

Hitachi Solution for Oracle RAC Database using Hitachi VSS Block Storage

Reference Architecture Guide

© 2024 Hitachi Vantara LLC. All rights reserved.

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including copying and recording, or stored in a database or retrieval system for commercial purposes without the express written permission of Hitachi, Ltd., or Hitachi Vantara LLC (collectively "Hitachi"). Licensee may make copies of the Materials provided that any such copy is: (i) created as an essential step in utilization of the Software as licensed and is used in no other manner; or (ii) used for archival purposes. Licensee may not make any other copies of the Materials. "Materials" mean text, data, photographs, graphics, audio, video and documents.

Hitachi reserves the right to make changes to this Material at any time without notice and assumes no responsibility for its use. The Materials contain the most current information available at the time of publication.

Some of the features described in the Materials might not be currently available. Refer to the most recent product announcement for information about feature and product availability, or contact Hitachi Vantara LLC at <https://support.hitachivantara.com/en-us/contact-us.html>.

Notice: Hitachi products and services can be ordered only under the terms and conditions of the applicable Hitachi agreements. The use of Hitachi products is governed by the terms of your agreements with Hitachi Vantara LLC.

By using this software, you agree that you are responsible for:

1. Acquiring the relevant consents as may be required under local privacy laws or otherwise from authorized employees and other individuals; and
2. Verifying that your data continues to be held, retrieved, deleted, or otherwise processed in accordance with relevant laws.

Notice on Export Controls. The technical data and technology inherent in this Document may be subject to U.S. export control laws, including the U.S. Export Administration Act and its associated regulations, and may be subject to export or import regulations in other countries. Reader agrees to comply strictly with all such regulations and acknowledges that Reader has the responsibility to obtain licenses to export, re-export, or import the Document and any Compliant Products.

Hitachi and Lumada are trademarks or registered trademarks of Hitachi, Ltd., in the United States and other countries.

AIX, DB2, DS6000, DS8000, Enterprise Storage Server, eServer, FICON, FlashCopy, GDPS, HyperSwap, IBM, OS/390, PowerHA, PowerPC, S/390, System z9, System z10, Tivoli, z/OS, z9, z10, z13, z14, z15, z16, z/VM, and z/VSE are registered trademarks or trademarks of International Business Machines Corporation.

Active Directory, ActiveX, Bing, Excel, Hyper-V, Internet Explorer, the Internet Explorer logo, Microsoft, Microsoft Edge, the Microsoft corporate logo, the Microsoft Edge logo, MS-DOS, Outlook, PowerPoint, SharePoint, Silverlight, SmartScreen, SQL Server, Visual Basic, Visual C++, Visual Studio, Windows, the Windows logo, Windows Azure, Windows PowerShell, Windows Server, the Windows start button, and Windows Vista are registered trademarks or trademarks of Microsoft Corporation. Microsoft product screen shots are reprinted with permission from Microsoft Corporation.

All other trademarks, service marks, and company names in this document or website are properties of their respective owners.

Copyright and license information for third-party and open source software used in Hitachi Vantara products can be found in the product documentation, at <https://www.hitachivantara.com/en-us/company/legal.html>.

Feedback

Hitachi Vantara welcomes your feedback. Please share your thoughts by sending an email message to SolutionLab@HitachiVantara.com. To assist the routing of this message, use the paper number in the subject and the title of this white paper in the text.

Revision history

Changes	Date
Initial release	March 2024

Reference Architecture Guide

This paper demonstrates a reference configuration for running Oracle database on Hitachi Advanced Server with Hitachi Virtual Storage Software block (VSS block). It leverages the latest capabilities of data storage integrations to create, protect, and manage on-premises clusters. These are relevant to deployment configurations with Hitachi Advanced Server in a hybrid converged/hyperconverged configuration and VSS block on bare metal configurations.

A key element in a successful deployment is having a robust and flexible infrastructure that meets the wide variety of requirements in a highly dynamic environment. Hitachi Vantara configured Oracle RAC database on compute nodes with Oracle Linux and VSS block as storage nodes that provide highly available and stable block storage infrastructures for database technologies.

This reference architecture also provides the reference design for a build-your-own Oracle RAC cluster on an Oracle Linux operating system environment using Hitachi Virtual Storage Software block. While this reference design is specific to Hitachi VSS block storage solutions, it can easily be used for building your own database with other Hitachi Virtual Storage Platform solutions.

The intended audience of this document is IT administrators, system architects, consultants, and sales engineers to assist in planning, designing, and implementing Hitachi Virtual Storage Software block running Oracle Database.

Oracle RAC cluster overview

Oracle RAC cluster is a powerful, object-relational database system. Oracle database with the RAC option allows multiple servers to access common single shared storage. In our solution we used Virtual Storage Software block (VSS block) as a storage device.

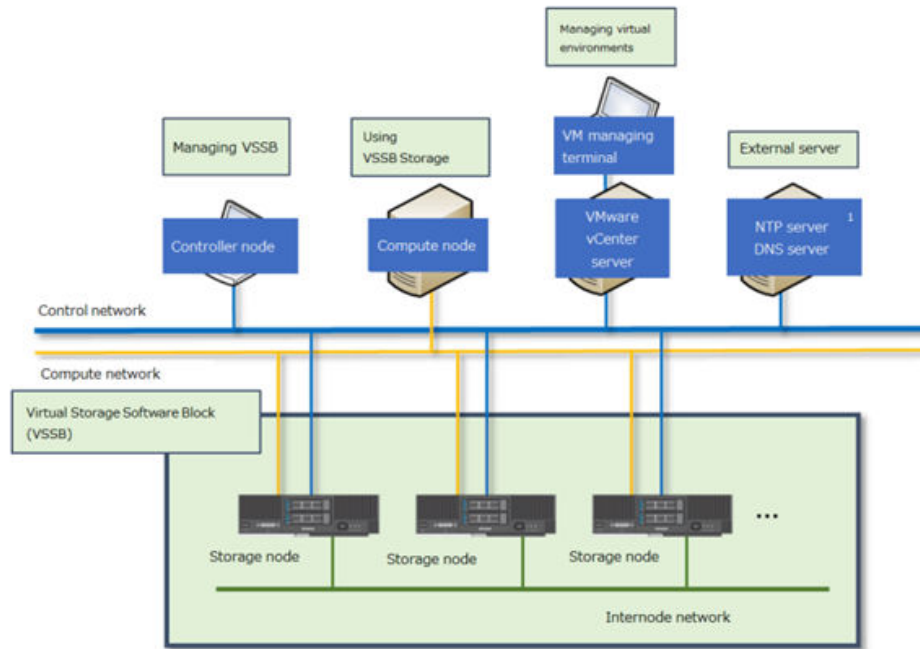
Oracle RAC comes with proven architecture that delivers reliability, high availability, data integrity, scalability, and high-performance features. Oracle RAC also supports the primary/standby architecture where data can be stored at multiple locations depending on geographically separated in case of disaster.

In this reference architecture, we implemented Hitachi Solution for Databases for an Oracle Real Application Clusters database using VSS block storage.

Hitachi Virtual Storage Software block overview

Hitachi Virtual Storage Software block is a storage software product that builds and sets up a virtual storage system from multiple general-purpose servers. The Virtual Storage Software block storage system is two-layer software-defined storage (SDS) in which storage nodes are separated from the compute nodes (disaggregated storage).

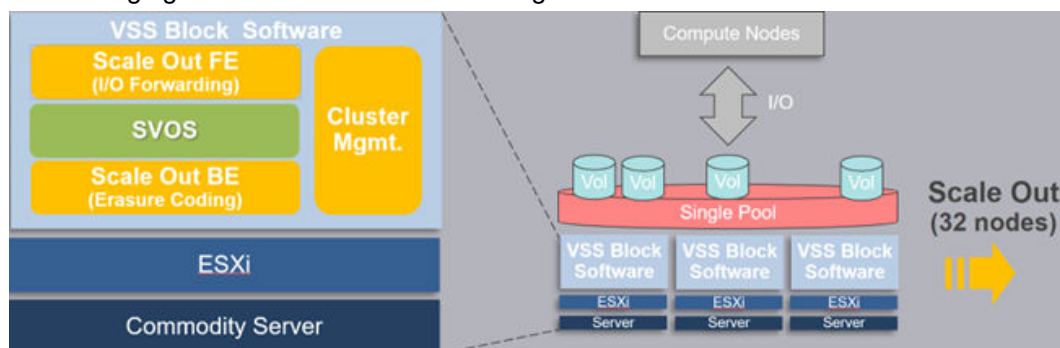
The following figure shows the configuration of Virtual Storage Software block.



1. Storage node and NTP/DNS server can be configured in different subnets.
For another subnet configuration, the NTP/DNS server must be reachable via the Default Gateway.

To maximize return on investment (ROI) and deliver long-term scaling, Hitachi has redesigned the Storage Virtualization Operating System (SVOS RF) to run on Hitachi Advanced Server. By running virtualized, you can integrate your core enterprise storage platform with your existing hypervisor. This gives you the ability to run the latest distributed application alongside traditional business applications on a scale-out infrastructure, allowing you to scale performance and capacity independently to consolidate even more workloads.

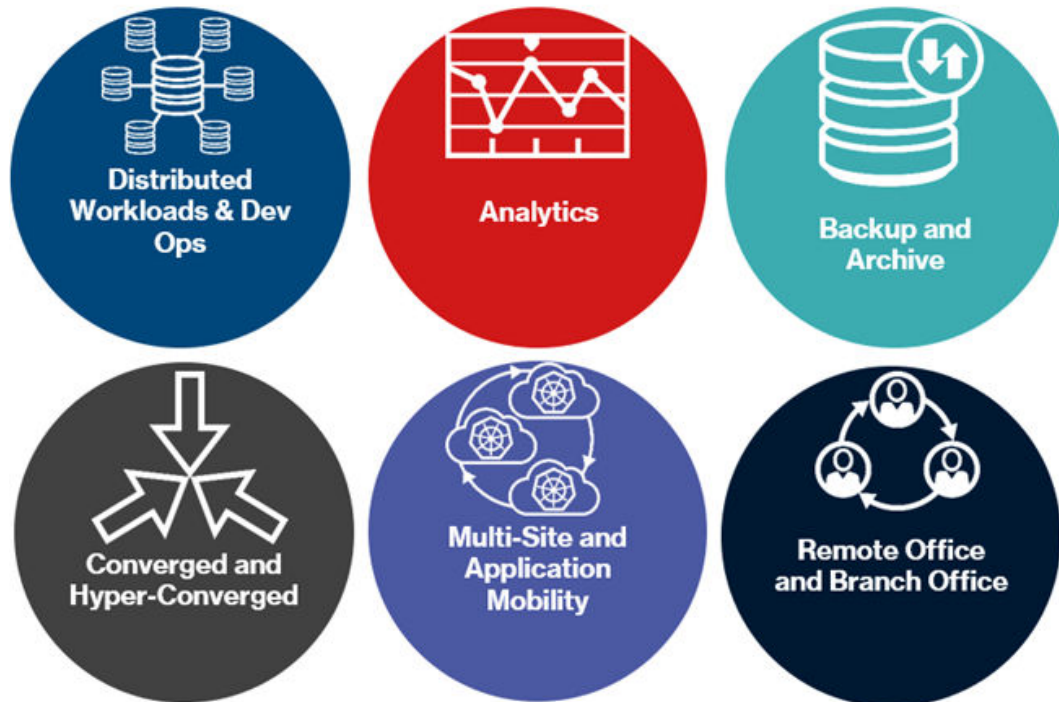
The following figure shows an architectural diagram of VSS block.



The Hitachi SDS offering is a true enterprise solution. With years of research and development, our patented Hitachi Polyphase Erasure Coding brings our legendary reliability into the SDS marketplace. By allowing data to be balanced across the entire cluster, SDS reduces the cost and complexity of management while offering the most resilient solution. Integration with a hypervisor allows you to automate and simplify daily tasks and guarantee uptime.

Use cases

To modernize the digital core, Virtual Storage Software block provides an SDS option along with Virtual Storage Platform (VSP) to drive customer expectations of data mobility for multi-site environments, expanding on Hitachi enterprise storage capabilities to continue to deliver enterprise reliability and performance you can expect from the Hitachi brand.

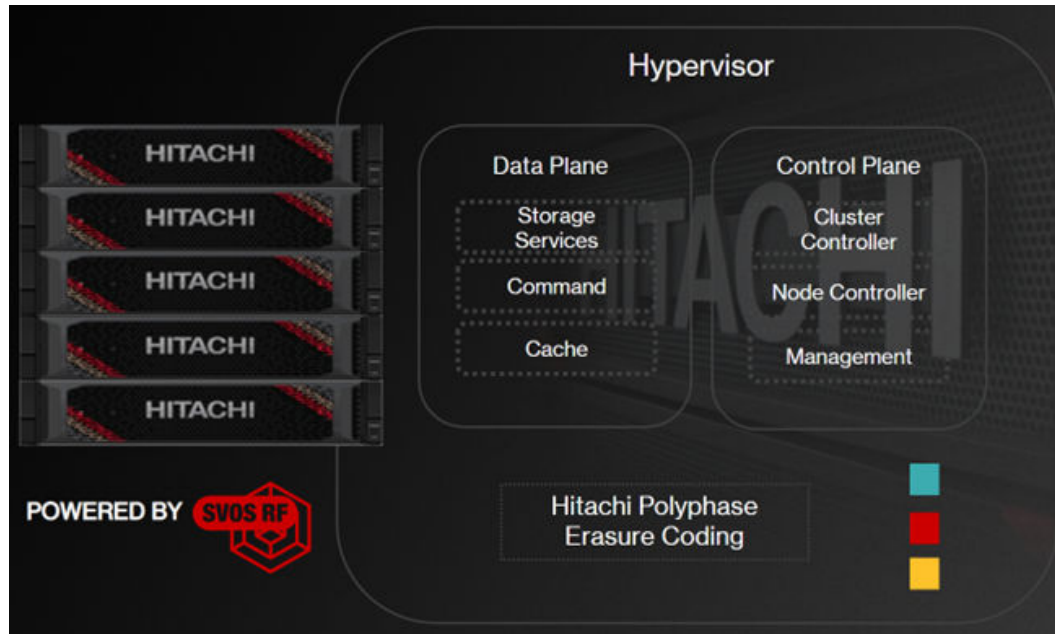


Key features

Traditional storage solutions for distributed containerized workloads such as OpenShift are too complex and have numerous limitations such as a limited number of volumes and automatic provisioning of storage to meet the dynamic needs of containers. The Hitachi SDS solution targets these limitations by abstracting storage software from its hardware and providing key features that enable a cloud-native storage solution tightly integrated into container orchestration frameworks optimized for distributed workloads.

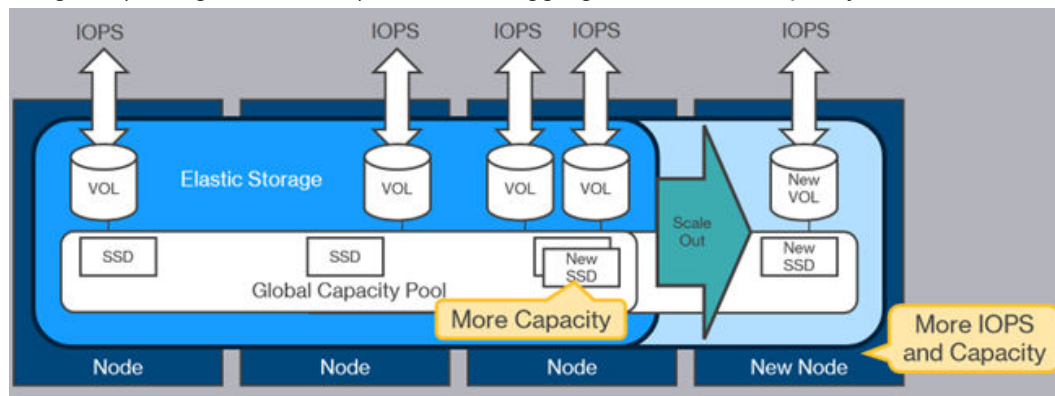
Distributed storage

Running on x86 servers inside your hypervisor and powered by the Hitachi Storage Virtualization Operating System RF (SVOS RF) continues to be on the leading edge of storage virtualization, pools, tiering, deduplication or compression, and storage replication technologies. It demonstrates our ability to evaluate an underlying hardware technology and conceptualize its implementation across different industries before bringing it to market.



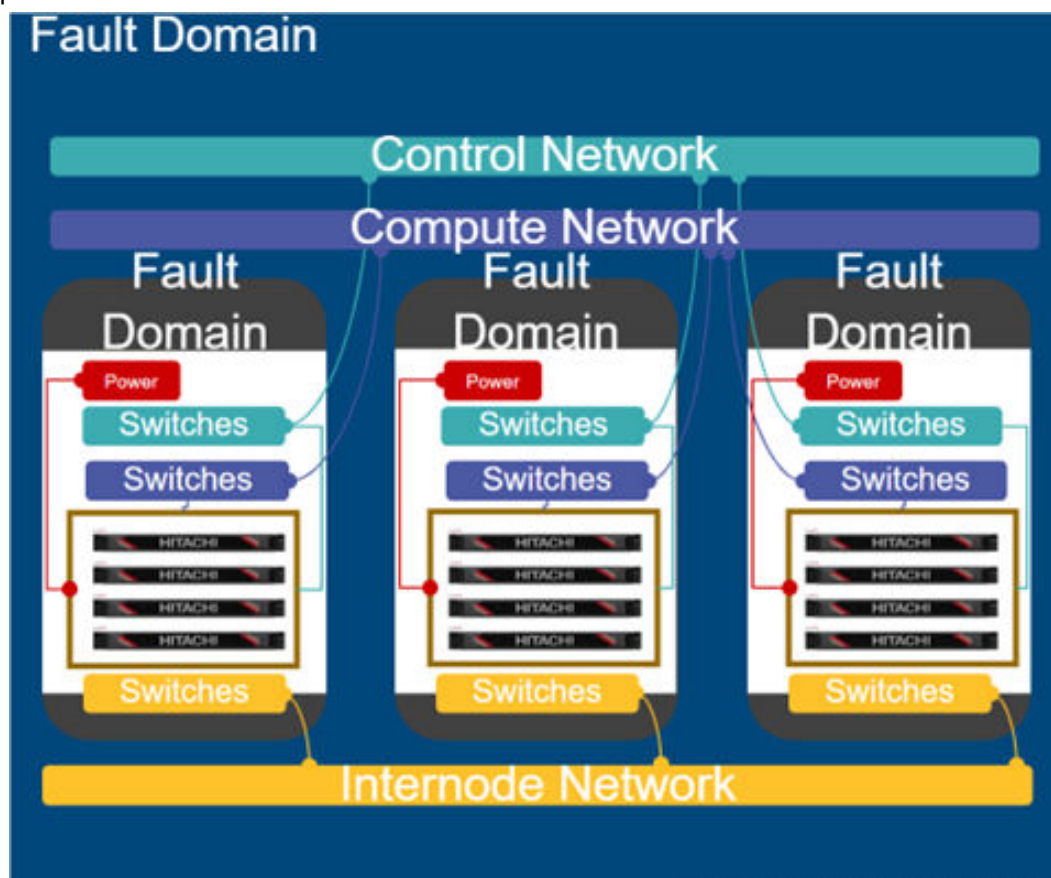
Scale-out by design

Linearly scale performance and capacity starting from 37 TB all the way up to 3.38 PB using a disaggregated (non-HCI) storage cluster. Resources can be added to an existing cluster by scaling out (adding more nodes) to increase aggregate IOPS and capacity.



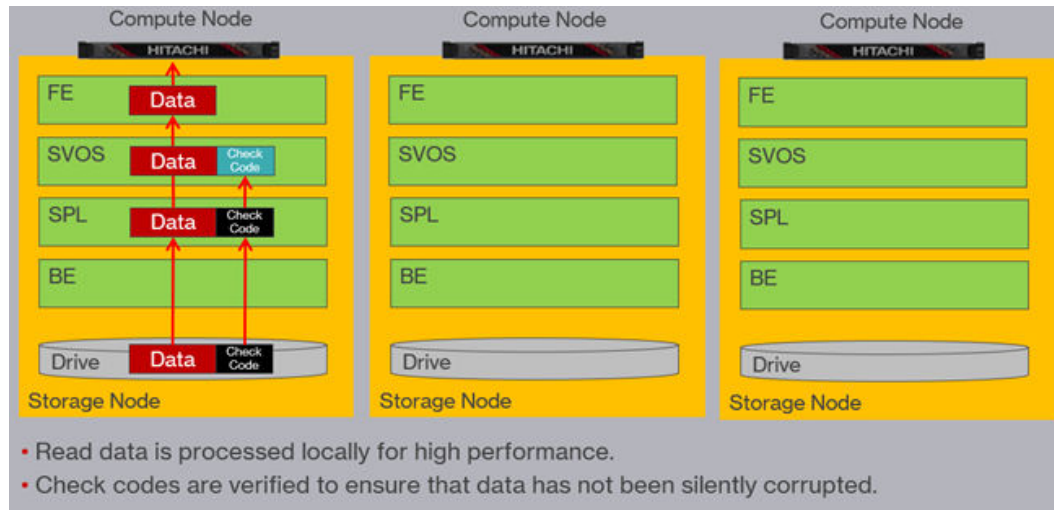
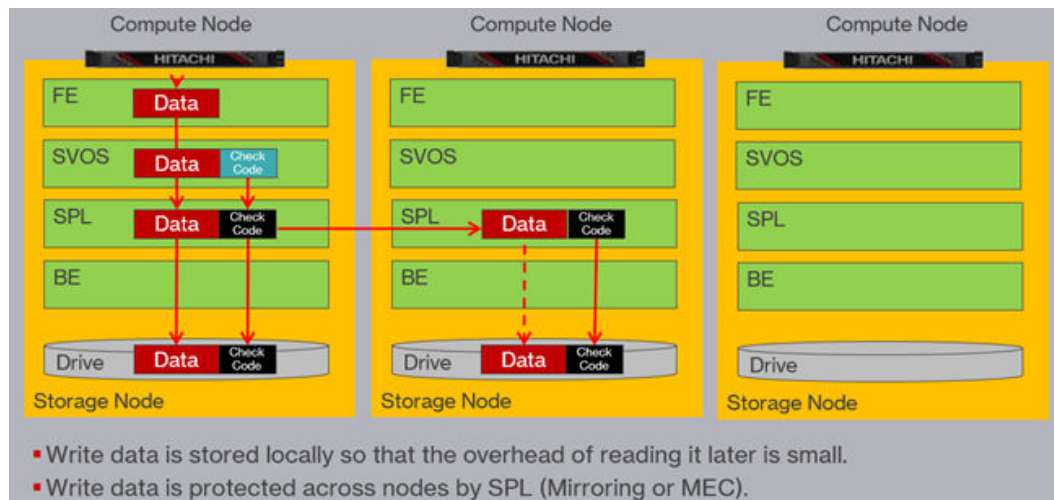
Fault domains

Fault domains are a group of storage nodes that share hardware such as the same power supply and network switches. Having fault domains enables storage systems to continue operations if a failure occurs in one of the fault domains.



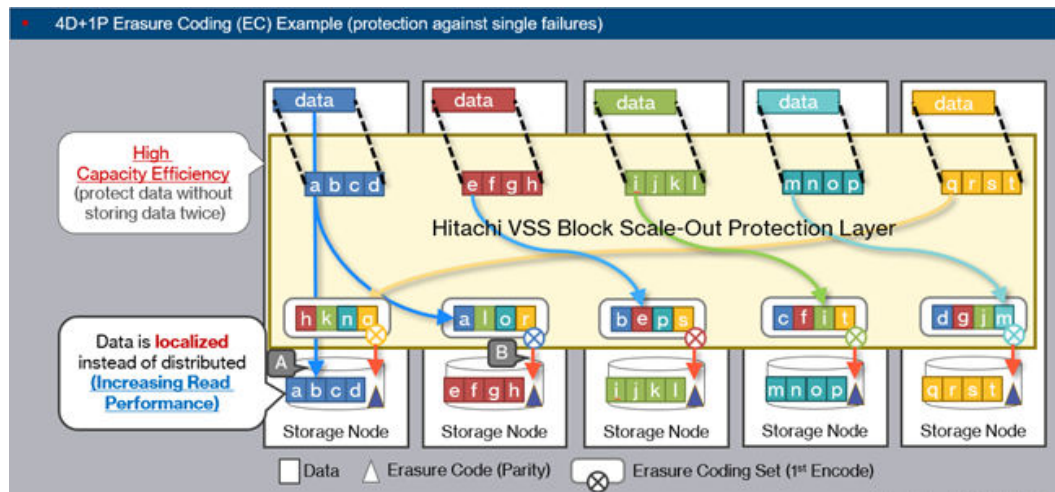
I/O flow read/write optimization

Hitachi VSS block is designed to take full advantage of data localization to improve performance for both reads and writes.

I/O read optimization**I/O write optimization****Flexible Data Protection**

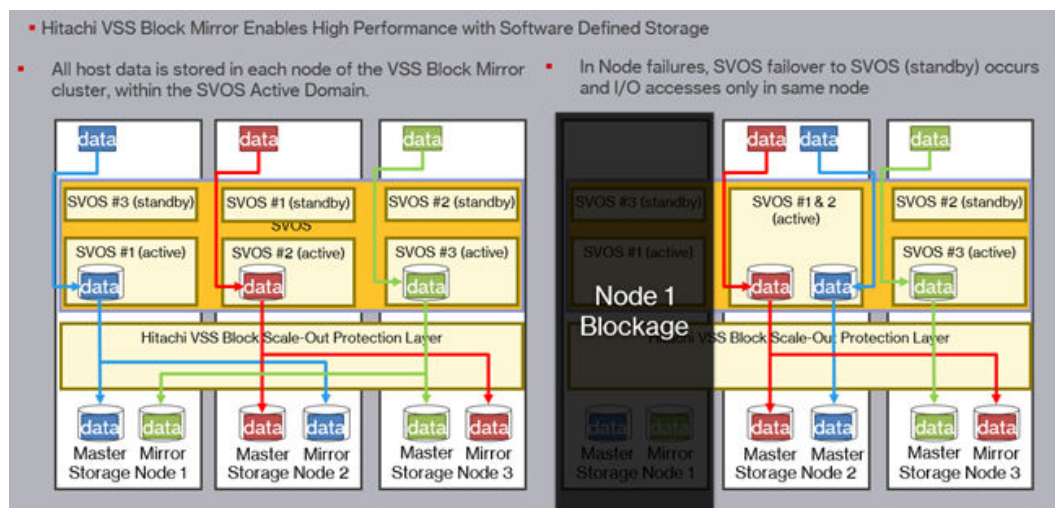
Hitachi Polyphase Erasure Coding (HPEC) features a different approach to adapting software data protection. Hitachi carefully evaluated erasure coding through years of research and developed a unique patented approach (US# 10,185,624 B2 and 10,496,479 B2). This approach improved crucial elements of data and parity placement, which efficiently improved latency within Hitachi Polyphase Erasure Coding, accelerated read performance, and lowered storage resource overhead. These necessary enhancements provide a place where your data can be synthesized so you can create more value, while increasing the reliability of a Hitachi-based software-defined storage solution.

HPEC can be configured in mirror disk configuration, which protects against a single node failure.



Mirroring

Mirroring mode is also available where all data is stored multiple times resulting in low-capacity efficiency but high read and write performance, which is ideal when performance matters.

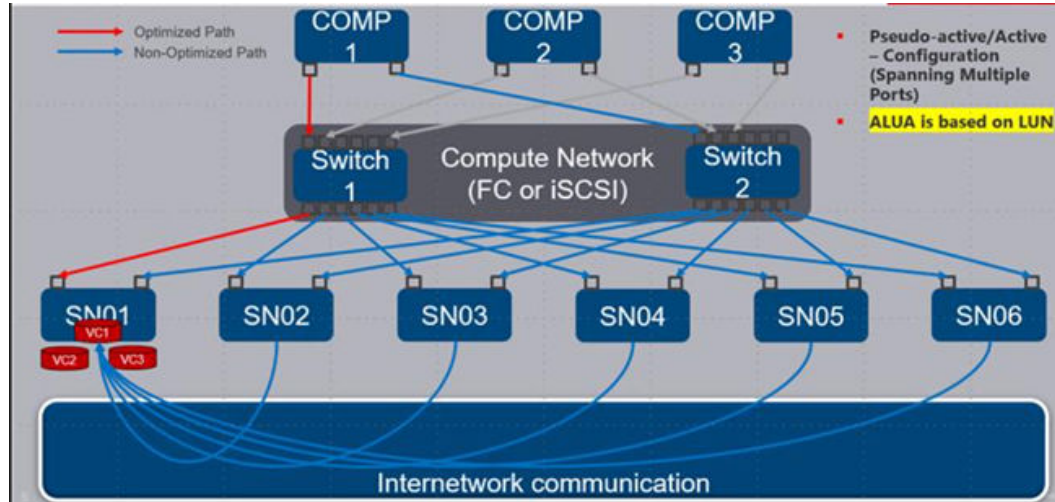


Standard front-end protocols and multipathing

Proprietary host drivers are not required because standard 32 Gbps Fibre Channel and 10/25 Gbps iSCSI protocols are natively supported. OS Native Multipath (Device Mapper, MPIO) is supported.

Hitachi Polyphase Asymmetric Logical Unit Access (ALUA)

Using Pseudo Active/Active configuration, I/O can be sent to ports on any storage processor that defines the ownership of the LUN and therefore its optimized path. Non-owner processors can send and receive I/O to a given LUN as well. However, traffic must first be sent from the optimized path. In addition, true Pseudo Active/Active I/O can also be sent from a back-end channel, which in our case can be the VSS block internode network.



Compute host I/O forwarding

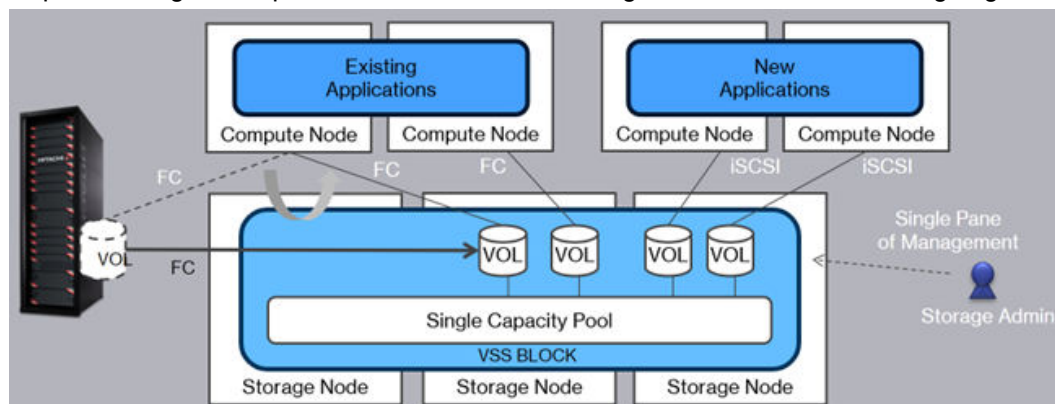
I/O forwarding is available so compute hosts can perform I/O operations to non-priority paths in case of failures.

Reserve rebuild capacity

The policy for fixed rebuild capacity can be specified so rebuilds are always possible and are not being used for user data such as with a variable policy. Also, the number of drives for rebuild capacity can be specified. For example, in a mirrored configuration with three nodes, four disks per node, and the fixed rebuild policy set to two drives, two drive failures can be sustained before the node is lost.

Seamless data migration across Hitachi storage systems

With HPEC-enabled minimally disruptive migration from existing Hitachi storage systems using Fibre Channel, copying data is offloaded to the storage layer so compute workloads are unimpeded. Single data paths can be cut over ensuring I/O continues even during migrations.



Auto-recovery

By default, storage nodes perform self-diagnosis and auto-recovery from a server failure caused by software components (such as firmware or drivers) or by a temporary failure of the internode network.

Storage-ready nodes

Hitachi VSS block ready nodes are a quick, simple to deploy, and reliable way of building your new distributed storage platform. They combine pre-configured and validated Hitachi HA servers powered by Intel® Xeon® processors, and the software-defined storage ability of VSS block. Remove the guesswork of building an SDS infrastructure and accelerate your deployment of new applications.

Hitachi Advanced Server HA810

Purpose-built for better performance, high density and power efficiency, this general-purpose dual-processor server delivers a balance of compute and storage capacity with the flexibility to power a wide range of solutions and applications.



HA810-A

▪ Single node, 1U ▪ Dual Socket, up to 10 cores/CPU ▪ Up to 128GB RAM ▪ iSCSI = 2x 10/25GbE dual port Network card ▪ FC= 1 X 10/25GbE dual port Network card & 1 X 32 dual port HBA ▪ 2x 100-240v 1600W PSU ▪ Integrated Hitachi Remote Ops Monitoring	▪ Min 3 nodes max 32 in cluster ▪ Max Logical Capacity ▪ Mirror : 24 TB / Node ▪ HPEC(4D+1P) : 39 TB / Node ▪ HPEC(4D+2P) : 32 TB / Node ▪ Drives Supported ▪ 1.6 TB x 2 SFF SSD (Boot) ▪ 6.4 TB x4 (up to 8) SFF SSD (Storage)
--	--

Hitachi Advanced Server HA820

Adaptable for diverse workloads and environments, the secure 2U HA820 delivers world-class performance with the right balance of expandability and scalability. Designed for supreme versatility and resiliency while being backed by a comprehensive warranty makes it ideal for multiple environments from containers to cloud to big data.



HA820-A

▪ Single node, 2U ▪ Dual Socket, up to 10 cores/CPU ▪ Up to 128GB RAM ▪ iSCSI = 2 or 4x10/25GbE dual port Network card ▪ FC= 1 or 2X10/25GbE dual port Network card & 1 X 32 dual port HBA ▪ 2x 100-240v 1600W PSU ▪ Integrated Hitachi Remote Ops Monitoring	▪ Min 3 nodes max 32 in cluster ▪ Max Logical Capacity ▪ Mirror : 74 TB / Node ▪ HPEC(4D+1P) : 105 TB / Node ▪ HPEC(4D+2P) : 99 TB / Node ▪ Drives Supported ▪ 1.6 TB x 2 SFF SSD (Boot) ▪ 6.4 TB x4 (up to 24) SFF SSD (Storage) ▪ 1.6 TB x4 (up to 24) SFF SSD (Storage)
---	--

Hitachi Advanced Server HA820 G2

HA820-G2-A Compression Ready



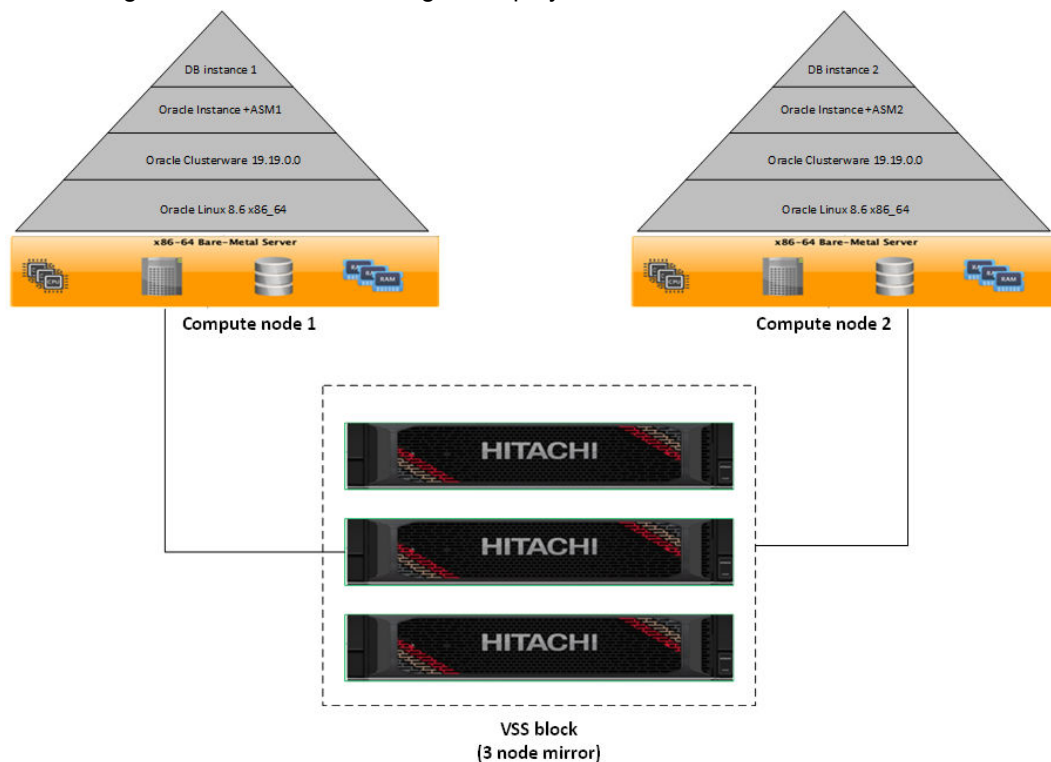
▪ Single node, 2U ▪ Dual Socket, up to 8 or 24 cores/CPU ▪ Up to 256GB RAM ▪ iSCSI = 2 or 4x10/25GbE dual port Network card ▪ FC= 1 or 2X10/25GbE dual port Network card & 1 X 32 dual port HBA ▪ 2x 100-240v 1600W PSU ▪ Integrated Hitachi Remote Ops Monitoring	▪ Min 3 nodes max 32 in cluster ▪ Max Logical Capacity ▪ Mirror : 74 TB / Node ▪ HPEC(4D+1P) : 105 TB / Node ▪ HPEC(4D+2P) : 99 TB / Node ▪ Drives Supported ▪ 1.9 TB x 2 SFF SSD (Boot) ▪ 1.6 TB x4 (up to 24) SFF SSD (Storage) ▪ 3.2 TB x4 (up to 24) SFF SSD (Storage) ▪ 6.4 TB x4 (up to 24) SFF SSD (Storage)
--	--

Solution overview

This solution uses 3 node mirror configurations of VSS block storage to allocate block storage on 2 compute nodes that are configured as Oracle RAC clusters to validate Hitachi Solution for Oracle RAC Database.

This solution also offloads ASM disk mirroring from compute nodes to storage nodes.

The following illustration shows the logical deployment architecture.



Solution components

These are the key hardware and software components used for this environment.

Hardware components

The following tables list the versions of hardware and software tested in this reference architecture.

The solution tested used specific features based on the following hardware. You can use either Hitachi Advanced Server DS120 G1 or G2, DS220 G1 or G2, DS225, DS240, HA810 G1 or G2, HA820 G1 or G2, or any qualified server platform.

Vendor	Hardware	Detail Description	Version	Quantity
Hitachi Vantara	HA820 server (VSS block storage nodes)	2 × Intel Xeon Gold 6226 12-core, 2.70GHz processors 12 × 16 GB DIMMs, 192 GB memory 2 × 1.6 TB SAS SSDs for boot 12 × 800 GB SAS SSDs	iLO: 2.55 BIOS: U30 v2.54	3
Hitachi Vantara	DS220 G2 server	2 × Intel(R) Xeon(R) Platinum 8368 CPU @ 2.40GHz, 76 Core(s), 152 Logical Processor(s) 32 (32 GB * 8) DIMM DDR4 Synchronous Registered (Buffered) 3200 MHz 1 × 256 GB NVMe SSD for boot	Firmware: 3.64.06 BIOS: S5XH3A19.H00	2
Cisco	Nexus 93180YC-FX3	48 × 10/25 GbE 6 × 40/100 Gbps	- BIOS: 07.65 - NXOS 9.3.5	2
Cisco	Nexus C92348GC-X	1 GE 48-Port Gb Ethernet Cisco switch	NXOS 9.3(3)	1

Software components

The following table lists the key software components.

Software	Version	Function
Compute Nodes		
Oracle RAC	19.19	Oracle Grid Software on compute nodes
Oracle Linux	8.6 UEK6	Operating system for compute nodes
Storage Nodes		
Hitachi Storage Virtualization Operating System RF	8.x	Operating System
Hitachi Virtual Storage Software block	1.12	VSS block software

Solution design

This is a detailed solution example of Hitachi Advanced Server, Hitachi Virtual Storage Software block, and clusters set up to run Oracle RAC database 19c.

Solution infrastructure components

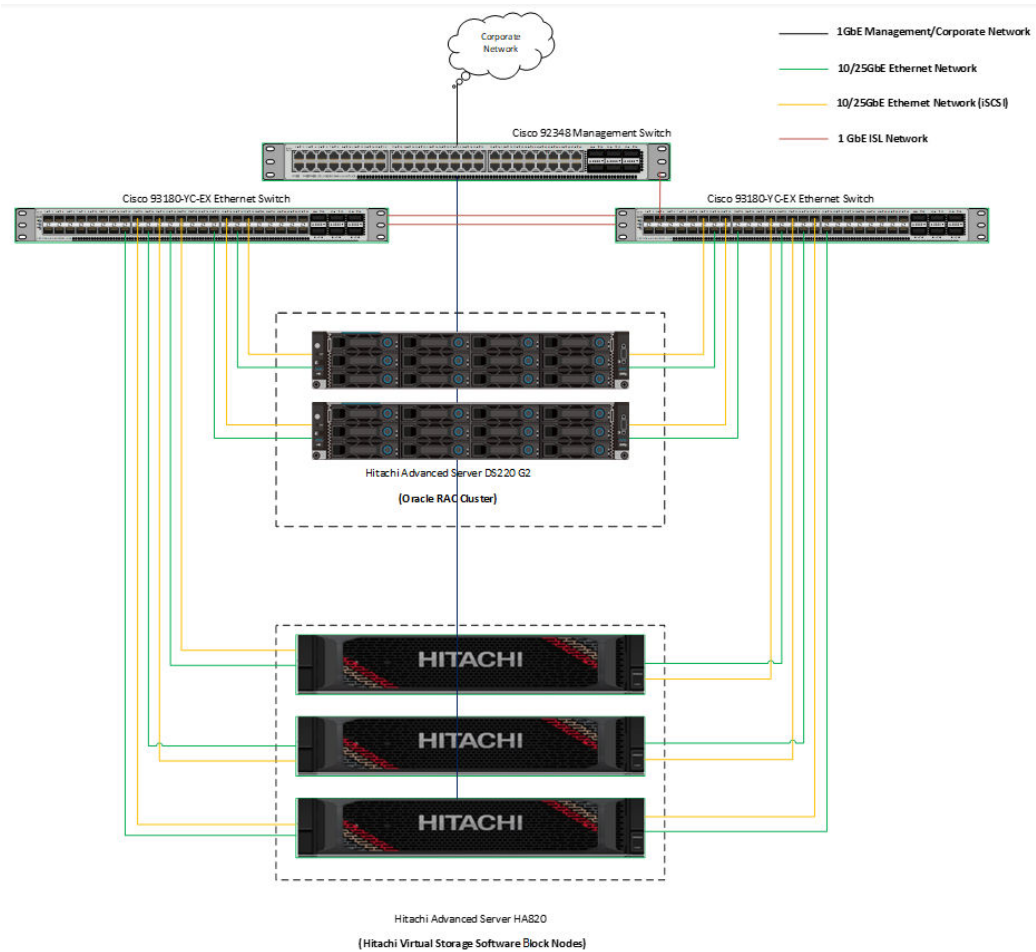
The following hardware components are used for testing:

- Compute nodes
 - 2 × Hitachi Advanced Server DS220 G2 for cluster compute nodes
 - 1 × Cisco 93180YC-FX Ethernet switch
 - 1 × Cisco Nexus C92348GC-X management switch
- Storage nodes
 - 3 × Hitachi Advanced Server HA820 G2 for VSS block storage nodes
 - 1 × Cisco Nexus 93180YC-FX Ethernet switch
 - 1 × Cisco Nexus C92348GC-X management switch

The following figure shows the high-level physical architecture diagram of this architecture.



Note: An existing VSS block 3-node cluster was used for testing purposes. The compute nodes and VSS block storage nodes are in different racks. For simplification, the diagram shows direct connectivity to the same ToR switch.



Note: The network layout shown was used for testing purposes only. For a production environment and complete configuration of VSS block required and the recommended network components are as follows:

- Four Cisco 93180YC-FX Ethernet switches
- Two Cisco 92348GC-X management switches

Storage configuration

Two Cisco 92348GC-X management switches are used for the storage configuration.

Log in at the following URL: <https://10.xx.xx.xx/hsds/#/login>.

Name	IP Address
HA820 VSS block storage Node 1	Control: 10.76.47.51 Compute: 192.168.111.51 Internode: 192.168.110.51
HA8200 VSS block storage Node 2	Control: 10.76.47.52 Compute: 192.168.111.52 Internode: 192.168.110.52
HA820 VSS block storage Node 3	Control: 10.76.47.53 Compute: 192.168.111.53 Internode: 192.168.110.53

The solution uses an iSCSI network for SAN connectivity for compute. Volumes are provisioned from VSS block storage.

ASM disk group	Total number of volumes	Volume sizes (GB)	Purpose
DATA	1	200 GB	OLTP Application Tablespaces, System Sysaux, Undo, Temp
REDO	2	15 GB	Online Redo Logs and Control Files
OCR	3	10 GB	Oracle Cluster Registry and Voting Disk

Network configuration

The following table lists allocated IP addresses.

Name	IP Address	Notes
Oracle RAC Node 1	BMC: 10.76.xx.xx OS: 10.76.xx.xx Private: 192.168.xx.xx iSCSI: 192.168.111.xxx	Configure virtual IP and SCAN-IP as per standard grid configuration.
Oracle RAC Node 2	BMC: 10.76.xx.xx OS: 10.76.xx.xx Private: 192.168.xx.xx	Configure virtual IP and SCAN-IP as per standard grid configuration.

Name	IP Address	Notes
	iSCSI: 192.168.111.xx	
Cisco 9k Switch	10.76.xx.xx	-

Solution validation

The following high-level activities must be completed to validate this solution to install Oracle Grid Software on compute nodes and VSS block for storage nodes, which is already configured.

- Add compute nodes on VSS block storage.
- Configure multipathing and discover storage on compute nodes.
- Validate test cases.

Add compute nodes to VSS block

To provision VSS block storage, compute nodes need to be added to VSS block.

Note the iSCSI Initiator Name for each compute node can be found in the `/etc/iscsi/initiatorname.iscsi` file as shown.

```
[root@ds220g276 ~]# cat /etc/iscsi/initiatorname.iscsi
InitiatorName=ign.1988-12.com.oracle:f616752b9c82
[root@ds220g276 ~]#
```

Verify that the `iscsid` service is running on each compute node.

```
[root@ds220g275 ~]# systemctl status iscsid
● iscsid.service - Open-iSCSI
   Loaded: loaded (/usr/lib/systemd/system/iscsid.service; enabled; vendor preset: disabled)
   Active: active (running) since Thu 2024-02-08 05:05:02 EST; 1 day 2h ago
     Docs: man:iscsid(8)
```

Use the identified iSCSI Initiator Name to add the compute node to VSS block Storage. Log in to the VSS block web UI and select Register Compute Node.

URL : <https://10.xx.xx.xx/hsds/#/login>

Login : admin/*****

Register Compute Node

COMPUTE NODE NAME
pgsql-ocpmaster1

OS TYPE
Linux

WWNS (Example: 0000000100000001)
+ Add WWN

ISCSI INITIATOR NAMES (Example: iqn.1992-01.com.company:server)
iqn.1994-05.com.redhat:15c569d0659
+ Add iSCSI Initiator Name

Cancel + Submit and add another compute node Submit

After adding all the compute nodes, your compute nodes should look like the following.

SELECT	NAME	VIRTUAL PRIVATE STORAGE	OS TYPE	PORT CONNECTIONS	TOTAL	USED	ID
	ds220g275.unfnet.local	(system)	LINUX	3	260.00 GiB	140.13 GiB	9eb075b0e-4b36-40b4-b...
	ds220g276.unfnet.local	(system)	LINUX	3	260.00 GiB	140.13 GiB	c79e60c4-43e3-4f28-a...

Configure multipathing and discover storage on compute nodes

After compute nodes have been added to VSS block storage, we must discover VSS block storage from compute nodes so that any volumes created on VSS block are visible on compute nodes.

On compute nodes, do following.

Procedure

1. Enable multipathing.

```
[root@ds220g275 ~]# vi /etc/multipath.conf
defaults {
    user_friendly_names yes
    find_multipaths yes
}
devices {
    device {
        vendor HITACHI
        product Hi-SDS
        path_grouping_policy group_by_prio
        prio alua
        path_checker readsector0
        no_path_retry 6
        dev_loss_tmo infinity
    }
}

#systemctl status multipathd
#systemctl start multipathd
#systemctl enable multipathd
```

2. Locate the IP addresses of VSS block storage nodes, for example 192.168.111.51/52/53, The default port is 3260.

```
[root@ds220g275 ~]# iscsiadm -m discovery -t sendtargets -p 192.168.111.51:3260
192.168.111.51:3260,0 iqn.1994-04.jp.co.hitachi:rsd.sph.t.14b07.000
[root@ds220g275 ~]#
[root@ds220g275 ~]# iscsiadm -m discovery -t sendtargets -p 192.168.111.52:3260
192.168.111.52:3260,1 iqn.1994-04.jp.co.hitachi:rsd.sph.t.14b07.001
[root@ds220g275 ~]#
[root@ds220g275 ~]# iscsiadm -m discovery -t sendtargets -p 192.168.111.53:3260
192.168.111.53:3260,2 iqn.1994-04.jp.co.hitachi:rsd.sph.t.14b07.002
[root@ds220g275 ~]#
```

3. Verify VSS block storage node discovery on compute nodes.

- a. Go to the `/var/lib/iscsi/nodes` directory. Three files will be created with the `iqn` number of storage nodes that we added in the previous step.

```
[root@ds220g275 ~]# cd /var/lib/iscsi/nodes
[root@ds220g275 nodes]#
[root@ds220g275 nodes]# ls -lrt
total 0
drw-----. 3 root root 35 Jan 25 03:55 iqn.1994-
04.jp.co.hitachi:rsd.sph.t.14b07.000
drw-----. 3 root root 35 Jan 25 03:57 iqn.1994-
04.jp.co.hitachi:rsd.sph.t.14b07.001
drw-----. 3 root root 35 Jan 25 03:57 iqn.1994-
04.jp.co.hitachi:rsd.sph.t.14b07.002
[root@ds220g275 nodes]#
```

- b. Log in to the `iqn` numbers of the VSS block storage node.

```
[root@ds220g275 ~]# iscsiadm -m node -T iqn.1994-
04.jp.co.hitachi:rsd.sph.t.14b07.000 -l
Logging in to [iface: default, target: iqn.1994-
04.jp.co.hitachi:rsd.sph.t.14b07.000, portal: 192.168.111.51,3260]
Login to [iface: default, target: iqn.1994-
04.jp.co.hitachi:rsd.sph.t.14b07.000, portal: 192.168.111.51,3260]
successful.
[root@ds220g275 ~]#
[root@ds220g275 ~]# iscsiadm -m node -T iqn.1994-
04.jp.co.hitachi:rsd.sph.t.14b07.001 -l
Logging in to [iface: default, target: iqn.1994-
04.jp.co.hitachi:rsd.sph.t.14b07.001, portal: 192.168.111.52,3260]
Login to [iface: default, target: iqn.1994-
04.jp.co.hitachi:rsd.sph.t.14b07.001, portal: 192.168.111.52,3260]
successful.
[root@ds220g275 ~]#
[root@ds220g275 ~]# iscsiadm -m node -T iqn.1994-
04.jp.co.hitachi:rsd.sph.t.14b07.002 -l
Logging in to [iface: default, target: iqn.1994-
04.jp.co.hitachi:rsd.sph.t.14b07.002, portal: 192.168.111.53,3260]
Login to [iface: default, target: iqn.1994-
04.jp.co.hitachi:rsd.sph.t.14b07.002, portal: 192.168.111.53,3260]
successful.
[root@ds220g275 ~]#

[root@ds220g275 ~]# iscsiadm -m session
tcp: [1] 192.168.111.51:3260,0 iqn.1994-
04.jp.co.hitachi:rsd.sph.t.14b07.000 (non-flash)
tcp: [2] 192.168.111.52:3260,1 iqn.1994-
04.jp.co.hitachi:rsd.sph.t.14b07.001 (non-flash)
tcp: [3] 192.168.111.53:3260,2 iqn.1994-
```

```
04.jp.co.hitachi:rsd.sph.t.14b07.002 (non-flash)
[root@ds220g275 ]#
```

Here we can see iSCSI sessions that were created on compute nodes that are connected to VSS block storage.

c. Create a volume on VSS block, for example OCR_11.

SELECT	NAME	VIRTUAL PRIVATE S.	STATUS SUMMARY	STATUS	SNAPSHOT STATUS	TOTAL	USED	CAPACT.	NUMBE.	SNAPSHOTS	VOLUME TYPE	RICENAME	ID
+													
☐	DATA1	(system)	Normal	Normal	-	200.00 GiB	144.62 GiB	Disab...	2	Not exist	Normal	Ora_DATA1	56ac1...
☒	OCR_11	(system)	Normal	Normal	-	15.00 GiB	-	Disab...	0	Not exist	Normal	OCR_11	1214...

d. Rescan the volume from a compute node and verify the LUN.

```
[root@ds220g275 ]# /usr/bin/rescan-scsi-bus.sh
Scanning SCSI subsystem for new devices
Scanning host 0 for SCSI target IDs 0 1 2 3 4 5 6 7, all LUNs
Scanning for device 23 0 0 1 ...
OLD: Host: scsi23 Channel: 00 Id: 00 Lun: 01
      Vendor: HITACHI   Model: Hi-SDS           Rev: 8401
      Type:   Direct-Access           ANSI SCSI revision: 05
3 new or changed device(s) found.
      [21:0:0:0]
      [22:0:0:0]
      [23:0:0:0]
0 remapped or resized device(s) found.
0 device(s) removed.
[root@ds220g275 ]#

[root@ds220g275 nodes]# fdisk -l

Disk /dev/sdc: 15 GiB, 16106127360 bytes, 31457280 sectors
Units: sectors of 1 * 512 = 512 bytes
Sector size (logical/physical): 512 bytes / 512 bytes
I/O size (minimum/optimal): 65536 bytes / 65536 bytes
[root@ds220g275 nodes]#
```

The following shows that we configured multipathing with the LUN showing three paths to VSS block.

```
[root@ds220g275 nodes]# multipath -ll
mpatha (360060e8114b070006014b07000000004) dm-6 HITACHI,Hi-SDS
size=15G features='1 queue_if_no_path' hwhandler='1 alua' wp=rw
|-+- policy='service-time 0' prio=50 status=active
|  '- 23:0:0:1 sde      8:64  active ready running
`-+- policy='service-time 0' prio=10 status=enabled
|  21:0:0:1 sdc       8:32  active ready running
|  22:0:0:1 sdd       8:48  active ready running
[root@ds220g275 nodes]#

[root@ds220g275 ~]# iscsiadm -m session --rescan
```

```
Rescanning session [sid: 1, target: iqn.1994-04.jp.co.hitachi:rsd.sph.t.14b07.000, portal: 192.168.111.51,3260]
Rescanning session [sid: 2, target: iqn.1994-04.jp.co.hitachi:rsd.sph.t.14b07.001, portal: 192.168.111.52,3260]
Rescanning session [sid: 3, target: iqn.1994-04.jp.co.hitachi:rsd.sph.t.14b07.002, portal: 192.168.111.53,3260]
[root@ds220g275 ~]#
```

Next steps

Create additional volumes as needed. We configured volumes for OCR, DATA, TEMP, REDO, and Disk to set up Oracle RAC. Follow standard steps to install Oracle RAC clusters referring to Oracle documentation.



Note: Follow these steps for all compute nodes to discover VSS block volumes.

Install Oracle RAC software on compute nodes

Oracle cluster software installation is not in the scope of this document. Follow the standard steps from Oracle support.

Solution test cases

In this solution we performed several tests to verify the functionality and stability of the cluster.

Functionality tests

This section provides details of the functional test details that were used for validation.

Creating required DB volumes on VSS block storage and allocating to Oracle RAC nodes over iSCSI

In this test case we have created some volumes (OCR/DATA/REDO, etc.) to install Oracle cluster software and database.

The screenshot shows the VSP One SDS Block Administrator interface. The top navigation bar includes tabs for EVENT, MONITOR, PROTECTION DOMAINS, STORAGE NODES, DRIVES, COMPUTE NODES, VOLUMES, STORAGE POOLS, and PORTS. The 'VOLUMES' tab is active. Below the navigation bar, there is a filter section with a 'Select All' button and a '0 selected' indicator. A table of volumes is displayed with columns: SELECT, NAME, VIRTUAL PRIVAT..., STATUS SUMMAR..., STATUS, SNAPSHOT STAT..., TOTAL, USED, CAPA..., NUMB..., SNAPSHOTS, VOLUME TYPE, NICKNAME, and ID. The table contains 8 rows of data, each representing a volume named DATA0 through DATA5. Each volume is 200.00 GiB in total size and 146.63 GiB in used space. The status is 'Normal' and the volume type is 'Normal'.

SELECT	NAME	VIRTUAL PRIVAT...	STATUS SUMMAR...	STATUS	SNAPSHOT STAT...	TOTAL	USED	CAPA...	NUMB...	SNAPSHOTS	VOLUME TYPE	NICKNAME	ID
☐	DATA0	(system)	Normal	Normal	-	200.00 GiB	146.63 GiB	Disa...	2	Not exist	Normal	ORA_DATA0	ca51...
☐	DATA1	(system)	Normal	Normal	-	200.00 GiB	146.63 GiB	Disa...	2	Not exist	Normal	ORA_DATA1	874...
☐	DATA10	(system)	Normal	Normal	-	200.00 GiB	146.63 GiB	Disa...	2	Not exist	Normal	ORA_DATA10	62a...
☐	DATA11	(system)	Normal	Normal	-	200.00 GiB	146.67 GiB	Disa...	2	Not exist	Normal	ORA_DATA11	a88...
☐	DATA2	(system)	Normal	Normal	-	200.00 GiB	146.63 GiB	Disa...	2	Not exist	Normal	ORA_DATA2	781...
☐	DATA3	(system)	Normal	Normal	-	200.00 GiB	146.63 GiB	Disa...	2	Not exist	Normal	ORA_DATA3	d9d...
☐	DATA4	(system)	Normal	Normal	-	200.00 GiB	146.63 GiB	Disa...	2	Not exist	Normal	ORA_DATA4	0d2...
☐	DATA5	(system)	Normal	Normal	-	200.00 GiB	146.63 GiB	Disa...	2	Not exist	Normal	ORA_DATA5	561...

```
[root@ds220g275 ]# /usr/bin/rescan-scsi-bus.sh
Scanning SCSI subsystem for new devices
Scanning host 0 for SCSI target IDs 0 1 2 3 4 5 6 7, all LUNs
Scanning for device 0 1 123 0 ...
OLD: Host: scsi0 Channel: 01 Id: 123 Lun: 00
new or changed device(s) found.
    [21:0:0:0]
    [22:0:0:0]
    [23:0:0:0]
0 remapped or resized device(s) found.
0 device(s) removed.
```

From the following images we can see LUNs created on VSS block, and because of multipathing, we can see both compute nodes have 3 paths to each LUN.

```
[root@sd220g275 ~]# multipath -ll
REDO02 (360060e8114b070006014b0700000001e) dm-22 HITACHI,Hi-SDS
size=20G features='1 queue_if_no_path' hwhandler='1 alua' wp=rw
|-+- policy='service-time 0' prio=50 status=active
| '- 22:0:0:16 sdat 66:208 active ready running
|-+- policy='service-time 0' prio=10 status=enabled
| '- 21:0:0:16 sdaz 67:48 active ready running
| '- 23:0:0:16 sdbg 67:160 active ready running
DATA01 (360060e8114b070006014b07000000009) dm-12 HITACHI,Hi-SDS
size=200G features='1 queue_if_no_path' hwhandler='1 alua' wp=rw
|-+- policy='service-time 0' prio=50 status=active
| '- 23:0:0:4 sdv 65:80 active ready running
|-+- policy='service-time 0' prio=10 status=enabled
| '- 21:0:0:4 sdo 8:224 active ready running
| '- 22:0:0:4 sdl 8:176 active ready running
REDO01 (360060e8114b070006014b0700000001d) dm-21 HITACHI,Hi-SDS
size=20G features='1 queue_if_no_path' hwhandler='1 alua' wp=rw
|-+- policy='service-time 0' prio=50 status=active
| '- 22:0:0:15 sdap 66:144 active ready running
|-+- policy='service-time 0' prio=10 status=enabled
| '- 21:0:0:15 sdau 66:224 active ready running
| '- 23:0:0:15 sdbe 67:128 active ready running
DATA00 (360060e8114b070006014b07000000008) dm-8 HITACHI,Hi-SDS
size=200G features='1 queue_if_no_path' hwhandler='1 alua' wp=rw
|-+- policy='service-time 0' prio=50 status=active
| '- 23:0:0:3 sds 65:32 active ready running
|-+- policy='service-time 0' prio=10 status=enabled
| '- 22:0:0:3 sdj 8:144 active ready running
| '- 21:0:0:3 sdk 8:160 active ready running
TEMP01 (360060e8114b070006014b07000000019) dm-26 HITACHI,Hi-SDS
size=210G features='1 queue_if_no_path' hwhandler='1 alua' wp=rw
```

```
[root@sd220g276 ~]# multipath -ll
REDO02 (360060e8114b070006014b0700000001e) dm-23 HITACHI,Hi-SDS
size=20G features='1 queue_if_no_path' hwhandler='1 alua' wp=rw
|-+- policy='service-time 0' prio=50 status=active
| '- 22:0:0:16 sdaw 67:0 active ready running
|-+- policy='service-time 0' prio=10 status=enabled
| '- 23:0:0:16 sdbc 67:96 active ready running
| '- 21:0:0:16 sdbb 67:80 active ready running
DATA01 (360060e8114b070006014b07000000009) dm-12 HITACHI,Hi-SDS
size=200G features='1 queue_if_no_path' hwhandler='1 alua' wp=rw
|-+- policy='service-time 0' prio=50 status=active
| '- 23:0:0:4 sdn 8:208 active ready running
|-+- policy='service-time 0' prio=10 status=enabled
| '- 21:0:0:4 sds 65:32 active ready running
| '- 22:0:0:4 sdo 8:224 active ready running
REDO01 (360060e8114b070006014b0700000001d) dm-21 HITACHI,Hi-SDS
size=20G features='1 queue_if_no_path' hwhandler='1 alua' wp=rw
|-+- policy='service-time 0' prio=50 status=active
| '- 22:0:0:15 sdat 66:208 active ready running
|-+- policy='service-time 0' prio=10 status=enabled
| '- 21:0:0:15 sdax 67:16 active ready running
| '- 23:0:0:15 sdaz 67:48 active ready running
DATA00 (360060e8114b070006014b07000000008) dm-7 HITACHI,Hi-SDS
size=200G features='1 queue_if_no_path' hwhandler='1 alua' wp=rw
|-+- policy='service-time 0' prio=50 status=active
| '- 23:0:0:3 sdm 8:192 active ready running
|-+- policy='service-time 0' prio=10 status=enabled
| '- 22:0:0:3 sdl 8:176 active ready running
| '- 21:0:0:3 sdp 8:240 active ready running
TEMP01 (360060e8114b070006014b07000000019) dm-26 HITACHI,Hi-SDS
size=210G features='1 queue_if_no_path' hwhandler='1 alua' wp=rw
|-+- policy='service-time 0' prio=50 status=active
| '- 21:0:0:20 sdbl 67:240 active ready running
|-+- policy='service-time 0' prio=10 status=enabled
| '- 22:0:0:20 sdbf 67:144 active ready running
| '- 23:0:0:20 sdbm 68:0 active ready running
TEMP00 (360060e8114b070006014b07000000018) dm-25 HITACHI,Hi-SDS
size=210G features='1 queue_if_no_path' hwhandler='1 alua' wp=rw
|-+- policy='service-time 0' prio=50 status=active
| '- 23:0:0:19 sdbk 67:224 active ready running
|-+- policy='service-time 0' prio=10 status=enabled
| '- 22:0:0:19 sdbd 67:112 active ready running
| '- 21:0:0:19 sdbj 67:208 active ready running
DATA09 (360060e8114b070006014b0700000000d) dm-20 HITACHI,Hi-SDS
size=200G features='1 queue_if_no_path' hwhandler='1 alua' wp=rw
```

Outcome: All the mapped volumes are visible on the compute nodes through 3 paths each on both RAC nodes.

ASM disk creation

In this test scenario, we used the volume we created on VSS block to create an ASM disk.

```
[root@ds220g275 ~]# oracleasm status
Checking if ASM is loaded: yes
Checking if /dev/oracleasm is mounted: yes
[root@ds220g275 ~]# oracleasm createdisk DATA01 /dev/mapper/DATA1
Writing disk header: done
Instantiating disk: done
[root@ds220g275 ~]# oracleasm createdisk REDO1 /dev/mapper/REDO1
Writing disk header: done
Instantiating disk: done
[root@ds220g275 ~]# oracleasm createdisk REDO2 /dev/mapper/REDO2
Writing disk header: done
Instantiating disk: done
[root@ds220g275 ~]# oracleasm scandisks
Reloading disk partitions: done
Cleaning any stale ASM disks...
Scanning system for ASM disks...
[root@ds220g275 ~]# oracleasm listdisks
DATA01
REDO1
REDO2
[root@ds220g275 ~]#
```

We can observe that the disk was successfully created from the ASM instance as well.

```
SQL> select NAME ,PATH , HEADER_STATUS, mount_status, state from v$asm_disk;

NAME                                PATH                                HEADER_STATU MOUNT_S STATE
-----
                                /dev/mapper/REDO2                  PROVISIONED  CLOSED  NORMAL
                                /dev/mapper/REDO1                  PROVISIONED  CLOSED  NORMAL
                                /dev/mapper/DATA1                  PROVISIONED  CLOSED  NORMAL
OCR_0002                            /dev/mapper/OCR03                  MEMBER       CACHED  NORMAL
OCR_0000                            /dev/mapper/OCR01                  MEMBER       CACHED  NORMAL
OCR_0001                            /dev/mapper/OCR02                  MEMBER       CACHED  NORMAL

6 rows selected.
```

Outcome: ASM disk creation is successful on the compute node for VSS block volumes.

Test shared disk functionality for Oracle RAC setup

In this test we provisioned LUNs to both compute nodes and performed write operations using peakmarks workload. We restarted the instance as well. The following is a data loading operation performed on the database.

```
BENCH#VSSDB1 SQL> exec pmk.loaddb

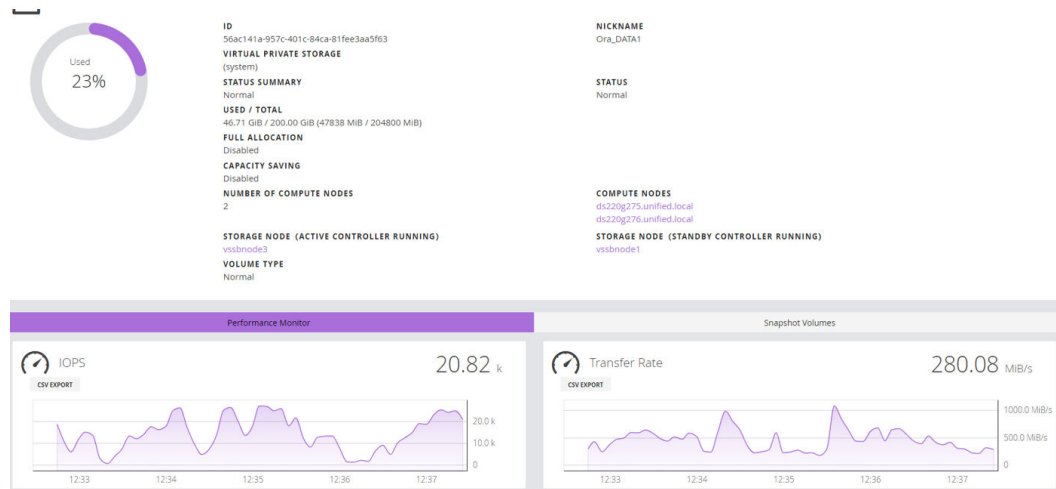
peakmarks Benchmark Software. Copyright (c) 2011 - 2022 peakmarks Ltd. All rights reserved.
Release.....: 9.6
Build.....: 220201

peakmarks command.....: pmk.loaddb

Load benchmark database.....: completed with 0 Oracle errors, 0 peakmarks errors, 11.92 minutes elapsed time
PL/SQL procedure successfully completed.

BENCH#VSSDB1 SQL> BENCH#VSSDB1 SQL>
```

On the VSS block storage GUI, we can monitor I/O and volume space utilization during this operation.



To test stability we bounced database instances at a time. The database continued to function from a surviving node in the cluster.

```
[oracle@ds220g275 ~]$ date
Thu Feb 8 02:21:46 EST 2024
[oracle@ds220g275 ~]$ srvctl status database -d VSSDB -v -f
Instance VSSDB1 is running on node ds220g275. Instance status: Open.
Instance VSSDB2 is running on node ds220g276. Instance status: Open.
[oracle@ds220g275 ~]$ srvctl stop instance -i VSSDB1 -d VSSDB -o immediate
[oracle@ds220g275 ~]$ srvctl status database -d VSSDB -v -f
Instance VSSDB1 is not running on node ds220g275
Instance VSSDB2 is running on node ds220g276. Instance status: Open.
[oracle@ds220g275 ~]$ srvctl start instance -i VSSDB1 -d VSSDB
[oracle@ds220g275 ~]$ srvctl status database -d VSSDB -v -f
Instance VSSDB1 is running on node ds220g275. Instance status: Open.
Instance VSSDB2 is running on node ds220g276. Instance status: Open.
[oracle@ds220g275 ~]$

[oracle@ds220g276 ~]$ date
Thu Feb 8 02:25:31 EST 2024
[oracle@ds220g276 ~]$ srvctl status database -d VSSDB -v -f
Instance VSSDB1 is running on node ds220g275. Instance status: Open.
Instance VSSDB2 is running on node ds220g276. Instance status: Open.
[oracle@ds220g276 ~]$ srvctl stop instance -i VSSDB2 -d VSSDB -o immediate
[oracle@ds220g276 ~]$ srvctl status database -d VSSDB -v -f
Instance VSSDB1 is running on node ds220g275. Instance status: Open.
Instance VSSDB2 is not running on node ds220g276
[oracle@ds220g276 ~]$ srvctl start instance -i VSSDB2 -d VSSDB
[oracle@ds220g276 ~]$ srvctl status database -d VSSDB -v -f
Instance VSSDB1 is running on node ds220g275. Instance status: Open.
Instance VSSDB2 is running on node ds220g276. Instance status: Open.
[oracle@ds220g276 ~]$
```

Outcome: Oracle Database is started on the RAC cluster and no errors on write operations are observed because the disk is shared on both RAC nodes.

Add a new LUN to an existing disk group

In this test scenario, we added a new LUN from VSS block to an existing disk group in ASM. We created DATA2 LUN of 200 GiB size on VSS block.

SELECT	NAME	VIRTUAL PRIVATE...	STATUS SUMMAR...	STATUS	SNAPSHOT STAT...	TOTAL	USED	CAPA...	NUMB...	SNAPSHOTS	VOLUME TYPE	NICKNAME	ID
	DATA1	(system)	Normal	Normal	-	200.00 GiB	134.81 GiB	Disa...	2	Not exist	Normal	Ora_DATA1	56ac...
	DATA2	(system)	Normal	Normal	-	200.00 GiB	-	Disa...	0	Not exist	Normal	DATA2	92d...

We updated udev rules and reloaded, and then updated the multipath configuration. We then scanned the new LUN from the compute nodes.

```
[root@ds220g275 ~]# /usr/bin/rescan-scsi-bus.sh
Scanning SCSI subsystem for new devices
Scanning host 0 for SCSI target IDs 0 1 2 3 4 5 6 7, all LUNs
Scanning for device 0 1 123 0 ...
OLD: Host: scsi0 Channel: 01 Id: 123 Lun: 00
Vendor: QCT Model: D52BQ-2U Rev: 0340
Type: Enclosure ANSI SCSI revision: 06
Scanning host 1 for SCSI target IDs 0 1 2 3 4 5 6 7, all LUNs
Scanning host 2 for SCSI target IDs 0 1 2 3 4 5 6 7, all LUNs
Scanning host 3 for SCSI target IDs 0 1 2 3 4 5 6 7, all LUNs
Scanning host 4 for SCSI target IDs 0 1 2 3 4 5 6 7, all LUNs

DATA2 (360060e8114b070006014b07000000007) dm-12 HITACHI,Hi-SDS
size=200G features='1 queue_if_no_path' hwhandler='1 alua' wp=rw
|-+- policy='service-time 0' prio=50 status=active
| '- 21:0:0:6 sdu 65:64 active ready running
`-+- policy='service-time 0' prio=10 status=enabled
|- 22:0:0:6 sdv 65:80 active ready running
`- 23:0:0:6 sdw 65:96 active ready running
DATA1 (360060e8114b070006014b07000000004) dm-11 HITACHI,Hi-SDS
size=200G features='1 queue_if_no_path' hwhandler='1 alua' wp=rw
|-+- policy='service-time 0' prio=50 status=active
| '- 23:0:0:3 sdr 65:16 active ready running
`-+- policy='service-time 0' prio=10 status=enabled
|- 21:0:0:3 sdl 8:176 active ready running
`- 22:0:0:3 sdh 8:112 active ready running
[root@ds220g275 rules.d]#
```

The new LUN is visible in the ASM instance as well.

```
SQL> !date
Thu Feb 8 04:18:54 EST 2024

SQL> select NAME , PATH , HEADER_STATUS, mount_status, state, header_status from v$asm_disk;

NAME                                PATH                                HEADER_STATU MOUNT_S STATE    HEADER_STATU
-----
DATA01_0000                        /dev/mapper/DATA2                  CANDIDATE    CLOSED   NORMAL    CANDIDATE
REDO02_0000                        /dev/mapper/DATA1                  MEMBER       CACHED   NORMAL    MEMBER
OCR_0002                          /dev/mapper/OCR02                  MEMBER       CACHED   NORMAL    MEMBER
OCR_0001                          /dev/mapper/OCR03                  MEMBER       CACHED   NORMAL    MEMBER
OCR_0000                          /dev/mapper/OCR01                  MEMBER       CACHED   NORMAL    MEMBER
REDO01_0000                        /dev/mapper/REDO1                  MEMBER       CACHED   NORMAL    MEMBER

7 rows selected.

SQL> alter diskgroup DATA01 add disk '/dev/mapper/DATA2' rebalance power 2;

Diskgroup altered.

SQL> select * from v$asm_operation;

GROUP_NUMBER OPERA PASS      STAT      POWER    ACTUAL    SOFAR    EST_WORK    EST_RATE EST_MINUTES ERROR_CODE
-----
2 REBAL COMPACT WAIT        2         2         0         0         0         0
2 REBAL REBALANCE RUN        2         2        655      18495      3113         5
2 REBAL REBUILD DONE        2         2         0         0         0         0
```

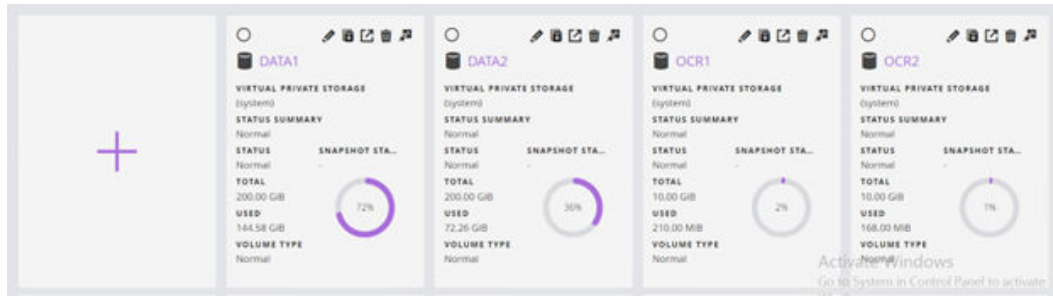
The new LUN is part of the disk group.

```
SQL> select NAME , PATH , HEADER_STATUS, mount_status, state, header_status from v$asm_disk;

NAME                                PATH                                HEADER_STATU MOUNT_S STATE    HEADER_STATU
-----
DATA01_0000                        /dev/mapper/DATA1                  MEMBER       CACHED   NORMAL    MEMBER
REDO02_0000                        /dev/mapper/REDO2                  MEMBER       CACHED   NORMAL    MEMBER
OCR_0002                          /dev/mapper/OCR03                  MEMBER       CACHED   NORMAL    MEMBER
OCR_0001                          /dev/mapper/OCR02                  MEMBER       CACHED   NORMAL    MEMBER
OCR_0000                          /dev/mapper/OCR01                  MEMBER       CACHED   NORMAL    MEMBER
REDO01_0000                        /dev/mapper/REDO1                  MEMBER       CACHED   NORMAL    MEMBER
DATA01_0001                        /dev/mapper/DATA2                  MEMBER       CACHED   NORMAL    MEMBER

7 rows selected.
```

The following GUI image of VSS block shows DATA2 includes the new LUN, has some data, and is being used.



The ASM instance shows that the new LUN has data written on it.

```
SQL> select GROUP_NUMBER,DISK_NUMBER,NAME,PATH,MOUNT_STATUS,OS_MB,TOTAL_MB,FREE_MB from v$asm_disk;
```

GROUP_NUMBER	DISK_NUMBER	NAME	PATH	MOUNT_S	OS_MB	TOTAL_MB	FREE_MB
2	0	DATA01_0000	/dev/mapper/DATA1	CACHED	204800	204800	130792
4	0	REDO02_0000	/dev/mapper/REDO2	CACHED	15360	15360	13664
1	2	OCR_0002	/dev/mapper/OCR03	CACHED	10240	10240	10124
1	1	OCR_0001	/dev/mapper/OCR02	CACHED	10240	10240	10116
1	0	OCR_0000	/dev/mapper/OCR01	CACHED	10240	10240	10008
3	0	REDO01_0000	/dev/mapper/REDO1	CACHED	15360	15360	13664
2	1	DATA01_0001	/dev/mapper/DATA2	CACHED	204800	204800	130808

7 rows selected.

Outcome: As we can see, when we add new disk to disk group, data is spread over the new disk. We can observe it in the space utilization of the new disk. Successful rebalancing of data happens over volumes.

Perform volume expansion to validate online expansion functionality

In this test case, volume expansion was performed on VSS block while the database was online, and transactions were occurring. Previously the volume size was 200 GiB, and it was expanded to 250 GiB.

At the same time we initiated transactions in the database using peakmarks workload.

```
BENCH@VSSDB1 SQL> !date
Wed Feb 14 06:29:16 EST 2024

BENCH@VSSDB1 SQL> exec pmk.run

peakmarks Benchmark Software. Copyright (c) 2011 - 2022 peakmarks Ltd. All rights reserved.
Release.....: 9.6
Build.....: 220201
Customer.....: Hitachi Vantara, Santa Clara
peakmarks command.....: pmk.run_manual
Test(s) found in mbc.cfg.....: 1
Benchmark run completed.....: Oracle errors 0, peakmarks errors 0, elapsed time 11.25 min
PL/SQL procedure successfully completed.
```

In the following image, we can see that transactions were successfully completed while at the same time we resized the ASM disk.

```
SQL> !date
Wed Feb 14 06:33:52 EST 2024
SQL> select NAME ,PATH , HEADER_STATUS, mount_status, state, header_status,OS_MB,TOTAL_MB,FREE_MB from v$asm_disk;
```

NAME	PATH	HEADER_STATU	MOUNT_S	STATE	HEADER_STATU	OS_MB	TOTAL_MB	FREE_MB
DATA01_0000	/dev/mapper/DATA1	MEMBER	CACHED	NORMAL	MEMBER	256000	204800	48004
REDO02_0000	/dev/mapper/REDO2	MEMBER	CACHED	NORMAL	MEMBER	15360	15360	13664
REDO01_0000	/dev/mapper/REDO1	MEMBER	CACHED	NORMAL	MEMBER	15360	15360	13664
OCR_0000	/dev/mapper/OCR01	MEMBER	CACHED	NORMAL	MEMBER	10240	10240	10084
OCR_0001	/dev/mapper/OCR02	MEMBER	CACHED	NORMAL	MEMBER	10240	10240	10120
OCR_0002	/dev/mapper/OCR03	MEMBER	CACHED	NORMAL	MEMBER	10240	10240	10124

```

6 rows selected.

SQL> alter diskgroup DATA01 resize all;

Diskgroup altered.

SQL> select NAME ,PATH , HEADER_STATUS, mount_status, state, header_status,OS_MB,TOTAL_MB,FREE_MB from v$asm_disk;
```

NAME	PATH	HEADER_STATU	MOUNT_S	STATE	HEADER_STATU	OS_MB	TOTAL_MB	FREE_MB
DATA01_0000	/dev/mapper/DATA1	MEMBER	CACHED	NORMAL	MEMBER	256000	256000	99204
REDO02_0000	/dev/mapper/REDO2	MEMBER	CACHED	NORMAL	MEMBER	15360	15360	13664
REDO01_0000	/dev/mapper/REDO1	MEMBER	CACHED	NORMAL	MEMBER	15360	15360	13664
OCR_0000	/dev/mapper/OCR01	MEMBER	CACHED	NORMAL	MEMBER	10240	10240	10084
OCR_0001	/dev/mapper/OCR02	MEMBER	CACHED	NORMAL	MEMBER	10240	10240	10120
OCR_0002	/dev/mapper/OCR03	MEMBER	CACHED	NORMAL	MEMBER	10240	10240	10124

The database is up and running.

INSTANCE_NUMBER	INSTANCE_NAME	HOST_NAME	VERSION_FULL	STARTUP_TIME	STATUS	ACTIVE_ST
1	VSSDB1	ds220g275.unified.local	19.19.0.0.0	08-FEB-2024 05:06:14	OPEN	NORMAL
2	VSSDB2	ds220g276.unified.local	19.19.0.0.0	08-FEB-2024 05:06:51	OPEN	NORMAL

```

SYS@VSSDB1 SQL> !date
Wed Feb 14 06:46:56 EST 2024

```

Outcome: Volume expansion is successful while database is up and running with live transactions ongoing.

High availability tests

This section details the high availability tests that were used for validation. Compute nodes configured as clusters (Oracle RAC) and three storage nodes were configured as a VSS block mirror configuration.

Compute node failure test — Shut down one RAC node at a time

In this test case we did an unplanned compute node shutdown at a time to test the impact on the cluster.

```

[root@ds220g275 ~]# date
Thu Feb  8 02:28:44 EST 2024
[root@ds220g275 ~]#
[root@ds220g275 ~]# /u01/app/19c/grid_home/bin/crsctl check cluster -all
*****
ds220g275:
CRS-4537: Cluster Ready Services is online
CRS-4529: Cluster Synchronization Services is online
CRS-4533: Event Manager is online
*****
ds220g276:
CRS-4537: Cluster Ready Services is online
CRS-4529: Cluster Synchronization Services is online
CRS-4533: Event Manager is online
*****
[root@ds220g275 ~]#
[root@ds220g275 ~]# reboot

```

From the following image we can see that the surviving node in the cluster continues to function.

```
[root@ds220g276 ~]# date
Thu Feb  8 02:31:11 EST 2024
[root@ds220g276 ~]# /u01/app/19c/grid_home/bin/crsctl check cluster -all
*****
ds220g276:
CRS-4537: Cluster Ready Services is online
CRS-4529: Cluster Synchronization Services is online
CRS-4533: Event Manager is online
*****
[root@ds220g276 ~]#
```

```
sys@VSSDB2 SQL> !date
Thu Feb  8 02:31:50 EST 2024
sys@VSSDB2 SQL> select a.INSTANCE_NUMBER,a.INSTANCE_NAME,HOST_NAME,VERSION_FULL,STARTUP_TIME,STATUS,ACTIVE_STATE from gv$instance a;
-----
INSTANCE_NUMBER  INSTANCE_NAME          HOST_NAME                VERSION_FULL    STARTUP_TIME        STATUS    ACTIVE_ST
-----
                2  VSSDB2                ds220g276.unified.local  19.19.0.0.0      08-FEB-2024 02:26:52 OPEN      NORMAL
```

Now the second node is restarted.

```
[root@ds220g276 ~]# date
Thu Feb  8 02:55:03 EST 2024
[root@ds220g276 ~]# /u01/app/19c/grid_home/bin/crsctl check cluster -all
*****
ds220g275:
CRS-4537: Cluster Ready Services is online
CRS-4529: Cluster Synchronization Services is online
CRS-4533: Event Manager is online
*****
ds220g276:
CRS-4537: Cluster Ready Services is online
CRS-4529: Cluster Synchronization Services is online
CRS-4533: Event Manager is online
*****
[root@ds220g276 ~]#
```

This image shows that the second node is not available, yet the instance on node 1 continues to function.

```
sys@VSSDB2 SQL> !date
Thu Feb  8 02:31:50 EST 2024
sys@VSSDB2 SQL> select a.INSTANCE_NUMBER,a.INSTANCE_NAME,HOST_NAME,VERSION_FULL,STARTUP_TIME,STATUS,ACTIVE_STATE from gv$instance a;
-----
INSTANCE_NUMBER  INSTANCE_NAME          HOST_NAME                VERSION_FULL    STARTUP_TIME        STATUS    ACTIVE_ST
-----
                2  VSSDB2                ds220g276.unified.local  19.19.0.0.0      08-FEB-2024 02:26:52 OPEN      NORMAL
```

After both compute nodes reboot successfully, they re-joined the cluster.

```
[oracle@ds220g276 ~]$ srvctl status database -d VSSDB -v -f
Instance VSSDB1 is running on node ds220g275. Instance status: Open.
Instance VSSDB2 is running on node ds220g276. Instance status: Open.
[oracle@ds220g276 ~]$ ps -ef|grep tns
root          922      0 03:00 ?        00:00:00 [netns]
grid         10954      1 03:02 ?        00:00:00 /u01/app/19c/grid_home/bin/tnslsnr LISTENER -no_crs_notify -inherit
grid         11010      1 03:02 ?        00:00:00 /u01/app/19c/grid_home/bin/tnslsnr LISTENER SCAN1 -no_crs_notify -inherit
grid         11021      1 03:02 ?        00:00:00 /u01/app/19c/grid_home/bin/tnslsnr ASMNET1LSNR_ASM -no_crs_notify -inherit
oracle       18982    14567  0 03:07 pts/0    00:00:00 grep --color=auto tns
[oracle@ds220g276 ~]$ uptime
03:07:55 up 7 min, 1 user, load average: 1.46, 0.73, 0.34
[oracle@ds220g276 ~]$
```

The database instance is up and running on both nodes.

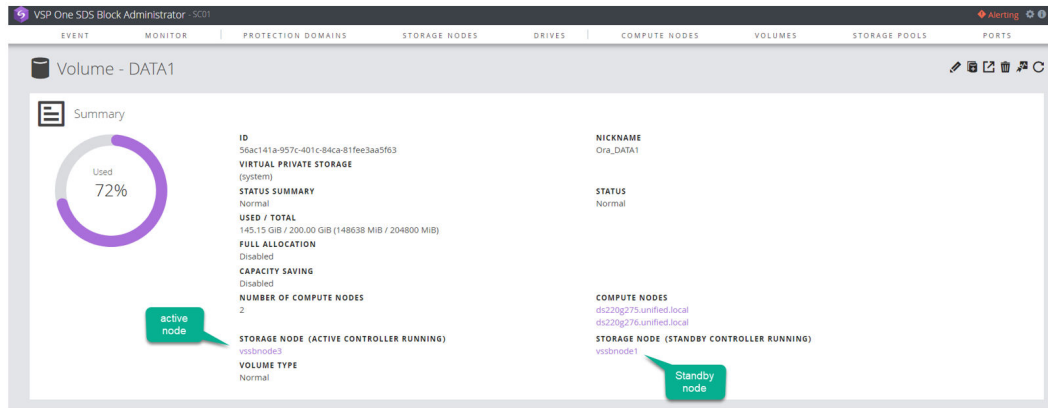
```
-----
INSTANCE_NUMBER  Name                Server                Oracle
Version          Startup              Status    ACTIVE_ST
-----
                2  VSSDB2            ds220g276.unified.local  19.19.0.0.0  08-FEB-2024 03:02:29    OPEN      NORMAL
                1  VSSDB1            ds220g275.unified.local  19.19.0.0.0  08-FEB-2024 02:35:31    OPEN      NORMAL
sys@VSSDB2 SQL> !uptime
03:09:31 up 8 min, 1 user, load average: 0.41, 0.58, 0.33
```

Outcome: When we shut down a compute node, database services continue to keep running from the active node without any issues.

Storage node failure test — shut down one node from the VSS block storage cluster

If a storage node is down due to failure or maintenance, regular operations continue because of the highly available design of VSS block. With our mirror configuration, regular operations will continue with one storage node failure.

As we can see in the following image, the DATA1 volume is present on `vssbnode3` which is an active node, whereas `vssbnode1` is a standby node.



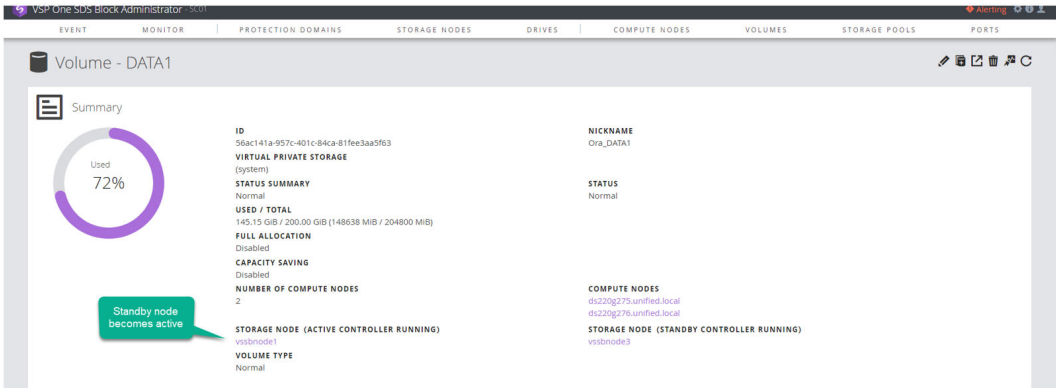
`vssbnode3` was shut down to perform an unplanned shutdown of one of the storage nodes on which the DATA1 volume is present.

SELECT	NAME	STATUS SUMMARY	STATUS	FAULT DOMAIN	ROLE	ATTRIBUTES	NUMBER OF REBUILD...	ID
☐	vssbnode1	Normal	Ready	SC01-PD01-FD01	Master	(none)	1	b44f2643-9e2d-48...
☐	vssbnode2	Normal	Ready	SC01-PD01-FD01	Master	(none)	1	29e370fe-ef3e-4db...
☐	vssbnode3	Error	TemporaryBlockageProcessInProgress	SC01-PD01-FD01	Master	(none)	1	68ae44f2-c072-419...

As shown in the following screenshots, for a 200 GB data disk, one path is faulty out of three paths after an unplanned shutdown of a VSS block node. The database cluster is healthy and continues to run without any issue.

The volume is accessible through other available paths on the compute nodes because of the multipathing configuration, and data is still accessible from standby nodes even if the primary node on the storage cluster is down.

```
[root@ds220g275 ~]# date
Wed Feb 14 04:46:59 EST 2024
[root@ds220g275 ~]#
[root@ds220g275 ~]# multipath -ll /dev/mapper/DATA1
DATA1 (360060e8114b070006014b07000000004) dm-11 HITACHI,Hi-SDS
size=200G features='1 queue_if_no_path' hwhandler='1 alua' wp=rw
|+- policy='service-time 0' prio=0 status=enabled
|  '- 23:0:0:3 sdr      65:16 failed faulty running
|+- policy='service-time 0' prio=30 status=active
|  '- 22:0:0:3 sdi      8:128 active ready running
|  '- 21:0:0:3 sdk      8:160 active ready running
[root@ds220g275 ~]#
[root@ds220g275 ~]# /u01/app/19c/grid_home/bin/crsctl check cluster -all
*****
ds220g275:
CRS-4537: Cluster Ready Services is online
CRS-4529: Cluster Synchronization Services is online
CRS-4533: Event Manager is online
*****
ds220g276:
CRS-4537: Cluster Ready Services is online
CRS-4529: Cluster Synchronization Services is online
CRS-4533: Event Manager is online
*****
[root@ds220g275 ~]#
```



Outcome: Using the VSS block GUI we can verify that standby node `vssbnode1` is now active node and there is no impact to the cluster.

Summary of test cases

Test Case	Steps Performed	Status	Remark
Create required DB volumes on VSS block storage and allocate to Oracle RAC nodes over iSCSI	- Configure multipathing on compute nodes. - Add compute nodes on VSS block storage using IQN number.	Passed	We successfully integrated compute nodes with VSS block storage over an iSCSI channel.

Test Case	Steps Performed	Status	Remark
	- Create volumes on VSS block storage and allocate to compute nodes. - Scan volumes on compute nodes.		
ASM disk creation	- Discover volumes on compute nodes. - Allocate disks using Oracle ASM.	Passed	Disk created successfully using <code>oracleasm</code> .
Test shared disk functionality for Oracle RAC setup	Create a volume, share it to both compute nodes.	Passed	Volume create and successfully shared, using shared disk Oracle grid configured successfully.
Add new LUN to existing disk group	- Create a new volume and share it to compute nodes. - Configure udev, multipathing. - Add new disk to exiting disk group in ASM.	Passed	New disk visible and rebalance operation successful in ASM. Data has been spread on new disk.
Volume expansion - Perform Volume expansion to validate online expansion functionality	- Expand existing volume on VSS block. - Rescan LUN on compute nodes. - Reload/restart multipath service.	Passed	Volume resize performed successfully while database is up and running with live transactions.
Compute node failure test – Shut down one RAC node at a time	Bring down compute nodes one at a time.	Passed	Cluster is up and running successfully even if one node is unavailable.
Storage node failure test – Shut down 1 node form the VSSB storage cluster	- In VSS block, bring one node down. - Check volume status, volume failover to the other node.	Passed	Bring one storage node offline through an unplanned shut down, volume failover to standby node. No impact on the database.

Conclusion

Hitachi Advanced Server and Hitachi Virtual Storage Software block together build powerful a solution to implement Oracle RAC database with zero single point of failure and they help offload ASM disk mirroring to backend storage (VSS block).

This reference architecture highlights recommended components for deploying Oracle RAC Database in a Hitachi infrastructure environment (Hitachi Advanced Server or VSS block) while taking advantage of various Hitachi data storage integrations to achieve a highly resilient and protected platform to clusters at scale.

Hitachi Vantara

Corporate Headquarters
2535 Augustine Drive
Santa Clara, CA 95054 USA



HitachiVantara.com/contact