memory, disk I/O, network I/O, and disk storage.

- Any business, compliance, or technical constraints that must be considered.

- The list of destination servers.

Administrators can create their own custom scenarios that best suit their situation. Alternatively, they can choose from a set of preconfigured consolidation scenarios provided in the product, representing conservative, aggressive, and medium consolidation schemes. Once created, the various scenarios can be executed and the results can be evaluated to choose the best consolidation strategy. Administrators can also change the various parameters in a scenario, rerun the scenario, and view the new results.

Consolidation Constraints Consolidation constraints are business, corporate, or technical constraints that must be taken into account when consolidating. The constraints represent the restrictions placed on the consolidation process for business, legal, or compliance reasons. There may be constraints on the way in which workloads on the source servers can be combined, or there may be constraints on the way in which destination servers can be used as targets for consolidation.

For source servers, there are two categories of constraints: inclusive and exclusive. Inclusive constraints bring potential consolidation candidates into the consolidation scope based on target properties. Exclusive constraints

exclude servers from the consolidation scope based on target properties. Some examples of exclusive constraints are

- **Location** Laws or regulations may mandate that specific servers must remain within a specific location, such as a particular datacenter location or within a specific geographical area. Sometimes, country-specific regulations or data privacy rules may dictate that resources must remain within the national boundaries.

- **Lifecycle status** Test and production servers should not be consolidated on common hardware.

- **Application** For businesses that host applications for several customers, there may be contractual obligations to keep separate the servers and data for those customers.

- **Security** Certain machines may have specific security requirements.

For destination servers, there are three categories of consolidation constraints: utilization, related, and exclusiveness. Utilization constraints limit the amount of server resources that may be utilized. Related constraints dictate that some servers are associated and therefore should be consolidated in the same server pool or server zone. Exclusiveness constraints prevent certain source servers from being consolidated on the same destination server due to compliance and technical reasons. The following are some examples of scenarios where servers should not be consolidated on the same destination server:

- **Department** Broker and dealer systems should be on different machines.

- **Lifecycle status** Servers utilized within different stages of the application lifecycle, such as testing and production should not be mixed together.

- **Affinity** Servers used for the middleware and database tiers should be consolidated separately.

- **Topology** Database Real Application Cluster nodes should not be on the same physical server because doing that will not provide fault tolerance against physical server failure.

■ **High availability**   Machines with specific HA requirements, such as multiple web servers or application servers, should not be on the same physical machine.

■ **Disaster recovery**   Machines supporting disaster recovery strategies or other application-level availability features should not be placed with other regularly used systems.

■ **Security**   Servers in different security zones should not be mixed together.

Choosing the Best Consolidation Strategy   You can define new scenarios and rerun existing scenarios with new input parameters and compare the results. This iterative process helps you to obtain the optimized consolidation scenario that is generated by comparing various factors and weighing different trade-offs. For each scenario that's executed, Oracle Enterprise Manager provides the following information:

■ **Destinations**   The list of destination servers to which the source servers will be consolidated. Resource configuration and calculated utilization are shown for each destination server.

■ **Ratio**   The ratio of source servers to destination servers. By default, the Consolidation Planner tries to "fit" source servers into as few destination servers as possible.

■ **Confidence**   The percentage of the data collected for source servers that meets the source server usage requirements defined in the scenario. This value is aggregated for all source servers defined with the project.

■ **Violations**   The number of violations of technical or business constraints defined in the scenario.

■ **Exclusions**   The number of source servers that do not have a qualified mapping to a destination server. These are source servers that exceed the capacity of available destination servers. This metric is applicable only if auto-mapping of source servers to destination servers is used.

A different set of constraints may result in a different optimal scenario. You can modify the constraints to come up with different scenario results. Comparing various results lets you determine the consolidation strategy that best meets your requirements. These strategies include:

- Matching source server workload requirements with destination servers to achieve required application performance after consolidation

- Fitting source server requirements with each destination server's available resources as tightly as possible, so you can get maximum usage of destination server capacity

- Balancing the source server workload across all available destination servers

In summary, Oracle Enterprise Manager 12*c* Consolidation Planner helps you with the cloud planning process by allowing you to explore various options for reusing existing physical hardware, leveraging server virtualization, and procuring new hardware such as Oracle Exadata or Oracle Exalogic. If done correctly, consolidation, as a first step to building a cloud architecture, can go a long way toward providing improved performance, ease of management, improved technology, and lower costs. Once the technology portfolio for the cloud is identified, the hardware and software components need to be modeled into a cloud architecture that addresses key business requirements.

Cloud Architecture

Without an optimized architecture that is built on a standardized portfolio, with standardized interfaces, the cloud provider will not be able to effectively provide at an attractive price the quality of service (QoS) and the level of agility required to keep its cloud consumers happy. Conversely, a cloud consumer of shared services doesn't need to know about the underlying cloud infrastructure, but should be aware of the available services and the resource model exposed in the GUI or API.
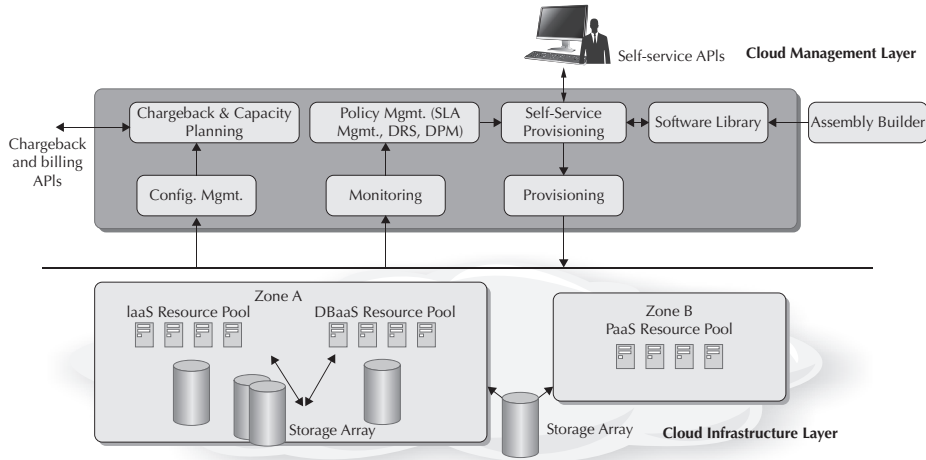
FIGURE 2-5. *Oracle Enterprise Manager cloud components*

Figure 2-5 depicts the high-level cloud schematic, which consists of two layers. The upper layer, the cloud management layer, is responsible for handling user interactions with the cloud. The lower layer, the cloud infrastructure layer, deals with virtual and physical resource management. The Oracle Enterprise Manager Cloud Control 12*c* Virtualization Plug-in provides the underlying resource management for Oracle VM today.

Cloud Management Layer Components

Oracle Enterprise Manager handles cloud user authentication, user/role management, software library management, provisioning requests, policy definition and enforcement, monitoring and reporting, and other aspects of managing the cloud. In addition, you can use a client-side tool such as Oracle VM Template Builder or Oracle Virtual Assembly Builder (OVAB) to build and modify Oracle VM Templates or Oracle Virtual Assemblies.

Following is a brief description of various components included in cloud management:

■ **Self-Service Provisioning**   As the name indicates, this component allows authorized end users (developers, testers, and so forth) to provision applications and resources on demand in the cloud. In addition to a self-service portal GUI, a web services API is available to allow integration of external applications. To supplement self-service provisioning, the cloud management layer includes the following:

  ■ **Oracle VM Template Builder/Oracle Virtual Assembly Builder**   These tools can be used to design and bundle various services to be published and deployed in the cloud. An *Oracle VM Template* is a package that comprises an operating system and other application components that can be deployed as a single virtual machine. An *Oracle Virtual Assembly* extends that concept. A Virtual Assembly is a collection of virtual appliances that comprises a multitier application that can be deployed across several virtual machines (for example, a web application with a web server, an application server, and a database server). Assembly Builder can be used to modify the prepackaged Virtual Assemblies shipped by Oracle. It also integrates with Oracle Enterprise Manager Software Library so that Virtual Assemblies can be staged within the Software Library directly from Assembly Builder.

  ■ **Software Library**   This is the repository for managing all prepackaged and user-defined Oracle VM Templates and Virtual Assemblies, provisioning profiles (in case of physical servers), and associated scripts. The Software Library can be co-located either within the same infrastructure as the management server or remotely, closer to the end targets of the cloud.

■ **Chargeback and Capacity Planning**   This component is responsible for providing chargeback, trending, and other information related to resource consumption and can integrate with external billing systems.

■ **Policy Management**   This component is responsible for enforcing user-defined policies. These policies could be as simple as determining initial placement of VMs or reserving capacity to meet a future need (VM reservation), or could be as sophisticated as dynamically scaling applications up or down to maintain a certain level of business service. This includes Dynamic Resource Scheduling (DRS) and Dynamic Power Management (DPM) capabilities.

■ **Monitoring, Configuration Management, and Compliance Management**   These components provide comprehensive post-deployment operations management, such as performance and configuration management solutions for the cloud.

Cloud Infrastructure Layer Components

The cloud infrastructure layer consists of hardware and software components that are used to build the cloud environment. This includes the various software, server, storage, and networking components organized into resource pools and cloud zones. As discussed earlier, Oracle Enterprise Manager's Consolidation Planner can help an administrator identify the right resources to be consolidated for use in the cloud.

The cloud infrastructure layer consists of cloud resources, modeled such that the physicality is uniformly abstracted regardless of the nature of service—IaaS or PaaS. Oracle offers technologies to support both physical and virtual environments. In each of these cases, the cloud comprises multiple subclouds, called *zones.* Each zone consists of a pool of physical servers, representing a logical entity such as a datacenter location, a department, or a certain level of availability. In terms of implementation, zones may be either of the following:

■ An Oracle VM Zone, representing a bunch of servers with the Oracle VM hypervisor installed

■ A PaaS Infrastructure Zone, which can comprise multiple physical hosts (with Oracle Enterprise Manager agent installed) or a preexisting Oracle VM Zone that is used by higher-order services like PaaS

When a service is exposed to the end consumer, the physical resources within the zone are all concealed from a user-interaction perspective. The end user simply selects the zone to deploy the service onto.

■ Zones encapsulate physical resources, including attached storage and network resources.

■ Zones can serve as a boundary for chargeback and metering, because Oracle Enterprise Manager allows assignment of charge plans at the zone level.

- Zones can be exposed either directly as an Oracle VM Zone directly or as a PaaS Infrastructure Zone.

- Members of a zone are selected by cloud administrators.

- Within a zone, there are groupings of homogeneous resources, called pools.

An Oracle VM Zone is made up of virtual server and storage pools. A virtual server pool can contain a maximum of 32 Oracle VM Servers. There must be at least one virtual server in the server pool. One of the servers acts as a master server for the server pool.

All virtual servers in a server pool should have CPUs in the same CPU family, but the configuration of the CPUs can differ, such as the number of cores. If the CPUs are not in the same CPU family, some operations, such as live migration, may fail. Other hardware components on the host computer may also differ, such as the amount of RAM, the number and size of disk drives, and so on. All the servers in a pool should share a common file system (either a NAS export or a mounted LUN).

A storage pool can consist of external storage elements such as storage servers, exported file systems, and LUNs. An external storage element is created on dedicated storage hardware, such as a server configured for NAS offering NFS shares. The server on which the storage element is present must be accessible by the Oracle VM Servers through a Fiber Channel or Ethernet network.

Similarly, a PaaS Infrastructure Zone may comprise database or middleware pools, representing hosts with homogeneous software installations, such as a particular database or a middleware version, so that the placement within the pool can be done transparently and the same QoS can be guaranteed independent of the placement.

To summarize, resource pools

- Are typically homogeneous

- Are defined against a single zone

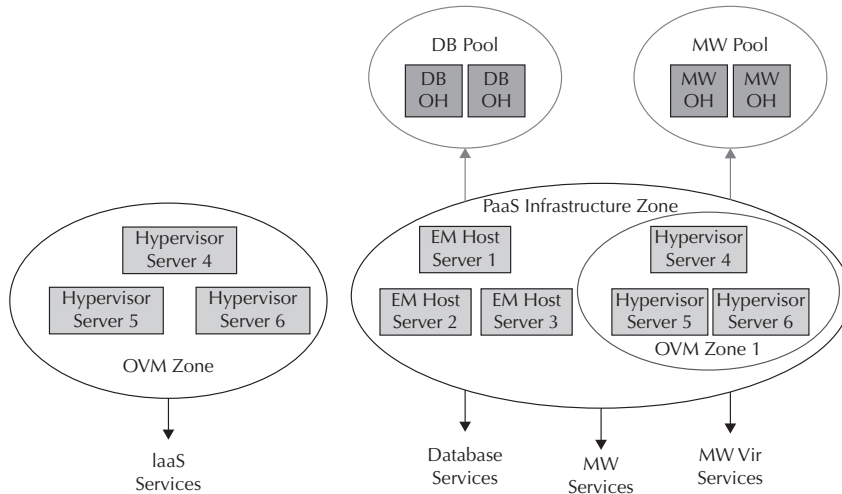 - Underlying physical resource shared across resource pools

FIGURE 2-6. *Cloud resource model*

- ■ Have membership criteria set by integrators

  - ■ Based on software requirements for a particular service (for example, 10.3.5.0 Middleware homes or a bunch of 10.2.0.4 RAC clusters)

- ■ Can have certain limits imposed by the underlying technology (for example, number of schemas supported within a database)

The organization of zones and pools is depicted in Figure 2-6.

Cloud Infrastructure Planning

Now that the cloud resource models have been defined, cloud administrators can use the concepts to plan their transition strategy. Cloud can be delivered in the form of discrete services, such as OS infrastructure (IaaS), and runtime platforms (PaaS), such as databases (DBaaS) or middleware (MWaaS). In addition, tailored services can be offered for testing (TaaS) or delivering packaged software as services (SaaS). Server virtualization is suited to deliver

IaaS, but it may fall short in delivering PaaS. Whether to choose server virtualization or build the cloud directly on physical servers depends on several factors:

- **Type of service   The type of service offered** (IaaS, DBaaS, or PaaS) may determine the underlying infrastructure required.

- **Isolation**   Do the tenants need to be isolated, so that they can be patched and upgraded independently? VMs offer the maximum degree of isolation because everything from the guest operating system and up is isolated to a tenant.

- **Nature of the application**   Owing to performance reasons, especially related to I/O, certain workloads, such as production databases, are best run on physical hosts. Similarly, some applications may be more conducive to nonclustered pools.

- **Skill set of the administrator**   The skill set of the person that plans and delivers the cloud, such as a database administrator (DBA) or a system administrator (sysadmin) may determine the technology they are most comfortable working with

- **Manageability**   VMs offer isolation, but may also result in sprawl. Administrators often may find it difficult to maintain thousands of VMs from a compliance perspective. Other models, such as a RAC-based DBaaS, offer better consolidation and could be easier to manage.

The good news is that Oracle Enterprise Manager supports different deployment models for each of these use cases. It offers a variety of architectural choices to the cloud provider to support all the aforementioned service types, in both virtual and physical environments. Figure 2-7 depicts some of the choices an administrator has available. You can deploy multitier applications on virtualized servers or deploy databases and applications on pre-created platforms, whether physical or virtual. Even within the realm of virtualization, Oracle offers choices! You can choose virtualization based on Oracle VM Server for x86, Oracle VM Server for SPARC, or Solaris Containers
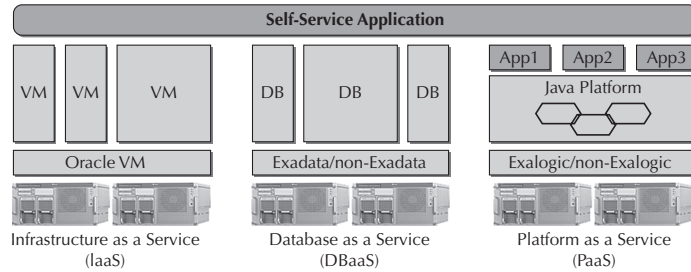
FIGURE 2-7. *Various architectural choices mapped to cloud services*

(though this book covers only Oracle VM Server for x86, which is supported by Oracle Enterprise Manager Cloud Control 12*c*).

Oracle Enterprise Manager 12*c* comes with a Service Catalog that is rich with service templates for these different technologies. Using Oracle Enterprise Manager 12*c*, a cloud provider can create one or more of these services and characterize them based on size (total CPU, memory, etc.), QoS, version, and other attributes that are inherently part of the Oracle Enterprise Manager's discovery and collection process.

So, what actions does the administrator need to perform to set up the cloud? As previously discussed, administrators are in the business of cloud enablement, for which Oracle Enterprise Manager 12*c* ships with out-of-the-box roles such as cloud administrator and self-service administrator. Oracle Enterprise Manager also provides out-of-the-box automation for these roles. The cloud administrator can set up the cloud infrastructure, such as servers, storage, and database resource pools, and group them into zones. The self-service administrator then defines access controls, quotas, retirement policies, and the charge plans for the service and releases the service in the self-service catalog. Figure 2-8 shows the various activities an administrator performs for setting up a cloud.
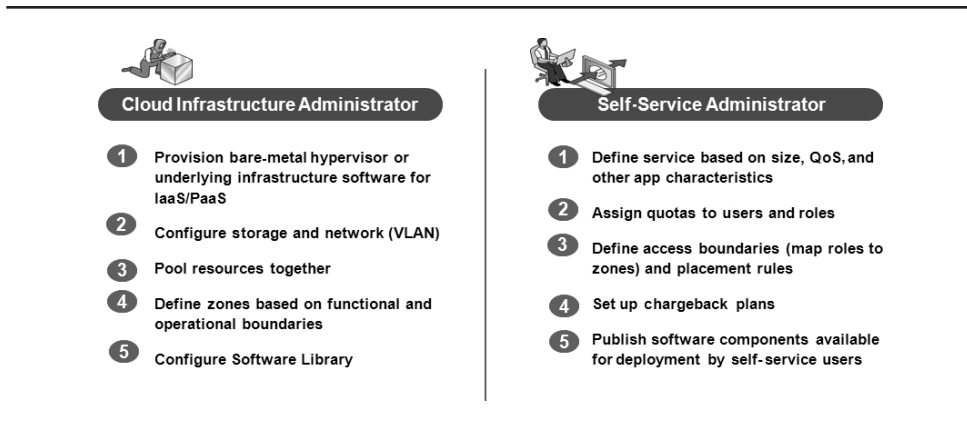
Cloud Infrastructure Administrator

1 Provision bare-metal hypervisor or underlying infrastructure software for IaaS/PaaS

2 Configure storage and network (VLAN)

3 Pool resources together

4 Define zones based on functional and operational boundaries

5 Configure Software Library

Self-Service Administrator

1 Define service based on size, QoS, and other app characteristics

2 Assign quotas to users and roles

3 Define access boundaries (map roles to zones) and placement rules

4 Set up chargeback plans

5 Publish software components available for deployment by self-service users

FIGURE 2-8. *Oracle Enterprise Manager–enabled cloud setup activities*

Summary

In this chapter, you were introduced to cloud computing concepts and the various reasons why planning is an important and unavoidable aspect of building a cloud. You learned about

■ Automated discovery of IT assets

■ The IaaS and PaaS cloud service delivery models and how the different consolidation technologies from Oracle can serve them

■ The cloud resource models and the concepts of zones and pools, which will be very important in subsequent chapters