White Paper

The Effect of Priorities on LUN Management Operations

Applied Technology

Abstract

This white paper describes the effect of each of the four Priorities (ASAP, High, Medium, and Low) on overall EMC® CLARiiON® performance in executing LUN Management Operations. The LUN Management Operations are migrate, rebuild, and bind. In addition, this paper provides a method for estimating the duration of these operations and the effects on CLARiiON performance for each priority.

November 2008

Part Number H4153.2

Table of Contents

Executive summary

You have a lot of leeway when you decide how long LUN management operations such as migrate, rebuild, and bind can take to execute. The duration selected for LUN operations uses more or less of the CLARiiON's resources, which has a corresponding effect on system performance. There are three resources to manage on an EMC® CLARiiON® that affect overall system performance: storage processor (SP) CPUs, back-end buses, and RAID group hard disks. You need to balance the desired LUN operation's duration with the available resources.

The main factors affecting the duration of a LUN operation are:

- Priority: Low, Medium, High, and ASAP (As Soon As Possible.)
- Background workload: Storage system utilization.
- Underlying LUN RAID group type: mirror or parity.
- RAID group disk type: Drive rpm speed and interface (for example, Fibre Channel or SATA)
- Cache settings (can have an impact on a LUN migration)

Priority has the largest effect on an operation's duration and hence system performance. There are four priorities (Low, Medium, High, and ASAP) for LUN Migration, rebuild, background verify, and bind.

The Low, Medium, and High priorities economically use the system's resources, but have the longest duration. The ASAP priority accomplishes the operation in the shortest period of time. It is not uncommon for "ASAP" rates and resource allocations to be as much as 20 or more times higher than economical rates and resource allocations. However, it also reallocates storage system resources and has a significant impact on overall system performance and especially application response times. *This reallocation can adversely affect the overall system performance.* Beginning in release 28, the ASAP setting is no longer the default speed for any LUN operations. EMC strongly recommends, for reasons shown within this white paper, that you do not use ASAP as a LUN management operation priority unless there are no response time sensitive applications running on your array that might be impacted.

EMC strongly encourages users in production environments to plan ahead, and use the economical Low, Medium, and High priorities for LUN management. These priorities make the most efficient use of the CLARiiON's resources and ensure the highest overall storage system performance.

Introduction

This white paper is about performance planning. It describes how to estimate the duration and storage system resource utilization for the prioritized LUN operations: migrations, rebuilds, and binds. Using the information provided, users can plan to avoid the "ASAP Effect" on overall system performance. This paper was formerly titled *The Influence of Priorities on EMC CLARiiON LUN Management Operations.*

This paper assumes familiarity with CLARiiON storage system fundamentals. If you need to refresh your knowledge of these concepts, please review the *EMC CLARiiON Fibre Channel Storage Fundamentals* white paper.

The examples and data given in this paper are for the most common use cases. Users seeking information on special cases, such as storage system performance tuning, should refer to the *EMC CLARiiON Performance and Availability* white paper.

Finally, this paper assumes the user is familiar with the use of EMC's Navisphere® Array Manager and the operations discussed. It discusses only the performance issues involved with these LUN operations. No attempt is made to describe using Navisphere to achieve the results described in this paper. Users should refer to the Navisphere Online Help for this "how to" information.

Examples, tables, and figures

Unless otherwise stated, the examples and information included in this paper apply to the CLARiiON CX4-960 with FLARE® release 28 installed. The CX4-960 is the most powerful model CLARiiON storage system. Users of other CLARiiON CX4 models or of the CX3 series on earlier releases of FLARE can expect longer durations and greater resource utilizations in the operations described in this paper.

The CX4-960 used to document this paper's results was configured with 73 GB 15k rpm Fibre Channel, 400 GB 10k rpm Fibre Channel, and 1 TB 7200 rpm SATA hard drives. RAID group types and sizes are as follows:

- Five-disk (4+1) RAID 5
- Six-disk (3+3) and eight-disk (4+4) RAID 1/0
- Eight-disk (6+2) and ten-disk (8+2) RAID 6

The hard drives were distributed across the CLARiiON's DAEs according to the recommendations found in the *EMC CLARiiON Performance and Availability* white paper. All the LUNs in the examples were owned by the same SP. The other SP was idled during all testing. Available read cache was set at 1,000 MB and write cache was set at 3,000 MB, with an 8 KB Page Size configured. High and Low Watermarks of 80:60 were set. All test data was recorded using Navisphere Analyzer and internal EMC diagnostic tools.

Audience

This white paper is intended for system engineers, EMC partners, members of EMC and partner sales and professional services, and engineers wanting to understand the effect of priority on LUN operations.

Terminology

Background Verify (BV): A prioritized SP background process that verifies a hard disk's parity sectors with the data.

LUN migration: Movement of data from one LUN to another or changing a LUN's underlying RAID group type.

Priority: There are four priorities: ASAP, High, Medium, and Low.

Rebuild priority: The relative rate of the automatic rebuild operation when a hot spare replaces a failed disk. There are four priorities: ASAP, High, Medium, and Low.

Response time: The interval of time the storage system takes to respond to a host request I/O operation measured in milliseconds by Navisphere Analyzer.

Verify priority: The relative rate at which BV executes. There are four priorities: ASAP, High, Medium, and Low.

Performance planning resources

The availability of three key storage system resources needs to be considered when scheduling the prioritized LUN operations of migrate, rebuild, and bind. They are SP CPUs, the storage system back-end bus, and hard disks. In a production environment, the effect of the current and future workload on these resources also needs to be understood.

The CX4-960 storage system used for illustration in this paper is a computer-based device with one quad-core CPU per SP. The metric for measuring an SP's CPU utilization is percentage CPU utilization. System performance is affected by the number of tasks executing on an SP's CPU. In a production environment, the percentage of SP CPU utilization directly affects observed response time. Scheduled LUN operations can use considerable CPU resources. For example, scheduling CPU intensive operations such as more than one ASAP LUN migration during the system's "busy hour" could adversely affect system response time.

The back-end bus and hard disks constitute the *back end*. The number of disks per DAE, the RAID type (mirrored or parity), the type of disks (Fibre Channel, SATA, or SAS), and the distribution of the LUN's information on the disks all affect system performance. This paper uses the Navisphere Analyzer Percentage Disk Utilization metric for measuring the back-end utilization. In a production environment, LUNs composed of RAID groups having high disk utilization can be adversely affected by prioritized operations. For example, scheduling more than one simultaneous back-end intensive operation, such as ASAP LUN binds, during the system's "busy hour" could adversely affect system response time.

Workload is always the primary driver of system response time. Maintaining the Service Level Agreement (SLA) sets the "ceiling" for all resource utilization. If too many resources are diverted from production to LUN operations, the system's overall response can increase to an unacceptable level. When scheduling prioritized LUN operations, you need to know the current resource utilization to anticipate the effects of the additional demands.

In a production environment, you need to understand the CPU utilization and back-end utilization of the CLARiiON under its workload. The difference between these measurements determines the system resources available for workload expansion and the execution of the operations such as rebuilds, migrations, and binds. Recommended thresholds of operation and how to calculate resource availability using basic configuration information are found in the *EMC CLARiiON Best Practices for Fibre Channel Storage* white paper.

Prioritized LUN operations

The three primary operations performed on LUNs are migrations, rebuilds, and binds. LUN migrations and rebuilds are the two LUN operations that can be prioritized. The priority chosen for these operations has a significant effect on the duration of the operation and the amount of SP resources devoted to completing it. Related to LUN binds are background verifies (BVs). BVs are prioritized automatically, can be configured at bind time, and can have a significant effect on CLARiiON resources when the BV executes.

LUN migrations

A LUN migration is the scheduled movement of one LUN's data to another LUN.

Why migrate?

LUNs are typically migrated to increase storage capacity or to improve the storage system's performance.

The most common migration is the simple increase in the size of a LUN, or a LUN expansion. In this case, a LUN is increased in size to meet the owning application's need for additional storage capacity. For example, a LUN composed of a five-disk (4+1) RAID 5 group with a capacity of 100 GB may be migrated to a larger LUN on another RAID group, allowing for expansion by the host file system.

A less common migration occurs during storage system tuning. Balancing hard drive utilization over more spindles increases a storage system's throughput. In this drive utilization case, a LUN whose underlying RAID group or groups is resident on hard drives that experience sustained heavy utilization is moved to hard drives with a lower utilization.

Another performance tuning migration involves changing the underlying RAID type or the number of hard drives within a LUN's current RAID group type. For example a LUN created from a RAID 1/0 group may be migrated to a LUN created from a RAID 5 group to use fewer hard disks while maintaining the same capacity.

Migration overview

In a migration there is a *source* LUN and a *destination* LUN. The destination LUN must be equal to or greater in capacity to the source. Data is copied from the source LUN to the destination LUN. The source LUN is available for use during the copy. The migration can be cancelled at any time without affecting the source LUN's data. When the copy is complete, if the destination LUN is larger than the source LUN, the additional new capacity is initialized. After initialization, the destination LUN is renamed for the source LUN, and assumes all its properties. Note the RAID type or capacity of the LUN may now be different from its original. The former source LUN is left unbound after the migration.

The resource initially most affected in a LUN migration is the SP CPU. LUN migrations are a cached operation. As long as the cache does not fill, the drive type, drive speed, RAID type, and RAID size of the LUN have no effect on performance. The cache is not likely to fill except in the case of an ASAP priority migration. The full effect of priorities on LUN migrations is discussed next.

Migration priority

There are four migration priorities: Low, Medium, High, and ASAP. These priorities set the rate of the migration and the subsequent utilization of SP resources. To mitigate the possible adverse effects of an "ASAP" migration on customer applications, the current FLARE release 28 default migration rate was changed to "high."

Economical priorities

The Low, Medium, and High priorities are set up to economically use SP resources. The migration rate is throttled so production environment performance is unaffected by the operation. The Low priority takes the longest time to complete a migration; the High priority takes less time. The migration rates for these priorities are the same for all CLARiiON models. The following table shows the rate of migration in MB/s for each of these settings.

Table 1. Low, Medium, and High migration rates

Priority	Rate MB/s
Low	0.9
Medium	1.6
High	3.4

ASAP priority

With ASAP priority, the LUN migration completes in the least amount of time. It also uses considerable SP resources. The expedited migrations of the ASAP priority have different rates dependent on the CLARiiON model hosting the migration. Mirror RAID types migrate more quickly than parity RAID types. The following table shows the rate of migration in MB/s for each RAID type at the ASAP setting.

Table 2. ASAP migration rates

Operation	Mirror RAID rate MB/s	Parity RAID rate MB/s
Migration	185	170

Whenever the size of a LUN is grown using LUN Migration, there will be a second phase of the migration process called the "expansion" phase. This expansion phase runs more slowly than the migration phase rate. The expansion phase is affected by the CLARiiON's cache settings. Larger cache settings slow the operation down. For Parity RAID groups, this expansion rate can range from 12 MB/s to 40 MB/s depending on your cache settings, RAID type, and disk layout. For Mirrored RAID types, expansion rates can be as low as 6 MB/s. The section "Best practices for migrations" on page 10 has more information .

Migration duration

The duration of a LUN migration depends on the migration priority (rate) and the size of the LUN being migrated.

The typical migration is a two-step process: copy and expansion (sometimes called "initialization"). In the most common LUN migration, the LUN expansion, data from a smaller LUN is first copied to a new larger LUN. Then the additional new storage of the larger LUN is initialized. Priority applies to both steps. Storage is expanded at a different rate than it is copied; the expansion rate is dependent on the RAID group type (parity or mirror) and the write cache settings, and the expansion rate is slower than the copy rate. No CLARiiON bus bandwidth is consumed during the expansion phase. The performance effects of the expansion are restricted to the RAID group disks of the LUN being expanded. Note this means all LUNs sharing the RAID group of the expanding LUN are affected by the expansion.

The following describes the calculation to determine the duration of a LUN migration in hours.

Basic migration time calculation

- Time: Duration of LUN migration
- Source LUN Capacity: Size of the source LUN in GB
- Migration Rate: Rate of copy from the source LUN to destination LUN from Table 1 or Table 2, depending on the selected migration priority
- Destination LUN Capacity: Size of the destination LUN in GB

- Initialization Rate: Speed at which new additional storage is initialized in MB/s (Table 2 for ASAP, or else omit)

Time = (Source LUN Capacity * Migration Rate) +
((Destination LUN Capacity – Source LUN Capacity) * Initialization Rate)

Note that the typical migration has two parts, the copy and the expansion (sometimes referred to as “initialization.”). If there is no expansion required, then the expansion/initialization phase does not occur.

Example

How many hours will it take for a LUN expansion from a five-disk (4+1) RAID 5 on a 100 GB LUN to a five-disk RAID 5 on a 200 GB LUN at an ASAP migration priority on a CLARiiON CX4-960? The example will use 15k rpm disks.

- Time: Duration of LUN migration in hours
- Source LUN Capacity: 100 GB
- Migration Rate: 170 MB/s (from Table 2)
- Destination LUN Capacity: 200 GB
- Expansion Rate: 12 MB/s (from Table 2 supporting text)

Time = (100 GB * ((1/170 MB/s) * 1024 MB/GB * (1/3600 sec/hr))) + ((200 GB – 100 GB) * ((1/12 MB/s) * 1024 MB/GB * (1/3600 sec/hr))) = 2.53 hr

Note from the example, the storage initialization takes about 90 percent of the migration’s total duration as the expansion rate is much slower than the copy rate.

It is informative to compare the durations for the other three priority types for the same example. The much lower migration rates used in the Low, Medium, and High priorities will result in longer durations. The following table shows the result of the calculation for the example using each of the priorities.

Table 3. LUN expansion example all priorities

Priority	Duration (Hrs)
Low	32
Medium	18
High	9
ASAP	2.5

Migration resource utilization

The following figure shows the SP percent utilization and hard disk utilization for the typical ASAP priority LUN migration on an idle CX4-960. The two LUNs involved are a 50 GB source LUN resident on a five-disk (4+1) Fibre Channel RAID 5 group and a 100 GB destination LUN on a separate similarly configured RAID 5 group on a separate DAE. The CLARiiON is idle, except for the migration.

In the graph, the destination LUN RAID group hard drive utilization is shown in red. The Source LUN RAID group Hard Drive Utilization is shown in orange. Note that the CX4-960 has four CPU cores per SP, with a total of eight cores on two CPUs for the storage system. LUN operations have SP CPU 0 affinity. The SP CPU 0 utilization is shown in dotted dark blue. The SP CPU 1 utilization is shown in dashed light blue. These graph conventions are maintained throughout this paper, except where noted.

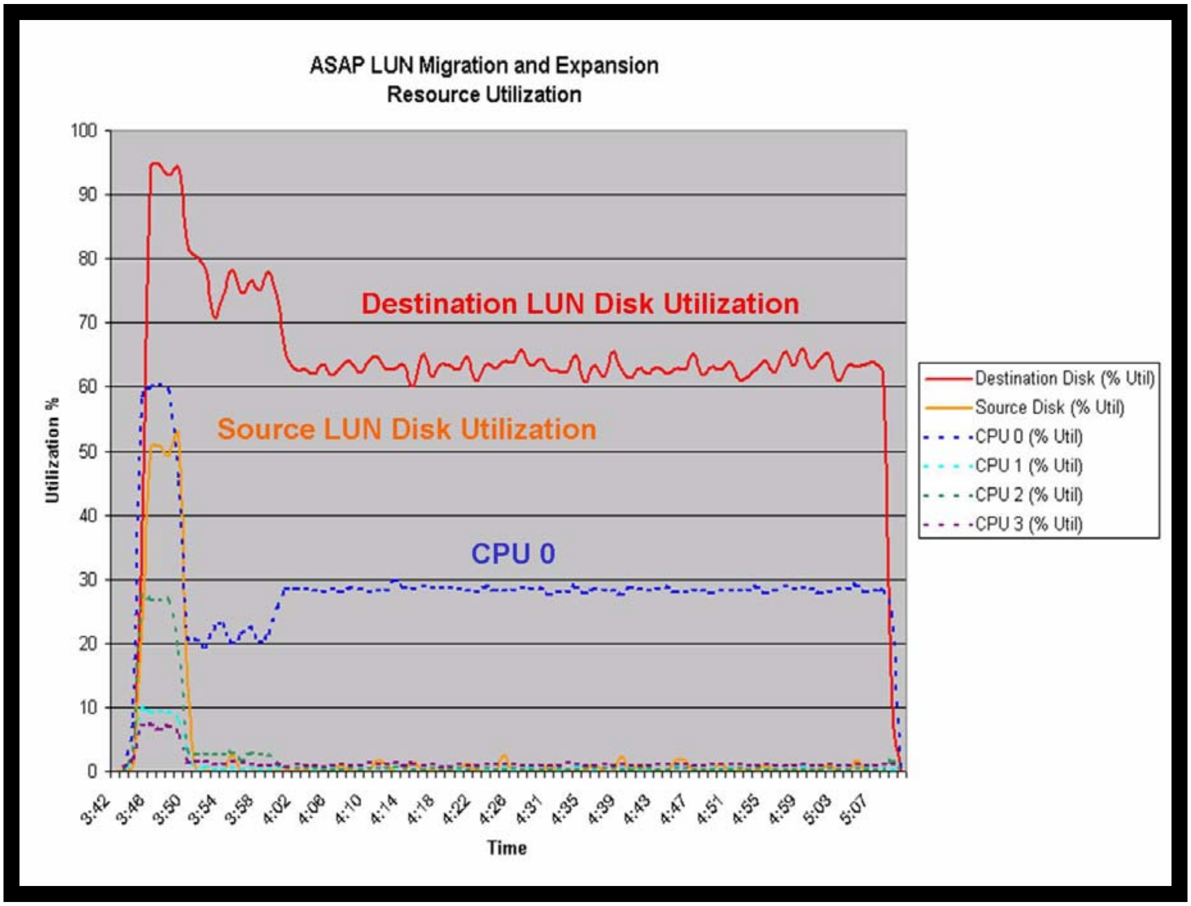

Figure 1. ASAP LUN migration resource utilization

From the graph it can be seen that during the copy phase of the migration, the CPU handling the migration is at 60 percent utilization. The CPU moves down to around 30 percent utilization during the expansion phase of the operation. Also, the destination RAID group's hard drives are at a high utilization for the duration of the migration. When you schedule an ASAP migration, you should be aware of the initial and continued SP CPU utilization's effect on the SP's performance, and overall high disk utilization on LUNs sharing the destination LUN's RAID group.

Best practices for migrations

ASAP LUN migrations should be avoided. If the destination LUN cannot write as fast as the source can read, the write cache will fill for the duration of the migration. This will cause forced flushes. Forced flushes result in poor response times for the other LUNs on the same SP.

An effective way to ensure that an ASAP LUN Migration does not overwhelm the SP and cause performance degrading "forced flushes" to cache is to set the cache write-aside to 511 (the default is 2048) at the destination LUN. This must be done *before* the migration operation is initiated, since the target will become unavailable once it starts. In this case, the 4+1 RAID 5 device will migrate at about 40 MB/s (144 GB/hr) with no effect on the write cache.

An alternative solution is for the write cache to be disabled at the destination LUN. However, as the migration proceeds, host writes to the source must be mirrored to the (write-disabled) target drive. Unless speed is important, consider running at the High priority setting instead.

When migrating to a destination LUN that has a greater capacity than its source, an internal consistency operation (similar to background zeroing) occurs in the “new” space on the destination LUN. On the 4+1 RAID 5 mentioned previously, it runs at a best possible speed of about 40 MB/s (144 GB/hr). This speed may vary based upon drive type and other factors. To avoid the time penalty associated with this operation, consider using the concatenate operation under metaLUNs to create the correct sized source LUN, and then copy it to its destination. Alternatively, copy the source to the same-sized destination, and then perform a LUN expansion of the destination.

LUN rebuilds

A rebuild replaces a failed hard disk within a RAID group with an operational disk. Note one or more LUNs may be bound to the RAID group with the failed disk. All LUNs affected by the failure are rebuilt.

Why rebuild?

A LUN rebuild is a data integrity feature of the CLARiiON. A rebuild restores a LUN to its fully assigned number of hard drives using an available hot spare should a hard drive in one of RAID groups fail. The rebuild restores the data protection capability of the LUN’s underlying mirror or parity RAID.

Rebuild overview

When a hard drive exceeds a predefined error threshold or fails outright, the LUN contents of the failed disk are *rebuilt* from the LUN’s RAID group parity data or mirror data onto an available hot spare. A rebuild can be started either automatically by FLARE or manually. It is most common for failing disks to be detected by FLARE and rebuilt automatically. A rebuild has two steps: rebuild and equalization.

A hot spare is a hard drive reserved on the CLARiiON as a replacement for failed or failing hard drives. Not all CLARiiON hard drives are compatible replacements for each other. FLARE algorithmically selects the best rebuild hot spare from the available pool of replacement hot spares to be the rebuild hot spare. During the rebuild step, the failed drive is rebuilt to the rebuild hot spare.

In the rebuild step, the contents of the failed drive are rebuilt LUN by LUN to an available hot spare. If the failed drive belongs to a parity RAID group (RAID types 3, 5, or 6) the contents of the failed drive are reconstructed (from the RAID group’s parity data, and data from the surviving drives) onto the rebuild hot spare. For mirrored RAID types (R1, R1/0), the failed drive is rebuilt by copying the data from its mirroring drive to the rebuild hot spare. This operation is a disk-to-disk copying of data. Note that copying from a mirror RAID requires less SP resources (primarily bandwidth) than rebuilding it from a parity RAID.

A few rebuild facts follow:

- RAID groups have one or more LUNs on them.
- LUNs are rebuilt one by one.
- The greater the utilized capacity of the RAID group the longer it will take to rebuild.
- LUNs are owned by one SP.
- Each LUN is rebuilt by its owning SP.

If a hot spare is unavailable, the RAID group runs in a degraded state until the failed drive is replaced. When the failed drive is replaced, the contents of the failed drive are rebuilt directly from the surviving members of the RAID group onto the replacement drive.

In the equalization step, the contents of the rebuild hot spare are copied to a replacement drive. The equalization step takes place after the rebuild step has completed and the failed drive is replaced with a new replacement drive. Some time can elapse between the rebuild and equalize steps. During equalization, FLARE copies the contents of the rebuild hot spare to the replacement drive. This is a disk-to-disk copying of data. The hot spare used in the rebuild is returned to the pool of available hot spares for reuse after the copy. The RAID group is considered fully rebuilt when the equalization step is complete.

Performance is generally not adversely affected when using a fully rebuilt hot spare. However, until a RAID group is completely rebuilt with a bound hard drive, the LUN is not in a normal state.

The resources that are most affected in a LUN rebuild are the rebuilding LUN's disks. LUN rebuilds are an uncached operation. The drive type, drive speed, RAID type, and RAID size of the LUN have a measurable effect on performance.

Rebuild priority

There are four rebuild priorities: Low, Medium, High, and ASAP. These priorities set the rate of the rebuild and the subsequent utilization of SP resources. The ASAP rate completes a rebuild in the least amount of time. It also uses the most CLARiiON resources. The Low rate is the slowest and uses the least resources.

The following tables show the rebuild rates in MB/s for each of the priorities for a single 15k Fibre Channel hard drive that is located in the same 4 GB/s bused DAE as its RAID group. The rebuild and equalization rates are for an eight-disk RAID 1/0 mirrored RAID and a five-disk RAID 5 parity RAID.

Economical priorities

The rebuild rates for the Low, Medium, and High priorities are throttled so production environment performance is unaffected by their operation.

Table 4. Economical mirrored and parity RAID Fibre Channel rebuild rates

Priority	Mirrored RAID rate MB/s per disk	Parity RAID rate MB/s per disk
Low	2.0	2.0
Medium	6.1	6.1
High	12.2	12.2

Note these priority rebuild rates are the same for each RAID type (parity or mirrored). The equalization rates for Low, Medium, and High priorities are equal to the rebuild rates.

ASAP priority

The ASAP priority completes a rebuild in the least amount of time. These expedited rebuilds have different rates dependent on the RAID type, the hard drive type, and the number of drives in the RAID group. The following table shows RAID type dependent rates.

Table 5. ASAP mirrored and parity RAID Fibre Channel rebuild rates

Operation	Mirrored RAID rate MB/s per disk	Parity RAID rate MB/s per disk
Rebuild	83	73
Equalization	82	82

The equalization rate for economical priorities is now scaled back to the same rates as the rebuilds for corresponding priorities. It is no longer the same as the equalization rate for ASAP.

Other factors affecting ASAP rebuild

There are additional factors, besides priority and RAID type, that have a measurable effect on ASAP rebuild duration. These factors apply because of the uncached nature of the operation. The following factors should also be taken into account when estimating the time and resource utilization of a rebuild:

- Drive type and speed
- RAID type
- RAID group size

SATA hard drives rebuild at 75 percent of the rate of 15k rpm Fibre Channel drives. To calculate the rebuild rate for a SATA type hard drive, multiply the appropriate priority rebuild rate (in these tables) by 1.33. The 10k rpm Fibre Channel hard drives rebuild at 65 percent of the 15k rpm Fibre Channel drives. To calculate the rebuild rate for 10k Fibre Channel drive, multiply by 1.54.

Parity type RAID 6 groups generally require a little more time to rebuild than parity type RAID 5 groups with the same number of drives. This difference in time decreases as the number of drives in the group increases. Larger parity groups also build more slowly than smaller groups. Use the following table to calculate the effects of RAID group size on rebuild rate.

Table 6. ASAP Parity RAID rebuild rate by RAID group size

# of Data Disks* in RAID group	RAID 5 MB/s	RAID 6 MB/s
4	57 to 74**	51 to 74**
8	35 to 43**	30 to 74**
10	28 to 36**	25 to 36**
14	24 to 26**	22 to 26**

*Data Disks refers to the equivalent number of "data containing" drives in the LUN. In parity RAID types, not all disk capacity in the RAID group is available for user data (data disks.) For example, "4 data disks" means a five-drive (4+1) RAID 5 (where the equivalent of one disk is reserved for parity.) Similarly, it would mean a six-drive (4+2) RAID 6 LUN (the equivalent of two disks are reserved for data protection), or an eight-drive (4+4) RAID 1/0 LUN (where every data drive is mirrored).

**A range of rebuild rates is possible for LUNs depending upon physical location of the disks within the LUN. For example, a LUN that contains disks spread across multiple back-end buses will obtain faster rebuild speeds than a LUN that has all of its disks in the same bus. The range presented here attempts to account for that factor.

For parity RAID types smaller than six disks, use the parity value given in Table 6 for four data disks.

Rebuild duration

The duration of a LUN rebuild is dependent on the rebuild priority (rate), RAID type, size of the failed disk being rebuilt, the type of drive being rebuilt, and the number of drives in the failed drive's RAID group. The back-end bus speed and even the number of buses upon which the RAID group is distributed also contribute, by spreading the load across more buses. Note the effect of the type of drive and number of drives in the failed drive's RAID group only applies in ASAP priority rebuilds.

The following describes the calculation to determine the duration of a rebuild in hours.

- Time: Duration of rebuild
- Failed Hard Drive Capacity: RAID group capacity utilization * hard drive size in GB
- Rebuild Rate: If priority is ASAP, use the time listed in Table 5 or Table 6, otherwise use Table 4

- Disk Type and Speed Adjustment: Speed adjustment to Tables 5 thru 7 for non-15k rpm Fibre Channel hard drives
- Equalization Rate: Speed at which the hot spare is copied to replacement for a failed disk.

Time = ((Failed Hard Drive Capacity * Rebuild Rate) * Disk Type and Speed Adjustment) + (Failed Hard Drive Capacity * Equalization Rate)

Note the rebuild has two parts — the rebuild and the equalization. Manual replacement of the failed hard drive must occur before equalization. This calculation assumes "instantaneous" replacement.

Example

How many hours will it take to rebuild a 500 GB SATA drive that is part of a fully bound and utilized, six-disk (4+2) RAID 6 group at the ASAP priority? Assume a quick replacement of the failed hard drive allowing a seamless equalization. Assume this LUN is bound with all of its disks on the same bus and enclosure.

- Time: Duration of rebuild in hours
- Failed Hard Drive Capacity: 500 GB
- Rebuild Rate: 74 MB/s (from Table 6)
- Disk Type and Speed Adjustment: 1.33 for SATA
- Equalization Rate: 63.1 MB/s for ASAP, or same as rebuild rate for economical rates.

Time = (500 GB * ((1/74 MB/s) * 1024 MB/GB * (1/3600 sec/Hrs) * 1.33)) + (500 GB * ((1/63.1 MB/s) * 1024 MB/GB * (1/3600 sec/Hrs.) = 4.84 Hrs

Note, from the example, the equalization takes a little less than 50 percent of the rebuild's total duration. The actual total hours vary depending upon whether or not the drives in the LUN are spread across multiple buses. If they are all on the same bus, then it will take a little longer as the rebuild rate will decrease, per Table 6.

The much lower rebuild rates used in the Low, Medium, and High priorities result in longer durations. It is informative to compare the durations for these priority types with the ASAP example. Table 7 shows the result of the calculation for the example using each of the priorities.

Table 7. Rebuild RAID 6 example with all priorities

Priority	Duration (Hrs)
Low	108
Medium	36
High	18
ASAP	5

Note that this chart applies to the same example configuration noted previously — a 500 GB SATA drive that is part of a fully bound and utilized six-disk (4+2) RAID 6 group bound in a single enclosure on a single bus.

Rebuild resource utilization

The following charts show the SP percent utilization and bandwidth utilization for the typical ASAP priority rebuild on an idle storage system. Except where noted, the rebuilds shown in the following figures do not use hot spares, and only show the resource utilizations of the rebuild phase.

Figure 2 shows the resource utilization for a RAID 1/0 ASAP rebuild. The example LUN is a 100 GB RAID 1/0 of six (3+3) 73 GB Fibre Channel hard drives. This example uses a hot spare.

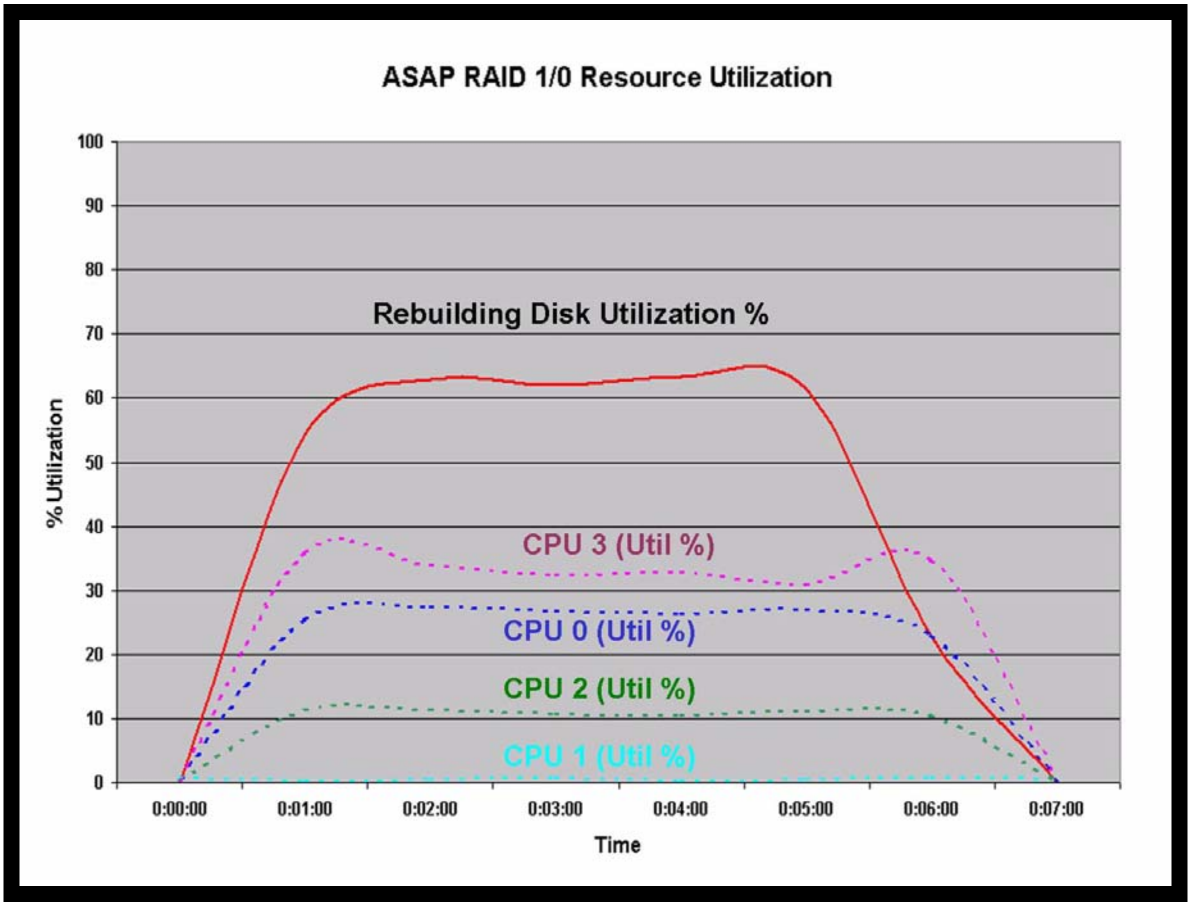

Figure 2. ASAP rebuild for a Fibre Channel RAID 1/0 group LUN

A mirror RAID type has an entirely different utilization profile from parity RAID types. On an idle system, with a mirror RAID LUN experiencing a disk failure, only the RAID group's hard drive that is duplicating the failed drive shows utilization. The other hard drives in the LUN are effectively idle. In the example shown, the mirroring hard drive writes out its contents to an available hot spare, with a high utilization.

In Figure 3, a RAID 5 (4+1) group of 750 GB SATA drives with a single 100 GB LUN is rebuilt at ASAP priority on an otherwise idle storage system. This example does not use a hot spare. (No equalization is required.)

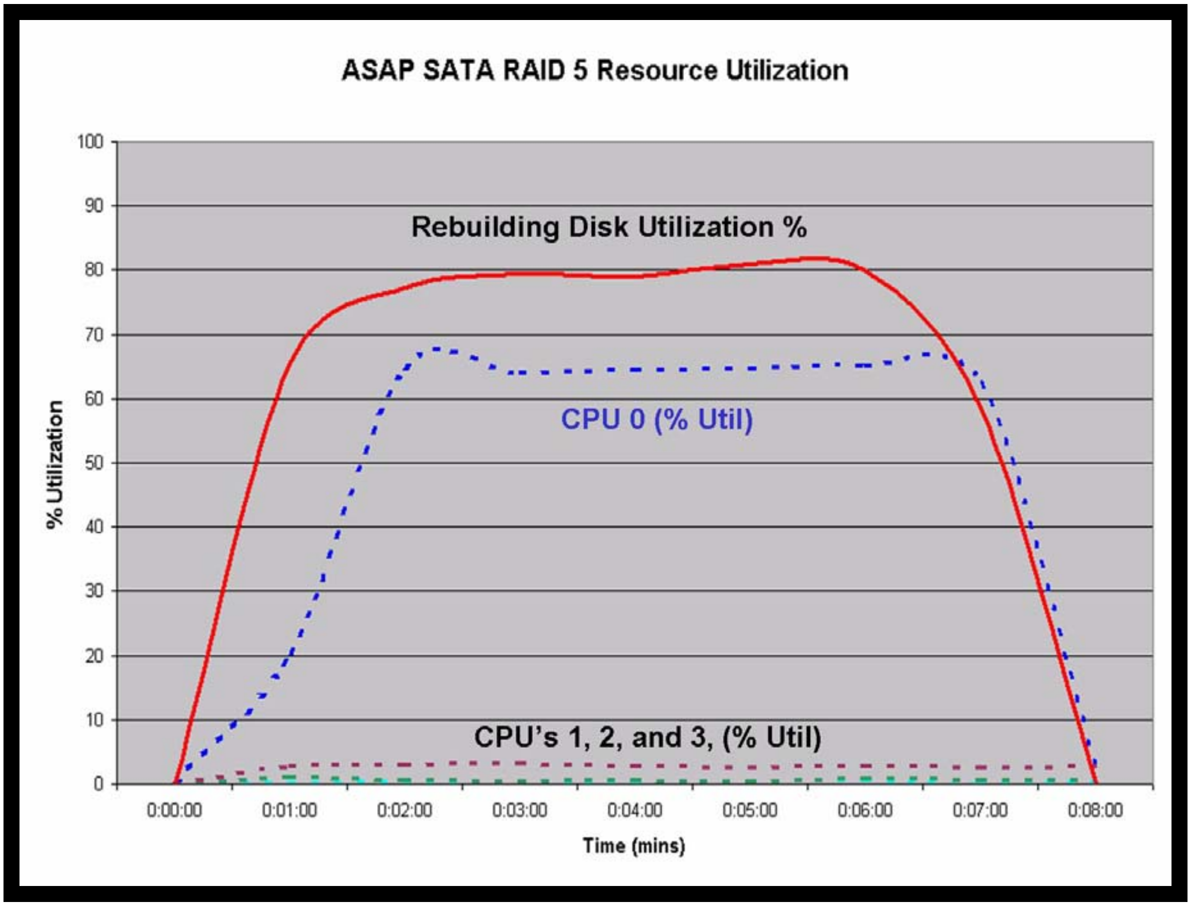

Figure 3. ASAP rebuild for a SATA RAID 5 group LUN

The average disk utilization of the LUN during the rebuild is about 80 percent. Note the higher CPU utilization over the Mirrored RAID rebuild shown in Figure 2. This is the result of parity operations in progress.

In Figure 4, a RAID 5 (4+1) group of 73 GB 15k rpm Fibre Channel drives with a single 100 GB LUN is rebuilt at ASAP priority on an otherwise idle storage system. This example does not use a hot spare. (No equalization is required.)

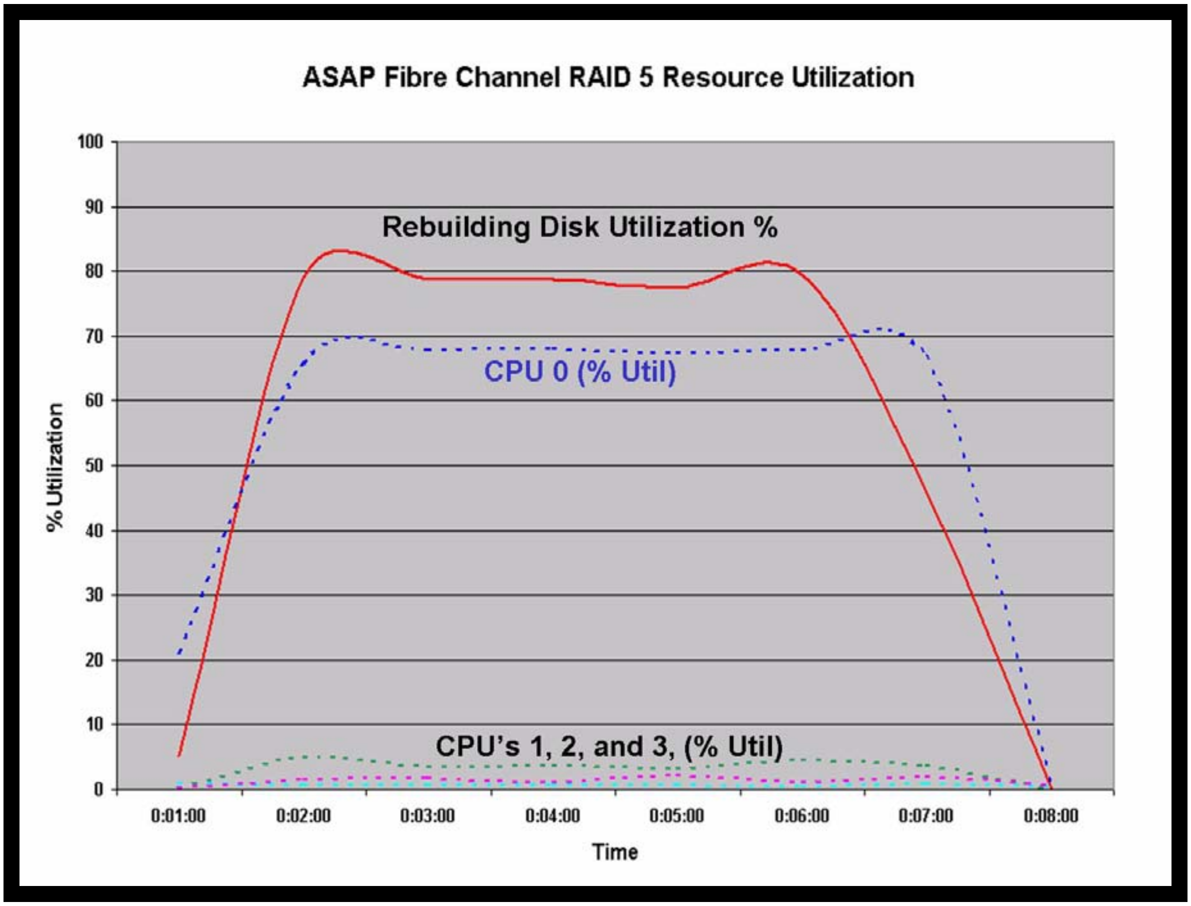

Figure 4. ASAP rebuild Fibre Channel RAID 5 group LUN

The average disk utilization of the LUN during the rebuild is about 80 percent. CPU 0 is utilized at around 75 percent throughout the operation. Because an ASAP rebuild can be so CPU intensive, the current rebuild priority defaults to “High” for new LUNs when they are created.

Comparing the LUN hard disk utilizations between SATA and Fibre Channel shows the LUN hard disk utilization is approximately the same; however the CPU Utilization is slightly higher on Fibre Channel.

Further comparing the LUN hard disk utilization between RAID 1/0 shown in Figure 2 and either RAID 5 example (Figures 3 and 4) shows the subordinate effect of RAID group type (mirror versus parity) on prioritized operations. In this case, RAID 1/0 rebuilds have a smaller effect on CPU utilization, and RAID 5 rebuilds have a larger effect on CPU utilization. However, rebuilds have a significant effect on RAID group hard drive utilization in all cases. When you schedule an ASAP rebuild, you should be aware of the performance effect that high RAID group hard disk utilization will have on *all* the LUNs of the RAID group’s hard disks.

Rebuild changes for release 28

Note there have been several changes to the Rebuild process for FLARE release 28. These changes have a large effect on the time required to rebuild.

The rebuild priority default value when binding a LUN is now "High." This is a change from recent prior releases of FLARE where the setting was "ASAP."

The rate at which rebuilds proceed at the High, Medium, and Low settings has been increased. LUNs rebuilding at the High and Medium settings will complete four times as fast as they did at the same setting in prior releases, and LUNs rebuilding at the Low setting will complete twice as fast. The ASAP rate remains the same. All operational time calculations quoted earlier in this paper are applicable to release 28.

LUN binds

A bind creates LUNs on a RAID group.

Why bind?

A bind is an information organization, data security, and data integrity feature of CLARiiON. Binding a LUN involves the preparation of allocated storage space. This preparation is particularly important when storage capacity is being reallocated for reuse. This reuse of storage includes erasing any previous data found on the hard drives, and the setting of parity and metadata for the storage.

The Fastbind feature of FLARE makes LUNs immediately available to the user after they are created. In a Fastbind, the bind takes place as a background process, allowing immediate use of the storage. The information-organizing aspects of LUNs are not discussed in this section.

Bind overview

LUNs are bound after RAID groups are created. LUNs are available for use immediately after they are bound. However, the bind is not strictly complete until after all the bound storage has been prepared and verified. Depending on the LUN size and verify priority, these two steps (preparation and verification) may take several hours.

During the preparation step, the storage allocated to the LUN is overwritten with binary zeroes. These zeroes erase any previous data from the storage and set up for the parity calculation. When zeroing is complete, parity and metadata is calculated for the LUN sectors.

Verification involves a background verify (BV). A BV is a reading of the LUN’s parity sectors and verification of their contents. A BV is executed by default after a bind. This default can be manually overridden by selecting “no initial verify” during the bind operation to make the bind faster.

The resources most affected by a LUN bind are the back end and the disks.

Bind priority

Strictly speaking, there is no bind priority. The preparation step of a bind has a fixed execution rate. This rate differs slightly between different model disk drives. The verification step is prioritized; however there are only two effective rates. When a CLARiiON is idle, BV will make use of free cycles, effectively running at a higher priority than requested either manually at the initiation of the BV, or as the default verify priority specified when the LUN was bound.

Bind duration

LUNs are immediately available for use after a bind (this is referred to as a *fastbind*). However, all the operations associated with a bind may not complete for some time. The duration of a LUN bind is dependent on:

- LUN’s bind time background verify priority (rate)
- Size of the LUN being bound
- Type of drives in the LUN’s RAID group
- Potential disabling of initial verify on bind
- State of the storage system (Idle or Load)
- Position of the LUN on the hard disks of the RAID group

From this list, priority, LUN size, drive type, and verification selection have the greatest effect on duration. Table 8 shows the idle SP bind rates for available drive types assuming verification. *Note: Your bind rate may vary slightly by disk type/model.

Table 8. Average bind rates

Disk Type	Bind Rate MB/s at ASAP	Bind Rate MB/s at High	Bind Rate MB/s at Medium (default)	Bind Rate MB/s at Low
Fibre Channel	83	7.54	5.02	4.02
SATA	61.7	7.47	5.09	3.78

The following describes the calculation to determine the duration of a bind with verification in hours.

- Time: Duration of bind
- Bound LUN Capacity: Size of LUN in GB
- Bind Rate: Drive type dependent rate of bind from Table 8

Time = Bound LUN Capacity * Bind Rate

Example

How many hours will it take to bind a 100 GB LUN on a five-disk (4+1) RAID 5 group composed of 15k rpm 73 GB Fibre Channel drives with a default verify priority set to ASAP?

- Time: Duration of rebuild in hours
- Bound LUN Capacity: 100 GB
- Bind Rate: 83.0.0 MB/s (from Table 8)

Time = 100 GB * ((1/83.0 MB/s) * 1024 MB/GB * (1/3600 sec/hrs)) = 0.34 hrs

Bind resource utilization

The following chart shows the CPU percent utilization and hard drive utilization for the typical ASAP bind on an idle CX4-960. In the example, a 100 GB LUN is bound to a RAID 5, five-disk (4+1) RAID group of Fibre Channel drives.

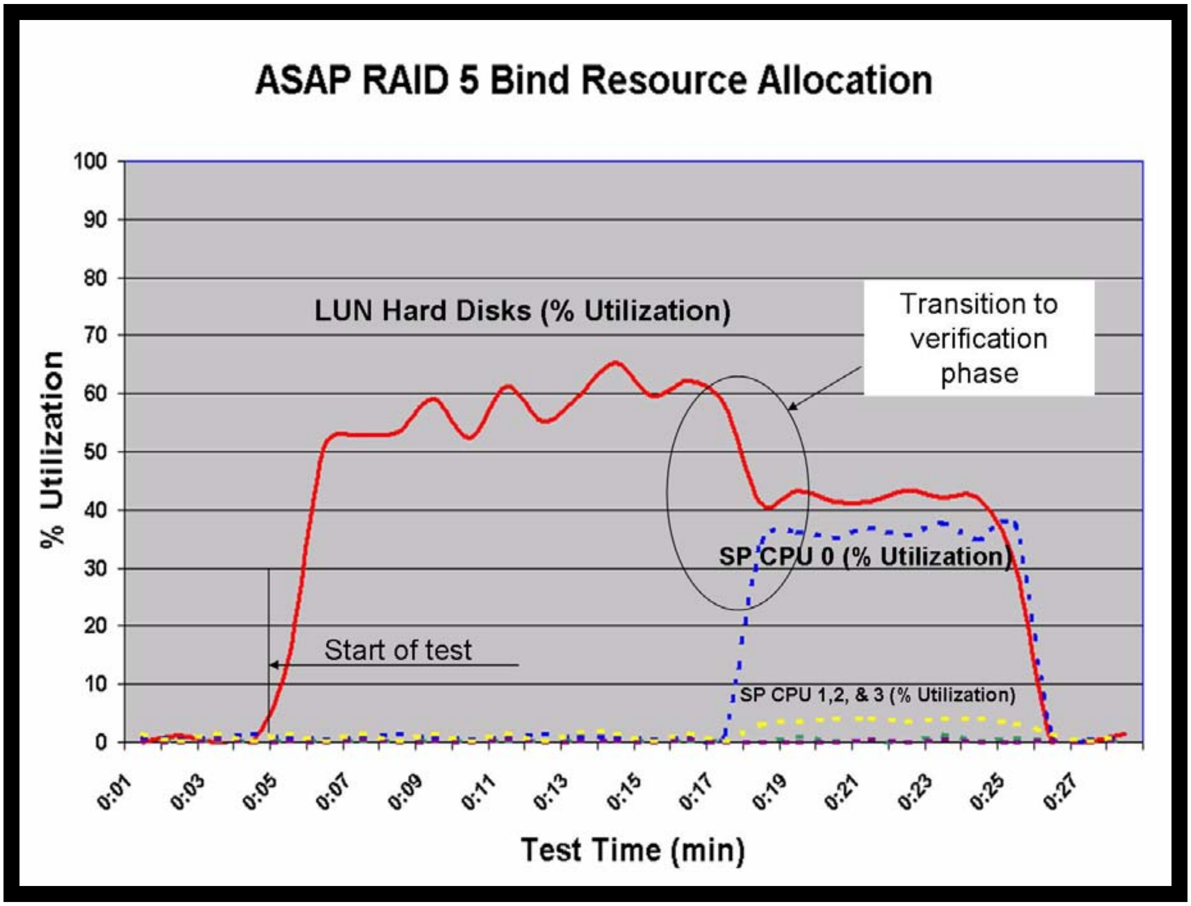

Figure 5. RAID 5 ASAP LUN bind resource utilization

The transition from the bind preparation step to the verification step can be observed in the graph about 10 minutes after the bind starts (the circled area found at the 17-minute mark on the timeline.) The preparation step occurs first. It uses the most LUN disk resources, averaging approximately 60 percent. The verification step follows. It uses significantly more CPU resources than the previous step while maintaining slightly lower disk utilization. Overall CPU resources are primarily focused on CPU 0, which remains around 35 percent utilization throughout the verify process. CPU 0 is responsible for the data integrity checking that occurs in this phase of the bind. When you schedule a bind, you should be aware of the performance effect that high RAID group hard disk utilization will have on *all* the LUNs of the RAID group's hard disks.

Bind changes for release 28

Note there have been changes to the Bind process for FLARE release 28. These changes have an effect on the time required to bind, as well as future operations that may occur on the LUN, such as a background verify. In release 28, the initial and default "background verify" rate for a LUN has been changed from "ASAP" to "Medium."

Cumulative effects on performance

The previous sections have shown the effects of priority on an operation's duration and its use of the owning SP's CPU utilization and LUN RAID group hard disk utilization on an idle storage system. The following figures show the effect of an ASAP prioritized operation on a simulated production workload from two perspectives: dependent and independent LUNs. Dependent LUNs reside on the same disks as the online transaction processing (OLTP) workload. Independent LUNs reside on different disks from the OLTP workload.

The LUN migration with expansion example is used in this section. It is a common operation and its results are typical of executing “ASAP” prioritized operations. Beginning in FLARE release 28, the default priorities of LUN operations no longer default to ASAP due to the dramatic impact it can have on application response times, as shown in the figures in this section.

In the figures, a simulated OLTP load is created on the storage system. Eighteen 50 GB, five-disk (4+1), RAID 5, 15k rpm Fibre Channel disk LUNs are participating in the workload. This *background* load exercises the SP CPUs to about 12 percent and the hard disks to about 80 percent disk utilization on average.

Independent LUN effects on performance

Figure 6 shows the resource utilization for CPUs and hard disks for an independent LUN migration. An independent LUN migration is where the LUN being migrated is not participating in a background OLTP workload. In the figure, a 20 GB independent Source LUN is migrated to a 40 GB Destination LUN. The figure shows disk utilization for both the OLTP LUN’s hard disks, and also the independent migration LUN’s hard disks. The graph makes it clear that the migration operation that occurs on independent disks does not affect the utilization of the disks involved in the OLTP load.

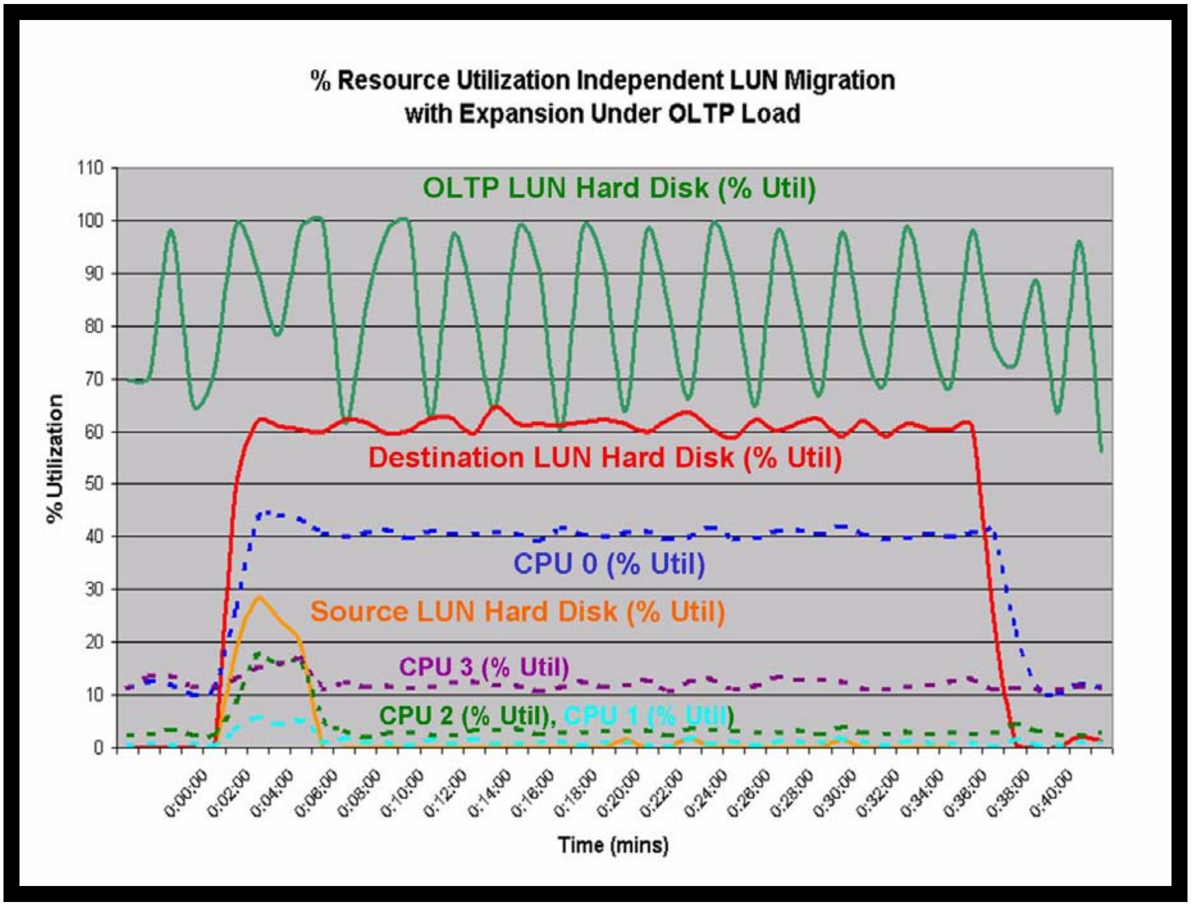

Figure 6. Percentage of resource utilization for an ASAP LUN migration with an independent LUN under an OLTP load

The average hard disk utilization for the OLTP workload shown in green is about 80 percent, and remains unaffected by the independent LUN migration operation. It fluctuates primarily due to cache flushing. The figure shows that a modest effect on CPU utilization occurs when the LUN is migrated. CPU Utilization starts around 12 percent, briefly rises, and then settles to about 40 percent. The source LUN disk utilization rises for a short while, and the destination LUN's utilization shows a sustained spike in activity for the duration of the migration.

Conclusion

Application disk utilization is not impacted by ASAP LUN Migrations taking place on other independent disks.

- OLTP Section: Green lines – The fluctuations come from write cache flushing operations. They do not change during the migration.
- Operation LUNs: Orange and Red lines – Show the duration of the migration operation with respect to the 30 percent bump in CPU utilization. Does not impact OLTP disk utilization.

One of the most important metrics for users in an OLTP environment is the response time. Therefore the LUN response time metric needs to be studied to fully explore the effects of the migration on overall performance. This metric is the time in milliseconds the storage system takes to respond to a host I/O.

Figure 7 shows the maximum response time for all LUNs participating in the OLTP load but not involved in the LUN migration. The red line shows the effects of the migration operation on the OLTP load max response time. Response time increases by over 40 percent during the operation. A 40 percent increase on OLTP application response times during this ASAP independent LUN migration is significant and is likely to be noticeable.

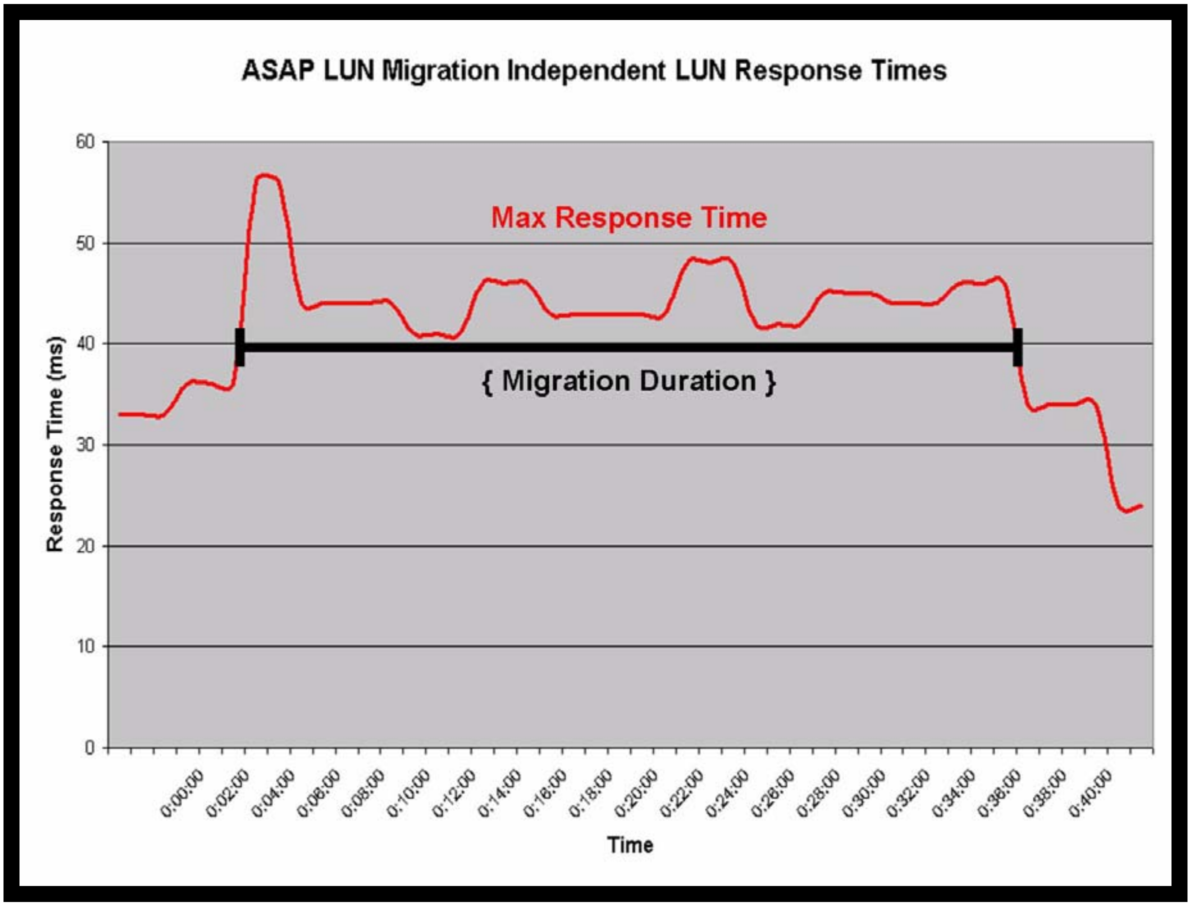

Figure 7. ASAP LUN migration independent LUN response time

Of further interest is the relative resource utilizations using the other available priorities. Figure 8 shows the hard disk utilization for the independent LUN example for RAID 5 LUN to RAID 5 LUN migration at the four available priorities. In other words, it repeats the operation shown in Figure 6 at the Low, Medium, and High priorities.

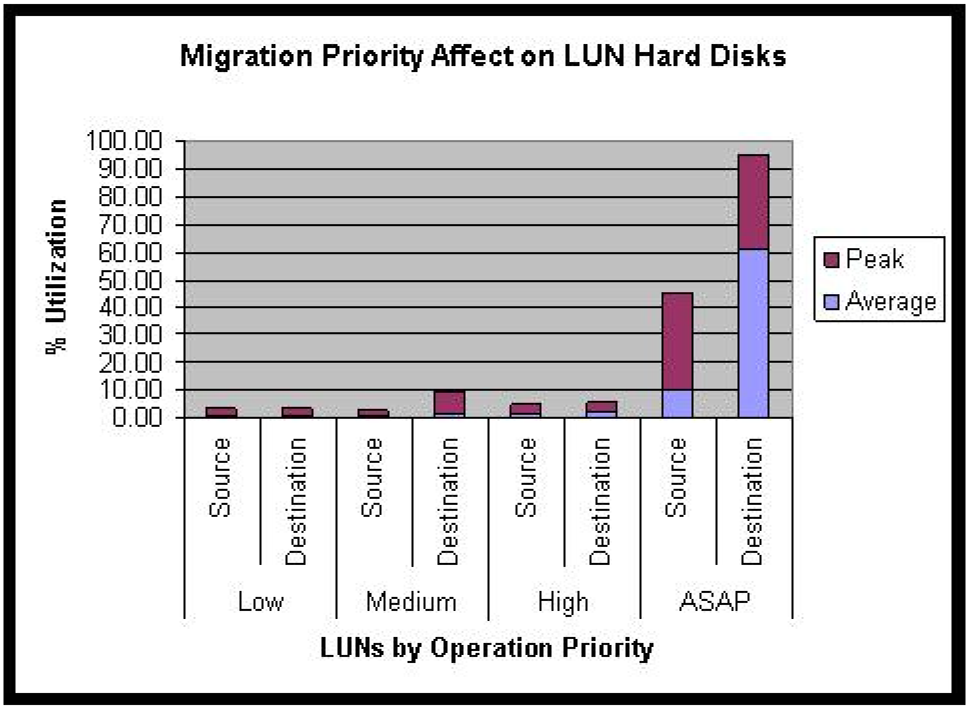

Figure 8. Migration priority effect on LUN hard disks

Compare the hard disk utilizations for the different priorities. The figure shows that the more economical priorities (Low, Medium, High) use very little of the disk resource during a migration. ASAP migrations use a significant amount of the disk resource.

The example of this LUN migration is similar to the results of all prioritized operations. Using the economical priorities substantially reduces both the CPU and hard disk resources used in a prioritized operation.

In summary, ASAP migration and expansion of an *independent* LUN have only a modest overall effect on overall storage system resource utilization. However, note that more than one LUN can reside on an underlying RAID group. *Users need to be aware of the utilization of all the LUNs on the destination LUN's RAID group.* The effects of the priority chosen for a LUN migration will be felt by all the LUNs in the same RAID group(s) as the LUNs being migrated. In addition, scheduling more than one LUN management operation at the same time can consume a lot of the available SP CPU resources. It is advisable to perform independent LUN operations one at a time. Consider the following recommendations:

- Do not use ASAP priority on LUN operations when application performance is a consideration.
- Whenever practical, stagger the LUN management operations such that only one LUN is migrating, rebuilding, or binding at a time, especially at higher priorities.

Dependent LUN effects on performance

The following example demonstrates th*e* ASAP Effect. The ASAP Effect shows how overall system performance can be adversely affected by an ASAP priority LUN management operation.

Figure 9 shows the resource utilization for CPUs and hard disks for a dependent LUN migration. A dependent LUN migration is where the LUN being migrated is participating in the OLTP workload. In the figure, the storage system is performing an ASAP 20 GB RAID 5 (4+1) LUN migration to a 40 GB RAID 5 (4+1) LUN while servicing an OLTP workload.

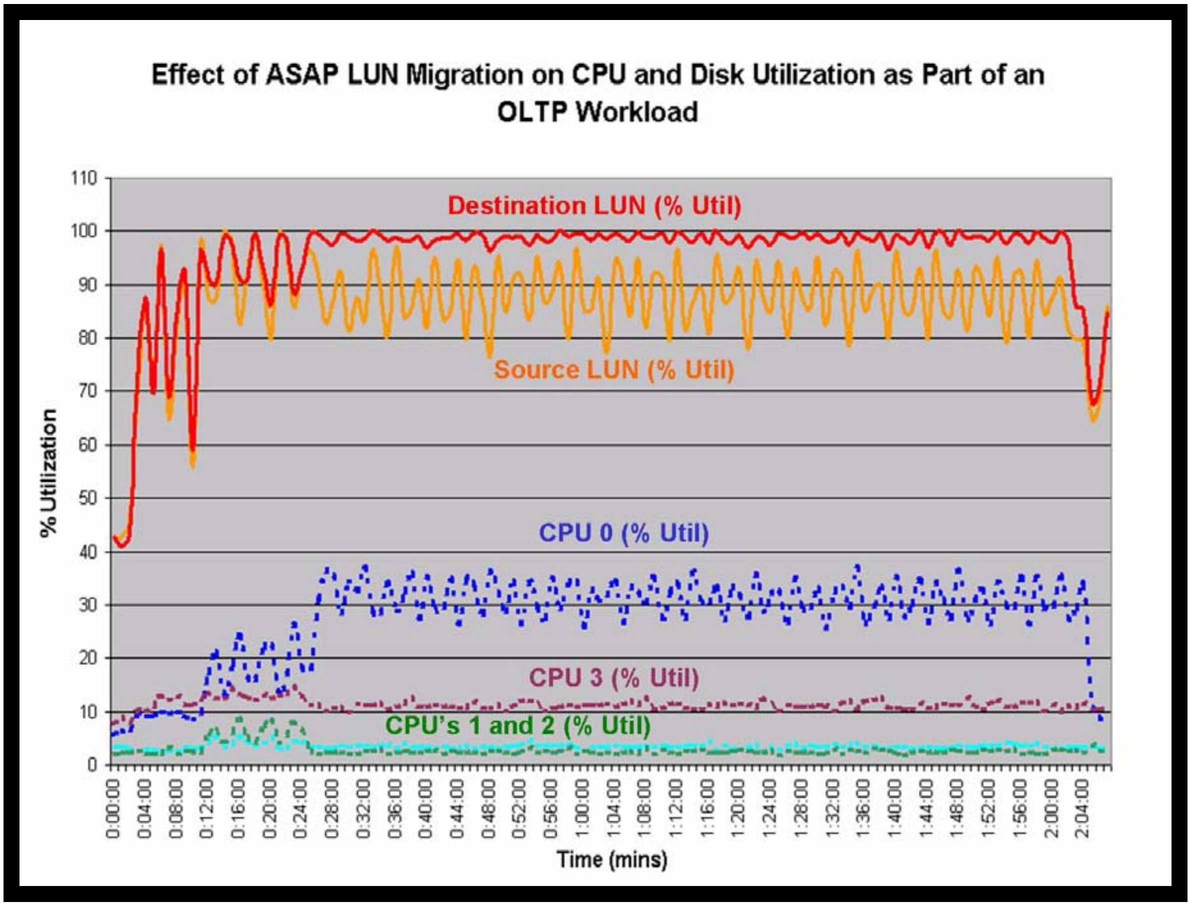

Figure 9. ASAP LUN migration resource utilization under load

Before executing the migration, the disk utilization is under 45 percent, and the CPU Utilization is at slightly under 15 percent. The figure shows there is an effect on CPU utilization when the LUN is migrated. (CPU utilization increases to about 30 percent.) In addition, the OLTP LUN hard disk utilization appears to be affected. The figure shows the destination hard disk utilization jump from around 80 percent to fully utilized. These increases may at first appear modest, but they require a review of a second metric, response time

Figure 10 shows the total LUN response time for all LUNs on the storage system for the OLTP load with dependent LUN migration.

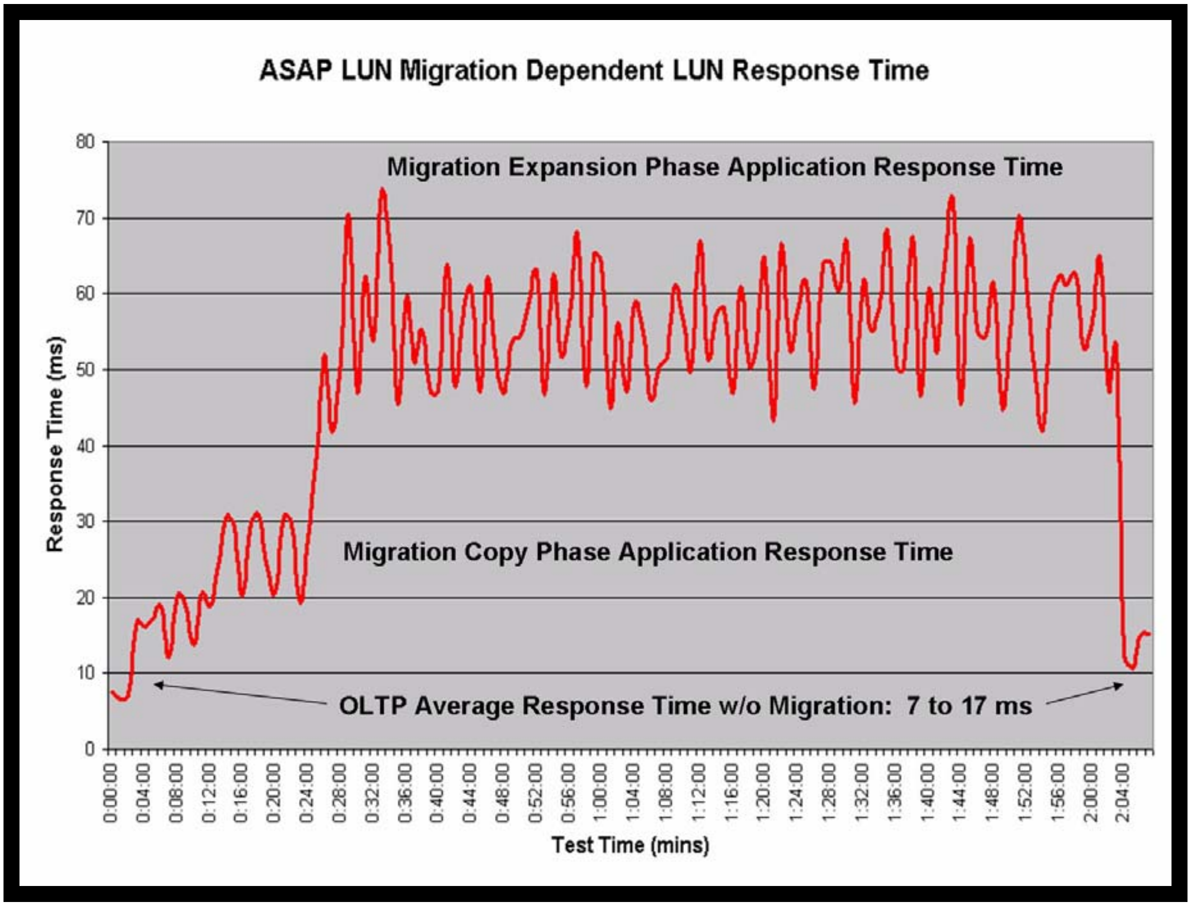

Figure 10. ASAP LUN migration dependent LUN response time

An average response time for the baseline OLTP load is about 12 ms. While the participating OLTP LUN is being migrated, the figure shows the LUN response time has increased to an average of almost 60 ms. The migration of a dependent LUN has a significant adverse effect on the OLTP application's overall response time. In this case, response times quadrupled. In some cases, the "ASAP" Effect can be even more dramatic.

In summary, a user executing an ASAP LUN migration on a dependent LUN is allocating storage system resources *away* from the production workload. It is advisable to plan ahead, and use the economical LUN operation priorities when performing LUN management operations on dependent LUNs.

Conclusion

There are three resources to manage on a CLARiiON that affect overall system performance. They are:

- Storage processor CPU
- Back-end bus
- Hard disks

The speed that LUN management operations such as migrations, rebuilds, and binds execute is dependent on the following factors in order of impact:

- Priority
- Workload
- Underlying RAID type

- RAID group's hard drive type
- Cache settings

Priority is the primary influence on the speed of these LUN operations, but the other factors can contribute in varying degrees. Choosing to execute LUN management operations at the ASAP level completes them in the shortest amount of time. However, the ASAP priority places the greatest demand on the CLARiiON's resources. To mitigate the potential adverse effects of ASAP prioritized operations on customer applications, beginning in release 28 of FLARE the default priorities of most operations have been changed to "High," or "Medium."

The current application workload and its resource utilization must also be factored into a decision to perform ASAP priority operations. In particular, scheduling ASAP operations at the same time with a production load can have an adverse effect on overall system performance.

Users should use the ASAP priority level with care. Scheduling operations at ASAP priority or leaving ASAP as a default for automatic operations such as rebuild can result in an unanticipated diversion of CLARiiON resources. In a production environment, application performance will be degraded while the operation completes. Users can avoid this "ASAP Effect" by planning ahead, and using the economical LUN management priorities

References

- *EMC ControlCenter Navisphere Analyzer Version 6.X Administrator's Guide* (July 2002)
- *EMC CLARiiON Performance and Availability – Applied Best Practices* white paper (October 2008)
- *EMC CLARiiON Fibre Channel Storage Fundamentals* white paper (September 2007)