

Expanding Oracle Private Cloud Appliance Using Oracle ZFS Storage Appliance

ORACLE WHITE PAPER | JANUARY 2018





Table of Contents	0
Introduction	1
Why Use Oracle ZFS Storage Appliance with Oracle Private Cloud Appliance?	2
Cabling	2
Management	3
10Gb Ethernet	3
InfiniBand	4
Additional InfiniBand Expansion	5
External Oracle ZFS Storage Appliance Configuration	6
Oracle Integrated Lights Out Manager (Oracle ILOM) Configuration	6
Management Interfaces	7
Data Interfaces	9
Active/Passive Cluster	9
Active/Active Cluster	10
Pool Setup	12
NFS	13
iSCSI	13
Active/Passive Cluster	14
Active/Active Cluster	14
Project and LUNs	15
Oracle VM Configuration	15
NFS	15



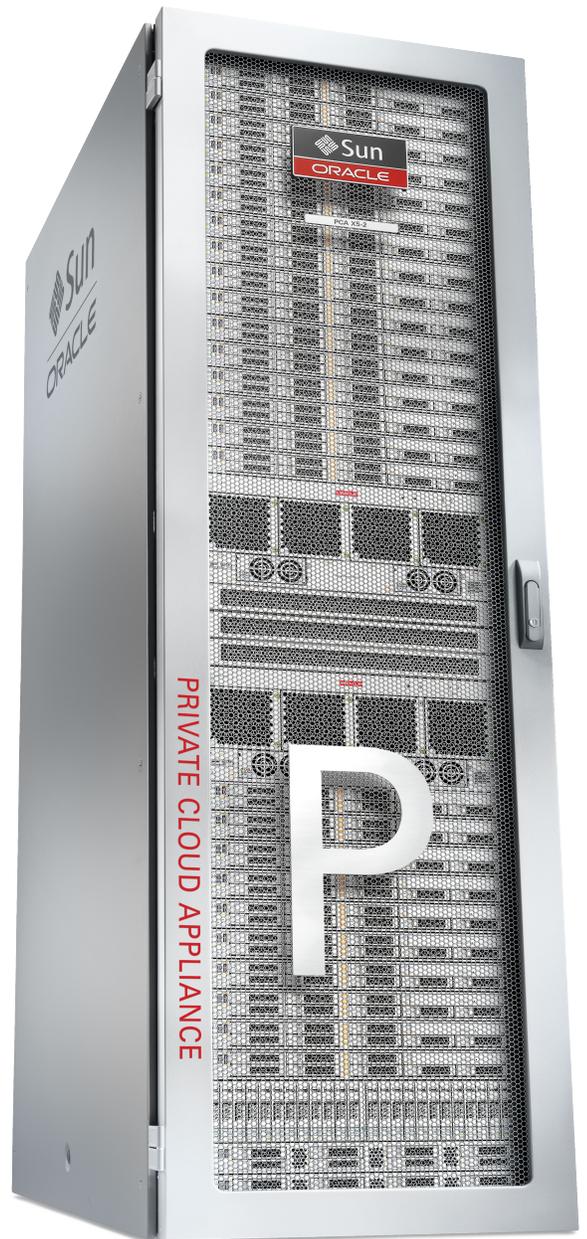
iSCSI	17
Additional Use Cases	20
Best Practices	21

Introduction

Oracle Private Cloud Appliance is a converged infrastructure solution that combines preconfigured networking, servers, and storage into a convenient package that system administrators can easily deploy into an existing data center. Each Oracle Private Cloud Appliance arrives in a single rack with up to 25 servers (compute nodes), as well as multiple InfiniBand and Ethernet switches. Oracle Private Cloud Appliance has been designed to scale dynamically using built-in software automation to accommodate the expansion of its server and storage space. Combined with the virtualization power of Oracle VM, Oracle Private Cloud Appliance provides the perfect general-purpose solution for rapidly and easily bringing online a new rack into a cloud environment.

The storage within Oracle Private Cloud Appliance is a small Oracle ZFS Storage Appliance (the Oracle ZFS Storage ZS-ES model). The internal storage is configured for resiliency and availability rather than performance or scale. It should be considered as the PCA "system disk", and only for VM workloads with light I/O requirements. It contains two clustered 1U controllers and a single disk tray supplied with twenty 900 GB SAS data drives. This design is limiting for high-performance database workloads as well as for backup and recovery needs. To address these requirements, Oracle Private Cloud Appliance has been qualified for use with an additional external Oracle ZFS Storage Appliance. This document describes how to increase the storage capacity of Oracle Private Cloud Appliance by adding a new rack containing a larger Oracle ZFS Storage Appliance cluster.

This document augments the instructions in the Installation Guide, Chapter 8 Extending Oracle Private Cloud Appliance - External Storage and should be used in conjunction. Please take particular note of section "8.3 Adding External InfiniBand Storage" at https://docs.oracle.com/cd/E83758_01/E83753/html/install-extend-storage-ipoib.html



Why Use Oracle ZFS Storage Appliance with Oracle Private Cloud Appliance?

Oracle ZFS Storage Appliance is an ideal solution for storage expansion of Oracle Private Cloud Appliance. It has been co-engineered with Oracle Private Cloud Appliance to maximize performance and efficiency while reducing deployment risk and total cost of ownership. As an engineered storage expansion, Oracle ZFS Storage Appliance offers the following capabilities to Oracle Private Cloud Appliance customers:

- » Oracle ZFS Storage Appliance provides extremely high performance for applications and workloads deployed on Oracle Private Cloud Appliance. It is optimized for Input/Output Operations per Second (IOPS)–intensive workloads, such as OLTP databases, as well as for bandwidth-driven workloads including data warehousing, business intelligence analytics, and video processing. Oracle ZFS Storage Appliance is powerful enough to run a diverse set of workloads concurrently by leveraging the Oracle Private Cloud Appliance's InfiniBand network.
- » Oracle ZFS Storage Appliance also comes with superior storage analytics, which allow customers to visualize and drill down into specific workloads to understand where congestion occurs and why. It can even allow them to examine and manage the storage aspects of Oracle Private Cloud Appliance environments all the way down to the VM level.
- » Oracle ZFS Storage Appliance provides scalable capacity Private Cloud Appliance customers. It is offered in multiple configurations to address different application needs and can expand up to 9 PB.
- » Oracle ZFS Storage Appliance reduces risk by automating storage management using Oracle Enterprise Manager, so customers have fewer storage systems to integrate and manage. It also lowers risk by providing leading fault-monitoring and self-healing capabilities, and by simplifying setup and management through its DTrace Analytics feature.
- » Oracle ZFS Storage Appliance reduces complexity because its large DRAM and flash cache–based architecture is more efficient in serving the I/O from large virtualized environments. In addition, Oracle Database's unique Hybrid Columnar Compression feature, when used in conjunction with Oracle Private Cloud Appliance, reduces the amount of storage needed for data warehouses. And, it enables customers to lower total cost of ownership because they need fewer systems that cost less and are easier to manage.
- » Up to three Oracle ZFS Storage Appliances may be connected via Infiniband to a single Oracle Private Cloud Appliance with up to eight Infiniband connections per ZFS Storage Appliance, enabling additional storage performance and data segregation.

Oracle ZFS Storage Appliance is ideal for expanding Oracle Private Cloud Appliance storage by utilizing the Oracle ZFS Storage Appliance intelligent caching capabilities. This provides the I/O performance of DRAM (1000x faster than flash) at the cost and scalability of a disk-based storage solution.

The following sections outline the steps required for expanding Oracle Private Cloud Appliance using Oracle ZFS Storage Appliance. Either the Oracle ZFS Storage ZS3-2, ZS4-4, ZS5-2 or ZS5-4 appliance can be utilized depending on the capacity demands of the environment.

Cabling

Oracle Private Cloud Appliance currently supports external connectivity to an Oracle ZFS Storage Appliance cluster using InfiniBand and/or 10Gb Ethernet for storage data traffic. Ethernet connections should also be made between the Oracle ZFS Storage Appliance cluster management ports and an Ethernet switch connected to a customer's network infrastructure. The two Oracle Switch ES1-24 switches located in U21 of the Oracle Private Cloud Appliance should *not* be used for Oracle ZFS Storage Appliance connectivity.

Both InfiniBand and 10Gb Ethernet connections may be utilized. Connectivity for physical disks (LUNs) and Oracle VM repositories is best provided via the InfiniBand connection for highest performance. 10Gb Ethernet can also be used for this purpose, but InfiniBand is recommended for performance and simpler deployment.

NFS shares and iSCSI targets mounted by the guest virtual machine are accessed via 10Gb Ethernet connections. The many available 10Gb Ethernet ports on the PCA may be utilized for individual storage pools, VLANs, or other approaches to segregate data and networks. **Note:** Guest VMs have no network interfaces to the InfiniBand network, and access storage only as virtual or physical disks presented through the hypervisor. Ethernet connectivity is required for guest VMs access to NFS or iSCSI (the VM acts as the NFS client or iSCSI initiator).

While the Oracle Private Cloud Appliance and Oracle ZFS Storage Appliance support Fiber Channel (FC) connectivity, it is not recommended as optimum performance can be achieved with InfiniBand or 10Gb Ethernet connectivity.

Management Network

The network management port on both controllers of each external Oracle ZFS Storage Appliance should be connected to a customer-provided Ethernet switch. NET0 on both controllers should also be connected to this switch for a total of four connections to your data center network

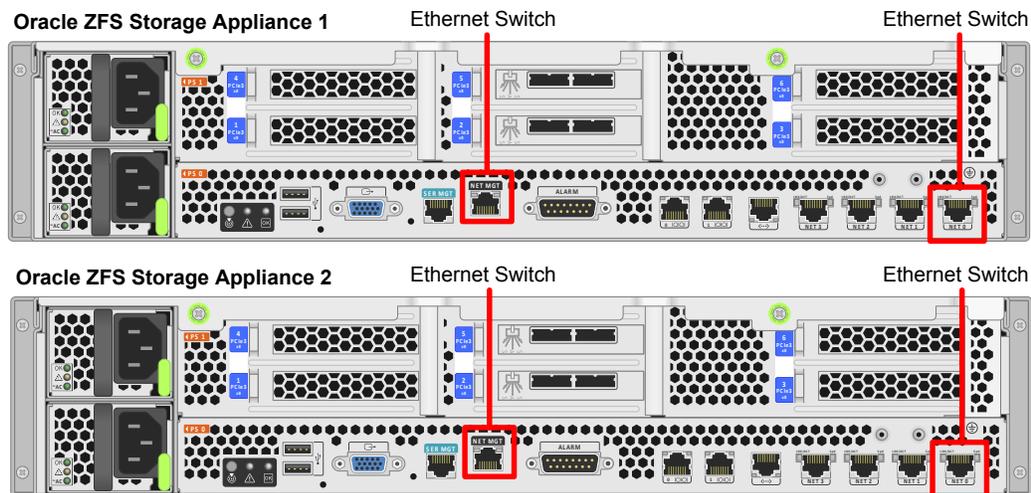


Figure 1 Connecting the Oracle ZFS Storage Appliance controllers

10Gb Ethernet

The 10Gb Ethernet network ports on the Oracle F1-15 Fabric Directors should be used for Ethernet based storage access. The external ZFS Storage Appliance should be connected to the data center 10Gb Ethernet infrastructure as per the Networking Best Practices with Oracle ZFS Storage Appliance whitepaper. Connectivity to the PCA can be via existing customer Data Center networks or dedicated 10Gb Ethernet networks for storage. These networks may be created as per the Network Customization section of the Oracle Private Cloud Appliance Administrators Guide.

Additional 10Gb Ethernet ports may be required in the Oracle ZFS Storage Appliance. The level of redundancy implemented is at the discretion of the implementer, however, one connection from each controller to each network fabric is recommended. See Figure 2.

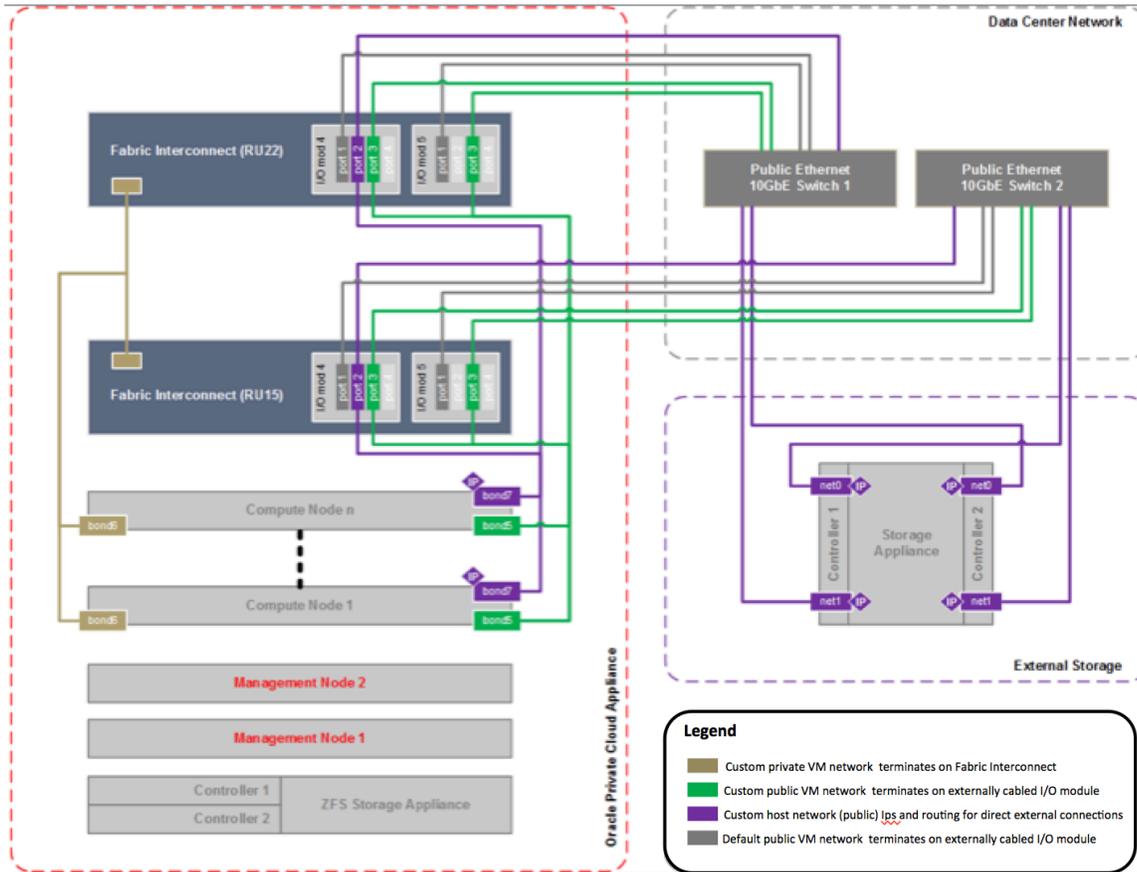


Figure 2 Connecting external storage via 10Gb Ethernet

InfiniBand

Four InfiniBand connections should be made between the two Oracle ZFS Storage Appliance controllers and two Oracle Fabric Interconnect F1-15 switches. Oracle utilizes redundant switches and network connections to avoid any single point of failure. Figure 3 illustrates which ports should be utilized on the switches and their paths to the Oracle ZFS Storage Appliance controllers.

If additional redundancy is desired and space permits, two more InfiniBand HCAs may be added to the Oracle ZFS Storage Appliance and cabled with an additional 4 cables as detailed in the following section.

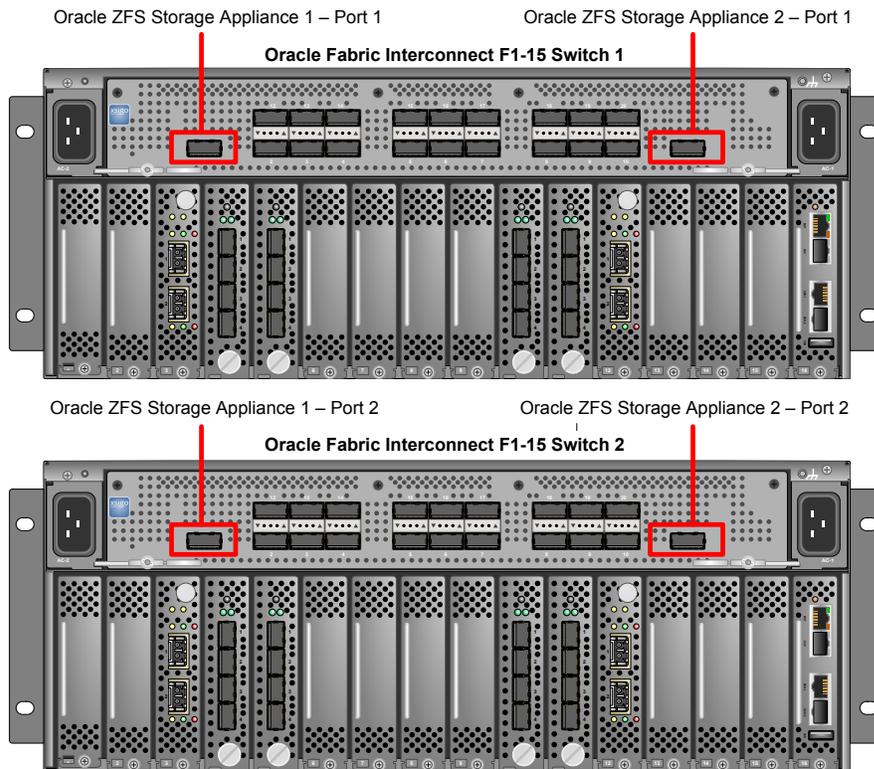


Figure 3 Connecting the Oracle Fabric Interconnect F1-15 switches

Additional InfiniBand Expansion

If higher performance is required between Oracle Private Cloud Appliance and the external Oracle ZFS Storage Appliance, two Sun Data Center InfiniBand Switch 36 (NM2-36) may be added to the ZFS rack. Both of these switches should have two connections each into the Oracle Fabric Interconnect F1-15 devices. They should occupy the same ports as dictated in Figure 2. The external Oracle ZFS Storage Appliance should then be connected to the two Sun Data Center InfiniBand Switch 36 (NM2-36), up to eight connections for the cluster, four per controller. This storage configuration will require two InfiniBand HCA's per storage controller. These connections should be diversified between the additional switches to avoid a single point of failure. This is the preferred configuration for additional expansion.

If additional ZFS appliances are desired, up to a total of three may be connected to the PCA. They may use the same configurations of four or eight Infiniband ports. You may connect the second and third ZFS Storage appliances to secondary Infiniband switches as outlined above:

	Controller 1 Port 1	Controller 2 Port 1	Controller 1 Port 2	Controller 2 Port 2	Controller 1 Port 3	Controller 2 Port 3	Controller 1 Port 4	Controller 2 Port 4
Storage Appliance A	F1-15 1 Port 1	F1-15 1 Port 11	F1-15 2 Port 1	F1-15 2 Port 11	External IB Switch 1	External IB Switch 1	External IB Switch 2	External IB Switch 2
Storage Appliance B	F1-15 1 Port 5	F1-15 1 Port 6	F1-15 2 Port 5	F1-15 2 Port 6	External IB Switch 1	External IB switch 1	External IB Switch 2	External IB Switch 2
Storage Appliance C	F1-15 1 Port 12	F1-15 1 Port 13	F1-15 2 Port 12	F1-15 2 Port 13	External IB Switch 1	External IB switch 1	External IB Switch 2	External IB Switch 2

Open Infiniband ports on the F1-15's may be used in place of an external switches. If this configuration is required, port mapping is:

	Controller 1 Port 1	Controller 2 Port 1	Controller 1 Port 2	Controller 2 Port 2	Controller 1 Port 3	Controller 2 Port 3	Controller 1 Port 4	Controller 2 Port 4
Storage Appliance A	F1-15 1 Port 1	F1-15 1 Port 11	F1-15 2 Port 1	F1-15 2 Port 11	F1-15 1 Port 15	F1-15 1 Port 16	F1-15 2 Port 15	F1-15 2 Port 16
Storage Appliance B	F1-15 1 Port 5	F1-15 1 Port 6	F1-15 2 Port 5	F1-15 2 Port 6	F1-15 1 Port 7	F1-15 1 Port 17	F1-15 2 Port 7	F1-15 2 Port 17
Storage Appliance C	F1-15 1 Port 12	F1-15 1 Port 13	F1-15 2 Port 12	F1-15 2 Port 13	F1-15 1 Port 4	F1-15 1 Port 14	F1-15 2 Port 4	F1-15 2 Port 15

External Oracle ZFS Storage Appliance Configuration

Oracle ZFS Storage Appliance supports both an active/active and an active/passive cluster configuration. The configurations slightly differ in their network and storage layouts. An active/active configuration provides the highest performance, while an active/passive configuration provides the most stable and predictable environment during a storage failover event.

Oracle Integrated Lights Out Manager (Oracle ILOM) Configuration



On both controllers, a serial connection must be made to the RJ-45 serial management port. Serial access should use the following settings:

- » 9600 baud rate
- » 8N1: Eight data bits, no parity, one stop bit
- » No flow control, no hardware control, no software control

Once a connection has been established, log in to the console on each Oracle ZFS Storage Appliance controller using username `root` and password `changeme`.

Next, use the following commands to set up the Oracle ILOM network interfaces:

```
-> cd /SP/network
```

```
-> set pendingipaddress=192.168.150.100
```

Note: This is an example IP address. Oracle Private Cloud Appliance customers should provision an address within their own network for this connection.

```
-> set pendingipnetmask=255.255.255.0
```

```
-> set pendingipgateway=192.168.150.1
```

Note: This is an example gateway address. Oracle Private Cloud Appliance customers should provide their own gateway address.

```
-> set commitpending=true
```

After completion, ping each Oracle ILOM interface using any compute node on Oracle Private Cloud Appliance to verify that the interfaces are working correctly.

Management Interfaces

Create a VNIC and its management interface for each controller. Next, set default routes, configure services, and set up clustering.

1. Log in to the Oracle ZFS Storage Appliance console through Oracle ILOM.

The default username is `root` and the password is `changeme`.

2. Before clustering can be set up, reset each Oracle ZFS Storage Appliance controller to its factory conditions by issuing the following command from the CLI on each controller:

```
ZFS:> maintenance system factoryreset
```

3. After both controllers have rebooted, log in to the console of Oracle ZFS Storage Appliance controller 1 and enter a new password in the Setup screen. Do not enter networking information now; it will be entered later.

4. Create a VNIC for each controller:

```
ZFS:maintenance system setup net> datalinks
```

```
datalinks> vnic
```

```
datalinks> set links=ixgbe0
```

Note: Use `ixgbe0` for an Oracle ZFS Storage ZS3-2 or ZS4-4 controller or `igb0` for an Oracle ZFS Storage ZS3-4 controller.

```
datalinks> commit
```

```
datalinks> vnic
datalinks> set links=ixgbe0
datalinks> commit
```

5. Create the management interface for each VNIC:

```
ZFS:maintenance system setup net datalinks> cd ..
net> interfaces
interfaces> ip
ip> set links=vnic1
ip> set v4addrs=ip_address_zfs_1/subnet_mask
```

Example: set v4addrs=192.168.150.100/24

```
ip> commit
interfaces> ip
ip> set links=vnic2
ip> set v4addrs=ip_address_zfs_2/subnet_mask
ip> commit
```

6. Destroy any system-created interfaces:

```
interfaces> destroy ixgbe0
```

Note: Use `ixgbe0` for an Oracle ZFS Storage ZS3-2 or ZS4-4 controller or `igb0` for an Oracle ZFS Storage ZS3-4 controller.

7. Create default routes for both management interfaces using a customer-supplied gateway address:

```
ZFS:maintenance system setup net interfaces> cd ..
net> routing
routing> create
routing> set destination=0.0.0.0
routing> set mask=0
routing> set gateway=gateway_ip_address
routing> set interface=vnic1
routing> set family=IPv4
routing> commit
routing> create
routing> set destination=0.0.0.0
routing> set mask=0
routing> set gateway=gateway_ip_address
routing> set interface=vnic2
routing> set family=IPv4
routing> commit
routing> done
net> done
```

8. Set up DNS using a customer-supplied DNS server:

```
dns> set domain=domain_name
dns> set servers=dns_ip_address
dns> done
```

9. Set up NTP using a customer-supplied NTP server:

```
ntp> set servers=ntp_ip_address
ntp> commit
ntp> done
```

10. Set up any needed AD, LDAP, or NIS server, or else type the following command:

```
directory> done
```

11. Bypass the storage setup for now:

```
storage> done
```

12. Set up “Phone Home” capability for Oracle Support:

```
support> scrk
scrk> set soa_id=oracle_support_username
scrk> set soa_password=password
scrk> done
```

13. Configure clustering:

```
ZFS:> configuration cluster setup
cabling> done
identity> set nodename=zfs_2_hostname
identity> set password=changeme
identity> done
```

14. Assign the second VNIC to the second Oracle ZFS Storage Appliance controller:

```
ZFS:configuration cluster> resources
resources> select net/vnic2
net/vnic2> set owner=zfs_2_hostname
net/vnic2> commit
resources> commit
```

15. If an active/active cluster configuration is desired, type **Y** at the prompt to failback—otherwise, choose **N**.

Data Interfaces

The storage data interfaces should be created using InfiniBand ports `ibp0` and `ibp1` on Oracle ZFS Storage Appliance. The network layout will differ based on the cluster configuration chosen in the previous section.

Active/Passive Cluster

1. Create a partition key of `ffff` for each InfiniBand device:

```
ZFS:> configuration net datalinks partition
partition> set links=ibp0
partition> set pkey=ffff
partition> set linkmode=cm
partition> commit
datalinks> partition
partition> set links=ibp1
partition> set pkey=ffff
partition> set linkmode=cm
partition> commit
```

If the storage configuration has been provisioned with supplemental InfiniBand HCA's and switches outlined in section [Additional InfiniBand Expansion](#), please execute the following additional commands:

```
partition> set links=ibp2
partition> set pkey=ffff
partition> set linkmode=cm
partition> commit
datalinks> partition
partition> set links=ibp3
partition> set pkey=ffff
partition> set linkmode=cm
partition> commit
```

2. Create an interface on each datalink:

```
ZFS:configuration net datalinks> cd ..
net> interfaces
interfaces> ip
ip> set links=pffff_ibp0
ip> set v4addrs=0.0.0.0/8
ip> commit
interfaces> ip
ip> set links=pffff_ibp1
ip> set v4addrs=0.0.0.0/8
ip> commit
```

If the storage configuration has been provisioned with supplemental InfiniBand HCA's and switches outlined in section [Additional InfiniBand Expansion](#), please execute the following additional commands:

```
interfaces> ip
ip> set links=pffff_ibp2
ip> set v4addrs=0.0.0.0/8
ip> commit
interfaces> ip
ip> set links=pffff_ibp3
ip> set v4addrs=0.0.0.0/8
ip> commit
```

3. Build an IPMP group using both interfaces with two virtual IP addresses:

```
ZFS:configuration net interfaces> ipmp
ipmp> set links=pffff_ibp0,pffff_ibp1
ipmp> set v4addrs=192.168.40.242/24,192.168.40.243/24
ipmp> commit
```

If the storage configuration has been provisioned with supplemental InfiniBand HCA's and switches outlined in section [Additional InfiniBand Expansion](#), please execute the following commands instead:

```
ZFS:configuration net interfaces> ipmp
ipmp> set links=pffff_ibp0,pffff_ibp1,pffff_ibp2,pffff_ibp3
ipmp> set v4addrs=192.168.40.242/24,192.168.40.243/24
ipmp> commit
```

Active/Active Cluster

1. Create a partition key of ffff for each InfiniBand device:

```
ZFS1:> configuration net datalinks partition
partition> set links=ibp0
partition> set pkey=ffff
partition> set linkmode=cm
partition> commit
partition> set links=ibp1
partition> set pkey=ffff
partition> set linkmode=cm
partition> commit
```

If the storage configuration has been provisioned with supplemental InfiniBand HCA's and switches outlined in section [Additional InfiniBand Expansion](#), please execute the following additional commands:

```
ZFS1:> configuration net datalinks partition
partition> set links=ibp2
partition> set pkey=ffff
```

```
partition> set linkmode=cm
partition> commit
partition> set links=ibp3
partition> set pkey=ffff
partition> set linkmode=cm
partition> commit
```

2. Create an interface on each datalink if you do not have additional InfiniBand switches:

```
ZFS1:configuration net datalinks> cd ..
net> interfaces
interfaces> ip
ip> set links=pffff_ibp0
ip> set v4addrs=192.168.40.242/24
ip> commit
ZFS2:configuration net interfaces> ip
ip> set links=pffff_ibp1
ip> set v4addrs=192.168.40.243/24
ip> commit
```

3. If the storage configuration has been provisioned with supplemental InfiniBand HCA's and switches outlined in section [Additional InfiniBand Expansion](#), please execute the following commands to create IPMP interfaces:

```
ZFS1:> configuration net interfaces
interfaces> ip
ip> set links=pffff_ibp0
ip> set v4addrs=0.0.0.0/8
ip> commit
interfaces> ip
ip> set links=pffff_ibp1
ip> set v4addrs=0.0.0.0/8
ip> commit

ZFS2:> configuration net interfaces
interfaces> ip
ip> set links=pffff_ibp2
ip> set v4addrs=0.0.0.0/8
ip> commit
interfaces> ip
ip> set links=pffff_ibp3
ip> set v4addrs=0.0.0.0/8
ip> commit

ZFS1:configuration net interfaces> ipmp
ipmp> set links=pffff_ibp0,pffff_ibp2
ipmp> set v4addrs=192.168.40.242/24
ipmp> commit

ZFS2:configuration net interfaces> ipmp
ipmp> set links=pffff_ibp1,pffff_ibp3
ipmp> set v4addrs=192.168.40.243/24
ipmp> commit
```

If additional Oracle ZFS Storage appliances are to be added, the IP address range of 192.168.20.242-249 is available. The following table is a suggestion of IP addresses to be used in place of those in the previous section:

	ZFS Controller 1	ZFS Controller 2
ZFS Storage Appliance #2	192.168.40.242	192.168.40.243
ZFS Storage Appliance #3	192.168.40.246	192.168.40.247

Pool Setup

Regardless of the cluster configuration, the following storage pool configuration should be used when creating new pools inside of Configuration → Storage:

- » Mirrored or RAIDZ1 storage data profile
 - » Use Mirrored for a ZFS pool that will be used for Oracle VM repositories. LUNs presented directly to a VM as a "physical disk" can use whichever pool type is recommended for that workload just as if running "bare metal" without PCA or Oracle VM. Consult the ZFS whitepapers for appropriate recommendations
 - » Choose RAIDZ1 for streaming or backup workloads, such as Oracle Recovery Manager.
- » Striped write log devices
- » Striped read cache
- » Single pool for active/passive configurations
- » Two pools for active/active configurations, one per controller
- » Write log and read cache devices are highly recommended for Oracle VM workloads.

NFS

Oracle VM requires either NFS or iSCSI for protocol connectivity. The following steps outline the NFS configuration on Oracle ZFS Storage Appliance. Refer to the next section if iSCSI is more desirable. Repositories may be presented on NFS or iSCSI.

1. Create a project called "PCA":

```
ZFS:> shares project PCA
```

2. Set the default user and group to root:

```
PCA> set default_group=root
PCA> set default_user=root
```

3. Ensure root access is available to all the Oracle Private Cloud Appliance subnets:

```
PCA> set
sharenfs="sec=sys,rw=@192.168.4.0/24:@192.168.40.0/24,root=@192.168.4.0/24:@192.168.40.0/24"
PCA> commit
```

4. Create a sample share to be imported to Oracle VM:

```
ZFS:> shares select PCA
PCA> filesystem testshare
PCA> commit
```

5. Create any additional NFS shares needed to expand Oracle VM's storage capacity.

6. Repeat this procedure for the second controller if an active/active cluster configuration is being used.

Note: Changing the Database Record Size of the NFS share containing Oracle VM repositories from the default of 128K can have significant **negative** performance impact, *even if changed back to 128K at a later point in time*

If additional Oracle ZFS Storage Appliances are in use, repeat as necessary.

iSCSI

Each compute node on Oracle Private Cloud Appliance has an IQN identifier that must be manually added to Oracle ZFS Storage Appliance.

1. Extract the IQN from a root shell on a Oracle Private Cloud Appliance compute node:

```
# cat /etc/iscsi/initiatorname.iscsi
InitiatorName=iqn.1988-12.com.oracle:974da248268c
```

2. Add the IQN to Oracle ZFS Storage Appliance controller 1:

```
ZFS:> configuration san iscsi initiators
initiators> create
initiator-000> set alias=PCA_computenode_hostname
initiator-000> set initiator=iqn.1988-12.com.oracle:974da248268c
```

Note: Use the IQN value discovered on the compute node. This is just an example IQN.

```
initiator-000> commit
```

3. Repeat the steps above for every compute node.

4. Create an initiator group and add all the IQNs:

```
initiators> groups
groups> create
group-000> set name=PCA
group-000> set initiators=iqn.1,iqn.2,iqn.3,iqn.4
```

Note: Use the Tab key to add each IQN value added in Step 2 and Step 3

```
group-000> commit
```

Next, create an iSCSI target and an iSCSI target group for an active/passive cluster, or create two iSCSI targets and two iSCSI target groups for an active/passive cluster.

Active/Passive Cluster

1. Create an iSCSI target:

```
ZFS:> configuration san iscsi targets
targets> create
targets> set alias=PCA
targets> set interfaces=ipmp0
targets> commit
```

2. Create an iSCSI target group with the previously created iSCSI target:

```
targets> groups
groups> create
group-000> set name=PCA
group-000> set targets=iqn_initiator_string
```

Note: The IQN value can be discovered using the Tab key with tabbed completion.

```
group-000> commit
```

If additional Oracle ZFS Storage Appliances are in use, repeat as necessary.

Active/Active Cluster

Alternatively, create two iSCSI targets and two iSCSI target groups for an active/active cluster:

1. Create an iSCSI target on Oracle ZFS Storage Appliance controller 1:

```
ZFS1:> configuration san iscsi targets
targets> create
target-000> set alias=PCA-1
target-000> set interfaces=ibp0
target-000> commit
```

2. Create an iSCSI target group with the previously created iSCSI target:

```
targets> groups
groups> create
group-000> set name=PCA-1
group-000> set targets=iqn_initiator_string
```

Note: The IQN value can be discovered using the Tab key with tabbed completion.

```
group-000> commit
```

3. Create an iSCSI target on Oracle ZFS Storage Appliance controller 2:

```
ZFS2:> configuration san iscsi targets
targets> create
target-000> set alias=PCA-2
target-000> set interfaces=ibp1
target-000> commit
```

4. Create an iSCSI target group with the previously created iSCSI target:

```
targets> groups
groups> create
group-000> set name=PCA-2
group-000> set targets=iqn_initiator_string
```

Note: The IQN value can be discovered using the Tab key with tabbed completion.

```
group-000> commit
```

If additional Oracle ZFS Storage Appliances are in use, repeat as necessary.

Project and LUNs

After creating the initiator and target groups, an Oracle Private Cloud Appliance project and LUNs should be added for Oracle VM to access.

1. Create a project called "PCA":

```
ZFS:> shares project PCA  
PCA> commit
```

2. Create a sample LUN to be imported to Oracle VM:

```
ZFS:> shares select PCA  
PCA> lun testlun  
PCA/testlun> set initiatorgroup=PCA  
PCA/testlun> set targetgroup=PCA
```

Note: Use PCA-1 or PCA-2 instead of PCA if an active/active cluster configuration is being used.

```
PCA/testlun> set volsize=1T  
PCA/testlun> commit
```

3. Create any additional LUNs needed to expand Oracle VM's storage capacity.

If additional Oracle ZFS Storage Appliances are in use, repeat as necessary.

Oracle VM Configuration

Oracle VM Manager can be accessed by logging into <https://manager-vIP:7002/ovm/console>, where *manager-vIP* is the virtual IP address of the Oracle VM management console.

NFS

1. Select the **Storage** tab.



Figure 4 Selecting the Storage tab

2. Click the **Discover File Server** icon.



Figure 5 Clicking the Discover File Server icon

3. In the window that opens, enter the following and click **Next**:

Name: zfssa_1
Access Host (IP) Address: 192.168.40.242

The screenshot shows a configuration window titled "Discover a File Server". On the left, there is a navigation pane with "File Server Parameters" selected. The main area contains several fields: "Storage Plug-in" (Oracle Generic Network File System), "Name" (zfssa_1), "Access Host (IP) Address" (192.168.40.242), "Admin Host", "Admin Username", and "Admin Password". There is also a "Uniform Exports" checkbox which is checked, and a "Description" text area. At the bottom right, there are "Cancel" and "Next" buttons.

Figure 6 Entering a name and IP address

4. Move all of the Oracle Private Cloud Appliance compute nodes from the **Available Admin Server(s)** section to the **Selected Admin Server(s)** section. Click **Next**.

The screenshot shows a configuration window with tabs for "Configuration", "Admin Servers", and "Refresh Servers". The "Admin Servers" tab is active. Below the tabs, there is a text box: "Select the Server(s) that can be used for administrative access to this File Server. Required when using a Network File System in a clustered Server Pool." Below this, there are two list boxes: "Available Admin Server(s)" (empty) and "Selected Admin Server(s)" containing a list of server IDs: ovcacn07r1, ovcacn08r1, ovcacn09r1, ovcacn26r1, ovcacn10r1, ovcacn11r1, ovcacn12r1, ovcacn13r1, ovcacn14r1, and ovcacn27r1. Between the lists are navigation arrows: a right arrow, a double right arrow, a left arrow, and a double left arrow. At the bottom right, there are "Cancel" and "OK" buttons.

Figure 7 Selecting the servers that can be used for administrative access

5. Click **Next** in the Select Refresh Servers screen. Oracle Private Cloud Appliance compute nodes do not need to be added.
6. Select the test share presented.



Figure 8 Selecting the test share

7. Click **Finish** to complete.
8. Repeat these steps for Oracle ZFS Storage Appliance controller 2 if an active/active cluster configuration is being used. Use IP address 192.168.40.243 when repeating Step 3.

The Oracle ZFS Storage Appliance should now be found in the File Server directory tree. Its shares can be added as a new storage repository for Oracle VM.

iSCSI

5. Select the **Storage** tab.



Figure 9 Selecting the Storage tab

6. Click the **Discover SAN Server** icon.



Figure 10 Clicking the Discover SAN Server icon

7. Enter the Oracle ZFS Storage Appliance controller's host name in the **Name** field and set **Storage Type** to iSCSI Storage Server. Click **Next**.

The screenshot shows the 'Discover SAN Server' configuration window. On the left, a navigation pane lists steps: Discover SAN Server (selected), Access Information (if required), Set Storage Name (if required), Add Admin Servers, and Manage Access Group. The main area contains the following fields:

- Name:** zfssa_1
- Description:** (empty text area)
- Storage Type:** iSCSI Storage Server
- Storage Plug-in:** Oracle Generic SCSI Plugin
- Plug-in Private Data:** (empty text area)
- Admin Host:** (empty text area)
- Admin Username:** (empty text area)
- Admin Password:** (empty text area)

'Cancel' and 'Next' buttons are located at the bottom right of the window.

Figure 11 Entering the Oracle ZFS Storage Appliance controller's host name

8. Enter IP address 192.168.40.242 in the **Access Host** field and click **OK**.

The screenshot shows the 'Discover SAN Server' configuration window with a 'Create Access Host' dialog box open. The dialog box contains the following fields:

- Access Host:** 192.168.40.242
- Access Port:** 3260
- Access Username:** (empty text area)
- Access Password:** (empty text area)

An information icon and text state: 'Enable Chap for all Access Hosts to enable the Username and Password.' 'Cancel' and 'OK' buttons are located at the bottom right of the dialog box.

Figure 12 Entering the IP address

5. Move all of the Oracle Private Cloud Appliance compute nodes from the **Available Server(s)** section to the **Selected Server(s)** section. Click **Next**.

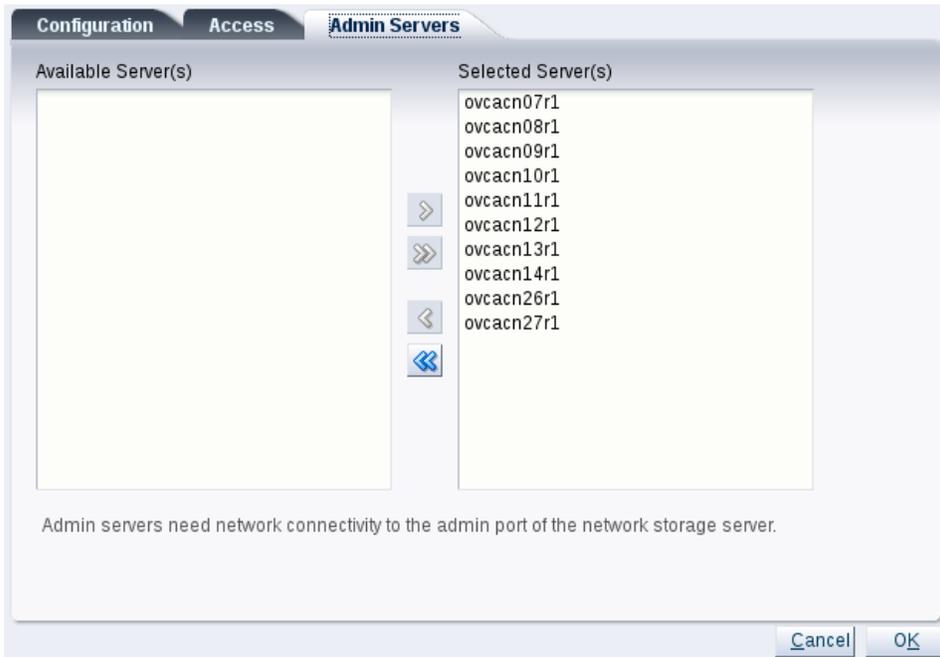


Figure 13 Moving the compute nodes to the Selected Server(s) section

6. Click the **Edit Access Group** icon (pencil) to edit the selected “Default access group.”

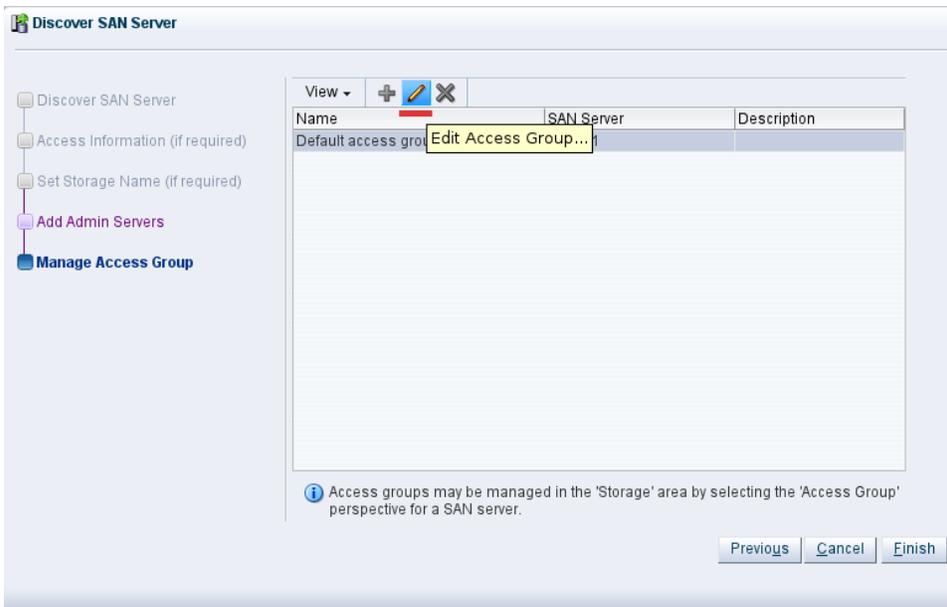


Figure 14 Clicking the Edit Access Group icon

7. Click the **Storage Initiators** tab and move all initiators from the **Available Storage Initiators** section to the **Selected Storage Initiators** section. Click **OK**.

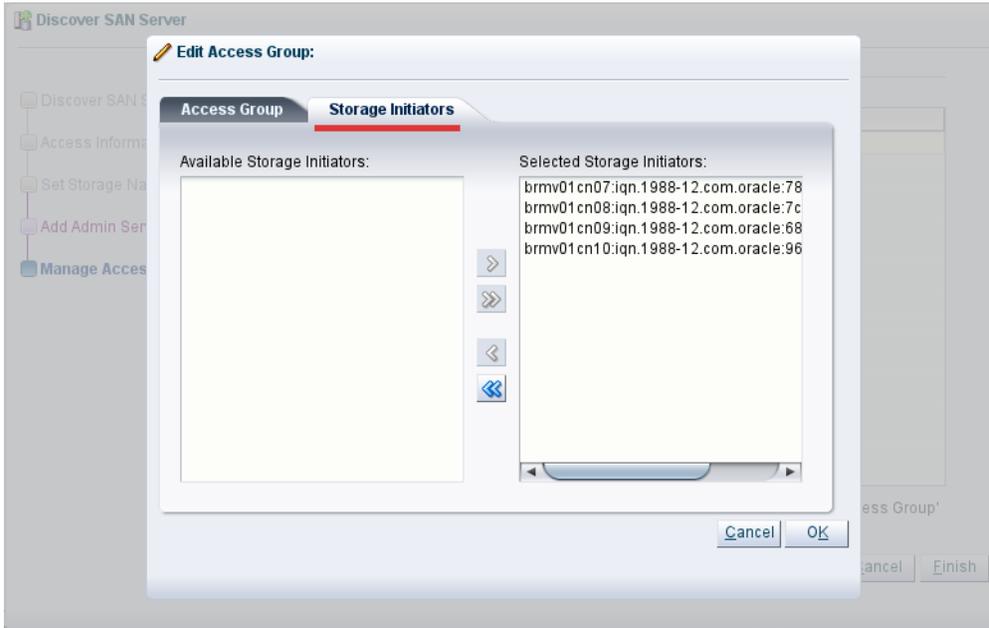


Figure 15 Moving the initiators to the Selected Storage Initiators section

8. Click **Finish** to complete.
9. Repeat these steps for Oracle ZFS Storage Appliance controller 2 if an active/active cluster configuration is being used. Use IP address 192.168.40.243 when repeating Step 3.

Oracle ZFS Storage Appliance should now be found in the SAN Server directory tree. Its LUNs can be added as a new storage repository for Oracle VM.

Networking Storage Reports and Resources Jobs						
View Perspective: Physical Disks						
Name	Event Severity	Size (GiB)	Server	Status		
▷ SUN (9)	Informational	80.0	ovcacn07r1, ovcacn08...	online		
▷ SUN (10)	Informational	4096.0	ovcacn26r1	online		
▷ SUN (11)	Informational	2048.0	ovcacn12r1, ovcacn14r1	online		
▷ SUN (12)	Informational	2048.0	ovcacn11r1, ovcacn13r1	online		

Figure 16 Oracle ZFS Storage Appliance in the SAN Server directory tree

Additional Use Cases

Oracle ZFS Storage Appliance can also be used for general-purpose storage with Oracle VM. Table 1 documents the connectivity options and protocols for this integration. Please note that the guest virtual machines of Oracle VM cannot use Oracle ZFS Storage Appliance for general NFS with IP over InfiniBand (IPoIB). This use case should be implemented using 10 GbE.

TABLE 1: GENERAL-PURPOSE CONNECTIVITY FOR ORACLE VM

Connectivity	Protocol	Oracle VM Hypervisor (dom0)	Oracle VM Guests (domU)
Storage IP over InfiniBand (IPoIB) network	NFS	Yes	No
Storage IP over InfiniBand (IPoIB) network	iSCSI	Yes	No
Public 10 GbE network	NFS	Yes with >PCA version 2.2.1 host networks	Yes
Public 10 GbE network	iSCSI	Yes with >PCA version 2.2.1 host networks	Yes

Best Practices

The Oracle Private Cloud Appliance and ZFS Storage Appliance offer a great many options and features; This section describes how to using these best practices can increase availability and performance of the Oracle PCA and Oracle ZFS Storage Appliance when used together.

ZFS Configuration changes best practices for PCA

Networking

Tuning networks for performance can often be a tricky and tedious task. However, after much testing and evaluation, it has been determined that the MTU size is one of the most important tunable parameters for a ZFS Storage Appliance attached to a PCA. Please set the MTUs as follows:

InfiniBand	MTU	64k
10Gb Ethernet	MTU	9000

NFS

The Database Record Size property in the ZFS Storage Appliance.

The ZFS Storage appliance allows the administrator to tune a large number of parameters in the storage appliance. One of these is the Database Record Size. The Database Record size specifies a suggested block size for files in the file system. This property is only valid for filesystems and is designed for use with database workloads that access files in fixed-size records. The system automatically tunes block sizes according to internal algorithms optimized for typical access patterns.

For databases that create very large files but access them in small random chunks, these algorithms may be suboptimal. Specifying a record size greater than or equal to the record size of the database can result in significant performance gains. Use of this property for general-purpose file systems is strongly discouraged, and may adversely affect performance.

The default record size is 128 KB. The size specified must be a power of two greater than or equal to 512 and less than or equal to 1 MB. Changing the file system's record size affects only files created afterward; existing files and received data are unaffected.

NOTE: If block sizes greater than 128K are used for projects or shares, replication of those projects or shares to systems that don't support large block sizes will fail.

The Database record size setting can be bypassed by the Oracle Intelligent Storage Protocol. Instead of using the record size defined in the file system the Oracle Intelligent Storage Protocol can use the block size value provided by the Oracle Database NFSv4 client. The block size provided by the Oracle Database NFSv4 client can only be applied when creating a new database files or table. Block sizes of existing files and tables will not be changed. For more information, see [Oracle Intelligent Storage Protocol](#) .

When creating repositories for Oracle VM on NFS shares, the default Database Record Size of 128KB should be retained and *never changed*. Changing this parameter can greatly degrade performance. If this parameter is changed from the default of 128KB to a larger value and then changed back, all records created with the non default setting will retain that size and performance will be impacted. Changing this from 128KB to 1MB results in a typical degradation of 7-8x, even after being changed back.

Compression and Deduplication

Compression can be a very useful addition if the source material compresses well. Using the LZJB compression method can improve performance while potentially saving space

The use of deduplication, particularly in tandem with compression, is strongly discouraged. Once deduplication is enabled on a ZFS share, it will be in effect for all written blocks even if turned off later. To remove replication, the best practice is to copy the contents of a deduplicated share to a new ZFS share.

Access control

Best practices for security recommend restricting access to NFS shares holding repositories to the PCA alone and other NFS shares only to the VMs or networks needing to have access. All others should be disallowed. This is set in the Protocols tab of share properties.

All other access properties should be as restrictive as possible.

The screenshot shows the Oracle VM NFS share configuration interface. The breadcrumb is 'pool0/local/default/vms'. The 'Protocols' tab is selected. The share name is 'vms'. The path is 'nas06.web9.com/export/vms'. The share mode is 'Read/write'. There are checkboxes for 'Disable setuid/setgid file creation' and 'Prevent clients from mounting subdirectories'. The anonymous user mapping is 'nobody'. The character encoding is 'default'. The security mode is 'Default (AUTH_SYS)'. There is an 'Inherit from project' checkbox. The 'Usage' section shows 1.0% of 5.25T referenced data and 56.4G total space. The 'Static Properties' section shows compression ratio of 1.00x, case sensitivity of Mixed, reject non UTF-8 of yes, and normalization of None. The 'NFS Exceptions' table is as follows:

TYPE	ENTITY	ACCESS MODE	CHARSET	ROOT ACCESS
Network	192.168.4.0/24	Read/write	default	<input checked="" type="checkbox"/>
Network	192.168.6.0/24	Read/write	default	<input type="checkbox"/>
DNS Domain	myexample.com	Read/write	default	<input type="checkbox"/>
DNS Domain	yourexample.com	Read/write	default	<input type="checkbox"/>
Network	192.168.8.0/24	Read/write	default	<input checked="" type="checkbox"/>

Figure 17 Share Access By Protocol

Synchronous Write Bias

For best performance, set this to 'Latency'.

Virus Scan



Virus scanning should be done from a single point, either inside the VM or within the storage. Performing virus scanning in more than one place can result in disk contention.

iSCSI

Write Cache Behavior

This setting controls whether the LUN caches writes. With this setting off, all writes are synchronous and if no log device is available, write performance suffers significantly. Turning this setting on can therefore dramatically improve write performance, but can also result in data corruption on unexpected shutdown unless the client application understands the semantics of a volatile write cache and properly flushes the cache when necessary. Consult your client application documentation before turning this on.

This should be used with great caution. Best practice is to set this to off. Using write cache is strongly recommended.

Volume Size

Do not change the volume size once the LUN is being used. Changing the size of a LUN while actively exported to clients may yield undefined results. It may require clients to reconnect and/or cause data corruption on the filesystem on top of the LUN. Check best practices for your particular iSCSI client before attempting this operation.

Thin Provisioning

Controls whether space is reserved for the volume. By default, a LUN reserves exactly enough space to completely fill the volume. This ensures that clients will not get out-of-space errors at inopportune times. This property allows the volume size to exceed the amount of available space. When set, the LUN will consume only the space that has been written to the LUN. While this allows for thin provisioning of LUNs, most filesystems do not expect to get "out of space" from underlying devices, and if the share runs out of space, it may cause instability and/or data corruption on clients.

When not set, the volume size behaves like a reservation excluding snapshots. It therefore has the same pathologies, including failure to take snapshots if the snapshot could theoretically diverge to the point of exceeding the amount of available space.

If the consumers are expected to utilize the entire space, or there are charge back policies requiring provisioning of actual space, do not use thin provisioning.

Other Settings

NFS vs iSCSI

Whether to use NFS or iSCSI is a frequently asked question, which can greatly depend on several factors.

NFS is a very mature protocol offering inherently shared filesystem storage with locking across heterogeneous clients and networks. NFS file systems are managed at the server and provided to client machines at a high level. NFS can be cached at several locations in the client-server chain, each of which may be a tunable parameter. Depending on the version of NFS being used, there may be varying degrees of write through caching. It is recommended to use NFS v4 whenever possible.

iSCSI is an inherently block oriented protocol intended for serving out virtual block devices to a single client. Additional layers, such as ASM or OCFS2 can allow sharing of data or file systems across clients, but it is not an inherent part of the protocol. Oracle VM and PCA use OCFS2 to coordinate sharing of iSCSI data across servers.



The single client nature of iSCSI places the filesystem manipulation on the client side. The server simply provides blocks. Caching on the client side can lead to data loss if the client fails before buffers are flushed.

In the end, the familiarity of the storage and systems administrators will, in general, determine the choice of iSCSI vs NFS. However, the following may be taken into consideration

Repositories

- » Oracle VM repositories may reside on NFS, iSCSI, or Fiber Channel LUNs.
- » Fiber Channel connectivity from the ZFS Storage appliance to Oracle PCA is not recommended. 40Gb InfiniBand and 10Gb Ethernet offer better performance than the 8Gb Fiber Channel connections in the PCA platform
- » NFS repositories offer the ability to examine the repository structure through any client mounting the NFS file system and flexibility of access.
- » iSCSI repositories offer the ability to examine the repository structure through any client to which the LUN is presented and mounted. This will require an OCFS2 mount on the client. While inherent to each PCA compute node, this may not be the case with external clients.
- » iSCSI repositories in the context of PCA and Oracle ZFS Appliance directly attached via InfiniBand offer inherently good performance.
- » The use of iSCSI allows thin cloning and snapshots, which can make creating new VMs quick and space-efficient
- » There are a limited number of LUNs, which may be presented to a VM and to compute nodes. See Table 3.4 at https://docs.oracle.com/cd/E83758_01/E89780/html/pcarelnotes-maxconfig.html

Virtual disks vs physical disks vs directly presented iSCSI LUNs

Oracle VM guest (VMs) may have disks presented as virtual or physical disks from the OVM repository or as iSCSI LUNs directly presented to the guest OS. In general, the following use cases would apply:

- Virtual Disks
 - General purpose VM storage.
 - Boot and OS files.
- Physical disks
 - Storage requiring additional performance
 - Databases
 - Real time logging or data collection
 - Storage requiring thin cloning
 - Storage on external SAN
- Directly presented iSCSI LUNs
 - Storage requiring additional performance
 - Databases
 - Real time logging or data collection
 - Storage requiring thin cloning

- Storage on external SAN

Using Direct NFS with the Oracle Private Cloud Appliance and ZFS Storage Appliance

Direct NFS Client integrates the NFS client functionality directly in the Oracle database software to optimize the I/O path between Oracle and the NFS server. This integration can provide significant performance improvements.

Direct NFS requires:

- NFS servers must have write size values (wtxmax) of 32768 or greater to work with Direct NFS Client.
- NFS mount points must be mounted both by the operating system kernel NFS client and Direct NFS Client, even though you configure Direct NFS Client to provide file service.

If Oracle Database cannot connect to an NFS server using Direct NFS Client, then Oracle Database connects to the NFS server using the operating system kernel NFS client. When Oracle Database fails to connect to NAS storage through Direct NFS Client, it logs an informational message about the Direct NFS Client connect error in the Oracle alert and trace files.

- Follow standard guidelines for maintaining integrity of Oracle Database files mounted by both operating system NFS and by Direct NFS Client.

As such, the guest OS must have access to the NFS server itself, in this case the ZFS Storage appliance. In order to implement this, the ZFS Storage Appliance must be on a 10Gb Ethernet network that the guest has access to. This may be a storage network common to all guests or a dedicated storage network for database servers. For maximum performance, a network accessible only to guests running the Oracle database with 10Gb Ethernet network with dedicated ports on the PCA and dedicated NFS shares on the Oracle ZFS Storage Appliance may be implemented.

Direct NFS is independent of the nature of PCA and should be tuned as if the system was not running in a VM.

See Also:

- [Oracle Database Performance Tuning Guide](#) for performance benefits of enabling Parallel NFS and Direct NFS dispatcher
 - <https://docs.oracle.com/database/122/LADBI/about-direct-nfs-client-mounts-to-nfs-storage-devices.htm>
- [Oracle Automatic Storage Management Administrator's Guide](#) for guidelines on managing Oracle Database data files created with Direct NFS Client or kernel NFS
- [Oracle ZFS Storage Appliance Administration Guide – Oracle Intelligent Storage Protocol](#) for additional information on using OISP to improve performance.
- https://docs.oracle.com/cd/E27998_01/html/E48433/integration__oracle_intelligent_storage_protocol.html

Connecting a single Oracle ZFS Storage Appliance to multiple Oracle Private Cloud Appliances.

The advantages of connecting a single high availability, high performance storage device such as an Oracle ZFS Storage Appliance between multiple Oracle Private Cloud Appliances can dramatically multiply the capabilities and functionality of the Oracle Private Cloud Appliance.

A single ZFS Storage Appliance can connect to a maximum of one Oracle PCA through Infiniband due to IP addressing restrictions. In order to connect a more than one PCA to the single ZFS, we must use 10Gb Ethernet.

That can be dedicated Ethernet segments, VLANs, or shared across a common 10Gb Ethernet structure as shown below. For best performance, use both InfiniBand and Ethernet, with IB connection to the PCA requiring maximum performance.

Unfortunately, due to IP address space collision and InfiniBand fabric contention, it is not possible to connect a single Oracle ZFS Storage Appliance to more than one Oracle Private Cloud Appliances using Infiniband. Only one ZFS to PCA Infiniband connection is supported.

While the Oracle Private Cloud Appliance cannot be directly attached to an Oracle Exadata Database Machine or Exalogic, it is possible to connect an Oracle ZFS Storage Appliance to both the Oracle Private Cloud Appliance and Oracle Exadata Database Machine using separate InfiniBand cards in the Oracle ZFS Appliance.

Infiniband and Shared 10Gb Ethernet Implementation (Recommended)

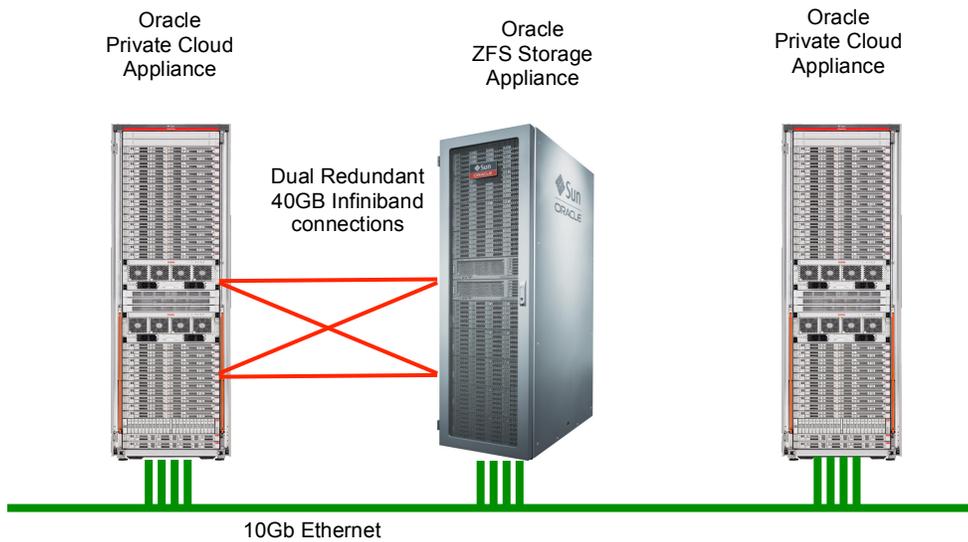


Figure 18 - Infiniband plus 10Gb Shared Ethernet

Shared 10Gb Ethernet Implementation

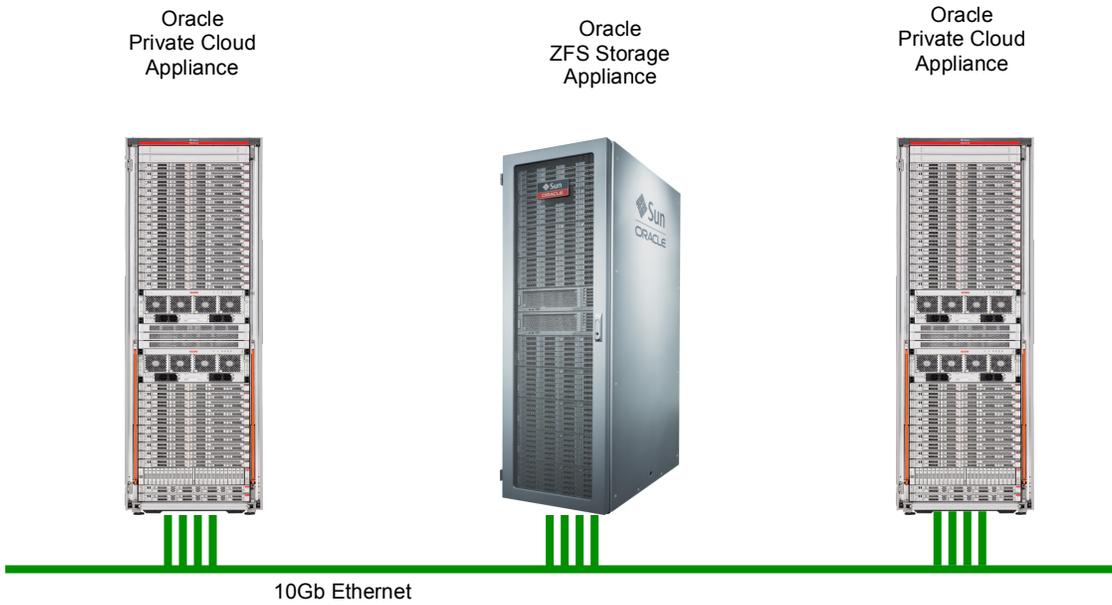


Figure 19 - Shared 10Gb Ethernet only

If desired and Ethernet ports are available on the ZFS Appliance, separate 10Gb Ethernet networks may be implemented to provide segregation.

Segregated 10Gb Ethernet Implementation

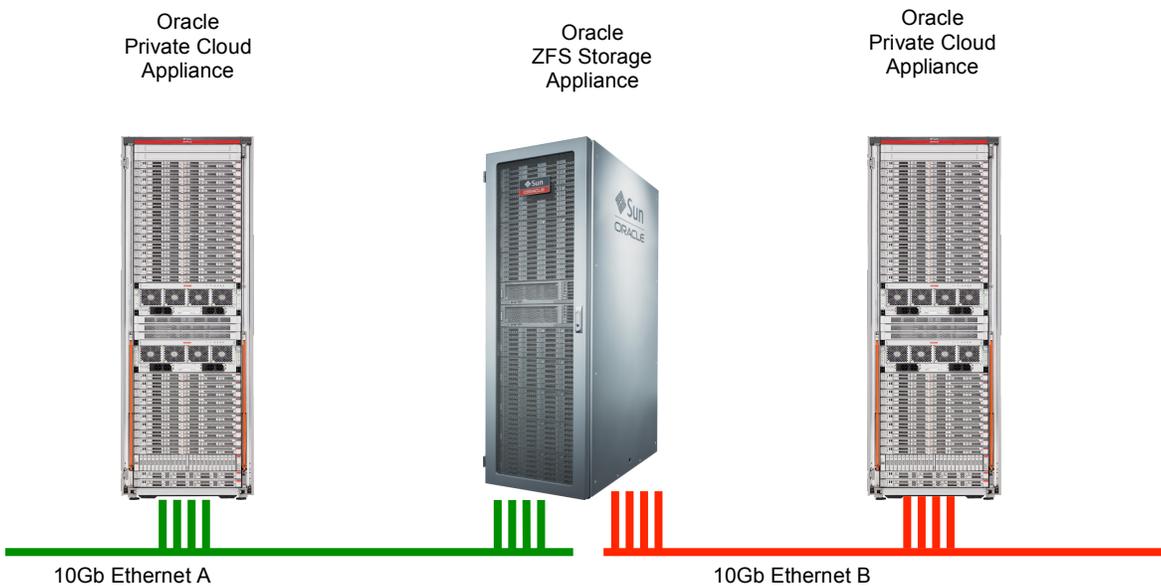


Figure 20 – Segregated 10Gb Ethernet

Learn More

For more information about Oracle virtualization products, visit <http://www.oracle.com/us/technologies/virtualization>, call +1.800.ORACLE1 to speak to an Oracle representative, or visit the web resources below.

RESOURCES

Oracle Private Cloud Appliance Product	https://www.oracle.com/servers/private-cloud-appliance/index.html
Oracle ZFS Storage Appliance Product	https://www.oracle.com/storage/nas/index.html
Oracle virtualization overview	http://www.oracle.com/us/technologies/virtualization/overview/index.html
Optimizing Oracle VM Server for X86 Performance	http://www.oracle.com/technetwork/server-storage/vm/ovm-performance-2995164.pdf
"Oracle VM 3: Architecture and Technical Overview" white paper	http://www.oracle.com/us/technologies/virtualization/ovm3-arch-tech-overview-459307.pdf
Oracle virtualization documentation	http://docs.oracle.com/en/virtualization
Oracle VM Release 3.4 documentation	http://docs.oracle.com/cd/E64076_01/index.html
Architectural Overview of the Oracle ZFS Storage Appliance	http://www.oracle.com/technetwork/.../o14-001-architecture-overview-zfsa-2099942.pdf
Xen hypervisor project	http://www.xenproject.org/users/virtualization.html
Oracle Private Cloud Appliance Documentation Library	1) https://docs.oracle.com/cd/E83758_01/
"Oracle VM 3: 10GbE Network Performance Tuning" white paper	http://www.oracle.com/technetwork/server-storage/vm/ovm3-10gbe-perf-1900032.pdf
Additional Oracle VM White Papers	http://www.oracle.com/technetwork/server-storage/vm/overview/index.html



Oracle Corporation, World Headquarters

500 Oracle Parkway
Redwood Shores, CA 94065, USA

Worldwide Inquiries

Phone: +1.650.506.7000
Fax: +1.650.506.7200

CONNECT WITH US

-  blogs.oracle.com/oracle
-  facebook.com/oracle
-  twitter.com/oracle
-  oracle.com

Integrated Cloud Applications & Platform Services

Copyright © 2015, 2016, 2017, 2018 Oracle and/or its affiliates. All rights reserved. This document is provided *for* information purposes only, and the contents hereof are subject to change without notice. This document is not warranted to be error-free, nor subject to any other warranties or conditions, whether expressed orally or implied in law, including implied warranties and conditions of merchantability or fitness for a particular purpose. We specifically disclaim any liability with respect to this document, and no contractual obligations are formed either directly or indirectly by this document. This document may not be reproduced or transmitted in any form or by any means, electronic or mechanical, for any purpose, without our prior written permission.

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

Intel and Intel Xeon are trademarks or registered trademarks of Intel Corporation. All SPARC trademarks are used under license and are trademarks or registered trademarks of SPARC International, Inc. AMD, Opteron, the AMD logo, and the AMD Opteron logo are trademarks or registered trademarks of Advanced Micro Devices. UNIX is a registered trademark of The Open Group. 0615

Expanding Oracle Private Cloud Appliance Using Oracle ZFS Storage Appliance
January 2018
Author: Paul Johnson
2nd Edition: Bob Bownes