Exadata X8M Backup with ZFS Storage Appliance

Oracle Exadata Database Machine X8M data protection best practices with Oracle ZFS Storage Appliance ZS7-2

WHITE PAPER / JANUARY 9, 2020
DISCLAIMER

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

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

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

Disclaimer........................................................................................................................................... 2

Executive Overview ................................................................................................................................. 5

Introduction ........................................................................................................................................... 8

Configuring the Network ....................................................................................................................... 9

- Option 1: Sharing Exadata X8M Switches ......................................................................................... 9
- Option 2: Dedicated Top of Rack (ToR) Switches............................................................................... 10
- Switch Configuration Best Practices ............................................................................................... 11
- vPC Configuration Steps .................................................................................................................... 12
- Multiple Backup Networks ............................................................................................................... 15

Configuring the Oracle ZFS Storage Appliance .................................................................................. 18

- Choosing a Controller ...................................................................................................................... 18
- Choosing the Correct Disk Shelves .................................................................................................... 18
- Choosing a Storage Profile ................................................................................................................ 19
- Configuring the Storage Pools .......................................................................................................... 20
- Configuring the Projects and Shares ............................................................................................... 21
- Configuring the NFS Server ............................................................................................................. 24
- Configuring IO Cards ......................................................................................................................... 24
- Configuring Data Path Network ........................................................................................................ 25
- Choosing Direct NFS Client ............................................................................................................. 28

Configuring the Oracle Exadata Database Machine ........................................................................... 30
Configuring Backup Network.................................................................30
Configure Backup Shares ........................................................................31
Configuring Direct NFS Client .................................................................32
Configuring Oracle RMAN Backup Services ...........................................33
Preparing Oracle Database for Backup..................................................33
Oracle RMAN Configuration...................................................................34
Conclusion .........................................................................................37
EXECUTIVE OVERVIEW

Protecting the mission-critical data that resides on Oracle Exadata Database Machine (Oracle Exadata) is a top priority. The Oracle ZFS Storage Appliance is ideally suited for this task due to superior performance, enhanced reliability, extreme network bandwidth, powerful features, simplified management, and cost-efficient configurations.

The Oracle Exadata X8M product line introduces a new and improved networking infrastructure built on high-speed, low latency 100Gb Ethernet optimized for RDMA over Converged Ethernet (RoCE). This paper focuses on network attachment and configuration best practices when using an Oracle ZS7 Storage Appliance to provide data protection services for a RoCE-based Oracle Exadata Database Machine. More comprehensive general-purpose guidelines for using ZFS Storage in an Exadata environment are provided in My Oracle Support document 2087231.1 “Guidelines when Using ZFS Storage in an Exadata Environment.”

The backup network is comprised of bondeth1 (25Gb) interfaces on the Exadata compute nodes and data path interfaces on the ZS7-2 controllers. There are two network attachment options for the backup network in this solution.

1. Directly connect the backup network onto the internal Exadata 100Gb Ethernet leaf switches.
2. Attach the backup network using dedicated 100Gb Ethernet switches

These networking options are simple to implement and manage. Either option can be used to provide a reliable data protection solution with a high level of redundancy and excellent performance.

Directly connecting to the leaf switches will eliminate the need to acquire and manage dedicated top of rack switches. This option will provide superior ROI and tighter integration. The switches are managed as part of the Oracle Exadata HW stack and software levels are upgraded during the patching process. This option requires two 25Gb Ethernet cards in each ZS7-2 controller and specific port configurations which are detailed in the network configuration section.

Using dedicated top of rack switches has the benefits of isolating the backup network and providing a greater level of control. Dedicated switches can support higher levels of backup throughput for large configurations due to the ability to use jumbo frames. Top of rack switches are not managed as part of the Oracle Exadata HW stack and will require occasional software upgrades. This option provides flexibility to use 10Gb, 25Gb or 40Gb interfaces on the ZFS Storage Appliance.

The following graph shows maximum sustainable backup and restore performance that a ZS7-2 with 6 high-capacity disk shelves can support when the backup network is directly connected to the internal
Exadata 100Gb leaf switches. Physical throughput rates were measured at the network level for backup and restore workloads between Oracle Exadata and Oracle ZFS Storage Appliance systems.

![Graph showing maximum sustainable backup and restore throughput of Oracle ZFS Storage ZS7-2 when connected to Oracle Exadata X8M internal leaf switches.](image)

Figure 1. Maximum sustainable backup and restore throughput of Oracle ZFS Storage ZS7-2 when connected to Oracle Exadata X8M internal leaf switches

Oracle ZFS Storage Appliance ZS7-2 configurations with more than six high-capacity disk shelves can support higher restore rates. When the backup network is deployed using dedicated top of rack switches the Oracle ZFS Storage Appliance can support higher backup rates with the use of jumbo frames to improve network efficiency with Oracle Recovery Manager (RMAN) workloads and reduce CPU overhead. For general-purpose performance characterization of RMAN workloads across a complete range of ZS7-2 configurations, please reference My Oracle Support document 2087231.1 “Guidelines when Using ZFS Storage in an Exadata Environment.”

These are complete, real-world results using Oracle Database 19c and an online transactional processing (OLTP) database that was populated with sample customer data in a sales order-entry schema. Advanced Row Compression, an Oracle Database feature, was used at the database level to align with best-practice recommendations for customers that are running OLTP workloads. These throughput rates were not obtained using a database or input/output (I/O) generator test tool, which can be misleading. Also, they were not projected based on low-level system benchmarks. The backup and restore performance data collected for this document was measured using level 0 backup and restore operations for an otherwise idle Oracle Database. When accounting for database-level compression or incremental
backup strategies, effective backup rates that are much higher than the physical rates documented here are routinely observed.

The rates shown in figure 1 represent the maximum database backup/restore throughput that an Oracle ZS7-2 with one to six high-capacity disk shelves is capable of sustaining. Achieving these rates is contingent on having an Oracle Exadata X8M configuration large enough and with enough available resources to support this performance. Small configurations and configurations with a heavy concurrent workload will have difficulty achieving maximum throughput rates that the storage can support.

High performance is an important consideration when choosing a solution to protect Oracle Exadata. The following technologies make it possible for Oracle ZFS Storage Appliance systems to achieve these backup and restore rates:

» **Oracle Recovery Manager Integration** – Oracle Recovery Manager (Oracle RMAN) is a highly parallelized application that resides within Oracle Database and optimizes backup and recovery operations. Oracle ZFS Storage Appliance systems are designed to integrate with Oracle RMAN by utilizing up to 3,000 concurrent threads that distribute I/O across many channels spread across multiple controllers. This improves performance dramatically with sequential large block streaming I/O workloads that are typical for most backup and restore situations.

» **Oracle Database’s Direct NFS Client feature** – The optimized Direct NFS Client feature is an aggressive implementation that allocates individual TCP connections for each Oracle Database process, in addition to reducing CPU and memory overhead, by bypassing the operating system and writing buffers directly to user space.

» **1 MB Record Sizes** – Oracle ZFS Storage Appliance systems enable larger 1 MB record sizes. This reduces the number of input/output operations per second (IOPS) that are required to disk, preserves the I/O size from Oracle RMAN buffers to storage, and improves performance of large-block sequential operations.

» **Hybrid Storage Pools** – Oracle ZFS Storage Appliance systems have an innovative Hybrid Storage Pool (HSP) architecture that utilizes dynamic storage tiers across memory, flash, and disk. The effective use of dynamic random-access memory (DRAM) and enterprise-class software specifically engineered for multilevel storage is a key component that facilitates the superior performance of Oracle ZFS Storage Appliance systems.

The performance benefits of the Oracle ZFS Storage Appliance are well documented and independently verified. Oracle periodically publishes Storage Performance Council’s SPC-1 and SPC-2 benchmark results, as well as Standard Performance Evaluation Corporation’s SPEC SFS benchmark results to demonstrate performance results for the Oracle ZFS Storage Appliance. Combine this with the powerful features, simplified management, and Oracle-on-Oracle integrations, and it is easy to see why these systems are a compelling solution for protecting mission-critical data on Oracle Exadata.
INTRODUCTION

Database, system, and storage administrators are faced with a common dilemma when it comes to backup and recovery of Oracle Database instances—how to back up more data, more often, in less time, and within the same budget. Moreover, practical challenges associated with real-world outages mandate that data protection systems be simple and reliable to ensure smooth operation under compromised conditions. The Oracle ZFS Storage Appliance helps administrators meet these challenges by providing cost-effective and high-bandwidth storage systems that combine the simplicity of the NFS protocol with ZFS-enhanced disk reliability. Through Oracle ZFS Storage Appliance technology, administrators can reduce the capital and operational costs associated with data protection while maintaining strict service-level agreements with end customers.

Oracle ZFS Storage Appliance systems are easy-to-deploy unified storage systems uniquely suited for protecting data contained in Oracle Exadata. With native 25 GbE and 40 GbE connectivity, they are an ideal match for Oracle Exadata X8M. These high-bandwidth interconnects reduce backup and recovery time, as well as reduce backup application costs and support fees, compared to traditional NAS storage systems. With support for both traditional tiered and incrementally updated backup strategies, Oracle ZFS Storage Appliance systems deliver enhanced storage efficiency that can further reduce recovery time and simplify system administration.

Deploying Oracle ZFS Storage Appliance systems for protecting the mission-critical data that resides on Oracle Database on Oracle Exadata requires that backup window and recovery time objectives (RTOs) be met to ensure timely recovery in the event of a disaster.

This document focuses on networking guidelines for setting up an Oracle ZFS Storage Appliance ZS7-2 for optimal backup and recovery of Oracle Databases running on an Oracle Exadata X8M Database Machine. Selecting the right backup strategy, understanding encryption options and when to deploy a deduplication solution are all important considerations. Guidelines to help understand and make these deployment decisions are provided in My Oracle Support document 2087231.1 “Guidelines when Using ZFS Storage in an Exadata Environment.” This solution brief illustrates the networking guidelines by utilizing a reference architecture with a standard RMAN backup strategy with no encryption or deduplication requirements.

This paper addresses the following topics:
» Configuring the network
» Configuring the Oracle ZFS Storage Appliance
» Configuring the Oracle Exadata X8M
CONFIGURING THE NETWORK

This section provides two alternatives for deploying the backup network. It documents specific cabling guidelines and networking best practices to support a high performance and highly available data protection solution. The backup network is comprised of bondeth1 (25Gb) interfaces on the Exadata compute nodes and data path interfaces on the ZS7-2 controllers.

**Option 1: Sharing Exadata X8M Switches**

In this network configuration option, the backup network shares the Oracle Exadata X8M internal leaf switches. These are Cisco 9336C-FX2-OR 100GbE switches. The backup network consumes a handful of available ports using TCP/IP while most ports are dedicated for the Exadata internal network using RDMA over Converged Ethernet (RoCE).

Directly connecting to the leaf switches eliminates the need to acquire and manage dedicated top of rack switches for the backup network. This option will provide superior ROI and tighter integration. The switches are managed as part of the Oracle Exadata HW stack and software levels are upgraded during the patching process.

The backup network will always use 25Gb Ethernet cards in the database nodes. This option also requires the use of 25Gb Ethernet in the storage controllers with two cards installed in each ZS7-2 controller. The configuring IO cards section of this document provides more details on the installation and setup of network cards on the storage.

There are a limited number of ports available on the Exadata X8M switches. Ports 4 through 33 are reserved for the RoCE internal network. These ports should be reserved even in a partial rack configuration to enable the option of future expansion. 4x 25Gb QSFP28 to SFP28 breakout cables are used to connect the backup network.

Figure 2 depicts the network cabling diagram when the backup network is configured on the internal Exadata X8M switches.

![Backup Network Sharing Exadata X8M Switches](image)
The following table documents the port mapping details of this network solution.

**TABLE 1. BACKUP NETWORK PORT MAPPING SHARED X8M SWITCHES**

<table>
<thead>
<tr>
<th>Group</th>
<th>Switch 1</th>
<th>Switch 2</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eth1/1/1</td>
<td>Controller 1 PCIe 4 port 1</td>
<td>Controller 1 PCIe 4 port 2</td>
<td>25G Link via 4x25G breakout</td>
</tr>
<tr>
<td>Eth1/1/2</td>
<td>Controller 1 PCIe 8 port 1</td>
<td>Controller 1 PCIe 8 port 2</td>
<td>25G Link via 4x25G breakout</td>
</tr>
<tr>
<td>Eth1/1/3</td>
<td>Controller 2 PCIe 4 port 1</td>
<td>Controller 2 PCIe 4 port 2</td>
<td>25G Link via 4x25G breakout</td>
</tr>
<tr>
<td>Eth1/1/4</td>
<td>Controller 2 PCIe 8 port 1</td>
<td>Controller 2 PCIe 8 port 2</td>
<td>25G Link via 4x25G breakout</td>
</tr>
<tr>
<td>Eth1/2/1</td>
<td>db01 backup port 1</td>
<td>db01 backup port 2</td>
<td>25G Link via 4x25G breakout</td>
</tr>
<tr>
<td>Eth1/2/2</td>
<td>db02 backup port 1</td>
<td>db02 backup port 2</td>
<td>25G Link via 4x25G breakout</td>
</tr>
<tr>
<td>Eth1/2/3</td>
<td>db03 backup port 1</td>
<td>db03 backup port 2</td>
<td>25G Link via 4x25G breakout</td>
</tr>
<tr>
<td>Eth1/2/4</td>
<td>db04 backup port 1</td>
<td>db04 backup port 2</td>
<td>25G Link via 4x25G breakout</td>
</tr>
<tr>
<td>Eth1/3/1</td>
<td>db05 backup port 1</td>
<td>db05 backup port 2</td>
<td>25G Link via 4x25G breakout</td>
</tr>
<tr>
<td>Eth1/3/2</td>
<td>db06 backup port 1</td>
<td>db06 backup port 2</td>
<td>25G Link via 4x25G breakout</td>
</tr>
<tr>
<td>Eth1/3/3</td>
<td>db07 backup port 1</td>
<td>db07 backup port 2</td>
<td>25G Link via 4x25G breakout</td>
</tr>
<tr>
<td>Eth1/3/4</td>
<td>db08 backup port 1</td>
<td>db08 backup port 2</td>
<td>25G Link via 4x25G breakout</td>
</tr>
<tr>
<td>Eth1/34</td>
<td>Peer link (1 of 2)</td>
<td>Peer link (1 of 2)</td>
<td>100G Link</td>
</tr>
<tr>
<td>Eth1/35</td>
<td>Peer link (2 of 2)</td>
<td>Peer link (2 of 2)</td>
<td>100G Link</td>
</tr>
<tr>
<td>Eth1/36</td>
<td>Uplink</td>
<td>Uplink</td>
<td>100G Link</td>
</tr>
<tr>
<td>mgmt0</td>
<td>Keepalive</td>
<td>Keepalive</td>
<td>1G copper</td>
</tr>
</tbody>
</table>

**Option 2: Dedicated Top of Rack (ToR) Switches**

Placing the backup network on dedicated switches installed in the top of the rack has the benefits of isolating the backup network from other workloads and providing a greater level of control. Dedicated switches can support higher levels of backup throughput for large configurations due to the ability to use jumbo frames. Top of rack switches are not managed as part of the Oracle Exadata HW stack and will require occasional software upgrades.

The backup network will always use 25Gb Ethernet cards in the database nodes. This option supports the use of 25Gb or 40Gb Ethernet in the storage controllers. The recommended configuration is two cards installed in each ZS7-2 controller.

25Gb would use QSFP28 to 4x SFP28 breakout cables to form four 25Gb links on a 100Gb port. 40Gb would use QSFP28 to QSFP28 standard cables. There are plenty of switch ports available to accommodate this solution. The configuring IO cards section of this document provides more details on the installation and setup of network cards on the storage.

Figure 3 depicts the network cabling diagram when the backup network is configured on top of rack switches. This example uses 40Gb Ethernet cards in the storage controllers. 25Gb Ethernet cards with breakout cables can also be used.
Switch Configuration Best Practices

UPLINK

The 100gb uplink port is not required but is an optional recommendation to provide network services for the backup network. It can also be leveraged for WAN-based solutions such as archival cloud backup or storage replication. The network uplink should be placed on port 36.

LACP CONFIGURATION

For optimal performance and availability, it is recommended to use a Link Aggregation Control Protocol (LACP) group that spans both switches.

A port-channel configuration on the switch provides the basis to aggregate multiple interfaces together and then perform load balancing. However, it is limited to ports on the same switch. For optimal availability it is recommended to have connections to redundant switches.

The Cisco switches have a feature called Virtual Port-Channel (vPC) that provides the ability to configure a port-channel that spans multiple switches. vPC requires two 100 Gb peer links that are configured on ports 34 and 35. vPC also requires a keepalive connection which should be configured on the management port (mgmt0). vPC is used to configure LACP groups that span both switches for the four ports on each ZS7-2 controller and the bondeth1 backup network ports on the database nodes.
The following table shows the vPC configuration for a backup network deployed on shared Exadata X8M switches in a full rack configuration.

TABLE 2. VIRTUAL PORT CHANNEL (VPC) CONFIGURATION (SHARED X8M SWITCHES – FULL RACK)

<table>
<thead>
<tr>
<th>Port-Channel</th>
<th>Switch 1</th>
<th>Switch 2</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Po1(SU)</td>
<td>Eth1/34(P), Eth1/35(P)</td>
<td>Eth1/34(P), Eth1/35(P)</td>
<td>vPC peer link (LACP)</td>
</tr>
<tr>
<td>Po10(SU)</td>
<td>Eth1/1/1(P), Eth1/1/2(P)</td>
<td>Eth1/1/1(P), Eth1/1/2(P)</td>
<td>ZS7-2 Controller 1 (LACP)</td>
</tr>
<tr>
<td>Po11(SU)</td>
<td>Eth1/1/3(P), Eth1/1/4(P)</td>
<td>Eth1/1/3(P), Eth1/1/4(P)</td>
<td>ZS7-2 Controller 2 (LACP)</td>
</tr>
<tr>
<td>Po12(SU)</td>
<td>Eth1/2/1(P)</td>
<td>Eth1/2/1(P)</td>
<td>db01 backup network (LACP)</td>
</tr>
<tr>
<td>Po13(SU)</td>
<td>Eth1/2/2(P)</td>
<td>Eth1/2/2(P)</td>
<td>db02 backup network (LACP)</td>
</tr>
<tr>
<td>Po14(SU)</td>
<td>Eth1/2/3(P)</td>
<td>Eth1/2/3(P)</td>
<td>db03 backup network (LACP)</td>
</tr>
<tr>
<td>Po15(SU)</td>
<td>Eth1/2/4(P)</td>
<td>Eth1/2/4(P)</td>
<td>db04 backup network (LACP)</td>
</tr>
<tr>
<td>Po16(SU)</td>
<td>Eth1/3/1(P)</td>
<td>Eth1/3/1(P)</td>
<td>db05 backup network (LACP)</td>
</tr>
<tr>
<td>Po17(SU)</td>
<td>Eth1/3/2(P)</td>
<td>Eth1/3/2(P)</td>
<td>db06 backup network (LACP)</td>
</tr>
<tr>
<td>Po18(SU)</td>
<td>Eth1/3/3(P)</td>
<td>Eth1/3/3(P)</td>
<td>db07 backup network (LACP)</td>
</tr>
<tr>
<td>Po19(SU)</td>
<td>Eth1/3/4(P)</td>
<td>Eth1/3/4(P)</td>
<td>db08 backup network (LACP)</td>
</tr>
</tbody>
</table>

MTU

The Maximum Transmission Unit (MTU) setting helps determine the largest packet size that can be communicated over the network. The default MTU is 1500 bytes. Jumbo frames (9000 bytes) reduce the level of IP fragmentation and allow for more efficient processing of large streaming workloads with lower CPU overhead.

When directly connecting the backup network to the Oracle Exadata X8M internal clustering switches, the backup ports should be configured to use the default MTU of 1500.

When placing the backup network on dedicated switches, the ports should be configured to support jumbo frames.

IP networks perform path MTU discovery and auto-negotiate the MTU size between the client, switches and server. Client interfaces using a smaller MTU can still communicate with switches and servers that support jumbo frames.

vPC Configuration Steps

This section provides an example on the steps necessary to configure and optimize the switches using vPC and LACP for maximum performance and redundancy. There are three steps in the configuration process. First, enabling the switches for vPC and establishing peer links. Second, configuring LACP and port-channel groups for the storage controllers. Third, configuring LACP and port-channel groups for the database nodes. The example provided here uses Virtual Local Area Network (VLAN) 478 for the backup network.

VIRTUAL PORT-CHANNEL CONFIGURATION

```bash
switch1
switch1# conf t
switch1(config)# feature vpc
switch1(config)# feature lacp
switch1(config)# spanning-tree vlan 2744-3967 priority 4096
switch1(config)# vrf context peerkeepalive
switch1(config-if)# vpc domain 4
switch1(config-vpc-domain)# peer-switch
```
switch1(config-vpc-domain)# role priority 3000
switch1(config-vpc-domain)# system-priority 100
switch1(config-vpc-domain)# peer-keepalive destination <SWITCH2_IP> source <SWITCH1_IP>
switch1(config-vpc-domain)# delay restore 10
switch1(config-vpc-domain)# peer-gateway
switch1(config-vpc-domain)# ipv6 nd synchronize
switch1(config-vpc-domain)# ip arp synchronize
switch1(config)# interface port-channel1
switch1(config-if)# description switch2 Po1 e1/34-35
switch1(config-if)# switchport mode trunk
switch1(config-if)# switchport trunk native vlan 999
switch1(config-if)# switchport trunk allowed vlan 478,999
switch1(config-if)# spanning-tree port type network
switch1(config-if)# no lacp suspend-individual
switch1(config-if)# vpc peer-link
switch1(config-if)# int e1/34-35
switch1(config-if-range)# description vpc_peerlinks switch1 e1/34-35
switch1(config-if-range)# channel-group 1 mode active

switch1# conf t
switch1(config)# interface breakout module 1 port 1 map 25g-4x
switch1(config)# interface port-channel 10
switch1(config-if)# description zs_ctlr1-net1/2
switch1(config-if)# switchport access vlan 478

switch2
switch2(config)# feature vpc
switch2(config)# feature lacp
switch2(config)# spanning-tree vlan 2744-3967 priority 8192
switch2(config)# vrf context peerkeepalive
switch2(config-if)# vpc domain 4
switch2(config-vpc-domain)# role priority 6000
switch2(config-vpc-domain)# system-priority 100
switch2(config-vpc-domain)# peer-keepalive destination <SWITCH1_IP> source <SWITCH2_IP>
switch2(config-vpc-domain)# delay restore 10
switch2(config-vpc-domain)# peer-gateway
switch2(config-vpc-domain)# ipv6 nd synchronize
switch2(config-vpc-domain)# ip arp synchronize
switch2(config)# interface port-channel1
switch2(config-if)# description switch2 Po1 e1/34-35
switch2(config-if)# switchport mode trunk
switch2(config-if)# switchport trunk native vlan 999
switch2(config-if)# switchport trunk allowed vlan 478,999
switch2(config-if)# spanning-tree port type network
switch2(config-if)# no lacp suspend-individual
switch2(config-if)# vpc peer-link
switch2(config-if)# int e1/34-35
switch2(config-if-range)# description vpc_peerlinks switch2 e1/34-35
switch2(config-if-range)# channel-group 1 mode active

LACP CONFIGURATION FOR STORAGE CONTROLLERS

switch1
switch1# conf t
switch1(config)# interface breakout module 1 port 1 map 25g-4x
switch1(config)# interface port-channel 10
switch1(config-if)# description zs_ctlr1-net1/2
switch1(config-if)# switchport access vlan 478
switch1(config-if)# mtu 1500
switch1(config-if)# speed 25000
switch1(config-if)# duplex full
switch1(config-if)# vpc 10
switch1(config-if)# interface port-channel 11
switch1(config-if)# description zs_ctlr2-net1/2
switch1(config-if)# switchport access vlan 478
switch1(config-if)# mtu 1500
switch1(config-if)# speed 25000
switch1(config-if)# duplex full
switch1(config-if)# vpc 11
switch1(config)# interface ethernet 1/1/1-2
switch1(config-if-range)# description zs_ctlr1-p1
switch1(config-if-range)# lacp rate fast
switch1(config-if-range)# fec off
switch1(config-if-range)# channel-group 10 force mode active
switch1(config-if-range)# exit
switch1(config)# interface ethernet 1/1/3-4
switch1(config-if-range)# description zs_ctlr2-p1
switch1(config-if-range)# lacp rate fast
switch1(config-if-range)# fec off
switch1(config-if-range)# channel-group 11 mode active

switch2
switch2# conf t
switch2(config)# interface breakout module 1 port 1 map 25g-4x
switch2(config)# interface port-channel 10
switch2(config-if)# description zs_ctlr1-net1/2
switch2(config-if)# switchport access vlan 478
switch2(config-if)# mtu 1500
switch2(config-if)# speed 25000
switch2(config-if)# duplex full
switch2(config-if)# vpc 10
switch2(config-if)# interface port-channel 11
switch2(config-if)# description zs_ctlr2-net1/2
switch2(config-if)# switchport access vlan 478
switch2(config-if)# mtu 1500
switch2(config-if)# speed 25000
switch2(config-if)# duplex full
switch2(config-if)# vpc 11
switch2(config-if)# interface ethernet 1/1/1-2
switch2(config-if-range)# description zs_ctlr1-p2
switch2(config-if-range)# fec off
switch2(config-if-range)# lacp rate fast
switch2(config-if-range)# channel-group 10 mode active
switch2(config-if-range)# exit
switch2(config)# interface ethernet 1/1/3-4
switch2(config-if-range)# description zs_ctlr2-p2
switch2(config-if-range)# lacp rate fast
switch2(config-if-range)# fec off
switch2(config-if-range)# channel-group 11 mode active
LACP CONFIGURATION FOR DATABASE NODES

switch1
switch1(config)# interface breakout module 1 port 2-3 map 25g-4x
switch1(config)# interface port-channel 12
switch1(config-if)# description dbnode1
switch1(config-if)# switchport access vlan 478
switch1(config-if)# mtu 1500
switch1(config-if)# speed 25000
switch1(config-if)# duplex full
switch1(config-if)# vpc 12
switch1(config-if)# interface ethernet 1/2/1
description dbnode1-net3
switch1(config-if)# lacp rate fast
switch1(config-if)# fec off
switch1(config-if)# channel-group 12 mode active

switch2
switch2(config)# interface breakout module 1 port 2-3 map 25g-4x
switch2(config)# interface port-channel 12
switch2(config-if)# description dbnode1
switch2(config-if)# switchport access vlan 478
switch2(config-if)# mtu 1500
switch2(config-if)# speed 25000
switch2(config-if)# duplex full
switch2(config-if)# vpc 12
switch2(config-if)# interface ethernet 1/2/1
description dbnode1-net4
switch2(config-if)# lacp rate fast
switch2(config-if)# fec off
switch2(config-if)# channel-group 12 mode active

-----------Repeat steps above for all DB nodes with port-channel 13 to 19-----------

Multiple Backup Networks

The Oracle ZFS Storage Appliance ZS7-2 can serve as a backup target for multiple environments. VLANs should be used to isolate the backup networks.

This section covers guidance when using a single ZS7-2 to provide data protection for databases residing on two independent Oracle Exadata X8Ms. These guidelines are not applicable for multi-rack Exadata configurations. When backing up multi-rack systems it is recommended to use dedicated top of rack switches.

SHARING EXADATA X8M SWITCHES

When the backup network is placed on the Exadata X8M switches additional 25GbE network cards should be added to the storage controllers. The configuring IO cards section later in this document provides additional details. This solution supports up to a 2:1 backup target ratio. Figure 4 provides a networking diagram to support data protection for two Oracle Exadata Database Machines when the backup networks share X8M switches.
DEDICATED TOP OF RACK SWITCHES

When the backup network is placed on dedicated switches, no additional networking cards are required for the storage controllers to support multiple backup networks.

This solution supports up to a 4:1 backup target ratio. Figure 5 provides a networking diagram to support data protection for two Oracle Exadata Database Machines when the backup networks reside on dedicated top of rack switches.

Figure 4. Network diagram when the backup network is sharing the Exadata X8M switches (2:1 backup ratio)
Figure 5. Network diagram with dedicated top of rack switches (2:1 backup ratio)
CONFIGURING THE ORACLE ZFS STORAGE APPLIANCE

This section provides best practices for optimizing an Oracle ZFS Storage Appliance system to provide Oracle Database protection in an Oracle Exadata X8M environment with a standard RMAN backup strategy.

Choosing a Controller

Oracle ZFS Storage Appliance systems are currently available in two models: Oracle ZFS Storage ZS7-2 mid-range and Oracle ZFS Storage ZS7-2 High-End. The following table provides details for each model.

TABLE 3. ORACLE ZFS STORAGE APPLIANCE DETAILS

<table>
<thead>
<tr>
<th>Features</th>
<th>Oracle ZFS Storage ZS7-2 Mid-Range</th>
<th>Oracle ZFS Storage ZS7-2 High-End</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPU Cores</td>
<td>72</td>
<td>96</td>
</tr>
<tr>
<td>DRAM</td>
<td>1 or 2 TB</td>
<td>3 TB</td>
</tr>
<tr>
<td>Maximum Read-Optimized Flash</td>
<td>Up to 491 TB</td>
<td>Up to 1.4 PB</td>
</tr>
<tr>
<td>Raw Storage Capacity</td>
<td>24 TB to 5.3 PB Scalability</td>
<td>24 TB to 16 PB Scalability</td>
</tr>
<tr>
<td>High-Availability Cluster Option</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Focus</td>
<td>Mid-range</td>
<td>Scalability</td>
</tr>
</tbody>
</table>

Please refer to oracle.com/storage/nas/index.html for the latest Oracle ZFS Storage Appliance model specifications.

Oracle ZFS Storage ZS7-2 High-End is a flagship product that offers maximum levels of scalability, CPU, and DRAM. This is a highly scalable platform that can support up to 16 PB of raw storage capacity.

Oracle ZFS Storage ZS7-2 Mid-Range is a cost-efficient model that can still achieve high levels of throughput and redundancy but does not provide the same level of scalability that the high-end model does.

A ZS7-2 high-end model is recommended for Exadata backup.

Choosing the Correct Disk Shelves

Oracle ZFS Storage Appliance systems include a configurable number of disk shelves. A standard RMAN backup use case should use high-capacity (DE3-24C) disk shelves. Oracle Storage Drive Enclosure DE3-24C features high-capacity 14 TB disks. Each disk shelf contains 24 disks and can be configured with optional write-optimized flash. (Up to four disks per disk shelf can be replaced with solid-state drive [SSD] write-flash accelerators.) Oracle ZFS Storage ZS7-2 can be customized based on disk shelf and write-optimized flash requirements. A minimum of 3 disk shelves should be included in the configuration to provide full redundancy. This will allow for a No Single Point of Failure (NSPF) storage configuration that can tolerate the loss of a disk shelf.

For standard RMAN backup use cases it is recommended to include write-flash accelerators at a minimum ratio of 4 per every 4 disk shelves. For example, an 8 disk shelf configuration would include 6 shelves of 24 14 TB drives and 2 shelves of 20 14 TB drives and 4 write-flash accelerators. If standard RMAN backup is the primary use case but the ZS7-2 will also be handling other use cases that generate a significant number of synchronous writes, then it is recommended to include write-flash accelerators at a minimum ratio of 4 per every 2 disk shelves.
Choosing a Storage Profile

When a storage profile is selected to protect Oracle Exadata, mirrored, single-parity, and double-parity profiles are all worthy of consideration. The following table provides a comparison of the storage profiles. For a standard RMAN backup use case it is recommended to use a double parity storage profile.

**TABLE 4. STORAGE PROFILE COMPARISON**

<table>
<thead>
<tr>
<th></th>
<th>Usable Capacity</th>
<th>Advantages</th>
<th>Negatives</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mirrored</td>
<td>42.2%</td>
<td>Restore performance</td>
<td>Costly</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Maximum protection</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Maximum flexibility</td>
<td></td>
</tr>
<tr>
<td>Single Parity</td>
<td>69.3%</td>
<td>Backup performance</td>
<td>Limited redundancy</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Moderate flexibility</td>
<td></td>
</tr>
<tr>
<td>Double Parity</td>
<td>76.7%</td>
<td>Streaming performance</td>
<td>Limited IOPS</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Most efficient</td>
<td></td>
</tr>
</tbody>
</table>

Note: Usable capacity accounts for raw capacity lost due to parity, spares, and file system overhead, as well as small amounts of space lost on each disk due to operating system (OS) overhead, drive manufacturer overhead, and scratch space reservations. This will vary slightly depending on the size of the storage pool; this example assumes a configuration with four disk shelves.

**DOUBLE PARITY**

Double parity provides the best usable capacity and performs as well as single parity for large streaming I/O, which is typical for standard Oracle RMAN workloads. It accomplishes this by utilizing a wide stripe width. The width varies at the time of storage pool creation depending on the number of disks in the configuration, but it ranges up to 14 disks. As a result, the number of vdevs in a double-parity storage pool is far fewer than with mirrored or single parity profiles. The ability to handle IOPS-intensive workloads is severely diminished.

Double parity is recommended when Oracle ZFS Storage Appliance systems are 100 percent dedicated to large sequential workloads, such as traditional Oracle RMAN backup and restore workloads. It is not advisable for use cases such as cloning for DevTest provisioning or utilizing an incrementally updated backup strategy. Mirrored or single-parity profiles are more flexible for handling additional use cases that might result in heavier disk IOPS with lower latencies. Figure 6 reflects raw disk capacity distribution for different storage profiles.

![Figure 6. Raw disk capacity distribution](image)
Configuring the Storage Pools

In most situations it is recommended to configure two storage pools with one primary on each controller. This allows for an active/active backup environment that leverages the CPU, memory and networking resources of both ZS7 controllers. This provides better performance and ROI.

In unusual situations with strict performance SLAs that must be maintained even during failure scenarios it may be recommended to configure a single storage pool and enable an active/standby backup environment where one of the ZS7 controllers will sit idle waiting to take over for the active controller.

In an active/active backup environment each storage pool should be configured with half of the data disk drives in each Oracle Storage Drive Enclosure DE3-24C disk shelf. Each storage pool should also include half of the write-flash accelerators in each DE3-24C disk shelf. This allows for maximum performance and redundancy.

It is recommended to select the No Single Point of Failure (NSPF) option when configuring the storage pool. This ensures that the loss of an entire disk shelf does not compromise the availability of data. To enable NSPF a minimum of three disk shelves is needed for double-parity profiles.

Follow these steps to configure storage pools for an active/active standard backup deployment using the ZFS Storage Appliance’s Browser User Interface (BUI).

Build a storage pool that will be primary on controller 1:

1. Access the BUI on controller 1, select Configuration > Storage and configure a storage pool by clicking the plus icon next to Available Pools.

2. Provide a descriptive name for the new storage pool. Controller 1 will have primary ownership for this storage pool. In this example the storage pool is named “BACKUP1”. Click apply to move to the device allocation screen.

3. Select half of the data devices and half of the log devices from each disk shelf to be included in the storage pool. This method of device allocation will provide optimal performance and redundancy. Select commit to move to the storage profile configuration screen.

4. Choose a data profile of double parity with NSPF. For maximum redundancy it is recommended to select a log profile of mirrored. Select commit to complete the configuration and build the new storage pool.
**Build a storage pool that will be primary on controller 2:**

1. Access the BUI on **controller 2**, select **Configuration > Storage** and configure a storage pool by clicking the plus icon next to Available Pools.

2. **Provide a descriptive name** for the new storage pool. Controller 2 will have primary ownership for this storage pool. In this example the storage pool is named “BACKUP2”. **Click apply** to move to the device allocation screen.

3. Select the remaining data and log devices **from each disk shelf** to be included in the storage pool. **Select commit** to move to the storage profile configuration screen.

4. **Choose a data profile** of double parity with NSPF. For maximum redundancy it is recommended to select a log profile of mirrored. **Select commit** to complete the configuration and build the new storage pool.

**Configuring the Projects and Shares**

A project provides a single access and control point for managing filesystems (shares). Projects can be used for grouping logically related shares together. Shares within a project typically share common settings. Quotas can be enforced at the project level in addition to the share level.

It is recommended to use a project for managing the Oracle Exadata backup solution. Projects for different storage pools can share the same name.

**Create a project for the BACKUP1 storage pool:**

1. Access the BUI on **controller 1**, select **Shares > Projects** and create a new project by clicking the plus icon next to Projects.

2. **Provide a descriptive name** for the new project such as “bkup_x8m”.

3. **Click apply** to create the new project.

**Repeat this process** to create a project on the BACKUP2 storage pool by accessing the BUI on **controller 2**.
Project settings can be configured so that any filesystem (share) created within the project with inherit settings from the project.

**Optimize the new project settings for a standard RMAN backup use case:**

1. **Access the BUI on controller 1,** select Shares > Projects > “bkup_x8m” (project name) > General.

2. **Configure the mountpoint.** A mountpoint of “/export/zs1” was used in this example.

3. **Select LZ4 compression.** LZ4 compression should be enabled on the storage. It provides additional benefit when combined with Oracle Database compression by reducing the bandwidth to back-end disk with only minor impact to Oracle ZFS Storage Appliance ZS7-2 CPU utilization.

   Physical network throughput is typically increased when using LZ4 since SAS bandwidth and disk utilization are often limiting factors.

   With Advanced Row Compression enabled for an OLTP Oracle Database, LZ4 typically provides additional space savings in the range of 1.8x to 2.4x.

4. **Set cache device usage to “do not use cache devices”**. Read-optimized flash should not be used for caching standard Oracle RMAN workloads because there is little benefit from storing Oracle RMAN backup sets in cache. Moreover, the level 2 ARC is not intended for streaming workloads.

5. **Set synchronous write bias to throughput.** This is a share setting that controls behavior for servicing synchronous writes. It can be optimized for latency or throughput.

   All writes are initially written to the ZFS adaptive replacement cache (ARC), regardless of whether they are asynchronous, synchronous, latency-optimized, or throughput-optimized. Also, all writes are copied from the ARC to the storage pool.

   An asynchronous write returns an acknowledgement to the client after the write to ARC is complete. When synchronous writes are optimized for throughput, an acknowledgement is not returned until the write is copied to the storage pool.

   When synchronous writes are optimized for latency, an additional copy is written to persistent storage so that acknowledgements can be returned to the client faster. When write-optimized flash is configured in the storage pool, it is used as the persistent storage for latency-sensitive synchronous writes.

   Standard RMAN backup use cases generate bandwidth-sensitive workloads and writes are mostly asynchronous.

6. **Configure the record size to 1 MB.** The record size influences the size of back-end disk I/O.

   Optimal settings depend on the network I/O sizes used by the application—in this case, Oracle RMAN. Standard Oracle RMAN workloads with Direct NFS Client generate large 1 MB writes and reads at the network layer. In this case, a 1 MB record-size setting should be used.

   The ability to use large record sizes has significant advantages, such as increased throughput performance, which is critical for bandwidth-intensive workloads. Other benefits include reduced utilization of controller CPU resources.

   In recent years, HDD capacities have grown as quickly as ever, yet the IOPS these disks can deliver has leveled off. Oracle RMAN workloads often generate datasets on the TB scale, with only a small frequency of read-backs. As such, caching is not an optimal solution for handling IOPS. Maximizing the throughput and limiting the IOPS to disk are important factors for achieving the best performance from the backup solution. Oracle RMAN standard backup strategies enable this by delivering large, multichannel network I/O that greatly benefits from large record sizes on the filesystem.
7. Configure permissions for the oracle user (1001) and dba group (1002) with file access of 750. Reference figure 8 for an example.

8. Click apply to configure the default settings for all shares in this project.

**Repeat this process** to optimize settings for the other storage pool’s project by accessing the BUI on **controller 2**.

---

**Inherited Properties**

- **Mountpoint**: /export/zs1
- **Read only**: 
- **Update access time on read**: checked
- **Non-blocking mandatory locking**: 
- **Data deduplication**: checked
- **Data compression**: LZ4 (Fast)
- **Checksum**: Fletcher4 (Standard)
- **Cache device usage**: Do not use cache devices
- **Synchronous write bias**: Throughput
- **Database record size**: 1M
- **Additional replication**: Normal (Single Copy)
- **Virus scan**: 
- **Prevent destruction**: 
- **Restrict ownership change**: 

**Default Settings**

<table>
<thead>
<tr>
<th>FILESYSTEMS</th>
<th>LUNS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>User</strong></td>
<td>1001</td>
</tr>
<tr>
<td><strong>Group</strong></td>
<td>1002</td>
</tr>
<tr>
<td><strong>Permissions</strong></td>
<td>R W X R W X</td>
</tr>
</tbody>
</table>

---

**Configure the NFS share mode:**

1. Navigate to **Shares > Projects > “bkup_x8m”** (project name) > **Protocols** and set the NFS share mode to **read/write**.

2. Ensure that the NFS protocols settings match figure 9 and then **click apply**.

3. **Repeat this step** for the other project.

---

**Figure 8. Setting default share settings for each project**

**Figure 9. NFS protocols settings for “bkup_x8m” project**

The next step is to create the shares that will be mounted and accessed by the NFS clients.
Create a filesystem (share) for the BACKUP1 storage pool:
1. Access the BUI on controller 1 and select Shares > Projects > “bkup_x8m” (project name) > Shares.
2. Create a filesystem by clicking on the plus icon next to Filesystems.
3. Assign the filesystem a unique name. This example creates a filesystem named “bkup_x8m_1”. The filesystem will inherit the properties that were configured for the “bkup_x8m” project. No further configuration is required.
4. Click apply to create the filesystem.

Create a filesystem (share) for the BACKUP2 storage pool:
1. Access the BUI on controller 2 and select Shares > Projects > “bkup_x8m” (project name) > Shares.
2. Create a filesystem by clicking on the plus icon next to Filesystems.
3. Assign the filesystem a unique name. This example creates a filesystem named “bkup_x8m_2”.
4. Click apply to create the filesystem.

Two filesystems have been created, one on each storage pool.

Configuring the NFS Server
Start the NFS service and optimize settings:
1. Access the BUI and navigate to Configuration > Services and select NFS
2. Ensure that the service is enabled and online.
3. Set the maximum supported version to NFSv4.1.
4. Configure the maximum # of server threads to 3000.
5. Click apply to commit the changes.

It is only required to perform this step on one controller.

Configuring IO Cards
It is recommended that the Oracle ZFS Storage Appliance ZS7-2 high-end system be configured with four SAS-3 cards (slots 2, 3, 9 and 10) and two dual port 25Gb Ethernet cards (slots 4 and 8) in each controller.

If the ZS7-2 will be providing data protection services for two backup networks residing on different switches, then two additional 25Gb Ethernet cards can be installed in slots 5 and 7 to avoid sending backup data over uplink ports and allow for direct connection into the secondary backup network.

Figure 10 provides an example of the recommended PCIe configuration.
Figure 10. Oracle ZFS Storage Appliance ZS7-2 high-end PCIe configuration

Configuring Data Path Network

The data path network is represented by the blue 25Gb SFP28 ports in figure 11.

These interfaces are connected to 100Gb switches using 4x25Gb breakout cables. Port 1 of each card is connected to switch 1 and port 2 of each card is connected to switch 2. Reference the network configuration chapter for detailed port mapping.

A Link Aggregation Control Protocol (LACP) group is used to provide full HA redundancy. The Oracle Direct NFS Client can provide a level of HA, but currently relies on the kernel NFS mount for opening or creating files. LACP is required to provide full HA in all situations.

A Virtual Network Interface Card (VNIC) is defined on the Oracle ZFS Storage Appliance to provide secondary access to the physical link aggregation group. The physical datalink will provide data path services for one of the storage pools while the VNIC will provide data path services for the other storage pool.

Storage pools and the data path network are clustered resources and will transfer ownership during a controller failover.

Create an LACP aggregation of four 25Gb ports:

1. Access the BUI on controller 1, select Configuration > Network > Configuration and build a new datalink by clicking the plus icon next to the Datalinks column.

2. Provide a descriptive name for the new datalink. This datalink will provide data path for the BACKUP1 storage pool. Controller 1 has primary ownership for this storage pool. In this example the datalink’s name is 4p-lacp-1.
3. **Select the Maximum Transmission Unit (MTU).** The MTU setting helps determine the maximum packet size that can be communicated over the network.

   The default MTU is 1500 bytes. Jumbo frames (9000 bytes) reduce the level of IP fragmentation and allow for more efficient processing of large streaming workloads with lower CPU overhead. Larger MTUs improve performance for backup networks.

   When directly connecting the backup network to the Oracle Exadata X8M internal clustering switches, the backup ports will use an MTU of 1500. When using ToR switches for the backup network, the ports will use an MTU of 9000.

   IP networks will always perform path MTU discovery and auto-negotiate the MTU size between the client, switches and server. Client interfaces using a smaller MTU can still communicate with switches and servers that support jumbo frames.

4. **Select the LACP Aggregation checkbox and add all four 25Gb devices.** Reference figure 12 for an example.

5. **Configure the policy to L3 and L4.** This manages load balancing across the underlying network devices by hashing based on IP as well as (TCP/UDP) port numbers.

6. **Set the timer to short.** This manages the frequency of LACP control messages. A short setting will match the configuration on the Oracle Exadata compute nodes.

7. **Click apply** to create the LACP aggregation datalink.

---

Create an VNIC of the physical link aggregation group:

1. Access the BUI on controller 2, select Configuration > Network > Configuration and build a new datalink by clicking the plus icon next to the Datalinks column.

2. **Provide a descriptive name** for the new datalink. This datalink will provide data path for the BACKUP2 storage pool. Controller 2 has primary ownership for this storage pool. In this example the datalink’s name is 4p-ltcp-2.
3. **Configure the MTU.** Match the MTU with the previously created LACP aggregation datalink.

4. **Select the VNIC checkbox** under properties and **add the previously created LACP aggregation datalink.** Reference figure 13 for an example.

5. **Click apply** to create the VNIC datalink.

### Network Datalink

<table>
<thead>
<tr>
<th>Properties</th>
<th>VLAN</th>
<th>VNIC</th>
<th>IP Partition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name</td>
<td>4p-lacp-2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max Transmission Unit (MTU)</td>
<td>1500</td>
<td>Custom</td>
<td>9000</td>
</tr>
<tr>
<td>VLAN ID</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Datalinks</th>
<th>21 available</th>
<th>LACP Aggregation</th>
</tr>
</thead>
<tbody>
<tr>
<td>4p-lacp-1</td>
<td>0:e7:cf:4b:ad</td>
<td></td>
</tr>
<tr>
<td>igb0-1</td>
<td>0:10:e0:de:89:d4</td>
<td></td>
</tr>
</tbody>
</table>

Figure 13. Create a VNIC for data path

Create a network interface for the first storage pool:

1. **Access the BUI on controller 1**, select **Configuration > Network > Configuration** and **build a new interface** by clicking the plus icon next to the Interfaces column.

2. **Provide a descriptive name** for the new interface. This interface will provide data path for the BACKUP1 storage pool. Controller 1 has primary ownership for this storage pool. In this example the interface’s name is 4p-lacp-1.

3. **Assign a unique IP address to this interface.** This will be the address used by the Oracle Exadata compute nodes to access the first storage pool.

4. **Select the physical LACP group (4p-lacp-1) as the datalink associated with this interface.** Reference figure 14 as an example.

5. **Click apply** to create the interface.
Create a network interface for the second storage pool:

1. Access the BUI on controller 2, select Configuration > Network > Configuration and build a new interface by clicking the plus icon next to the Interfaces column.

2. Provide a descriptive name for the new interface. This interface will provide data path for the BACKUP2 storage pool. Controller 2 has primary ownership for this storage pool. In this example the interface’s name is 4p-lacp-2.

3. Assign a unique IP address to this interface.

4. Select the VNIC LACP group (4p-lacp-2) as the datalink associated with this interface.

5. Click apply to create the interface.

Enable adaptive routing for the multihoming policy when using LACP to ensure that outbound traffic from an Oracle ZFS Storage Appliance system is balanced over the network links and IP addresses. Access the BUI, select Configuration > Network > Routing, and then select option multihoming=adaptive.

Choosing Direct NFS Client

The Direct NFS Client feature of Oracle Database is highly recommended for all Oracle RMAN workloads between Oracle Exadata and Oracle ZFS Storage Appliance systems, and it is required to achieve optimal performance.

Direct NFS Client is a custom NFS client that resides within the Oracle Database kernel and provides several key advantages:

- Significantly reduces system CPU utilization by bypassing the OS and caching data just once in user space with no second copy in kernel space
- Boosts parallel I/O performance by opening an individual TCP connection for each Oracle Database process
- Distributes throughput across multiple network interfaces by alternating buffers to multiple IP addresses in a round-robin fashion
- Provides high availability (HA) by automatically redirecting failed I/O to an alternate address
These advantages enable increased bandwidth and reduced CPU overhead.

No additional steps are required on Oracle ZFS Storage Appliance systems to enable Direct NFS Client.

ORACLE INTELLIGENT STORAGE PROTOCOL

Oracle Intelligent Storage Protocol, a feature of Oracle ZFS Storage Appliance systems, was introduced to interact with Direct NFS Client in Oracle Database 12c. It enables Oracle Database–aware storage by dynamically tuning record size and synchronous write bias on Oracle ZFS Storage Appliance systems. This simplifies the configuration process and reduces the performance impact of configuration errors. Hints are passed from the Oracle Database kernel to the Oracle ZFS Storage Appliance system. These hints are interpreted to construct a workload profile to dynamically optimize storage settings.

Oracle Intelligent Storage Protocol is an optional protocol that requires NFSv4 and SNMP. In the current implementation, a properly configured environment that adheres to the best practices in this document performs equally well without Oracle Intelligent Storage Protocol. For instructions on how to enable Oracle Intelligent Storage Protocol, see My Oracle Support document 1943618.1 "Oracle ZFS Storage Appliance: How to Enable Oracle Intelligent Storage Protocol (OISP)."
CONFIGURING THE ORACLE EXADATA DATABASE MACHINE

This section provides best practices for configuring an Oracle Exadata X8M when using an Oracle ZFS Storage Appliance ZS7-2 high end system to perform Oracle Database protection with a standard RMAN backup strategy.

Configuring Backup Network

The backup network is represented by the blue 25Gb SFP28 ports in figure 15 and figure 16.

These interfaces are connected to 100Gb switches using 4x25Gb breakout cables. Port 1 is connected to the top switch and port 2 is connected to the bottom switch. Reference the network configuration chapter for detailed port mapping.

**External Connectivity for X8M-2 DB Workloads**

*Quarter rack and larger configurations*

![Network Connectivity Diagram](image)

- 2x SFP28 10G/25G BCM57414
- eth3
- eth4
- eth0
- re0
- re1
- eth5
- eth6
- 2x QSFP28 HCA RoCE
- Optional: P/N7118016 - Oracle Dual Port 25 Gb Ethernet Adapter (2x SFP28 10/25G Ports)

**Figure 15. Network connectivity for X8M-2 quarter rack or larger database nodes**

**External Connectivity for X8M-2 DB Workloads**

*Eighth rack configurations*

![Network Connectivity Diagram](image)

- 2x SFP28 10G/25G BCM57414
- re0
- re1
- eth3
- eth4
- eth0
- eth1
- eth2
- ILOM (1G NET MGT)
- HOST admin (1G NET 0)
- RJ45 (Net 3/4) OR SFP28 (Net1/2) BCM57417 LOM

**Figure 16. Network connectivity for X8M-2 eighth rack database nodes**

The backup network is configured as bondeth1 interface on each database node. This interface is comprised of eth3 and eth4 and can be configured using either Linux bonding in active-backup (mode 1) or Link Aggregation Control Protocol (LACP). It is recommended to use LACP to load balance across both underlying interfaces. The configuration steps in this solution brief assume the use of LACP. Oracle Exadata Deployment Assistant (OEDA) can assist with the configuration of the backup network.

LACP should be configured to use a policy of layer3+4 and a fast polling rate. The default MTU is 1500 bytes. Jumbo frames (9000 bytes) reduce the level of IP fragmentation and allow for more efficient processing of large streaming workloads with lower CPU overhead.
Larger MTUs improve performance for backup networks. When directly connecting the backup network to the Oracle Exadata X8M internal clustering switches, the backup ports will use an MTU of 1500. When using dedicated top of rack (ToR) switches for the backup network it is recommended to configure the switch ports for jumbo frames.

Optimal settings for bondeth1 are shown in this example:

```
~$ cat /proc/net/bonding/bondeth1
Bonding Mode: IEEE 802.3ad Dynamic link aggregation
Transmit Hash Policy: layer3+4 (1)
MII Status: up
MII Polling Interval (ms): 100
Up Delay (ms): 200
Down Delay (ms): 200

802.3ad info
LACP rate: fast
Min links: 0
Aggregator selection policy (ad_select): stable

Slave Interface: eth3
MII Status: up
Speed: 25000 Mbps
Duplex: full
<...
Slave Interface: eth4
MII Status: up
Speed: 25000 Mbps
Duplex: full
```

Configure Backup Shares

The following mount options are recommended for shares dedicated to standard RMAN backup use cases:

```
rw,_netdev,hard,rsize=1048576,wsize=1048576,tcp,nfsvers=4,timeo=600 0 0
```

Direct NFS Client does not utilize NFS mount options. However, setting the proper mount options is recommended to follow Oracle Database requirements and to improve performance and functionality if Direct NFS Client is not available and the system reverts to NFS.

Backup shares should be mounted on all Oracle Database nodes.

Mount the Oracle ZFS Storage Appliance backup shares:

1. **Edit the /etc/fstab file** and add entries like the following three-line example
   
   ```
   # added for zs7-2 with a backup workload
   192.168.x.11:/export/zs1/bkup_x8m_1 /zfssa/zs1/bkup_x8m_1 nfs
rw,_netdev,hard,rsize=1048576,wsize=1048576,tcp,nfsvers=4,timeo=600 0 0
   192.168.x.12:/export/zs1/bkup_x8m_2 /zfssa/zs1/bkup_x8m_2 nfs
   rw,_netdev,hard,rsize=1048576,wsize=1048576,tcp,nfsvers=4,timeo=600 0 0
   ```

   In this example, 192.168.xxx.11 is the data path IP address for the “BACKUP1” storage pool and 192.168.xxx.12 is the data path IP address for the “BACKUP2” storage pool. The export path of the
filesystems are /export/zs1/bkup_x8m_1 and /export/zs1/bkup_x8m_2. The mount points on the compute nodes are /zfssa/zs1/bkup_x8m_1 and /zfssa/zs1/bkup_x8m_2.

2. Copy these changes to the other compute nodes.

3. Create the /zfssa base directory on all compute nodes (mkdir /zfssa) and change ownership to the oracle user and dba group (chown oracle:dba /zfssa).

4. As the oracle user, create the mount points used in the fstab on all compute nodes.

   ~]$ mkdir -p /zfssa/zs1/bkup_x8m_1
   ~]$ mkdir -p /zfssa/zs1/bkup_x8m_2

5. Mount the backup shares on all compute nodes.

   mount /zfssa/zs1/bkup_x8m_1; mount /zfssa/zs1/bkup_x8m_2

Configuring Direct NFS Client

In Oracle Database 12c and later, Direct NFS Client is enabled by default.

Confirm that Direct NFS Client is enabled by checking the Oracle Database alert log for an Oracle Disk Manager (ODM) message after Oracle Database startup:

Oracle instance running with ODM: Oracle Direct NFS ODM Library Version 6.0

Direct NFS Client activity can also be confirmed by SQL query:

SQL> select * from v$dnfs_servers;

Database v$ dnfs views will only be populated if Direct NFS Client IO has occurred since the last database start.

For a complete list of recommended patches, see My Oracle Support document 1495104.1 "Recommended Patches for Direct NFS Client."

OPTIONAL ORANFSTAB

The oranfstab file is an optional configuration file that is required to use advanced features of Direct NFS. The Direct NFS client will still function without an oranfstab file and will use NFSv3 from a single source IP address to a single destination IP address. It is not required for this solution.

The oranfstab file configures load spreading of Direct NFS Client connections over multiple addresses on an Oracle ZFS Storage Appliance system (represented by "path") or multiple addresses on Oracle Exadata for Oracle Database (represented by "local"). This is unnecessary if LACP groups are in use since a single source and destination address will be used and load balancing across multiple underlying interfaces will be handled by the Link Aggregation Control Protocol.

The file is created in $ORACLE_HOME/dbs/oranfstab and applies to all Oracle Database instances that share ORACLE_HOME. When changes are made to the oranfstab file, the database should be bounced for the changes to take effect. NFSv4 is required to take advantage of Oracle Intelligent Storage Protocol (OISP). To enable NFSv4 for the Direct NFS, configure an oranfstab like the following example.

server: <STORAGE_NODE1>
local: 192.168.x.101 path: 192.168.x.11
nfs_version: nfsv4
export: /export/zs1/bkup_x8m_1 mount: /zfssa/zs1/bkup_x8m_1
server: <STORAGE_NODE2>
local: 192.168.x.102 path: 192.168.x.12
nfs_version: nfsv4
export: /export/bkup_x8m_2 mount: /zfssa/zs1/bkup_x8m_2
Configuring Oracle RMAN Backup Services

Oracle RMAN backup services should be created and used to balance Oracle RMAN workloads across all Oracle Exadata compute nodes. Spreading a backup across multiple Oracle Real Application Clusters (Oracle RAC) nodes improves performance, increases parallel tasks, and reduces utilization load on any single component. Oracle RMAN backup services are automatically migrated to other Oracle Exadata servers for Oracle Database in the Oracle RAC cluster when the preferred instance is unavailable.

The configuration steps below assume an admin-managed RAC database on an Exadata X8M-2 half rack with four Oracle RAC nodes. Hulk is the database name.

The syntax is: `srvctl add service -d <db_name> -r <preferred instance> -a <alternate instance(s)> -s <name for newly created service>`

```
[oracle@<HOSTNAME> ~]$ srvctl add service -d hulk -r hulk1 -a hulk2,hulk3,hulk4 -s hulk_bkup1
[oracle@<HOSTNAME> ~]$ srvctl add service -d hulk -r hulk2 -a hulk1,hulk3,hulk4 -s hulk_bkup2
[oracle@<HOSTNAME> ~]$ srvctl add service -d hulk -r hulk3 -a hulk1,hulk2,hulk4 -s hulk_bkup3
[oracle@<HOSTNAME> ~]$ srvctl add service -d hulk -r hulk4 -a hulk1,hulk2,hulk3 -s hulk_bkup4
```

```
[oracle@<HOSTNAME> ~]$ srvctl start service -d hulk -s hulk_bkup1
[oracle@<HOSTNAME> ~]$ srvctl start service -d hulk -s hulk_bkup2
[oracle@<HOSTNAME> ~]$ srvctl start service -d hulk -s hulk_bkup3
[oracle@<HOSTNAME> ~]$ srvctl start service -d hulk -s hulk_bkup4
```

```
[oracle@<HOSTNAME> ~]$ srvctl status service -d hulk
Service hulk_bkup1 is running on instance(s) hulk1
Service hulk_bkup2 is running on instance(s) hulk2
Service hulk_bkup3 is running on instance(s) hulk3
Service hulk_bkup4 is running on instance(s) hulk4
```

When the database is restarted the RMAN backup services should be rebalanced. This can be accomplished with:

```
[oracle@<HOSTNAME> ~]$ srvctl stop service -d hulk; srvctl start service -d hulk
```

Preparing Oracle Database for Backup

ARCHIVELOG MODE

Archiving of the online redo logs is enabled when Oracle Database is configured to operate in “archivelog” mode. Benefits of using archivelog mode include

» Protection is provided in the event of media failure.
» Oracle Database transactions that occurred after the most recent backup can be recovered.
» Backups can be performed while Oracle Database is open and active.
» Inconsistent backups can be used to restore Oracle Database.

It is recommended that Oracle Database run in archivelog mode. It is also recommended that Data Guard is used to provide site recovery services. If Data Guard is not in use the archivelogs should be duplexed or multiplexed with the primary copy on Oracle Exadata storage and an optional copy on the Oracle ZFS Storage Appliance system.

BLOCK-CHANGE TRACKING

Block-change tracking is an Oracle RMAN feature that records changed blocks within a datafile. The level 0 backup scans the entire datafile, but subsequent incremental backups rely on the block-change tracking file to scan just the blocks that have been marked as changed since the last backup.
It is recommended to enable block-change tracking to improve performance for incremental backups. If the chosen backup strategy includes only full or level 0 backups, block-change tracking should not be enabled.

**Oracle RMAN Configuration**

**COMPRESSION**

Oracle RMAN compression is suitable with Oracle ZFS Storage Appliance systems only when network bandwidth is limited. RMAN compression should not be used with local backups. It is far more efficient to rely on compression at the database level and at the storage level using LZ4. RMAN compression is recommended for remote backups where network bandwidth is a limiting factor.

Oracle RMAN compression should not be used with Transparent Data Encryption (TDE) because it causes the data to be unencrypted, compressed, and re-encrypted during the backup session. This creates a major bottleneck for backup performance and places an enormous strain on Oracle Database CPU resources. Do not use RMAN encryption. TDE tablespace encryption is the recommended method.

**OPTIMIZING CHANNELS**

Determining the number of Oracle RMAN channels to use is an important aspect of tuning a backup solution. When Oracle RMAN opens a new channel, it allocates a new set of input and output buffers. Each channel can take a datafile or a section of a datafile and process the backup or restore job in parallel to work being done by other channels. Channels can be assigned to different nodes in the Oracle RAC cluster, and can have different backup destinations, with shares potentially owned by different Oracle ZFS Storage Appliance controllers.

Additional channels increase scalability and can provide significantly improved performance, more-efficient resource utilization, load balancing across Oracle Database nodes, a more robust HA architecture, and workload spreading between storage controllers.

As hardware limits are approached, allocating additional Oracle RMAN channels provides diminishing returns. It is not recommended to over-allocate channels because there is no performance gain, despite additional memory and CPU resources allocated for more Oracle RMAN buffers and added complexity in the form of more backup pieces being created.

Determining the recommended number of channels for a configuration depends on the hardware factor that will limit overall performance in an optimally configured solution. Performance limiting components could be many things, including Oracle Exadata or the network, HDD, CPU, or SAS resources. Thorough testing is always recommended when implementing major changes in a production environment. However, the following table provides guidance for how many Oracle RMAN channels to configure in a traditional Oracle RMAN backup strategy for each hardware configuration.

This table assumes an Oracle Exadata X8M-2 and an Oracle ZFS Storage ZS7-2 with storage balanced across both controllers. It assumes that network and SAS bandwidth are not limiting factors, that the best practices in this document are implemented, and that there are no other significant concurrent workloads during the backup window. When following these sizing guidelines, CPU resources consumed on the X8M database nodes for RMAN workloads will always be less than 5%.

**TABLE 5. SUGGESTED RMAN CHANNELS PER CONFIGURATION FOR A STANDARD BACKUP STRATEGY**

<table>
<thead>
<tr>
<th>Channels per</th>
<th>Channels per</th>
<th>Channels per</th>
<th>Channels per</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oracle Exadata</td>
<td>Oracle Exadata</td>
<td>Oracle Exadata</td>
<td>Oracle Exadata</td>
</tr>
<tr>
<td>Eighth Rack</td>
<td>Quarter Rack</td>
<td>Half Rack</td>
<td>Full Rack</td>
</tr>
<tr>
<td>1 Disk Shelf</td>
<td>8</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>2 Disk Shelves</td>
<td>8</td>
<td>12</td>
<td>12</td>
</tr>
</tbody>
</table>
When Oracle RMAN channels are configured or allocated, they should be alternated across the Oracle RAC nodes and storage shares.

SECTION SIZE

Enabling highly parallel Oracle RMAN workloads is critical for achieving optimal performance and resource utilization from the backup solution. One challenge is when a very large datafile is encountered. If it is processed by a single Oracle RMAN channel, throughput slows significantly, and other hardware resources in the environment sit idle while waiting for the outlier datafile processing to be completed.

Oracle RMAN’s solution to this problem lies in its ability to break up large files into smaller pieces that can be processed in parallel by multiple channels. This is called multisection support and is determined by the section size parameter. It is recommended to set the section size to 100 gigabytes (100G).

FILESPERSET PARAMETER

The filesperset parameter determines how many datafiles or sections of datafiles are included in each backup set. When multiple input files are read to create a single backup set, it can improve performance, particularly when the read or copy phases are limiting factors. The default filesperset setting is 64; however, this is detrimental for single-file or partial Oracle Database restore operations because the entire backup set will be read back, even though only a small section is used. Also, an excessively large filesperset setting can impact the load balancing and performance scaling properties of Oracle RMAN. The objective is to have all Oracle RMAN channels effectively utilized throughout the backup. If there is a limited number of datafiles or data sections, it might not be possible to create full backup sets on every channel.

As a general practice, it is recommended to set the filesperset parameter to 1. Testing has shown that this provides excellent performance while load balancing across all channels. If deduplication is enabled, it is a requirement to set filesperset to 1. Including multiple files or sections in the same backup piece diminishes deduplication benefits.

SNAPSHOT CONTROL FILE

The snapshot control file is not part of the backup. It is a temporary file used by the Oracle RMAN process. It should be placed on local shared Exadata storage. Reference the sample run block for an example.

SAMPLE RUN BLOCK

Here is a sample run block for a weekly level 0 backup that can be included as part of an incremental backup strategy. This example assumes an Oracle Exadata X8M-2 half rack backing up to both controllers of an Oracle ZFS Storage Appliance configured with four disk shelves. Hulk is the name of the database used in this example. RMAN backup services are used to evenly spread channels across all four RAC nodes. Channels are alternated between the two storage shares with one owned by each controller. The first file handles persistent configurations and only needs to be run initially and again after any changes to RMAN settings. The second file would be run during every backup cycle.

RUN
{
CONFIGURE CHANNEL 1 DEVICE TYPE DISK CONNECT 'sys/passwd@<SCAN_ADDRESS>/hulk_bkup1' FORMAT '/zfs/zs1/bkp_x8m_1/backup_%d_%U';
CONFIGURE CHANNEL 2 DEVICE TYPE DISK CONNECT 'sys/passwd@<SCAN_ADDRESS>/hulk_bkup2' FORMAT
'/zfsa/zsl/bkup_x8m_2/backup_%d_%U';
CONFIGURE CHANNEL 3 DEVICE TYPE DISK CONNECT 'sys/passwd@<SCAN_ADDRESS>/hulk_bkup3' FORMAT
'/zfsa/zsl/bkup_x8m_1/backup_%d_%U';
CONFIGURE CHANNEL 4 DEVICE TYPE DISK CONNECT 'sys/passwd@<SCAN_ADDRESS>/hulk_bkup4' FORMAT
'/zfsa/zsl/bkup_x8m_2/backup_%d_%U';
CONFIGURE CHANNEL 5 DEVICE TYPE DISK CONNECT 'sys/passwd@<SCAN_ADDRESS>/hulk_bkup1' FORMAT
'/zfsa/zsl/bkup_x8m_1/backup_%d_%U';
CONFIGURE CHANNEL 6 DEVICE TYPE DISK CONNECT 'sys/passwd@<SCAN_ADDRESS>/hulk_bkup2' FORMAT
'/zfsa/zsl/bkup_x8m_2/backup_%d_%U';
CONFIGURE CHANNEL 7 DEVICE TYPE DISK CONNECT 'sys/passwd@<SCAN_ADDRESS>/hulk_bkup3' FORMAT
'/zfsa/zsl/bkup_x8m_1/backup_%d_%U';
CONFIGURE CHANNEL 8 DEVICE TYPE DISK CONNECT 'sys/passwd@<SCAN_ADDRESS>/hulk_bkup4' FORMAT
'/zfsa/zsl/bkup_x8m_2/backup_%d_%U';
CONFIGURE CHANNEL 9 DEVICE TYPE DISK CONNECT 'sys/passwd@<SCAN_ADDRESS>/hulk_bkup1' FORMAT
'/zfsa/zsl/bkup_x8m_1/backup_%d_%U';
CONFIGURE CHANNEL 10 DEVICE TYPE DISK CONNECT 'sys/passwd@<SCAN_ADDRESS>/hulk_bkup2' FORMAT
'/zfsa/zsl/bkup_x8m_2/backup_%d_%U';
CONFIGURE CHANNEL 11 DEVICE TYPE DISK CONNECT 'sys/passwd@<SCAN_ADDRESS>/hulk_bkup3' FORMAT
'/zfsa/zsl/bkup_x8m_1/backup_%d_%U';
CONFIGURE CHANNEL 12 DEVICE TYPE DISK CONNECT 'sys/passwd@<SCAN_ADDRESS>/hulk_bkup4' FORMAT
'/zfsa/zsl/bkup_x8m_2/backup_%d_%U';
CONFIGURE CHANNEL 13 DEVICE TYPE DISK CONNECT 'sys/passwd@<SCAN_ADDRESS>/hulk_bkup1' FORMAT
'/zfsa/zsl/bkup_x8m_1/backup_%d_%U';
CONFIGURE CHANNEL 14 DEVICE TYPE DISK CONNECT 'sys/passwd@<SCAN_ADDRESS>/hulk_bkup2' FORMAT
'/zfsa/zsl/bkup_x8m_2/backup_%d_%U';
CONFIGURE CHANNEL 15 DEVICE TYPE DISK CONNECT 'sys/passwd@<SCAN_ADDRESS>/hulk_bkup3' FORMAT
'/zfsa/zsl/bkup_x8m_1/backup_%d_%U';
CONFIGURE CHANNEL 16 DEVICE TYPE DISK CONNECT 'sys/passwd@<SCAN_ADDRESS>/hulk_bkup4' FORMAT
'/zfsa/zsl/bkup_x8m_2/backup_%d_%U';
CONFIGURE DEVICE TYPE DISK PARALLELISM 16;
CONFIGURE CONTROLFILE AUTOBACKUP ON;
CONFIGURE CONTROLFILE AUTOBACKUP FORMAT FOR DEVICE TYPE DISK TO '/zfsa/zsl/bkup_x8m_1/ctrl_%F';
CONFIGURE SNAPSHOT CONTROLFILE NAME TO '+'RECOCI/hulk/snapcf_hulk.f';
CONFIGURE ARCHIVELOG DELETION POLICY TO BACKED UP 1 TIMES TO DISK;
}

RUN
{
BACKUP AS BACKUPSET
  SECTION SIZE 100G
  INCREMENTAL LEVEL 0
  DATABASE
  FILESPERSET 1
  TAG 'bkup_weekly_L0';
}
CONCLUSION

Selecting the right backup solution for Oracle Exadata is an important decision. Costly third-party alternatives provide poor ROI and cannot support high-performance environments. Competitive offerings are inflexible and do not address all the customer’s needs.

Oracle Exadata X8M leverages a new and improved networking infrastructure built on high-speed, low latency 100Gb Ethernet. The solutions in this paper leverage the Oracle ZS7-2 using either X8M internal switches or dedicated top of rack switches to provide a robust, fault-tolerant and high-performance data protection architecture.

Oracle ZFS Storage Appliance systems have proven to be an ideal solution for protecting the mission-critical data that resides on Oracle Exadata. Powerful features combined with custom Oracle-on-Oracle integrations enable a wide range of Oracle RMAN backup strategies. These provide outstanding performance and flexibility unmatched by third-party solutions.

Extreme restore throughput helps satisfy even the most stringent RTOs. Archive log multiplexing delivers recovery points of 20 minutes or less. Oracle Intelligent Storage Protocol, HCC, LZ4 storage compression, large 1 MB record sizes, data deduplication, and Direct NFS Client provide unique advantages when protecting Oracle Database.

In addition to data protection benefits, an Oracle Exadata backup solution using Oracle ZFS Storage Appliance systems provides many other advantages, such as low-cost, high-performance storage for unstructured data that resides outside of Oracle Database. It is easy to see why Oracle ZFS Storage Appliance systems offer an excellent solution for protecting Oracle Exadata.