



# Oracle ZFS Storage - High Availability User Guide

---

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

October 2023 | Version 4.11  
Copyright © 2023, Oracle and/or its affiliates  
Public

# CONTENTS

<b>Purpose Statement</b>	<b>4</b>
<b>Disclaimer</b>	<b>4</b>
<b>Introduction</b>	<b>5</b>
Overview of Configuration Steps	6
OCI Requirements for ZFS-HA Clusters	7
ZFS Compute Instance and Network Requirements	7
Example - OCI Configuration for a ZFS-HA Cluster	9
Second Example – Simplified OCI Configuration for a ZFS-HA Cluster	11
<b>First Steps</b>	<b>11</b>
<b>Installation Checklist</b>	<b>12</b>
<b>Run The ZFS Storage Deployment Tool From OCI Marketplace</b>	<b>13</b>
<b>Configure the Deployment Tool Variables</b>	<b>14</b>
<b>Apply the Stack</b>	<b>18</b>
<b>Set a Password for ZFS Administration</b>	<b>19</b>
<b>Connect to the Browser User Interface (BUI)</b>	<b>21</b>
Active/Active Clustering	21
ZFS-HA Networking Recommendations	22
ZFS-HA Network -> Configuration -> Routing	22
ZFS-HA Network -> Configuration -> Datalinks	22
ZFS-HA Network -> Configuration->Interfaces	22
<b>Share An SMB Filesystem</b>	<b>23</b>
<b>Share An NFS Filesystem</b>	<b>26</b>
<b>Deep Dive - Cluster Configuration Overview</b>	<b>28</b>
High Availability Clustering	28
Active/Passive Clustering	29
Active/Active Clustering	30
Clustered Instance Terminology	31
Clustered Configuration Operation	31
<b>OS8.8.X Documentation Specific to ZFS-HA</b>	<b>32</b>
Automatically Expanding a ZFS-HA Storage Pool	32
Set Automatic Expansion of a pool using the BUI	32
Set Automatic Expansion of a pool using the CLI	33
Set Automatic Expansion of a pool using the RESTful API	33
Manually Expanding a ZFS-HA Storage Pool	34
Expanding Existing Block Volumes	34
Creating New Block Volumes	36
Adding New Block Volumes to a Pool	38
Adding Clustered Interfaces	40
Overview - Active/Active Cluster	40
Overview - Active/Passive Cluster	40
1 - Attach New VNICs In the OCI Console	40
2 - Assign Secondary IP Addresses	42
3 - Reboot Both Controllers	42
4 – Run Takeover/Failback	42
5 – Edit New Interfaces	43
About Devices, Datalinks, and Interfaces	43
6 – Reboot and Rebalance	44

<b>Upgrading Your ZFS-HA Instance</b>	<b>46</b>
<b>ZFS-HA System Notes</b>	<b>47</b>
Networking	47
ZFS-HA Network Routing	47
ZFS-HA Network Datalinks	47
ZFS-HA Network Interfaces	47
Block Storage Notes	47
System Boot Disk	47
Storage Pools	47
Boot and Block Volume Backups	47
Backing Up the ZFS Configuration	48
Submitting a Service Request	49
<b>Documentation and Security References</b>	<b>50</b>
<b>Installation Notes</b>	<b>50</b>
Root User Configuration	50
Known Issues	50

## 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-HA Storage Deployment Tool.

The Oracle ZFS-HA 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. There is no charge for the use of the Deployment Tool.

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.

## 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.

## INTRODUCTION

Oracle is uniquely positioned to provide products and services that run 24/7 either on-premises or in the cloud and our expertise helps us optimally operate our engineered products in Oracle's cloud infrastructure.

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-HA 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 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
<b>VM.Standard2.4</b>	60GB	4.1 Gbps	256 MB/s	192 MB/s	Tens
<b>VM.Standard2.8</b>	120GB	8.2 Gbps	512 MB/s	384 MB/s	Hundred
<b>VM.Standard2.16</b>	240GB	16.4 Gbps	1025 MB/s	768 MB/s	Few Hundred
<b>VM.Standard2.24</b>	320GB	24.6 Gbps	1537 MB/s	1150 MB/s	Hundreds
<b>VM.Standard3.Flex</b>	512GB	32 Gbps	2000 MB/s	1500 MB/s	Thousand
<b>VM.Standard.E4.Flex</b>	1024GB	40 Gbps	2500 MB/s	1875 MB/s	Thousands
<b>BM.Standard2.52</b>	768GB	25x2 Gbps	3125 MB/s	2343 MB/s	Thousands
<b>BM.Standard3.64</b>	1024GB	50x2 Gbps	6250 MB/s	4687 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-HA Storage Deployment Tool and contains the following sections:

- Run the Oracle ZFS-HA Storage Deployment Tool from OCI Marketplace
- Configure the Deployment Tool Variables
- Apply the Deployment Tool 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
- APIs for ZFS Storage in OCI – The “Oracle ZFS Storage - High Availability API Guide” provides additional management APIs used developed specifically for Oracle ZFS Storage - High Availability version.

## OCI 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 and Network Requirements

- A dynamic group must be created in **the default identity domain of the root compartment** of the tenancy. This group will be used to grant access to the compartment the ZFS-HA compute instances and block volumes will be created in. This will be referred to as the cluster compartment. The access to the cluster compartment is granted with a rule that points to the cluster compartment's Oracle Cloud Identifier (OCID), such as in the following example. (Note that the example OCID here is much shorter than an actual OCID.)

```
All {instance.compartment.id = 'ocid1.compartment.oc1..aaaa5ldkqg'}
```

Note that if your tenancy is not using Identity Domains, the dynamic group must be created in the root compartment of the tenancy.

For more information, see [Create a Dynamic Group and Matching Rules](#).

- An Identity Policy must be added in the **root compartment** to allow the ZFS-HA compute instances in the cluster compartment to manage the resources needed for the cluster processes. The statements in this policy define entities, permission, and scope of access. The rules for this Identity Policy may grant access to the dynamic group or it may use a group created in the **root compartment** of the tenancy that defines a list of users who will have administrative privileges. The use of a group is not covered in this document; instead, it is assumed that a dynamic group will be used to grant access. The privileges required in either case are:

```
read instance-images in tenancy
manage instances in compartment
manage console-histories in compartment
inspect vnic-attachments in compartment
manage volume-attachments in compartment
manage volumes in compartment
```

Additionally, the following privileges must be given for the compartments that contain each Virtual Cloud Network (VCN) that the instances will use.

```
use vnics in compartment
read private-ips in compartment
- OR -
use private-ips in compartment
```

Any compartment with a VCN that contains VNICs with secondary IP addresses that will be moved between the controller instances at a resource takeover or failback event must have an identity policy rule to “use private-ips” rather than a rule allowing reads. In the example given in the section **Example OCI Configuration for a ZFS-HA Cluster** later in this document, this rule would be needed for the `z_nas_networks` compartment.

These policy rules must be tailored to allow access to the correct compartments and VCNs in your tenancy. 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](#).

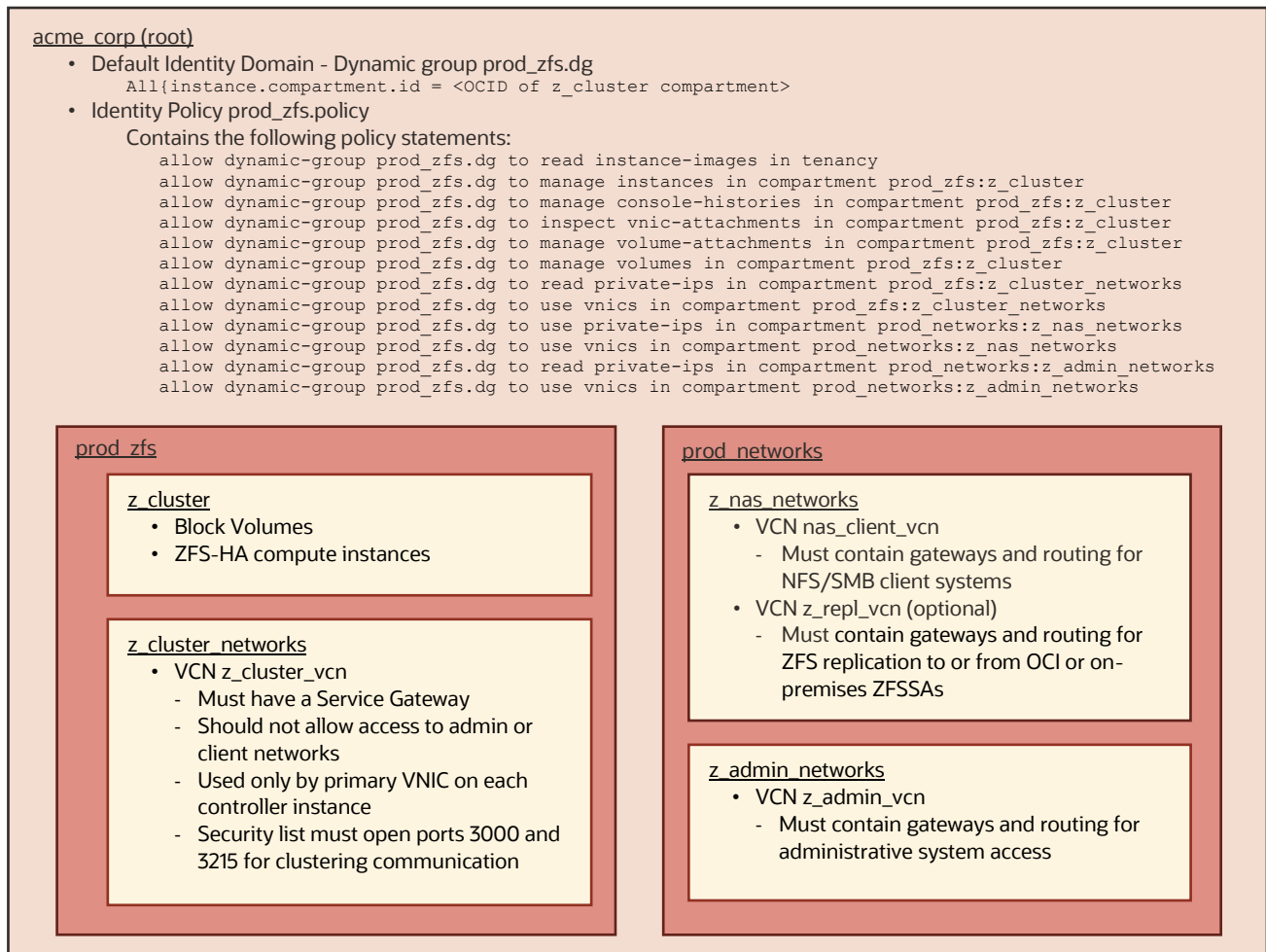
- The Deployment Tool automates the attachment of the ZFS-HA's Virtual Network Interface Cards (VNICs) to VCNs for clustering, administrative, and NAS client connectivity. These VCNs have different requirements, shown here.
  - All VCNs described below and any subnets under them must enable the use of DNS by checking the **"Use DNS hostnames in this VCN/SUBNET"** box. This is the default when creating VCNs or subnets and should not be overridden.
  - Cluster connectivity: The primary VNIC on each ZFS-HA instance is reserved for Block Volume I/O and clustering resources between the two instances. It may also be used for accessing other Oracle services such as ZFS cloud backups to OCI object storage.
    - **There must be an OCI Service Gateway on this VCN for "All <region> Services in Oracle Services Network".**
    - **The security list for this VCN must have ports 3000 and 3215 open to allow communication between the clustered instances.**
    - **The default Route Table for this VCN must have a rule with the target type of "Service Gateway", the destination service of "All <region> Services...", and the target service gateway the name of the service gateway in this VCN.**
    - **Do not create a subnet in this network, but instead identify a subnet range for the tool to create.** The Deployment Tool will create its own subnet in 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 effect on the overall performance of the ZFS-HA system.
  - NAS client connectivity: This VCN must have appropriate gateways, firewall settings, and routing for the client networks. Any OCI compute instance or on-premises computer that accesses data in an ZFS-HA share is considered a client. A subnet must be created under this VCN before the Deployment Tool is run.
  - Administrator access: This VCN must have appropriate gateways, firewall settings, and routing for administrative access to the ZFS-HA controller instances, including CLI, REST, and BUI access. It is recommended that the Admin interface be on a separate VCN for security, but it is not a requirement. A subnet must be created under this VCN before the Deployment Tool is run.
  - Replication (optional): A feature of Oracle ZFS Storage Appliances, whether in 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 cloned to be mountable shares on those appliances. While most installations will use the NAS client VCN for replication, it may be advantageous to designate a separate VCN for this traffic to keep it off the VCNs used for client and administrative traffic.

If this is desired, a pair of clustered interfaces can be created and assigned to a new VCN (see **Adding Clustered Interfaces** later in this guide). This VCN must have appropriate gateways and routing to reach the other appliances involved in ZFS replication. Note that replication network configuration is not a part of the Deployment Tool. Additional Identity Policy rules may be needed for VNIC and private-ip access if additional VCNs or compartments are used.



## Example - OCI Configuration for a ZFS-HA Cluster

The following block diagram shows the OCI components and their relative locations within the tenancy based on the requirements above.



The large outer box represents the tenancy's root compartment. It contains `prod_zfs.dg`, a dynamic group in the default identity domain of the root compartment that points to the `prod_zfs` compartment, and `prod_zfs.policy`, the Identity Policy granting the privileges needed to the dynamic group `prod_zfs.dg`. Note that if your tenancy is not using Identity Domains, the dynamic group must be added in the root compartment.

The inner boxes represent the `prod_zfs` and `prod_networks` compartments. `Prod_zfs` contains two inner compartments. These are `z_cluster`, which contains the ZFS-HA instances and the block volumes used for the storage pools, and `z_cluster_networks`, which contains the VCN used for block volume I/O and cluster traffic.

The `prod_networks` compartment which holds two compartments which hold the networking components used by the corporation to control traffic in and out of the tenancy boundaries. Within this compartment is the `z_nas_networks` compartment which in turn holds `nas_client_vcn`, which is used to define NAS client access throughout the tenancy. Another compartment in `prod_networks` is `z_admin_networks`, which contains `z_admin_networks`, the VCN used to access the admin interfaces on the ZFS-HA instances.

Please note that it is not necessary for the `nas_client_vcn` nor the `z_admin_vcn` to be in their own compartments. The example above illustrates how to use sub-compartments for the VCNs, but many installations will use an existing Networking compartment for the NAS and administrative access. Whatever the configuration, the important point is to have the correct Policy Statements for the compartments and VCNs used.

It's worth taking a closer look at the Identity Policy `prod_zfs.policy` to break down the rules being used and what they apply to. The first rule is the only one global to the entire tenancy. Its purpose is to allow the creation of instances in `prod_zfs` and to probe other instances for their configuration so that the ZFS-HA Deployment Tool can properly set IP addresses and other variables:

```
allow dynamic-group prod_zfs.dg to read instance-images in tenancy
```

The next set of rules give the Deployment Tool permission to create and use block volumes and compute instances in the compartment `z_cluster`. These rules also allow a clustered instance to take over resources such as shared IP addresses and the storage pool should the other instance become unavailable.

```
allow dynamic-group prod_zfs.dg to manage instances in compartment prod_zfs:z_cluster
allow dynamic-group prod_zfs.dg to manage console-histories in compartment prod_zfs:z_cluster
allow dynamic-group prod_zfs.dg to inspect vnic-attachments in compartment prod_zfs:z_cluster
allow dynamic-group prod_zfs.dg to manage volume-attachments in compartment prod_zfs:z_cluster
allow dynamic-group prod_zfs.dg to use volumes in compartment prod_zfs:z_cluster
```

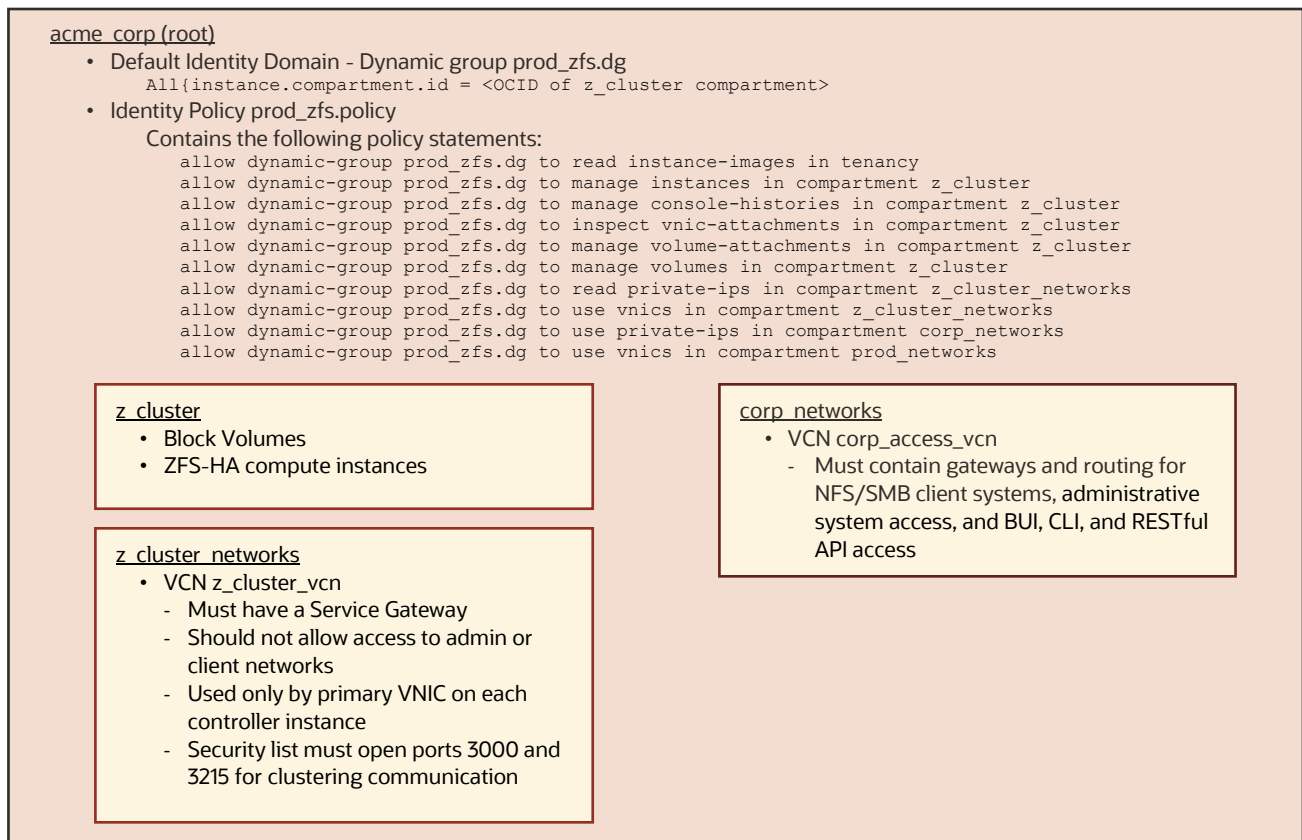
The final set of rules gives the ZFS-HA instances in the `prod_zfs` compartment privileges to the networking resources in the appropriate compartments. Note that access is given to the compartments that contain the `z_nas_networks` and `z_admin_networks` VCNs separately. Giving access to the `prod_networks` compartment may have inadvertently allowed access to other resources that should have remained restricted. It is always recommended to create policy rules to grant the lowest level of privilege possible while maintaining functionality.

```
allow dynamic-group prod_zfs.dg to read private-ips in compartment prod_networks:z_cluster_networks
allow dynamic-group prod_zfs.dg to use vnics in compartment prod_networks:z_cluster_networks
allow dynamic-group prod_zfs.dg to use private-ips in compartment prod_networks:z_nas_networks
allow dynamic-group prod_zfs.dg to use vnics in compartment prod_networks:z_nas_networks
allow dynamic-group prod_zfs.dg to read private-ips in compartment prod_networks:z_admin_networks
allow dynamic-group prod_zfs.dg to use vnics in compartment prod_networks:z_admin_networks
```

Note that we have given the `z_nas_networks` compartment the slightly higher privilege to “use private-ips”. This is so the secondary IP addresses used for client mount points can be moved between instances as needed.

## Second Example – Simplified OCI Configuration for a ZFS-HA Cluster

The following block diagram shows an example placing the `z_cluster` and `z_cluster_networks` compartments in the root compartment and using an existing compartment and VCN for NAS and administrative connectivity. This configuration is not recommended and is presented here only for clarification.



Because of the unique requirements of the two compartments used for the ZFS cluster, they remain separate in this example.

## 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

The Installation Checklist on the next page will help in organizing this information so that it is easily available when it is time to run the Deployment Tool.

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.

## 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	*

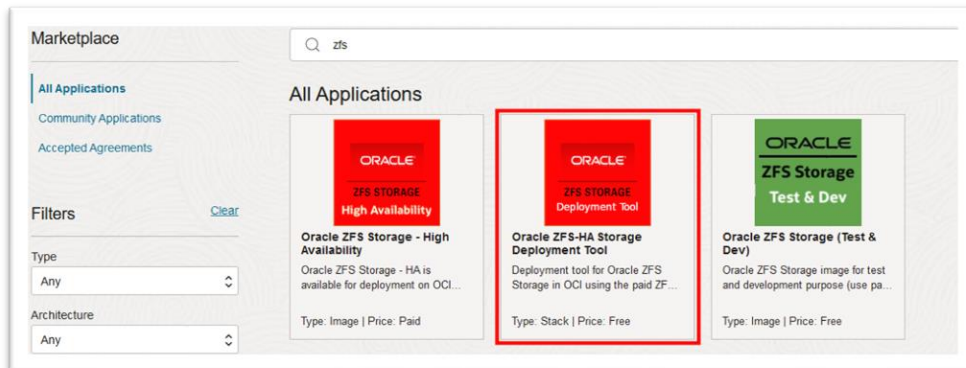
\* IP Addresses will be assigned from chosen Subnet range if not defined

### 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)	

## RUN THE ZFS STORAGE DEPLOYMENT TOOL FROM OCI MARKETPLACE

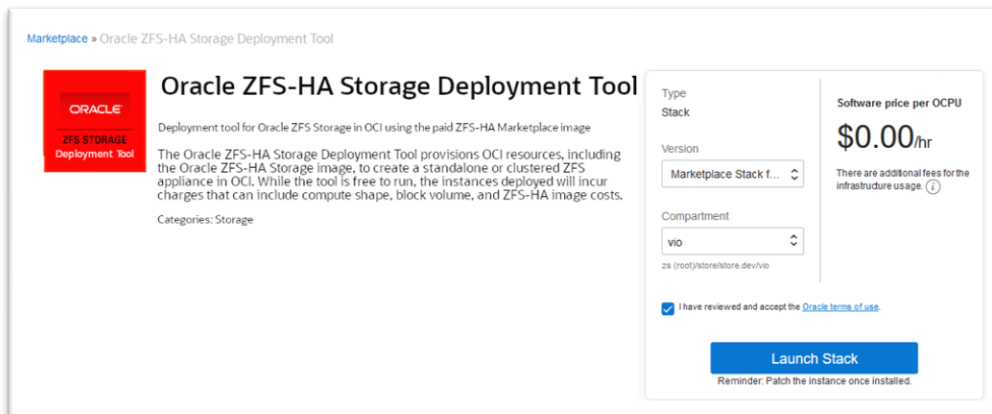
1. Log in to your OCI tenancy and go to the Marketplace and search All Applications for ZFS Storage images. Select the Oracle ZFS-HA Storage Deployment Tool.



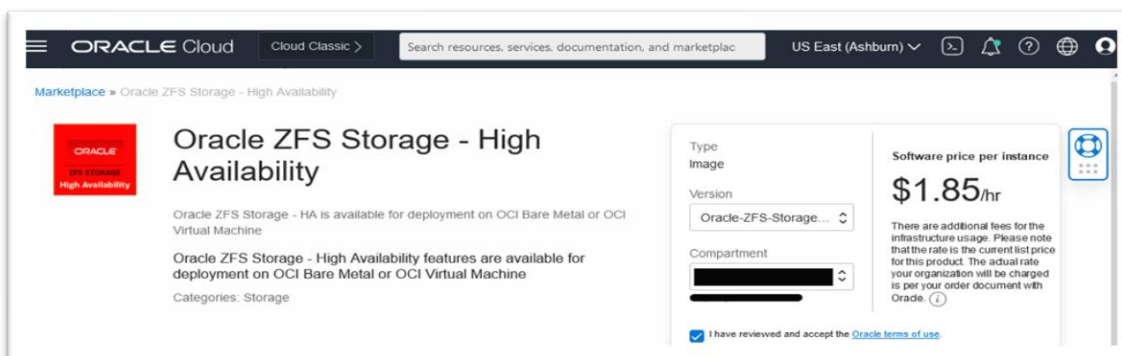
2. Choose the version from the pulldown menu. Except for special cases, the latest version is recommended and is the default choice. As of OS8.8.59, a unified image for both Bare Metal (BM) and Virtual Machine (VM) shapes is provided. In some limited cases a previously released image may be desired, 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  
v10  
zs (root)/store/store.dev/v10

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. Choose SingleHead when high availability is not desired and a single controller is sufficient. Some variables listed here will not appear if SingleHead or 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. The recommended shape (E4.Flex) and configuration (8 OCPU/128GB) is a good starting point for its performance and throughput. When using a Flex shape, the OCPU and memory can be changed even after the ZFS-HA cluster is up and running.

**Storage Configuration and Placement**

Storage Configuration  
Active/Active  
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 supports only SingleHead configuration.

Number of OCPUs *Optional*  
8  
For Active/Active configuration, at least 5 OCPUs is required, 8 OCPUs is recommended.

Memory Size (GBs) *Optional*  
128  
At least 80 GBs is required, 128 GBs is recommended.

### NOTES:

- For Flex shapes, note that the total bandwidth that will be available to the instance is tied to the OCPU count with each OCPU adding 1Gbps to the overall available bandwidth until the maximum bandwidth for the shape is reached as shown in the chart [Network Bandwidth Expectations for NFS/SMB Clients](#).
- An Active/Active cluster 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. An Active/Active cluster will not run on a VM.Standard2.4 shape.

- 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. Running the cluster instances in separate Availability Domains is not supported.

Storage Name  
jh-zfs  
Base hostname for ZS 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  
IZbs-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.

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

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.
  1. 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.
  2. 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.  
**Do not use a CIDR block that has already been assigned to a subnet in this VCN.**

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.
  1. 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.

### Networking Configuration

☐ Admin Network custom configuration  
Check this box to allow users to configure NAS Admin Network access VNICs. Uncheck this box for auto assignment, and to configure from the NAS admin VCN compartment and subnets(This is a default option and recommended).

Compartment of NAS Admin Network VCN  
  
Compartment where NAS Admin Network VCN is configured.

Subnet in NAS Admin Network ⓘ  
  
Subnet where to configure admin access VNICs.

You may optionally enter IP addresses from within the chosen subnet's range for the VNICs that will be created by checking the box for Admin Network Custom Configuration. Any VNICs that do not have an IP address assigned here will have IP addresses automatically assigned from the chosen subnet's range.

### Networking Configuration

☒ Admin Network custom configuration  
Check this box to allow users to configure NAS Admin Network access VNICs. Uncheck this box for auto assignment, and to configure from the NAS admin VCN compartment and subnets(This is a default option and recommended).

Compartment of NAS Admin Network VCN  
  
Compartment where NAS Admin Network VCN is configured.

Subnet in NAS Admin Network ⓘ  
  
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.

IP Address for Admin Access VNIC on Secondary *Optional*  
  
Private IP address to assign for admin access on the secondary instance (e.g., 10.0.1.13). Leave blank for auto assignment.

IP Address for Unused Admin Access VNIC on Secondary *Optional*  
  
Private IP address to assign for unused admin access on the secondary instance (e.g., 10.0.1.14). Leave blank for auto assignment.

2. From the pulldown menus, choose the Compartment and Subnet for the NAS Data network, which is used by clients to mount shares.

☐ Data Network custom configuration  
Check this box to allow users to configure NAS Data Network access VNICs. Uncheck this box for auto assignment, and to configure from the NAS data VCN compartment and subnets(This is a default option and recommended).

Compartment of NAS Data Network VCN  
  
Compartment where NAS Data Network VCN is configured.

Subnet in NAS Data Network ⓘ  
  
Subnet where to configure data access VNICs and IP address for NAS data IO.



As with the Admin network, you may optionally enter IP addresses from within the chosen subnet's range for the VNICs that will be created by checking the box for Data Network Custom Configuration. Any VNICs that do not have an IP address assigned here will have IP addresses automatically assigned from the chosen subnet's 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.

☒ **Data Network custom configuration**  
 Check this box to allow users to configure NAS Data Network access VNICs. Uncheck this box for auto assignment, and to configure from the NAS data VCN compartment and subnets (This is a default option and recommended).

**Compartment of NAS Data Network VCN**  
 Choose...  
 Compartment where NAS Data Network VCN is configured.

**Subnet in NAS Data Network** ⓘ  
 There was an error retrieving options.  
 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.

**IP Address for Data Access VNIC on Secondary** *Optional*  
 Private IP address to assign for data access on the secondary instance (e.g., 10.0.3.14). Leave blank for auto assignment.

The screen shots above were taken while setting up a cluster using the default VM.Standard.E4.Flex shape and active/passive cluster configuration. These options will vary when different shapes or cluster types are chosen.

- 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 and each pool must have a minimum of two volumes. Only one pool will be created for an Active/Passive pool.

**Data Volume Configuration**

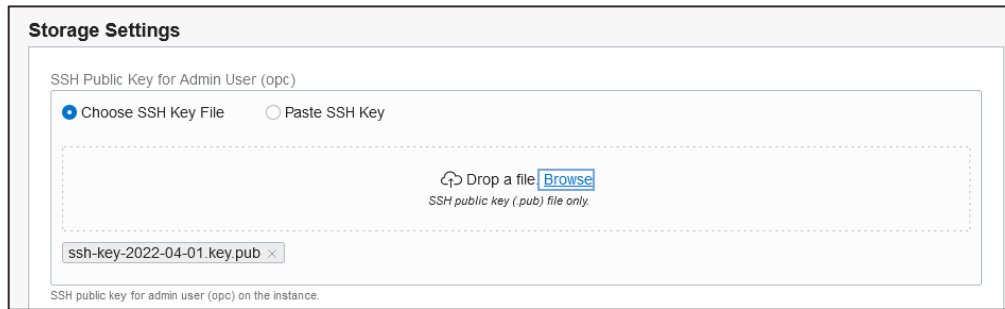
**Number of Block Volumes for Primary Storage Pool**  
 2  
 Number of block volumes (data disks) to create the primary storage pool. Max 32 block volumes in total per storage.

**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).

**NOTE: The size and number of block volumes affects performance.** OCI Block Volumes improve in performance as they grow, and tops out at 1TB. In addition, ZFS storage pool performance increases as the number of block volumes is increased, topping out at around ten volumes. Keep this in mind when determining the number and size of the block volumes used to create a storage pool.

- 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)

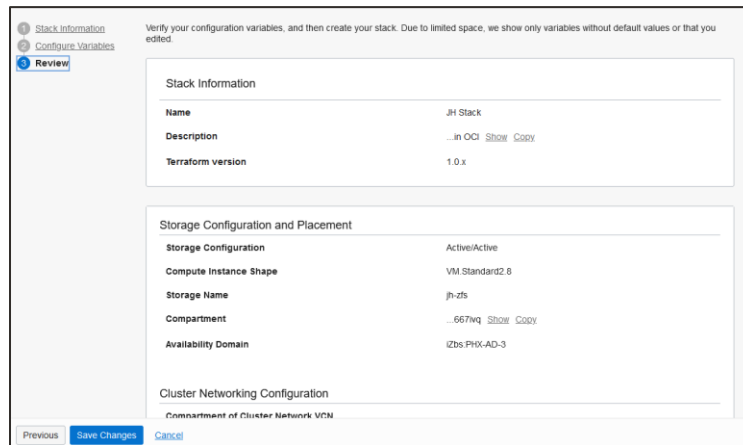
☒ Choose SSH Key File
 ☐ Paste SSH Key

Drop a file [Browse](#)  
SSH public key (.pub) file only

ssh-key-2022-04-01.key.pub ×

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.



1 Stack Information  
2 Configure Variables  
3 **Review**

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

**Stack Information**

Name	JH Stack
Description	...in OCI <a href="#">Show</a> <a href="#">Copy</a>
Terraform version	1.0.x

**Storage Configuration and Placement**

Storage Configuration	Active/Active
Compute Instance Shape	VM.Standard2.8
Storage Name	jh-zfs
Compartment	...667nq <a href="#">Show</a> <a href="#">Copy</a>
Availability Domain	UDb-PHX-AD-3

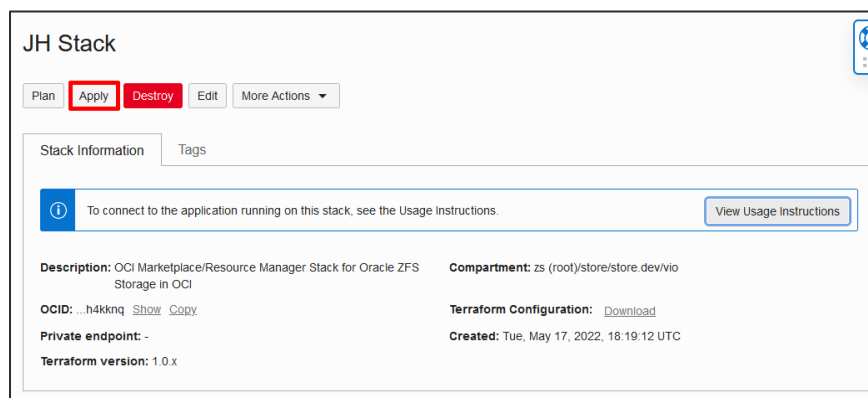
**Cluster Networking Configuration**

Compartment 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.



**JH Stack**

Plan **Apply** Destroy Edit More Actions ▾

Stack Information Tags

**Info** To connect to the application running on this stack, see the Usage Instructions. [View Usage Instructions](#)

**Description:** OCI Marketplace/Resource Manager Stack for Oracle ZFS Storage in OCI **Compartment:** zs (root)/store/store.dev/vio

**OCID:** ...h4kknq [Show](#) [Copy](#) **Terraform Configuration:** [Download](#)

**Private endpoint:** - **Created:** Tue, May 17, 2022, 18:19:12 UTC

**Terraform version:** 1.0.x

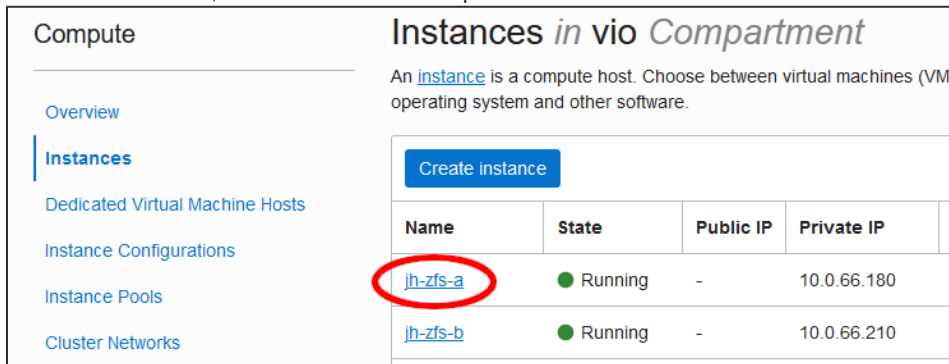
It will take approximately five minutes for the stack to build the ZFS-HA cluster and may take another 5 minutes to complete the booting process the initial clustering configuration.

## 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:

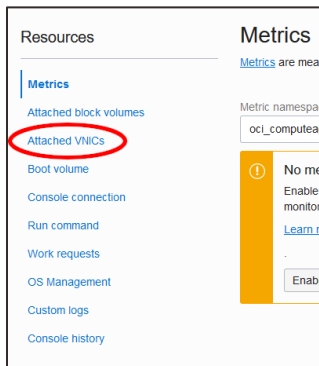
- List the instances in the compartment and select the instance with the name given in the variable configuration with “-a” at the end, as shown in this example:



The screenshot shows the OCI Compute console. On the left, the 'Instances' tab is selected. On the right, a table lists instances in the compartment. The instance 'jh-zfs-a' is circled in red.

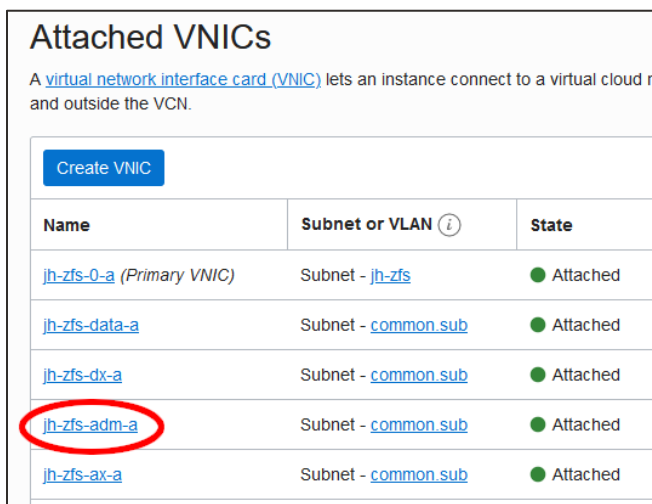
Name	State	Public IP	Private IP
<a href="#">jh-zfs-a</a>	Running	-	10.0.66.180
<a href="#">jh-zfs-b</a>	Running	-	10.0.66.210

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



The screenshot shows the OCI Instance details page. On the left, the 'Resources' menu is expanded, and 'Attached VNICS' is highlighted with a red circle.

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



The screenshot shows the OCI Attached VNICS page. A table lists the attached VNICS. The VNIC 'jh-zfs-adm-a' is circled in red.

Name	Subnet or VLAN ⓘ	State
<a href="#">jh-zfs-0-a (Primary VNIC)</a>	Subnet - <a href="#">jh-zfs</a>	Attached
<a href="#">jh-zfs-data-a</a>	Subnet - <a href="#">common.sub</a>	Attached
<a href="#">jh-zfs-dx-a</a>	Subnet - <a href="#">common.sub</a>	Attached
<a href="#">jh-zfs-adm-a</a>	Subnet - <a href="#">common.sub</a>	Attached
<a href="#">jh-zfs-ax-a</a>	Subnet - <a href="#">common.sub</a>	Attached

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



- 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](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
```

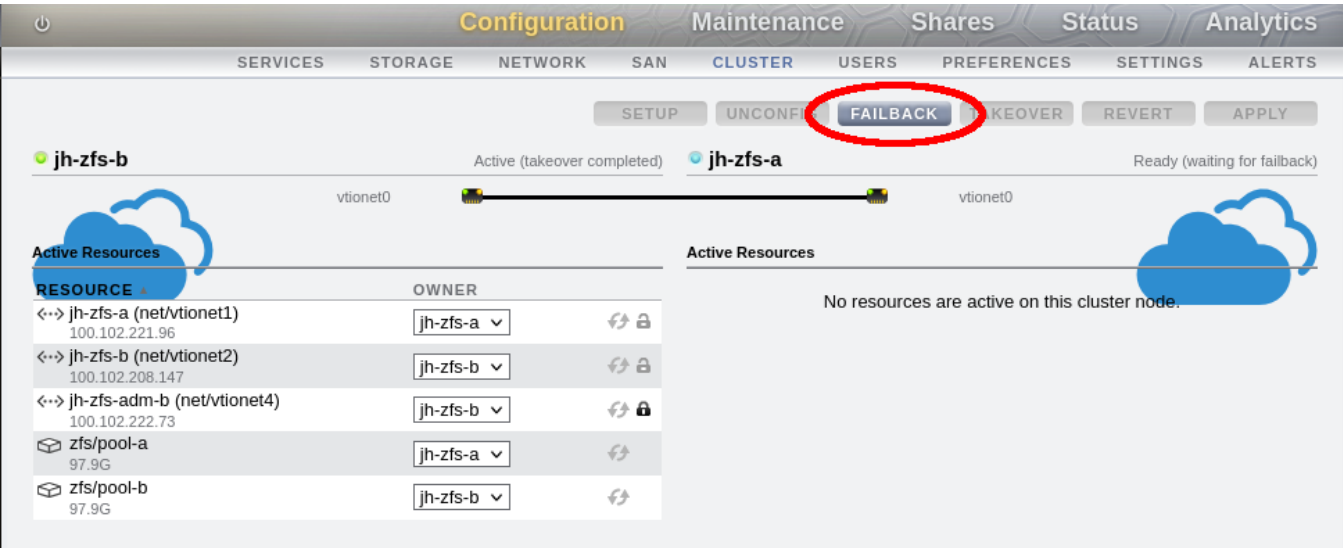
# CONNECT TO THE BROWSER USER INTERFACE (BUI)

Log in to the BUI by connecting your browser to [https://<primary\\_admin\\_address>:215](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.

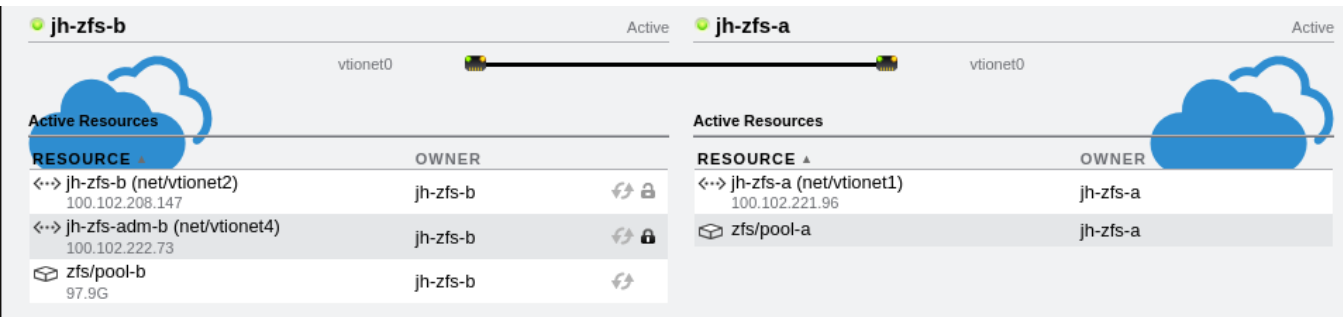
The BUI may now be used to create shares or perform other ZFS Appliance administration tasks.

## 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.



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



## ZFS-HA Networking Recommendations

The following are recommendations that refer to settings within the ZFS-HA images. The Deployment Tool may have configured some or all of these, but it is recommended that they be verified after deployment.

### ZFS-HA Network -> Configuration -> Routing

- It is recommended to set the multihoming model to strict.

### ZFS-HA Network -> Configuration -> Datalinks


- Link Speed, Link Duplex and Flow Control should all be set to Auto.
- The 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 have their MTU set to 9000 (jumbo frames) by default for best performance. In some cases, especially when data will traverse Wide Area Network links such as when replicating to or from an on-premises ZFS Storage Appliance, this may cause connections to hang if any step of the path is not set to use jumbo frames. In these cases, an MTU of 1500 is recommended.

### ZFS-HA Network -> Configuration->Interfaces

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

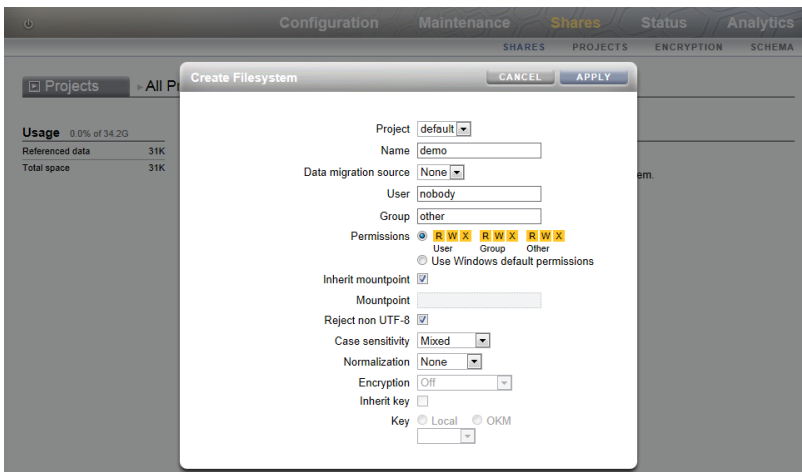
## SHARE AN SMB FILESYSTEM


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](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.



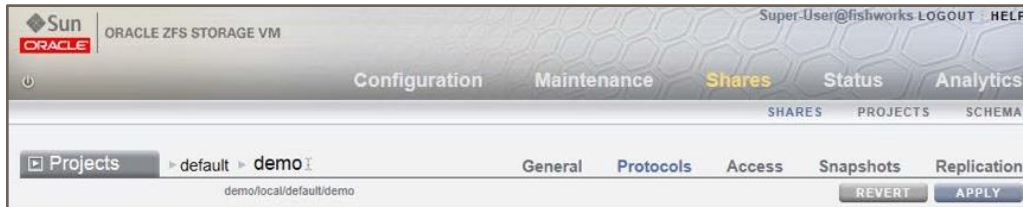
2. Name the filesystem and change the permissions for Group and Other to allow anyone to read, write, and execute on the filesystem. In this example, the filesystem is named demo. The filesystem is part of the default project. Click APPLY to save the changes.



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

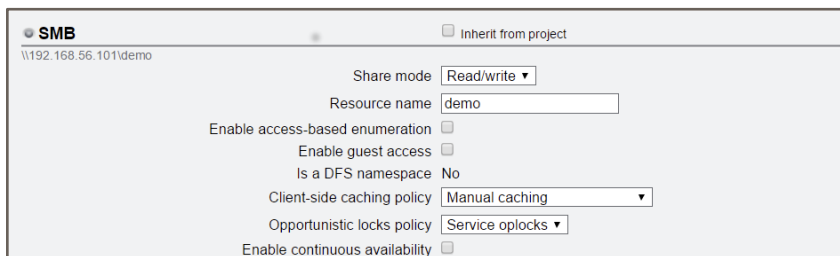


- Click Protocols.



- 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.

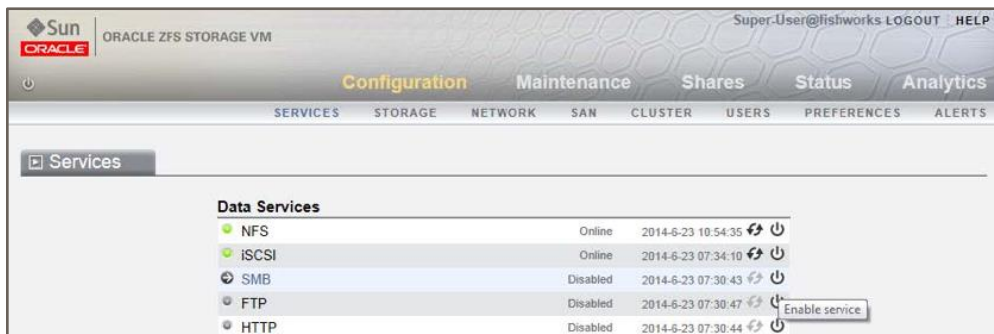
In this example, the Resource Name is demo. Click APPLY to save the changes.



**Note:** In the SMB section of the Protocol screen, the status of the SMB service is shown with either a green, amber, or grey circle. Beneath that is a possible path to mount the share at, but the IP address shown will be the address of the interface the browser connected to. This is almost certainly incorrect, since the best practice would be to connect the browser through the administrative interface, not the client data interface. You should verify that the correct address is used when accessing the share from a data client.




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

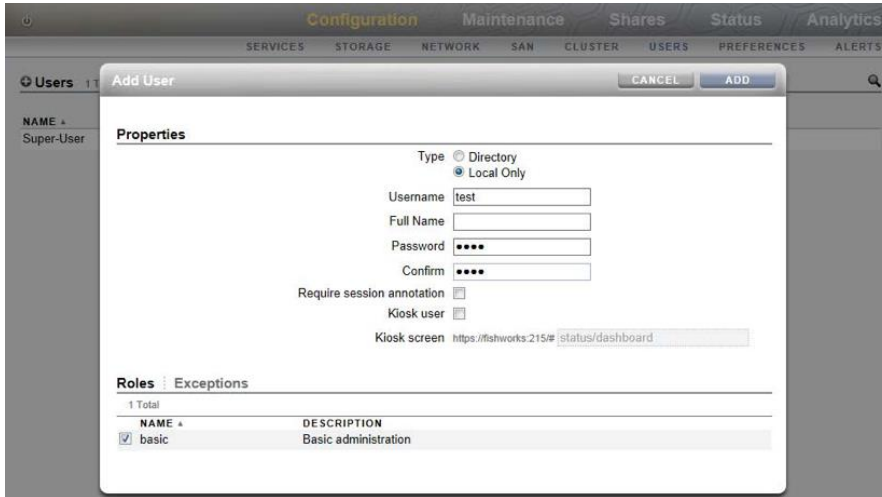


The state will change from Disabled to Online.



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 BUI. The dialog has a 'Properties' tab and a 'Roles' section. In the 'Properties' tab, the 'Type' is set to 'Local Only'. The 'Username' field contains 'test'. The 'Full Name' field is empty. The 'Password' and 'Confirm' fields are masked with dots. There are checkboxes for 'Require session annotation' and 'Kiosk user'. The 'Kiosk screen' field contains the URL 'https://fishworks.215#status/dashboard'. In the 'Roles' section, there is a table with one role: 'basic' with the description 'Basic administration'.

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.

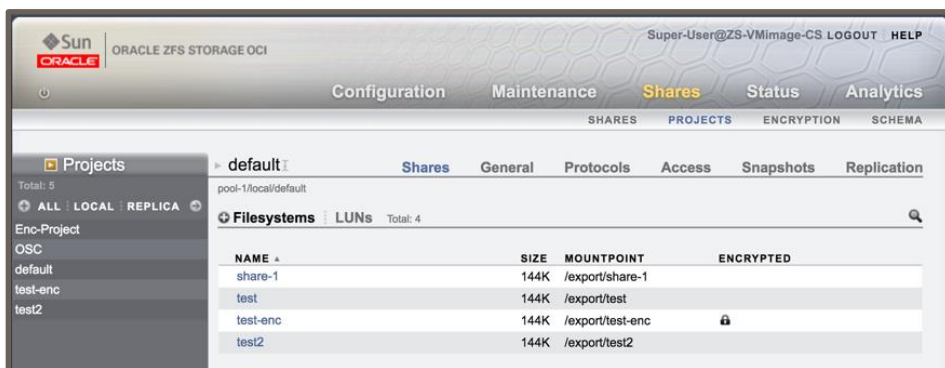
**Note:** When looking at the Protocol screen of a share in the BUI, the mount point given will use the IP address (or FQDN if it resolves in DNS) of the administrative interface. This is not correct – the NAS client address or FQDN should be substituted.

## SHARE AN NFS FILESYSTEM

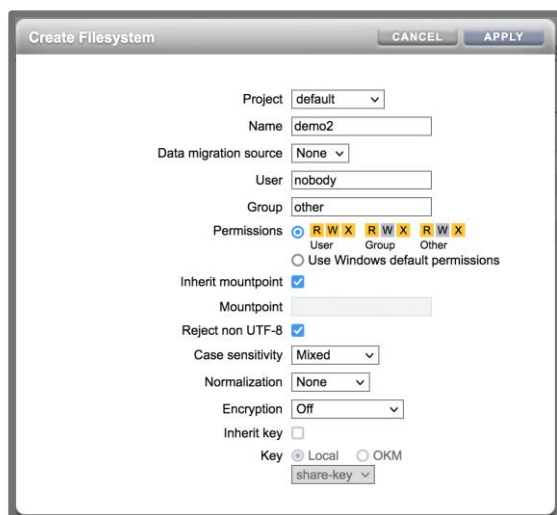
Complete the following steps to set up a simple filesystem share over NFS to share with an NFS client or clients. Begin by logging in to the BUI by connecting your browser to [https://<primary\\_admin\\_address>:215](https://<primary_admin_address>:215) and using “opc” as the username with the password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


1. Navigate to the Shares screen.

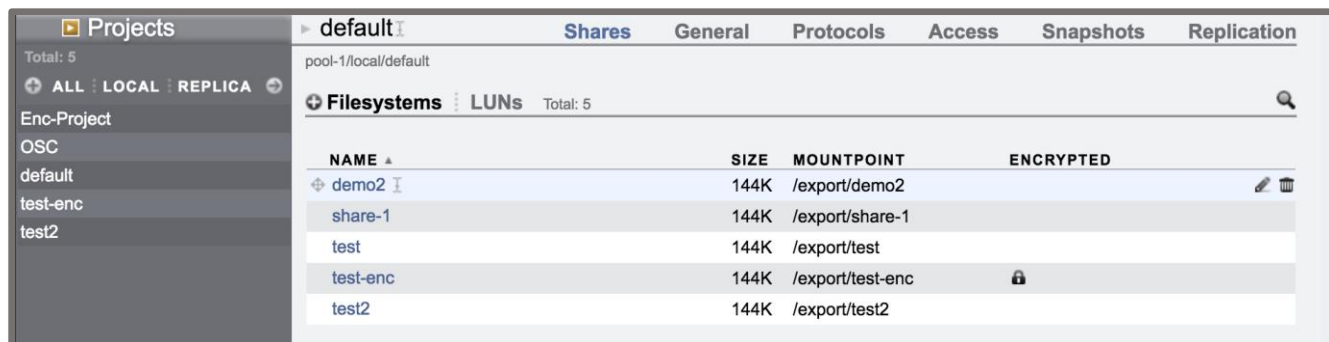
Click the add item icon  next to Filesystems to create a new filesystem. Projects provide an administrative point for filesystems so you can set properties at the project level that are inherited by filesystems within the project. The system includes the default project.



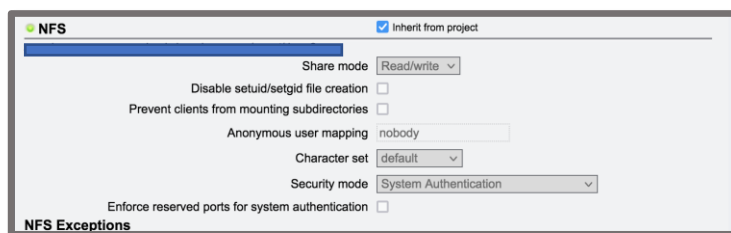
2. Name the filesystem and change the permissions to match the user/group requirements. In this example, the filesystem is named demo2. The filesystem is part of the default project. Click APPLY to save the changes.



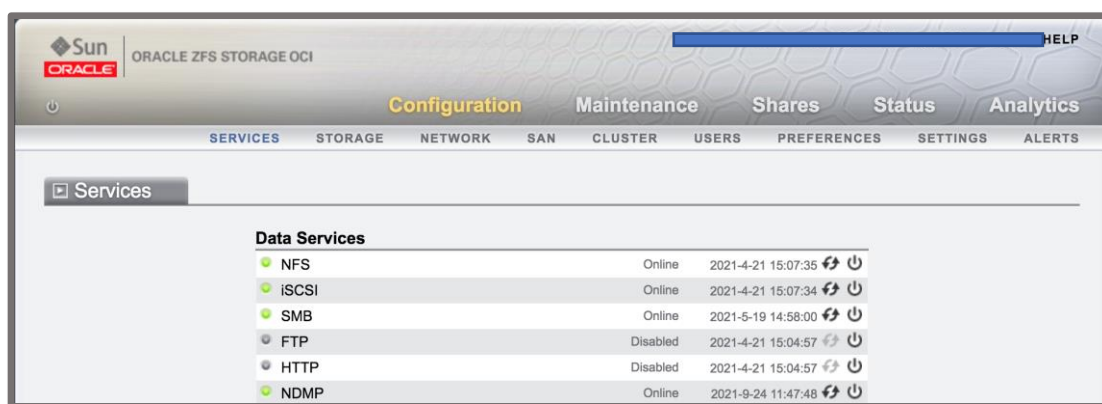
3. In the Shares screen, mouse over the entry for the new filesystem and click the edit icon  to edit the filesystem attributes.



- 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.



- 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.

**Note:** When looking at the Protocol screen of a share in the BUI, the mount point given will use the IP address (or FQDN if it resolves in DNS) of the administrative interface. This is not correct – the NAS client address or FQDN should be substituted.

## DEEP DIVE - 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.

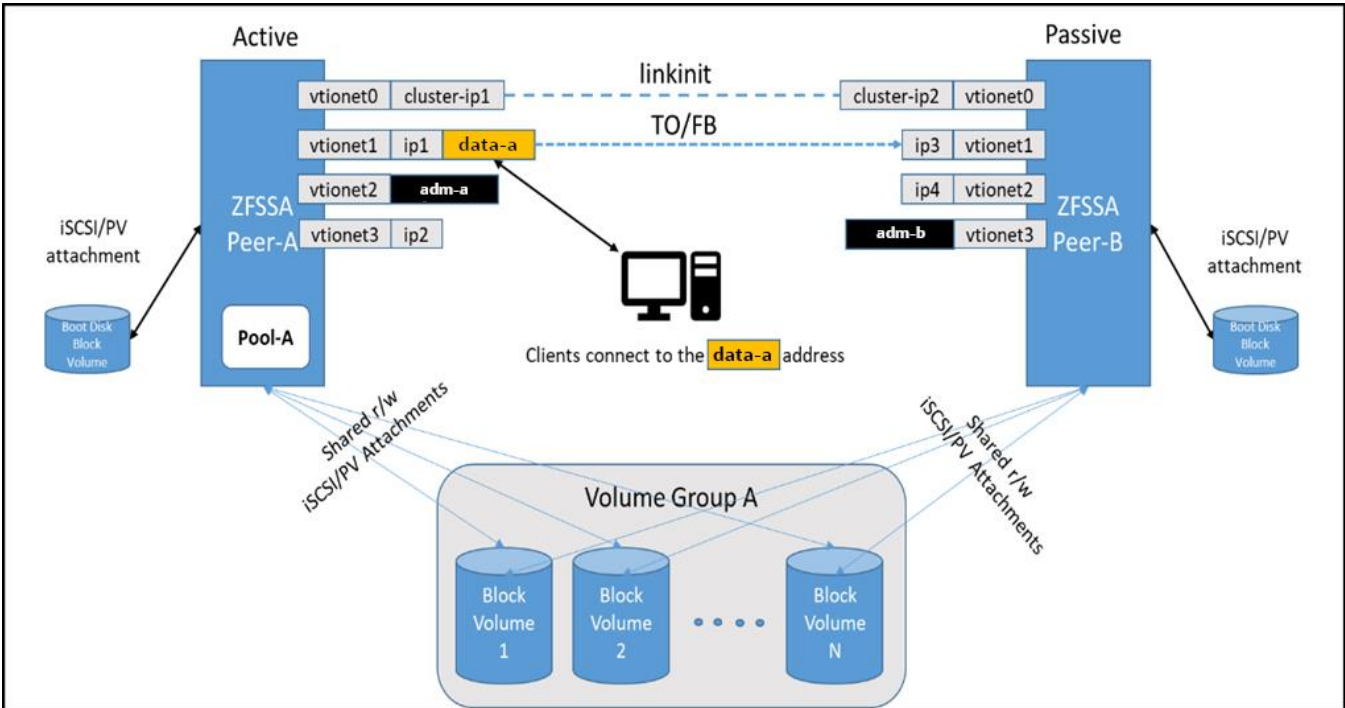
**NOTE:** The Primary VNIC (Virtual Network Interface Card) is the first network interface in a Compute instance. Additional VNICs may be added and are referred to as Secondary VNICs. Each VNIC is assigned an IP address at its creation and is referred to as the primary IP address. Additional IP address 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.

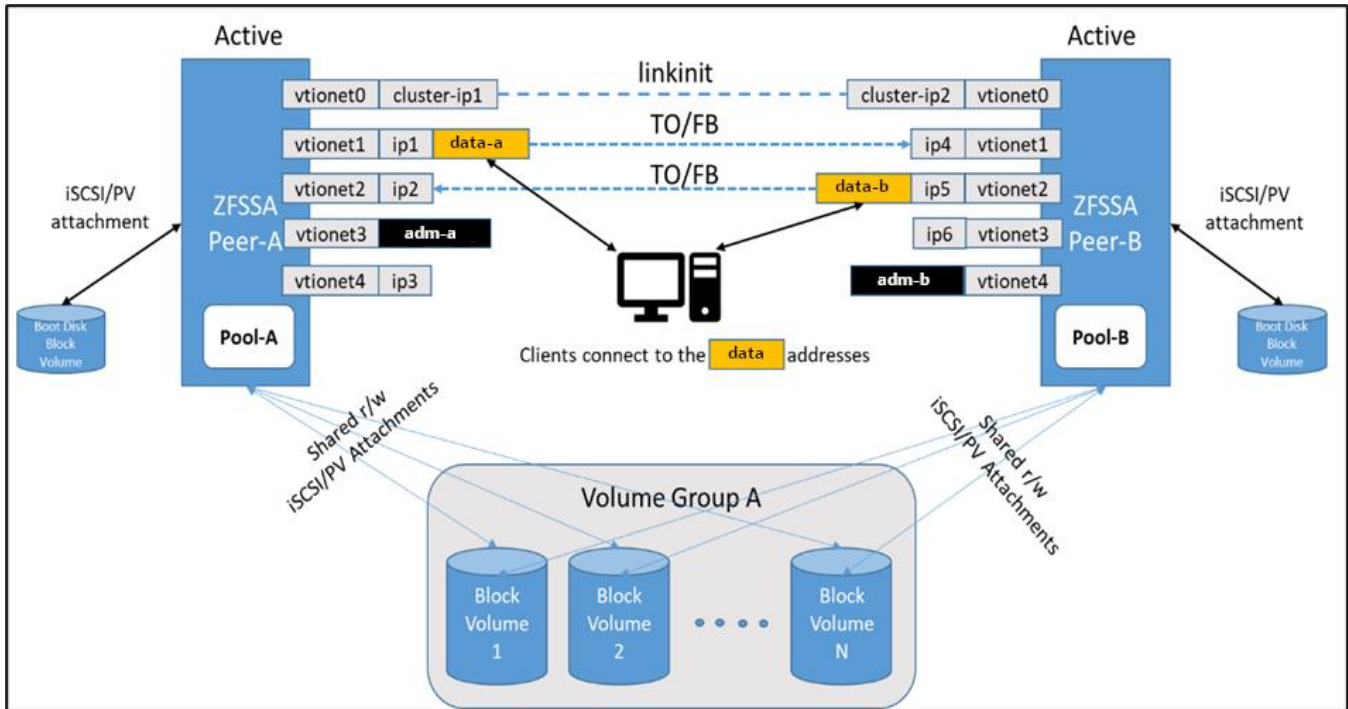
VNIC	ZFS INSTANCE A (ACTIVE)	ZFS INSTANCE B (PASSIVE)	USAGE
<b>vtionet0 (Primary VNIC)</b>	cluster-ip1	cluster-ip2	Used for cluster I/O only
<b>vtionet1 - primary IP</b>	ip1 - Placeholder	ip3 - Placeholder	Primary IP addresses are locked to the instance and cannot move.
<b>vtionet1 - secondary IP</b>	data-a address	unassigned	Floating IP address used for client access to Pool-A. Always assigned to the active controller
<b>vtionet2 – primary IP</b>	adm-a address	ip4 - Placeholder	Private Administrative access to Node A (B unused)
<b>vtionet3 – primary IP</b>	ip2 - Placeholder	adm-b address	Private Administrative access to Node B (A unused)

In this example, a secondary IP address, data-a, is assigned to vtionet1 on Node A, the active controller. In the event that Node B is made active, the data-a address will automatically move to vtionet1 on Node B. Clients attached to the data-a 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 data 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
<b>vtionet0 (Primary VNIC)</b>	cluster-ip1	cluster-ip2	Used for cluster I/O only
<b>vtionet1 - primary IP</b>	ip1 - Placeholder	ip4 - Placeholder	Primary IP addresses are locked to the instance and cannot move.
<b>vtionet1 - secondary IP</b>	data-a address	unassigned	Floating IP address used for client access for Pool-A.
<b>vtionet2 - primary IP</b>	ip2 - Placeholder	ip5 - Placeholder	Primary IP addresses are locked to the instance and cannot move.
<b>vtionet2 - secondary IP</b>	unassigned	data-b address	Floating IP address used for client access for Pool-B.
<b>vtionet3 - primary IP</b>	mgmt-ip1	ip6 - Placeholder	Private Administrative access to Node A (B unused)
<b>vtionet4 - primary IP</b>	ip3 - Placeholder	mgmt-ip2	Private Administrative access to Node B (A unused)



In this example, secondary IP addresses, data-a and data-b, are assigned to vtionet1 on Node A and vtionet2 on Node B. If either node becomes inactive, the data-a and data-b 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 data-a 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
<b>Resource Type</b>	
<b>Singleton</b>	Known by both instances but only active on one instance. (Storage Pools and NAS IP)
<b>Private</b>	Only available and active on one instance. (Administration Network Interface)
<b>Replicate</b>	Resource known by both heads. (Service configuration)
<b>Symbiote</b>	Follows other resources (Replications actions follow storage pool)
<b>Clustered State</b>	
<b>Unconfigured</b>	Clustering is not configured.
<b>Owner</b>	Clustering is configured. This active instance owns the storage and data resources.
<b>Stripped</b>	Clustering is configured. This passive instance does not control any shared resources.
<b>Clustered</b>	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.

# OS8.8.X DOCUMENTATION SPECIFIC TO ZFS-HA

This section contains information that applies only to ZFS-HA instances. This information cannot be found in the online documentation for the Oracle ZFS Storage Appliance.

## Automatically Expanding a ZFS-HA Storage Pool

A pair of ZFS-HA clustered instances can support up to 32 block volumes of up to 32TB each, for a total of 1PB of usable storage space across both controller instances. Most installations are configured with much smaller storage footprints, and over time may need to expand the pools to provide more storage. This section documents the process of automatically expanding an existing pool.

**NOTE:** The Automated Capacity Scaling feature described below was released in OS8.8.57. Instances running an older version of the software are encouraged to upgrade by following the steps in the section “[Upgrading Your ZFS-HA Instance](#)”. New instances should be created using an image based on OS8.8.57 or above.

Automatic Capacity Scaling, when enabled, will monitor the pool’s used capacity. If it reaches a 80%, the pool will be expanded by a combination of expanding existing block volumes and adding new block volumes to the pool until its used capacity drops to 60%. Optionally, a maximum limit can be set to prevent the pool from being expanded beyond this limit. Because ZFS storage pools cannot be reduced in size, this limit can help prevent unexpected jumps in OCI Block Volume storage charges.

Automatic Capacity Scaling may be disabled at any time. The limit may also be changed but cannot be set below the current pool size.

## Set Automatic Expansion of a pool using the BUI

For each storage pool listed in the Configuration->Storage screen of the ZFS-HA BUI, Automated Capacity Scaling may be enabled by checking the appropriate box.

ORACLE ZFS STORAGE OCI

Oracle Public Cloud User@jh-zfs-a LOGOUT HELP

ConfigurationMaintenanceSharesStatusAnalytics

SERVICESSTORAGENETWORKSANCLUSTERUSERSPREFERENCESSETTINGSalerts

About Storage Configuration

Storage is configured in pools that are characterized by their underlying data redundancy, and provide space that is shared across all filesystems and LUNs.

During the configuration process, you will select which devices to allocate to a storage pool and the redundancy profile most appropriate to your workload, balancing performance, availability, and capacity.

Importing storage will search all devices attached to the system for existing pool configurations, from which you can select one as the system pool. This option is used to migrate pools between systems, and in some cases can recover pools that were destroyed inadvertently.

Available Pools

IMPORT

HOST : POOL	DATA PROFILE	LOG PROFILE	STATUS	ERRORS	ENCRYPTED
jh-zfs-a:pool-a	Striped	-	Online	0	off

jh-zfs-a:pool-a

ADDREMOVEUNCONFIGREVERTAPPLY

Pool Namepool-a

Data ProfileStriped

Log Profile-

Pool StatusOnline

Data ErrorsNo known persistent errors

Scrub Schedule30 days

Scrub StatusScrub completed: 0 errors

2023-8-19 08:23:28 (0h0m)

SCRUB

Automated Capacity Scaling

Limit10T

Allocation

Data2.93T

Data + Reserve3 disks

Spare0 disks

Log0 disks

Cache0 disks

Meta0 disks

Device Status

0 errors

No device faults have been detected in the storage pool.



## Set Automatic Expansion of a pool using the CLI

From the main prompt after logging in with SSH, enter “configuration storage”. At this point you may give the command “list” to show the pools and their settings.

```
jh-zfs-a:> configuration storage
jh-zfs-a:configuration storage> list
Properties:
            pool = pool-a
            status = online
            errors = 0
            owner = jh-zfs-a
            profile = stripe
            log_profile = -
            cache_profile = -
            meta_profile = -
            autoscale = false
            autoscale_limit = 0
            scrub = scrub completed after 0h0m with 0 errors at 2023-8-19 12:23:28
            scrub_schedule = 30 days
            async_destroy_reclaim_space = 0
            encryption = off
```

Use the “select” command to choose a pool to work with if there are multiple pools.

In the example below, the autoscale property is set to true, and a limit that the pool may scale to is set to 10T.

```
jh-zfs-a:configuration storage> set autoscale=true
            autoscale = true (uncommitted)
jh-zfs-a:configuration storage> set autoscale_limit=10T
            autoscale_limit = 10T (uncommitted)
jh-zfs-a:configuration storage> commit
```

## Set Automatic Expansion of a pool using the RESTful API

The autoscale and autoscale\_limit properties shown in the above section can also be listed and set with calls to the RESTful API as shown in the following examples:

A pool’s properties can be listed with a GET call to “api/storage/v1/pools/<poolname>” as shown in this example:

```
curl -XGET -k -u root:password https://<ip_address>:215/api/storage/v1/pools/pool-a | json_pp
{
  "pool": {
    "asn": "3b545811-68cc-46de-8000-a80e869f920a",
    "async_destroy_reclaim_space": 0,
    "autoscale": "True",
    "autoscale_limit": "1099511627776.0",
    [ snip ]
  }
}
```

The pool’s properties can be set with a PUT call. An example of turning autolimit on and setting autoscale\_limit:

```
curl -XPUT -H "Content-Type: application/json" -k -u root:password \
  https://<ip_address>:215/api/storage/v1/pools/pool-a/edit \
  -d '{"autoscale": true, "autoscale_limit": "10T"}'
```

Full documentation on working with the RESTful API can be found in the [Oracle ZFS Storage Appliance RESTful API Guide, Release OS8.8.x](#)

## Manually Expanding a ZFS-HA Storage Pool

Pools may also be expanded manually by creating new OCI block volumes and adding them to the pool, or by expanding existing ones.

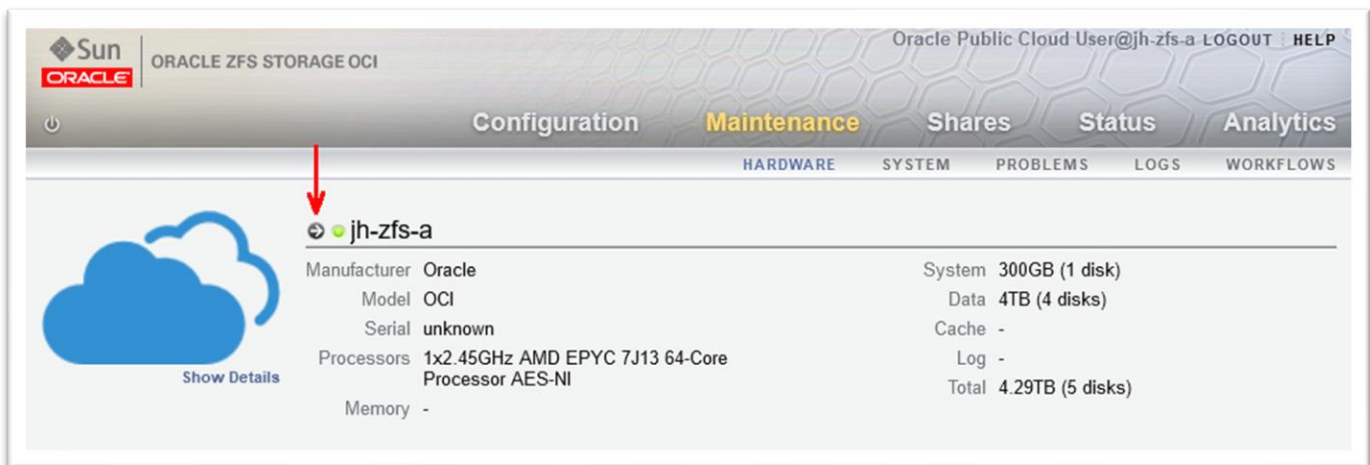
Both can be done on ZFS-HA instances without the need to configure anything through an OCI console.

### Expanding Existing Block Volumes

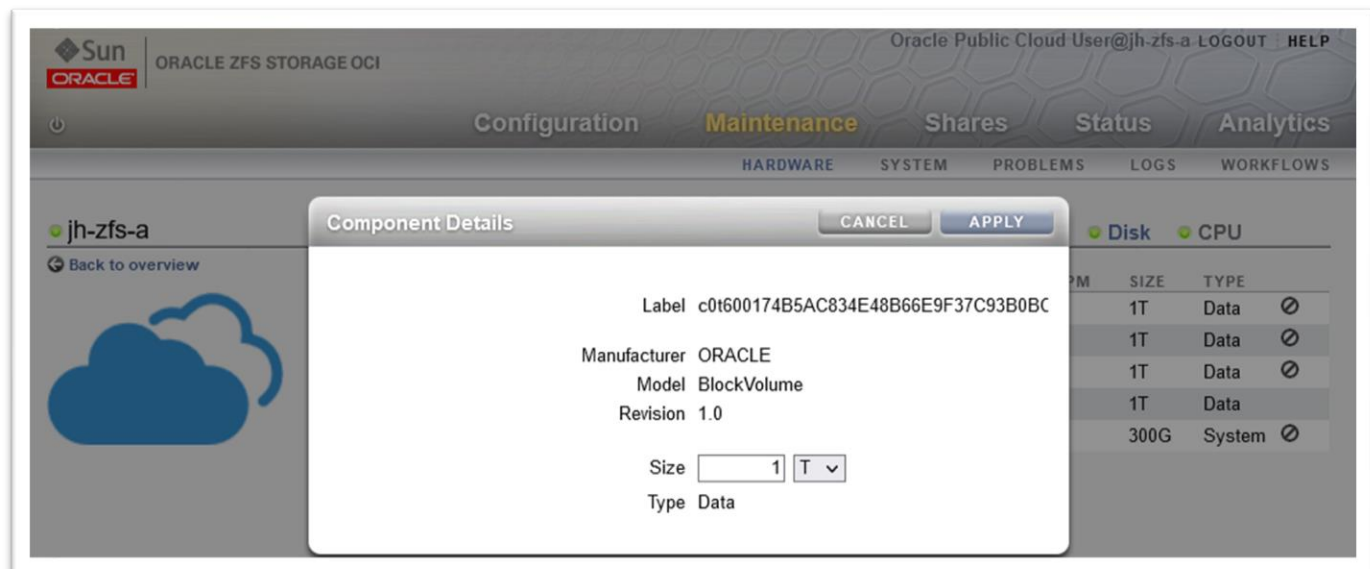
When expanding an existing block volume, the pool will be grown automatically.

### Expanding a Block Volume using the BUI

Navigate to the Maintenance → Hardware screen and click on the arrow to the left of the instance name to bring up the detail view.



Click on the green circle of a listed disk to bring up the Component Details window.



Here you may increase the size of a disk up to 32T. When the Apply button is clicked, the volume size is increased and integrated into the pool.

Note that while the boot volume, listed as "System" in the detail view, may be enlarged, there is generally no reason to do so.

## Expanding a Block Volume using the CLI

From the main prompt after logging in with SSH, enter “configuration hardware” and then “select chassis-000” to affect the controller, then “select disk”. Then use “list” to list the block volumes connected to the system. Use the select command to choose the volume to resize. Use the “set” command to set the size, which can range from 50G to 32T, then commit the change.

```
jh-zfs-a:maintenance chassis-000 disk> select disk-004
jh-zfs-a:maintenance chassis-000 disk-004> ls
Properties:
    label = c0t60F51204A1A34E099587F38929751240d0
    present = true
    faulted = false
    manufacturer = ORACLE
    model = BlockVolume
    revision = 1.0
    size = 1T
    type = data
    use = data
    device = c0t60F51204A1A34E099587F38929751240d0
    offline = false

jh-zfs-a:maintenance chassis-000 disk-004> set size=2T
    size = 2T (uncommitted)
jh-zfs-a:maintenance chassis-000 disk-004> commit
```

## Expanding a Block Volume using the RESTful API

The attached volumes can be listed with a GET call to `https://<ip_address>:215/api/hardware/v1/chassis/chassis-000/disk`. Example:

```
curl -XGET -H "Content-Type: application/json" -k -u root:password \
https://<ip_address>:215/api/hardware/v1/chassis/chassis-000/disk | json_pp
{
  "disk": [
    {
      "label": "c0t600174B5AC834E48B66E9F37C93B0BC8d0",
      "present": true,
      "faulted": false,
      "manufacturer": "ORACLE",
      "model": "BlockVolume",
      "revision": "1.0",
      "size": 1099511627776,
      "type": "data",
      "use": "data",
      "device": "c0t600174B5AC834E48B66E9F37C93B0BC8d0",
      "offline": false,
      "href": "/api/hardware/v1/chassis/chassis-000/disk/disk-000"
    },
    [ snip ]
  ]
}
```

Expand the volume with a PUT call and specify the new size:

```
curl -XPUT -H "Content-Type: application/json" -k -u root:password \
https://<ip_address>:215/api/hardware/v1/chassis/chassis-000/disk/disk-002 -d '{"size": "2T"}' | json_pp
{
  "disk":
  {
    "href": "/api/hardware/v1/chassis/chassis-000/disk/disk-002",
    "label": "c0t6054CEFD72DB49FAAF8A7D41E9AB1D23d0",
    "present": true,
    "faulted": false,
    "manufacturer": "ORACLE",
    "model": "BlockVolume",
    "revision": "1.0",
    "size": 2199023255552,
    "type": "data",
    "device": "c0t6054CEFD72DB49FAAF8A7D41E9AB1D23d0"
  }
}
```

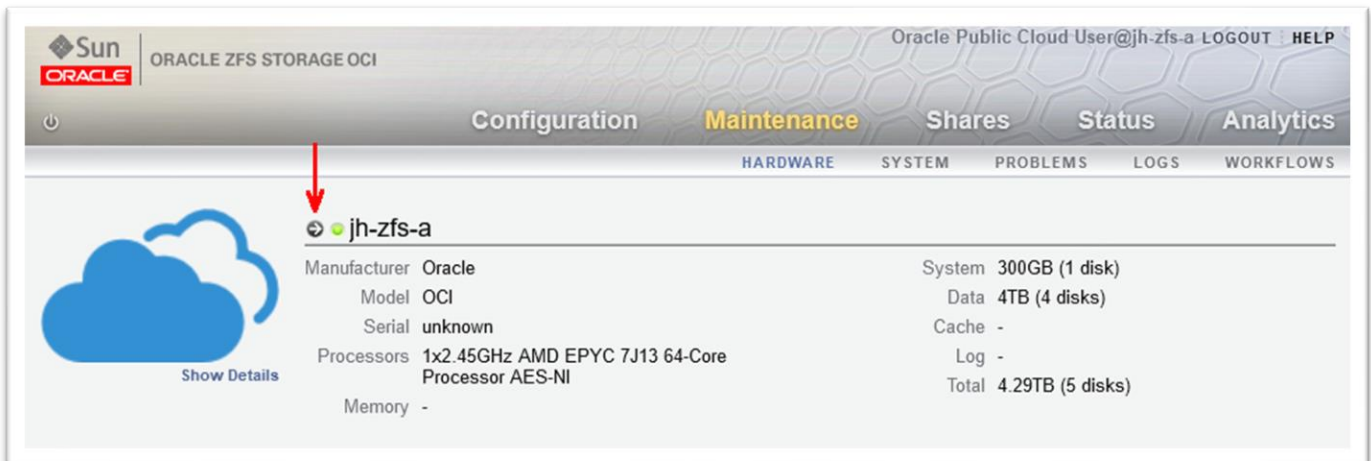
Full documentation on working with the RESTful API can be found in the [Oracle ZFS Storage Appliance RESTful API Guide, Release OS8.8.x](#)

## Creating New Block Volumes

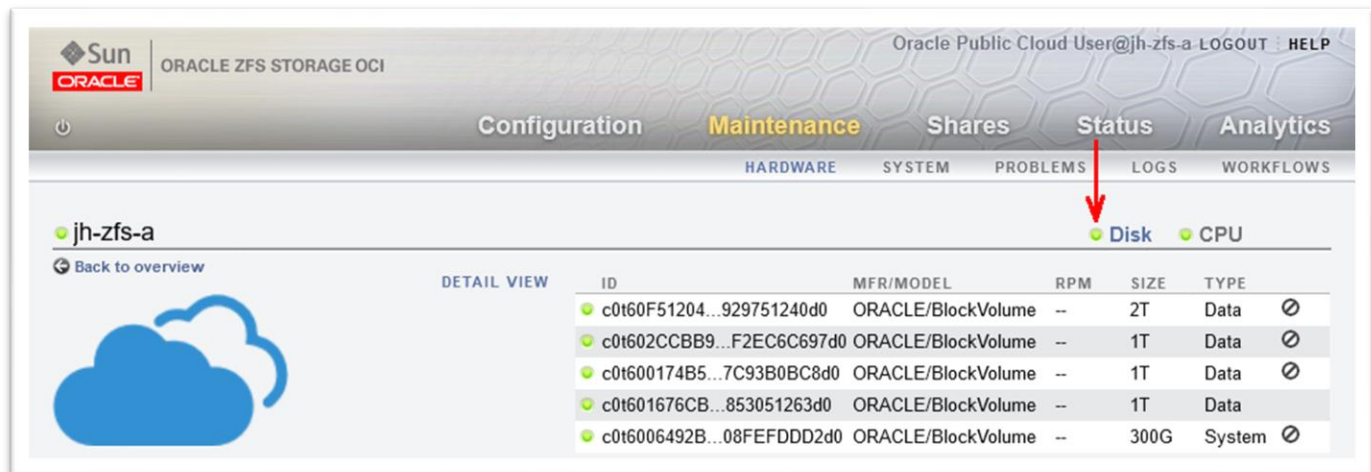
After creating a new block volume, it must be added to the pool manually. This step is explained in the next section.

### Creating a Block Volume using the BUI

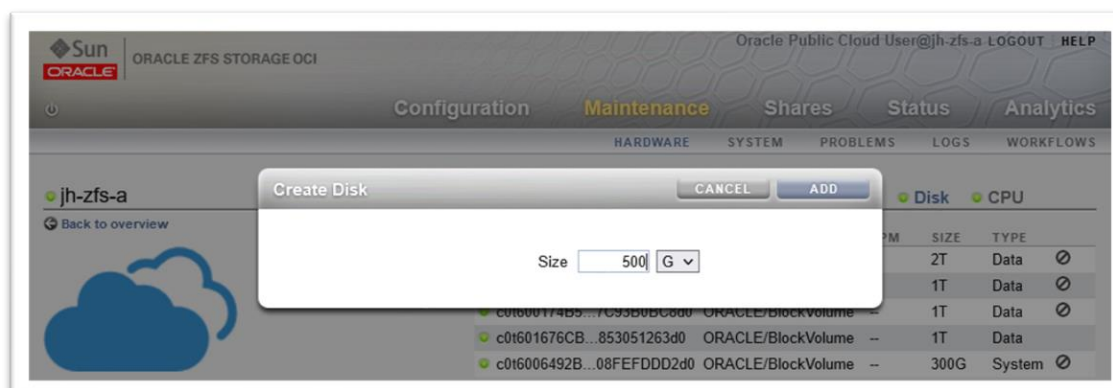
Navigate to the Maintenance → Hardware screen and click on the arrow to the left of the instance name to bring up the detail view.



In the detail view, click on the green circle for Disk to bring up the Add Disk window.



Enter the size of the new volume, then click the Add button to create the disk.



## Creating a Block Volume using the CLI

From the main prompt after logging in with SSH, enter “configuration hardware” and then “select chassis-000” to affect the controller, then “select disk”. Use the “create” command to add the new volume, then set the size of the new volume to between 50G and 32T, and commit to add the volume.

```
jh-zfs-a:maintenance chassis-000 disk> create
jh-zfs-a:maintenance chassis-000 disk create (uncommitted)> set size=50G
size = 50G
jh-zfs-a:maintenance chassis-000 disk create (uncommitted)> commit
```

## Creating a Block Volume using the RESTful API

Create a new block volume with a POST call and specify the size:

```
curl -XPOST -H "Content-Type: application/json" -k -u root:password \
https://<ip_address>:215/api/hardware/v1/chassis/chassis-000/disk -d '{"size": "50G"}' | json_pp
{
  "disk": [
    {
      "label": "c0t60D08AE0B8D44AD69A00451701055761d0",
      "present": true,
      "faulted": false,
      "manufacturer": "ORACLE",
      "model": "BlockVolume",
      "revision": "1.0",
      "size": 295279001600,
      "type": "data",
      "use": "system",
      "device": "c0t60D08AE0B8D44AD69A00451701055761d0",
      "offline": false,
      "href": "/api/hardware/v1/chassis/chassis-000/disk/disk-000"
    },
    {
      "label": "c0t60036C55D381426FAB5487581748332Bd0",
      "present": true,
      "faulted": false,
      "manufacturer": "ORACLE",
      "model": "BlockVolume",
      "revision": "1.0",
      "size": 53687091200,
      "type": "data",
      "device": "c0t60036C55D381426FAB5487581748332Bd0",
      "href": "/api/hardware/v1/chassis/chassis-000/disk/disk-001"
    }
  ]
}
```

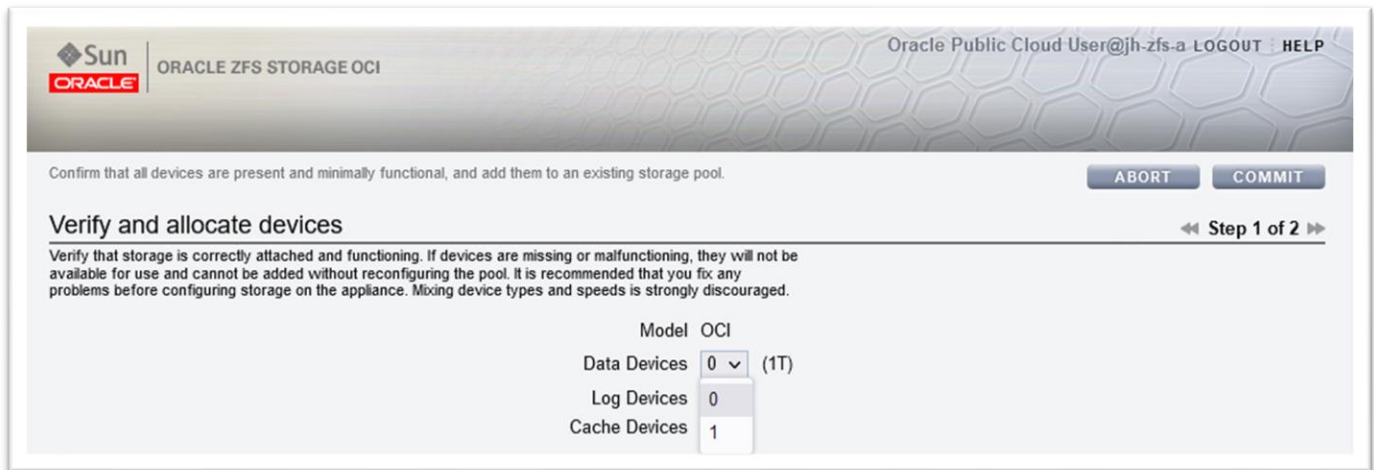
Full documentation on working with the RESTful API can be found in the [Oracle ZFS Storage Appliance RESTful API Guide, Release OS8.8.x](#)

## Adding New Block Volumes to a Pool

After creating a new block volume, it must be added to the pool. This section shows how to perform this action. The default storage profile is striped, since Block Volumes are already redundant at the OCI level. Note that if you have a different storage profile, two or more volumes must be added at the same time depending on the profile.

### Adding New Block Volumes to a pool using the BUI

In the BUI, navigate to the Configuration → Storage screen and click the Add button. Use the pulldown menu for Data Devices and choose the number of volumes to add to the pool, then click Apply.



Oracle Public Cloud User@jh-zfs-a LOGOUT HELP

ORACLE ZFS STORAGE OCI

Confirm that all devices are present and minimally functional, and add them to an existing storage pool. [ABORT] [COMMIT]

### Verify and allocate devices

Verify that storage is correctly attached and functioning. If devices are missing or malfunctioning, they will not be available for use and cannot be added without reconfiguring the pool. It is recommended that you fix any problems before configuring storage on the appliance. Mixing device types and speeds is strongly discouraged.

	Model	OCI
Data Devices	0	(1T)
Log Devices	0	
Cache Devices	1	

### Adding New Block Volumes to a pool using the CLI

From the main prompt after logging in with SSH, enter “configuration storage”. If you have a single pool, which is the default ZFS-HA configuration, the pool name is not displayed, but it is selected. If you have multiple pools, a default pool is selected and displayed. If this is not the pool to which you want to add the device, enter `set pool=` and specify another online pool.

```
jh-zfs-a:> configuration storage
jh-zfs-a:configuration storage> show
Properties:
    pool = pool-a
    status = online
    errors = 0
    owner = jh-zfs-a
    profile = stripe
    log_profile = -
    cache_profile = -
    meta_profile = -
    autoscale = false
    autoscale_limit = 0
    scrub = scrub completed after 0h0m with 0 errors at 2023-8-19 12:23:28
    scrub_schedule = 30 days
    async_destroy_reclaim_space = 0
    encryption = off
```

Enter “add”. You can use the command “show” to view the device count available. In this example, we see that zero new data volumes out of a possible one have been allocated to the pool:

```
jh-zfs-a:configuration storage verify> show

System  OCI

State   ok
Data Disks  0 of 1 (1T)
Log Disks   0 of 0
Cache Disks 0 of 0
```



Use the “set” command to specify the number and type of drive to add:

```
jh-zfs-a:configuration storage verify> set data=1
data = 1
```

Enter “done” to complete the process. You may also enter the “abort” command to interrupt the process and stop the new volumes from being added.

## Adding New Block Volumes to a pool using the RESTful API

Add the previously created block volumes to a pool with a PUT call and specify the number of volumes to add as data volumes. The new pool status will be returned:

```
curl -XPUT -H "Content-Type: application/json" -k -u
root:zfsisfun! https://<ip_address>:215/api/storage/v1/pools/pool-a/add -d '{"data": 1}' | json_pp

{
  "pool" : {
    "name" : "pool-a",
    "vdev" : [
      {
        "label" : "c0t60F51204A1A34E099587F38929751240d0",
        "type" : "disk",
        "chassis" : "jh-zfs-a",
        "state" : "healthy"
      },
      {
        "type" : "disk",
        "label" : "c0t600174B5AC834E48B66E9F37C93B0BC8d0",
        "chassis" : "jh-zfs-a",
        "state" : "healthy"
      },
      {
        "type" : "disk",
        "label" : "c0t602CCBB9D82843C8B7B5A25F2EC6C697d0",
        "chassis" : "jh-zfs-a",
        "state" : "healthy"
      },
      {
        "state" : "healthy",
        "chassis" : "jh-zfs-a",
        "type" : "disk",
        "label" : "c0t601676CBF59C434B9A44BBB853051263d0"
      }
    ],
    [ pool properties snipped ]
  }
}
```

Full documentation on working with the RESTful API can be found in the [Oracle ZFS Storage Appliance RESTful API Guide, Release OS8.8.x](#)

## Adding Clustered Interfaces

When using ZFS Storage in OCI, it may be advantageous to spread workloads across multiple VNICs. One possible use case detailed here is to send or receive ZFS replication traffic across a new set of interfaces. This requires configuration both in the OCI console and on the ZFS-HA instances. Configuration of the ZFS-HA instances will be done using the ZFS Browser User Interface (BUI) which is accessed via the IP address of the admin interface on port 215.

This section covers the configuration of interfaces on an active/active cluster. In the case of an active/passive cluster, differences will be noted.

### Overview - Active/Active Cluster

1. Attach two new VNICs, one as main, the other as standby, on both instances in the OCI console
2. Add secondary IP addresses in OCI console to the main VNICs on each instance
3. Reboot both nodes one at a time
4. Run Takeover/Failback on node A
5. Edit new interfaces on each node
  - Change name
  - Change IP address to secondary IP address from step 2
6. Reboot and rebalance the clustered resources

### Overview - Active/Passive Cluster

1. Attach a new VNIC on both instances in OCI console
2. Add a secondary IP address to the new VNIC on the active node (usually A) in OCI console
3. Reboot passive then active nodes
4. Run Takeover/Failback on node A
5. Edit new interface on active node
  - Change name
  - Change IP address to secondary IP address from step 2
6. Reboot and rebalance the clustered resources

## 1 - Attach New VNICs In the OCI Console

We start by adding VNICs to the ZFS-HA instances in the OCI cluster. When adding VNICs, use the naming convention `<clustername>-<type>-[a|b]`. The example cluster name here is `jh-zfs` and the type is `repl` (for replication). Your configuration will vary.

Find the A and B instances in the OCI console and open a browser tab for each. Click on Attached VNICs in the Resources sidebar to access the correct screen. Click Create VNIC to add a new VNIC.

### Active/Active Cluster

Each instance will have two interfaces added.

Add two new VNICs to each controller. On the A controller, name the first VNIC `jh-zfs-repl-a` and the second `jh-zfs-rx-a`. (The rx VNICs are there as a holding point to take over in case of failure.)

On the B controller, name the first VNIC `jh-zfs-rx-b` and the second `jh-zfs-repl-b` – *note the name reversal*.

This is important to keep the naming consistency correct within the ZFS controllers.

### Active/Passive Cluster

Each instance will have one interface added.



Add a new VNIC to each controller. Name the VNIC on the A node jh-zfs-repl-a and jh-zfs-rx-a on the B node.

## Adding VNICs

Resources

Metrics

Attached block volumes

**Attached VNICs**

Boot volume

Console connection

Run command

Work requests

OS Management

Custom logs

Console history

### Attached VNICs

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

Create VNIC

Name	Subnet or VLAN <sup>(i)</sup>	State	FQDN <sup>(i)</sup>	VLAN tag
<a href="#">jh-zfs-0-b</a> (Primary VNIC)	Subnet - <a href="#">jh-zfs</a>	● Attached	jh-zfs-0-b... <a href="#">Show</a> <a href="#">Copy</a>	2295
<a href="#">jh-zfs-dx-b</a>	Subnet - <a href="#">common_sub</a>	● Attached	jh-zfs-dx-... <a href="#">Show</a> <a href="#">Copy</a>	1474
<a href="#">jh-zfs-data-b</a>	Subnet - <a href="#">common_sub</a>	● Attached	jh-zfs-dat... <a href="#">Show</a> <a href="#">Copy</a>	2417
<a href="#">jh-zfs-ax-b</a>	Subnet - <a href="#">common_sub</a>	● Attached	jh-zfs-ax-... <a href="#">Show</a> <a href="#">Copy</a>	1108
<a href="#">jh-zfs-adm-b</a>	Subnet - <a href="#">common_sub</a>	● Attached	jh-zfs-adm... <a href="#">Show</a> <a href="#">Copy</a>	2580

Show

When adding VNICs, choose the appropriate compartment/VCN/subnet. These will probably be the same as the data VNICs. You may also set up a separate VCN, but that's outside the scope of this documentation.

If using a separate VCN, it should be in the same compartment as the data VNICs since the new VNICs will have secondary IP addresses that can move between instances at failover/takeover events. This requires slightly elevated permissions for the compartment containing the VNICs. If a different compartment is used, it must use an Identity Policy rule to “use private-ips” rather than just reading them. It is not recommended to use the same compartment as the cluster or admin VCNs as they do not require the elevated access.

### VNIC information

Name *Optional*

Select a virtual cloud network in **Networks** [\(Change Compartment\)](#)

Network

Normal setup: subnet

The typical choice when adding a VNIC to an instance. ✓

Advanced setup: VLAN

Only for experienced users who have purchased the Oracle Cloud VMware Solution.

Select a subnet in **Networks** [\(Change Compartment\)](#)

☐ Use network security groups to control traffic (optional) <sup>(i)</sup>

☐ Skip source/destination check <sup>(i)</sup>

Assigned IP addresses are usually fine and public IPs are not required. Verify these choices with your OCI network team.

## 2 - Assign Secondary IP Addresses

### Active/Active Cluster

Assign secondary IP addresses to the new repl-a and repl-b VNICs on each controller. Note these IP addresses and the subnet mask for later. These must be assigned only to the repl VNICs and NOT the rx VNICs!

### Active/Passive Cluster

Assign a secondary IP address to the new repl-a VNIC on the active controller and note the IP and subnet mask for later.

Resources

Metrics

IPv4 Addresses

IPv4 Addresses

Assign Secondary Private IP Address

Private IP Address	Public IP Address	Fully Qualified Domain Name	Assigned
192.168.216.216 (Primary IP)	(Not Assigned)	-	Thu, Dec 22, 2022, 17:48:50 UTC

Showing 1 item

## 3 - Reboot Both Controllers

Reboot the nodes to make the new VNICs visible. Reboot one, allow time for it to come back up, then reboot the other. In the case of an active/passive cluster, reboot the passive instance first. There is no preferred order for an active/active cluster.

## 4 – Run Takeover/Failback

When both have come back up, redistribute the resources if needed using Configuration->Cluster->Failback/Takeover from the ZFS BUI.

ConfigurationMaintenanceSharesStatusAnalytics

SERVICESSTORAGENETWORKSANCLUSTERUSERSPREFERENCESSETTINGSALERTS

SETUPUNCONFIGFAILBACKTAKEOVERREVERTAPPLY

jh-zfs-a

Active (takeover completed)

jh-zfs-b

Ready (waiting for failback)

Active Resources

RESOURCE	OWNER
jh-zfs-a (net/vtinet1)	jh-zfs-a
jh-zfs-b (net/vtinet2)	jh-zfs-b
jh-zfs-adm-a (net/vtinet3)	jh-zfs-a
zfs/pool-a	jh-zfs-a
zfs/pool-b	jh-zfs-b

Active Resources

No resources are active on this cluster node.


## 5 – Edit New Interfaces

Start by connecting to the BUI of node A and navigate to the Configuration->Network screen. New devices, datalinks, and interfaces will have been automatically added to this screen corresponding to the VNICs that have been added to the instance.

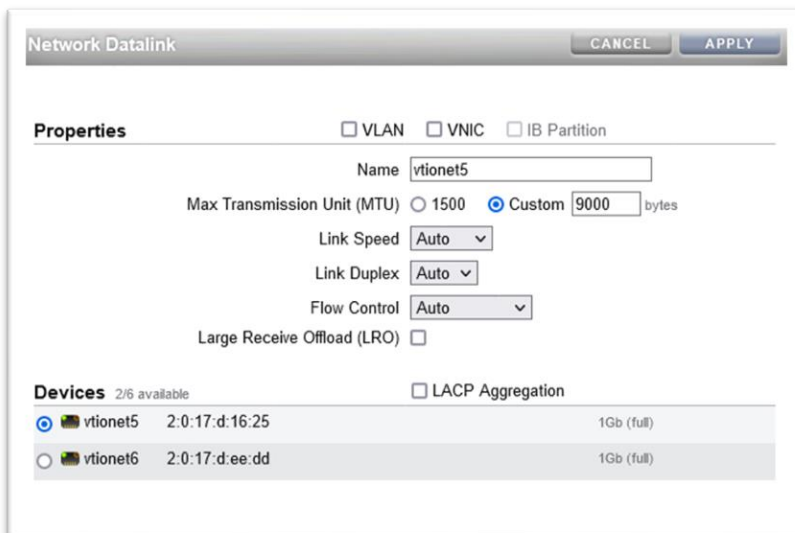
### About Devices, Datalinks, and Interfaces

Network devices represent the VNICs that are attached to a VM and which connect to VCNs in your tenancy. Network devices are created by the system and have no configurable settings. The devices will have a *vtionet* label. Note that the devices will all have a listed speed of 1Gb even though they will use the full bandwidth allowed by the compute shape.

Datalinks are Layer 2 objects for sending and receiving packets for specific network devices. Use datalinks to apply settings such as MTU to network devices. Datalinks can correspond 1:1 with a device, or you can define VLAN datalinks composed of other devices and datalinks. Datalinks are required to complete network configuration, even if the datalinks do not apply specific settings to network devices.

Edit the datalink by clicking on the edit icon  and review the MTU size. An MTU of 9000 gives the best performance and is recommended if all the clients are within OCI. Note that if the client is located across a Wide Area Network (WAN), such as when replicating to or from an on-premises Oracle ZFS Storage Appliance, all steps must support jumbo frames or the connection may hang. (See <https://docs.oracle.com/en-us/iaas/Content/Network/Troubleshoot/connectionhang.htm> for more details.) In these cases, an MTU size of 1500 is recommended on both the source and target datalinks.

Press Apply to add the datalink.



Interfaces are Layer 3 objects for IP, configuring IP addresses and other properties for datalinks. The interfaces created when the new VNICs were attached to the instance are not configured correctly for high-availability clustering the way the data client network is. Edit the interfaces depending on your cluster type.

### Active/Active Cluster

On the A node, the two new interfaces, usually named *vtionet5* and *vtionet6*, correspond to the *repl-a* and *rx-a* devices. Click on the edit icon for the first interface, usually *vtionet5*.

Change the name of the interface to be consistent with our naming convention – here, we use *jh-zfs-repl-a*. Set the static IP Address/Mask to be the same as the **secondary IP address** you assigned to the *repl-a* VNIC in the OCI console – **do not use the primary IP address!** The netmask must be the same as the subnet you chose to add the VNIC to in the OCI console. Refer back to the instance details in the OCI console if needed. Press Apply to save the changes and close the window.

If an MTU change is needed as described in the section “About Devices, Datalinks, and Interfaces”, edit both new datalinks to change it.

Finally, click the Apply button on the main Network screen to save both the datalink and interface changes. Do not click Cancel or the modifications will not be saved even if apply had been clicked in the datalink or interface windows.

Next, go to the Configuration->Network screen on the B controller (jh-zfs-b). Note that the interface for vtionet5 now reflects the name change made on the A instance. Edit the second (usually vtionet6) interface using the steps above. Use the **secondary IP address** that was assigned to the **repl-b** VNIC created in the OCI console, **not the primary IP**.

If the MTU was changed for the new datalink on node A, change the MTU here for the second new datalink shown (vtionet6).

Click Apply on the Network screen to complete the configuration.

## Active/Passive Cluster

On the A node, the new interface, usually named vtionet4, corresponds to the repl-a device. Click on the edit icon new interface.

Change the name of the interface to be consistent with our naming convention – here, we use jh-zfs-repl-a. Set the static IP Address/Mask to be the same as the **secondary IP address** you assigned to the repl-a VNIC in the OCI console – **do not use the primary IP address!** The netmask must be the same as the subnet you chose to add the VNIC to in the OCI console. Refer back to the instance details in the OCI console if needed. Press Apply to save the changes and close the window.

If an MTU change is needed as described in the section “About Devices, Datalinks, and Interfaces”, edit the new datalink to change it. You must change the MTU in the new datalinks on both nodes if a change is needed.

## 6 – Reboot and Rebalance

The final step in the process is to reboot both controllers one more time, again starting with the passive node if there is one, followed by performing a failover or takeback action if needed.

On the Configuration->Cluster screen, we will see the new replication interfaces as active on their respective controllers. Configuration is complete!

We can see from the Configuration->Cluster screen that the new network interfaces are shared and use the IP addresses we assigned.

ConfigurationMaintenanceSharesStatusAnalytics

SERVICESSTORAGENETWORKSANCLUSTERUSERSPREFERENCESSETTINGSALERTS

SETUPUNCONFIGFAILBACKTAKEOVERREVERTAPPLY

jh-zfs-aActive

jh-zfs-bActive

Active Resources

RESOURCE ^	OWNER	
<-> jh-zfs-a (net/vtinet1) 1986.1982.2148.2182	jh-zfs-a	↔ 🔒
<-> jh-zfs-adm-a (net/vtinet3) 1986.1982.2148.2148	jh-zfs-a	↔ 🔒
<-> jh-zfs-repl-a (net/vtinet5) 1986.1982.2147.412	jh-zfs-a	↔ 🔒
zfs/pool-a 1.95T	jh-zfs-a	↔

Active Resources

RESOURCE ^	OWNER	
<-> jh-zfs-b (net/vtinet2) 1986.1982.2148.214	jh-zfs-b	
<-> jh-zfs-repl-b (net/vtinet6) 1986.1982.2148.1482	jh-zfs-b	
zfs/pool-b	jh-zfs-b	

When setting up ZFS replication on other instances to use this cluster as a target, use the IP addresses of these replication interfaces to identify the appropriate target.

## UPGRADING YOUR ZFS-HA INSTANCE

The ZFS-HA controller instances use the same Operating System (OS) releases as the hardware ZFS Storage Appliances. Your ZFS-HA instances should be updated on a regular basis to enable new features and apply bug fixes.

Note that upgrading the software on a single controller system will incur a few minutes of downtime while the system reboots to activate the new code. This time will be slightly higher on a bare metal instance than on a virtual machine instance.

New versions of the OS are released monthly. Release notes and downloads of the OS as well as upgrade instructions are available at [My Oracle Support Doc ID 2021771.1](#).

Because the releases are the same for both the hardware and cloud based ZFS systems, the documentation may refer to service processor (SP) upgrades or device firmware. These do not apply to ZFS-HA instances in OCI and may be ignored.

Once a system is upgraded, you can choose to apply deferred updates at a time of your choosing. Deferred updates can provide additional features or fixes to the system but they do remove the possibility of rolling back an update. It is generally considered a good idea to run a few days with the OS upgrade before applying the deferred updates. Deferred updates can be applied without any downtime or service interruption.

## ZFS-HA SYSTEM NOTES

### Networking

#### ZFS-HA Network Routing

- It is recommended to set the multihoming model to strict. This is the default when the Deployment Tool configures the cluster.

#### ZFS-HA 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-HA Network Interfaces

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

### Block Storage Notes

#### 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'

#### 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.
- It is suggested to create a volume group containing all data disks for each storage pool.
- 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.

#### Boot and Block Volume Backups

While it is possible to use OCI's Boot and Block Volume Backup services to create snapshots of either the storage pool block or boot volumes, it is not recommended. Using these services will require the entire ZFS-HA cluster be shut down to ensure that the snapshots would be usable.

It is recommended instead that the storage pools be backed up in at least one of two ways:

- Using [ZFS Remote Replication](#) to copy share or project snapshots to another ZFS Storage instance, whether an on-premises ZFS Storage Appliance or other ZFS-HA instances in OCI
- To OCI Object Storage using the ZFS appliance's built-in Cloud service, which leverages ZFS snapshots for object storage backups. See the documentation on [Configuring Cloud Backups](#) for details on enabling and using this service.

It is also recommended that rather than boot volume backups, the ZFS configuration be backed up and downloaded for safekeeping.



## Backing Up the ZFS Configuration

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

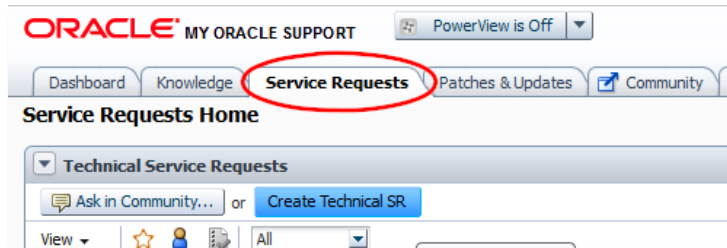
- From the Appliance BUI, go to Maintenance→System.
- Under the Configurations section, click Backup.
- This will create a backup of the Appliance configuration, that can be downloaded and stored separately for recover purposes.

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

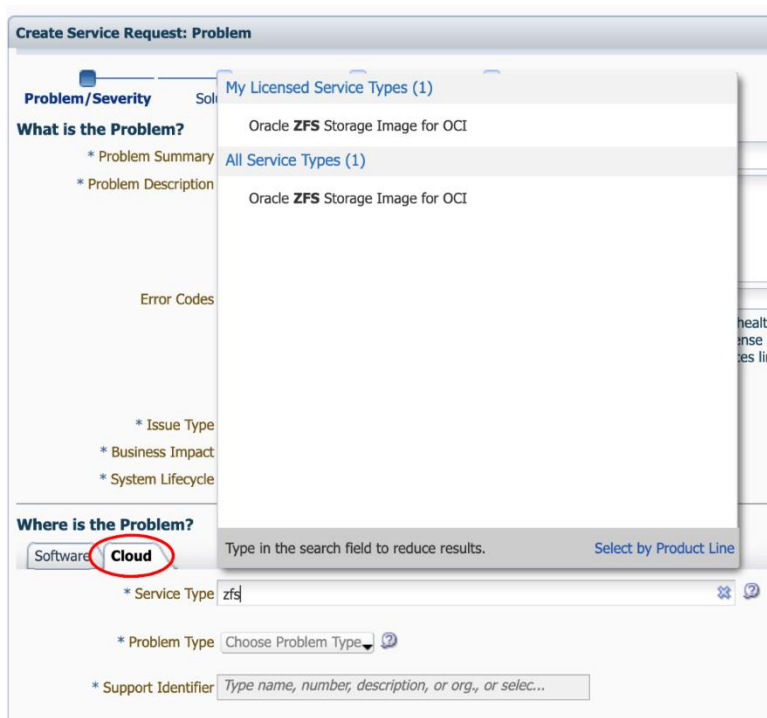
## Submitting a Service Request

ZFS-HA instances are eligible for support. The steps below will walk you through submitting a service request (SR) for your instance in the event of a problem with your installation.

- Log into the My Oracle Support portal at <https://support.oracle.com>. Create an account if you do not already have one.
- Click on the “Service Requests” tab along the top of the page, then click the “Create Technical SR” button.



- Enter a Problem Summary and Problem Description in the appropriate fields. Note the character length restrictions in each text box.
- From the pull-down menus, select the appropriate responses for Issue Type, Business Impact, and System Lifecycle.
- In the “Where is the Problem” section, click on the “Cloud” tab and then type “ZFS” in the Service Type field. “Oracle ZFS Storage Image for OCI” will be displayed. Select it, then choose the closest fit for Problem Type from the pulldown menu. Click in the Support Identifier field and choose the appropriate selection from the options chosen.



- Click the “Next” button to continue through the next screens, adding any relevant information as needed. On the last screen, confirm your contact information including the Contact Method, then press the “Submit” button.

## DOCUMENTATION AND 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](#)
- [Oracle ZFS Storage Appliance RESTful API 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.

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 a saved Terraform 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)
- When data connections are made across Wide Area Networks (WANs) that do not support jumbo frames at each step along the network path, the connections may hang. (See <https://docs.oracle.com/en-us/iaas/Content/Network/Troubleshoot/connectionhang.htm> for more detail.) In these cases it is best to set the MTU on the ZFS-HA datalink used and at the on-premises client to 1500.

### CONNECT WITH US

Call +1.800.ORACLE1 or visit [oracle.com](https://www.oracle.com).

Outside North America, find your local office at [oracle.com/contact](https://www.oracle.com/contact).

 [blogs.oracle.com](https://blogs.oracle.com)

 [facebook.com/oracle](https://facebook.com/oracle)

 [twitter.com/oracle](https://twitter.com/oracle)

Copyright © 2023, 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

ZFS-HA Quick Start Guide v4.10  
September 2023  
Author: Joe Hartley

