An Oracle White Paper
September 2010

Architecting the Oracle VM solution using the Oracle Sun ZFS Storage Appliances and Oracle Sun Servers

ORACLE

Introduction

The impact of virtualization within the datacenter has shaken IT at its core — changing the dated deployment model of assigning one physical server for each and every application. With virtualization and its ability to install multiple operating systems (OSs) and run them independently on a single server, IT can effectively consolidate hundreds of servers while optimizing resources and simplifying the infrastructure. This consolidation and optimization of resources clearly helps IT reduce datacenter costs by lowering physical server count, energy expenses, and administrative overhead. Additionally, virtualized environments help IT reduce complexity, improve resiliency, and provide high availability.

While servers can be consolidated to improve utilization and reduce server count, storage does not necessarily follow suit. Storage demand, even though it can be a shared resource in a virtualized environment, continues to grow and shows no signs of being able to be reduced. In fact, virtual environments require network storage for virtual machine high availability and live migration. A successful virtualization initiative must take into consideration both server virtualizations as well as proper selection of storage.

With an emphasis on storage, this document is a primer to understanding the technical value of Sun hardware in Oracle VM environments. In particular, this paper provides key guidelines for storage architects to plan for Oracle virtualization utilizing a reference configuration based on Oracle's Sun ZFS Storage Appliances and Sun Fire servers.

Benefits of Oracle's Virtualization

Oracle's virtualization strategy aims to simplify, manage, and support the enterprise stack from applications to disks for Oracle and non-Oracle products. Oracle VM is a key technology to fulfill this strategy — offering a highly available and high performance hypervisor that has a cost-effective zero license with 24x7 premier quality support for the full range of Oracle products. From a storage perspective, Sun ZFS Storage Appliance complements Oracle VM virtualization for the data of virtualized applications and operating environments. The Sun ZFS Storage Appliance is designed to automatically enhance Oracle VM environments with Flash speed and become cost-effective with the license-free model for key data storage software. Companies choosing both of these technologies can greatly benefit from a holistic software and hardware stack that is designed for performance, availability, and simplicity.

About Oracle VM

Oracle VM is a free, next-generation server virtualization and management solution available for download and use that makes enterprise applications — both Oracle and non-Oracle applications — easier to deploy, manage, and support. There are no licensing costs for Oracle VM; the only cost is the cost of supporting the technology.

As shown in Figure 1, Oracle VM consists of one or more Oracle VM Servers for x86 that are grouped into a Server Pool and managed by Oracle VM Manager. Oracle VM Server allows hardware resources — CPUs, memory, networks and storage — to become flexible resource pools from which virtual machines are built. Oracle VM Manager provides an easy-to-use graphical interface to create and manage virtual infrastructure. Oracle VM is backed by the Oracle worldwide technical support organization and is fully certified for all Oracle applications.

Virtual machines (VMs) share physical servers but behave like independent physical servers. Each virtual machine created with Oracle VM has its own virtual CPUs, network interfaces, storage, and operating system. Oracle VM Manager lets you create, clone, share, configure, boot and migrate Oracle virtual machines.

Oracle VM Manager

Manager connects with one or more Oracle VM Servers to create, configure, and manage virtual resources. Virtual machines can be created and deployed using shares of physical resources such as CPU, memory, network and disk.

Oracle VM Manager's capabilities streamline tasks that are otherwise highly manual and time-intensive to significantly reduce data center complexity. For example, users can consolidate many physical servers using virtual machines, quickly create new virtual machines to deploy operating systems and applications using Oracle VM Templates, and reduce the cost and complexity of VM failover.

Oracle VM Server

A self contained virtualization environment designed to provide an agile, secure, and server-based platform for running guest virtual machines. Based on open technology, Oracle VM Server includes a VM agent which enables the communication channel with the Oracle VM Manager for management purposes. One or more servers are configured in a single "*server pool*" with one server pool master assigned. That is done for high availability purposes where if a VM server fails, another server takes over. The other roles the server can be assigned are the utility server, and virtual machine server.

Oracle VM Templates

Oracle VM Templates provide an innovative approach to deploying a fully configured software stack by offering pre-installed and pre-configured software images. Oracle VM Templates of many key Oracle products are available for download, including Oracle Database, Oracle Enterprise Linux, Oracle Fusion Middleware, etc., thus providing the benefit of gold images of stacks that are well tested, configured, and popular.

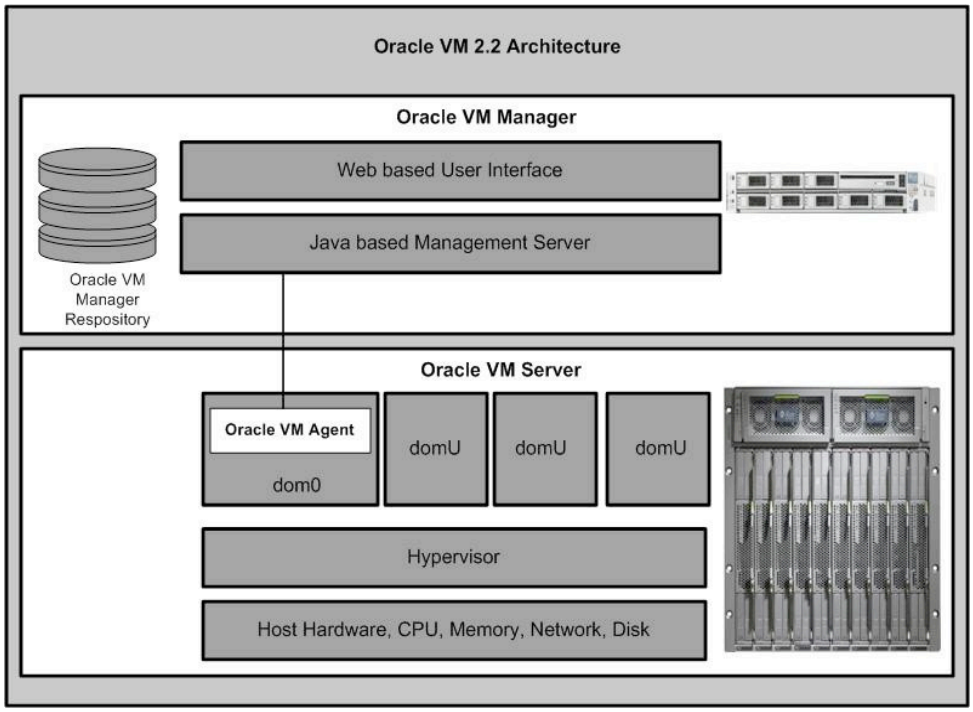

Figure 1. Oracle VM architecture

For additional information on Oracle VM, refer to the resources section at the end of this document.

About Sun ZFS Storage Appliance

The Sun ZFS Storage Appliance offers new innovations in storage that includes Flash based hybrid storage pools, enterprise-class data services, massive scale, and a choice of storage protocols, while delivering significant cost savings. These systems feature a common, easy-to-use management interface that requires no additional training and has the industry's most comprehensive analytics environment to help isolate and resolve issues to minimize business impact.

Oracle offers four models of Sun ZFS Storage Appliance to meet the scalability and availability needs of today's demanding applications. These models include the 7120, 7320/7320C, 7420/7420C, and 7720. All of them utilize a common storage software foundation. The models excluding 7120 offer up to 2TB of read cache which enables the appliances to have a response time typically in low single digit milliseconds. The write Flash on all four platforms provide response time for the synchronous writes with < 1ms response time.

The new Sun ZFS Storage Appliance platforms offer faster CPUs, bigger flash cache, larger storage capacity, better throughput, and broader bandwidth to meet the storage requirements of mission critical applications.

The following two sections are intended to help storage architects drive the design decision for protocol choice and capacity planning.

Storage Protocol Choice for Oracle VM

Sun ZFS Storage Appliance supports multiple protocols such as NFS, Common Internet File System (CIFS), Internet Small Computer System Interface (iSCSI), InfiniBand (IB), and Fibre Channel (FC). Invented by Sun, NFS is one of the most successful examples of open network file sharing. It is designed to be both vendor neutral and operating system-type neutral, and integrates file access, file locking, and mount protocols into a single, unified protocol to ease traversal through a firewall and to improve security. NFS offers the utmost simplicity in attaching storage in Oracle VM virtualization environments over Ethernet. NFS on Sun ZFS Storage Appliance scales to many concurrent I/O threads due to the innovative architecture design, thus enabling 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 also user data on the ZFS Storage System. While additional SAN infrastructure is required to use FC interface, NFS and iSCSI can run on the existing LAN infrastructure over 1Gb or 10Gb links.

If choosing FC or iSCSI, it is recommended to use OCFS2 clustered filesystem over the LUNs.

Storage Capacity and Architecture Planning

To meet a variety of needs for capacity, price, and performance, Sun ZFS Storage Appliance is available in different configurations that include a high availability (HA) cluster configurations such as the Sun ZFS Storage 7320 System, Sun ZFS Storage 7420 System, and Sun ZFS Storage 7720 System. Under a clustered configuration, the head node resources fail over to the surviving node in the event of node crash, thus helping to ensure near continuous availability. Key hardware component specifications of the Sun ZFS Storage Appliance are listed in Table 1.

Table 1: Sun ZFS Storage Appliance Models

Platform	Storage Capacity	Processor	Memory (DRAM)	Write Optimized SSD	Read Optimized SSD	Cluster Option
Sun ZFS Storage 7120	Up to 60 x 2TB SAS Disks [120TB]	1 x Quad Core Intel Westmere EP E5620 @ 2.4GHz	Up to 36GB	Up to 96GB	N/A	N
Sun ZFS Storage 7320 (details are per controller)	Up to 96 x 2TB SAS Disks [192TB]	2 x Quad Core Intel Westmere EP E5620 @ 2.4GHz	Up to 72GB	Up to 16 x 18GB	Up to 4 x 512GB	Y
Sun ZFS Storage 7420 (details are per controller)	Up to 576 x 2TB SAS Disks [1.1PB]	4 x 6C Intel Nehalem EX E7530 @ 1.86GHz [or] 4 x 8C Intel Nehalem EX X7550 @ 2GHz	Up to 512GB	Up to 96 x 18GB	Up to 4 x 512GB	Y
Sun ZFS Storage 7720	Expandable racks. Each Rack 720 TB	4 x 8C Intel Nehalem EX X7550 @ 2GHz	Up to 512GB per controller	2 x 18GB per cage	Up to 4 x 512GB per controller	Y

Sun ZFS Storage High Availability

Clustering with Sun ZFS Storage 7320C, 7420C, or 7720 Appliances provide an optimal solution for sites with stringent availability requirements. With *active/passive* clustered configurations of the Sun ZFS Storage 7320 System, the Sun ZFS Storage 7420 System, or the Sun ZFS Storage 7720 System, a second server node functions as a hot spare — providing failover and protection in the event of the loss of the primary server node. Active/active configurations are also supported in a configuration that serves two different storage pools to the clients.

Sun ZFS Storage Appliance Benefits for Oracle VM

Hybrid Storage Pool Using Flash Media

Sun ZFS Storage Appliance implements a Hybrid Storage Pool (HSP) architecture (Figure 3) that dramatically reduces storage management by dynamically and intelligently caching active data over multiple tiers of storage media — greatly benefiting virtualized environments. Data is automatically placed and serviced from two tiers of cache and supported by high capacity physical disks where:

- The first tier is DRAM primary cache (ZFS – Adaptive Replacement Cache (ARC)) with response time < 1 ms
- The second tier is a read-optimized flash drive for the ZFS L2ARC (an extension to the ARC) with response time < 5 ms
- Physical disk with response times < ~50ms

The hybrid storage pool provides a solid state response time with spinning disk capacity, providing an ideal virtualized environment as both image and user data are automatically cached and served for optimal performance.

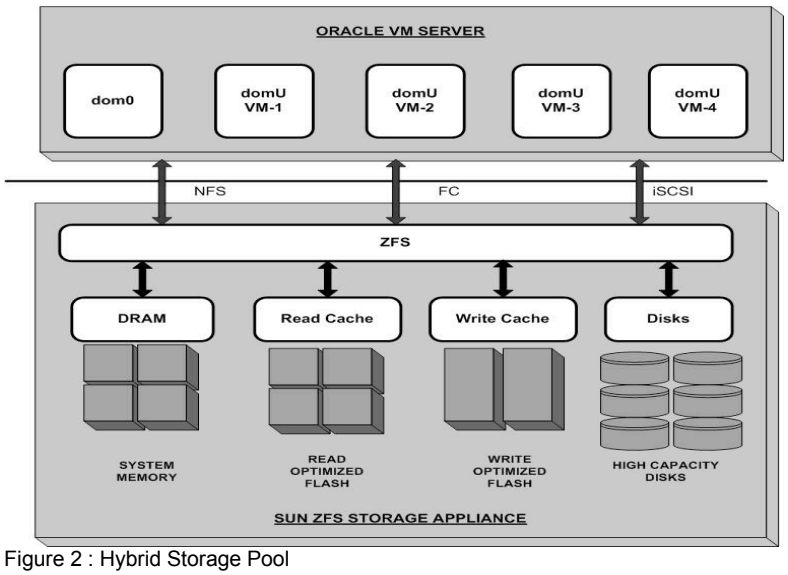

Figure 2 : Hybrid Storage Pool

Data Protection

Sun ZFS Storage Appliance includes a set of comprehensive data protection features that provide instant snapshots, cloning, rollback, and remote replication — making data protection and maintenance tasks easier and restores faster.

- *Snapshots*: Snapshots are point-in-time copies of data. An unlimited number of snapshots can be taken from the source. The snapshot mechanism uses a copy-on-write method and does not occupy initial space until data is written. As new data write, the old copy writes to the snapshot. Each snapshot can be made visible and can be accessed read-only.
- *Cloning*: A clone is a read-write copy of the snapshot. Clones can be created from a snapshot and can be used in a VM environment where the boot images can be cloned many times and accessed. Snapshots and clones don't require any space allocation when they are created but will start consuming space as when the data is changed.
- *Rollback*: As a faster mechanism for restores, shares can be rolled back from snapshots — bringing all data in the share back to the point-in-time when the snapshot was taken. Cloning and rollback of OS images in the VM environment are an excellent use of this feature.
- *Remote Replication*: Data is asynchronously replicated from one ZFS Storage System (source) to one or more remote sites (targets) for the purpose of disaster recovery in the event of failure at the primary site or scheduled maintenance. This feature allows for easy failover, role reversal, and failback operations as well as the ability to utilize target sites for test, development, and backup purposes.

Analytics

The Solaris Dynamic Tracing (DTrace) based advanced graphical user interface can be used to monitor and analyze the performance of Sun ZFS Storage Appliance. Analytics provide real time statistics for throughput, capacity, and utilization analytics for items such as network, CPU, memory (ARC), flash (L2ARC), and storage. In the Oracle VM environment, individual virtual machine activities can be monitored.

For some samples,

a) Figure 3 shows the graphical representation of number of operations performed over the FC luns.

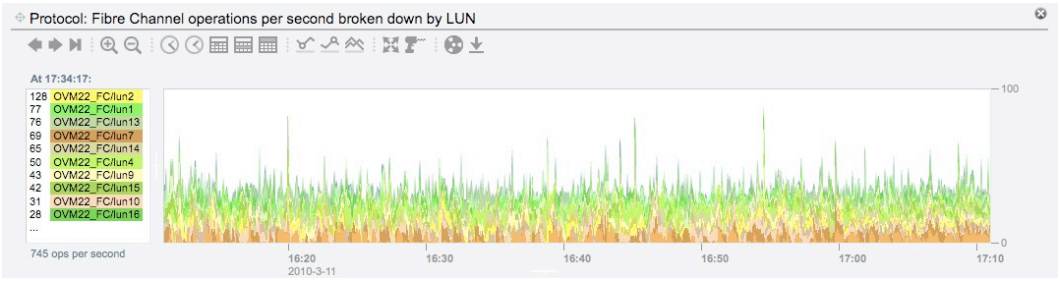

Figure 3:Analytics screenshot of Fibre channel IOPS broken down by LUN

b) Figure 4 shows which clients are accessing the storage using NFSv3 protocol

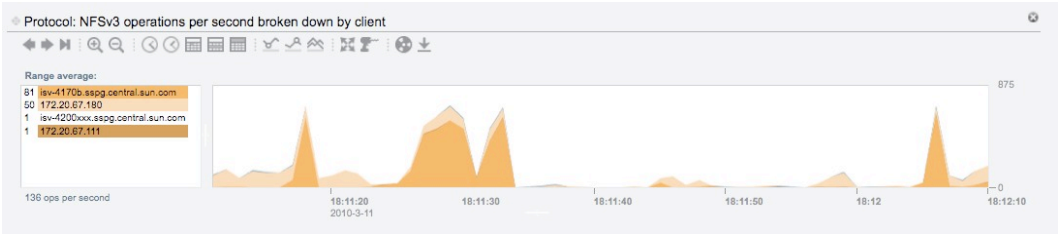

Figure 4 :. NFSv3 operations

c) Figure 5 shows the analytics screenshot regarding which files are being accessed.

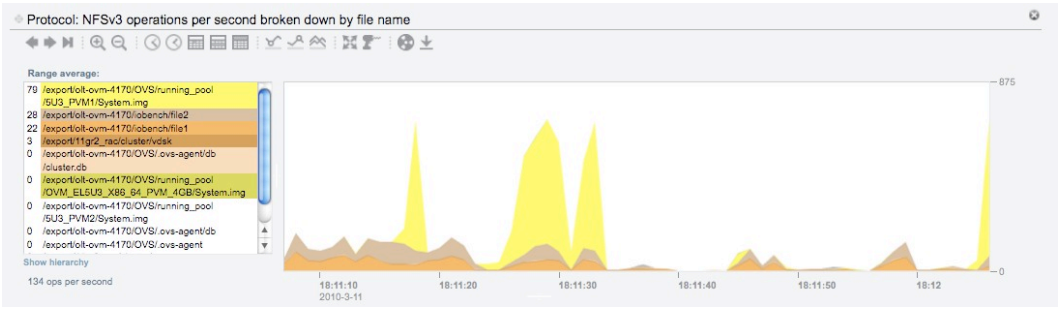

Figure 5 :.NFSv3 operations broken down by file name

Reference Architectures

This section provides two Oracle VM deployment scenarios using Sun Servers and Sun ZFS Storage appliance. Architecture 1 discusses a typical enterprise deployment model using high end servers and storage. Architecture 2 describes small and medium business deployment using medium configured servers and storage.

Architecture - 1

An example reference architecture, illustrated in figure 6, offers high performance, availability, reliability, and manageability specifically for database virtual machines. While these requirements can be met in numerous ways, this reference architecture implements an optimal approach. The following sections detail the servers component and virtual machine allocation, storage configuration, project allocation with one of the high end model Sun ZFS Storage 7420 Clustered System, and network configuration.

Either blade-based or rack-mounted Sun x64 servers can be used to run the Oracle VM infrastructure in a large, high availability configuration. As an example, either a series of AMD Opteron™ or Intel® Xeon® processor-based servers (such as multiple Intel Xeon processor-based Sun Fire™ X4170 servers) or a Sun Blade™ 6000 server (with 10 Sun Blade X6270 server modules) can be configured to run the Oracle VM software. Each server is attached to a clustered Sun ZFS Storage shared storage that is accessed using NFS, or via FC/iSCSI lun enabling additional features such as live migration, high availability, distributed resources scheduling, etc. Set of servers identified as part of a single server pool is recommended to configured in a similar way so that the service level agreement can be met during fail-overs or migration.

Oracle VM can be installed on bare metal x64 systems and is supported on current model Sun x64 rack-mount or blade servers with multi-core CPUs, a minimum of 32 GB of memory such as :

- Sun Fire X4170 server
- Sun Fire X4270 server
- Sun Fire X4450 server
- Sun Fire X4440 server
- Sun Fire X4600 M2 server
- Sun Blade x64 server modules with AMD or Intel processor blades (such as Sun Blade 6000 and Sun X6240 or Sun X6270 server modules)

The Sun Blade server modules include two 1 GbE NICs. Two shared 10 GbE ports per blade server can be installed using two Sun Blade 6000 Virtualized Multi-Fabric Network Express Modules. The two 1 GbE network ports for each blade can be used for the management purposes. The remaining two 10 GbE interfaces can be used for NFS services to the Sun ZFS Storage 7420

System for the virtual machines, supplying plenty of bandwidth and cable aggregation for virtual machine data. For NFS and iSCSI access, 10 GbE ports are used.

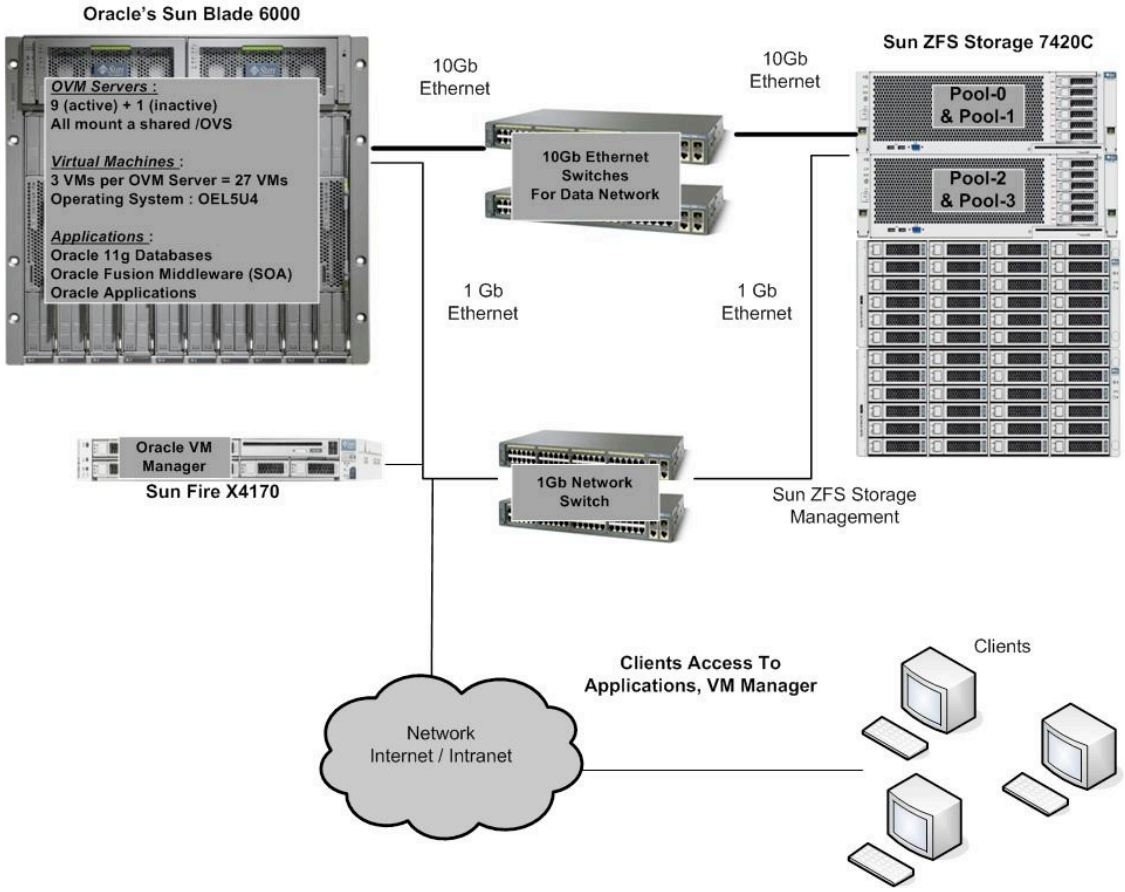

Figure 6 : HA Architecture for Oracle VM environment

For each virtual machine, the number of virtual CPUs and memory is dependent on the kind of applications that are deployed. In this example architecture, minimum of 4GB and 2 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 4 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 and the type and number of network links are some of the critical aspects to consider for a successful deployment. For example, for certain IO load over a single 1Gb link, the network could be saturated first before either the server or the storage. In such case, adding more links or moving to one or more10Gb link may provide a better performance. If the CPUs are pegged running multiple virtual machines on the same host, either add additional servers to the server pool and live migrate the virtual machines to balance the Pool or the Oracle VM infrastructure can be moved from the mid-range server to an enterprise class server.

For the high availability and optimal performance, Sun ZFS Storage 7420 clustered system is used as the storage platform. Few storage pools are created as per Table 2 :

Table 2 : Oracle VM Pool and Projects for High-performance

Pool Name	RAID	Sun ZFS Storage 7420 Cluser Head	System / Virtual Machine	Purpose
pool-0	RAID-Z2	HEAD-1	Dom0	Accessed from domain 0 for storing the Oracle VM storage repository. This contains the OS images from which the virtual machines are launched.
pool-1	MIRRORED	HEAD-1	Virtual machines that runs the database	For storing database files and database binaries. One or more projects are created to cater for 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 the binaries, 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.

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

Pool-0 : One OVS 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, more number of disks are to be allocated for this pool – as the random reads (db sequential reads) benefit from more disks during read-miss from the cache scenarios.

Pool-2 : One or more projects are created for storing the various middleware components – 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).

In this example architecture, all the 27 virtual machines are active and the various applications access Sun ZFS Storage 7420 clustered system via redundant data switch. This end-to-end high available architecture provide a high performance infrastructure for the demanding enterprise needs.

Refer to the mid-level deployment architecture for an example file system configuration and mapping details.

Architecture - 2

In this example, an mid-level Oracle VM high available architecture is deployed by placing the Oracle's application, middleware, and the database layer in a single Oracle VM server. This is more suitable towards small to medium deployments where the application processing and IO requirement can be met with mid-range servers and storage. In Figure 7, the Sun ZFS Storage 7320 file system mapping to the NFS mount points in the various virtual machines are depicted. In this example, one Sun Fire X4170 is used as the Oracle VM 2.2 server. Three OEL5U4 virtual machines are configured in the Oracle VM server. One virtual machine is for the database layer, one virtual machine for the middleware layer and one virtual machine for the application layer.

Four projects are created in the Sun ZFS Storage 7320 clustered system. Two storage pools *pool-0* (RAIDZ2) and *pool-1* (Mirrored) are created. The storage pool *pool-0* is for storing the OVS repositories and OS images. *Pool-1* is for the user data. Pool-0 is active on node-1 of the Sun ZFS Storage 7320 Clustered head. Pool-1 is active on node-2.

To enable high availability of the Oracle VM, another Sun Fire X4170 is included in this architecture as part of the same server pool. The server can be used for the live migration or activated Oracle VM Server in the event of failure of the production Sun Fire X4170 server.

In the event of any failure in the production Oracle VM server, the virtual machines are restarted in the second Oracle VM server. . Since all the OS images are stored on the shared storage, there is no complexity in terms of additional configuration is required in bringing up the new virtual machine(s) in this server.

For load sharing condition, one or more virtual machines are migrated to the second node using the *live migration* feature. Since the user data and the OVS repositories are stored in the shared storage, by executing a simple xm command, the virtual machine is migrated from one server to another and thus enhances the performance.

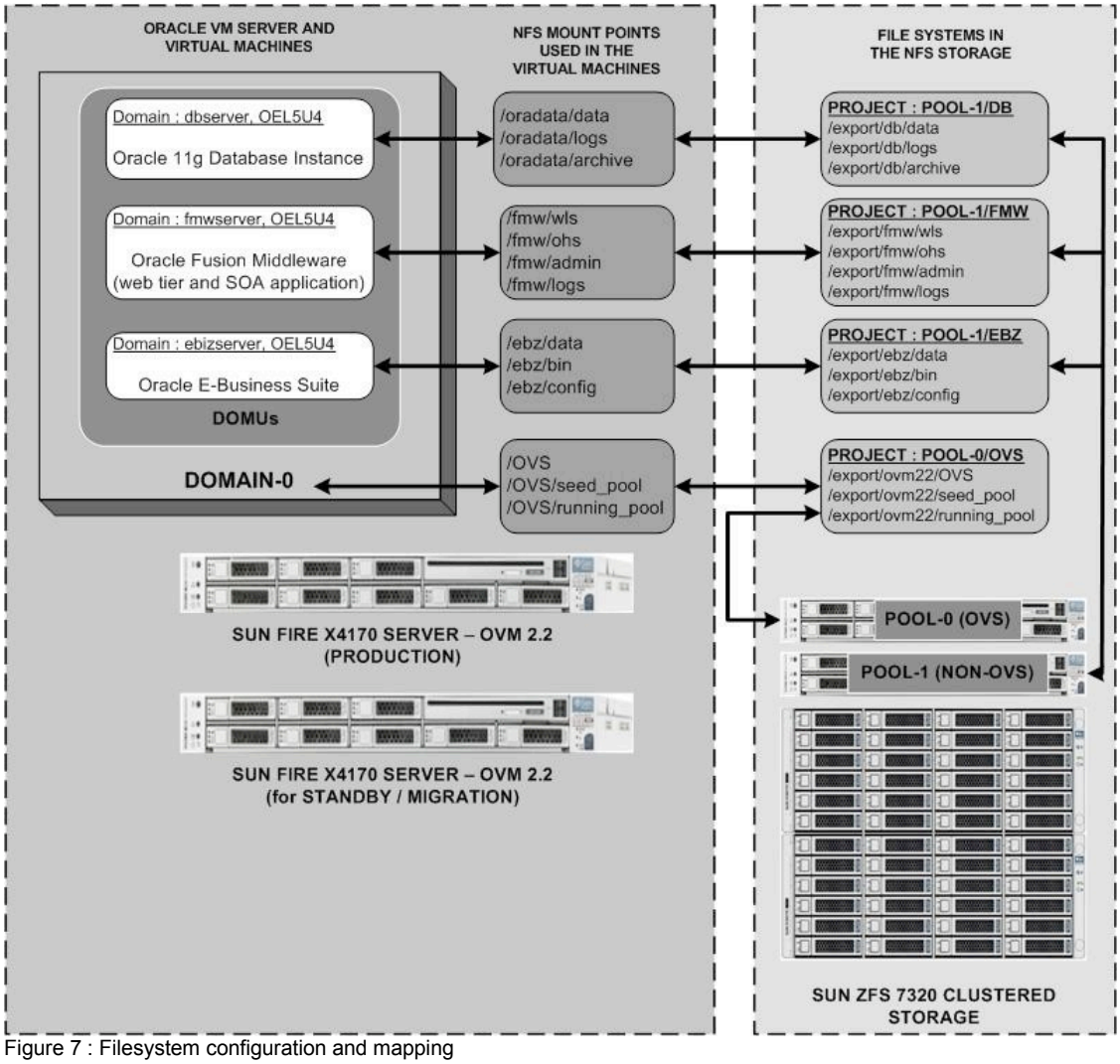

Figure 7 : Filesystem configuration and mapping

Storage Configuration Recommendations

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

The high-performance storage capabilities in this configuration results from using solid state drives and flash acceleration in some models of System — enabling rapid write capabilities for fast data placement on the storage pool and optimizing I/O rates. Both Sun ZFS Storage 7420 and Sun ZFS Storage 7320 models have read optimized and write optimized flash devices which enable excellent response time and throughput for the demanding virtualized environments. This boosts VM cache

performance especially because of the low latency and high performance possible with Flash 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 for the OVS repository is recommended. OVS share is created under the project. The share 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.

Conclusion

Together, Oracle VM and Sun ZFS Storage and x86 servers help achieve the performance, scalability, and simplicity goals of virtualization.

- NFS simplicity though a choice of protocols within the storage architecture
- Faster installation using VM and VM Templates with Hybrid Storage Pools
- Enhanced performance tuning with DTrace Analytics
- Clustered storage for high availability
- Zero-cost of included data services software and connectivity protocol

Sun ZFS Storage Appliance is an ideal solution to help create high performance, scalable Oracle Enterprise Linux and Oracle Virtual Manager based environments. To more fully understand the economics and benefits of the combined solution, contact an Oracle sales representative and ask about obtaining a TCO analysis profile.

Resources

TABLE 2. ADDITIONAL INFORMATION FOR THE TECHNOLOGIES MENTIONED

RESOURCE	URL
Sun ZFS Storage Appliance	http://www.oracle.com/us/products/servers-storage/storage/unified-storage/index.html
Architect Center: Virtualization	http://www.oracle.com/technology/tech/virtualization/index.html?origref=http://www.oracle.com/us/technologies/virtualization/index.htm
Sun ZFS Storage Appliance for Oracle VM	http://www.oracle.com/us/technologies/virtualization/index.html
Validated Configurations	http://www.oracle.com/technology/tech/linux/validated-configurations/index.html
Sun ZFS Storage Appliance and Oracle Databases	http://wikis.sun.com/display/BigAdmin/Using+Sun+Storage+7000+Unified+Storage+Systems

Oracle VM Solution using Oracle's Sun ZFS Storage Appliances and Oracle Sun Servers
September 2010
Author: Sridhar Ranganathan
Contributing Authors: Chris Barclay, Sandeep Bhalerao, Matthew Miller

Oracle Corporation
World Headquarters
500 Oracle Parkway
Redwood Shores, CA 94065
U.S.A.

Worldwide Inquiries:
Phone: +1.650.506.7000
Fax: +1.650.506.7200
oracle.com

Oracle is committed to developing practices and products that help protect the environment

SOFTWARE. HARDWARE. COMPLETE.