



Using Oracle Automatic Storage Management with Pillar Data Systems Axiom Storage System

Best practices for deploying Oracle Database 10g on Pillar storage

March 2006

Table of Contents

Introduction	2
Purpose	2
Pillar Axiom Storage System Overview	3
Quality of Service (QoS)	3
Advanced Backup and Recovery	4
Axiom LUN Provisioning	4
Axiom Path Manager (APM)	5
Oracle Database 10g	6
Automatic Storage Management	6
ASMLIB Option	6
Discovery	7
Oracle Flash Recovery Area	7
Multiple Databases Sharing One Common Storage Pool	8
Best Practice Recommendations	10
Conclusion	11

Introduction

Most companies face an increasing challenge with rapidly growing demands on their system requirements while at the same time looking to lower costs. These conflicting demands are often solved by looking to GRID computing for reduction of both cost and management overhead without compromising the performance and availability of these systems.

With the release of the Oracle Database 10g, Oracle introduced several new features that simplify storage management for the database environment. The two features that this paper focuses on are Automatic Storage Management (ASM) and the Flash Recovery Area. When used together, these new features offer significant savings in both the cost to deploy an Oracle database 10g environment and reduction in the ongoing management overhead.

Combining these new features in the Oracle Database 10g with unique storage features found in the Pillar Data Systems' Pillar Axiom storage system enables additional cost savings. This paper provides an overview of the Pillar Axiom architecture and explains how its unique Quality of Service (QoS) feature complements the performance and ease of management provided by the Oracle database.

This paper shares best practices for deploying these new technologies in a simple, easy-to-manage manner. Oracle and Pillar have provided advancements in their technology that, when combined, can return significant benefits to performance and management overhead.

Purpose

This paper is written for both storage and database administration staff. It explains how to use Pillar Axiom block storage with Oracle 10g ASM to achieve optimal performance. As a technical paper, it is intended to provide both an overview and technical details that are the basis for following best practices.

The paper starts with an overview of the Oracle and Pillar technologies. Next, it addresses best practice recommendations. Finally, the paper provides additional information in the appendix that shares details on a specific installation.

Pillar Axiom Storage System Overview

The Pillar Axiom™ storage system delivers enterprise-class high availability for Oracle Database 10g environments. The system combines cost-effective serial ATA (SATA) disk drive technology with an intelligent Quality of Service (QoS) that delivers high performance at much lower price than what other storage vendors provide. The Pillar Axiom system is designed to eliminate single points of failure with redundant system components:

- The Pilot is a dual-redundancy policy controller that performs the management function for Pillar Axiom storage systems
- The Slammer is a dual-redundancy storage controller that virtualizes the storage pool for Pillar Axiom storage systems and moves and manages data.
- The Brick is a storage enclosure that house two RAID controllers and thirteen disk drives, including one hot-swappable spare

The modular architecture of the Pillar Axiom system allows you to quickly replace any failed components without disrupting or slowing system performance. Bricks are available with low-cost SATA disk drives organized into two sets of six-disk RAID-5 groups. A local hot spare disk drive is available to replace a failed disk, allowing RAID rebuilds to begin immediately and restore data in hours instead of days.

Sophisticated software layered on top of the hardware also assures high availability. Software processes on each Slammer control unit (CU) constantly communicate on status and performance. The Pillar Axiom system uses a double-safe write system and secures the I/O in battery-backed, non-volatile RAM (NVRAM), so that the I/O is safe in case of external power loss. The Pillar Axiom system's redundant CU architecture secures the I/O in both CUs before it is acknowledged as a complete transaction. The write is secured in two places, so only a catastrophic system event can affect write integrity.

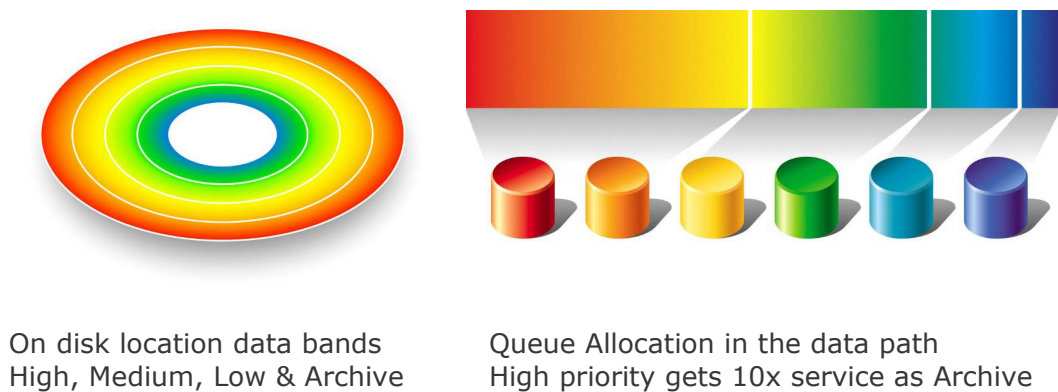
Quality of Service (QoS)

In most database layouts, disk drive I/O performance is established by the physical disk drive location of the data blocks to be accessed. The outer tracks of a disk drive deliver the highest data rates for two reasons:

- The head does not need to move as far to access a given number of blocks
- The outer tracks move more data past the head because of the larger circumference

With SATA disk drives, data transfer rates are about 100 percent higher on the outer section of a disk drive than on the inner section. The Pillar Axiom system supports four tiers on a single storage platform. The QoS also includes queue prioritization so that high priority reads and writes are given 10 times more CPU cycles than archive priority. In other words, QoS is a combination of where the data location on the disk drive and the priority it is given throughout the data path. A set of policies are created to govern the QoS for each filesystem or LUN. These policies determine how data is laid out on the disk drive so that data is laid out on the portion of the disk that best supports its performance priority.

Diagram 1: Quality of Service = Disk Zones and Queue Prioritization



Advanced Backup and Recovery

The Pillar Axiom storage system offers rapid, non-disruptive backup and recovery options for both NAS and SAN. These operate in conjunction with Recovery Manager (RMAN) as well as other Oracle backup solutions. For NAS, the Pillar Axiom storage system offers data protection flexibility by supporting both file-level snapshots and Network Data Management Protocol (NDMP). For SAN, the Pillar Axiom storage system provides snapshot backup and restore capability at the LUN level.

The Pillar Axiom system complements the Oracle Database 10g Grid benefits by providing high availability and optimal performance with lower management and acquisition costs. The system also enables sharing of storage resources across both NAS and SAN attached storage interfaces. Both NAS and SAN can be combined on the same system with a filesystem device being placed on top of the block-based storage. Each Slammer is a redundant active/active pair of control units (CUs) for NAS attachment that serve as the front-end caching and buffering system. The write cache is mirrored across CUs to provide both high availability and high performance. Each Slammer is either NAS or SAN yet shares the back-end storage pool with storage allocated from both SAN and NAS Slammers.

Storage is provisioned through the Pilot, which runs Axiom Storage Manager software. As storage is allocated for either a LUN or a filesystem, the administrator selects criteria regarding the access pattern and Quality of Service (QoS). That places the device in one of four classes of service which leverage both the location on the disk drives (outer tracks being higher performance than inner) and the queuing priority in the data path (both the Slammer and Brick). This capability ensures that targets or filesystems do not interfere with the performance of others.

Axiom LUN Provisioning

The Axiom Storage Manager user interface can be leveraged to provision LUNs with different QoS to optimize database performance. During LUN creation, the administrator can define the LUN profile level of high, medium, low, or archive for each target. As the available space in one storage tier runs out, it is allowed to grow into the storage tiers above its existing priority. So a LUN set to archive QoS can grow into low and then into medium if the low QoS band runs out. However, since the LUN cannot grow into the lower QoS bands, it will be ensured an equal or better performance as it grows.

Axiom Path Manager (APM)

The Axiom Path Manager provides a multi-path capability for LUNs defined on the Pillar Axiom system. It is an OS driver that provides both channel failover protection and performance enhancement in that it distributes I/O across multiple channels. It is a software option provided with the Pillar Axiom system at no extra charge and therefore reduces total deployment cost.

Oracle ASM and Pillar APM software driver complement each other. By system design, a Pillar Axiom LUN has four Fibre Channel paths for read/write. The storage administrator may restrict channel paths to a LUN using CLI commands or the Axiom Storage Manager GUI. Oracle ASM relies on the OS drivers, like the Axiom Path Manager, to resolve multiple LUN paths to a device. The Pillar APM driver loaded on a Linux machine masks multiple SCSI controller paths to a LUN. The driver load balances I/O traffic directed to a given LUN.

Oracle Database 10g

The Oracle Database 10g™ provides several new features that make management of database environment much easier than in previous releases. Automatic Storage Management (ASM) and the Flash Recovery Area are two new features in Oracle Database 10g that simplify, automate, and optimize storage management for the database.

Automatic Storage Management

Automatic Storage Management (ASM) is a feature in Oracle Database 10g that provides the database administrator with a simple storage management interface that is consistent across all server and storage platforms. As a vertically integrated filesystem and volume manager purpose-built for Oracle database files. ASM combines the performance of direct asynchronous I/O with the easy management of a filesystem. ASM provides capability that saves the database administrators (DBAs) time and provides flexibility to manage a dynamic database environment with increased efficiency.

The primary benefit of ASM is that the DBA has significantly more control of the storage resources on which the database is deployed. The DBA manages a few disk groups instead of hundreds or possibly thousands of data files stored on many volumes. The DBA is also less dependent on the system administrator needs less interactions to provision additional storage. Server or storage administrators create the LUNs to the specifications requested by the DBA and set permissions so that Oracle database processes can have access to read and write. The DBA may create ASM disk groups using newly provisioned LUNs and/or add LUNs to existing ASM disk groups as required.

Although ASM does offer the ability to provide redundancy (mirroring) protection, this protection can be provided by an external disk array such as the Pillar Axiom. Another benefit of ASM is its ability to stripe data across a number of channels or disks and maintain even distribution of extents (stripes) as storage configurations change. This unique ASM capability is complementary to the I/O distribution provided within the Axiom storage layer. Double striping through ASM and the Axiom provide equal or better performance than either one in isolation.

ASM is not a general-purpose filesystem and can be used only for Oracle data files, redo logs, and control files. Files in ASM can be created and named automatically by the database (by use of the Oracle Managed Files feature) or manually by the DBA. RMAN is the primary interface to backing up ASM databases. RMAN or ^{third} party backup software that leverages the RMAN interface can be used for backup and recovery.

ASMLIB Option

Automatic Storage Management (ASM) has an extension interface library called ASMLIB. It is a storage management interface for ASM that enables more efficient and capable access to disk groups on the Linux platform.

As an extension to the core ASM features, Oracle has developed a kernel-level ASMLIB driver on Linux to address some platform-specific issues. Using this driver is

optional. It is considered best practice for Linux because it provides persistent binding and avoids having to set privileges that might be confused if a reboot finds devices in a different sequence.

ASMLIB includes:

- Oracleasm: the ASM library
- Oracleasm-support: utilities needed to administer ASMLIB
- Oracleasm: a kernel module for the ASM library

Each Linux distribution has its own set of ASMLIB packages, and within each distribution, each kernel version has a corresponding oracleasm package.

The ASM instance is created before disk groups are constructed.

Discovery

The ASMLIB driver scans and discovers LUNs on all available paths on the host. After discovery finishes, the new devices are marked for use by ASMLIB. Each ASM device is given a unique name.

A disk group is also given a unique name when it is created. Disk group names should reflect function and not storage class (QoS) type because they could be changed over time. ASM can stripe I/O across disk group members in 1M stripe-width size. It is a good idea to employ optimization techniques on the Linux machine to ensure that database I/O size closely matches the I/O buffer size of the operating system.

Previously discovered Pillar Axiom LUNs can be added manually to the intended disk group. As a best practice, all disks in a diskgroup should have the same Pillar Axiom performance profile.

Oracle Flash Recovery Area

Flash Recovery Area is another new feature in Oracle Database 10g. It is a self-managed pool of storage configured to hold full backups, archived logs, and incremental backups. The database manages the deletion of older files to make space for newer backups, using available space as efficiently as possible.

The Pillar Axiom medium and low QoS LUNs are recommended for the fastest restore times possible. Typically, the Flash Recovery Area would be two to three times the size of the database work area, depending on how far back in time the customer may wish to recover. Therefore, the QoS for Recovery Area diskgroup might best be set at the archive QoS band so it is able to grow into the low QoS to meet the required capacity.

Multiple Databases Sharing One Common Storage Pool

As a validation of Oracle 10g databases on a Pillar Axiom SAN storage system, a joint project using a combination of production databases and test databases on the same Pillar Axiom storage system. This project demonstrates the advantage of the Pillar Axiom QoS technology. The set up included two disk groups for each of these two databases:

1. Database work area
2. Flash Recovery Area

For the production database, the work area on an ASM disk group was comprised of high-priority LUNs and the recovery area on a disk group of medium-priority LUNs. For the test database, the work area on an ASM disk group was comprised of the low-priority LUNs and the recovery area for the test database was configured on an ASM disk group comprised of archive-priority LUNs.

Both databases were put under load using a workload generator called Swing Bench. That workload enables one to drive a sales application with a variable number of end users and a mixed workload of read vs. writes. The configuration was run with 50 end users and a 50/50 read-to-write ratio. The Swing Bench application displays transactions per minute as well as the rate of several types of transactions taking place within the application environment.

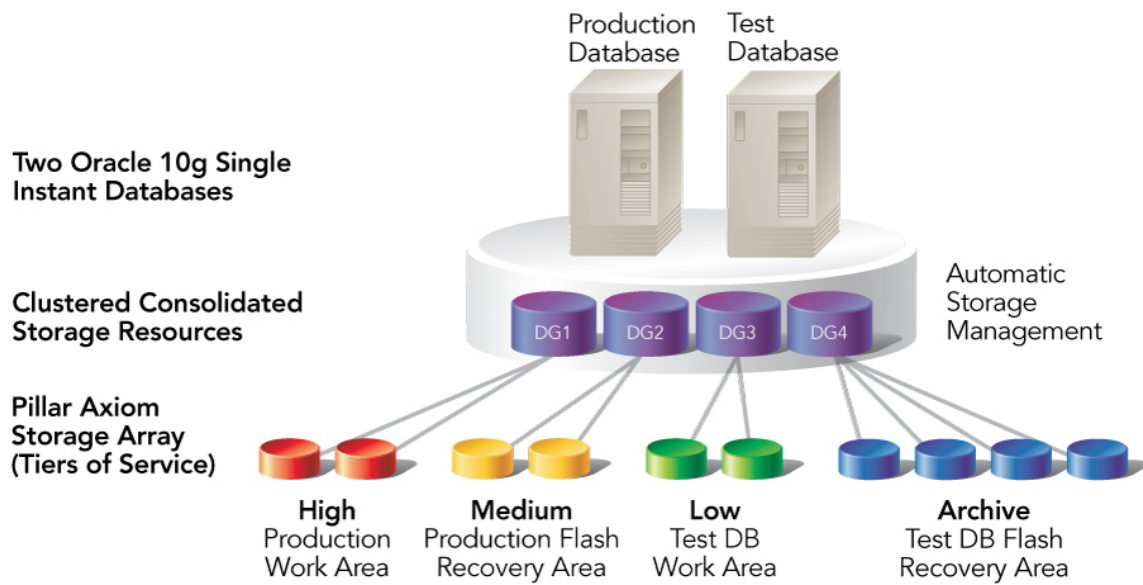
The interesting observation, which highlights how the QoS works, was that the production database was running the same workload as the test database. While the production database was doing about 6,000 TPM, the test database was doing about 4,000 TPM. To further highlight how the QoS works, when the production database was shut down, the test database increased to about 5,200 TPM.

This project was done with the Oracle 10g database release 2 on the Pillar Axiom system running firmware release 1.3. It served as a validation of the use of both ASM and the Flash recovery area and highlighted the benefit of combined QoS and Oracle 10g grid features.

It is not considered a best practice to mix LUNs of archive-priority with high-priority disks in the same disk group, unless the administrator is migrating a disk group from one storage class to another. In that case, it is a temporary state while the DBA is doing an ASM disk group add disk/drop disk to make the transition. In this project, the unique ASM feature was used to migrate the Flash Recovery Area from medium QoS disks to archive QoS disks. This was done while the database was up and running under the load of about 6,000 TPM using the Grid Control interface to Automatic Storage Management to manage the ASM disk groups and leveraging the Axiom Storage Manager software to monitor the storage system. This project highlighted the agility and ease of management that this combination of the two technologies provides. It was an impressive display that not only validated the combination, but also showed the capability of both these technologies to provide a powerful virtualization and an easy provisioning process.

The following diagram shows this set up described above.

Diagram 2: Multi Databases on Tiered Storage



Best Practice Recommendations

Pillar and Oracle recommend the following best practices when deploying Oracle databases using ASM on Pillar Axiom storage systems:

- Configure ASM disk groups with external redundancy using the Pillar Axiom Storage Manager
- Use more than one LUN per disk group
- Use Axiom Path Manager multi-path software for high availability
- Configure disk groups with storage targets of the same performance and I/O capability as others in the same disk groups
- Use multiple disk groups to leverage the multiple tiers of QoS for optimal performance or multi-tenancy

It is considered a best practice for ASM to offload data redundancy when it is available on an external storage system. Not mirroring on the host reduces CPU overhead and additional CPU cycles are available to enhance the database operation.

Having more than one LUN per disk group enables host-level I/O queuing to help avoid contention. Host-based striping also contributes to the overall I/O distribution of the database system by spreading the load across multiple channels. Combining these two forms of striping leverages going wide at multiple points in the stack and enables fewer bottlenecks that might slow performance.

The Axiom Path Manager (APM) driver provides channel load balancing and fail over functions for the Pillar Axiom system. Multiple redundant channel configurations provide higher resiliency and protection for the Oracle database.

It is also a best practice to make sure that all devices in a given disk group have the same QoS, capacity, and performance characteristics. Otherwise, performance will degrade to the lowest common denominator.

The Pillar Axiom QoS can enable multi-tenancy capability of four QoS storage priorities that can be applied to help the Oracle 10g databases can reach higher levels of performance. If there is a single database on one Pillar Axiom system, one can build two disk groups with the following associations:

- Database Work Area placed in a disk group built with medium QoS devices
- Flash Recovery Area on a disk group built with archive QoS devices

For most deployments, it is recommended to leverage different Oracle files onto different QoS targets. With Oracle ASM, it is considered a best practice to create disk groups with LUNs of the same QoS. If there are multiple databases running on a single Pillar Axiom system, the QoS can be leveraged to ensure that one database Service Level Agreement (SLA) is not adversely impacted by the activity of another database. The configuration described in the Appendix is an example of how this might be set up.

Storage provisioning on the Pillar Axiom system is further simplified by adding capacity with the Pillar Axiom auto-grow feature. When used for database disk groups, this can be done without the storage administrator's intervention. The

database administrators would issue the resize command to have the disk group discover the extra capacity. However, if the Pillar Axiom auto-grow feature is set on a LUN device, it is best to have all devices in the same disk group have this set. Therefore, when the DBA issues the command to resize all the devices in the diskgroup, they all will be expanded at the same time and avoid any unneeded rebalancing.

One final best practice is to name disk groups based on their function and not their storage QoS. Because one could add disks of a new QoS tier and drop the disks to change a disk group from one storage class to another, the QoS class would not always be the same over an extended period of time. This can be done without having to take the database offline.

Conclusion

Automatic Storage Management and the Flash Recovery Area provide significant savings of time, cost, and resources for Oracle Database 10g environments. These features provide significant additional cost savings and ease of management when combined with the Quality of Service feature of the Pillar Axiom storage system. This paper has provided several examples of best practices for extending these benefits further.

Together, Oracle and Pillar offer technology that, when combined, enables new ways to reduce costs of deployment and ongoing management of the database and storage grid. Using ASM and the Flash Recovery Area provide the DBA with new ways to reduce time spent managing storage resources for the database. Leveraging these two features with the Pillar Axiom multiple QoS tiers achieves greater performance optimization and reduces costs.

The best practices outlined in this paper provide several ways for further optimization of performance while also leveraging the cost savings and reduction of the management cycles of the DBA, system and storage administrators. These practices address the challenge of reducing costs while supporting rapid growth demands on the database environment.

Copyright © 2006 Pillar Data Systems and Oracle. All Rights Reserved.

Pillar Data Systems, Pillar Axiom, and the Pillar logo are all trademarks of Pillar Data Systems. Other company and product names may be trademarks of their respective owners.

The information in this publication is provided "as is" and is subject to change without notice.