White Paper

Leveraging EMC FAST Cache with Oracle OLTP Database Applications

Applied Technology

Abstract

This white paper introduces EMC's latest groundbreaking technology, FAST Cache, and how users can leverage it with Oracle database applications. This white paper covers several use cases that were lab-tested and the results from those tests. This paper also discusses best practices for implementing FAST Cache with Oracle database applications.

August 2010

Part Number h8018

Table of Contents

Executive summary

EMC® FAST Cache technology introduces an extra layer of cache using Flash drives between DRAM cache and rotating spindles. The database application hot and cold data is automatically identified and hot data is cached by this newly created cache layer on Flash drives. This technology provides very low latencies to frequently accessed data, thus improving the overall application response times and significantly reducing the investment in Flash drives. In 2008 EMC introduced the Flash drive technology to its midrange storage arrays as a high-performing tier of storage. Although using Flash drive technology has many advantages, users face challenges such as the following for implementing the drives as a tier:

- Users must move the entire dataset to a Flash-based tier to take full advantage of the drives.
- The task of identifying hot data to move to Flash drives is not only complex but also repetitive because of following factors and trends:
    - Constant data temperature changes, that is, the hot data keeps changing over time
    - Massive data consolidation driven by virtualization
    - Data explosion where the customer datasets are growing by 60 percent every year

FAST Cache technology eliminates the need for manually identifying hot data by automating the task. By focusing the use of Flash drives toward the most frequently accessed data, the investment made toward Flash drives proves to be beneficial since it delivers optimal magnitude of application service improvement, and significantly reduces total cost of ownership (TCO) of the solution. EMC FAST Cache technology is immediately available on both the Celerra® unified storage platform and CLARiiON® CX4® platform.

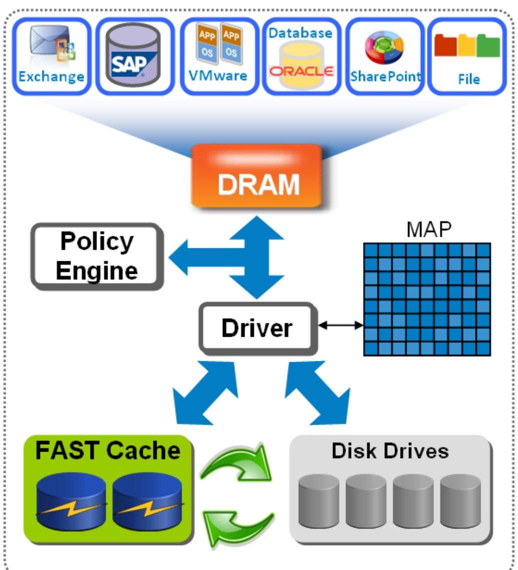

Figure 1. FAST Cache

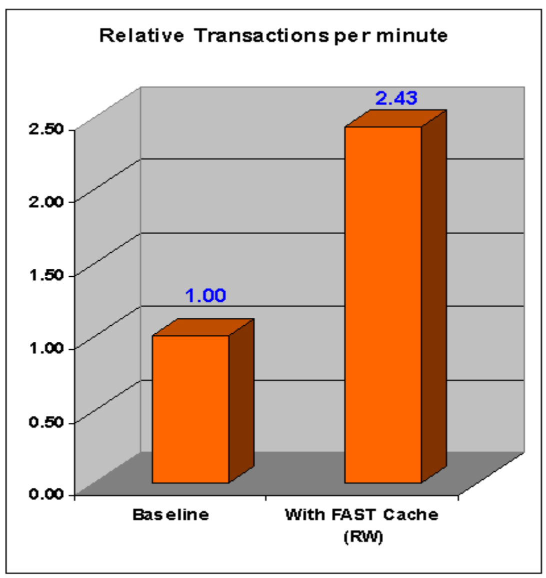

Figure 2. Benefits of FAST Cache

The need for FAST Cache

Online Transaction Processing (OLTP) database applications tend to be mission-critical and usually have stringent I/O latency requirements. Traditionally, these OLTP databases are deployed on a huge number of rotating Fibre Channel (FC) spindles (a process known as short-stroking) to meet the low I/O latency requirement. Therefore, when OLTP databases are deployed on a huge number of rotating spindles, the effective capacity utilization of these spindles is very low, thereby increasing the cost of the solution. Over the last few decades, the disk drive technology has not evolved as much as CPU technology. In real-world applications, even the fastest rotating spindle with a 15k rpm speed performs only around 180 to 200 random 8K IOPS in most application workloads. Flash drives are a disruptive technology that significantly improve IOPS/$ but come at a relatively higher $/GB cost. As long as OLTP databases are small enough to fit on a fewer number of Flash drives, these drives provide the best possible $/TPM (transactions per minute) for database applications. Once the database outgrows the space available in those few drives, database administrators and storage administrators must spend valuable time and energy identifying hot and cold data in order to redistribute the data to the right tiers. Unfortunately, the data pruning exercise is a repetitive process, sometimes involving unwanted downtime to execute depending on the technology used. Storage technologies like CLARiiON virtual LUN migration can alleviate this problem to some extent but requires careful database partitioning and layout.

EMC FAST Cache technology is part of the Fully Automated Storage Tiering (FAST) suite of EMC products that have two primary goals:

- Reduce the need to buy more Flash drives to keep up with database growth.
- Automatically and nondisruptively migrate hot and cold data between the available storage tiers, thereby improving the effective storage utilization

FAST Cache technology creates a faster medium for storing frequently accessed data on Flash drives. This significantly reduces the need to buy more Flash drives and provides Flash latencies to frequently accessed data. The hot/cold data is cached in, and cold data flushed out of FAST Cache automatically and transparently depending on data usage patterns, thus eliminating the need for users to manually classify the hot/cold data. Once the hot data is cached by FAST Cache, the FAST Cache handles the majority of I/O and the underlying rotating spindles will receive very few IOPS. This effect enables these possibilities:

- You can deploy the database on fewer rotating spindles
- You can deploy the database on slower spindles like 10k rpm drives or SATA drives
- You can utilize more of the rotating spindles' available capacity without significantly impacting the overall application latency

Introduction

This white paper introduces EMC FAST Cache technology and explains how it can be leveraged by Oracle OLTP applications. This paper discusses in detail the positive effects of deploying EMC FAST Cache. Some of these benefits include improved TPM and lowered latency. It also documents the test configurations and performance results of various tests conducted at EMC engineering labs. The testing procedure involved establishing the baseline by running the database benchmark on pure rotating Fibre Channel drives, comparing these baseline numbers with application metrics after adding the FAST Cache to the storage subsystem, and enabling it on all database LUNs where database files reside.

Audience

This white paper is intended for:

- Database and storage administrators
- Database and storage architects
- Anybody interested in Oracle database performance

Technology overview

EMC Celerra unified storage

EMC Celerra unified storage platforms combine an IP storage enclosure and best-in-class, native CLARiiON storage using NAS, iSCSI, and Fibre Channel in a single packaged solution. Having built on a robust platform like CLARiiON, Celerra unified storage inherits all the high-availability (like five 9s) characteristics of a CLARiiON storage platform along with support for almost all of the latest technologies available on CLARiiON. The EMC unified storage platform also supports the latest generation of disk drive technologies including Flash drives, Fibre Channel drives, and SATA II.

FAST Cache technology is supported on all models of the Celerra unified storage platform except NX4:

	NX4	NS-120	NS-480	NS-960
Availability	Failover	Failover	Advanced Failover (N+1/N+M)	Advanced Failover (N+1/N+M)
Number of X-Blades	1 or 2	1 or 2	2 or 4	2–8
Connectivity	NAS, iSCSI, Fibre Channel	NAS, MPFS, iSCSI, Fibre Channel	NAS, MPFS, iSCSI, Fibre Channel	NAS, MPFS, iSCSI, Fibre Channel
Storage	CLARiiON	CLARiiON	CLARiiON	CLARiiON
Maximum usable IP storage capacity (X-Blade/system)	16 TB/ 32 TB	32 TB/ 64 TB	64 TB/ 192 TB	128 TB/ 740 TB

Figure 3. Celerra unified storage models

EMC CLARiiON CX4

The EMC CLARiiON CX4 series with UltraFlex™ technology is based on breakthrough architecture and extensive technological innovation, to provide a very competitive midrange storage solution. The CX4 is the fourth-generation CX series, and continues EMC's commitment to maximizing customer's investments in CLARiiON technology because it ensures that existing resources and capital assets are optimally utilized as customers adopt new technology.

CLARiiON CX4 systems support the latest generation of disk drive technologies like Flash drives, 4 Gb/s FC drives for high performance, and SATA II for high capacity. CLARiiON CX4 is the first midrange storage system to support all of these types of disk drive technologies. The CLARiiON CX4 with the latest FLARE® release and FAST suite of software provides maximum performance and tiered storage functional flexibility. Although this paper does not contain a complete introduction to the CX4 series, you can obtain more documents that provide this information in the "References" section. A few major features of the CLARiiON CX4 series are listed next.

All four models of CLARiiON CX4 listed in Figure 4 support FAST Cache technology:

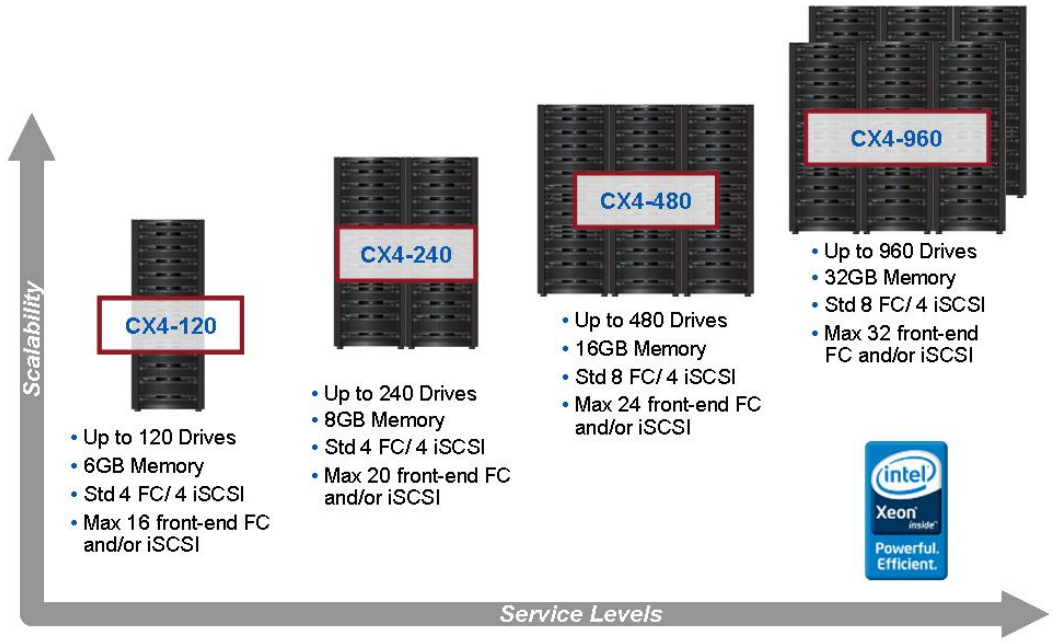

Figure 4. CLARiiON CX4 models

Most of the data points discussed in this paper are in the CLARiiON CX4 context but still apply to the native block implementation of any Celerra unified storage platform.

EMC Flash drive technology

The enterprise-class Flash drives supported by CLARiiON CX4 and the Celerra unified storage platform are constructed with a nonvolatile semiconductor NAND Flash memory and are packaged in a standard 3.5-inch disk drive form factor used in existing CLARiiON disk drive array enclosures. These drives are especially well suited for latency-sensitive applications that require consistently low read/write response times. Flash drives also benefit from the advanced capabilities that CLARiiON provides like local and remote replication, virtual LUN migration, Quality of Service Manager, and availability features like redundant failover, hot sparing, and multiple RAID levels. Several studies regarding the applicability of CLARiiON Flash drive technology to various enterprise applications have been conducted. For more information about these studies, see the "References" section.

EMC FAST Cache technology

In midrange storage systems, the DRAM caches are relatively small in proportion to the active data being accessed by the host application. As a result, the data residency in DRAM is shorter as new data takes the place of the older data. The ability to expand the size of DRAM may be both disruptive and cost-prohibitive. FAST Cache technology builds a near DRAM speed caching area for frequently accessed data on Flash drives, thereby allowing the hotter data to stay longer on a faster medium. FAST Cache tracks the data temperature at a 64 KB chunk size and caches the chunks in the Flash tier once its temperature reaches a certain threshold. After a data chunk gets promoted to FAST Cache, the subsequent accesses to that chunk of data will be served at Flash latencies. Eventually when the data temperature cools down, the data chunks are evicted from FAST Cache and are replaced by new, hot data. The FAST Cache uses a simple LRU (least re-used) mechanism to evict the data chunks.

FAST Cache is built on the premise that the overall applications' latencies can improve when most frequently accessed data is maintained on a relatively smaller sized, but faster storage medium, such as

Flash drives. FAST Cache technology identifies the frequently accessed data and moves it to Flash drives automatically and nondisruptively. The data movement is completely transparent to applications, thereby making this technology application-agnostic and management-free. It is as simple as turning on/off a switch. For example, you can enable or disable the FAST Cache on any LUN/pool simply by selecting or clearing the LUN/Storage pool cache checkbox settings as shown in Figure 5.

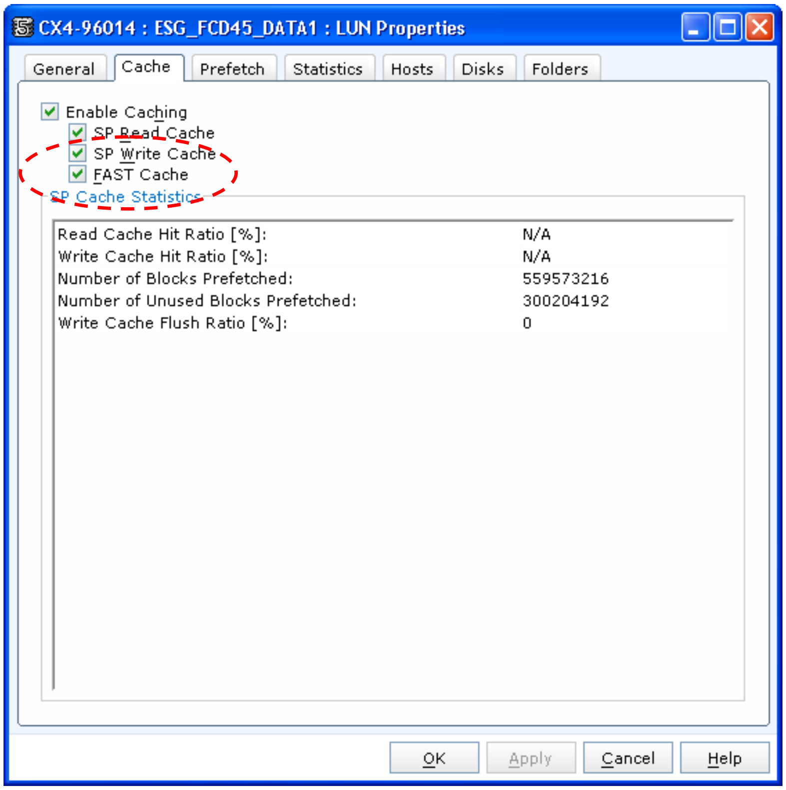

Figure 5. How to enable FAST Cache

The LUN/pool-level granularity enables FAST Cache to be selectively enabled on a few LUNs/pools within a storage array depending on the application performance requirements and SLAs.

There are several distinctions to EMC FAST Cache technology, including:

- FAST Cache is created on a persistent medium like Flash drives, which can be accessed by both storage processors. This approach has several advantages:
  - FAST Cache does not consume any extra PCI-E slots inside the storage processor.
  - In the event of one storage processor failure, the surviving storage processor can simply continue accessing FAST Cache that is already populated by the failed storage processor.
  - In the event of a complete power outage, upon power restoration both storage processors can simply reload the cache from a persistent medium instead of repopulating it from scratch by observing the data access patterns again.
- Users can create a FAST Cache without incurring any array downtime. The process involves simply selecting the Flash drives that are going to be part of FAST Cache and does not require any array downtime.

Applying FAST Cache technology with an Oracle database

Users need to thoroughly understand the application workload characteristics before implementing any storage technology, including FAST Cache. While comprehending basic information like workload I/O characteristics is important, it is vital that users understand the active working set of the workload for right-sizing the FAST Cache. The following sections explain in detail how to determine the application I/O characteristics and application working set.

Application I/O characteristics

The relative improvement from any faster storage technology is dependent on whether the application is bottlenecked by the storage subsystem. For example, you can use Oracle AWR (Automated Workload Repository) reports to identify potential I/O bottlenecks with any Oracle application. There are several white papers and documents that describe how to use database performance information to determine whether an application is I/O bound. Generally, most Oracle DBAs and storage administrators turn to the I/O subsystem because this solution is simpler than restructuring application to avoid I/O subsystem bottleneck. To set the right expectations, you should thoroughly analyze database-level performance before investing time and additional resources (such as deploying more Flash drives) to accelerate the storage subsystem.

Understanding the database working set

This step is critical to maximize the benefits from FAST Cache. Generally, FAST Cache is a very suitable technology for workloads with small random I/O and relatively small working sets that fit in the FAST Cache nicely. Most Oracle OLTP databases tend to be highly random in nature with small working sets compared to the total database size. Normally, databases have datasets with varied I/O patterns because of following reasons:

- OLTP databases tend to be temporal in nature, as the most recent data is more important than older data. This is generally referred to as a working set of a database.
- The relative importance of data changes from object to object. Some tables may be accessed more often than some other tables.
- The number of IOPS per gigabyte size of an object, also known as object intensity, changes quite significantly. The best example is an index compared with a table in a database. The relative IOPS received by a database block occupied by an index object are very high compared to the IOPS received by a database block consumed by a table object.

The following figures track the read I/Os received by every 1 GB slice of the database LUN from two different environments. The test database used for characterizing FAST Cache in the EMC lab is a 1.2 TB Oracle database deployed on a 2 TB ASM diskgroup created on 8 x 250 LUNs. Figure 6 represents the read I/O hit rate of a 250 GB LUN from the benchmark test in the lab. Similarly, Figure 7 represents the read I/O hit rate of a 600 GB LUN from a production environment of a customer.

The 250 GB LUN from the benchmarking environment in Figure 6, almost has 30 percent to 40 percent of the slices receiving very high I/O, whereas the 600 GB customer LUN in Figure 7 has only around 10 percent of data active. These figures clearly indicate the following:

- The various parts of dataset have varied temperatures.
- The real-world applications tend to have smaller working sets compared to total database size.
- The benchmarking environment used in the lab is aggressive with a large working set. If FAST Cache can show improvement with a benchmark used in the lab, the real-world applications should see similar or even better improvements with their applications.

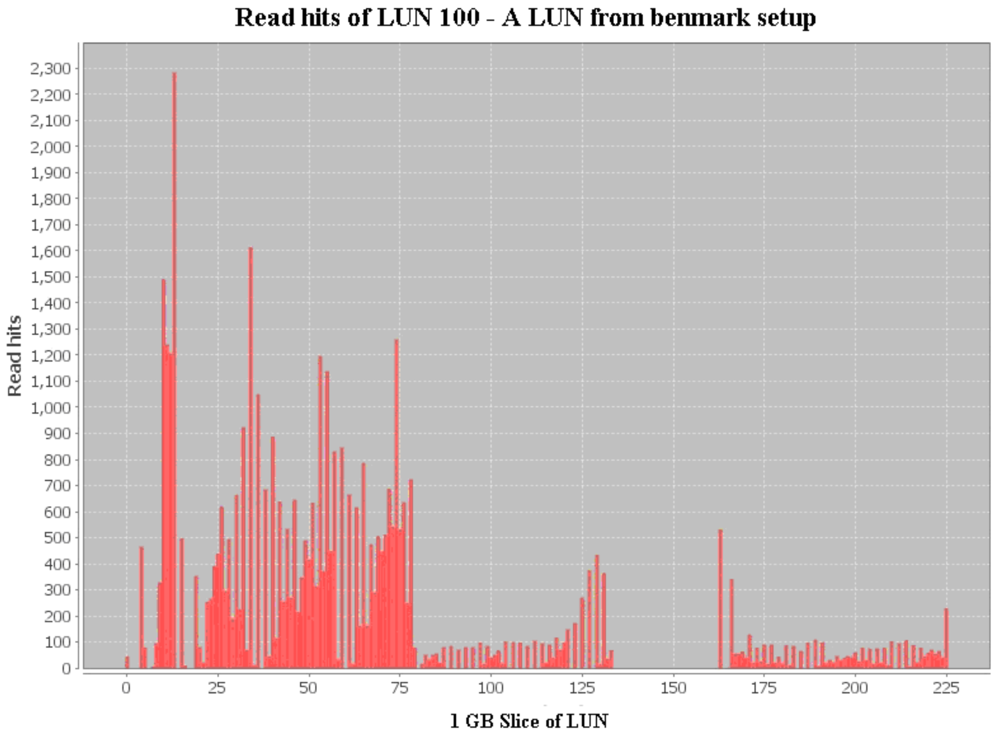

Figure 6. Slice hit chart from the benchmark test

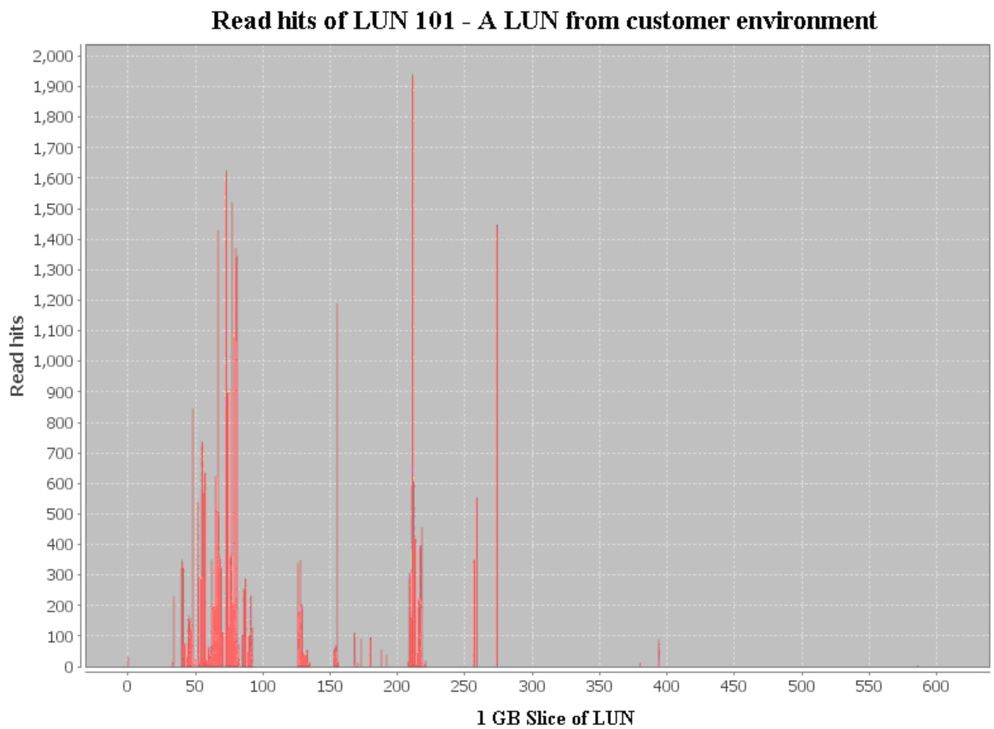

Figure 7. Slice hit chart from the customer

Benchmark setup

To characterize FAST Cache with Oracle OLTP workloads, the following benchmark setup is used: The benchmark testing is done on a single-instance Oracle 11g R2 database deployed on a CLARiiON CX4-960 storage array and a Dell R900 server. The hardware configuration is shown in Figure 8.

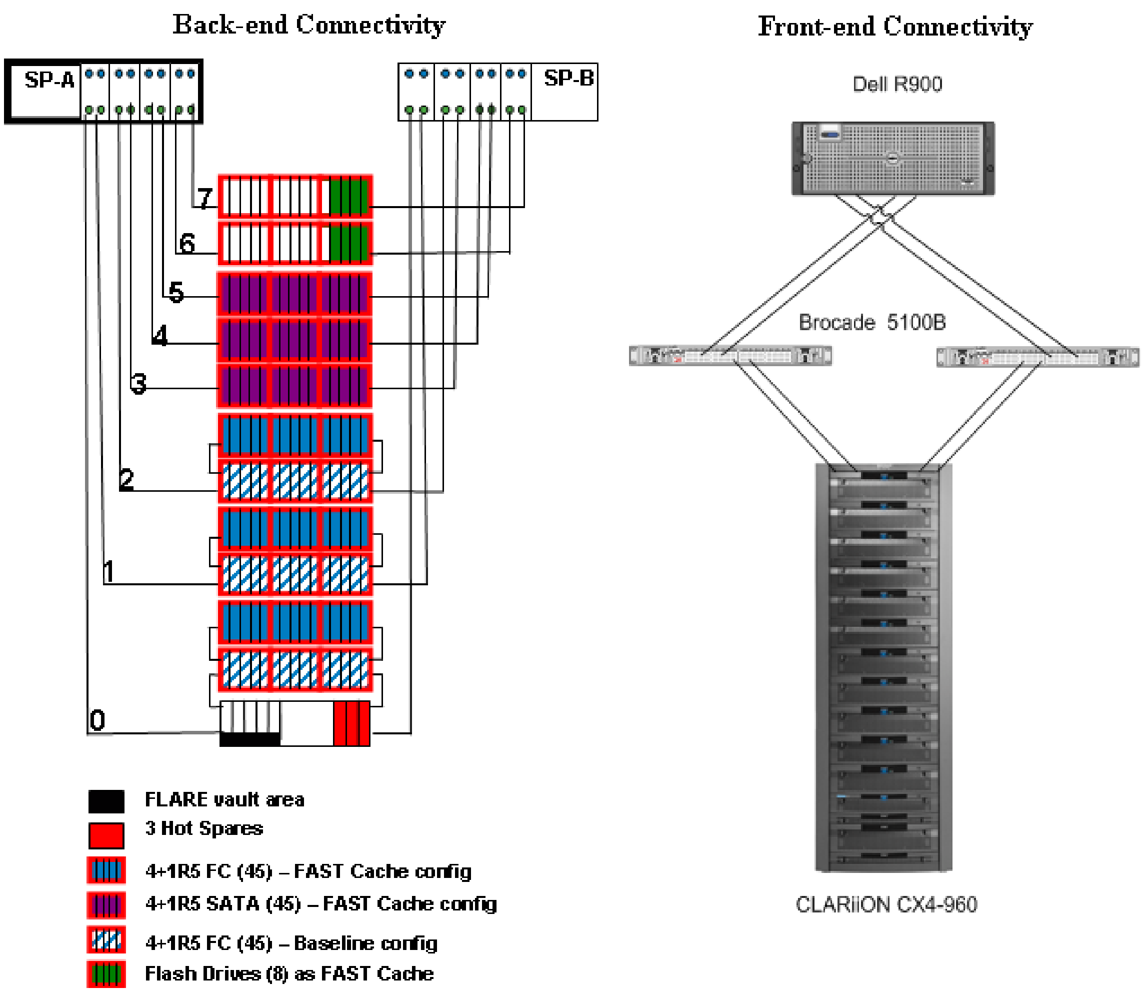

Figure 8. Hardware setup

In the configuration above, all the available eight back-end buses on the CX4-960 are used just to ensure that the benchmark is not bus-bound. There is no real requirement to utilize all back-end buses to realize FAST Cache benefits. Configure your back-end buses depending on your I/O requirement and following EMC best practices for bus and disk layout.

Workload description

The database workload simulates an Order Entry system. The database is laid out on Oracle ASM diskgroups and the underlying LUN layout is chosen so that both storage processors handle an equal number of LUNs. The Oracle ASM does a pretty good job of distributing workload against all available LUNs within an ASM diskgroup, thereby loading both storage processors equally.

Generally, most OLTP databases tend to have a read/write mix of 80/20. To ensure that the sub-LUN FAST technology can handle workloads involving a high write mix, EMC created an aggressive test scenario with a 60/40 read/write mix. The Oracle OLTP database benchmark used in the test has the characteristics shown in Table 1.

Table 1. Workload characteristics

Property	Value / Description
Database size	1.2 TB
Database version	Oracle 11g R2 single instance
Storage type	Oracle ASM
Storage container size	2.0 TB – ASM diskgroup created on 8 x 250 GB or 4 x 500 GB LUNs
Read / Write ratio	60 / 40
Database metric	TPM – Transactions per minute
Working set	Around 200 to 300 GB of data receiving most of the I/O
Number of concurrent users	100

EMC kept consistent workload characteristics for all the use cases described below so users can analyze the application metrics.

Use case description

EMC conducted the following tests to understand the impact of FAST Cache on Oracle OLTP databases and to compare the impact of FAST Cache on the workload using the application-level and database-level metrics. The goal of these use cases is to show that FAST Cache can significantly reduce the number of IOPS received by rotating spindles once the hot data is cached by FAST Cache. The reduction in IOPS of a rotating spindle count opens up several possibilities either to create a configuration with a fewer number of rotating Fibre Channel drives or to completely replace the rotating Fibre Channel drives with SATA drives. The following use cases are crafted to demonstrate this very basic point.

- **Baseline:** The baseline metric is established on a 45 x 600 GB Fibre Channel drives.
- **Use case1:** Add FAST Cache with 8 x 73 GB Flash drives to the baseline environment.
- **Use case2:** Repeat use case1 where 45 x 600 GB Fibre Channel drives are replaced by 25 x 600 GB Fibre Channel drives and front-ended by FAST Cache created on 8 x 73 GB Flash drives.
- **Use case3:** Repeat use case1 where 45 x 600 GB Fibre Channel drives are replaced by 45 x 2 TB SATA drives and front-ended by FAST Cache created on 8 x 73 GB Flash drives.

Baseline

The baseline was established on an all-Fibre Channel rotating drive configuration. There were a total of 45 600 GB disks used in this setup with Oracle data files deployed on 40 drives and database online redo logs on five drives. Figure 9 depicts the exact database and Oracle ASM layout.

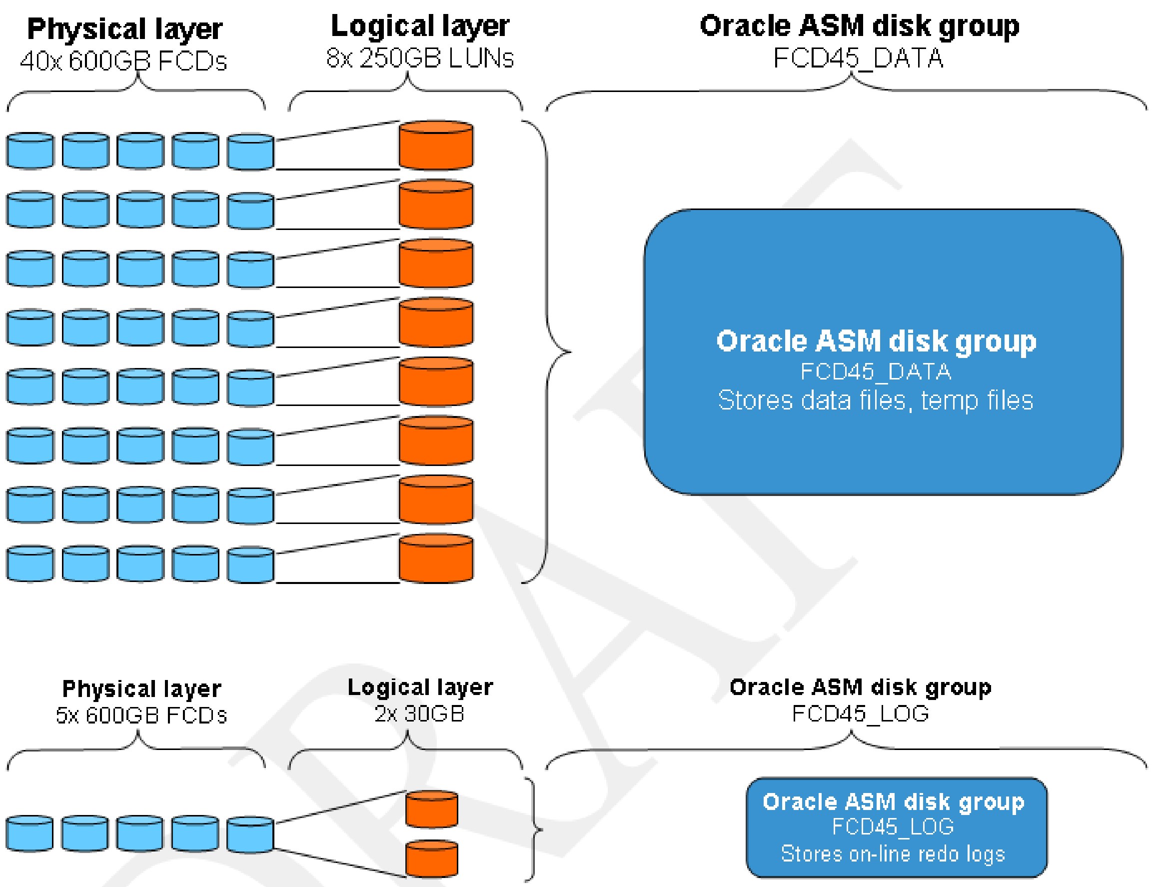

Figure 9. Baseline disk and database layout

Use case1

The database layout in this use case is identical to the baseline scenario except that FAST Cache is created on 8 x 73 GB Flash drives and is enabled on the entire data file LUNs. *FAST Cache is not enabled on the online redo log LUNs because you will yield better results by enabling FAST Cache on the data file LUNs*. In this configuration, the frequently used data is cached by FAST Cache, resulting in improved latencies to the most accessed data and thereby enhancing overall application performance. This configuration obtained 2.43 times the number of transactions at one-third the latency when compared to the baseline number of transactions. This is a significant return on a small investment made into FAST Cache.

A close analysis of the performance data from underlying rotating spindles reveals that once the hot data gets cached by FAST Cache, the 40 rotating drives with Oracle data files on them are only receiving around 27 IOPS each. The reduced number of IOPS can therefore be met by fewer rotating Fibre Channel drives, or for that matter, even by SATA drives. Thorough analysis and sizing should be done to determine the exact number of drives. Figure 10 shows the reduction in IOPS count as more and more hot data gets cached by FAST Cache.

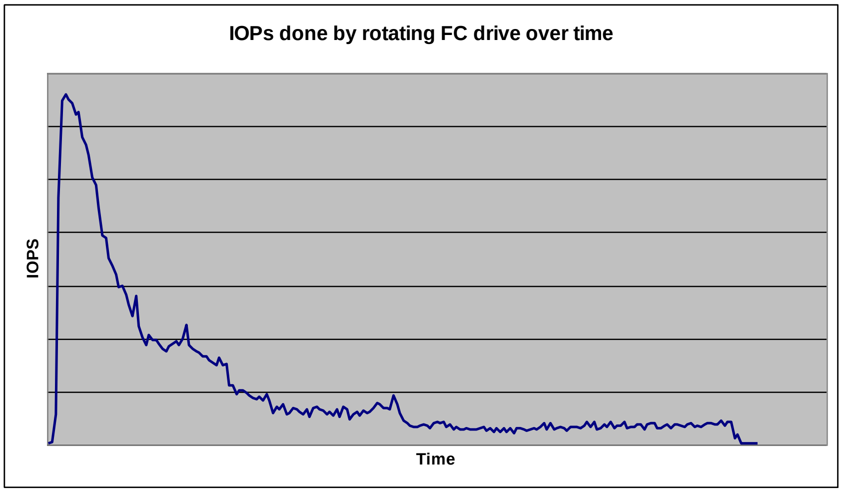

Figure 10. Rotating Fibre Channel drive IOPS over time

Use case2

Based on the performance data from use case1, the 45 underlying rotating FC drives are replaced with 25 FC drives of the same size and speed. This change reduces disk costs by almost 50 percent. You can lower the costs even further with reductions in the data center footprint, power, and cooling. Reducing the number of rotating Fibre Channel drives certainly decreases the capacity you can use, with OLTP databases; however, the entire capacity of drives was never used in real-world deployments to take advantage of short-stroking and to guarantee very low latencies. FAST Cache technology enables applications to tap into the available capacity of these drives for the first time. By implementing FAST Cache, the rotating medium receives only a small fraction of overall I/Os and the application will not be majorly impacted by those few extra disk seeks. This allows more workloads to be consolidated to the same set of available spindles. A careful analysis needs to be done to determine the right number of rotating spindles before attempting consolidation. Figure 11 shows the exact disk and database layout used for use case2.

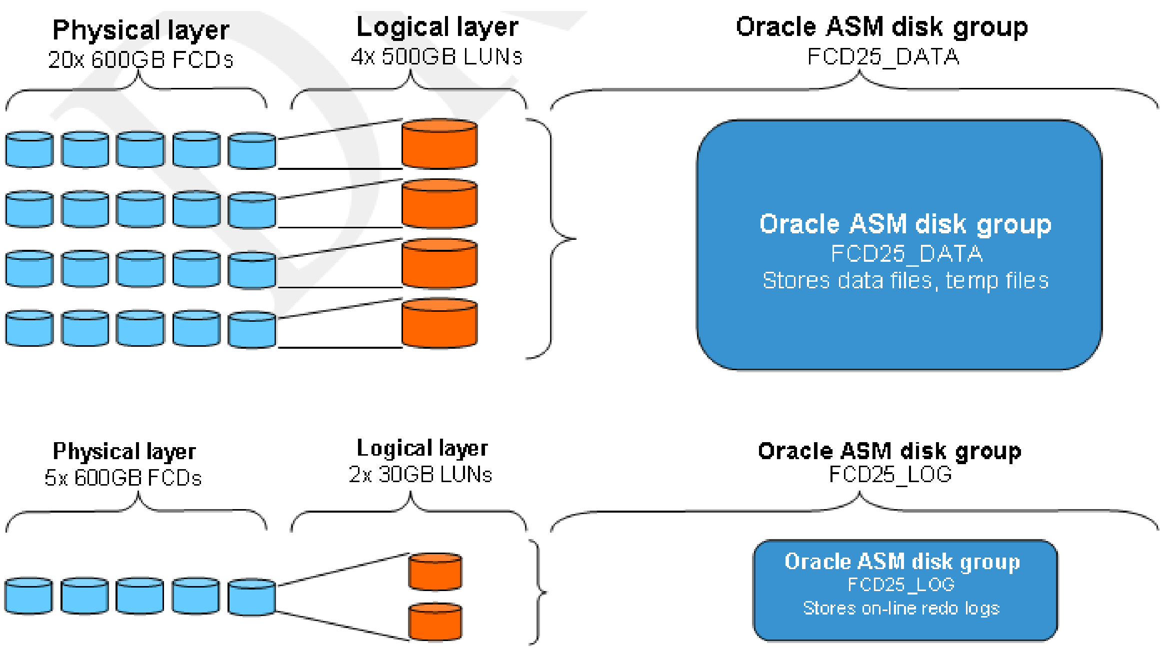

Figure 11. Use case2 disk and database layout

The 25 x 600 GB rotating Fibre Channel configuration with FAST Cache yields the same transactions per minute as that of the 45 x 600 GB Fibre Channel drive configuration with FAST Cache. This is a huge reduction in cost, which already justifies the added cost of FAST Cache. Finally, this solution achieves 2.43 times the number of transactions compared to the baseline configuration.

Use case3

The database layout in this use case is identical to that of use case1, except that the database data files are now laid out on 45 x 2 TB SATA drives instead of 45 x 600 GB rotating Fibre Channel drives. FAST Cache is also created on 8 x 73 GB Flash drives and is enabled on all data file LUNs and *not on the online redo log LUNs* because of the reasons cited earlier. The benchmark tests reveal that the application performs the same amount of transactions per minute even with this SATA configuration compared to use case1 where 45 x 600 GB Fibre Channel drives are used, This is an exceptional improvement in the TCO of the system given the expanded space and the lower price of SATA. Though the reliability of SATA drives seems to be one of the major concerns of production DBAs and storage administrators, EMC's tests indicate that the SATA drives are almost as reliable as Fibre Channel drives. Users can consider the alternate RAID options like RAID 6 or RAID 1/0 if reliability is a concern. The added parity cost of RAID 6 is not a major issue since only a fraction of I/Os are targeted at the rotating medium once FAST Cache starts absorbing the majority of the I/O.

Results summary

Figure 12 represents the relative transactions per minute compared to the baseline measurement.

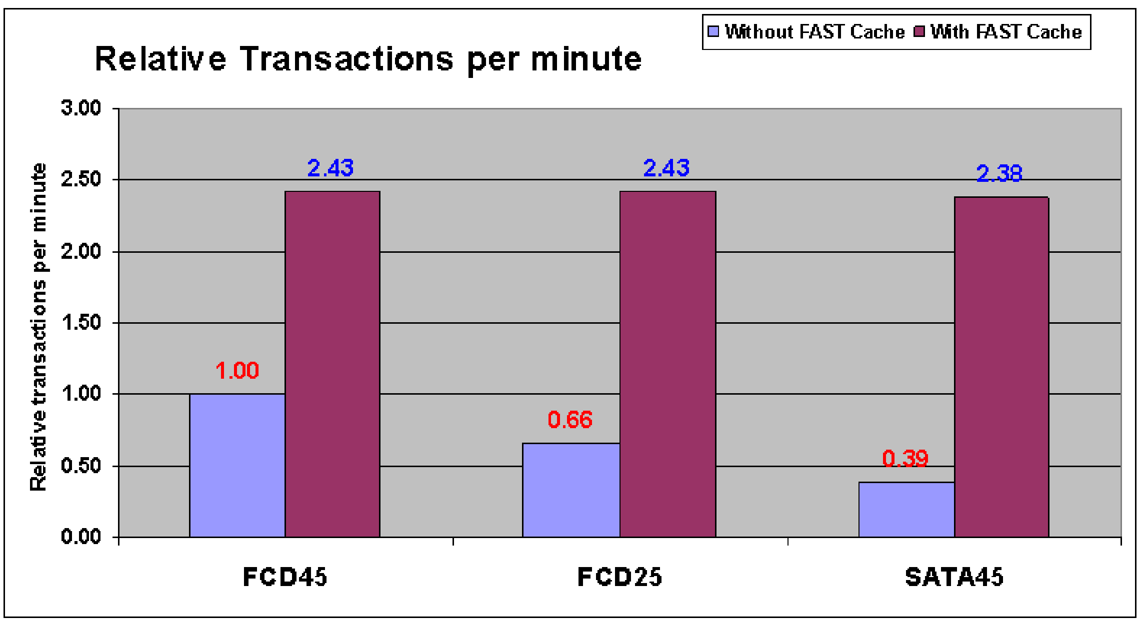

Figure 12. Relative application performance

This figure indicates that irrespective of underlying spindles, once the hot data is cached by FAST Cache, every configuration supports the same amount of transactions per minute. It is important to note that the application is able to perform almost 2.5 times the number of transactions at one-third the latency. The increased transactions and reduced latency allow the application to scale much better and improve the end-user experience significantly.

Table 2 shows the Oracle wait statistics from the Oracle AWR reports. It is important to note that with the configurations involving FAST Cache, the number of waits is higher because the database performed almost 2.5 times the number of transactions when compared to the baseline. The key metric to track is the "db file sequential read" latency, which represents the random small block I/O latency that the Oracle database incurs at every wait. There is a significant reduction in this value from 14 ms prior to implementing FAST Cache to 5 ms after implementing it. Similarly, the "log file sync" is higher in the FAST Cache configurations, once again justified by the increased number of transactions. In all the use cases, FAST Cache is never enabled for online redo log LUNs. The table illustrates that the DRAM cache and very few other dedicated rotating spindles can handle the sequential nature of the database log I/O. Even when the logs are on just five rotating spindles by themselves, the log latency is always under 3 ms.

Table 2. Oracle wait events from use cases

Baseline – 45x FCD

Event`	Waits	Time (s)	Avg (ms)	Time
db file sequential read	23,587,547	332,923	14	88.5
db file parallel read	638,376	32,504	51	8.6
CPU time		6,254		1.7
log file sync	1,467,120	1,738	1	0.5
log file parallel write	1,358,604	1,062	1	0.3

45x FCD + 8x73GB as FAST Cache

Event	Waits	Time (s)	Avg (ms)	Time
db file sequential read	61,974,617	315,080	5	82.9
db file parallel read	1,706,375	33,122	19	8.7
CPU time		19,226		5.1
log file sync	3,891,237	9,359	2	2.5
log file parallel write	1,938,577	2,479	1	0.7

25x FCD + 8x73GB as FAST Cache

Event	Waits	Time (s)	Avg (ms)	Time
db file sequential read	62,248,605	315,144	5	82.9
db file parallel read	1,696,368	32,107	19	8.4
CPU time		20,112		5.3
log file sync	3,863,427	9,428	2	2.5
log file parallel write	1,972,955	2,427	1	0.6

45x SATA + 8x73GB as FAST Cache

Event	Waits	Time (s)	Avg (ms)	Time
db file sequential read	63,197,071	315,353	5	83.1
db file parallel read	1,730,796	30,801	18	8.1
CPU time		19,818		5.2
log file sync	3,945,794	10,397	3	2.7
log file parallel write	1,922,923	2,469	1	0.7

Conclusion

FAST Cache technology significantly reduces the TCO of a database solution by reducing the investment for the Flash tier and improving capacity utilization of the rotating drives. The application-agnostic nature of FAST Cache technology makes it applicable to any application with a low I/O latency requirement. Even though FAST Cache technology can be used with any application with any type of I/O pattern, it is especially suited for applications with small random I/Os and relatively small working sets compared to the total database size. Different models of CLARiiON arrays support various sizes of FAST Cache, therefore the right sizing of FAST Cache is very important to maximize the benefit. The following best practices should be followed to realize the full advantage of FAST Cache:

- Disable FAST Cache on LUNs that do not require it.
- Size FAST Cache appropriately depending on the application active dataset.
- Disable FAST Cache on LUNs where Oracle online redo logs reside. Enabling FAST Cache on database online redo logs may or may not help, depending on workload characteristics. The relative gains of enabling FAST Cache on online redo logs will be small when compared to enabling FAST Cache on LUNs with Oracle database files.
- Never enable FAST Cache on archive logs because these files are never overwritten and rarely read back unless the database needs recovery.

References

The following white papers can be found on Powerlink:

- *EMC CLARiiON and Celerra Unified FAST Cache - A Detailed Review*
- *An Introduction to EMC CLARiiON and Celerra Unified Platform Storage Device Technology*
- *Leveraging EMC CLARiiON CX4 with Enterprise Flash Drives for Oracle Database Deployments*
- *Leveraging EMC CLARiiON CX4 with Enterprise Flash Drives for SAP Deployments*
- *Maximize the Performance Benefit from Enterprise Flash Drive (EFD) by Sharing through Virtual Provisioning in EMC CLARiiON CX4*
- *CLARiiON Virtual LUN Technology*
- *EMC CLARiiON Virtual Provisioning*
- *EMC CLARiiON CX4 Model 960 Networked Storage System* (specification sheet)
- *Introduction to the EMC CLARiiON CX4 Series Featuring UltraFlex Technology*
- *EMC CLARiiON Best Practices for Performance and Availability: Release 30.0 Firmware Update*

Further information can be found in the EMC Unified Storage for Oracle Database 11g solutions documentation on Powerlink.