EMC Symmetrix DMX-4 Enterprise Flash Drives with IBM DB2 on z/OS

Applied Technology

Abstract

This white paper examines the use of DB2 on z/OS with enterprise Flash drives including use cases, performance characteristics, and general guidelines for database file placement.

January 2009
# Table of Contents

**Executive summary** ................................................................. 4  
**Introduction** ............................................................................. 4  
  Audience ..................................................................................... 4  
**Technology overview** ................................................................. 4  
  Symmetrix DMX-4 ............................................................................... 4  
  Enterprise Flash drives ................................................................... 5  
**DB2 on z/OS systems on enterprise Flash drives** ...................... 7  
  DB2 on z/OS environments best suited for enterprise Flash drives ......................................................... 7  
  DB2 on z/OS Flash drive recommendations ................................................. 7  
  DB2 log and table space placement ................................................................. 7  
  RAID protection ............................................................................. 7  
**DB2 and Flash drive testing and analysis** ................................. 7  
  Testing configuration ....................................................................... 8  
  z/OS volume setup ......................................................................... 8  
  DB2 transaction workload ................................................................. 8  
  DMX-4 configuration ........................................................................ 9  
  Testing results .............................................................................. 9  
**Selected placement of DB2 table spaces** ................................. 12  
**Use cases** .................................................................................. 12  
**Conclusion** ............................................................................... 14  
**References** ................................................................................ 14
Executive summary
EMC has enhanced the latest release of Symmetrix® Enginuity™ version 5773 to integrate enterprise-class Flash drives into the Symmetrix DMX-4 storage array. With Flash drives, EMC has created a new ultra-performance storage tier, “Tier 0,” that removes previous performance limitations imposed by magnetic disk drives.

For years, the most demanding enterprise applications have been limited by the performance of magnetic disk media. Tier 1 performance in storage arrays has been constrained by the physical limitations of hard disk drives. Enterprise Flash drives for Tier 0 deliver unprecedented performance and response times, which are benefits well suited for demanding DB2 on z/OS configurations.

Enterprise Flash drives dramatically increase performance for latency-sensitive applications like DB2 on z/OS. Enterprise Flash drives, also known as solid state drives (SSD), contain no moving parts, which removes much of the storage latency delay associated with traditional magnetic disk drives. A Symmetrix DMX-4 with enterprise Flash drives can deliver single-millisecond application response times and up to 30 times more I/O operations per second (IOPS) than traditional Fibre Channel hard disk drives (HDD). Additionally, because there are no mechanical components, Flash drives consume significantly less energy than hard disk drives. When replacing a larger number of HDDs with a lesser number of enterprise Flash drives, energy consumption can be reduced by up to 98 percent for a given IOPS workload.

The high-performance characteristics of enterprise Flash drives eliminate the need for organizations to purchase large numbers of traditional hard disk drives, while only utilizing a small portion of their capacity to satisfy the IOPS requirements of DB2 on z/OS systems. The practice of underutilizing a hard disk drive for increased performance is commonly referred to as short-stroking. Enterprise Flash drives can increase DB2 on z/OS database performance and eliminate the need to short-stroke drives, thus keeping storage footprint and power consumption to a minimum and reducing total cost of ownership (TCO).

Introduction
This white paper addresses the advantages of deploying enterprise Flash drives in a DB2 on z/OS database environment, including the performance, power savings, and consolidation benefits. Use cases and best practices will be examined, and performance examples comparing Flash drives to HDD are also discussed.

Audience
This white paper is intended for DB2 z/OS database administrators and storage architects who want to understand how enterprise Flash drives can be deployed with DB2 on z/OS. Data center managers who are concerned with increasing power demands and lack of adequate power feeds and decreasing space availability may also benefit from reading this.

Technology overview
Symmetrix DMX-4
The Symmetrix DMX-4 system is the next generation in the Symmetrix DMX™ series and extends EMC’s leadership in the high-end enterprise storage market. The DMX-4 delivers support for the latest generation of disk drive technologies including Flash drives and 4 Gb/s FC drives for high performance, along with SATA II drives for high capacity requirements. The Symmetrix DMX-4 is the first and only high-end storage system to support all of these new generations of disk drive technologies, allowing for unprecedented flexibility in performance and tiered storage functionality.
Enterprise Flash drives

The enterprise-class EMC Flash drives are constructed with nonvolatile semiconductor NAND Flash memory. They are packaged in a standard 3.5-inch disk drive form factor and include a Fibre Channel interface used in existing Symmetrix DMX-4 drive array enclosures. The enterprise Flash drives used in the DMX-4 differ significantly from the solid state technology used in consumer electronics, particularly in their performance and reliability characteristics. To satisfy enterprise-level drive requirements NAND single level cell Flash technology was made more robust with dynamic wear leveling functions, which ensures all cells in the Flash memory are used evenly, to minimize the risk of “wear-out” common to less advanced Flash devices. Additionally, enterprise Flash drives include bad block remapping and multi-bit error correction. Because of these reliability enhancements and the fact that the drive has no moving parts the life expectancy of the Flash drive exceeds that of traditional hard disk drives.

The lack of moving parts also offers additional benefits. As previously discussed, enterprise Flash drives can deliver unprecedented performance, especially well suited for applications that require consistently low read/write response times. Additionally enterprise Flash drives consume considerably less power when compared to traditional rotating media.

The following figures offer an introductory overview to the IOPS and power benefits offered by enterprise Flash drives. Figure 3 shows the relative IOPS rates for various drive technologies that can be expected to deliver suitable response times for DB2 on z/OS. It can be seen that while short-stroking conventional hard disk drives can deliver slightly more IOPS over a fully stroked drive, the performance attained is only a fraction of what is possible with enterprise Flash drive technology.
Figure 3. IOPS for different drive types

Figure 4 outlines, in addition to the performance benefits, the power savings that can be achieved when comparing enterprise Flash drives against traditional hard disk drives. The replacement of a larger number of short-stroked HDDs with a reduced number of enterprise Flash drives can deliver a significant reduction in power and cooling requirements. The savings in power will clearly depend on the exact configuration, and to what extent the traditional HDDs have been short-stroked, as well as the total number of HDDs being replaced. For more detailed information, consult with an EMC representative for specific power usage derived from the EMC Power Calculator.

Figure 4. Flash drive vs. HDD power consumption
DB2 on z/OS systems on enterprise Flash drives

DB2 on z/OS environments best suited for enterprise Flash drives

DB2 on z/OS administrators, in certain environments, have found it necessary to short-stroke traditional hard disk drives to achieve the required number of IOPS at the desired latencies. The downside of short-stroking hard disk drives is the total number of drives as well as the total number of storage arrays are increased significantly. This increased number of components causes not only higher power consumption but places a stress on both cooling requirements and data center floor space. The introduction of enterprise Flash drives to such an environment can lead to both increased performance and also a consolidation of storage units. A reduction in the total number of drives and arrays will extend the power, cooling, and space savings benefits. The following paragraphs describe DB2 on z/OS environments best suited for enterprise Flash drives from a performance and consolidation perspective.

All I/O requests from a host are serviced by the Symmetrix from its global cache, regardless of the underlying disk technology. Under normal operating conditions, write requests are written directly to cache and incur no delay due to physical disk access. For read requests, if the requested data is in global cache, either because of a recent read or write, or due to sequential prefetch, the request is immediately serviced without accessing disk. This is referred to as a read hit. If the requested data is not in global cache (an event referred to as a read miss) the Symmetrix must retrieve the data from disk. The result of a read miss is an increased I/O response time due to the mechanical delays of hard disk drives. Enterprise Flash drives are therefore ideally suited for workloads with random small block I/O requests that exhibit low Symmetrix cache hit rates.

DB2 on z/OS Flash drive recommendations

DB2 log and table space placement

The DB2 active log workload, under normal processing activities, is sequential write activity. Since writes are serviced from Symmetrix cache as previously discussed, the writes to the DB2 active log will not generally benefit from Flash drives. In an effort to efficiently utilize the Flash drive space with the most appropriate workload, it is reasonable to place the DB2 active logs onto HDD. Placing the log workload onto HDD will allow more DB2 table space datasets, the best candidate for use on Flash drives, to occupy the drives thereby optimizing the use of Flash drive space.

RAID protection

It is recommended to use RAID 5 as the protection mechanism for the database volumes when deployed on enterprise Flash drives. RAID 5 is the optimal trade-off of protection, cost, and performance for enterprise Flash drives. If performance is paramount and cost a lesser factor, RAID 1 protection with Flash drives can offer improved performance.

While RAID 5 volumes have additional overhead of recalculating parity on write operations, this penalty is negligible in comparison to the high performance characteristics of the enterprise Flash drives. On the other hand, with the active log workload on HDD, it is recommended for the log volumes to be in a RAID 1 or RAID 10 configuration.

DB2 and Flash drive testing and analysis

In order to demonstrate the performance differential between HDD and enterprise Flash drives two sets of databases within a single DB2 subsystem were set up. One of the sets and its ICF catalog were deployed onto the hard drives. The other identical construct was set up on enterprise Flash drives on the same

1 IBM DB2 performance personnel use the term “Synchronous I/O” for a read miss.
DMX-4 with its own ICF catalog. The location of the catalog with its related volumes enabled a volume based restore for repetitive testing.

Testing configuration

It should be noted that RAID 10 protection schemes were used for both source and target volumes, which is contrary to the general recommendation that RAID 5 be used for the enterprise Flash drives. RAID 10 was utilized for these tests in order to make both the HDD and FD configurations as identical as possible.

The two configurations were made as similar as possible in order to highlight the relative performance differential. The following table describes the two database configurations:

Table 1. Comparing hard drive and Flash drive configurations

<table>
<thead>
<tr>
<th>Configuration aspect</th>
<th>Hard Drives</th>
<th>Flash Drives</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protection</td>
<td>RAID 10</td>
<td>RAID 10</td>
</tr>
<tr>
<td>Drive count</td>
<td>24</td>
<td>24</td>
</tr>
<tr>
<td>Drive size</td>
<td>146 GB</td>
<td>146 GB</td>
</tr>
<tr>
<td>Drive attachment</td>
<td>2 Gb</td>
<td>2 Gb</td>
</tr>
<tr>
<td>Drive rpm</td>
<td>10k</td>
<td>n/a</td>
</tr>
<tr>
<td>Channels</td>
<td>6 x FICON</td>
<td></td>
</tr>
<tr>
<td>Volume occupancy</td>
<td></td>
<td>90%</td>
</tr>
<tr>
<td>Instrumentation</td>
<td>RMF (one minute intervals) and STP</td>
<td></td>
</tr>
<tr>
<td>Processor</td>
<td></td>
<td>Z9</td>
</tr>
<tr>
<td>Operating system</td>
<td></td>
<td>z/OS 1.8</td>
</tr>
<tr>
<td>DB2 version</td>
<td></td>
<td>8.1</td>
</tr>
<tr>
<td>Write percentage</td>
<td></td>
<td>15%</td>
</tr>
<tr>
<td>Random read percentage</td>
<td></td>
<td>85%</td>
</tr>
<tr>
<td>Dynamic HyperPAV</td>
<td></td>
<td>On</td>
</tr>
</tbody>
</table>

z/OS volume setup

Both the Flash drive and the hard drive configuration were deployed using:

- 50 Mod 9 Volumes
- 46 Mod 27 Volumes

Multiple databases were manually deployed to get precise control of where all the datasets were placed in order to be able to replicate the database configuration from the HDD to the Flash drive. Each volume was filled to approximately 90 percent such that the total occupied space for data and indexes on each set of drives was about 1.3 TB.

The DB2 catalog, BSDS, active logs, and archive logs were all placed on different physical drives from the Flash drives and HDD under analysis. The subsystem was constructed this way in order to make these components fixed for both sets of drives and to become a non-factor in the performance measurements.

DB2 transaction workload

The transaction workload was generated using 32 batch jobs running simultaneously. Each batch job simulated online users and processed an input dataset with approximately 75,000 transactions in it. Each transaction processed tables in an order-entry warehouse system. Each transaction fulfilled orders, adjusted
quantities, stock levels, and delivery information. The input transactions were generated with randomized keys resulting in a very random access pattern to the table spaces involved.

**DMX-4 configuration**

A DMX-4 4500 was used for the testing with a 64 GB cache. As there were no other workloads running on the DMX competing for cache resources and with the DB2 subsystem being a relatively small system (1.4 TB), it was decided to use Dynamic Cache Partitioning to limit the DB2 test activity to only a small portion of the total available cache. With this cache setup it was possible to get a read-miss rate that was reasonable for a random workload.

Running DB2 in such a DMX cache-constrained configuration more closely represents what DB2 would be faced with in the real world where DB2 would be competing with many other applications for cache slots.

**Testing results**

Figure 5 compares the response time measured by RMF for the two identical workloads. The Average response time over the duration of the jobs for the hard drives was approximately 15 ms while the response time from the Flash drives was about 2.5 ms. The transaction workload on the hard drives completed in about 80 minutes while the workload on the Flash drives completed in about 26 minutes.

![Response Time Comparison](image)

**Figure 5. Response time comparison between Flash drives and HDD**

Figure 6 shows the corresponding IOPS for the two workloads. Note that the Flash drives’ workload performs about 10,800 IOPS while the HDD runs with about 2,800 IOPS. While the same number of read and write operations are executed for each workload, that is to say the area under the graph is the same for both, the actual job duration is reduced significantly since there is less time spent waiting on I/O.

It should be noted that the Flash drives are able to deliver much higher I/O rates than those achieved in the tests. This capability can be seen by examining the “busyness” of the drives and how much headroom they have to perform more work. This comparison is shown in a later section in this paper.
**Figure 6. IOPS comparison between Flash drives and HDD**

Figure 7 contrasts the different CPU consumption of the two activities. It is important to realize that when a task is not waiting for I/O (or waiting minimally) it has more time to process the data. This will always lead to an increase in CPU%. This does not necessarily translate into more CPU consumed but simply the same amount of CPU consumed in a much shorter period of time. Essentially, as with the previous graph, the area under the graph lines is identical for both workloads.

**Figure 7. CPU consumption comparison between Flash drives and HDD**

It is important to realize that while replacing a HDD with an enterprise Flash drive will greatly assist with random read misses, if there are other I/O bottlenecks in the system, the impact may not be as dramatic as expected. The mainframe I/O can be broken up into four components:

- **IOSQ** (time spent queuing in the host)
- **PEND** (time spent waiting for a path to the device)
- **DISC** (time spent waiting for data access)
• **CONN** (time spent during data transfer)

Replacing HDD with EFD will primarily show improvement in DISC time — the time that the host is waiting for the storage subsystem to retrieve the data. The other three factors may show some related improvement by having faster disk access, but not to the same level. Figure 8 contrasts the disconnect time measured by RMF for the two workloads. On average the Flash drives showed disconnect times of 1.5 ms while the HDD had DISCONNECT times of about 13 ms. This is almost an order of magnitude level of improvement.

![Disconnect Times (MS)](image)

**Figure 8. Disconnect times for HDD and FD**

So far all the data presented for review of the performance of the DB2 workload has come from RMF. One metric that cannot be measured with RMF is the “busyness” of the drives. This is because RMF does not have the notion of the physical disk but rather only those volumes presented as slices of the physical disk. For this reason the EMC tool STP Navigator was used to measure the busyness of the disks while running on the DMX service processor.

It is important to understand how busy the disks are under such a workload because that shows how much more headroom the drives have until they reach their maximum ability to deliver I/O with an acceptable response time.

In an ideal situation it would have been desirable to drive the IOPS to the Flash drives to a much higher level. The practical aspects of overcoming the following determined the workload that was eventually measured:

- DB2 buffer pool hits
- Front-end cache hits
- High host CPU utilization

Figure 9 shows that during the test the 24 HDDs are running at 85%-95% of their capability. In fact, there is very little room there for any more I/O activity without severely elongated response times. However, the Flash drives are only running at 10%-15% of their capability and are almost idling when delivering 11,000 IOPS or close to 500 IOPS per drive.
Selected placement of DB2 table spaces

While it is instructive to understand how a DB2 system can benefit from having all its data on Flash drives, it may not be practical from an economic point of view. If a DB2 subsystem exists on both hard drives and Flash drives, it would be better if one could locate on the Flash drives only the table spaces that experienced a large number of random read misses. This way there is a better return on investment from the Flash drives.

The determination of which table spaces have a high read-miss rate can be done by the examination of the SMF 42 subtype 6 records. These records provide the information on whether I/Os are hits or misses in the DMX cache. It is normal that very large datasets that are randomly accessed will have a high miss rate. So choosing the datasets with the highest miss rates might be a good approach except that the Flash drive space will be consumed very quickly if the largest datasets are all selected. The better choice would be to choose the datasets with the highest “miss density.” This means those datasets with the highest number of misses per MB of storage consumed. The miss density can be calculated from the RMF 42 subtype 6 records.

Customers can engage EMC personnel to help determine the best datasets to be placed on enterprise Flash drives. EMC staff can use internal tools to process and analyze the SMF data to assist customers in efficient and effective deployment of enterprise Flash drives. EMC finds that most customers know which DB2 tables are most critical. They also know through monitoring tools or buffer displays which tables are randomly read. This knowledge in addition to the EMC analysis will lead to optimal use of Flash drives.

Use cases

In general, customers can justify the purchase of Flash drives for increased performance, reduced footprint, and lower energy costs. There are some specific use cases where this justification is even easier. Here are a few examples:

- The length of time to process the batch window is growing and growing and the window now overlaps into normal online transaction time. Every conceivable tweak to the application has been made, other than a major application rewrite. Every DB2 and z/OS change has been made to reduce the duration of the batch run. There are no more “tweaks.” In order to handle overlaps in batch and transaction periods, the customer ends up overprovisioning host resources.
Flash drives might be the solution in this case. Flash can shorten the time to run batch without the exhaustive and costly rewrite of the application. By eliminating the batch/online overlap through EFDs, overprovisioning host resources would not be necessary, providing significant cost savings.

- Performance is good most of the time and the Service Level Agreements (SLA) are being met except during market open. During market open the sheer volume of accesses is causing the system to degrade. SLAs to the end user are not being met. This can result in hundreds of thousands of dollars in fines based on these agreements.

Flash drives may help in these situations by placing the table spaces with high demands on Flash drives and avoiding the penalties associated with not meeting SLAs.

- To meet the required performance during the day it was necessary to buy far more HDDs than were needed based on the actual storage capacity of the drives themselves. The additional unused space on the drives was wasted by short-stroking the drives. Additional frames were needed to accommodate all the additional drives.

Flash drives can really help reduce cost here by providing the IOPS needed for the application without wasting disk space and without the additional power and cooling requirements of the extra storage frames.

- Your critical application is suddenly experiencing response issues. Customer satisfaction is negatively impacted. Initial analysis indicates that DASD response time has degraded significantly. There are no obvious reasons for the DASD response time degradation. In-depth analysis will have to be done.

Flash drives may help. As a stopgap measure, moving the datasets on the stressed DASD to a Flash drive device could improve response enough to get by the immediate issue. In-depth analysis to discover the root cause of the problem can proceed without further impacting customer satisfaction.
Conclusion
Incorporation of Flash drives into Symmetrix DMX-4 with Enginuity 5773 provides a new Tier 0 storage layer that is capable of delivering very high I/O performance at a very low latency, which can dramatically improve storage performance for DB2 on z/OS. Magnetic disk drive technology no longer defines the performance boundaries for mission-critical storage environments. The costly approach of spreading workloads over dozens or hundreds of underutilized disk drives is no longer necessary. Depending on the DB2 environment, Flash drives can lead to increased performance and savings in power, cooling, and data center floor space requirements.

Symmetrix now combines the performance and power efficiency of Flash drive technology with traditional disk drive technology in a single array managed with a single set of software tools, to deliver advanced functionality, ultra-performance, and expanded storage tiering. Also, by utilizing EMC storage management expertise, users can identify the table spaces with the most demanding access requirements and place these on the Flash drives to get the best ROI for their investment in this technology.

References
The following white papers can be found on EMC.com and Powerlink®:

- Implementing EMC Symmetrix DMX-4 Flash Drives with OLTP Databases
- EMC Symmetrix DMX-4 Ultra-Performance Tier 0 Using Flash Drives
- EMC Symmetrix DMX-4 Flash Drives with Microsoft Exchange
- EMC Symmetrix DMX-4 Enterprise Flash Drive Solution for Initiate Systems Master Data Hub
- Implementing EMC Symmetrix DMX-4 Flash Drives with Oracle Databases

The following feature sheet can be found on Powerlink:

- Flash Solid State Drives for Symmetrix DMX-4