Best Practices for Optimizing Storage for Oracle Automatic Storage Management with Oracle FS1 Series Storage

ORACLE®

Table of Contents

Introduction

Oracle Automatic Storage Management capability of Oracle Database might conflict with SAN or NAS storage system management operations. This paper describes the coengineering advantage of using Oracle FS1 Series storage systems with Oracle Automatic Storage Management to drive performance at flash speed and latency, automate storage provisioning, and simplify management. In addition, the paper includes best practices for leveraging application profiles and the QoS Plus feature of Oracle FS1 Series storage systems. QoS Plus delivers quality of service (QoS) capabilities that automatically provision storage systems to get and maintain optimum performance and efficiency using flash in the Oracle Automatic Storage Management environment.

Further, the paper describes how to configure a system comprising Oracle Database 12*c* with Oracle Automatic Storage Management on Oracle Linux, using an Oracle FS1 Series flash storage system. It assumes that you know how to configure each of the individual components (Oracle Database 12*c*, Oracle Automatic Storage Management, Oracle Linux, and Oracle FS1 Series storage systems), but you need to know the best practices for making these components work well together. The primary aim of this document is to give you the settings that will quickly get you to 80 to 90 percent performance. Of course, your particular environment and applications are different from those outlined in this paper; hence, you will need to perform a final tuning of settings that can give you the best performance.

A secondary aim of this paper is to give you detail about **how** the Oracle team determined these configuration settings. Then you can repeat these tests in your environment, adjusting the workloads to more closely match your applications.

While the testing outlined in this paper was undertaken against the Oracle Database 12*c* release, the testing and results should be equally applicable for Oracle Database 11*g* Release 2.

The Test Environment

Figure 1 shows the physical setup. Starting from the left, it shows an Oracle FS1-2 flash storage system connected to a 16 Gb/sec Fibre Channel (FC) switch. The red lines indicate multipathing FC connections. The switch is connected to two Sun Server X4-2 servers from Oracle that run Oracle Database 12*c* as an Oracle Real Application Clusters (Oracle RAC) cluster. A third server is used for running workload generators. This is connected to the database servers by standard Ethernet/IP networking.
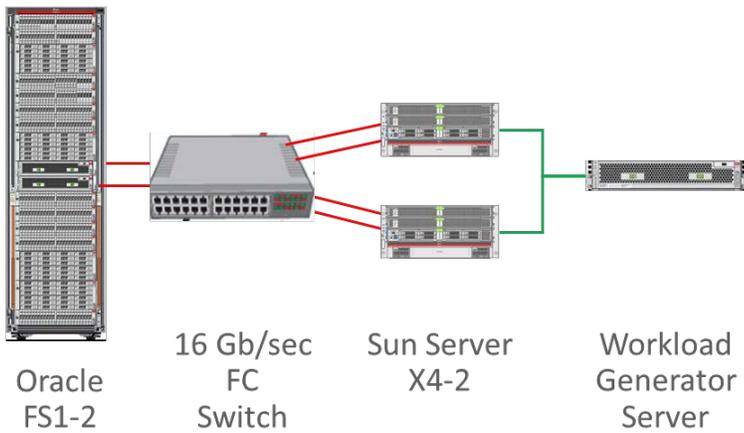
Figure 1. Physical setup

Figure 2 shows the same setup but from a logical, rather than physical, perspective. Starting at the bottom of the stack, the Oracle FS1-2 storage system appears with its four types of storage media and two controllers. The DBA or storage administrator creates virtual disks (LUNs) on the array, and these are what the next layer up (Oracle Automatic Storage Management) sees. Oracle Automatic Storage Management groups the LUNs into disk groups and presents these to Oracle Database 12*c*. Above the database layer, there is an IP load balancer that evenly distributes the database requests coming from the workload generators to the nodes of the Oracle RAC cluster.
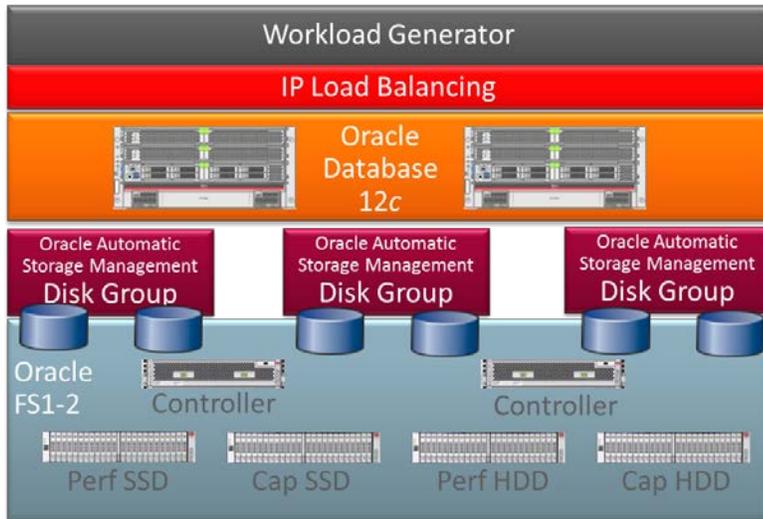

Figure 2. Logical setup

In terms of workload generators, two very different pieces of software are used. The first is a program called Swingbench (http://dominicgiles.com/swingbench.html). This is a load generator designed to simulate the operation of a real application and to stress test Oracle Database. It consists of a load generator, as well as tools for scheduling and coordinating multiple instances of the load generator when deployed for very large workloads. As Figure 3 shows, it includes a nice GUI for ease of use.
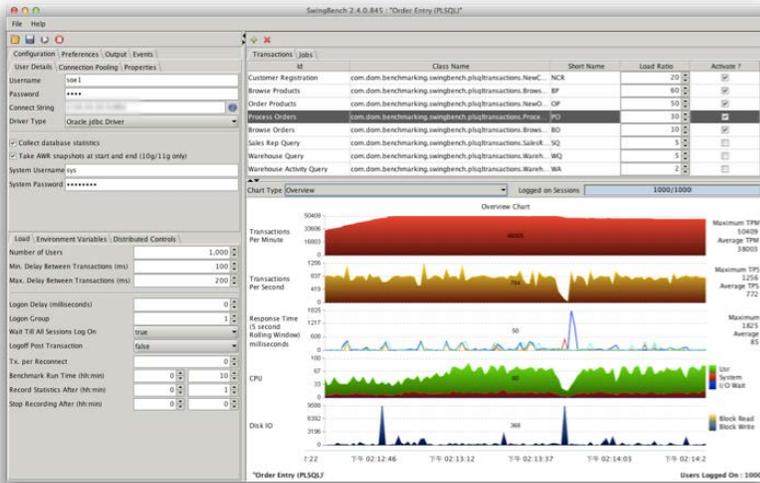
Figure 3. Swingbench screenshot

The second tool is Orion. This is an Oracle tool that is included as part of the standard installation package with Oracle Database 12*c* (and earlier releases). It is a tool for predicting the I/O performance of Oracle Database without having to install the database. It is designed to simulate the database I/O workloads using the same software stack as Oracle Database and can even simulate the effects of striping performed by Oracle Automatic Storage Management. As Figure 4 shows, while Swingbench is a GUI-oriented tool, Orion is a text-based one; its results can be imported into a spreadsheet for analysis.

Orion executes tests at different I/O load levels in order to measure performance metrics such as MBPS, IOPS, and I/O latency. Unlike Swingbench, it operates at the Oracle Automatic Storage Management level—see Figure 5. As such, it can be used to assess the I/O performance of individual Oracle Automatic Storage Management disk groups, rather than the performance of the database as a whole.
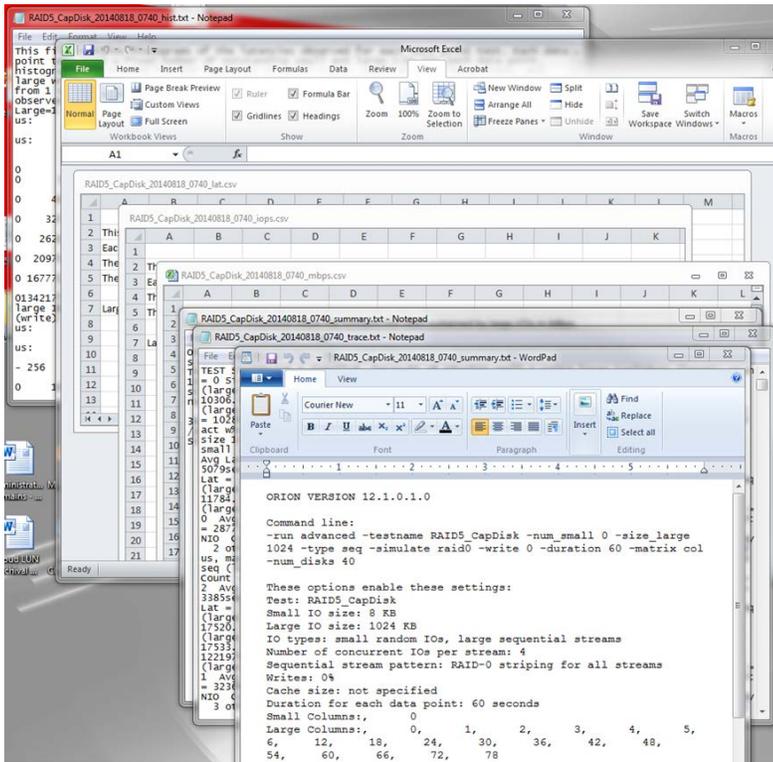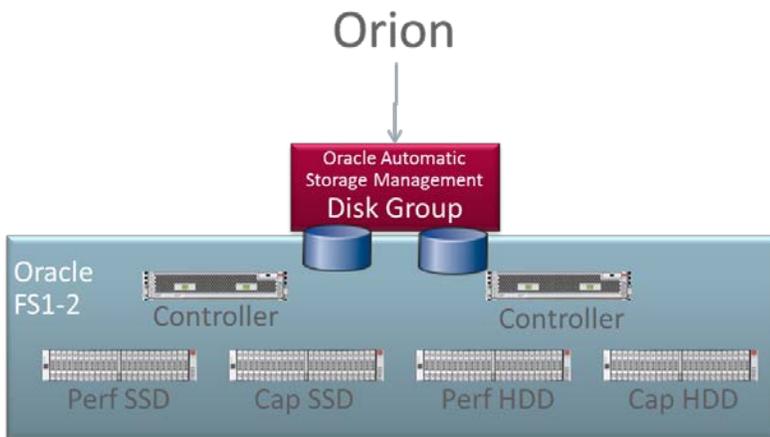
Figure 4. Orion screenshot

Figure 5. Orion's position in the I/O stack

Swingbench and Orion are complementary tools. During testing, the Oracle team went back and forth between the two tools. Swingbench told us about the I/O operations being generated by the whole system. On the basis of this information, the team defined the necessary properties of the individual Oracle Automatic Storage Management disk groups. The team tested these disk groups, one by one, using Orion. Then the team returned to Swingbench to retry the performance with the new disk groups and iterate.

Best Practices

The best practice recommendations fall into three distinct groups:

» **Oracle Automatic Storage Management disk group recommendations**
What disk groups should the DBA create and how many LUNs should be in each disk group?

» **QoS Plus recommendations**
For each of the LUNs created (and assigned to Oracle Automatic Storage Management disk groups), what settings for the Oracle FS1-2 storage system (so called "QoS Plus" settings) should be assigned?

» **Storage domain and autotiering recommendations**
How should the Oracle Automatic Storage Management disk groups be distributed among Oracle FS1-2 storage domains? Are any setting changes required to make the Oracle FS1-2 autotiering facility work effectively?

Oracle Automatic Storage Management Disk Groups Recommendations

Figure 6 shows the role of Oracle Automatic Storage Management. The left side of this figure shows the traditional role it has played in the case of servers with direct-attached disks. In this case, Oracle Automatic Storage Management had the role of a volume manager (managing the striping and mirroring of data across the disks as well as carving up the disk pool into virtual disks and merging these into disk groups for use by the database) and a file system (for the easy naming and management of the database's files).
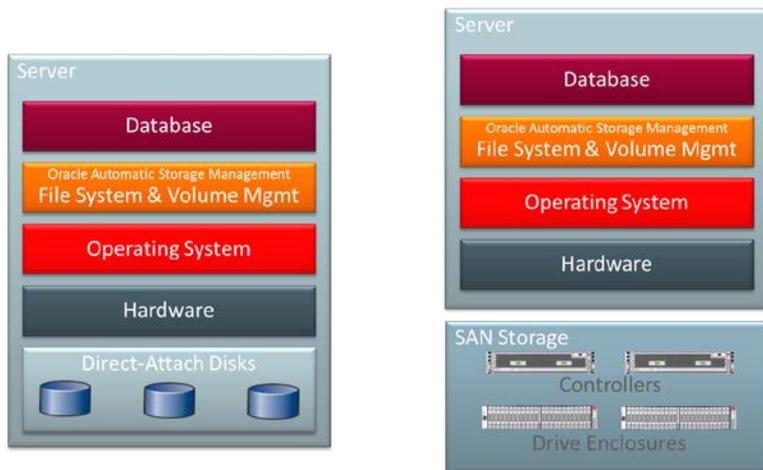

Figure 6. The role of Oracle Automatic Storage Management

This paper focuses on the situation shown on the right side of the figure. Here, there are no direct-attached disks; instead, the storage needs of the server and database are met through shared FC-attached SAN storage. In this case, many of the functions of a volume manager are undertaken by the SAN controllers; virtual disks are carved out of the large pool of contained storage and the reliability, redundancy, and performance of this storage is managed by a portfolio of techniques (RAID, striping, QoS, advanced schedulers, and so on). In addition, such controllers (and in particular in this case, the Oracle FS1-2 storage system) contain large caches and tiered storage (performance solid-state drives [SSDs], capacity SSDs, performance hard disk drives [HDDs], and capacity HDDs).

The standard Oracle recommendations are to create two Oracle Automatic Storage Management disk groups (+DATA and +FRA) with four disks per disk group. However, these recommendations are for the situation shown on the left side of Figure 6, that is, when the redundancy is being managed by Oracle Automatic Storage Management and when all disks have the same storage characteristics. These recommendations need to be reconsidered in the case of an intelligent, high-function SAN storage controller.

Oracle Database 12*c* has many processes that drive I/O to the storage infrastructure. These various processes have very different I/O patterns. For example, the REDO logs are primarily a small sequential write pattern, while archive logging is a mixed large sequential read/write pattern. In addition, the application driving the database can have a profound impact; a transactional application will generate a small workload of random write operations, while a decision-support or warehousing application will generate a large workload of sequential read operations. From a purely storage perspective, it makes sense to direct such different I/O patterns to distinct Oracle Automatic Storage Management disk groups, with the storage properties of the individual disk groups fine-tuned to their respective I/O patterns. However, the manageability of such an environment needs to be considered. Tens of such disk groups create an extremely complex environment. Hence, Oracle's recommendation balances the needs of I/O performance (matching the I/O patterns generated by the various database processes to the storage characteristics of the disk groups) against the need for easy management (keeping the number of disk groups to a small number).

Recommendation #1

Use three disk groups: +DATA (for tablespaces and temp), +REDO (for redo logs and control files), and +FRA (for the fast recovery area, archive logs, and backup sets).

Figure 7 shows sample output for one test that was done to examine the impact of the number of LUNs (virtual disks provisioned from the Oracle FS1-2 storage system) assigned to an Oracle Automatic Storage Management disk group. Many such tests were undertaken for different kinds of workloads. This result is typical. It shows a significant performance improvement in going from a single LUN to multiple LUNs. But going beyond two LUNs shows little or no additional benefit.

Many DBAs and/or storage administrators may have developed procedures to use multiple LUNs per disk group. They should continue to use their tried-and-tested procedures as long as they use at least two LUNs per disk group, and the chosen number of LUNs is always a multiple of two (see Recommendation #2.
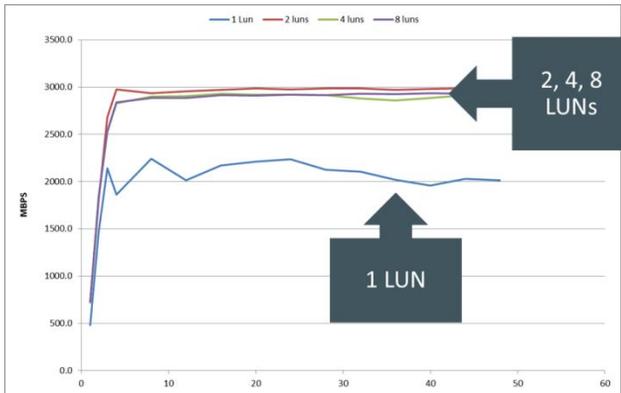
Figure 7. Variance of I/O performance for a sequential workload as the number of LUNs in an Oracle Automatic Storage Management disk group is changed

Recommendation #2

Use two LUNs (or multiples of two LUNs) per Oracle Automatic Storage Management disk group.

Recommendation #3

LUNs in Oracle Automatic Storage Management disk groups should be balanced across the two Oracle FS1-2 controllers.

The Oracle FS1-2 storage system has two controllers. Each LUN that is provisioned on the Oracle FS1-2 storage system is assigned to a primary controller, with that controller serving its I/O requests and the other controller being called upon only in failure scenarios.

The Oracle FS1-2 storage system has an autoassign capability that, as new LUNs are provisioned, assigns those LUNs in a balanced manner to one or the other controller. Using this facility, in almost all circumstances, results in a well-balanced system (all of the various LUNs being assigned equally between the controllers). However, in this case, when the administrator is always creating LUNs in pairs, it is good practice to explicitly assign each member of the pair to a different controller.

Figure 8 shows a screenshot of the provisioning GUI for the Oracle FS1-2 storage system, highlighting how a controller can be explicitly assigned to a newly created LUN. The same operation also can be accomplished via a command-line interface for scripting purposes.
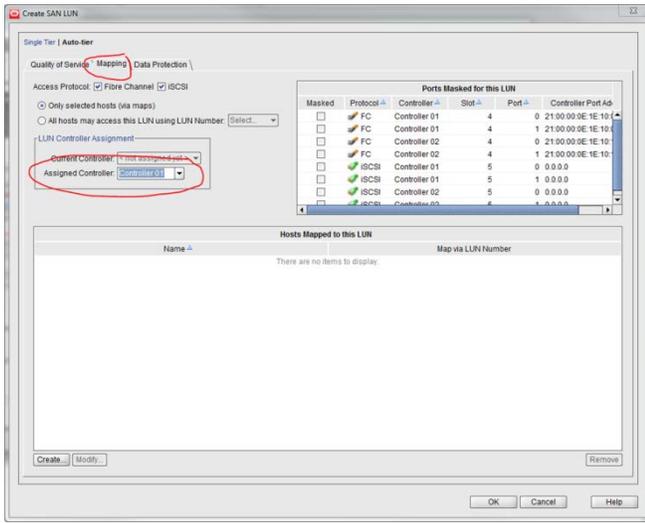
Figure 8. Screenshot of LUN creation on the Oracle FS1-2 storage system

QoS Plus Recommendations

QoS Plus refers to the settings assigned to LUNs, as they are provisioned, that tune the performance characteristics of the LUNs. As mentioned above, the performance characteristics of the Oracle Automatic Storage Management disk groups (and, hence, of the LUNs that comprise the disk groups) should be matched to the I/O patterns of the database elements assigned to the various disk groups.

Recommendation #4

Configure the QoS Plus settings of LUNs created for Oracle Automatic Storage Management as shown in Table 1.

TABLE 1. QOS SETTINGS

Application Profile						
Name	RAID Level	Read-Ahead	Priority	Stripe Width	Writes	Preferred Storage Classes
ASM_DATA_OLTP	Mirrored	Conservative	High	Autoselect	Back	Perf Disk, Perf SSD
ASM_DATA_DSS	Single parity	Aggressive	High	Autoselect	Back	Perf Disk, Perf SSD
ASM_REDO	Single parity	Normal	Premium	All	Back	Perf Disk
ASM_FRA	Double parity	Aggressive	Archive	Autoselect	Back	Cap Disk

Note that for +DATA, the online transaction processing (OLTP) and decision-support system (DSS) workloads are distinguished because of their very different I/O patterns.

+REDO gets a performance gain from using SSDs. However, the I/O pattern of redo logs is a sequential workload, for which HDDs have very good performance. SSDs outperform HDDs, but the performance boost relative to the extra cost of such media, in general, does not warrant the increased cost; hence, the recommendation shown above is to use performance HDDs. Nevertheless, if the performance boost is required and SDDs are used, the Oracle team recommends changing the writes from write-back to write-through.

Table 1 shows many QoS Plus settings, so the task of assigning numerous settings to each LUN as each LUN is provisioned might seem arduous. In fact, it is an easy task. The various QoS Plus settings for each LUN (part of an Oracle Automatic Storage Management disk group) are collected together into an *application profile*. As shown in Figure 9, these can be selected from a single drop-down box at LUN provisioning time.
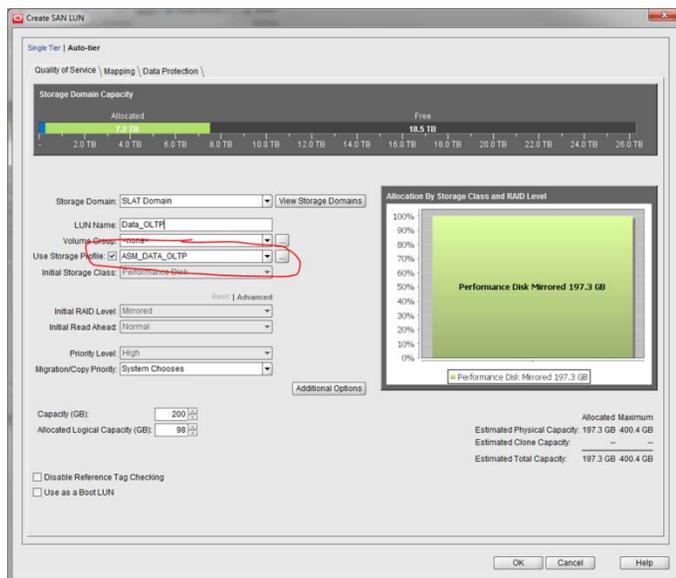

Figure 9. Assigning an application profile to a LUN

Storage Domain and Autotiering Recommendations

The Oracle FS1-2 storage system has a facility called *storage domains*. Storage domains isolate data within drive enclosure (DE) virtual resource pools. You can combine DEs (HDDs and/or SSDs) in any way to form a domain within the Oracle FS1-2 storage system, or you can use all the DEs as a single domain.

The data stored in a storage domain is physically segregated in that domain within the Oracle FS1-2 storage pool. No data in one domain can trespass into another domain. This prevents data comingling for customers who wish to isolate data within the Oracle FS1-2 storage system. In addition, it also prevents contention between workloads—the I/O operations directed to one storage domain do not interfere with the I/O operations directed to another domain.

Recommendation #5

Separate Oracle Automatic Storage Management disk groups into separate Oracle FS1-2 storage domains.

Create three storage domains corresponding to the +DATA, +REDO, and +FRA disk groups. The +FRA domain should contain only capacity HDDs. The +REDO domain should contain performance HDDs (however, as discussed earlier, if you want a performance boost, add performance SSDs). The +DATA domain should contain performance HDDs, capacity SSDs, and performance SSDs.

Recommendation #6

For each storage domain, add SSD capacity based upon the skew tipping point for the data assigned to that domain.

When constructing the +DATA storage domain, how much SSD capacity (given how expensive it is relative to HDD capacity) should you add relative to the amount of HDD capacity? For almost all applications, only a fraction of the data is actually "hot"; that is, in a typical application, a small fraction of the data tends to generate almost all of the I/O operations. This fraction varies from application to application, but it is typically between 5 and 15 percent. This is called *data skew*.

Figure 10 depicts data skew in a data set. If all the data in this set were equal (uniformly accessed), this graph would be a straight line joining (0,0) to (100,100). However, in this case, the data is heavily skewed, with 15 percent (in terms of capacity) generating 85 percent of the I/O operations per second (IOPS). The point where the sum of the capacity and IOPS equals 1 is referred to as the *skew tipping* point.

The skew tipping point leads to a rule of thumb. To estimate the fraction of SSD capacity you should add to a storage domain, first determine the skew tipping point for your application. Then add that SSD capacity (in the case of Figure 10, this is 15 percent).
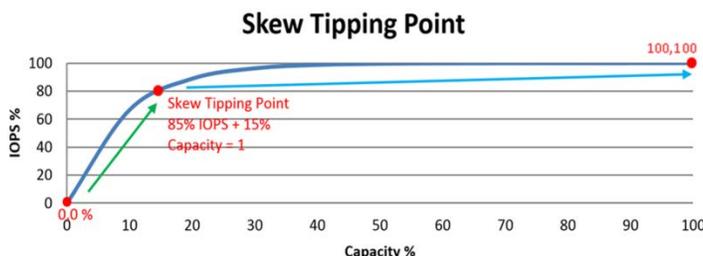

Figure 10. Data skew in a data set and the skew tipping point

A key point here is that you need to determine the *quantity* of data skew in your data set, but you don't need to identify the specific data that is hot (in any case, that varies over time). This is where the Oracle FS1-2 storage system's autotiering facility comes into action. It measures the I/O operations going to the various parts of your data set and automatically moves the hot data to SSDs. And over time, it continues to migrate hot data up to and cold data down from the SSDs. This gives you cost-optimal use of the various storage tiers and maintains that optimal setting over time as your data's hot spots change.

Summary

Much expertise exists about how to configure Oracle Database to work effectively with direct-attached storage, and tools such as Oracle Automatic Storage Management have encoded key parts of the best practices and eased the administrative burden. However, the introduction of the Oracle FS1-2 SAN controller, with its multiple storage tiers, QoS Plus, storage domains, and large cache changes these calculations.

This paper described key best practices for building a system of Oracle Database 12*c*, Oracle Automatic Storage Management, Oracle Linux, and the Oracle FS1-2 storage system. By following these recommendations (duplicated below for quick reference), you will be able to build a baseline system quickly that has excellent performance. By following the techniques outlined (building a test environment and using the same software), you will be able to perform the necessary fine-tuning to adapt the practices for your specific environment.

Recommendation #1

Use three disk groups: +DATA (for tablespaces and temp), +REDO (for redo logs and control files), and +FRA (for the fast recovery area, archive logs, and backup sets).

Recommendation #2

Use two LUNs (or multiples of two LUNs) per Oracle Automatic Storage Management disk group.

Recommendation #3

LUNs in Oracle Automatic Storage Management disk groups should be balanced across the two Oracle FS1-2 controllers.

Recommendation #4

Configure the QoS Plus settings of LUNs created for Oracle Automatic Storage Management as shown in Table 2:

TABLE 2. QOS SETTINGS

Application Profile						
Name	RAID Level	Read-Ahead	Priority	Stripe Width	Writes	Preferred Storage Classes
ASM_DATA_OLTP	Mirrored	Conservative	High	Autoselect	Back	Perf Disk, Perf SSD
ASM_DATA_DSS	Single parity	Aggressive	High	Autoselect	Back	Perf Disk, Perf SSD
ASM_REDO	Single parity	Normal	Premium	All	Back	Perf Disk
ASM_FRA	Double parity	Aggressive	Archive	Autoselect	Back	Cap Disk

Recommendation #5

Separate Oracle Automatic Storage Management disk groups into separate Oracle FS1-2 storage domains.

Recommendation #6

For each storage domain, add SSD capacity based upon the skew tipping point for the data assigned to that domain.

CONNECT WITH US

blogs.oracle.com/oracle

facebook.com/oracle

twitter.com/oracle

oracle.com

Hardware and Software, Engineered to Work Together

Best Practices for Optimizing Storage for Oracle Automatic Storage Management with Oracle FS1 Series
January 2015, Version 1.0

Oracle is committed to developing practices and products that help protect the environment