

Oracle Database Deployments with EMC CLARiiON AX4 Storage Systems

Applied Technology

Abstract

This white paper investigates configuration and replication choices for Oracle Database deployment with EMC® CLARiiON® AX4 storage systems, including the Fibre Channel and iSCSI models, and SnapView™ and MirrorView™ replication technologies. AX4 storage systems offer a cost-effective and highly functional solution for Oracle Database deployments.

July 2008

Copyright © 2008 EMC Corporation. All rights reserved.

EMC believes the information in this publication is accurate as of its publication date. The information is subject to change without notice.

THE INFORMATION IN THIS PUBLICATION IS PROVIDED “AS IS.” EMC CORPORATION MAKES NO REPRESENTATIONS OR WARRANTIES OF ANY KIND WITH RESPECT TO THE INFORMATION IN THIS PUBLICATION, AND SPECIFICALLY DISCLAIMS IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Use, copying, and distribution of any EMC software described in this publication requires an applicable software license.

For the most up-to-date listing of EMC product names, see EMC Corporation Trademarks on EMC.com.

All other trademarks used herein are the property of their respective owners.

Part number H5502

Table of Contents

Executive summary	4
Introduction	4
Audience	4
Terminology	4
CLARiiON AX4 storage systems and Oracle Databases	6
Test environment	6
Test methodology	6
Fibre Channel models and iSCSI models	7
ORION small and big I/O simulation tests	8
OAST OLTP workload test	8
SnapView and MirrorView replication support	9
Conclusion	11
References	11

Executive summary

With the introduction of EMC® CLARiiON's entry-level storage system – CLARiiON® AX4 – Oracle Database deployments targeted at small-to-medium businesses can be done at a price significantly lower than other products, while offering the same scalability and functionality of the more powerful storage models. With AX4, users can provide a versatile and cost-effective solution for organizations looking to deploy the Oracle Database with storage systems.

This white paper analyzes how CLARiiON AX4 storage systems can be configured and used in Oracle Database deployments, as well as demonstrates how the Oracle Database performs under different storage settings. The storage settings explored are for:

- Fibre Channel and iSCSI protocols
- Replication using SnapView™ and MirrorView™

Introduction

The CLARiiON AX4 storage system is designed with two key objectives: ease of use and affordability. Most components are hot swappable with minimal configuration required, and storage management can be done through a simple and intuitive graphical web interface. Also, its low-cost approach makes it an ideal choice for entry-level storage environment and virtualized server environments.

Leveraging Intel Xeon processors, the AX4 storage system is offered as a Fibre Channel (FC) or an iSCSI model, where iSCSI opens up an opportunity for further cost reduction by leveraging the available IP network. FC requires FC switches and HBAs but offers a high-bandwidth data-transfer option suitable for large size I/O workloads. The two FC front-end ports or two iSCSI front-end ports support the following:

- For Fibre Channel, 1 Gb/s, 2 Gb/s, and 4 Gb/s Fibre Channel with auto-negotiation
- For iSCSI, 10 Mb/s, 100 Mb/s, and 1 Gb/s iSCSI with auto-negotiation

Navisphere's array-based replication software products are optional features that extend AX4's functionality. For data replication within the same array, SnapView can be used. For replication to remote arrays, MirrorView can be used. For remote replication with non-EMC arrays, SAN Copy™ can be used. Other features available on AX4 are Replication Manager, RepliStor®, and Navisphere® Analyzer.

Audience

This white paper is intended for EMC employees, customers, partners, IT managers, storage architects, database administrators, and consultants who are looking to deploy Oracle Database systems with EMC CLARiiON AX4 storage systems.

Terminology

Bandwidth: The average amount of read/write data in megabytes that is passed through the storage system per second.

Disk-array enclosure (DAE): A CLARiiON term for the shelf enclosure that houses disk drives, power supply, and link control cards.

Host bus adapter (HBA): A module that plugs into the server and acts as the interface between the computer's internal bus and the network connection.

Logical unit number (LUN): A unique identifier that is used to distinguish among logical storage objects in a storage system.

MirrorView: The CLARiiON array software that supports continual and automatic propagation of changed data from one CLARiiON system to another CLARiiON system that are both on the same storage area network.

MirrorView/Asynchronous (MV/A): Changes to production data are batched up and periodically sent across from the CLARiiON storing the production data to the secondary in a manner that guarantees that the protection array data is always from a single point in time, matching the production data content from some time point in the past.

MirrorView/Synchronous (MV/S): A mode of change propagation under MirrorView where the change to production data is not acknowledged back to the server until the propagation is guaranteed. This ensures no possibility of data loss under normal MirrorView operation.

Online transaction processing (OLTP): A type of information processing with emphasis on responsiveness to satisfy many concurrent user requests, where the I/O pattern for these requests is typically small and random in nature. Each request is considered to be a transaction.

Oracle Automated Stress Testing (OAST): An automated test suite designed to build an OLTP type workload to stress and validate the robustness of systems using Oracle Database. It creates tables, performs stress test runs, and outputs transaction-related performance data.

RAID group (RG): A CLARiiON term for a collection of fault-tolerant disks that are bound into a group. One or more LUNs can be created from the available space in a RAID group to be used by applications for storing data that must survive certain physical disk failures within the group.

Redundant Array of Independent Disks (RAID): RAID technology groups separate disks into one logical unit number (LUN). Data is distributed across these disks to improve reliability and performance. The AX4 series supports RAID levels 1/0, 3, and 5.

Response time: The average time in milliseconds that it takes for one request to pass through the storage, including any waiting time.

Serial Advanced SCSI (SAS): A high-performance I/O technology that uses a serial point-to-point interface instead of a parallel bus interface, as with parallel SCSI.

Serial Advanced Technology Attachment (SATA): An enhancement of ATA (Advanced Technology Attachment) that uses thinner cables and provides better performance over parallel drive technology.

SAN Copy: The software that allows the content of a CLARiiON LUN in one array to be bitwise-copied to another LUN that the CLARiiON system can reach on the SAN. This includes the ability to bitwise-refresh the content to the target LUN after the first complete copy (incremental SAN Copy).

SnapView: The software to create point-in-time exact bit copies of a LUN content within a single CLARiiON array.

SnapView clones: A clone is a complete copy of a source LUN. You specify a source LUN when you create a clone group. The source LUN can be any conventional LUN, any destination LUN of a SAN Copy, or any MirrorView primary or secondary image. The copy of the source LUN begins when you add a clone LUN to the clone group. The SnapView software assigns each clone a clone ID. The ID remains with the clone until you remove the clone from the group.

SnapView snapshots: A snapshot is a virtual LUN that allows a secondary server to view a point-in-time copy of a source LUN. You determine the point in time when you start a SnapView session. The session keeps track of the source LUN's data at a particular point in time.

Storage area network (SAN): A Fibre Channel or iSCSI storage network used to connect servers and shared storage systems.

TCP/IP offload engine (TOE): A NIC with TOE offloads processing of TCP/IP segments from the host CPU. For systems with high CPU usage, a TOE may provide better performance.

Throughput: The average number of read/write requests in I/Os that is passed through the storage per second.

Transactions per minute (TPM): The measured number of transactions completed by an OLTP database over a 1-minute interval.

CLARiiON AX4 storage systems and Oracle Databases

The CLARiiON AX4 model offers configuration choices that enhance flexibility and functionality. These choices also affect cost and performance levels of Oracle Database deployments, which are critical factors in decision making. The configuration choices tested in the context of Oracle Databases are:

- FC connectivity and iSCSI connectivity
- Remote replication using SnapView and MirrorView

Test environment

Server machines: Dell PowerEdge 6850

OS	RHEL4 U4 x86_64 (2.6.9-42.ELsmp)
CPU	2 x Xeon MP 3.16 GHz
Memory	8 GB
Network adapter	Intel PRO/1000PT Gigabit NIC
FC adapter	QLogic 4 GB FC HBA
Naviagent/CLI	naviagentcli-6.26.5.0.71-1
PowerPath [®]	EMCpower.LINUX-5.0.0-156

Server machines for remote replication: 2 x Dell PowerEdge 2850

OS	RHEL4 U4 x86 (2.6.9-34.ELsmp)
CPU	2 x Xeon 3.2 GHz
Memory	8 GB
Network adapter	Intel PRO/1000PT Gigabit NIC
FC adapter	QLogic 4 GB FC HBA
Naviagent/CLI	naviagentcli-6.26.5.0.71-1
PowerPath	EMCpower.LINUX-5.0.0-156

Storage systems: EMC CLARiiON AX4-5 (FC), AX4-5i (iSCSI)

Processors	2
Memory size	1 GB per SP
Number of disks	60 SAS 146 GB @ 15k rpm
Base software	03.23.050.3.526

Database software: Oracle 11.1.0

I/O load simulation	Oracle ORION
Transaction stress test	Oracle OAST

Test methodology

A CLARiiON AX4-5 storage system was used for the FC connection, and an AX4-5i storage system was used for the iSCSI connection. Each LUN was created from five disk spindles with RAID 5 (4+1) configuration. Oracle Automatic Storage Management (ASM) was used to manage disk groups for the database, as well as EMC PowerPath for multi-path load balancing. For 8 KB and 1 MB I/O load simulation, the Oracle ORION tool was used. For transactional workload simulation, the Oracle OAST database was created on the storage system using six LUNs. The layout in Figure 1 on page 7 illustrates the LUN configuration for all OAST tests performed in this paper. For replication testing, RMDATA3 and RMDATA4 were converted to one FLASH diskgroup for the flash recovery area and one ARCHIVE diskgroup for archive logging.

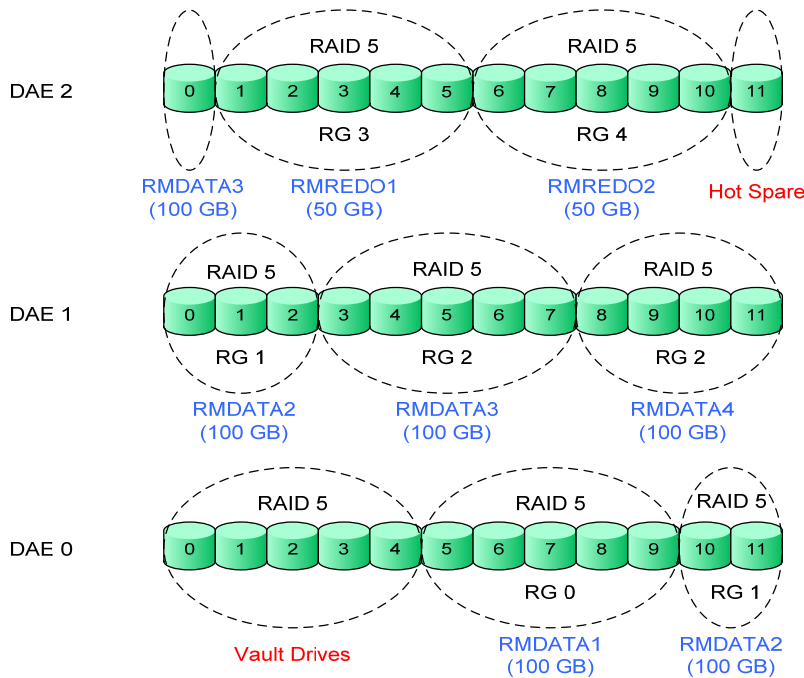


Figure 1. Disk layout for the Oracle OAST test

Fibre Channel models and iSCSI models

The decision to use the Fibre Channel or iSCSI protocol may be affected by the following conditions:

- Current network environment
- Current budget and staff skill sets
- Application performance requirement
- Disaster recovery plans including remote replication

Fibre Channel offers up to four times better application performance than iSCSI but needs an FC SAN environment including FC HBAs on servers. With iSCSI, users can minimize costs by leveraging the existing Ethernet network and NIC ports on servers. The relatively higher CPU usage level required by iSCSI can be offloaded by using an iSCSI HBA that includes a TCP/IP offload engine (TOE). Because the AX4 storage system does not support remote replication through iSCSI, deployments requiring remote replication need to use FC models. Also, for workloads requiring bandwidth of 100MB/s or more, FC is usually the preferred choice.

The following details how Oracle Database performs under FC models and iSCSI models. The ORION tool was used to measure storage I/O performance. The result shows average I/O performances for driving a mixture of 8 KB small I/O and 1 MB big I/O workloads against an AX4-5 system using a configuration ranging from five to 55 disk drives. Small I/O is based on random read operation and the big I/O is based on sequential read operation. The comparison results are relative, where the FC results are scaled to be 100 percent. The ORION I/O simulation test results are shown in Figure 2 on page 8. For the small I/O tests (8 KB), simulating the more typical OLTP type deployments, the key metric is the I/O operations processed per second (IOPS). For the large test (1 MB), simulating the common data scan type I/O in Decision Support Systems (DSS) workloads, the key metric is sustained MB/s of data delivered by the storage system.

ORION small and big I/O simulation tests

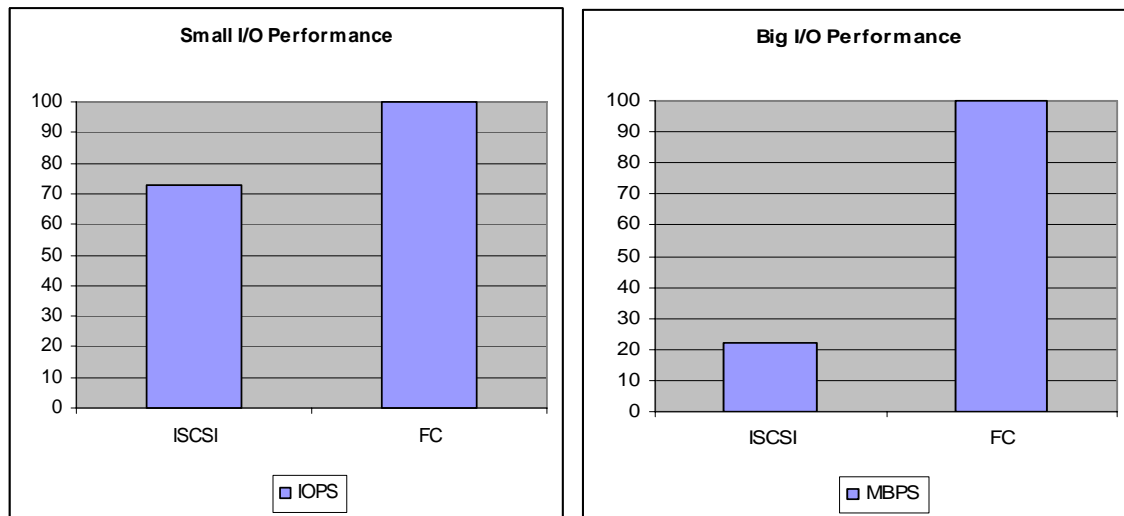


Figure 2. ORION small and big I/O simulation tests

The test results suggest that the FC connection performs 27 percent better than iSCSI on average for small I/Os. For big I/Os, the differences between FC and iSCSI is even larger, with FC supporting up to four times higher bandwidth than iSCSI for data stored in the same LUNs.

As ORION is purely an I/O subsystem stressing workload, where the limiting factor to the performance observed is primarily what the storage system can sustain, the small I/O IOPS comparisons suggest that while iSCSI software drivers carry a higher resource overhead by comparison to FC drivers, when the workload is pushing to the limit of what the storage system can sustain, iSCSI driver overhead becomes less of a factor.

In contrast, for typical test workloads that try to push large I/Os (such as 1 MB), the FC HBAs, driver protocols, higher wire bandwidth, and AX4-5 front side driver operation combine to allow significantly higher MB/s to be sustained.

OAST OLTP workload test

The OAST runs on top of the Oracle Database to measure transactional performance that is closer to the real database deployment environment. For the test, a warehouse size of 500 was created and 100 users were simulated for the duration of 1 hour. The TPM and CPU usage numbers output by OAST is not intended for any benchmark use but should be noted as a relative performance comparison between FC and iSCSI. The OAST OLTP workload test results are shown in Figure 3 on page 9.

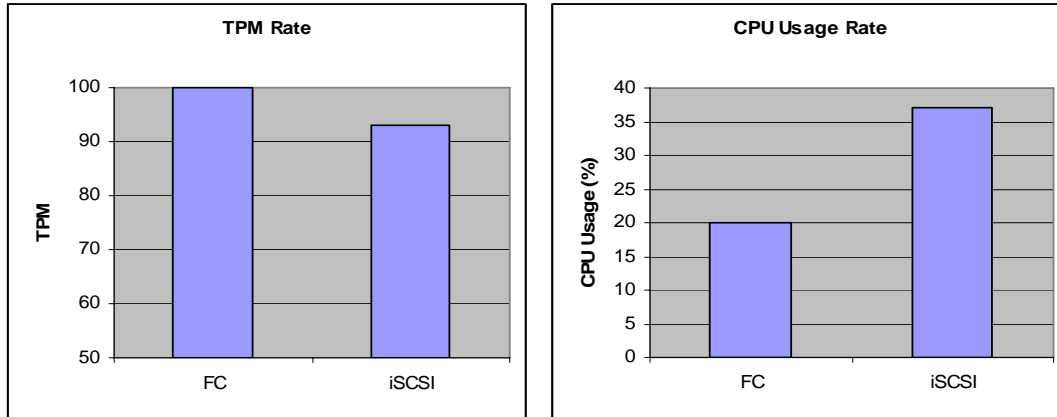


Figure 3. OAST OLTP workload test

When compared to iSCSI, the FC connection yielded about a 10 percent higher TPM rate. The server CPU usage rate was about 17 percent higher for iSCSI, meaning that more CPU cycles were needed in order to process the iSCSI protocol. As more spindles are used, the TPM performances are expected to increase, and the TPM difference between FC and iSCSI is also expected to increase.

SnapView and MirrorView replication support

With the optional Navisphere Management Suite, most of the replication software available on the CLARiiON CX series systems is available on the AX4 storage systems. That replication software includes SnapView, MirrorView, SAN Copy, Replication Manager, and RepliStor. The benefits of using the CLARiiON storage-based replication technologies in Oracle Database deployments are as follows:

- Enhance production-site business continuity support
- Enhance remote-site disaster protection and recovery support
- Offload the backup process to satisfy service level requirements of the operation
- Rapid database cloning for repurposing

Two of the most widely used replication tools were tested: SnapView, for a point-in-time snapshot and cloning of data within the same CLARiiON storage array, and MirrorView, for disaster recovery mirroring between two CLARiiON storage arrays. Testing was focused on ensuring consistency and restoring from a point-in-time backup of the database. Oracle-based replication technology such as RMAN or hot backup was not used.

The test was carried out using Fibre Channel models and SAS disk drives. Using six RAID 5 LUNs managed by Oracle ASM, a database that simulates about 400,000 rows of bank transactions was created. Then, snapshots, clones, and remote mirrored pairs were created on the appropriate primary and secondary storage arrays with ongoing updates to the database. Figure 4 on page 10 illustrates how the test was set up for replication.

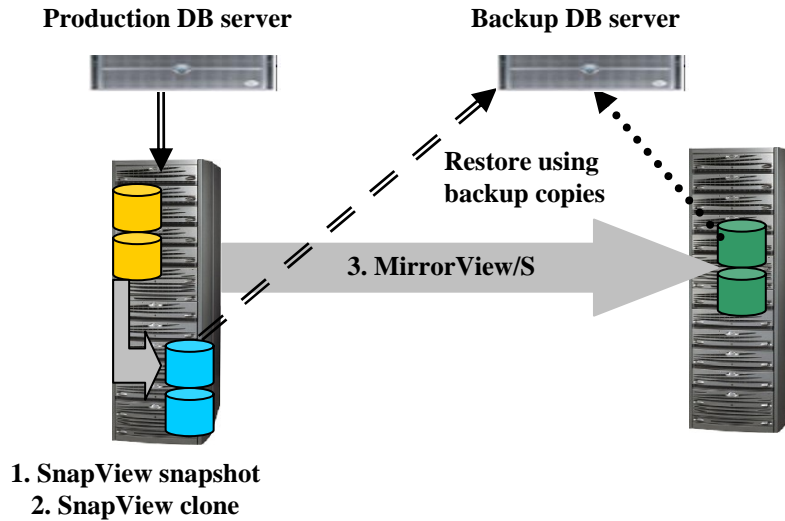


Figure 4. SnapView and MirrorView test setup

To check the validity of replicated data, the committed System Change Number (SCN) of the database was recorded and compared. The detailed steps and results of SnapView snapshots, SnapView clones, and MirrorView/Synchronous (MV/S) operations are as follows:

1. Enable archive logging and flashback in Oracle Database.
2. While the production database is being updated, create a restore point for flashback in Oracle Database.
3. For the SnapView snapshot, create and start a snapshot sessions of all LUNs except ARCHIVE.
For the SnapView clone, create clone groups and fracture when synchronized except ARCHIVE.
For MV/S, create a remote mirror and fracture when synchronized except ARCHIVE.
4. Perform a log switch in Oracle Database to flush the latest updates and execute replication for ARCHIVE.
5. For the SnapView snapshot, add snapshots to the backup host storage group and activate the snapshots.
For the SnapView clone, add clones to the backup host storage group.
For MV/S, add secondary images to the backup host storage group and promote the secondary images.
6. Scan and mount the disks to the backup host.
7. Issue flashback to the previously created restore point and check status.
8. Issue recovery (roll-forward) from the archive log and check status.
9. As shown in the results table, all replication features were restored in the backup host successfully.

RESULTS			TIMES RUN
SnapView snapshot	SnapView clone	MV/S	
Successfully restored and rolled forward	Successfully restored and rolled forward	Successfully restored and rolled forward	5

With the advanced replication features of the CLARiiON AX4 storage systems, a simplified consistent backup can be created without using host-based backup technologies. This relieves host servers from high server resource demand during backup and ensures that the database performance is not impacted for service. Depending on your deployment environment, SAN Copy, Replication Manager, RepliStor, and the appliance-based RecoverPoint can also be used with the AX4 models.

Conclusion

With CLARiiON AX4 storage systems, Oracle Database deployments can be done with either the high-performing FC models or the cost-effective iSCSI models. By utilizing advanced AX4 replication software, storage resources can be leveraged to make backups within or outside the storage system. Further choices like SAS or SATA drive offerings, and Navisphere Analyzer and Quality of Service Manager tools can enhance flexibility and performance monitoring capability for Oracle Database deployments. For more affordable and reliable Oracle Database deployments, the CLARiiON AX4 storage system proves to be a highly recommended choice.

References

EMC product information can be found at the following locations on EMC.com:

- *EMC CLARiiON AX4 Storage System* datasheet
<http://www.emc.com/collateral/hardware/data-sheet/H4097-emc-clariion-ax4-ds.pdf>
- *Reducing Storage Costs and Complexity with EMC CLARiiON AX4* white paper
<http://www.emc.com/collateral/hardware/white-papers/H4111-reduce-stor-cost-cmplxty-emc-clariion-ax4-wp.pdf>
- EMC SnapView product page
<http://www.emc.com/products/detail/software/snapview.htm>
- EMC MirrorView product page
<http://www.emc.com/products/detail/software/mirrorview.htm>