ORACLE®

ZFS STORAGE
APPLIANCE

An Oracle Technical White Paper

May 2014

How to Set Up Oracle Solaris Kernel Zones Using Oracle ZFS Storage Appliance LUNs

ORACLE®

Table of Contents

Introduction

With the release of Oracle Solaris 11.2, the operating system's virtualization functionality has been further improved, adding strong support for private cloud solutions. This makes Oracle Solaris 11.2 a perfect platform for hardware and software consolidation. When consolidating individual compute resources into a pool of shared resources, it is natural to also consolidate all the storage capacity used by those individual devices into a single storage subsystem supporting all common storage services like the Oracle ZFS Storage Appliance.

The Oracle ZFS Storage Appliance combines advanced hardware and software architectures to provide a multiprotocol storage subsystem that enables you to simultaneously run a variety of application workloads and offer advanced data services. Distinguished by first-class performance characteristics – demonstrated by test results of the industry-standard benchmarks like SPC-1, SPC-2 and SPECsfs – the Oracle ZFS Storage Appliance provides a perfect platform to consolidate and centrally manage all data capacity services from the various applications and operating system instances.

This document provides a basic overview of Oracle's virtualization options, showing where the Oracle Solaris Zones feature fits in the product range. From there, more detail is given about the Solaris Zones features, followed by the benefits of using the newly introduced Oracle Solaris Kernel Zones feature in Oracle Solaris release 11.2 in combination with the Oracle ZFS Storage Appliance.

Understanding Virtualization Model Options

Server virtualization offers many benefits, such as increased hardware utilization – which lowers operational and investment costs – increased security, and increased flexibility. The main goal for most organizations employing server virtualization, or compute resource sharing, is consolidating hardware.

Two primary characteristics define virtualization models: resource flexibility and execution isolation. The two characteristics are diametrically opposed, as the more resource flexibility offered, the less isolation, and vice versa. Each virtualization model offers a certain balance between the two characteristics. Oracle offers a choice of three different models of virtualization:
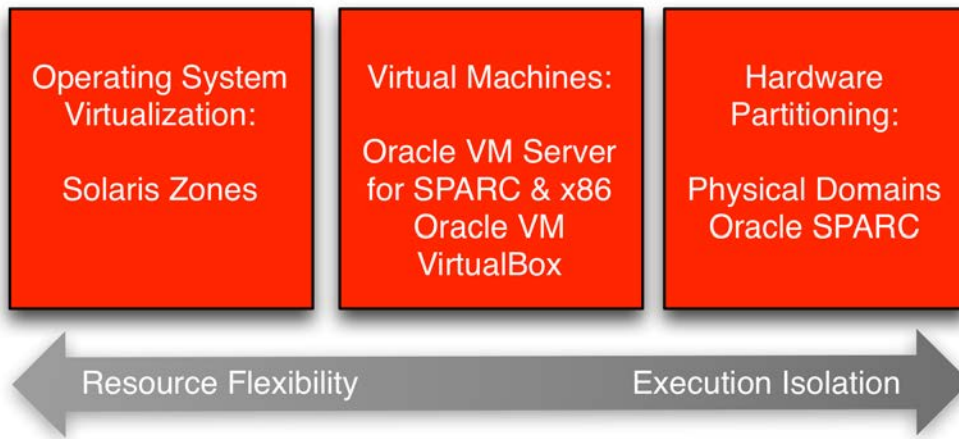

Figure 1. Virtualization model options and their relative positions on the scale of functional characteristics

- **Operating System (OS) virtualization** – A single global OS instance offering one or more isolated execution environments, with each execution environment offering a virtual private copy of the OS. The right amount of resources can easily be allocated to such a private copy in a more efficient way than virtual machines or physical domains do.

  The Oracle Solaris Zone product offers this virtualization model. With the new Kernel Zone function, organizations can achieve execution isolation similar to the virtual machine models.

- **Virtual Machines (VMs)** – A hardware resource can run multiple virtual machines under the control of a firmware hypervisor or piece of control software, creating the illusion for the VM that it is running on its own hardware. Each virtual machine can run a different type of operating system. This model offers better execution isolation than OS virtualization but is less flexible in resource allocation.

  Oracle VM Servers for SPARC, Oracle VM Servers for x86, and Oracle VM VirtualBox provide this type of virtualization model.

- **Physical domains also known as hardware partitions –** Physical domains offer hardware separation among different instances of operating systems. Each OS instance runs directly on its own allocated set of hardware resources, offering bare metal performance. This model offers the best form of execution isolation but is much less flexible in resource allocation.

  Dynamic Domains in Oracle SPARC M-Series and Oracle's Sun Blade 6000 modular systems offer hardware partitioning.

Note that there is no restriction to the exclusive use of just one virtualization model. A combination of virtualization models can be used to maximize workload density. For instance, you could use a SPARC server to create a number of physical domains and run an Oracle Solaris Zone in each of them.

Another option is to use the new Oracle Solaris Kernel Zone and create a number of Oracle Solaris Zones in them. The Kernel Zone provides OS isolation from the Global Zone, thus giving the Zones within the Kernel Zone complete OS separation from the Global Zone. The following sections in the document provide more details on Oracle Solaris Zones.

Using Oracle Solaris Zones for OS Virtualization

Oracle Solaris Zones partitioning is an OS virtualization method providing an isolated and secure environment for running applications. A Zone provides application execution isolation, meaning there is no interaction possible between processes in different Zones, nor access to physical resources of the platform the Zones are running on.

Oracle Solaris uses two types of Zones: Global and non-Global Zones. The Global Zone runs the default Oracle Solaris operating system as it is booted from the bare metal machine, and has control over all processes. A Global Zone always exists even when no other zones are configured. Non Global Zones, or simply Zones, are created, configured, and administered inside the Global Zone.

Zones provide an abstraction layer that separates applications from the physical attributes of a machine, like physical network ports and storage device links. The administrator in the Global Zone has full access to all physical attributes and resources of the hardware and allocates them through logical mappings to the different Zones on the machine. Resources are allocated to a Zone matching the Zone's application requirements, like amount of memory and CPUs needed.

After the initial installation of Oracle Solaris on a server, the Global Zone is present and is the starting point for the administrator to set up the Zone configuration on the machine. A Zone is created within the Oracle Solaris OS instance in the Global Zone providing a virtualized OS environment.

Because the Oracle Solaris Zone architecture is not using a hypervisor, Zones offer a near native performance to applications.

Types of Zones

In the Global Zone, three different types of non-Global Zones are supported:

- Oracle Solaris default Zones – Provides standard OS isolation environment. The Oracle Solaris version running in the Zone uses the same package repository as the Oracle Solaris in the Global Zone, meaning the Oracle Solaris OS in the Zone is always of the same version as the Oracle Solaris OS version in the Global Zone.

- Oracle Solaris 10 Zones – This type of Zone is designed to accommodate applications requiring an Oracle Solaris 10 environment. The Oracle Solaris version in the Zone must match the version in the Global Zone, but it emulates an Oracle Solaris 10 environment.

- Oracle Solaris Kernel Zones – Kernel Zones are a new addition to the Oracle Solaris Zone environment. They can be created using a different version of Oracle Solaris than the one in the Global Zone and as such provide a fully independent kernel and user environment within the Zone. This means an increased kernel separation between the Zone and the Global Zone. Kernel Zones can be used to host non-Kernel Zones, creating a nested Zone structure. The Kernel Zone acts as a Global Zone for the nested Zones.

Zones can be set up as read only, so that the root file system is set up as read only. Zones that have been given additional datasets still have full read/write access to those datasets.

Working with Zones and Storage Consolidation

When consolidating individual data center components into a virtualized environment, it is logical to also include a consolidation of the various individually used storage resources into a single managed storage subsystem. This requires a storage subsystem that supports multiple connection protocols to be able to fulfill the variety of connection requirements from the different hosts and applications.

The Oracle ZFS Storage Appliance is a perfect fit for such an environment, offering both block and network file system types of connections that can be used concurrently. Storage pools can be created with performance and reliability characteristics matching the requirements for the different applications that can be present in the same virtualized host environment.
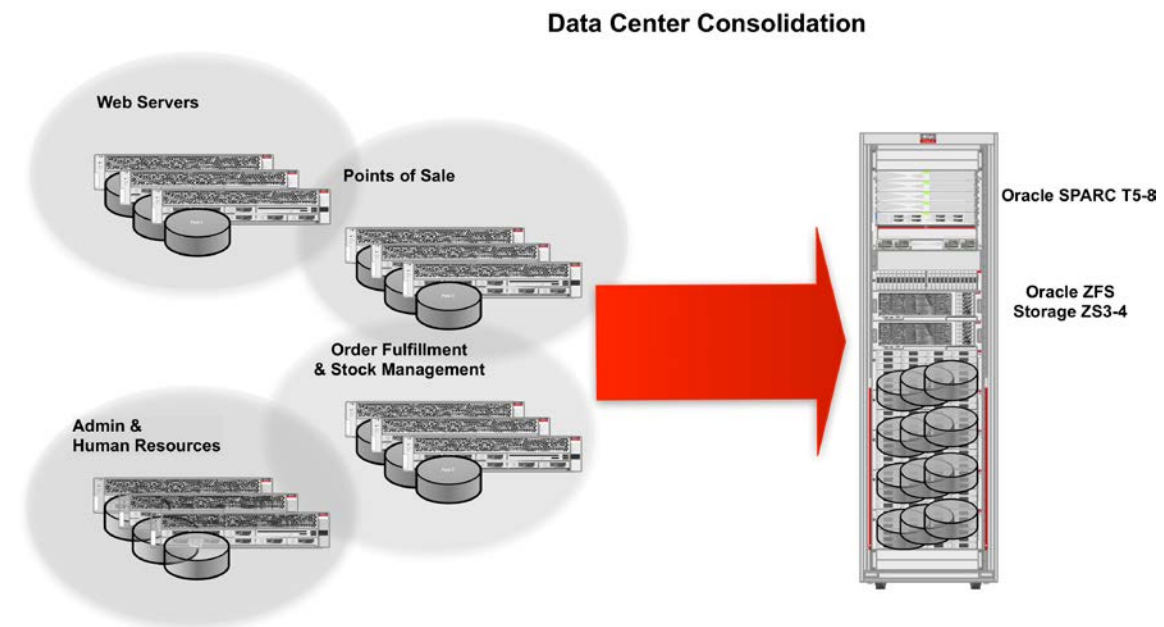

Figure 2. Data center consolidation with the Oracle ZFS Storage Appliance

Oracle Solaris Zones can be hosted on any block-based storage as long as ZFS storage pools can be created on them from the Global Zone.

Each Zone can be given a dedicated amount of storage to host the application data and the local file system needed for the Zone. Additionally, external file systems like CIFS (common internet file system) and NFS, connected through the network, can be mounted within a Zone.

The capabilities provided with Kernel Zones go a step further; an administrator of a Kernel Zone has greater flexibility in configuring and managing the Zone than a default Solaris Zone administrator. The administrator can manage storage private to the Zone, create and destroy ZFS pools, and configure iSCSI and CIFS (common internet file system) services. This makes the Oracle ZFS Storage Appliance a perfect fit for such consolidation environments.

Managing Private Storage Block Devices in Zones

During Zone configuration, you can use the device resource to give the Zone its own storage space. This can be a local block device or a LUN made available through iSCSI or Fibre Channel.

Deploying iSCSI or Fibre Channel Protocols

The choice of which block-level protocol and infrastructure to use depends on local policy and any existing connections between the Oracle ZFS Storage Appliance and the Oracle Solaris host.

Fibre Channel SAN deployment requires installation, configuration and management of Fibre Channel switches, cabling, and host bus adapters (HBAs), but it offers potentially more consistent performance and segregation between standard network traffic and data access traffic. Fibre Channel protocol provides deterministic performance at the network level with bandwidth but for the price of added administration overhead.

iSCSI deployment can require minimal additional configuration other than the usual networking connectivity between storage and server. The iSCSI protocol comes at a performance cost, though, as the standard network and data access traffic contends for resources on the same network. In practice, it may be advisable to provide a second network over which only iSCSI data is routed, keeping user and data traffic separate. iSCSI can be deployed over standard Ethernet interfaces but specialized TCP/IP offload engine (T.O.E.) cards are available to reduce the load on the processors.

The choice of which protocol and, therefore, which network infrastructure to employ is outside the scope of this document.

Setting Up a Kernel Zone Using Oracle ZFS Storage Appliance

Start from a machine that has Oracle Solaris version 11.2 or higher installed. The first decision to take is on which device the kernel boot and root file system will be installed. When no location information is specified in the Kernel Zone creation step, the new Kernel Zone's root file system will be installed in the Global Zone's root pool.

```
/dev/zvol/{dsk,rdsk}/<rootpool>/VARSHARE/zones/<zonename>/.
```

For various reasons, that root pool may not be the desired location. Storage can be made available to a Kernel Zone through block devices such as raw disks, ZFS volumes and lofi (loop file interface) devices. The administrator of a Kernel Zone can modify or create partition tables and file systems, and mount those file systems on any device that is delegated to the Kernel Zone.

If you would like to have the option of migrating a Zone, consider allocating some space for the Zone's root file system on the Oracle ZFS Storage Appliance. This paper's example uses an iSCSI LUN configured on an Oracle ZFS Storage Appliance.

Unlike the other type of Zones, the Kernel Zone root file system is not available for access in the Global Zone.

The configuration process consists of the following steps.

1. Create one or more LUNs on the Oracle ZFS Storage Appliance.

2. Check if the host hardware meets the requirements for supporting the Kernel Zone functionality.

3. Configure a new Kernel Zone. In this step, the LUN(s) from the Oracle ZFS Storage Appliance that will hold the Kernel Zone's root file system are specified.

4. Create the new Kernel Zone.

5. Perform the Oracle Solaris post installation steps after the initial boot of the new Kernel Zone.

Preliminary Tasks to Kernel Zone Installation

The following tasks correspond to steps 1 and 2.

Creating LUNs on the Oracle ZFS Storage Appliance

Once the choice of protocol to use to export LUNs on the Oracle ZFS Storage Appliance has been made, it is necessary to create the LUNs on the Oracle ZFS Storage Appliance. LUNs are added to the Kernel Zone specified in the device resource during the Kernel Zone creation procedure.

A tutorial on creating iSCSI LUNs for use in an Oracle Solaris Environment is available at:

http://www.oracle.com/technetwork/server-storage/sun-unified-storage/documentation/iscsi-quickstart-v1-2-051512-1641594.pdf

An equivalent Fibre-Channel LUN tutorial is available at:

http://www.oracle.com/technetwork/articles/servers-storage-admin/o12-012-fc-lun-solaris-1534203.html

Whichever block protocol is used, the LUNs should be created with a large block size to allow the ZFS cache to operate efficiently. Ultimately, the block size should be matched to the file system work profile; this will differ according to which applications will reside within the Zones.

In this example, the Zone root zpool is created on a single LUN. Once the LUN has been configured and presented to the Oracle Solaris Environment, it will be represented by device nodes in `/dev/dsk` and in `/dev/rdsk`.

Following the example from the iSCSI tutorial mentioned earlier, the LUN will be accessible through `/dev/dsk/c0t600144F08F1579D600004E1C2E250001d0` with the corresponding `/dev/rdsk` entry.

Checking Prerequisites

For SPARC systems, the minimum hardware requirements are, a SPARC T4 (or newer) series server running system firmware version 8.5.0 or newer, or SPARC T5 (or newer) series, M5 or M6 server running system firmware version 9.2.0 or newer.

For x86, Kernel Zones require Intel Nehalem (or newer) or AMD Barcelona (or newer) based x86 machines with CPU virtualization (that is, VT-x) enabled in the BIOS. Specifically, Kernel Zones on x86 require Extended/Nested Page Table support, also referred to as EPT, NPT, or RVI (Rapid Virtualization Indexing).

In addition, a system needs a minimum of 8 GB of physical RAM. Tuning the settings of the ZFS ARC cache is recommended. See the document on this topic listed in References in the Appendix.

Before starting to create a Zone, use the tool `virtinfo` (1M) to verify that the appropriate hardware, firmware and OS levels are in place to support the use of Kernel Zones.

Enable Oracle Solaris multipathing for iSCSI and Fibre Channel, and use the related `c0..` Oracle Solaris device path that is created from the enabled multipathing. The Global Zone can determine the device GUID from this. Use `stmsboot` to enable multipathing, which is disabled by default in Oracle Solaris.

```
root@solaris# stmsboot –D iscsi –D fp –e
WARNING: This operation will require a reboot.
Do you want to continue ? [y/n] (default: y) y
The changes will come into effect after rebooting the system.
Reboot the system now ? [y/n] (default: y)
```

Deploying a New Kernel Zone

The next three operations correspond to steps 3 to 5 previously outlined.

Configuring a Kernel Zone

First connect to the Oracle Solaris server that will host the zone to a CLI session, log in as a valid user, and assume the 'root' role.

```
login: an
Password: xxxxxxxxxxxx
Last login: DDD MMM DD HH:MM:SS from sgdproxy
an@solaris$ su
Password: xxxxxxxxxxxxxxxx
```

Before starting to create a Zone, use the tool `virtinfo` (1M) to verify that the appropriate hardware, firmware and OS levels are in place to support the use of Kernel Zones.

```
root@solaris# virtinfo
NAME            CLASS
```

```
logical-domain          current
non-global-zone  supported
kernel-zone      supported
logical-domain          supported
```

Next, look up the URI (Uniform Resource Identifier) for the iSCSI LUN using the suriadm tool.

```
root@solaris# suriadm lookup-uri -t iscsi
/dev/dsk/c0t600144F08F1579D600004E1C2E250001d0
iscsi://appliancenode.my.com/luname.naa.600144F08F1579D600004E1C2E250001
iscsi://appliance.my.com/target.iqn.1986-03.com.sun:02:4ec2b85a-9b5d-e02c-ee9c-
bdd55452e7de,lun.0
```

The URI will be used to specify the new Kernel Zone boot device.

Create the Zone configuration using the zonecfg (1M) command. In this example, setting the *autoboot* variable to true sets the Zone to automatically boot when the host machine is rebooted.

```
root@solaris# zonecfg -z mykz1
zonecfg:mykz1> create -t SYSsolaris-kz
zonecfg:mykz1> set autoboot=true
zonecfg:mykz1> remove device id=0
zonecfg:mykz1> add device
zonecfg:mykz1:device>set id=0
zonecfg:mykz1:device>
storage=iscsi://appliancenode.my.com/luname.naa.600144F08F1579D600004E1C2E250001
d0
zonecfg:mykz1:device> set bootpri=0
zonecfg:mykz1:device> end
zonecfg:mykz1>commit
zonecfg:mykz1>exit
```

Note that when creating a Kernel Zone, the configuration will already contain a device with id=0 to be used as the device upon which to create the Zone rpool. Simply adding devices with a bootpri attribute set will add these devices to the rpool as a mirror.

```
device:
    match not specified
    storage: dev:/dev/zvol/dsk/rpool/VARSHARE/zones/mykz1/disk0
    id: 0
    bootpri: 0
```

In order to change the default device used for the rpool, change the URL in the storage attribute of the device with id=0 or simply delete the device and create a new one.

Note in the example that the device path from the Global Zone in the storage attribute is used; that is, the iSCSI LUN is picked up by the Global Zone. If you plan to migrate a Zone, it is a good idea to use a storage URI – in this case, an iscsi URI like the following for reference to the LUN in the storage attribute:

```
iscsi://<ZFSSAhostname>/luname.naa.600144F08F1579D600004E1C2E250001
```

Add any LUNs for use by the application environment to the configuration using the add device match resource property or add device set storage resource property. Do not set the bootpri attribute for these devices.

If some devices already have one or more zpools configured on them, you can import them within the Zone once the Zone has been brought up.

Create a mirrored rpool by adding more devices and setting the bootpri attribute.

Creating a Kernel Zone

Use the following commands to create the Kernel Zone.

```
root@solaris# zoneadm -z mykz1 install
Progress being logged to /var/log/zones/zoneadm.20140419T032707Z.mykz1.install
pkg cache: Using /var/pkg/publisher.
 Install Log: /system/volatile/install.5368/install_log
 AI Manifest: /tmp/zoneadm4798.dAaO7j/devel-ai-manifest.xml
  SC Profile: /usr/share/auto_install/sc_profiles/enable_sci.xml
Installation: Starting ...

        Creating IPS image
        Installing packages from:
            solaris
                origin:  http://ipkg/solaris11/dev/
        The following licenses have been accepted and not displayed.
        Please review the licenses for the following packages post-install:
          consolidation/osnet/osnet-incorporation
        Package licenses may be viewed using the command:
          pkg info --license <pkg_fmri>

DOWNLOAD                            PKGS        FILES    XFER (MB)   SPEED
Completed                         549/549  76929/76929  680.9/680.9  8.4M/s

PHASE                                   ITEMS
Installing new actions             104278/104278
Updating package state database             Done
Updating package cache                       0/0
Updating image state                        Done
Creating fast lookup database               Done
Installation: Succeeded
        Done: Installation completed in 354.38 seconds.
root@solaris#
```

You can use zpool status during the zoneadm execution to verify the device(s) being used for rpool of the new Zone.

Performing Post Zone Installation Configurations

Once the new Kernel Zone has booted, log in to finish the installation by following the Oracle Solaris sysconfig (1M) process steps. This will configure system hostname, network, time zone and locale, date and time, user and root accounts, name services, keyboard layout, and support.

```
root@solaris# zlogin -C mykz1
[Connected to zone 'mykz1' console]
Loading smf(5) service descriptions: 220/220
…..
                     System Configuration Tool
```

```
System Configuration Tool enables you to specify the following
configuration parameters for your newly-installed Oracle Solaris 11
system:
- system hostname, network, time zone and locale, date and time, user
  and root accounts, name services, keyboard layout, support

System Configuration Tool produces an SMF profile file in
/etc/svc/profile/sysconfig/sysconfig-20140419-034040.

How to navigate through this tool:
- Use the function keys listed at the bottom of each screen to move
  from screen to screen and to perform other operations.
- Use the up/down arrow keys to change the selection or to move
  between input fields.
- If your keyboard does not have function keys, or they do not
  respond, press ESC; the legend at the bottom of the screen will
  change to show the ESC keys for navigation and other functions.

 F2_Continue  F6_Help  F9_Quit
```

If any devices were added to the Kernel Zone that already contained zpools, use the command `zpool import` to bring them into the `mykz1` Kernel Zone. This feature makes application environment migration very simple when the old application environment was already using a ZFS file system.

If storage was not consolidated, it might be worthwhile to connect the old systems to the Oracle ZFS Storage Appliance and migrate the data onto it. The migrated ZFS file systems can be exported from the old systems before they are brought down. Next, add the LUNs of the ZFS file system now on the Oracle ZFS Storage Appliance to the Zones used to consolidate all the application environments. Add the LUNs as devices to the Zone corresponding to the application migrated onto it and import the ZFS file systems.

Conclusion

The Kernel Zone feature in Oracle Solaris provides a very flexible environment for hosting multiple application instances. Complete consolidation solutions can be rapidly deployed in combination with the Oracle ZFS Storage Appliance and the data migration features of the Oracle ZFS file system.

Appendix: References

See the following resources for additional information relating to the products covered in this document.

References to Sun ZFS Storage Appliance, Sun ZFS Storage 7000, and ZFS Storage Appliance all refer to the same family of Oracle ZFS Storage Appliance products. Some cited documentation may still carry these legacy naming conventions.

Oracle ZFS Storage Appliance Documentation Library, including Installation, Analytics, Customer Service, and Administration guides:
http://www.oracle.com/technetwork/documentation/oracle-unified-ss-193371.html

The *Oracle ZFS Storage Appliance Administration Guide* is also available through the Oracle ZFS Storage Appliance help context.
The Help function in Oracle ZFS Storage Appliance can be accessed through the browser user interface.

Oracle ZFS Storage Appliance Product Information
http://www.oracle.com/us/products/servers-storage/storage/nas/overview/index.html

Oracle ZFS Storage Appliance White Papers and Subject-Specific Resources
http://www.oracle.com/technetwork/server-storage/sun-unified-storage/documentation/index.html

Oracle ZFS Storage 7000 System Administration Guide
http://docs.oracle.com/cd/E27998_01/pdf/E48433.pdf

Oracle ZFS Storage Appliance Documentation Library 2013.1.0, including
the Administration Guide
http://docs.oracle.com/cd/E27998_01/index.html

The *Oracle ZFS Storage Appliance Administration Guide* is also available through the Oracle ZFS Storage Appliance help context.
The Help function in Oracle ZFS Storage Appliance can be accessed through the browser user interface.

Oracle Solaris 11.2 Information Library
http://docs.oracle.com/cd/E36784_01/index.html

"Tuning the Host ZFS ARC Cache on a Kernel Zone Host"
http://docs.oracle.com/cd/E36784_01/html/E37629/gnwox.html

ORACLE®

How to Set Up Oracle Solaris Kernel Zones Using
Oracle ZFS Storage Appliance LUNs
May 2014 Version 1.0
Author: Application Integration Engineering,
Peter Brouwer

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