



Oracle ZFS Storage - High Availability Quick Start Guide

Configuration of an Oracle ZFS Storage - High Availability Instance
in Oracle Cloud Infrastructure (OCI) Using the Deployment Tool

November 2022 | Version 4.3.1
Copyright © 2022, Oracle and/or its affiliates
Public

PURPOSE STATEMENT

This document provides step-by-step instructions for configuring an Oracle ZFS Storage - High Availability (ZFS-HA) instance in OCI using the Oracle ZFS Storage Deployment Tool.

For more details on creating a ZFS-HA cluster manually or for details on the APIs for ZFS Storage in OCI, refer to the “Oracle ZFS Storage - High Availability in OCI User Guide”.

The Oracle ZFS Storage Deployment Tool will always create a ZFS-HA system using the ZFS High-Availability Marketplace Image. **The use of this image is not free and will incur a cost of \$1.85 per hour per compute instance.** This cost is in addition to the compute shape and block volume storage charges. The Deployment Tool cannot be used to deploy the Oracle ZFS Storage image, which is limited in the shapes it supports but which has no image use cost. There is no charge for the use of the Deployment Tool.

DISCLAIMER

This document in any form, software or printed matter, contains proprietary information that is the exclusive property of Oracle. Your access to and use of this confidential material is subject to the terms and conditions of your Oracle software license and service agreement, which has been executed and with which you agree to comply. This document and information contained herein may not be disclosed, copied, reproduced or distributed to anyone outside Oracle without prior written consent of Oracle. This document is not part of your license agreement, nor can it be incorporated into any contractual agreement with Oracle or its subsidiaries or affiliates.

This document is for informational purposes only and is intended solely to assist you in planning for the implementation and upgrade of the product features described. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon in making purchasing decisions. The development, release, and timing of any features or functionality described in this document remains at the sole discretion of Oracle.

Due to the nature of the product architecture, it may not be possible to safely include all features described in this document without risking significant destabilization of the code.

TABLE OF CONTENTS

Purpose Statement	2
Disclaimer	2
Introduction	4
Overview of Configuration Steps	5
Requirements for ZFS-HA Clusters	6
ZFS Compute Instance Requirements	6
Network Requirements	6
Cluster Configuration Overview	8
High Availability Clustering	8
Active/Passive Clustering	9
Active/Active Clustering	10
Clustered Instance Terminology	11
Clustered Configuration Operation	11
First Steps	11
Get The ZFS Storage Deployment Tool From OCI Marketplace	12
Configure the Deployment Tool Variables	13
Apply the Stack	17
Set a Password for ZFS Administration	17
Connect to the Browser User Interface (BUI)	19
Active/Active Clustering	19
Share An SMB Filesystem	20
Share An NFS Filesystem	23
Upgrading Your ZFS Storage Instance	25
ZFS in OCI Instance Best Practices	25
Network Best Practices	25
ZFS Storage in OCI Network Routing	25
ZFS Storage in OCI Network Datalinks	25
ZFS Storage in OCI Network Interfaces	25
Block Storage Best Practices	25
System Boot Disk	25
Storage Pools	25
Backup of ZFS Configuration	26
Block Volume Backups	26
Security References	26
Installation Notes	26
Known Issues	26
Appendix A – Installation Checklist	27

INTRODUCTION

Oracle is uniquely positioned to provide products and services that run 24/7 either on-premises or in the cloud and so has the expertise to optimally run our own products in Oracle's own cloud.

The Oracle ZFS Storage - High Availability (ZFS-HA) Marketplace Image provides cloud-based NAS storage and replication services to enable on-premises ZFS Storage customers to migrate data and apps from on-premises to OCI. Oracle ZFS Storage - High Availability instances provide both protocol services and performance for data migration, replication, and sharing.

Two OCI Compute instances running the ZFS-HA image can be clustered together to create a highly available operating environment providing file and storage services in the event of a single instance node failure. Both active/active and active/passive modes are supported. Each instance can detect that the peer instance is unavailable and take over servicing the peer's data pools.

This document covers in detail how to provision a ZFS-HA cluster in OCI using the ZFS-HA image and the "Oracle ZFS Storage Deployment Tool". This tool is a Terraform stack which automates the process of creating and configuring the ZFS-HA compute instances, the Virtual Network Interface Cards (VNICs), and IP addressing needed to build a full ZFS-HA cluster.

The cluster can also be built and created manually. Refer to the "Oracle ZFS Storage - High Availability in OCI User Guide" for more details, as well as information on using APIs for ZFS Storage in OCI.

The Oracle ZFS Storage – High Availability image in OCI can be configured as a Bare Metal (BM) or Virtual Machine (VM) instance to support the following use cases:

- Create a DR site in OCI rather than building out a second on-premises facility by replicating data to a ZFS-HA instance in OCI as a replication target from an on-premises ZFS Storage Appliance and reverse the replication back to on-premises as needed
- Share data from a ZFS-HA instance in OCI over NFS, SMB, or cross protocols back to on-premises
- Migrate and host application storage workloads using similar protocols as your on-premises deployments
- Migrate data to OCI over NFSv3, NFSv4, SMB or cross protocols with AD integration using an Oracle ZFS Storage – High Availability instance as a storage gateway

Sharing data and replicating data can be hosted in the following ways:

- Cloud to Cloud
- On-premises to Cloud
- Cloud to on-premises

Review the following summary of supported shapes and recommended number of NFS and SMB clients to determine the best shape for your requirements.

Network Bandwidth Expectations for NFS/SMB Clients

Shape	Max Memory	Max Network Bandwidth	Max Client Bandwidth	Typical Sustained Bandwidth	Number of Clients
VM.Standard2.4	60GB	4.1 Gbps	256 MB/s	192 MB/s	Tens
VM.Standard2.8	120GB	8.2 Gbps	512 MB/s	384 MB/s	Hundred
VM.Standard2.16	240GB	16.4 Gbps	1025 MB/s	768 MB/s	Few Hundred
VM.Standard2.24	320GB	24.6 Gbps	1537 MB/s	1150 MB/s	Hundreds
VM.Standard3.Flex	512GB	32 Gbps	2000 MB/s	1500 MB/s	Thousand
VM.Standard.E4.Flex	1024GB	40 Gbps	2500 MB/s	1875 MB/s	Thousands
BM.Standard2.52	768GB	25x2 Gbps	3125 MB/s	2343 MB/s	Thousands

Notes:

- The Flex shapes listed require a minimum number of OCPUs to have enough VNICs allocated for High Availability clustering.
 - Active/Active configurations require a minimum of five (5) OCPUs.
 - Active/Passive configurations require a minimum of four (4) OCPUs.
- Typical sustained workload mix with 50% read / 50% write.
- Number of clients depends on the desired throughput available to each client. If more throughput is needed per client then fewer clients should be used.
- A bare metal (BM) or virtual machine (VM) instance requires only one volume for operation. You can add more volumes to increase storage capacity for your needs.
- Maximum block volume capacity per instance is 1024TB based on maximum OCI volumes size of 32TB and the OCI limit of 32 volume attachments.
- Detailed shape specifications are available at [OCI Shapes](#).

Overview of Configuration Steps

This guide describes the steps to configure Oracle ZFS Storage as a compute instance in Oracle's Cloud Infrastructure (OCI) using the Oracle ZFS Storage Deployment Tool and contains the following sections:

- Get the Oracle ZFS Storage Deployment Tool from OCI Marketplace
- Configure the Stack Variables
- Apply the Stack
- Set a Password for ZFS Administration
- Share an SMB Filesystem
- Share an NFS Filesystem

For more information, see the following references:

- [Oracle ZFS Storage Appliance - Release OS8.8.x](#) - General ZFS Storage administration information
- "Oracle ZFS Storage - High Availability in OCI User Guide" – Manual configuration in detail and APIs used for ZFS Storage in OCI
- APIs for ZFS Storage in OCI – The final section in the "Oracle ZFS Storage - High Availability in OCI User Guide" provides additional management APIs used developed specifically for Oracle ZFS Storage - High Availability version.

Requirements for ZFS-HA Clusters

Many of the requirements for provisioning a ZFS-HA cluster are handled by the Deployment Tool , but some items must be in place before the Stack can be run. Review the following sections to identify requirements for provisioning a ZFS-HA cluster.

ZFS Compute Instance Requirements

- Network resources may be in a separate compartment than the compute instances and volumes. This separation is recommended but are not required.
- A dynamic group must be created in the **root compartment** of the tenancy which contains the cluster compartment as described below. In the examples below, it is assumed that the cluster compartment is in the tenancy's root compartment. It is possible for this compartment to be within other compartments in the tenancy. In such a case, care should be taken with the syntax to correctly identify the location of the cluster compartment. (For more information, see [Create a Dynamic Group and Matching Rules](#))
- Identity Policies must be added to allow the ZFSSA compute instances in the dynamic group to manage the cluster.

The following policies are all required for cluster management. These examples use the following compartment names:

`z_cluster`: Compartment for Compute Instances and Block Volumes
`z_cluster_networks`: Compartment for clustering network resources
`z_nas_networks`: Compartment for data access network resources
`z_admin_networks`: Compartment for access to the administration network resources
`z_dgroup`: Dynamic group in the tenancy's **root compartment**.

```
allow dynamic-group z_dgroup to manage instances in compartment z_cluster
allow dynamic-group z_dgroup to manage console-histories in compartment z_cluster
allow dynamic-group z_dgroup to inspect vnics-attachments in compartment z_cluster
allow dynamic-group z_dgroup to manage volume-attachments in compartment z_cluster
allow dynamic-group z_dgroup to use volumes in compartment z_cluster
allow dynamic-group z_dgroup to read private-ips in compartment z_cluster_networks
allow dynamic-group z_dgroup to use vnics in compartment z_cluster_networks
allow dynamic-group z_dgroup to read private-ips in compartment z_admin_networks
allow dynamic-group z_dgroup to use vnics in compartment z_admin_networks
allow dynamic-group z_dgroup to use private-ips in compartment z_nas_networks
allow dynamic-group z_dgroup to use vnics in compartment z_nas_networks
allow dynamic-group z_dgroup to read instance-images in tenancy
```

If the compartment or VCNs in your implementation are grouped differently, you must ensure that these policies are modified to allow access to the correct compartments and VCNs. Failure to do so will result in issues in creating the cluster and in resource takeover/failback within the cluster.

For more information on Identity Policies, see [Write Policies for Dynamic Groups](#)

Network Requirements

The ZFS-HA clustering operates across Virtual Cloud Networks (VCNs) in OCI for various purposes.

- Cluster connectivity: `z_cluster_vcn`
 - Traffic in this VCN includes the block volume I/O, the clustering I/O between the two ZFS-HA controllers, and possibly other Oracle services such as ZFS cloud backups to OCI object storage.
 - **There must be an OCI Service Gateway on this VCN.**
 - This VCN should not allow access to storage administrators or NAS clients. Sharing cluster and block volume traffic on the same VCN as the administrator or data traffic can have a negative affect on the overall performance of the ZFS-HA system.

NAS client connectivity: `z_nas_vcn`

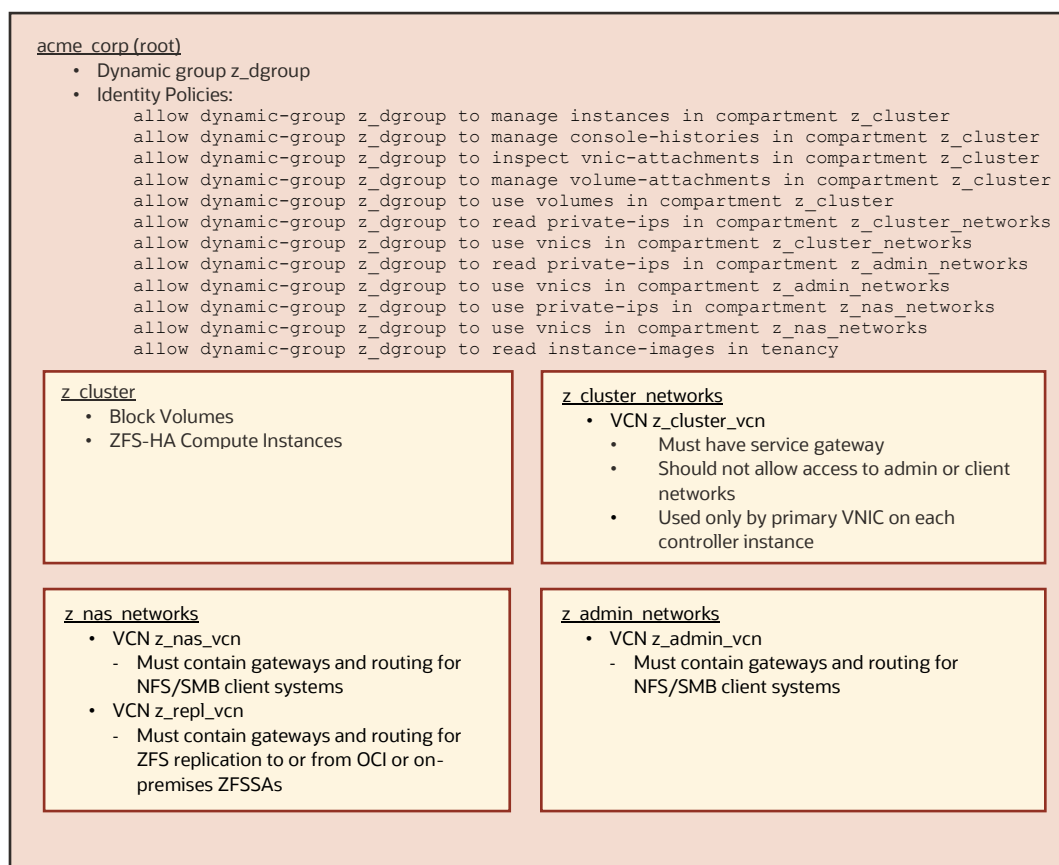
- This VCN must have appropriate gateways, firewall settings, and routing for the client networks. Any compute instance or on-premises computer that accesses data in an ZFS-HA share is considered a client.

- Administrator access: `z_admin_vcn`
 - This VCN must have appropriate gateways and routing for administrative access to the ZFS-HA controller instances, including CLI, REST, and BUI access.
- Replication (optional): `z_repl_vcn`
 - A feature of Oracle ZFS Storage Appliances, whether on OCI or on-premises, is the ability to replicate ZFS snapshots, which capture a share or project's data at the specific point-in-time that the snap is taken. These snapshots can be replicated to other appliances and restored.
 - This VCN must have appropriate gateways and routing to reach the other appliances.
 - It is recommended that additional VNICS be created and assigned to this VCN to separate traffic. These VNICS should be configured like the VNICS used for NAS access, with floating IP addresses.
 - The replication network configuration is not automated with the Oracle ZFS Storage Deployment Tool at this time.
 - If a separate replication network is used and is in a different compartment from the other networks (for example, `z_repl_networks`), two rules must be added to the dynamic group `z_dgroup`:


```
allow dynamic-group z_dgroup to use private-ips in compartment z_repl_networks
allow dynamic-group z_dgroup to use vnics in compartment z_repl_networks
```

 These rules are not required if `z_repl_vcn` is in any of the three network compartments shown below. Whichever compartment it is in must have a rule to use private-ips rather than just read them

The following block diagram shows the required OCI components and their relative locations within the tenancy based on the examples above. The large outer box represents the tenancy's root compartment, and the inner boxes represent child compartments for `z_cluster`, `z_cluster_networks`, `z_admin_networks`, and `z_nas_networks`. If your tenancy uses a different allocation of these resources, the Identity Policies must be modified to reflect your tenancy's configuration. Work with your OCI administrator configure these components in your tenancy.



CLUSTER CONFIGURATION OVERVIEW

This section gives an overview of how two OCI ZFS-HA instances are connected, either in an Active/Active or Active/Passive configuration. When configured as Active/Passive, one instance is active providing data services and one instance is passive, performing no data operations but available for operation if the active instance becomes unavailable.

Active/Passive configuration behavior:

- The primary data pool or pools are configured and running on the active instance.
- If the active instance fails, the primary data pool(s) are exported and imported on the passive instance and NAS IP addresses are migrated.
- The passive instance becomes the active instance until the active instance is recovered.
- When using Flex shapes, a minimum of four (4) OCPUs and 64GB of memory must be provisioned.

Active/Active configuration behavior:

- A minimum of two data pools are required. The total storage for both pools is equal to the number of storage volumes that can be attached to one compute instance. Each node is owner of its own pool(s) and services NAS clients via an IP address tied to those pool(s).
- If the either instance fails, the peer data pool(s) are exported and imported on the working instance and NAS IP addresses are migrated.
- The working instance will now serve both pools from both nodes and may operate in a degraded state since it now must serve those pools with only one node instead of two nodes.
- When using Flex shapes, a minimum of five (5) OCPUs and 80GB of memory must be provisioned.

Takeover behavior:

- Estimated failover time between instances is 70-90 seconds
- Orchestration software transitions the following components when takeover occurs back to the active instance:
 - Secondary IP addresses
 - Public IP addresses
 - Storage volumes

High Availability Clustering

A virtual cluster link (VIO) is used to cluster two ZFS Storage High Availability instances. The Primary VNICs on each instance are used for the link over which the cluster heartbeats occur. The cluster quorum is determined by OCI compute instance metadata properties.

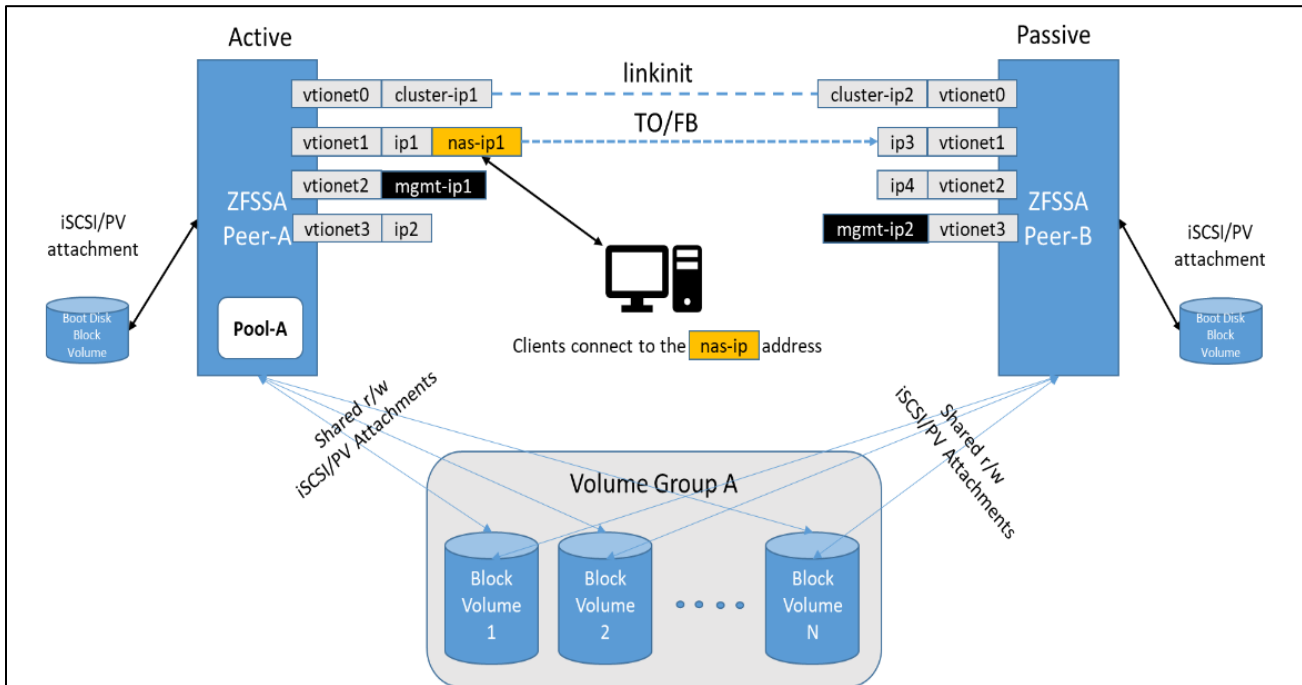
NOTE: The Primary VNIC (Virtual Network Interface Card) is the first network interface in a Compute instance. Additional VNICs may be added and are referred to as Secondary VNICs. Each VNIC is assigned an IP address at its creation and is referred to as the primary IP address. Additional IP address may be assigned to it. These are referred to as the VNIC's secondary IP addresses. Care should be taken not to confuse secondary VNICs with secondary IP addresses.

In a cluster, the following applies to the VNICs:

- The Primary VNIC is used for the VIO link as well as storage volume I/O and OCI API calls. This VNIC is often setup on a private subnet with no access to NAS clients or storage administrators but check with your tenancy administrator for the proper IP subnets and addresses to use.
- Secondary VNICs are configured by the stack to supply access to NAS clients and storage administrators.
- Configuration changes are synchronized across instances.

Active/Passive Clustering

In an Active/Passive cluster, all resources are controlled by a single ZFS-HA compute instance, the Active controller. If the Active controller suffers a failure or an administrator performs a Takeover function, the Passive controller takes over the shared resources such as the storage pool and the nas-ip address.



In an Active/Passive configuration with a single pool, four VNICs are required on each ZFS-HA instance as shown in the table below.

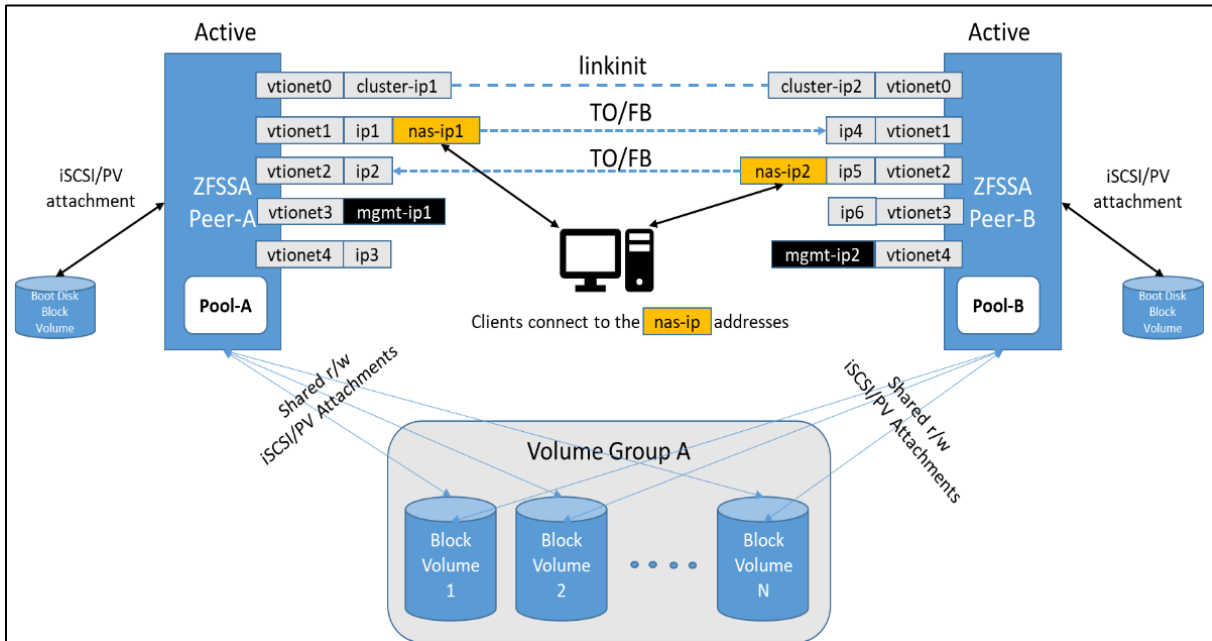
VNIC	ZFS INSTANCE A (ACTIVE)	ZFS INSTANCE B (PASSIVE)	USAGE
vtinet0 (Primary VNIC)	cluster-ip1	cluster-ip2	Used for cluster I/O only
vtinet1 - primary IP	ip1 - Placeholder	ip3 - Placeholder	Primary IP addresses are locked to the instance and cannot move.
vtinet1 - secondary IP	nas-ip1 address	unassigned	Floating IP address used for client access to Pool-A. Always assigned to the active controller
vtinet2 - primary IP	mgmt-ip1	ip4 - Placeholder	Private Administrative access to Node A (B unused)
vtinet3 - primary IP	ip2 - Placeholder	mgmt-ip2	Private Administrative access to Node B (A unused)

In this example, a secondary IP address, nas-ip1, is assigned to vtinet1 on Node A, the active controller. In the event that Node B is made active, the nas-ip1 address will automatically move to vtinet1 on Node B. Clients attached to the nas-ip1 address will have a brief interruption but will continue to be connected to the storage pool when the takeover by Node B is complete.

While only one IP address is used on each controller to connect for management purposes, two VNICs are created for cluster management reasons. One on each controller will always remain unused.

Active/Active Clustering

In an Active/Active cluster, resources are shared across both ZFS-HA compute instances. If either controller suffers a failure or an administrator performs a Takeover function, all shared resources such as the storage pools and the nas-ip addresses are moved to the remaining Active controller.



In an Active/Active configuration with two pools, five VNICs are required on each ZFS-HA instance as shown in the table below.

VNIC	ZFS INSTANCE A	ZFS INSTANCE B	USAGE
vtionet0 (Primary VNIC)	cluster-ip1	cluster-ip2	Used for cluster I/O only
vtionet1 - primary IP	ip1 - Placeholder	ip4 - Placeholder	Primary IP addresses are locked to the instance and cannot move.
vtionet1 - secondary IP	nas-ip1 address	unassigned	Floating IP address used for client access for Pool-A.
vtionet2 - primary IP	ip2 - Placeholder	ip5 - Placeholder	Primary IP addresses are locked to the instance and cannot move.
vtionet2 - secondary IP	unassigned	nas-ip2 address	Floating IP address used for client access for Pool-B.
vtionet3 - primary IP	mgmt-ip1	ip6 - Placeholder	Private Administrative access to Node A (B unused)
vtionet4 - primary IP	ip3 - Placeholder	mgmt-ip2	Private Administrative access to Node B (A unused)

In this example, secondary IP addresses, nas-ip1 and nas-ip2, are assigned to vtionet1 on Node A and vtionet2 on Node B. If either node becomes inactive, the nas-ip1 and nas-ip2 addresses will be assigned to vtionet1 and vtionet2 on the same controller, respectively. The remaining active instance will also control both storage pools. Clients attached to the nas-ip1 address will have a brief interruption but will continue to be connected to the storage pool when the takeover by Node B is complete.

Clustered Instance Terminology

A resource is a physical or virtual object that is present and possibly active on one or both cluster heads. Resources are managed by storage administrators who can set which instance owns the resource when clustered.

Term	Description
Resource Type	
Singleton	Known by both instances but only active on one instance. (Storage Pools and NAS IP)
Private	Only available and active on one instance. (Administration Network Interface)
Replicate	Resource known by both heads. (Service configuration)
Symbiote	Follows other resources (Replications actions follow storage pool)
Clustered State	
Unconfigured	Clustering is not configured.
Owner	Clustering is configured. This active instance owns the storage and data resources.
Stripped	Clustering is configured. This passive instance does not control any shared resources.
Clustered	Clustering is configured in an active/active configuration.

Clustered Configuration Operation

- OCI API commands are issued from each clustered ZFS instance to manage OCI compute, storage, and network resources.
- OCI principal authentication is used to issue OCI API commands.
- All ZFS cluster resources must be in the same OCI availability domain and the same dynamic group.
- All storage volumes will be mounted as shareable on both ZFS instances.
- Network interfaces configured as singletons must use secondary IP addresses so they can be migrated.

FIRST STEPS

If your organization does not have an Oracle Cloud Infrastructure (OCI) account already, one can be set up at <https://www.oracle.com/cloud/>. Note that the Oracle ZFS Storage – High Availability image is not available as part of the Oracle Cloud Free Tier.

This guide assumes that usable compartments, virtual cloud networks (VCN), and subnets have already been created within the OCI tenancy. An administrator for your OCI tenancy will authorize resources in a specified compartment for you to use.

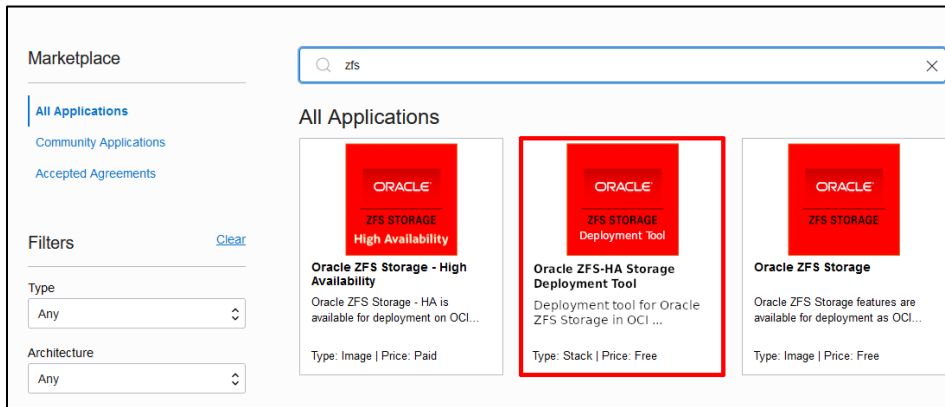
The following information will be needed to configure the OCI compute instance.

1. OCI Compartment IDs
2. VCN Compartments and Names
3. Subnet Compartments and Names

You will also need an SSH client to do the initial configuration and know how to configure the SSH client to use ssh key authentication. An SSH key pair must be generated before stating the Stack configuration process.

GET THE ZFS STORAGE DEPLOYMENT TOOL FROM OCI MARKETPLACE

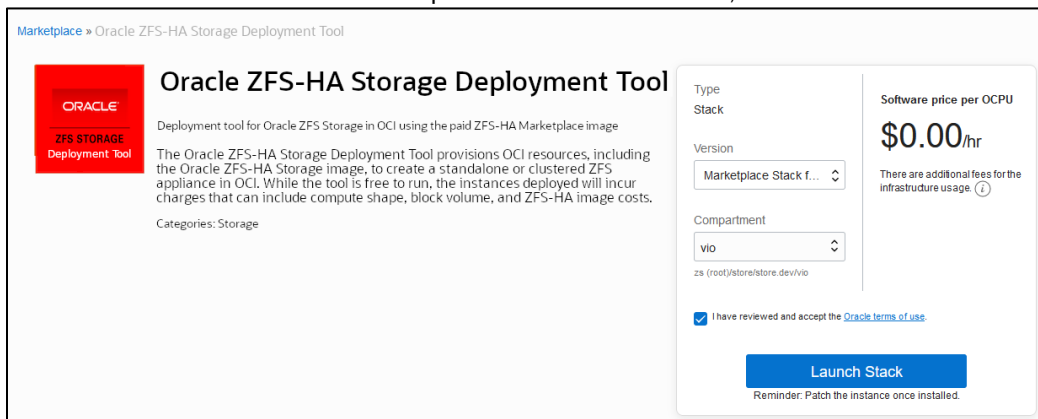
1. Log in to your OCI tenancy and go to the Marketplace and search All Applications for ZFS Storage images. Select the Oracle ZFS Storage Deployment Tool.



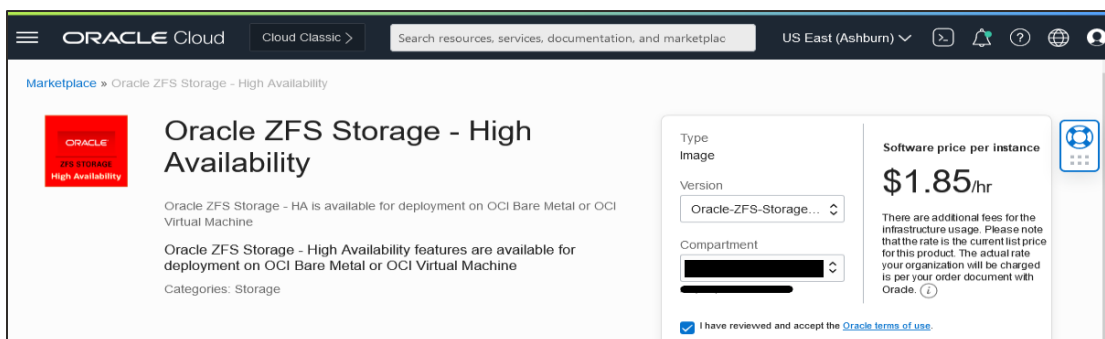
2. Choose either the Bare Metal (BM) or the Virtual Machine (VM) version from the pulldown menu depending on what shape the instance should run on. The default is the latest Virtual Machine stack. In some limited cases a previously released image may be desired, so these images are shown in the pulldown menu. Ensure that the correct type and version of the image is selected for your use case.

Select the appropriate compartment for your tenancy to run the Compute instances in.

Read the Overview and review and accept the terms and conditions, then click “Launch Stack”.



Note that while the Deployment Tool is free to use, it will deploy a pair of instances using the Oracle ZFS Storage – High Availability image. Each ZFS-HA instance has an hourly charge in addition to the Compute shape and Block Storage charges incurred, as shown below.



CONFIGURE THE DEPLOYMENT TOOL VARIABLES

- Enter a name for the Stack and optionally add a description or tags. Neither the Compartment name nor Terraform version can be changed on this screen. Click Next.

Create Stack

1 Stack Information 2 Configure Variables 3 Review

Your application will launch as part of a stack that includes the infrastructure resources required to ensure that the application deploys and runs properly.

Name *Optional*
JH Stack

Description *Optional*

Create in compartment
vio
za (root)/store/store-dev/vio

Terraform version
1.1.x

0.11.x is no longer supported. [What Terraform versions are supported by Resource Manager?](#)

Tags

Optional tags to organize and track resources in your tenancy. [How do I use tags?](#)

Tag Namespace	Tag Key	Tag Value
None (add a free-form tag)		

Next Cancel

- Configure the variables in the “Storage Configuration and Placement” section.
 - Use the pulldown menu to select the type of cluster desired. An Active/Active cluster will create two storage pools, while an Active/Passive cluster will create only one. Some variables listed here will not appear if an Active/Passive cluster is chosen.
 - Choose the Compute Instance Shape from the pulldown menu. If a Flex shape is chosen, enter the number of OCPUs and the Memory Size in GBs. Note that an Active/Active cluster will not run on a VM.Standard2.4 shape. Also note that when using Flex shapes, Active/Active clusters must have a minimum of five (5) OCPUs and 80GB of memory; Active/Passive clusters must have a minimum of four (4) OCPUs and 64GB of memory.

Storage Configuration and Placement

Storage Configuration

Active/Passive

Select a type for storage configuration. HA Solution: Active/Passive or Active/Active, Non-HA Solution: SingleHead.

Compute Instance Shape

VM.Standard.E4.Flex

Compute instance shape to use for ZS OCI instances. Select a shape supported by the image. VM.Standard2.4 shape does not support Active/Active configuration.

Number of OCPUs *Optional*

4

Number of OCPUs to allocate for ZS OCI instance.

Memory Size (GBs) *Optional*

60

Memory size in GBs to allocate for ZS OCI instance.

- In Storage Name, choose a name for the cluster. The two Compute instances will use this name plus “-a” or “-b” appended to it as the hostnames for the instance. As an example, if ‘zfsha’ is entered here, the two Compute instances will be named ‘zfsha-a’ and ‘zfsha-b’.
- In Compartment, choose the compartment in which the Compute instances and block volumes will be placed.
- In Availability Domain, choose the AD in your region to place the compute instances and block volumes.

- Choose the Fault Domains in which to run each instance using the pulldown menus. The instances must run in separate Fault Domains.

Storage Name

p-zfs

Base hostname for ZFS OCI instances and their resources. For cluster, a is appended for primary, b for secondary. Use alphanumeric characters and hyphen("-") only. Cannot end with hyphen.

Compartment

vio

Compartment where to place the storage.

Availability Domain

UZbs-PHX-AD-3

Availability domain where to place the storage.

Fault Domain for Primary

FAULT-DOMAIN-1

Fault domain to place the primary instance.

Fault Domain for Secondary

FAULT-DOMAIN-2

Fault domain to place the secondary instance.

More on regions and availability domains may be found at <https://docs.oracle.com/en-us/iaas/Content/General/Concepts/regions.htm>.

- Configure the variables in the “Cluster Networking” section.
 - In the Cluster Networking Configuration section, use the pulldown menus to choose the Compartment and Subnet to be used for the network the iSCSI and VIO clustering traffic will run on.
 - Enter an unused CIDR block for the Cluster network. A subnet will be created for this block, and the IP addresses used by the Primary VNICS on each Compute Instance will be assigned within this block.

Cluster Networking Configuration

Compartment of Cluster Network VCN

store.dev

Compartment where Cluster Network VCN is configured.

Cluster Network VCN (Non-NAS Network)

zsoci

Cluster Network VCN (Non-NAS Network) where to create a subnet for cluster network.

Subnet CIDR Block for Cluster Network

10.0.66.0/24

CIDR Block to configure the subnet for cluster network (e.g., 10.0.101.0/24).

- Configure the Variables in the Networking Configurations.
 - From the pulldown menus, choose the Compartment and Subnet for the Admin network, which is used to access the browser (BUI) and command line (CLI) interfaces of the ZFS-HA instances. You may optionally enter IP addresses from within the chosen subnet’s range for the VNIC’s IP addresses. If these are not entered, IP addresses will be automatically assigned from the subnet range.

Networking Configuration

Compartment of NAS Admin Network VCN

Networks

Compartment where NAS Admin Network VCN is configured.

Subnet in NAS Admin Network ⓘ

common.sub (Regional)

Subnet where to configure admin access VNICs.

IP Address for Admin Access VNIC on Primary *Optional*

Private IP address to assign for admin access VNIC on the primary instance (e.g., 10.0.1.11). Leave blank for auto assignment.

IP Address for Unused Admin Access VNIC on Primary *Optional*

Private IP address to assign for unused admin access VNIC on the primary instance (e.g., 10.0.1.12). Leave blank for auto assignment.

- From the pulldown menus, choose the Compartment and Subnet for the NAS Data network, which is used by clients to mount shares.
You may optionally enter IP addresses from within the chosen subnet's range for the VNIC's IP addresses. If these are not entered, IP addresses will be automatically assigned from the subnet range. Note that the "NAS Data IO" address is the one that will be used by the clients and may be moved between the cluster instances depending on which instance is active for the pool the address is associated with.

Compartment of NAS Data Network VCN

Networks

Compartment where NAS Data Network VCN is configured.

Subnet in NAS Data Network ⓘ

common.sub (Regional)

Subnet where to configure data access VNICs and IP address for NAS data IO.

IP Address for Data Access VNIC on Primary *Optional*

Private IP address to assign for data access VNIC on the primary instance (e.g., 10.0.3.11). Leave blank for auto assignment.

IP Address for NAS Data IO on Primary *Optional*

IP address for NAS data traffic on the primary instance (e.g., 10.0.3.12). Leave blank for auto assignment.

- In the Data Volume Configuration, enter the number and size of the Block Volumes used for each storage pool. In an Active/Active cluster, the two pools do not need to be the same size. Note that there is a limit of 32 volumes across both pools. Only one pool will be created for an Active/Passive pool.

Data Volume Configuration

Number of Block Volumes for Primary Storage Pool

2

Number of block volumes (data disks) to create the primary storage pool. Max 32 block volumes in total per storage.

Number of Block Volumes for Secondary Storage Pool

2

Number of block volumes (data disks) to create the secondary storage pool. The secondary storage pool is created only for Active/Active cluster configuration.

Block Volume Size (GBs)

50

Size of each block volume in GBs: 50GB - 32768GB(32TB).

- In the storage section, choose an SSH public key file or paste an SSH public key. There is no need to modify the User Init Data fields. Click Next when complete.

Storage Settings

SSH Public Key for Admin User (opc)

☒ Choose SSH Key File ☐ Paste SSH Key

Drop a file [Browse](#)

SSH public key (.pub) file only.

ssh-key-2022-04-01.key.pub x

SSH public key for admin user (opc) on the instance.

- Verify the values that have been entered for the variables. Click Previous if changes need to be made. When complete, click Save changes.

Stack Information

Verify your configuration variables, and then create your stack. Due to limited space, we show only variables without default values or that you edited.

Configure Variables

Review

Name	JH Stack
Description	...in OCI Show Copy
Terraform version	1.0.x

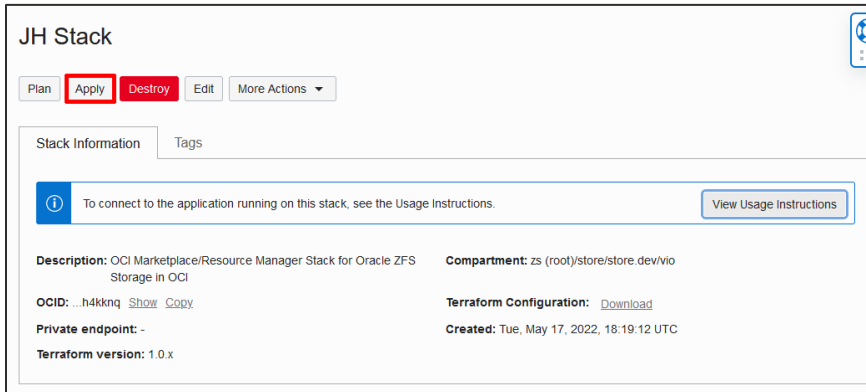
Storage Configuration	Active/Active
Compute Instance Shape	VM.Standard2.8
Storage Name	jh-zfs
Compartment	...667hvq Show Copy
Availability Domain	IZb6-PHX-AD-3

Cluster Networking Configuration
Commitment of Cluster Network VCN

Previous Save Changes Cancel

APPLY THE STACK

To begin the process of creating the stack, click the Apply button. If you wish to review the Terraform execution plan before applying the stack, click the Plan button and review the log from that action.



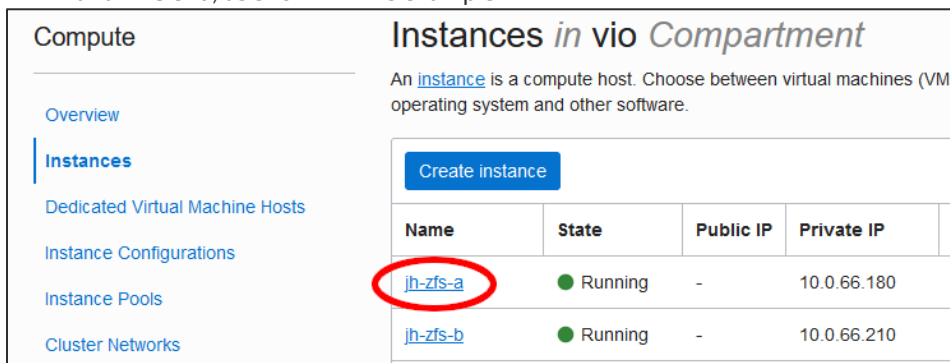
It will take approximately five minutes for the stack to build the ZFS-HA cluster.

SET A PASSWORD FOR ZFS ADMINISTRATION

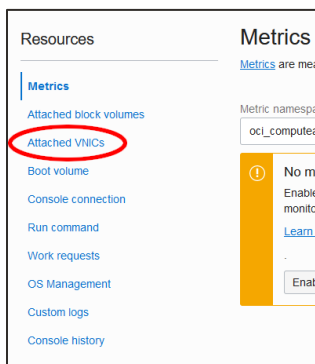
Once the Stack has been applied and completed successfully, a password must be set to allow administrators to manage the storage on the cluster via either the BUI or CLI. Connect to the primary ZFS-HA instance by using SSH to connect to the Admin VNIC on the primary ZFS-HA Compute instance.

This address may have been configured as the “IP Address for Admin Access VNIC on Primary” variable in the stack. If the address was automatically assigned, it may be found with the following steps:

- List the instances in the compartment and select the instance with the name given in the variable configuration with “-a” at the end, as shown in this example:



- On the Instance details screen, scroll down until the Resources menu is shown on the left side of the screen and choose “Attached VNICs”.



- In the list of Attached VNICs, find the VNIC that ends with “-adm-a” and select it.

Attached VNICs

A [virtual network interface card \(VNIC\)](#) lets an instance connect to a virtual cloud network and outside the VCN.

[Create VNIC](#)

Name	Subnet or VLAN ⓘ	State
jh-zfs-0-a (Primary VNIC)	Subnet - jh-zfs	● Attached
jh-zfs-data-a	Subnet - common.sub	● Attached
jh-zfs-dx-a	Subnet - common.sub	● Attached
jh-zfs-adm-a	Subnet - common.sub	● Attached
jh-zfs-ax-a	Subnet - common.sub	● Attached

- Find the VNIC’s Private IP address and copy it.

VNIC Information Tags

VNIC Information

OCID: ...iclsrq [Show](#) [Copy](#)

Created: Thu, May 19, 2022, 17:27:56 UTC

Compartment: zs (root)/Networks

Subnet: [common.sub](#)

Primary IP Information

Private IP Address: [100.104.21.251](#)

Private IP OCID: ...olnt2q [Show](#) [Copy](#)

Assigned: Thu, May 19, 2022, 17:27:53 UTC

Network Security Groups: None [Edit](#)

- Using your tool of choice, SSH to the address copied in the above step. Use the private part of the SSH key applied in the Stack variable configuration process and use “opc” as the user.

The instance includes the `opc` user by default. The `opc` account has all authorizations enabled and can be used to configure the storage appliance. If root user access is needed, see https://support.oracle.com/knowledge/Sun%20Microsystems/2811414_1.html.

You can transition to a full administrative-capability root account once you have logged in as the `opc` user if you need full administrative access to the instance.

Run the commands as shown in this example, using your own password where the asterisks are shown:

```
ssh -i .ssh/opc opc@100.104.21.251
```

```
jh-zfs-a:> configuration users
jh-zfs-a:configuration users> select opc
jh-zfs-a:configuration users opc> set initial_password
Enter new initial_password: *****
Re-enter new initial_password: *****
Initial_password - (set) (uncommitted)
jh-zfs-a:configuration users opc> commit
jh-zfs-a:configuration users> exit
```

CONNECT TO THE BROWSER USER INTERFACE (BUI)

Log in to the BUI by connecting your browser to https://<primary_admin_address>:215 and using “opc” as the username with the password entered in the previous section. Note that because a self-signed certificate is used for HTTPS encryption, your browser may identify the site as a security risk. You may safely continue to the site.

The BUI may now be used to create shares or perform other ZFS Appliance administration tasks.

Active/Active Clustering

NOTE: If this is an Active/Active cluster, the secondary ZFS-HA instance will have control of all shared resources. Log into the BUI of the secondary ZFS-HA instance, navigate to the Configuration->Network screen, and click the Failback button to move the shared resources owned by the primary instance to where they belong.

The screenshot shows the ZFS Appliance BUI Configuration - Network screen. The top navigation bar includes Configuration, Maintenance, Shares, Status, and Analytics. Below this, there are tabs for SERVICES, STORAGE, NETWORK, SAN, CLUSTER, USERS, PREFERENCES, SETTINGS, and ALERTS. The CLUSTER tab is selected, and the FAILBACK button is circled in red. Below the navigation bar, there are buttons for SETUP, UNCONF, FAILBACK, KEOVER, REVERT, and APPLY. The main content area shows two cluster nodes: jh-zfs-b (Active (takeover completed)) and jh-zfs-a (Ready (waiting for failback)). A diagram shows a network topology with vtinet0 interfaces connected by a line. Below the diagram, there are two sections for Active Resources. The jh-zfs-b section lists resources and their owners, while the jh-zfs-a section shows no resources are active on this cluster node.

RESOURCE	OWNER
<--> jh-zfs-a (net/vtinet1) 100.102.221.96	jh-zfs-a
<--> jh-zfs-b (net/vtinet2) 100.102.208.147	jh-zfs-b
<--> jh-zfs-adm-b (net/vtinet4) 100.102.222.73	jh-zfs-b
zfs/pool-a 97.9G	jh-zfs-a
zfs/pool-b 97.9G	jh-zfs-b

After the Failback has completed, all resources will be on the appropriate instances, as shown here.


The screenshot shows the ZFS Appliance BUI Configuration - Network screen after the Failback operation has completed. The top navigation bar and tabs are the same as in the previous screenshot. The main content area shows two cluster nodes: jh-zfs-b (Active) and jh-zfs-a (Active). A diagram shows a network topology with vtinet0 interfaces connected by a line. Below the diagram, there are two sections for Active Resources. The jh-zfs-b section lists resources and their owners, while the jh-zfs-a section lists resources and their owners.

RESOURCE	OWNER
<--> jh-zfs-b (net/vtinet2) 100.102.208.147	jh-zfs-b
<--> jh-zfs-adm-b (net/vtinet4) 100.102.222.73	jh-zfs-b
zfs/pool-b 97.9G	jh-zfs-b

RESOURCE	OWNER
<--> jh-zfs-a (net/vtinet1) 100.102.221.96	jh-zfs-a
zfs/pool-a	jh-zfs-a

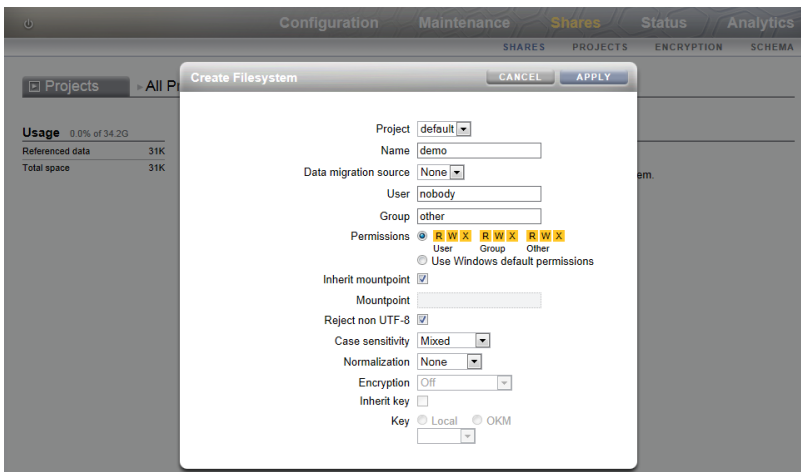
SHARE AN SMB FILESYSTEM


Complete the following steps to set up a simple filesystem share over Server Message Block (SMB) with Windows user access. Begin by logging in to the BUI by connecting your browser to https://<primary_admin_address>:215 and using “opc” as the username with the password entered in the section “Set a Password for ZFS Administration”.

1. Navigate to the Shares screen.
Click the add item icon  next to Filesystems to create a new filesystem.



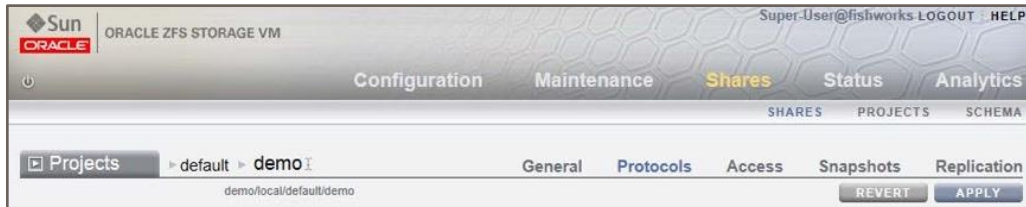
2. Name the filesystem and change the permissions for Group and Other to allow anyone to read, write, and execute on the filesystem. In this example, the filesystem is named demo. The filesystem is part of the default project. Click APPLY to save the changes.



3. In the Shares screen, mouse over the entry for the new filesystem and click the edit icon  to edit the filesystem attributes.

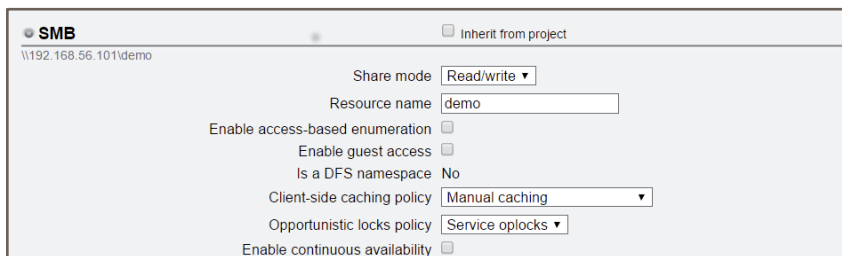


- Click Protocols.

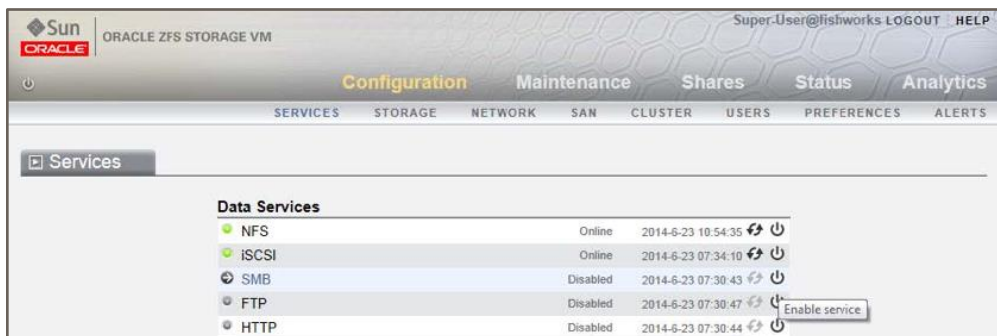


- In the SMB section, clear the checkbox for Inherit from project, select Read/Write Shareable in the Share mode drop-down list, and set the Resource Name.

In this example, the Resource Name is demo. Click APPLY to save the changes.




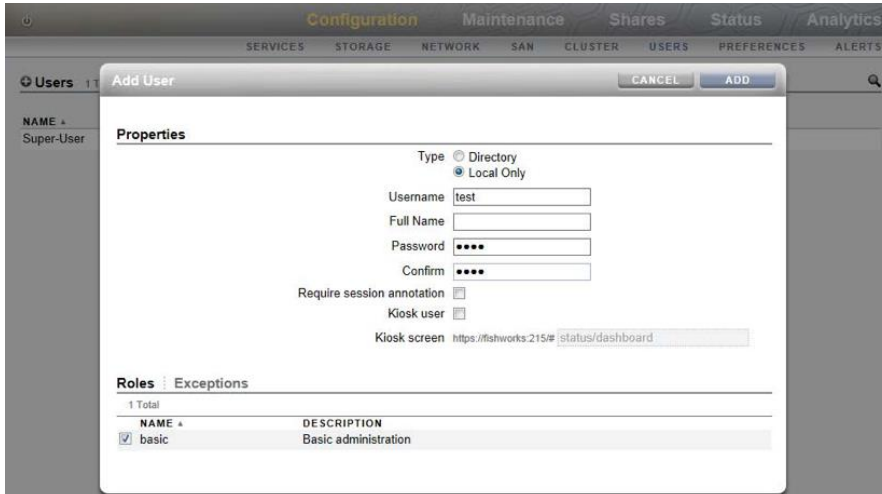
- Select the Configuration tab to access the Configuration Services screen.
- Enable the SMB service by clicking the power icon.



The state will change from Disabled to Online.

8. Configure a user with access to the filesystem share.

- a. Click USERS in the navigation bar, and click the add item icon  next to Users to create a new user.
- b. Select Local Only, set the Username and Password, and click ADD. Log out of the BUI by clicking LOGOUT near the top of the screen.



The screenshot shows the 'Add User' dialog box in the Oracle ZFS Storage Management console. The dialog has a 'Properties' tab and a 'Roles' section. The 'Properties' section includes fields for Username (test), Full Name, Password, and Confirm. The 'Type' is set to 'Local Only'. There are checkboxes for 'Require session annotation' and 'Kiosk user'. The 'Kiosk screen' field contains the URL 'https://fishworks.215#status/dashboard'. The 'Roles' section shows a table with one role, 'basic', which is selected.


NAME	DESCRIPTION
basic	Basic administration

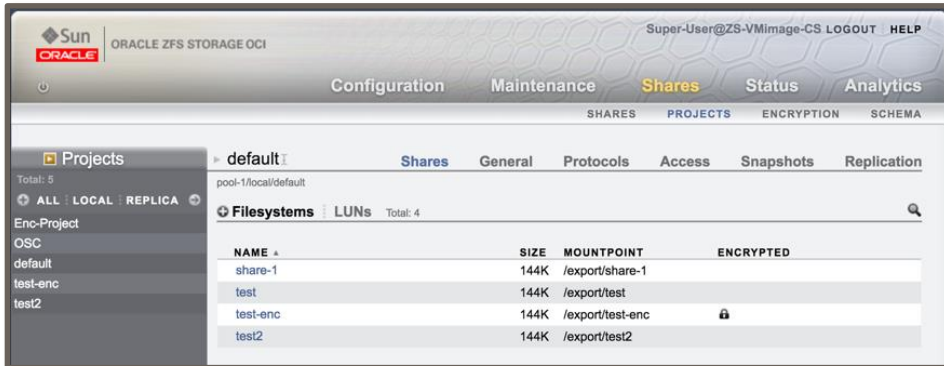
9. From a Windows client, connect to the NAS Data IO of your ZFS Storage instance, and log in with the credentials you set in step 8 to access the shared filesystem.

SHARE AN NFS FILESYSTEM

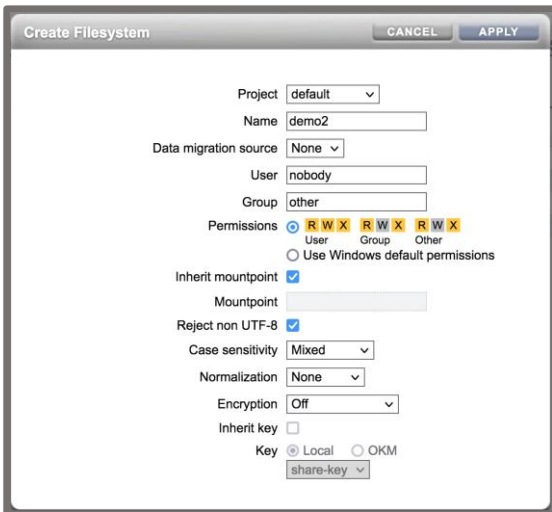
Complete the following steps to set up a simple filesystem share over NFS to share with an NFS client or clients. Begin by logging in to the BUI by connecting your browser to https://<primary_admin_address>:215 and using “opc” as the username with the password entered in the section “Set a Password for ZFS Administration”.

1. Navigate to the Shares screen.

Click the add item icon  next to Filesystems to create a new filesystem. Projects provide an administrative point for filesystems so you can set properties at the project level that are inherited by filesystems within the project. The system includes the default project.



2. Name the filesystem and change the permissions to match the user/group requirements. In this example, the filesystem is named demo2. The filesystem is part of the default project. Click APPLY to save the changes.



Project: default

Name: demo2

Data migration source: None

User: nobody

Group: other

Permissions: ☒ User ☒ Group ☒ Other

☐ Use Windows default permissions

☒ Inherit mountpoint

Mountpoint:

☒ Reject non UTF-8

Case sensitivity: Mixed


Normalization: None

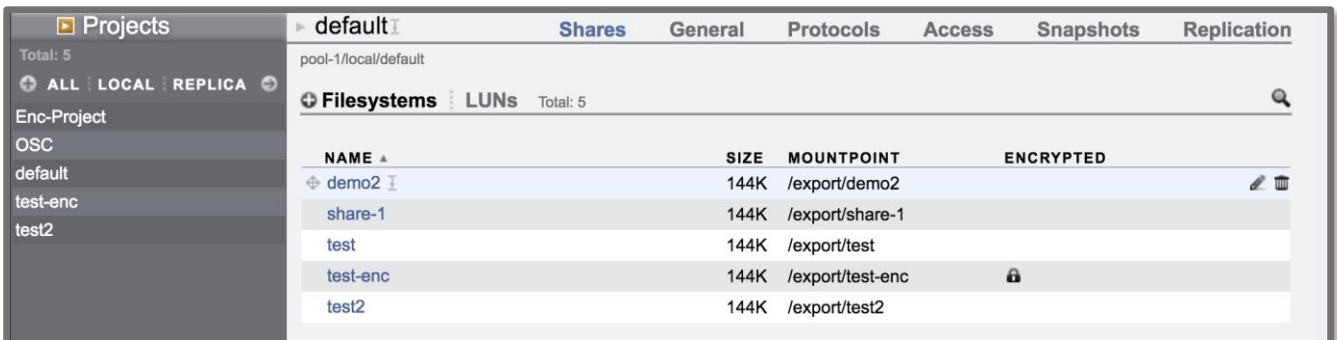
Encryption: Off

☐ Inherit key

Key: ☒ Local ☐ OKM

share-key

3. In the Shares screen, mouse over the entry for the new filesystem and click the edit icon  to edit the filesystem attributes.



- Click Protocols. In the NFS section, set the Share mode to Read/write in the pulldown menu, if it is not inherited from the project. Click APPLY.

NFS ☒ Inherit from project

Share mode: Read/write

Disable setuid/setgid file creation: ☐

Prevent clients from mounting subdirectories: ☐

Anonymous user mapping: nobody

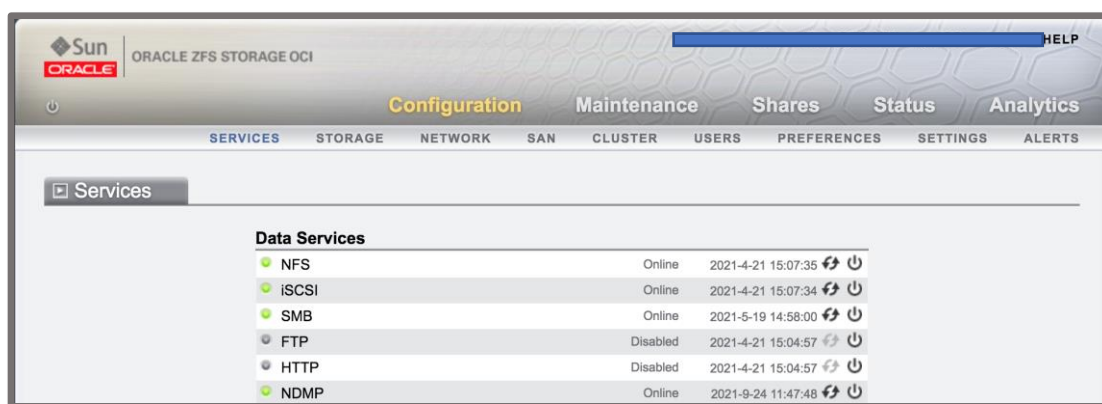
Character set: default

Security mode: System Authentication

Enforce reserved ports for system authentication: ☐

NFS Exceptions

- Select the Configuration tab to access the Configuration Services screen.
- Enable the NFS service by clicking the power icon if it is not already enabled.



- Mount the filesystem over NFS with syntax similar to the following:

```
% mount -t nfs <NAS_Data_IO_address>:/export/demo2 /mnt
```

Use the IP address assigned to the NAS Data IO IP address on the primary ZFS instance.

UPGRADING YOUR ZFS STORAGE INSTANCE

When new ZFS Storage images are available from the OCI Marketplace listing, you can upgrade your running instances. For information about upgrading your ZFS Storage instances, see the following doc: Oracle Support Document 2817714.1 (How to Upgrade a ZFS Storage on Oracle Cloud Infrastructure (OCI) Marketplace Deployment) can be found at: <https://support.oracle.com/epmos/faces/DocumentDisplay?id=2817714.1>.

ZFS IN OCI INSTANCE BEST PRACTICES

Network Best Practices

ZFS Storage in OCI Network Routing

- It is recommended to set the multihoming model to strict.
- Create a network IPv4 route on the primary network interface with the destination set to 169.254.0.0/16 for iSCSI traffic to increase network throughput.

ZFS Storage in OCI Network Datalinks

- Link Speed, Link Duplex and Flow Control should all be set to Auto.
- Link speed for VM instances will be reported as 1GB but will actually use the full amount of bandwidth allocated to the instance. (See known issues)
- All network datalinks should have the MTU set to 9000 for best performance.

ZFS Storage in OCI Network Interfaces

- The primary network interface used for iSCSI traffic should not be modified because it can cause a system panic. (See known issues)
- Consider using separate subnets for storage administrators and NAS clients for enhanced security.
- NAS client interfaces should uncheck 'Allow Administration' for enhanced security.

Block Storage Best Practices

System Boot Disk

- System disk contains read only OS image, logs, core dumps and configurations.
- Configuration data can be backed up using 'Maintenance System Configs'
- Does not include OS image, logs, core dumps, replication or share data.
- Logs and core dumps can be saved using 'Maintenance System Bundles'
- Entire system disk can be backed up using OCI boot volume backups.

Storage Pools

- Pool disks contain all configuration data under 'Shares'
- All disks in each pool should be same size especially if they are under 800GB.
- All data disks in each pool should have the same performance settings.
- Suggest creating a volume group containing all data disks for each storage pool.
- Block volume backups must use volume groups to keep pool data consistent.
- For best system resource usage, it is recommended to have only one pool per VM.
- All data disks provided by OCI have multiple copies so striped pools provide data protection. ZFS will detect bit rot but data will have to be restored from backup if bit rot is detected.
- Consider backing up data disks or using a single-parity or mirrored storage profile to protect against bit rot or a block volume outage.

Backup of ZFS Configuration

We recommend that after your ZFS Storage in OCI instance is configured, that you create a backup of the configuration with the following steps:

- From the Appliance BUI, go to Maintenance→System.
- Under the Configurations section, click Backup.
- This will create a backup of the Appliance configuration, that can be downloaded and stored separately for recover purposes.

For information about the configuration backup content, what is included and what is not included, see [Backing Up the Configuration](#).

Block Volume Backups

OCI Block Volume Service allows you to create snapshots of both boot volume and block volumes.

- Boot Volume snapshots
 - <https://docs.oracle.com/en-us/iaas/Content/Block/Tasks/backingupabootvolume.htm>
- Block Volume Backups
 - <https://docs.oracle.com/en-us/iaas/Content/Block/Concepts/blockvolumebackups.htm>

SECURITY REFERENCES

For information about setting permissions on shares and recommended security practices, see the following references:

- [Access Control Lists for Filesystems](#)
- [Oracle® ZFS Storage Appliance Security Guide. Release OS8.8.x](#)

INSTALLATION NOTES

Root User Configuration

You will need to configure the root user to perform some tasks such as taking a configuration backup, configuring replication, or even logging in remotely as root or using the `su` command to become root.

To enable root login over ssh, from the Appliance BUI, go to the Configuration tab to reach the Configuration Services screen. Under Remote Access, select the ssh service. From the ssh service screen, enable Permit root login.

For more detailed configuration information, see [My Oracle Support Doc ID 2811414.1](#).

Known Issues

- Virtual Machine instances will show network devices speed as 1Gb even though it will use the full bandwidth allowed by the compute shape. (32749253 - VNICs speed is mentioned as 1G at CLI/BUI though VNIC effective bandwidth is more)
- Destroying the cluster with the stack fails when detaching VNICs from the Compute instances. The workaround is to terminate the cluster nodes from the OCI console and rerun the “Destroy” action.
- If a new OCI VNIC is added to a running ZFS Storage in OCI VM, a reboot is required before the network device can be used. (32518670 - Adding an additional VNIC to the OCI ZFS-HA VM fails)
- If OCI VNICs are added before a VM instance finishes its first boot there is a chance the instance will hang. The workaround is to wait for system to finish booting before adding OCI VNICs and then reboot the instance to pick up the new VNICs. (34045542 - ZFSSA hangs if OCI VNICs are added while booting large VM shapes)
- Failback might fail to migrate back the secondary IP address to its owner head. This happens when the DNS service is setup with DNS servers other than 169.254.169.254. (34064814 - failback sometimes might not return the secondary IP address) A general workaround is to add 169.254.169.254 as an additional DNS server. If a system is already in this state without an NAS IP working, the following command can be used to retry the IP migration:

```
python -mak.nori.viorpc cluster.clustered
```
- After VIO clustering is done with VM instances, any new iSCSI disk attachment will automatically attach the same disk as PARAVIRTUALIZED to the peer head. (34086951 - OCI volume attach type should match peer's type) The workaround is to manually reattach disks as iSCSI while the node is in STRIPPED state.

APPENDIX A – INSTALLATION CHECKLIST

ZFS-HA Configuration and Placement

OCI Region (Select from OCI console before getting Stack)	
Stack Name	
Cluster Type	Active/Active or Active/Passive
Bare Metal or Virtual Machine	
Shape	
Storage (appliance host) Name	
Compute and Block Storage Compartment	
Availability Domain	
Fault Domain 1	
Fault Domain 2	
SSH Key File Location	

Cluster Network Configuration

Compartment of Cluster Network VCN	
Cluster Network VCN	
Cluster Network subnet CIDR block	

Admin Network Configuration

Compartment of NAS Admin Network VCN	
Subnet in NAS Admin Network	
IP Address for Admin Access VNIC on Primary	*
IP Address for Unused Admin Access VNIC on Primary	*
IP Address for Admin Access VNIC on Secondary	*
IP Address for Unused Admin Access VNIC on Secondary	*

* IP Addresses will be assigned from chosen Subnet range if not defined

NAS Data Network Configuration

Compartment of NAS Data Network VCN	
Subnet in NAS Data Network	
IP Address for Data Access VNIC on Primary	*
IP Address for Pool-a Data IO on Primary (used by the clients)	*
IP Address for Unused Data Access VNIC on Primary	*
IP Address for Data Access VNIC on Secondary	*
IP Address for Pool-b Data IO on Secondary	*
IP Address for Unused Data Access VNIC on Secondary	*

* IP Addresses will be assigned from chosen Subnet range if not defined

Data Volume Configuration

Number of Block Volumes for Primary Storage Pool	(32 Max across all pools)
Number of Block Volumes for Secondary Storage Pool	
Block Volume Size (50GB - 32768GB)	

CONNECT WITH US

Call +1.800.ORACLE1 or visit oracle.com.

Outside North America, find your local office at oracle.com/contact.

 blogs.oracle.com

 facebook.com/oracle

 twitter.com/oracle

Copyright © 2022, 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. 0120

User Guide

November 2022

Author: Joe Hartley, zfssa-storage-feedback_ww_grp@oracle.com