Intel® Cloud Builders Guide: Cloud Design and Deployment on Intel® Platforms–Oracle

Cloud Design and Deployment on Intel® Platforms: Oracle

Intel® Xeon® Processor 5500 Series

Intel® Xeon® Processor 5600 Series

AUDIENCE AND PURPOSE

This reference architecture outlines an enterprise cloud infrastructure deployment using Oracle's complete stack with servers, storage, network, and software components. The Oracle* Optimized Solution for Enterprise Cloud Infrastructure makes use of pre-configured Oracle VM virtual machines, that contain optimized configurations for running applications on Oracle's Sun Blade 6000 Modular Systems with Intel® Xeon® processor based blade servers and Oracle's Sun ZFS Storage Appliances. Using the contents of this paper, which includes detailed scenarios and screen shots, should significantly reduce the learning curve for building and operating your first cloud computing infrastructure.

IT organizations typically spend many weeks to plan, architect, and deploy a multi-vendor solution: the process is not only time consuming, but also error-prone, making it hard to get the best value from the investment. This document offers a much simpler approach that speeds deployment and reduces risk. It is a single-vendor solution for the entire hardware and software stack and can be deployed in hours or days rather than weeks.

Because the creation and operation of a cloud can require integration and customization to existing IT infrastructure and business requirements, it is not expected that this paper can be used "as-is." For example, adaptation to an existing network and identification of management requirements are out of scope for this paper. Therefore, it is expected that the user of this paper will make adjustments to the design to meet specific customer requirements. This paper is assumed to be a starting point for that journey.

www.oracle.com/goto/cloud

June 2011

Table of Contents

Executive Summary

This document describes how the Oracle* Optimized Solution for Enterprise Cloud Infrastructure can be utilized to provide an enterprise cloud computing deployment. It covers deployment of software, hardware, storage, and network components and is intended to serve as a practical guide to help organizations get up and running quickly while maximizing the potential benefits of Oracle VM for quickly building a cloud deployment using Oracle's Sun Blade servers based on Intel® Xeon® processors and Sun ZFS Storage Appliances.

Oracle Optimized Solution for Enterprise Cloud Infrastructure provides a complete, integrated solution to dramatically reduce the time required to deploy Infrastructure-as-a-Service clouds, and deliver lower total cost of ownership and improved productivity for enterprises and cloud service providers. Oracle Advanced Customer Services offers a suite of services that maximizes the value of Oracle Enterprise Cloud Infrastructure investments. Through a lifecycle services approach, Oracle experts can help architect customized solutions that reduce deployment risk and ensure operational readiness.

Introduction

The Oracle Optimized Solution for Enterprise Cloud Infrastructure is used to build the Intel® Cloud Builders test configuration with Oracle's complete infrastructure stack with hardware and software components (Figure 1). The optimized solution makes use of pre-configured virtual machines that contain optimized configurations for running applications on Oracle's Sun Blade Modular Systems with Intel Xeon processors, Oracle's Sun ZFS Storage Appliances, and Intel® architecture based Storage Servers. The hardware components have been tested with Oracle VM by the respective product teams, and the Intel Cloud Builders scenarios have been implemented by Oracle EMEA (Europe, Middle East, Africa) and Intel.

Oracle VM Templates are used to simplify and accelerate deployment of the Oracle Optimized Solution software stack. For the Intel Cloud Builders test cases, pre-installed VM template versions of Oracle Linux* and Oracle VM Manager were used. These were deployed on Oracle VM Server, which runs directly on Oracle's Sun Blade servers. Best practices for optimizing the environment are built into the templates, which define configurations for all components of the software stack. Oracle Secure Global Desktop is utilized for remote desktop access.

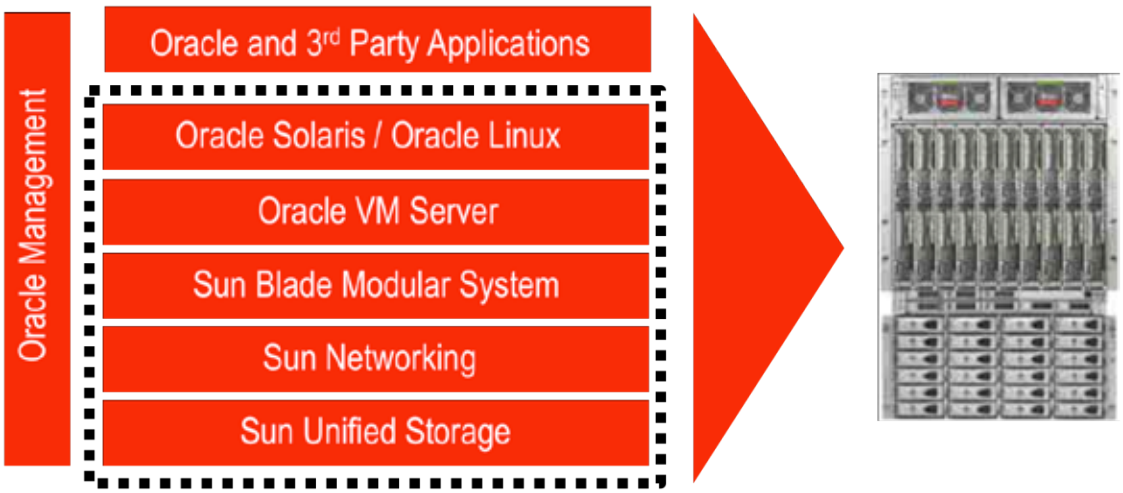


Figure 1: Oracle* Optimized Solution for Enterprise Cloud Infrastructure provides a complete hardware and software stack that can accelerate deployment and reduce the risk of errors

The Oracle Optimized Solution for Enterprise Cloud Infrastructure is designed, deployed, and operated as a single system allowing customers to deploy applications in hours instead of days. Tests show drastic reductions in virtual machine deployment times, with significantly lower cost of acquisition, support and operating costs.

Complementary to the Oracle Optimized Solution for Enterprise Cloud Infrastructure, users can then simplify and accelerate application or database deployments using pre-defined, customizable Oracle VM Templates which contain ready-to-run software configurations that can be up and running in hours as opposed to weeks.

Oracle VM

To fully understand the rationale behind the described best practices and their implications, it is important to review some of the key concepts and components of the Oracle VM solution. For x86 servers, Oracle VM consists of Oracle VM Server for x86 and Oracle VM Manager.

Oracle VM Server for x86

Oracle VM Server for x86 installs directly on server hardware including Intel Xeon processors that support PAE (Physical Address Extension) and does not require a host operating system. It requires HVM (Hardware Virtual Machine) support (e.g. Intel® Virtualization Technology) on the underlying hardware platform in order to run HVM guests. The Oracle VM Agent for server management is installed with Oracle VM Server.

Oracle VM Manager

Oracle VM Manager is a Java-technology-based management server. It uses an Oracle database as its management repository, which can be installed either on the management server or a separate server. Oracle Database Express Edition (XE), Standard Edition (SE), Enterprise Edition (EE), and Real Application Clusters (RAC) are supported as the management repository. Oracle VM Manager manages the virtual machines running on Oracle VM Server through the Oracle VM Agent. For Intel Cloud Builders tests we have installed it as a VM, using the Oracle VM Manager Template.

Although not discussed in the Oracle Optimized Solution for Enterprise Cloud Infrastructure, Oracle VM Servers can alternatively be managed using the Oracle Enterprise Manager Grid Control Oracle VM Management Pack, which is licensed separately from Oracle VM. The Oracle VM Management Pack provides Oracle VM management functionality in a similar manner to Oracle VM Manager, but is integrated into Oracle Enterprise Manager Grid Control. This enables management and monitoring of Oracle VM environments alongside all other Oracle and third-party products managed by Enterprise Manager. This integrated management approach offers deeper insights into the health, performance, and configuration of the hardware and software stack.

Oracle VM software stack can be deployed in private and public clouds. This paper discusses how Oracle VM software stack is used to create a private cloud. Oracle customers can also run their production workloads in public clouds using Oracle VM. The public cloud support allows customers to add or shed resources as needed, paying only for resources used and freeing up scarce engineering resources from running technology infrastructure. Amazon EC2 now uses Oracle VM Server to run Oracle software in the public cloud. In September 2010 Amazon Web Services* and Oracle announced Certification and Support of Oracle Applications, Middleware, and Databases on Amazon Elastic Compute Cloud using Oracle VM.

Oracle VM Templates

Oracle VM Server provides the ability to rapidly and easily deploy a pre-built, pre-configured, pre-patched guest virtual machine (or multiple machines depending on the application). The guest VM can contain a complete software solution along with the operating system and related software infrastructure. These guest VMs, called Oracle VM Templates, are available from Oracle's E-Delivery Website and are ready to download and run. Already configured for production use, Oracle VM Templates can save users days or weeks learning to install and configure a sophisticated product such as Siebel CRM or Oracle Enterprise Manager Grid Control. Instead, users can simply download and start the VM to begin using it right away.

Within the downloadable Oracle VM Templates, Oracle software is configured in the same manner as if the software had been installed and patched manually. The package and patch inventories are completely standard and up-to-date so that no changes to operations procedures are required to maintain the instances over time. Accordingly, Oracle VM Templates can also be fully customized post-install and then re-saved as "golden image" Templates in Oracle VM.

Oracle VM Templates can serve as a user's enterprise deployment standard to minimize risks and variation across multiple instance deployments. An up-to-date list of available Oracle VM Templates can be found on the Oracle VM Templates Website.

Sun Blade 6000 Modular Systems

Sun Blade 6000 Modular Systems from Oracle provide an open modular architecture that is optimized for performance, efficiency, and density. With a choice of server modules including the latest Intel Xeon processors, organizations can select the platforms that best match their applications or existing infrastructure, without worrying about vendor lock-in. The blade servers also support a choice of operating systems as well as industry-standard PCI Express* (PCIe) I/O modules, expandable storage, transparent networking, and consolidated management.

Oracle blades are optimized to run Oracle VM, Oracle Linux, and Oracle Solaris*. The result is a modular IT infrastructure architecture that serves the needs of the data center and the goals of the business while protecting existing investments into the future.

Sun Blade X6270 M2 and X6275 M2 10GbE Server Modules

Within the Sun Blade 6000 Modular System Chassis, the recommended blade server modules for the Oracle Optimized Solution for Enterprise Cloud Infrastructure are the Sun Blade X6270 M2 and the dual-node X6275 M2 10 Gigabit Ethernet (10GbE) server modules.

Oracle's Sun Blade X6270 M2 server module is the best fit for mid-range virtualization workloads that require a balanced blade. It supports up to 144GB of main memory and up to 12 CPU cores with the high-speed six-core Intel Xeon processor 5600 series. The Sun Blade X6270 M2 server module offers balanced performance and superior energy efficiency while reducing the total cost and complexity of a virtualized infrastructure.

Sun Blade X6275 M2 10GbE server modules are ideal for large virtualization infrastructures that require the high density of a dual-node blade and support for 10 Gigabit Ethernet to reduce I/O bottlenecks. The Sun Blade X6275 M2 10GbE server module is Oracle's highest compute density blade server with support for a total of 24 CPU cores and up to 192GB of main memory in a single dual-node blade. Combining the Sun Blade X6275 M2 10GbE with the high performance, low latency Sun Blade 6000 Ethernet Switched Network Express Module (NEM) 24p 10GbE delivers the required bandwidth to eliminate potential network bottlenecks that can be experienced in virtualized environments.

Oracle Sun Blade 6000 Ethernet Switched NEM 24p 10GbE

The Oracle Sun Blade 6000 Ethernet Switched NEM 24p 10GbE simplifies data center deployments while simultaneously delivering the highest attainable throughput with 10GbE switching technology. With its combination of advanced cut-through switching, novel use of QSFP for uplinks, and line rate quality of service, the switched NEM 24p provides the ideal solution for virtualization and network storage. The solution scales to four chassis, offering 75 percent fewer cables and eliminating the need for expensive external switching.

In the Oracle Optimized Solution for Enterprise Cloud Infrastructure, the Sun Blade 6000 Modular System is equipped with a redundant pair of switched NEM 24p interconnects. Each is dedicated to one head of the Sun ZFS Storage Appliance cluster. This solution leverages the bandwidth of 10GbE while ensuring no single points of failure. It also offers numerous unused 10GbE uplinks per NEM for seamless integration into the data center.

The high performance and reduced cabling offered by the Sun Blade 6000 Ethernet Switched NEM 24p 10GbE make it ideal for the virtualized architecture of the Oracle Optimized Solution for Enterprise Cloud Infrastructure.

Key benefits include:

- Simplified network topology with embedded 10GbE switching, eliminating the need for external switching
- Reduced costs and simplified management with 4:1 cable reduction and simple scalability to four chassis
- Seamless integration into existing data centers with the industry's highest density uplinks

Sun ZFS Storage Appliances

Sun ZFS Storage Appliances provide all the benefits of unified storage in an easy-to-use appliance package to help organizations dramatically simplify their storage deployment and management while also reducing costs. These appliances change the economics of storage by using a high-performance Hybrid Storage Pool (HSP) architecture that combines the performance characteristics of SSDs with high-capacity hard disk drives (HDDs). The SSDs provide read and write cache, enabling higher performance than traditional storage architectures at up to 35 percent less cost.

Oracle's Unified Storage systems also provide unmatched simplicity and ease-of-use through an intuitive and powerful browser user interface (BUI) for simplified management. Revolutionary business analytics functionality allows administrators to quickly diagnose and resolve performance issues in production systems.

Fast I/O Throughput for Virtualized Environments

Sun ZFS Storage Appliances support many concurrent I/O threads due to an innovative architecture that leverages SSDs, multiple CPU cores, and 10 Gigabit

Ethernet connectivity. This high level of I/O throughput enables Sun ZFS Storage Appliances to support more virtual machines without sacrificing service levels.

While the appliances support many different storage I/O protocols including Common Internet File System (CIFS), Internet Small Computer System Interface (iSCSI), InfiniBand* (IB) Technology, and Fibre Channel (FC), NFS over Ethernet is recommended for attaching the storage in the Oracle Optimized Solution for Enterprise Cloud Infrastructure. This is due the extreme simplicity of NFS and the fact that it scales well on Sun ZFS Storage Appliances.

Flexibility Through Configuration Choices

To meet a variety of needs for capacity, price, and performance, Sun ZFS Storage Appliances are available in different configurations that include high availability (HA) cluster configurations for the Sun ZFS Storage Appliance 7320, 7420, and 7720 system models. In a clustered configuration, the head node resources failover to the surviving node in the event of node crash, thus helping to ensure near continuous availability. All systems come bundled with the same software including data protocols, replication, compression, and Oracle Solaris Dynamic Tracing (DTrace) analytics software for system troubleshooting and performance optimization.

Intel® Cloud Builders Usage Scenarios

Intel Cloud Builders test cases explore key aspects in the management and usage of a cloud computing environment. Oracle has implemented the Intel Cloud Builders test cases using Oracle VM Server to illustrate how Oracle VM can support these requirements. Oracle VM Manager was used throughout to manage the physical servers and virtual machines. The Intel Oracle Bid Support Centre uses Oracle VM Server and Oracle Secure Global Desktop to deliver a cloud computing environment to partners and customers across EMEA. Oracle Secure Global Desktop remote access to the Intel Cloud Builders test environment is illustrated as an example – this enables Oracle VMs to be accessed from anywhere.

Please refer to the Technical Review section for details of how the use cases have been implemented using Oracle VM.

Key Deployment Concepts

From a deployment perspective, multiple Oracle VM Servers are grouped into server pools as shown in Figure 2. Every server in a given pool has access to shared storage, which can be NFS, SAN (Fibre Channel), or iSCSI storage. This allows VMs associated with the pool to start and run on any physical server within the pool. Typically the server chosen is the server that has the most resources available, or a server that closely matches the resource requirements of the VM. Given the uniform access to shared storage mounted under the /OVS directory in which all resources will reside, VMs may also be securely Live Migrated or automatically started or restarted across any servers in the pool. In the Intel Cloud Builders test lab, Oracle's Sun ZFS Storage Appliances offer managed storage for Oracle VM Servers to share storage via iSCSI and OCFS2.

VMs are associated with a given server in the pool. This association is made dynamically at power-on and is based on load balancing algorithms or on a user-defined list of named servers called the Preferred Server list. The Preferred Server list identifies which servers are to be used as a host for specific VMs. When VMs are powered-off and not running, they are not associated with any particular physical server. Powered-off VMs simply reside in the shared pool storage.

As a result of this architecture, VMs can easily start-up, power-off, migrate, and/or restart without being blocked by the failure of any individual server or by the failure of multiple servers, as long as there are adequate resources in the pool to support the requirements for all VMs to run concurrently.

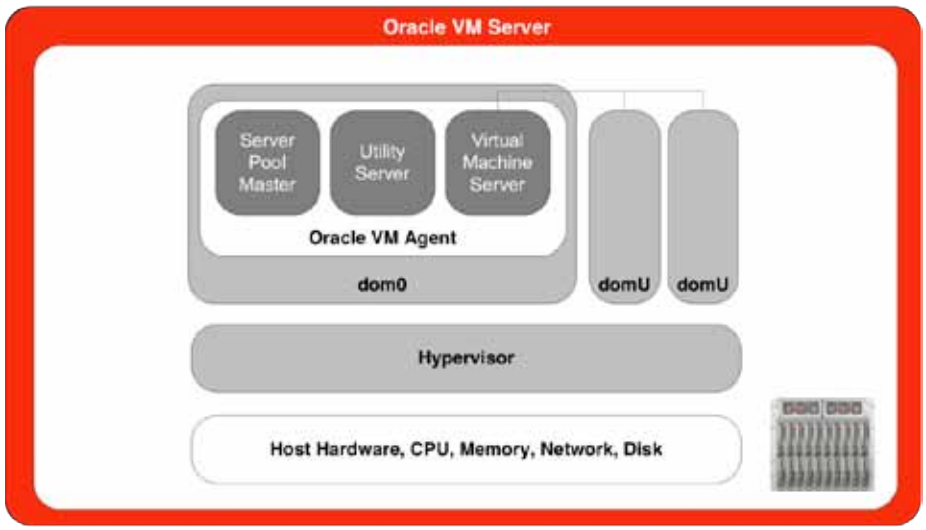


Figure 2: Oracle* VM Server Architecture

As shown in Figure 2, there are different roles of Oracle VM servers in the pool. These roles, which are defined below, are implemented by the agent running on the Oracle VM server. Multiple roles can be combined in a single server, but there is only one Server Pool Master in the server pool at any given time.

1. Server Pool Master

In each server pool, there is exactly one Server Pool Master at any given time. It coordinates a number of the activities of the pool, particularly when that action requires coordination across multiple servers.

This includes such key activities as coordinating Secure Live Migration actions as well as Guest VM HA (high availability) auto-restart, and power-on actions amongst others.

All of the Oracle VM agents in the pool communicate directly with the Server Pool Master, which, in turn, communicates with the Oracle VM Manager. This architecture provides a number of benefits including high management scalability in large environments as well as higher availability at the Oracle VM Manager instance level because functionality is distributed and isolated to minimize the impact of any single failure. For example, a management server outage does not prevent the ability of the pool master(s) to complete Secure Live Migration tasks or to automatically fail over/restart failed VMs.

Similarly, a pool master outage on one pool does not affect the operation of another pool. By default, the first server added to a pool is assigned the role of the Server Pool Master. With Oracle VM 2.2, the Server Pool Master can perform auto-failover when a server pool is HA enabled and a virtual IP address is set for the server pool.

2. Utility Server

In each server pool, there can be one or multiple servers designated with the Utility Server role. Utility Servers are responsible for performing resource-intensive copy or transfer activities in the pool. For example, cloning activities, creating VMs from Oracle VM Templates, or saving existing VMs as templates all involve either copying an existing VM image or moving an image from one directory to another. Depending on the size of the files involved, this activity can be resource-intensive. It may be desirable to off-load this activity from servers that are hosting production VMs so as to minimize or eliminate any service level impact. Multiple servers in a pool can be designated as Utility Servers to provide better load balancing and availability as described later in this paper.

3. Virtual Machine Server

Virtual Machine Servers are simply servers that host VMs. At least one server in a pool must be a Virtual Machine Server. Depending on total server capacity requirements, it may or may not be separate from the Server Pool Master and/or the Utility Server(s) in the pool as described later in this document.

4. Server Pool Planning

There are a large number of considerations when planning the virtual infrastructure and one size does not fit all. This section provides some considerations and guidelines to help develop a plan that is well suited to an organization's unique requirements.

It may be helpful to think of the server pool as if it were one big server with an aggregate amount of CPU, memory, storage, and network bandwidth. As such, planning for deploying VMs into a pool is much like planning for a server consolidation. It involves deciding how much aggregate capacity is needed to support normal and peak workloads as well as what types of workloads are appropriate to share to the pool or server. Workload profiles should be considered in addition to how predictable or unpredictable the workloads may be.

There are also some significant similarities between server pool planning and physical server planning with regards to node (physical server) size versus overall pool size. For example, in some cases it is better to have relatively fewer but larger servers in a pool. In other cases, a greater number of relatively small servers or blades are a better fit. Both deployments may provide the same aggregate CPU, memory, storage, and bandwidth but the implications of the deployment in a pool can be different.

For Oracle VM 2.2 and later, each server pool must have its own shared storage resources that can be accessed by Oracle VM servers within the same pool. A separate server pool must have its own separate shared storage.

5. High Availability Planning

Oracle VM provides the following features to provide maximum up time for VMs running in server pools:

- Guest VM HA — Auto-restart on server or VM failure.
- Secure Live Migration — Move VMs off of servers that are undergoing planned maintenance.
- Automatic pool load balancing on VM start-up — At VM power-on, an algorithm dynamically assigns the host server for the VM in order to load balance, but also to avoid a down server blocking VM start.

Figure 3 highlights how these availability features are typically deployed. The features should be used to collectively

maximize up time of guest VMs running in a pool. The details of these features and capabilities are the subject of another white paper titled, "Oracle VM – Creating & Maintaining a Highly Available Environment for Guest VMs".

The following section summarizes best practices and considerations for planning server pools.

6. Recommendations for Deploying Highly Available Server Pools

The following steps are recommended to achieve high availability:

- Enable the "High Availability (HA)" option at both the server pool level and individual VM level to ensure VMs are automatically restarted after an unplanned failure. If the Server Pool Master fails in a High Availability set up of Oracle VM, another Oracle VM Server is automatically selected from the server pool to act as the Server Pool Master.
- Plan to use the Secure Live Migration feature to migrate VMs in support of planned events like server maintenance to prevent any service outage.
- Plan for enough excess capacity in aggregate across the pool to support running all VMs to an appropriate service level even when one or more servers in the pool are out of service. Up to 32 nodes are supported within a server pool.
- Plan for multiple server pools when there is a need for more than 32 physical servers.
- Enable HA and set up virtual IP (VIP) for the server pool to take advantage of the Server Pool Master auto-failover to minimize risk and thus maximize availability of the management services.
- Configure two or more dedicated and load-balancing Utility Servers to provide continuity of service in the event of a utility server outage. There can be multiple Utility Servers in a pool and Utility Servers are automatically load balancing. If one Utility Server fails, the remaining Utility Server(s) can continue without any disruption of utility services (VM cloning, VM import, Create VM from Templates, etc.).

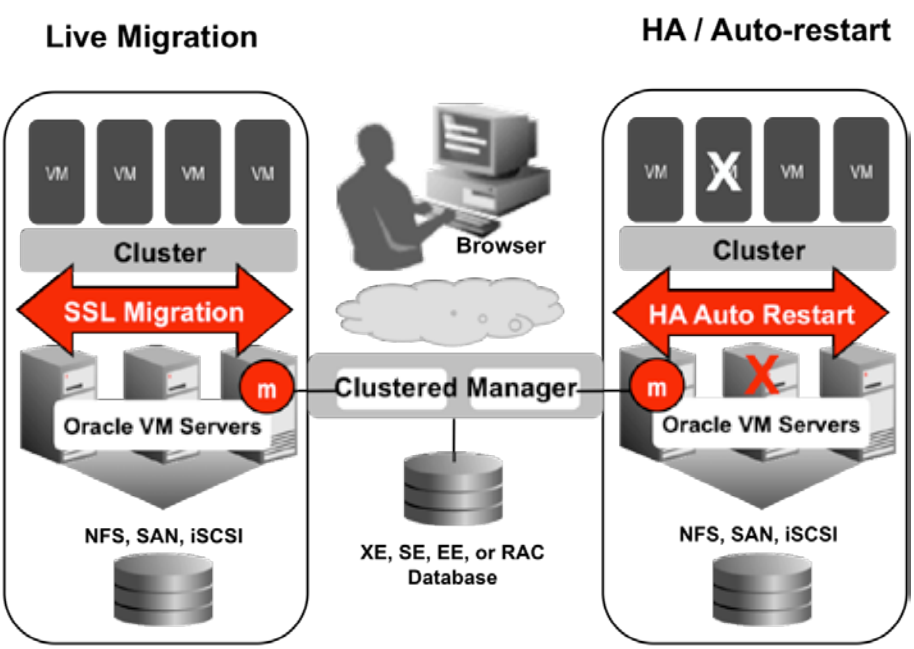


Figure 3: High availability helps minimize downtime due to planned or unplanned events

7. Pool Capacity, Performance, and Scalability

Capacity planning for a server pool is similar to capacity planning for a physical server. However, the following additional considerations are also important when planning capacity for a server pool:

Plan extra capacity to support Guest HA/ auto-restart. There should be sufficient capacity to support hosting additional VMs on relatively fewer machines in the event that one or more of the servers fails and its VMs end up being restarted on the remaining, healthy servers, if only temporarily.

Plan for extra capacity to support Live VM Migration during planned events. When performing maintenance on a server (or servers) in the pool, Live Migration allows administrators to migrate VM(s) to another server in the pool without interrupting service. To take advantage of this capability, there should be enough excess capacity in aggregate across the pool so that a server can be taken offline (after migrating its VMs) without inappropriately impacting service levels.

8. Determining How Many Servers or VMs a Pool Should Contain

Before creating a server pool, you need to consider how many physical Oracle VM Servers are to be included in the server pool, and what functions(s) each physical Oracle VM Server is to perform. The more guest virtual machines you run in the server pool, the more resources these guest virtual machines will consume, and so the more physical Oracle VM servers may be required to provide sufficient resources for the server pool.

A server pool is scalable. If you find a server pool does not have sufficient resources, such as CPU or memory, to run the virtual machines, you can expand the server pool by adding more Virtual Machine Servers. The number of servers or VMs that are ideal for a pool depends on a number of factors that can vary greatly between data centers and deployments. There is no single correct answer to this question, but there are several factors that should influence such decisions. Some considerations are described in the remaining points below.

9. Storage Topologies, Performance, and Implementation

Oracle VM pools require that all servers in a pool have shared access to the same storage so that VMs can be moved around easily. This means that server pools must use shared storage such as NFS, OCFS2 (Oracle's Cluster File System), or SAN (iSCSI or FC) storage. In the Oracle Optimized Solution for Enterprise Cloud Infrastructure, Oracle's Sun ZFS Storage Appliances are used to provide NFS or OCFS2 shared storage. Both the physical make-up of the storage devices and the scalability of the file systems used will dictate how many servers are practical for a given shared-storage pool without adversely affecting I/O performance.

When evaluating how many servers can share a given storage topology, the following questions should be considered:

- How much I/O will each server generate and can the throughput and latency needs be accommodated through the designated NIC or HBA ports?
- How much I/O can the storage device or devices support?
- What are the HA needs (bonding or multipath)?
- Are there any application requirements for directly accessed storage?

Of course, the answers to these storage questions depend on the I/O environment. Is the application I/O intensive? What is the average size of an I/O request?

Determining a realistic number of storage nodes in a storage cluster requires considering similar questions:

- How much I/O will each storage cluster node support?
- How much I/O will each server generate to/from the clustered file system?
- How many servers will be accessing the clustered file system?

10. Workload Profile

Is the workload flat and stable or variable and peaky? The ideal VM has relatively low utilization and is very flat and stable with minimal peaks. These types of VMs can often be very tightly consolidated, with a large number of VMs per server. Since they are very predictable they require little excess capacity or headroom to be available to accommodate unexpected peaks.

The next best scenario is when consolidating multiple VMs that may have peaks, but the peaks are very predictable in both timing and magnitude. For instance, some VMs may contain applications that always peak at the end of the week or end of the month. If those VMs could be consolidated with other VMs that peak at exactly the opposite time – workloads that peak at the beginning of the week or month– they can potentially be packed fairly tightly to maximize the number of VMs per server and per pool.

The worst case scenario is when the VMs are highly variable in load and in timing. In this situation, it is likely that a comparatively large amount of extra headroom will be needed on the servers. Thus fewer VMs will be able to be accommodated per server and per pool.

11. Service Level Support Strategy

Sometimes service level objectives may dictate that there is enough planned excess capacity to support normal service levels even if all the VMs peak at the same time. This is the most conservative option, but also the most expensive since it requires extra hardware that may not be utilized much of the time.

Another alternative is to plan for the average load and accept any performance hit based on resource contention if there is too much of a peak. This certainly reduces hardware expense, but may not provide acceptable service levels if the workloads are too unpredictable or are mission critical. As a result, many data centers plan their capacity to support some percentage of the aggregate peak load. For example, they may plan for 40-60 percent of the peak above average. This is often a good compromise between meeting service levels and having reasonable hardware utilization. However, it clearly depends on how critical service levels are for the given workloads.

12. Dynamic Server Pools

Understanding that the server pool is dynamic, keep in mind that VMs are not associated with any one physical server until they are powered-on and placed on a server. Typically they are placed on the server that currently has the most memory free. Based on HA events or Live Migrations, VMs can be moved around within the pool. Thus a given VM can end-up sharing a server with any other VM in the pool unless policies are implemented to restrict which servers can host which VMs. This means that the capacity plan really needs to be at the pool level, not the individual server level. So it may be best to consider keeping

highly volatile VMs in their own pool or restricting them to a subset of the pool where a relatively large amount of excess capacity can be maintained for handling unpredictable peaks.

Conversely, highly predictable VMs should be restricted to a separate pool where the resources can be very tightly planned for high utilization without the need for much excess capacity.

13. Server Affinity Policies

Consider the impact of VM placement. Oracle VM allows administrators to implement policies on an individual VM basis to restrict which servers are allowed to host that particular VM. Preferred Server lists are used to implement these policies. They contain a list of explicitly named servers that are allowed to host specific guest VMs within a server pool.

Preferred Server policies are typically implemented to assure that two components of the same application stack are not hosted on the same server in order to maximize availability. These policies are respected for all actions within the pool, including Live Migrations, Power-On, and HA/Auto-Restart. One consideration in using Preferred Server policies is that it can have the effect of increasing the number of servers required in the server pool to support HA and Live Migration.

Additionally, for any server in the pool, the Virtualization Server role can be removed to prevent that server from hosting any VMs, although this is typically only done to create dedicated Utility Servers. To give a simplistic example, when deploying a VM application with two clustered components, it is a good idea to create a Preferred Server policy for each VM. Such a policy assures that both of the nodes would never be on the same physical server at the same time. Yet it is still important to make sure there is enough capacity to fail-over and to generally support peak loads. It is also important to have the option to Live Migrate either component (VM) to another server to support server maintenance without any application down time. In this simplistic case, there would need to be a minimum of four nodes since each of the two VMs must have two servers available. The first server would act as the primary server for running the first VM and another server would be required for migrating or re-starting the first VM in the event of a failure. The second VM would similarly require its own two servers for primary and migration/re-start. If the restriction that both VMs can never operate on the same server at the same time is lifted, the two VMs could be hosted on just two servers and still have HA and migration support.

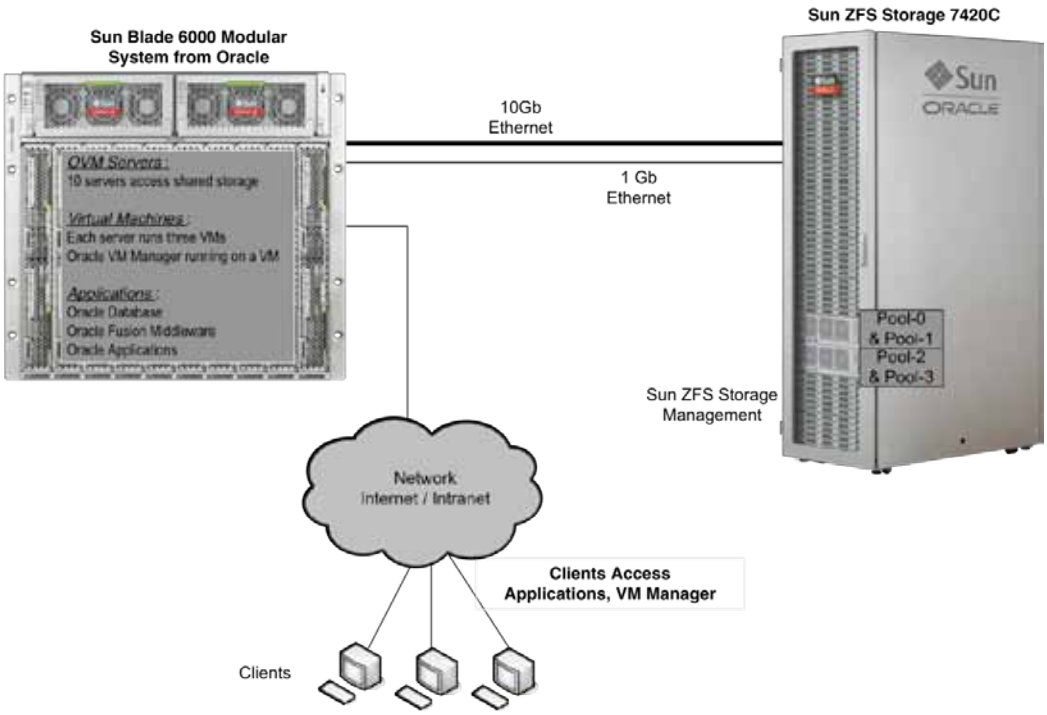


Figure 4: Oracle* Optimized Solution for Enterprise Cloud Infrastructure

Test Bed Blueprint Overview

Test Bed Design Considerations

This section shows an Oracle VM deployment scenario using Sun Blade 6000 and Oracle's Sun ZFS Storage Appliances. It's a typical enterprise deployment model using high-end Intel Xeon processor based servers and storage.

The Oracle Optimized Solution for Enterprise Cloud Infrastructure illustrated in Figure 4 offers high performance, availability, reliability, and manageability and is specifically intended for database virtual machines. While these requirements can be met in numerous ways, this reference architecture implements an optimal approach. The Intel Cloud Builders test lab environment closely follows these guidelines as explained in detail below.

The following sections detail the server components, virtual machine allocation, storage configuration, and project allocation with Oracle's Sun ZFS Storage Appliance storage and network configuration.

Hardware Description

Both blade-based and rack-mounted Oracle Sun x86 servers using Intel Xeon processors can be used to run the Oracle VM infrastructure in a large, high availability configuration. As an example, Oracle's Sun Blade 6000 Modular system with 10 Sun Blade X6270 M2 or 5 Sun Blade X6275 M2 server modules can be configured to run the Oracle VM software. Each server is attached to a clustered Intel® Storage Server that is accessed using NFS, or via FC/iSCSI LUN. This approach enables additional features such as live migration, high availability, distributed resources scheduling to be used for the VMs in the server pool. Sets of servers that are identified as part of a single server pool are recommended to be configured in a similar way so that the service level agreements can be met during failovers or migration.

Each Sun Blade server module must be fitted with a Sun Intel 32599 Dual 10GbE PCIe 2.0 Fabric Expansion Module (FEM). The FEM supplies each server with two shared 10GbE ports. The two 10GbE interfaces are used by VMs to access NFS services on the Oracle's Sun ZFS Storage Appliances. Two 10GbE interfaces supply plenty of bandwidth and cable aggregation for virtual machine data. For NFS and iSCSI access, 10GbE ports are also used.

Intel® Cloud Builders Lab Hardware

For the Intel Cloud Builders lab we utilized five Sun Blade X6270 M2 server modules, plus a single Sun ZFS Storage 7420 Appliance for Oracle VM shared storage, hosted at the Oracle Solution Centre at Oracle, Scotland.

The five Sun Blade X6270 M2 server modules were each configured with two Intel® Xeon® processor X5680, 72GB memory, 2 x 300GB 10K RPM internal SAS-2 disks.

The Oracle VM Pool Master and Utility Server were assigned to the first Sun Blade X6270 M2 server module in the Sun Blade 6000 chassis, and all five Oracle Sun Blade X6270 M2 server modules were used as Oracle VM Server Virtual Machine Servers, hosting VMs.

Oracle's Sun Blade 6000 modular system is ideal for virtualized application environments and application infrastructure products. Powered by the six-core Intel Xeon processor 5600 series, they leverage the flexible I/O architecture of the Sun Blade 6000 chassis, making it the easiest blade system to deploy. Built on Intel® microarchitecture, codenamed Nehalem the Intel® Xeon processor 5600 series expands the benefits of virtualization beyond consolidation with innovations that can help boost performance, increase consolidation ratios, and enable servers of different generations to be combined in the same virtualized server pool, improving virtual machine failover, load balancing, and disaster recovery capabilities, making it an excellent choice for virtualization based cloud deployments.

For the main Intel Cloud Builders tests we used the Sun ZFS Storage 7420 Appliance to provide shared storage for Oracle VM, enabling High Availability and Live Migration functionality for the VMs.

We implemented equivalent infrastructure at both the Intel Oracle Bid Support Centre (Intel, Ireland) and the Oracle Solution Centre (Oracle, Scotland). At Oracle we used the OSC's existing shared Oracle Secure Global Desktop deployment, whereas at the BSC we installed our own. In a production environment this would either be replicated or configured as high availability VMs depending on requirements. In both cases our software installation and configuration work was done remotely, using the browser based administration consoles of the blade server chassis, and SGD.

The Intel Cloud Builders lab environment has 1Gb but not 10Gb networking, and five blade servers.

Virtual Machine Configurations

For each virtual machine, the number of virtual CPUs and memory can be configured to address the specific requirements of the applications that are deployed. In the lab setup, a minimum of 4GB memory and two virtual CPUs are configured for each of the middleware and application virtual machines. For the virtual machine that hosts the database, 16GB of memory and four virtual CPUs are allocated. Note that the number of virtual CPUs can be tuned in a live system.

Choosing the number of virtual machines per VM server as well as the type and number of network links are also critical considerations for a successful deployment. For example, for certain I/O loads over a single 1GbE link, the network could be saturated first before either the server or the storage. In such cases, adding more links or moving to one or more 10GbE links may provide a better performance. If the CPUs are pegged running multiple virtual machines on the same host, then either add additional servers to the server pool and live migrate the virtual machines to balance the Pool, or move the Oracle VM infrastructure from the midrange server to an enterprise class server.

For high availability and optimal performance an Intel architecture based Sun ZFS Storage 7420 Appliance is used as the storage platform. The servers are multi-core and the number of CPU cores dictates the number of VMs that can be assigned to each server.

Table 1 shows the recommended configuration of four distinct storage pools and the purpose of each pool. Further details about the recommended pools are provided below.

Each storage head is set to be active for two pools. This enables both load-sharing and high availability for all storage pools. In the event of a failure in HEAD-1, HEAD-2 will take over ownership of pool-0 and pool-1 as well as continuing to serve the clients.

- Pool-0 — One Oracle VM project is created in pool-0 to store Oracle VM storage repositories. This is accessed from the dom0 of the Oracle VM server.
- Pool-1 — One project is created to share the ORACLE_HOME across the various database instances. Additionally, one project per database is created. Compared to the rest of the pools, a greater number of disks are to be allocated for this pool. This is because the random reads (db sequential reads) benefit from more disks during scenarios when there is a read miss from the cache.
- Pool-2 — One or more projects are created for storing the various middleware components. Projects might be dedicated to items such as Web server, SOA binaries, Tlogs, admin binaries and so on.
- Pool-3 — One or more projects are created for storing the various application binaries and configurations for the Oracle E-Business Suite, Oracle Siebel, Oracle Peoplesoft and so on.

The shares (file systems) created under these projects are mounted from the various virtual machines. For security purposes, access to certain projects and file systems can be restricted to specific clients (virtual machines).

Storage Configuration for High Performance with Oracle VM				
Pool Name	**Raid configuration**	**SUN ZFS Storage 7420 Cluster Head**	**System / VM**	**Purpose**
pool-0	RAID-Z2	HEAD-1	Dom0	For storing the Oracle VM storage repository. This pool is accessed from domain 0 and contains the OS images from which the virtual machines are launched.
pool-1	MIRRORED	HEAD-1	Virtual machines that run the database	For storing database files and database binaries. One or more projects are created to cater to each database instance.
pool-2	MIRRORED or RAID-Z2	HEAD-2	Virtual machines that run the Oracle Fusion Middleware	For storing the middleware components, including binaries and configuration files.
pool-3	MIRRORED or RAID-Z2	HEAD-2	Virtual machines that run the Oracle Applications	For storing the application layer components – including the binaries, configuration files.

Table 1

Software Description

The following software was installed in the Intel Cloud Builders test lab environment:

Technical Review

Intel® Cloud Builders Use Case Details

Oracle implemented the Intel Cloud Builders use cases to demonstrate how Oracle VM Server and Oracle Secure Global Desktop can meet the real world tasks required to deploy a scalable private cloud. The use cases were completed by using standard Oracle products without modification or extension:

- Create users, user group for business unit, and business unit administrator accounts
- Register and configure business unit resources (inc. physical servers)
- Create new users and view users
- Configure and use virtual machine templates
- View and modify virtual machine network settings
- Create services by virtual machine creation
- Scale out service by cloning virtual machines
- Add additional physical servers to the business unit cluster
- Permit user access to virtual machines through portal
- Monitor the state of services and virtual machines
- Remove service by terminating and optionally deleting virtual machines
- Automated failover to deal with hardware failure
- Handling requests to scale deployed services
- Decommissioning services and virtual machines
- Distributed application (data analysis) test by using multiple VMs

Intel® Cloud Builders Test Lab Environment Installed Software	
System Component	**Software Component**
Virtualization Platform	Oracle VM Server 2.2.1 x86-64 ▪ Download here ▪ Oracle VM is server virtualization software that fully supports both Oracle and non-Oracle applications and delivers more efficient performance
VM & Deployment Management	Oracle VM Manager Template v2.2 ▪ Download here ▪ Installed as a VM for Intel Cloud Builders, but can be installed natively (includes Oracle Express as configuration DB)
Remote Access	Oracle Secure Global Desktop 4.6 ▪ Download here (previously Sun Secure Global Desktop) ▪ Oracle Secure Global Desktop is utilized to allow remote desktop access to the lab infrastructure, Oracle VM Manager and the virtual machines themselves
VM Operating System for Compute Nodes	Oracle Linux* 5u5 – using Oracle VM Template OEL5u5 x86-64 PVM ▪ Download here ▪ Important note: Other Operating Systems can be installed
Storage Node	Sun ZFS Storage 7420 Appliance version '2010.08-17.1.1,1-1,16'.

Table 2

Assumptions

1. Oracle VM components (Server and Manager) have been installed correctly
2. Any required Oracle VM Templates have been downloaded and imported using Oracle VM Manager (Oracle Linux PVM template is used in the example)
3. An Oracle VM administrator is available to create the Business Unit Admin accounts as illustrated in the test cases (or the Oracle VM administrator credentials are available)

Tests and Results

Intel® Cloud Builders Use Case Summary

Create Business Unit Admin Accounts

Login to the Oracle VM Manager Web interface using the URL set when Oracle VM Manager was installed.

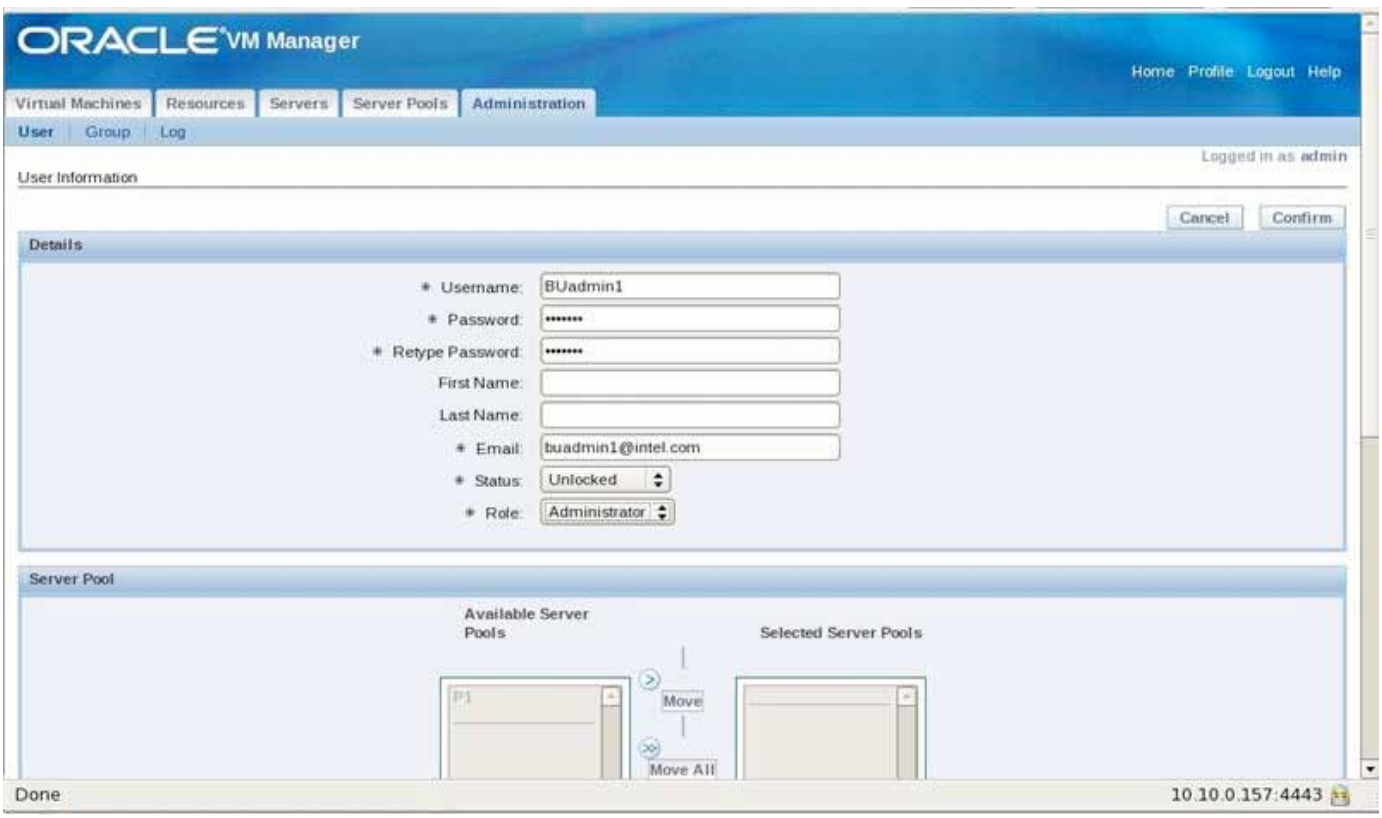
e.g. https://ovmm.#####.com:8888/OVS default user "admin", password set during install process of OVM Manager. You can also access by the IP of your Oracle VM Manager instance, instead of by name.

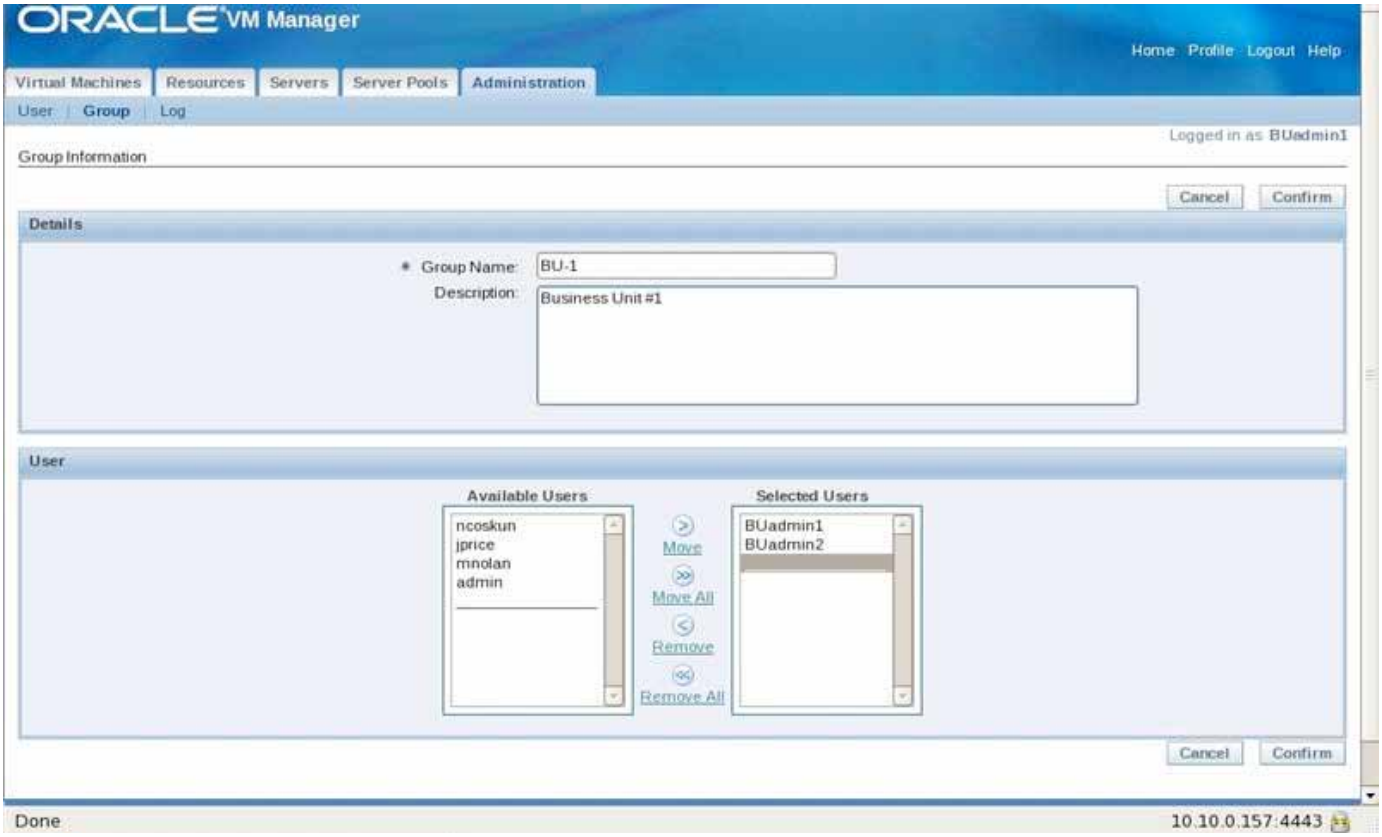
Click **Administration** tab, then **User**. Click **Create** to create a new user.

Complete new user details for each required BU admin user.

Create User Group for Business Unit Users

We can also create a user group, BU-1, for our business unit and put our new BUadmin users in it.

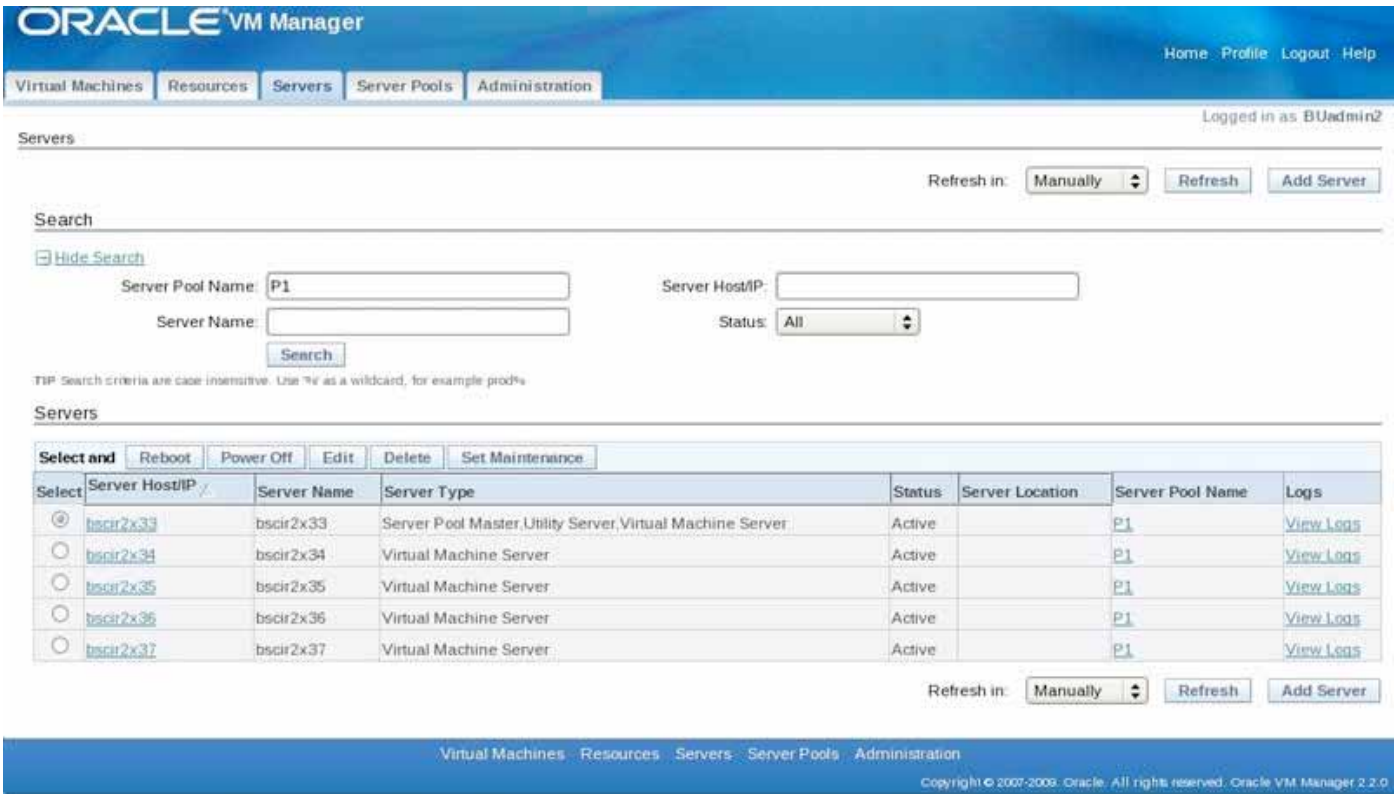



ORACLE VM Manager
Home Profile Logout Help
Virtual Machines
Resources
Servers
Server Pools
Administration
User | Group | Log
Logged in as BUadmin1
Group Information
Cancel
Confirm
Details
* Group Name: BU-1
Description: Business Unit #1
User
Available Users
ncoskun
jprice
mnolan
admin
Move
Move All
Remove
Remove All
Selected Users
BUadmin1
BUadmin2
Cancel
Confirm
Done
10.10.0.157:4443

Register/Configure BU Cluster Resources

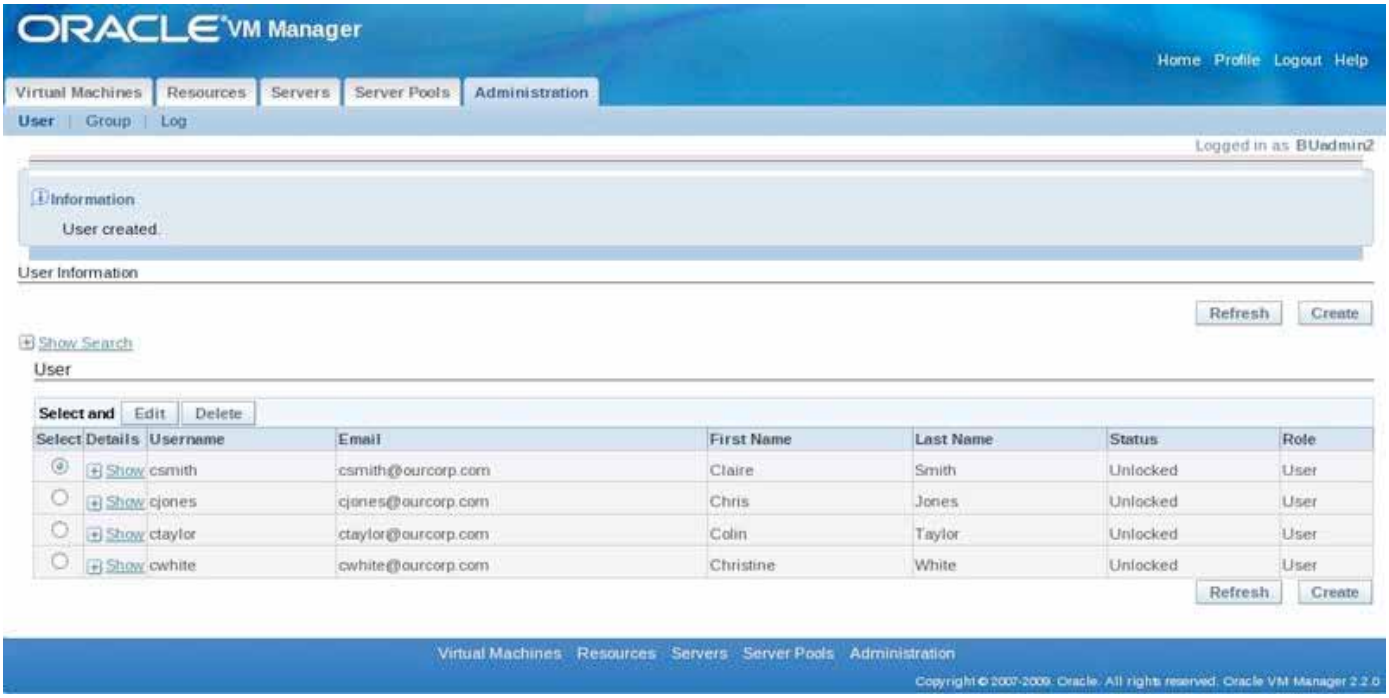
Now, logging in as one of the BUadmin accounts just created, **BUadmin2**, we can see that the BU admin can access the physical server resources of the server pool, **P1**, which has been assigned to the BU. From here servers can be added to, or removed from the pool. Servers can be shutdown, rebooted, or put into maintenance mode.

In the screen below, you can see that bscir2x33 is running additional components, Server Pool Master and Utility Server, to support the OVM Server Pool. In our lab five blades have been added to pool **P1**. Servers can be added and removed as required. The BUadmins can create new VMs, add and remove CPU cores, and virtual local and shared disks to the VMs as required. VM settings including network config, scheduling priority can be defined, VMs may be reallocated, live migrated to other servers, set for HA, cloned, etc.



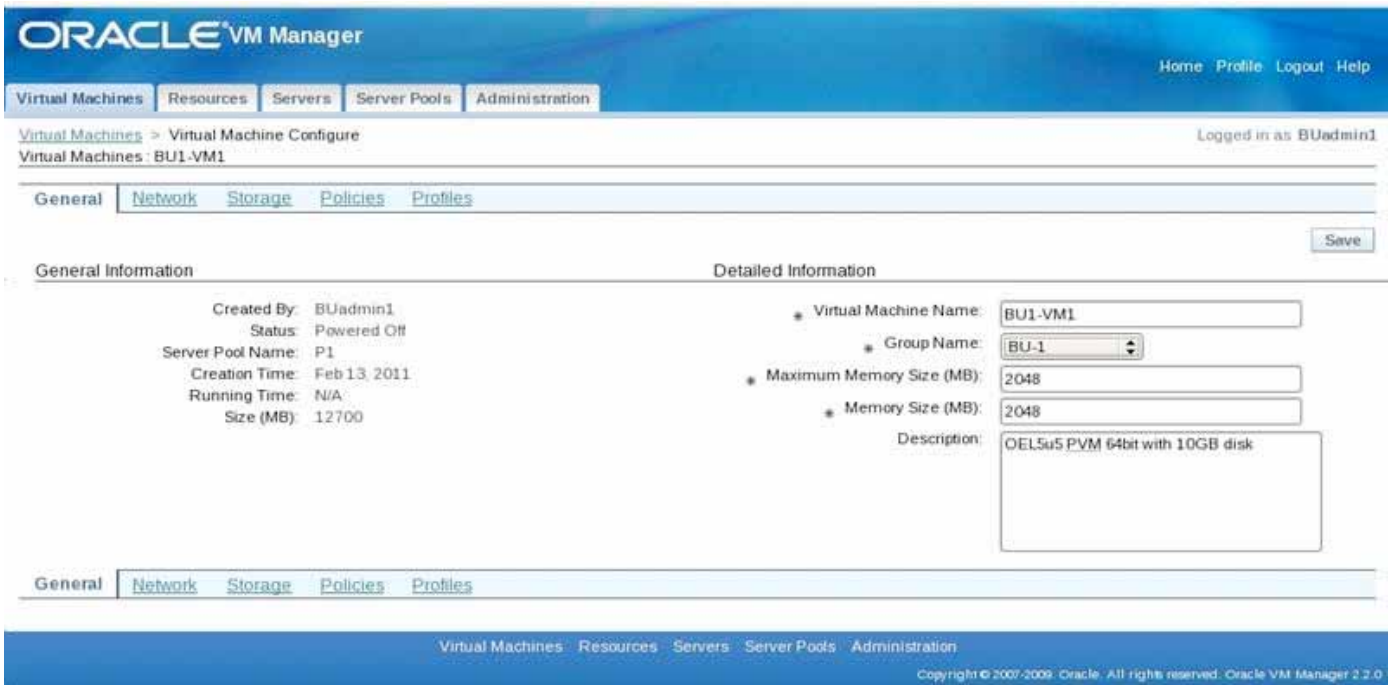
Create New Users/View Users

Both BUadmin1 and BUadmin2 can create new users as necessary. Here we create four new users to illustrate, one at a time, from Administration > Users. After creation the user list—filtered with "c%" in the search field— looks like this:

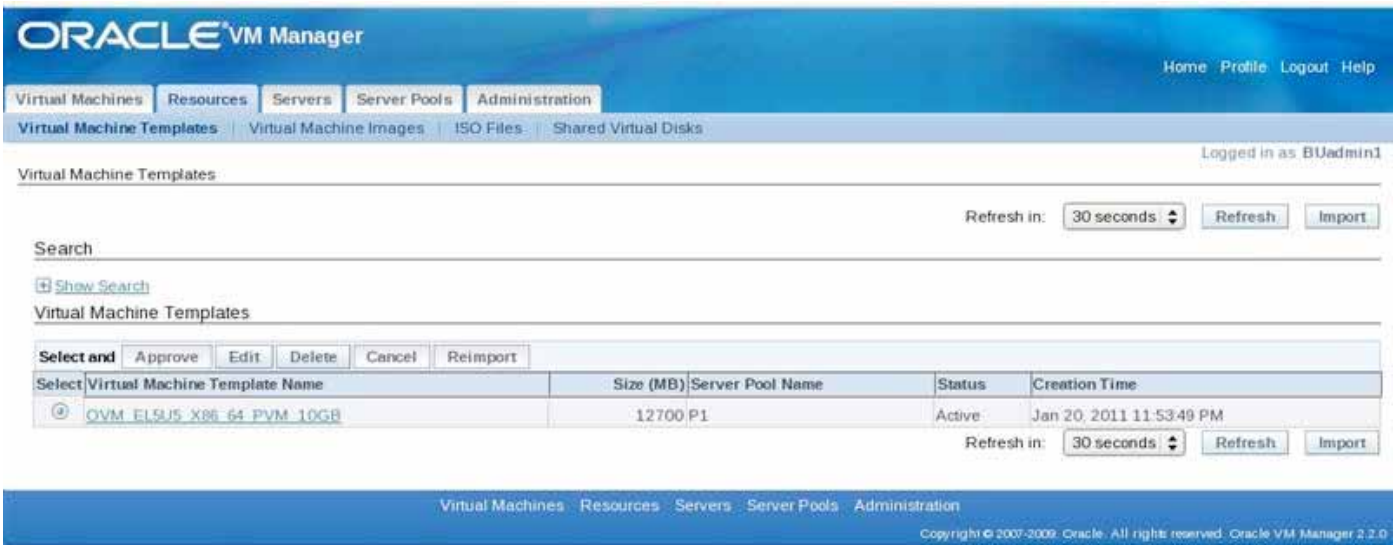


Configure Virtual Machine Templates

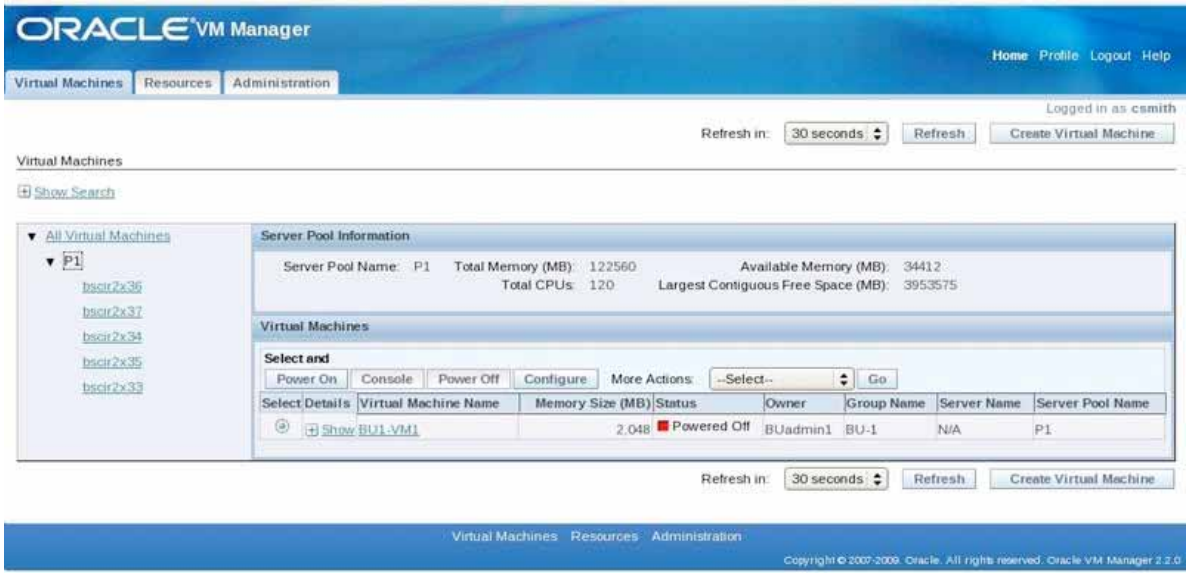
The BU admins can create new VMs, either by importing existing Virtual machines from elsewhere, or by creating new VMs from Oracle VM templates. VM templates for Oracle products may be downloaded directly from Oracle, or may be produced by developers to build a local library. They allow new VM instances to be created easily by admins from a standardised VM configuration (template). In this case we have produced a new VM **BU1-VM1** from the Oracle Enterprise Linux VM template that has been imported into our Oracle VM installation. Note that we assigned the new VM instance to group BU-1 when we created it, making it available to users in the **BU-1** group, such as **csmith**.



BU admins can import templates and make them available to users of a server pool.



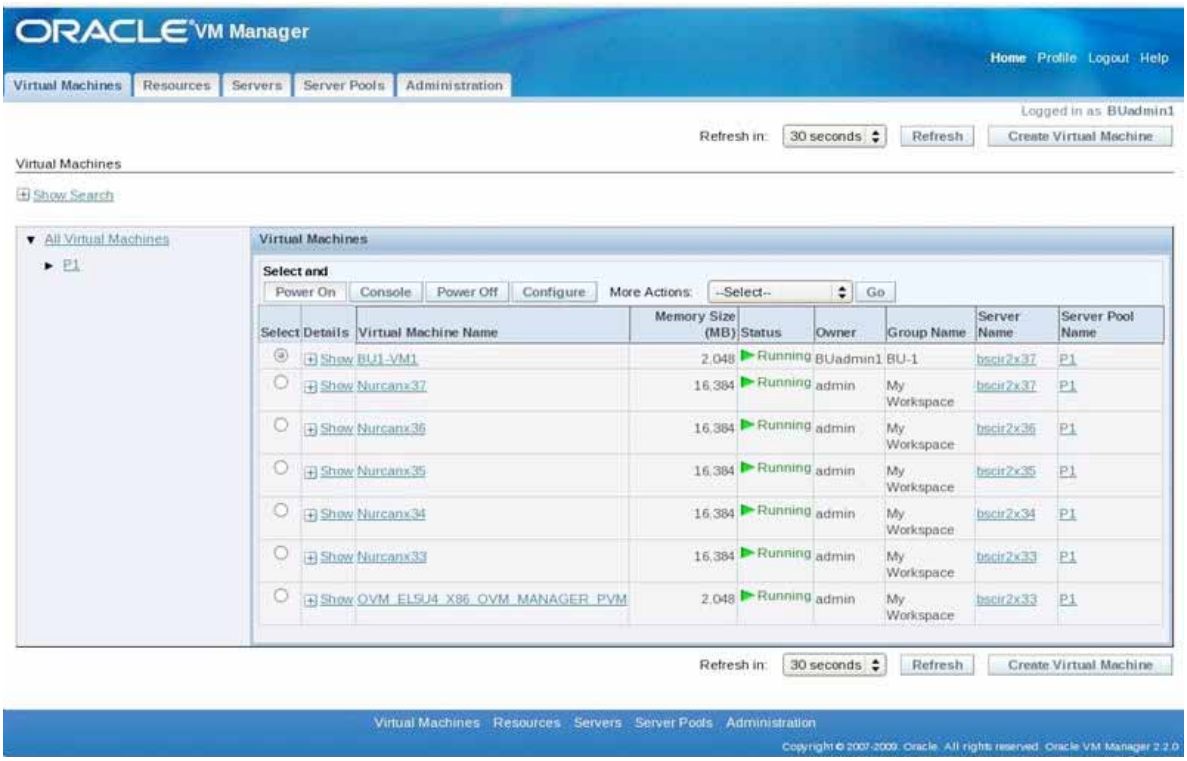
Our end users can login and see the VMs that we created for the BU, BU-1. Here user **csmith** sees the VM created by the BU admins.



Configure Network

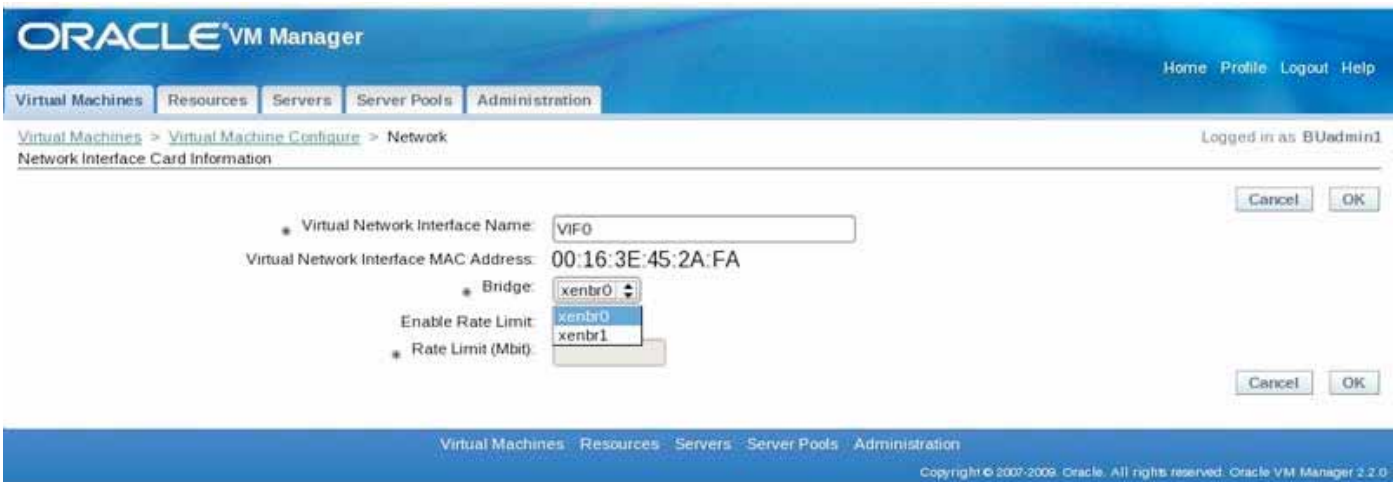
BU admins can see the status of current VMs, modify servers, and reconfigure network settings for the physical servers in their pool(s) and their VMs.

Here we select the BU1-VM1 VM.



Clicking on the network tab for the VM shows current network details for the VM. Network interfaces may be added, deleted, and modified for the VM:

Here we change the Xen* network bridge for example. Additional networks can be added as necessary.



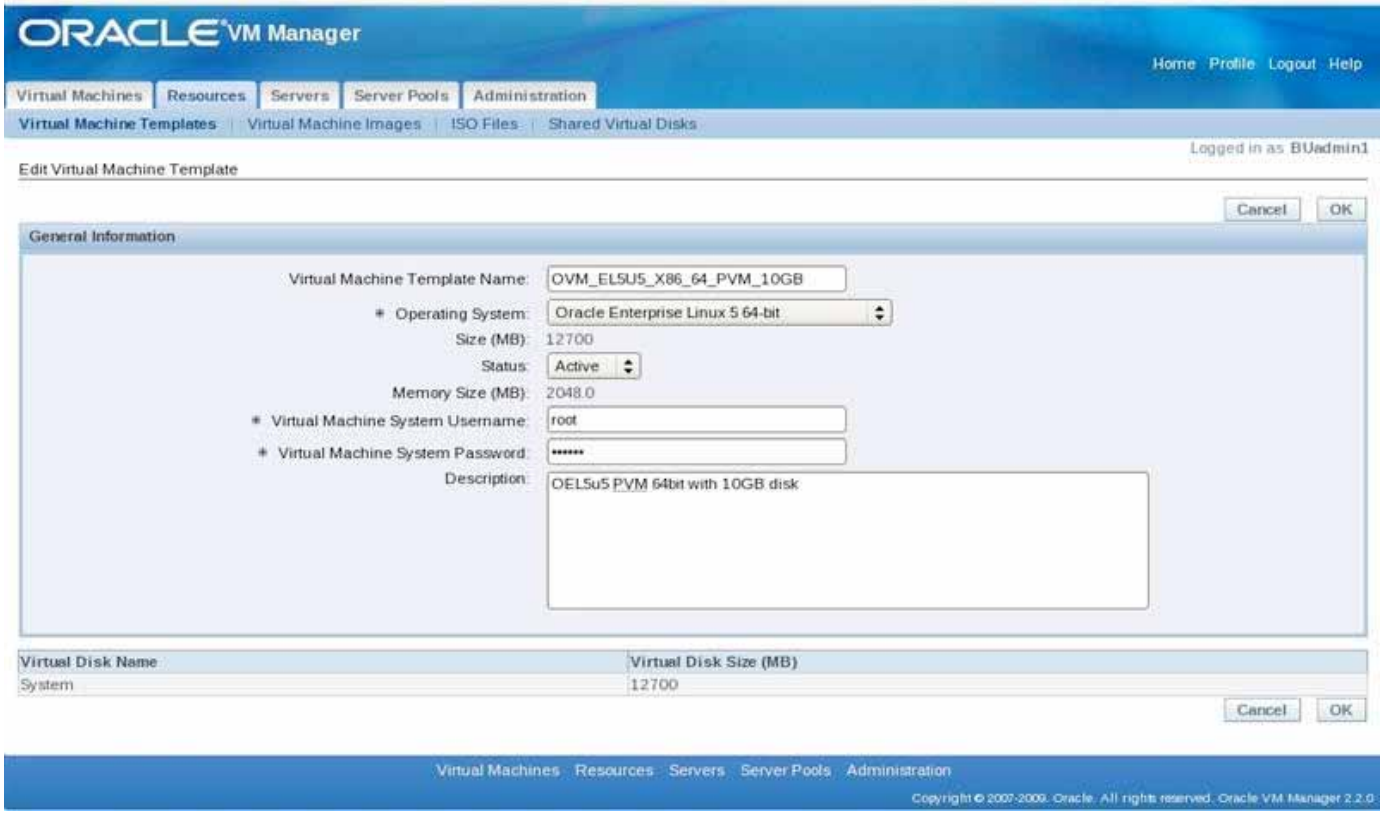
We can change the number of virtual CPUs (cores) assigned to the VM, relative scheduling priority compared with other VMs , scheduling cap, and whether the VM is configured for HA or not, amongst other settings.

Cost management facilities for VMs are planned for future release.

IP settings (including static, DHCP, gateway, DNS, etc.) can be configured to be set at first startup of the VM by definition within VM template scripts, or may be modified within the VM itself.

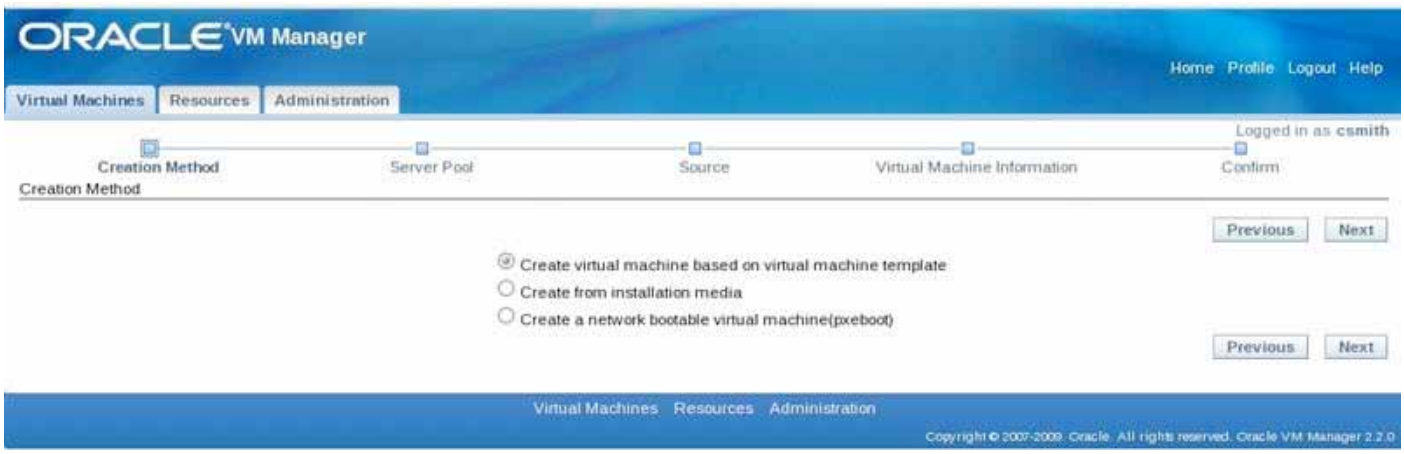
Create Service Instance (instantiate VMs)

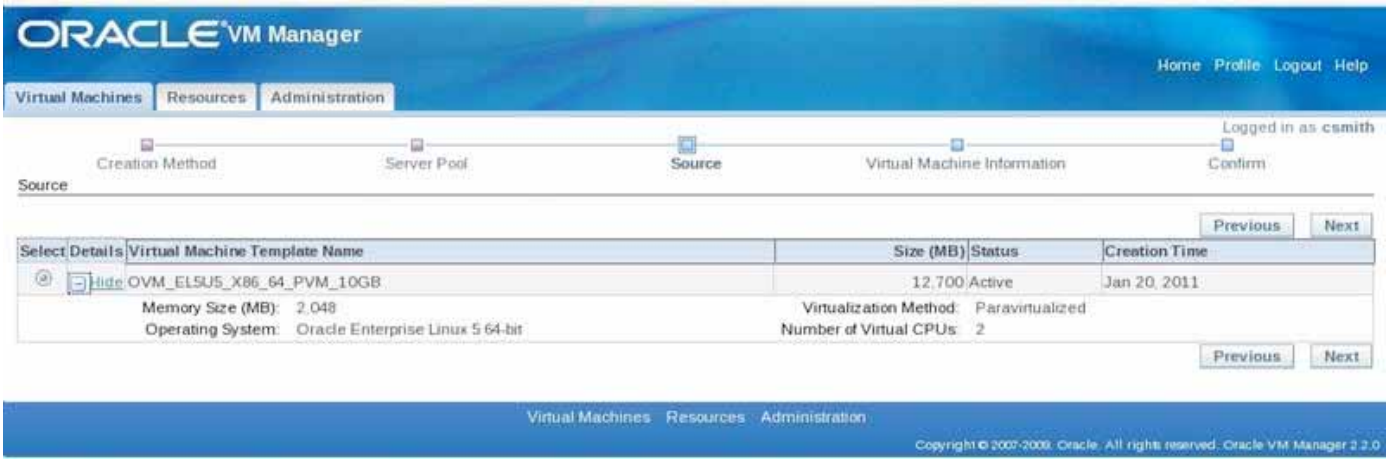
BU admins can examine available VM templates, import, edit, activate, and deactivate others.

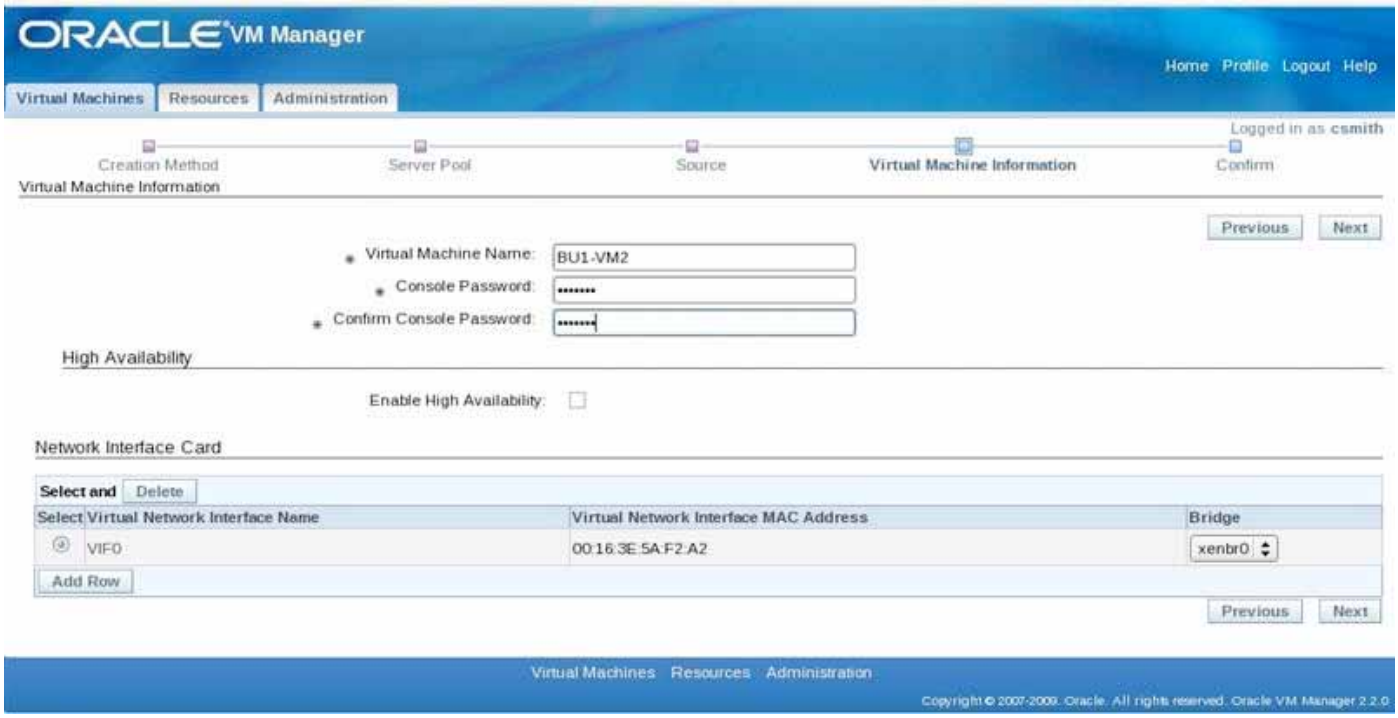


Here we create a new VM from a template by clicking "Create Virtual Machine", including specifying create from template, choose a server pool that the VM can run on, the template that we want to use, etc.

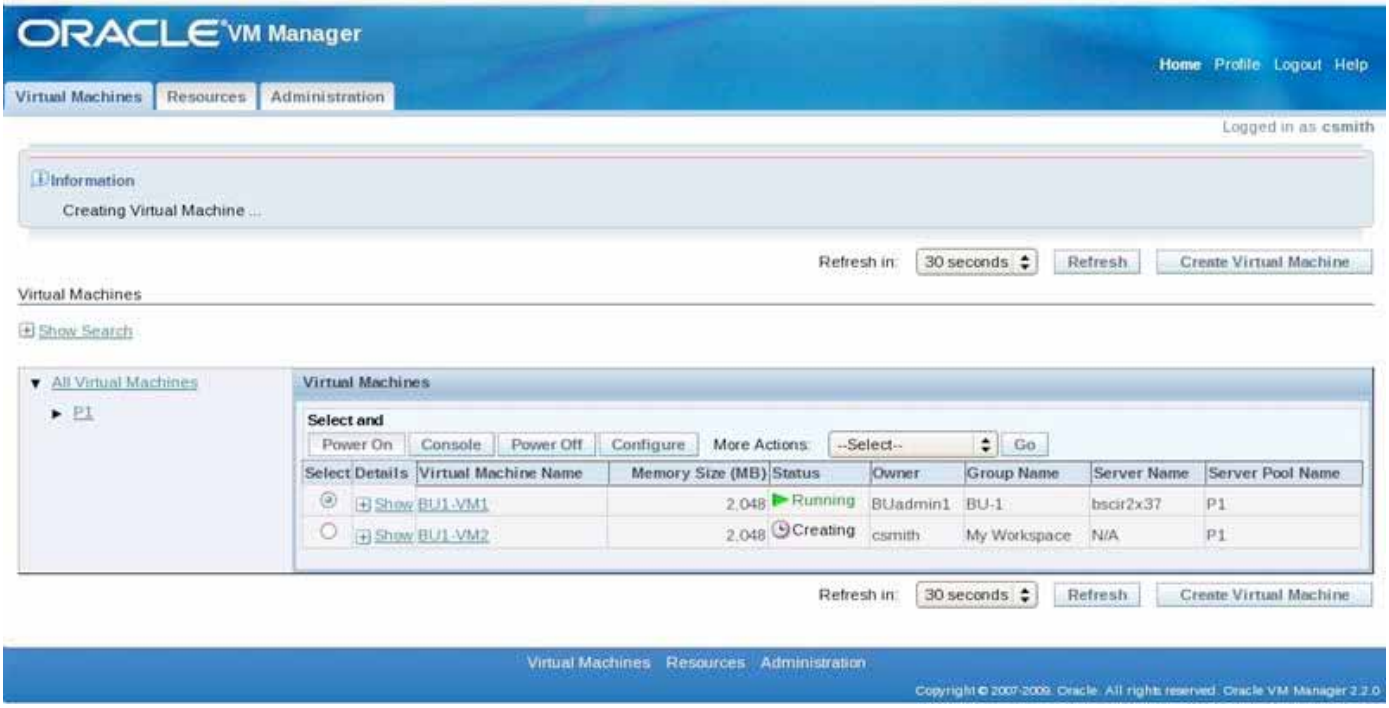
We set a name for the new VM and a password for connecting by VNC to the VM console and we select a network bridge from the interfaces available in the VM.



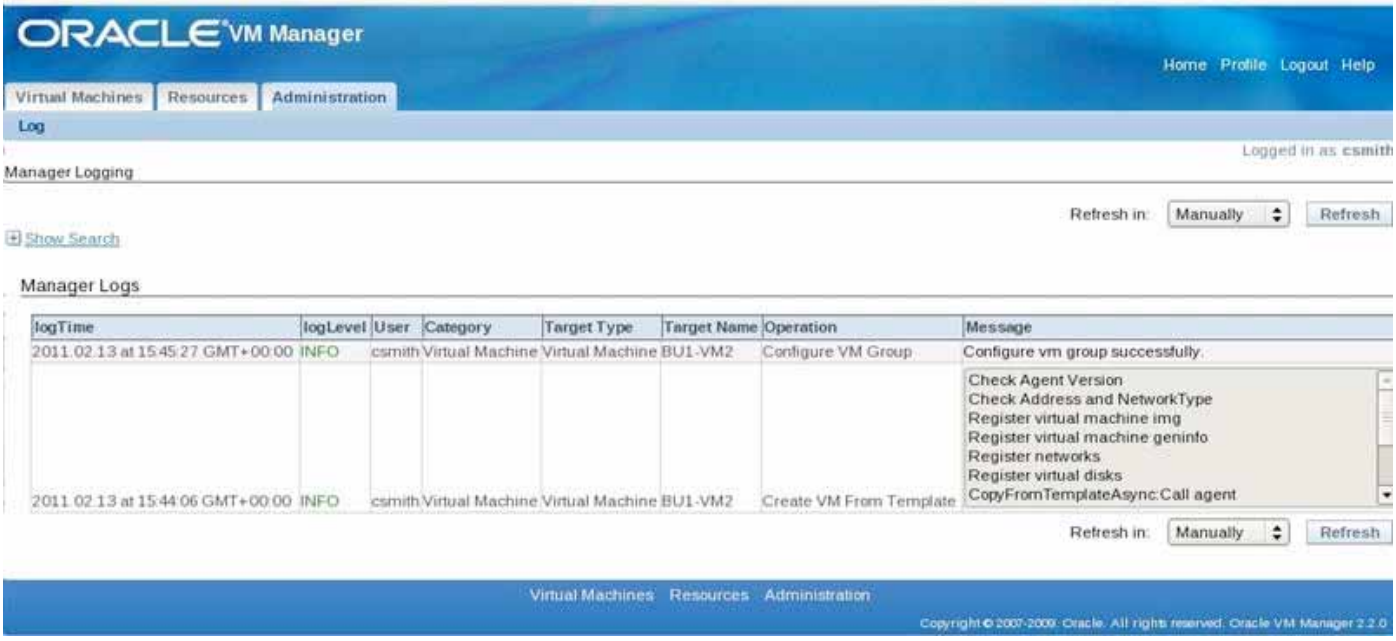

ORACLE VM Manager
Home Profile Logout Help
Virtual Machines Resources Administration
Logged in as csmith
Creation Method
Server Pool
Source
Virtual Machine Information
Confirm
Source
Previous Next
Select Details Virtual Machine Template Name Size (MB) Status Creation Time
Hide OVM_EL5U5_X86_64_PVM_10GB 12,700 Active Jan 20, 2011
Memory Size (MB): 2,048
Operating System: Oracle Enterprise Linux 5 64-bit
Virtualization Method: Paravirtualized
Number of Virtual CPUs: 2
Previous Next
Virtual Machines Resources Administration
Copyright © 2007-2009, Oracle. All rights reserved. Oracle VM Manager 2.2.0


ORACLE VM Manager
Home Profile Logout Help
Virtual Machines Resources Administration
Logged in as csmith
Creation Method
Server Pool
Source
Virtual Machine Information
Confirm
Virtual Machine Information
Previous Next
Virtual Machine Name: BU1-VM2
Console Password:
Confirm Console Password:
High Availability
Enable High Availability:
Network Interface Card
Select and Delete
Select Virtual Network Interface Name Virtual Network Interface MAC Address Bridge
VIF0 00:16:3E:5A:F2:A2 xenbr0
Add Row
Previous Next
Virtual Machines Resources Administration
Copyright © 2007-2009, Oracle. All rights reserved. Oracle VM Manager 2.2.0

After confirmation, creation of the VM image then starts.



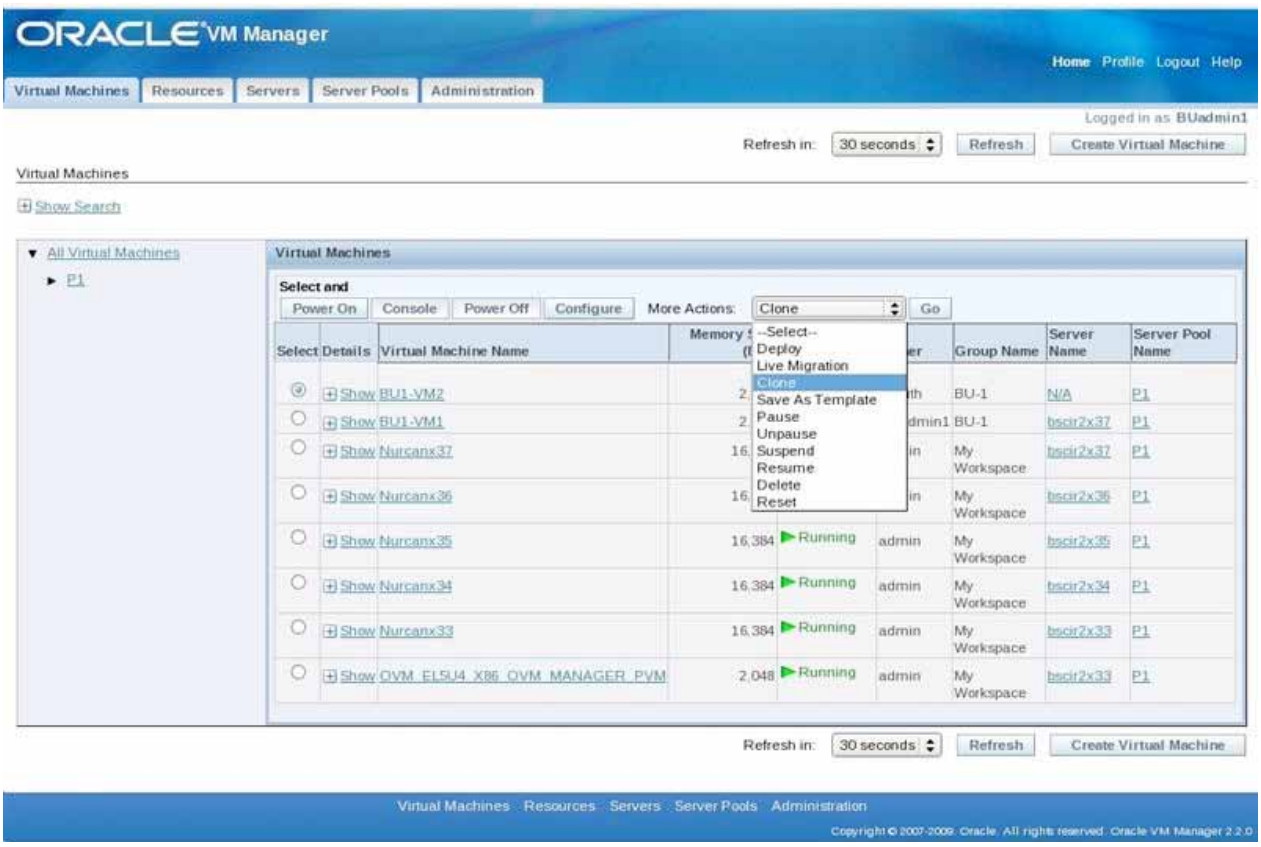
Later we can see that the VM was created successfully by the status, or by examining its logs.

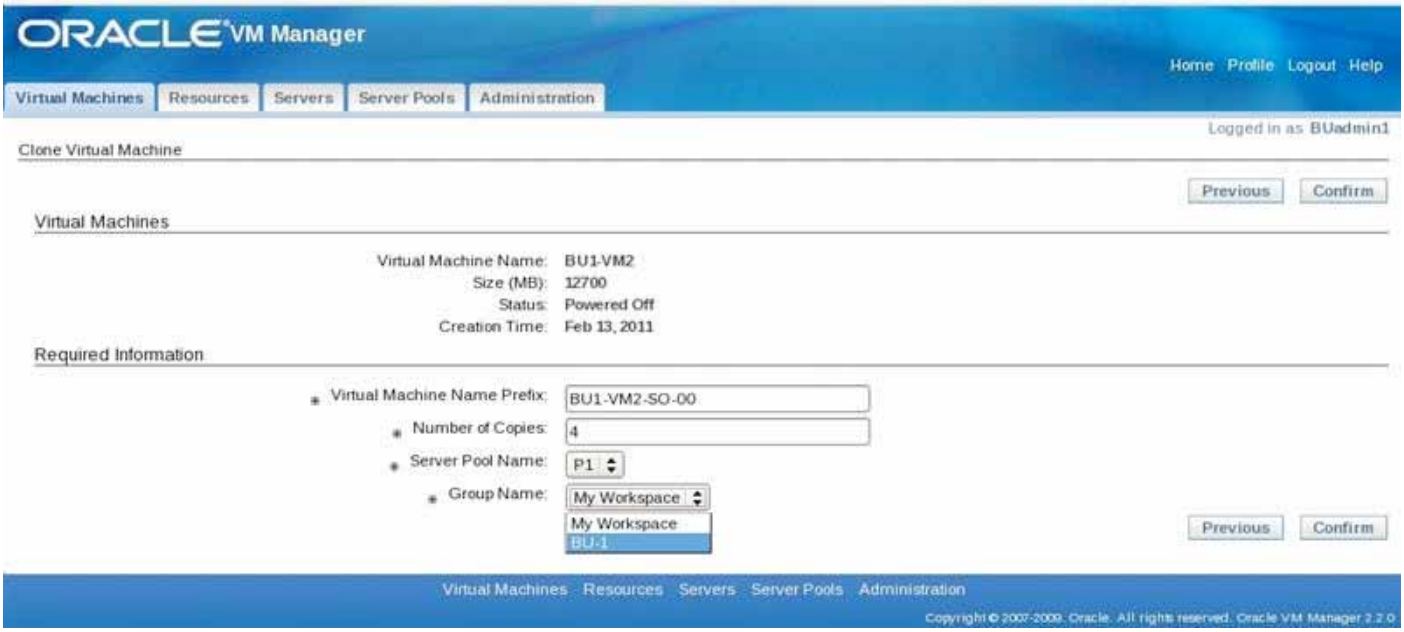


Additional virtual disks can also be created for the VM, or shared virtual disks attached/detached as required.

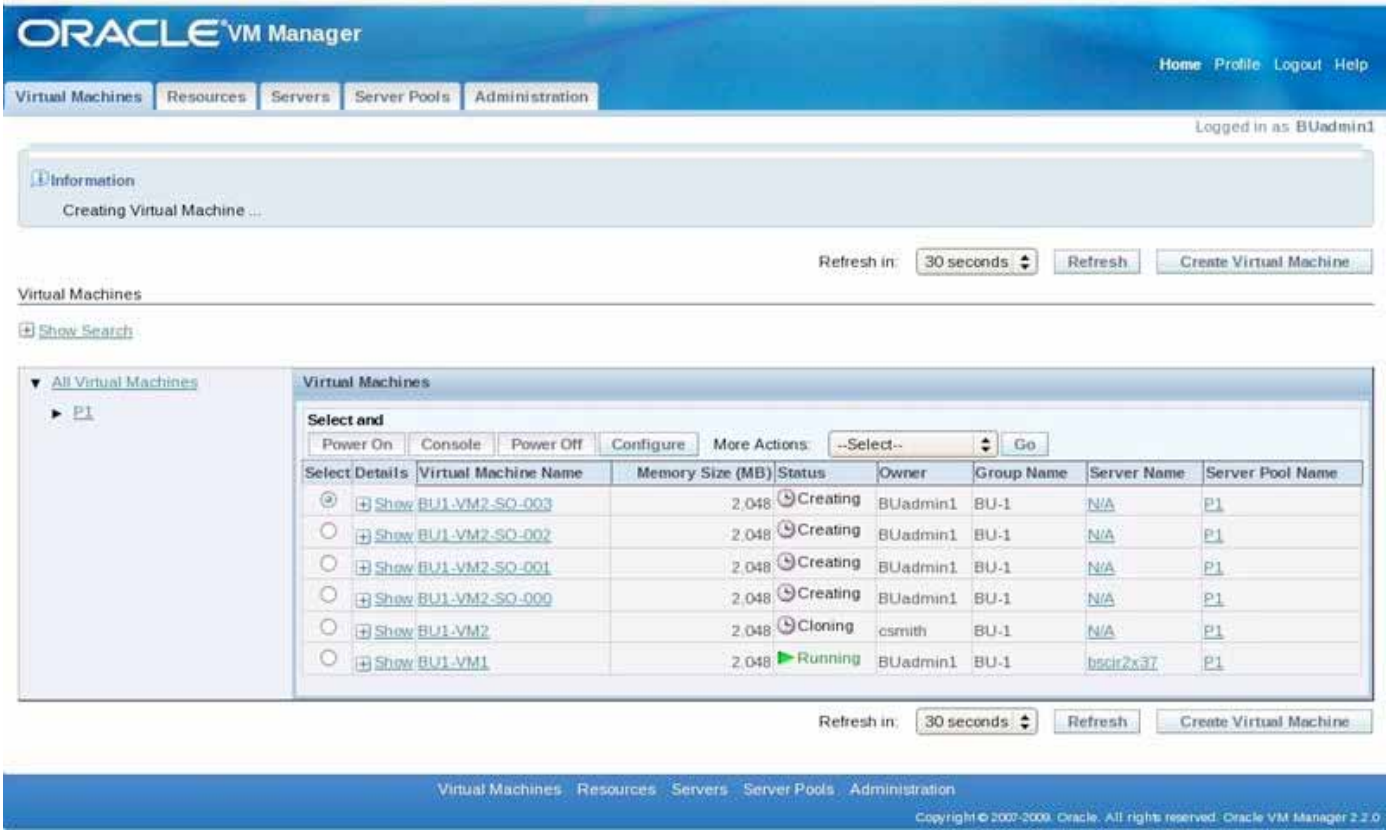
Create Multiple Virtual Machines (scale out service)

Logged in as user BUadmin1, we start by selecting the VM that was created in earlier, and selecting cloning, then a prefix name for the clones.





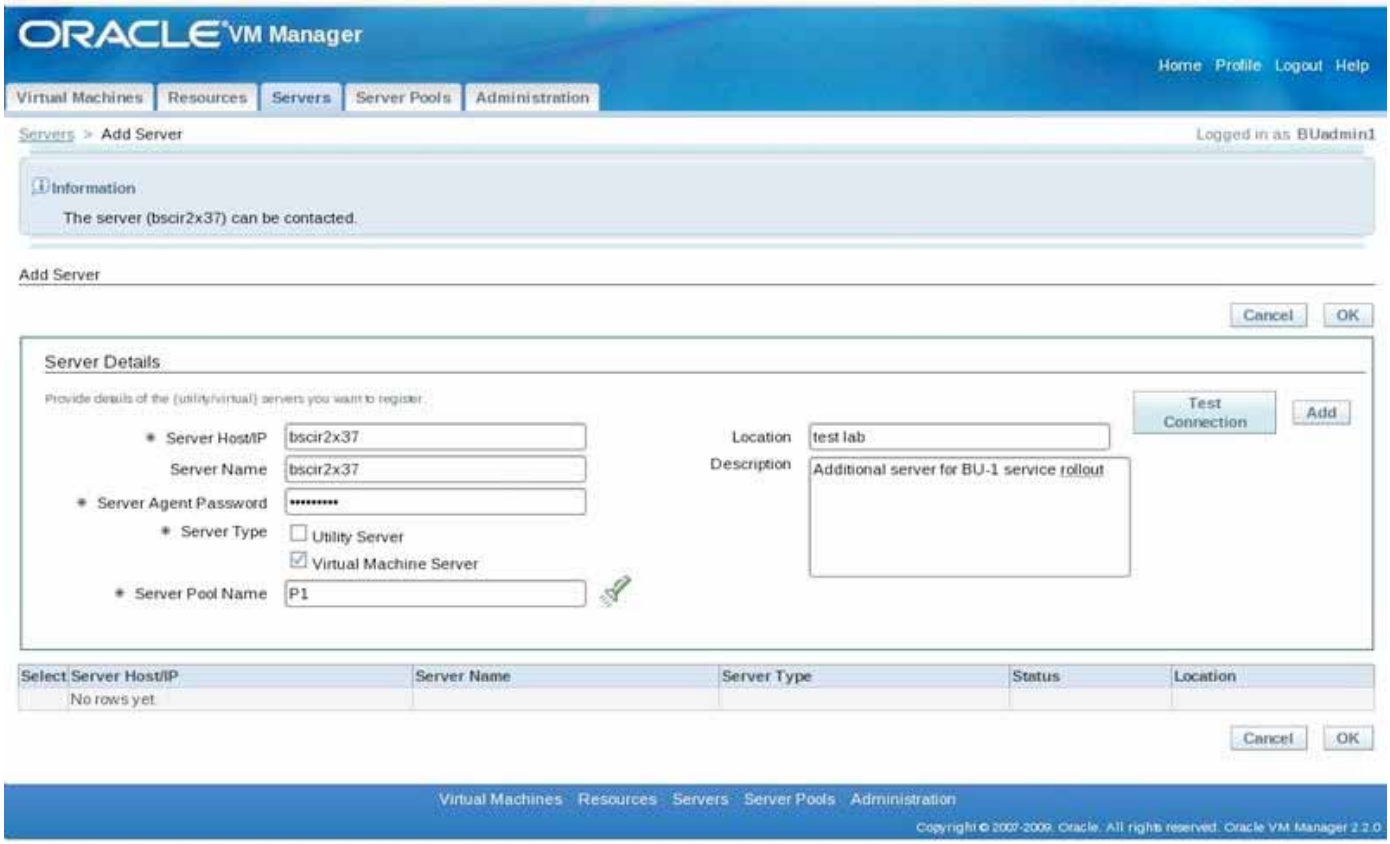
After hitting **Confirm,** we see the clone VMs being created.



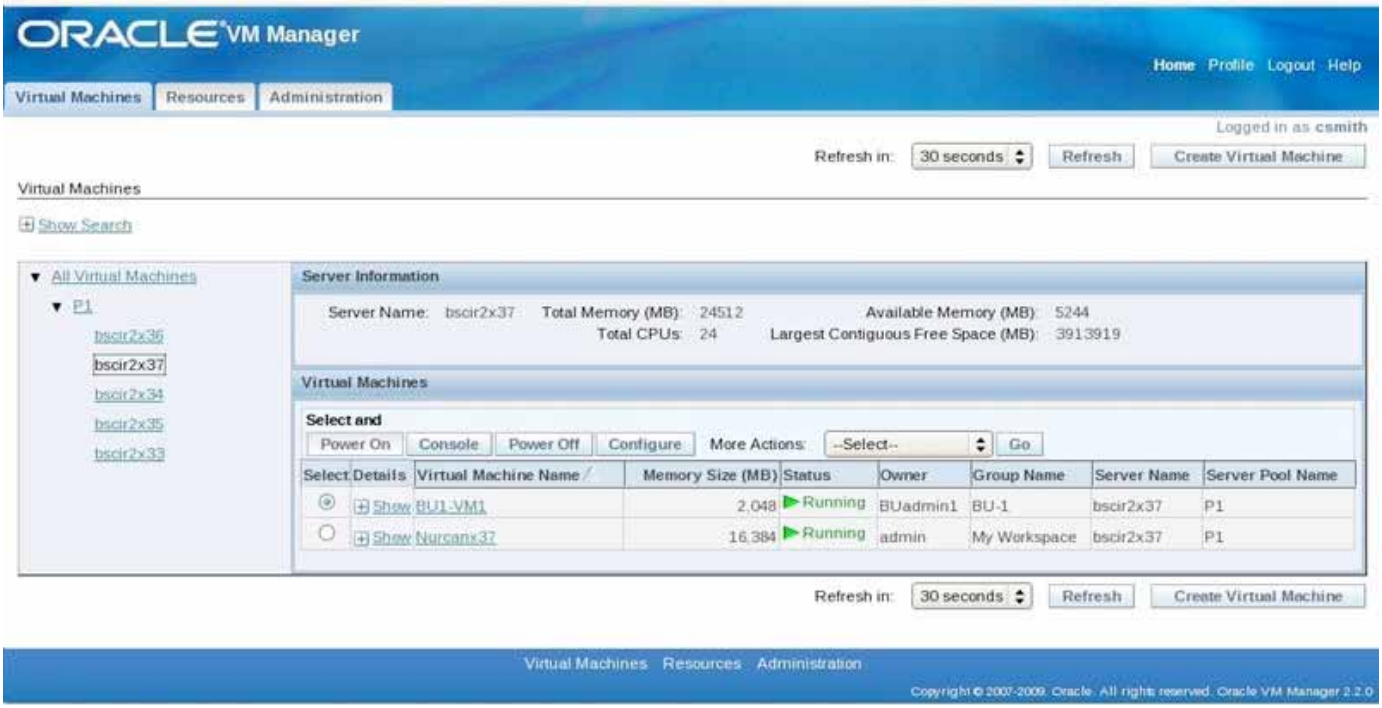
Add Bare Metal Capacity to Existing BU Cluster

Here we show the addition of server bscir2x37 to the existing cluster. The server is first installed with a vanilla install of Oracle VM Server. Then we use Oracle VM Manager to add it into the cluster – it has actually already been added here, but this makes no difference in illustrating the process. Starting from the Servers tab, press Add Server.

Complete the details of the server to be added, and test the connection to the OVM server agent.



Clicking on the server link shows available RAM, CPUs, and other info relating to the new server, plus the VMs that are running on it.

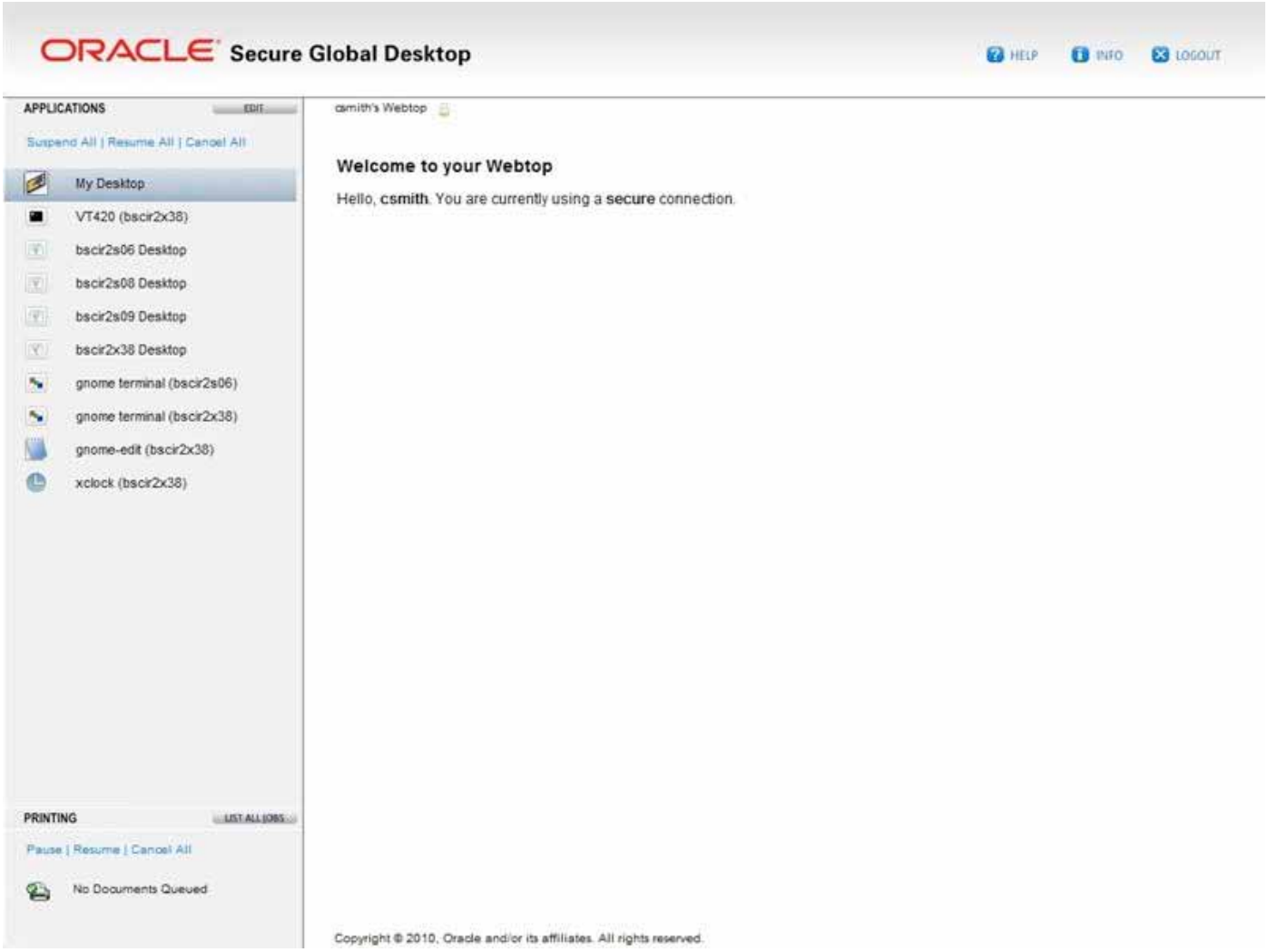


User Access to Virtual Machines Using Web Portal

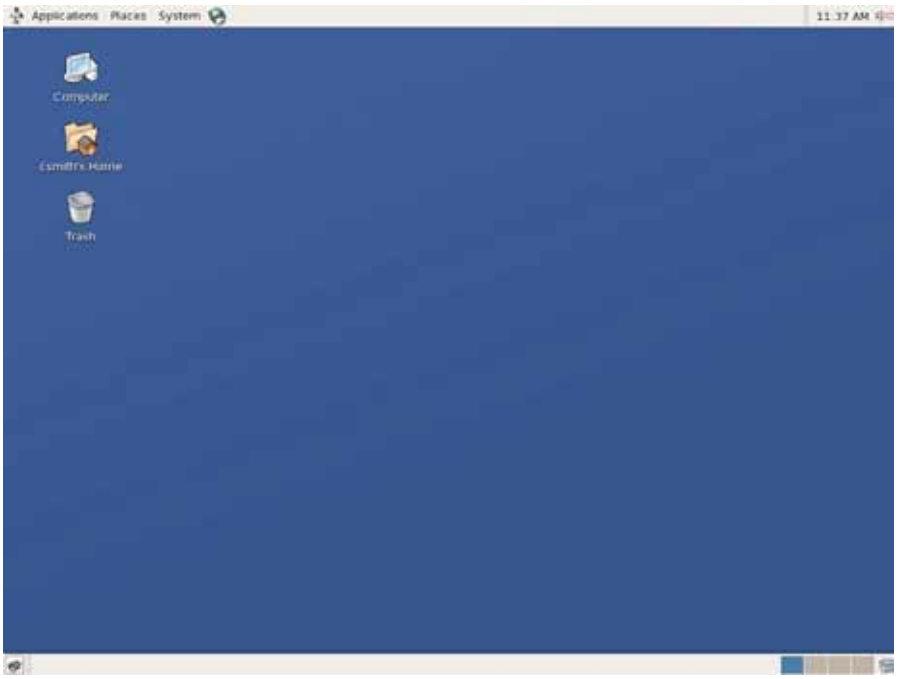
Oracle VM Manager can be deployed with https enabled on a public server.

We have used Oracle Secure Global Desktop to enable remote access to our hosted Oracle VM environment, including Oracle VM Manager and the VMs themselves. This was installed on a separate server, but we could alternatively have installed the Oracle Secure Global Desktop template for Oracle VM, giving us a pre-installed instance.

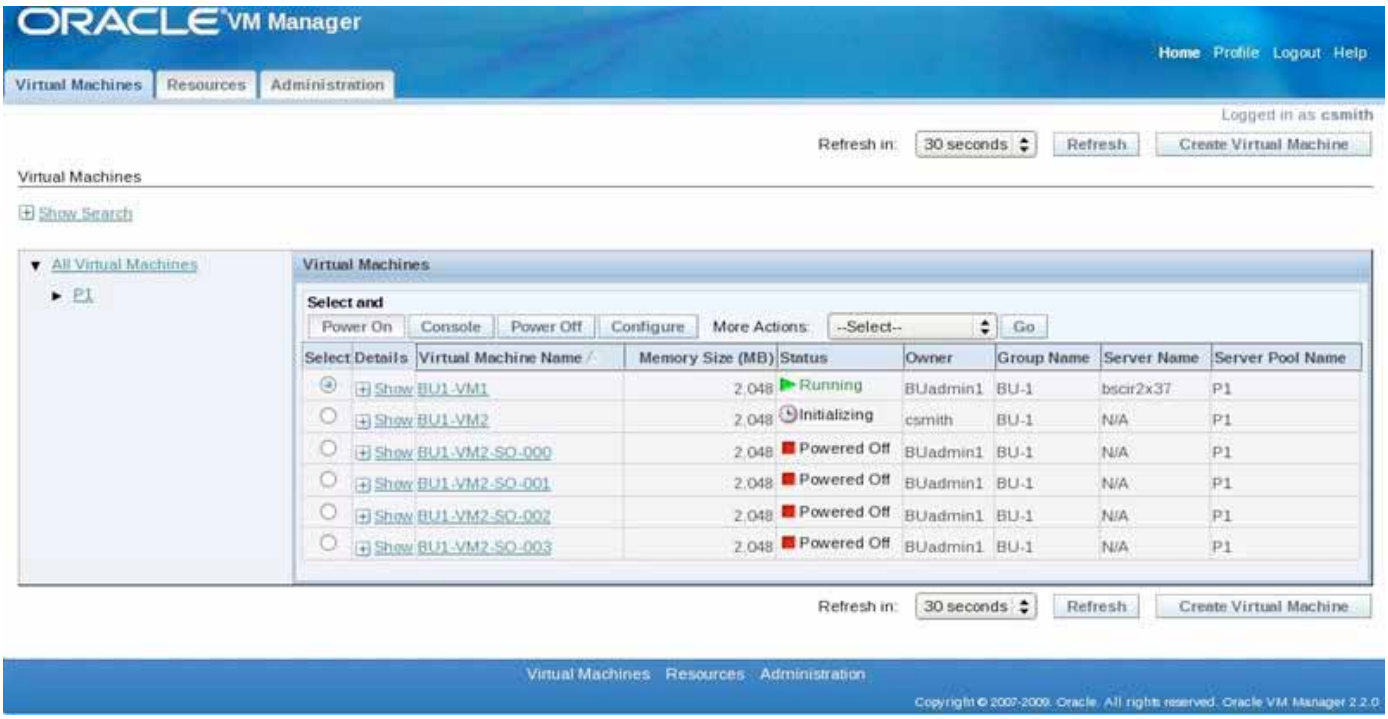
The SGD login is accessed from the public IP or registered name of its host server.


ORACLE Secure Global Desktop
HELP
INFO
LOGOUT
APPLICATIONS
EDIT
Suspend All | Resume All | Cancel All
My Desktop
VT420 (bscir2x38)
bscir2s06 Desktop
bscir2s08 Desktop
bscir2s09 Desktop
bscir2x38 Desktop
gnome terminal (bscir2s06)
gnome terminal (bscir2x38)
gnome-edit (bscir2x38)
xclock (bscir2x38)
PRINTING
LIST ALL JOBS
Pause | Resume | Cancel All
No Documents Queued
csmith's Webtop
Welcome to your Webtop
Hello, csmith. You are currently using a secure connection.
Copyright © 2010, Oracle and/or its affiliates. All rights reserved.

We just select the default "My Desktop" to access an Oracle Enterprise Linux User Desktop on the Oracle Secure Global Desktop server.

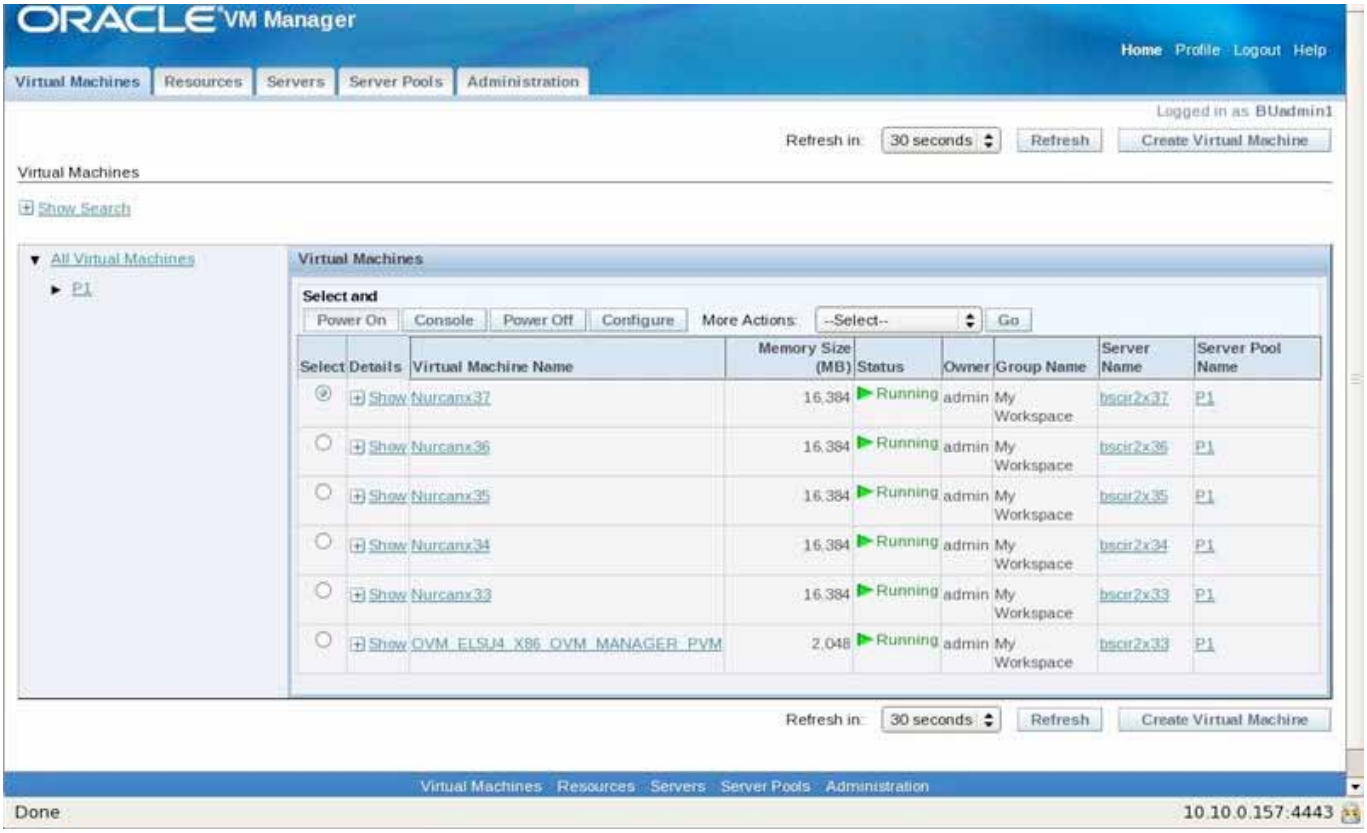


From there we can login to the OVM Manager as normal, through a browser. An SGD administrator can create panel shortcuts for direct access to desktop or terminal sessions on other VMs or physical machines.



Monitor State of a Service

Oracle VM users can login to the portal and view the status of the VMs and servers to which they have been granted access. They can create new VMs and disk quota within the limits of what is permitted by the servers to which they have been granted access. Oracle VM monitors the status of all VMs for which High Availability is enabled and will automatically restart these on available servers within the pool, if they should fail. Also the logs of the individual VMs can be viewed through Oracle VM Manager, showing the status of VMs over time. Extended monitoring facilities are planned for future release.



ORACLE VM Manager

Home Profile Logout Help

Virtual Machines | Resources | Administration

Log

Logged in as csmith

Manager Logging

Refresh in: Manually Refresh

Show Search

Manager Logs

logTime	logLevel	User	Category	Target Type	Target Name	Operation	Message
2011.02.14 at 05:08:28 GMT+00:00	INFO	csmith	Virtual Machine	Virtual Machine	BU1-VM2	Power On VM	Start - /OVS/running_pool/52_BU1-VM2 PowerOn successfully. Update server -bscir2x36 Update Port - 5901.0 Update VM Status - Running Configure CPU Cap Set CPU Cap successfully
2011.02.13 at 15:45:27 GMT+00:00	INFO	csmith	Virtual Machine	Virtual Machine	BU1-VM2	Configure VM Group	Configure vm group successfully.
2011.02.13 at 15:44:06 GMT+00:00	INFO	csmith	Virtual Machine	Virtual Machine	BU1-VM2	Create VM From Template	Check Agent Version Check Address and NetworkType Register virtual machine img Register virtual machine geninfo Register networks Register virtual disks CopyFromTemplateAsync:Call agent

Refresh in: Manually Refresh

Virtual Machines Resources Administration

Done 10.10.0.157:4443

Terminate VMs (remove service)

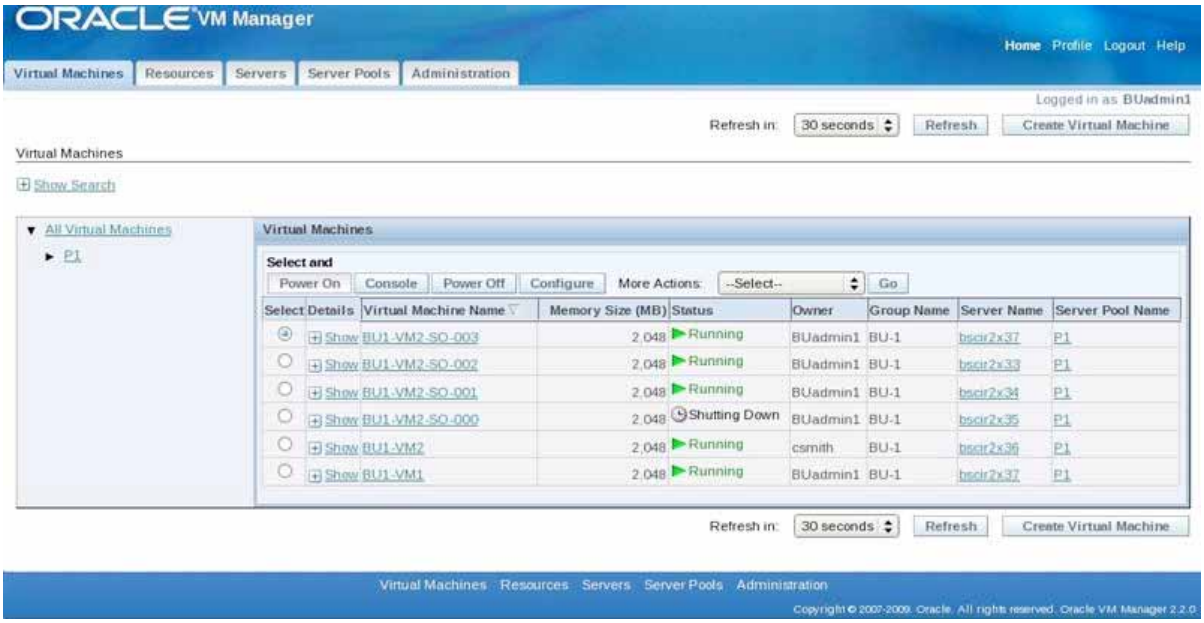
Services may be removed by shutting down and removing the appropriate VMs.

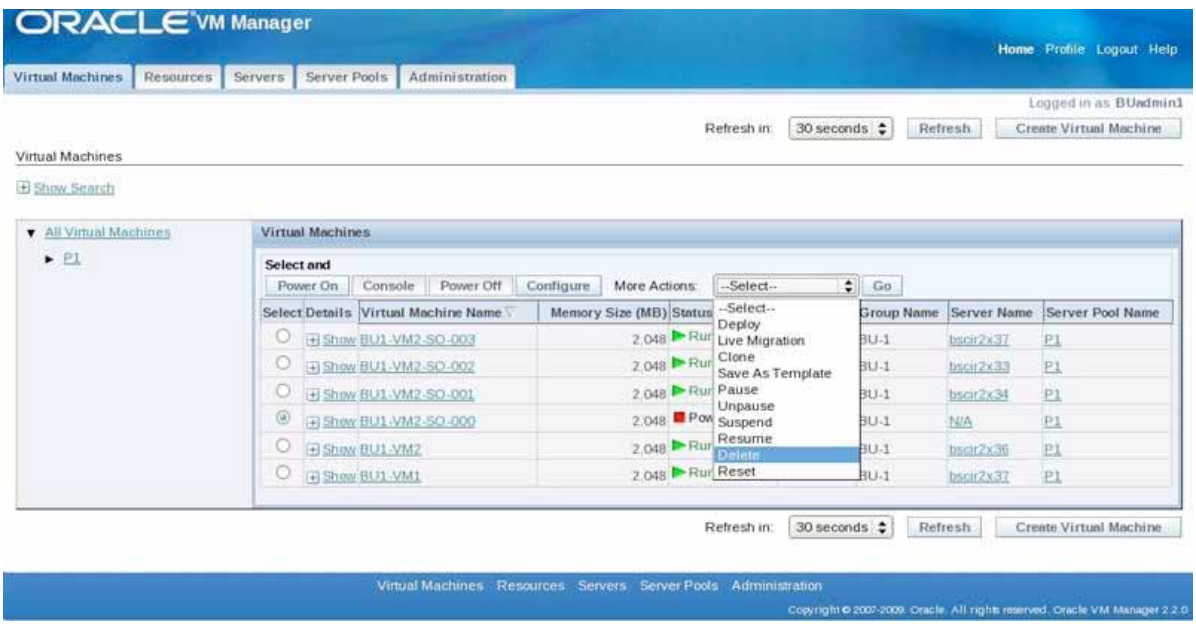
Note that Oracle PaaS (Platform as a Service) offerings will offer service level administration on Oracle VM.

Below we see the current status of the hosted VMs.

We have clicked **Power Off** for one of the VMs of the BU1-VM2-SO service.

Once powered off, we can select to delete the VM instance. Note the checkbox to delete the VMs files from the Oracle VM shared storage.





Balance Compute Resource Utilization

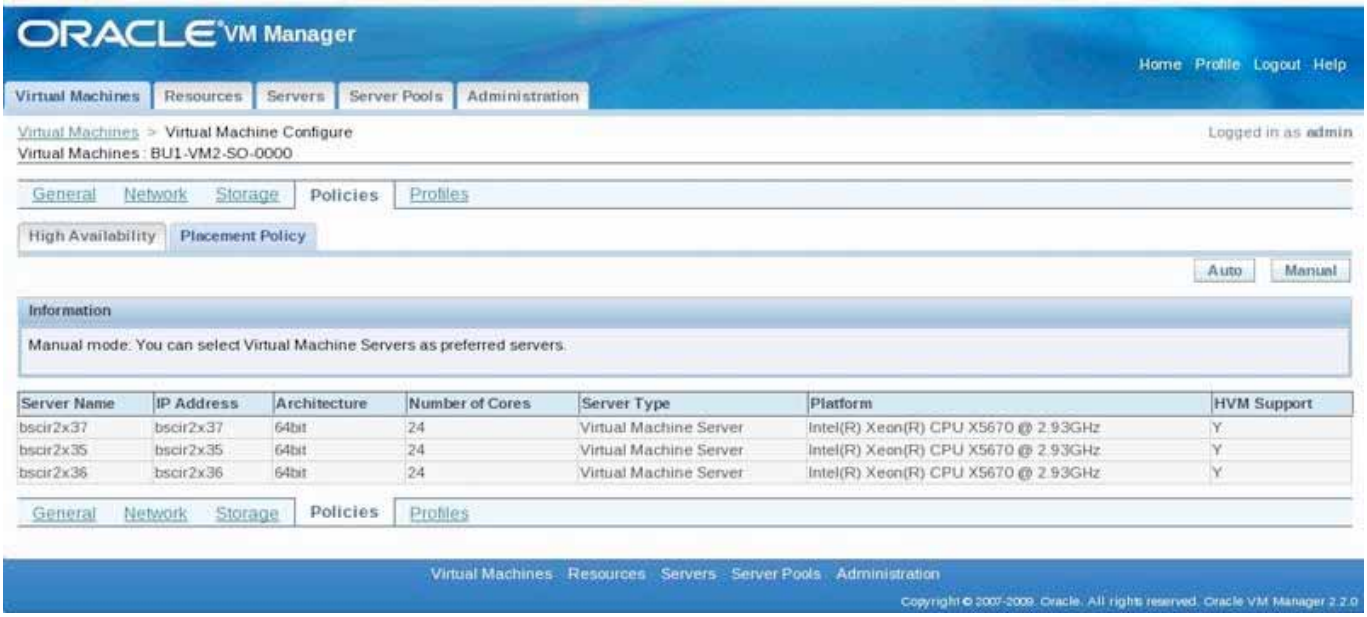
Oracle VM automatically allocates VMs to servers within a pool of servers based on their availability and compute usage at allocation time. Oracle VM will only allocate VM workloads to servers within a pool that have sufficient resources. If insufficient resources are available on one preferred VM then it will choose another, either by auto-allocation, or from within a list of preferred servers.

It is possible to assign "preferred" server(s) for VM startup. Oracle VM will try to use these servers to start the VM if they are available and have sufficient resources.

In the example below we try to start VM BU1-VM2-SO-0000.

We can see that the VM has been configured with three preferred servers.

The BU admin has effectively defined that one of these should be used to run the VM.



We click to Power On BU1-VM2-SO-0000, then we see that the VM has now started on physical server bscir2x36.

ORACLE VM Manager

Home Profile Logout Help

Virtual Machines | Resources | Servers | Server Pools | Administration

Logged in as BUadmin1

Refresh in: 30 seconds Refresh Create Virtual Machine

Virtual Machines

Show Search

- All Virtual Machines
  - P1
    - bscir2x36
    - bscir2x37
    - bscir2x34
    - bscir2x35
    - bscir2x33

Server Pool Information

Server Pool Name: P1 Total Memory (MB): 122560 Available Memory (MB): 44652
Total CPUs: 120 Largest Contiguous Free Space (MB): 3917744

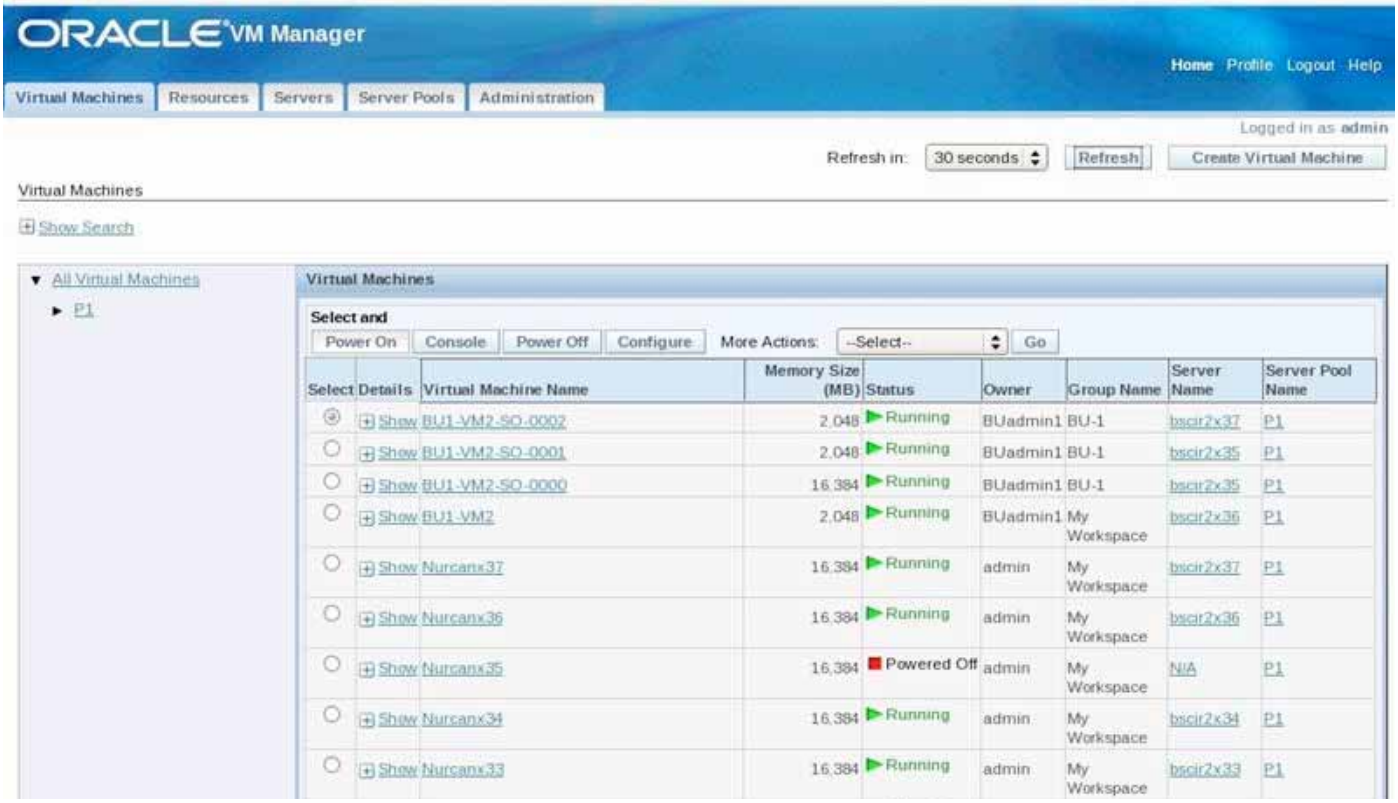
Virtual Machines

Select and: Power On | Console | Power Off | Configure | More Actions: --Select-- Go

Select	Details	Virtual Machine Name	Memory Size (MB)	Status	Owner	Group Name	Server Name	Server Pool Name
◉	Show	BU1-VM2-SO-0002	2,048	Running	BUadmin1	BU-1	bscir2x37	P1
○	Show	BU1-VM2-SO-0001	2,048	Running	BUadmin1	BU-1	bscir2x35	P1
○	Show	BU1-VM2-SO-0000	16,384	Running	BUadmin1	BU-1	bscir2x36	P1
○	Show	BU1-VM2	2,048	Running	BUadmin1	My Workspace	bscir2x36	P1
○	Show	Nurcanx37	16,384	Running	admin	My Workspace	bscir2x37	P1
○	Show	Nurcanx36	16,384	Powered Off	admin	My Workspace	N/A	P1
○	Show	Nurcanx35	16,384	Powered Off	admin	My Workspace	N/A	P1
○	Show	Nurcanx34	16,384	Running	admin	My Workspace	bscir2x34	P1
○	Show	Nurcanx33	16,384	Running	admin	My Workspace	bscir2x33	P1

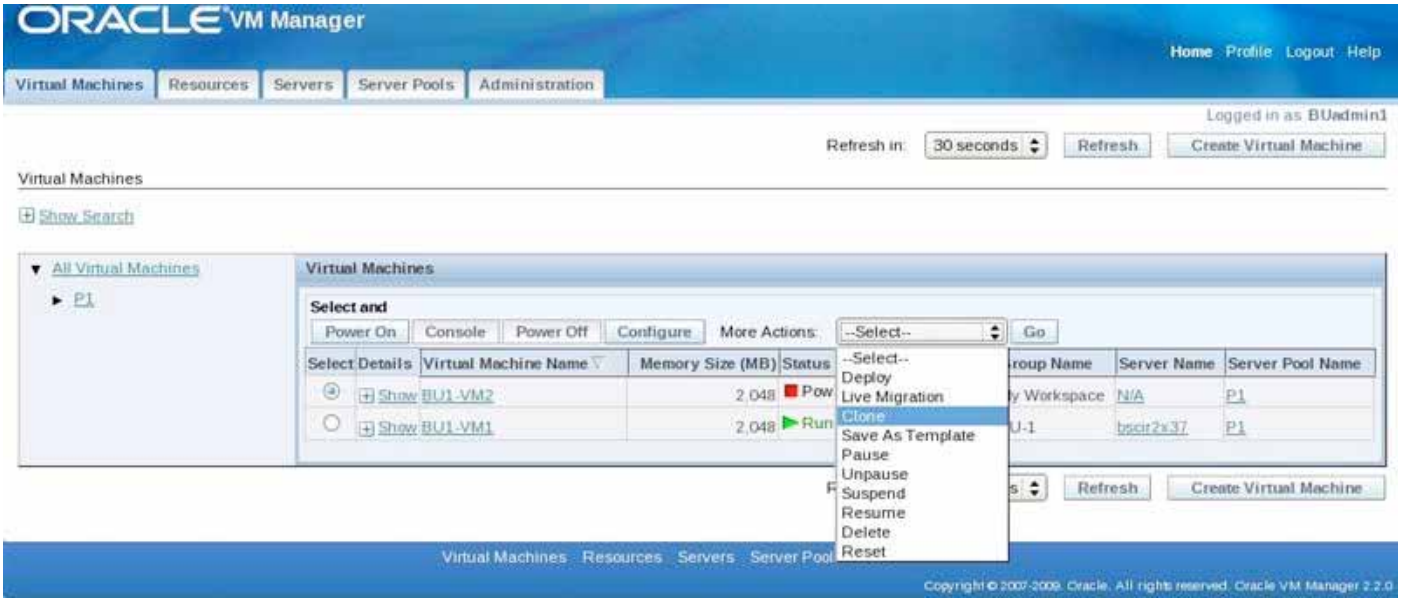
If at another time there is not enough free RAM on bscir2x36 to run our VM, Oracle VM chooses another server from the list of preferred servers for the VM, and starts it on bscir2x35 instead.

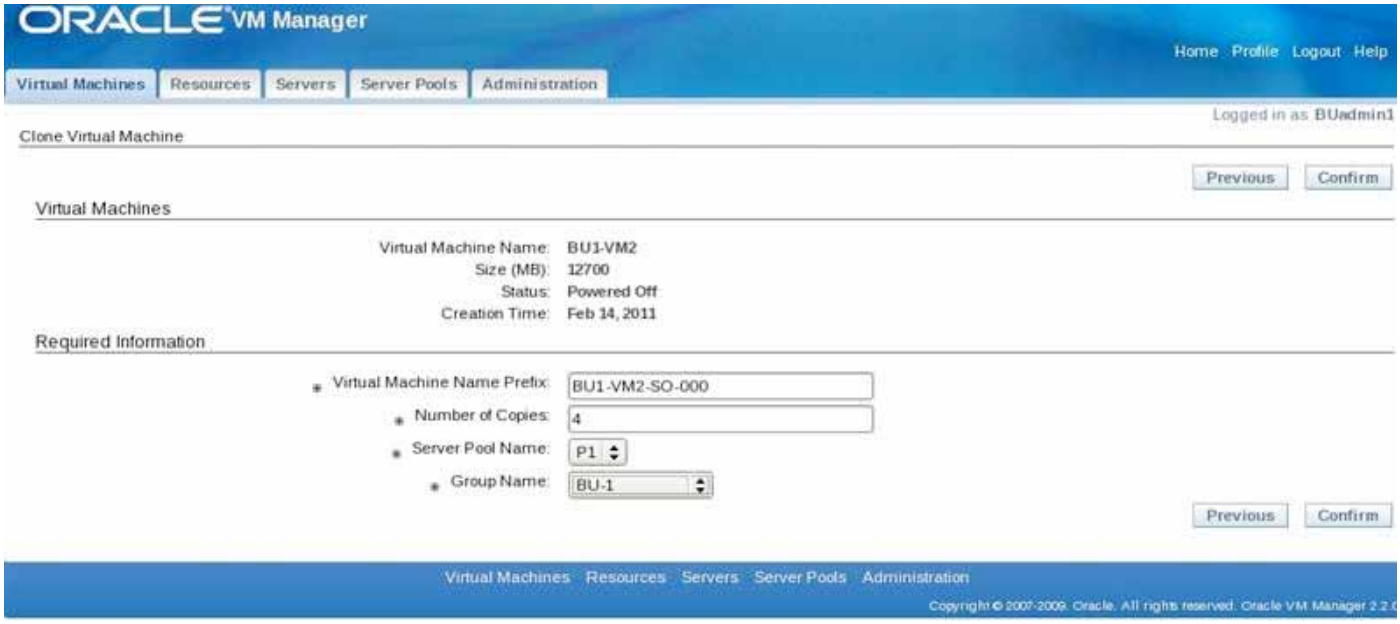
Note that VMs can be live migrated from one physical server to another without loss of service, using Oracle VM Manager.

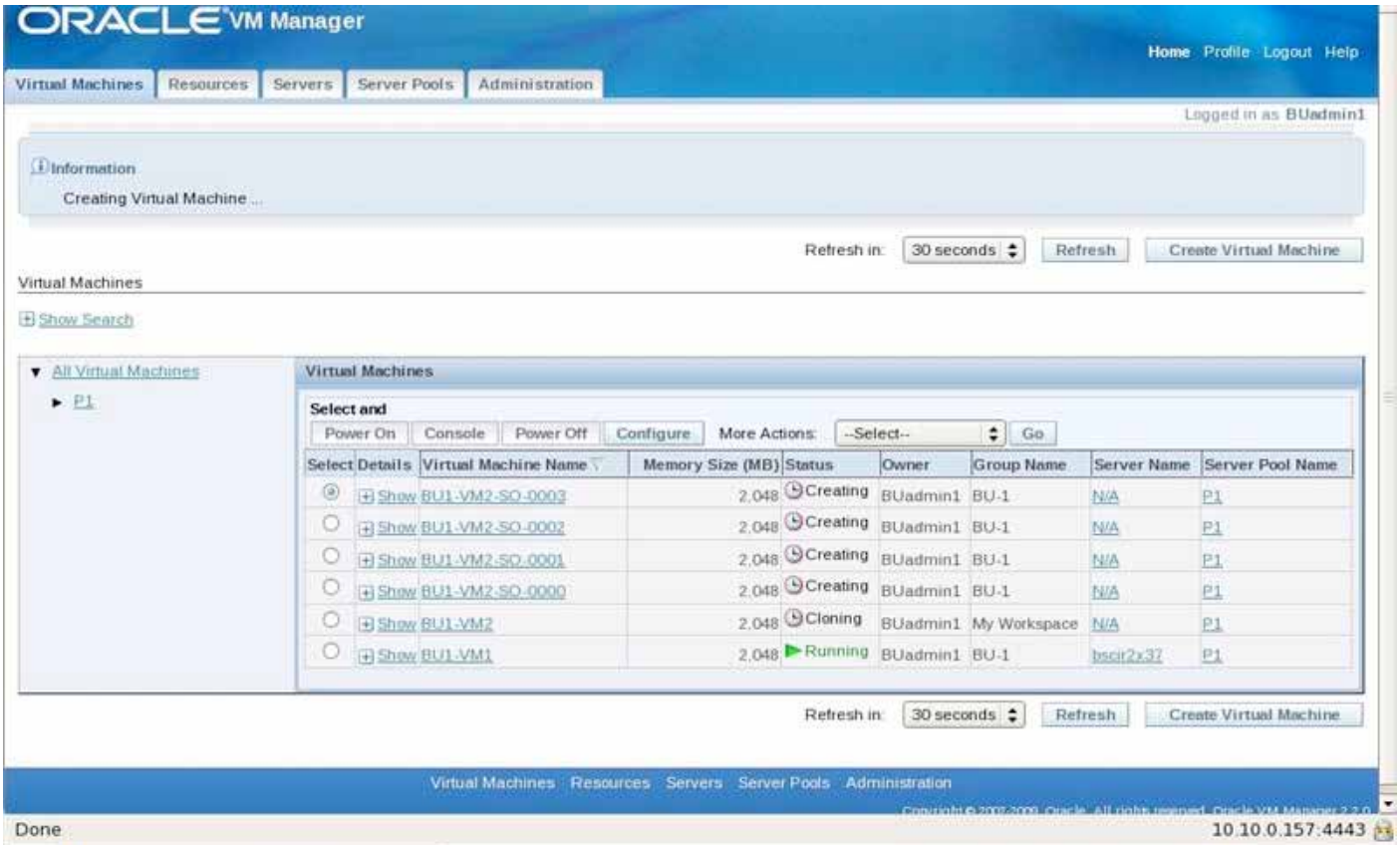


Create Multiple Virtual Machines (scale out service)

In this example we have created a VM from a template, and then we choose to clone it to produce a scaled out service.

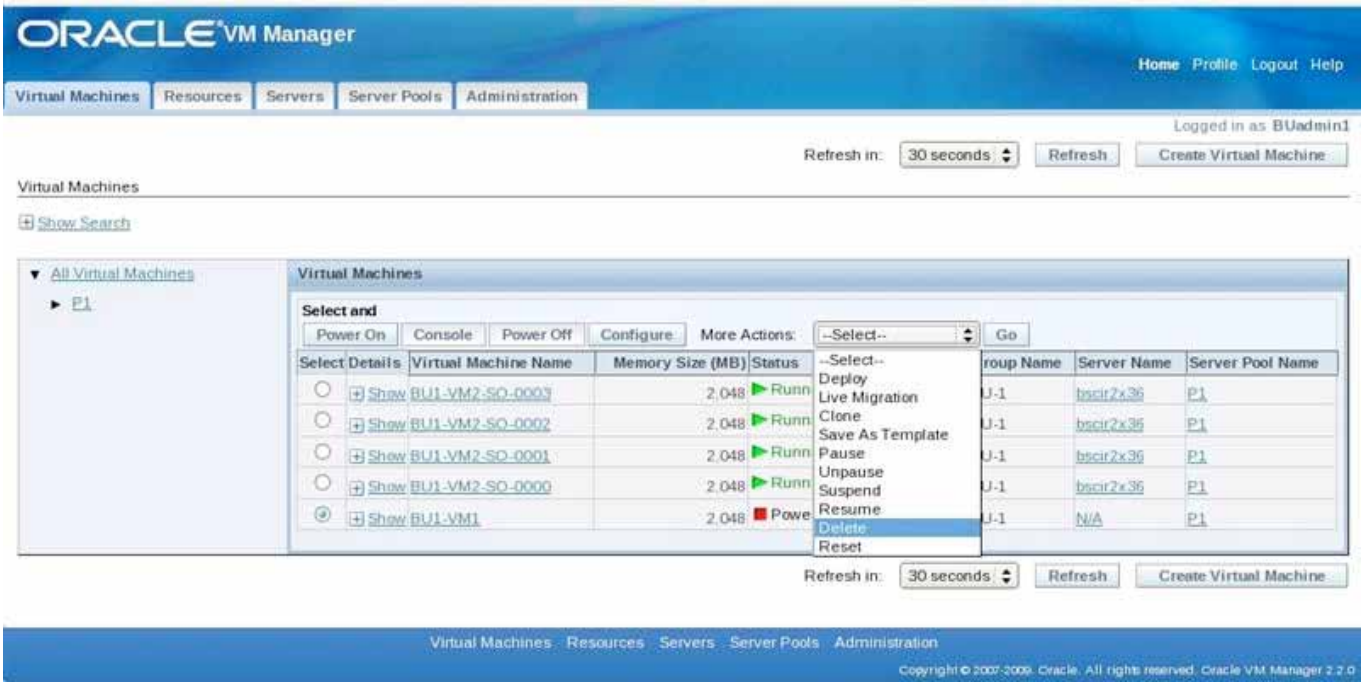


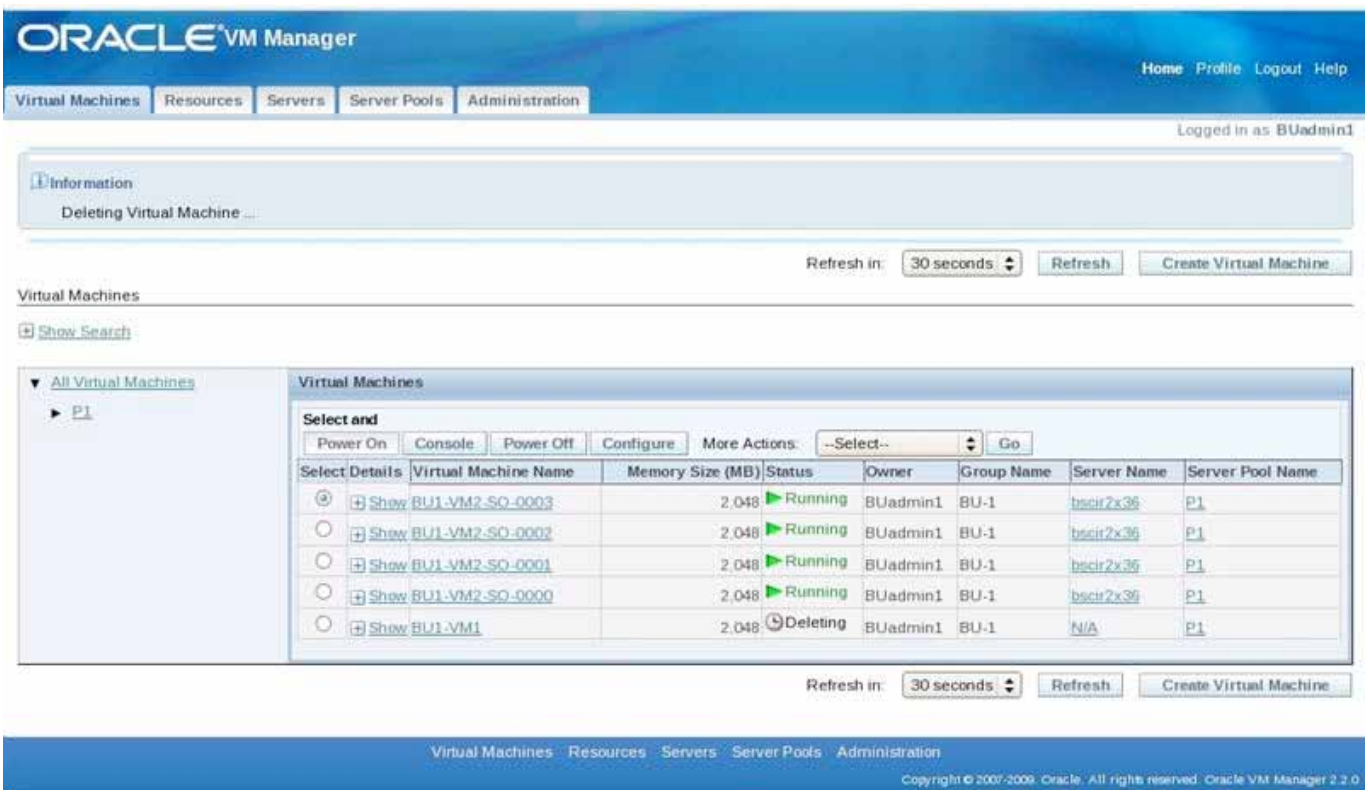



ORACLE VM Manager
Home Profile Logout Help
Virtual Machines Resources Servers Server Pools Administration
Logged in as BUadmin1
Information
Creating Virtual Machine ...
Refresh in: 30 seconds Refresh Create Virtual Machine
Virtual Machines
Show Search
All Virtual Machines
P1
Virtual Machines
Select and
Power On Console Power Off Configure More Actions: --Select-- Go
Select Details Virtual Machine Name Memory Size (MB) Status Owner Group Name Server Name Server Pool Name
Show BU1-VM2-SO-0003 2,048 Creating BUadmin1 BU-1 N/A P1
Show BU1-VM2-SO-0002 2,048 Creating BUadmin1 BU-1 N/A P1
Show BU1-VM2-SO-0001 2,048 Creating BUadmin1 BU-1 N/A P1
Show BU1-VM2-SO-0000 2,048 Creating BUadmin1 BU-1 N/A P1
Show BU1-VM2 2,048 Cloning BUadmin1 My Workspace N/A P1
Show BU1-VM1 2,048 Running BUadmin1 BU-1 bscir2x37 P1
Refresh in: 30 seconds Refresh Create Virtual Machine
Virtual Machines Resources Servers Server Pools Administration
Done
10.10.0.157:4443

Remove Virtual Machines (scale down service)

The BU admin is able to stop and then delete virtual machines.





Failure Scenario (Oracle VMs can be configured for HA, also live migration)

We can show that a physical server can fail and the BU cluster remains operational.

Oracle VM will restart VMs on alternate physical servers. It can also live migrate running VMs to alternate physical servers without loss of service.

Oracle VM HA relies on having shared storage configured, which was done as part of the Oracle VM installation for the lab environment.

We then need to enable HA for the Server Pool. The Server Pools > Edit Server Pool > Check button confirms whether the installation has been configured to support HA. If so then we just need to check the box to enable HA.

We then enable HA for each VM that requires it.

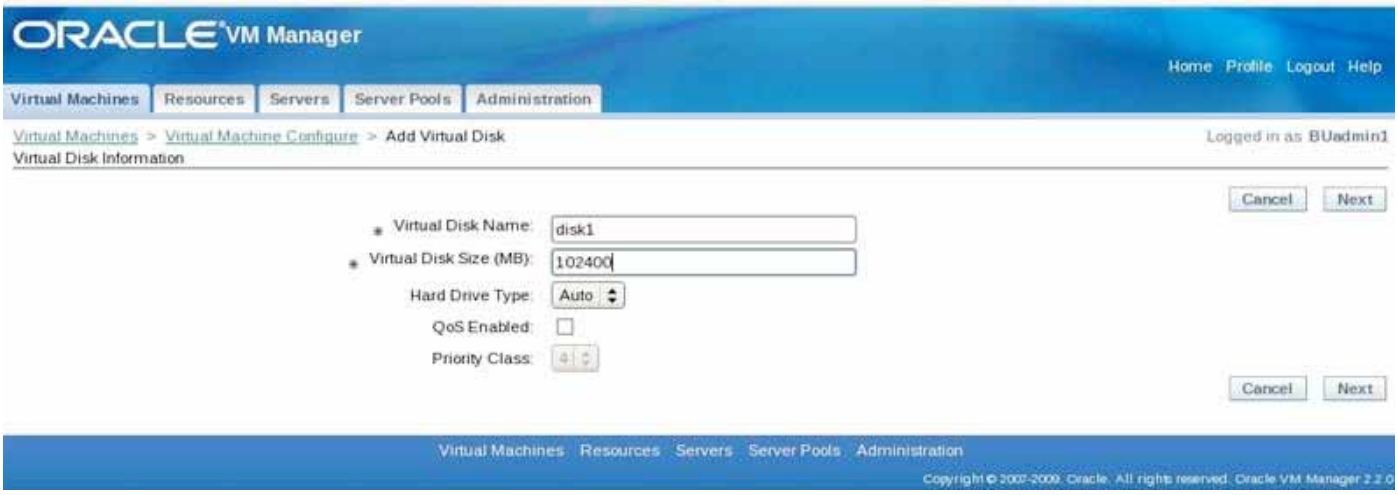
In a server failure scenario Oracle VM detects that the server is no longer available and after a short while restarts the VMs that were on it, choosing alternative server(s) from the server pool.

Service Change Request/Scale Service

We can login as the BU admin and add additional storage and compute resources.

We can add new physical servers to existing pools or create new pools of new servers.

We can add new virtual disks and networks. After creation of the virtual disk it can be formatted and made available by the guest OS.

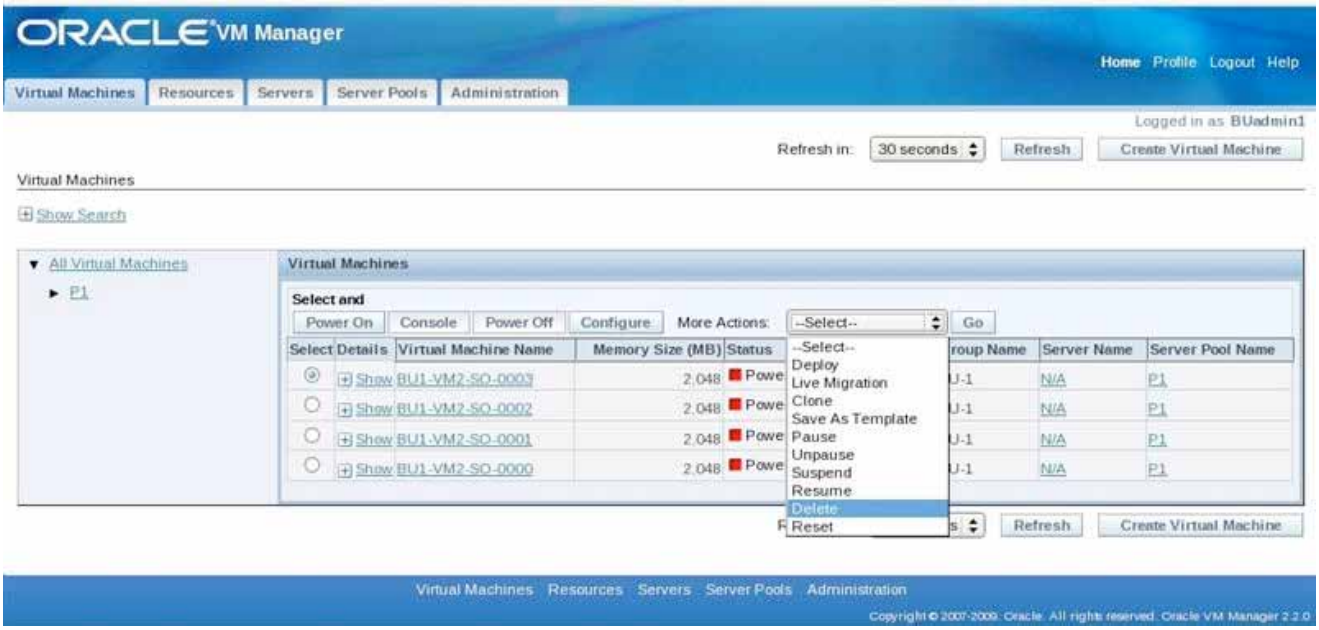


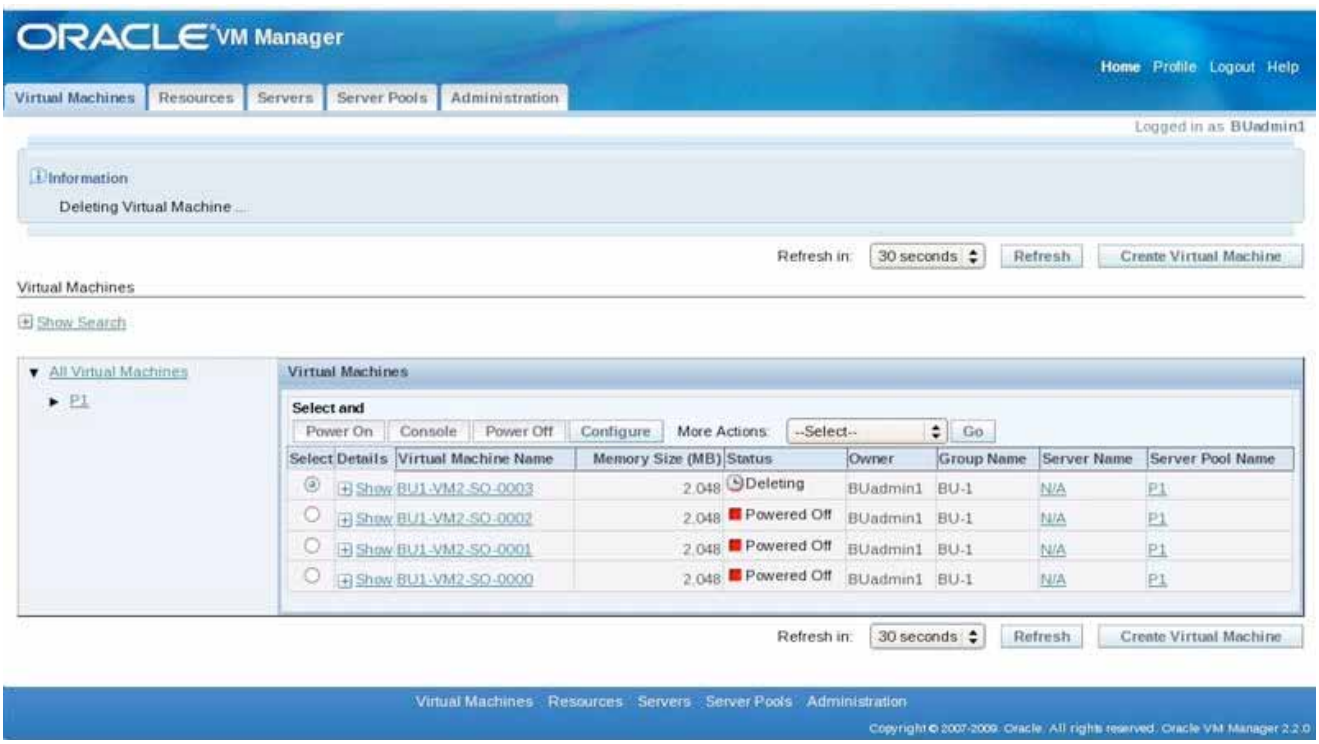
Decommission Services

BU admins can decommission deployed services by removing VMs.

Here we delete a VM after powering off, similarly for all VMs that we wish to decommission.

We check the box to physically remove the VM files from Oracle VM.





Distributed Application (Data Analysis) Test Case

Hadoop* is becoming increasingly attractive to a broad spectrum of organizations as a framework for running computational workloads to process big data across large clusters of servers. Cloud customers can use private and public clouds to process their big data with minimal investment by using a pay as you go approach. We tested Hadoop TeraSort and Word Count benchmarks in our test bed to show the feasibility of using virtual environments to run Hadoop applications.

For more information on Hadoop, please refer to:

- http://hadoop.apache.org
- http://hadoop.apache.org/common/docs/r0.20.2/index.html

Background & VM portability:

We initially ran these Hadoop tests using Intel processor based blade server hardware at the Intel Oracle Bid Support Centre at Intel in Ireland. We then copied a single VM from this environment to the Intel Cloud Builders test lab environment at Oracle, Scotland, where Intel processor based Sun Blade servers were used as detailed in the hardware description. We then cloned this single VM to provide the additional required instances and modified the hostnames, static IPs, gateway, and DNS settings accordingly. As a result we did not have to re-install the extra software components on each VM, saving time and effort during the test process. The VMs ran without modification on both environments, showing the migration and compatibility of Oracle VM instances on different brands of Intel Xeon processor based hardware.

Note that no attempt was made to optimize the Oracle VM installation for Hadoop performance.

Better performance would be expected by the specification of SSD devices for storage, 10Gb Ethernet between nodes, and direct utilization of storage devices rather than the default file-mapped storage, plus other potential optimizations which could be applied.

The main purpose of this test is to show that Hadoop can be successfully implemented on Oracle VM, and additional tuning of its performance is outside the scope of the Intel Cloud Builders test.

Additional comments on the use of Solid State Drives with Hadoop can be found in the Storage section.

Hadoop Platform Configuration:

We tested Hadoop by using a four node cluster. Each node is implemented by using an Oracle VM instance. For the Hadoop tests we chose not to use the Oracle's Sun ZFS Storage Appliance, but instead allowed each blade server to use its own local disks, as this was considered more appropriate for the Hadoop use cases which require local storage.

The following shows the test platform configuration steps that need to be done to prepare the cluster to run the Hadoop test cases:

- Four blade servers each have Oracle VM server 2.2.1 installed; Oracle VM Manager 2.2 was installed as a VM on the first node.
- Each blade server has a single Oracle Enterprise Linux 5u5 x86-64 Paravirtualized Virtual Machine running on it (PVM).
- In OVM Manager pools P1 - P4 were created to allow each Oracle VM Server to use its own local storage, but managed from a single OVM Manager instance. Each pool contained one physical server.
- Each VM uses standard Oracle VM file based storage for its system disk, and therefore has its own directory under /OVS/running_pool. We did not use shared storage for these tests, but instead allowed each OVM Server to use storage on its own local physical host (local disks). This option was chosen as being closest to the requirement of Hadoop, where we prefer local storage for each node.
- The standard OEL5 firewall was disabled on all VMs for the purpose of the test. (Note that some tests may require changes to the config of the firewall on the physical server hosting each VM; do not disable the firewall on any server that has Internet connectivity.)
- All VMs were given static IPs and configured to utilise the local DNS server and gateway. (Changes can be made by modifying the network config directly, or through "setup" at the OEL5 command line.)
- Each VM was given two additional virtual disks by adding local disks to each VM through Oracle VM Manager; resulting devices are visible under /proc/partitions. These virtual disks are actually represented as files on the actual storage devices of the physical host system – this is the standard approach in Oracle VM.
- Each virtual disk was then formatted using fdisk to create a new partition and then formatted as ext3 using mkfs.ext3.
- Each disk was then configured to automount by editing /etc/fstab in each VM.
- A test user, "nurcan" was created on each VM, for the test scripts. This user had no password. (This was done to simplify test execution in a lab environment, but should not be done in any live deployment.)
- SSH config on each VM was modified to

allow the test scripts to SSH to each VM as user nurcan, by adding "AllowUsers nurcan" to /etc/ssh/sshd_config. "PermitEmpty Passwords yes" was also added to allow the test script to run. (This should not be set on any VM or server in a live deployment, or on any server that has connection to the Internet.)

The platform configuration steps above should be executed in an automated fashion on a cloud portal based on the user's cluster specifications. In our test cases, we executed this custom configuration manually.

Hadoop Test Software Stack:

We installed the following software on every node in the test cluster (except when VMs were cloned):

- Jdk-6u23-linux-x64 (open source)
- Hadoop-0.20.2 (open source)
- Python-3.1.2 (open source)
- GnuPlot-4.4.2 (open source)
- Intel® Benchmark Install and Test Tool (Intel tool)

Hadoop Configuration and Test Case Execution:

The following describes Hadoop software configuration and Hadoop test case execution steps:

1. A four node Hadoop cluster is created. One of the nodes is designated as master node. We used the other three nodes as datanodes
2. Configured Hadoop to use two virtual disks to store HDFS data and map-reduce files on each node
3. Configured Hadoop to use the installed java home
4. Configured Hadoop nodes by using ulimit command to increase the limits for the user resources
5. Initialized Hadoop Distributed File System (HDFS) system by formatting namenode
6. Started Hadoop servers by using Hadoop "start-all.sh" script
7. Executed steps 8-16 to test Hadoop TeraSort benchmark
8. Generated test data for the benchmark
9. Started sar and iostat agents on each node
10. Ran the benchmark
11. Stopped sar and iostat agents on each node
12. Created performance plots that shows CPU, memory, disk, and network utilizations for each node
13. Collected performance data from all nodes
14. Created cluster performance metrics that shows average CPU and disk utilization for the cluster
15. Created cluster performance metrics that shows disk and network throughput for the cluster
16. Created cluster performance metrics that shows device I/O rate and total memory utilizations for the cluster
17. Repeated steps 8-16 to test Hadoop Word Count Benchmark
18. Stopped Hadoop servers by using Hadoop "stop-all.sh" script

The following are the screenshots captured during some of the test steps described above.

Figure 5: Hadoop* Installation Directory on the Oracle* Cluster Nodes

```
1280x1024 Desktop                                   LogMeIn - Remote Session
File  Edit  View  Terminal  Tabs  Help
[nurcan@nurcanx91 scripts]$ ls /tmp/install
bitt-1.0        gnuplot-4.4.2.tar     jdk       Python-3.1.3       tmp
bitt-1.0.tar    hadoop-0.20.2         jdk.tar   Python-3.1.3.tar
gnuplot-4.4.2   hadoop-0.20.2.tar.gz   mon       results
[nurcan@nurcanx91 scripts]$ ls /tmp/install/hadoop-0.20.2
bin           docs                        ivy           README.txt
build.xml     hadoop-0.20.2-ant.jar       ivy.xml       src
c++           hadoop-0.20.2-core.jar      lib           webapps
CHANGES.txt   hadoop-0.20.2-examples.jar  librecordio
conf          hadoop-0.20.2-test.jar      LICENSE.txt
contrib       hadoop-0.20.2-tools.jar     NOTICE.txt
[nurcan@nurcanx91 scripts]$ cat /tmp/install/hadoop-0.20.2/conf/masters
nurcanx91
[nurcan@nurcanx91 scripts]$ cat /tmp/install/hadoop-0.20.2/conf/slaves
nurcanx92
nurcanx93
nurcanx90
[nurcan@nurcanx91 scripts]$
```

Figure 6: Format Hadoop* Namenode

```
1280x1024 Desktop                                   LogMeIn - Remote Session
File  Edit  View  Terminal  Tabs  Help
[nurcan@nurcanx91 scripts]$ /tmp/install/hadoop-0.20.2/bin/hadoop namenode -form
at
11/03/03 09:12:59 INFO namenode.NameNode: STARTUP_MSG:
/************************************************************
STARTUP_MSG: Starting NameNode
STARTUP_MSG:   host = nurcanx91/10.129.2.97
STARTUP_MSG:   args = [-format]
STARTUP_MSG:   version = 0.20.2
STARTUP_MSG:   build = https://svn.apache.org/repos/asf/hadoop/common/branches/b
ranch-0.20 -r 911707; compiled by 'chrisdo' on Fri Feb 19 08:07:34 UTC 2010
************************************************************/
11/03/03 09:12:59 INFO namenode.FSNamesystem: fsOwner=nurcan,nurcan
11/03/03 09:12:59 INFO namenode.FSNamesystem: supergroup=supergroup
11/03/03 09:12:59 INFO namenode.FSNamesystem: isPermissionEnabled=true
11/03/03 09:12:59 INFO common.Storage: Image file of size 96 saved in 0 seconds.
11/03/03 09:12:59 INFO common.Storage: Storage directory /tmp/install/tmp/hadoop
-nurcan/dfs/name has been successfully formatted.
11/03/03 09:12:59 INFO namenode.NameNode: SHUTDOWN_MSG:
/************************************************************
SHUTDOWN_MSG: Shutting down NameNode at nurcanx91/10.129.2.97
************************************************************/
[nurcan@nurcanx91 scripts]$
```

Figure 7: Start Hadoop* Servers

```
1280x1024 Desktop                                   LogMeIn - Remote Session
File  Edit  View  Terminal  Tabs  Help
[nurcan@nurcanx91 scripts]$ /tmp/install/hadoop-0.20.2/bin/start-all.sh
starting namenode, logging to /tmp/install/hadoop-0.20.2/bin/../logs/hadoop-nurc
an-namenode-nurcanx91.out
nurcanx90: starting datanode, logging to /tmp/install/hadoop-0.20.2/bin/../logs/
hadoop-nurcan-datanode-nurcanx90.out
nurcanx93: starting datanode, logging to /tmp/install/hadoop-0.20.2/bin/../logs/
hadoop-nurcan-datanode-nurcanx93.out
nurcanx92: starting datanode, logging to /tmp/install/hadoop-0.20.2/bin/../logs/
hadoop-nurcan-datanode-nurcanx92.out
nurcanx91: starting secondarynamenode, logging to /tmp/install/hadoop-0.20.2/bin
/../logs/hadoop-nurcan-secondarynamenode-nurcanx91.out
starting jobtracker, logging to /tmp/install/hadoop-0.20.2/bin/../logs/hadoop-nu
rcan-jobtracker-nurcanx91.out
nurcanx90: starting tasktracker, logging to /tmp/install/hadoop-0.20.2/bin/../lo
gs/hadoop-nurcan-tasktracker-nurcanx90.out
nurcanx93: starting tasktracker, logging to /tmp/install/hadoop-0.20.2/bin/../lo
gs/hadoop-nurcan-tasktracker-nurcanx93.out
nurcanx92: starting tasktracker, logging to /tmp/install/hadoop-0.20.2/bin/../lo
gs/hadoop-nurcan-tasktracker-nurcanx92.out
[nurcan@nurcanx91 scripts]$
```

Figure 8: Generate Hadoop* Test Data

```
1280x1024 Desktop                                   LogMeIn - Remote Session
File  Edit  View  Terminal  Tabs  Help

[nurcan@nurcanx91 scripts]$ /tmp/install/hadoop-0.20.2/bin/hadoop jar /tmp/insta
ll/hadoop-0.20.2/hadoop-0.20.2-examples.jar teragen -D mapred.map.tasks=100 1000
00000 tera
Generating 100000000 using 100 maps with step of 1000000
11/03/03 09:28:22 INFO mapred.JobClient: Running job: job_201103030927_0001
11/03/03 09:28:23 INFO mapred.JobClient:  map 0% reduce 0%
11/03/03 09:28:38 INFO mapred.JobClient:  map 2% reduce 0%
11/03/03 09:28:39 INFO mapred.JobClient:  map 3% reduce 0%
11/03/03 09:28:41 INFO mapred.JobClient:  map 5% reduce 0%
11/03/03 09:28:44 INFO mapred.JobClient:  map 8% reduce 0%
11/03/03 09:28:47 INFO mapred.JobClient:  map 12% reduce 0%
11/03/03 09:28:50 INFO mapred.JobClient:  map 15% reduce 0%
11/03/03 09:28:54 INFO mapred.JobClient:  map 17% reduce 0%
11/03/03 09:28:56 INFO mapred.JobClient:  map 21% reduce 0%
11/03/03 09:28:57 INFO mapred.JobClient:  map 22% reduce 0%
11/03/03 09:28:59 INFO mapred.JobClient:  map 23% reduce 0%
11/03/03 09:29:00 INFO mapred.JobClient:  map 24% reduce 0%
11/03/03 09:29:02 INFO mapred.JobClient:  map 27% reduce 0%
11/03/03 09:29:03 INFO mapred.JobClient:  map 28% reduce 0%
11/03/03 09:29:06 INFO mapred.JobClient:  map 29% reduce 0%
11/03/03 09:29:08 INFO mapred.JobClient:  map 32% reduce 0%
11/03/03 09:29:09 INFO mapred.JobClient:  map 33% reduce 0%
11/03/03 09:29:11 INFO mapred.JobClient:  map 35% reduce 0%
```

Figure 9: Run TeraSort Benchmark

```
1280x1024 Desktop                                   LogMeIn - Remote Session
File  Edit  View  Terminal  Tabs  Help
[nurcan@nurcanx91 scripts]$ /tmp/install/hadoop-0.20.2/bin/hadoop jar /tmp/insta
ll/hadoop-0.20.2/hadoop-0.20.2-examples.jar terasort -D mapred.reduce.tasks=4 te
ra tera-sort
11/03/03 09:35:22 INFO terasort.TeraSort: starting
11/03/03 09:35:22 INFO mapred.FileInputFormat: Total input paths to process : 10
0
11/03/03 09:35:23 INFO util.NativeCodeLoader: Loaded the native-hadoop library
11/03/03 09:35:23 INFO zlib.ZlibFactory: Successfully loaded & initialized nativ
e-zlib library
11/03/03 09:35:23 INFO compress.CodecPool: Got brand-new compressor
Making 4 from 100000 records
Step size is 25000.0
11/03/03 09:35:23 INFO mapred.JobClient: Running job: job_201103030927_0002
11/03/03 09:35:24 INFO mapred.JobClient:  map 0% reduce 0%
11/03/03 09:35:34 INFO mapred.JobClient:  map 1% reduce 0%
11/03/03 09:35:35 INFO mapred.JobClient:  map 3% reduce 0%
11/03/03 09:35:39 INFO mapred.JobClient:  map 5% reduce 0%
11/03/03 09:35:41 INFO mapred.JobClient:  map 8% reduce 0%
11/03/03 09:35:43 INFO mapred.JobClient:  map 10% reduce 0%
11/03/03 09:35:47 INFO mapred.JobClient:  map 12% reduce 0%
11/03/03 09:35:49 INFO mapred.JobClient:  map 14% reduce 0%
11/03/03 09:35:53 INFO mapred.JobClient:  map 14% reduce 1%
11/03/03 09:35:54 INFO mapred.JobClient:  map 15% reduce 2%
```

Figure 10: Run Hadoop* to Display DFS Directories

```
1280x1024 Desktop                                   LogMeIn - Remote Session
File  Edit  View  Terminal  Tabs  Help
11/03/03 09:46:54 INFO mapred.JobClient:     FILE_BYTES_READ=16434149612
11/03/03 09:46:54 INFO mapred.JobClient:     HDFS_BYTES_READ=10003276900
11/03/03 09:46:54 INFO mapred.JobClient:     FILE_BYTES_WRITTEN=26624342992
11/03/03 09:46:54 INFO mapred.JobClient:     HDFS_BYTES_WRITTEN=10000000000
11/03/03 09:46:54 INFO mapred.JobClient:   Map-Reduce Framework
11/03/03 09:46:54 INFO mapred.JobClient:     Reduce input groups=100000000
11/03/03 09:46:54 INFO mapred.JobClient:     Combine output records=0
11/03/03 09:46:54 INFO mapred.JobClient:     Map input records=100000000
11/03/03 09:46:54 INFO mapred.JobClient:     Reduce shuffle bytes=10166370072
11/03/03 09:46:54 INFO mapred.JobClient:     Reduce output records=100000000
11/03/03 09:46:54 INFO mapred.JobClient:     Spilled Records=261022672
11/03/03 09:46:54 INFO mapred.JobClient:     Map output bytes=10000000000
11/03/03 09:46:54 INFO mapred.JobClient:     Map input bytes=10000000000
11/03/03 09:46:54 INFO mapred.JobClient:     Combine input records=0
11/03/03 09:46:54 INFO mapred.JobClient:     Map output records=100000000
11/03/03 09:46:54 INFO mapred.JobClient:     Reduce input records=100000000
11/03/03 09:46:54 INFO terasort.TeraSort: done
[nurcan@nurcanx91 scripts]$ /tmp/install/hadoop-0.20.2/bin/hadoop dfs -ls /user/
nurcan
Found 2 items
drwxr-xr-x   - nurcan supergroup          0 2011-03-03 09:35 /user/nurcan/tera
drwxr-xr-x   - nurcan supergroup          0 2011-03-03 09:46 /user/nurcan/tera-s
ort
[nurcan@nurcanx91 scripts]$
```

Figure 11: Stop Hadoop* Servers

```
1280x1024 Desktop                                   LogMeIn - Remote Session
File  Edit  View  Terminal  Tabs  Help
[nurcan@nurcanx91 scripts]$ /tmp/install/hadoop-0.20.2/bin/stop-all.sh
stopping jobtracker
nurcanx90: stopping tasktracker
nurcanx92: stopping tasktracker
nurcanx93: stopping tasktracker
stopping namenode
nurcanx90: stopping datanode
nurcanx93: stopping datanode
nurcanx92: stopping datanode
nurcanx91: stopping secondarynamenode
[nurcan@nurcanx91 scripts]$
```

Conclusion

Most IT organizations don't want to spend weeks building a multi-vendor virtualized infrastructure. Yet until now, they've had little choice because running IT without virtualization is too costly. The Oracle Optimized Solution for Enterprise Cloud Infrastructure offers the means to greatly simplify deployment and management of a virtualized cloud infrastructure while also reducing risk. Organizations can take full advantage of the hardware and software components in the Oracle Optimized Solution for Enterprise Cloud Infrastructure, thereby realizing greater business benefits from their virtualized infrastructure.

Oracle's innovative technology and open-systems approach make modular systems attractive across a broad set of applications and activities — from consolidating infrastructure through virtualization to deploying dynamic enterprise applications or cloud computing. Oracle's Sun Blade 6000 Modular System and Sun ZFS Storage Appliances provide the promised advantages of modular architecture while retaining essential flexibility for how technology is deployed and managed.

The entire software stack is pre-configured, pre-optimized, and ready-to-run on the recommended hardware platforms. Because all the components have already been tested and integrated together and are supported together, there is less risk of deployment delays or production downtime. With the ability to add compute, storage and networking resources independently, the system can be scaled in multiple dimensions to match evolving changes in demand for applications.

The Intel® Cloud Builders Usage Scenarios section described in this paper shows how Oracle Optimized Solution for Enterprise Cloud Infrastructure can be deployed with configuration of Oracle VM and Oracle Secure Global Desktop to provide a virtualized enterprise-class environment. Oracle VM Server for x86 is a fully supported platform for Oracle software deployment and third party software deployment, runs on Oracle and third party hardware, and is free to use.

Intel and Oracle have implemented a joint Cloud Computing facility, the Intel Oracle Bid Support Center, which provides an example of a private cloud hosted computing environment that is accessible to our partners, customers, and our own staff from anywhere. It is built on Oracle VM Server, Oracle VM VirtualBox, and Oracle Secure Global Desktop.

Oracle Optimized Solution for Enterprise Cloud Infrastructure also simplifies the support structure. Oracle Advanced Customer Services offers a suite of services that maximizes the value of your Oracle Enterprise Cloud infrastructure investments. Through a lifecycle services approach, Oracle experts can help architect customized solutions that reduce deployment risk and ensure operational readiness. There is no need to negotiate multiple support contracts with different vendors, saving time during procurement. Daily operations are also simplified because one vendor can be called for support regardless of which hardware or software component has caused the problem. A single support ticket with Oracle means that customer IT teams do not have to spend their valuable time trying to isolate the right system component before calling Oracle for support.

Additional Considerations

Storage

Intel-based Storage Servers support multiple protocols such as NFS, Common Internet File System (CIFS), Internet Small Computer System Interface (iSCSI), InfiniBand (IB) Technology, and Fibre Channel (FC).

NFS offers the utmost simplicity in attaching storage in Oracle VM virtualization environments over Ethernet. NFS on Intel-based Storage Servers scale to many concurrent I/O threads due to the innovative architecture design of Intel Xeon processors. This high I/O throughput enables more VM stacks to perform I/O without sacrificing service levels.

Architects can also choose FC or iSCSI protocols for storing and accessing both Oracle VM and user data on Oracle's Sun ZFS Storage Appliances. While additional SAN infrastructure is required to use FC interface, NFS and iSCSI can run on the existing LAN infrastructure over 1GbE or 10GbE links. If choosing FC or iSCSI, it is recommended to use OCFS2 clustered file system over the LUNs.

Oracle's Sun ZFS Storage Appliances offer many choices in terms of RAID layout to address capacity, protection, and performance. Mirrored or triple mirrored protection is the recommended RAID layout for Oracle VM storage repository and user data. However, depending on the capacity requirement and the service level agreement, RAID Z2 (double parity RAID) can also be deployed for storage repositories.

The high-performance storage capabilities can be implemented by using Intel solid state drives (SSDs) in some models of the Oracle's Sun ZFS Storage Appliances. Intel Enterprise SSDs enable rapid write capabilities for fast data placement in the storage pool. They optimize I/O rates by providing a fast buffer for reads and/or writes. Both Oracle's Sun ZFS Storage Appliances have read-optimized and write-optimized SSDs, which enable excellent response time and throughput for demanding virtualized environments. This especially boosts VM cache performance because of the low latency and higher performance possible with SSD or flash memory technology. These platforms also have clustering capabilities to provide high availability for the storage.

To address the synchronous write performance of Oracle VM, it is strongly recommended to have two or more write-optimized flash devices per storage pool. To fully utilize the hybrid storage pool model with shorter read response time, it is recommended to use two or more read-optimized flash devices.

One project per Oracle VM server pool with dedicated storage repository is recommended. The shared Oracle VM storage repository (/OVS) could be a file system or a LUN that is seen from the Oracle VM servers in the server pool.

For storing the structured user data such as databases, which are accessed from the virtual machine, it is recommended to match the share record size with the application block size for optimal performance (for example 8kB for OLTP databases). The unstructured data and binary files can be stored on the shares with 128kB.

Intel® Solid-State Drives (SSDs) should be used to improve data intensive application performance by reducing I/O bottlenecks and to minimize power consumption in high density computing environments. Data intensive distributed applications should be optimized by using compression. The compression overhead can be minimized by using optimized libraries such as Intel® Integrated Performance Primitives (Intel® IPP).

The following references explain optimization techniques that can be implemented by using Intel SSDs and compression libraries:

- http://software.intel.com/en-us/articles/hadoop-and-hbase-optimization-for-read-intensive-search-applications/
- http://www.slideshare.net/cloudera/hw09-optimizing-hadoop-deployments
- http://www.cloudera.com/videos/hw10_video_optimizing_hadoop_workloads

Scalability

The following guidelines are recommended for designing and sizing the server pool:

- Plan for excess resource capacity at the pool level to support advanced features such as Live Migration and guest VM HA.
- When determining the number of nodes in a server pool, consider storage topologies and their characteristics as well as network requirements, workload characteristics, and HA needs.
- Consider that using Preferred Server VM placement policies may increase the number of nodes required in a server pool.
- Plan excess capacity according to business requirements for meeting peak loads versus only a proportion of the peak load.
- Memory capacity is the most critical resource. I/O capacity and then CPU capacity are the second and third priorities respectively. CPU and I/O should be balanced given that I/O activity is often CPU intensive.
- The amount of memory required for all running VMs must never exceed the amount physically available on the server(s) in the pool. However, the total amount of memory required by all VMs assigned to the pool (running + powered

off) may exceed the total physical memory since powered off VMs do not consume memory.

- CPU over commitment is supported (e.g. more virtual CPUs configured than physically present).
- The over commitment ratio is dependent upon workload requirements. However, the recommendation is to keep the ratio of virtual CPU to physical CPU at 2:1 or less.
- Determining the best physical server node size for a pool is a combination of factors depending on workload characteristics as well as the operational and budgetary issues of the data center.
- Use identically configured server nodes throughout the pool to support consistent performance and feature set regardless of individual server failure(s). In order to support the best agility and flexibility, it is also recommended that every server node in the pool be identical in capacity and configuration. This approach ensures that no matter where live migration or (re-)start occurs, the same performance and features are uniformly available.

Networking

Oracle VM provides flexible networking configuration that can be used as-is or modified to meet customers' business requirements.

The management domain (Dom0) of the Oracle VM server has direct access to the physical devices, it exports a subset of the devices in the systems to guest domains (DomU). A virtual device driver (also known as front-end driver) appears to the guest operating system as a real device. The network configuration is the same way and will look like a regular host with MAC address, IP address, etc. It can receive I/O requests from its kernel, but since it does not have direct access to the hardware, it must pass those requests to the back-end driver running in Dom0. The back-end driver, in turn, receiving the I/O requests, validates them for safety and isolation. Then it proxies them to the real device. When the I/O operation completes, the back-end driver notifies the front-end driver that the operation was successful and it's ready to continue. The front-end driver then reports the I/O completion to the guest OS kernel.

Oracle Linux and Oracle Solaris come with paravirtualized I/O drivers for improved network throughput and higher disk I/O. Moreover, Oracle VM Windows* PV Drivers, signed by Microsoft* for the Windows Logo Program, are available to improve Windows guest I/O throughput.

In the Oracle Optimized Solution for Enterprise Cloud Infrastructure, the Sun Blade 6000 Modular System is equipped with a redundant pair of switched NEM (Network Express Module) 24p interconnects. Each is dedicated to one head of the Oracle's Sun ZFS Storage Appliances. This solution leverages the bandwidth of 10GbE while ensuring no single points of failure. It also offers numerous unused 10GbE uplinks per NEM for seamless integration into the datacenter.

The Oracle VM server network configuration and data center topology should provide enterprise grade redundancy and failover. This is typically achieved by providing dual paths and switches using bonding on virtual machine servers. These settings can be configured in the Oracle Optimized Solution for Enterprise Cloud Infrastructure.

Additional details about Oracle VM network configurations can be found on the third-party document called, "The Underground Oracle VM Manual".

The Intel Oracle BSC environment has 1Gb networking, however Intel ran four cables between each Intel architecture based blade cabinet and the BSC network switch. This had the benefit of redundancy of the link and also increased the available bandwidth between the blades and the network switch. In our environment this quadrupled the speed when transferring large Oracle VM image files between servers – this reduced transfer and backup times significantly which was very useful for the largest VMs (up to 300GB each). We did not perform further analysis once this benefit was achieved, so this is described for information only.

Processor Description

Sun Blade Server Module Features		
Specification Category	**Sun Blade X6270 M2 Server Modules**	**Sun Blade X6275 M2 10GbE Server Modules**
CPU Power Power	▪ Up to 12 cores ▪ Up to two six-core high-speed Intel® Xeon® Processor 5600 series CPUs	▪ Up to 24 cores ▪ Two compute nodes per blade ▪ Two six-core high-speed Intel® Xeon® Processor 5600 series CPUs per compute node
Main Memory	▪ 18 DIMM slots ▪ Up to 144GB total memory	▪ 12 DIMM slots per compute node ▪ Up to 96GB memory per compute node or a total of 192GB per blade
I/O	▪ Two embedded 10/100/1000-T Ethernet ports for standard GbE networking ▪ Total 282Gb/sec I/O bandwidth to easily handle VM networking requirements	▪ Two models: Integrated 10GbE or 1GbE networking ▪ Support for high IOPS Sun Flash Modules

Table 3

The following sections describe key Intel® processor technologies that enable implementation of virtualized multi-tenant data centers:

Parallelism

Intel-based Enterprise servers include multiple processors and multiple computation engines to enable high-performance computing. Two socket servers are the typical server model for most server applications. Four socket systems are used for the server workloads that require the higher performance. The number of processors per server can be increased if more processing power is needed for the server workload.

Intel® Turbo Boost Technology

Intel® Turbo Boost Technology allows processors to deliver higher speed execution on demand using available power to run at a higher frequency.

Intel® Hyper-Threading Technology

Many server applications lend themselves to parallel, multi-threaded execution. Intel® Hyper-Threading Technology enables simultaneous multi-threading within each processor core, up to two threads per core. Hyper-threading reduces computational latency, making optimal use of every clock cycle. For example, while one thread is waiting for a result or event, another thread is executing in that core, to maximize the work from each clock cycle.

Intel® QuickPath Technology

Intel® QuickPath Technology is a scalable, shared memory architecture that delivers a high memory bandwidth to enable top performance for bandwidth-intensive applications. It provides high speed point-to-point connections between processors, and between processors and the I/O hub. Each processor has its own dedicated memory that it accesses directly through an Integrated Memory Controller. In cases where a processor needs to access the dedicated memory of another processor, it can do so through a high-speed Intel® QuickPath Interconnect (Intel® QPI) that links all the processors.

Intel® Intelligent Power Technology

Within a single server, Intel® Intelligent Power technology minimizes power consumption when server components are not fully utilized. Integrated Power Gates allow individual idling cores to be reduced to near-zero power independent of other operating cores. Automated Low-Power States automatically put processor and memory into the lowest available power states that will meet the requirements of the current workload. Processors are enhanced with more and lower CPU power states, and the memory and I/O controllers have new power management features.

Intel® Virtualization Technology

Intel® Virtualization Technology (Intel® VT) enhances virtualization performance with new hardware assist capabilities across the following elements of the server:

- Processor: Intel VT for IA-32, Intel® 64 and Intel® Architecture (Intel® VT-x) provides hardware assisted page-table management, allowing guest OS more direct access to the hardware and reducing compute-intensive software translation from the VMM. Intel VT-x also includes Intel VT FlexMigration and Intel VT FlexPriority, which are capabilities for flexible workload migration and performance optimization across the full range of 32-bit and 64-bit operating environments.
- Chipset: Intel VT for Directed I/O (Intel® VT-d) helps speed data movement and eliminates much of the performance overhead by giving designated virtual machines their own dedicated I/O devices, thus reducing the overhead of the VMM in managing I/O traffic.
- Network Adapter: Intel VT for Connectivity (Intel® VT-c) enhances server I/O solutions by integrating extensive hardware assists into the I/O devices that are used to connect servers to the data center network and storage infrastructure. Two technologies comprise Intel VT-c: Virtual Machine Device Queues, which accelerates throughput and reduces the load on the VMM and server processors, and PCI-SIG* SR-IOV, which delivers near-native throughput and provides dedicated, direct connectivity between VMs and hardware resources.

Intel® AES New Instructions (Intel® AES-NI) Technology

Intel® Advanced Encryption Standard Technology enables robust encryption without needing additional appliances and increased performance overhead. This technology improves CPU performance for encryption by as much as 52 percent for secure Internet transactions and allows broader use of encryption throughout the data center.

Intel® Trusted Execution Technology (Intel® TXT)

Intel® Trusted Execution Technology (Intel® TXT) addresses a critical security need for all server deployments, especially virtualization and cloud-based use models by helping protect servers prior to OS launch or hypervisor launch. Intel® TXT complements other malware protections such as anti-virus and intrusion detection to help ensure that only trusted software is on the platform. VMs on trusted platforms are also protected, so they can easily be migrated onto other trusted platforms or created pools of platforms with trusted hypervisors.

Intel Web Resources

Intel Web Resources	
Intel® Xeon® Processors	http://www.intel.com/xeon
Intel® Solid-State Drives	http://www.intel.com/design/flash/nand/index.htm
Intel® Integrated Performance Primitives (Intel® IPP)	http://software.intel.com/en-us/articles/intel-ipp/
Intel Optimization References	http://software.intel.com/en-us/articles/hadoop-and-hbase-optimization-for-read-intensive-search-applications/ http://www.slideshare.net/cloudera/hw09-optimizing-hadoop-deployments http://www.cloudera.com/videos/hw10_video_optimizing_hadoop_workloads
Intel Cloud Builders	http://www.intel.com/cloudbuilders
Intel Oracle Bid Support Center	http://software.intel.com/file/18807

Table 4

Oracle Web Resources

For more information, visit the Web resources listed in Table 5 and the white papers listed in Table 6.

Oracle* Web Resources	
Web Resource Description	**Web Resource URL**
Oracle* Cloud	http://www.oracle.com/goto/cloud
Oracle Optimized Solutions	http://www.oracle.com/us/solutions/optimized-solutions-171609.html
Oracle Optimized Solution for Enterprise Cloud Infrastructure	http://www.oracle.com/us/solutions/enterprise-cloud-infrastructure-405145.html
Oracle Sun Blade Modular Systems	http://www.oracle.com/us/products/servers-storage/servers/blades/index.html
Oracle Sun x86 systems	http://www.oracle.com/us/products/servers-storage/servers/x86/index.html
Sun Networking	http://www.oracle.com/us/products/servers-storage/networking/
Oracle Solaris*	http://oracle.com/solaris
Oracle Linux*	http://oracle.com/linux
Oracle Virtualization	http://oracle.com/virtualization
Oracle VM Templates	http://www.oracle.com/technetwork/server-storage/vm/templates-101937.html
Sun ZFS Storage Appliances	http://www.oracle.com/us/products/servers-storage/storage/unified-storage/
Oracle Enterprise Manager	http://www.oracle.com/us/products/enterprise-manager/
Oracle Enterprise Manager Ops Center	http://www.oracle.com/us/products/enterprise-manager/opscenter/
Oracle Premier Support for Systems	http://www.oracle.com/us/support/systems/premier/
Oracle Advanced Customer Services	http://www.oracle.com/us/support/advanced-customer-services/index.html
Amazon Web Services* and Oracle Announce Certification and Support of Oracle Applications, Middleware and Databases on Amazon Elastic Compute Cloud using Oracle VM	http://www.oracle.com/us/corporate/press/173480
Oracle on Amazon EC2	http://aws.amazon.com/solutions/global-solution-providers/oracle/
Oracle VM – Creating & Maintaining a Highly Available Environment for Guest VM	http://www.oracle.com/us/technologies/026999.pdf
The Underground Oracle VM Manual	http://itnewscast.com/underground-oracle-vm-manual
Oracle VM Templates Website	http://www.oracle.com/technetwork/server-storage/vm/templates-101937.html
Production Workloads in Public Clouds	http://aws.amazon.com/solutions/global-solution-providers/oracle/

Table 5

Related White Papers	
White Paper Title	**Web URL**
Architecting Oracle VM solution with Sun ZFS Storage Appliances and Sun Servers	http://www.oracle.com/technetwork/articles/systems-hardware-architecture/vm-solution-using-zfs-storage-174070.pdf
Sun Blade 6000 Modular Systems from Oracle	http://www.oracle.com/technetwork/articles/systems-hardware-architecture/sb6000-architecture-163892.pdf
Creating and Using Oracle VM Templates: The Fastest Way to Deploy Any Enterprise Software	http://www.oracle.com/us/technologies/027001.pdf
Oracle VM - Creating & Maintaining a Highly Available Environment for Guest VMs	http://www.oracle.com/us/products/servers-storage/servers/blades/index.html

Table 6

Follow Oracle x86 Infrastructure technology and solutions at:

- Oracle systems blog: http://blogs.oracle.com/infrared/
- Oracle InfraRed on Twitter: http://twitter.com/ORCL_infrared/
- Oracle InfraRed on LinkedIn: http://www.linkedin.com/
- Oracle InfraRed on Facebook: http://www.facebook.com/pages/Oracle-Infrared/
- Oracle InfraRed on OracleMix: https://mix.oracle.com/groups/

Glossary

Intel® Data Center Manager (Intel DCM):

Intel DCM is an SDK from Intel that provides policy based tools for managing power in the data center. Used in conjunction with Intel® Intelligent Node Manager and integrated with management consoles, Intel DCM provides benefits such as increased rack density within space, power and cooling constraints through fine-grained power control, and reduced capital costs by right-sizing power and cooling infrastructure based on actual power consumption trends.

Data Center Manageability Interface (DCMI):

DCMI Specifications are derived from Intelligent Platform Management Interface (IPMI) 2.0, which has been widely adopted by the computing industry for server management and system-health monitoring. The DCMI specifications define a uniform set of monitoring, control features, and interfaces that target the common and fundamental hardware management needs of server systems that are used in large deployments within data centers, such as Internet Portal Data Centers. This includes capabilities such as secure power and reset control, temperature monitoring, event logging, and others.

Hardware profile:

Represents a non-overlapping, logical group of nodes that are provisioned with the identical set of kernel, network configuration, disk partitioning, and device modules.

Infrastructure as a Service (IaaS):

Infrastructure as a Service is the delivery of technology infrastructure —such as network, storage, and compute—as a service, typically through virtualization. Users subscribe to this virtual infrastructure on demand as opposed to purchasing servers, software, data center space, or network equipment. Billing is typically based on the resources consumed.

Software profile:

Defines non-overlapping, logical groups of nodes that are provisioned with identical sets of software.

Power Manager:

The function that is responsible for managing the power utilization in the cloud data center.

VM:

Virtual Machine.

Intel® Virtualization Technology (Intel® VT):

Provides comprehensive hardware assists, which boost virtualization software performance and improve application response times. Intel VT reduces demands placed on virtualization software so that you can consolidate more applications and heavier workloads per server to get better value from your server and software investments.

Disclaimers

Intel processor numbers are not a measure of performance. Processor numbers differentiate features within each processor family, not across different processor families. See www.intel.com/products/processor_number for details.

Intel® Turbo Boost Technology: Requires a system with Intel® Turbo Boost Technology capability. Consult your PC manufacturer. Performance varies depending on hardware, software and system configuration. For more information, visit http://www.intel.com/technology/turboboost

Intel® Hyper-Threading Technology: Requires an Intel® HT Technology enabled system, check with your PC manufacturer. Performance will vary depending on the specific hardware and software used. Not available on Intel® Core™ i5-750. For more information including details on which processors support HT Technology, visit http://www.intel.com/info/hyperthreading

Intel® Trusted Execution Technology: No computer system can provide absolute security under all conditions. Intel® Trusted Execution Technology (Intel® TXT) requires a computer system with Intel® Virtualization Technology, an Intel TXT-enabled processor, chipset, BIOS, Authenticated Code Modules and an Intel TXT-compatible measured launched environment (MLE). Intel TXT also requires the system to contain a TPM v1.s. For more information, visit http://www.intel.com/technology/security

Intel® Virtualization Technology: Intel® Virtualization Technology requires a computer system with an enabled Intel® processor, BIOS, virtual machine monitor (VMM). Functionality, performance or other benefits will vary depending on hardware and software configurations. Software applications may not be compatible with all operating systems. Consult your PC manufacturer. For more information, visit http://www.intel.com/go/virtualization

Intel® AES-NI requires a computer system with an AES-NI enabled processor, as well as non-Intel software to execute the instructions in the correct sequence. AES-NI is available on select Intel® processors. For availability, consult your reseller or system manufacturer. For more information, see http://software.intel.com/en-us/articles/intel-advanced-encryption-standard-instructions-aes-ni/

INFORMATION IN THIS DOCUMENT IS PROVIDED IN CONNECTION WITH INTEL® PRODUCTS. NO LICENSE, EXPRESS OR IMPLIED, BY ESTOPPEL OR OTHERWISE, TO ANY INTELLECTUAL PROPERTY RIGHTS IS GRANTED BY THIS DOCUMENT. EXCEPT AS PROVIDED IN INTEL'S TERMS AND CONDITIONS OF SALE FOR SUCH PRODUCTS, INTEL ASSUMES NO LIABILITY WHATSOEVER, AND INTEL DISCLAIMS ANY EXPRESS OR IMPLIED WARRANTY, RELATING TO SALE AND/OR USE OF INTEL PRODUCTS INCLUDING LIABILITY OR WARRANTIES RELATING TO FITNESS FOR A PARTICULAR PURPOSE, MERCHANTABILITY, OR INFRINGEMENT OF ANY PATENT, COPYRIGHT OR OTHER INTELLECTUAL PROPERTY RIGHT. UNLESS OTHERWISE AGREED IN WRITING BY INTEL, THE INTEL PRODUCTS ARE NOT DESIGNED NOR INTENDED FOR ANY APPLICATION IN WHICH THE FAILURE OF THE INTEL PRODUCT COULD CREATE A SITUATION WHERE PERSONAL INJURY OR DEATH MAY OCCUR.

Intel may make changes to specifications and product descriptions at any time, without notice. Designers must not rely on the absence or characteristics of any features or instructions marked "reserved" or "undefined." Intel reserves these for future definition and shall have no responsibility whatsoever for conflicts or incompatibilities arising from future changes to them. The information here is subject to change without notice. Do not finalize a design with this information.

The products described in this document may contain design defects or errors known as errata which may cause the product to deviate from published specifications. Current characterized errata are available on request. Contact your local Intel sales office or your distributor to obtain the latest specifications and before placing your product order. Copies of documents which have an order number and are referenced in this document, or other Intel literature, may be obtained by calling 1-800-548-4725, or by visiting Intel's Web site at www.intel.com.