ORACLE

EXADATA

An Oracle White Paper
December 2013

Exadata Smart Flash Cache Features and the Oracle Exadata Database Machine

ORACLE

Flash Technology and the Exadata Database Machine

The Oracle Exadata Database Machine is engineered to be the highest performing and most available platform for running the Oracle Database. Exadata is a modern architecture featuring scale-out industry-standard database servers, scale-out intelligent storage servers, and an extremely high speed InfiniBand internal fabric that connects all servers and storage. Unique software algorithms in Exadata implement database intelligence in storage, PCI based flash, and InfiniBand networking to deliver higher performance and high capacity at lower costs than other platforms. Exadata runs all types of database workloads including Online Transaction Processing (OLTP), Data Warehousing (DW) and consolidation of mixed workloads. Simple and fast to implement, the Exadata Database Machine powers and protects your most important databases and is the ideal foundation for a consolidated database cloud.

One of the key enablers of this is the Exadata Smart Flash Cache hardware and the intelligent Oracle Exadata Storage Server Software that drives it. The Exadata Smart Flash Cache feature of the Exadata Storage Server Software intelligently caches database objects in flash memory, replacing slow, mechanical I/O operations to disk with very rapid flash memory operations. The Exadata Storage Server Software also provides the Exadata Smart Flash Logging feature to speed database log I/O. Exadata Smart Flash Cache is one of the essential technologies, that enables a full rack Oracle Exadata Database Machine to process up to 2,660,000 random I/O operations per second (IOPS), and scan data at a rate of up to 100 GB/second. This performance scales linearly by adding more Exadata racks to the configuration.

Oracle Database 11g: The First Flash Optimized Database

Oracle's Exadata Smart Flash Cache features are unique. Exadata Flash storage is not a disk replacement – Exadata software intelligence determines how and when to use the Flash storage, and how best to incorporate Flash into the database as part of a coordinated data caching strategy. Scale out Exadata storage enables the benefits of flash performance to be delivered all the way to the application. Traditional storage arrays have many internal and network bottlenecks that prevent realizing the benefits of flash. Flash can be added to storage arrays, but they cannot deliver much of the potential performance to applications.

A full rack Exadata Database Machine delivers up to 100 GB/second of bandwidth from flash. This is dramatically higher than other solutions, and that is on uncompressed data. Combine this bandwidth with Hybrid Columnar Compression and offload processing and the effective bandwidth is much higher. The aggregate bandwidth of disks in a traditional storage array exceeds what the array controllers can handle. Most storage arrays are bottlenecked at small single digit GB/second of bandwidth. Adding flash to these systems will not improve their bandwidth. It will only increase the severity of the bottleneck. Traditional storage arrays cannot deliver the benefits of flash to a data warehouse, or OLTP batch and reporting workloads, where superior sequential bandwidth is required to meet the business requirement.

Traditional storage arrays are good at servicing random IOPS issued by simple disks. They can keep up with the random IOPS because there are relatively few of them in most OLTP systems. When flash, which is orders of magnitude faster than regular disks, is added in to the system, traditional storage arrays become bottlenecked. For example, a high-end storage subsystem delivers about 120,000 IOPS. An Exadata Database Machine delivers up to 2,660,000 IOPS at the database level. The Exadata software can simultaneously scan from Flash and disk to maximize bandwidth. That means 2,660,000 IOPS in flash, through the storage servers, across the network, and into the database servers. Storage arrays are already bottlenecked, and adding flash just exposes the bottlenecks even more.

Oracle is using flash PCIe cards in Exadata – not flash disks. While it is easy to add flash disks into an existing storage subsystem without having to change anything else, the potential of the technology is not realized. Disk controllers and directors were never designed to keep up with the performance that flash disks enable. By using flash PCIe cards, Oracle's solution does not have a slow disk controller limiting flash performance. Exadata storage delivers close to 1.33 GB/sec of throughput from each flash card and scales that performance linearly across the 4 cards in every Exadata Storage Server. Traditional storage arrays do not allow flash cards to be added to the system. Their architecture would need to be redesigned to avoid the disk controller limitations.

Oracle has implemented a smart flash cache directly in the Oracle Exadata Storage Server. The Exadata Smart Flash Cache holds frequently accessed data in very fast flash storage while most of the data is kept in very cost effective disk storage. This happens automatically without the user

having to take any action. The Oracle Flash Cache is smart because it knows when to avoid trying to cache data that will never be reused or will not fit in the cache. The Oracle Database and Exadata storage allow the user to provide directives at the database table, index and segment level to ensure that specific data is retained in flash. Tables can be moved in and out of flash with a simple command, without the need to move the table to different tablespaces, files or LUNs like you would have to do with traditional storage with flash disks.

With the Write Back Flash Cache feature, the Exadata Smart Flash Cache also caches database block writes. Write caching eliminates disk bottlenecks in large scale OLTP and batch workloads. The flash write capacity of a single full rack Exadata Database Machine X4-2 exceeds 1,960,000 8K write I/Os per second. The Exadata Write cache is transparent, persistent, and fully redundant. The I/O performance of the Exadata Smart Flash Cache is comparable to dozens of enterprise disk arrays with thousands of disk drives.

Exadata's Smart Flash Cache is designed to deliver flash-level IO rates, throughput, and response times for data that is many times larger than the physical flash capacity in the machine by automatically moving active data that is experiencing heavy IO activity into flash, while leaving cold data that sees infrequent IO activity on disk. It is common for hit rates in the Exadata Smart Flash Cache to be over 90%, or even 98% in real-world database workloads even though flash capacity is more than 10 times smaller than disk capacity. Such high flash cache hit rates mean that Exadata Smart Flash Cache provides an effective flash capacity that is often 10 times larger than the physical flash cache. For example, a full rack Exadata Database Machine X4-2 often has an effective flash capacity of 440 TB.

The Exadata Smart Flash Cache is also used to reduce the latency of log write I/O eliminating performance bottlenecks that might occur due to database logging. The time to commit user transactions is very sensitive to the latency of log writes. Also, many performance critical database algorithms such as space management and index splits are also very sensitive to log write latency. Today Exadata storage speeds up log writes using the battery backed DRAM cache in the disk controller. Writes to the disk controller cache are normally very fast, but they can become slower during periods of high disk IO. Smart Flash Logging takes advantage of the flash memory in Exadata storage to speed up log writes.

Flash memory has very good average write latency, but it has occasional slow outliers that can be one or two orders of magnitude slower than the average. The idea of the Exadata Smart Logging is to perform redo writes simultaneously to both flash memory and the disk controller cache, and complete the write when the first of the two completes. This literally gives Exadata the best of both worlds. The Smart Flash Logging both improves user transaction response time, and increases overall database throughput for IO intensive workloads by accelerating performance critical database algorithms.

Exadata includes Hybrid Columnar Compression and enables much higher levels of compression than was ever possible before. This has a large number of benefits including greatly reducing the cost of storing large amounts of data, and increasing the speed at which data can be scanned. It

also is very synergistic with Exadata's flash technology. By compressing data by a factor of ten times or more, Oracle fits ten times more data into flash. This means that flash becomes much more effective than the same flash capacity in any other product.

Oracle flash technology is tightly integrated into the Exadata end-to-end architecture—it is fully integrated into the database storage hierarchy – DRAM, flash and disks. It is not a bolt-on accelerator that the user has to manually manage and optimize. The Exadata Smart Flash technology delivers the bandwidth and IOPS required of the most demanding applications without placing a burden on the database and system administrators.

Exadata uses only enterprise grade flash that is designed by the flash manufacturer to have high endurance. Exadata is designed for mission critical workloads and therefore does not use consumer grade flash that can potentially experience performance degradations or fail unexpectedly after a few years of usage. The enterprise grade flash chips used in Exadata X4 have an expected endurance of 10 years or more for typical database workloads.

The automatic data tiering between RAM, flash and disk implemented in Exadata provides tremendous advantages over other flash-based solutions. When third-party flash cards or flash disks are used directly in database servers, the data placed in flash is only available on that server since local flash cannot be shared between servers. This precludes the use of RAC and limits the database deployment to the size of a single server handicapping performance, scalability, availability, and consolidation of databases. Any component failure, like a flash card, in a single server can lead to a loss of database access. Local flash lacks the intelligent flash caching and Hybrid Columnar Compression provided in Exadata and is much more complex to administer.

Real world experience has shown that server local flash cards and flash disks can become crippled without completely failing leading to database hangs, poor performance, or even corruptions. Flash products have been seen to intermittently hang, exhibit periodic poor performance, or lose data during power cycles, and these failures often do not trigger errors or alerts that would cause the flash product to be taken offline. Worse, these issues can cause hangs inside the Operating System causing full node hangs or crashes. Exadata software automatically detects and bypasses poorly performing or crippled flash. When an unusual condition is detected, Exadata will automatically route I/O operations to alternate storage servers.

Many storage vendors have recognized that the architecture of their traditional storage arrays inherently bottleneck the performance of flash and therefore have developed new flash-only arrays. These flash-only arrays deliver higher performance than traditional arrays but give up the cost advantages of smart tiering of data between disk and flash. Therefore the overall size of data that benefits from flash is limited to the size of expensive flash. Exadata smart flash caching often provides flash level performance for data that is 10 times larger than physical flash since it automatically keeps active data that is experiencing heavy IO activity in flash while leaving cold data that sees infrequent IO activity on low-cost disk. Database and Flash Cache Compression further extend the capacity of Exadata flash. Third party flash arrays will also not benefit from Exadata Hybrid Columnar Compression.

Exadata not only delivers much more capacity than flash-only arrays, it also delivers better performance. Flash-only storage arrays cannot match the throughput of Exadata's integrated and optimized architecture with full InfiniBand based scale-out, fast PCI flash, offload of data intensive operations to storage, and algorithms that are specifically optimized for database

Exadata Smart Flash Cache Hardware

Exadata systems use the latest PCI flash technology rather than flash disks. PCI flash greatly accelerates performance by placing flash directly on the high speed PCI bus rather than behind slow disk controllers and directors. Each Exadata Storage Server includes 4 PCI flash cards with a total capacity of 3.2 TB of flash memory. A Full Rack Exadata Database Machine includes 56 PCI flash cards providing 44.8 TB of flash memory. The flash modules used in Exadata X4 have an expected endurance of 10 years or more for typical database data. This solid state storage delivers dramatic performance advantages with Exadata storage. It implements automatic caching of database reads and writes and can do over 1,960,000 8K Flash Writes per second and 2,660,000 8K Flash Reads per second in a full rack X4 database machine.

Sun Flash Accelerator F80 PCIe Card

One of the key enablers of Exadata's extreme performance is the Exadata Smart Flash Cache hardware and the intelligent Oracle Exadata Storage Server Software that drives it. The Exadata Smart Flash Cache feature of the Exadata Storage Server Software intelligently caches database objects in flash memory, replacing slow, mechanical I/O operations to disk with very rapid flash memory operations. The Exadata Storage Server Software also provides the Exadata Smart Flash Logging feature to speed database log I/O. Exadata Smart Flash Cache is one of the essential technologies of the Oracle Exadata Database Machine that enables the processing of up to 2,660,000 random 8K I/O operations per second (IOPS), and the scanning of data within Exadata storage at up to 100 GB/second.

The performance that the Exadata Smart Flash Cache provides at the database level for the various Exadata X4 configurations is shown in the following table.

	Exadata Database Machine X3-8 Full Rack	Exadata Database Machine X4-2 Full Rack	Exadata Database Machine X4-2 Half Rack	Exadata Database Machine X4-2 Quarter Rack	Exadata Database Machine X4-2 Eighth Rack
Raw Flash Data Bandwidth (without data compression)	Up to 100 GB/sec	Up to 100 GB/sec	Up to 50 GB/sec	Up to 21.5 GB/sec	Up to 10.7 GB/sec
Database Flash Read IOPS [1]	Up to 1,500,000	Up to 2,660,000	Up to 1,330,000	Up to 570,000	Up to 285,000
Database Flash Write IOPS [1]	Up to 1,000,000	Up to 1,960,000	Up to 980,000	Up to 420,000	Up to 210,000
Raw Disk Data Bandwidth • **High Capacity Disk** • **High Performance Disk** **(without data compression)**	Up to 20 GB/sec 24 GB/sec	Up to 20 GB/sec 24 GB/sec	Up to 10 GB/sec 12 GB/sec	Up to 4.5 GB/sec 5.2 GB/sec	Up to 2.25 GB/sec 2.6 GB/sec
Database Disk IOPS [1] • **High Capacity Disk** • **High Performance Disk**	Up to 32,000 50,000	Up to 32,000 50,000	Up to 16,000 25,000	Up to 7,000 10,800	Up to 3,500 5,400

Exadata Database Machine X4-2 and X3-8 I/O Performance

Exadata Storage Server Software and the Flash Cache Hardware

There are two key features of the Exadata Storage Server Software that leverage the Exadata Flash hardware and make the Exadata Database Machine such a fast system on which to deploy the Oracle Database. First is Exadata Smart Flash Cache which provides the capability to stage active database objects in flash. Second is the Exadata Smart Flash Logging which speeds the

critical function of database logging. Lastly, the deployment of the Oracle Database requires mission critical resilience and the Exadata Storage Server Software in conjunction with the Oracle Database provides that.

Exadata Smart Flash Cache: Flash for Database Objects

Exadata Smart Flash Cache provides an automated caching mechanism for frequently-accessed data in the Exadata Database Machine. It is a write-back cache which can service extremely large numbers of random reads and writes and improves the responsiveness of OLTP applications deployed on the Database Machine.

Automated Management of the Exadata Smart Flash Cache

The Exadata Smart Flash Cache holds frequently accessed data in very fast flash storage while most of the data is kept in very cost effective disk storage. This happens automatically without the user having to take any action. The Oracle Flash Cache is smart because it knows when to avoid trying to cache data that will never be reused or will not fit in the cache.

When the database sends a read or write request to Exadata Storage Server, it includes additional information in the request about whether the data is likely to be read again and therefore whether it should be cached. Based on the information the database sends, the Exadata Storage Server Software intelligently decides which data is likely to be re-read, and is worth caching, versus data that would just waste cache. Random reads and writes against tables and indexes are likely to have subsequent reads and normally will be cached and have their data delivered from the flash cache. In addition, Exadata Smart Flash Cache is persistent across Exadata Storage Server restarts and will not require any warm up period.

Knowing what not to cache is of great importance to realize the performance potential of the cache. For example, when writing backups or to a mirrored copy of a block, the software avoids caching these blocks. Since these blocks will not be re-read in the near term there is no reason to devote valuable cache space to these objects or blocks. Only the Oracle Database and Exadata Storage Server software have this visibility and understand the nature of all the I/O operations taking place on the system. Having the visibility through the complete I/O stack allows optimized use of the Exadata Smart Flash Cache hardware to store only the most relevant data.

With the Exadata Storage Server Software 11.2.3.3.0 and above, the Exadata Smart Flash Cache software automatically caches objects read, in large table scans, in the flash cache based on how frequently the objects are read. Previously, the default behavior was to bypass the flash cache for such large sequential scans. The new algorithm takes into account the size of the object, the frequency of access of the object, and the frequency of access to data is displaced in the cache by the object, and the type of scan being performed by the database. Depending on the flash cache size, and the other concurrent workloads, all or only part of the table or partition is cached. For objects that are larger than the size of the flash cache, the software automatically caches only a portion of the object and eliminates thrashing the flash cache. This new feature largely eliminates

the need for manually pinning tables in flash cache. In earlier releases, database administrators had to mark an object as KEEP to have it cached in flash cache for large scan workloads.

This feature primarily benefits table scan intensive workloads such as Data Warehouses and Data Marts. Random I/Os such as those performed for Online Transactional Processing (OLTP) continue to be cached in the flash cache the same way as in earlier releases.

On top of the capacity benefits provided by smart caching, the Exadata Storage Server Software 11.2.3.3 and above, introduces the Exadata Smart Flash Cache Compression feature, which dynamically increases the capacity of the flash cache, in X3 and X4 systems, by transparently compressing user data as it is loaded into the flash cache. This allows much more data to be kept in flash memory, and further decreases the need to access data on disk drives. The compression and decompression operations are completely transparent to the application and database. Exadata Smart Flash Compression leverages hardware acceleration to deliver zero performance overhead for compression and decompression, even when running at rates of millions of I/Os per second or 100s of Gigabytes per second.

Flash cache compression benefits vary based on the compressibility of the user data. Tables that are uncompressed will see the largest benefits. Indexes will also typically compress very well. Exadata Smart Flash cache compression will also provide significant flash cache space expansion on top of the benefits already provided by OLTP and Basic table compression. OLTP applications will often see the overall logical size of the flash cache double even if they use OLTP compression. Tables that use Hybrid Columnar Compression or LOB compression will see minimal additional compression since these are already very highly compressed formats. With flash cache compression turned on, a full rack Exadata Database Machine X4-2 provides more than 88 TB of logical flash cache capacity (before database level compression is factored in).

Consolidating multiple databases onto a single Exadata Database Machine is a cost saving solution for customers. With Exadata Storage Server Software 11.2.2.3 and above, the Exadata I/O Resource Manager (IORM) can be used to enable or disable use of flash for the different databases running on the Database Machine. This empowers customers to reserve flash for the most performance critical databases.

The Oracle Database and Exadata storage optionally allow the user to provide directives at the database table, index and segment level to ensure that specific data is retained in flash. Tables can be retained in flash without the need to move the table to different tablespaces, files or LUNs like you would have to do with traditional storage and flash disks.

User Management of the Exadata Smart Flash Cache

There are two techniques provided to manually use and manage the cache. The first enables the pinning of objects in the flash cache. The second supports the creation of logical disks out of the flash for the permanent placement of objects on flash disks.

Pinning Objects in the Flash Cache

Preferential treatment over which database objects are cached is also provided with the Exadata Storage Server Software and Oracle Database. For example, objects can be pinned in the cache and always be cached, or an object can be identified as one which should never be cached. This control is provided by the new storage clause attribute, CELL_FLASH_CACHE, which can be assigned to a database table, index, partition and LOB column.

There are three values to which the CELL_FLASH_CACHE attribute can be set. DEFAULT specifies the cache used for a DEFAULT object is automatically managed as described in the previous section. NONE specifies that the object will never be cached. KEEP specifies the object should be kept in cache, once it is there.

For example, the following command could be used to direct that pages from the table CUSTOMERS remain in Exadata Smart Flash Cache, once they are there:

```
ALTER TABLE customers STORAGE (CELL_FLASH_CACHE KEEP)
```

This storage attribute can also be specified when the table is created.

The Exadata Storage Server will cache data for the CUSTOMERS table and will keep it in flash while other tables that KEEP has not been specified for will be aged out of cache. In the normal case where the CUSTOMERS table is spread across many Exadata Storage Servers, each Exadata cell will cache its part of the table in its own flash. Generally there should be more flash cache available than the objects KEEP is specified for. This leads to the table being completely cached over time.

If KEEP has been specified for an object, and it is accessed via an offloaded Smart Scan, the object is kept in and scanned from cache. Another advantage of the flash cache is that when an object that is kept in the cache is scanned, the Exadata software will simultaneously read the data from both flash and disk to get a higher aggregate scan rate than is possible from either source independently. With Exadata Storage Server Software version 11.2.3.3.0 and above, the Exadata software larger eliminates the need for marking objects with the KEEP attribute, as it automatically caches data read during a Smart Scan.

Creating Flash Disks Out of the Flash Cache

When an Exadata cell is installed, by default, the flash is assigned to be used as Flash Cache (or for Smart Logging) and user data is automatically cached using the default caching behavior. Optionally, a portion of the cache can be reserved and used as logical flash disks. These flash disks are treated like any Exadata cell disk in the Exadata cell except they actually reside and are stored as non-volatile disks in the cache. For each Exadata cell the space reserved for flash disks is allocated across sixteen (16) cell disks – 4 cell disks per flash card. Grid disks (the logical disks that reside on physical cell disks) are created on these flash-based cell disks and the grid disks are assigned to an Automatic Storage Management (ASM) diskgroup. The best practice would be to reserve the same amount of flash on each Exadata cell for flash disks and have the ASM diskgroup spread evenly across the Exadata cells in the configuration just as you would do for

regular Exadata grid disks. This will evenly distribute the flash I/O load across the Exadata cells and flash.

These high-performance logical flash disks can be used to store frequently accessed data. To use them requires advance planning to ensure adequate space is reserved for the tablespaces stored on them. In addition, backup of the data on the flash disks must be done in case media recovery is required, just as it would be done for data stored on conventional disks. In most situations using this functionality offers limited benefit for the overhead of setting up and managing the flash this way. But this option might be useful for highly write intensive workloads where the disk write rate exceeds what the disks can process.

Exadata Smart Flash Logging: Flash for Database Logging

In an OLTP workload, fast response time for database log writes is crucial. The Database Administrator (DBA) configures redo log groups and mirrored log files for availability but slow disk performance can have a negative impact on redo log write wait time and system performance – log writes wait for the write to the slowest disk to complete. Additionally, disk drives themselves can experience occasional "hiccups" in performance. These spikes can have a huge impact on database performance. In addition, flash technology can have similar performance hiccups due to erase cycles or wear leveling. An approach that deals with these issues, and others, has been provided in the Exadata Database Machine.

Smart Flash Log differentiates Exadata for OLTP workloads, and is another example of how Exadata optimizes database performance by engineering improvements across the software and hardware stack. The Smart Flash Logging requires Exadata Storage Software version 11.2.2.4 or later, and Oracle Database version 11.2.0.2 with Bundle Patch 11, Oracle Database version 11.2.0.3 with Bundle Patch 1, or a later version.

Smart Flash Logging leverages the flash hardware in the Exadata Database Machine. Smart Flash Logging is much more than simply placing the redo log in flash; duplexed and mirrored log files exist for important reasons. Just adding a flash-based log to the redo log group will not solve the problems mentioned above, namely that the database will end up waiting for log writes to the slowest device whether that device is a slow disk or slow flash. In addition, customers have to set up their redo log groups and mirrored log files to provide maximum availability. Exadata Smart Flash Logging has been designed such that no changes are required to this configuration to reap the benefits of low latency log writes. In essence this feature is transparent to the database software configuration and just as importantly database recovery.

Smart Flash Logging works as follows. When receiving a redo log write request, Exadata will do parallel writes to the on-disk redo logs as well as a small amount of space reserved in the flash hardware. When either of these writes has successfully completed the database will be immediately notified of completion. If the disk drives hosting the logs experience slow response times, then the Exadata Smart Flash Cache will provide a faster log write response time. Conversely, if the Exadata Smart Flash Cache is temporarily experiencing slow response times

(e.g., due to wear leveling algorithms), then the disk drive will provide a faster response time. This algorithm will significantly smooth out redo write response times and provide overall better database performance.

The Exadata Smart Flash Cache is not used as a permanent store for redo data – it is just a temporary store for the purpose of providing fast redo write response time. The Exadata Smart Flash Cache is a cache for storing redo data until this data is safely written to disk. The Exadata Storage Server comes with a substantial amount of flash storage. A small amount is allocated for database logging and the remainder will be used for caching user data. The best practices and configuration of redo log sizing, duplexing and mirroring do not change when using Exadata Smart Flash Logging. Smart Flash Logging handles all crash and recovery scenarios without requiring any additional or special administrator intervention beyond what would normally be needed for recovery of the database from redo logs. From an end user perspective, the system behaves in a completely transparent manner and the user need not be aware that flash is being used as a temporary store for redo. The only behavioral difference will be consistently low latencies for redo log writes.

By default, 512 MB of the Exadata flash is allocated to Smart Flash Logging. Relative to the 3.2 TB of flash in each Exadata cell this is an insignificant investment for a huge performance benefit. This default allocation will be sufficient for most situations. Statistics are maintained to indicate the number and frequency of redo writes serviced by flash and those that could not be serviced, due to, for example, insufficient flash space being allocated for Smart Flash Logging. For a database with a high redo generation rate, or when many databases are consolidated on to one Exadata Database Machine, the size of the flash allocated to Smart Flash Logging may need to be enlarged. In addition, for consolidated deployments, the Exadata I/O Resource Manager (IORM) has been enhanced to enable or disable Smart Flash Logging for the different databases running on the Database Machine, reserving flash for the most performance critical databases.

Mission Critical Availability of the Exadata Smart Flash Cache

The hardware used for the Exadata Smart Flash Cache is very reliable but all hardware is subject to failure. Spreading the flash cache across 4 PCIe cards mitigates some of this risk. If there is a failure of one of the flash cards, the Exadata Storage Server Software automatically detects the loss of the card and takes the failed portion of the flash cache offline. During this process the Exadata cell continues to operate and serve data from the remaining cache. So, while performance might be reduced because there is less flash cache available to service I/O requests, the system keeps running without interruption or data loss. This allows replacement of the failed flash card to be deferred until a convenient time when the Exadata cell can be taken offline and flash card replaced. After the card is replaced the Exadata Storage Server Software automatically detects the presence of the new card and automatically starts using the additional flash cache. If there were "dirty" blocks in the failed flash card, which were not yet written to disk, then the Exadata Storage Server software in conjunction with ASM will automatically retrieve the mirrored copies from the other storage cells to recover the latest copy of the data.

If logical flash disks have been placed in the flash and one of the flash PCIe cards fails, again the impact of the failure is minimized by the Exadata Storage Server Software and ASM. If a flash card failure occurs, the 4 flash cell disks on the failed flash card are automatically taken offline and I/O to those disks are serviced from the mirrored extents stored in flash on other Exadata cells. Eventually an ASM re-balance would occur to re-silver the data across the unaffected flash disks. Once the faulty card is replaced, the flash disks will automatically be added back into the ASM diskgroup and a re-balance will be performed reestablishing the normal configuration.

Conclusion

The Exadata Smart Flash Cache is the power behind the OLTP functionality of the Exadata Database Machine. It delivers unprecedented IOPS for the most demanding database applications and can more than double the scan rate for data in warehouse or reporting applications, as well as providing special support for the critical database logging function. By knowing what data to cache and how to automatically manage the cache, the Oracle Database, with the Exadata Smart Flash Cache, is the first and only flash enabled database.

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

Exadata Smart Flash Cache and the Oracle Exadata Database Machine
December 2013
Author: Mahesh Subramaniam
Contributing Authors: Caroline Johnston, Juan Loaiza, Tim Shetler, Kesavan Srinivasan, Kodi Umamageswaran

Oracle Corporation
World Headquarters
500 Oracle Parkway
Redwood Shores, CA 94065
U.S.A.

Worldwide Inquiries:
Phone: +1.650.506.7000
Fax: +1.650.506.7200

oracle.com

Hardware and Software, Engineered to Work Together