Deploy Hitachi Unified Compute Platform Select for Oracle Database using Oracle Database 11g R2 Enterprise Edition in a Medium-sized Solution

Reference Architecture Guide

By Kishore Daggubati

February 7, 2013
Feedback

Hitachi Data Systems welcomes your feedback. Please share your thoughts by sending an email message to SolutionLab@hds.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.
# Table of Contents

**Solution Overview** ..........................................................2

**Key Solution Components** ......................................................5

- Hitachi Compute Blade 2000 ....................................................7
- Hitachi Unified Storage VM .....................................................7
- Hitachi Dynamic Provisioning ..................................................8
- Hitachi Storage Navigator ....................................................9
- Oracle Linux ..........................................................................9
- Oracle Database ....................................................................9
- Fusion-io ioDrive2 ............................................................9

**Solution Design** ................................................................10

- Storage Architecture ..........................................................13
- Server and Application Architecture .....................................19
- SAN Architecture ..............................................................22
- Network Architecture ..........................................................23

**Engineering Validation** ........................................................25

- Test Methodology ..............................................................25
- Test Results .........................................................................27

**Conclusion** ..........................................................................32
Deploy Hitachi Unified Compute Platform Select for Oracle Database using Oracle Database 11g R2 Enterprise Edition in a Medium-sized Solution

Reference Architecture Guide

This reference architecture guide shows how the medium-sized Hitachi Unified Compute Platform Select for Oracle Database environment provides high performance in an integrated solution for an Oracle infrastructure. There is a comparison of results using and not using PCI-e flash memory.

This validated solution integrates servers, storage systems, and storage software. It provides reliability, high availability, scalability, and performance while processing small-scale to large-scale OLTP workloads on a single-dedicated server running Oracle Database 11g R2 with Oracle Enterprise Linux 5, Update 5.

Use this reference architecture as a resource to design an infrastructure that meets your infrastructure requirements and budget.

You benefit from this reference architecture if you are a database administrator, storage administrator, or have the responsibility to plan and deploy Oracle Database 11g R2 solutions. You need familiarity with the following:

- Hitachi Unified Storage VM
- Hitachi Compute Blade 2000
- Storage area networks
- Oracle Database 11g Release 2
- Oracle Automatic Storage Management
- Fusion-io ioDrive2

**Note** — Testing of this configuration was in a lab environment. Many things affect production environments beyond prediction or duplication in a lab environment. Follow the recommended practice of conducting proof-of-concept testing for acceptable results in a non-production, isolated test environment that otherwise matches your production environment before your production implementation of this solution.
Solution Overview

The Hitachi Unified Compute Platform Select for Oracle Database medium-sized solution implements a high performance environment for Oracle Database 11g Release 2 for typical on-line transaction processing workloads. Tailor your implementation of this solution to meet your specific needs.

This reference architecture guide discusses deploying this solution with and without PCI-e flash acceleration. In both cases, there are two server blades connected using an SMP interface connector hosting one Oracle database server. This shows the performance benefits of using PCI-e flash memory devices.

This reference architecture discusses the following:

- Hitachi Compute Blade 2000 using two server blades
- Hitachi Unified Storage VM
- SAN infrastructure
Figure 1 on page 3 shows the infrastructure without PCI-e flash acceleration. Figure 2 on page 4 shows the infrastructure using PCI-e flash acceleration.
Figure 2

Hitachi Compute Blade 2000

1 Gb/sec LAN Switch Modules (Total 2)

8 Gb/sec Fibre Channel Switch Modules (Total 2)

Fusion-io ioDrive2 PCIe Flash Cards (Total 4)

Two Controllers with sixteen 8 Gb/sec Fibre Channel Target Ports

Hitachi Unified Storage VM
Key Solution Components

Table 1 on page 5 and Table 2. Reference Architecture Software Components have the key hardware and software components used in this reference architecture.

**Table 1. Reference Architecture Hardware Components**

<table>
<thead>
<tr>
<th>Component</th>
<th>Description</th>
<th>Version</th>
<th>Quantity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Server chassis</td>
<td>Hitachi Compute Blade 2000</td>
<td>Firmware Version A0195-C-6443</td>
<td>1</td>
</tr>
<tr>
<td>Server blades</td>
<td>Model GVAX57A1 (X57-A1), each configured as follows:</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>- Intel Xeon X7560 processors, 2.26 GHz,</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- 2 × 8-core physical processors</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- 96 GB RAM using 4 GB DIMMs</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- 1 dual port Fibre Channel card (Mezzanine Slot 0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- 2 × 1 Gb/sec Ethernet NICs</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- (Used only with architecture using PCI-e flash card) 2 Fusion-io ioDrive2</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- 1.2 TB MLC</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Version 3.1.5, Boot version 0.0.3.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- In PCI-e expansion Slot 0 and PCI-e expansion Slot 1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 1. Reference Architecture Hardware Components (Continued)

<table>
<thead>
<tr>
<th>Component</th>
<th>Description</th>
<th>Version</th>
<th>Quantity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Storage system</td>
<td>Hitachi Unified Storage VM, configured as follows:</td>
<td>73-01-01-00/00</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>▪ 144 × 600 GB SAS 10k RPM drives in 6 trays with 24 disks in each tray, for about 86 TB raw capacity</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>▪ 2 main storage blades, each with:</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>▪ 2 front-end connectivity modules</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>▪ 4 × 8 Gb/sec Fibre Channel ports, 8 ports total</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>▪ 2 back-end connectivity modules</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>▪ 8 × 6 Gb/sec SAS links each, 16 links total</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>▪ 2 MP blades, each with one 8-core Intel XEON processor, 2.33 GHz</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>▪ 55 GB cache</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>▪ One front-end port on each controller connects to Hitachi Compute Blade 2000.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCI-e flash cards</td>
<td>Fusion-io ioDrive2</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>▪ Only used when using flash acceleration</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>▪ 1.2 TB PCI-e flash cards</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SAN connectivity</td>
<td>6 port, 8 Gb/sec Fibre Channel switch modules</td>
<td>V 642b</td>
<td>2</td>
</tr>
<tr>
<td>Symmetric multiprocessing (SMP)</td>
<td>SMP connector creates one server from two GVAX57A1 server blades</td>
<td>N/A</td>
<td>1</td>
</tr>
</tbody>
</table>

**Note** — This reference architecture uses X57-A1 (GVAX57A1) server blades with 4 GB DIMMs. You may upgrade to X57-A2 server blades using Intel Xeon Processor E7-8870. In addition, you may upgrade the 4 GB DIMMs to 8 GB DIMMs.
Table 2. Reference Architecture Software Components

<table>
<thead>
<tr>
<th>Component</th>
<th>Description</th>
<th>Version</th>
</tr>
</thead>
<tbody>
<tr>
<td>Operating system</td>
<td>Oracle Enterprise Linux</td>
<td>5 Update 5, Red Hat-compatible kernel</td>
</tr>
<tr>
<td>Volume manager and file system software</td>
<td>Oracle Automatic Storage Management</td>
<td>11g R2, 11.2.0.1.0</td>
</tr>
<tr>
<td>Multipath software</td>
<td>Device Mapper Multipath (DM-Multipath) of Red Hat Enterprise Linux for the Red Hat Enterprise Linux 5 release</td>
<td>0.4.7-34.el5</td>
</tr>
<tr>
<td>Database software</td>
<td>Oracle</td>
<td>11g R2, 11.2.0.1.0</td>
</tr>
<tr>
<td>Storage management software</td>
<td>Hitachi Storage Navigator</td>
<td>73-01-01/00</td>
</tr>
<tr>
<td></td>
<td>Hitachi Dynamic Provisioning</td>
<td>Microcode dependent</td>
</tr>
<tr>
<td>I/O calibration software</td>
<td>Orion</td>
<td>11.1</td>
</tr>
<tr>
<td>Load generator software</td>
<td>Swingbench</td>
<td>2.4.0.845</td>
</tr>
<tr>
<td>Database client communication software</td>
<td>Oracle Net Services</td>
<td>11g R2, 11.2.0.1.0</td>
</tr>
</tbody>
</table>

**Hitachi Compute Blade 2000**

*Hitachi Compute Blade 2000* is an enterprise-class blade server platform. It features the following:

- A balanced system architecture that eliminates bottlenecks in performance and throughput
- Configuration flexibility
- Eco-friendly power-saving capabilities
- Fast server failure recovery using a N+1 cold standby design that allows replacing failed servers within minutes

**Hitachi Unified Storage VM**

*Hitachi Unified Storage VM* is an entry-level enterprise storage platform. It combines storage virtualization services with unified block, file, and object data management. This versatile, scalable platform offers a storage virtualization system to provide central storage services to existing storage assets.

Unified management delivers end-to-end central storage management of all virtualized internal and external storage on Unified Storage VM. A unique, hardware-accelerated, object-based file system supports intelligent file tiering and migration, as well as virtual NAS functionality, without compromising performance or scalability.
The benefits of Unified Storage VM are the following:

- Enables the move to a new storage platform with less effort and cost when compared to the industry average
- Increases performance and lowers operating cost with automated data placement
- Supports scalable management for growing and complex storage environment while using fewer resources
- Achieves better power efficiency and with more storage capacity for more sustainable data centers
- Lowers operational risk and data loss exposure with data resilience solutions
- Consolidates management with end-to-end virtualization to prevent virtual server sprawl

**Hitachi Dynamic Provisioning**

On Hitachi storage systems, Hitachi Dynamic Provisioning provides wide striping and thin provisioning functionalities.

Using Dynamic Provisioning is like using a host-based logical volume manager (LVM), but without incurring host processing overhead. It provides one or more wide-stripping pools across many RAID groups. Each pool has one or more dynamic provisioning virtual volumes (DP-VOLs) of a logical size you specify of up to 60 TB created against it without allocating any physical space initially.

Deploying Dynamic Provisioning avoids the routine issue of hot spots that occur on logical devices (LDEVs). These occur within individual RAID groups when the host workload exceeds the IOPS or throughput capacity of that RAID group. Dynamic provisioning distributes the host workload across many RAID groups, which provides a smoothing effect that dramatically reduces hot spots.

When used with Hitachi Unified Storage VM, Hitachi Dynamic Provisioning has the benefit of thin provisioning. Physical space assignment from the pool to the dynamic provisioning volume happens as needed using 42 MB pages, up to the logical size specified for each dynamic provisioning volume. There can be a dynamic expansion or reduction of pool capacity without disruption or downtime. You can rebalance an expanded pool across the current and newly added RAID groups for an even striping of the data and the workload.
Hitachi Storage Navigator

Hitachi Storage Navigator enables essential management and optimization functions. Using Java agents, Storage Navigator runs on most browsers. A command line interface is available.

Use Storage Navigator for the following:

- Pool creation and expansion
- LUN creation and expansion
- Online microcode updates and other system maintenance functions
- Performance metrics

You need Storage Navigator to take advantage of the full features of Hitachi Unified Storage VM.

Oracle Linux

Oracle Linux is an enterprise-class operating system built and tested to run Oracle hardware, databases, and middleware. It is fully compatible with the Red Hat Enterprise Linux kernel.

Oracle Database

Oracle Database is optimized for use with Oracle products. It uses Oracle Database Automatic Storage Management (ASM), combining the features of a volume manager and an application-optimized file system for database files. ASM is part of the grid infrastructure component in Oracle Database.

Fusion-io ioDrive2

Fusion-io ioDrive 2 combines VSL and ioMemory into an ioMemory platform on a PCI-e flash card. The ioMemory platform provides consistent low latency access for mixed workloads with 15-microsecond access latency, 1.5 GB/sec bandwidth, over 275,000 read IOPS (512 bytes) and over 800,000 write IOPS (512 bytes).

This solution uses Fusion-io ioDrive2 as part of the default configuration when using PCI-e flash acceleration. They are high performance block devices used to accelerate Oracle database input and output performance.
Solution Design

This is the detailed description of the reference architecture environment implementing a medium-sized environment for Hitachi Unified Compute Platform Select for Oracle Database.

The following are the two configurations in this reference architecture:

- **With PCI-e Flash Acceleration** — The configuration with PCI-e flash acceleration has Fusion-io ioDrive2 cards. The database storage layout uses Fusion-io ioDrive2 devices for the preferred mirror read and the storage disks for the mirror.

- **Without PCI-e Flash Acceleration** — The configuration without PCI-e flash acceleration consists of the same components as the reference architecture, except without using Fusion-io ioDrive2 cards. Instead, the database layout only uses the storage disks with external redundancy for ASM mirroring.

The description of the hardware components used by Hitachi Data Systems for testing is in “Key Solution Components” on page 5. Specific infrastructure configuration includes the following:

- **Server** — This is a single server node, consisting of two X57-A1 server blades connected using an SMP connector for one logical Oracle Database server.

- **Storage System** — There are LDEVs mapped to each port that are presented to the server as LUNs.

- **SAN Fabric** — There are two zones created on each switch module to zone the two mezzanine Fibre Channel ports on each server blade and the four storage host ports.
Figure 3 on page 11 illustrates the reference architecture configuration without using PCI-e flash acceleration. Figure 4 on page 12 illustrates the reference architecture configuration using PCI-e flash acceleration.
Figure 4
Storage Architecture

This describes the storage architecture of this reference architecture. It takes into consideration Hitachi Data Systems and Oracle recommended practices for the deployment of database storage design.

Storage Configuration

This reference architecture uses RAID groups and storage pools created with Hitachi Dynamic Provisioning on Hitachi Unified Storage VM.

Figure 5 on page 13 shows the RAID groups, dynamic provisioning pools, and host groups for the architecture not using PCI-e flash acceleration. Figure 6 on page 14 shows the RAID groups, dynamic provisioning pools, and host groups for the architecture using PCI-e flash disks.

![Hitachi Unified Storage VM](image)

Figure 5
Table 3 has the details of the RAID groups for the Oracle on-line and archived redo logs.

Table 3. RAID Groups Used for the Oracle Online and Archived Redo Logs

<table>
<thead>
<tr>
<th>Parity Group</th>
<th>Purpose</th>
<th>RAID Level</th>
<th>Drive Type</th>
<th>No. of Drives</th>
<th>Capacity (GB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-1</td>
<td>Oracle on-line redo logs</td>
<td>RAID-10 (2D+2D)</td>
<td>600 GB 10k RPM SAS</td>
<td>4</td>
<td>1000</td>
</tr>
<tr>
<td>2-2</td>
<td>Oracle on-line redo logs</td>
<td>RAID-10 (2D+2D)</td>
<td></td>
<td>4</td>
<td>1000</td>
</tr>
<tr>
<td>2-3</td>
<td>Oracle archived redo logs</td>
<td>RAID-10 (2D+2D)</td>
<td></td>
<td>4</td>
<td>1000</td>
</tr>
</tbody>
</table>
Table 4 has the details for the LDEVs created in these RAID groups.

### Table 4. LUNs Used for the Oracle Online and Archived Redo Logs

<table>
<thead>
<tr>
<th>Parity Group</th>
<th>LDEVs</th>
<th>LDEV Size (GB)</th>
<th>Purpose</th>
<th>Storage Port</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-1</td>
<td>00:00:01</td>
<td>1000</td>
<td>Oracle online redo logs and Oracle control file</td>
<td>1A, 1B, 2A, 2B</td>
</tr>
<tr>
<td>2-2</td>
<td>00:00:02</td>
<td>1000</td>
<td>Oracle online redo logs and Oracle control file</td>
<td>1A, 1B, 2A, 2B</td>
</tr>
<tr>
<td>2-3</td>
<td>00:00:03</td>
<td>1000</td>
<td>Oracle archived redo logs</td>
<td>1A, 1B, 2A, 2B</td>
</tr>
</tbody>
</table>

Table 5 has the details for the dynamic provisioning pool.

### Table 5. Dynamic Provisioning Pool

<table>
<thead>
<tr>
<th>Dynamic Provisioning Pool ID</th>
<th>Parity Groups</th>
<th>RAID Level</th>
<th>Drive Type</th>
<th>No of Drives</th>
<th>Pool Capacity</th>
</tr>
</thead>
<tbody>
<tr>
<td>ora_datapool_01</td>
<td>1-1 — 1-12</td>
<td>RAID-10 (2D+2D) ×2</td>
<td>600GB 10k RPM SAS</td>
<td>120</td>
<td>30 TB</td>
</tr>
</tbody>
</table>

Table 6 has the details for the virtual volumes created from the dynamic provisioning pool without PCI-e acceleration.

### Table 6. Dynamic Provisioning Pool LUN Information Without PCI-e Acceleration

<table>
<thead>
<tr>
<th>Dynamic Provisioning Pool ID</th>
<th>LDEVs</th>
<th>LDEV Size (GB)</th>
<th>Purpose</th>
<th>Storage Port</th>
</tr>
</thead>
</table>
| 001                          | 00:00:22 — 00:00:3E | 200 | ▪ Oracle system  
▪ Sysaux  
▪ Undo  
▪ Temp  
▪ OLTP application tablespaces | 1A, 1B, 2A, 2B |
Table 7 has the details for the virtual volumes created from the dynamic provisioning pool with PCI-e acceleration.

Table 7. Dynamic Provisioning Pool LUN Information with PCI-e Acceleration

<table>
<thead>
<tr>
<th>Dynamic Provisioning Pool ID</th>
<th>LDEVs</th>
<th>LDEV Size (GB)</th>
<th>Purpose</th>
<th>Storage Port</th>
</tr>
</thead>
<tbody>
<tr>
<td>001</td>
<td>00:00:22 – 00:00:39</td>
<td>200</td>
<td>2nd failure group to host application tablespace</td>
<td>1A, 1B, 2A, 2B</td>
</tr>
<tr>
<td>001</td>
<td>00:00:3A – 00:00:3E</td>
<td>200</td>
<td>Oracle system</td>
<td>1A, 1B, 2A, 2B</td>
</tr>
</tbody>
</table>

Represent the Fusion-io ioDrive2 as a regular block device on fioa-fiod. Each device is 1.2 TB. Use these devices for the following:

- 1st failure group to host application tablespace
- Oracle ASM preferred mirror read

Database Layout

The database layout design uses recommended practices from Hitachi Data Systems for Hitachi Unified Storage VM for small random I/O traffic, such as the ones in OLTP transactions. The layout also takes into account the Oracle ASM best practices when using Hitachi storage.

Base the storage design for database layout needs on the requirements of a specific application implementation. The design can vary greatly from one implementation to another. The components in this solution set have the flexibility for use in various deployment scenarios to provide the right balance between performance and ease of management for a given scenario.

- **Data and Indexes Table Space** — Assign one dynamic provisioning pool for the application data and indexes. The allocated capacity is 30 TB. The small file table space consists of several 30GB data files. Set the tablespace to a small initial size with auto extend enabled to maximize storage utilization from thin provisioning.

- **Temp Table Space** — Place TEMP table space in this configuration in the Data and Indexes ASM diskgroup. Quite a number of small file tempfiles are created within one single small TEMP tablespace. Limit the size of each small file tempfile to 30 GB.
• **Undo Table Space** — Place UNDO table space in this configuration in the Data and Indexes ASM diskgroup. Quite a number of small file undo datafiles are created within one single small UNDO tablespace. Limit the Size of each small undo datafile to 30 GB.

• **Online Redo Logs** — Assign two RAID groups for the database instance.

• **Size Settings** — Set the database block size to 8 KB. Set ASM allocation unit to 1 MB.

• **ASM Preferred Mirror Read** — Configure the ASM Preferred Mirror Read for this reference architecture using one ASM disk group with two ASM failure groups. The first failure group uses the four block devices from PCI-e flash cards. The second failure group uses LUNs or disks from the provisioned ASM disks from the SAN disks.

Table 8 lists the disk mappings from the LUNS to the operating system devices and to the ASM disk groups for the reference architecture without PCI-e flash acceleration.

<table>
<thead>
<tr>
<th>LUN</th>
<th>OS device /dev/…</th>
<th>ASM disk</th>
<th>ASMDG</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>sda</td>
<td>N/A</td>
<td>N/A</td>
<td>Operating system and Oracle software binaries</td>
</tr>
<tr>
<td>1</td>
<td>360060e8013275e005020275e00000001p1</td>
<td>RGDISK01</td>
<td>REDODG01</td>
<td>Online REDO log group and control file</td>
</tr>
<tr>
<td>2</td>
<td>360060e8013275e005020275e00000002p1</td>
<td>RGDISK11</td>
<td>REDODG11</td>
<td>Online REDO log group and control file</td>
</tr>
<tr>
<td>3</td>
<td>360060e8013275e005020275e00000003p1</td>
<td>ARDISK01</td>
<td>ARCHDG</td>
<td>Archived REDO log file</td>
</tr>
<tr>
<td>4 – 32</td>
<td>360060e8013275e005020275e00000022p1 - 360060e8013275e005020275e0000003ep1</td>
<td>DADISK01 – DADISK29</td>
<td>DATADG</td>
<td>Systr</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Undo</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Temp</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Application data</td>
</tr>
</tbody>
</table>
Table 9 lists the disk mappings from the LUNS to the operating system devices and to the ASM disk groups for the reference architecture with PCI-e flash acceleration.

<table>
<thead>
<tr>
<th>LUN</th>
<th>OS device /dev/...</th>
<th>ASM disk</th>
<th>ASMDG</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>sda</td>
<td>N/A</td>
<td>N/A</td>
<td>Operating system and Oracle software binaries</td>
</tr>
<tr>
<td>1</td>
<td>360060e8013275e005020275e000000001p1</td>
<td>RGDISK01</td>
<td>REDODG01</td>
<td>Online REDO log group and control file</td>
</tr>
<tr>
<td>2</td>
<td>360060e8013275e005020275e00000002p1</td>
<td>RGDISK11</td>
<td>REDODG11</td>
<td>Online REDO log group and control file</td>
</tr>
<tr>
<td>3</td>
<td>360060e8013275e005020275e00000003p1</td>
<td>ARDISK01 – ARDISK03</td>
<td>ARCHDG</td>
<td>Archived REDO log file</td>
</tr>
<tr>
<td>N/A</td>
<td>fioa – fiod</td>
<td>FIODISK1 – FIODISK4</td>
<td>PFMDG</td>
<td>ASM preferred mirror read with 1st failure group — FGFIO</td>
</tr>
<tr>
<td>4 - 27</td>
<td>360060e8013275e005020275e000000022p1 - 360060e8013275e005020275e00000030p1</td>
<td>DADISK01 – DADISK24</td>
<td></td>
<td>ASM preferred mirror read with 2nd failure group — FGSAN</td>
</tr>
<tr>
<td>28 - 32</td>
<td>360060e8013275e005020275e00000003ap1 - 360060e8013275e005020275e0000003ep1</td>
<td>DADISK25 – DADISK29</td>
<td>DATADG</td>
<td>Sys, Undo, Temp</td>
</tr>
</tbody>
</table>
ASM Failure Group Configuration

An ASM failure group consists of a group of disks dependent on a single point of failure, such as a controller. Locate each mirror in a different failure group.

ASM normal redundancy is a two-way mirroring, requiring two failure groups. In the configuration with PCI-e flash acceleration, the Fusion-io ioDrive2 devices are under ASM control and make up one failure group. The storage disks on Hitachi Unified Storage VM form the other failure group.

All local disks belong to the same failure group.

When using PCI-e flash acceleration, set the following:

- The Oracle initialization parameter `ASM_PREFERRED_READ_FAILURE_GROUPS` to the ASM diskgroup
- The failure group to the Fusion-io ioDrive2 devices

Server and Application Architecture

The reference architecture uses two server blades in the Hitachi Compute Blade 2000 chassis for the database server. It includes a single logical database server instance running on the two server blades connected through an SMP connector.

There are 32 combined CPU cores with 192 GB RAM. This provides the compute power for Oracle Database to handle complex database queries and a large volume of transaction processing in parallel. See “Key Solution Components” on page 5 for hardware details.

This reference architecture does not use a third party volume manager. It uses Oracle Database 11g R2 ASM volume manager software.

To monitor and manage the database, the database servers, and the storage, this solution uses the following:

- Oracle Database 11g R2 Automatic Workload Repository (AWR) reports
- Hitachi Storage Navigator
- O/S IOSTAT and VMSTAT
Figure 7 on page 20 shows the software stack for the reference architecture without PCI-e flash acceleration. Figure 8 on page 21 shows the software stack with PCI-e flash acceleration.
Figure 8
SAN Architecture

Map the provisioned LDEVs to multiple ports on Hitachi Unified Storage VM. These LDEV port assignments provide multiple paths to the storage system from the host for high availability.

The environment used two Fibre Channel switch modules installed in the Hitachi Compute Blade 2000 chassis. This provides scalability and high availability. “Key Solution Components” on page 5 has host configuration details.

The database server used four Fibre Channel ports, with two ports from the mezzanine card on each server blade. This provides a four-path connection for all LUNs mapped to the server.

The environment used two Fibre Channel switch modules to provide redundant paths for all Unified Storage VM LUNs.

Table 10 provides the zoning details for the SAN. Use the same zoning whether your implementation is with or without PCI-e flash acceleration.

**Table 10. SAN Switch Architecture**

<table>
<thead>
<tr>
<th>Server</th>
<th>HBA Ports</th>
<th>Switch Zone</th>
<th>Storage Port</th>
<th>Switch</th>
</tr>
</thead>
<tbody>
<tr>
<td>Database server</td>
<td>B0-HBA1-1</td>
<td>BS2K_13_B0_HBA1_1_ASE46_142_1A</td>
<td>1A</td>
<td>5300-05</td>
</tr>
<tr>
<td></td>
<td>B0-HBA1-2</td>
<td>BS2K_13_B0_HBA1_2_ASE46_142_1B</td>
<td>1B</td>
<td>5300-06</td>
</tr>
<tr>
<td></td>
<td>B1-HBA1-1</td>
<td>BS2K_13_B1_HBA1_1_ASE46_142_2A</td>
<td>2A</td>
<td>5300-05</td>
</tr>
<tr>
<td></td>
<td>B1-HBA1-2</td>
<td>BS2K_13_B1_HBA1_2_ASE46_142_2B</td>
<td>2B</td>
<td>5300-06</td>
</tr>
</tbody>
</table>

Hitachi Data Systems recommends the use of dual SAN fabrics, multiple HBAs, and host-based multipathing software when deploying this reference architecture. You need at least two paths to provide to critical applications the redundancy required for the following:

- Database hosts connected to two independent SAN fabrics
- SAN fabric to two different controllers of the I/O subsystem
When designing your SAN architecture, follow these recommended practices for a secure, high-performance, and scalable database deployment:

- Use at least two HBAs and place them on different I/O buses within the server. This distributes the workload over the PCI-e bus architecture of the server.

- Use dual SAN fabrics, multiple HBAs, and host-based multipathing software in a business-critical deployment. Connecting two or more paths from the database servers to two independent SAN fabrics provides the redundancy required for critical applications.

- Zone your fabric to handle multiple, unique paths from HBAs to storage ports. Use single initiator zoning. Use at least two Fibre Channel switch fabrics to provide multiple, independent paths to Hitachi Unified Storage VM to prevent configuration errors from disrupting the entire SAN infrastructure.

- For large bandwidth requirements that exceed the port capability of a single HBA, do the following:
  - Use additional HBAs.
  - Use the round robin load-balancing setting for Oracle Enterprise Linux Device Mapper.

**Network Architecture**

This reference architecture has four on-board 1 Gb/sec NIC ports for different types of data traffic. The two NIC ports on each blade connect to two internal 1 Gb/sec Ethernet switches in the chassis.

Figure 9 on page 24 shows the network configuration for the reference architecture environment.
Engineering Validation

This describes the validation of this medium-sized Hitachi Unified Compute Platform Select for Oracle Database solution in this reference architecture for functionality and performance. Testing involved a configuration with and without PCI-e flash acceleration.

Test Methodology

This is the methodology used to validate this solution.

Workload

Testing included test iterations to run simulated and synthetic workloads using the following:

- **Oracle ORION** — This simulates the Oracle database server I/O patterns without installing Oracle or creating a database. Unlike other I/O calibration tools, ORION simulates Oracle Database I/O workloads by design. When used for testing, ORION determined the I/O throughput and bandwidth for the server configuration and the workloads.

- **Swingbench** — This performs a stress test on an Oracle database. The Swingbench load factor is ~313, which fills 1 TB data order entry (OE) application data.

The database size for this configuration is 1 TB.

The database block size is 8 KB. The ASM AU setting is 1 MB, based on Oracle best practice in Metalink Note 810484.1.

The ASM diskgroup configuration is in “ASM Failure Group Configuration” on page 19.

These were the simulated workloads for testing:

- 90% read transactions, 10% write transactions
- 65% read transactions, 35% write transactions

Testing Procedure

To validate the solution, this testing was done on the reference architecture:

- Run the testing procedure on the environment without PCI-e flash acceleration to establish the performance baseline data.
- Run the testing procedure on the environment with PCI-e flash acceleration to provide data to compare to the baseline data.
This is the testing procedure used to test the environment:

- **Small Random Reads Test** — Measured the IOPS using ORION
- **I/O Latency for Small Random Reads Test** — Measured the latency for small random reads using ORION
- **Large Sequential I/O Test** — Measured the throughput of large sequential I/Os using ORION
- **Stress Test** — Use Swingbench to run synthetic OLTP read- and write-intensive workload testing to identify the following:
  - Transactions per second (TPS)
  - Database IOPS
  - Throughput for a given workload
- **SQL SELECT Test** — Perform a full table scan to measure performance results using SQL statements.
  - Measure the throughput when doing a SQL SELECT statement against a 124 GB table (approximately 4.2 billion rows) that was populated using Swingbench. The defined workloads were not applicable for this test.

### Data Gathering

Performance statistics were collected at the following levels:

- **Storage**
  - Hitachi Storage Navigator collected storage performance data
- **Operating System**
  - “iostat” and “vmstat” collected operating system statistics
- **Database**
  - Oracle Automatic Workload Repository report for database performance
  - Swingbench for application-level statistics for the number of executed transactions and response times
- **Standalone tests on the drives/LUNS**
  - ORION for the throughput and IOPS performance for a given set of workload
Test Results

This summarizes the key observations from the test results for the medium-sized Hitachi Unified Compute Platform Select for Oracle Database with and without PCI-e flash acceleration.

Small Random Reads Test

The IOPS increased when using PCI-e flash acceleration over the baseline established without flash acceleration by the following:

- 97% for the 90% read workload
- 104% for the 65% read workload

Figure 10 shows the relative increase in IOPS when using PCI-e flash acceleration over the baseline established without using PCI-e flash acceleration.

![ORION - Relative Increase in IOPS](image-url)

Figure 10
I/O Latency for Small Random Reads Test

For both simulated workloads, using PCI-e flash acceleration decreased latency by 94% from the baseline results established without using PCI-e flash acceleration.

Figure 11 shows the relative decrease in latency when using PCI-e flash acceleration from the baseline established without using PCI-e flash acceleration.

0ns per second when using PCI-e flash acceleration over the baseline established without using PCI-e flash acceleration.
Large Sequential I/O Test

The throughput increased when using PCI-e flash acceleration over the established baseline without using PCI-e flash acceleration by the following:

- 93% for the 90% read workload
- 43% for the 65% read workload

Figure 12 shows the relative increase in throughput when using PCI-e flash acceleration over the baseline established without using PCI-e flash acceleration.

![Figure 12](image-url)
Stress Test

The transactions per second increased for various OLTP workloads when using PCI-e flash acceleration over the baseline established without using PCI-e flash acceleration by the following:

- 272% for the 90% read workload
- 106% for the 65% read workload

Figure 13 shows the relative increase in transactions per second when using PCI-e flash acceleration over the baseline established without using PCI-e flash acceleration.

![Swingbench - Relative Increase in TPS](image)

*Figure 13*
SQL SELECT Test

The throughput increased for a SQL SELECT against a 124 GB table when using PCI-e flash acceleration over the baseline established without flash acceleration by 80%. The two configurations had the following throughputs:

- 2,958.70 MB/sec without using PCI-e flash acceleration
- 5,332.80 MB/sec using PCI-e flash acceleration

Figure 14 shows the throughput when using PCI-e flash acceleration over the baseline established using without PCI-e flash acceleration.
Conclusion

Testing this reference architecture demonstrates a high performance Oracle database server for OLTP workloads. The medium-sized Hitachi Unified Compute Platform Select for Oracle Database does the following:

- ORION test results show the PCI-e flash acceleration improves the following:
  - IOPS and I/O latency of small Oracle random I/Os
  - Throughput of large Oracle sequential I/Os
- Swingbench test results show the PCI-e flash acceleration increases database throughput of OLTP workloads.

Using this reference architecture as a guide, you can design a larger infrastructure that meets your needs by adding the following:

- Additional server blades in the Hitachi Compute Blade 2000 chassis
- Fast MP server blades in the Hitachi Compute Blade 2000 chassis
- Larger 8 GB DIMMs in the Hitachi Compute Blade 2000 chassis
- Additional disks and host ports on Hitachi Unified Storage VM
- Storage monitoring and database management using Hitachi storage plug-ins for Oracle Enterprise Manager
For More Information

Hitachi Data Systems Global Services offers experienced storage consultants, proven methodologies and a comprehensive services portfolio to assist you in implementing Hitachi products and solutions in your environment. For more information, see the Hitachi Data Systems Global Services website.

Live and recorded product demonstrations are available for many Hitachi products. To schedule a live demonstration, contact a sales representative. To view a recorded demonstration, see the Hitachi Data Systems Corporate Resources website. Click the Product Demos tab for a list of available recorded demonstrations.

Hitachi Data Systems Academy provides best-in-class training on Hitachi products, technology, solutions and certifications. Hitachi Data Systems Academy delivers on-demand web-based training (WBT), classroom-based instructor-led training (ILT) and virtual instructor-led training (vILT) courses. For more information, see the Hitachi Data Systems Services Education website.

For more information about Hitachi products and services, contact your sales representative or channel partner or visit the Hitachi Data Systems website.