An Oracle White Paper
April 2010

Sun Reference Architecture for Oracle 11*g* Grid

ORACLE

Executive Overview

Intended for executives evaluating the use of horizontally scalable grids for database application deployment, as well as technical staff looking to design, implement, and tune database applications in a grid computing environment, this white paper provides an overview of database deployment challenges and options, and describes how Oracle's Sun Reference Architecture for Oracle 11*g* Grid delivers on those requirements. An architectural overview and descriptions of the hardware and software components used in the reference architecture are also included.

Introduction

In the endless quest to do more with less, grid computing has emerged as one of the most effective tools in the IT arsenal. One of the newest and most effective uses of grid computing is for enterprise database implementations.

Enterprise grid and database software must be deployed on solid infrastructure. but the choices can be daunting. Sun Reference Architecture for Oracle 11*g* Grid helps ease this effort. Designed around proof-of-concept deployments and tested and validated configurations, Sun Reference Architecture for Oracle 11*g* Grid provides a high-performance, highly available, and reliable database infrastructure in a cost-effective grid environment. This reference architecture also includes failover mechanisms and redundancy to eliminate single points of failure. By using these advanced products and technologies, customers can implement robust, scalable database platforms that maintain or increase service levels.

Built on the Oracle Solaris operating system, Sun Reference Architecture for Oracle 11*g* Grid benefits from that operating system's performance advantages for database services, world-record benchmark results, and optimizations—yielding faster

performance, without requiring changes to existing applications. This reference architecture provides an ideal platform for a powerful database grid solution.

One of the greatest benefits of Sun Reference Architecture for Oracle 11*g* Grid is its ability to integrate complementary elements from both Oracle and third-party vendors. Vendor interoperability fosters a more-adaptable enterprise and enables computing resources and bandwidth to be reprovisioned to meet changing business demands. By combining low-cost servers with other architecture components, Sun Reference Architecture for Oracle 11*g* Grid balances price, performance, and availability. The result is a flexible, scalable, and powerful solution that can help reduce costs throughout the solution lifecycle.

Information Technology Challenges and Grid Computing

Today's IT executives face mounting pressure to deploy reliable, high-performance database solutions within severely constrained capital and expense budgets. Expectations are high for horizontally scalable solutions that can maintain or increase application availability, reliability, and performance, while supporting service levels and simultaneously reducing TCO.

These requirements demand integrated solutions that are becoming increasingly complex to design, implement, and maintain, and can require specialized technical and business expertise. Rapid changes in business cycles and increasing demand for compute power are prompting IT organizations to look for alternative solutions.

Today, grid computing is shifting from concept to reality in the form of a viable computing framework. IT organizations are now turning to the grid computing model to adapt to changing business needs, deliver greater efficiency, minimize costs, and provide investment protection and rapid ROI. With expectations high and the outcome of such deployments critical, IT organizations should answer the following questions in evaluating a grid computing solution:

- Is low-cost grid architecture a viable option?
- What cost factors are involved in scaling a grid horizontally?
- How do additional connections, software licensing fees, and the implementation and management of more-complex systems affect solution costs?
- Can the grid solution provide a reliable, horizontally scalable platform for running enterprise-grade commercial applications at acceptable service levels?
- What components does the grid architecture include?
- How are applications distributed across servers?
- Where are the potential bottlenecks?
- In a database cluster, how does the system recover from a component failure?
- With respect to interconnect technology, how do Gigabit Ethernet (GbE), 10 GbE, and InfiniBand compare in terms of latency, bandwidth, and overhead?
- How can Oracle products add value to the grid solution?
- How do Oracle's Sun server products complement Oracle Real Application Clusters (Oracle RAC) deployments?

Sun Reference Architectures

Sun Reference Architectures are integrated solutions that combine hardware and software products and incorporate third-party offerings from top-tier software and networking equipment

vendors. These solutions include a set of recommended components that are sized, tested, tuned, and documented to work together, delivering specific functionality and reducing the complexity, costs, and risks of deploying new technology.

Key Benefits of Sun Reference Architecture for Oracle 11*g* Grid

Providing a horizontally scalable solution that delivers high performance and premium-quality service at low cost, Sun Reference Architecture for Oracle 11*g* Grid offers the following benefits:

- **Proven solution portfolio.** Sun Reference Architecture for Oracle 11*g* Grid provides a blueprint for deploying Oracle RAC 11*g* in a grid computing environment—formulating best practices for an Oracle RAC 11*g* infrastructure with other best-of-breed third-party products within a grid context.
- **Maximum availability.** Sun Reference Architecture for Oracle 11*g* Grid maximizes database availability by leveraging Oracle RAC 11*g* and Oracle Solaris Cluster, and by deploying redundant components to eliminate single points of failure. Using a clustered grid infrastructure, Sun Reference Architecture for Oracle 11*g* Grid employs built-in load-balancing and failover mechanisms to minimize interruption of service levels due to individual component failure. Oracle RAC 11*g* enables databases to span multiple physical servers, eliminating the server as a single point of failure and ensuring high availability for applications and users.
- **Maximum reliability.** The implementation of Oracle Solaris Cluster enhances the reliability and availability of Oracle RAC 11*g* and improves the architecture's resiliency. Oracle Solaris Cluster 3.2 Advanced Edition offers the advantages of a solid, well-proven cluster framework that integrates tightly with Oracle Solaris 10 kernel for the quick detection and handling of failures. Combined with an effective I/O fencing mechanism and full redundancy and load balancing of interconnect traffic, Oracle Solaris Cluster software helps ensure a secure and dependable Oracle environment on Oracle Solaris 10.
- **Maximum performance.** The low-cost, high-performance SPARC Enterprise servers with UltraSPARC T2 and UltraSPARC T2 Plus processors deliver the core computing power behind the reference architecture. These servers can be equipped with as many as 128 compute threads per server, providing extraordinary versatility and setting a new standard for server availability, price/performance, and reduced ownership cost.
- **Flexible scalability.** System infrastructures must allow organizations to increase system capacity and throughput by scaling up or scaling out—deploying larger servers, adding a number of small servers, or both. The ability of Oracle RAC 11*g* to span the database across multiple physical servers provides the flexible scaling so vital to the architecture's design.
- **Minimal costs.** IT organizations are always looking for ways to reduce implementation costs, especially those associated with acquisition, deployment, and maintenance. Sun Reference

Architecture for Oracle 11*g* Grid helps organizations meet these goals by taking advantage of low-cost, robust commodity hardware and software.

- **Simplified management.** One challenge of grid computing is managing the additional complexity of a grid infrastructure. Comprehensive management framework solutions, including Oracle Enterprise Manager 10*g* Grid Control and Oracle Enterprise Manager Ops Center Manager software, are incorporated in the reference architecture. These tools provide automated self-management capabilities that can be leveraged to manage grid complexity, simplify support, and drive down operating costs.

The Design of Sun Reference Architecture for Oracle 11*g* Grid

Sun Reference Architecture for Oracle 11*g* Grid was designed with several key requirements in mind, including reliability, high availability, resiliency, cost effectiveness, scalability, and incorporation of best-of-breed products. By adhering to these fundamental criteria, Sun Reference Architecture for Oracle 11*g* Grid provides a database grid ideally suited for enterprise applications. (See Figure 1.)

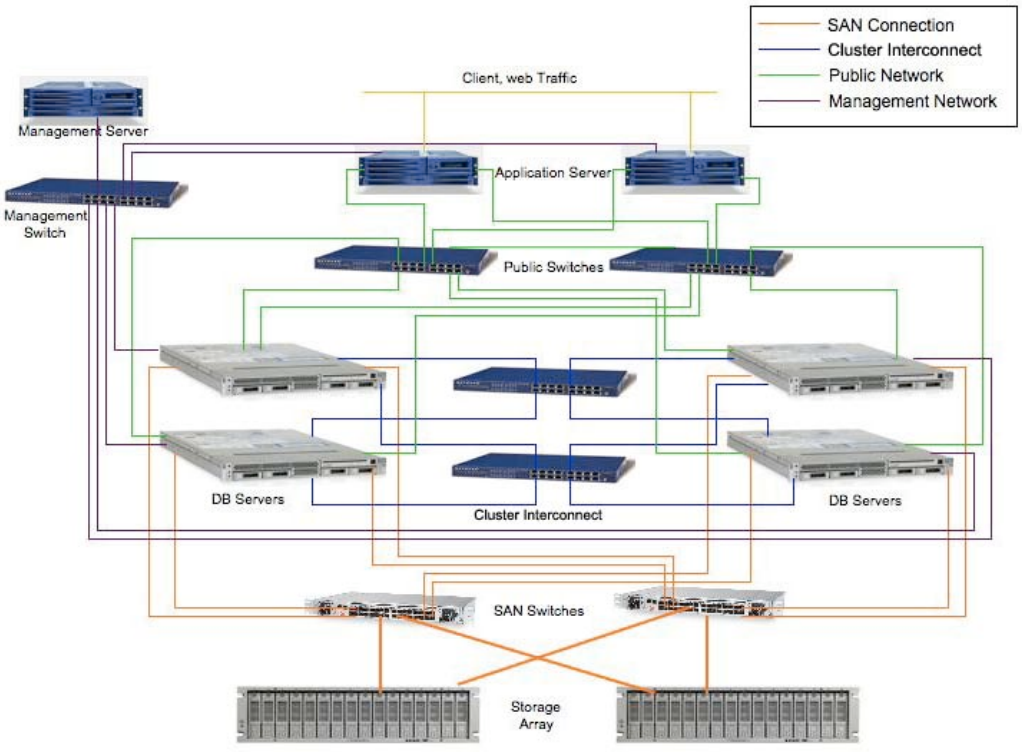

Figure 1. This illustrates an implementation of Sun Reference Architecture for Oracle 11*g* Grid.

Reference Architecture Components

Sun Reference Architecture for Oracle 11*g* Grid combines quality products from top-tier vendors. (See Table 1.) By using fully redundant components, the reference architecture delivers a highly robust, available, and reliable cluster grid infrastructure at a low acquisition cost.

TABLE 1. HARDWARE AND SOFTWARE COMPONENTS IN SUN REFERENCE ARCHITECTURE FOR ORACLE 11*g* GRID

HARDWARE COMPONENTS	SOFTWARE COMPONENTS
SPARC Enterprise servers with UltraSPARC T2 and UltraSPARC T2 Plus processors (database tier)	Oracle Solaris 10 05/08 (Update 5)
Sun Fire server with UltraSPARC processor (application and management tier)	Oracle 11*g* R1 Clusterware and Database
Oracle's StorageTek 6140 Fibre Channel Array (configured with dual RAID controllers)	• Oracle Solaris Cluster 3.2 U1 (2/08) Advanced Edition for Oracle RAC software • Oracle's StorageTek QFS 4.6.5
InfiniBand server switches	Oracle Enterprise Manager 10*g* Grid Control Release 4
SAN switches	Oracle Enterprise Manager Ops Center 1.1
10 GbE switches	
GbE switches	
Terminal console	

SPARC Enterprise Servers

SPARC Enterprise servers with CoolThreads technology deliver higher performance in less space and using less power than competitive servers.[1] They are based on UltraSPARC T2 and UltraSPARC T2 Plus processors, the industry's first massively threaded systems-on-a-chip (SoC), offering breakthrough performance and energy efficiency.

[1] Testing of Sun Reference Architecture for Oracle 11*g* Grid was focused on Oracle's Sun servers equipped with UltraSPARC T2 and UltraSPARC T2 Plus processors.

CoolThreads Technology

Based on chip multithreading technology, third-generation CoolThreads technology can support up to 128 threads on just one rack unit (1U), providing increased computational density within constrained envelopes of power and cooling. Very high levels of integration help reduce latency, lower costs, and improve security and reliability. Balanced system design provides support for a wide range of application types—from Web services to high-performance computing.

SPARC Enterprise T5120 Server

Based on the UltraSPARC T2 processor, the SPARC Enterprise T5120 server is the world's first 64-thread, general-purpose server.

SPARC Enterprise T5140 Server

The SPARC Enterprise T5140 server is the world's first dual-socket, general-purpose server powered by the third-generation UltraSPARC T2 Plus SoC processor. It comes with up to 128 compute threads per server.

StorageTek 6140 Fiber Channel Array

Oracle's StorageTek 6140 Fiber Channel Array delivers high performance, high availability, and reliability in one economical package through affordable enterprise-class features and functionality such as dual hot-swap power and cooling, hot-swap redundant RAID controllers with mirrored cache, hot-swap disk drives, global and local hot sparing, dynamic logical unit number expansion, dynamic capacity expansion, nondisruptive firmware code loading, and remote status monitoring. Simplifying management and configuration, the array provides pay-as-you-grow scalability to protect your investment. Use of the StorageTek 6140 Fiber Channel Array in the reference architecture affords reliability and impressive I/O performance in a small footprint.

Oracle Solaris 10

Supported on more than 1,000 x86 and SPARC platforms, Oracle Solaris 10 delivers the performance, stability, and security that enterprise grid environments demand. With more available applications than for any other open operating system, you can use a single Oracle Solaris 10 instance to span your entire enterprise: the Web tier, the data warehouse, and the most demanding technical compute applications.

The latest version of Oracle Solaris offers new features, including virtualization, volume management support, and advanced real-time application debugging (DTrace). Oracle Solaris OS can scale both vertically and horizontally, and offers a broad hardware platform—important capabilities for enterprise grid solutions.

Oracle Solaris Cluster Advanced Edition for Oracle Real Application Clusters

Oracle Solaris Cluster Advanced Edition software for Oracle RAC provides enterprises with additional high availability for Oracle RAC on Oracle Solaris 10. In addition to providing I/O fencing to help guarantee data integrity, this product suite leverages two key components of Oracle Solaris Cluster software to deliver high performance. The StorageTek QFS shared file system helps eliminate administrative overhead and provides maximum data management and throughput for the most data-intensive applications. The volume manager feature in Oracle Solaris provides partitioning and automatic device relocation for data mirroring and high availability. Together, these components provide a high-performance clustered file system for traditional file system management and administration. In addition, Oracle Solaris Cluster Advanced Edition for Oracle RAC is tightly integrated with Oracle Cluster Ready Services and Oracle Solaris to help minimize application downtime and maximize application performance.

Oracle Solaris Cluster high-availability functionality supports active/active configurations. Solaris Cluster uses a single logical private interconnect (clprivnet0), where two or more interfaces are combined for high throughput and to stripe traffic across all the interfaces.

Oracle Enterprise Manger Ops Center

Oracle Enterprise Manger Ops Center is an integrated management tool that makes it possible to manage thousands of Microsoft Windows, Linux, and Oracle Solaris x86 and SPARC systems. Administrators can discover, provision, update, and monitor both virtual and physical systems—all from one graphical interface.

Oracle Enterprise Manager Ops Center is also designed to enable the remote management of heterogeneous environments. With its state-of-the-art, browser-based Ajax graphical user interface, there's no compromise between user experience and portability. Virtual machines and the physical hardware they run on can be managed from anywhere through a standard browser. Oracle Enterprise Manger Ops Center software is designed to work in existing datacenters, even in companies that aren't currently running Sun technology.

Oracle Real Application Clusters 11*g*

Oracle RAC 11*g* is the database designed for grid implementations. A proven technology that allows multiple low-cost servers to perform like a single large server, Oracle RAC forms a key foundation for enterprise database grids. With Oracle RAC, even very high-end systems can be constructed from small, low-cost clusters made from standard commodity parts. Running Oracle RAC on a cluster provides the highest level of database availability along with flexibility in scaling. If a node in the cluster fails, the Oracle software continues running on the remaining nodes. If more processing power is necessary, new nodes can be added to the cluster easily. As a result, organizations can deploy low-cost solutions while reaping the benefits of high availability and adaptability.

Oracle Enterprise Manager 10*g* Grid Control

Oracle Enterprise Manager 10*g* Grid Control software allows administrators to easily manage business applications, end-user services, and the entire grid infrastructure. It contains improved tools such as new service modeling, broader support for service protocols, and comprehensive policies and templates, enabling the management of grids as a single entity. Graphical service topologies and service dashboards provide the right information to high-level decision-makers and line-of-business managers for more-effective planning, while automated provisioning and patching functionality give IT administrators the power to make efficient, error-free changes.

InfiniBand Technology

Serving as a high-performance, switched-fabric interconnect, InfiniBand fosters greater scalability and improved manageability. It can also help accelerate database server speed and provide an added boost to clustered databases.

Reliable Datagram Sockets (RDS) over InfiniBand can provide a horizontally scalable, high-performance alternative to traditional vertical scaling for enterprises using Oracle Database 11*g* and Oracle RAC.

10 Gigabit Ethernet Technology

Oracle's 10 Gigabit Ethernet networking technology is one of the industry's first network interfaces designed to accelerate multithreaded clustered application performance by optimizing I/O throughput in environments that use parallel threads. It enables systems to

- Improve use of compute resources
- Minimize system I/O latency
- Leverage existing wiring infrastructure
- Take advantage of more network bandwidth

SPARC enterprise chip multithreading–based servers come with network ports, which improve the Oracle RAC cache fusion traffic over interconnect.

Connecting the Network, Servers, Applications, and Storage

Interconnect latency and throughput are essential elements in the scalability of a database grid cluster. Internode communication and cache fusion traffic can benefit from low latency, high throughput, and low system overhead.

Deploying InfiniBand further aids integration of the elements of Sun Reference Architecture for Oracle 11*g* Grid because InfiniBand leverages the high-performance RDS protocol, a very low-latency Remote Direct Memory Access–based protocol driver. The RDS protocol can be used to create a unified fabric for interprocess communication.

Ten GbE networking technology is designed to help accelerate clustered applications that require high throughput and low latency for performance.

Sun Reference Architecture for Oracle 11*g* Grid is validated with InfiniBand and 10 GbE interconnect technologies. Tests have shown the benefit of using high-bandwidth and low-latency interconnect over traditional GbE technology. For more on this topic, see the following section, "Performance Characterization."

Figure 2 illustrates the two key network connection areas: the private interconnect and public network. Using high-throughput (InfiniBand or 10 GbE) technologies for the private interconnects between the database servers accelerates cache fusion traffic and Oracle RAC internode communication, and results in very low latency. The public network connection enables traffic between the application and database servers.

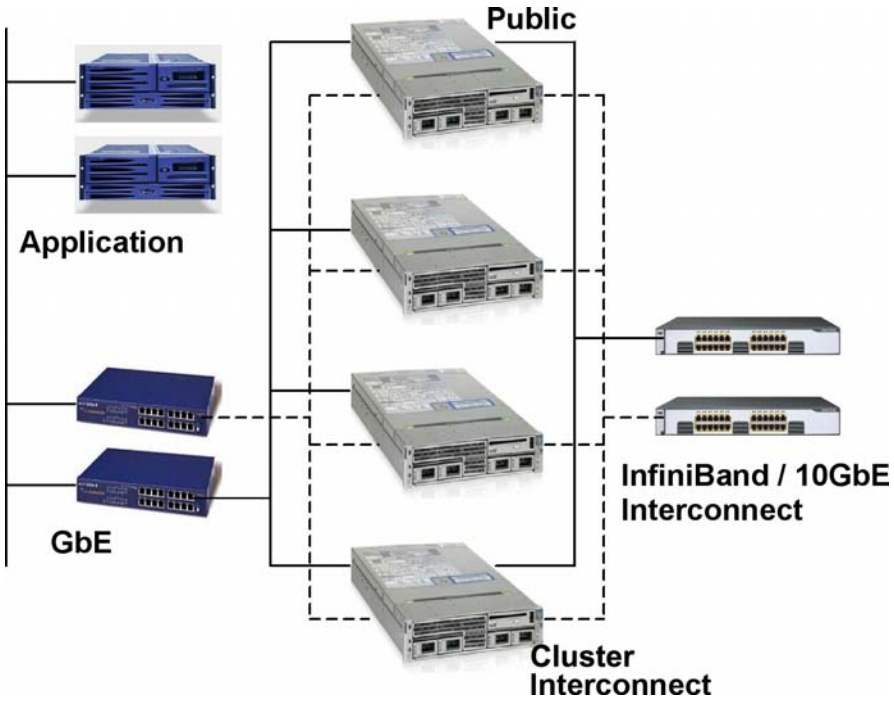

Figure 2. The private interconnect and the public network serve as the two key network connection areas.

The management network is reserved for telnet access, basic health and monitoring of components, connection to the service processor module, and monitoring and communication by the Oracle Enterprise Manager 10*g* Grid Control and Oracle Enterprise Manger Ops Center Manager software. (See Figure 3.) The terminal console is connected to the built-in management port of the servers to enable console access.

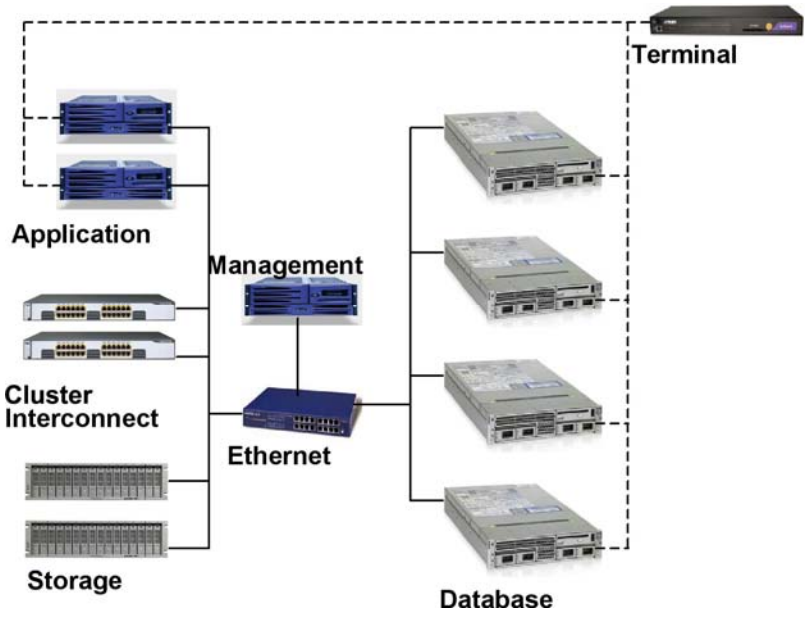

Figure 3. The terminal console is connected to the built-in management port of the servers to enable console access.

All four database nodes share a pair of Sun StorEdge 6140 Fiber Channel Arrays via two SAN switches. (See Figure 4.) This configuration, together with multiplexed I/O, helps provide full redundancy in the event of failures.

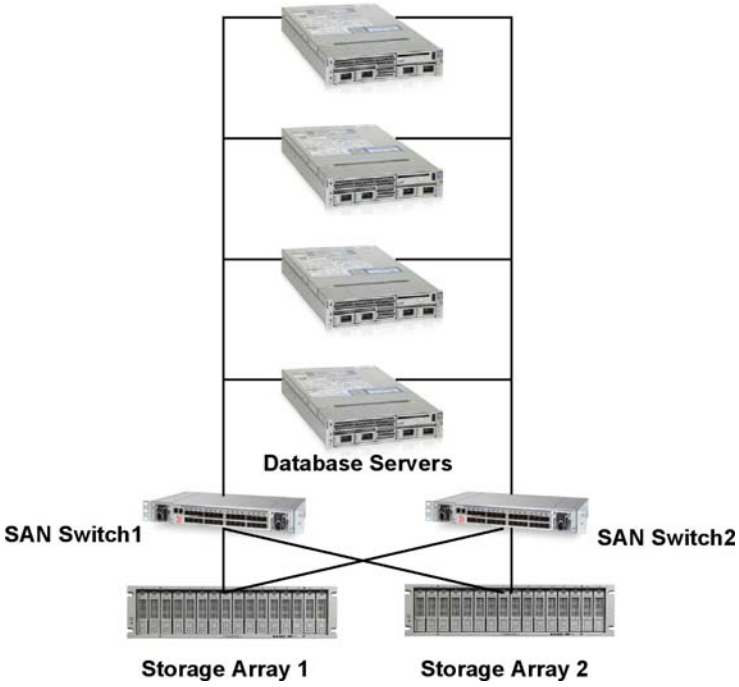

Figure 4. The database servers share the two storage arrays through the SAN switches.

Application Logical Architecture

Figure 5 provides a representation of the logical structure of Sun Reference Architecture for Oracle 11*g* Grid. The management software stack meets the critical need for a grid management framework by leveraging software from Oracle Enterprise Manager 10*g* Grid Control, Oracle Enterprise Manger Ops Center, and Oracle Solaris Cluster management software. In addition, other key components form the foundation of the logical architecture, including the Oracle Solaris 10, Oracle Solaris Cluster Advanced Edition for Oracle RAC software, and Oracle RAC 11*g*. Oracle Clusterware provides the database cluster framework and is enhanced by the use of the Oracle Solaris Cluster software, a robust and proven cluster framework that's tightly integrated with the Oracle Solaris kernel. Oracle data file storage options can be raw devices through Oracle Automatic Storage Management, the Oracle Solaris Cluster file system with StorageTek QFS shared file system software, or both.

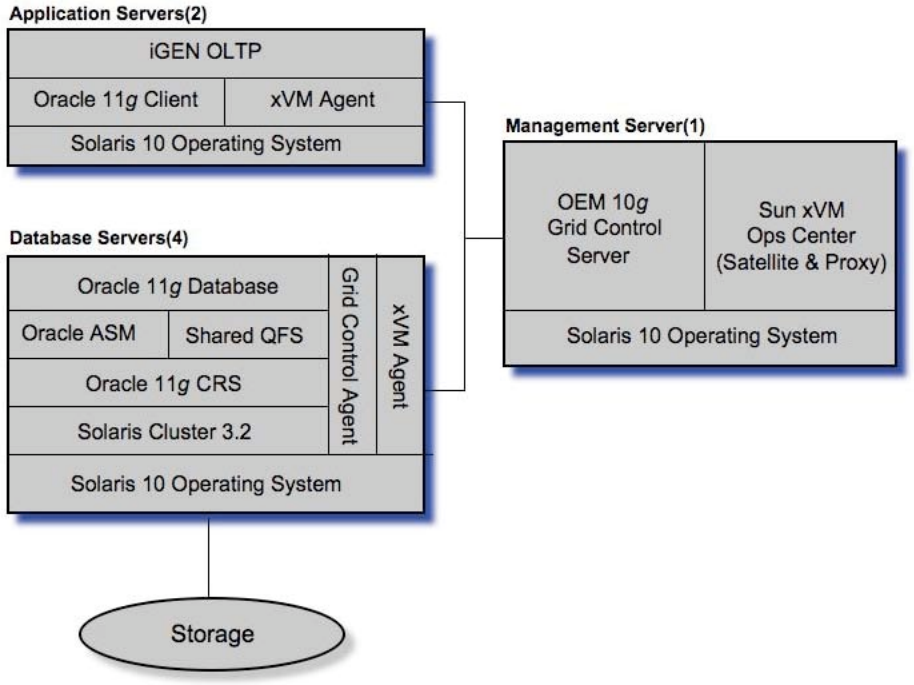

Figure 5. The diagram illustrates the logical structure of Sun Reference Architecture for Oracle 11*g* Grid.

System Components

Table 2 details the reference architecture hardware and software components. Hardware and software specifications are for models and releases available at the time the reference architecture was developed. For updates on available products, please refer to individual vendors' Web sites.

The reference architecture was designed, integrated, tested, and tuned for arbitrary workloads generated with the iGEN-OLTP, a benchmark suite that mimics an online customer order-entry

application. In addition, the reference architecture was validated and tested with various interconnect technologies, including GbE, InfiniBand, and 10 GbE. The Oracle cluster interconnect plays a critical role in overall system performance. For more on this topic, see the following section, "Performance Characterization."

TABLE 2. REFERENCE ARCHITECTURE COMPONENTS

SERVER TYPE	Database servers	• 4 x SPARC Enterprise servers 5120 (1 x UltraSPARC T2 processor, 32 GB RAM, 2 x 146 GB drives) • 4 x SPARC Enterprise servers 5140 (2 x UltraSPARC T2 Plus processor, 32 GB RAM, 2 x 146 GB drives)
	Management servers	• 1 x Sun Fire (8 x UltraSPARC processor, 8 GB RAM, 2 x 146 GB disks)
	Application servers	• 6 x Sun Fire (8 x UltraSPARC processor, 8 GB RAM, 2 x 146 GB disks)
	Storage arrays	• 2 x Sun StorageTek 6140 Fiber Channel Array
	Interconnect switches	• 2 x GbE switch—12 port • 2 x InfiniBand switch—12 port • 2 x 10 GbE switch—12 port
STORAGE	Management switch	• 24-port GbE switch
	SAN switches	• 2 x 16-port Sun SAN switch
NETWORK	Adapter cards (for database servers)	• 4 x dual-port GbE cards • 4 x dual-port Fibre Channel cards • 4 x dual-port InfiniBand cards
	Terminal console	• 16-port terminal console
	OS	• Oracle Solaris 10
	Cluster software	• Oracle Clusterware • Oracle Solaris Cluster Advanced Edition for Oracle RAC (SC3.2U1)
	Database software	• Oracle 11*g* Database
	Management software	• Oracle Enterprise Manager 10*g* Grid Control Release 4 • Oracle Enterprise Manger Ops Center 1.1

SOFTWARE	Volume manager software	• Oracle Automatic Storage Management • Oracle Solaris volume manager and StorageTek QFS Shared File System 4.6
	Workloads	• iGEN-OLTP • Oracle 11*g* Client
	Miscellaneous	• Oracle 10*g* Grid Control Agent

Performance Characterization

This section describes the methods used to test the stability and performance of Sun Reference Architecture for 11*g* Grid, and explains some of the benchmarking results from the testing. These tests included both functional tests to validate the reference architecture and performance tests to show Oracle RAC scalability with Sun servers horizontally.

Workload

The iGEN-OLTP workload benchmark was used in the reference architecture illustrated previously in this white paper. An internally developed benchmark toolkit, iGEN-OLTP is based on real customer workloads. It simulates a global order system that includes four continent-based schema, all accessed via central views. The transactions contained various SQL statements: read-only selects, update, and insert operations.

Data Collection

During the reference architecture's testing, OS and Oracle performance statistics were gathered and then used to tune the architecture to achieve the maximum workload. The iGEN tool collects statistics such as number of transactions, transactions per second, and response time.

Test Matrix

Sun Reference Architecture for Oracle 11*g* Grid was validated and tested on various cluster interconnect technologies such as GbE, InfiniBand with Reliable Datagram Sockets (RDS), and 10 GbE. The reference architecture solution was also validated and tested on various storage options for Oracle RAC. Tests were performed to compare the storage option available for Oracle Cluster, including Oracle Automatic Storage Management and Sun Shared QFS.

The tests were performed to stress the servers at different workload (CPU use) levels by increasing the number of connections to the database. The purpose of the testing was to show how servers perform at different workloads, such as 50 percent CPU use, 75 percent CPU use, and 100 percent CPU use.

The testing was conducted on a four-node Oracle RAC. The Oracle RAC was scaled a node at a time, from one node up to four nodes, to show how the server scales horizontally (scales out) in an Oracle RAC grid architecture.

The tests were performed on the chip multithreading family of servers (SPARC Enterprise 5120 and SPARC Enterprise 5220 servers).

Test Summary

The following text and diagrams summarize the test results.

Interconnect Tests

Cluster interconnect is the most important factor for Oracle RAC performance and scalability. Interconnects with high-bandwidth, low-interconnect latency and low CPU usage are always preferred solutions for Oracle RAC grid architecture. Sun Reference Architecture for Oracle 11*g* Grid was tested over different interconnects. Some of the results are shown in Figures 6 and 7.

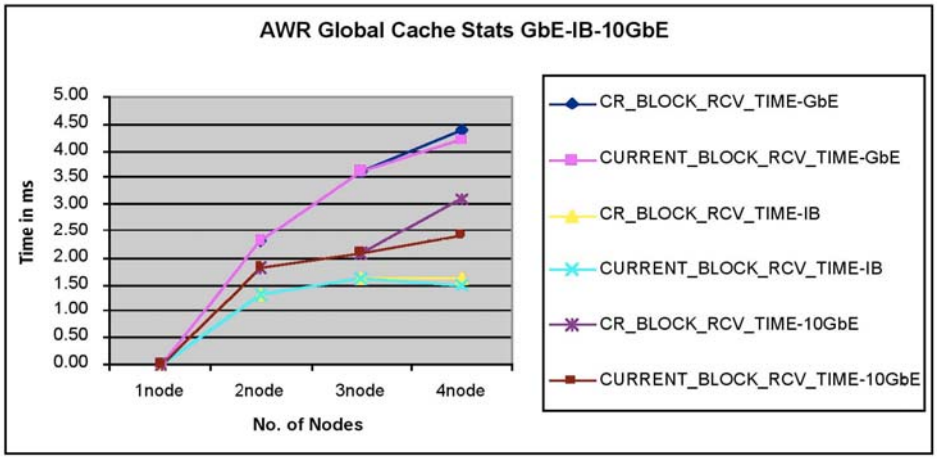

Figure 6. As the number of interconnect nodes increases, the minutes of processing time also increases.

The interconnect tests showed that

- 10 GbE and InfiniBand with RDS improve overall response time[2]
- Cache latency at cluster interconnect is one-half with 10 GbE interconnect compared with GbE interconnect

[2] The response time also depends on other components such as I/O and CPU.

- Cache latency at cluster interconnect is one-third with InfiniBand and RDS compared with GbE interconnect
- Server CPU use is 10 to 20 percent lower with 10 GbE and InfiniBand compared with GbE interconnect

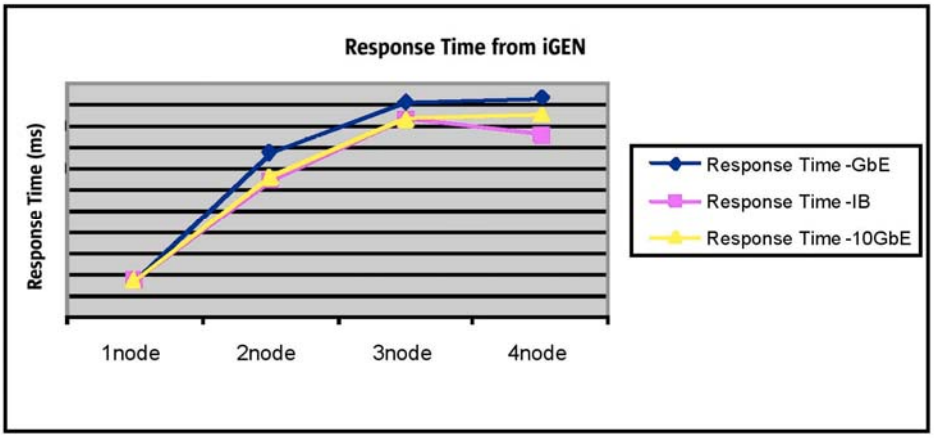

Figure 7. The graph shows the response time for iGEN.

Oracle Automatic Storage Management vs. Sun Shared QFS Tests

Oracle RAC grid implementations require shared storage. Oracle Automatic Storage Management and Sun Shared QFS are the options considered best for Sun Reference Architecture for Oracle 11*g* Grid.

Figures 8 and 9 show the performance differences for Oracle Automatic Storage Management and Sun Shared QFS.

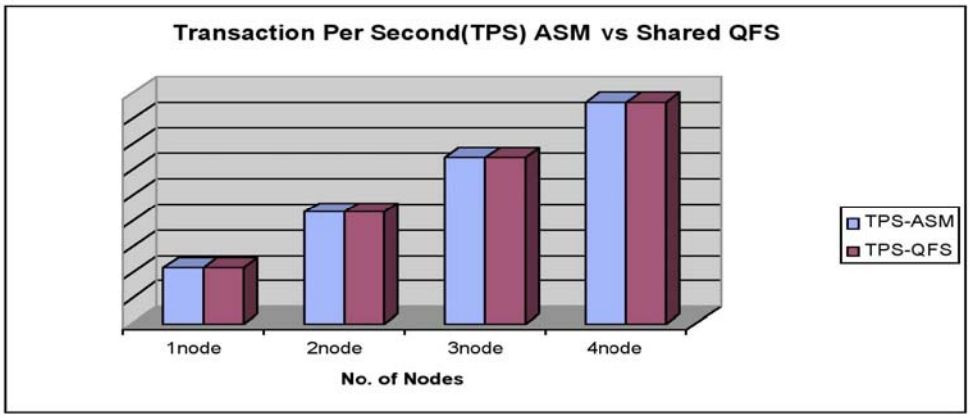

Figure 8. The chart shows the throughput for both Oracle Automatic Storage Management and Sun Shared QFS.

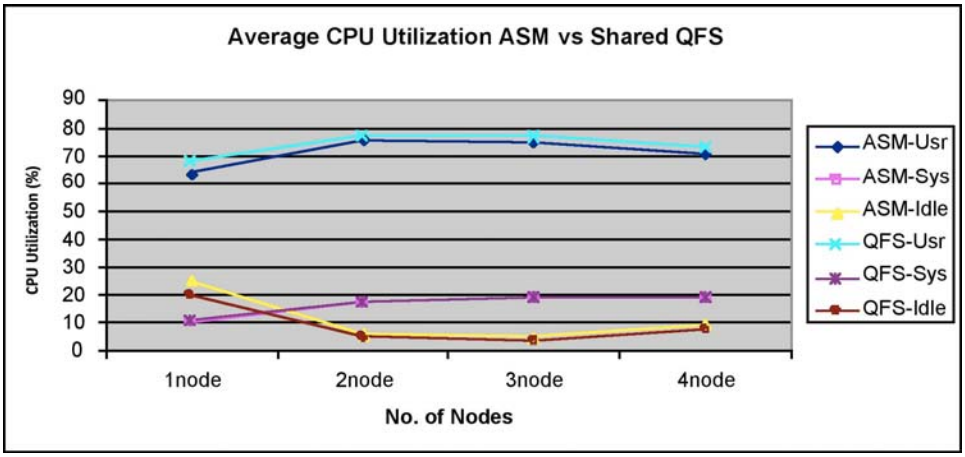

Figure 9. The chart shows server resources for Oracle Automatic Storage Management and Sun Shared QFS.

The testing revealed the following:

- Performance results for Sun Shared QFS and Oracle Automatic Storage Management are comparable.
- Sun Shared QFS extends the storage option for Oracle RAC with ease of management. OS Utility can be used to monitor or manage the file system.
- Throughput and response time are nearly the same for Oracle Automatic Storage Management and Sun Shared QFS.
- There is not much system overhead using Sun Shared QFS compared to Oracle Automatic Storage Management.

Oracle Real Application Clusters Scaling

Oracle RAC was scaled a node at a time, from one node up to four nodes. Figure 10 shows the server scaling results achieved in testing. The baseline was created on one node by looking at the overall CPU use, and the user count was increased to achieve higher CPU use. The scalability factor was calculated based on TPS achieved.

Scalability Factor = TPS Achieved / TPS Baseline (single node)

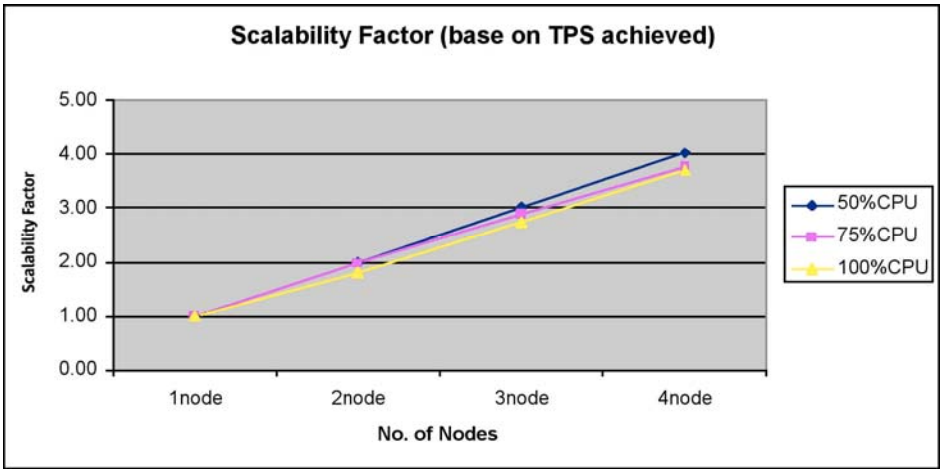

Figure 10. The scalability factor for Sun Reference Architecture for Oracle 11*g* Grid is based on TPS achieved.

Testing demonstrated that Sun servers showed linear scaling in the Oracle RAC grid architecture.

Storage Best Practices

The Sun Shared QFS layout and file creation options are the keys to high database performance. Sun Shared QFS has two file allocation schemes available: an ms file system type and an ma file system type.

The ma file system type was used for Sun Reference Architecture for Oracle 11*g* Grid, where Sun Shared QFS is installed on multiple partitions in which metadata is written to mm devices and data can be written to md devices. Both devices should share the same spindle.

While creating the file system with the `sammkfs` command, specify 1 MB DAU size; this reduces the I/O and increases the application performance.

Enabling the SAMAIO for Oracle RAC data files provides a performance boost by decreasing CPU use. You can use `setfa` with the -q option to enable it.

Network Best Practices

The following subsections describe the best practices employed in the reference architecture's network design.

Virtual Local Area Network

Virtual LANs (VLANs) are logical groupings of network devices. Using VLANs to segment different types of traffic to specific subnets provides improved throughput, manageability, application separation, and security. This is optional, but for large and complex networking situations, it helps to separate VLANs for applications, storage, and Oracle RAC traffic.

Driver Patch

You should always have the latest updated drivers on all network devices on the system. For Sun Reference Architecture for Oracle 11*g* Grid, 10 GbE cards were used with XAUI interfaces for cluster interconnect. The following patches are recommended:

- 127755-01
- 125891-01
- 125476-02
- 127127-11
- 138048-01

Network nnd Tuning

Network tuning plays an important role in overall system performance. The following are tuning recommendations specific to Sun Reference Architecture for Oracle 11*g* Grid:

- ndd -set /dev/tcp tcp_smallest_anon_port 2048
- ndd -set /dev/tcp tcp_conn_req_max_q 81920
- ndd -set /dev/tcp tcp_conn_req_max_q0 81920
- ndd -set /dev/tcp tcp_xmit_hiwat 800000
- ndd -set /dev/udp udp_max_buf 4194304
- ndd -set /dev/udp udp_xmit_hiwat 262144
- ndd -set /dev/udp udp_recv_hiwat 262144

Database Server Best Practices

Solaris Update and Firmware

Sun Reference Architecture for Oracle 11*g* Grid is built on Solaris 10 Update 5. Before deploying the grid, you should verify the version shipped with the server and determine whether it's the latest production version supported. Make sure you have the latest firmware on the system. For the chip multithreading servers, the latest firmware available—7.1.4.a—was installed.

Memory

Sun Reference Architecture for Oracle 11*g* Grid was tested with servers using 32 GB of total RAM memory. Please refer to the database server documentation to determine the total memory required for your implementation. Oracle recommends that you configure your servers with the maximum memory feasible to meet your scalability and performance needs.

Intimate Shared Memory is an Oracle Solaris facility that enables large pages (4 MB) and locking pages in memory. The only action needed from a database point of view is to define the maximum memory that Oracle application should use. This is defined by Oracle parameter memory_max_target.

Sun Shared QFS Mount Options

Sun Shared QFS is one of the important components of the reference architecture, and tuning the Sun Shared QFS mount options improves overall database performance.

Table 3 describes some important mount options.

TABLE 3. MOUNT OPTIONS

OPTION	VALUE	DESCRIPTION
Stripe	1	This sets the stripe width for the file system to n disk allocation units. Set the 1 for DAUs >= 128 K. For our testing, the DAU size was set to 1 M.
mh_write		Enables simultaneous reads and writes to the same file for multiple hosts.
Qwrite		Enables simultaneous reads and writes to the same file from different threads.
Forcedirectio		Specifies direct I/O as the default I/O mode (meaning the data is transferred directly between user's buffer and disk).
Rdlease	600	The read release time option specifies the maximum number of seconds that the file can be read before reacquiring the read release; the default is 30.
Wrlease	600	The write release time option specifies the maximum number of seconds that the file can be written before reacquiring the write release; the default is 30.
Aplease	600	The append release time option specifies the maximum number of seconds that the file can be appended before reacquiring the append release; the default is 30.

Database Configuration Best Practices

LMS Binding

Set the LMS processes to the optimum level using ADDM to maximize this value. During the testing of Sun Reference Architecture for Oracle 11*g* Grid, eight LMS processes were found to be optimal for driving the work and improving the cache fusion traffic. The parameter was set using the gc_server_processes. During testing, it was found that creating the process set and binding the LMS process to the process set improved the interconnect traffic. Use `psrset -c` to create the process set and `psrset -b` to bind the LMS process ids to the process set. Binded the process set at 12:25. Figure 11 illustrates the results.

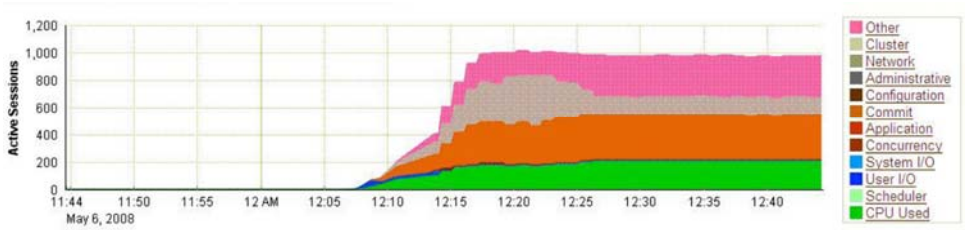

Figure 11. The figure shows the average active sessions for Sun Reference Architecture for Oracle 11*g* Grid.

Conclusion

Today's IT managers are turning to grid computing to deliver the capabilities they require while remaining within their budgets. However, coordinating the hardware, software, and networking equipment involved can be a daunting task. Sun Reference Architecture for Oracle 11*g* Grid makes things easier. Composed of Oracle and third-party products, Sun Reference Architecture for Oracle 11*g* Grid is designed to reduce the risks, uncertainty, and costs associated with implementing an Oracle database within a grid computing environment.

Sun Reference Architecture for Oracle 11*g* Grid consists of recommended integrated hardware and software stacks for a proven grid database solution. Along with this architecture, Oracle provides a best practices framework for obtaining optimal performance, availability, and resiliency.

The most compelling benefits of Oracle's Sun Reference Architecture for Oracle 11*g* Grid come not from its individual components, but from its ability to integrate complementary elements—whether from Oracle or another vendor. The combination of SPARC Enterprise servers with InfiniBand or 10 GbE technology ensures higher database throughput, while the adoption of Oracle Solaris Cluster Advanced Edition for RAC enhances the reliability and availability of Oracle RAC 11*g*. The result is a solution that can help deliver the highest service levels and satisfy the most demanding business requirements of a database grid.

Sun Reference Architecture
For Oracle 11*g* Grid
January 2010

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

Copyright © 2009, 2010, Oracle and/or its affiliates. All rights reserved.
This document is provided for information purposes only and the contents hereof are subject to change without notice. This document is not warranted to be error-free, nor subject to any other warranties or conditions, whether expressed orally or implied in law, including implied warranties and conditions of merchantability or fitness for a particular purpose. We specifically disclaim any liability with respect to this document and no contractual obligations are formed either directly or indirectly by this document. This document may not be reproduced or transmitted in any form or by any
means, electronic or mechanical, for any purpose, without our prior written permission.

Oracle and Java are registered trademarks of Oracle and/or its affiliates. Other names may be trademarks of their respective owners.

AMD, Opteron, the AMD logo, and the AMD Opteron logo are trademarks or registered trademarks of Advanced Micro Devices. Intel and Intel Xeon are trademarks or registered trademarks of Intel Corporation. All SPARC trademarks are used under license and are trademarks or registered trademarks of SPARC International, Inc. UNIX is a registered trademark licensed through X/Open Company, Ltd. 0110

SOFTWARE. HARDWARE. COMPLETE.