SPARC SUPERCLUSTER

An Oracle Technical White Paper
November 2012

Best Practices for Deploying Oracle Solaris Zones with Oracle Database 11*g* on SPARC SuperCluster

Introduction

Consolidation is a core enabler for deploying Oracle databases on public and private clouds that are built on the Oracle SPARC SuperCluster engineered system.

Complemented with integrated virtualization in the form of Oracle VM Server for SPARC and Oracle Solaris Zones, the SPARC SuperCluster simplifies and accelerates the consolidation of database elements onto a single highly available and scalable platform. Because the entire environment is engineered and optimized to work together, IT organizations can get application and database services up and running faster, whether they are rolling out new deployments or consolidating an older, existing deployment.

This paper provides best practices for setting up and managing consolidated instances of Oracle Database 11*g* Release 2 in a virtualized environment on the SPARC SuperCluster.

SPARC SuperCluster Technologies for Database Consolidation

The SPARC SuperCluster is a complete, pre-engineered, and pretested high-performance enterprise infrastructure solution that is faster and easier to deploy than a collection of individual servers. This general-purpose engineered system combines innovative Oracle technology—the computing power of Oracle's SPARC T4-4 servers, the performance and scalability of Oracle Solaris, the optimized database performance of Oracle Database 11*g* accelerated by Oracle Exadata Storage Servers, and a high-bandwidth, low-latency InfiniBand network fabric—into a scalable, engineered system that is optimized and tuned for consolidating enterprise applications.

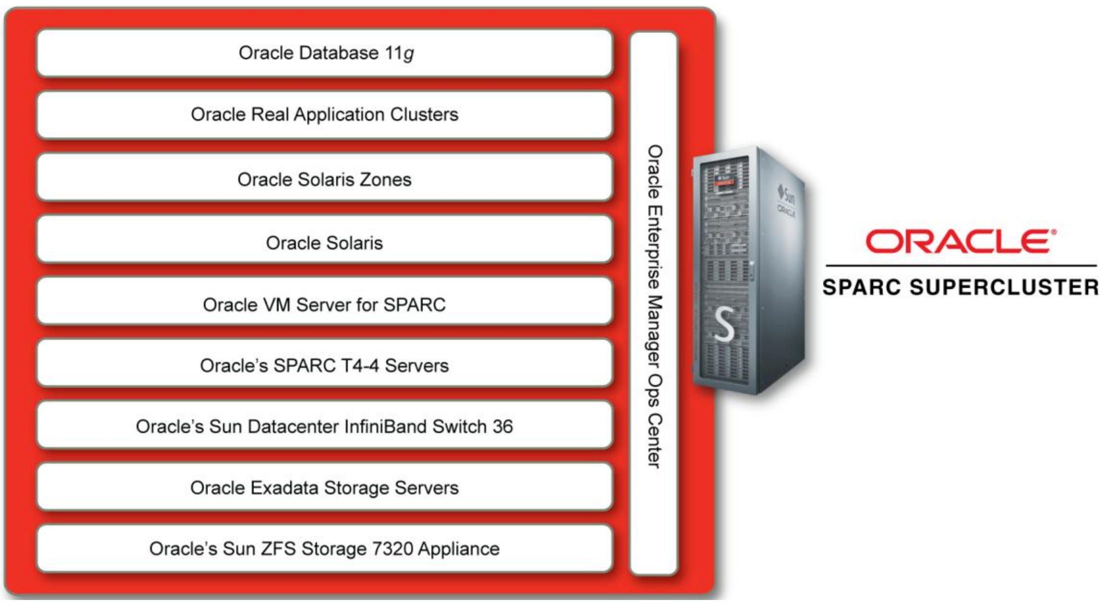


Figure 1. Oracle technology used in the SPARC SuperCluster engineered system.

All components within the SPARC SuperCluster system, including SPARC T4-4 servers, Oracle's Sun ZFS Storage Appliances, and Oracle Exadata Storage Servers, are interconnected over a fully redundant InfiniBand fabric. Built-in virtualization enables consolidation and helps ensure that applications are isolated from one another and remain highly available, virtually eliminating resource contention and service disruption. With the SPARC SuperCluster, the service instances, application servers, and Oracle Database software are all consolidated on the system, eliminating much of the integration effort and deployment time typically associated with clustered solutions and providing other benefits, such as the following:

- Simplified deployment results from consolidation of one or more Oracle databases coupled with the use of Oracle Solaris Zones.
- Lower physical infrastructure complexity and maintenance costs result from deploying fewer physical servers.

- Agile virtualization and configuration of services streamlines business processes.
- Distribution of system resources provides for higher server utilization and lower infrastructure costs.

A full-rack SPARC SuperCluster configuration consists of four SPARC T4-4 servers and six Oracle Exadata Storage Servers, whereas a half-rack configuration consists of two SPARC T4-4 servers and three Oracle Exadata Storage Servers.

Each SPARC T4-4 server contains four SPARC T4 processors, each with eight cores (eight vCPUs per core), operating at 3 GHz clock speed, with 1 TB of memory, six 600 GB internal SAS2 disk drives, and two 300 GB internal SSD disk drives.

Each of the Oracle Exadata Storage Servers in the SPARC SuperCluster are available with 12 high-performance or high-capacity disks. High-performance disks feature 15,000 RPM 600 GB capacity disk drives, whereas high-capacity disks feature 7,200 RPM 3 TB capacity disk drives. In addition, four of Oracle's Sun Flash Accelerator F40 PCIe cards—each with 400 GB space, for a total capacity of 1.6 TB of flash memory—are included in each of the Oracle Exadata Storage Servers. Additional capacity and bandwidth can be achieved with the help of Exadata Storage Expansion Racks.

SPARC T4-4 Domains in the SPARC SuperCluster

The applications that run on the SPARC T4-4 server in the SPARC SuperCluster run in one of three types of logical domains:

- **Database Domain**: A domain dedicated to running Oracle Database 11*g* Release 2, using Oracle Exadata Storage Servers for database storage. Access to Oracle Exadata Storage Servers is provided only through the Database Domain.
- **Application Domains**: One type of Application Domain is dedicated to running applications on Oracle Solaris 11. This domain is well suited to host Oracle Exalogic software in addition to general enterprise applications and middleware. The other type of Application Domain is dedicated to running applications on Oracle Solaris 10.

It is possible and supported to configure multiple domains of the same type on each of the SPARC T4-4 servers in the SPARC SuperCluster.

Database Services Within the SPARC SuperCluster

The database domains in the SPARC SuperCluster provide an optimal solution for all database workloads, ranging from scan-intensive data warehouse applications to highly concurrent online transaction processing (OLTP) applications or mixed application workloads.

By delegating SQL processing to the Oracle Exadata Storage Servers, all the disks can operate in parallel, reducing database server CPU consumption while using much less bandwidth to move data between storage and the database domains.

With the combination of smart Exadata Storage Server Software, Oracle Database software, and the latest industry-standard hardware components, the Database Domain is well suited for consolidating multiple databases into a single grid. Delivered as a complete preoptimized, and preconfigured package of software, servers, and storage, the Database Domain can be implemented quickly and can support large-scale business applications without the need for time-consuming configuration and tuning.

Database Consolidation on the SPARC SuperCluster

Consolidation allows organizations to increase the utilization of IT resources so that idle cycles can be minimized. This in turn lowers costs because fewer resources are required to achieve the same outcome. For example, applications that experience peak load at different times of the day can share the same hardware, rather than using a dedicated hardware that will be idle during the non-peak periods.

Database consolidation can be achieved in many different ways depending on the systems and circumstances involved. Running multiple application schemas in a single database, hosting multiple databases on a single platform, or a hybrid of these two configurations are all valid forms of database consolidation.

Oracle Exadata Storage Servers, when combined with Oracle Database 11*g* Release 2, provide an optimized environment for Oracle Data Warehouse and OLTP database workloads, and the balanced storage grid infrastructure provides an ideal platform for database consolidation. On the other hand, the massive scalability and performance of the SPARC T4-4 server makes it suitable for consolidating different applications from different departments in an organization onto the SPARC SuperCluster system, which enables cost reduction by supporting a higher density of databases through multi-tenancy configurations and makes “Database as a Service” or “Database Cloud” a possibility on the SPARC SuperCluster. Within the Database Domain on each of the SPARC T4-4 servers, Oracle Solaris Zones provide a complete runtime environment to multiple databases. Each zone provides full resource containment and control, fault, operational, and security isolation to ensure database systems do not hamper one another's access to resources or impact execution.

Oracle Solaris Zones on the SPARC SuperCluster

Oracle Solaris provides a unique partitioning technology called Oracle Solaris Zones that enable the creation of a virtualized operating system environment within a single instance of Oracle Solaris. These virtual environments appear to the application and its administrators as complete and dedicated environments, but Oracle Solaris Zones technology restricts the resources available to the environment and isolates the environment from other environments on the same server. This isolation helps enhance security and reliability, since processes running in one zone are prevented from interfering with processes running in another zone. This enables a very efficient and manageable consolidation, making Oracle Solaris Zones technology an excellent foundation for cloud deployments.

Oracle Solaris Zones can be deployed in all three types of domains, including the Database Domain, on the SPARC SuperCluster. The zones must run Oracle Solaris 11 in Database Domains.

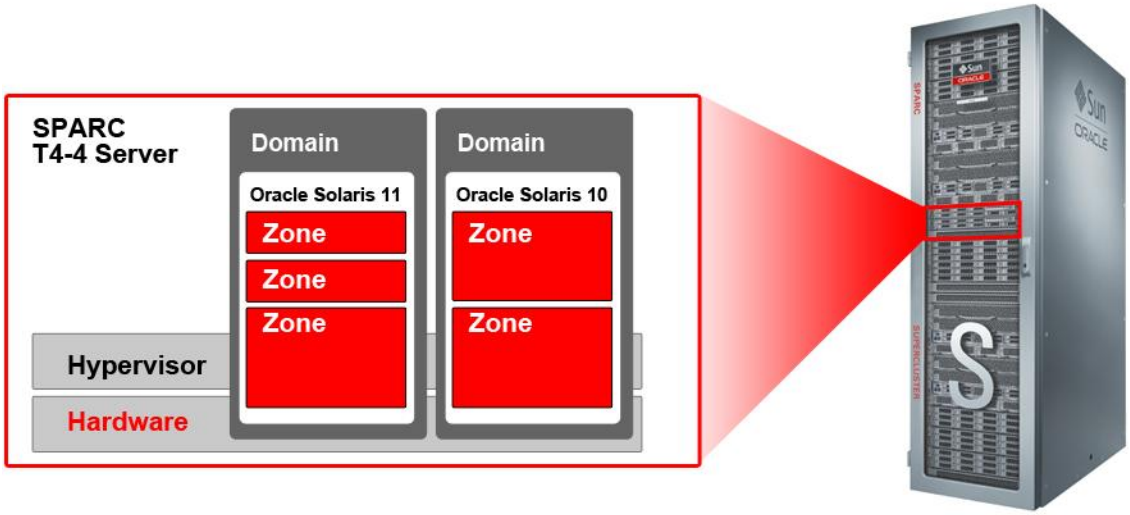


Figure 2. Oracle Solaris Zones on a SPARC T4-4 server.

Resource management tools provided with Oracle Solaris enable administrators to allocate resources, such as CPUs, to specific applications or zones. Each Oracle Solaris Zone contains a complete resource-controlled environment that allows you to allocate resources, such as CPU, memory, networking, and storage. Resource pools provide the capability to separate workloads so that consumption of CPU resources does not overlap. In addition, the dynamic features of resource pools enable administrators to adjust system resources in response to changing workload demands.

In the SPARC SuperCluster, zones offer an additional level of virtualization for consolidating both Oracle Real Application Clusters (Oracle RAC) and non-Oracle RAC databases. It may also be possible to deploy different Oracle Database 11*g* releases and patch levels in different zones, provided those releases are supported with the installed version of Exadata Storage Server Software.

Common Zone Deployment Scenarios

The SPARC SuperCluster offers layered, optimized virtualization. Logical domains form the lower layer, offering domains with near zero performance overhead, and Oracle Solaris Zones comprise the second layer above the domains. The combination allows customers to deploy flexible and effective virtualization to meet their business needs.

Zones provide an excellent deployment environment to satisfy all of the following requirements:

- Database instances require isolation for fault, operation, network, and security manageability and for resource allocation.
- The environment hosting the database instance requires a separate identity, for example, an independent host name and IP address.
- The hosting environment needs to be managed independently of other application environments, for example, booting, shutting down, and patching without impacting other applications.
- More than one Oracle RAC instance is needed in a single database domain.

Best Practices for Deploying Oracle Database 11*g* Using Oracle Solaris Zones

This section outlines best practices and recommendations based on the assumption that multiple databases will be consolidated on the SPARC SuperCluster engineered system using multiple Database Domains and Oracle Solaris Zones on each of the SPARC T4-4 servers.

Before consolidating multiple databases on a SPARC SuperCluster, customers are advised to evaluate the sizing requirements by extrapolating the current CPU, memory, I/O, and network bandwidth utilization rates and obtaining future growth projections from the business. The evaluation should include planning for the number of schemas per database, the Oracle Exadata storage grid (disk group) configuration per database, the number of databases per zone, the number of zones per database domain, the number of Database Domains, and the number of Oracle RAC and non-Oracle RAC database configurations in the SPARC SuperCluster.

It is possible to modify and reconfigure the resource allocations with some downtime later. But it is recommended that customers plan ahead, whenever possible, to gather all the requirements, which will reduce downtime and result in simpler management.

In order to meet changing business needs, based on the available hardware resources, customers are free to grow or shrink the number of zones and Database Domains as long as the deployed applications adhere to the SPARC SuperCluster recommended practices and requirements.

Choosing the Number of Database Domains

SPARC SuperCluster v1.1 supports as many as four database domains per SPARC T4-4 server. Each database domain is assigned its own portion of physical I/O resources for optimal performance. Oracle Solaris 11 is required for all the Database Domains.

Customers are encouraged to plan carefully and choose the number of database domains per SPARC T4-4 server that they might need to host their databases. The number of databases, multi-tenancy or multi-instance configurations, and the requirements of availability and scalability, fault, operational, network, security, and resource isolation are some of the factors that influence the number of database domains in a SPARC SuperCluster.

The recommendation is to spread critical, performance-sensitive databases across the database domains created on multiple SPARC T4-4 servers using Oracle Solaris Zones virtualization.

Oracle Solaris Zones

CPU Allocation

For optimal database performance, the recommendation is to allocate a minimum of one whole CPU core per zone. Partial CPU core allocation is not supported. At this rate, each SPARC T4-4 server with one or more Database Domains in a SPARC SuperCluster can support up to 32 zones When all 32 local zones are deployed, one of the zones must be configured with no CPU resource controls so that one whole CPU core can be shared between the local zone and the global zone. If there is no real need to configure and use 32 zones on a SPARC T4-4 server, allocate more CPU cores to the zones where performance-critical databases are planned to run.

If CPU capping is not enforced for some or all of the zones, it is still not recommended to create more zones than the available number of CPU cores in the Database Domain.

Using the Database Resource Manager features, such as instance caging, database administrators can gain more granular control over the resources allocated for each of the databases at the zone level.

Use the `zonestat`(1) utility to monitor the resource utilization of the currently running zones on any SPARC T4-4 server.

Physical Memory Allocation

With 1 TB of physical memory available on each of the SPARC T4-4 servers in the SPARC SuperCluster, it is possible to allocate a minimum of 32 GB of memory to each of the zones running in Database Domains. To minimize remote memory accesses, it is recommended that physical memory be allocated to each of the zones in multiples of 16 GB to align with the size of the memory banks configured on the system. However, customers are free to allocate memory based on the requirements of the databases running in different zones.

To override the default system-wide limit of 1/4 of physical memory for the maximum shared memory segment size, create and configure an Oracle Solaris project with the desired value for `max-shm-memory` resource control in each of the zones.

The following example configures a maximum of 512 GB for the shared memory segment size, which allows the Oracle Database SGA size to be set to as high as 512 GB. In this example, it was assumed that the `oracle` user will have the required privileges to start the Oracle database.

```
# projadd -p 100 -K "project.max-shm-memory=(priv,512G,deny)" orashmmax
# useradd -g dba -d /export/oracle -s /bin/bash -u 1001 -K project=orashmmax oracle
```

Using Dynamic Intimate Shared Memory

Dynamic Intimate Shared Memory (DISM) is not supported for use on the SPARC SuperCluster's Oracle Solaris environments in instances other than the Oracle Automatic Storage Management instance. If you have a current service contract, you can refer to the *SPARC SuperCluster T4-4 Owner's Guide* for the details and the instructions for disabling DISM.

Network Configuration

The following paragraphs describe the required network configuration to enable running Oracle databases on the SPARC SuperCluster using Oracle Solaris Zones virtualization. On the SPARC SuperCluster, the network configuration is usually set up automatically in zones using a SPARC SuperCluster–specific zone deployment tool, such as `ssc_exavm`. Refer to the "Zone Deployment Using the `ssc_exavm` Tool" section for details about the `ssc_exavm` tool.

Create each Oracle Solaris Zone with two interfaces on the 1 GbE management network, two interfaces on the 10 GbE client access network, and two interfaces on the private InfiniBand (IB) network, as shown in the following example.

```
# dladm show-link
LINK                         CLASS     MTU    STATE    OVER
net2                         phys      1500   up       --
net3                         phys      1500   up       --
n4_1075_bondib0              part     65520   up        ?
n5_1075_bondib0              part     65520   up        ?
net0                         vnic      1500   up        ?
net1                         vnic      1500   up        ?
```

The management network (net2, net3) are virtual network devices (vnet) on the SPARC T4-4 server exposed to the Oracle Solaris Zones as exclusive IP types. In order to facilitate vnets to zones, create multiple vnets in the parent Database Domain using the `addvnet.sh` script.

```
# dladm show-phys |grep vnet
net2 Ethernet up 0 unknown vnet0
net3 Ethernet up 0 unknown vnet1
```

The client access network is exposed to Oracle Solaris Zones using virtual network interfaces (vnic), which are created in the domain as exclusive IP types.

```
# dladm show-vnic
LINK                OVER     SPEED  MACADDRESS        MACADDRTYPE       VID
etc20n1d4z1/net0    net0     10000  2:8:20:bc:55:5b   random            0
etc20n1d4z1/net1    net1     10000  2:8:20:79:17:d2   random            0
```

The private IB network interfaces are IB partitions using the FFFF partition that is used for private interconnect and storage cell communication.

```
# dladm show-vnic
LINK                OVER     SPEED  MACADDRESS        MACADDRTYPE       VID
etc20n1d4z1/net0    net0     10000  2:8:20:bc:55:5b   random            0
etc20n1d4z1/net1    net1     10000  2:8:20:79:17:d2   random            0
```

Using Oracle Solaris IP Multipathing (IPMP), the network interfaces above are grouped into IPMP groups to provide fault-tolerance in active/passive fashion:

```
# ipmpstat -g
GROUP       GROUPNAME   STATE     FDT       INTERFACES
bondib0     bondib0     ok        --        n5_1075_bondib0 (n4_1075_bondib0)
bondeth0    bondeth0    ok        --        net0 (net1)
bondmgt0    bondmgt0    ok        --        net2 (net3)
```

The following is the usage description of the created IPMP interfaces:

- `bondmgto0` is the management network used for administrative access to the Oracle Solaris Zone that is an entry point to each Oracle Solaris Zone via SSH.
- `bondeth0` is the client access network providing access to the database. This network is used by Oracle RAC single client access name (SCAN) listeners and virtual IP addresses for establishing client connections to the database.
- `bondib0` is the private network utilizing the InfiniBand network. This network is used for the private interconnect (usage includes global cache data blocks and heartbeat) and for the communication to the Oracle Exadata Storage Servers.

```
# ipadm show-addr
ADDROBJ            TYPE     STATE        ADDR
lo0/v4             static   ok           127.0.0.1/8
bondeth0/v4        static   ok           10.129.201.224/24
bondmgt0/v4        static   ok           10.129.183.213/24
bondib0/v4         static   ok           192.168.10.75/22
lo0/v6             static   ok           ::1/128
```

IP Address Requirements

The actual number of IP addresses required depends on the number of Oracle RAC and non-Oracle RAC database instances, Oracle Solaris Zones, and Database Domains needed to consolidate the databases in a SPARC SuperCluster environment.

For example, a typical four-node Oracle RAC configuration may need three SCAN addresses for the cluster, four client access addresses on the 10 GbE network, and four management addresses on the 1 GbE network in addition to the IP addresses required by the zones and database domains. Private IP addresses will be used for the IB network interfaces that are used for private interconnect and storage cell communication.

Recommended TCP/IP Tunable Parameters

TABLE 1. TCP/IP TUNABLE PARAMETER VALUES

PARAMETER NAME	VALUE
`tcp_conn_req_max_q`	16384
`tcp_conn_req_max_q0`	16384

The tunable parameter `tcp_conn_req_max_q` sets the default maximum number of pending TCP connections for a TCP listener waiting to be accepted by `accept()`. If it is tuned too high, that can open up the system for a SYN flood DOS attack. 128 is the default value.

The parameter `tcp_conn_req_max_q0` sets the default maximum number of incomplete (three-way handshake not yet finished) pending TCP connections for a TCP listener.1024 is the default value.

Examples:

```
/usr/sbin/ndd -set /dev/tcp tcp_conn_req_max_q 16384
/usr/sbin/ndd -set /dev/tcp tcp_conn_req_max_q0 16384
```

Disk Space Allocation

Each SPARC T4-4 server in the SPARC SuperCluster consists of six 600 GB internal SAS2 disk drives. In a mirrored configuration, approximately 1.8 TB of disk space will be available for all the Oracle Database Domains and the Oracle Solaris Zones that host the Grid Infrastructure software (which is part of Oracle Database) and Oracle Database software.

Assuming 32 zones are populated across multiple Database Domains on a single SPARC T4-4 server to host a large number of database instances, each zone will have access to a little over 50 GB of local disk space, which may or may not be sufficient to hold the zone root file system, Grid Infrastructure, Oracle Database home directory, and any log files and traces generated by the database and other software that the customer may want to install in a zone. Hence, the recommendation is to create sufficiently large iSCSI volumes on the Sun ZFS Storage 7320 Appliance and use those for the zone root and other software, including Oracle Database.

Zone Root on the Sun ZFS Storage Appliance

The Sun ZFS Storage 7320 Appliance in the SPARC SuperCluster provides tens of terabytes of disk space, making it a suitable target to be used as the zone root for Oracle Solaris Zones deployed on SPARC SuperCluster v1.1.

A script, `iscsi-lun.sh`, has been developed to handle the creation and deletion of iSCSI targets, target groups, initiators, and initiator groups and to create LUNs on the Sun ZFS Storage Appliance. If you have a current service contract, you can refer to the "SPARC SuperCluster T4-4 Zones With Oracle Database on Database Domains Configuration Guide" white paper for details.

Another recommended practice is to keep the Oracle Database home directories and the zone root on separate file systems or ZFS data sets wherever possible. This recommendation translates to a minimum of two iSCSI LUNs per Oracle Solaris Zone.

Customers have the flexibility of allocating the available local disk space to any of the zones, as needed. The space allocation need not be uniform across all zones in Database Domains on each of the SPARC T4-4 servers.

ZFS File System

In the SPARC SuperCluster, ZFS is the default root file system for the non-global zones created in Database Domains. The combination of the ZFS file system and the Oracle Solaris Image Packaging System (IPS) provides the ability to precisely control installed packages and the de-duplication of data. Both of these capabilities significantly reduce the impact of multiple zone installations on a system.

Hosting Oracle Database binaries on the ZFS file system enables faster deployment of Oracle Database and faster setting up of another virtual environment by cloning the ZFS file system that is hosting it. Another well-known advantage with ZFS is the ability to grow the file system dynamically and transparently when the file system is nearly full.

Zone Deployment Using the `ssc_exavm` Tool

Oracle Solaris Zones can be created and deployed in Database Domains on the SPARC SuperCluster using tools provided by the operating system, such as `zoneadm`(1M), or by using `ssc_exavm`, a tool specific to the SPARC SuperCluster .

The `ssc_exavm` tool leverages the Deployment Assistant (a feature of Oracle Exadata) and `exavmgen` tools to deploy zones in Database Domains of the SPARC SuperCluster.

At a high level, there are three steps involved in creating and configuring zones using the `ssc_exavm` tool in any of the database domains in a SPARC SuperCluster:

1. Based on user input, such as the number of zones to be created in each of the database domains, the Assistant Questionnaire tool (a feature of Oracle Exadata) generates the configuration scripts.
2. The `exavmgen` tool accepts two inputs—the configuration scripts generated in the previous step along with another input file named `exavmgen.dat`—and generates two output files in XML format.

   Zone creation inputs—such as the zone root, the desired CPU policy (dynamic or static), memory and swap allocations, and interface names for the client access, management access, and IB interfaces—are part of the `exavmgen.dat` file. The `exavmgen.dat` file is provided by means of a template file. Customers just need to edit the template file with actual inputs.
3. The `ssc_exavm` tool creates and configures the zones using the two output files generated in the previous step. In addition to creating the zones, the `ssc_exavm` tool also takes care of setting up the required network configuration and resource allocation and keeps the zones in a state that it is ready for the Oracle Database set up.

The `ssc_exavm` tool creates two ZFS data sets: one for the zone root file system and the other to install the Grid Infrastructure and Oracle Database software.

The `ssc_exavm` tool also eliminates a lot of error-prone manual steps and saves time in creating and preparing the virtual environments (zones) so they are ready for the database setup. Hence, the general recommendation is to use the supported `ssc_exavm` tool instead of the OS tools to deploy zones in the database domains of a SPARC SuperCluster.

The `ssc_exavm` tool will patch the local zones to the same patch level as that of the global zone. A subsequent patching process (package updates) would involve patching of the global zone without disabling the patching of local zones.

Faster Multi-Zone Deployment Using `ssc_exavm`

The zone deployment tool, `ssc_exavm`, provides an option to create a zone template that can be used to create several zones quickly using the ZFS cloning feature. The zone template approach coupled with ZFS cloning can save a significant amount of time.

Operating System Tunable Parameters

In a SPARC SuperCluster environment, the recommended way of installing the required Oracle Database software to set up a two-node or four-node cluster database is to use Oracle's OneCommand utility. When executed, the OneCommand utility automatically tunes the required parameters for the Oracle Database software to function optimally.

However, when setting up multiple Oracle RAC or non-Oracle RAC databases on the SPARC SuperCluster, the required Oracle Database software has to be installed manually. In such a case, the `ssctuner` utility, which is specific to the SPARC SuperCluster, helps set up the required tunable parameters automatically in a SPARC SuperCluster environment.

The `ssctuner` Utility

The `ssctuner` utility runs as an operating system service on all domains including the Database Domains to monitor and tune the following in real time:

- `/etc/system` parameters
- TCP/IP `ndd` parameters
- `/kernel/drv/sd.conf`
- `/kernel/drv/ibd.conf`

In addition, `ssctuner` periodically checks for the use of DISM or suboptimal NFS mount options and modifies those options as needed.

If one or more of the parameters that `ssctuner` manages need to be controlled manually, consider disabling those specific `ssctuner` components.

When installing Oracle Database software manually, ensure that the System V IPC resource controls described in Table 2 were set to the recommended value in each zone or Database Domain where the software is being installed.

System V IPC Resource Controls

If the kernel parameters shown Table 2 were not set by the `ssctuner` utility, use the Oracle Solaris 11 Resource Control Facility to set them. The values set should be greater than or equal to the recommended values shown in Table 2.

TABLE 2. RECOMMENDED VALUES FOR RESORUCE CONTROLS

RESOURCE CONTROL NAME	RECOMMENDED VALUE
`max-file-descriptor`	131072
`max-sem-nsems`	1024
`max-sem-ops`	512
`max-msg-ids`	4096
`max-sem-ids`	65535
`max-shm-ids`	4096
`max-tasks`	131072
`max-shm-memory`	Greater than or equal to the database's SGA size

ZFS Adaptive Replacement Cache (ARC)

In general, ZFS tries to consume as much virtual memory as it can and release it as needed. Limit the ZFS cache to 2 GB in each zone by adding the following line to the `/etc/system` file:

```
set zfs:zfs_arc_max=2147483648
```

Database Grid Configuration

When running clustered (Oracle RAC) or non-clustered database configurations in a SPARC SuperCluster, the required setup for the Grid Infrastructure, which includes Oracle Clusterware and Oracle Automatic Storage Management, is to have one cluster per virtualized environment in which multiple databases can run.

In the case of Oracle Solaris Zones, each zone should have its own Grid Infrastructure irrespective of the number of zones configured per Database Domain on a SPARC T4-4 server in the SPARC SuperCluster. This configuration is simpler to manage and more fault tolerant, and it allows applications to have full access to the database CPU and memory bandwidth, if required or desired.

Each Oracle RAC cluster requires its own disk groups.

Oracle Exadata Storage Grid (Disk Group) Configuration

When running Oracle databases on the SPARC SuperCluster, the recommended storage configuration is to have one Oracle Exadata storage grid for each virtualized environment. A virtualized environment can run only one grid infrastructure, which can run multiple databases. To achieve maximum isolation, allow a maximum of one database per virtualized environment.

An 11*g*R2 storage grid contains only Oracle Exadata Storage Servers for the storage grid and is configured with Oracle Automatic Storage Management high- and/or normal-redundancy disk groups. In general, high redundancy is recommended for the disk groups to provide maximum protection for planned and unplanned outages. Alternatively, though not as secure as high redundancy, it is possible to configure the DATA disk group with high redundancy while having the RECO disk group on normal redundancy or vice versa.

SPARC SuperCluster configurations with no zones can have multiple clusters, but in separate Oracle Database Domains. Each cluster requires its own set of disk groups and can have one or more databases.

Disk Group Isolation

On the SPARC SuperCluster, multiple databases running in separate Oracle RAC clusters can share the disks in all Oracle Exadata Storage Servers, but they cannot share the same Oracle Automatic Storage Management disk groups. In other words, the same set of Oracle Automatic Storage Management disk groups (DATA and RECO) cannot be shared across multiple Grid Infrastructure homes. The requirement is to create a separate pair of DATA and RECO disk groups for each cluster by partitioning the available storage in all Oracle Exadata Storage Servers in such a way that each cluster will have its own storage in the form of unique Oracle Automatic Storage Management disk groups. Storage partitioning has to be completed before installing and configuring the database software in zones.

Allocating disk groups to each application database for DATA and RECO disk groups requires proper planning for space requirements. Once the space requirements are understood for each of the databases, the size for each grid disk can be calculated by dividing the required space by the available number of disks. Each Oracle RAC cluster having its own pair of unique DATA and RECO disk groups implies a similar set of grid disks for each Oracle Automatic Storage Management disk group on each of the Oracle Exadata Storage Servers.

The Oracle Cluster Registry and voting disks for every cluster can be stored in a high-redundancy disk group—a new, unique disk group created just for that purpose or an existing disk group, such as DATA, in the same cluster.

It is a recommended practice to reserve some space on all Oracle Exadata Storage Servers for future growth of existing databases and to consolidate additional databases.

In order to achieve a high level of I/O performance, configure each disk group using equal sized grid disks on every Oracle Exadata Storage Server. Also, while creating grid disks, specify a lower offset for the DATA disk groups so that the grid disks will be created on the outer tracks of the physical disks which, consequently, results in slightly improved I/O performance.

Disk group isolation is achieved either by sharing storage across multiple Oracle RAC clusters or by dedicating the storage to an Oracle RAC cluster, as explained in the next section.

Shared Storage Approach

One way to achieve disk group isolation is by creating Oracle Automatic Storage Management disk groups on different grid disks that span across all Oracle Exadata Storage Servers in a SPARC SuperCluster configuration. This shared approach lowers administration costs by reducing the number of storage systems that need to be maintained. Another benefit is that sharing leads to efficient use of storage, both from a space and a bandwidth standpoint.

Performance problems that arise by running multiple databases on shared storage can be mitigated by using I/O Resource Management (IORM). I/O Resource Management allows workloads and databases to share Oracle Exadata Storage Servers according to user-defined policies. To manage workloads within a database, administrators can define database resource plans using Database Resource Manager. To manage multiple databases, define inter-database plans.

The rest of this section illustrates the shared storage approach with an example.

Let us assume that an organization is consolidating four application databases on a full-rack SPARC SuperCluster using Oracle RAC clusters with a few other databases to be consolidated at a later date. With the assumption that one of the applications, APP2, store files in the database, raw disk space requirements for the initial four databases are shown in Table 3.

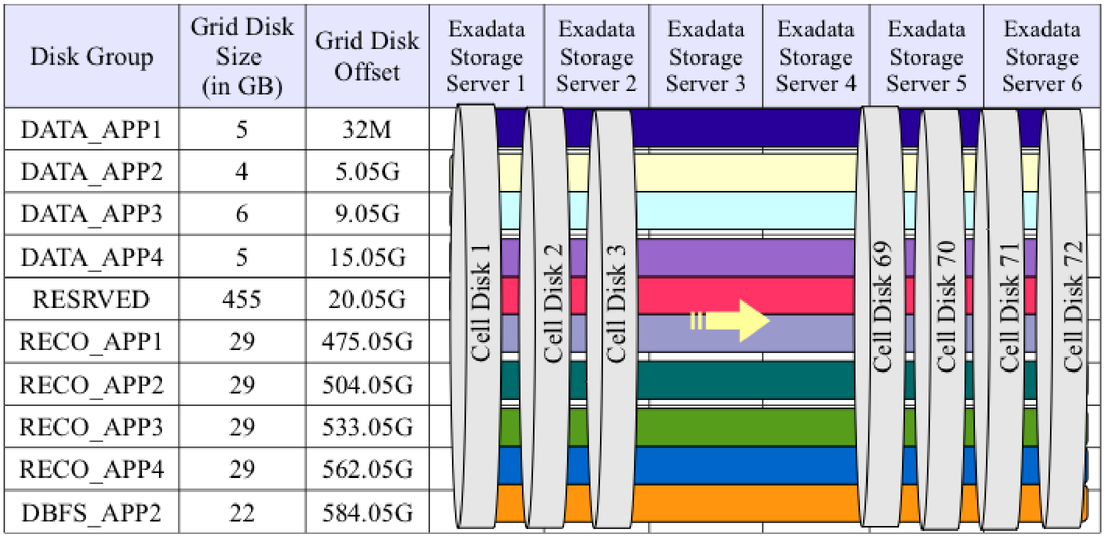
The disk space requirements take into account the high redundancy for DATA and the normal redundancy for RECO and DBFS by appropriately multiplying the actual requirement by 3 and 2, respectively. Unless constrained by the available disk space, files can be stored in the DATA disk group of the cluster. For demonstration purposes, this example proceeds with the assumption that the APP2 application's files are stored in a DBFS disk group in the same cluster.

TABLE 3. EXAMPLE OF SHARED STORAGE APPROACH

APPLICATION	DATA DISK GROUP SIZE (IN GB)		RECO DISK GROUP SIZE (IN TB)	DBFS DISK GROUP SIZE (IN TB)	TOTAL (IN TB)
	ACTUAL	WITH ~10% PROJECTED GROWTH			
APP1	270	300	2	0	2.3
APP2	225	250	2	1.5	3.75
APP3	375	410	2	0	2.41
APP4	300	330	2	0	2.33
TOTAL					10.8

In this example, there are six Oracle Exadata Storage Servers, each with twelve 600 GB SAS high-performance disks for a total capacity 43 TB of raw storage. The application raw disk space requirements include a 10 percent growth factor for the existing databases. With these space requirements, it was determined that the full-rack SPARC SuperCluster has sufficient disk space to host four clusters.

Dividing each of the raw sizes by the total number of available disks in the full-rack SPARC SuperCluster (72) will give the required size for each grid disk, as shown in Figure 3, which depicts the disk group layout. To keep it simple, disk groups that hold each cluster's Oracle Cluster Registry and voting disks were omitted, and only the disk groups pertaining to application databases are shown in Figure 3.



Disk Group	Grid Disk Size (in GB)	Grid Disk Offset	Exadata Storage Server 1	Exadata Storage Server 2	Exadata Storage Server 3	Exadata Storage Server 4	Exadata Storage Server 5	Exadata Storage Server 6
DATA_APP1	5	32M						
DATA_APP2	4	5.05G						
DATA_APP3	6	9.05G						
DATA_APP4	5	15.05G						
RESRVED	455	20.05G						
RECO_APP1	29	475.05G						
RECO_APP2	29	504.05G						
RECO_APP3	29	533.05G						
RECO_APP4	29	562.05G						
DBFS_APP2	22	584.05G						

Figure 3. Example disk group layout.

Based on these sizing requirements, the grid disks and then the Oracle Automatic Storage Management disk groups are created using the `create griddisk` and `create diskgroup` commands.

The following sample commands illustrate the steps involved in creating the grid disks and disk groups for the APP1 and APP2 cluster databases specified in the example. Similar commands can be used to create the grid disks and disk groups for the remaining databases, APP3 and APP4.

Step 1: Create the grid disks for each cluster using the CellCLI interface on all the Oracle Exadata Storage Servers:

```
cellcli -e create griddisk all harddisk prefix=DATA_APP1, size=5G, offset=32M
cellcli -e create griddisk all harddisk prefix=DATA_APP2, size=4G, offset=5.05G
cellcli -e create griddisk all harddisk prefix=RESRVED, size=455G, offset=20.05G
cellcli -e create griddisk all harddisk prefix=RECO_APP1, size=29G, offset=475.05G
cellcli -e create griddisk all harddisk prefix=RECO_APP2, size=29G, offset=504.05G
cellcli -e create griddisk all harddisk prefix=DBFS_APP2, size=22G, offset=584.05G
```

Step 2: As a user with the `asmadmin` role, on any Oracle Automatic Storage Management instance, create the Oracle Automatic Storage Management disk groups for each cluster:

```
create diskgroup DATA_APP1 high redundancy disk 'o/*/DATA_APP1*'
       attribute 'content.type'='data','cell.smart_scan_capable'='true',
       'compatible.rdbms'='11.2.0.2','compatible.asm'='11.2.0.3','au_size'='4M';

create diskgroup DATA_APP2 high redundancy disk 'o/*/DATA_APP2*'
       attribute 'content.type'='data', 'cell.smart_scan_capable'='true',
       'compatible.rdbms'='11.2.0.2','compatible.asm'='11.2.0.3','au_size'='4M';

create diskgroup RECO_APP1 normal redundancy disk 'o/*/RECO_APP1*'
       attribute 'content.type'='recovery', 'cell.smart_scan_capable'='true',
       'compatible.rdbms'='11.2.0.2','compatible.asm'='11.2.0.3','au_size'='4M';

create diskgroup RECO_APP2 normal redundancy disk 'o/*/RECO_APP2*'
       attribute 'content.type'='recovery', 'cell.smart_scan_capable'='true',
       'compatible.rdbms'='11.2.0.2','compatible.asm'='11.2.0.3','au_size'='4M';

create diskgroup DBFS_APP2 normal redundancy disk 'o/*/DBFS_APP2*'
       attribute 'content.type'='system', 'cell.smart_scan_capable'='true',
       'compatible.rdbms'='11.2.0.2','compatible.asm'='11.2.0.3','au_size'='4M';
```

Note that the reserved cell disk space was allocated as grid disks between DATA and RECO, but no associated Oracle Automatic Storage Management disk groups were created at this point; that was left as is for future considerations.

Dedicated Storage Server Approach

Another way to achieve disk group isolation is by dedicating Oracle Exadata Storage Servers to the Oracle Automatic Storage Management disk groups of a cluster. In order to ensure the ability to restore full redundancy after a storage server failure, a minimum of three Oracle Exadata Storage Servers is recommended for the cluster's Oracle Automatic Storage Management disk groups.

When a storage system is dedicated to a single database, it results in predictable performance and isolation of disk groups and application databases. However, this practice may lead to unused I/O bandwidth and space for some databases if the database's peak load and size are not properly estimated. Also, it limits to a very few the number of Oracle RAC databases that can be consolidated on the SPARC SuperCluster.

For example, on a full-rack SPARC SuperCluster, two Oracle RAC clusters can be configured: one with the Oracle Automatic Storage Management disk groups DATA1, RECO1, and a unique disk group to hold the Oracle Cluster Registry and voting disks residing on Oracle Exadata Storage Servers 1 to 3, and the second with the Oracle Automatic Storage Management disk groups DATA2, RECO2, and another unique disk group to hold Oracle Cluster Registry and voting disks residing on Oracle Exadata Storage Servers 4 to 6. An external Exadata Storage Expansion Rack may help in creating more Oracle RAC clusters in a dedicated storage server approach.

Oracle Database 11*g* Release 2 supports a maximum number of 63 Oracle Automatic Storage Management disk groups per storage grid.

Resizing Grid Disks

It is not uncommon to run into situations that may ultimately lead to the resizing of storage grid disks. For example, an under-allocated DATA disk group in a cluster may need to be resized to increase the allocated disk space to meet the demand for more space for application data.

If you have a current service contract, you can refer to the “Resizing Storage Grid Disks” section under the “Maintaining Oracle Exadata Database Machine and Oracle Exadata Storage Expansion Rack” chapter in *Oracle Exadata Database Machine Owner's Guide 11g Release 2 (11.2)* for an overview of different methods and instructions for resizing the grid disks.

Oracle RAC Configuration

On a half-rack SPARC SuperCluster with two SPARC T4-4 servers, it is possible to create multiple two-node Oracle RAC databases, whereas on a full-rack SPARC SuperCluster, it is possible to create multiple two-node or four-node Oracle RAC databases.

The recommendation is to use zones of identical configuration in creating two- or four-node Oracle RAC databases. For example, when creating a two-node clustered database in a half rack SPARC SuperCluster, if zone 1 (Z1) in the database domain (D1) on SPARC T4-4 server node 1 (N1) is identical in configuration to zone 1 (Z1) in the database domain (D1) on the second SPARC T4-4 server node (N2), configure Z1D1N1 and Z1D1N2 as the database nodes in the two-node Oracle RAC database configuration.

It is recommended that the `ssc_exavm` and OneCommand utilities be used to deploy a single two-node or four-node Oracle RAC database using zones in Database Domains on the SPARC SuperCluster.

Note that the OneCommand utility can be used only when the plan is to deploy only one Oracle RAC database in a SPARC SuperCluster environment. The OneCommand utility is not suitable for deploying multiple Oracle RAC deployments in the SPARC SuperCluster. Hence, multi-cluster deployments using the OneCommand utility are not supported.

When multiple Oracle RAC clusters are planned in the SPARC SuperCluster, plan and partition the available disk space to create a unique set of Oracle Automatic Storage Management disk groups for each cluster before installing the required Oracle Database software in respective zones running in Database Domains.

Follow the *Oracle Real Application Clusters Installation Guide 11g Release 2 (11.2) for Linux and UNIX* to set up Oracle RAC on the SPARC SuperCluster. The Oracle RAC installation and configuration steps have to be repeated for each of the clustered database setups.

Securing the Databases in SPARC SuperCluster

The SPARC SuperCluster platform provides a comprehensive set of security controls that can be employed to meet the challenging security demands of database deployments. It is critical to understand that each capability offers an opportunity to be layered with the others to create reinforced security postures for the target database deployments. These capabilities are grouped into four distinct categories: secure isolation, access control, cryptographic services, and monitoring and auditing.

Secure Isolation

The SPARC SuperCluster platform enables secure isolation at the workload, network, database, and storage levels, allowing organizations the flexibility to implement various isolation policies and strategies based upon their target database deployment.

The SPARC SuperCluster platform supports a variety of workload isolation strategies for secure database deployments, each with its own set of capabilities and trade-offs. The recommended isolation strategies are based on adopting techniques such as POSIX, Oracle Solaris Zones, Oracle VM Server for SPARC domains or a hybrid approach combining different techniques.

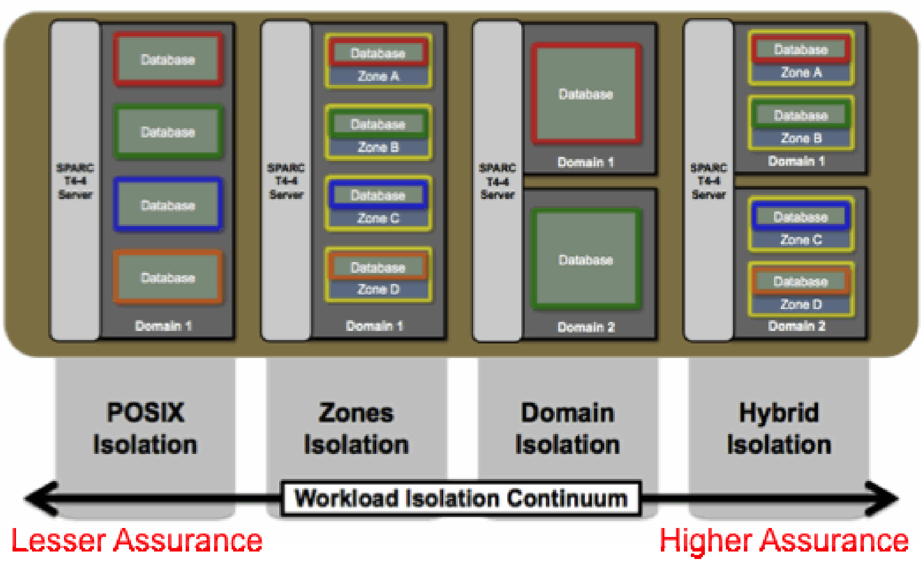


Figure 4. Techniques for secure workload isolation.

In POSIX isolation, multiple databases run on a single operating system, each with its own POSIX credentials (a UNIX user and a group identifier) that are used by the underlying operating system when making access control decisions. This ensures that the processes and the files of a database are protected from other users and databases that may exist on the same system.

Oracle Solaris Zones enable deploying individual databases or groups of databases into one or more virtualized containers that run on top of a single operating system image. This approach improves upon POSIX isolation by creating a stronger security boundary around its deployed services. The Oracle Solaris kernel ensures that zones operate with reduced privileges. As a result, database services running in a zone will not be able to see, manipulate, or adversely impact other services that may be deployed on the system. The use of Oracle Solaris Zones is strongly recommended for service isolation and containment regardless of other methods used.

Oracle VM Server for SPARC domains provide a hypervisor-mediated isolation that can be used to securely isolate database instances. Oracle VM Server for SPARC domains are created as Database Domains in the SPARC SuperCluster to facilitate database deployments, and each runs its own unique instance of the Oracle Solaris operating system. Access to physical resources is mediated by the built-in hypervisor in SPARC T4 processors. This ensures that the database services running in a domain are logically isolated from the database services running in other domains.

Database deployments in the SPARC SuperCluster can use a hybrid approach, leveraging POSIX, zones, and database domains to create resilient isolation architectures. With a combination of these technologies, organizations can limit the number of operating system instances that have to be managed while at the same time increasing database density without having to sacrifice performance or security.

Access Control

The SPARC SuperCluster platform supports a flexible, layered access control architecture covering every layer of the stack and supporting a variety of roles including end users and system and database administrators. This enables organizations to define policies that protect databases and the underlying compute, storage, and network infrastructure on which those services run. At the operating system and virtualization layers, access control begins with a reduced set of services being exposed to the network. By reducing the number of entry points through which a system can be accessed, the number of access control policies can also be reduced and more easily maintained over the life of the system. Within the Oracle Solaris operating system, access controls are implemented using a combination of POSIX permissions and the Oracle Solaris role-based access control (RBAC) facility.

The POSIX user and group identifiers form the basis for the most primitive access control decisions in the Oracle Solaris operating system. Assigned to each user is a unique user identifier, a primary group identifier and, optionally, additional groups to which the user belongs. Based on these values, the user is able to perform a variety of process and file management operations. For this reason, it is recommended that databases with different database administrators operate using distinct POSIX credentials. This helps to ensure that databases and their database administrators do not have overlapping permissions to access each other's processes, files, and services. Oracle Solaris also includes a fine-grained, role-based access control facility. Oracle Solaris RBAC is a collection of capabilities that enable grouping authorizations, operations, and privileges. These rights can be selectively assigned to individuals based on the need to satisfy least-privilege requirements.

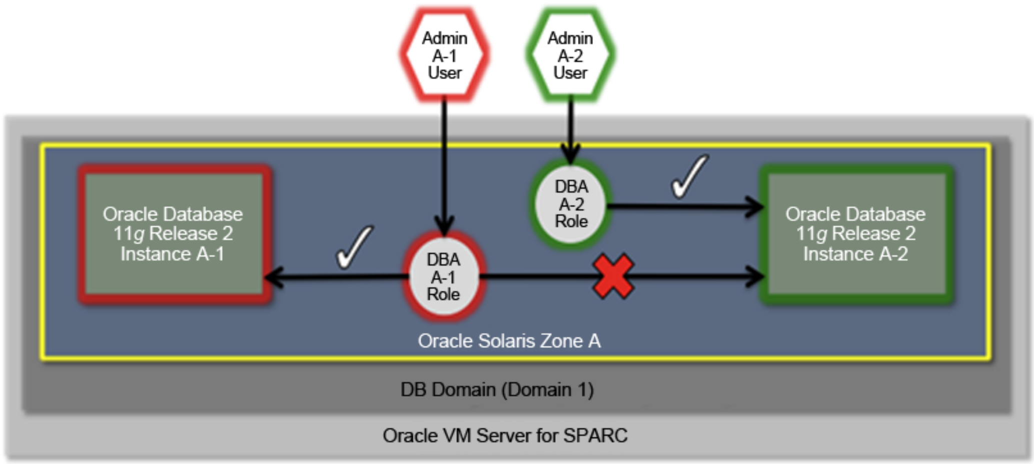


Figure 5. Example scenario of using Oracle Solaris RBAC.

When consolidating databases onto the SPARC SuperCluster, Oracle Solaris RBAC should be used to segment the rights and privileges of individual database and system administrators. In Figure 6, for example, Oracle Solaris roles are used to group the rights and privileges associated with two different database administrators. Each administrator is able to manage only their own respective databases through a combination of POSIX and role-based access controls. This approach offers organizations a scalable way to manage access because multiple users can be assigned to roles to centralize the access management policy while promoting strong accountability and least privilege.

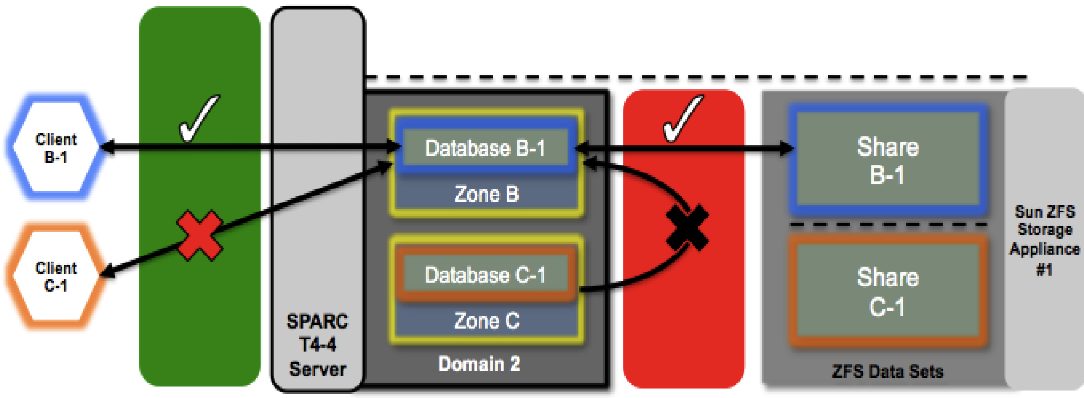


Figure 6. Example scenario of using Oracle Solaris roles to group rights and privileges.

In addition, organizations can also choose to implement a host-based firewall, such as the Oracle Solaris IP Filter service, to control access to the network services. For example, the IP Filter supports stateful packet filtering to filter packets by IP address, port, protocol, network interface, and traffic direction. These capabilities are useful in a SPARC SuperCluster where variety of inbound and outbound network traffic flows through many network interfaces. On the SPARC SuperCluster, IP Filter can be configured in Database Domains or it can be operated from within an Oracle Solaris Zone. This allows the network access control policy to be enforced in the same operating system container where database services are offered.

For example, each of the Oracle Solaris Zones shown in Figure 6 can communicate with the Sun ZFS Storage Appliance using IP over InfiniBand across a shared InfiniBand partition. In a database consolidation scenario, traffic flowing over this partition would be expected to be limited to database domains or zones communicating with the Sun ZFS Storage Appliance, likely over the NFS protocol. Consequently, organizations can easily install a network access control policy that limits the communication accordingly. This simple step will ensure that an individual domain or zone is not able to communicate with others over this shared IP network. However, they will still be able to initiate communication and use the NFS services of the Sun ZFS Storage Appliance.

Another reason why organizations should consider host-based firewall technologies is to limit outbound communication. While this scenario is not often considered, it is important to be able to limit outbound network communications because they can be used as a means of subverting more traditional inbound filtering practices. In a database consolidation scenario, the amount of outbound network activity will likely be minimal and can easily be categorized so that a policy can be created that limits communications to specific network interfaces and destinations. All other traffic would be denied and logged as part of a "default deny" policy to block unauthorized communications, both inbound and outbound.

Cryptographic Services

The SPARC SuperCluster platform has been designed to support the high-performance encryption needs of modern IT environments. For consolidated database architectures, where cryptography figures into nearly every aspect of the architecture, the SPARC SuperCluster platform and its supporting software enable organization's database deployments to meet the security and compliance objectives without having to sacrifice performance.

Workload and Network Cryptographic Services

The compute nodes in the SPARC SuperCluster utilize the SPARC T4 processor, which enables the acceleration of 16 industry-standard cryptographic algorithms. Together, those algorithms support cryptographic needs including encryption, random number generation, and the calculation and verification of digital signatures and message digests. All of these capabilities are made available to databases and services running on the Oracle Solaris operating system to protect information at rest, in transit, and in use.

In consolidated database architectures, each database leverages these capabilities to easily support strong cryptographic protections, including defenses against cache-based and timing side-channel attacks, for its communications and data without sacrificing performance. In addition, at the operating system level, cryptographic hardware acceleration is enabled by default for most of the core services, including Secure Shell, IPsec/IKE, and ZFS. This allows organizations to leverage improved security and performance not only for their core database services but also for any administrative or support activities that must be completed on those servers, even when they are run inside an Oracle Solaris Zone.

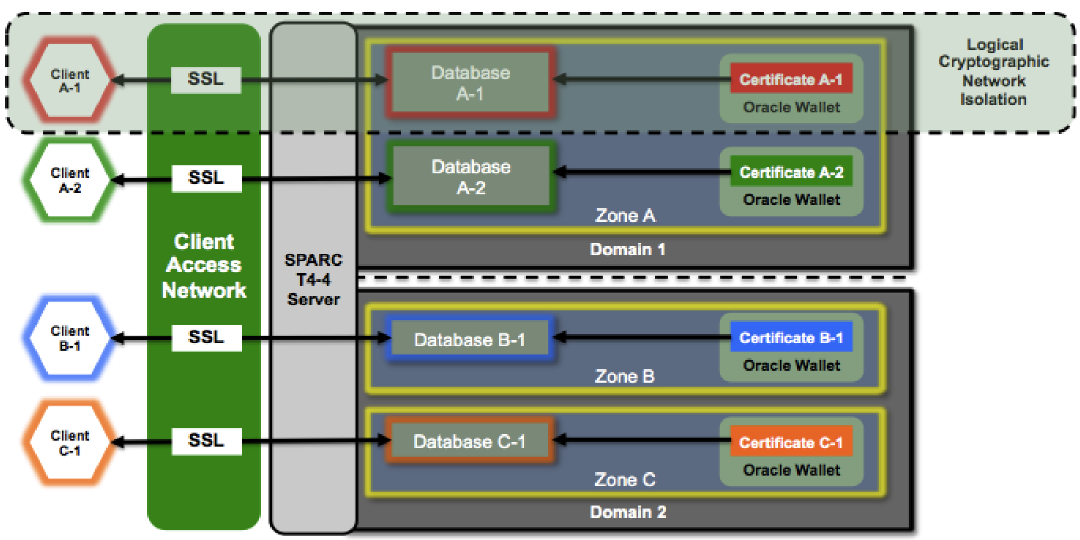


Figure 7. Example configuration of logical cryptographic network isolation.

Given the importance of privacy and compliance mandates, organizations considering consolidated database architectures should strongly consider the use of cryptography to protect the information flowing to and from their databases. This will ensure that data is not exposed to unauthorized parties. The Oracle Advanced Security option supports both Oracle native and SSL/TLS encryption methods to protect information in transit. Further, using individual SSL certificates for each instance or cluster allows organizations to essentially create logical cryptographically isolated network boundaries that protect data even when it must flow over a shared network or interface.

Database Cryptographic Services

Once information has been received over the network, it is then processed and stored by the database. This information at rest must also be protected to help organizations comply with their security policies and compliance mandates. The Oracle Advanced Security option also includes functionality to safeguard information at rest. Specifically, the Transparent Data Encryption capability of Oracle Database allows organizations to seamlessly encrypt application data using one of two methods: column-based encryption and tablespace encryption.

On the SPARC SuperCluster, Oracle Database software enables the database to automatically use the underlying hardware cryptographic acceleration capabilities for tablespace encryption without the need for additional end-user configuration.

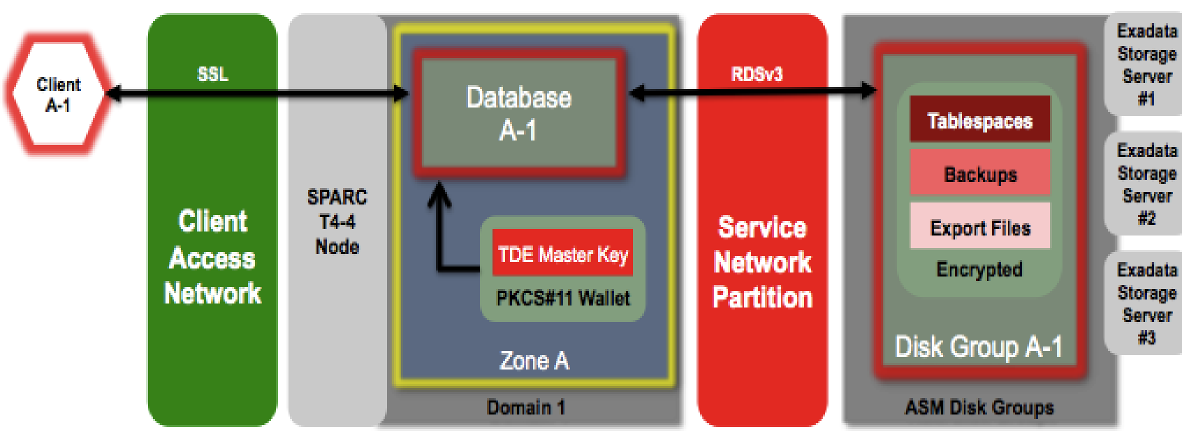


Figure 8. Example of database cryptographic services.

On the SPARC SuperCluster, Transparent Data Encryption can also leverage an Oracle Solaris PKCS#11 wallet as a key store to securely protect the master key. Using the Oracle Solaris PKCS#11 wallet will automatically engage the SPARC T4 hardware-assisted cryptographic acceleration for any master key operations. This allows the SPARC SuperCluster platform to significantly improve the performance of encryption and decryption operations associated with database backups using Oracle Recovery Manager (Oracle RMAN), exports using Oracle Data Pump, and redo logs using Oracle Active Data Guard. When using Oracle RAC to support availability and performance objectives, it is necessary to either copy or share the Oracle wallet so that each database instance in a cluster has access to the same master key. Organizations using a shared-wallet approach can leverage the storage subsystems in the SPARC SuperCluster platform, such as Oracle Automated Storage Management Cluster File System or a data set shared over NFS from the Sun ZFS Storage Appliance, to create a directory that can be shared across all of the nodes in a database cluster. Using a shared, centralized key store can help organizations better manage, maintain, and rotate the keys in clustered database architectures, because the keys will be synchronized across each of the nodes in the cluster.

Monitoring and Auditing

The SPARC SuperCluster platform provides the ability to monitor actions taken by users and administrators as well as activity detected by system services. The comprehensive monitoring capabilities of the SPARC SuperCluster platform enable organizations to obtain monitoring and audit information at the workload, network, database, and storage layers, which helps to ensure that activity can be detected no matter where it happens.

Oracle Solaris Audit Facility

Actions taken by users and services on the Oracle Solaris operating system are recorded by the Oracle Solaris audit facility. Enabled by default in the Oracle Solaris 11 operating system, the Oracle Solaris audit facility enables the creation of flexible policies that can track system, administrator, and end-user activity. For example, the Oracle Solaris audit facility can record system boot and shutdown requests, system console and network logins, attempts to perform privileged activities, and even more granular system level functions such as those related to process and file creation or deletion.

The Oracle Solaris audit facility can be configured at a system or per-user level, allowing organizations to set system-level policies that best reflect their requirements. Organizations can choose to record all actions regardless of the outcome or only those that succeed or fail. The Oracle Solaris audit facility includes metadata for each record produced, including the real and effective user and group identifier of the actor, the action attempted and whether it was successful, the date and time of the event, and a wealth of other information. Configured well, the Oracle Solaris audit facility is a critical capability for organizations attempting to develop a baseline for normal typical system activity, detect unauthorized activity or attempts to circumvent security controls, and collect evidence in support of audit or incident response requirements. Because Oracle VM Server for SPARC services and Oracle Solaris Zones each run on the Oracle Solaris operating system, events associated with the configuration, management, and usage of the Database Domains and zones are also audited by this facility.

The Oracle Solaris audit facility includes information related to network communications as well in bare metal Oracle Solaris operating system, Database Domains, and Oracle Solaris Zones. For example, the audit facility can record information related to incoming and outgoing network connections, including when services listen for and accept new connections for system or database services. By centralizing this information within the Oracle Solaris audit facility, this information can be more readily correlated with other system activities to develop a more complete picture of a situation.

In addition, the Oracle Solaris IP Filter capability can selectively record both inbound and outbound network communications.

For more information on security guidelines for database consolidation scenarios, refer to the Oracle technical white paper titled "Secure Database Consolidation Using the SPARC SuperCluster T4-4 Platform."

Oracle Exadata Database-Scoped Security

By default, all database and Oracle Automatic Storage Management instances have access to all storage cell grid disks. In order to restrict grid disk (disk group) access to only the authorized users by isolating the Automatic Storage Management disk groups from each other, Oracle Exadata database-scoped security can be configured for specific database clients and grid disks. Since each application database has its own operating system account and own `ORACLE_HOME` facilitates in a database consolidation scenario on the SPARC SuperCluster, configuring database-scoped security for each application database and its associated grid disks in turn restricts disk group access.

If you have a current service contract, you can refer to the "Configuring Security for Oracle Exadata Storage Server Software" chapter in the *Oracle Exadata Storage Server Software User's Guide* for detailed instructions about setting up database-scoped security on Oracle Exadata Storage Servers.

Example Database Consolidation Scenarios

As shown in the Figure 9, the virtualization facilities of the SPARC T4 platform Oracle Solaris provide a great deal of flexibility in how an organization can consolidate multiple Oracle Database 11*g* Release 2 databases onto the SPARC SuperCluster engineered system. The choices range from the very simple case, where the Oracle Database 11*g* Release 2 software is installed on a single database domain without the use of Oracle Solaris Zones (not shown in Figure 9), all the way up to the use of eight Oracle Solaris Zones per database domain, where one whole CPU core is allocated per zone and with each zone having an independent installation of the Oracle Database software and its own Oracle Automatic Storage Management disk group. Each zone enables multiple database instances running per zone.

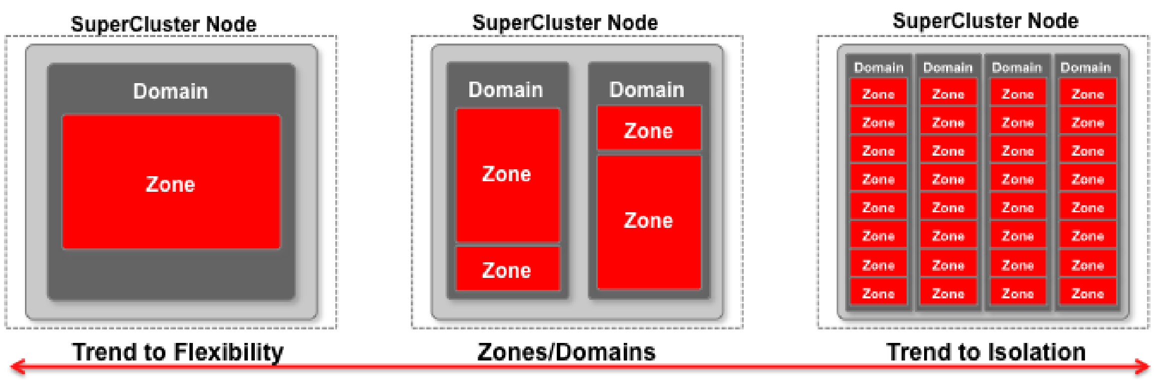


Figure 9. An example showing possible combinations of virtualization options on SPARC SuperCluster compute node.

Refer to the introductory paragraphs and the "Common Zone Deployment Scenarios" section in this document for information about the advantages of deploying zones in database domains on the SPARC SuperCluster.

Consolidation Example Using Half-Rack SPARC SuperCluster

This example uses a half-rack SPARC SuperCluster where an organization has consolidated the databases from four departments onto the Database Domains of the SPARC SuperCluster. The databases in this example are all two-node Oracle RAC configurations, and placing the databases of each department onto different Database Domains offer a high degree of operational and data isolation. Because each department will typically have a number of different application databases, Oracle Solaris Zones are used to separate and isolate the databases within the department, thus allowing patching and maintenance to be done separately even within the department. One Oracle Automatic Storage Management disk group is configured for each zone. Each zone has one installation of Oracle Database 11*g* Release 2, but could contain many instances, which would be typical in a database cloud deployment where privileged end users can request instances to be automatically created.

Figure 10 illustrates the layout of the two-node Oracle RAC databases across the database domains and Oracle Solaris Zones. It features four database domains with 8/8/4/2 zones, respectively, within each database domain.

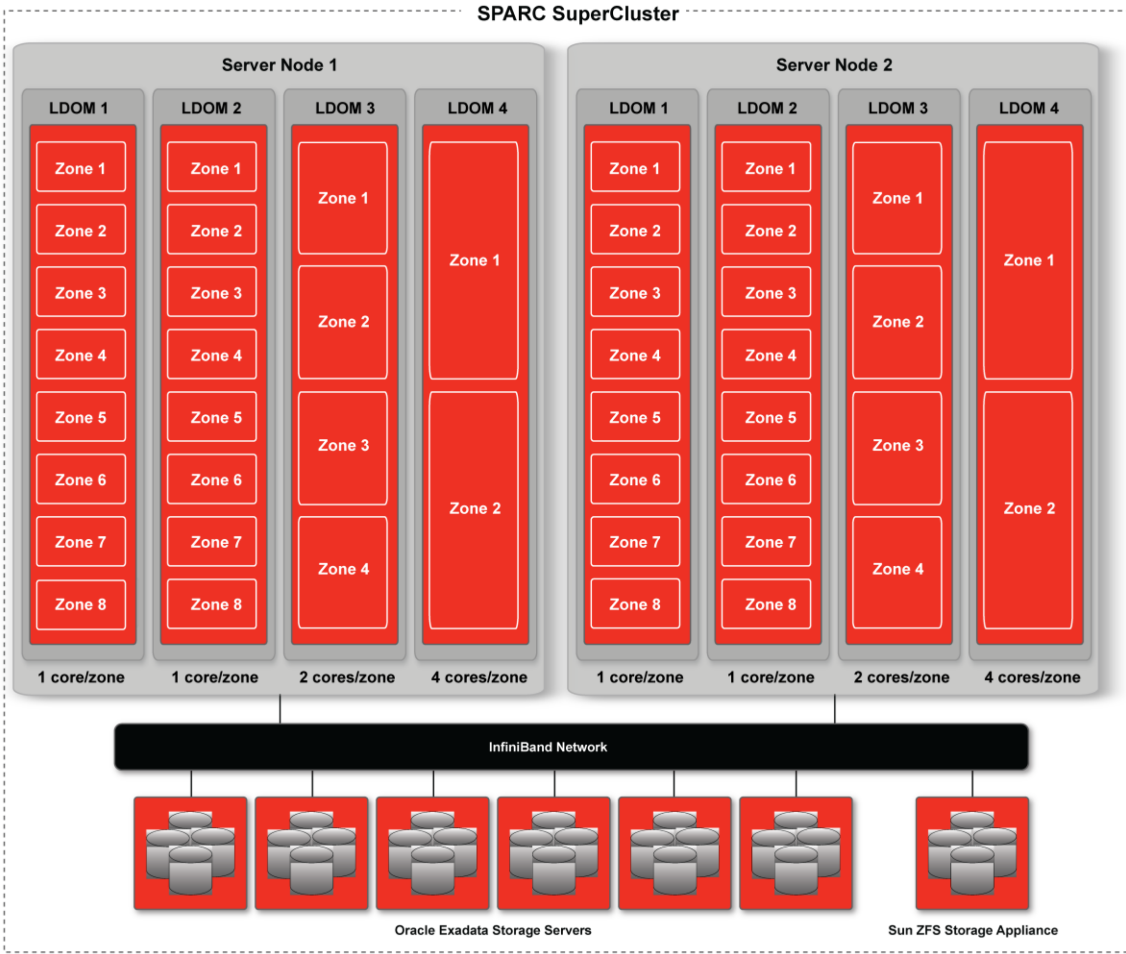


Figure 10. Layout of the two-node Oracle RAC configuration example.

Capacity (CPU and Memory) Planning

Assuming the use of four Database Domains for the consolidation of the departmental databases, the number of virtual CPUs (vCPUs), the amount of memory available for each database, and the number of possible database instances are illustrated in Table 4. The databases can be a mixture of single-instance Oracle Database 11*g* Release 2 database or two-node Oracle RAC clusters. It should be clear that the two-node Oracle RAC clusters would require a corresponding Oracle RAC instance within a zone on the second SPARC T4-4 server.

TABLE 4. CAPACITY ACCORDING TO DEPOLYMENT TYPE

DEPLOYMENT TYPE	ZONES PER DOMAIN	MEMORY PER ZONE	CORES AND VCPUS PER ZONE	11GR2 DATABASE INSTANCES PER ZONE	11GR2 DATABASE INSTANCES PER DATABASE DOMAIN
Highly isolated	8	32	1 core 8 vCPUs	8 * 4 GB	64 * 4 GB
Mixed	4	64	2 cores 16 vCPUs	4 * 16 GB	16 * 16 GB
Flexible	2	128	4 cores 32 vCPUs	4 * 32 GB	8 * 8 GB

Oracle Exadata Storage Server Capacity Planning

Consider an organization that is consolidating the databases of four major departments onto a half-rack SPARC SuperCluster as described previously. Figure 11 illustrates three Oracle Exadata Storage Servers with a total 36 disk drives.

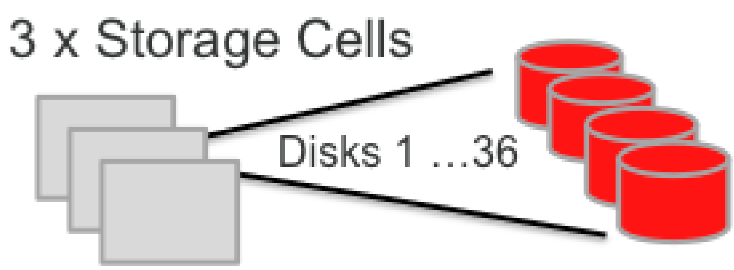


Figure 11. Three Oracle Exadata Storage Servers with 36 disk drives.

Assuming the use of the high-performance 600 GB, 15,000 RPM drives, approximately 21 TB of raw capacity is available before the storage is configured with appropriate striping and mirroring according to the requirements. In the consolidation setup described in previous sections, there are 22 Oracle Solaris Zones, each hosting a two-node Oracle RAC database. Each of those Oracle RAC clusters has a unique Oracle Automatic Storage Management disk group. And prior to the creation of the disk groups, the disks were striped and triple-mirrored for high redundancy, leaving approximately (21 TB / 3) / 22 = ~ 325 GB of uncompressed capacity available per two-node Oracle RAC database. If more disk space is required, there are several options, such as opting for normal redundancy instead of high redundancy, choosing the high-capacity drives in each of the Oracle Exadata Storage Servers, configuring the Exadata Storage Expansion Rack for additional capacity, or choosing the full-rack SPARC SuperCluster configuration rather than the half-rack configuration.

Consolidation Example Using Full-Rack SPARC SuperCluster

Assume that a large organization is consolidating 10 different Oracle databases on a full-rack SPARC SuperCluster. The organization requires 2 out of the 10 databases to be highly available and highly scalable to handle a large number of database sessions that map to its core business and 4 databases to be highly available and moderately scalable to handle intermittent bursts of business activity. The remaining 4 databases require large amounts of storage space to maintain a repository of data for reporting and data analysis, with less emphasis on high availability and scalability.

Another requirement is the ability to patch the highly available databases independently.

Given these high-level requirements, the organization can leverage various technologies available in the SPARC SuperCluster, ranging from virtualized environments (Oracle Solaris Zones) to Oracle Exadata Storage Servers, to meet its goals. The rest of this section outlines one of many possible solutions to meet the consolidation objective.

The solution proposes deploying a mix of clustered and non-clustered databases in database domains of the SPARC SuperCluster with a mix of high-performance and high-capacity SAS disks in the Oracle Exadata Storage Servers for the data store.

Number of Database Domains and Zones

A full-rack SPARC SuperCluster consists of four SPARC T4-4 servers and six Oracle Exadata Storage Servers. It is possible to create multiple two-node or four-node Oracle RAC 11*g* Release 2 databases on a full-rack SPARC SuperCluster. Per the requirements, the customer can create two four-node Oracle RAC databases, four two-node Oracle RAC databases, and four single-instance databases in the full-rack SPARC SuperCluster.

One way of achieving the other requirement of having the ability to patch different databases independently, is to isolate each database instance using Oracle Solaris Zones virtual environments. Even if only one database instance runs in a database domain, running it in a zone has advantages. For example, database patches can be applied and reverted at will using ZFS snapshots, another database can be consolidated in the same database domain at a later date without making any changes to the existing database running in the zone, and the database binaries can be shared with another database that may run in another zone in the same database domain, if needed.

In the case of both the two-node and four-node Oracle RAC database configurations, the `ssc_exavm` tool can deploy the zones in the desired Database Domains.

On the SPARC SuperCluster, Database Domains are dedicated to run the Oracle Database software, which has access to the Oracle Exadata Storage Servers for database storage. Each SPARC T4-4 server in the SPARC SuperCluster can host up to four database domains. With the recommendation of at least one CPU core per zone, up to 32 zones can be deployed in different database domains combined on a single SPARC T4-4 server. Since there are four SPARC T4-4 servers, the four-node Oracle RAC configuration can be set up using all four SPARC T4-4 servers with one Database Domain dedicated on each of the servers for optimal performance and availability. Similarly, the two-node Oracle RAC configuration can be set up on any two SPARC T4-4 servers with one Database Domain dedicated on each of the servers. The four single-instance databases can be deployed on each of the four SPARC T4-4 servers at one instance per server. In all these cases, the database instances run in zones deployed on respective Database Domains on each of the SPARC T4-4 servers.

In summary, the solution proposes a total of 16 Database Domains and 20 zones, as shown in Figure 12, to deploy two four-node database clusters, four two-node database clusters, and four single-instance databases on a full-rack SPARC SuperCluster.

Since the number of proposed zones is only five per SPARC T4-4 server and the servers have 4 physical processors, 32 cores and 1 TB of physical memory, depending on the nature of the database workloads that are being consolidated, the customer can allocate CPU and memory resources to each of those zones in such a way that the final configuration provides the optimal performance.

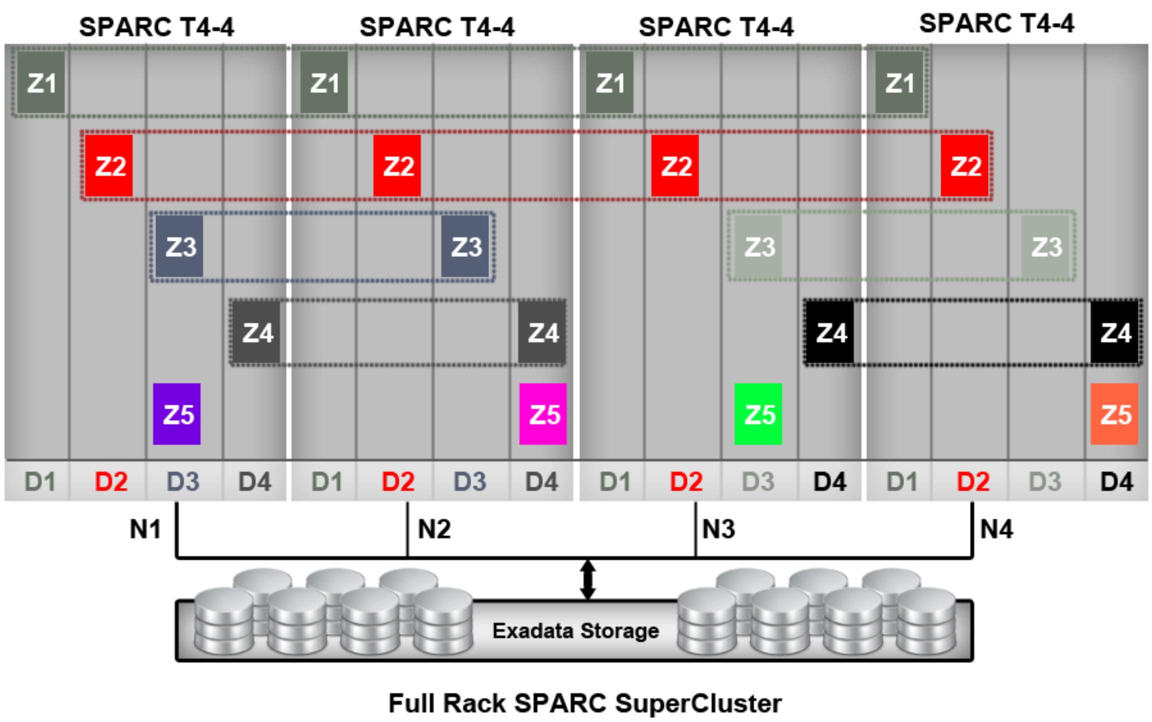


Figure 12. Example full-rack SPARC SuperCluster configuration.

If all four SPARC T4-4 server nodes are denoted by N1, N2, N3, and N4, all four database domains on each SPARC T4-4 server are denoted by D1, D2, D3, and D4, and each of the zones per database domain are denoted by Z1, Z2, Z3, Z4, and Z5, the proposed configuration would look like the following.

Two four-node Oracle RAC database configurations:

1) Z1D1N1, Z1D1N2, Z1D1N3, Z1D1N4

2) Z2D2N1, Z2D2N2, Z2D2N3, Z2D2N4

Four two-node Oracle RAC database configurations:

1) Z3D3N1, Z3D3N2

2) Z3D3N3, Z3D3N4

3) Z4D4N1, Z4D4N2

4) Z4D4N3, Z4D4N4

Four single-instance Oracle Database configurations

1) Z5D3N1

2) Z5D4N2

3) Z5D3N3

4) Z5D4N4

Oracle Exadata Storage Servers and Disk Groups

Each of the Oracle Exadata Storage Servers includes twelve 600 GB high-performance disks or twelve 3 TB high-capacity disks. Mixing Oracle Exadata Storage Servers that have high-performance and high-capacity disks in the same SPARC SuperCluster rack is not currently supported. Hence, choosing the right type of disks is critical in meeting the customer requirements. If the requirement is faster I/O operations, the recommendation is to choose high-performance disks over high-capacity disks. If the available capacity is not sufficient to host the data repository specified in the example, external an Exadata Storage Expansion Rack can be configured for additional capacity.

For the sake of simplicity, let us assume that the customer chooses the option of high-performance disks in Oracle Exadata Storage Servers. Each high-performance disk has a capacity of 600 GB, out of which approximately 570 GB is usable by the application. In such a case, the total available storage space for all the planned databases under a highly redundant configuration is approximately (24 disks * 570 GB per disk) = 13.3 TB. A normal redundant configuration would provide access to (36 disks * 570 GB per disk) = ~ 20 TB.

Since the proposed solution recommends database isolation by suggesting multiple Oracle RAC and non-Oracle RAC database configurations, the requirement of unique disk groups for each cluster can be fulfilled by sharing the storage on all Oracle Exadata Storage Servers across multiple Oracle RAC clusters. In this shared approach, unique disk groups will be created for each of the clusters on grid disks that span all 72 cell disks in all Oracle Exadata Storage Servers on the full-rack SPARC SuperCluster. Multiple databases sharing the same set of disks could potentially lead to I/O contention, which could be mitigated by using I/O Resource Management.

To achieve database isolation in this consolidation example, 10 pairs of unique DATA and RECO disk groups need to be created after careful planning of the disk space requirements for each of the databases by considering the growth factor, plans for additional consolidation, and redundancy options that best suits the business requirements.

Summary

The considerable computing resources of the SPARC T4-4 Servers coupled with the highly optimized Oracle Exadata Storage Servers to run a variety of Oracle Database workloads make Oracle SPARC SuperCluster a viable, flexible database consolidation platform. Integrated virtualization technologies—Oracle VM Server for SPARC and Oracle Solaris Zones—complemented with enterprise management software and resource management tools accelerate the consolidation of database elements on public and private clouds that are built on the SPARC SuperCluster engineered system.

The SPARC SuperCluster offers layered, optimized virtualization—logical domains forming the lower layer and zones comprising the second layer on top of logical domains—with each layer providing isolation at various levels, enabling customers to deploy flexible and effective virtualization to meet their business needs. On the SPARC SuperCluster, Oracle Solaris Zones offer an additional level of virtualization that is suitable for databases requiring isolation for fault, operational, network, security, manageability and resource allocation. Follow the best practices documented in this paper to achieve maximum stability, maximum availability, and predictable performance when supporting multiple applications and databases on the SPARC SuperCluster using shared resources in a consolidated environment.

References

See the following resources:

- "A Technical Overview of the Oracle SPARC SuperCluster T4-4"
- "Oracle Exadata Database Machine Consolidation: Segregating Databases and Roles," a technical white paper
- "Backup and Recovery Performance and Best Practices using the Sun ZFS Storage Appliance with the Oracle Exadata Database Machine," a technical white paper
- "Backup and Recovery Performance and Best Practices for Exadata Cell and Oracle Exadata Database Machine," a technical white paper

In addition, if you have a current service contract, see the following resources:

- "SPARC SuperCluster T4-4 Zones With Oracle Database on Database Domains Configuration Guide" white paper
- *SPARC SuperCluster T4-4 Owner's Guide*
- *Oracle Exadata Database Machine Owner's Guide 11g Release 2 (11.2)*
- *Oracle Exadata Storage Server Software User's Guide*

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

Best Practices for Deploying Oracle Solaris Zones with Oracle Database 11*g* on SPARC SuperCluster
November 2012, Version 1.0
Authors: Giri Mandalika, Sujeet Vasudevan, Ramin Moazeni, and Ramesh Nagappan

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

Hardware and Software, Engineered to Work Together