

Networking Best Practices with Oracle ZFS Storage Appliance


ORACLE WHITE PAPER | MAY 2018





Contents

Introduction	1
Oracle ZFS Storage Appliance Network Architecture	2
Hardware Components	2
Applying Networking Configuration Concepts to Oracle ZFS Storage Appliance	3
Devices as Layer One of the OSI Model	3
Datalinks as Layer Two of the Network Stack	4
Interfaces as the Top Network Layer	4
Routing and Gateways	4
Link Aggregation Control Protocol (LACP)	4
IP Networking Multipathing (IPMP)	5
Network Virtualization and Virtual Networks	5
Clustering Concepts	5
Initial Setup and Network Configuration for Oracle ZFS Storage Appliance	6
How It All Fits Together	6
Network Design Considerations and Best Practices	10
Topology	11
Simple Setup for Network Storage	11
Desktop User Home Directory	11
Application Database Servers	12
Direct Network Connections	12
Routing Tables	13
Route Flapping	14



Security	14
Scalability and Availability	15
LACP	15
IPMP	16
Comparing LACP and IPMP	16
Configuring IPMP	17
Configuring LACP	18
Multiple Switch Link Aggregation	19
VLANs and VNICs	19
MTU Recommendations	19
Replication	21
Hardware Changes	21
Remote Access	22
Configuration Recommendations for Specific Protocol Settings	22
NFS Protocol	22
InfiniBand Protocol	22
Cluster Networking	22
Direct Network Connections	24
Cluster Networking Best Practices	26
Troubleshooting Network Configuration Problems	28
Appendix A: Services and Associated IP Ports	30
Appendix B: Accessing the Oracle ZFS Storage Appliance Console Through the Oracle ILOM Server	31



Setting Up a Serial Connection to the Oracle ILOM Server	31
Accessing the Oracle ZFS Storage Appliance Console	31
Serial RJ45 Signal Definitions	32
Appendix C: References	33



Introduction

The Oracle ZFS Storage Appliance family combines advanced hardware and software architecture in multiprotocol storage subsystems that enable users to simultaneously run a variety of application workloads and offer advanced data services. First-class performance characteristics are illustrated by the results of the industry standard benchmarks like SPC-1, SPC-2, and SPECsfs.

Oracle ZFS Storage Appliance uses Ethernet, InfiniBand, and Fibre Channel (FC) connectivity to offer file- and block-based data storage services. While Fibre Channel connectivity is straightforward, network IP connectivity using Ethernet and InfiniBand can become complex, requiring informed design and configuration to simplify ongoing management and provide optimal overall performance for the network storage subsystem.

This document provides guidance and best practices on how to integrate an Oracle ZFS Storage Appliance system into a network infrastructure, monitor its functioning, and troubleshoot any operational network problems.

A related document addresses the design and operation of Oracle ZFS Storage Appliance in a Fibre Channel-based infrastructure. See “Appendix C: References,” for further details.

Oracle ZFS Storage Appliance Network Architecture

The following sections describe both the network architecture of Oracle ZFS Storage Appliance and its options, and the possible setup and configuration methods and options available.

Hardware Components

Oracle ZFS Storage Appliance architecture is designed for flexibility, offering a number of network connection options. File-based services can be configured using Ethernet and InfiniBand network adapter options. Block-based (LUN) services can use either network (iSCSI) or Fibre Channel adapter options.

Separate network and serial ports are provided for diagnostic and remote access purposes.

Each hardware node has four built-in LAN network ports.

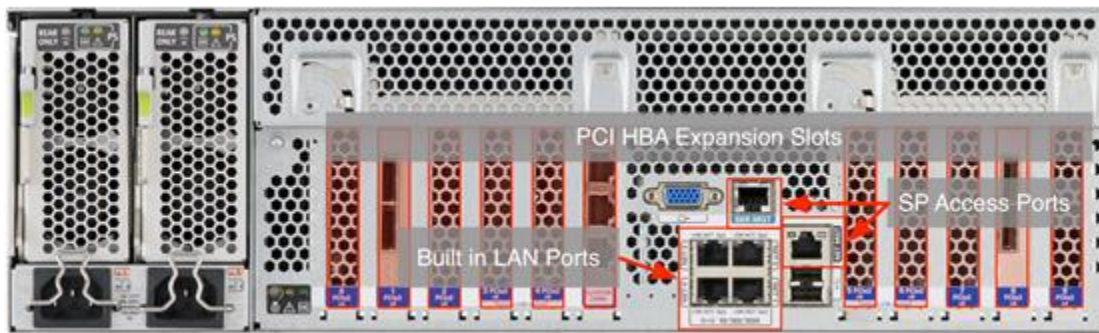


Figure 1. Network hardware components and connections

Further network connections can be added using the peripheral component interconnect (PCI) host bus adapter (HBA) expansion slots. Consult the hardware documentation for the specific model for available Ethernet, InfiniBand, and Fibre Channel HBAs.

Newly added network host bus adapters are automatically recognized and shown in the network configuration window of the Oracle ZFS Storage Appliance BUI, shown in the following figure. Take care not to move existing adapters when adding new HBAs to the system, and always ensure that cluster partners have matching cards in the same slots.

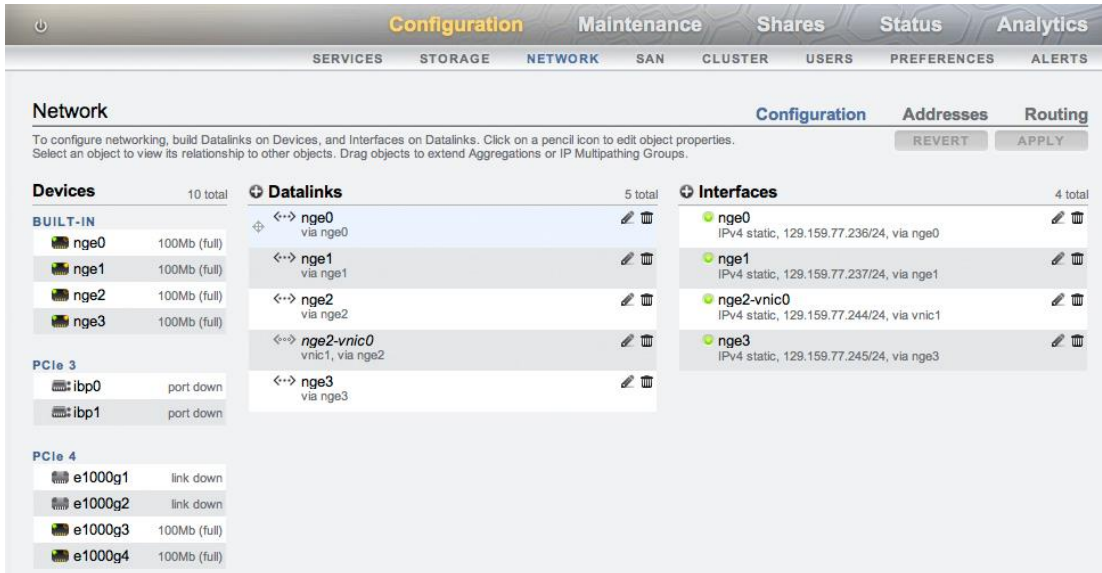


Figure 2. Network configuration screen in the Oracle ZFS Storage Appliance BUI

The physical network ports from the HBAs are shown as devices on the left-hand side of the BUI window along with their status.

Applying Networking Configuration Concepts to Oracle ZFS Storage Appliance

Oracle ZFS Storage Appliance presents the network stack to the user using the first three layers of the Open Systems Interconnection (OSI) model: devices, datalinks, and interfaces. Basing a network stack this way provides a very flexible configuration environment in which any combination of virtual, redundancy, and link aggregation options can be used to configure a reliable and well-performing networking architecture.

Devices as Layer One of the OSI Model

Devices represent the physical layer (layer one) of the OSI model and are the basic building blocks for the configuration network stack of Oracle ZFS Storage Appliance. Devices represent the physical ports of Ethernet HBAs or the IP on InfiniBand (IPoIB) partitions for InfiniBand HBAs.



Figure 3. Device icons and status indicators displayed in the Oracle ZFS Storage Appliance BUI

During the Oracle ZFS Storage Appliance startup, the devices are automatically recognized and presented in the Network Configuration screen of the BUI. The status of each device is shown by two LEDs in the icon, along with its negotiated speed and duplex mode. The left LED shows the connection status and the right LED shows network traffic activity. These device objects form the basis for the network connection configuration in Oracle ZFS Storage Appliance.

Datalinks as Layer Two of the Network Stack

Datalinks represent the datalink layer (layer two) of the OSI model and are the next building blocks in the network stack. Datalinks can either be built on top of devices on a one-to-one basis to form a physical datalink object or connected to multiple devices to form an aggregated connection to create a logical datalink object. Aggregated connections are built using the Link Aggregation Control Protocol (LACP). LACP allows you to combine multiple physical devices into a single logical connection to increase throughput. An LACP group is presented as a single logical datalink object. To be able to use LACP, an Oracle ZFS Storage Appliance system must be connected to a network switch that supports LACP.

Virtual datalink objects can be built on top of physical or logical datalink objects using the datalink virtual local area network (VLAN) or virtual network interface card (VNIC) configuration properties. Multiple virtual datalink objects can be created on a single physical datalink object. This layering of datalink objects enables you to implement very flexible virtual networking environments and fully utilize the physical device objects in an Oracle ZFS Storage Appliance clustered configuration.

The InfiniBand implementation in Oracle ZFS Storage Appliance is IP over InfiniBand (IB) rather than Ethernet over IB, which means that the configurations at layer two differ between IB and Ethernet. The datalink object's configuration reflect those differences. When configuring InfiniBand datalink objects, the options to create VLANs and VNICs are not present since they are Ethernet constructs. However, IB partition keys, which function similarly to VLANs, are available. You can create multiple InfiniBand datalink objects on top of an InfiniBand device object as long as the partition keys are unique for each InfiniBand datalink object sharing an InfiniBand device object.

Interfaces as the Top Network Layer

Interfaces represent the network layer (layer three) of the OSI model and are the building blocks at the top of the network configuration stack of the Oracle ZFS Storage Appliance system. They are used to configure the IP address and netmask properties of a connection for a specific datalink object. Both IPv4 and IPv6 protocols are supported. Multiple interface connections can be grouped into a multipath interface connection using the IP network multipathing (IPMP) mechanism.


Routing and Gateways

When IP packets are sent to a destination node with an IP address outside the subnet of the sender's IP address, information is required about which gateway(s) the packets have to be sent to in order to reach the destination IP address. Each sender needs the information to the next gateway. For an Oracle ZFS Storage Appliance system, it is sufficient to have information on the IP address of the nearest gateway for each of its configured subnets. Routing information for Oracle ZFS Storage Appliance is kept in a single routing table consisting of a collection of routing table entries. In a clustered configuration, both nodes share the same routing table.

Link Aggregation Control Protocol (LACP)

Oracle ZFS Storage Appliance supports the use of link aggregation in order to push network links throughput beyond the physical limitations of a single physical network interface card (NIC) port. All network devices participating in the LACP group must be connected to a network switch that supports LACP and has LACP enabled for the connected ports. Alternatively, direct network connections to a peer (server) that has LACP configured for those ports is supported as well. All devices grouped together using LACP form a datalink object in the Oracle ZFS Storage Appliance system. The LACP type datalink can be used to create virtual datalinks or create interface objects on top of it.

One word of caution on link aggregation: the potential theoretical aggregated network link speed might not be reached. Think of the supermarket checkout stations. If there are 10 stations open but no more than four customers arrive at the same time, the other six stations are idle. The same is true for aggregated network devices in an LACP



group. If there are four devices configured in an LACP group but no more than two network sessions are open at one point in time, only half of the available bandwidth will be utilized. You can use the LACP policy attribute to influence the method of load spreading. See the section "Comparing LACP and IPMP" for more details.

IP Networking Multipathing (IPMP)

Because IPMP is a more complex mechanism, it requires more conceptual and functional understanding. IPMP provides a mechanism for Oracle ZFS Storage Appliance to increase reliability and availability of network connections. This is done by configuring two or more interface objects into an IPMP group. An IPMP group consists of a number of interface objects and one or more logical data addresses. Interface objects in an IPMP can either be active or standby. Oracle ZFS Storage Appliance monitors all interface objects in an IPMP group for failures. If a failure is detected for an active interface object, a standby interface object is made active and IPMP automatically migrates all data IP addresses used on the failed interface object to one of the remaining available interface objects within the IPMP group.

IPMP uses two forms of failure detection: link based and probe based. For link-based failure detection, the IP interface driver link status is used to determine network connection failures.

Probe-based failure detection is handled by the IPMP daemon. It continuously checks connectivity of the interface objects in the IPMP group to the surrounding servers using the IP test address of each interface object in the IPMP group. The probing process finds the target systems to probe by going through the routing table to find the gateway for the subnet of the related interface object. If no gateways are found, an Internet Control Message Protocol (ICMP) multicast probe is used to find surrounding servers. The first five responding servers are used in the probing process.

The network interface objects in an IPMP group can be configured as active or passive. Active interfaces are assigned an IP data address from the IPMP group. Passive interfaces are treated as standby interfaces; if a failure is detected on an active interface, all data IP addresses on that interface are moved to the passive interface and the passive interface is made active.


Network Virtualization and Virtual Networks

Network virtualization is a widely accepted technology that is used by many customers to logically separate network data traffic and user/client access using VLAN functionality or to overcome restrictions of the number of physically available network segments and network ports. Oracle ZFS Storage Appliance can participate in VLANs which adhere to the 802.1q standard for VLAN tagging, or which are operated transparently by the switch by dedicating specific switch ports to specific VLANs. A virtual network layer also can be created within Oracle ZFS Storage Appliance using virtual network ports that are created on top of physical datalink objects.

When creating a virtual network layer, avoid overprovisioning of physical HBA ports bandwidth, which could create performance bottlenecks on physical network links.

Clustering Concepts

For high-availability environments, Oracle ZFS Storage Appliance can be configured using a two-node cluster configuration. Both nodes are identical in hardware configuration and can be configured in an active-active or active-passive mode of operation. In an active-passive mode of operation, all services are configured to run on the active node, while the passive node is in standby mode. The passive node takes over all services and network connections in case there is a catastrophic failure on the active node. In an active-active mode of operation, both nodes actively provide data services to clients. When a catastrophic failure occurs on one node, the other node takes over the data services and network connections of the failed node.



The two nodes monitor each other's status through a private cluster interconnect interface. The network configuration information is shared between the nodes and changes made on one node are automatically synchronized with the other node.

Active/Passive Cluster Operation Mode

In this mode, the active node serves all storage pools. All network connections are instantiated on the active node. In this mode the active node servers all have data services through the configured network or FC connections to the clients. A node role switch is triggered by either a failure on the active node or a manual user-initiated node failover action. The passive node is always in standby mode. Cluster resource sizing is based on the active node being able to handle all clients' I/O requests.

Active/Active Cluster Operation Mode

In this operational mode, both nodes are active. Each node is configured to serve a part of the storage. Note that network and storage resources cannot be shared between nodes; they only can be owned by one node at a time. When a failure brings one node down, the other node takes over the data services of the failed node. Client application I/O loads can be balanced between the nodes, so both nodes can be fully utilized.

Cluster resource sizing is based on both nodes handling client I/O requests and a reduced response during a failure when the remaining node has taken over the data services of the failing node.

Initial Setup and Network Configuration for Oracle ZFS Storage Appliance

A new Oracle ZFS Storage Appliance system does not contain any network configuration. The initial installation process will walk you through a procedure to provide Oracle ZFS Storage Appliance with an IP address, gateway, and DNS information to set up a single network port for gaining administrative access to the system. A connection to the Oracle Integrated Lights Out Manager (Oracle ILOM) server needs to be set up in order to complete this initial Oracle ZFS Storage Appliance installation process. See "Appendix B" for more information.

After the initial setup process is finished, the network configuration can be set up according to customer requirements using either the Oracle ZFS Storage Appliance CLI or BUI interface.

How It All Fits Together

The following configuration illustration, showing the network configuration options and capabilities of Oracle ZFS Storage Appliance, will be referenced throughout this paper. This example configuration uses most of the available configuration options and as such is not representative of a real-world configuration. The example consists of a local network with three subnets. A router provides the gateway function for these subnets to the bigger IP network.

Figure 4 shows the previously mentioned network stack layers and the building blocks available in each layer for an Oracle ZFS Storage Appliance system. The Oracle ZFS Storage Appliance system in the example is connected to one physical and two virtual subnets. The subnets are connected by a router to the outside world with a gateway address set up for each subnet. All network ports of Oracle ZFS Storage Appliance are connected to the same network switch. This switch supports the LACP protocol.

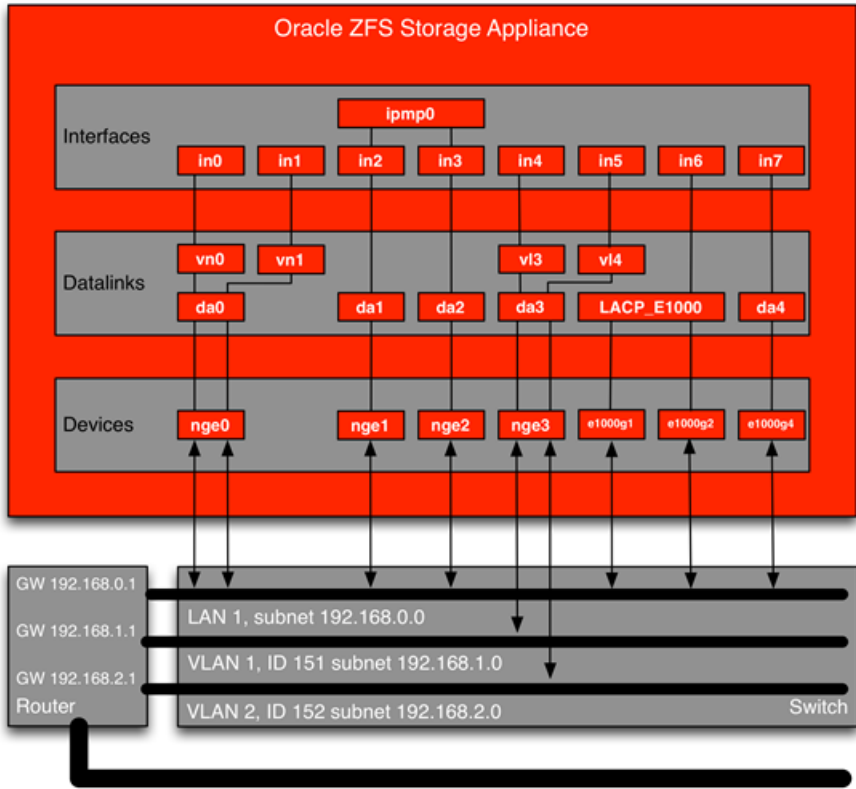


Figure 4. Network configuration options for Oracle ZFS Storage Appliance

The available network ports are shown in the devices layer of the network stack. In the next layer, a number of datalink objects have been created. There are two types of datalink objects: physical datalink and virtual datalink objects. The physical-type objects are directly connected to a device object. VNIC (vnx) and Virtual LAN (vlx) datalink objects are virtual-type objects and are linked to the physical datalink object. One physical-type object can have multiple virtual-type objects connected to it. Two devices have been aggregated into one datalink object using the LACP option.

At the top layer, interface objects have been created. These interface objects define the IP objects used to communicate to the outside world.

The following discussion provides more detail for some of the configuration options shown.

Figure 4, on the far right, shows the simplest configuration: a device (e1000g4), datalink (da4), and interface (in7) object forming a simple one-to-one relationship with each other. In7 has been given a static IP address and is using the datalink object da4. Under properties, administrative BUI/CLI access can be enabled or disabled as shown in figure 5.

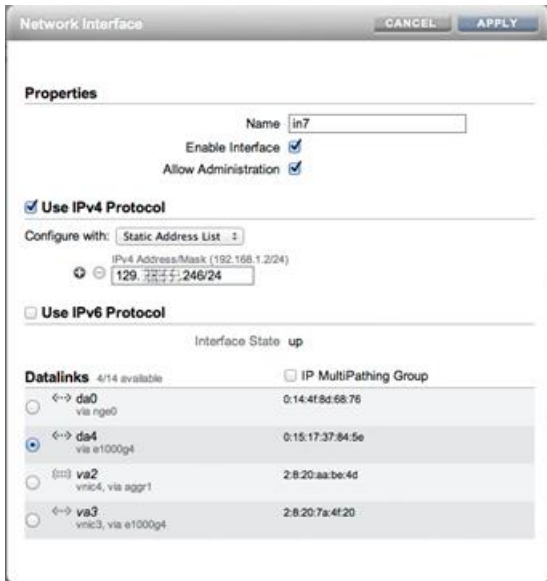


Figure 5. Simple interface object setup

Next, the virtual network interface controllers (VNICs) are set up. A VNIC is a virtual datalink object created on top of a physical type of datalink object. One physical datalink can have a number of VNICs associated with it. A maximum of eight VNICs per physical datalink object is recommended. The example shows two VNICs configured on one physical datalink. Each VNIC (vn0 and vn1) is associated with its own interface object (in0 and in1). Both interfaces share the same device (nge0) and consequently share the bandwidth available on that device.

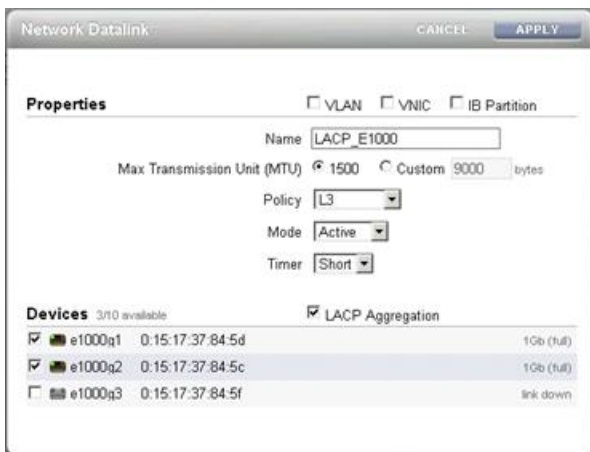


Figure 6. LACP datalink object setup

Link aggregation is set up for the device objects e1000g1 and e1000g2 by creating an LACP-enabled datalink and adding e1000g1 and e1000g2 to it. Only unallocated device objects can be added to an LACP datalink object.

The most complex object to set up and understand is the IPMP interface object. You start by creating simple interface objects. Each interface object is configured to use its own physical datalink object. Note that Oracle ZFS Storage Appliance allows you to create other IPMP group setups, such as IPMP groups using VNIC or VLAN datalink objects, for example. These types of configurations are not realistic and not recommended for use; as such, they are outside the scope of this document. Keep in mind the core function of IPMP: to mitigate the risk of losing

connection by losing link communication. ZFSSA nodes in clusters do not failover due to permanent network failures! This is very important to keep in mind.

Creating two VNICs on the same device port and building IPMP on top of them, for example, is a potential setup that does not make sense in light of that core IPMP function to provide a redundancy. IPMP is designed to provide redundancy when a physical link fails, no point creating an IPMP group on a single physical link using multiple VNICs on them.



Figure 7. IPMP interface object setup

Each interface object that is created to be part of the IPMP group can be given an IP address. This address functions as a test address for the IPMP daemon to detect network path failures and cannot be used by clients for accessing data services on Oracle ZFS Storage Appliance. To disable the IPMP probing mechanism for the current IPMP group, use IP address 0.0.0.0/8 for each interface object's IP address.

To create a new IPMP group, use the + icon next to "Interfaces" and select the IPMP option. Then select the participating (simple) interface objects that will be members of the IPMP group. Each selected interface can be set as an active or standby interface in the IPMP group.

One or more IP addresses can be assigned to the IPMP interface object. These data addresses are the IP addresses to be used by clients to access Oracle ZFS Storage Appliance data services. All IP test addresses used within the IPMP group need to be from the same subnet.

The following BUI screenshot shows the example configuration as implemented on an Oracle ZFS Storage Appliance system. You can easily identify all elements in the different layers used for a specific connection by simply mouse-clicking on one of the elements on the Network Configuration screen.

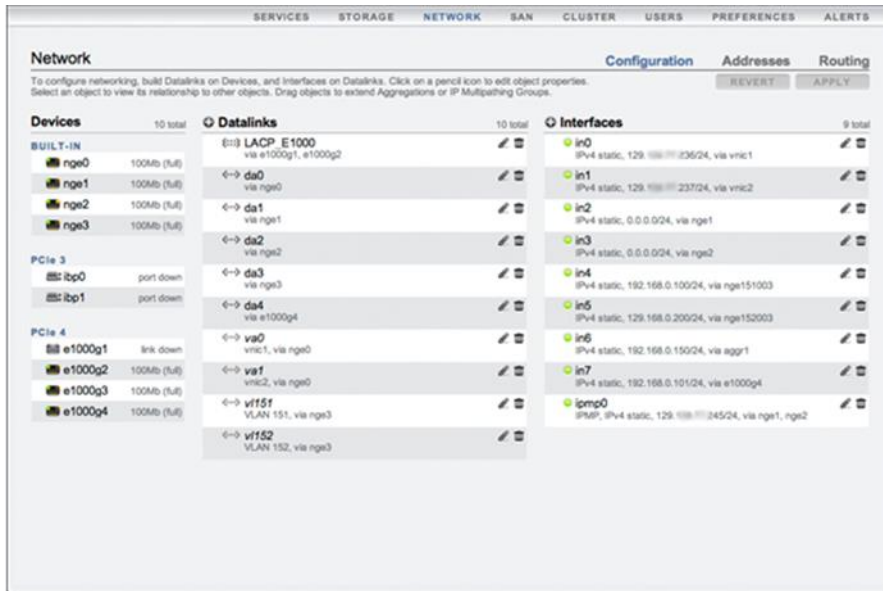


Figure 8. BUI Network screen for example configuration

Network Design Considerations and Best Practices

Many factors need to be considered before starting a network design incorporating storage subsystem data traffic. Key elements to scope according to customer and application(s) requirement are:

- » Reliability/availability
- » Scalability/performance
- » Security
- » Manageability
- » Available budget for both capital investment and operational costs

The first step to take is to determine the business requirements and translate those into IT-specific requirements. Reliability and availability information can be retrieved from a business continuity (BC) plan and derived from the service level agreements (SLAs) between the IT organization and the various business departments in the company.

Use the information from scoping the listed requirements to find the right balance between solution costs and the cost impact of loss of services that is acceptable for the business. It is equally important to keep all involved groups in the loop so that agreed-upon expectations are maintained and verified.

Avoid making designs too complex, or meant to solve problems that do not exist. Put yourself in the position of support engineers who are called out of bed at 4 a.m. on a weekend to deal with a major downtime escalation. They must be able to understand the design in order to properly troubleshoot it.

Topology

Network topologies can be looked at from both a physical and a logical perspective. The physical perspective deals with the various network components used and how they are wired together. The logical perspective deals with the data streams between the various network nodes. Especially when virtual network elements are used, the logical and physical views can be very different. Be careful to avoid oversubscribing the bandwidth capacity of physical components when mapping a number of logical data streams on top of them.

In larger configurations it is good practice to separate different types of data access, either on a physical level or logical level, depending on the infrastructure already present or specific customer requirements. For security reasons, administrative access can be separated from the data type services by using different subnets or VLANs for each of them.

Also, on larger systems, it makes sense to separate out connections to antivirus scanners so that the traffic between the antivirus scanners and the storage subsystems do not interfere with application data traffic.

In simple configurations where a single Oracle ZFS Storage Appliance system is used with one or two application servers, direct connections for data traffic between Oracle ZFS Storage Appliance systems and servers might be considered without using LAN network switches. By using 10 GbE network connections, dedicated high-speed datalinks are available for data access without the investment in a full switch-based 10 GbE LAN infrastructure. This isolates data traffic from the customer LAN infrastructure, thus providing an extra level of security, predictable performance, and cost reduction.

Simple Setup for Network Storage

The most common network topology used when deploying network storage is to connect both users and network storage subsystems to a corporate network. Often a separate administrative network/subnet is already present. If not, it is highly recommended to create one. This adds an extra level of security and reliability.

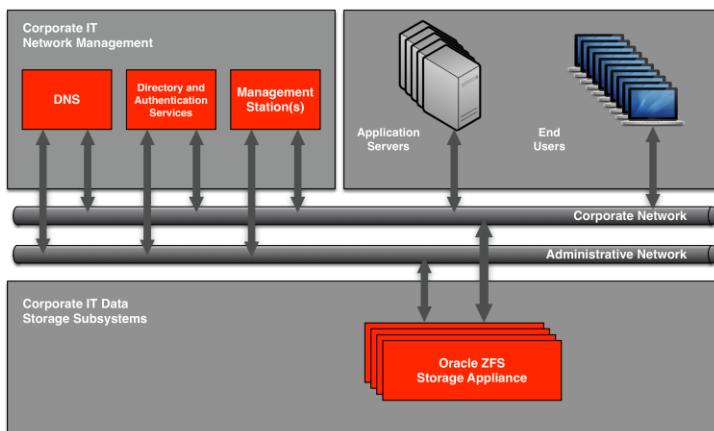


Figure 9. Most common network topology for network storage

On each node that needs administrative access, administrative access rights can be restricted to one or two network ports. Create redundancy for access to the corporate Domain Name System (DNS) and directory services by setting up routing tables on the storage subsystems for each subnet.

Desktop User Home Directory

Environments with lots of desktop users benefit from consolidating all home directory storage into centralized managed storage subsystems. As an extra benefit, data protection services can be centralized as well. Oracle ZFS

Storage Appliance offers antivirus services in combination with virus scanner servers as provided by well-known data security software vendors. File scan traffic can be isolated from the corporate network using a dedicated subnet for connections between antivirus scanner servers and Oracle ZFS Storage Appliance.

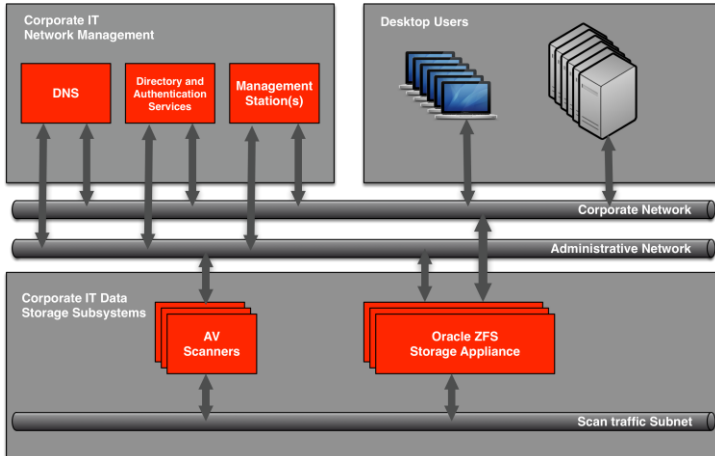


Figure 10. Home directory network topology setup

Application Database Servers

In environments where there is a significant amount of network traffic between the storage subsystem and application database servers, it makes sense to isolate that traffic by creating a separate application data subnet and route all data traffic between the application database servers and the storage subsystem through that subnet. Data traffic between end users and the application database server can still be handled through the main corporate network.

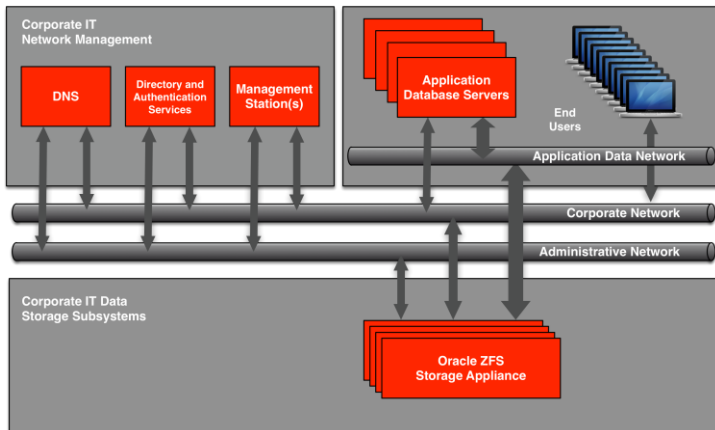


Figure 11. Application database network topology setup

Direct Network Connections

When using one or two application servers in combination with an Oracle ZFS Storage Appliance system, it makes sense to use direct connected 10 GbE connections, saving the cost and complexity of using 10 GbE network switches. When using Oracle Solaris on the servers, the Oracle Solaris IPMP functionality on the server can be used to create redundant connections to an Oracle ZFS Storage Appliance cluster configuration. Refer to the “Cluster Networking” section of this document for an example.

Routing Tables

Oracle ZFS Storage Appliance maintains an IP routing table. The routing table is used to determine which route to use in order to deliver an IP packet for a certain destination IP address. If the routing table entry contains gateway information, the next hop gateway is used to which the IP packet is sent.

Routing table entries can be automatically managed when using the Routing Information Protocol (RIP). RIP can be configured under the Oracle ZFS Storage Appliance BUI Services, System Settings tabs. Oracle ZFS Storage Appliance supports both RIPv1 and RIPv2 for IPv4, and RIPng for IPv6. Routes that are configured using these protocols are marked as type "dynamic" in the routing table.

Pay attention to the routing table entries when Oracle ZFS Storage Appliance is connected to different subnets and you want full control over which interfaces (and thus device ports) are used for data exchange between Oracle ZFS Storage Appliance and certain subnets. For instance, you would not want to have some data traffic being routed over an administrative subnet.

Routing of IP packets can be managed in two ways:

- » Managing the contents of the routing table, either by editing a routing table entry or by controlling the process of routing table information gathering.
- » Managing the way routing information is used to determine the optimal route for the IP packet to reach its destination.

An Oracle ZFS Storage Appliance system gathers routing information from different sources; the type of source determines if an entry can be changed or deleted. System, Dynamic Host Configuration Protocol (DHCP), and dynamic-type entries cannot be edited or deleted.

Dynamic routing information is gathered through RIP. By switching off RIP in the BUI's Services menu, these types of entries can be deleted but they have to be replaced by static-type entries that are provided by the administrator.

When using more than one IP interface in an Oracle ZFS Storage Appliance system, multiple equivalent routes may be possible for a certain destination and IP packets may arrive on one IP interface while the optimal return route might be hosted by another IP interface. Oracle ZFS Storage Appliance provides three different types of policies as potential methods to select the most optimal route for an outgoing IP packet. Set the policy type under the multihoming model property of the routing table, as shown in figure 12.



Figure 12. Selecting the multihoming model property for routing

TABLE 1. ROUTING MULTIHOMING PROPERTY OPTIONS

POLICY	DESCRIPTION
Loose	Do not enforce any binding between an IP packet and the IP interface used to send or receive it: 1) An IP packet will be accepted on an IP interface so long as its destination IP address is up on an Oracle ZFS Storage Appliance system. 2) An IP packet will be transmitted over the IP interface tied to the route that most specifically matches an IP packet's destination address, without any regard for the IP addresses hosted on that IP interface.

	If no eligible routes exist, drop the packet.
Adaptive	<p>Identical to loose policy, except prefer routes with a gateway address on the same subnet as the packet's source IP address:</p> <ol style="list-style-type: none"> 1) An IP packet will be accepted on an IP interface so long as its destination IP address is up on an Oracle ZFS Storage Appliance system. 2) An IP packet will be transmitted over the IP interface tied to the route that most specifically matches an IP packet's destination address. <p>If multiple routes are equally specific, prefer routes that have a gateway address on the same subnet as the packet's source address. If no eligible routes exist, drop the packet.</p>
Strict	<p>Require a strict binding between an IP packet and the IP interface used to send or receive it:</p> <ol style="list-style-type: none"> 1) An IP packet will be accepted on an IP interface so long as its destination IP address is up on that IP interface. 2) An IP packet will be transmitted only over an IP interface if its source IP address is up on that IP interface. To enforce this, when matching against the available routes, Oracle ZFS Storage Appliance will ignore any routes that have gateway addresses on a different subnet from the packet's source address. <p>If no eligible routes remain, drop the packet.</p>

For new configurations, the use of the strict policy is recommended. Using this option gives the expected results from the routing rules in the routing table.

When configuring active/active cluster configurations, you must configure a default route on each of the nodes. See the "Cluster Networking" section for further details.

Route Flapping

When using dynamic routing in your network infrastructure, "route flapping" may cause instability, resulting in reduced throughput and higher network service latency. Route flapping is caused by a router continuously flipping between routes as a result of malfunctioning cables, intermittent failing router ports, or high packet loss. This manifests itself as frequent updates in routing information for a certain destination. To diagnose route flapping, look at frequent updates in routing tables for a certain route, or a high level of packet drops at certain network ports on a router.

Reduce the risk of route flapping occurrences by setting route flapping protection policies in your routers (route dampening), when supported, and use IPMP and/or LACP to reduce the impact of malfunctioning cables and network ports.

Configuration errors such as wrong Maximum Transmission Unit (MTU) settings in one of the network components for a certain network path can cause high packet loss.

Security

As already mentioned, it is strongly advised to limit the number of network interfaces that have administrative access enabled. Only allow administrative access on interfaces designated for administrative access. Keep administrative-enabled ports on a separate subnet. To prevent being locked out from administrative access, make sure there is an alternative administrative access method available in case the main allocated administrative interface fails. Either have a second interface enabled or wire the Oracle ILOM management port of an Oracle ZFS Storage Appliance system into the administrative network, too. This enables the use of the Oracle ZFS Storage Appliance (virtual) console for storage access for administrative functions in case Oracle ZFS Storage Appliance cannot be reached using the primary administrative network interface. See "Appendix B: Accessing the Oracle ZFS Storage Appliance Console Through the Oracle ILOM Server" for how to configure Oracle ILOM access.

You can set up Oracle ZFS Storage Appliance to use user accounts and authentication through central LDAP or Active Directory services. Oracle ZFS Storage Appliance logs successful user login events as part of its audit functionality.

Set up access restrictions for block devices using the initiator and target group options. The target group defines over which network interface ports the block device (LUN) is made visible, and the initiator group defines which clients are granted access to a specific block device.

Access restrictions for NFS filesystems can be set up using the NFS Exceptions option under Protocols in the Shares section. The exception rules are applied to clients that match the specified criteria. If you do not want clients to have access rights besides the ones set up in the exception rules, specify “never” in the share mode option for NFS.



Figure 13. Setting up NFS access restrictions

More detailed information on the security aspects of setup and maintenance of Oracle ZFS Storage Appliance can be found in the Oracle ZFS Storage Appliance Security Guide.

Scalability and Availability

Scalability and availability requirements often are seen as the same. They often go hand in hand but there are differences. For instance, by using trunked network connections, bandwidth is increased and protection against the loss of a network link is created, but the loss of the network switch to which all trunked lines are connected still causes loss of connectivity. It is key to delineate what the scalability and availability requirements are and keep them separated. Once they are clear, you can see if technology used for scalability can be used to meet availability requirements and vice versa.

Oracle ZFS Storage Appliance offers two network services in this area: LACP for scalability and IPMP for availability. LACP offers some extra availability while IPMP offers some scalability.

LACP

LACP is implemented in layer 2 in the network stack. This layer is hardware dependent, and LACP is limited to the Ethernet type hardware, excluding the use of LACP on InfiniBand, for instance. Bandwidth aggregation is transparent, meaning no extra configuration steps are needed in Oracle ZFS Storage Appliance. Each (virtual) datalink object and/or interface object configured on top of an LACP-type datalink object has full use of the aggregated bandwidth of all device object members within that LACP datalink object.

LACP always operates on a peer-to-peer basis. LACP must be set up on both sides of the physical wires, either a network switch/router or, in case of a direct host connection, a host that runs an operating system that supports the LACP functionality in its network stack.

IPMP

Consider the following IPMP configuration rules when considering its use:

- » All IPMP test addresses within an IPMP group have to be within the same subnet.
- » There can be only one IPMP group used per IP test address subnet.
Note that test addresses do not have to be in the same subnet as the IP addresses used for the IPMP data interfaces. Also, non-routable addresses (like 192.168.x.x) can be used for test addresses as long as the discoverable addresses of the targets used by IPMP for the probing are in the same subnet as the test addresses in the IPMP group.
- » Only servers within the same L2 ICMP broadcast domain as Oracle ZFS Storage Appliance can participate in the IPMP probe error detection mechanism.

The following diagram illustrates the IPMP configuration rules.

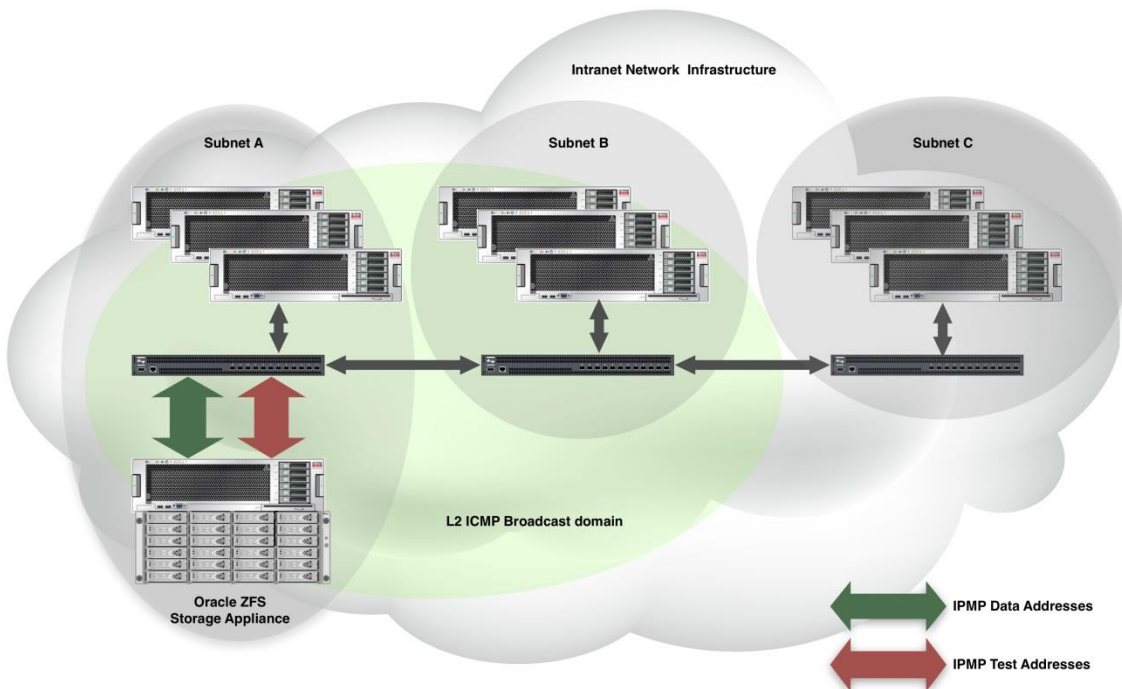


Figure 14. IPMP configuration

There is no restriction on the use of data addresses in an IPMP group. But when using IPMP probing error detection, understand that using data addresses that are outside the network area that is covered by the IPMP probing error detection will not be covered by the IPMP connection loss detection mechanism.

Comparing LACP and IPMP

Both LACP and IPMP group network interfaces share the same related features, but there are differences between the two protocols. A key difference is the network layer in which each is implemented. Based on this, LACP has less of a performance impact than IPMP, while IPMP has some restrictions on the use of IP test addresses. The following table details the differences.

TABLE 2. LACP COMPARED TO IPMP

Function	LACP	IPMP
----------	------	------

Network layer	Network layer 2	Network layer 3
Link-based failure detection	Yes	Yes
Probe-based failure detection	Based on Link Aggregation Control Protocol (LACP), targeting immediate peer host or switch.	ICMP-based, targeting any system across multiple levels of intervening layer 2 switches.
Max. network path covered	Network path between Oracle ZFS Storage Appliance and its LACP peer, being a switch or a server when direct connect used.	Network path between Oracle ZFS Storage Appliance and the network segment to the clients as detected by the ICMP broadcast responses.
Use of standby interfaces	No	Yes
Span multiple switches	There is no standard (RFC) for this; there are some switch vendor-specific implementations available.	Yes, as long as all test interface addresses within an IPMP group are part of the same subnet.
Symmetric/asymmetric configuration	Symmetric—requires a node (switch or server) that also supports LACP	Asymmetric—does not require a peer IPMP at the other side of the connection.
Performance impact	No significant impact.	Can consume up to 5% of the available bandwidth per link; will also consume some CPU cycles.
Link layer requirements	LACP RFC spec specific for hardware Ethernet layer.	L2 ICMP broadcast capable.
Load spreading support	Both inbound and outbound load spreading, full links bandwidth aggregation. Method of load spreading is determined by hashing method selected. See LACP hash mode table for more information.	Outbound load spreading by IPMP kernel driver depending on routing multihoming policy settings. Inbound load spreading depending on IP DNS name servers, name to IP address resolution mechanism. No bandwidth aggregation per network connection session.
Requires a connection to network switch	No, as long as the peer supports LACP. This can be an Oracle Solaris server, for instance. If the peer is a network switch, the switch needs to support LACP.	No, IPMP is peer independent.
DHCP restrictions	No, LACP runs on layer 2; DHCP on layer 3.	Yes, DHCP cannot be used for the IPMP data addresses. It is recommended to use fixed addresses for test addresses.

You can achieve higher levels of availability in a network infrastructure when using network switches that support Virtual Router Redundancy Protocol (VRRP), as described in RFC 5798. VRRP uses the concept of a virtual router in which a number of physical routers are grouped together. A group consists of a master router and a number of backup routers; this setup avoids the router being a single point of failure in a network infrastructure. A similar concept is offered by Cisco's proprietary Hot Standby Router Protocol (HSRP).

Configuring IPMP

The IPMP daemon uses a probing mechanism to detect network path failures. When configured, probing can be performed using test addresses; transitive probing is performed when no test addresses are configured (0.0.0.0/24).

The data IP address in the IPMP group is bound to one of the underlying network interfaces and, as a consequence, all inbound network traffic will come in through that specific network interface. Outbound packets are spread over all IPMP network interface members.

Enable adaptive routing for the multihoming policy when using IPMP to ensure that outbound traffic from Oracle ZFS Storage Appliance is balanced over the active IPMP data addresses, and hence network device ports.

DHCP cannot be used to provide IPMP data addresses. When using DHCP for the IPMP test addresses, it is recommended to set up DHCP to return a fixed address for each interface in the group.

Configuring LACP

When using LACP on Oracle ZFS Storage Appliance, make sure LACP is also set up at the other end of the “wire”—that is, ports of the router/switch to which Oracle ZFS Storage Appliance is connected must be set up for LACP. When using direct connections to a host, make sure the host operating system supports LACP.

As mentioned earlier, there is no guarantee that the aggregated bandwidth of all network ports in the LACP group will be fully utilized. Multiple sessions to/from client(s) are needed to be able to utilize all physical devices in the LACP group. The methodology for spreading the network connections over the network devices in the LACP group at Oracle ZFS Storage Appliance can be tuned by selecting a so-called hash policy that best fits your situation.

Use the same policy setting (L2, L3, L4) at both Oracle ZFS Storage Appliance and the network switch; using the same setting within the whole network segment is recommended. Not doing this can cause problems, especially when using a switch vendor’s multiswitch LACP implementation. The following table shows the available LACP hash policy options for Oracle ZFS Storage Appliance. The options described in the highlighted cells are the recommended settings.

TABLE 3. LACP LOAD BALANCING POLICY OPTIONS

LACP Policy	Single Client	Multiple Clients (< # of Combined Links)	Multiple Clients (> # of Combined Links)
L2 policy (MAC address)	Only one physical network port will be used to talk to a single client NIC.	The number of client MAC addresses will dictate the number of NICs that are able to be used; the remaining NICs will be unused.	The load will be spread out between links but no single client MAC will use more than one NIC at a time.
L3 policy (IP address)	Only one physical network port will be used to talk to each IP address. The client may add IP addresses to its NIC to spread the load if the application supports it.	The number of IP addresses will dictate the number of NICs that are able to be used; the remaining NICs will be unused.	The load will spread out between links but no single client IP address will use more than one NIC at a time.
L4 policy (TCP/UDP port)	Only one physical network port will be used per TCP/UDP connection. The number of application threads (layer 4 connections) will determine load spreading.	The load will spread out between links based on layer 4 connections in use.	The load will spread out between links based on TCP and UDP ports in use. All links can be used. No single layer 4 connection will use more than one NIC.

Analytics can be used to quickly check the utilization of the configured network devices as shown in figure 15. The example shows the network load on an LACP setup with device members e1000g1 and e1000g2. Two NFS clients are used with a single copy process running on each client, copying data from the client to the same NFS volume.



Figure 15. Using analytics to check LACP

Multiple Switch Link Aggregation

When using LACP, the switch to which an Oracle ZFS Storage Appliance system is connected through LACP, becomes a single point of failure (SPOF). Simply creating interconnections between network switches to provide multiple network paths between source and destination can lead to network loops, causing broadcast storms resulting in network instability. You can use Spanning Tree Protocol (STP) or Rapid Spanning Tree Protocol (RSTP) to avoid broadcast storms. STP/RSTP protocols can detect loss of connections and create alternative routes, but it can take up to 60 seconds for connections to be restored.

The IEEE802.1D standard, defining the LACP protocol, has not defined a standard for LACP usage with multiple switches. Vendors have implemented proprietary solutions that span LACP groups over two or more network switches. From the Oracle ZFS Storage Appliance side, such an LACP group appears to be formed by just one switch. As long as the vendor adheres to the IEEE802.1D LACP protocol standard for these types of solutions, Oracle ZFS Storage Appliance will work in such a setup. Such environments offer a more robust and stable network with faster recovery times than STP-based solutions.

Multiswitch link aggregation-type solutions have advantages over the use of IPMP, as they take out the monitoring and recovery functions for failed connections from Oracle ZFS Storage Appliance. The only caveat is that since these solutions are not covered in a standard specification, the Oracle ZFS Storage Appliance requirement is that any issues that might arise in such setups must be reproducible with a single-switch LACP configuration so that Oracle support can work on the issues.

VLANs and VNICs

The VNIC datalink feature was introduced with Oracle ZFS Storage Appliance software version 2013.1 to enable full utilization of all network device ports in an Oracle ZFS Storage Appliance cluster configuration.


While this full utilization could already be done by using the VLAN datalink function, the approach forced the use of a VLAN setup on the network switch, too.

VNICs can optionally be tagged with VLAN IDs and multiple VNICs can share the same VLAN tag on a device object, unlike VLAN datalink objects. For new configurations, use the 802.1q tagged VNICs instead of Oracle ZFS Storage Appliance VLAN datalink objects.

Do not use more than eight VNIC datalink objects per device port.

MTU Recommendations

When the components in the network infrastructure supports the use of larger frame sizes than the default ones (1.5 Kb) for network packets, you can probably improve performance by tuning the size of them upwards. Why would this



be the case? Using larger frame sizes means less packets will be transmitted over the network, so network drivers spend less time on transmitting and receiving packets, resulting in less CPU time used and an increase in network throughput.

For larger size frames to be beneficial, all network components must be able to handle the required frame size. This means that all NICs, network switches, and routers in the data path must support the requested frame size, otherwise packets will be dropped or still be broken up.

A second requirement is that ICMP messages, which are used to signal the maximum supported frame size, are not blocked in the data path between the sender and receiver. For example, a network switch will reply back in an ICMP message to the sender its maximum supported frame size so that the sender can reduce its frame size accordingly. When, for example, ICMP messages are blocked by a firewall, packets will be dropped and no proper connection can be established between sender and receiver.

Where supported, use large MTU frames in the datalink properties setting. Oracle ZFS Storage Appliance supports frame sizes up to 9000. As explained earlier, you have to ensure that all network elements (network switches, HBAs) used in the network infrastructure between Oracle ZFS Storage Appliance and the client support larger MTU settings.

In older network documentation, MTU frame sizes of 9000 are often referred to as jumbo frames. Investigate the maximum value for MTU that can be handled by the network components in your network infrastructure and which value is optimal for your specific application environment.

Using the `ping` command is an easy way to test whether the infrastructure between an Oracle ZFS Storage Appliance system and a server is set up for using larger frame. Unfortunately, the `ping` command options for this testing are not universal. Use the following, based on your operating system, to test for an MTU size of 9000:

For Oracle Solaris, use: `ping -D -s <ip address> 8972`

For Linux, use: `ping -M do -s 8972 <ip address>`

For Microsoft Windows


Server 2003 use ¹: `ping -f -l 8972 <ip address>`

In ESX shell, use: `vmkping -s 8972 <ipaddress>`

Almost all `ping` implementations do not account for the overhead of 28 bytes used, so the packet size to specify for `ping` is the MTU size -28, which is 8972 when testing for an MTU size of 9000. Most `ping` commands support the `-v` flag to provide more diagnostic information; check the manual (`man`) page on your OS platform for your particular situation.

```
~ # ping -f -l 9000 192.168.20.246
Pinging 192.168.20.246 with 9000 bytes of data:
Packet needs to be fragmented but DF set.
Packet needs to be fragmented but DF set.
Packet needs to be fragmented but DF set.
~ # ping -f -l 8972 192.168.20.246
```

¹ Some other versions of Microsoft Windows do take overhead into account.



Pinging 192.168.20.246 with 8972 bytes of data:

```
Reply from 192.168.20.246: bytes=8972 time<1ms TTL=255
```

```
Reply from 192.168.20.246: bytes=8972 time<1ms TTL=255
```

```
Reply from 192.168.20.246: bytes=8972 time<1ms TTL=255
```

When ping indicates problems with the required MTU sizes, a more detailed analysis of the behavior of the network components in the data path may be required. A tool like Wireshark can be used to test for ICMP messages indicating a network component not capable of handling the requested MTU size.

A final recommendation on MTU size: When working with a network infrastructure containing a mix of noncapable and capable large MTU size devices are present, keep those components separate from each other by using either VLANs or a different subnet for them. This separation makes it easier to set up routing tables that keep network traffic using large MTU settings restricted to network devices capable of handling them, thus avoiding the risk of packets still being broken up on the way.

Use the DTrace Analytics feature of Oracle ZFS Storage Appliance to help determine the effect of a changed MTU value and to find an optimal value.

Replication

Much has already been mentioned related to performance. The basic rule is: Keep it simple and do not overcomplicate things just because the technology offers various fancy options. Be aware that network switches may not scale to the maximum aggregated port bandwidth. You might have to add switches to provide for your required aggregated bandwidth requirements before a switch runs out of port connections.

In virtual environments, be careful not to oversubscribe the performance capabilities of the underlying physical resources. Again, the Oracle ZFS Storage Appliance analytics function can be of great help to determine any potential performance issues.

Keep monitoring for changes in application servers' configurations in terms of type and number of applications. An application version update might introduce different I/O network loads.

It might be worthwhile to check for any configured network protocols in your infrastructure that are not really used, like NetWare IPX, AppleTalk, and NetBUI. Limiting the number of protocols used on servers might help take away some unwanted network traffic. It also may make analyzing network traffic easier, because it reduces the need to sift through irrelevant data in network traces.

Hardware Changes

Oracle ZFS Storage Appliance has several hardware upgrade options, including increasing memory capacity, increasing NIC/HBA ports, and adding Readzilla devices. Consult the Oracle ZFS Storage Appliance Customer Service Manual to identify the proper part numbers and upgrade procedures for each of them.

One word of caution when adding NIC/HBA PCIe cards: The Oracle ZFS Storage Appliance Service Manual contains information about which PCIe slots can be used for particular types of PCIe cards. Never move already present PCIe cards to other PCIe slots. During the initial boot of an Oracle ZFS Storage Appliance system, or when PCIe cards have been added, the Oracle ZFS Storage Appliance system goes through a configuration process to identify the PCIe cards used and their locations, and enumerates the device NIC and IB ports accordingly. Moving PCIe cards on an already configured Oracle ZFS Storage Appliance will result in a change of device names and result in a nonconsistent network setup.

Remote Access

Configuring the Oracle ILOM network port into the corporate administrative LAN is highly recommended. A failure in a network configuration or physical component failure might result in losing administrative access to Oracle ZFS Storage Appliance. Via the Oracle ILOM console function, access can be gained to the Oracle ZFS Storage Appliance CLI interface. See "Appendix B: Accessing the Oracle ZFS Storage Appliance Console Through the Oracle ILOM Server" for how to set this up.

Configuration Recommendations for Specific Protocol Settings

The following sections address specific configuration recommendations for both NFS and InfiniBand protocols.

NFS Protocol

MOS note 359515.1 describes the recommended mount options when using Oracle files with NFS. You can find MOS notes by logging in through My Oracle Support at <http://support.oracle.com>.

InfiniBand Protocol

Multiple switches on the same subnet must be connected to each other to prevent multiple "subnet managers" being active on the same subnet.

VNICs and VLAN configuration options are not available for InfiniBand datalink objects. You could achieve an equivalent configuration by using multiple datalink objects in the same InfiniBand port, each with a different partition key (pkey) value. The same partition key cannot be assigned to multiple IPoIB partitions (as new data object instances) over the same ibp device object. You then have to use the subnet manager in the InfiniBand switch to set up the proper access rights for the partitions.

Cluster Networking

Consider the following principles and recommendations when designing a network for an Oracle ZFS Storage Appliance cluster.

A network design for a cluster using an active/passive mode of operation is straightforward. All network resources are configured on the active node, and the resources will fail over to the other node either during a forced role reversal by the user or when triggered by a catastrophic failure event on the current active node.

Designing a network configuration for a cluster operating in active/active mode requires more consideration, based on two fundamental concepts:

- » All configuration information for setting up an Oracle ZFS Storage Appliance system is shared between the two nodes. This requires both nodes to be identical in hardware configuration. Each node keeps a copy of its configuration information, so a change made in a setting on one node is automatically copied to the other node.
- » Network interface objects and storage pools can be only active/owned on one node at a time. For each resource, you can specify which node is the primary owner of the resource. This enables you to specify on which node the resource will be active when both nodes are active.

The cluster resource manager manages resources in a cluster. The resource manager determines what part of a network needs to be brought up on each node, based on how you have configured the node ownership of each of the resources.

The following figure 16 shows a basic network setup in an active/active cluster. Storage pool 1 is active on node A and storage pool 2 is active on node B.

Each node has two interface objects, while their equivalent objects on the other node are in standby mode. This means that on Node A, active data traffic flows through the network device ports igb0 and igb1 and on Node B through the network device ports igb2 and igb3. This way, only two of the four network ports are utilized on each cluster node.

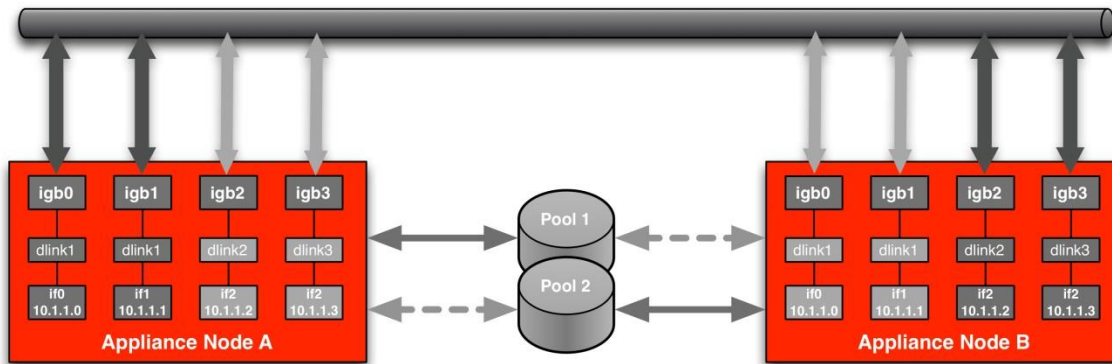


Figure 16. Basic cluster network configuration

To use a network device port on both nodes at the same time, create two VNICs² or VLANs using the same network device port.

By creating an interface object on each VNIC and giving each node ownership of one of the two interface objects, you have created an active interface on each node on the same network device port.

When one node takes over all the resources of the other node, both network interfaces will operate on the same network interface port, each with its own assigned IP address. When using a dedicated administrative interface, you do not want that IP address to fail over. For such situations, use the lock resource option in the cluster configuration setup. This option prevents the locked resource from being taken over by the remaining node during a cluster node failover process.

The VNIC configuration option in a cluster also can be used to create an IPMP group on each node without having to reserve network ports on the other node, as seen in figure 17.

Regarding the usage or creation of IPMP groups using VNICs or VLANs, it does not give any form of redundancy. As of IPMP, it is the only exception of the situation mentioned on page 6 as long as VNICs in the IPMP group are always located on different physical interfaces.

² This feature was introduced with Oracle ZFS Storage Appliance software version 2013.1.

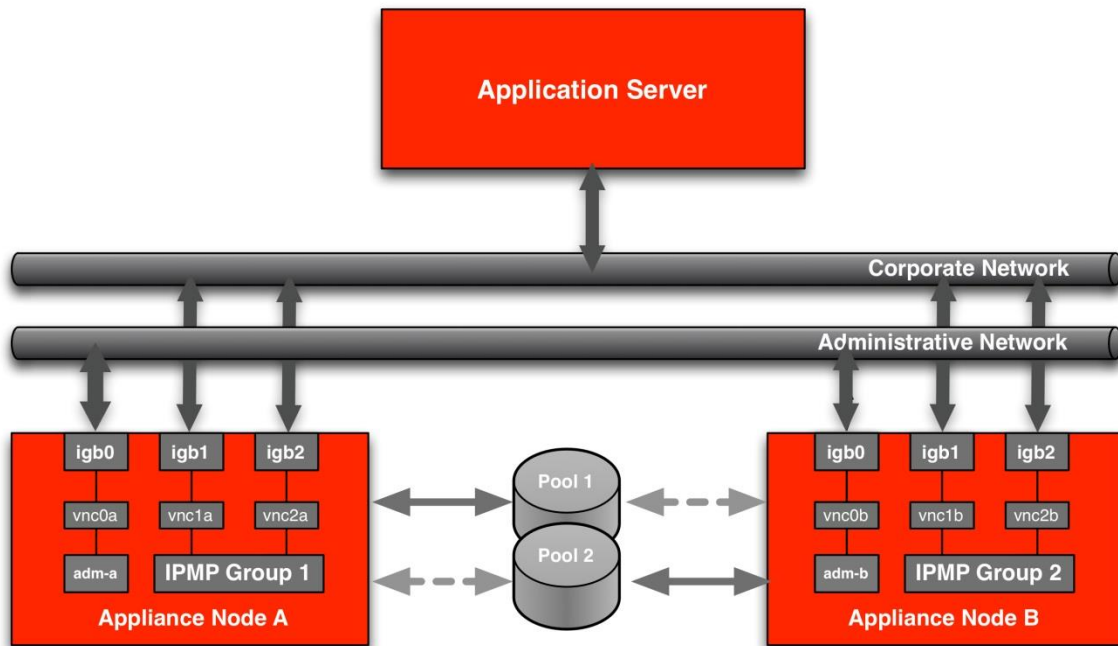


Figure 17. Oracle ZFS Storage Appliance cluster network connections using VNICs

Direct Network Connections

When using one or two application servers in combination with an Oracle ZFS Storage Appliance system, it makes sense to use direct connected 10 GbE connections, saving the cost and complexity of using 10 GbE network switches. When using Oracle Solaris on the servers, the Oracle Solaris IPMP functionality on the server can be used to create redundant connections to an Oracle ZFS Storage Appliance cluster configuration.

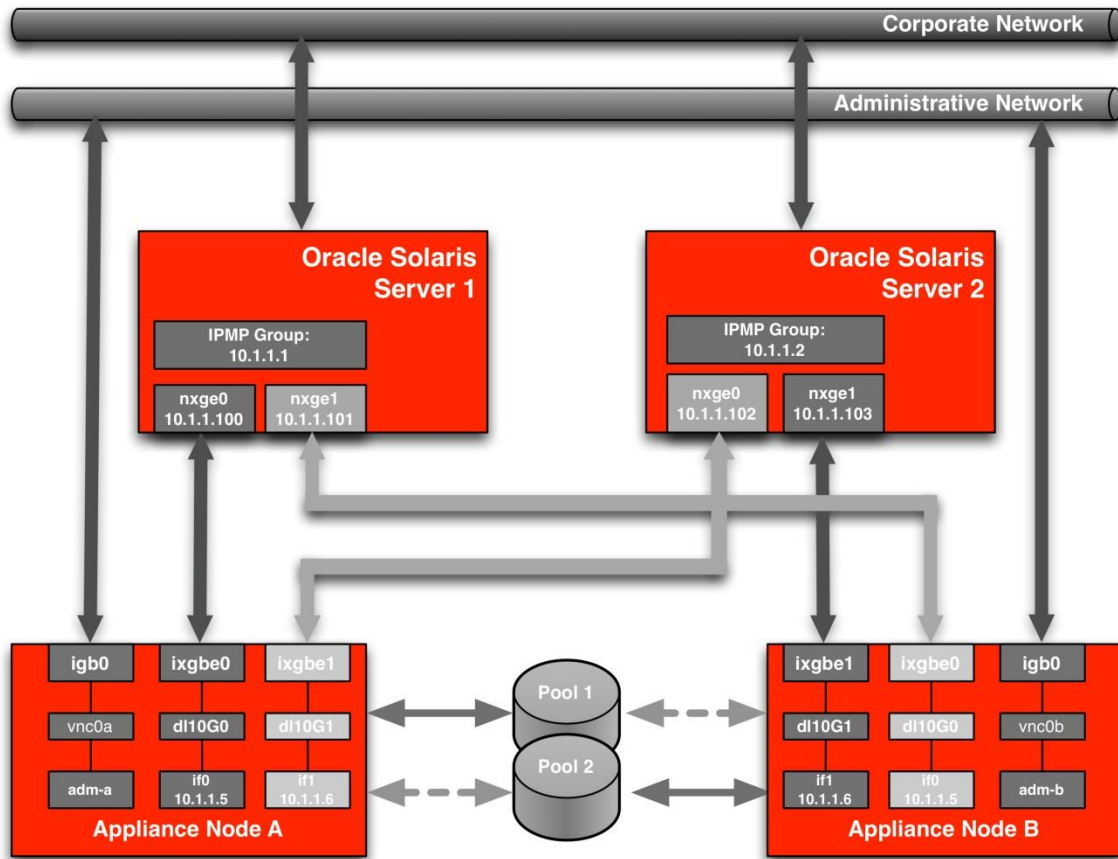


Figure 18. Direct 10 GbE connections

Figure 18 shows two servers, each using storage from a pool dedicated to a specific server. Each server has an IPMP group setup containing an active connection to the Oracle ZFS Storage Appliance node that owns the pool serving that server. When, during a disastrous event, an Oracle ZFS Storage Appliance node fails, its resources are taken over by the remaining node. The IPMP daemon of the server that lost its active node will detect the link failure and switch the communication over to the standby interface to connect to the remaining Oracle ZFS Storage Appliance node. Use test addresses so IPMP can detect a connection to the Oracle ZFS Storage Appliance node going down.

The configuration in the previous figure would not work for servers in a cluster configuration. Typically for a server cluster configuration, both servers must be able to access the same storage devices simultaneously. This can still be done using direct connected network links. Two IPMP groups on each Oracle Real Application Clusters (Oracle RAC) node are needed, each with one active and one passive link.

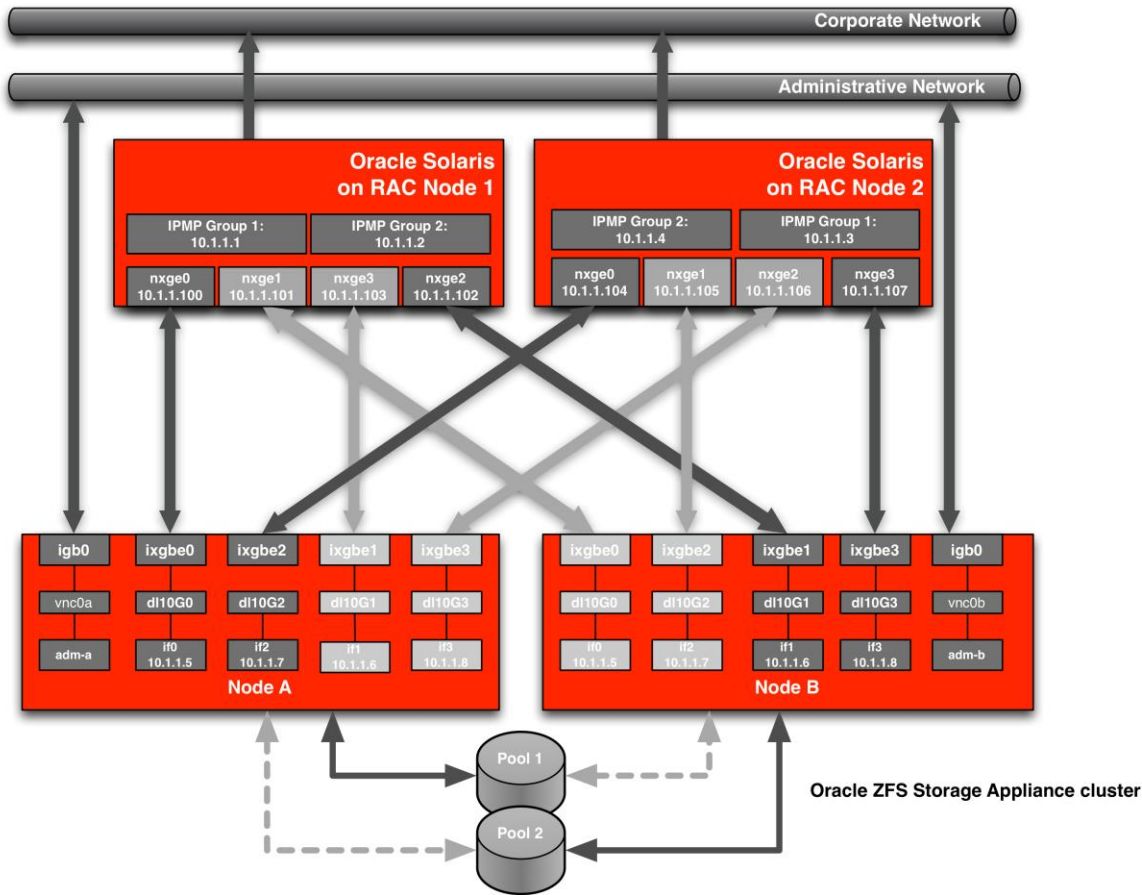


Figure 19. Direct 10 GbE connections with a two-node database Oracle RAC

Cluster Networking Best Practices

When setting up a cluster configuration, start with either two nodes containing factory default settings or use an existing Oracle ZFS Storage Appliance system and add a second Oracle ZFS Storage Appliance system containing factory default settings. After connecting the cluster interconnect cables, power up the new node and issue a join cluster command. Never try to use a join operation with an Oracle ZFS Storage Appliance system that already has been part of a cluster.

As mentioned before, make sure that both Oracle ZFS Storage Appliance products to be used in a cluster have exactly the same hardware configuration. The cluster join process does not enforce or check this. There are several reasons for using the same hardware configuration. Since configuration information is shared between the two nodes, a device network port on each node should identify the same physical HBA port on each node. By ensuring both nodes have the same amount of cache memory and solid-state drive (SSD) options, data services will behave the same on either node. Dealing with two identical nodes also simplifies the administration process.

When setting up the initial network configuration, make sure both nodes have a default route entry in the routing table. This will most likely be the network interfaces used for administrative access. Create one default route entry for each node on the initial node. Create a routing table entry of the type default for each of them. Lock the administrative interface for node A. After finishing setting up the resources in the cluster setup process, fail back the resources to node B. Log in on node B and lock the administrative network interface for node B. By locking

administrative interface(s) on each node, access to the node for administrative tasks is still available when the other node has taken over all services and resources.

The Oracle ZFS Storage Appliance online help manual describes the whole process in detail.



Figure 20. BU: cluster resources

In a cluster configuration, resources can be locked to a specific node (notice the lock symbol in the above figure 20. It is a best practice to lock the interface(s) used for administrative access. The administrative interface(s) will be taken over by the other node after a **TAKEOVER** action and, as a result, access to the other node is lost. It may make it difficult to upload support bundles.

Network and/or link failure events do not trigger an Oracle ZFS Storage Appliance node failover/takeover. Network path redundancy need to be designed using LACP and/or IPMP. See the section "Scalability and Availability" in this document.

Care should be taken when setting up an IPMP group using probing failure detection. As explained earlier, IPMP probing involves the discovery of up to five peer systems that the IPMP daemon uses to detect path failures. It is important that there are actual systems present besides the cluster peer node for this. If the cluster peer node is the only network target interface discovered, connectivity to the IPMP group may be lost during a failover.

It is a best practice to configure access to Oracle ILOM for each node in the cluster. If, for whatever reason, errors are made in the network configuration settings that might prevent a node from joining a cluster or results in an admin access lockout on one of the nodes, Oracle ILOM access always provides the access to the Oracle ZFS Storage Appliance CLI. You can use the CLI commands to restore normal administrative access. See "Appendix B: Accessing the Oracle ZFS Storage Appliance Console Through the Oracle ILOM Server."

Pay attention in setting up routing configuration on each node. For each active interface on each node, routing information must be available for IP traffic to reach its destination. The initial default route entry created by Oracle ZFS Storage Appliance might not always be the optimal setting. Often the first network port used during the setup procedure will be used as administrative access port by most users and a related entry in the routing table will be added with type default. In some cases this might result in data traffic being routed through the administrative interface.

A best practice is to use static configured IP addresses in cluster configurations. Do not use dynamic IP addressing through DHCP for cluster IP addresses. Since IP addresses can fail over, there is no one-to-one relationship between an IP address and a MAC address, so you cannot use DHCP to provide fixed addresses either.

Troubleshooting Network Configuration Problems

Networking problems exhibit themselves in many forms, ranging from network connectivity to network congestion problems. So before starting to change various configuration settings, it is important to understand the nature of the problem and have an indication of where it originates. Oracle ZFS Storage Appliance offers status and diagnostic information on various levels, and its logs are one of the first resources to consult.

Other sources of information are event logs and system messages on the client side. Check for any network-related errors or warning messages that might give a clue about the issue you are investigating.

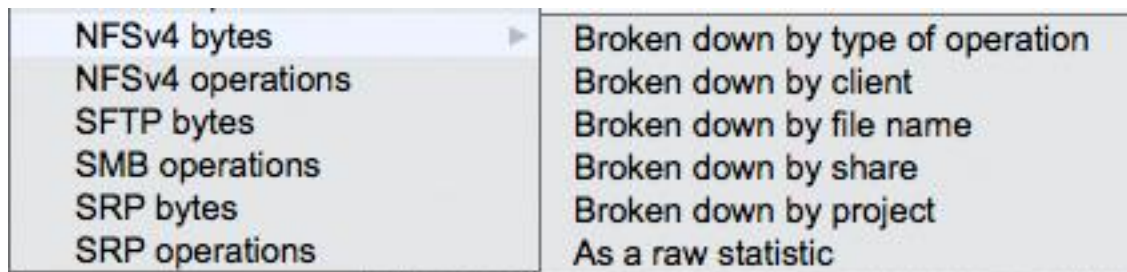


Alerts 19 Total				
ALERTS	FAULTS	SYSTEM	AUDIT	PHONE HOME
TIME	EVENT ID	DESCRIPTION		TYPE
2014-2-11 13:49:11	9b73aef0-a2c2-e435-a886-dbe801559767	The service processor has stopped responding to requests.		Minor alert
2014-2-5 15:19:42	e924f7c6-0bea-cb84-f798-97df373015a5	Full IP connectivity via interface e1000g0 has been established.		Minor alert
2014-2-5 15:19:41	722b6046-5eec-6e4d-ctb8-81dbfe15f666	Network connectivity via dataink e1000g0 has been established.		Minor alert

Figure 21. Accessing Oracle ZFS Storage Appliance logs

From the servers/clients side, check simple connectivity problems using the well-known `ping` and `traceroute` utilities. They can reveal connectivity issues, DNS, and/or routing problems. The `ping` utility also can help investigation of any latency-related performance problems.

Performance problems often turn out to be more difficult to deal with. Start walking through the whole connectivity chain to see if there is an element that could act as a bottleneck in the I/O traffic between the data repository and the client's application. Check for any element that could be "oversubscribed" in terms of network bandwidth capacity. The Oracle ZFS Storage Appliance Analytics option is a good tool to use. Start investigating the main components in the system, like load patterns per share, per client, or per network connections. Is a client or client's application behaving abnormally? Are all network ports fully utilized? Are storage pools reaching 80 percent to 90 percent usage levels?



NFSv4 bytes	Broken down by type of operation
NFSv4 operations	Broken down by client
SFTP bytes	Broken down by file name
SMB operations	Broken down by share
SRP bytes	Broken down by project
SRP operations	As a raw statistic

Figure 22. Analytics operations broken down by options example

Information to answer any of these questions can be quickly found using analytics.

For detailed performance analysis, Oracle's performance tools like `Vdbench`, a feature of Oracle ZFS Storage Appliance, and Storage Workload Analysis Tool (SWAT) can be used as load generation and analysis tools in combination with the appliance analytics. `Vdbench` can be used to set up specific workload definitions to be used on specific data shares on Oracle ZFS Storage Appliance to analyze system behavior.

The following flowchart provides some common issues and resources to consult related to these issues. Oracle's support website has numerous documents related to specific types of network problems, and describes how to diagnose and resolve them.

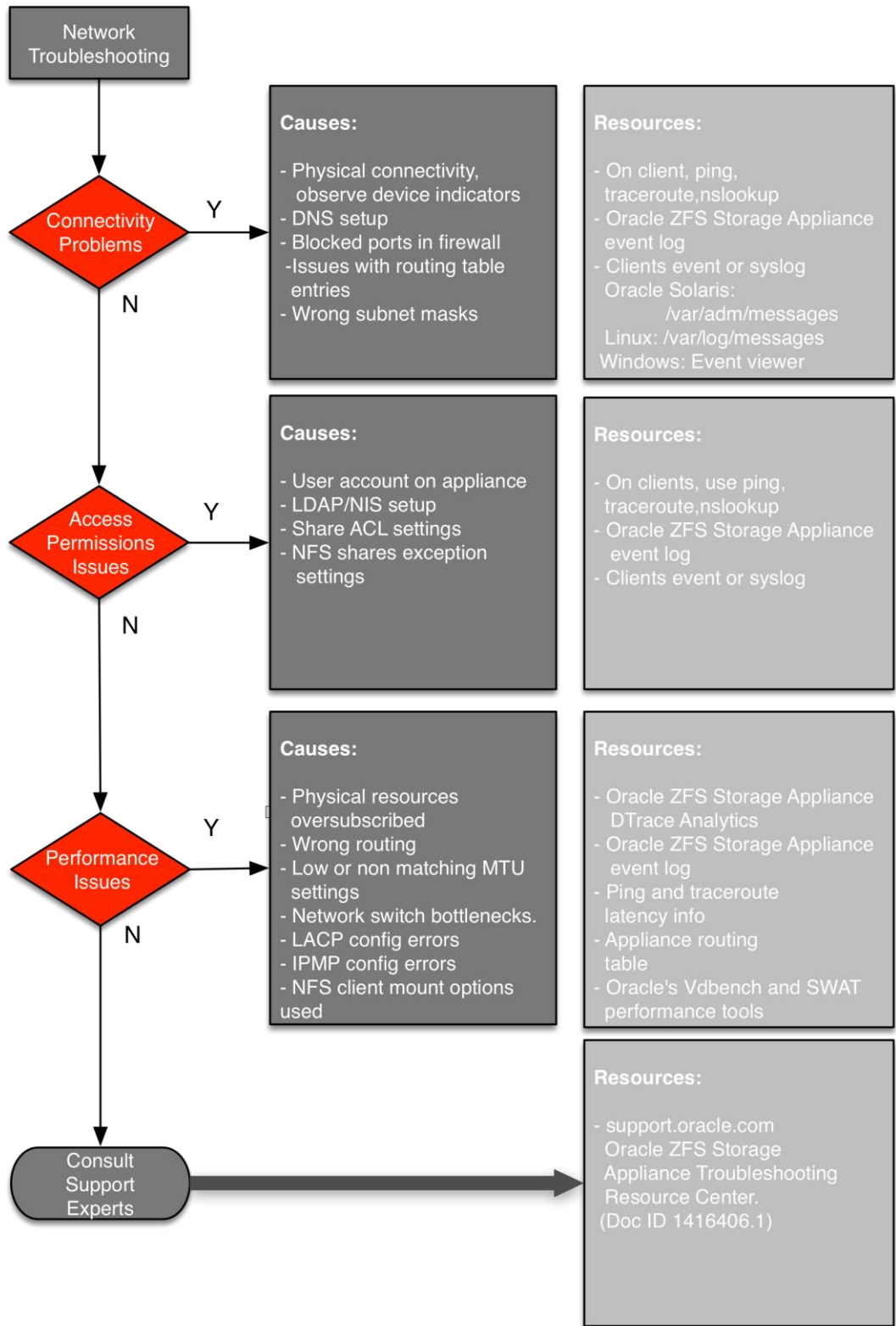


Figure 23. Oracle ZFS Storage Appliance troubleshooting flowchart

Appendix A: Services and Associated IP Ports

The following table lists the network ports used by Oracle ZFS Storage Appliance. If a firewall is present between the clients and Oracle ZFS Storage Appliance, make sure the ports for services used by the client(s) are unblocked in the firewall.

TABLE 4. ORACLE ZFS STORAGE APPLIANCE SERVICES AND ASSOCIATED IP PORT NUMBERS

Service Name	Description	Ports Used
FTP	Filesystem access through the FTP protocol	21
SSH	SSH for CLI access	22
DNS	Domain name service client	53
HTTP	Filesystem access through HTTP protocol	80
Kerberos	Kerberos Authentication Kerberos Change and Set Password (SET_CHANGE) Kerberos Change and Set Password (RPCSEC_GSS)	88
NFS	Filesystem access through NFSv3 and NFSv4 protocols	111 and 2049
SMB	Filesystem access through SMB protocol	137 NetBIOS Name Service 138 NetBIOS Datagram 139 SMB over NetBIOS 445 SMB over TCP
BUI	Browser user interface	215
Remote Replication	Remote Replication	216, 217
SFTP	Filesystem access through SFTP protocol	218
LDAP	Authentication of users and groups from an LDAP directory	389
Phone home	Product registration and support configuration	443
iSCSI	LUN access through the iSCSI protocol	3260 and 3205
NDMP	NDMP host service	10000

The Oracle ZFS Storage Appliance online help manual gives more details on the setup of firewall rules for use of Oracle ZFS Storage Appliance in such an environment. Find this information under Configuration, Services.

Appendix B: Accessing the Oracle ZFS Storage Appliance Console Through the Oracle ILOM Server

When an Oracle ZFS Storage Appliance system cannot be reached through the network, you can gain access to the Oracle ZFS Storage Appliance console prompt through the Oracle Integrated Lights Out Management (Oracle ILOM) server.

The Oracle ILOM server has a separate IP interface that is reachable through a WebGUI or `ssh` using the Oracle ILOM's IP address. The Oracle ILOM connection to the user's network is made through the network port labeled NET MGT.



Figure 24. Network management port

If the Oracle ILOM's IP address is unknown or not yet configured, the Oracle ILOM server can be reached through the serial port connection, labeled SER MGT. This enables an administrator to gain access to the Oracle ZFS Storage Appliance console when attempts to gain access through the IP network fail.

Setting Up a Serial Connection to the Oracle ILOM Server

To connect to Oracle ILOM using a serial connection, complete the following steps:

1. Attach a serial cable from a terminal, a serial terminal concentrator, or a PC running terminal emulation software to the Oracle ZFS Storage Appliance SER MGT port. The cable should be a length of 4.5 m or less.
2. Verify that your terminal or laptop is operational.
3. Configure the terminal device or the terminal emulation software to use the following settings:
 - » Set 8N1: eight data bits, no parity, one stop bit
 - » Set 9600 baud
 - » Disable software flow control (XON/XOFF)
 - » Disable hardware control
4. Verify that power is supplied to either PSU. If there is power applied to either PSU, then Oracle ILOM will be functional regardless of the power state of compute nodes.
5. Press Enter on the terminal device. A connection between the terminal device and Oracle ILOM is established. The Oracle ILOM login prompt is displayed.
6. Log in to the CLI using the default user name and the password (root and changeme). The Oracle ILOM default command prompt is displayed.
7. If Oracle ILOM has not been set up for network access, follow the procedure as described in the sections "Configuring a Static IP Address" or "Configuring Oracle ILOM to use DHCP" in the documentation that came with the Oracle ZFS Storage Appliance.

Note: Plan to allocate an IPv4 address for the Oracle ILOM network interface. IPv6 is not supported for the Oracle ILOM network interface.

Accessing the Oracle ZFS Storage Appliance Console

When using `ssh` or a serial connection to connect to Oracle ILOM, use the following command to access the Oracle ZFS Storage Appliance console:

```
> cd /SP/console  
> start
```

To return from the CLI shell to the Oracle ILOM prompt, press the following two keyboard keys: <Escape> and (. Do not forget to log out of the Oracle ZFS Storage Appliance console shell before terminating the Oracle ILOM console session to prevent unauthorized access to Oracle ZFS Storage Appliance through the Oracle ILOM console function.

When using the WebGUI, use the remote management option to start a redirected console session.

Serial RJ45 Signal Definitions



Figure 25. Serial management port connector

The following table shows the serial RS232 signals used and how to connect them to a serial port of a terminal or a terminal concentrator.

TABLE 5. SERIAL MANAGEMENT RJ45 CONNECTION SIGNALS

Pin	Signal Description	Pin	Signal Description
1	Request to Send	5	Ground
2	Data Terminal Ready	6	Receive Data
3	Transmit Data	7	Data Set Ready
4	Ground	8	Clear to Send

Appendix C: References

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

- » Oracle ZFS Storage Appliance Documentation Library, including Installation, Analytics, Customer Service, and Administration guides:
<http://www.oracle.com/technetwork/documentation/oracle-unified-ss-193371.html>
- » Oracle ZFS Storage Appliance Administration Guide is also available through Oracle ZFS Storage Appliance help context.
The Help function in Oracle ZFS Storage Appliance can be accessed through the browser user interface.
- » Oracle ZFS Storage Appliance Product Information
<http://www.oracle.com/us/products/servers-storage/storage/nas/overview/index.html>
- » Oracle ZFS Storage Appliance White Papers and Subject-Specific Resources
<http://www.oracle.com/technetwork/server-storage/sun-unified-storage/documentation/index.html>
Including: " Understanding the Use of Fibre Channel in the Oracle ZFS Storage Appliance"
<http://www.oracle.com/technetwork/server-storage/sun-unified-storage/documentation/o12-019-fclun-7000-rs-1559284.pdf>
- » Wikipedia's Virtual Router Redundancy Protocol Page
http://en.wikipedia.org/wiki/Virtual_Router_Redundancy_Protocol
- » RFC5789: Virtual Router Redundancy Protocol (VRRP) Version 3 for IPv4 and IPv6
<http://tools.ietf.org/html/rfc5798>
- » IPv6 Administration Guide
<http://docs.oracle.com/cd/E19683-01/817-0573/index.html>







Oracle Corporation, World Headquarters

500 Oracle Parkway
Redwood Shores, CA 94065, USA

Worldwide Inquiries

Phone: +1.650.506.7000
Fax: +1.650.506.7200

CONNECT WITH US

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

Integrated Cloud Applications & Platform Services

Copyright © 2014, 2018, Oracle and/or its affiliates. All rights reserved. This document is provided for information purposes only, and the contents hereof are subject to change without notice. This document is not warranted to be error-free, nor subject to any other warranties or conditions, whether expressed orally or implied in law, including implied warranties and conditions of merchantability or fitness for a particular purpose. We specifically disclaim any liability with respect to this document, and no contractual obligations are formed either directly or indirectly by this document. This document may not be reproduced or transmitted in any form or by any means, electronic or mechanical, for any purpose, without our prior written permission.

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

Intel and Intel Xeon are trademarks or registered trademarks of Intel Corporation. All SPARC trademarks are used under license and are trademarks or registered trademarks of SPARC International, Inc. AMD, Opteron, the AMD logo, and the AMD Opteron logo are trademarks or registered trademarks of Advanced Micro Devices. UNIX is a registered trademark of The Open Group. 0116

Networking Best Practices with Oracle ZFS Storage Appliance
May 2018 Version 3.0
Author: Peter Brouwer, Ulrich Conrad