An Oracle Technical White Paper
October 2011

Oracle Optimized Solution for Oracle Database Mission-Critical System Environments

Introduction 1
SPARC Enterprise M-Series Servers... 2
Non-Stop/Continuous Database Operations...................................... 4
Dynamic System Domains 5
Dynamic Reconfiguration of Hardware and Database 7
Oracle Solaris Cluster and Oracle Real Application Clusters.......... 8
Oracle Solaris Predictive Self-Healing and Redundant Interconnects 10
Evolution of Oracle's Multicore, Multithreaded Processor Design..... 10
Chip Multithreading 10
The SPARC T4 Processor 11
SPARC T4 Server Overview 12
SPARC T4 Server Virtualization 15
Oracle Solaris for Multicore Scalability.. 16
Oracle's Pillar Axiom 600 Storage System Overview........................ 18
Key Differentiating Technologies 18
Workload Scalability with Flash Technology 19
Understanding and Addressing I/O Performance Issues 19
Flash Technology... 20
Sun Storage F5100 Flash Array... 21
Using Flash to Extend the System Global Area 23
Test Results 24
Reference Architecture—Using the Sun Storage F5100 Flash Array as Storage for Database Indexes 25
System Performance Testing... 27
Database Smart Flash Cache 33
Conclusion 34
References.. 34
Appendix 35
Pillar Axiom 600 Storage System —Steps to Create LUNs and Assign Them to Hosts 35
Setting Up the Oracle Grid Infrastructure...................................... 38
Deploying Oracle Database 11*g* Release 2 45
Configuring Automatic Storage Management 52
Building an Oracle 11*g* Release 2 Database with the Database Configuration Assistant.. 56
Appendix Prerequisites and Checklists... 66

Introduction

Oracle is a complete, full-range solution provider, fully accountable for all aspects of a database system: hardware, operating system, database, data storage, business applications, and management tools. With innovative reliability, availability, and serviceability (RAS) features, Oracle's SPARC Enterprise M8000 and M9000 and SPARC T4 servers are particularly well suited for hosting and consolidating large mission-critical database systems. Key elements of this solution include the Oracle Solaris operating system, Oracle Database 11*g* Release 2, Oracle Enterprise Manager Ops Center, Oracle's Storage technology, Oracle's InfiniBand technology, and Oracle business applications.

In crafting this complete solution, Oracle delivers value in four key areas:

- **Simplification**. A complete system from Oracle eliminates common problems, and the finger pointing that can happen with multiple vendors, over issues such as hardware and software compatibility, systems integration, and support.

- **Non-stop (continuous) operations**. Systems, software, and applications keep running through component failures, service events, and upgrades with minimal downtime.

- **Workload scalability with flash technology**. Oracle solutions provide broad flexibility in how systems can grow cost effectively and with minimal disruption.

- **Investment protection**. Oracle solutions extend the life of computing resources to maximize the return on technology investments.

This paper focuses on the technical advantages of the Oracle Optimized Solution for Oracle Database, which allow the solution to deliver non-stop operations and impressive workload scalability. For further information about simplification and investment protection with this solution, refer to the companion business paper, "Oracle Optimized Solution for Oracle Database Mission-Critical System Environments—A Business White Paper."

SPARC Enterprise M-Series Servers

With symmetric multiprocessing (SMP) scalability supporting up to 64 processors, memory subsystems as large as 4 TB, and high-throughput I/O architectures, SPARC Enterprise M-Series servers from Oracle easily perform the heavy lifting required by consolidated database workloads. Furthermore, SPARC Enterprise M-Series servers run the powerful Oracle Solaris 10 and 11 operating systems and include innovative virtualization technologies at no extra cost. By offering Dynamic System Domains, Dynamic Reconfiguration, and Oracle Solaris Containers technology, the SPARC Enterprise M-Series servers bring sophisticated mainframe-class resource control to an open systems compute platform. Figure 1 is an illustration of the SPARC Enterprise M-Series servers for Oracle Optimized Solution for Oracle Database in mission-critical environments.

Figure 1. The SPARC Enterprise M-Series servers -- Oracle Optimized Solution for Oracle Database in mission-critical environments

The members of the SPARC Enterprise M-Series servers implemented for Oracle Optimized Solution for Oracle Database in mission-critical environments share characteristics that provide scalability, reliability, and flexibility to enterprises. The SPARC Enterprise M-Series servers all deliver exceptional throughput to software applications and databases with a balanced, highly scalable SMP design that utilizes the latest generation of SPARC64 processors connected to memory and I/O by a high-speed, low-latency system interconnect. Architected to reduce planned and unplanned downtime, SPARC Enterprise M-Series servers include advanced reliability, availability, and serviceability capabilities to avoid outages and reduce recovery time.

Design features that boost the reliability of SPARC Enterprise M-Series servers include:

- Advanced CPU integration—The SPARC64 VII+ is a quad-core processor, with each core featuring two-way simultaneous multithreading (SMT), versus many fewer cores and more primitive multithreading in earlier generations.
- Memory patrol—This feature periodically scans memory for errors, and prevents the use of faulty areas of memory before they can cause system or application errors.
- Memory mirroring—The memory system duplicates the data on write and compares the data on read to each side of the memory mirror.
- End-to-end data protection—The Jupiter interconnect has full ECC protection on system buses and in memory. SPARC Enterprise M8000 and M9000 servers feature degradable crossbar switches and redundant bus routes.
- Fault-resilient power options and hot-swappable components—These systems feature redundant, hot-swappable power supply and fan units, as well as the option to configure multiple CPUs, memory DIMMs, and I/O cards while the system is running. Redundant storage can be created using hot-swappable disk drives with disk-mirroring software. High-end servers also support redundant, hot-swappable service processors.
- Hardware redundancy —The SPARC Enterprise M-Series servers provide redundant power, redundant fans, redundant data paths, and the SPARC Enterprise M9000 server also provides a redundant system clock.

For very large mission-critical systems, SPARC Enterprise M8000 and M9000 servers deliver the massive processing power needed. These systems are compared and contrasted in Table 1. Additional information about the features and capacities across this server line can be found at http://www.oracle.com/us/products/servers-storage/servers/sparc-enterprise/.

TABLE 1. SPARC ENTERPRISE M8000 AND M9000 SERVER FEATURES

	SPARC ENTERPRISE M8000 SERVER	SPARC ENTERPRISE M9000 SERVER
ENCLOSURE	• One cabinet	• Up to two cabinets
PROCESSORS	• SPARC64 VII+ • 3.00 GHz • 12 MB L2 cache • Up to 16 quad-core chips	• SPARC64 VII+ • 3.00 GHz • 12 MB L2 cache • Up to 64 quad-core chips
MEMORY	• Up to 1 TB • 128 DIMM slots	• Up to 4 TB • 512 DIMM slots
INTERNAL I/O SLOTS	• 32 PCIe	• 128 PCIe
EXTERNAL I/O CHASSIS	• Up to 8 units	• Up to 16 units
INTERNAL STORAGE	• Serial Attached SCSI • Up to 16 drives	• Serial Attached SCSI • Up to 64 drives
DYNAMIC SYSTEM DOMAINS	• Up to 16	• Up to 24

Non-Stop/Continuous Database Operations

There is usually considerable commonality in the tasks undertaken by users connected to the same database instance, and users running transaction-based workloads, in particular, frequently access many of the same data blocks. For this reason, Oracle Database keeps frequently used data blocks in a cache in memory called the Buffer Cache, and it shares other frequently accessed information, such as table metadata and parsed (processed) SQL statements, in a second memory cache called the Shared Pool.

These memory caches are held in a single, shared memory to allow multiple users to access them concurrently. Shared memory also facilitates inter-process communication. Since shared memory is very heavily used in Oracle Database environments, it is important to optimize access to it and to minimize the amount of CPU consumed while referring to it. With this in mind, a specially tuned variant of System V Shared Memory, called Intimate Shared Memory (ISM), was introduced in Oracle Solaris many years ago. ISM has been widely used for database memory that is shared by Oracle Database software ever since.

With the use of ISM, Oracle Solaris and SPARC Enterprise M-Series servers are engineered to continue running through component failures, component replacement, and system expansion, but that is not enough. For nonstop database operations, systems must be capable of being divided into multiple fault-isolated servers, also known as Dynamic System Domains. Administrators must be able to move hardware resources, such as CPUs and memory, in and out of those domains using Dynamic Reconfiguration, and the database must be able to adapt to those changes using features in Oracle Database 11*g* Release 2, such as Automatic Memory Management.

In addition, the aforementioned is also made possible with enhancements that have been made to Oracle Solaris for nonstop database operations as well as for dynamic reconfigurations brought about by scheduled and unscheduled events based on customer needs and responses to hardware changes. The Oracle Solaris feature that provides responsiveness to adapting to these changes is Dynamic Intimate Shared Memory (DISM).

DISM provides shared memory with the same essential characteristics as ISM except that it is dynamically resizable. That means that DISM offers the performance benefits of ISM while allowing the shared memory segment to be dynamically resized, both for the sake of performance and to allow dynamic reconfiguration (for example, adding or removing memory from a system or a domain). This dynamic resizing can be scheduled or unscheduled based upon particular operating system and associated hardware events. For further information about DISM, refer to the Oracle white paper "Dynamic SGA Tuning of Oracle Database on Oracle Solaris with DISM."

Dynamic System Domains

Dynamic System Domains enable IT organizations to divide a single large system into multiple, fault-isolated servers—each running independent instances of the Oracle Solaris operating system, with access to designated I/O devices. When system components are exclusively dedicated to individual Dynamic System Domains, hardware or software faults in one domain remain isolated and unable to impact the operation of other domains. Each domain within a single server platform can run a different version of the Oracle Solaris operating system, making this technology extremely useful for pre-production testing of new or modified applications or for consolidation of multiple tiers of a database application.

Dynamic System Domains enable organizations to custom-tailor the compute capacity of SPARC Enterprise M-Series servers to meet specific enterprise needs. For instance, the SPARC Enterprise M9000 server can be configured as a single domain with up to 64 SPARC64 VII+ processors to host an exceptionally compute-intensive application. An organization with multiple databases that require isolation from one another might divide a single SPARC Enterprise M9000 server into as many as 24 isolated domains. Typical configurations involve three or four domains in a single SPARC Enterprise M-Series server. The number of domains configurable within each type of SPARC Enterprise server can be found in Table 1. Figure 2 is an illustration of Dynamic System Domains.

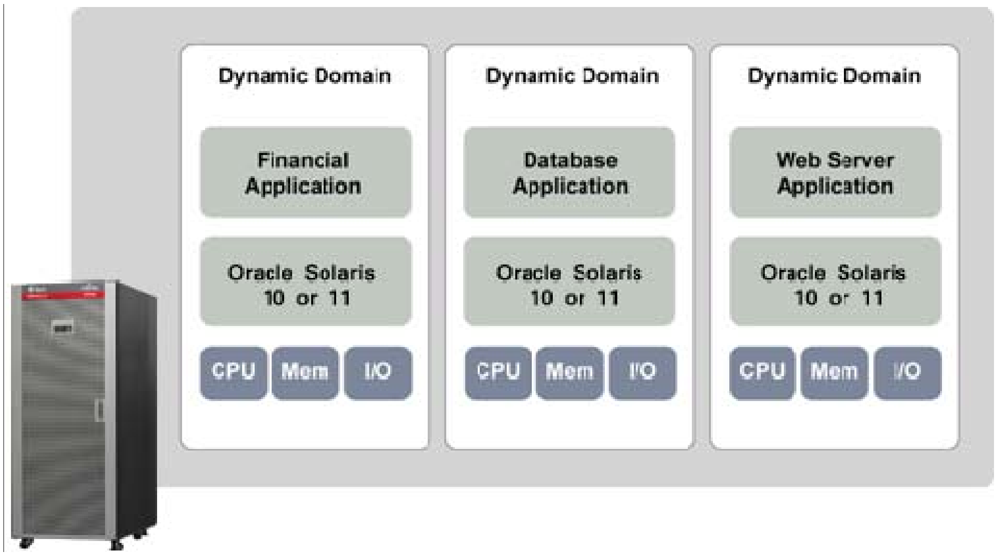


Figure 2. Organizations can run several different applications that require isolation from each other using Dynamic System Domains.

eXtended System Control Facility

The eXtended System Control Facility (XSCF) is at the center of remote monitoring and management capabilities in SPARC Enterprise M-Series servers. The XSCF is also a tool that system administrators use to configure Dynamic System Domains. The XSCF consists of a dedicated processor that is independent of the server system and runs the XSCF Control Package. The Domain to Service Processor Communication Protocol (DSCP) is used for communication between the XSCF and the server. The DSCP protocol runs on a private TCP/IP-based or PPP-based communication link between the service processor and each domain.

The XSCF regularly monitors the environmental sensors, provides early warning of potential error conditions, and executes proactive system maintenance procedures, as necessary. For example, the XSCF can initiate a server shutdown in response to temperature conditions that might cause physical system damage. The XSCF Control Package running on the service processor enables administrators to remotely control and monitor domains as well as the platform itself.

Using a network or serial connection to the XSCF, operators can effectively administer the server from anywhere on the network. Remote connections to the service processor run separately from the operating system and provide the full control and authority of a system console.

Redundant XSCF

Dual, redundant XSCFs are provided on SPARC Enterprise M8000 and M9000 servers. One XSCF is configured as active and the other is configured as a standby. The XSCF network between the two service processors enables the exchange of system management information. In case of failover, the service processors are already synchronized and ready to change roles.

Dynamic Reconfiguration of Hardware and Database

Dynamic Reconfiguration technology enables administrators to reallocate resources among Dynamic System Domains without interrupting critical systems. Using Dynamic Reconfiguration technology, administrators can perform physical or logical changes to system hardware resources while the server continues to execute applications. Removing and adding components, such as CPUs, memory, and I/O subsystems, from a running system helps reduce system downtime. In addition, Dynamic Reconfiguration makes maintenance and upgrades easier by eliminating the need for system reboots after hardware configuration changes.

Multiple Dynamic Reconfiguration operations can also execute simultaneously for efficient management of resources. This capability enables independent domain administrators to perform Dynamic Reconfiguration operations simultaneously without concern for the status of Dynamic Reconfiguration requests or executions in other domains.

Concurrent Maintenance

Administrators are increasingly challenged to carve out planned downtime windows to perform necessary service on essential systems. Concurrent maintenance refers to the ability to perform hardware configurations without impacting application availability. Since SPARC Enterprise M-Series servers are equipped with redundant components, failures no longer necessarily result in server downtime. However, the replacement of failed or downgraded components must occur at some point. Dynamic Reconfiguration enables maintenance operations to be completed while the system continues to operate. Even the assimilation of new components into running system domains need not interrupt critical services.

Automatic Dynamic Reconfiguration

Automatic Dynamic Reconfiguration in the Oracle Solaris operating system enables the execution of Dynamic Reconfiguration operations without interaction from a user. Automatic Dynamic Reconfiguration activities are triggered by pre-defined system events set by a system administrator. Implementations can include application-specific preparatory tasks before a Dynamic Reconfiguration operation, execution of error recovery actions during Dynamic Reconfiguration, and clean-up procedures after Dynamic Reconfiguration completion. For example, an Automatic Dynamic Reconfiguration event can be created to allow the automatic addition of a system board to a Dynamic System Domain when a business-critical application reaches full CPU utilization.

Dynamic Reconfiguration in Oracle Database 11*g* Release 2

Oracle Database 11*g* Release 2 includes several features that enable changes to be made to the instance configuration dynamically. For example, the dynamic SGA infrastructure can be used to alter an instance's memory usage. Dynamic SGA enables the size of the buffer cache, the shared pool, the large pool, and the process-private memory to be changed without shutting down the database instance. Oracle also provides transparent management of working memory for SQL execution by self-tuning the initialization runtime parameters that control allocation of private memory.

Another type of dynamic reconfiguration occurs when the Oracle Database polls the operating system to detect changes in the number of available CPUs and reallocates internal resources. In addition, some initialization parameters can be changed without shutting down the instance. The ALTER SYSTEM statement can be used to change the value of a parameter of the Oracle Database instance while the Oracle Database is actively running.

Oracle Solaris Cluster and Oracle Real Application Clusters

Oracle Solaris Cluster offers an extensive high availability and disaster recovery solution for Oracle Solaris environments, in physical and virtualized environments, for local and global datacenters. The software provides out-of-the box support for a large portfolio of applications and databases including Oracle Database and Oracle Real Application Clusters (Oracle RAC).

The Oracle Solaris Cluster framework monitors all the components of the environment including servers, the network, and storage, and it leverages the redundancy of the configuration to recover appropriately. Through application-specific module agents, it can start, stop, and monitor the health of the application and take corrective action to regain application availability upon failure. For example, it can restart the application or fail over the application to another healthy node.

Oracle RAC enables the deployment of database instances on multiple hosts that act on the single database residing on shared storage. The database is highly available, because access to the shared data is available as long as one Oracle RAC instance is online. Oracle RAC is scalable because new instances can be added on new hosts, up to the configuration limits of the cluster and interconnect.

The Appendix of this white paper has step-by-step deployment screenshots of Oracle RAC implemented upon two SPARC T4-2 servers.

Private Network and Cache Fusion

In an Oracle RAC database, multiple instances access a single set of database files. Each server in the cluster must be connected to a private network by way of a private interconnect. The interconnect serves as the communication path between nodes in the cluster to synchronize the use of shared resources by each instance.

Each database instance in an Oracle RAC database uses its own memory structures and background processes. When a data block located in the buffer cache of one instance is required by another instance, Oracle RAC uses Cache Fusion to transfer the data block directly between the instances using the private interconnect. This is much faster than having one database instance write the data blocks to disk and requiring the other database instance to reread the data blocks from disk. Cache Fusion technology enables the Oracle RAC database to access and modify data as if the data resided in a single buffer cache.

InfiniBand, Application Latency, and Fabric Convergence

Moving data between applications over a traditional network can be time consuming and can drain precious server resources. With traditional network technologies, data exchanges traverse the operating systems on both the source and destination servers, resulting in excessive application latency due to operating system calls, buffer copies, and interrupts. The Oracle Optimized Solution for Oracle Database employs InfiniBand networking for the Oracle RAC private interconnect. InfiniBand Remote Direct Memory Access (RDMA) technology provides a direct channel from the source application to the destination application, bypassing the operating systems on both servers. The InfiniBand channel architecture eliminates the need for the operating system intervention in network and storage communication. This provides a very high-speed, low-latency interface for more efficient and robust movement of data across the enterprise cluster. The use of InfiniBand networking also preserves server resources for other database processing.

InfiniBand technology delivers 40 Gb/sec connectivity with application-to-application latency as low as 1 microsecond. It has become a dominant interconnect fabric for high-performance enterprise clusters. InfiniBand provides ultra-low latency and near-zero CPU utilization for remote data transfers, making it ideal for high-performance clustered applications.

In addition to providing unrivaled access to remote application data, InfiniBand's industry-leading bandwidth enables fabric convergence, allowing all network, storage, and inter-process communication traffic to be carried over a single fabric. Converged fabrics aggregate the functions of dedicated, sole-purposed networks and alleviate the associated expense of building and operating multiple networks and their associated network interfaces. In addition, InfiniBand, along with Oracle Solaris, supports IPoIB (Internet Protocol over InfiniBand) for application and Oracle RAC database access, providing higher bandwidth and lower latency of access than can be achieved through Ethernet networking today.

The Appendix of this white paper has step-by-step deployment screenshots of Oracle RAC implemented upon two SPARC T4-2 servers which includes use of InfiniBand for the Private RAC interconnect communications.

Sun Datacenter InfiniBand Switch 36

The Sun Datacenter InfiniBand Switch 36 from Oracle offers low-latency, quad data rate (QDR), 40 Gb/sec fabric and cable aggregation for Oracle servers and storage. It supports a fully non-blocking architecture, and it acts as a self-contained fabric solution for InfiniBand clusters up to 36 nodes. The Sun Datacenter InfiniBand Switch 36 provides hardware support for adaptive routing, including InfiniBand 1.2 congestion control, which helps to eliminate fabric hotspots and to drive maximal throughput at the lowest-possible latencies. It is ideal for deployment with clustered databases and converged datacenter fabrics.

Advanced features support the creation of logically isolated subclusters, as well as traffic isolation and quality of service (QoS) management. The embedded InfiniBand fabric management module is enabled to support active/hot-standby dual-manager configurations, ensuring a seamless migration of the fabric management service in the event of a management module failure. The Sun Datacenter InfiniBand Switch 36 is provisioned with redundant power and cooling for high availability in demanding datacenter environments. When deployed in pairs, the InfiniBand fabric provides a highly available and resilient network for business-critical applications.

Oracle Solaris Predictive Self-Healing and Redundant Interconnects

A standard part of the Oracle Solaris 10 and 11 operating systems, Oracle Solaris Predictive Self-Healing software further enhances the reliability of all SPARC servers. The implementation of Oracle Solaris Predictive Self-Healing software for SPARC servers enables constant monitoring of all CPUs and memory. Depending upon the nature of the error, persistent CPU soft errors can be resolved by automatically taking a thread, a core, or an entire CPU offline. Similarly, memory page retirement capability enables memory pages to be taken offline proactively in response to repeated data access errors attributable to a specific memory DIMM. Redundant interconnects within SPARC Enterprise M9000 servers help ensure that even an unlikely problem with the passive backplane will not shut down database operations.

Evolution of Oracle's Multicore, Multithreaded Processor Design

While processor speeds continue to double every two years in accordance with Moore's law, memory speeds typically double only every six years. This growing disconnect is the result of memory suppliers focusing on density and cost, rather than speed, as their design center. As a result, memory latency now dominates much application performance, erasing even very impressive gains in clock rates.

Unfortunately, this relative gap between processor and memory speeds leaves ultrafast processors idle as much as 85% of the time, waiting for memory transactions to be completed. Ironically, as traditional processor execution pipelines get faster and more complex, the effect of memory latency grows. Worse, idle processors continue to draw power and generate heat. Grouping two or more conventional processor cores on a single physical die—creating multicore processors or chip multiprocessors—typically yields only slight improvements in performance since it replicates cores from existing (single-threaded) processor designs. However, the aggregate chip performance increases since multiple programs (or multiple threads) can be accommodated in parallel.

Chip Multithreading

Unlike complex single-threaded processors, multicore or multithreaded processors use the available transistor budget to implement multiple hardware multithreaded processor cores on a chip die. First introduced with the UltraSPARC T1 processor, multicore and multithreading take advantage of chip multiprocessing advances but add a critical capability—the ability to scale with threads rather than frequency.

Unlike traditional single-threaded processors and even most current multicore processors, hardware multithreaded processor cores allow rapid switching between active threads as other threads stall for memory. Figure 3 illustrates the difference between chip multiprocessing (CMP), fine-grained hardware multithreading, and multicore with multithreading in Chip Multithreading (CMT). The key to CMT is that each core in an Oracle multicore and multithreaded processor is designed to switch between multiple threads on each clock cycle. As a result, the processor's execution pipeline remains active doing real useful work, even as memory operations for stalled threads continue in parallel.

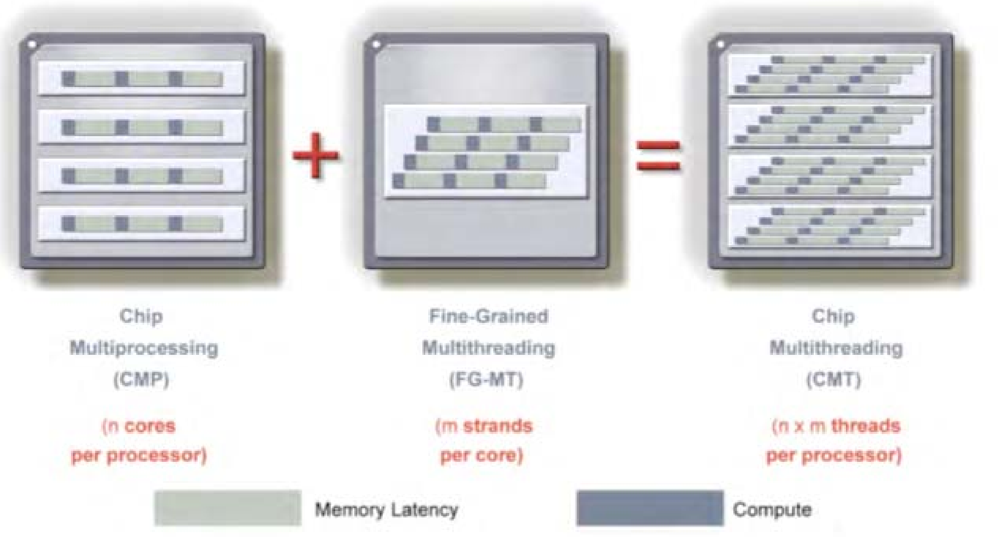


Figure 3. Oracle’s multicore/multithreaded approach combines CMP and hardware multithreading.

Oracle’s multicore/multithreading approach to processor design provides real value since it increases the ability of the execution pipeline to do actual work on any given clock cycle. Use of the processor pipeline is greatly enhanced because a number of execution threads now share its resources. The negative effects of memory latency are effectively masked, because the processor and memory subsystems remain active in parallel to the processor execution pipeline. Since these individual processor cores implement much-simpler pipelines that focus on scaling with threads rather than frequency, they are also substantially cooler and require significantly less electricity to operate.

The SPARC T4 Processor

Each SPARC T4 processor provides eight cores, with each core supporting up to eight threads (64 threads per processor). In addition, each core provides two integer execution pipelines, so a single SPARC core is capable of executing two threads at a time.

The SPARC T4 core architecture includes aspects that are conventionally associated with superscalar designs, such as out-of-order (OOO) instruction execution, sophisticated branch prediction, prefetching of instructions and data, deeper pipelines, three levels of caches, support for a much larger page size (2 GB), and multiple instruction dispatching. All these characteristics are included to improve single-thread execution, networking, and throughput performance.

The SPARC T4 processor has a three-level cache architecture. Levels 1 and 2 are specific to each core and are not shared with other cores. Level 3 is shared across all cores of a given processor. The SPARC T4 processor has Level 1 caches that consist of separate data and instruction caches; both are 16 KB per core. A single Level 2 cache, again per core, is 128 KB. The Level 3 cache is shared across all eight cores of the SPARC T4 processor and is 4 MB.

The SPARC T4 processor eliminates the need for expensive custom hardware and software development by integrating computing, security, and I/O onto a single chip.

Integrated Networking

The SPARC T4 processor provides integrated on-chip networking. All network data is supplied directly from and to main memory. Placing networking so close to memory reduces latency, provides higher memory bandwidth, and eliminates inherent inefficiencies of I/O protocol translation. The SPARC T4 processor provides two 10 gigabit Ethernet (GbE) ports with integrated serializer/deserializer (SerDes). Multiple DMA engines (16 transmit and 16 receive DMA channels) match DMAs to individual threads, providing binding flexibility between ports and threads.

Stream Processing Unit

The Stream Processing Unit (SPU) is implemented within the core as part of the pipelines themselves, and it operates at the same clock speed as the core. The SPU is designed to achieve wire-speed encryption and decryption on the processor's 10 GbE ports. The SPARC T4 processor supports twelve widely used cryptographic algorithms.

Integral PCI Express Generation 2 Support

SPARC T4 processors provide dual on-chip PCIe Generation 2 interfaces. Each operates at 5 Gb/sec per x1 lane bi-directionally through a point-to-point dual-simplex chip interconnect, meaning that each x1 lane consists of two unidirectional bit-wide connections. An integral Input/Output Memory Management Unit (IOMMU) supports I/O virtualization and process device isolation by using the PCIe BUS/Device/Function (BDF) number. The total theoretical I/O bandwidth (for a x8 lane) is 4 GB/sec. The actual realizable bandwidth is more likely to be approximately 2.8 GB/sec.

Power Management

Beyond the inherent efficiencies of Oracle's multicore/multithreaded design, the SPARC T4 processor incorporates unique power management features at both the core and memory levels of the processor. These features include reduced instruction rates, parking of idle threads and cores, and the ability to turn off clocks in both cores and memory to reduce power consumption.

SPARC T4 Server Overview

The SPARC T4 server is based on fifth-generation multicore multithreading technology. The SPARC T4 server can support one-, two-, and four-socket implementations, known as the SPARC T4-1, T4-2, and T4-4 servers respectively.

SPARC T4-1 Server Overview

The compact SPARC T4-1 server provides significant computational power in a space-efficient, low-power 2RU rackmount package. With a low price-to-performance ratio, a low acquisition cost, and tightly integrated high-performance 10 GbE, this server is ideally suited to the delivery of horizontally scaled transaction and Web services that require extreme network performance. The server is designed to address the challenges of modern datacenters with greatly reduced power consumption and a small physical footprint. Coupled with Oracle RAC, SPARC T4-1 servers provide a sound foundation for the Oracle Optimized Solution for Oracle Database.

The SPARC T4-1 server includes the following major components:

- One SPARC T4 processor with eight cores operating at 2.85 GHz
- Up to 256 GB of memory in 16 DDR3 DIMM slots (4 GB, 8 GB, and 16 GB DDR3 DIMMs supported)
- Four onboard 10/100/1000 Mb/sec Ethernet ports
- Dedicated low-profile PCIe slots (x8) with two combination 10 GbE Attachment Unit Interface (XAUI) or low-profile PCIe x4 slots
- Five USB 2.0 ports
- Eight disk drive slots that support SAS-2 commodity disk drives
- ILOM 3.0 system controller
- Two (N+1) hot-swappable, high-efficiency 1200 watt AC power supplies
- Six fan assemblies (each with two fans), under environmental monitoring and control, N+1 redundancy; fans are accessed through a dedicated top panel door

SPARC T4-2 Server Overview

The expandable SPARC T4-2 server is optimized to deliver transaction and Web services, including Java Platform Enterprise Edition application services, enterprise application services such as enterprise resource planning (ERP), customer relationship management (CRM), and supply chain management (SCM). The SPARC T4-2 server is superb for the Oracle Optimized Solution for Oracle Database in mission-critical environments when coupled with Oracle RAC.

With considerable expansion capabilities and integrated virtualization technologies, the SPARC T4-2 server is also an ideal platform for consolidated Tier 1 and Tier 2 workloads.

The SPARC T4-2 server includes the following major components:

- Dual SPARC T4 processors with eight cores per processor operating at 2.85 GHz
- Up to 512 GB of memory in 32 DDR3 DIMM slots (4 GB, 8 GB, and 16 GB DDR3 DIMMs)
- Four onboard 10/100/1000 Mb/sec Ethernet ports
- Ten dedicated low-profile PCIe slots

- One dedicated slot for 4x XAUI port (this port cannot be shared with any PCIe card)
- Five USB 2.0 ports
- Six available disk drives slots that support SAS-2 commodity disk drives
- ILOM 3.0 system controller
- Two (N+1) hot-pluggable/hot-swappable high-efficiency 2000 watt AC power supplies
- Six fan assemblies under environmental monitoring and control, N+1 redundancy

SPARC T4-4 Server Overview

With support for up to 256 threads, memory up to 1 TB, cryptographic acceleration, and integrated on-chip I/O technology, the SPARC T4-4 server is ideal for providing high throughput within significant power, cooling, and space constraints. With breakthrough levels of price/performance, this server is ideally suited as compute nodes within horizontally scaled environments for demanding mid-tier application server deployments or Web-tier and application-tier consolidation. In addition, when coupled with Oracle RAC, medium-to-large Oracle Databases are also serviced very well by these servers for the Oracle Optimized Solution for Oracle Database. The SPARC T4-4 server's large capacity presents opportunities for it to be used as a consolidation and virtualization server. Depending on the model selected, the SPARC T4-4 server features dual or quad-SPARC T4 processors. Uniquely setting this processor apart from past multicore processors is its ability to deliver a 5X performance increase on single-threaded workloads.

The SPARC T4-4 server includes the following major components:

- Two or four SPARC T4 processors with eight cores per processor operating at a clock speed of 3.00 GHz
- Up to 1.024 GB of memory in 64 DDR3 DIMM slots (4 GB, 8 GB, and 16 GB DDR3 DIMMs are supported)
- Four onboard 10/100/1000 Mb/sec Ethernet ports
- Sixteen x8 PCIe Generation 2 slots (via EMs)
- Eight XAUI ports via 2 QFSP quad connectors
- Four USB 2.0 ports
- Eight available disk drives slots that support SAS-2 commodity disk drives
- ILOM 3.0 system controller
- Four (N+N) hot-swappable, high-efficiency 2060 watt AC power supplies
- Five fan assemblies under environmental monitoring and control, 2 + 2 redundancy

SPARC T4 Server Virtualization

Oracle VM Server for SPARC is a virtualization technology supported on all Oracle CMT servers, including the SPARC T4 server. Oracle VM Server for SPARC leverages the built-in hypervisor to subdivide system resources down to the CPUs thread, cryptographic processor, memory, and PCI bus, by creating partitions called logical (or virtual) domains. Each logical domain runs in one or more dedicated CPU threads.

Oracle VM Server for SPARC delivers the following features and capabilities:

- Secure Live Migration—Migrate an active domain to another physical machine while maintaining application services to users with secure, wire-speed encryption capabilities for live migration.
- PCIe Direct I/O—Assign either individual PCIe cards or entire PCIe buses to a guest domain. This delivers native I/O throughput.
- Dynamic Reconfiguration (DR)—Allow computing resources to be dynamically added or removed on an active domain. CPUs, virtual I/O, cryptographic units, and memory can be added or removed on an active domain.
- Advanced RAS—The logical domain supports virtual disk mutipathing and failover, as well as faster network failover with link-based IP multipathing (IPMP) support. The logical domain can also handle path failure between an I/O domain and storage. Moreover, the domain is fully integrated with the Oracle Solaris Fault Management Architecture (FMA), which enables predictive self-healing.
- CPU Whole Core Allocation and Core Affinity—These capabilities enable organizations to optimize the assignment of virtual CPUs to deliver higher and more-predictable performance for all types of application workloads.
- CPU Dynamic Resource Management (DRM)—DRM enables a resource management policy and domain workload to trigger the automatic addition and removal of CPUs, which helps to better align business priorities.
- Physical-to-Virtual (P2V) Conversion—Quickly convert an existing SPARC server that runs the Solaris 8, Solaris 9, or Oracle Solaris 10 OS into a virtualized Oracle Solaris 10 image. Use this image to facilitate OS migration into the virtualized environment.
- CPU Power Management—Implement power saving by disabling each core on a SPARC T4 processor that has all of its CPU threads idle. Power management capability is further enhanced by providing CPU clock speed adjustment, memory power management, and power-limit settings to ensure that energy consumption is optimized with utilization.
- Advanced Network Configuration—Obtain more-flexible network configurations, higher performance, and scalability with the following features: Jumbo frames, VLANs, virtual switches for link aggregations, and network interface unit (NIU) hybrid I/O.

Oracle Solaris for Multicore Scalability

Oracle Solaris 10 and, when it is released, Oracle Solaris 11 are specifically designed to deliver the considerable resources of SPARC T4 processor-based systems. Oracle Solaris 10 and 11 have incorporated many features to improve application performance on Oracle's multicore/multithreaded architectures.

Accelerated Cryptography

Accelerated cryptography is supported through the cryptographic framework in Oracle Solaris as well as the SPARC T4 processor. The SPARC T4 processor permits access to cryptographic cypher hardware implementations through user-level instructions. The cyphers are implemented within the appropriate pipeline itself rather than as a co-processor. This means both a more efficient implementation of the hardware-based cyphers as well as no privilege-level changes, resulting in increased efficiency in cryptographic algorithm calculations. In addition, database operations can make more efficient use of the various cryptographic cyphers that are implemented within the instruction pipeline itself.

Critical Thread Optimization

Oracle Solaris 10 and 11 now have the ability to permit either a user or programmer to allow the Oracle Solaris Scheduler to recognize a "critical thread" by raising its priority to 60 or above through the either the Command Line Interface or system calls to a function If this is done, the thread will run by itself on a single core, garnering all the resources of that core for itself. To prevent resource starvation to other threads, this feature is disabled when there are more runnable threads than available CPUs.

Multicore/Multithreaded Awareness

Oracle Solaris 10 and 11 are aware of the SPARC T4 processor hierarchy, so the Oracle Solaris Scheduler can effectively balance the load across all the available pipelines. Even though it exposes each of these processors as 64 logical processors, Oracle Solaris understands the correlation between cores and the threads they support, and it provides a fast and efficient thread implementation.

Fine-Granularity Manageability

For the SPARC T4 processor, Oracle Solaris 10 and 11 have the ability to enable or disable individual cores and threads (logical processors). In addition, standard Oracle Solaris features, such as processor sets, provide the ability to define a group of logical processors and schedule processes or threads on them.

Binding Interfaces

Oracle Solaris allows considerable flexibility in that processes and individual threads can be bound to either a processor or a processor set, if needed.

Support for Virtualized Networking and I/O

Oracle Solaris contains technology to support and virtualize components and subsystems on the SPARC T4 processor, including support for the on-chip 10 GbE ports and PCIe interface. As a part of a high-performance network architecture, Oracle multicore/multithreaded-aware device drivers are provided so that applications running within virtualization frameworks can effectively share I/O and network devices.

Non-uniform Memory Access Optimization in Oracle Solaris

Memory is managed by each SPARC T4 processor on the SPARC T4-2 and T4-4 servers, so these implementations represent a non-uniform memory access (NUMA) architecture. In NUMA architectures, the time needed for a processor to access its own memory is slightly shorter than that required to access memory managed by another processor. Oracle Solaris provides technology that can specifically help to decrease the impact of NUMA on applications and improve performance on NUMA architectures:

- Memory Placement Optimization (MPO)—Oracle Solaris 10 uses MPO to improve the placement of memory across the physical memory of a server, resulting in increased performance. Through MPO, Oracle Solaris 10 works to help ensure that memory is as close as possible to the processors that access it, while still maintaining enough balance within the system. As a result, many database applications are able to run considerably faster with MPO.
- Hierarchical Lgroup Support (HLS)—HLS improves the MPO feature in Oracle Solaris. HLS helps Oracle Solaris optimize performance for systems with more-complex memory latency hierarchies. HLS lets Solaris OS distinguish between the degrees of memory remoteness, allocating resources with the lowest-possible latency for applications. If local resources are not available by default for a given application, HLS helps Oracle Solaris allocate the nearest remote resources.

Oracle Solaris ZFS

Oracle Solaris ZFS offers a dramatic advance in data management, automating and consolidating complicated storage administration concepts and providing unlimited scalability with the world’s first 128-bit file system. Oracle Solaris ZFS is based on a transactional object model that removes most of the traditional constraints on I/O issue order, resulting in dramatic performance gains. Oracle Solaris ZFS also provides data integrity, protecting all data with 64-bit checksums that detect and correct silent data corruption.

A Secure and Robust Enterprise-Class Environment

Existing SPARC applications continue to run unchanged on SPARC T4 platforms, protecting software investments. Certified multilevel security protects Oracle Solaris environments from intrusion. The Fault Management Architecture in Oracle Solaris means that elements such as Oracle Solaris predictive self-healing can communicate directly with the hardware to help reduce both planned and unplanned downtime. Effective tools, such as Oracle Solaris DTrace, help organizations tune their applications to get the most out of a system’s resources.

Oracle's Pillar Axiom 600 Storage System Overview

Oracle's Pillar Axiom 600 storage system delivers enterprise-class high availability for Oracle Database deployments. It is designed to eliminate single points of failure with redundant system components.

- The Pillar Axiom Pilot is a dual-redundancy policy controller that performs the management function for Pillar Axiom storage systems.
- The Pillar Axiom Slammer is a dual-redundancy storage controller that virtualizes the storage pool for Pillar Axiom storage systems and moves and manages data.
- The Pillar Axiom Brick is a storage enclosure that houses two RAID controllers and thirteen disk drives, including one hot-swappable spare. Bricks are available with SSD, FC, and SATA.

The modular architecture of the Pillar Axiom 600 storage system allows administrators to quickly replace any failed components without disrupting or slowing system performance. Bricks are organized into two sets of six-disk RAID-5 or RAID-10 groups. A local hot-spare disk drive is available to replace a failed disk, allowing RAID rebuilds to begin immediately and data to be restored in hours instead of days.

Sophisticated software layered on top of the hardware also ensures high availability. Software processes on each Slammer control unit (CU) constantly communicate on status and performance. The Pillar Axiom 600 storage system uses a double-safe write system and secures the I/O in battery-backed, nonvolatile RAM (NVRAM). The Pillar Axiom 600 storage system's redundant CU architecture secures the I/O in both CUs before it is acknowledged as a complete transaction.

Key Differentiating Technologies

The Pillar Axiom 600 storage system provides the following advantages:

- Modular scalability provides the ability to dynamically scale performance and capacity independently.
- To ensure Quality of Service (QoS), application prioritization and contention management enable multiple applications and databases to co-exist on the same storage system. A set of policies is created to govern the QoS for each file system or LUN.
- Storage domains prevent co-mingling of data and isolate customers, users, and applications to disparate storage enclosures.
- Application profiles provide application-oriented provisioning and storage management linked to application priority. These profiles determine how data is laid out on the disk drive to ensure that data is laid out on the portion of the disk that best supports the application's performance priority.
- Distributed RAID provides superior scalability and performance even during drive rebuilds by moving RAID into storage enclosures.
- The architecture lends itself to consolidating applications and databases in a simplified but very robust and reliable manner.

The Appendix of this white paper provides step-by-step deployment screenshots of the creation of Pillar Axiom 600 storage system SAN LUNs and the assignment of these LUNs to the Oracle Real Application Cluster SPARC T4-2 servers that were used to demonstrate the concepts discussed in this white paper.

Workload Scalability with Flash Technology

In general, databases run fastest when the entire working set can be kept in memory. This is sometimes possible, but most often is not, because main memory is very expensive, and useful working sets tend to be large, or at least larger than the amount of memory for which budget is available. The solution is flash technology, which is less expensive per unit of memory yet still provides solid-state's fast access times. The issue for most technology companies is how specifically to manage flash technology as an intermediate form of storage between the database system's main memory and its rotating disk storage.

SPARC Enterprise M-Series servers are highly scalable. Oracle Solaris provides the ability to scale linearly and reliably, as the SPARC Enterprise M-Series servers scales up to 4 terabytes of memory and up to 64 sockets, with 256 cores and 512 threads supported on the largest system, the SPARC Enterprise M9000 server. 9,248 disk drives can be directly attached to the SPARC Enterprise M9000 server, and even more are possible by using SAN technology.

Understanding and Addressing I/O Performance Issues

Although I/O patterns and performance can be a complex subject, the majority of I/O performance issues involve single-threaded latency, streaming data rate, or multithreaded asynchronous Input/Output Operations Per Second (IOPS).

Single-Threaded Latency

Single-threaded latency measures the total time required to complete a write operation or read operation. This total time required is sometimes referred to as service time. Usually measured in milliseconds for hard disk drives and microseconds for flash or memory-based storage, service time is the length of time from command initiation until I/O operation completion. Measurements that determine the service time offered by a particular device are generally taken while no other concurrent demands are placed on the storage subsystem and for operations on a relatively small or single block of data—512 bytes or a few kilobytes.

Since most block-based storage systems today utilize hard disk drive technology, service times for small blocks are limited by mechanical behavior—moving actuator arms over the correct track on a rotating disk (seek latency) or waiting for the correct block or sector of data to rotate under the head to be written or read (rotational latency). Today, even the best hard disk drives create total latencies of a few milliseconds for random block accesses. Commodity and consumer-grade hard drives can take over 10 milliseconds. In contrast, Sun FlashFire technology from Oracle can deliver average 8 KB read and write service times of about 400 microseconds, providing a ten-fold reduction in service time over hard disk drives.

From an application perspective, the impact of service time is often greatest for synchronous I/O operations that lack the benefit of caching. For instance, performance can stall in cases where the application, database, or operating system process thread is blocked or waiting for I/O completion response before allowing other processing to continue. Code sections that suffer the performance impact of long service times include database index lookups and writes to redo logs within online transaction workloads. In many cases, every microsecond of service time for synchronous I/O operations results in wait time for the database and its application.

Maximum Streaming Data Rate

Often measured in megabytes per second, the maximum streaming data rate employs very large data block sizes for the read or write operations. Each data block can be hundreds of kilobytes or even a few megabytes long. More than one concurrent I/O thread or queue is often needed to keep the storage subsystem busy enough to achieve its best performance for this data movement metric.

The rotational speed of the drive and the bit recording density of the drive platters limit the data rate. In some cases, the channel bandwidth of the disk controllers and storage interconnects are the performance limiters. For databases, stream performance bottlenecks can often be reached quickly in table scans in online analytical processing (OLAP), decision support systems (DSSs), and business intelligence data warehousing (BIDW) workloads.

Multithreaded, Asynchronous IOPS

There are cases where the saturation I/O rate of random accesses to the storage subsystem or hard disk drive limits database and application performance or scaling. Most databases use the Stripe And Mirror Everything (SAME) strategy to spread this type of I/O workload over as many hard disk drives as possible and as uniformly as possible. This technique seeks to reduce "hot spots" in the storage subsystem and can keep response time across the entire storage subsystem within asynchronous processing limits.

Modern enterprise databases often parallelize queries and update tablespaces using asynchronous I/O operations. Database consistency and integrity is maintained with less frequent synchronous checkpoints, while journals and logs are usually applied as part of application commitment requests. Even modest transactions or queries can generate tens or even hundreds of physical storage subsystem I/O operations. Since they are often asynchronous, the database and application can often continue processing in parallel with I/O completion.

Flash Technology

Sun FlashFire technology takes advantage of the capabilities of standard flash technology and adds advanced performance and reliability features. Utilizing devices with Sun FlashFire technology alongside traditional storage arrays can help organizations provide dramatic performance acceleration to meet the performance challenges of data-intensive computing.

Oracle's portfolio of products with Sun FlashFire technology include the following:

- Oracle's Sun Flash Modules are the building blocks for all other devices in the family of Oracle products with Sun FlashFire technology. Sun FlashFire Modules deliver ultra-high I/O performance with many enterprise features in an ultra-compact SODIMM sized form factor.
- Oracle's Sun Storage F5100 Flash Array utilizes Sun FlashFire technology to deliver over 1M IOPS in 1 rack unit (1U), improving database application response times two to five times, eliminating bottlenecks, and slashing energy consumption by up to 80% compared to spinning disks.

Sun Storage F5100 Flash Array

The Sun Storage F5100 Flash Array is well suited for inclusion in a hybrid storage architecture that includes Oracle's Sun Storage 6000 disk arrays. With this new approach to data storage, organizations can vastly accelerate the performance of enterprise database applications. The Sun Storage F5100 Flash Array is a simple, high-performance, eco-efficient solid-state storage solution. By incorporating enterprise-quality flash memory modules, the Sun Storage F5100 Flash Array can provide low latency at an affordable price, delivering fast access to critical data and enhancing the responsiveness of database applications—without requiring modifications to the application code.

The Sun Storage F5100 Flash Array accelerates databases with over 1 million IOPS performance and up to 100 times less power and space than traditional disk-based solutions for the same performance level. Exceeding the IOPS performance of over 3,000 disk drives, and more than four times the performance of other flash technology-based systems, the Sun Storage F5100 Flash Array provides a cost-effective building block that can deliver breakthrough value for high-performance database applications.

Using low-latency solid-state memory modules, the Sun Storage F5100 Flash Array delivers performance, capacity, and low power consumption in a compact enclosure. With as much as 2 terabytes in a rackmountable 1U chassis, it offers far greater capacity than a bank of solid-state disk devices at an affordable price.

One Sun Storage F5100 Flash Array is designed as four separate SAS domains that can be attached to a number of servers or Dynamic System Domains. The Sun Storage F5100 Flash Array supports a variety of configurations, so storage architects can design flexible, cost-effective solutions that complement the existing storage infrastructure and meet performance, capacity, and availability goals. In this way, the array helps to deliver substantial ROI for business-critical database applications. Table 2 provides a brief summary of the Sun Storage F5100 Flash Array features.

TABLE 2. SUN STORAGE F5100 FLASH ARRAY FEATURES

FEATURE	SUN STORAGE F5100 FLASH ARRAY
STORAGE DENSITY	• Random reads: 1.6 million IOPS (4 KB block size, 32 threads) • Random writes: 1.2 million IOPS (4 KB block size, 32 threads) • Sequential reads: 12.8 million IOPS (4 KB block size, 32 threads) • Sequential writes: 9.7 GB/sec (1 MB block size, 32 threads) • Read latency: 405 microseconds (4 KB block size, single thread) • Write latency: 282 microseconds (4 KB block size, single thread)
PERFORMANCE	• Sixteen x4 mini-SAS ports (4 x 3 Gb/sec mini-SAS ports per domain)
CONNECTIVITY	• Sixteen x4 mini-SAS ports (four x4 x 3 Gb/sec mini-SAS ports per domain)
RELIABILITY	• Redundant power supplies and fan modules • Supercomputer-backed DRAM • Storage Common Array Manager for system manager
POWER CONSUMPTION	• 2.1 Watts per Sun Flash Module; 300 Watts per fully populated array.

Innovative Storage System Design

With a rackmount design that complements other storage and server products from Oracle, the Sun Storage F5100 Flash Array addresses fast-access database storage requirements at new economic levels. The Sun Storage F5100 Flash Array offers:

- Low latency—Flash technology completes I/O operations in microseconds, placing flash technology between hard disk drives and DRAM in terms of access times. Because flash technology contains no moving parts, flash technology avoids the seek times and rotational latencies inherent in traditional hard disk drive technology. As a result, data transfers to and from the Sun Storage F5100 Flash Array are significantly faster than what electro-mechanical disk drives can provide. A single Sun Flash Module can provide thousands of IOPS for write operations and tens of thousands of IOPS for read operations, compared to only hundreds of IOPS provided by hard disk drives.
- High reliability—Reliability features help to increase availability and meet Service Level Agreement (SLA) targets. The Sun Storage F5100 Flash Array has a relatively small part count and is designed specifically for high reliability, availability, and serviceability (RAS). Sun Flash Modules incorporate features such as wear leveling, ECC, and block mapping, and they are subject to rigorous quality standards to sustain enterprise-level reliability. An Energy Storage Module (ESM) contains supercapacitor units that help to flush Sun Flash Module metadata safely from DRAM to flash memory in the event of a sudden power loss. Redundant power supplies and fans also enhance reliability, helping to provide continuous operation.

- Simplified management—Oracle's Sun Storage Common Array Manager software provides a consistent interface to administer storage products from Oracle, including the Sun Storage F5100 Flash Array. The software supplies an easy-to-use interface to perform administrative tasks such as configuration, firmware upgrades, maintenance, and device monitoring. Oracle Database 11*g* Release 2 automatically manages SGA data placement between memory and the Sun Storage F5100 Flash Array, removing the need for administrators to manage the placement of data objects.
- Flexible configurations and price points—The Sun Storage F5100 Flash Array can be deployed in virtually any situation that accepts a SAS-attached storage appliance. The array can be partially or fully populated, allowing storage architects to design the most cost-effective solution for the application at hand. Each SAS domain in the array holds up to 20 Sun Flash Modules that provide up to 480 GB of usable storage. Array configurations are available with 20 (480 GB), 40 (960 GB), or 80 (1.92 TB) flash modules. To expand capacity and performance, additional flash modules can be added as upgrades to a single domain or across all four domains. When all four array domains are fully populated in the 1U rackmount chassis, a single array provides a maximum capacity of 1.92 TB.
- Ultra-dense flash array packaging—With a 1U chassis, the Sun Storage F5100 Flash Array holds up to 80 Sun Flash Modules in a minimum amount of space, making flash technology easy to integrate into new and existing deployments. The innovative ultra-dense I/O design provides 24 GB/sec throughput with 64 SAS1 lanes that can deliver 3 Gb/sec via 16 four-wide SAS1 ports, allowing organizations to get the performance they need where they need it.

Using Flash to Extend the System Global Area

In Oracle Database software, the System Global Area (SGA) is a group of shared memory areas that are dedicated to an Oracle Database instance. Oracle Database processes use the SGA to store incoming data (data and index buffers) and internal control information that is needed by the database. Utilizing the Sun Storage F5100 Flash Array to extend the Oracle shared SGA can have a positive impact on performance levels.

Traditionally, the size of the SGA is limited by the size of the available physical memory. Oracle Flash Cache allows extension of the SGA size and caching beyond physical memory to a Sun Storage F5100 Flash Array.

To illustrate the performance of Oracle Flash Cache, Oracle created a benchmark workload that consisted of a high volume of SQL select transactions accessing a very large table in a typical business-oriented OLTP database. The database consisted of various tables, such as products, customers, orders, and warehouse inventory data. The warehouse inventory table alone was three times the size of the `db_buffer_size`—decreasing the probability of completing the majority of data retrievals from the database in-memory cache.

To obtain a baseline, throughput and response times were measured by applying the workload against a traditional storage configuration constrained by disk I/O demand. The workload was then executed with a Sun Storage F5100 Flash Array that was configured to contain an Extended SGA of incremental size. During each test, the in-memory SGA was limited to 25 GB. The Extended SGA was allocated on a "raw" Oracle Solaris Volume created with the Oracle Solaris Volume Manager on a set of Sun FlashFire Modules residing on the Sun Storage F5100 Flash Array. The benchmark test configuration is listed in Table 3.

TABLE 3. EXTENDED SGA CACHING ON SUN STORAGE 5100 FLASH ARRAY BENCHMARK CONFIGURATION

SERVER CONFIGURATION	• SPARC Enterprise 5000 Server • Eight SPARC64 VII 2.4 GHz quad-core processors • 96 GB memory
STORAGE CONFIGURATION	• Eight Sun Storage Arrays from Oracle • 12x 146 GB 15K RPM disks each (96 disks total) • One Sun Storage F5100 Flash Array
SOFTWARE CONFIGURATION	• Oracle Database 11*g* Release 2 • Oracle Solaris 10

Test Results

The tests results graphed in Figure 4 show the performance impact of increasing the Flash Cache size of the Extended SGA. As flash cache size increased, throughput scaled steadily and transaction response time decreased. In fact, scaling the Sun Storage F5100 Flash Array using the Oracle Extended SGA feature into the hundred-gigabyte range demonstrated an almost 500% performance improvement. Compared to traditional RAM-based solutions, the Sun Storage F5100 Flash Array using Oracle Extended SGA provides much faster I/O at a more reasonable cost than equivalently sized RAM.

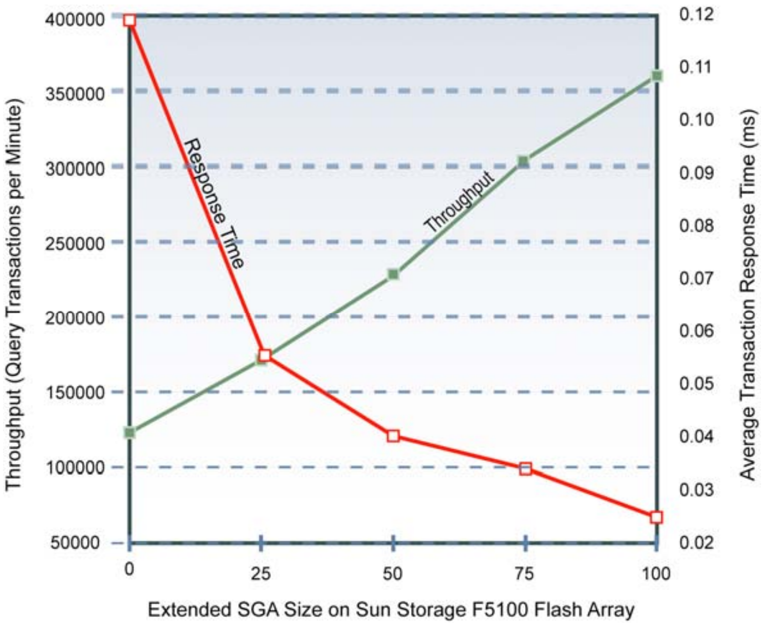


Figure 4. Utilizing the Sun Storage F5100 Flash Array to store an Oracle Extended SGA can significantly improve database performance.

Reference Architecture—Using the Sun Storage F5100 Flash Array as Storage for Database Indexes

Storing database indexes on flash storage can improve performance in two ways:

- Given a fixed workload, indexes on flash storage can reduce application response time
- Given a fixed desired response time, indexes on flash storage can increase the maximum supportable workload

A database can be segregated into three components, as shown in Figure 5:

- Production table data files
- Production index files
- Flash recovery area

All table data files are contained in the production files (marked *P* in the figure), all index files are contained in the production index files (marked *P'*), and all recovery-related files are contained in the flash recovery area (FRA, marked *F* in the figure). The online redo log and control files are multiplexed over *P* and *F*. Although this example uses the Sun Storage F5100 Flash Array for storage of database indexes, the array may also be used to support any database data file that requires the low-latency and high-bandwidth features of flash storage.[1]

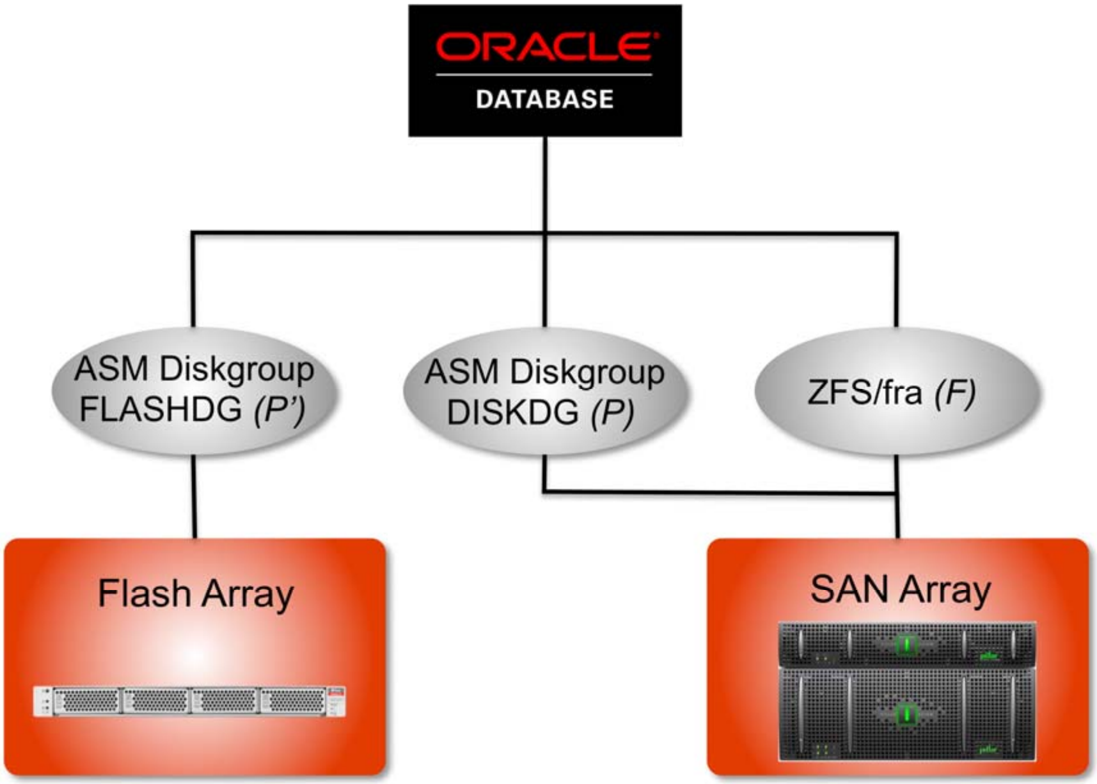


Figure 5. Architecture for using the Sun Storage F5100 Flash Array to store database indexes.

The Oracle Database 10*g* or 11*g* instance communicates with Oracle Automatic Storage Management (ASM) and Oracle Solaris ZFS for access to production and recovery data (see Table 4). Oracle ASM provides mirrored data protection (normal redundancy) for index files stored on the Sun Storage F5100 Flash Array. This storage target is designed to deliver ultra-low (1 to 3 ms) data access time for single-block reads at extreme access density (greater than 30 IOPS/GB) for a small subset of critical data that is needed early and often by important transactions.

[1] The study from which these results are excerpted can be found in Jeff Wright's June 2010 Oracle white paper titled "Accelerating Databases with Oracle's Sun Storage F5100 Flash Storage Array."

TABLE 4. SOFTWARE CONFIGURATION

SOFTWARE CONFIGURATION	DATA LAYOUT
• Oracle Solaris 10 5/09 • Oracle ASM with Oracle Database 10*g*/11*g* • Oracle ASM normal redundancy (flash) • Oracle ASM external redundancy (SAN) • Oracle Solaris ZFS dynamic striping (SAN)	• FLASHDG: Indexes • DISKDG: Online redo, table data • Oracle ASM normal redundancy (flash) • Oracle Solaris ZFS: online redo, FRA, archived redo

As part of this reference architecture, Oracle ASM implements striping (external redundancy) over hardware-mirrored data protection for database table data files stored on the Sun storage array. This storage target is designed to deliver low data access times (5 to 15 ms) at high data access rates (1–3 IOPS/GB) for mission-critical production data and log files.

For further detail regarding best practices and associated storage layout for Oracle Optimized Solution for Oracle Database in mission-critical environments, refer to the Oracle white paper “Oracle Optimized Solution for Oracle Databases: Storage Best Practices.” Oracle Solaris ZFS provides a scalable and feature-rich storage solution to help ensure efficient space utilization and enterprise-class data protection to completely protect the critical recovery components of Oracle Database, including online redo log, archived redo log, backup sets, and FRA.

System Performance Testing

This section presents the results of a database study that compared the performance of storage architectures taking advantage of Sun Storage F5100 Flash Arrays for indexes with storage architectures based only on traditional SAN disk.

This study compared:

- Client-side new order service time versus new order rate
- Oracle instance measured (STATSPACK) tablespace I/O read service time versus I/O rate

Method

The assessment method employed sought to determine the service time versus workload characteristics of the Sun Storage F5100 Flash Array in the context of Oracle Database supporting an online transaction processing (OLTP) application. The workload used in the test process was ramped from lightly loaded through application-level saturation. In the case of the Sun Storage F5100 Flash Array, throughput to store was limited by system bottlenecks strongly influenced by lock contention and read service times of database table data files stored on spinning disk. Although the test process accurately measures storage service time at fixed workloads, due to system-level performance constraints, the test process did not measure the maximum throughput the Sun Storage F5100 Flash Array can support.

The test application implements new-order, order status, payment, and stock-level transactions to a 1 TB database. Oracle Database 10*g* hosts the database application data. In order to generate capacity planning information valid over a broad range of conditions, no specific application or database tuning was done to the system. Consequently, the throughput data represented in this paper represents a conservative estimate for system performance, and application-specific tuning could result in throughput increases.

Client-Side New Order Response Time Versus New Order Rate

Measuring changes in application service time as a function of application workload for systems based on hybrid flash/disk technology compared to traditional disk-only technology shows the kinds of gains realized at the application level for applications constrained by service time from the storage devices. In this example, the average transaction executed eight writes and 25 reads. The reads are further distributed with 50% executed against indexes and 50% executed against tables.

Figure 6 shows the results of a small OLTP test system. In all cases, new order service time was significantly lower for the system with index files stored on flash devices compared to the system with all data files stored on traditional disk. At a workload of 2500 new order transactions per minute (TPM), the hybrid disk/flash system delivers transaction service times of 0.2 sec compared to 0.4 sec in a disk-only solution. This 50% improvement in service time is a direct result of dropping I/O read service time for the indexes to 1 ms in the case of the Sun Storage F5100 Flash Array compared to 15 ms in the case of the disk-only solution. At a fixed new order service time of 0.4 sec, the maximum supportable workload of the hybrid disk/flash system improves 36% to 3400 TPM compared to 2500 TPM in the disk-only system.

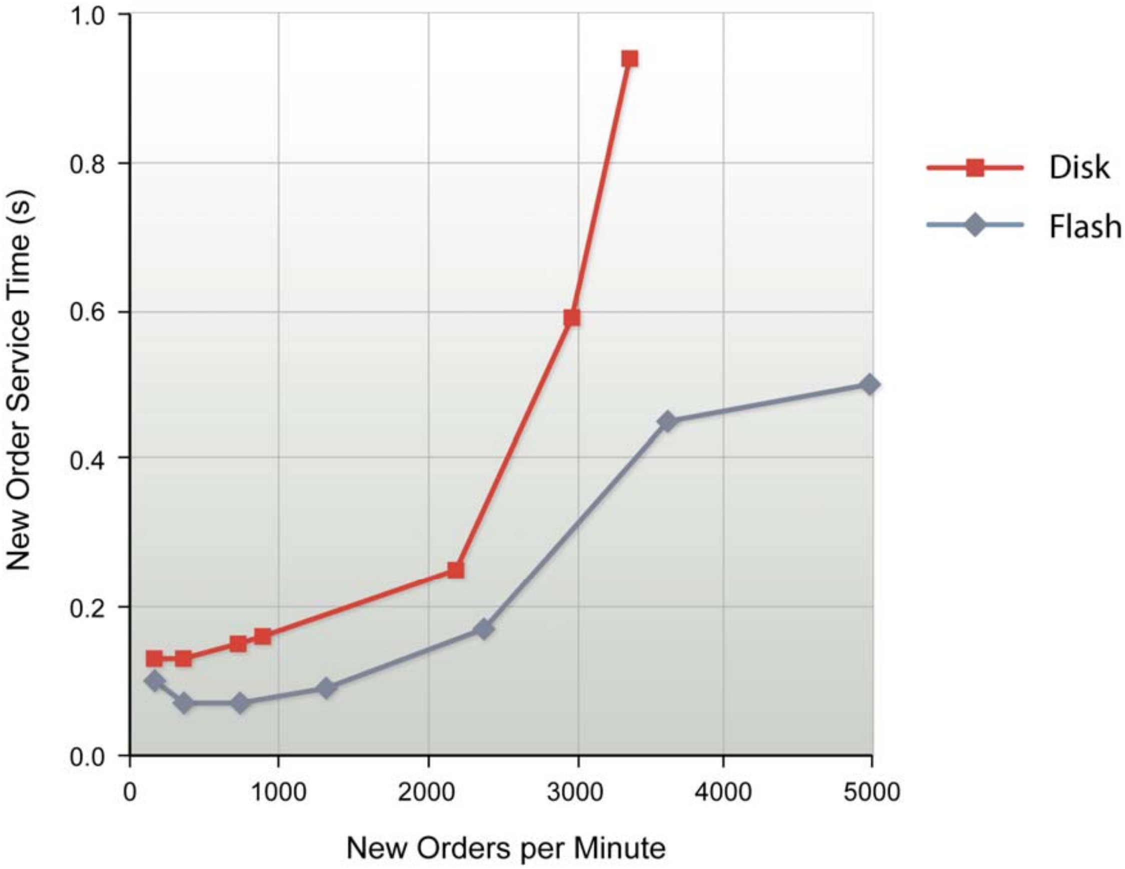


Figure 6. New order service times were reduced by 50% by using flash versus disk.

Oracle Tablespace I/O Read Service Time Versus I/O Rate

Read service time from media (database file sequential read) is the leading wait event influencing transaction response time for the OLTP test system. In the case of the hybrid flash/disk architecture, this average includes data from two different populations: reads from flash and reads from disk. In the test system used to generate this illustration, 40% of the read I/O came from the indexes, with 60% of the read I/O coming from the table data and the I/O service time for reads coming from the data files.

Adding a Sun Storage F5100 Flash Array to off-load index I/O processing from an existing disk system to flash-based storage improves system performance in two ways:

- Read service time from index files is reduced dramatically.
- Read service times for data files are reduced noticeably.

In the case of the index files, compared to a modestly loaded 15000 RPM disk drive, service time drops from 15 ms to 1 ms—an improvement of more than 90%. In the case of the table data files, because index processing has been moved to the flash device, there is less workload for the disk to support, so the spinning disk can get the remaining work done more quickly.

Figure 7 shows the Oracle-instance reported database file sequential read wait event verses the total front-end I/O rate. The front-end I/O rate is defined as the sum of the physical reads, physical writes, and transactions executed per second. The service time is defined as the average I/O service time over all tablespaces, including the data and indexes. In the case of the test application, where about 50% of the I/O is executed against the indexes and 50% of the I/O is executed against the data, the average service time is approximately the average of the service time to each tablespace. In the case of migrating from spinning disk to Sun Storage F5100 Flash Array technology, the nearly ten times reduction in service time to the index effectively halves the average service time for the system. In the lightly loaded case, average read service time drops from 6 ms to 3 ms, and as the disk begins to saturate, average read service time drops from 12 ms to 6 ms.

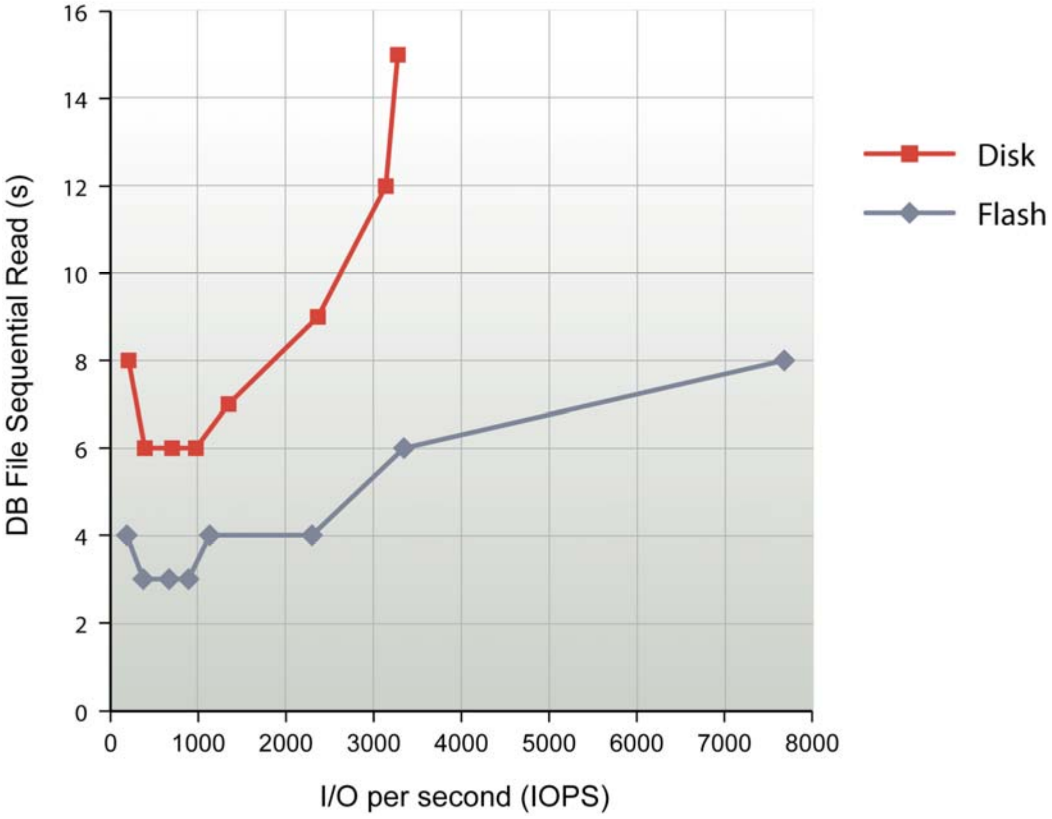


Figure 7. Small system results: IOPS for disk versus flash.

Scaling to Larger Systems

The test results presented so far represent a relatively small OLTP system based on Oracle Database. Further testing explored scalability to evaluate whether these improvements extend to larger, higher-capacity configurations.

Read Service Time Versus Workload per Sun Flash Module

A scalability test pushed the I/O rate to about 750 IOPS per Sun Flash Module with the I/O to the indexes tablespace. The results (Figures 8 and 9) show that for up to 750 IOPS per Sun Flash Module, the flash module consistently returns read I/O requests in 1–2 ms.

For comparison, the figures also include similar data taken for spinning disk. The comparison highlights the substantial improvement in response time and throughput as a result of including the Sun Storage F5100 Flash Array: an 85% reduction in I/O service time (2 ms versus 15 ms) at a 400% increase in throughput (750 IOPS versus 150 IOPS).

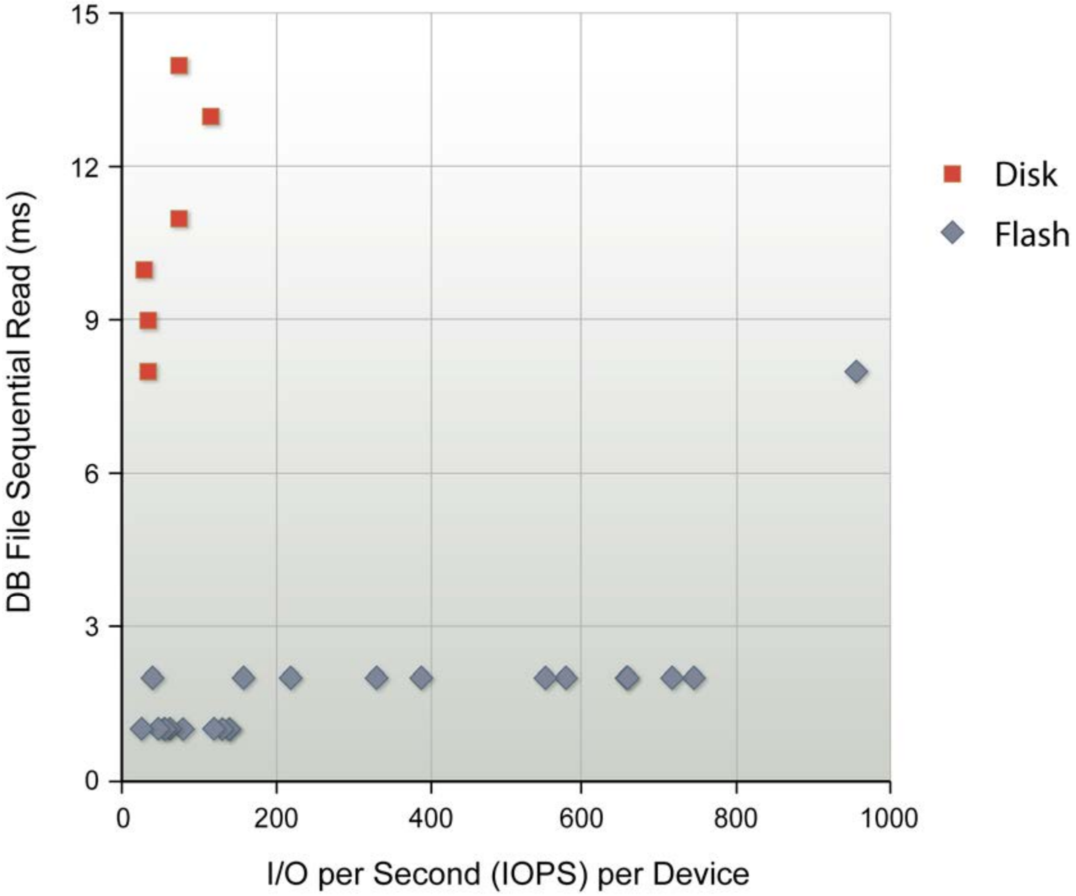


Figure 8. The Sun Storage F5100 Flash Array substantially improves response times.

Index Read Service Time Versus Storage Workload

An important measurement taken during the large-system test scalability study compared realized index read service time from the Sun Storage F5100 Flash Array and traditional spinning disk. Figure 9 shows index read service time as a function of the total front-end I/O rate (index + table) for an OLTP system running from 5,000 to 50,000 IOPS.

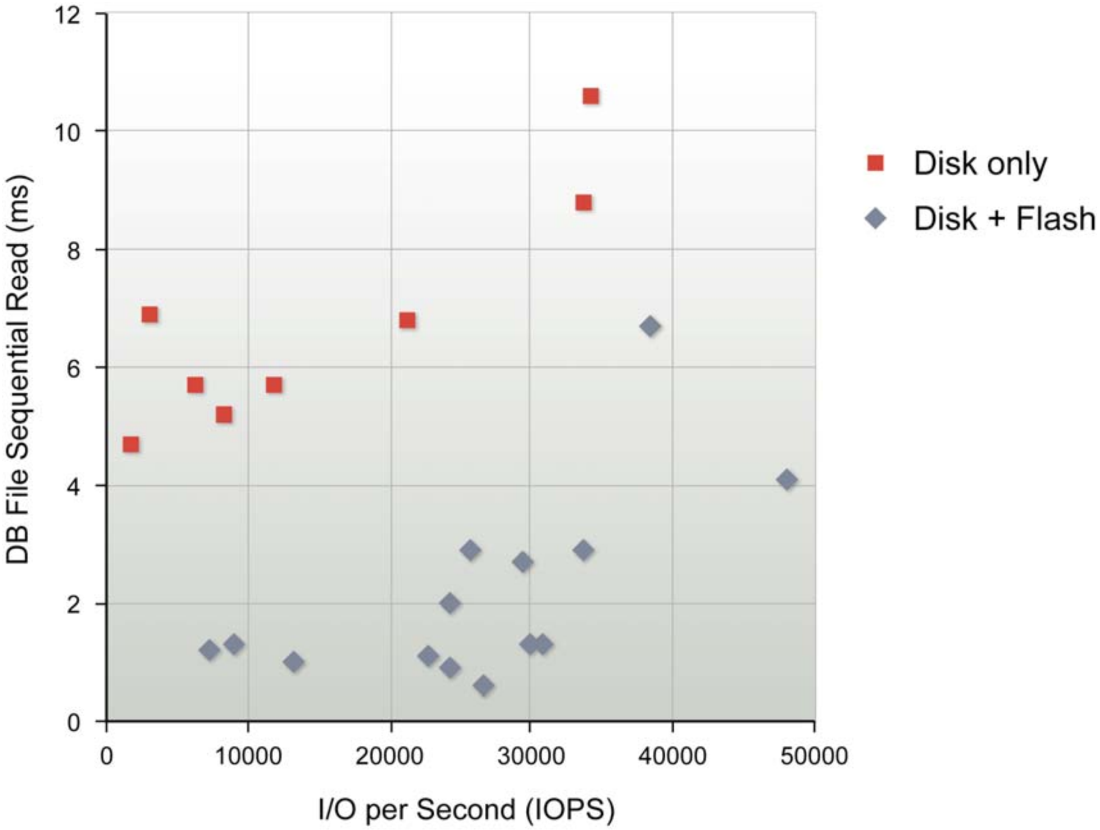


Figure 9.Combined disk and flash significantly improves index read service time over a broad range of I/O rates.

In this example, index read service time from the Sun Storage F5100 Flash Array measures 1 to 3 ms compared to 6–10 ms for the spinning disk drive. Performance gains of a 70% reduction in index I/O service time were achieved in this test case.

Summary of Performance Results When Using Flash for Database Indexes

The data collected in this study highlight five important points:

- At a fixed rate of 2500 new order transactions per minute (TPM), the new order service time dropped from 0.4 sec to 0.2 sec (a reduction of 50%) when indexes were moved to the Sun Storage F5100 Flash Array.
- At a fixed new order service time of 0.4 sec, maximum throughput increases from 2500 TPM to 3400 TPM (an increase of 36%).
- Average read service time for all I/O (data + indexes) dropped 50% when indexes were stored on flash.

- Read service times of 2 ms per Sun Flash Module can be realized at workloads up to 750 IOPS per Sun Flash Module, yielding an access density of over 30 IOPS/GB.
- When the workload was scaled out to 10,000 to 30,000 IOPS, I/O service times for indexes dropped 60–70% when the indexes were moved to the Sun Storage F5100 Flash Array compared to traditional SAN storage.

End-user response time savings for a specific implementation will depend on how much time the user's transaction spends waiting on I/O and how much time can be saved by storing some or all of the application data on the Sun Storage F5100 Flash Array.

Database Smart Flash Cache

Database Smart Flash Cache is available in Oracle Database 11*g* Release 2. It intelligently caches data from the Oracle Database, replacing slow mechanical I/O operations to disk with much faster flash-based storage operations. The Database Smart Flash Cache feature acts as a transparent extension of the database buffer cache using flash technology or solid-state drive (SSD). The flash acts as a Level 2 cache to the database buffer cache. If a process doesn't find the block it needs in the buffer cache, it has to perform a physical read from the second-level SGA buffer pool residing on flash. The read from flash will be quite fast, in the order of microseconds, when compared to performing a physical read operation from a traditional hard disk drive (HDD), which is completed in milliseconds.

Database Smart Flash Cache with flash storage gives administrators the ability to greatly improve the performance of Oracle databases by reducing the required amount of traditional disk I/O at a much lower cost than adding an equivalent amount of RAM. Software intelligence determines how and when to use the flash storage and how best to incorporate flash into the database as part of a coordinated data caching strategy to deliver much of the potential performance to applications.

The Database Smart Flash Cache technology allows frequently accessed data to be kept in very fast flash storage while most of the data is kept in very cost-effective disk storage. This happens automatically without the user having to take any action. Oracle Database Smart Flash is smart because it knows when to avoid trying to cache data that will never be reused or will not fit in the cache.

Random reads against tables and indexes are likely to have subsequent reads and normally will be cached and have their data delivered from the flash cache, if it is not found in the buffer cache. Scans, or sequentially reading tables, generally would not be cached since sequentially accessed data is unlikely to be subsequently followed by reads of the same data. Write operations are written through to the disk and staged back to cache if the software determines they are likely to be subsequently re-read. Knowing what not to cache is of great importance to realize the performance potential of the cache. For example, when writing redo or backups, or when writing to a mirrored copy of a block, the software avoids caching these blocks. Since these blocks will not be re-read in the near term, so there is no reason to devote valuable cache space to these objects or blocks.

In addition, Oracle Database allows the user to provide directives at the database table, index, and segment levels to ensure that Database Smart Flash Cache is used where desired. Tables can be moved in and out of flash with a simple command, without the need to move the table to different tablespaces, files, or LUNs as is often required for traditional storage with flash disks. Only Oracle Database has this visibility and understands the nature of all the I/O operations taking place on the system. Having the visibility through the complete I/O stack allows optimized use of Database Smart Flash Cache to store only the most frequently accessed data.

A white paper describing the results of testing with Oracle's Database Smart Flash Cache is listed in the “References” section.

Conclusion

Optimized database systems are critical to the success of business and government entities large and small. These systems must be highly available, and downtime must be minimized while cost-effectively optimizing performance and employee productivity. The union of Oracle and Sun has created one innovative company responsible for all the hardware and software components of a highly available, highly integrated, optimized database system. Oracle provides the technology to keep the database system running despite hardware and software errors, and provides a single source for ongoing services and support.

Flash technology also offers a cost-effective way to dramatically accelerate Oracle Database performance. Tests showed that flash technology resulted in a 50% reduction in service times and a 36% higher transaction rate for the OLTP test application.

References

For more information, visit the Web resources listed in Table 5.

TABLE 5. WEB RESOURCES FOR FURTHER INFORMATION

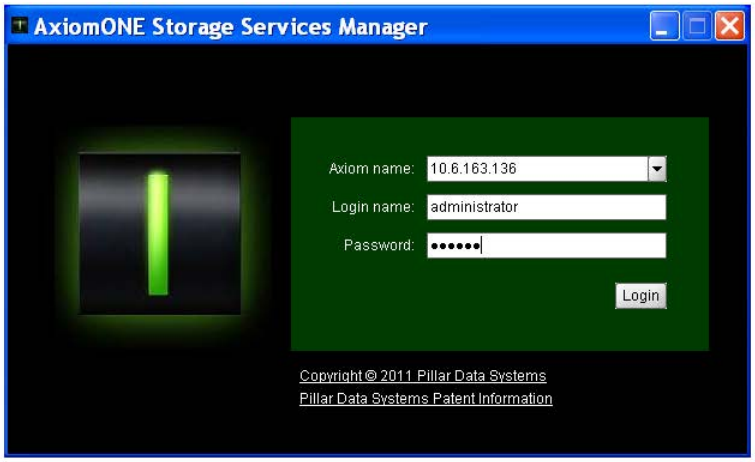
PRODUCT OR TECHNOLOGY	WEB RESOURCE URL
Sun Storage F5100 Flash Array	http://www.oracle.com/us/products/servers-storage/storage/disk-storage/043967.html
SPARC Enterprise M8000 Server	http://www.oracle.com/us/products/servers-storage/servers/sparc-enterprise/m-series/m8000/overview/index.html
SPARC Enterprise M9000 Server	http://www.oracle.com/us/products/servers-storage/servers/sparc-enterprise/m-series/m9000/overview/index.html
SPARC T4 Server	http://www.oracle.com/us/products/servers-storage/servers/sparc-enterprise/t-series/index.html
Oracle Solaris	http://www.oracle.com/solaris
Pillar Axiom 600 Storage Systems	http://www.pillardata.com
WHITE PAPERS	
Oracle Database Smart Flash Cache Results with SPARC Enterprise Midrange Servers	http://www.oracle.com/technetwork/articles/servers-storage-admin/smart-flash-cache-oracle-perf-361527.html

Appendix

This Appendix is provided as a reference to steps that were followed in the deployment of Oracle Real Application Clusters on Oracle's SPARC T4-2 servers, Pillar Axiom storage system, Oracle's InfiniBand technology, and Oracle Solaris 10. It is highly recommended that the entire Appendix be reviewed before attempting to utilize any component of this Appendix in attempts to deploy any of the associated software or hardware listed herein.

Pillar Axiom 600 Storage System —Steps to Create LUNs and Assign Them to Hosts

Log in to AxiomONE Storage Services Manager



Axiom name: 10.6.163.136

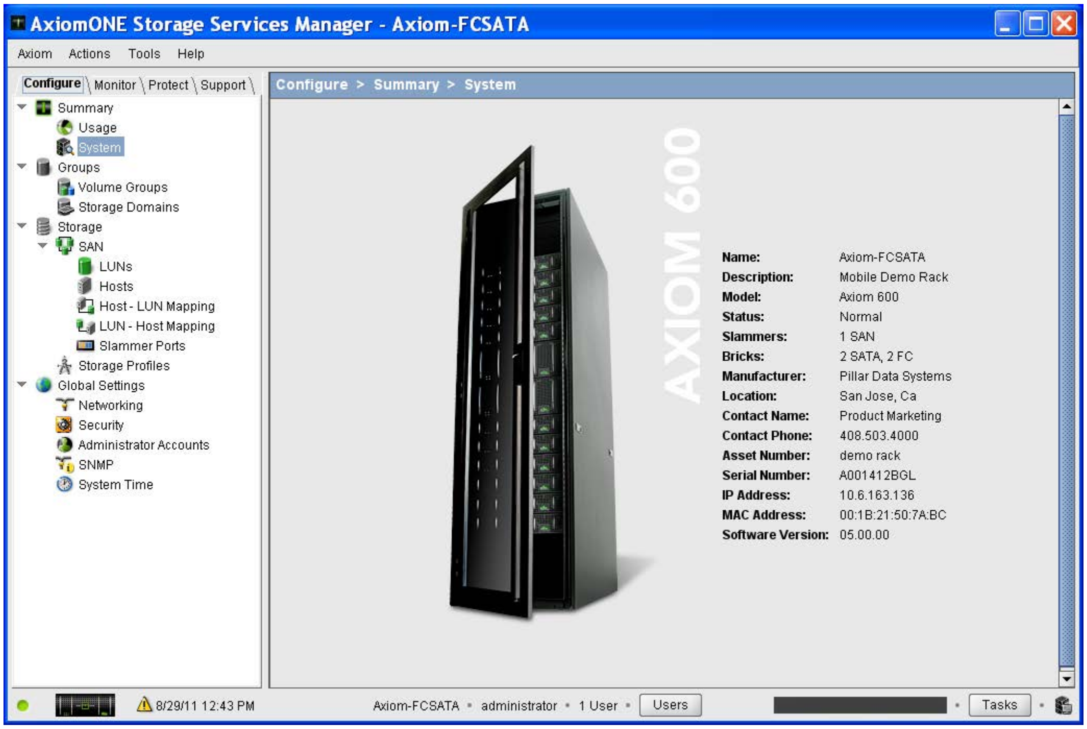
Login name: administrator

Password: ••••••

Copyright © 2011 Pillar Data Systems

Pillar Data Systems Patent Information

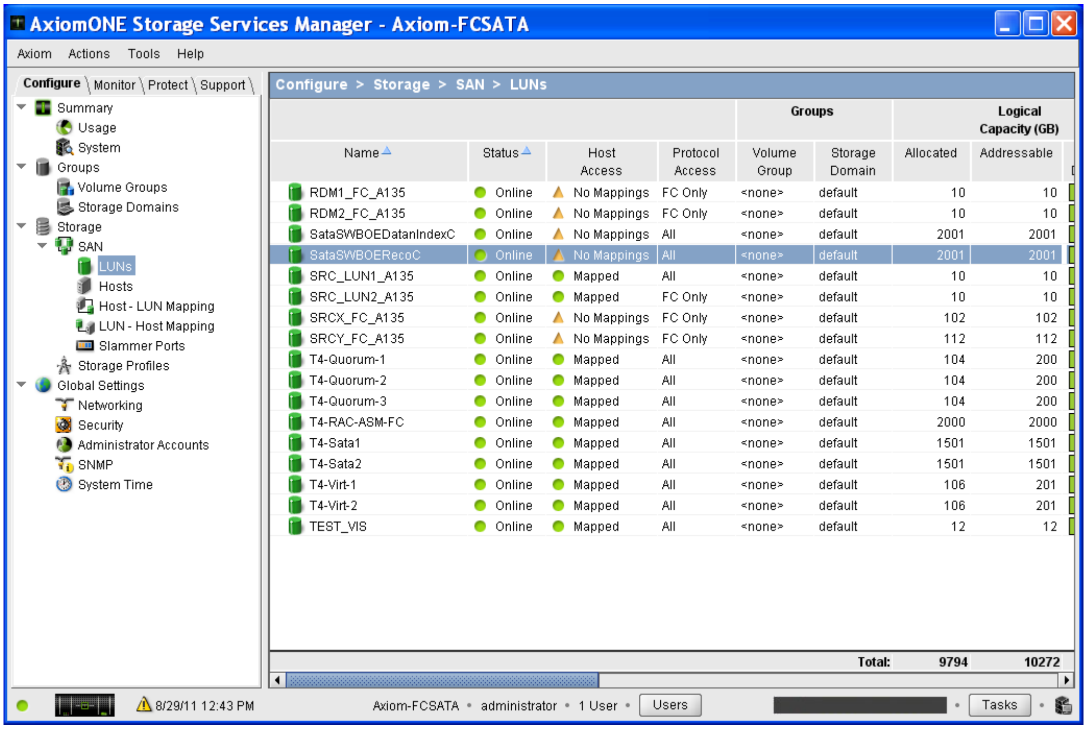
The Pillar Axiom 600 storage system summary is displayed.



Configure > Summary > System

Name:	Axiom-FCSATA
Description:	Mobile Demo Rack
Model:	Axiom 600
Status:	Normal
Slammers:	1 SAN
Bricks:	2 SATA, 2 FC
Manufacturer:	Pillar Data Systems
Location:	San Jose, Ca
Contact Name:	Product Marketing
Contact Phone:	408.503.4000
Asset Number:	demo rack
Serial Number:	A001412BGL
IP Address:	10.6.163.136
MAC Address:	00:1B:21:50:7A:BC
Software Version:	05.00.00

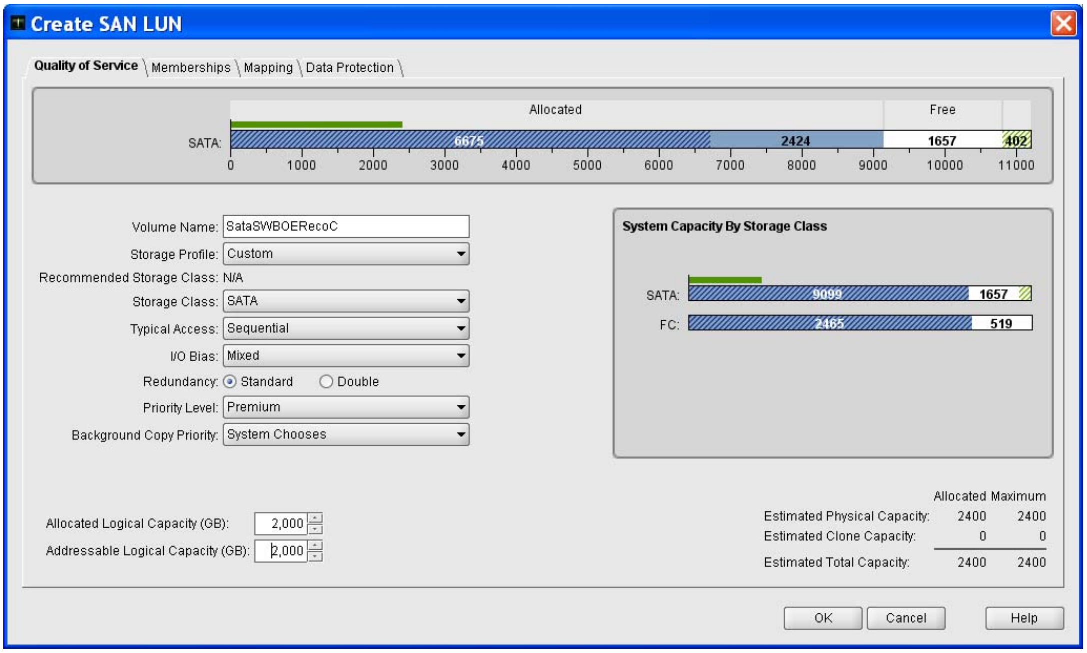
Select SAN LUNs. The existing Pillar Axiom 600 storage system SAN LUNs are displayed.



Configure > Storage > SAN > LUNs

				Groups		**Logical Capacity (GB)**	
Name	Status	Host Access	Protocol Access	Volume Group	Storage Domain	Allocated	Addressable
RDM1_FC_A135	● Online	⚠ No Mappings	FC Only	<none>	default	10	10
RDM2_FC_A135	● Online	⚠ No Mappings	FC Only	<none>	default	10	10
SataSWBOEDatanIndexC	● Online	⚠ No Mappings	All	<none>	default	2001	2001
SataSWBOERecoC	● Online	⚠ No Mappings	All	<none>	default	2001	2001
SRC_LUN1_A135	● Online	● Mapped	All	<none>	default	10	10
SRC_LUN2_A135	● Online	● Mapped	FC Only	<none>	default	10	10
SRCX_FC_A135	● Online	⚠ No Mappings	FC Only	<none>	default	102	102
SRCY_FC_A135	● Online	⚠ No Mappings	FC Only	<none>	default	112	112
T4-Quorum-1	● Online	● Mapped	All	<none>	default	104	200
T4-Quorum-2	● Online	● Mapped	All	<none>	default	104	200
T4-Quorum-3	● Online	● Mapped	All	<none>	default	104	200
T4-RAC-ASM-FC	● Online	● Mapped	All	<none>	default	2000	2000
T4-Sata1	● Online	● Mapped	All	<none>	default	1501	1501
T4-Sata2	● Online	● Mapped	All	<none>	default	1501	1501
T4-Virt-1	● Online	● Mapped	All	<none>	default	106	201
T4-Virt-2	● Online	● Mapped	All	<none>	default	106	201
TEST_VIS	● Online	● Mapped	All	<none>	default	12	12
					Total:	**9794**	**10272**

Select an action to create a Pillar Axiom 600 storage system SAN LUN and configure it.



Volume Name: SataSWBOERecoC

Storage Profile: Custom

Recommended Storage Class: N/A

Storage Class: SATA

Typical Access: Sequential

I/O Bias: Mixed

Redundancy: ● Standard ○ Double

Priority Level: Premium

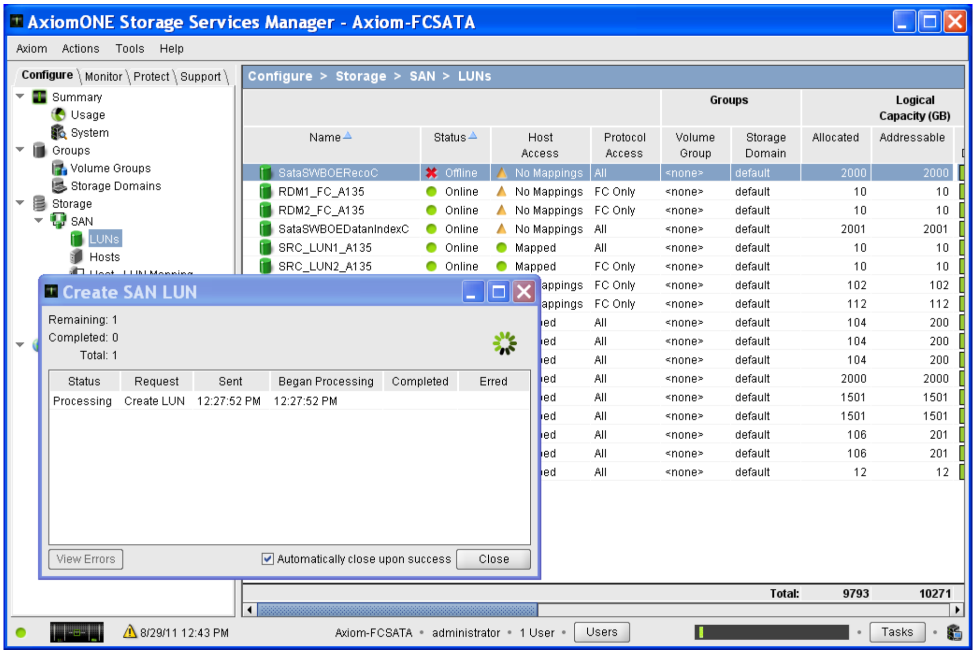
Background Copy Priority: System Chooses

Allocated Logical Capacity (GB): 2,000

Addressable Logical Capacity (GB): 2,000

	Allocated	Maximum
Estimated Physical Capacity:	2400	2400
Estimated Clone Capacity:	0	0
Estimated Total Capacity:	2400	2400

Select **OK** to finalize the creation of the Pillar Axiom 600 storage system SAN LUN.



Configure > Storage > SAN > LUNs

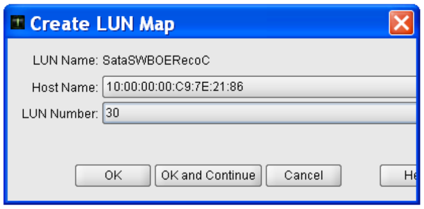
				Groups		Logical Capacity (GB)	
Name	Status	Host Access	Protocol Access	Volume Group	Storage Domain	Allocated	Addressable
SataSWBOERecoC	Offline	No Mappings	All	<none>	default	2000	2000
RDM1_FC_A135	Online	No Mappings	FC Only	<none>	default	10	10
RDM2_FC_A135	Online	No Mappings	FC Only	<none>	default	10	10
SataSWBOEDatanIndexC	Online	No Mappings	All	<none>	default	2001	2001
SRC_LUN1_A135	Online	Mapped	All	<none>	default	10	10
SRC_LUN2_A135	Online	Mapped	FC Only	<none>	default	10	10
		appings	FC Only	<none>	default	102	102
		appings	FC Only	<none>	default	112	112
		ed	All	<none>	default	104	200
		ed	All	<none>	default	104	200
		ed	All	<none>	default	104	200
		ed	All	<none>	default	2000	2000
		ed	All	<none>	default	1501	1501
		ed	All	<none>	default	1501	1501
		ed	All	<none>	default	106	201
		ed	All	<none>	default	106	201
		ed	All	<none>	default	12	12
					Total:	9793	10271

Remaining: 1
Completed: 0
Total: 1

Status	Request	Sent	Began Processing	Completed	Erred
Processing	Create LUN	12:27:52 PM	12:27:52 PM		

☑ Automatically close upon success

Create mapping for the Pillar Axiom 600 storage system SAN LUN for specific hosts through WWN.



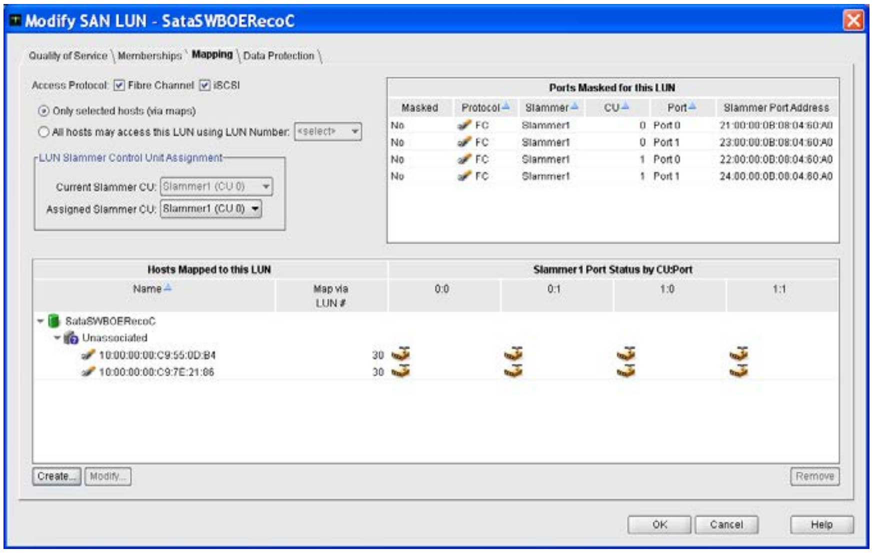
Create LUN Map

LUN Name: SataSWBOERecoC

Host Name: 10:00:00:00:C9:7E:21:86

LUN Number: 30

Once all hosts are defined and specific mapping is set up as desired, click **OK**.



Modify SAN LUN - SataSWBOERecoC

Access Protocol: ☑ Fibre Channel ☑ iSCSI

● Only selected hosts (via maps)

○ All hosts may access this LUN using LUN Number: <select>

LUN Slammer Control Unit Assignment

Current Slammer CU: Slammer1 (CU 0)

Assigned Slammer CU: Slammer1 (CU 0)

Ports Masked for this LUN

Masked	Protocol	Slammer	CU	Port	Slammer Port Address
No	FC	Slammer1	0	Port 0	21:00:00:0B:08:04:60:A0
No	FC	Slammer1	0	Port 1	23:00:00:0B:08:04:60:A0
No	FC	Slammer1	1	Port 0	22:00:00:0B:08:04:60:A0
No	FC	Slammer1	1	Port 1	24:00:00:0B:08:04:60:A0

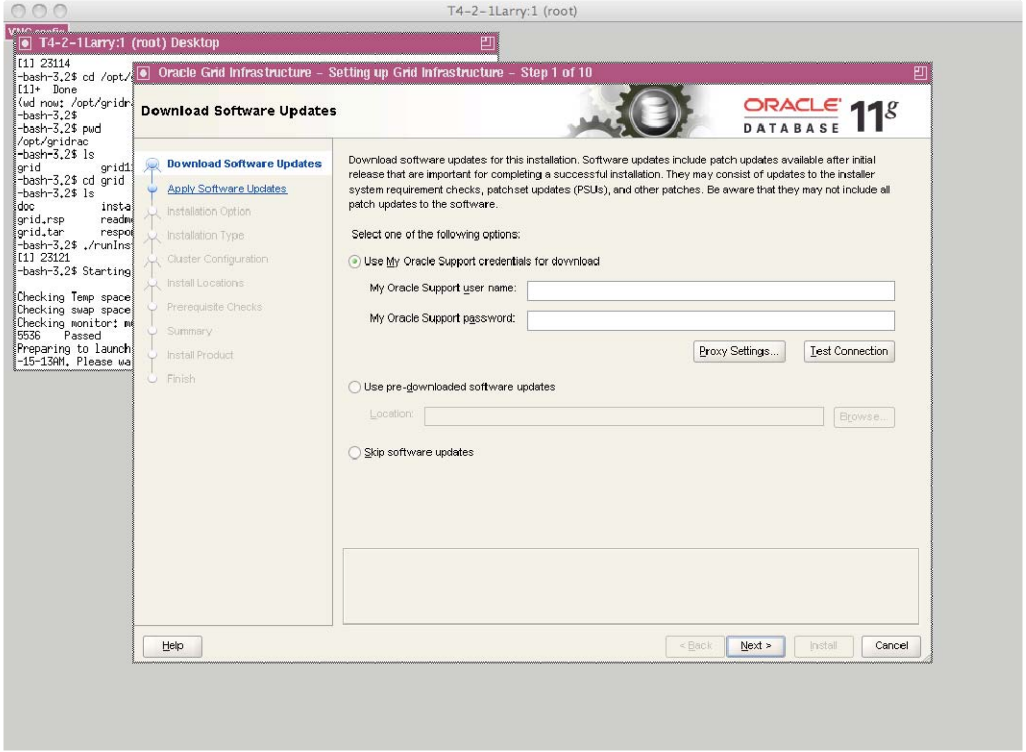
Hosts Mapped to this LUN		Slammer1 Port Status by CU:Port			
Name	Map via LUN #	0:0	0:1	1:0	1:1
SataSWBOERecoC					
Unassociated					
10:00:00:00:C9:55:0D:B4	30				
10:00:00:00:C9:7E:21:86	30				

Setting Up the Oracle Grid Infrastructure

The following screenshots show the step-by-step process for deploying the Oracle Database 11*g* Release 2 Grid Infrastructure used with two Oracle SPARC T4-2 servers in conjunction with this white paper.

These VNC-captured screenshots were gathered after an `su – oracle` command was issued followed by entering a `runInstaller` command to launch the GUI installer.

NOTE: The parameters entered here should be modified to fit organizational requirements. The screenshots are included to show how this specific deployment was configured.



```
[1] 23114
-bash-3.2$ cd /opt/
[1]+  Done
(wd now: /opt/gridr
-bash-3.2$
-bash-3.2$ pwd
/opt/gridrac
-bash-3.2$ ls
grid         grid1
-bash-3.2$ cd grid
-bash-3.2$ ls
doc          insta
grid.rsp     readm
grid.tar     respo
-bash-3.2$ ./runIns
[1] 23121
-bash-3.2$ Starting

Checking Temp space
Checking swap space
Checking monitor: m
5536    Passed
Preparing to launch
-15-13AM. Please wa
```

Download Software Updates

Download software updates for this installation. Software updates include patch updates available after initial release that are important for completing a successful installation. They may consist of updates to the installer system requirement checks, patchset updates (PSUs), and other patches. Be aware that they may not include all patch updates to the software.

Select one of the following options:

● Use My Oracle Support credentials for download

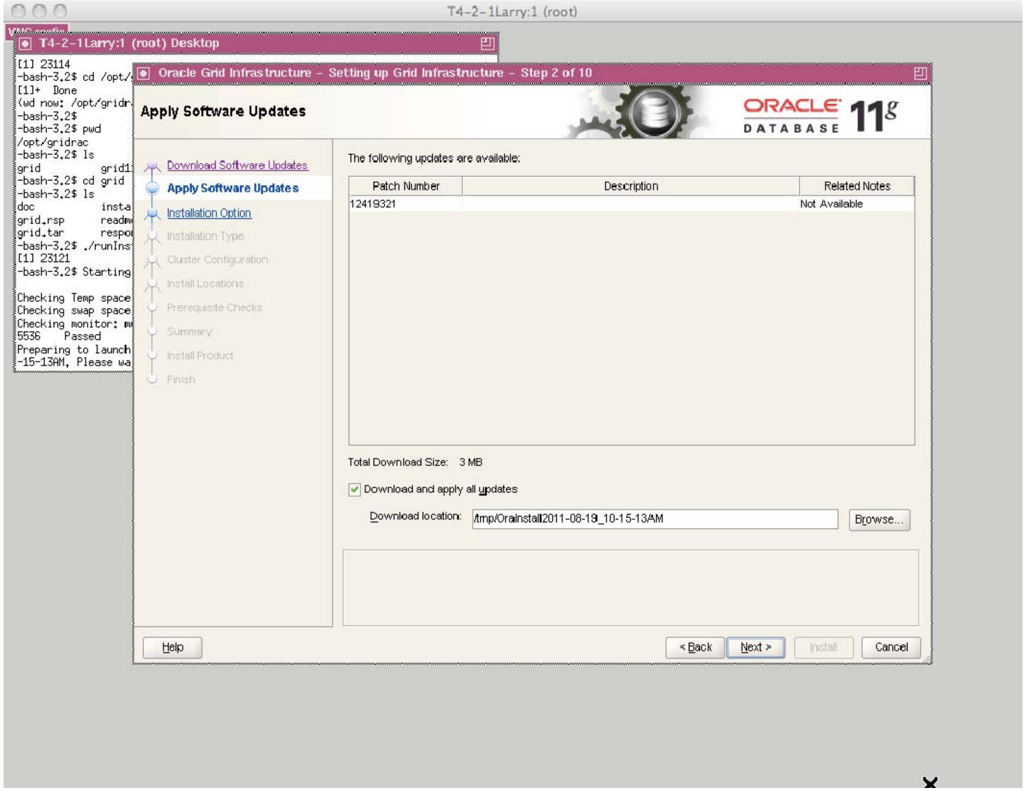
My Oracle Support user name:

My Oracle Support password:

○ Use pre-downloaded software updates

Location:

○ Skip software updates



```
[1] 23114
-bash-3.2$ cd /opt/
[1]+  Done
(wd now: /opt/gridr
-bash-3.2$
-bash-3.2$ pwd
/opt/gridrac
-bash-3.2$ ls
grid         grid1
-bash-3.2$ cd grid
-bash-3.2$ ls
doc          insta
grid.rsp     readm
grid.tar     respo
-bash-3.2$ ./runIns
[1] 23121
-bash-3.2$ Starting

Checking Temp space
Checking swap space
Checking monitor: m
5536    Passed
Preparing to launch
-15-13AM. Please wa
```

Apply Software Updates

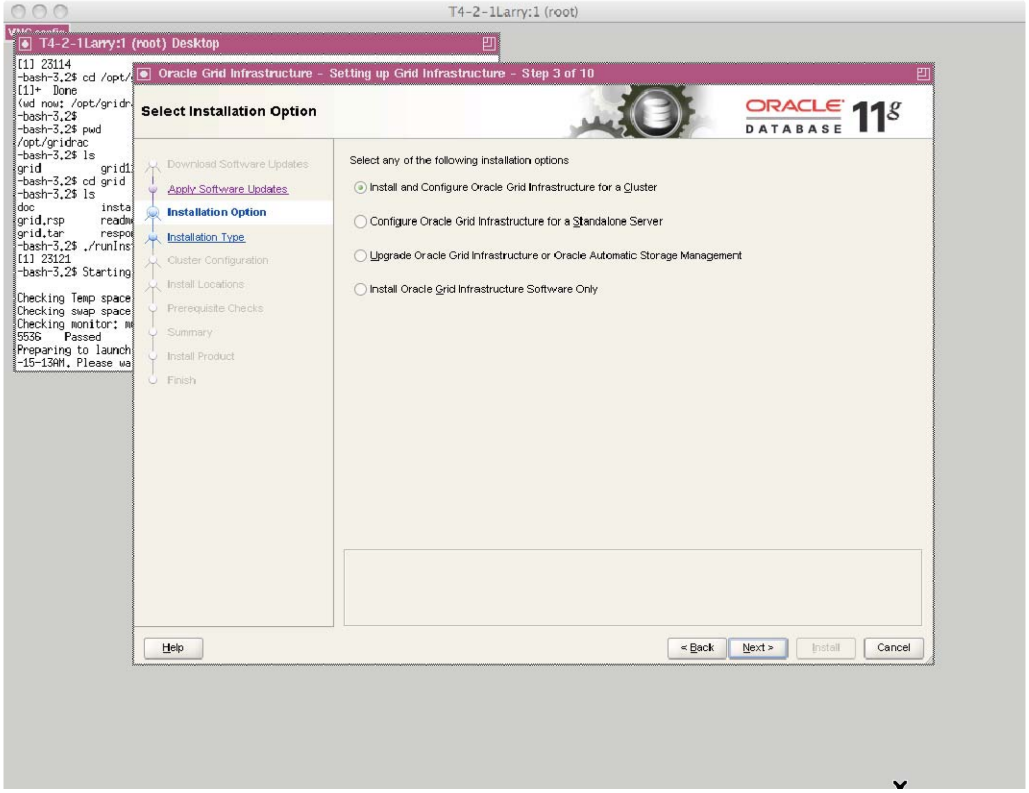
The following updates are available:

Patch Number	Description	Related Notes
12419321		Not Available

Total Download Size: 3 MB

☑ Download and apply all updates

Download location: /tmp/OraInstall2011-08-19_10-15-13AM



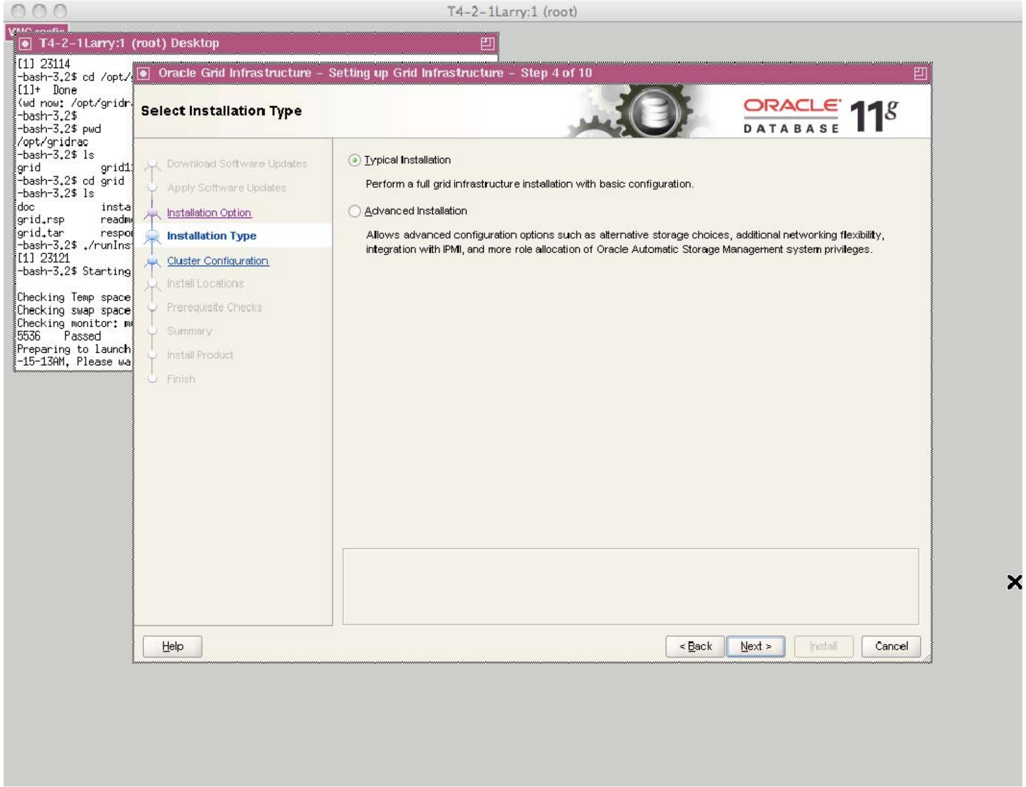
```
[1] 23114
-bash-3.2$ cd /opt/
[1]+  Done
(wd now: /opt/gridr
-bash-3.2$
-bash-3.2$ pwd
/opt/gridrac
-bash-3.2$ ls
grid           grid1
-bash-3.2$ cd grid
-bash-3.2$ ls
doc            insta
grid.rsp       readm
grid.tar       respo
-bash-3.2$ ./runIns
[1] 23121
-bash-3.2$ Starting

Checking Temp space
Checking swap space
Checking monitor: m
5536    Passed
Preparing to launch
-15-13AM. Please wa
```

Select Installation Option

Select any of the following installation options

- ● Install and Configure Oracle Grid Infrastructure for a Cluster
- ○ Configure Oracle Grid Infrastructure for a Standalone Server
- ○ Upgrade Oracle Grid Infrastructure or Oracle Automatic Storage Management
- ○ Install Oracle Grid Infrastructure Software Only



```
[1] 23114
-bash-3.2$ cd /opt/
[1]+  Done
(wd now: /opt/gridr
-bash-3.2$
-bash-3.2$ pwd
/opt/gridrac
-bash-3.2$ ls
grid           grid1
-bash-3.2$ cd grid
-bash-3.2$ ls
doc            insta
grid.rsp       readm
grid.tar       respo
-bash-3.2$ ./runIns
[1] 23121
-bash-3.2$ Starting

Checking Temp space
Checking swap space
Checking monitor: m
5536    Passed
Preparing to launch
-15-13AM. Please wa
```

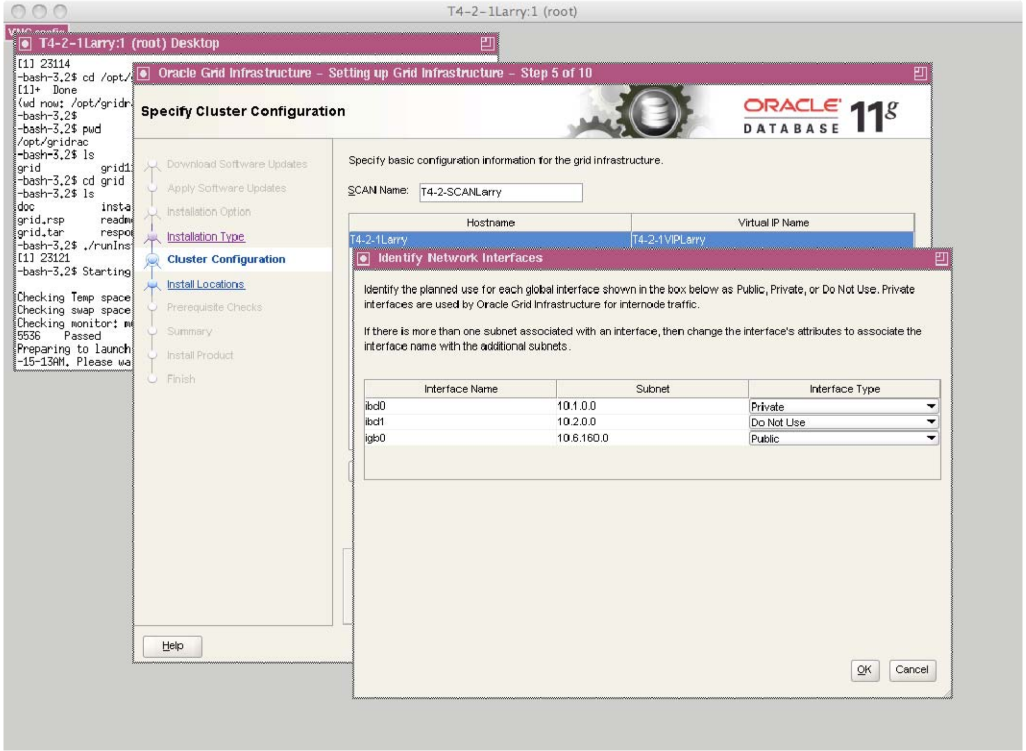
Select Installation Type

● Typical Installation

Perform a full grid infrastructure installation with basic configuration.

○ Advanced Installation

Allows advanced configuration options such as alternative storage choices, additional networking flexibility, integration with IPMI, and more role allocation of Oracle Automatic Storage Management system privileges.



```
[1] 23114
-bash-3.2$ cd /opt/
[1]+  Done
(wd now: /opt/gridr
-bash-3.2$
-bash-3.2$ pwd
/opt/gridrac
-bash-3.2$ ls
grid           grid1
-bash-3.2$ cd grid
-bash-3.2$ ls
doc            insta
grid.rsp       readm
grid.tar       respo
-bash-3.2$ ./runIns
[1] 23121
-bash-3.2$ Starting

Checking Temp space
Checking swap space
Checking monitor: m
5536     Passed
Preparing to launch
-15-13AM. Please wa
```

Specify Cluster Configuration

Specify basic configuration information for the grid infrastructure.

SCAN Name: T4-2-SCANLarry

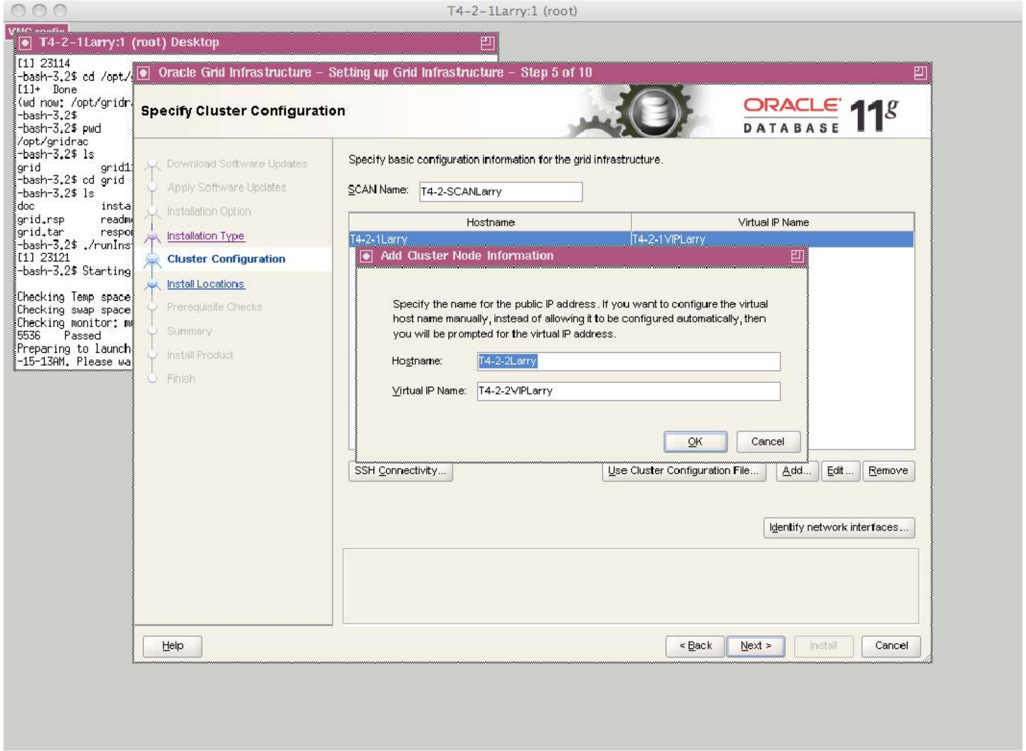
Hostname	Virtual IP Name
T4-2-1Larry	T4-2-1VIPLarry

Identify Network Interfaces

Identify the planned use for each global interface shown in the box below as Public, Private, or Do Not Use. Private interfaces are used by Oracle Grid Infrastructure for internode traffic.

If there is more than one subnet associated with an interface, then change the interface's attributes to associate the interface name with the additional subnets.

Interface Name	Subnet	Interface Type
ibd0	10.1.0.0	Private
ibd1	10.2.0.0	Do Not Use
igb0	10.6.160.0	Public



```
[1] 23114
-bash-3.2$ cd /opt/
[1]+  Done
(wd now: /opt/gridr
-bash-3.2$
-bash-3.2$ pwd
/opt/gridrac
-bash-3.2$ ls
grid           grid1
-bash-3.2$ cd grid
-bash-3.2$ ls
doc            insta
grid.rsp       readm
grid.tar       respo
-bash-3.2$ ./runIns
[1] 23121
-bash-3.2$ Starting

Checking Temp space
Checking swap space
Checking monitor: m
5536     Passed
Preparing to launch
-15-13AM. Please wa
```

Specify Cluster Configuration

Specify basic configuration information for the grid infrastructure.

SCAN Name: T4-2-SCANLarry

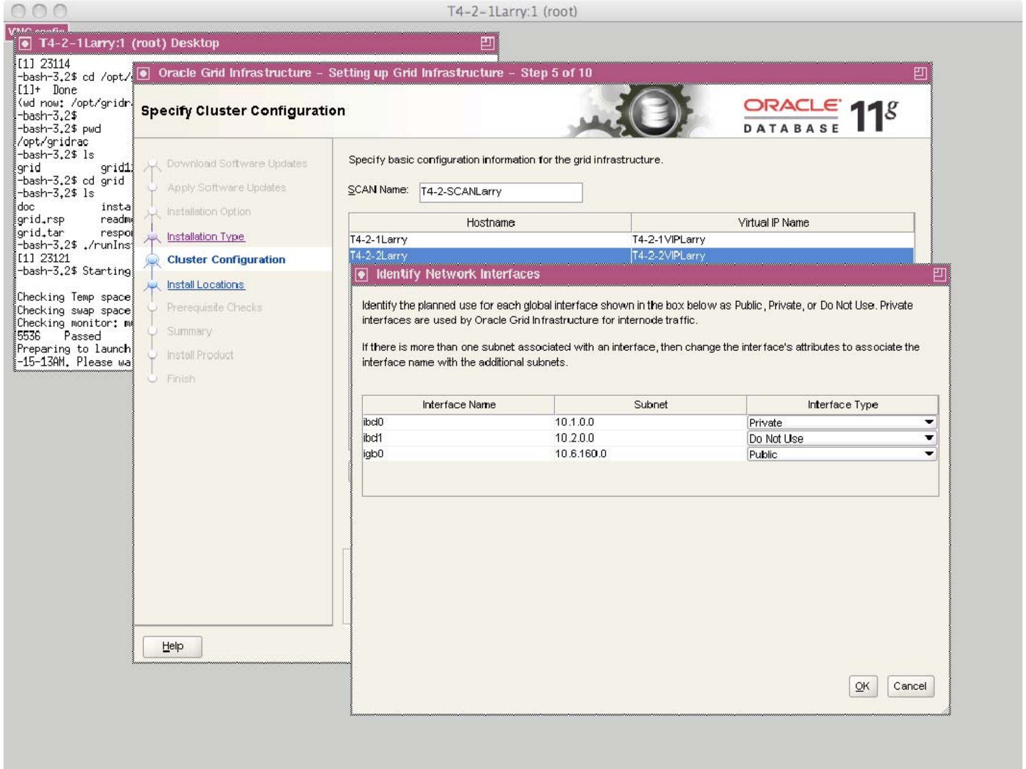
Hostname	Virtual IP Name
T4-2-1Larry	T4-2-1VIPLarry

Add Cluster Node Information

Specify the name for the public IP address. If you want to configure the virtual host name manually, instead of allowing it to be configured automatically, then you will be prompted for the virtual IP address.

Hostname: T4-2-2Larry

Virtual IP Name: T4-2-2VIPLarry



T4-2-1Larry:1 (root)

T4-2-1Larry:1 (root) Desktop

```
[1] 23114
-bash-3.2$ cd /opt/
[1]+  Done
(wd now: /opt/gridr
-bash-3.2$
-bash-3.2$ pwd
/opt/gridrac
-bash-3.2$ ls
grid          grid1
-bash-3.2$ cd grid
-bash-3.2$ ls
doc           insta
grid.rsp      readm
grid.tar      respo
-bash-3.2$ ./runIns
[1] 23121
-bash-3.2$ Starting

Checking Temp space
Checking swap space
Checking monitor: m
5536    Passed
Preparing to launch
-15-13AM. Please wa
```

Oracle Grid Infrastructure - Setting up Grid Infrastructure - Step 5 of 10

Specify Cluster Configuration

Specify basic configuration information for the grid infrastructure.

SCAN Name: T4-2-SCANLarry

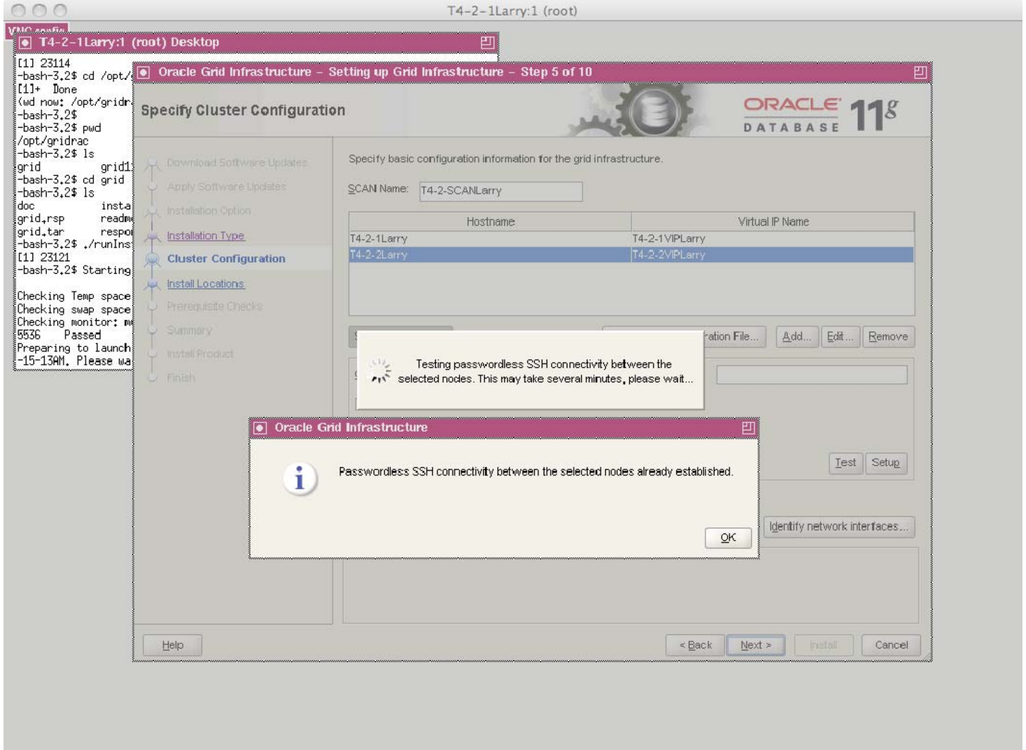
Hostname	Virtual IP Name
T4-2-1Larry	T4-2-1VIPLarry
T4-2-2Larry	T4-2-2VIPLarry

Identify Network Interfaces

Identify the planned use for each global interface shown in the box below as Public, Private, or Do Not Use. Private interfaces are used by Oracle Grid Infrastructure for internode traffic.

If there is more than one subnet associated with an interface, then change the interface's attributes to associate the interface name with the additional subnets.

Interface Name	Subnet	Interface Type
ibd0	10.1.0.0	Private
ibd1	10.2.0.0	Do Not Use
igb0	10.6.160.0	Public



T4-2-1Larry:1 (root)

T4-2-1Larry:1 (root) Desktop

```
[1] 23114
-bash-3.2$ cd /opt/
[1]+  Done
(wd now: /opt/gridr
-bash-3.2$
-bash-3.2$ pwd
/opt/gridrac
-bash-3.2$ ls
grid          grid1
-bash-3.2$ cd grid
-bash-3.2$ ls
doc           insta
grid.rsp      readm
grid.tar      respo
-bash-3.2$ ./runIns
[1] 23121
-bash-3.2$ Starting

Checking Temp space
Checking swap space
Checking monitor: m
5536    Passed
Preparing to launch
-15-13AM. Please wa
```

Oracle Grid Infrastructure - Setting up Grid Infrastructure - Step 5 of 10

Specify Cluster Configuration

Specify basic configuration information for the grid infrastructure.

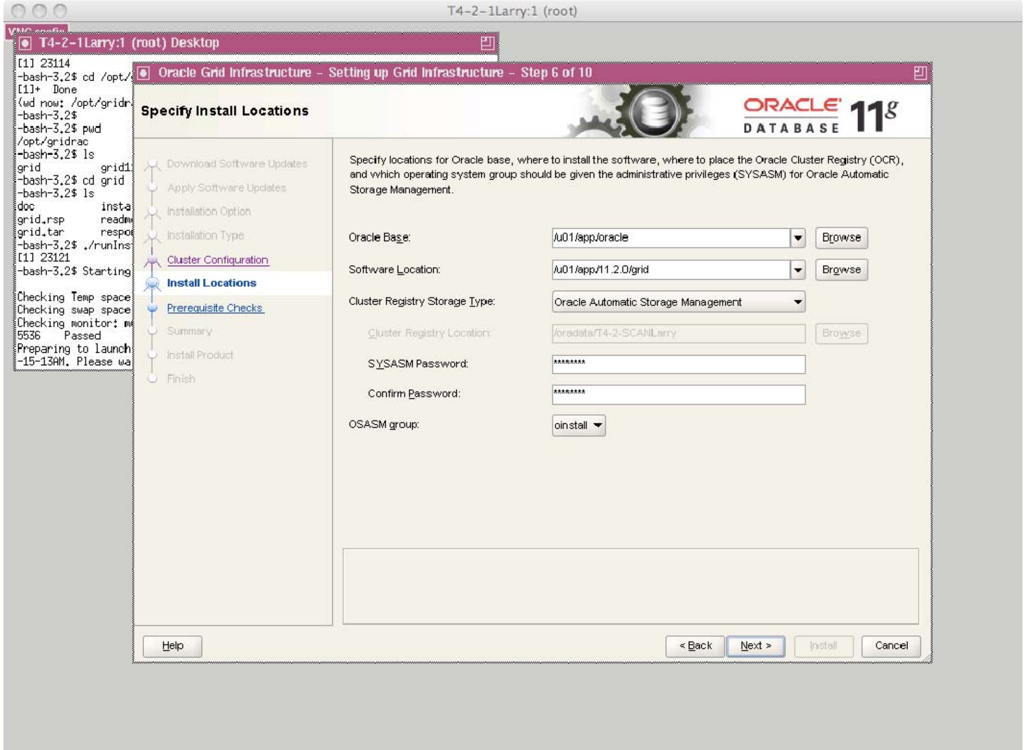
SCAN Name: T4-2-SCANLarry

Hostname	Virtual IP Name
T4-2-1Larry	T4-2-1VIPLarry
T4-2-2Larry	T4-2-2VIPLarry

Testing passwordless SSH connectivity between the selected nodes. This may take several minutes, please wait...

Oracle Grid Infrastructure

Passwordless SSH connectivity between the selected nodes already established.



```
[1] 23114
-bash-3.2$ cd /opt/
[1]+  Done
(wd now: /opt/gridr
-bash-3.2$
-bash-3.2$ pwd
/opt/gridrac
-bash-3.2$ ls
grid          grid1
-bash-3.2$ cd grid
-bash-3.2$ ls
doc           insta
grid.rsp      readm
grid.tar      respo
-bash-3.2$ ./runIns
[1] 23121
-bash-3.2$ Starting

Checking Temp space
Checking swap space
Checking monitor: m
5536    Passed
Preparing to launch
-15-13AM. Please wa
```

Oracle Grid Infrastructure - Setting up Grid Infrastructure - Step 6 of 10

Specify Install Locations

Specify locations for Oracle base, where to install the software, where to place the Oracle Cluster Registry (OCR), and which operating system group should be given the administrative privileges (SYSASM) for Oracle Automatic Storage Management.

Oracle Base: /u01/app/oracle

Software Location: /u01/app/11.2.0/grid

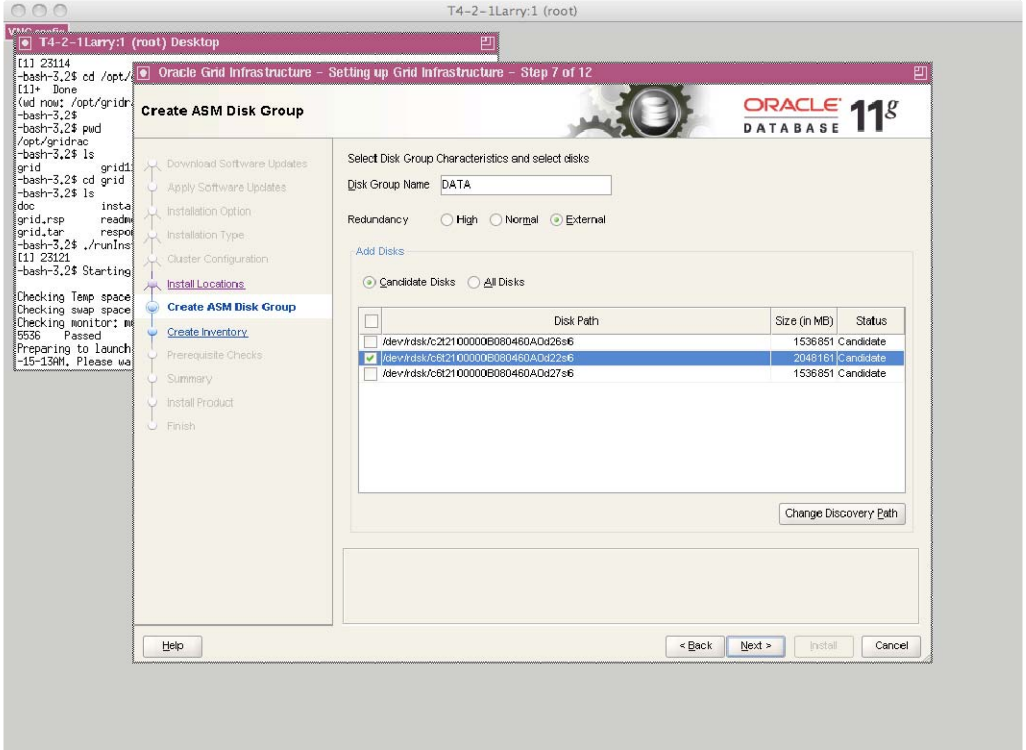
Cluster Registry Storage Type: Oracle Automatic Storage Management

Cluster Registry Location: /oradata/T4-2-SCANLarry

SYSASM Password: ********

Confirm Password: ********

OSASM group: oinstall



```
[1] 23114
-bash-3.2$ cd /opt/
[1]+  Done
(wd now: /opt/gridr
-bash-3.2$
-bash-3.2$ pwd
/opt/gridrac
-bash-3.2$ ls
grid          grid1
-bash-3.2$ cd grid
-bash-3.2$ ls
doc           insta
grid.rsp      readm
grid.tar      respo
-bash-3.2$ ./runIns
[1] 23121
-bash-3.2$ Starting

Checking Temp space
Checking swap space
Checking monitor: m
5536    Passed
Preparing to launch
-15-13AM. Please wa
```

Oracle Grid Infrastructure - Setting up Grid Infrastructure - Step 7 of 12

Create ASM Disk Group

Select Disk Group Characteristics and select disks

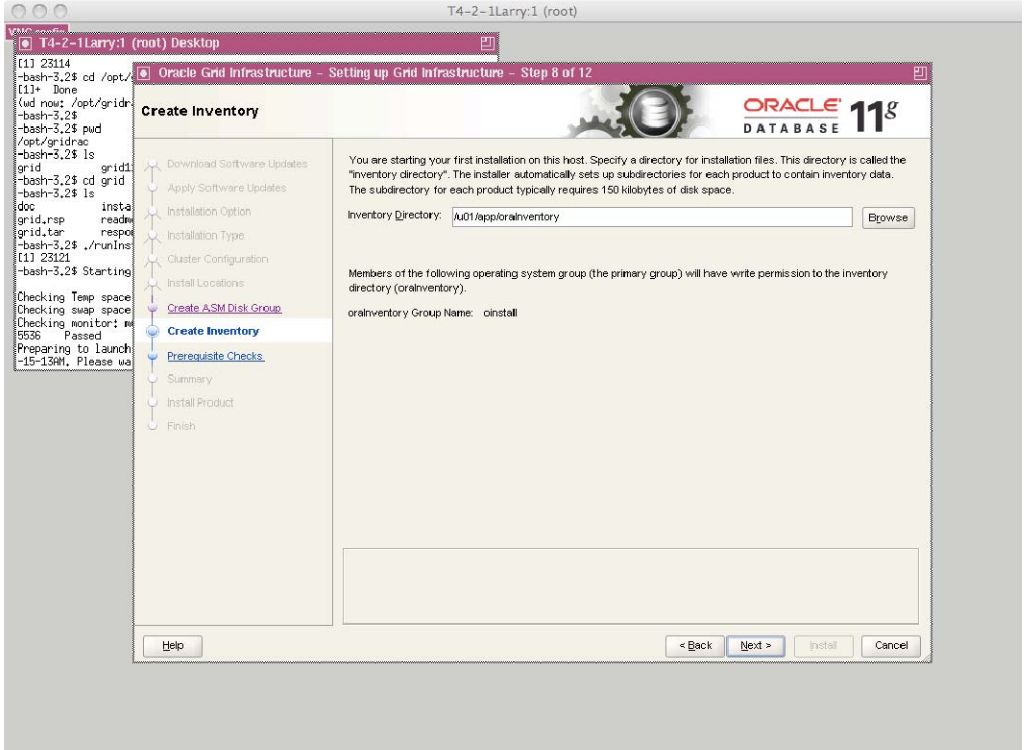
Disk Group Name: DATA

Redundancy: ○ High ○ Normal ● External

Add Disks

● Candidate Disks ○ All Disks

□	Disk Path	Size (in MB)	Status
□	/dev/rdsk/c2t2100000B080460A0d26s6	1536851	Candidate
☑	/dev/rdsk/c6t2100000B080460A0d22s6	2048161	Candidate
□	/dev/rdsk/c6t2100000B080460A0d27s6	1536851	Candidate



```
[1] 23114
-bash-3.2$ cd /opt/
[1]+  Done
(wd now: /opt/gridr
-bash-3.2$
-bash-3.2$ pwd
/opt/gridrac
-bash-3.2$ ls
grid          grid1
-bash-3.2$ cd grid
-bash-3.2$ ls
doc           insta
grid.rsp      readm
grid.tar      respo
-bash-3.2$ ./runIns
[1] 23121
-bash-3.2$ Starting

Checking Temp space
Checking swap space
Checking monitor: m
5536    Passed
Preparing to launch
-15-13AM. Please wa
```

Oracle Grid Infrastructure – Setting up Grid Infrastructure – Step 8 of 12

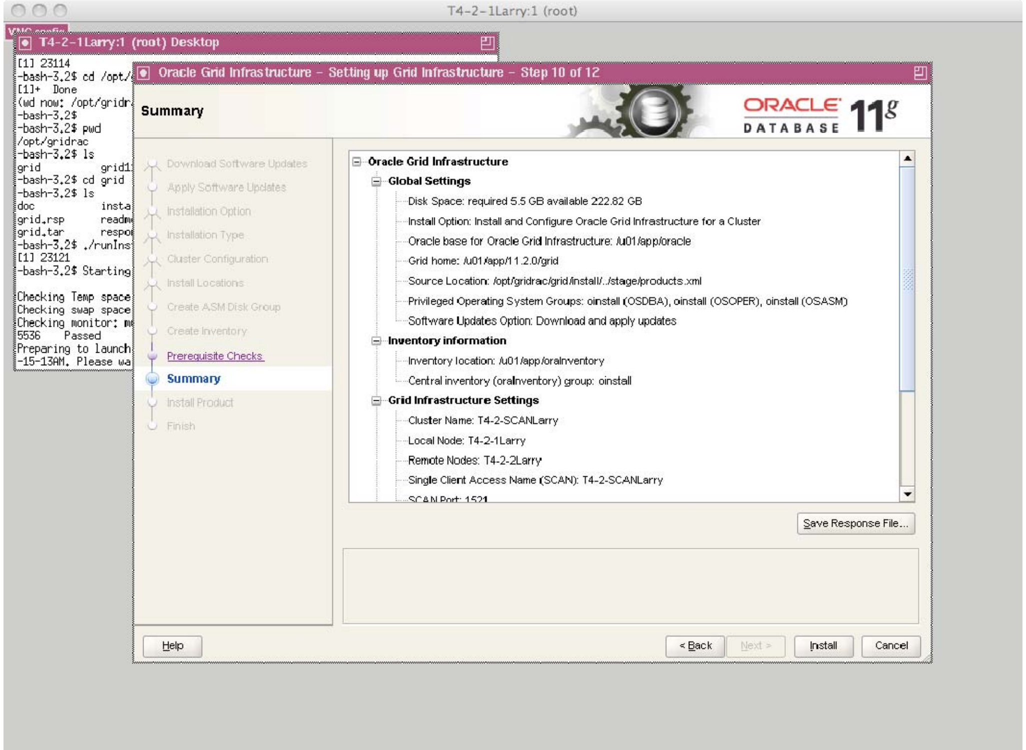
Create Inventory

You are starting your first installation on this host. Specify a directory for installation files. This directory is called the "inventory directory". The installer automatically sets up subdirectories for each product to contain inventory data. The subdirectory for each product typically requires 150 kilobytes of disk space.

Inventory Directory: /u01/app/oraInventory

Members of the following operating system group (the primary group) will have write permission to the inventory directory (oraInventory).

oraInventory Group Name: oinstall



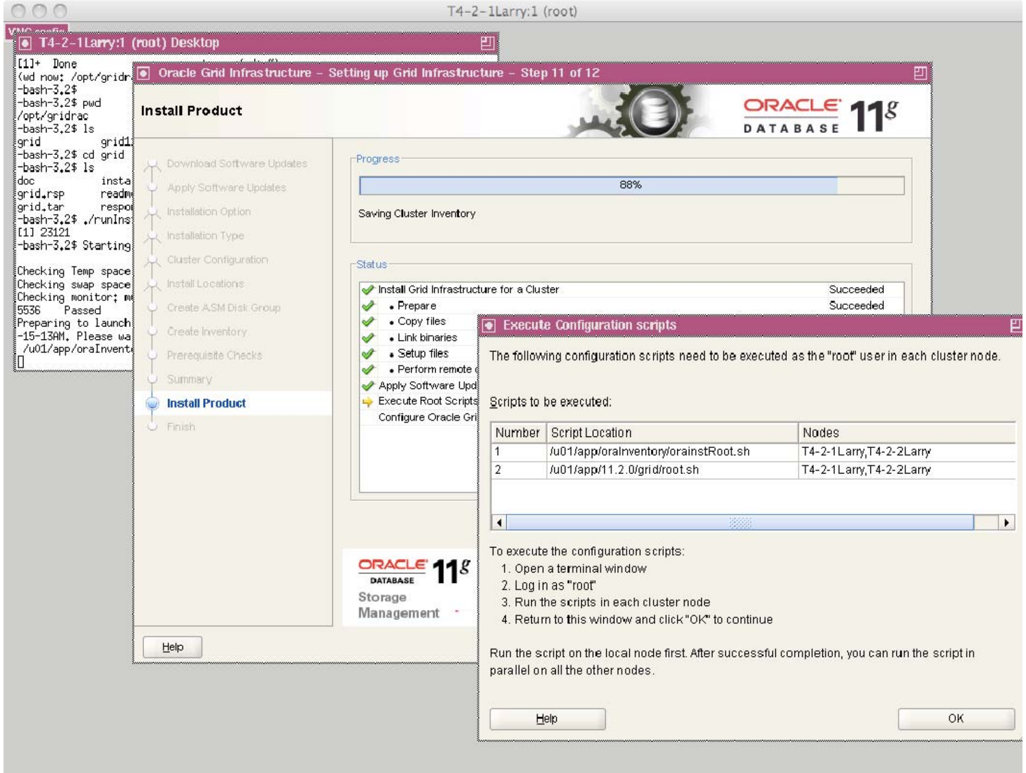
```
[1] 23114
-bash-3.2$ cd /opt/
[1]+  Done
(wd now: /opt/gridr
-bash-3.2$
-bash-3.2$ pwd
/opt/gridrac
-bash-3.2$ ls
grid          grid1
-bash-3.2$ cd grid
-bash-3.2$ ls
doc           insta
grid.rsp      readm
grid.tar      respo
-bash-3.2$ ./runIns
[1] 23121
-bash-3.2$ Starting

Checking Temp space
Checking swap space
Checking monitor: m
5536    Passed
Preparing to launch
-15-13AM. Please wa
```

Oracle Grid Infrastructure – Setting up Grid Infrastructure – Step 10 of 12

Summary

- **Oracle Grid Infrastructure**
  - **Global Settings**
    - Disk Space: required 5.5 GB available 222.82 GB
    - Install Option: Install and Configure Oracle Grid Infrastructure for a Cluster
    - Oracle base for Oracle Grid Infrastructure: /u01/app/oracle
    - Grid home: /u01/app/11.2.0/grid
    - Source Location: /opt/gridrac/grid/install/../stage/products.xml
    - Privileged Operating System Groups: oinstall (OSDBA), oinstall (OSOPER), oinstall (OSASM)
    - Software Updates Option: Download and apply updates
  - **Inventory information**
    - Inventory location: /u01/app/oraInventory
    - Central inventory (oraInventory) group: oinstall
  - **Grid Infrastructure Settings**
    - Cluster Name: T4-2-SCANLarry
    - Local Node: T4-2-1Larry
    - Remote Nodes: T4-2-2Larry
    - Single Client Access Name (SCAN): T4-2-SCANLarry
    - SCAN Port: 1521



```
[1]+  Done
(wd now: /opt/gridr
-bash-3.2$
-bash-3.2$ pwd
/opt/gridrac
-bash-3.2$ ls
grid          grid1
-bash-3.2$ cd grid
-bash-3.2$ ls
doc           insta
grid.rsp      readm
grid.tar      respo
-bash-3.2$ ./runIns
[1] 23121
-bash-3.2$ Starting

Checking Temp space
Checking swap space
Checking monitor; m
5536    Passed
Preparing to launch
-15-13AM. Please wa
 /u01/app/oraInvent
```

Saving Cluster Inventory

Install Grid Infrastructure for a Cluster — Succeeded
- Prepare — Succeeded
- Copy files
- Link binaries
- Setup files
- Perform remote
Apply Software Upd
Execute Root Scripts
Configure Oracle Gri

The following configuration scripts need to be executed as the "root" user in each cluster node.

Scripts to be executed:

Number	Script Location	Nodes
1	/u01/app/oraInventory/orainstRoot.sh	T4-2-1Larry,T4-2-2Larry
2	/u01/app/11.2.0/grid/root.sh	T4-2-1Larry,T4-2-2Larry

To execute the configuration scripts:
1. Open a terminal window
2. Log in as "root"
3. Run the scripts in each cluster node
4. Return to this window and click "OK" to continue

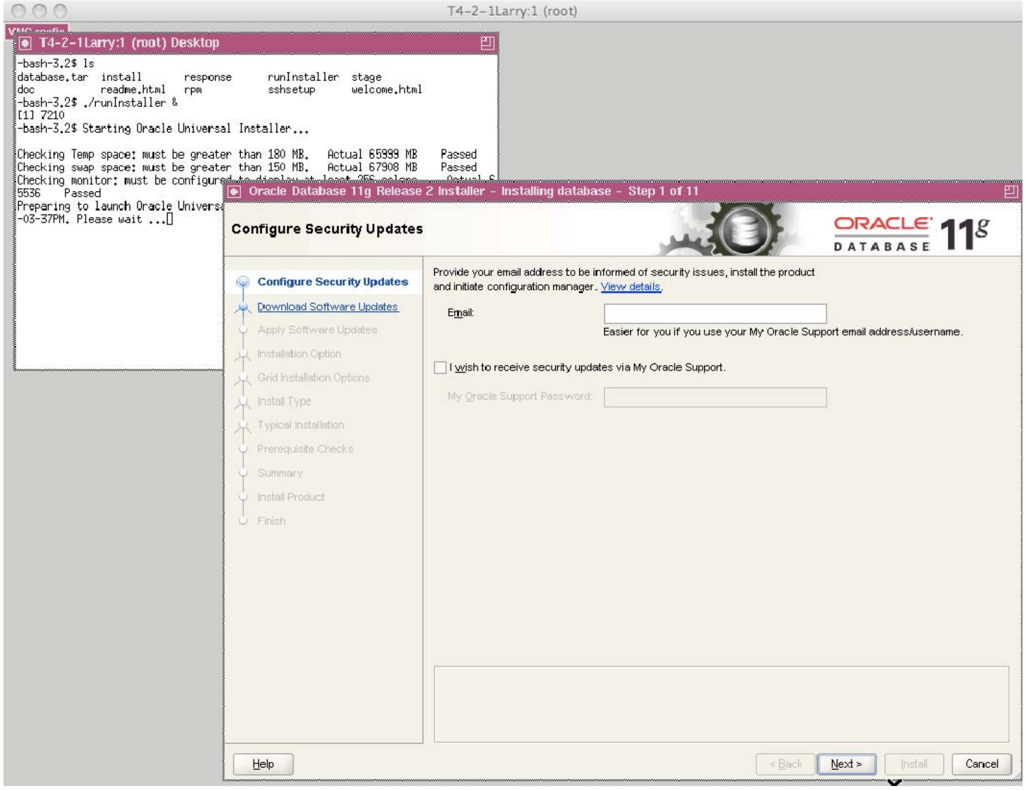
Run the script on the local node first. After successful completion, you can run the script in parallel on all the other nodes.

Deploying Oracle Database 11*g* Release 2

The following screenshots show the step-by-step process for deploying Oracle Database 11*g* Release 2 on top of the previously deployed Grid Infrastructure used in with this white paper.

These VNC-captured screenshots were gathered after an `su – oracle` command was issued followed by entering a `runInstaller` command to launch the GUI installer.

NOTE: The parameters entered here should be modified to fit organizational requirements. The screenshots are included to show how the specific deployment was configured.



```
-bash-3.2$ ls
database.tar  install       response      runInstaller  stage
doc           readme.html   rpm           sshsetup      welcome.html
-bash-3.2$ ./runInstaller &
[1] 7210
-bash-3.2$ Starting Oracle Universal Installer...

Checking Temp space: must be greater than 180 MB.   Actual 65999 MB    Passed
Checking swap space: must be greater than 150 MB.   Actual 67908 MB    Passed
Checking monitor: must be configured
5536    Passed
Preparing to launch Oracle Univers
-03-37PM. Please wait ...
```

Configure Security Updates

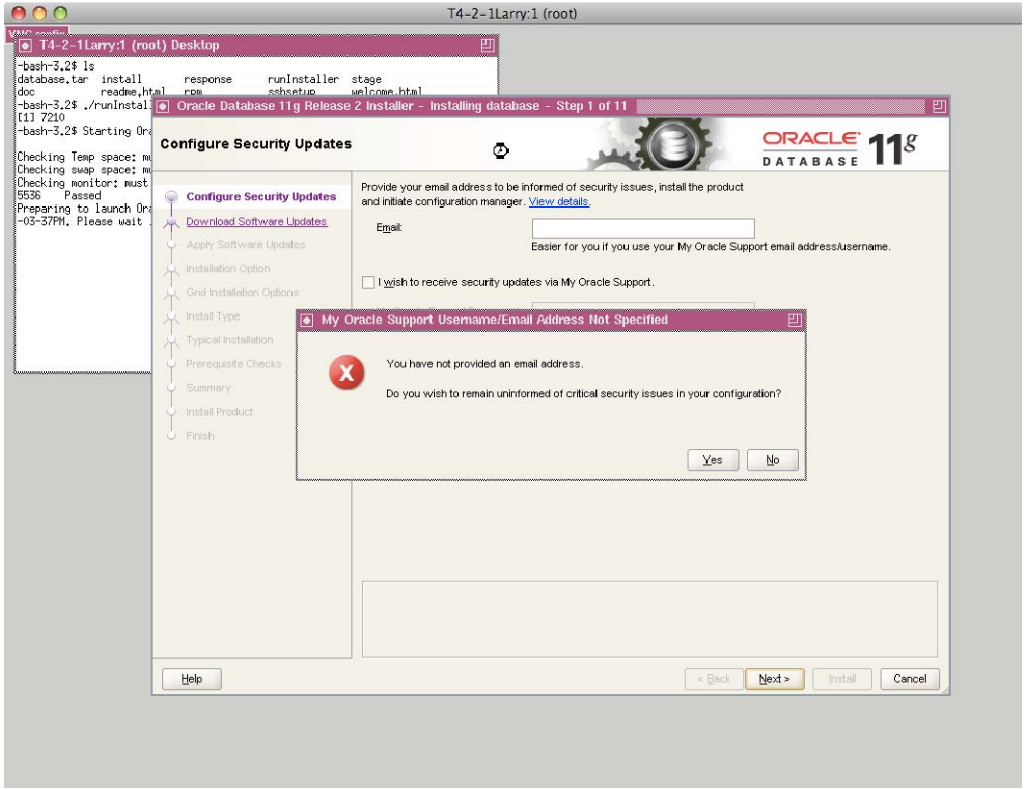
Provide your email address to be informed of security issues, install the product and initiate configuration manager. View details.

Email:

Easier for you if you use your My Oracle Support email address/username.

□ I wish to receive security updates via My Oracle Support.

My Oracle Support Password:



```
-bash-3.2$ ls
database.tar  install       response      runInstaller  stage
doc           readme.html   rpm           sshsetup      welcome.html
-bash-3.2$ ./runInstal
[1] 7210
-bash-3.2$ Starting Or

Checking Temp space: mu
Checking swap space: mu
Checking monitor: must
5536    Passed
Preparing to launch Or
-03-37PM. Please wait
```

Configure Security Updates

Provide your email address to be informed of security issues, install the product and initiate configuration manager. View details.

Email:

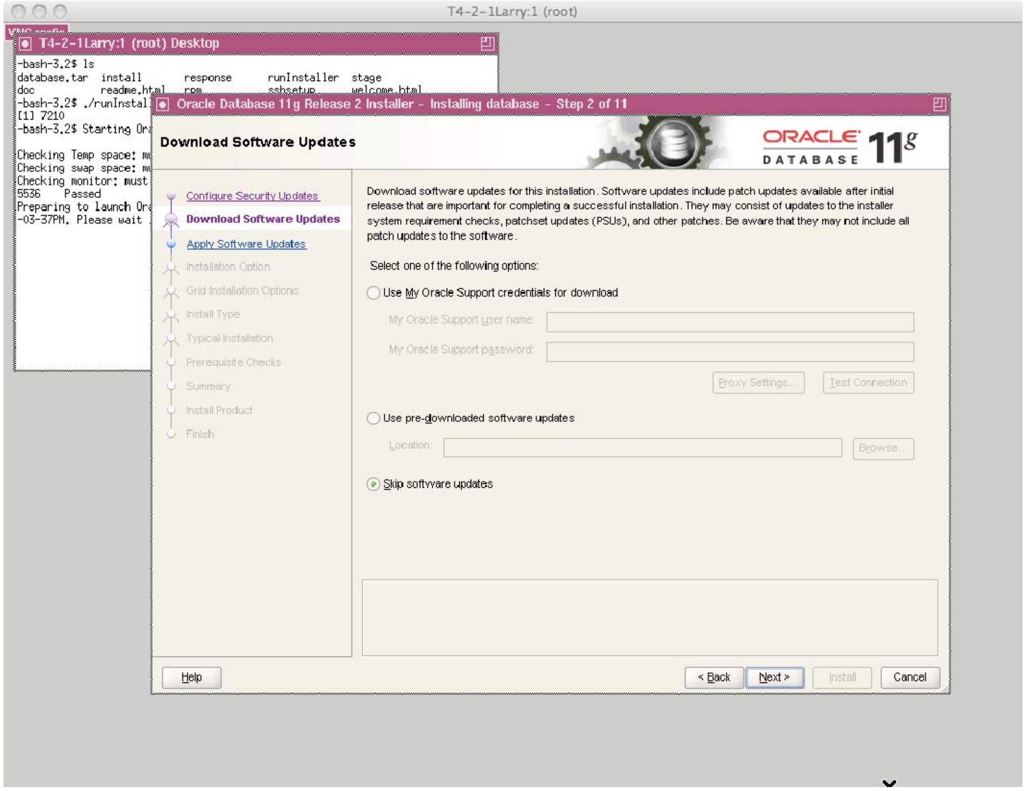
Easier for you if you use your My Oracle Support email address/username.

□ I wish to receive security updates via My Oracle Support.

My Oracle Support Username/Email Address Not Specified

You have not provided an email address.

Do you wish to remain uninformed of critical security issues in your configuration?



```
-bash-3.2$ ls
database.tar  install      response     runInstaller  stage
doc           readme.html  rpm          sshsetup      welcome.html
-bash-3.2$ ./runInstal
[1] 7210
-bash-3.2$ Starting Ora

Checking Temp space: mu
Checking swap space: mu
Checking monitor: must
5536    Passed
Preparing to launch Ora
-03-37PM. Please wait
```

Oracle Database 11g Release 2 Installer - Installing database - Step 2 of 11

Download Software Updates

Download software updates for this installation. Software updates include patch updates available after initial release that are important for completing a successful installation. They may consist of updates to the installer system requirement checks, patchset updates (PSUs), and other patches. Be aware that they may not include all patch updates to the software.

Select one of the following options:

○ Use My Oracle Support credentials for download

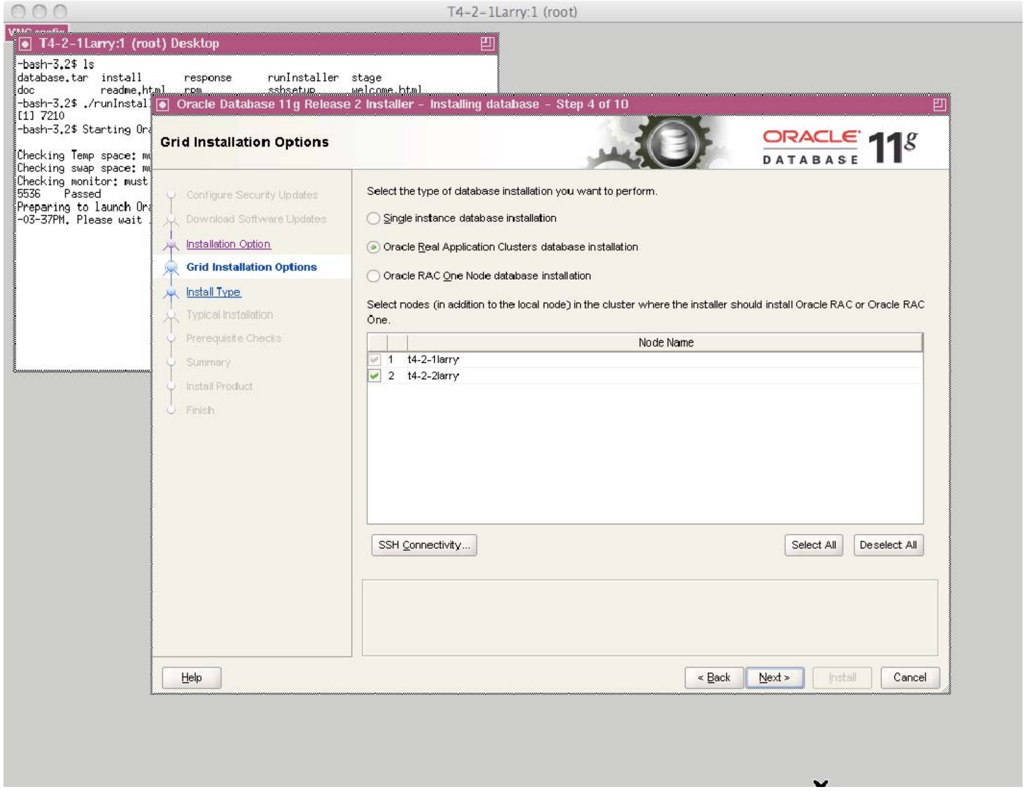
My Oracle Support user name:

My Oracle Support password:

○ Use pre-downloaded software updates

Location:

● Skip software updates



```
-bash-3.2$ ls
database.tar  install      response     runInstaller  stage
doc           readme.html  rpm          sshsetup      welcome.html
-bash-3.2$ ./runInstal
[1] 7210
-bash-3.2$ Starting Ora

Checking Temp space: mu
Checking swap space: mu
Checking monitor: must
5536    Passed
Preparing to launch Ora
-03-37PM. Please wait
```

Oracle Database 11g Release 2 Installer - Installing database - Step 4 of 10

Grid Installation Options

Select the type of database installation you want to perform.

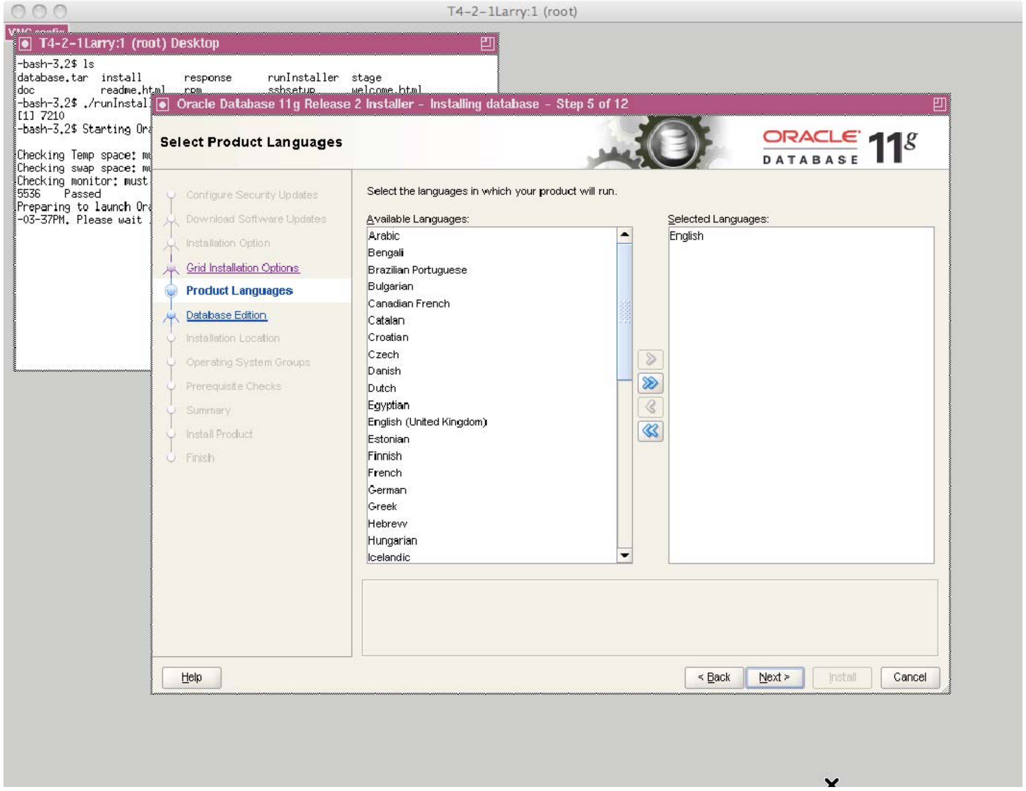
○ Single instance database installation

● Oracle Real Application Clusters database installation

○ Oracle RAC One Node database installation

Select nodes (in addition to the local node) in the cluster where the installer should install Oracle RAC or Oracle RAC One.

		Node Name
☑	1	t4-2-1larry
☑	2	t4-2-2larry



```
-bash-3.2$ ls
database.tar  install      response     runInstaller  stage
doc           readme.html  rpm          sshsetup      welcome.html
-bash-3.2$ ./runInstal
[1] 7210
-bash-3.2$ Starting Ora

Checking Temp space: mu
Checking swap space: mu
Checking monitor: must
5536    Passed
Preparing to launch Ora
-03-37PM. Please wait
```

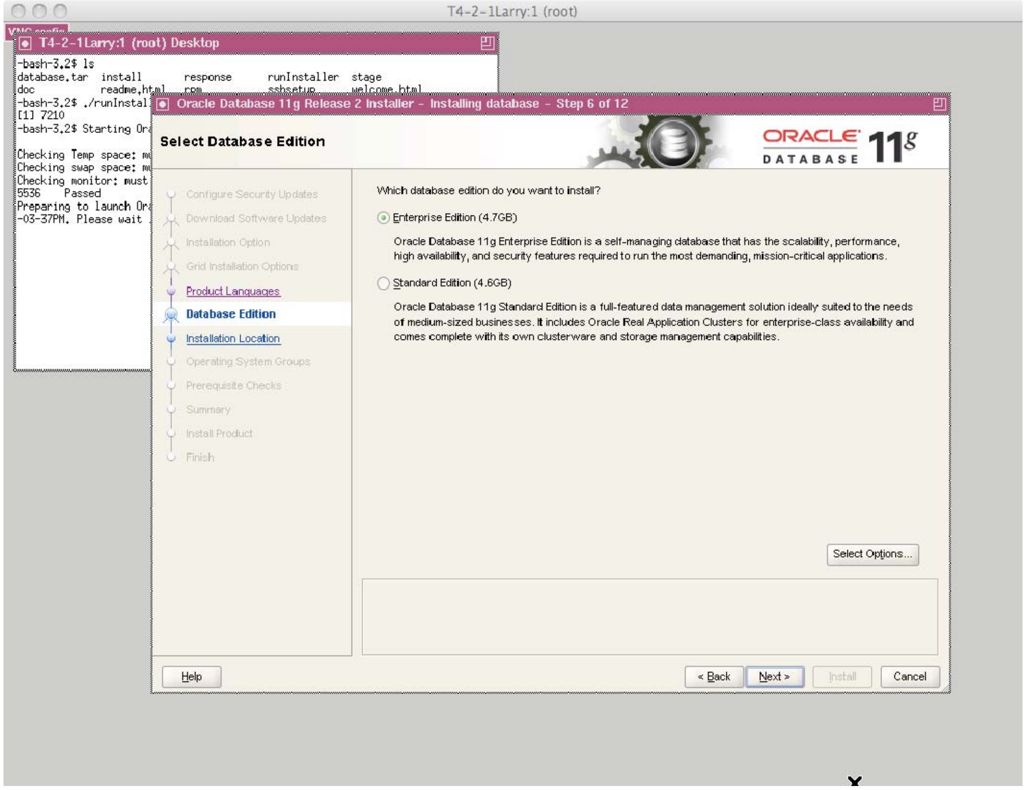
Oracle Database 11g Release 2 Installer - Installing database - Step 5 of 12

Select Product Languages

Select the languages in which your product will run.

Available Languages: Arabic, Bengali, Brazilian Portuguese, Bulgarian, Canadian French, Catalan, Croatian, Czech, Danish, Dutch, Egyptian, English (United Kingdom), Estonian, Finnish, French, German, Greek, Hebrew, Hungarian, Icelandic

Selected Languages: English



```
-bash-3.2$ ls
database.tar  install      response     runInstaller  stage
doc           readme.html  rpm          sshsetup      welcome.html
-bash-3.2$ ./runInstal
[1] 7210
-bash-3.2$ Starting Ora

Checking Temp space: mu
Checking swap space: mu
Checking monitor: must
5536    Passed
Preparing to launch Ora
-03-37PM. Please wait
```

Oracle Database 11g Release 2 Installer - Installing database - Step 6 of 12

Select Database Edition

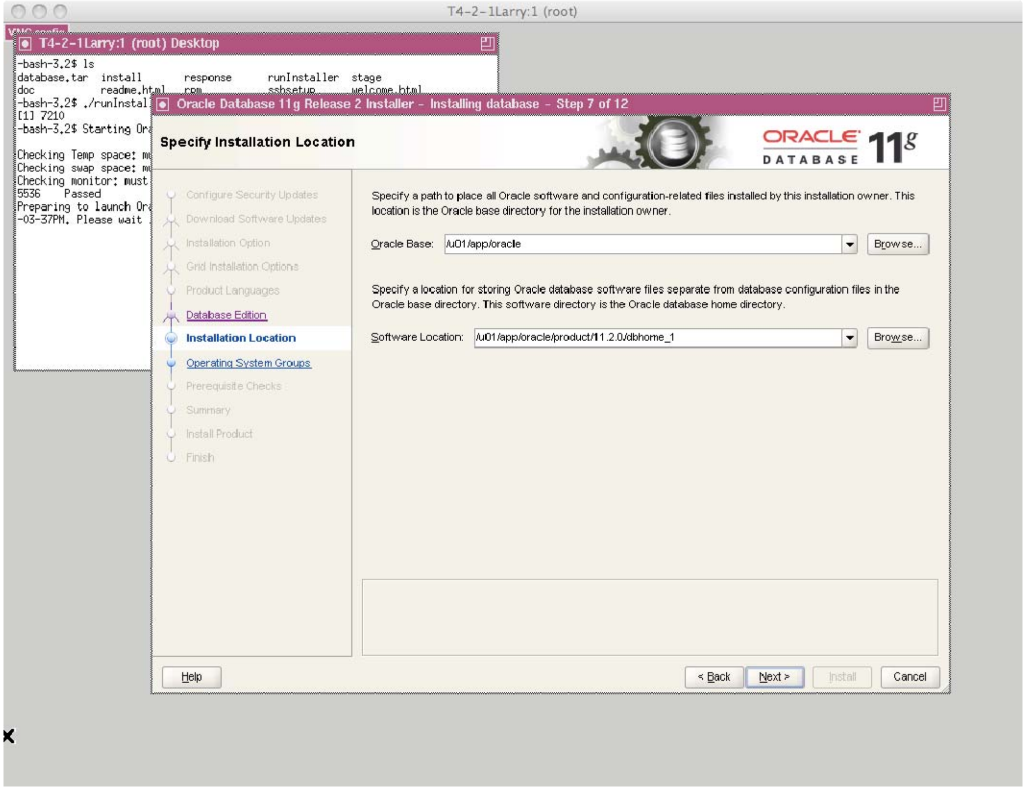
Which database edition do you want to install?

● Enterprise Edition (4.7GB)

Oracle Database 11g Enterprise Edition is a self-managing database that has the scalability, performance, high availability, and security features required to run the most demanding, mission-critical applications.

○ Standard Edition (4.6GB)

Oracle Database 11g Standard Edition is a full-featured data management solution ideally suited to the needs of medium-sized businesses. It includes Oracle Real Application Clusters for enterprise-class availability and comes complete with its own clusterware and storage management capabilities.



```
-bash-3.2$ ls
database.tar  install      response     runInstaller  stage
doc           readme.html  rpm          sshsetup      welcome.html
-bash-3.2$ ./runInstal
[1] 7210
-bash-3.2$ Starting Ora

Checking Temp space: mu
Checking swap space: mu
Checking monitor: must
5536    Passed
Preparing to launch Ora
-03-37PM. Please wait
```

Oracle Database 11g Release 2 Installer - Installing database - Step 7 of 12

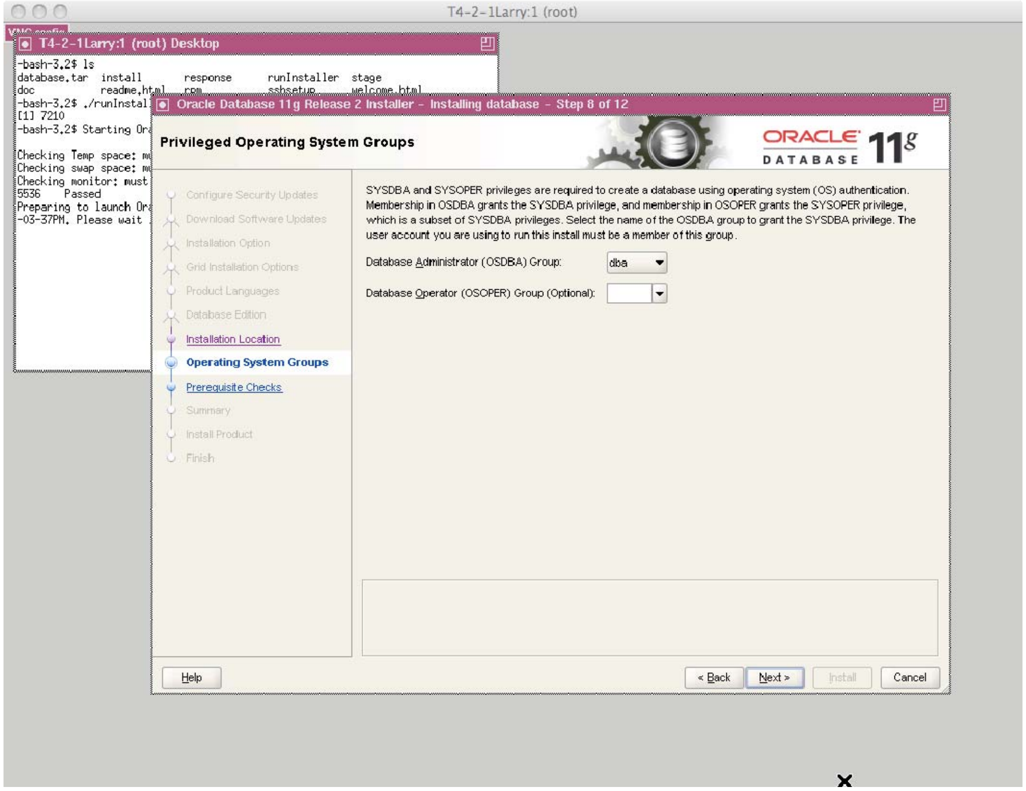
Specify Installation Location

Specify a path to place all Oracle software and configuration-related files installed by this installation owner. This location is the Oracle base directory for the installation owner.

Oracle Base: /u01/app/oracle

Specify a location for storing Oracle database software files separate from database configuration files in the Oracle base directory. This software directory is the Oracle database home directory.

Software Location: /u01/app/oracle/product/11.2.0/dbhome_1



```
-bash-3.2$ ls
database.tar  install      response     runInstaller  stage
doc           readme.html  rpm          sshsetup      welcome.html
-bash-3.2$ ./runInstal
[1] 7210
-bash-3.2$ Starting Ora

Checking Temp space: mu
Checking swap space: mu
Checking monitor: must
5536    Passed
Preparing to launch Ora
-03-37PM. Please wait
```

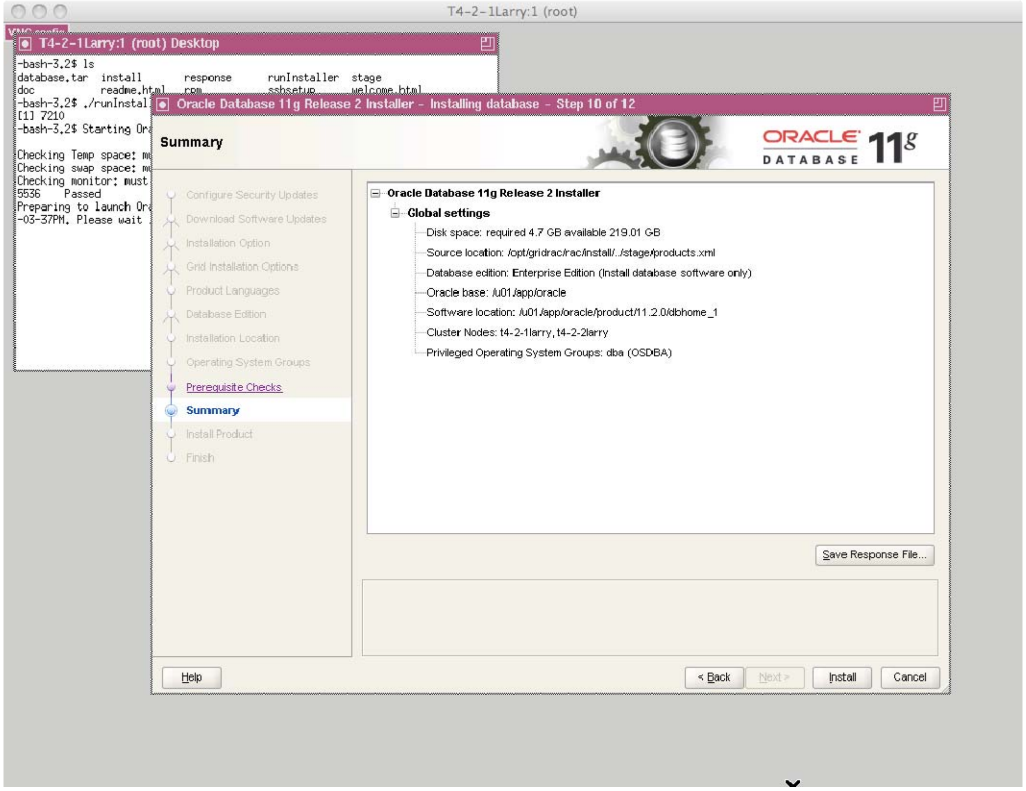
Oracle Database 11g Release 2 Installer - Installing database - Step 8 of 12

Privileged Operating System Groups

SYSDBA and SYSOPER privileges are required to create a database using operating system (OS) authentication. Membership in OSDBA grants the SYSDBA privilege, and membership in OSOPER grants the SYSOPER privilege, which is a subset of SYSDBA privileges. Select the name of the OSDBA group to grant the SYSDBA privilege. The user account you are using to run this install must be a member of this group.

Database Administrator (OSDBA) Group: dba

Database Operator (OSOPER) Group (Optional):



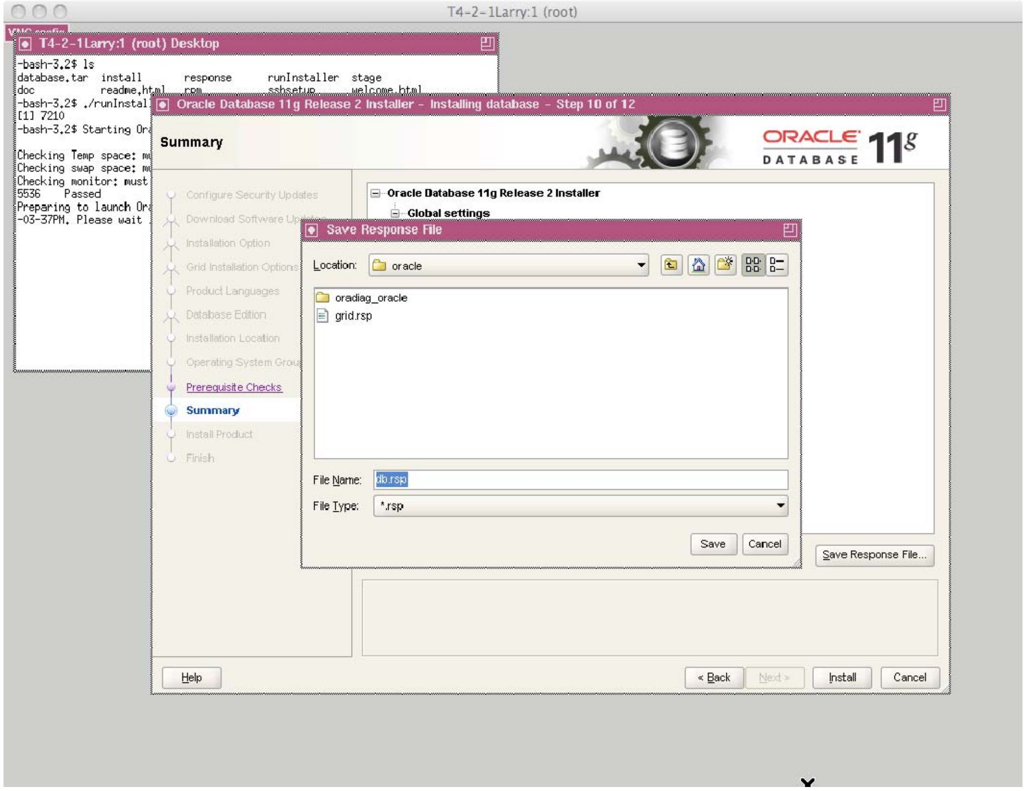
```
-bash-3.2$ ls
database.tar  install      response     runInstaller  stage
doc           readme.html  rpm          sshsetup      welcome.html
-bash-3.2$ ./runInstal
[1] 7210
-bash-3.2$ Starting Or

Checking Temp space: m
Checking swap space: m
Checking monitor: must
5536    Passed
Preparing to launch Or
-03-37PM. Please wait
```

Oracle Database 11g Release 2 Installer - Installing database - Step 10 of 12

Summary

- Oracle Database 11g Release 2 Installer
  - Global settings
    - Disk space: required 4.7 GB available 219.01 GB
    - Source location: /opt/gridrac/rac/install/../stage/products.xml
    - Database edition: Enterprise Edition (Install database software only)
    - Oracle base: /u01/app/oracle
    - Software location: /u01/app/oracle/product/11.2.0/dbhome_1
    - Cluster Nodes: t4-2-1larry, t4-2-2larry
    - Privileged Operating System Groups: dba (OSDBA)



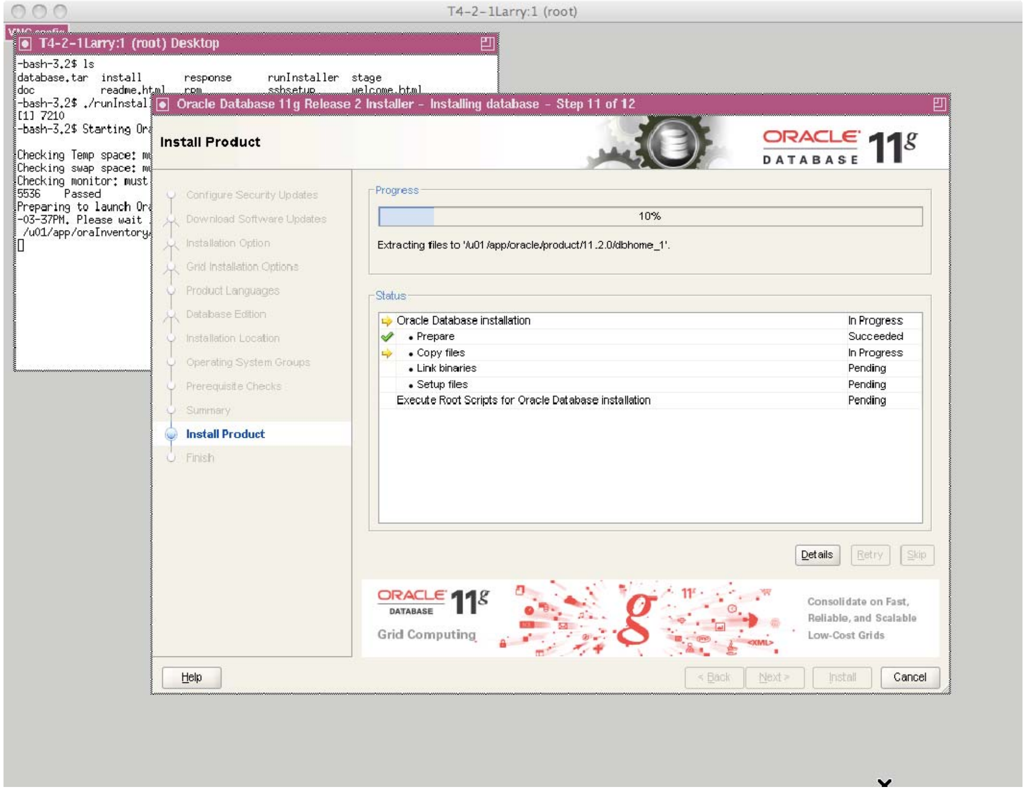
T4-2-1Larry:1 (root)
VNC config
T4-2-1Larry:1 (root) Desktop
Oracle Database 11g Release 2 Installer - Installing database - Step 10 of 12
Summary
ORACLE DATABASE 11g
Oracle Database 11g Release 2 Installer
Global settings
Save Response File

Location: oracle

oradiag_oracle
grid.rsp

File Name: db.rsp

File Type: *.rsp

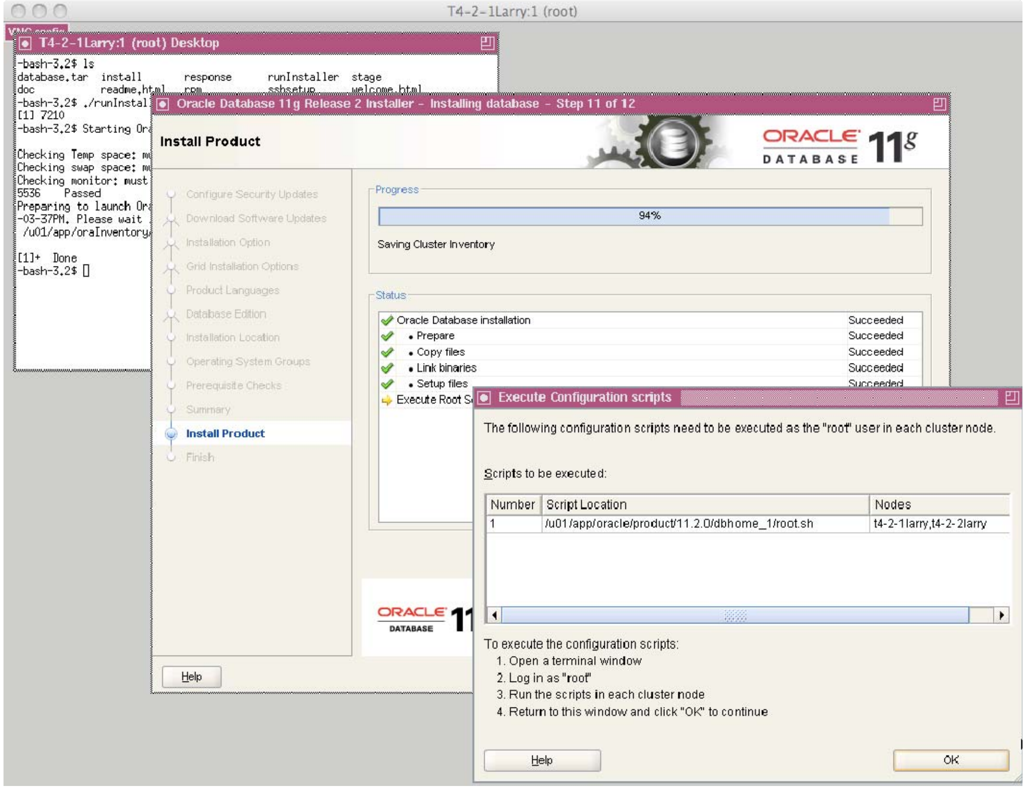


```
-bash-3.2$ ls
database.tar  install      response     runInstaller  stage
doc           readme.html  rpm          sshsetup      welcome.html
-bash-3.2$ ./runInstal
[1] 7210
-bash-3.2$ Starting Ora

Checking Temp space: mu
Checking swap space: mu
Checking monitor: must
5536    Passed
Preparing to launch Ora
-03-37PM. Please wait
 /u01/app/oraInventory
```

Extracting files to '/u01/app/oracle/product/11.2.0/dbhome_1'.

Oracle Database installation	In Progress
• Prepare	Succeeded
• Copy files	In Progress
• Link binaries	Pending
• Setup files	Pending
Execute Root Scripts for Oracle Database installation	Pending



```
-bash-3.2$ ls
database.tar  install      response     runInstaller  stage
doc           readme.html  rpm          sshsetup      welcome.html
-bash-3.2$ ./runInstal
[1] 7210
-bash-3.2$ Starting Ora

Checking Temp space: mu
Checking swap space: mu
Checking monitor: must
5536    Passed
Preparing to launch Ora
-03-37PM. Please wait
 /u01/app/oraInventory

[1]+  Done
-bash-3.2$
```

Saving Cluster Inventory

Oracle Database installation	Succeeded
• Prepare	Succeeded
• Copy files	Succeeded
• Link binaries	Succeeded
• Setup files	Succeeded
Execute Root S	

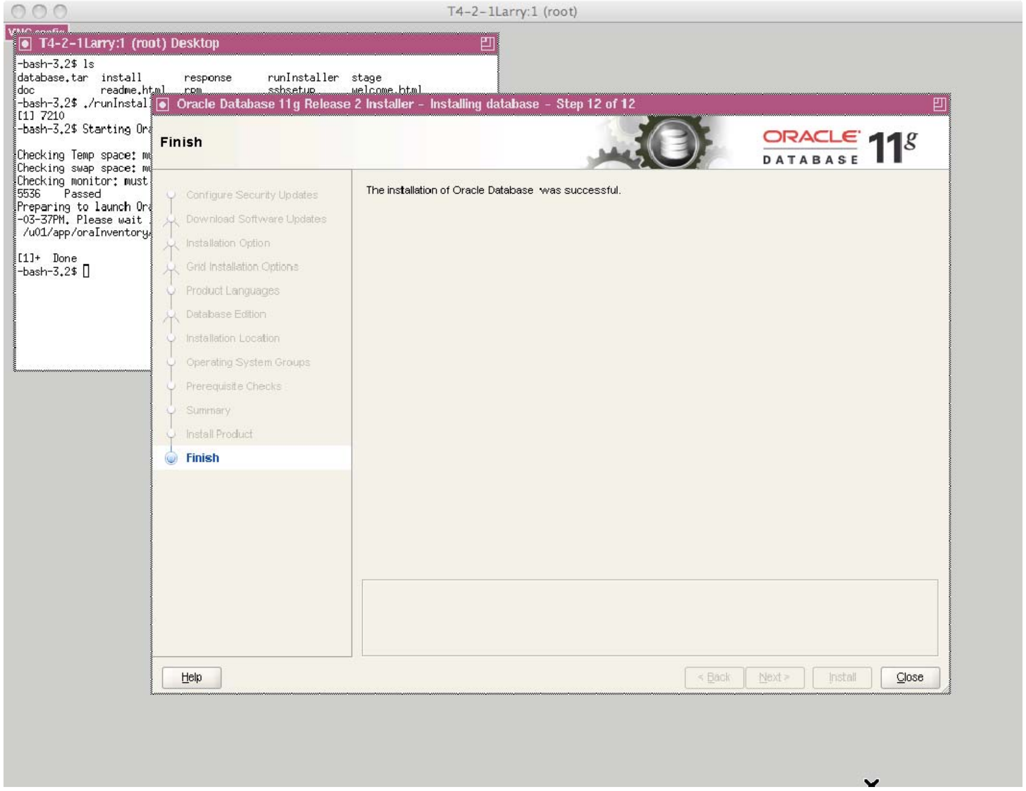
The following configuration scripts need to be executed as the "root" user in each cluster node.

Scripts to be executed:

Number	Script Location	Nodes
1	/u01/app/oracle/product/11.2.0/dbhome_1/root.sh	t4-2-1larry,t4-2-2larry

To execute the configuration scripts:

1. Open a terminal window
2. Log in as "root"
3. Run the scripts in each cluster node
4. Return to this window and click "OK" to continue



```
-bash-3.2$ ls
database.tar  install       response      runInstaller  stage
doc           readme.html   rpm           sshsetup      welcome.html
-bash-3.2$ ./runInstal
[1] 7210
-bash-3.2$ Starting Or

Checking Temp space: mu
Checking swap space: mu
Checking monitor: must
5536    Passed
Preparing to launch Or
-03-37PM. Please wait
 /u01/app/oraInventory

[1]+  Done
-bash-3.2$ 
```

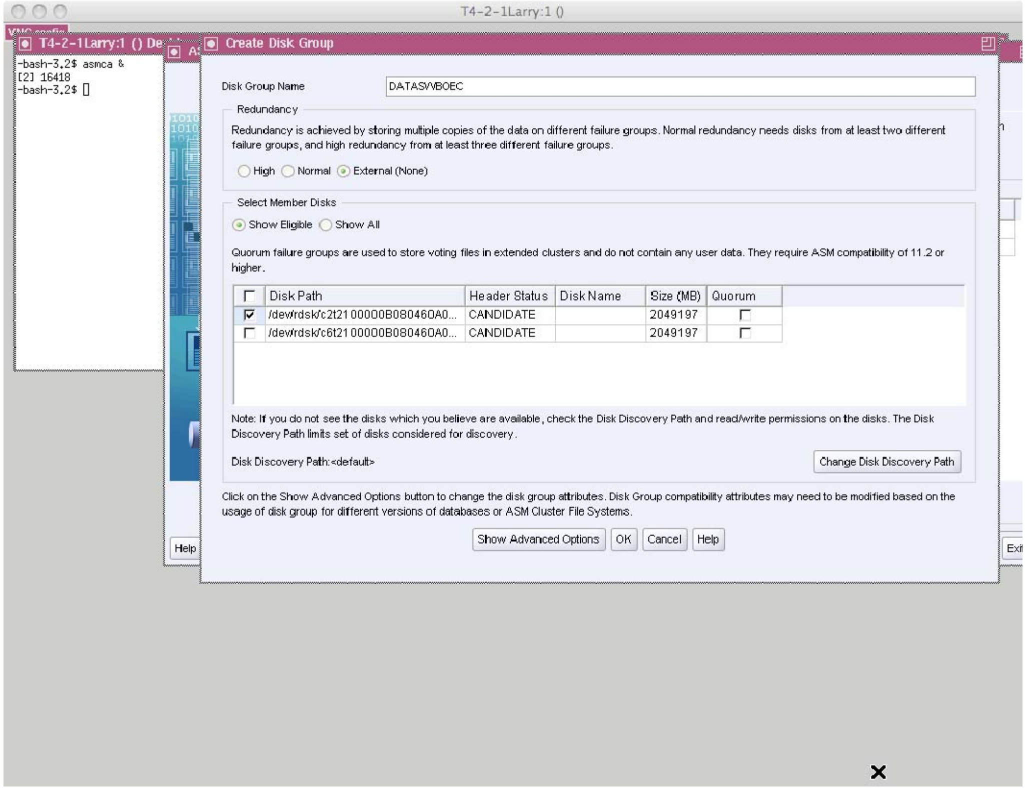
The installation of Oracle Database was successful.

Configuring Automatic Storage Management

The following screenshots are the steps taken to further configure Automatic Storage Management on the previously deployed Grid Infrastructure used with this white paper.

These VNC-captured screenshots were gathered after an `su – oracle` command was issued followed by entering an `asmca` command to launch the GUI ASM Configuration Assistant.

NOTE: The parameters entered here should be modified to fit organizational requirements. The screenshots are included to show how ASM was configured to utilize Pillar Axiom SAN LUNs deployed for the specific configuration used with this white paper.



```
-bash-3.2$ asmca &
[2] 16418
-bash-3.2$ []
```

Disk Group Name: DATASWBOEC

Redundancy

Redundancy is achieved by storing multiple copies of the data on different failure groups. Normal redundancy needs disks from at least two different failure groups, and high redundancy from at least three different failure groups.

○ High ○ Normal ● External (None)

Select Member Disks

● Show Eligible ○ Show All

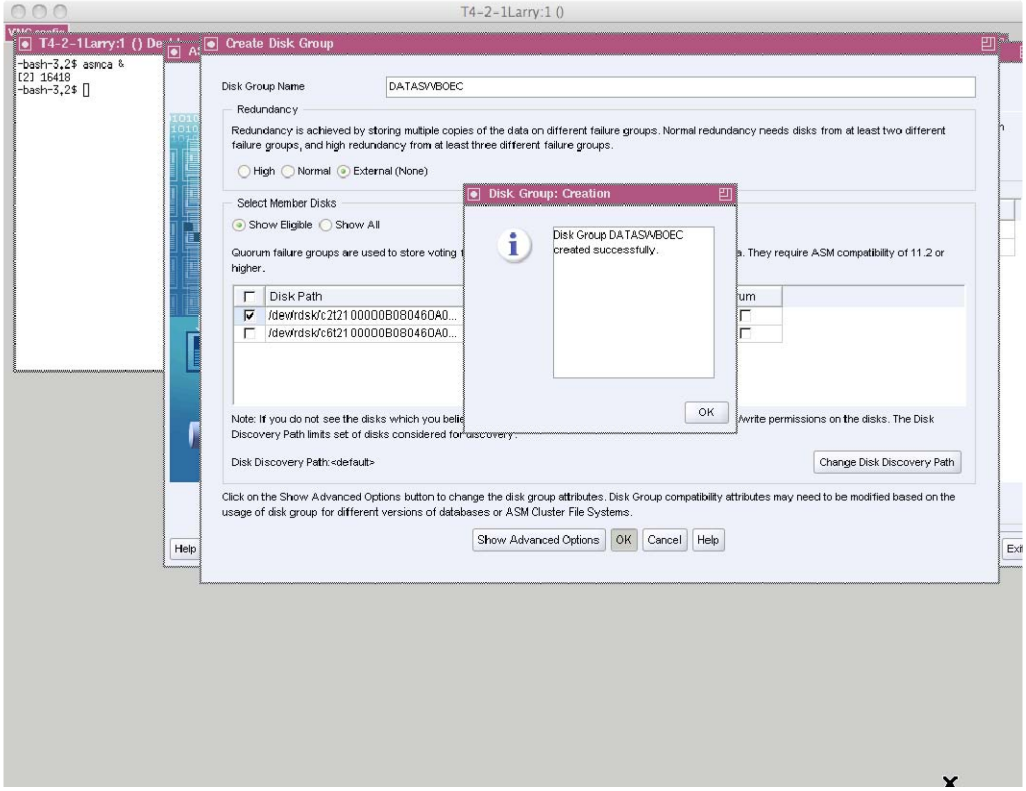
Quorum failure groups are used to store voting files in extended clusters and do not contain any user data. They require ASM compatibility of 11.2 or higher.

□	Disk Path	Header Status	Disk Name	Size (MB)	Quorum
☑	/dev/rdsk/c2t2100000B080460A0...	CANDIDATE		2049197	□
□	/dev/rdsk/c6t2100000B080460A0...	CANDIDATE		2049197	□

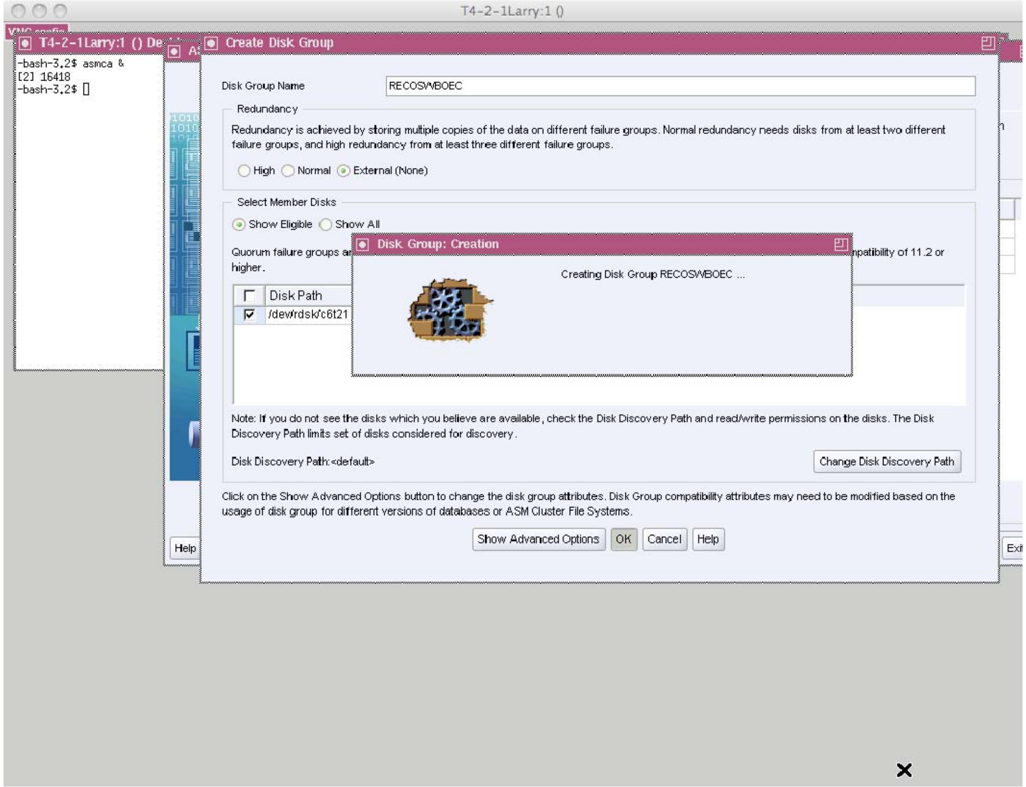
Note: If you do not see the disks which you believe are available, check the Disk Discovery Path and read/write permissions on the disks. The Disk Discovery Path limits set of disks considered for discovery.

Disk Discovery Path:<default>

Click on the Show Advanced Options button to change the disk group attributes. Disk Group compatibility attributes may need to be modified based on the usage of disk group for different versions of databases or ASM Cluster File Systems.



Disk Group DATASWBOEC created successfully.



```
-bash-3.2$ asmca &
[2] 16418
-bash-3.2$ 
```

Disk Group Name: RECOSWBOEC

Redundancy

Redundancy is achieved by storing multiple copies of the data on different failure groups. Normal redundancy needs disks from at least two different failure groups, and high redundancy from at least three different failure groups.

○ High ○ Normal ● External (None)

Select Member Disks

● Show Eligible ○ Show All

Quorum failure groups ar… …mpatibility of 11.2 or higher.

☐	Disk Path
☑	/dev/rdsk/c6t21

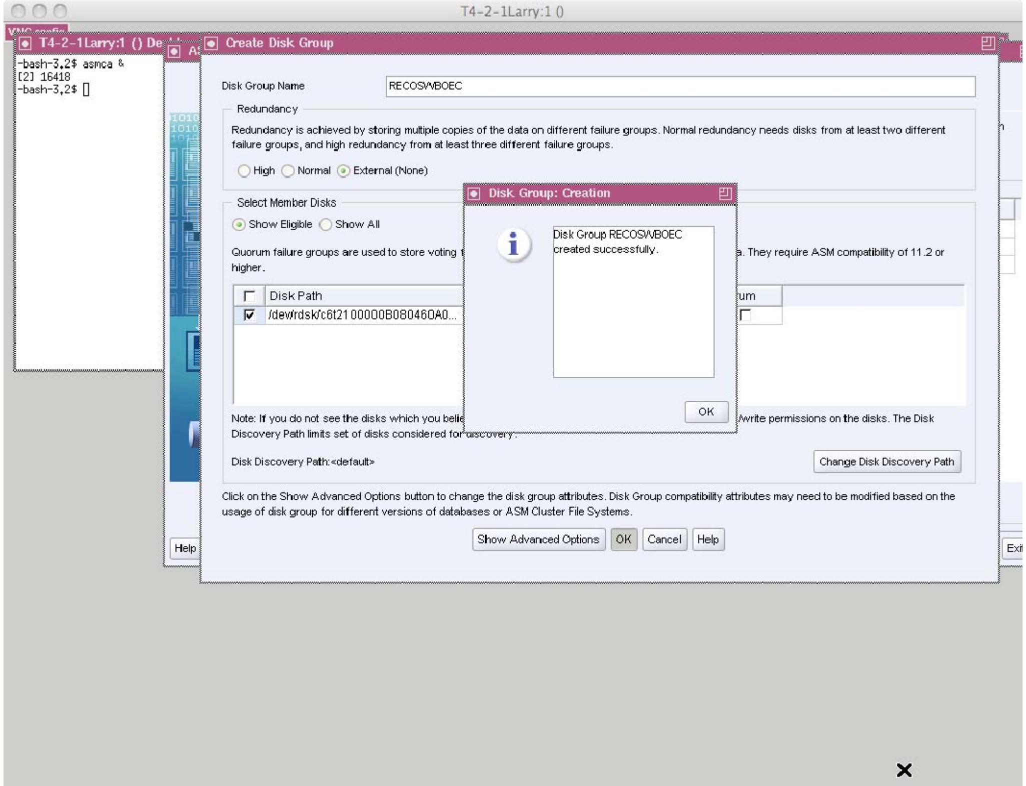
Disk Group: Creation

Creating Disk Group RECOSWBOEC ...

Note: If you do not see the disks which you believe are available, check the Disk Discovery Path and read/write permissions on the disks. The Disk Discovery Path limits set of disks considered for discovery.

Disk Discovery Path:<default>

Click on the Show Advanced Options button to change the disk group attributes. Disk Group compatibility attributes may need to be modified based on the usage of disk group for different versions of databases or ASM Cluster File Systems.



```
-bash-3.2$ asmca &
[2] 16418
-bash-3.2$ 
```

Disk Group Name: RECOSWBOEC

Redundancy

Redundancy is achieved by storing multiple copies of the data on different failure groups. Normal redundancy needs disks from at least two different failure groups, and high redundancy from at least three different failure groups.

○ High ○ Normal ● External (None)

Select Member Disks

● Show Eligible ○ Show All

Quorum failure groups are used to store voting … …a. They require ASM compatibility of 11.2 or higher.

☐	Disk Path	…rum
☑	/dev/rdsk/c6t21000000B080460A0...	☐

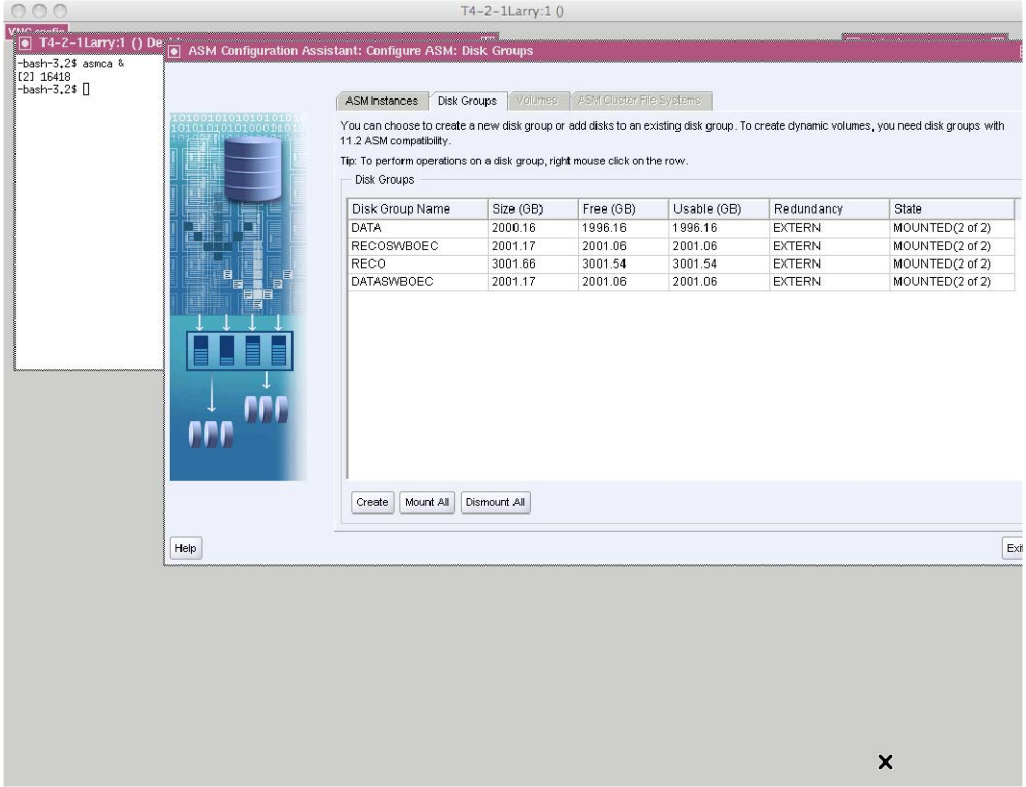
Disk Group: Creation

Disk Group RECOSWBOEC created successfully.

Note: If you do not see the disks which you belie… …/write permissions on the disks. The Disk Discovery Path limits set of disks considered for discovery.

Disk Discovery Path:<default>

Click on the Show Advanced Options button to change the disk group attributes. Disk Group compatibility attributes may need to be modified based on the usage of disk group for different versions of databases or ASM Cluster File Systems.



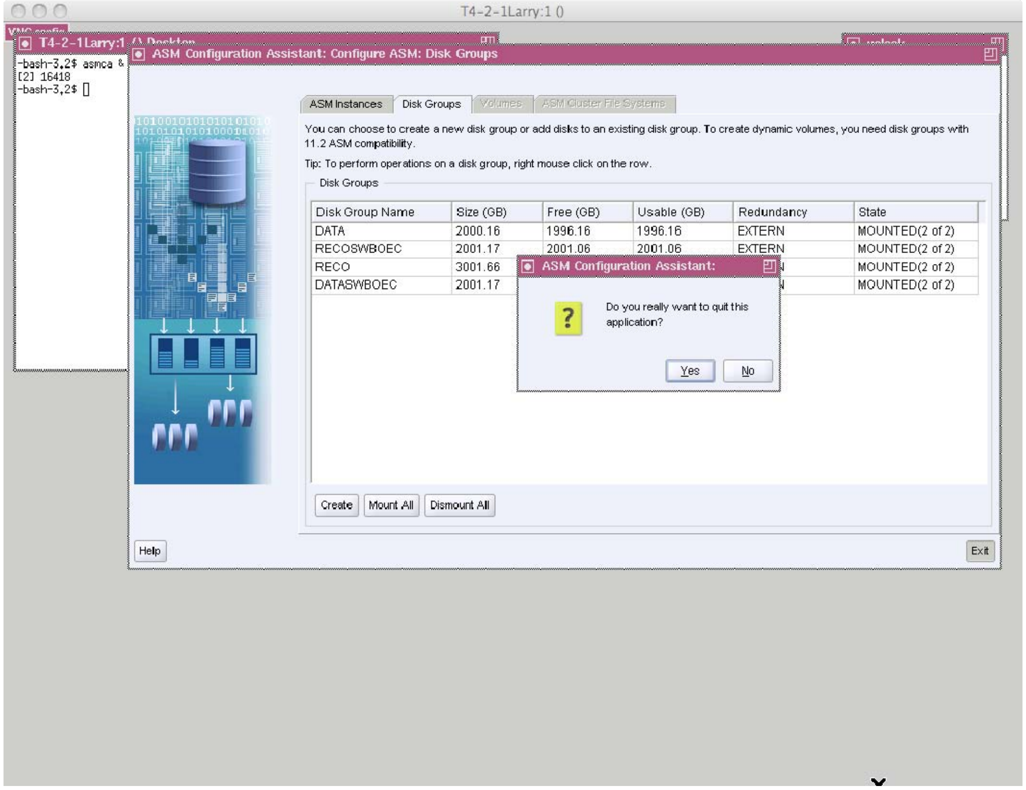
```
-bash-3.2$ asmca &
[2] 16418
-bash-3.2$ 
```

You can choose to create a new disk group or add disks to an existing disk group. To create dynamic volumes, you need disk groups with 11.2 ASM compatibility.

Tip: To perform operations on a disk group, right mouse click on the row.

Disk Groups

Disk Group Name	Size (GB)	Free (GB)	Usable (GB)	Redundancy	State
DATA	2000.16	1996.16	1996.16	EXTERN	MOUNTED(2 of 2)
RECOSWBOEC	2001.17	2001.06	2001.06	EXTERN	MOUNTED(2 of 2)
RECO	3001.66	3001.54	3001.54	EXTERN	MOUNTED(2 of 2)
DATASWBOEC	2001.17	2001.06	2001.06	EXTERN	MOUNTED(2 of 2)



```
-bash-3.2$ asmca &
[2] 16418
-bash-3.2$ 
```

You can choose to create a new disk group or add disks to an existing disk group. To create dynamic volumes, you need disk groups with 11.2 ASM compatibility.

Tip: To perform operations on a disk group, right mouse click on the row.

Disk Groups

Disk Group Name	Size (GB)	Free (GB)	Usable (GB)	Redundancy	State
DATA	2000.16	1996.16	1996.16	EXTERN	MOUNTED(2 of 2)
RECOSWBOEC	2001.17	2001.06	2001.06	EXTERN	MOUNTED(2 of 2)
RECO	3001.66			N	MOUNTED(2 of 2)
DATASWBOEC	2001.17			N	MOUNTED(2 of 2)

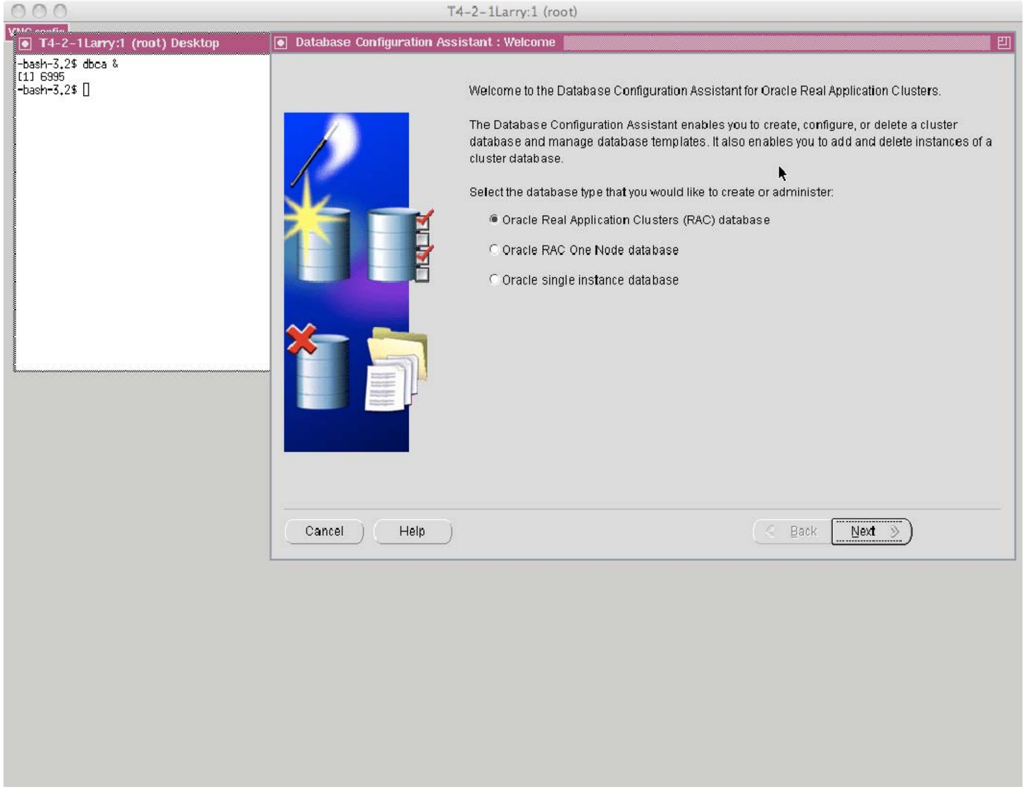
Do you really want to quit this application?

Building an Oracle 11*g* Release 2 Database with the Database Configuration Assistant

The following screenshots are the steps taken to further configure Automatic Storage Management on the previously deployed Grid Infrastructure used with this white paper.

These VNC-captured screenshots were gathered after an `su – oracle` command was issued followed by entering a `dbca` command to launch the GUI Database Configuration Assistant for Oracle Real Application Clusters.

NOTE: The parameters entered here should be modified to fit organizational requirements. The screenshots are included to show how `dbca` was used to build a custom database that is implemented on top of ASM on the associated Pillar Axiom SAN that was deployed for the specific configuration used with this white paper.


T4-2-1Larry:1 (root)
T4-2-1Larry:1 (root) Desktop

```
-bash-3.2$ dbca &
[1] 6995
-bash-3.2$ 
```

Database Configuration Assistant : Welcome

Welcome to the Database Configuration Assistant for Oracle Real Application Clusters.

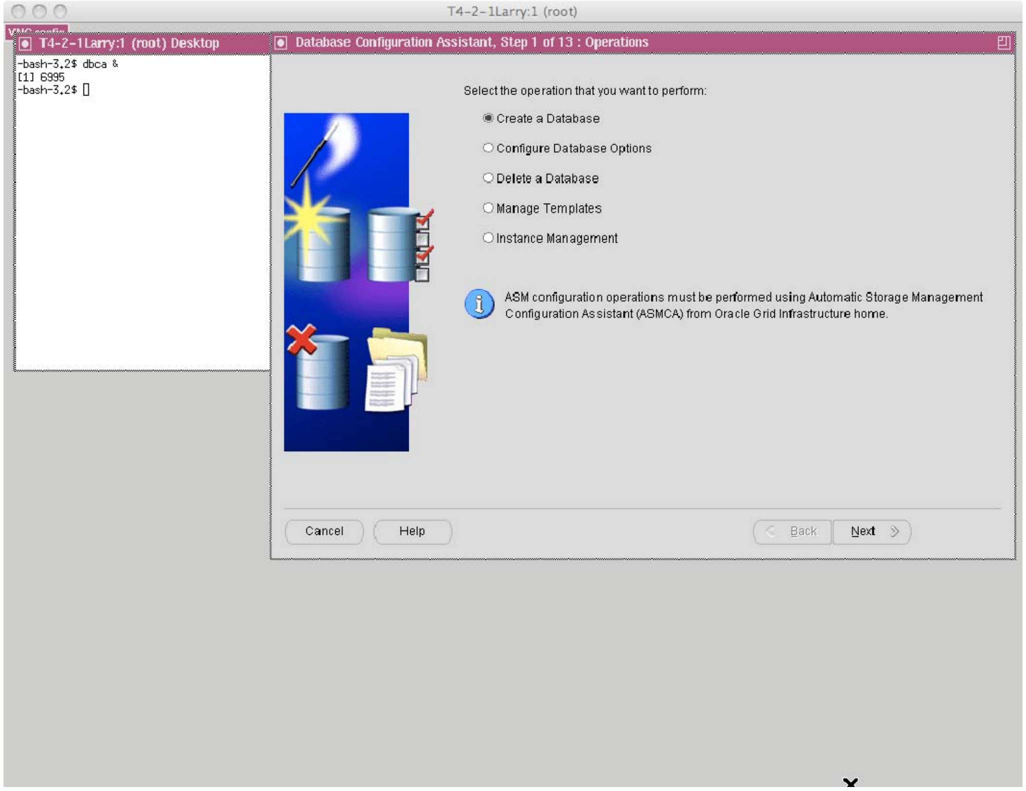
The Database Configuration Assistant enables you to create, configure, or delete a cluster database and manage database templates. It also enables you to add and delete instances of a cluster database.

Select the database type that you would like to create or administer:

● Oracle Real Application Clusters (RAC) database

○ Oracle RAC One Node database

○ Oracle single instance database



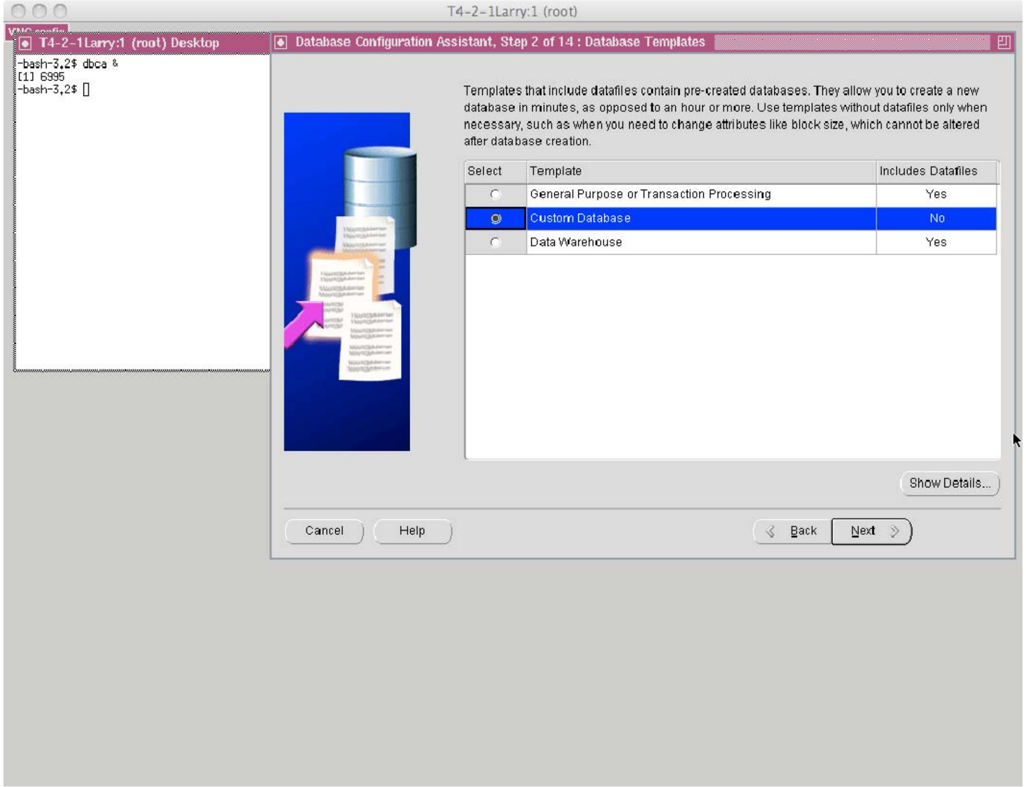
```
-bash-3.2$ dbca &
[1] 6995
-bash-3.2$ 
```

Database Configuration Assistant, Step 1 of 13 : Operations

Select the operation that you want to perform:

- ● Create a Database
- ○ Configure Database Options
- ○ Delete a Database
- ○ Manage Templates
- ○ Instance Management

ASM configuration operations must be performed using Automatic Storage Management Configuration Assistant (ASMCA) from Oracle Grid Infrastructure home.

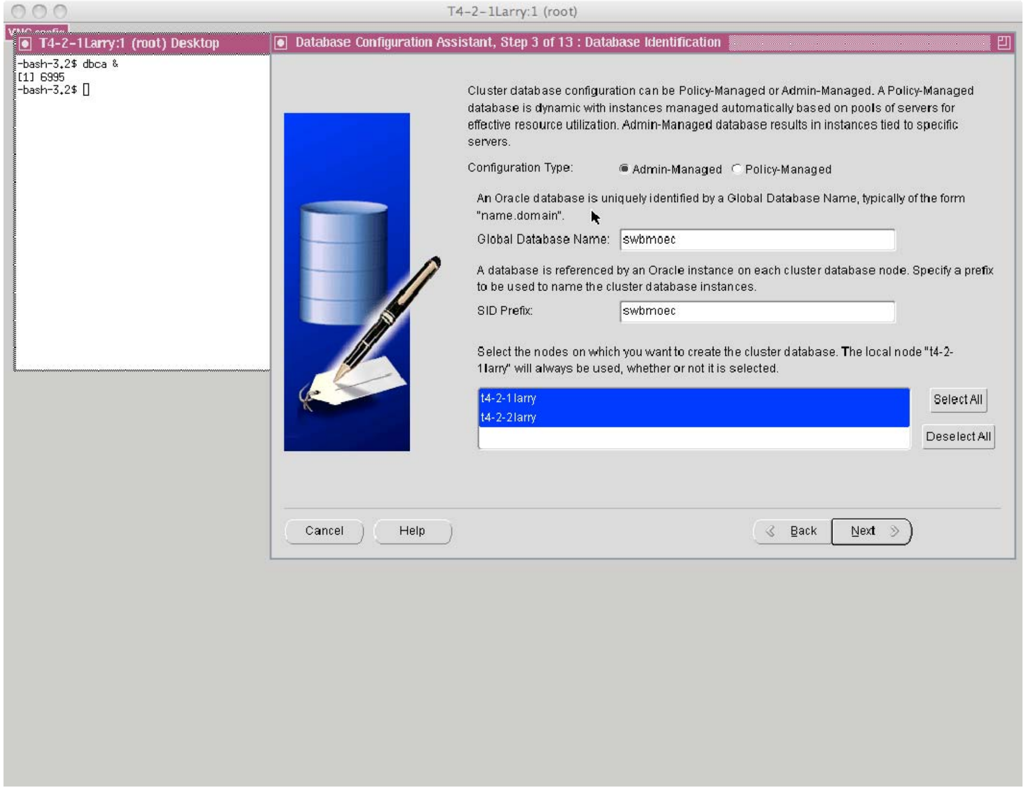


```
-bash-3.2$ dbca &
[1] 6995
-bash-3.2$ 
```

Database Configuration Assistant, Step 2 of 14 : Database Templates

Templates that include datafiles contain pre-created databases. They allow you to create a new database in minutes, as opposed to an hour or more. Use templates without datafiles only when necessary, such as when you need to change attributes like block size, which cannot be altered after database creation.

Select	Template	Includes Datafiles
○	General Purpose or Transaction Processing	Yes
●	Custom Database	No
○	Data Warehouse	Yes



```
-bash-3.2$ dbca &
[1] 6995
-bash-3.2$ []
```

Database Configuration Assistant, Step 3 of 13 : Database Identification

Cluster database configuration can be Policy-Managed or Admin-Managed. A Policy-Managed database is dynamic with instances managed automatically based on pools of servers for effective resource utilization. Admin-Managed database results in instances tied to specific servers.

Configuration Type: ● Admin-Managed ○ Policy-Managed

An Oracle database is uniquely identified by a Global Database Name, typically of the form "name.domain".

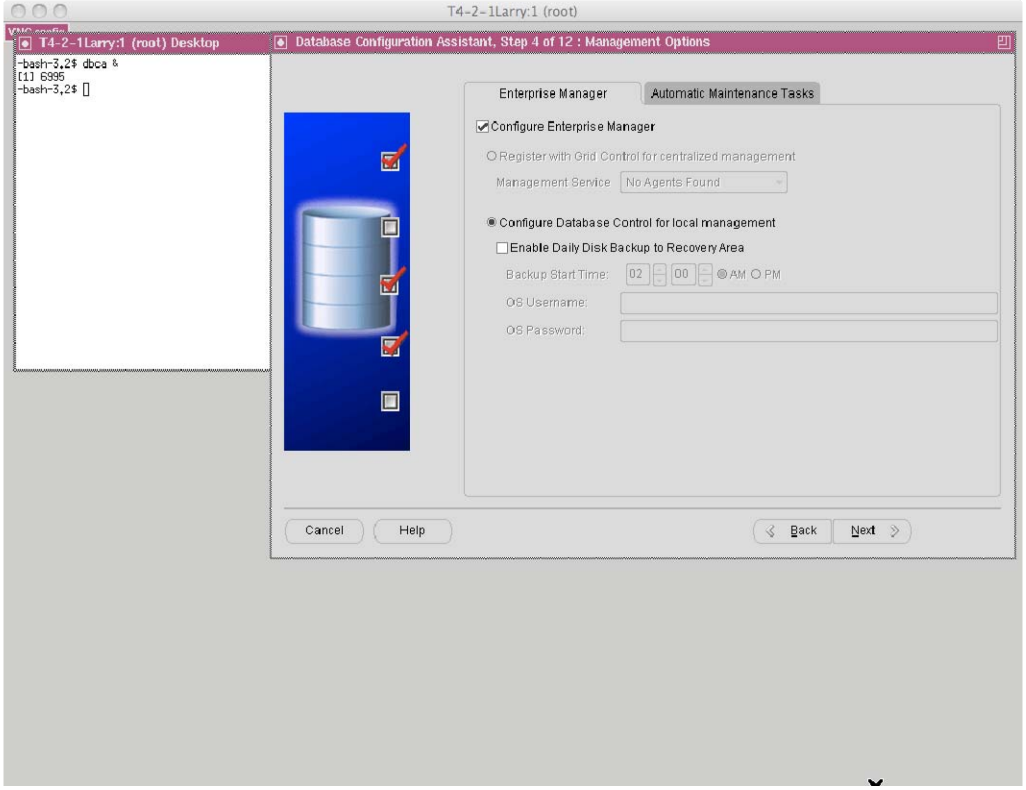
Global Database Name: swbmoec

A database is referenced by an Oracle instance on each cluster database node. Specify a prefix to be used to name the cluster database instances.

SID Prefix: swbmoec

Select the nodes on which you want to create the cluster database. The local node "t4-2-1larry" will always be used, whether or not it is selected.

t4-2-1larry
t4-2-2larry



```
-bash-3.2$ dbca &
[1] 6995
-bash-3.2$ []
```

Database Configuration Assistant, Step 4 of 12 : Management Options

☑ Configure Enterprise Manager

○ Register with Grid Control for centralized management

Management Service: No Agents Found

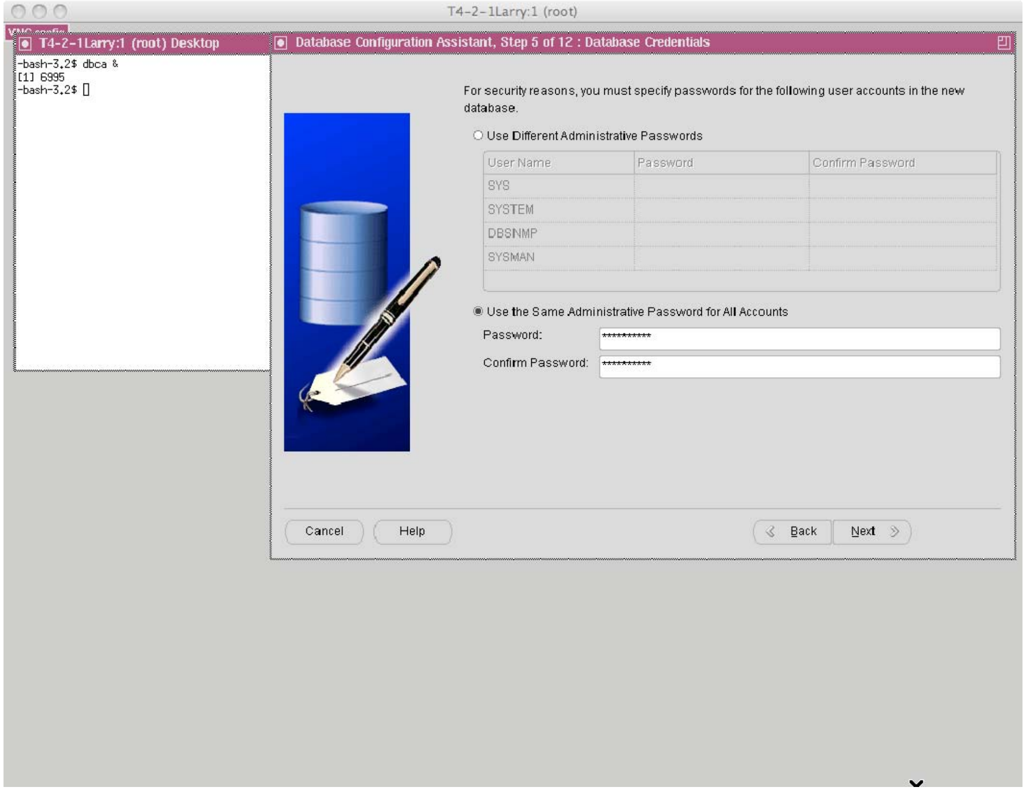
● Configure Database Control for local management

□ Enable Daily Disk Backup to Recovery Area

Backup Start Time: 02 00 ● AM ○ PM

OS Username:

OS Password:



```
-bash-3.2$ dbca &
[1] 6995
-bash-3.2$ []
```

Database Configuration Assistant, Step 5 of 12 : Database Credentials

For security reasons, you must specify passwords for the following user accounts in the new database.

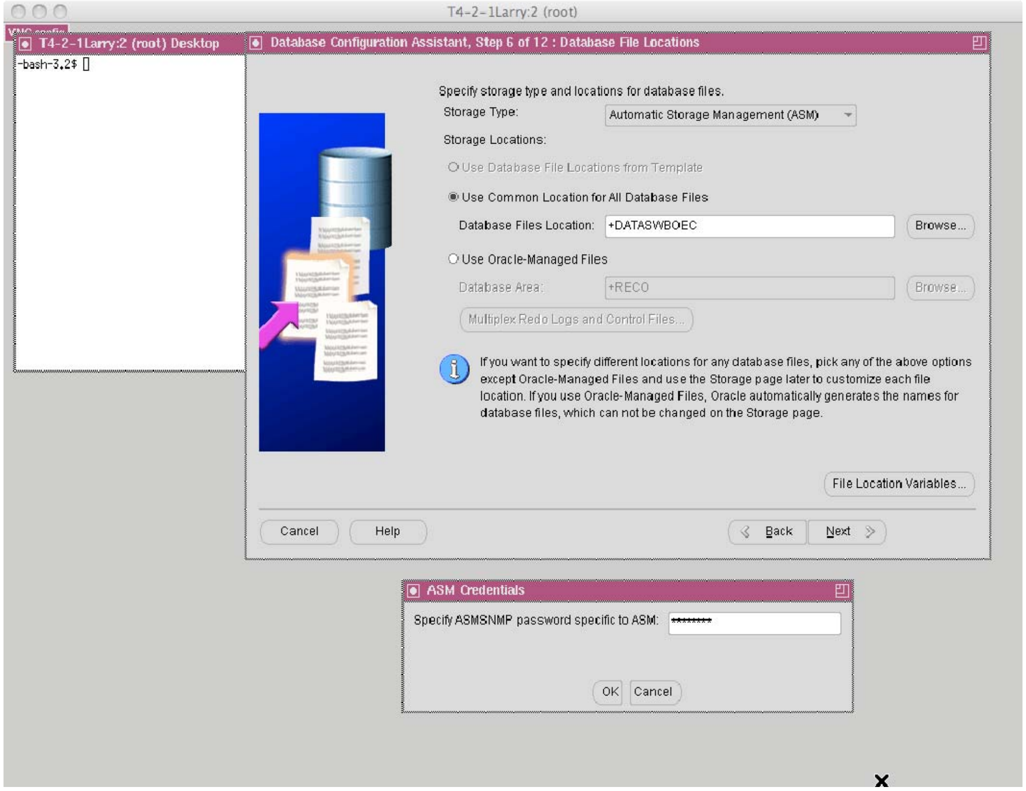
○ Use Different Administrative Passwords

User Name	Password	Confirm Password
SYS		
SYSTEM		
DBSNMP		
SYSMAN		

● Use the Same Administrative Password for All Accounts

Password: **********

Confirm Password: **********



```
-bash-3.2$ []
```

Database Configuration Assistant, Step 6 of 12 : Database File Locations

Specify storage type and locations for database files.

Storage Type: Automatic Storage Management (ASM)

Storage Locations:

○ Use Database File Locations from Template

● Use Common Location for All Database Files

Database Files Location: +DATASWBOEC

○ Use Oracle-Managed Files

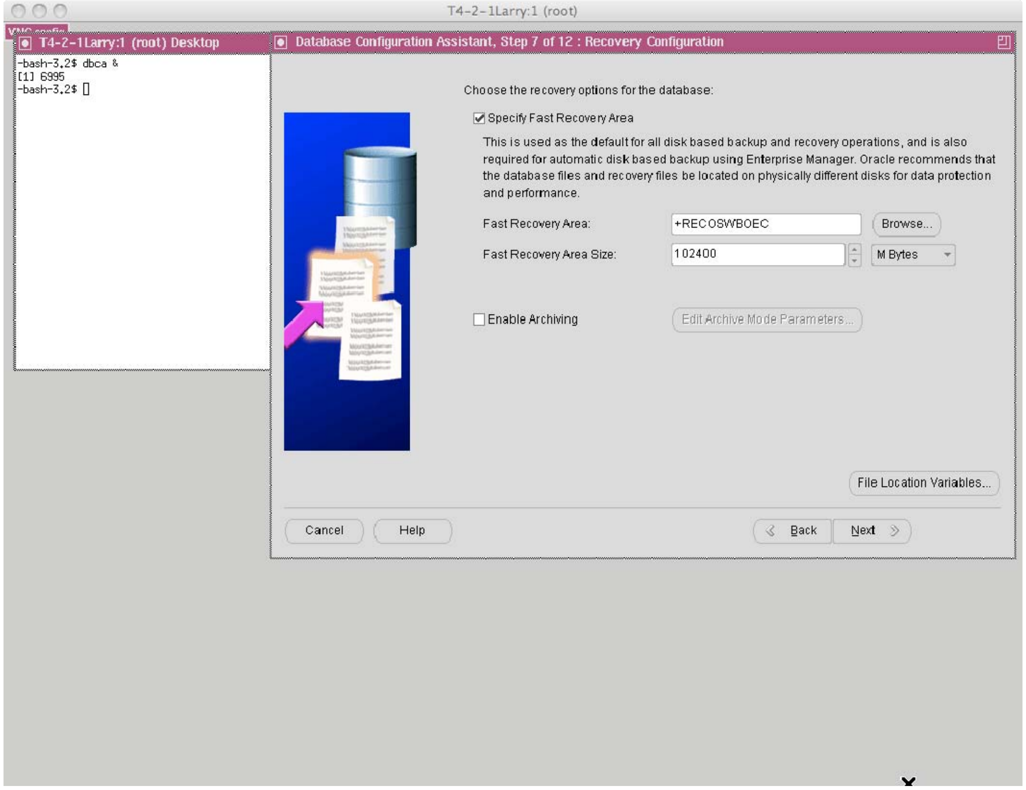
Database Area: +RECO

If you want to specify different locations for any database files, pick any of the above options except Oracle-Managed Files and use the Storage page later to customize each file location. If you use Oracle-Managed Files, Oracle automatically generates the names for database files, which can not be changed on the Storage page.

Browse...
Browse...
Multiplex Redo Logs and Control Files...
File Location Variables...
Cancel
Help
Back
Next

ASM Credentials

Specify ASMSNMP password specific to ASM: ********


T4-2-1Larry:1 (root)
T4-2-1Larry:1 (root) Desktop
Database Configuration Assistant, Step 7 of 12 : Recovery Configuration

```
-bash-3.2$ dbca &
[1] 6995
-bash-3.2$ ▯
```

Choose the recovery options for the database:

☑ Specify Fast Recovery Area

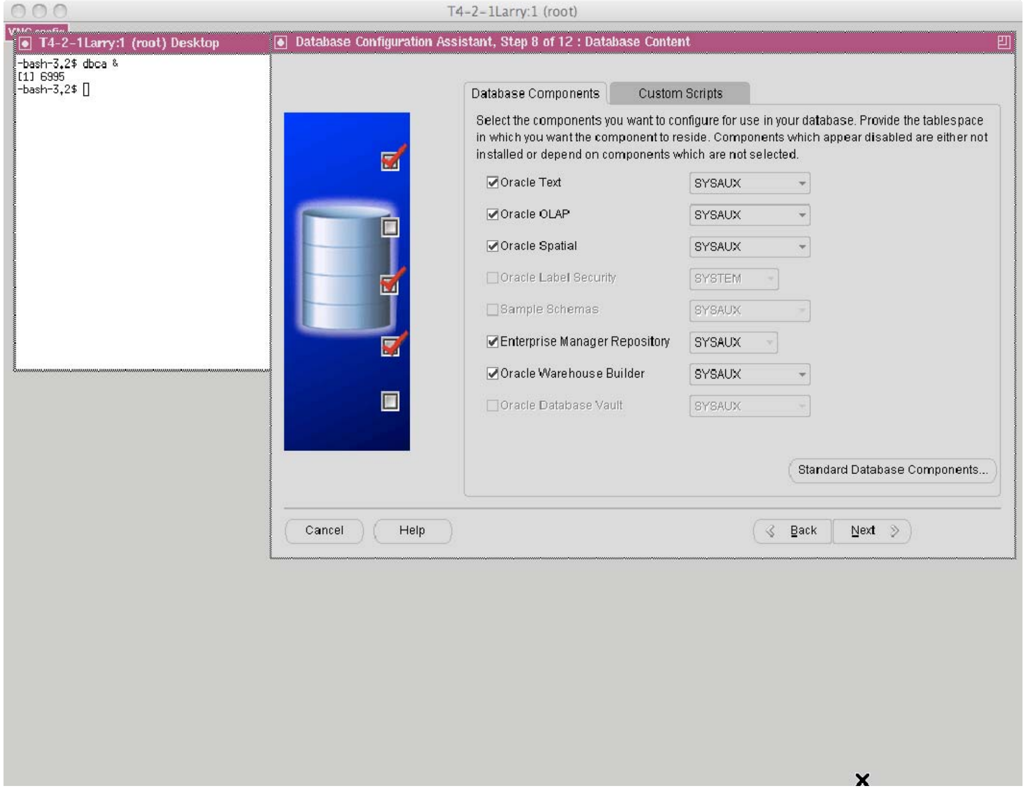
This is used as the default for all disk based backup and recovery operations, and is also required for automatic disk based backup using Enterprise Manager. Oracle recommends that the database files and recovery files be located on physically different disks for data protection and performance.

Fast Recovery Area: +RECOSWBOEC

Fast Recovery Area Size: 102400 M Bytes

□ Enable Archiving

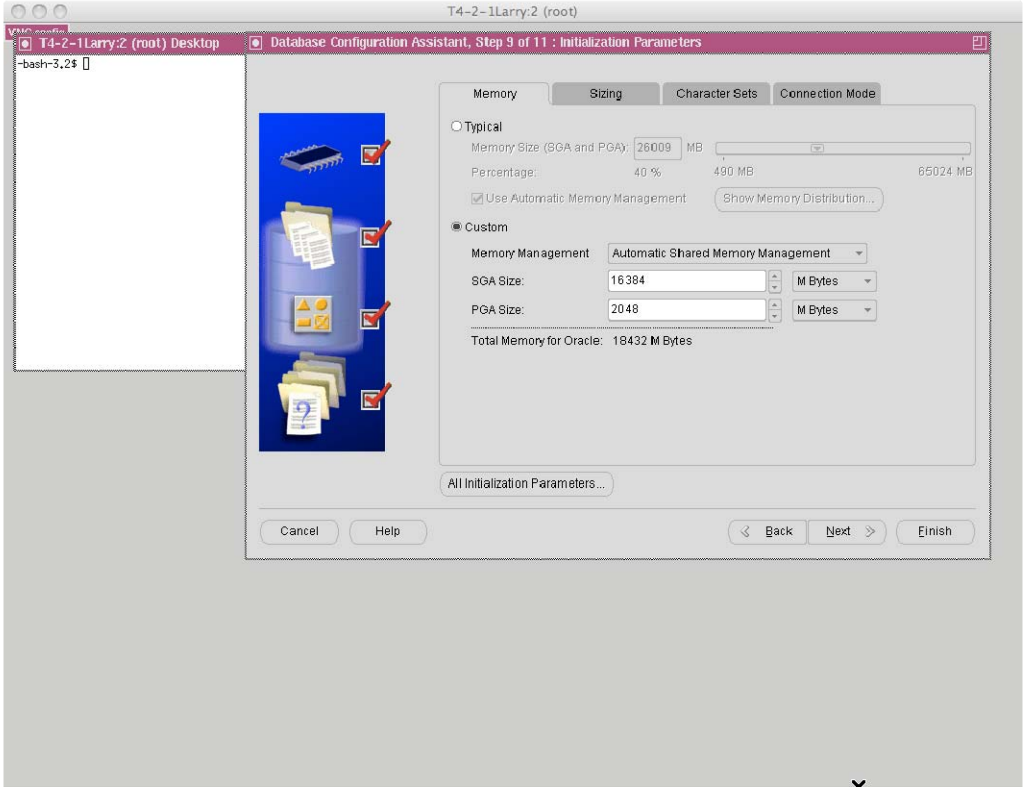
Browse...
Edit Archive Mode Parameters...
File Location Variables...
Cancel
Help
Back
Next


T4-2-1Larry:1 (root)
T4-2-1Larry:1 (root) Desktop
Database Configuration Assistant, Step 8 of 12 : Database Content
Database Components
Custom Scripts

```
-bash-3.2$ dbca &
[1] 6995
-bash-3.2$ ▯
```

Select the components you want to configure for use in your database. Provide the tablespace in which you want the component to reside. Components which appear disabled are either not installed or depend on components which are not selected.

☑ Oracle Text	SYSAUX
☑ Oracle OLAP	SYSAUX
☑ Oracle Spatial	SYSAUX
□ Oracle Label Security	SYSTEM
□ Sample Schemas	SYSAUX
☑ Enterprise Manager Repository	SYSAUX
☑ Oracle Warehouse Builder	SYSAUX
□ Oracle Database Vault	SYSAUX


T4-2-1Larry:2 (root)
T4-2-1Larry:2 (root) Desktop
Memory
Sizing
Character Sets
Connection Mode
Show Memory Distribution...
All Initialization Parameters...
Cancel
Help
Back
Next
Finish

Database Configuration Assistant, Step 9 of 11 : Initialization Parameters

```
-bash-3.2$
```

○ Typical

Memory Size (SGA and PGA): 26009 MB

Percentage: 40 % 490 MB 65024 MB

☑ Use Automatic Memory Management

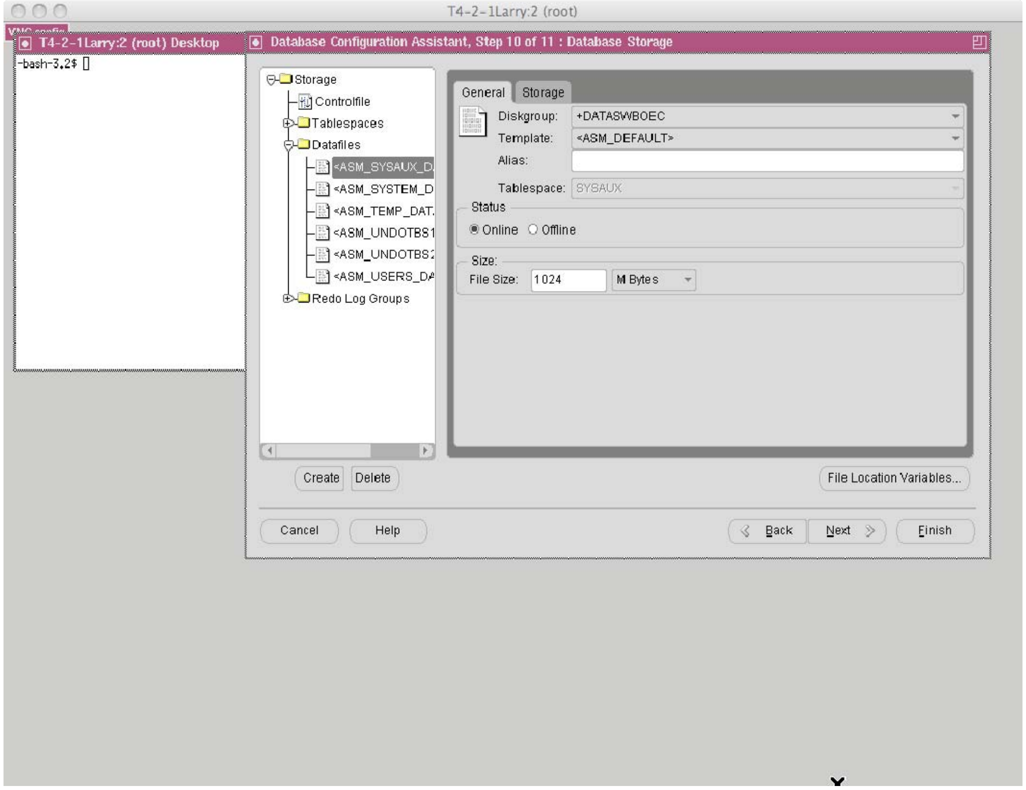
● Custom

Memory Management: Automatic Shared Memory Management

SGA Size: 16384 M Bytes

PGA Size: 2048 M Bytes

Total Memory for Oracle: 18432 M Bytes



Database Configuration Assistant, Step 10 of 11 : Database Storage

```
-bash-3.2$
```

Diskgroup: +DATASWBOEC

Template: <ASM_DEFAULT>

Alias:

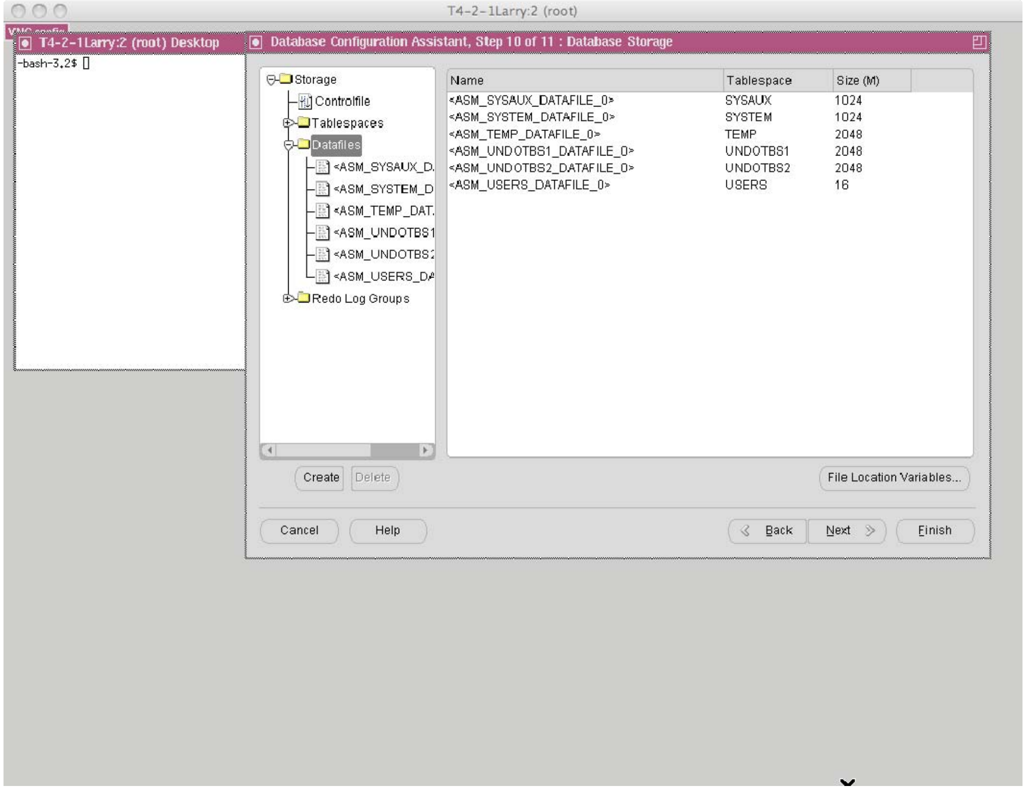
Tablespace: SYSAUX

Status

● Online ○ Offline

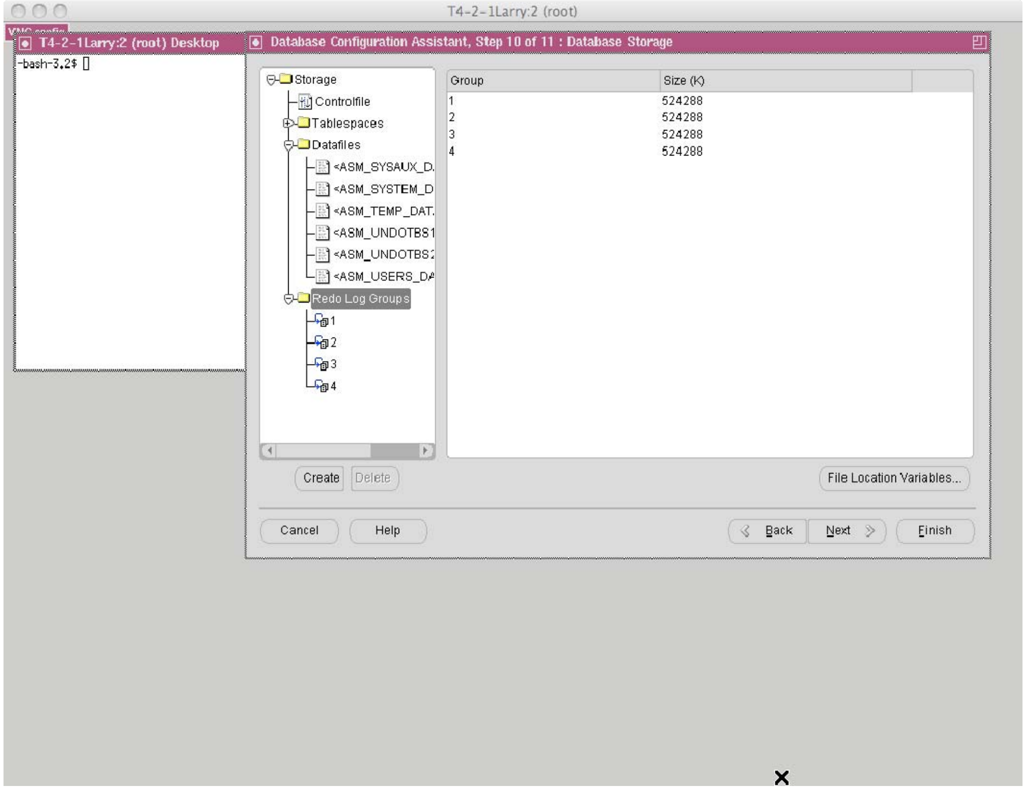
Size:

File Size: 1024 M Bytes



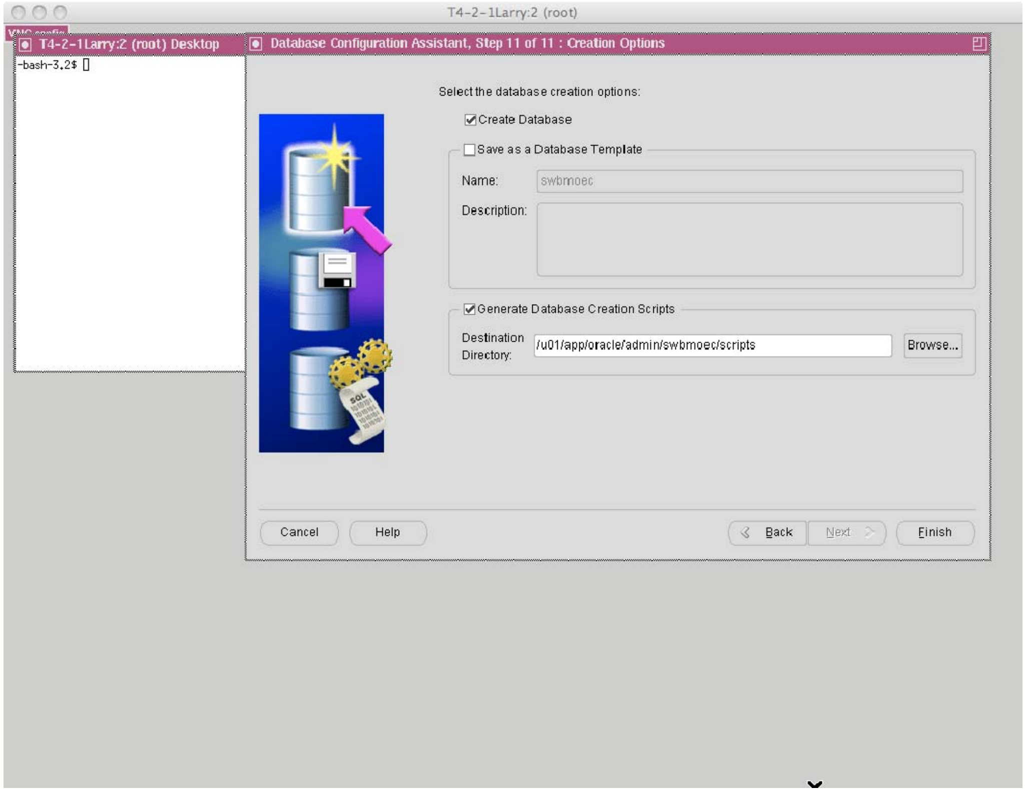
```
-bash-3.2$
```

Name	Tablespace	Size (M)
<ASM_SYSAUX_DATAFILE_0>	SYSAUX	1024
<ASM_SYSTEM_DATAFILE_0>	SYSTEM	1024
<ASM_TEMP_DATAFILE_0>	TEMP	2048
<ASM_UNDOTBS1_DATAFILE_0>	UNDOTBS1	2048
<ASM_UNDOTBS2_DATAFILE_0>	UNDOTBS2	2048
<ASM_USERS_DATAFILE_0>	USERS	16



```
-bash-3.2$
```

Group	Size (K)
1	524288
2	524288
3	524288
4	524288



Database Configuration Assistant, Step 11 of 11 : Creation Options

Select the database creation options:

☑ Create Database

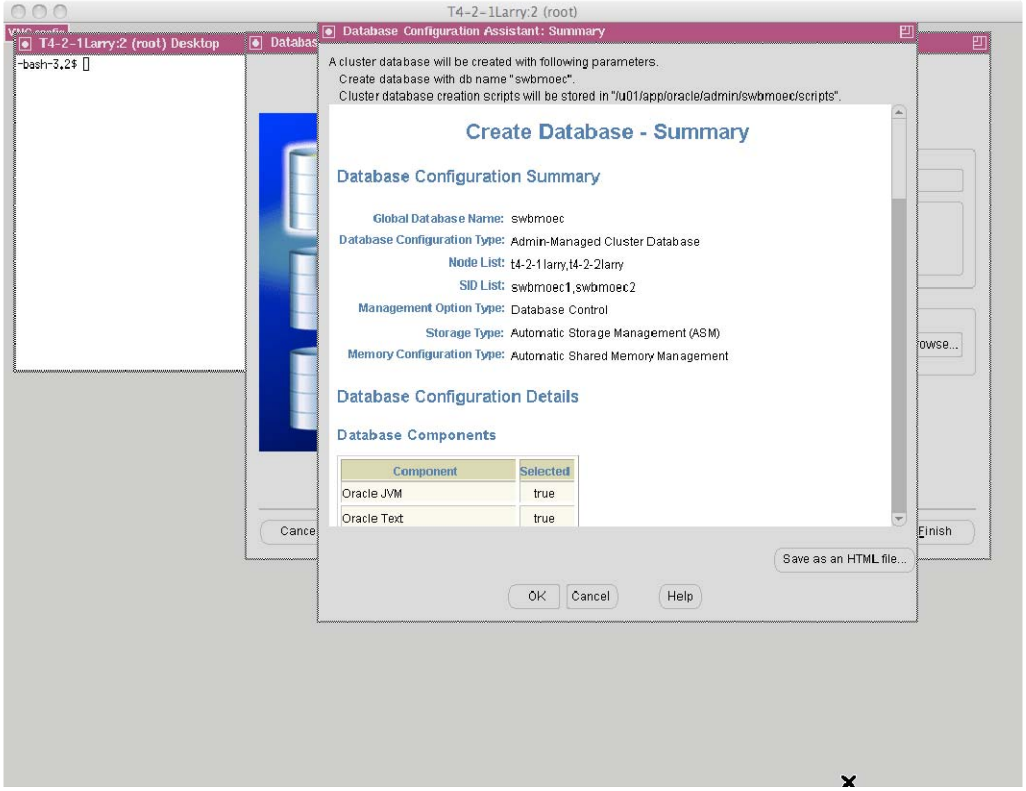
☐ Save as a Database Template

Name: swbmoec

Description:

☑ Generate Database Creation Scripts

Destination Directory: /u01/app/oracle/admin/swbmoec/scripts



Database Configuration Assistant: Summary

A cluster database will be created with following parameters.
Create database with db name "swbmoec".
Cluster database creation scripts will be stored in "/u01/app/oracle/admin/swbmoec/scripts".

Create Database - Summary

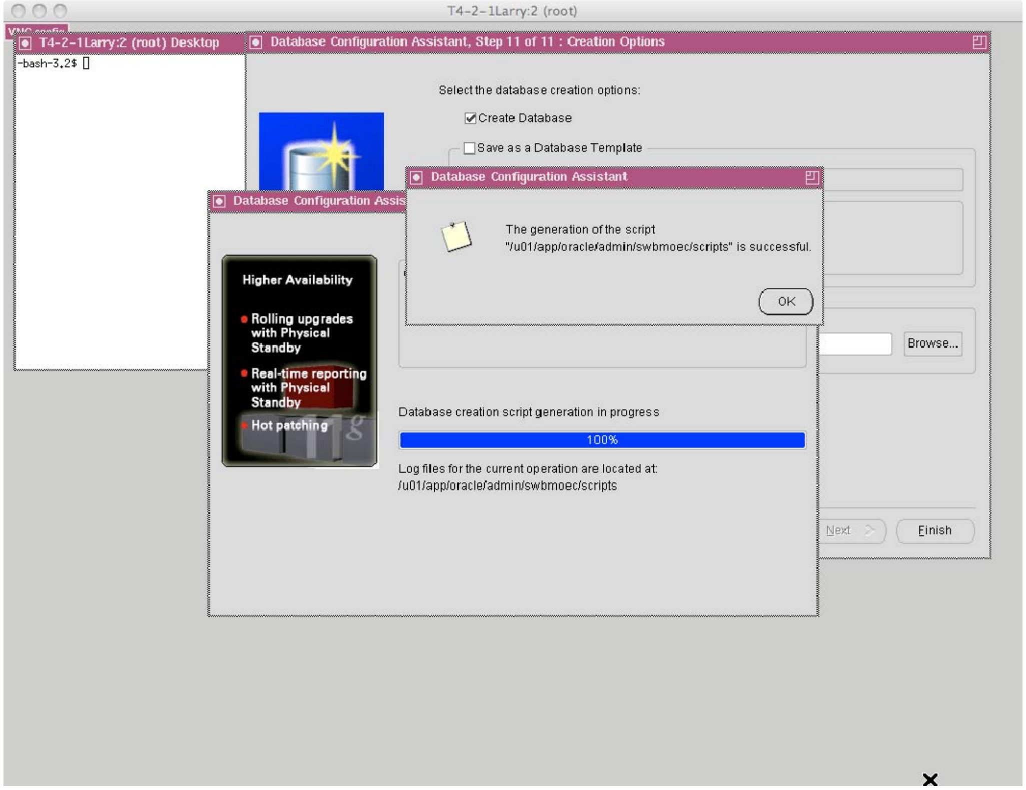
Database Configuration Summary

Global Database Name:	swbmoec
Database Configuration Type:	Admin-Managed Cluster Database
Node List:	t4-2-1larry,t4-2-2larry
SID List:	swbmoec1,swbmoec2
Management Option Type:	Database Control
Storage Type:	Automatic Storage Management (ASM)
Memory Configuration Type:	Automatic Shared Memory Management

Database Configuration Details

Database Components

Component	Selected
Oracle JVM	true
Oracle Text	true


T4-2-1Larry:2 (root)
T4-2-1Larry:2 (root) Desktop
Database Configuration Assistant, Step 11 of 11 : Creation Options
Database Configuration Assis
Database Configuration Assistant
Higher Availability
Rolling upgrades with Physical Standby
Real-time reporting with Physical Standby
Hot patching
11g
OK
Browse...
Next
Finish

```
-bash-3.2$ 
```

Select the database creation options:

☑ Create Database

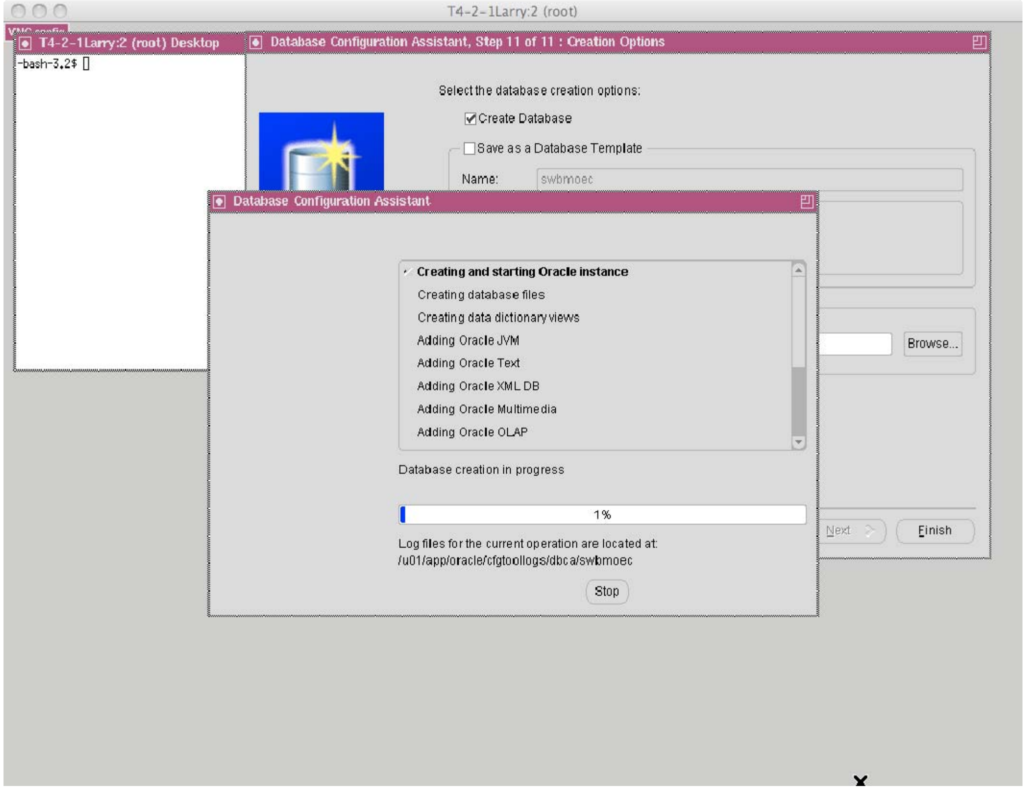
□ Save as a Database Template

The generation of the script "/u01/app/oracle/admin/swbmoec/scripts" is successful.

Database creation script generation in progress

100%

Log files for the current operation are located at:
/u01/app/oracle/admin/swbmoec/scripts


T4-2-1Larry:2 (root)
T4-2-1Larry:2 (root) Desktop
Database Configuration Assistant, Step 11 of 11 : Creation Options
Database Configuration Assistant
Browse...
Next
Finish
Stop

```
-bash-3.2$ 
```

Select the database creation options:

☑ Create Database

□ Save as a Database Template

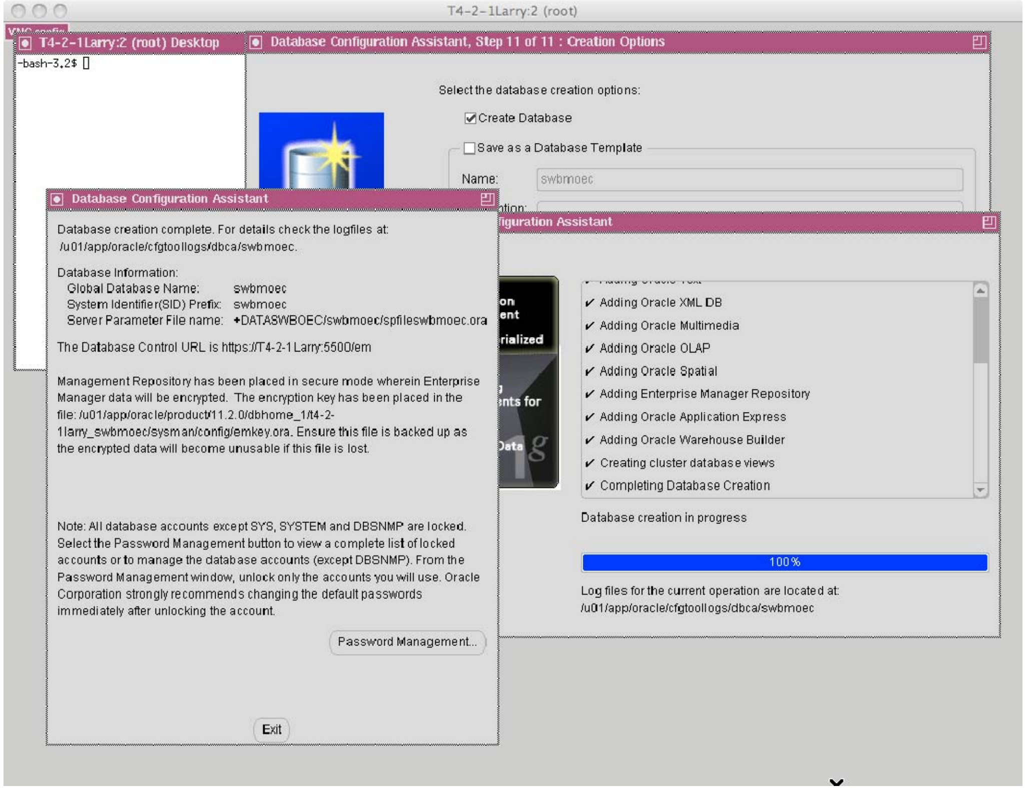
Name: swbmoec

- ✓ Creating and starting Oracle instance
- Creating database files
- Creating data dictionary views
- Adding Oracle JVM
- Adding Oracle Text
- Adding Oracle XML DB
- Adding Oracle Multimedia
- Adding Oracle OLAP

Database creation in progress

1%

Log files for the current operation are located at:
/u01/app/oracle/cfgtoollogs/dbca/swbmoec



T4-2-1Larry:2 (root)
T4-2-1Larry:2 (root) Desktop
-bash-3.2$
Database Configuration Assistant, Step 11 of 11 : Creation Options

Select the database creation options:

☑ Create Database

□ Save as a Database Template

Name: swbmoec

Database Configuration Assistant

Database creation complete. For details check the logfiles at:
/u01/app/oracle/cfgtoollogs/dbca/swbmoec.

Database Information:

Global Database Name:	swbmoec
System Identifier(SID) Prefix:	swbmoec
Server Parameter File name:	+DATASWBOEC/swbmoec/spfileswbmoec.ora

The Database Control URL is https://T4-2-1Larry:5500/em

Management Repository has been placed in secure mode wherein Enterprise Manager data will be encrypted. The encryption key has been placed in the file: /u01/app/oracle/product/11.2.0/dbhome_1/t4-2-1larry_swbmoec/sysman/config/emkey.ora. Ensure this file is backed up as the encrypted data will become unusable if this file is lost.

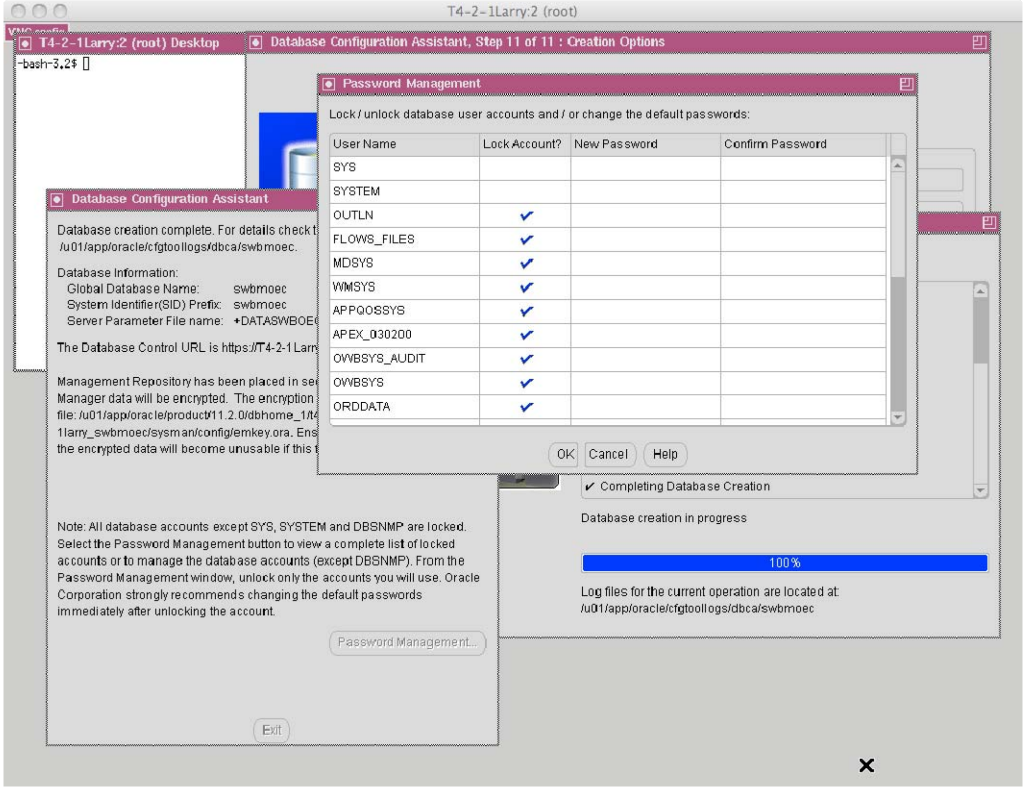
Note: All database accounts except SYS, SYSTEM and DBSNMP are locked. Select the Password Management button to view a complete list of locked accounts or to manage the database accounts (except DBSNMP). From the Password Management window, unlock only the accounts you will use. Oracle Corporation strongly recommends changing the default passwords immediately after unlocking the account.

- ✓ Adding Oracle XML DB
- ✓ Adding Oracle Multimedia
- ✓ Adding Oracle OLAP
- ✓ Adding Oracle Spatial
- ✓ Adding Enterprise Manager Repository
- ✓ Adding Oracle Application Express
- ✓ Adding Oracle Warehouse Builder
- ✓ Creating cluster database views
- ✓ Completing Database Creation

Database creation in progress

100%

Log files for the current operation are located at:
/u01/app/oracle/cfgtoollogs/dbca/swbmoec



T4-2-1Larry:2 (root)
T4-2-1Larry:2 (root) Desktop
-bash-3.2$
Database Configuration Assistant, Step 11 of 11 : Creation Options
Database Configuration Assistant
Password Management

Lock / unlock database user accounts and / or change the default passwords:

User Name	Lock Account?	New Password	Confirm Password
SYS			
SYSTEM			
OUTLN	✓		
FLOWS_FILES	✓		
MDSYS	✓		
WMSYS	✓		
APPQOSSYS	✓		
APEX_030200	✓		
OWBSYS_AUDIT	✓		
OWBSYS	✓		
ORDDATA	✓		

Database creation complete. For details check t
/u01/app/oracle/cfgtoollogs/dbca/swbmoec.

Database Information:
Global Database Name: swbmoec
System Identifier(SID) Prefix: swbmoec
Server Parameter File name: +DATASWBOE

The Database Control URL is https://T4-2-1Larr

Management Repository has been placed in se
Manager data will be encrypted. The encryption
file: /u01/app/oracle/product/11.2.0/dbhome_1/t4
1larry_swbmoec/sysman/config/emkey.ora. Ens
the encrypted data will become unusable if this f

Note: All database accounts except SYS, SYSTEM and DBSNMP are locked. Select the Password Management button to view a complete list of locked accounts or to manage the database accounts (except DBSNMP). From the Password Management window, unlock only the accounts you will use. Oracle Corporation strongly recommends changing the default passwords immediately after unlocking the account.

✓ Completing Database Creation

Database creation in progress

100%

Log files for the current operation are located at:
/u01/app/oracle/cfgtoollogs/dbca/swbmoec

Appendix Prerequisites and Checklists

Prior to deploying Oracle Database 11*g* Release 2 Grid Infrastructure and Oracle Database 11*g* Release 2 on Oracle Solaris, follow all Oracle recommended prerequisites and use the latest documentation that Oracle provides. At the time of this writing, there are a number of items that should be performed, as follows.

1) Verify system requirements:

• System must have >= 4 GB RAM. The requirement will vary based upon how much consolidation and the overall demands are placed upon the system.

```
–/usr/sbin/prtconf | grep “Memory size”
```

•Swap space =RAM (4 – 16 GB RAM) or 0.75xRAM (> 16 GB RAM)

```
–/usr/sbin/swap -s
```

```
–df -h
```

• Provide for > 1 GB/disk for Oracle Clusterware files (using ASM)

• Provide > 5.5 GB disk space for Grid home

• Provide > 1.5 GB for `/tmp`

2) Check network requirements:

• SCAN (Single Client Access Name) IP for the cluster

• Public IP addresses for each database node

• Private IP addresses for each database Node

• One VIP for each database node

• DNS entries for host resolution

3) Check operating system packages:

•Prerequisite Oracle Solaris packages needed:

```
pkginfo -i SUNWarc SUNWbtool SUNWhea \ SUNWlibC SUNWlibm SUNWlibms SUNWsprot \
  SUNWtoo SUNWi1of SUNWi1cs SUNWi15cs \ SUNWxwfnt SUNWcsl \
```

4) Create groups and users:

```
groupadd -g 1000 oinstall
groupadd -g 1031 dba
useradd -u 1101 -g oinstall -G dba -d /export/home/oracle \
-s /usr/bin/bash oracle
mkdir -p  /u01/app/11.2.0/grid
mkdir -p /u01/app/oracle
chown -R oracle:oinstall /u01
chmod -R 775 /u01/
```

5) Check storage ownership. Ensure the slices of the LUNs that will be candidates for use have appropriate ownership and access:

```
format
Snip
c2t2200000B080461B0d40 <Pillar-Axiom600-0000 cyl 51447 alt 2 hd 64 sec 128>
          /pci@500/pci@1/pci@0/pci@0/SUNW,emlxs@0/fp@0,0/ssd@w2200000b080461b0,28
chown oracle:oinstall /dev/rdsk/c2t2200000B080461B0d40s6
chmod 660 /dev/rdsk/c2t2200000B080461B0d40s6
```

6) Ensure the private network RAC InfiniBand interfaces are defined:

```
cfgadm –al
Ap_Id
Snip
hca:212800013E6A7E             InfiniBand-HCA      connected    configured   ok
hca:212800013E6B6E             InfiniBand-HCA      connected    configured   ok
ib                             InfiniBand-Fabric   connected    configured   ok
ib::212800013E6A7F,ffff,ipib   InfiniBand-VPPA     connected    configured   ok
ib::212800013E6B6F,ffff,ipib   InfiniBand-VPPA     connected    configured   ok
ib::daplt,0                    InfiniBand-PSEUDO   connected    configured   ok
ib::iser,0                     InfiniBand-PSEUDO   connected    unconfigured unknown
ib::rdsib,0                    InfiniBand-PSEUDO   connected    configured   ok
ib::rdsv3,0                    InfiniBand-PSEUDO   connected    configured   ok
ib::rpcib,0                    InfiniBand-PSEUDO   connected    configured   ok
ib::sdpib,0                    InfiniBand-PSEUDO   connected    configured   ok
ib::sol_uverbs,0               InfiniBand-PSEUDO   connected    configured   ok


ifconfig –a
Snip
ibd0: flags=1000843<UP,BROADCAST,RUNNING,MULTICAST,IPv4> mtu 2044 index 2
        inet 10.1.0.111 netmask ffffff00 broadcast 10.1.0.255
        ipib 0:0:0:49:fe:80:0:0:0:0:0:0:0:21:28:0:1:3e:6a:7f
```

In closing, ensure that the latest Oracle documentation is reviewed to ensure no steps are missed regarding any updates to hardware or software and to ensure the most up-to-date verification processes are followed prior to and during implementation of both hardware and software provided by Oracle.

ORACLE

Oracle Optimized Solution for
Oracle Database Mission-Critical System
Environments
October 2011, Version 3.0

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

Hardware and Software, Engineered to Work Together