White Paper

EMC Tiered Storage for Microsoft SQL Server 2008 Enabled by EMC CLARiiON CX4 and Enterprise Flash Drives

A Detailed Review

EMC Information Infrastructure Solutions

Abstract

This white paper demonstrates the efficiency and cost savings realized in a Microsoft SQL Server 2008 enterprise-class environment by utilizing the Enterprise Flash Drive technology capabilities of the EMC® CLARiiON® CX4 Series storage system.

October 2010

Part number: H7071.1

Table of Contents

Executive summary

Business case

The role of the database administrator covers a variety of tasks such as administration, database maintenance, manageability, availability, and security, and often now includes responsibility for SAN management. One of the most difficult and time-consuming tasks—and one of the biggest challenges for a Microsoft SQL database administrator—is performance tuning.

Poor performance can be the result of a combination of factors including poor database design, non-optimal Transact-SQL (T-SQL) coding, insufficient memory, database contention, and operating system constraints. As businesses follow best practices from Microsoft and their server vendors, they also expect their storage vendors to provide them with recommendations to ensure optimized performance on their specific platforms.

The focus of this white paper is to demonstrate potential solutions for some of the problems that can result from an inefficient storage design that does not meet the workload requirements of the SQL Server environment. Such problems can be the result of frequent disk sorts, full table scans, missing indexes, row chaining, and data fragmentation. In addition, there are situations where there might not be enough disk input/output operations per second (IOPS) provided to meet the demands of the SQL database.

To remedy this situation, the database administrator would have to add a large number of drives and employ a technique of **"**short-stroking" the drives that leads to an imbalance between IOPS' requirements and storage allocation. As this situation recurs over time, the eventual outcome is a great deal of wasted, unusable capacity, adding a significant environment cost to the solution.

In 2009, EMC introduced support for Enterprise Flash Drives (EFDs), which provided a better solution for the IOPS problem. EMC's EFDs provide many more IOPS than a single Fibre Channel (FC) drive, at a much lower power-consumption rate than a traditional hard drive. Therefore, a small number of EFDs can handle the IOPS workload of dozens of traditional FC drives.

EMC provides support for the deployment of the latest EFD technology drives, combined with FC and SATA drives in the same system. This enables customers to provide an optimal structure for the application data layout, where each tier of storage matches the I/O demands of the application data it hosts.

The solution described in this white paper shows how EMC tested five EFDs to provide a similar performance to 60 FC disks, utilizing a simulated online transaction processing (OLTP) high read-write ratio workload. This white paper presents the before and after test results and provides customers with some considerations for deploying their Microsoft SQL 2008 databases with EFDs.

This white paper also illustrates the benefits of implementing manual storage tiering in an enterprise-class SQL Server 2008 environment. It demonstrates how the introduction of EFDs has freed administrators from having to use large numbers of drives and short-stroke them to achieve higher-aggregate I/O. Traditional hard disk drives (HDDs) are associated with the mechanical delays of head seek and rotational latency. EFDs, however, have no moving parts and therefore no seek or rotational delays, which dramatically improves their ability to sustain a higher number of IOPS with very low overall response times.

Product solution

The EMC® CLARiiON® CX4 array supports the introduction of one of the industry's first EFD technologies, enabling organizations to facilitate the movement of critical, frequently accessed data to a maximum-performance storage tier, and less frequently accessed data to a lower-cost, higher-capacity storage tier.

This solution comprises the following elements:

- Tiering within the array by using different drive technologies:
  - High-performance EFDs
  - Traditional FC drives delivering low response times
  - High-capacity, low-cost SATA drives
- EMC Navisphere® Manager, a web-based management interface that also provides a simple way to configure LUN migration between storage tiers.
- EMC Navisphere Analyzer, which provides details of the performance for each application by monitoring performance before and after manual storage tiering implementation.

Key results

When employed as part of a tiering strategy, EFDs clearly show an excellent return on investment (ROI) by reducing the total cost of ownership (TCO) through their ability to sustain highly demanding data to a smaller number of physical drives, leading to:

- Increased application I/O performance
- Elimination of I/O bottleneck at the LUN level
- Reduced database disk latency
- Reduced number of spinning disks
  - Reduced power consumption
  - Reduced cooling requirements
- Elimination of the need to isolate disks for random workloads

Introduction

Introduction

This white paper illustrates the benefits of tiered storage in an active, enterprise-class, SQL Server OLTP-type environment. To demonstrate this, a performance baseline was determined on a 75,000-user (750 GB) TPC-E-type database, hosted on the FC storage tier. Prior to any customization of the environment, the hot database tables were identified and table partitioning was employed to isolate them to their own file groups.

As part of the implementation of a manual storage tiering policy, partitioned-table data files were created to hold the hot data table identified during the baseline testing.

Next, performance was baselined after the implementation of partition tabling to identify candidate files for promotion to EFD, which formed the top storage tier.

Lastly, performance was again measured against the baseline performance results that were previously determined.

This white paper includes the following sections:

Purpose

The purpose of this white paper is to:

- Determine the optimal FC disk configuration for the given OLTP workload.
- Demonstrate CLARiiON CX4 EFD scalability (by comparing EFD and FC performance).
- Demonstrate the performance benefits of tiered storage.

Scope

The scope of this white paper is to:

- Present an overview of the concepts and technologies in the solution.
- Document the baseline performance testing of an OLTP, TPC-E-type database running on a Microsoft SQL Server 2008 SP1 cluster, utilizing both FC and EFD storage tiers.

- Present the test results and consequent business benefits of the solution.

This white paper does not document the following:

- Supply and build of the physical environment
- Installation or patching of Microsoft Server 2008 R2
- Installation, patching, or Microsoft failover cluster configuration of Microsoft SQL Server 2008 SP1

The information in this paper is not intended to replace existing, detailed product implementation guides or best practices.

Audience

This white paper is intended for:

- Field personnel who are tasked with deploying EMC CLARiiON CX4 as the storage platform
- Customers, including IT planners, storage architects, and SQL database administrators
- EMC staff and partners, for guidance and the development of proposals

It is assumed that the reader is familiar with:

- Microsoft SQL Server 2008 and partition tables in a SQL Server environment
- EMC CLARiiON storage
- EMC Navisphere Manager (LUN migration technology)
- EMC Navisphere Analyzer

Terminology

This section defines terms used in this document.

Term	Definition
Bandwidth	The amount of data a storage system can process over time, which is measured in megabytes per second.
Disk Transfers/sec	Disk Transfers/sec is the rate of read and write operations on the disk.
Enterprise Flash Drives (EFD)	Also known as solid state drives (SSD), EFDs contain no moving parts, which removes the storage latency associated with traditional magnetic disk drives.
Logical unit number (LUN)	A unique identifier used to identify logical storage objects in a storage system.

Partitioned tables	Partitioning is a method of splitting up a large data set into smaller, more manageable chunks. All management operations on very large tables can be performed at a more granular level when the table is partitioned.
RAID 10	RAID method that provides data integrity by mirroring data onto another disk. This RAID type provides the greatest assurance of data integrity at the greatest cost in disk space.
RAID 5	RAID method where data is striped across disks in large stripes. Parity information is stored so data can be reconstructed, if necessary. One disk can fail without data loss. Performance is good for reads, but slower for writes.
Response time	The interval of time between submitting an I/O request and receiving a response.
Short-stroked drives	This is a technique where data is laid out on partially populated disks to reduce the spindle head movement and to provide higher IOPS at a very low latency.
SP	Storage processor on a CLARiiON storage system. On a CLARiiON storage system, a circuit board with memory modules and control logic that manages the storage-system I/O between the host's FC adapters and the disk modules.
Throughput	The number of individual I/Os the storage system can process over time, which is measured in I/Os per second.
Write penalty	The write penalty is inherent in RAID data protection techniques, which require multiple disk I/O requests for each application write request. For RAID 10, two I/O requests are required for each application request; for RAID 5, four I/O requests are required for each application request.

Overview of components

Introduction

This section identifies and briefly describes the components deployed in the solution environment. The components used are:

- EMC CLARiiON CX4-960
- EMC Navisphere Manager
- EMC Navisphere Analyzer
- EMC PowerPath®
- Microsoft Windows 2008 R2
- Microsoft SQL Server 2008

EMC CLARiiON CX4-960

The CLARiiON CX4-960 system is a high-end, enterprise storage array comprising a system bay that includes storage processor enclosures (SPEs), storage processors (SPs), disk array enclosures (DAEs) and separate storage bays that can scale up to 960 disk drives. The CX4-960 arrays support multiple drive technologies, including EFDs, FC drives, and serial advanced technology attachment (SATA) drives, and the full range of redundant array of independent disks (RAID) types.

EMC Navisphere Manager

Navisphere Manager is a centralized storage-system management tool for configuring and managing the CLARiiON CX4-960 storage systems. It provides basic functionality such as the discovery of storage systems, status and configuration information display, event management, and storage configuration and allocation. Navisphere Manager is a web-based user interface that enables the secure management of storage systems, locally on the same LAN or remotely over the Internet.

EMC Navisphere Analyzer

Navisphere Analyzer is a web-based tool, using a common browser that allows an administrator to graphically examine the performance characteristics of the logical and physical entities that make up a CX4-960 storage system. Analyzer supports immediate (realtime) data display, as well as the display of previously logged data. As a result, it is possible to do immediate comparisons, long-term trend analysis, and offsite performance troubleshooting and analysis.

EMC PowerPath

PowerPath is server-resident software that enhances performance and application availability. PowerPath works with the storage system to intelligently manage I/O paths, and supports multiple paths to a logical device. PowerPath provides automatic failover in the event of a hardware failure by automatically detecting the path failure and redirecting I/O to another path. PowerPath also provides dynamic multipath load balancing. It distributes I/O requests to a logical device across all available paths, thus improving I/O performance and reducing management time and downtime by eliminating the need to configure paths statically across logical devices.

Configuration

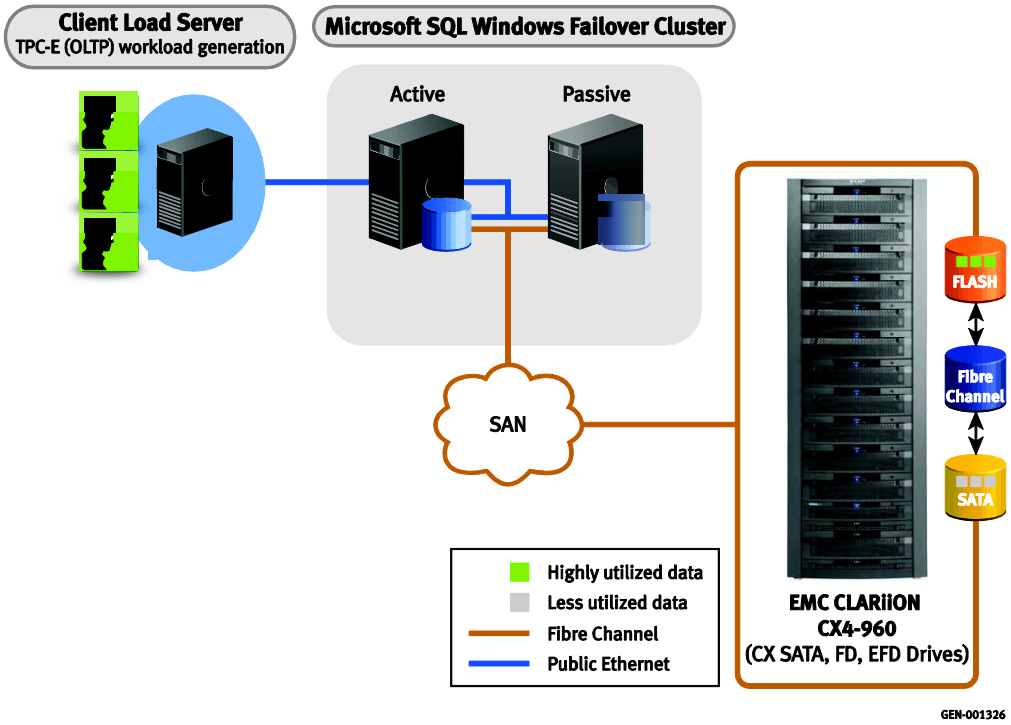
Overview of the physical components

This tiered storage design consists of the following physical components:

- A two-node active/passive Microsoft Windows failover cluster
- Storage connectivity was provided by an 8 Gb/s FC switch
- Network connectivity was provided by a 1 gigabit Ethernet (GbE) network switch
- A CLARiiON CX4-960 was attached to the servers through four front-end FC ports
- A CLARiiON CX4-960 provided access to the EFD, FC, and SATA storage

Physical environment

The following diagram illustrates the overall physical architecture of the environment.



Test configuration

This section describes the test configuration used.

SQL Server 2008 test configuration

The SQL Server 2008 test configuration is based on the following profile:

- Number of SQL users supported is 75,000
- Simulated user workload with 75 concurrent users
- User data: 750 GB

SQL Server test application

The SQL load test tool used in this environment simulates an OLTP workload. It comprises a set of transactional operations designed to exercise system functionalities in a manner representative of a complex OLTP-type application environment.

OLTP workloads

The OLTP application used to generate the user load in this test environment is based on the TPC Benchmark-E (TPC-E) standard. TPC-E testing includes a set of transactions that represent the processing activities. The database schema, data, population, transactions and implementation rules have been designed to be broadly representative of modern OLTP systems. TPC-E application models the activity of a brokerage firm that:

- Manages customer accounts
- Executes customer trade orders
- Tracks customer activity with financial markets

Hardware resources

The hardware used to validate the solution is listed in the following table.

Equipment	Quantity	Configuration
EMC CLARiiON CX4-960	1	105 x 450 GB 15k FC 10 x 400 GB EFD 15 x 750 GB 7.2k SATA
Fibre switch	2	FC 8 Gb switch 48-port
Network switch	1	Ethernet 1 Gb switch 48-port
Domain controller	1	2 core/4 GB RAM
TPC-E load servers	1	TPC-E load servers
SQL Server - active	1	24-core/128 GB RAM
SQL Server - passive	1	8-core/24 GB RAM

Software resources

The software used to validate the solution is listed in the following table.

Software	Version
EMC PowerPath	5.3
EMC FLARE®	4.29.000.5.003
EMC Navisphere Manager	6.29.0.6.34
EMC Navisphere Analyzer	6.29.0.6.34
Microsoft Windows Server 2008	Enterprise Edition R2
Microsoft SQL Server 2008	Enterprise SP1

Storage design considerations

Introduction to storage design

This section details storage design considerations for:

- Microsoft SQL Server 2008
- Microsoft Windows Server 2008 R2
- Storage system
- EMC CLARiiON CX4 Series
- RAID type

Microsoft SQL Server 2008 considerations

Like any other database, SQL Server manages data using database files and transaction log files. The database files are frequently accessed with random reads and writes. Transaction log files, on the other hand, generally consist of sequential write operations, with occasional reads of recently written transaction records. The SQL Server **tempDB** database is read/write-intensive and is used for storing temporary tables, temporary stored procedures, and sub-queries, and for sorting aggregate operations.

Best practices

The following best practices should be considered:

- Plan for storage performance, not for capacity
- Place SQL Server transaction log and database files on physically separate RAID groups
- Place transaction log files on RAID 10 volumes
- Place database files on a volume with an appropriate RAID type. Choose RAID 10 volumes for OLTP-type applications with highly active database files
- Use Microsoft failover cluster (MSFC) for high availability
- Make the SQL Server account part of an Active Directory domain
- Enable SQL Server to keep pages in memory
- Set your database file sizes and autogrow increments appropriately
- Plan your database filegroups based on your workload and requirements
- Consider table and index partitioning
- Plan the location, layout, and size of your tempdb
- Use defaults for processors and memory

For more information, refer to *EMC CLARiiON Database Storage Solutions: Microsoft SQL Server 2008 in Virtualized Environments - Best Practices Planning*.

Microsoft Windows Server 2008 considerations

This section details recommendations for the configuration of Windows Server 2008 for use with a Microsoft SQL Server instance.

- Only use hardware that is approved by Microsoft
- Use the latest verified network interface card (NIC) driver
- Use a dedicated VLAN for cluster heartbeat connectivity

Storage system considerations

Storage performance issues are most often the result of configuration issues with underlying storage devices. Storage performance is a vast topic that depends on workload, hardware, vendor, RAID level, cache size, stripe size, and so on. Many workloads are very sensitive to the latency of I/O operations. It is therefore important to have the storage devices configured correctly.

This section details recommendations for the configuration of your EMC storage system for use with Microsoft SQL Server 2008.

- Plan storage layouts for performance, not for capacity

  The most common error made while planning storage for Microsoft SQL Server is designing for capacity and not for performance, or I/Os per second (IOPS). To properly plan the disk layout there must be an estimate as to the number of IOPS that need to be supported on a sustained basis, the peak IOPS, and the duration of the peak.

  Many customers gather data while the application is running, then use a 90th percentile to determine the level that should be planned for. There are three primary variables used for determining the number of spindles for database storage:

  - IOPS (or sometimes MB/s, if a serial workload)
  - RAID level—when planning for performance, striped RAID 10 will require fewer spindles than RAID 5 for almost all read/write workloads. They are approximately equal in a read-only workload.
  - Latency goals
- Use Diskpart to align your LUNs for best performance
- Set the NTFS allocation unit to 64 KB
- Plan storage operations for minimal disruption to capacity and performance
- Be aware of the capacity and performance limitations of FLARE vault drives
- Bind at least one hot spare for every 30 drives
- Connect sufficient FC ports into the storage network
- Use multiple paths for high availability and improved performance

EMC CLARiiON CX4 Series considerations

The importance of understanding and planning the storage system cannot be underestimated since the storage system affects SQL Server performance more than any other factor. Performance improvement depends heavily on hardware characteristics such as disk layout, I/O characteristics, disk allocation units, and storage array caching practices. The following general guidelines should suffice for most midrange database designs.

- Increase CLARiiON prefetch settings to improve restore performance
- Balance the LUNs' design across SPs
- Name the LUNs for quick identification purposes
- Use the latest verified HBA driver
- Use EMC PowerPath on physical servers

Best practices for EFD

Due to the extremely high performance of EFDs, cache settings for EFDs do not follow traditional guidelines.

EFDs are extremely fast. When the read cache is enabled for the LUNs residing on the EFDs, the read-cache lookup for each read request can add significantly higher overhead when compared to FC drives in an application profile that is not expected to get many read cache hits at any rate. Therefore, it can be faster to disable the read cache to directly read the block from the EFD.

In scenarios where the CLARiiON CX4 is being shared by several applications and, especially, where it is deployed with slower SATA drives, the write cache may become fully saturated, placing the EFDs in a force flush situation, which adds latency. In these situations, it is better to write the block directly to the EFDs rather than to the write cache of the storage system.

However, in this solution environment, it was found that no significant overhead was added with the read cache enabled, so a decision was made to leave the read cache enabled for the database volumes.

In testing, it was found that the write cache did not become fully saturated and also there was a performance benefit to leaving the write cache enabled for the database volumes. From testing, it was seen that SQL transactions logs gain little advantage from being placed on EFDs, due to the sequential write nature of the transaction logs.

Best practices for FC

- Set the storage array cache block size to the default 8 KB
- Allocate sufficient read and write cache on the storage array
- Enable write caching for all volumes
- Enable read and write caching for SQL Server transaction log volumes
- Enable read and write caching for SQL Server transaction database volumes

Disk type considerations

- Use EFDs for high-performance and latency-sensitive applications
- Use FC drives for high-performance SQL Server applications
- Use SATA drives for low cost and large capacity

RAID type considerations

This section details recommendations for the RAID type.

- Use more disk drives in a RAID group to improve performance. Build RAID groups consisting of smaller disk drives rather than a few large drives. This enables more disks to work in parallel to read and write data, thus improving performance.
- Select a RAID type based on its performance, fault tolerance, cost, and application workload.
- It is recommended to distribute RAID groups across back-end buses as evenly as possible, in round-robin order, to define each RAID group to be on a separate bus. This is known as horizontal provisioning. RAID 10 groups used for highest availability can benefit from being distributed over two back-end buses.

RAID 10 gives excellent random read/write performance. RAID 10 is ideal for OLTP environments with lots of small random read/writes. This RAID type also provides good fault tolerance since it can survive the failure of up to half of the disks, provided one disk in each mirror image pair survives. Use RAID 10 if more than 30 percent of the I/O is small random writes and if budget permits.

RAID 5 gives excellent read performance, especially large sequential I/O. But it provides lower random write performance. It also delivers lower fault tolerance than RAID 10 since it can tolerate only one drive failure per RAID 5 LUN. Also, in the event of a drive failure, the time for the storage system to rebuild the content of the failed drive is longer than RAID 10. During the content rebuild of the failed drive, overall performance to the data will be impacted, though there will not be a loss of data access. RAID 5 costs less for the same storage capacity compared to RAID 10.

Storage design validation

Introduction to storage design validation

The section details the storage design validation as follows:

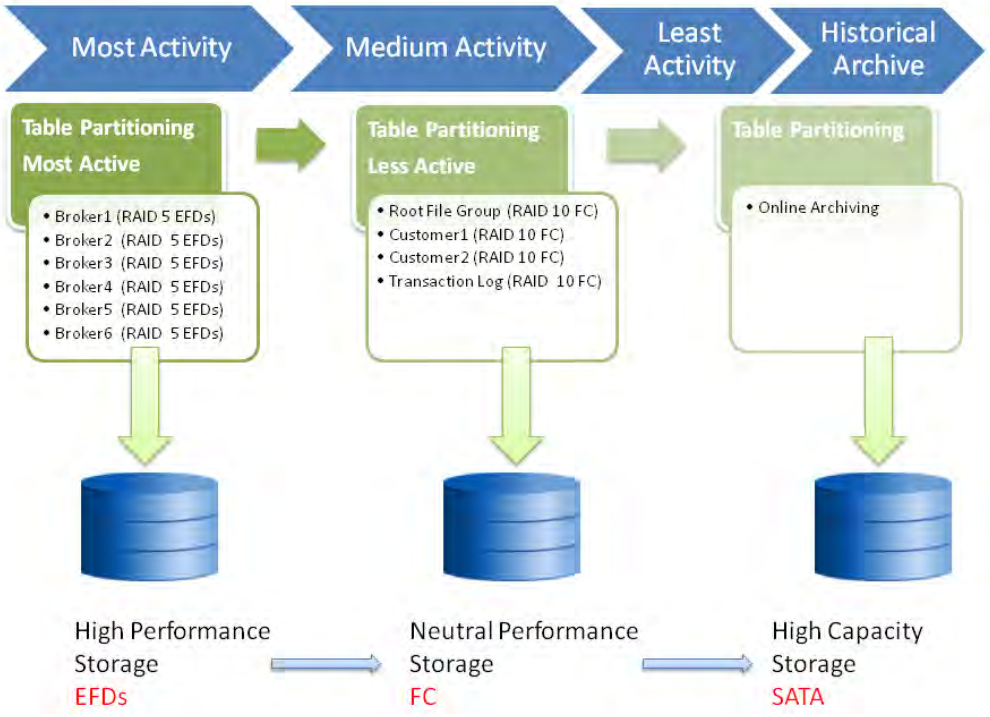
- Database performance requirements
- Production database spindle calculation
- Production log spindle calculation
- Utilities to monitor and identify design for tiered storage

Database performance requirements

Prior to designing a baseline storage layout, the first step is to determine the SQL application workload. In a production environment, the best way to obtain this information is to analyze the current application and database performance. Tools such as PerfMon and SQL Profiler can be used for trace purposes to investigate and analyze:

- Total host IOPS
- Percent read
- Percent write
- I/O sizes
- Capacity requirements for database files

The following image illustrates database performance.



Production database spindle calculation

Storage spindle design refers to the process of determining how many disk spindles in the CLARiiON CX4-960 are required to satisfy the host IOPS requirements of the application. Once the database performance requirements are collected, they can be plugged into the formula, which is used to calculate the number of spindles required to support the total host IOPS.

A total of 450 GB FC 15k rpm drives was used in the CLARiiON CX4-960, which are the typical FC drives supplied as part of a base configuration. Guidelines suggest that these can yield 180 IOPS per spindle when used in a random read-write environment, such as that found in an OLTP-type application. RAID 10 RAID groups were used for production databases to ensure faster rebuild times, better protection for data, and better read-write performance, as would normally be used in customer environments.

The spindle requirements were based on the following application workload.

- Total IOPS: 10,000
- Read-write ratio: 9/1
- Write penalty: 2 (RAID 10)
- Disk IOPS: 180

The formula was calculated as follows:

```
Number of Spindles = (Total IOPS x % read) + WP (Total IOPS x
% write)
_____________________________________________________________

Disk IOPS
```

That is:

```
(10000 x .10) +2 (10000 x.90)
-----------------------------------
180
```

60 spindles required for databases

In the test environment, there were six broker files in their own RAID 10 4+4 group, two customer files that shared a RAID 10 4+4 group, and one root file in a RAID 10 2+2 group. The root file holds the generic tables with the hot tables partitioned across the six broker files and the two customer files.

This number of disks satisfies the IOPS requirements and it must also satisfy the capacity requirements. Sixty mirrored disks results in 30 disks of effective storage space. Thirty 450 GB (around 402 GB of usable space) disks provide approximately 12 TB of storage space, which is more than enough space to handle the database files. Each RAID group has one LUN bound and presented to the SQL server as the storage for each of the database files.

Production log spindle calculation

The number of spindles required for log volumes is calculated slightly differently. Log transactions are mostly a series of sequential writes with minimal read activities. Given this, the RAID type should always be RAID 10, and the primary consideration for planning is the number of write IOPS the transaction log will need to manage, as this is crucial to SQL performance, with response times being a critical factor. For this white paper, the transaction log was planned to handle 50 GB of updates across an eight-hour window.

The number of write IOPS was configured using a projected update of 50 GB for this test configuration.

Where ...	Is ...
50 GB = 53,687,091,200 bytes and 8 KB = 8,192 bytes	53,687,091,200 ÷ 8192 = 6,553,600 8 KB updates to the transaction log during a normal eight-hour period

This can be converted to IOPS by dividing the number of updates that need to be supported across an eight-hour period (28,800 seconds) as follows:

```
6553600 ÷ 28800 = 227.55 IOPS
```

The requirement is 227.55 IOPS over an eight-hour period. With each spindle supporting 180 IOPS, four spindles in a RAID 10 2+2 group are required to support the transaction logs. It is also necessary to plan for peaks where a RAID 10 2+2 could sustain up to 350 IOPS with 100 percent writes.

Two RAID 10 2+2 groups were used for the four temp database data files, with two files on each group, and the temp database log placed on one of these groups also. Also, additional storage is required for operating system, SQL Server binary, and system databases. The OS and cluster require storage for the mount points and quorum volumes. These resources have very low IOPS and storage requirements, so the LUNs for those resources are provided by another RAID 5 4+1 group. Global hot spares were also accounted for in this configuration.

Utilities to monitor and identify design for tiered storage

The following utilities were used to monitor and identify the design for tiered storage:

- Microsoft Windows Performance Monitor
- EMC Navisphere Analyzer

Microsoft Windows Performance Monitor

The Windows OS comes with the Performance Monitor tool that collects a large number of system metrics to help identify performance problems. This section describes the statistics that can help identify a system with I/O subsystems that are causing performance problems.

- **Physical disks: Avg. Disk sec/Read**

  The **Avg. Disk sec/Read** performance counter indicates the average time, in seconds, of a read from the disk. The average value of the **Avg. Disk sec/Read** performance counter for the OLTP data file should be under 10 milliseconds. The

maximum value of the **Avg. Disk sec/Read** performance counter for the OLTP data file should not exceed 20 milliseconds.

For the SQL log file, the performance counter should be less than 5 milliseconds. Ideally, the counter should be less than 10 milliseconds for the log file: five is good, one is excellent.

- **Physical disks: Avg. Disk sec/Write**

  The **Avg. Disk sec/Write** performance counter indicates the average time, in seconds, of a write from the disk. The average value of the **Avg. Disk sec/Write** performance counter for the OLTP data file should be under 10 milliseconds. The maximum value of the **Avg. Disk sec/Write** performance counter for the OLTP data file should not exceed 20 milliseconds. For the SQL log file, the performance counter should be less than 5 milliseconds. Ideally, the counter should be less than 1 millisecond for the log file.

- **Physical disks: Average Disk Queue Length**

  Average Disk Queue Length is an estimate of requests on the physical or logical disk that are either in service or waiting for service.

- **Processor: % Processor time**

  Measuring the CPU activity of your SQL Server is a key way to identify potential CPU and I/O bottlenecks. The **Process Object: % Processor Time** counter is available for each CPU (instance) in the server, and measures the utilization of each individual CPU. Alternatively, the **Total instance** can be monitored, which provides the total overall CPU utilization for the SQL Server and is a good overall indicator of how busy the SQL Server is. For the SQL OTLP environment, the SQL Server CPU should be kept less than 80 percent.

Navisphere Analyzer

Navisphere Analyzer software is a host-based performance analysis tool that is intended to be used as a microscope to examine specific data in as much detail as necessary, to determine the cause behind a bottleneck and/or a performance issue. Once the cause has been isolated, Navisphere Analyzer is of further assistance in helping to assess whether fine tuning parameters of the array will solve the problem or whether hardware components, such as cache memory or disks, need to be added.

- **Disk Utilization**

  Utilization of all components should be 70 percent or less.

- **Total Throughput**

  The average number of host requests that are passed through the LUN per second, including both read and write requests.

- **Response Time (ms)**

  The average time, in milliseconds, that it takes for one request to pass through the disk, including any waiting time.

Physical drive configuration

The following table details the logical drives used in the test environment.

Drive type	Number of drives	Specifications	RAID type
EFDs	6	400 GB	RAID 5
FC	76	450 GB, 15k rpm	RAID 10

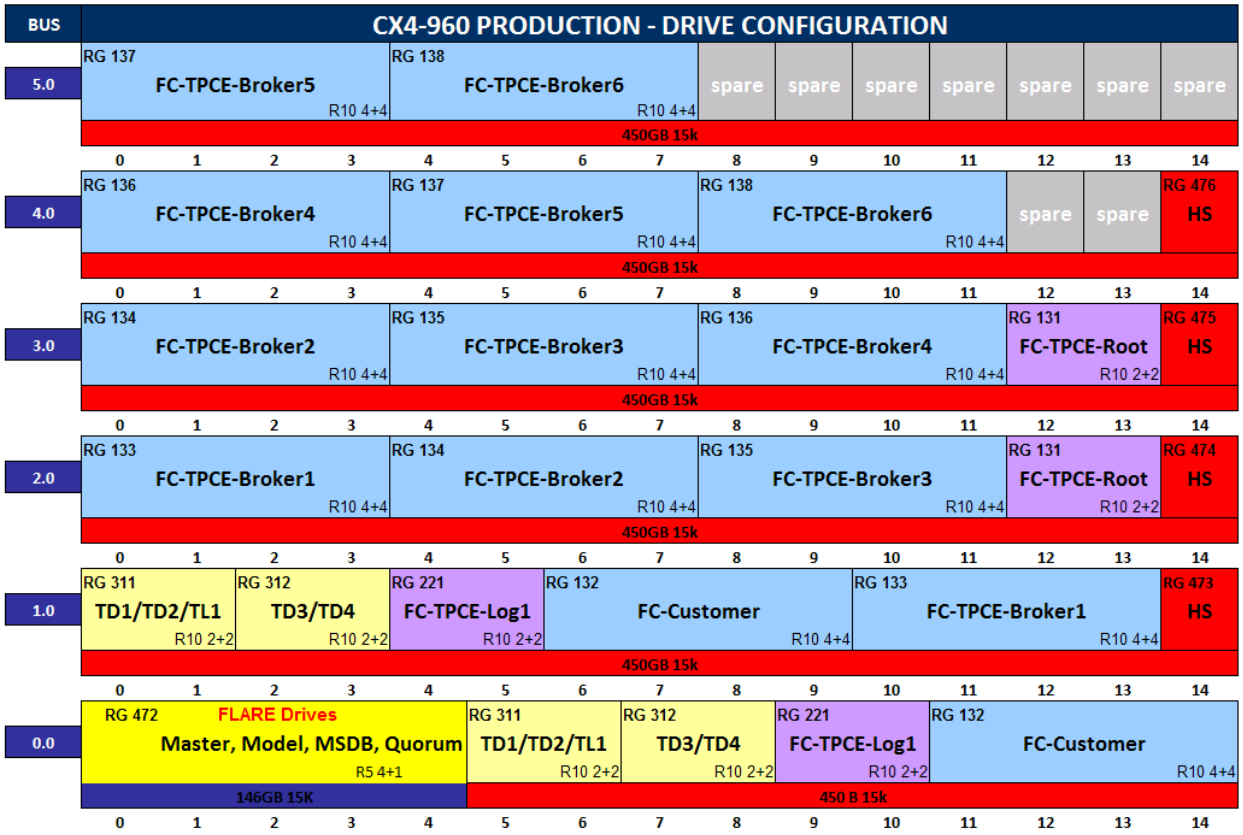
Logical drive configuration

The following table details the logical drives used in the test environment.

File group	Number of LUNs	Size of LUNs	Number of drives	Drive type	RAID type	RAID group
TPCE-Root	1	270 GB	4 (2+2)	FC	RAID 10	131
TPCE-Customer 1& 2	2	175 GB	8 (4+4)	FC	RAID 10	132
TPCE-Broker 1	1	175 GB	8 (4+4)	FC	RAID 10	133
TPCE-Broker 2	1	175 GB	8 (4+4)	FC	RAID 10	134
TPCE-Broker 3	1	175 GB	8 (4+4)	FC	RAID 10	135
TPCE-Broker 4	1	175 GB	8 (4+4)	FC	RAID 10	136
TPCE-Broker 5	1	175 GB	8 (4+4)	FC	RAID 10	137
TPCE-Broker 6	1	175 GB	8 (4+4)	FC	RAID 10	138
TPCE-Logs	1	800 GB	4 (2+2)	FC	RAID 10	221
TempDB 1-2	2	100 GB	4 (2+2)	FC	RAID 10	311
TempDB 3-4 & log file	3	100 GB	4 (2+2))	FC	RAID 10	312
Master/Model/MSDB	1	5 GB	5 (4+1)	FC	RAID 5	472
Quorum	1	5 GB				
Mount Point	1	1 GB				
MSDTC	1	5 GB				

Storage system layout A—initial FC layout diagram

The following image details the production layout of this environment. Each block represents a physical spindle. In the layout process, an effort was made to separate and spread the load as evenly as possible between the spindles, storage processors, and back-end bus.



Tiered storage design considerations

Introduction to tiered storage design

CLARiiON CX4 provides integrated support for both high-performance EFDs and cost-effective, lower-performance SATA drives within the same storage system. This type of integrated support from high-performance EFDs, midrange FC drives, and cost-effective SATA drives enables data to be stored and accessed in a uniform manner from the database software.

It is very common to move frequently accessed data to faster storage and move the less frequently accessed data to a cost-effective, low-performance storage option such as SATA. A regular way of deploying a single database over multiple tiers is by file type. For example, the archive log and backup images for the database can use SATA drives while the transaction logs and data file can use FC. It is then possible to place latency-sensitive data files on the EFDs.

The LUN migration feature is included in the EMC Navisphere Manager software. This is an online process that enables you to move the data in one LUN to another LUN. In the test environment, this technology could be employed to move SQL data files stored on FC to EFD and EFD to FC. During a LUN migration, Navisphere Manager copies the data from the source LUN to a destination LUN. After migration is complete, the destination LUN assumes the identity (World Wide Name and other identities) of the source LUN. Refer to *EMC CLARiiON Best Practices for Performance and Availability: Release 29 Firmware Update* to determine migration speeds as there will be some disruption in performance during the LUN migration.

However, to achieve optimum utilization of drive resources when the database is large, it may be better to place only the data that is accessed most frequently or has the most demanding latency requirements on the EFDs. For the SQL OTLP database, optimum utilization can be achieved by using table partitioning.

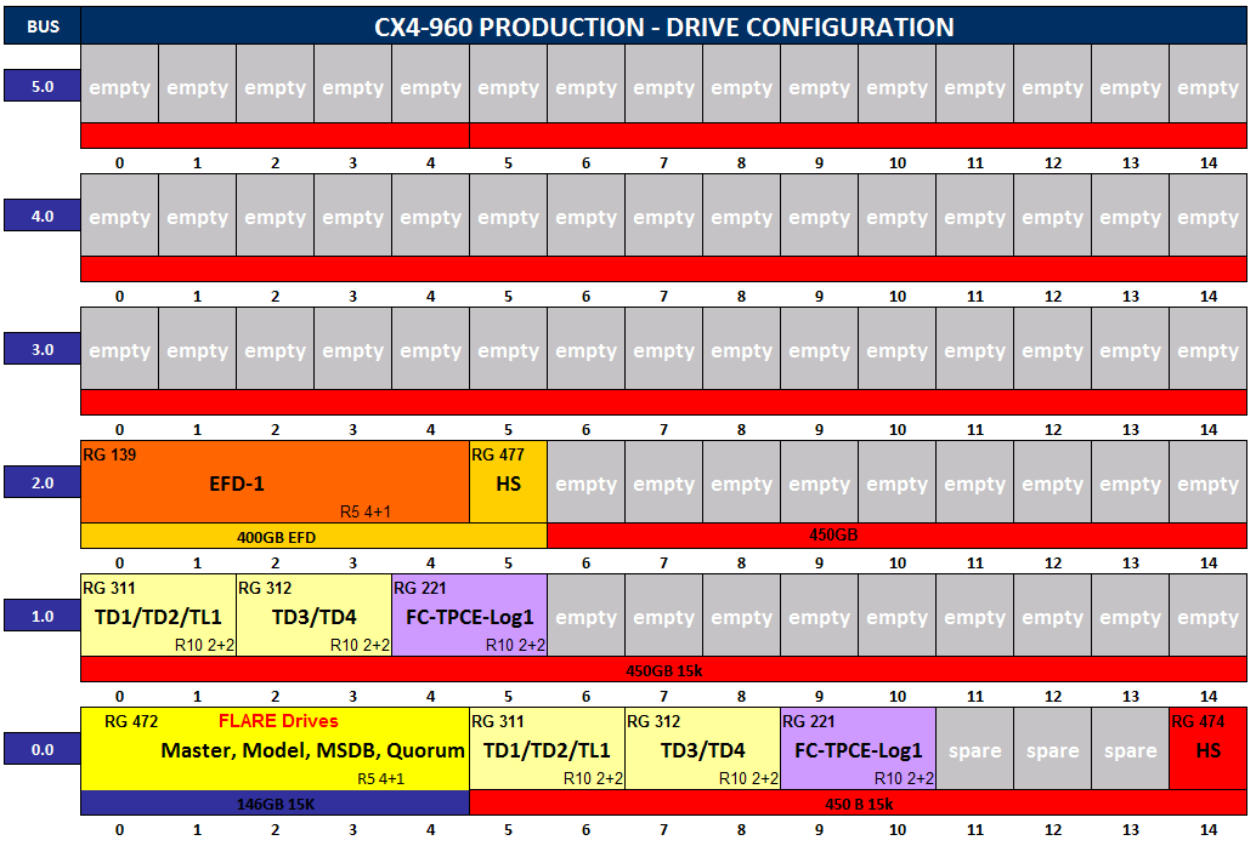
Disk type design considerations for tiered storage

Consider the following disk type design for tiered storage:

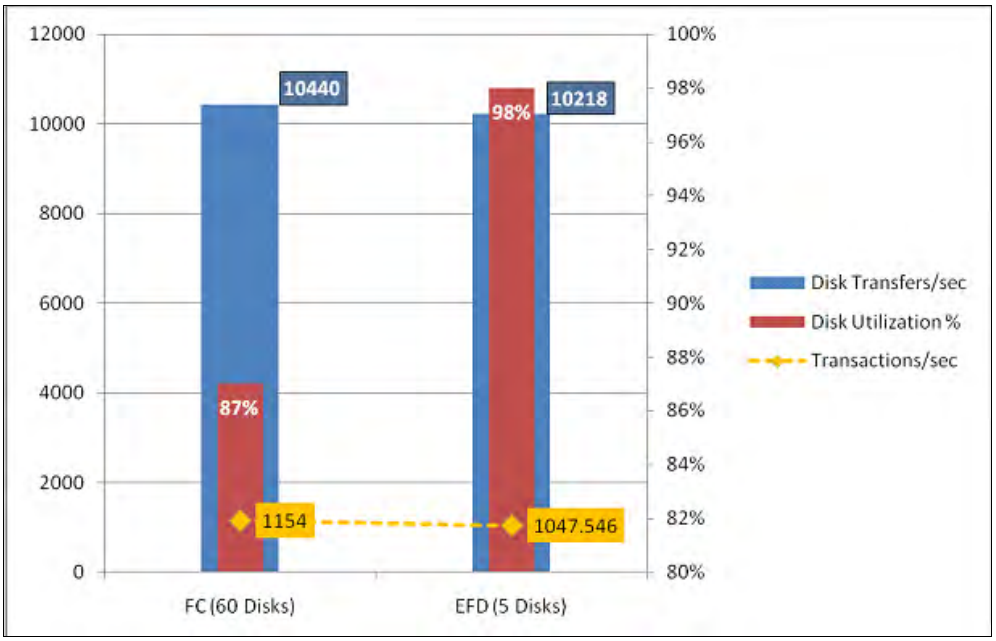
- Use EFDs for read-write heavy (hot) tables and latency-sensitive data for OLTP applications.
- Use FC drives with a large number of random I/O reads and writes for SQL Server applications.
- Use SATA drives for low cost and high capacity. Because SATA drives have lower I/O performance, they may not be suitable for a high-performance, OLTP-type application. SATA is ideal for storing aged data and backup images of database files for SQL Server applications.

Storage system layout B—all files moved to EFD

The following image details the production layout of this test environment with all files moved to EFD.

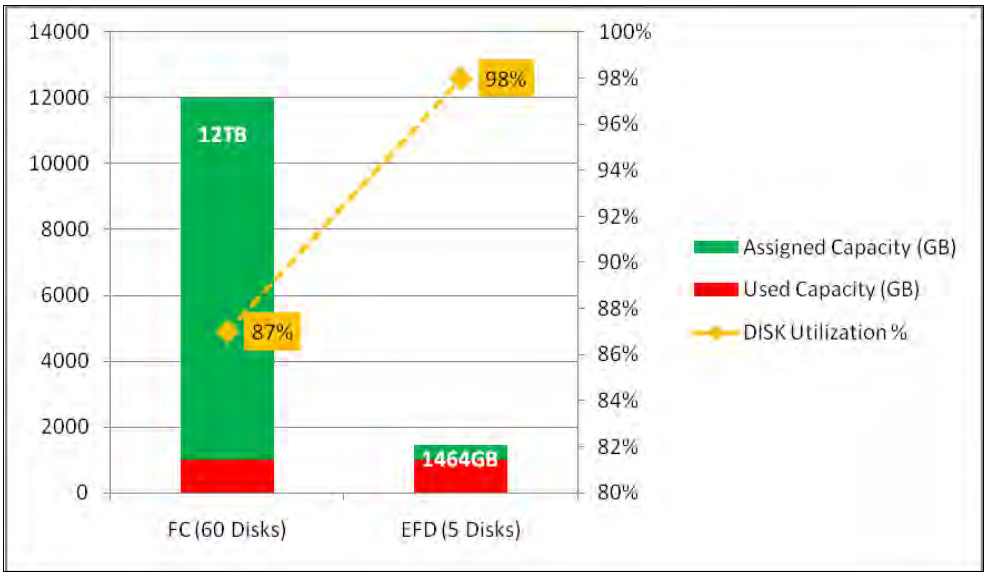


After moving all the data files, except the **tpce** database log file, to the EFD, performance was tested to compare results between FC and EFD as shown in the following chart.



It can be seen that when moving to EFD the LUN utilization increases. This is due to the significantly reduced number of spindles being assigned to handle the given workload of approximately 10,000 IOPS, produced during a normal workload run. There is minimal difference in the total IOPS and transactions/sec that a RAID 5, 4+1 EFD RAID group can handle, compared to the complex layout and high spindle count required by traditional FC disks.

The following chart graphically illustrates how five EFD disks providing 1.464 TB of usable storage is a better fit for a 750 GB database with ample room for growth.

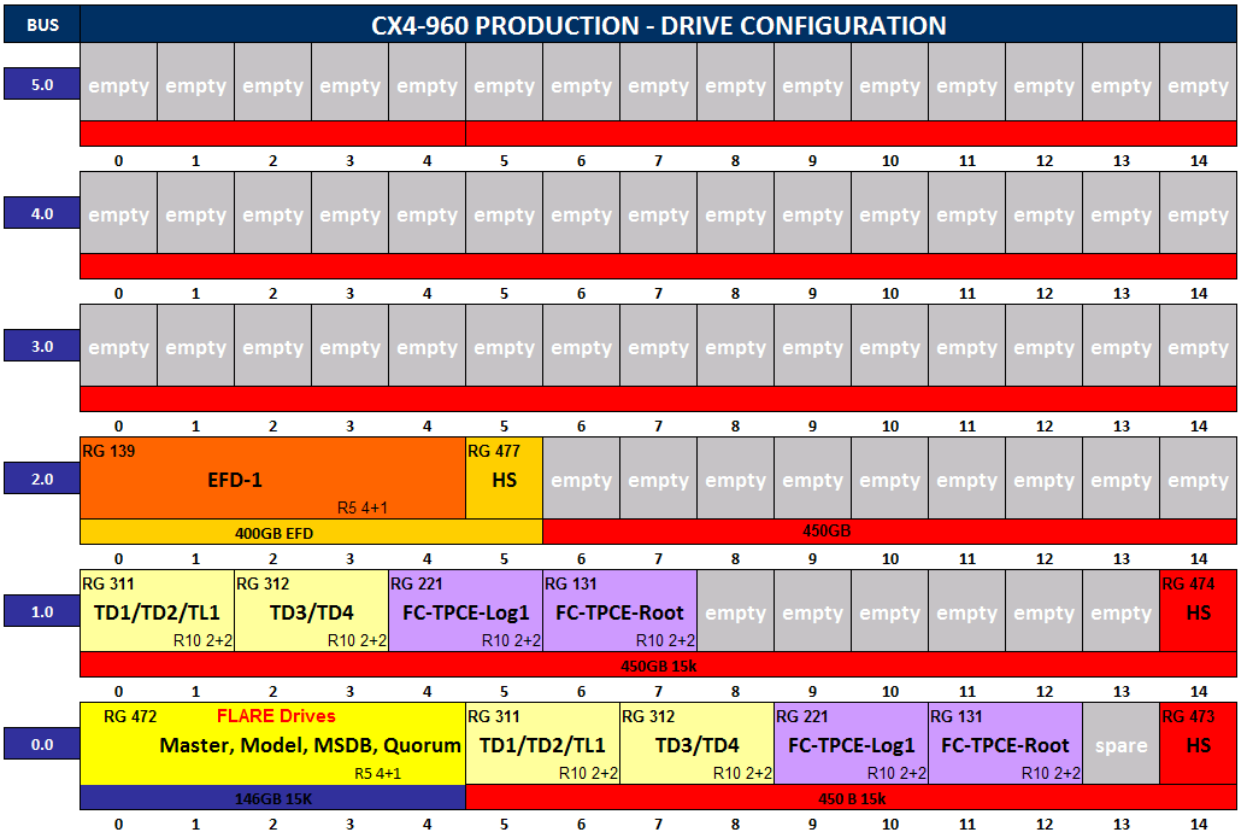


The equivalent 60 FC disks provide a usable capacity of 12 TB of storage. This leaves around 10.5 TB of unusable storage, since use of this storage risks a direct impact on application performance. This clearly shows how the proper use of EFD technology enables the CLARiiON CX4 Series to be used to its optimal performance level and provides greater ROI and reduced TCO.

The EFDs perform to a stated 98 percent disk utilization performance compared to 87 percent for FC, which again shows how the five EFDs are an excellent fit for an intensive OLTP-type workload. And given the nature of EFDs, they may continue to perform even under greater workloads.

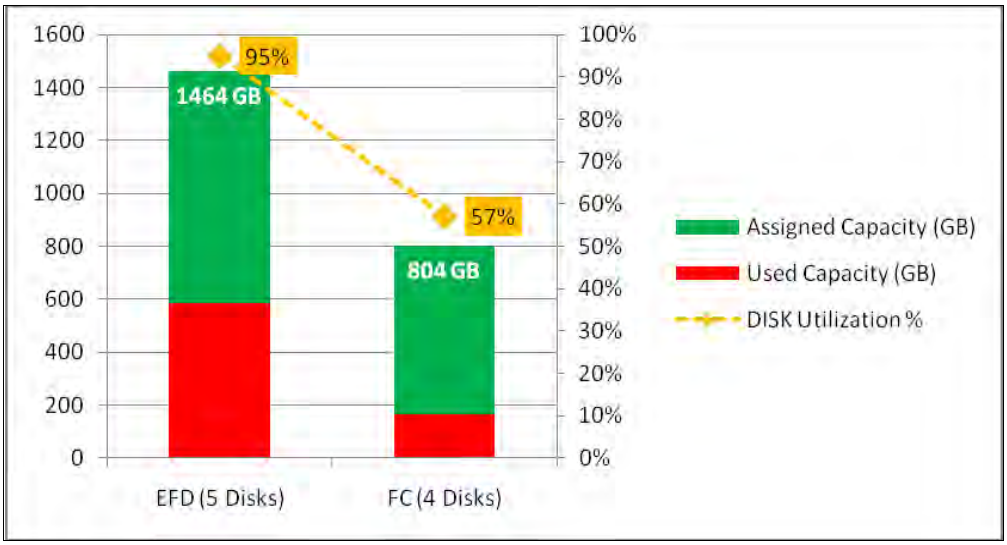
Storage system layout C—files with low IOP requirement moved to FC

The following image details the production layout of this environment with only partitioned table data files on EFD and all other files moved to FC.



The layout shows the less IOP-intensive LUNs (that is, TPCE-Root and TPCE-Customer 1& 2) moved back down to FC. This combination has little impact in terms of the overall IOPS and footprint on the CLARiiON CX4, with only four FC drives now being identified as adequate to host the data files.

The following chart shows the additional 200 GB of usable space now made available on the EFD through the process of manual tiering from the EFD to the FC drives.



Conclusion

Summary

This white paper highlights the efficiency and cost savings realized in a Microsoft SQL Server 2008 enterprise class environment by introducing and integrating the EFD technology capabilities of the CLARiiON CX4 Series storage system.

The functionality, testing, and observations documented in this white paper demonstrate and prove how:

- EFD technology can increase database throughput while reducing disk latency on the CLARiiON CX4 Series, while at the same time reducing the overall spindle count required to meet demanding application requirements. It provides:
  - Reduced TCO by placing the right data on the right tier at the right time
  - Optimization of storage resource utilization, performance, and availability
  - Reduced storage costs by using fewer drives, and lowering energy consumption and the storage footprint
- EMC's suite of storage management, monitoring, and reporting applications provides simplified management of a Microsoft SQL Server 2008 enterprise environment on the CLARiiON CX4 Series:
  - EMC Navisphere Analyzer collects statistics that assist the process of designing storage layout and monitoring CLARiiON system performance.
  - The use of table partitioning in a Microsoft SQL Server 2008 enterprise class environment provided a mechanism for partitioning and controlling where data resides in the storage tiers.

Key points

EFD technology greater reduces the impact of running a Microsoft SQL Server 2008 enterprise class environment on the CX4 Series storage array, utilizing storage resources to their maximum and ensuring that its footprint has minimal impact on the system. The table below summarizes the key points that this solution addresses.

Key Point	Solution objective
Reduced spindle count on the CLARiiON CX4 Series	EFD technology can greatly reduce the number of spinning disks required to meet the application demands on the CLARiiON CX4 Series storage array. EFD technology also greatly reduced the overall footprint required to meet application demand. This white paper demonstrates how five EFDs in a RAID 5 4+1 configuration proved sufficient to match 60 RAID 10 FC drives.
Reduced storage complexity	EFD technology negates the need to employ complex short-stroking strategies in the design layout that potentially can create unusable disk space that can impact performance.

Reduced energy consumption on the CLARiiON CX4 Series	With fewer spinning disks and no moving parts, EFDs provide reduced TOC and faster ROI through reduced running costs.

EFD read/write cache on and off

Since the introduction of support for EFD technology in the CLARiiON FLARE 28/29 code, the option of enabling the EFD read and write cache on a per-LUN basis can help to improve performance in some dedicated environments, where storage is not shared across many applications. This is highly specific to individual environments, however, and dependent on the type of applications workload being applied.

During testing, it was observed that with the read and write cache both enabled, no negative impact was found. In fact, with the read and write cache enabled on the database LUNs, a small performance improvement for database reads and writes was seen. During periods of routine database maintenance tasks, such as re-indexing, this could be advantageous.

It is recommended, therefore, that deep analysis and investigation be done before committing changes to this setting on the EFD LUNs in production environments.

Next steps

To learn more about this and other solutions contact an EMC representative or visit: www.emc.com.

References

White papers

For additional information, see the white papers listed below.

- *Implementing EMC CLARiiON CX4 with Enterprise Flash Drives for Microsoft SQL Server 2008 Databases – Applied Technology*
- *EMC Tiered Storage for Microsoft SQL Server 2008 Enabled by EMC Symmetrix VMAX with FAST—A Detailed Review*
- *EMC CLARiiON Database Storage Solutions: Microsoft SQL Server 2008 in Virtualized Environments - Best Practices Planning*
- *EMC CLARiiON Best Practices for Performance and Availability: Release 29 Firmware Update*

Other documentation

For additional information, see the documents listed below.

- *Technology Primer: Flash Drives for Business-Critical Storage*
- *Performance Analysis for TPC-E Test using Microsoft SQL Table and Index Partitioning on CLARiiON CX4-960*