

## **IBM® System Storage™ DS8700™ Performance with Easy Tier®**

**May 2010**

**Lee La Frese  
Kaisar Hossain  
Joseph Hyde  
Andrew W. Lin  
Bruce McNutt  
Christopher Sansone  
Leslie Sutton  
Yan Xu  
Yijie Zhang**

**Document WP101675**

**Systems and Technology Group  
© 2010, International Business Machines Corporation**

## **Notices, Disclaimer and Trademarks**

Copyright © 2010 by International Business Machines Corporation.

No part of this document may be reproduced or transmitted in any form without written permission from IBM Corporation. Product data has been reviewed for accuracy as of the date of initial publication. Product data is subject to change without notice. This information may include technical inaccuracies or typographical errors. IBM may make improvements and/or changes in the product(s) and/or programs(s) at any time without notice. References in this document to IBM products, programs, or services does not imply that IBM intends to make such products, programs or services available in all countries in which IBM operates or does business. THE INFORMATION PROVIDED IN THIS DOCUMENT IS DISTRIBUTED "AS IS" WITHOUT ANY WARRANTY, EITHER EXPRESS OR IMPLIED. IBM EXPRESSLY DISCLAIMS ANY WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE OR NON-INFRINGEMENT.

IBM shall have no responsibility to update this information. IBM products are warranted according to the terms and conditions of the agreements (e.g., IBM Customer Agreement, Statement of Limited Warranty, International Program License Agreement, etc.) Under which they are provided. IBM is not responsible for the performance or interoperability of any non-IBM products discussed herein. The performance data contained herein was obtained in a controlled, isolated environment. Actual results that may be obtained in other operating environments may vary significantly. While IBM has reviewed each item for accuracy in a specific situation, there is no guarantee that the same or similar results will be obtained elsewhere. Statements regarding IBM's future direction and intent are subject to change or withdraw without notice, and represent goals and objectives only. The provision of the information contained herein is not intended to, and does not grant any right or license under any IBM patents or copyrights. Inquiries regarding patent or copyright licenses should be made, in writing, to:

IBM Director of Licensing  
IBM Corporation  
North Castle Drive  
Armonk, NY 10504-1785  
U.S.A.

IBM, Easy Tier, FlashCopy, System Storage, Storage Tier Advisor Tool, DSMT, IBM Tivoli Storage Productivity Center, DS8000, DS8300, and DS8700 are trademarks of International Business Machines Corporation in the United States, other countries, or both. Other company, products or service names may be trademarks or service marks of others.

## **Acknowledgements**

The authors would like to thank the following colleagues for their comments and insight:

Lawrence Y. Chiu – IBM Research, Almaden, CA.

E. Allen Marin – IBM Systems & Technology Group, Boulder, CO.

Dietmar Noll – IBM Software Group, Germany.

Vic T. Peltz – IBM Systems & Technology Group, San Jose, CA.

Richard A. Ripberger – IBM Systems & Technology Group, Tucson, AZ.

David Sacks – IBM Systems & Technology Group, Chicago, IL.

Sonny E. Williams – IBM Systems & Technology Group, Tucson, AZ.

## **A Note to the Reader**

This White Paper assumes a familiarity with the general concepts of Enterprise Disk Storage Systems and the DS8000 product line. Readers unfamiliar with these topics should consult the References section at the end of this paper.

## **Table of Contents**

Acknowledgements.....	3
A Note to the Reader .....	3
Table of Contents.....	4
<b>1</b> Introduction .....	5
<b>1.1</b> Audience .....	6
<b>2</b> Introducing Easy Tier .....	7
<b>2.1</b> A High-Level Description of Easy Tier .....	7
<b>2.2</b> The Advantage of Using Easy Tier Automatic Mode.....	8
<b>2.3</b> Easy Tier Tools .....	9
<b>3</b> DS8700 Performance with Easy Tier Automatic Mode .....	13
<b>3.1</b> SPC-1 Performance .....	13
<b>3.2</b> DB2 Brokerage Workload Performance.....	16
<b>4</b> DS8700 Performance with Easy Tier Manual Mode .....	21
<b>5</b> Best Practices and Considerations.....	23
<b>5.1</b> z/OS Considerations.....	24
<b>5.2</b> IBM Tivoli Storage Productivity Center (TPC) Considerations .....	24
<b>6</b> Conclusions.....	25
<b>7</b> References.....	25
<b>8</b> Frequently Asked Questions .....	26
<b>9</b> Appendix .....	28

## 1 Introduction

The authors of this white paper had the privilege to be among the first to see Easy Tier “in action” while preparing the information presented here. In one key test, documented in a subsequent section of the paper, we placed 34 terabytes of SPC-1 data onto 96 SATA drives. We then started the SPC-1 test workload, while also enabling Easy Tier.

During the ensuing 24 hours, we watched the performance of the system gradually improve, until finally the system throughput achieved a level more than 3 times its initial value. No other action was needed on our part, other than to observe a significant I/O throughput increase with a complementary reduction in response time. The movement of a small, selected portion of the SPC-1 data, needed to accomplish such dramatic performance gains, was entirely automatic. Easy Tier first identified the most active data stored on the SATA drives, then gradually moved this data to a small amount of available SSD capacity, while also carefully limiting the load imposed by the data movement itself.

The ability for a system to improve itself in this manner is unique in the experience of the authors of this paper and was exciting to watch. In this paper, we would like to invite all DS8700 customers to evaluate the breakthrough technology in performance management that Easy Tier represents, and to see for themselves the gains experienced by their own applications. For those customers that have an urgent need to make some of their I/O bound applications run faster, we strongly recommend Easy Tier as a simple and powerful tool to achieve the needed performance improvements.

This paper includes Easy Tier results produced with SPC-1, as well as a study of the performance gains achieved in a DB2 Brokerage workload, both in an Open systems environment. In both cases, similar throughput gains of over three times were achieved. In the DB2 Brokerage results, the ability to process I/O faster led directly to increased transactions resulting in the ability to utilize the System p Power 7 servers more effectively.

To put Easy Tier into its proper perspective, consider other technologies that have made it possible to form a memory hierarchy by combining multiple types of storage or memory media. Examples include host virtual memory backed by paging, and the L1 and L2 caches provided on a typical processor chip. There appears to be every reason to believe that Easy Tier (and subsequent advances that build upon it) can become as prevalent a part of enterprise computing as the two other examples of memory hierarchies just cited. This potential comes from its capability to combine the best aspects of contrasting storage media. Maximum performance may be provided for the “hot” data that needs it, while large capacity, inexpensive disks may be used for data that does not need constant access.

Besides the obvious benefit of improved application performance, Easy Tier also offers the opportunity to shift to high capacity and less expensive storage media to store the bulk of the application data. In the SPC-1 test, 38 TB of addressable storage capacity were provided in a single DS8700 frame (96 Serial Advanced Technology Attachment (SATA) drives configured as RAID-10). We expect that many DS8700 customers will consider the ability to store more data into a more compact footprint, made possible with Easy Tier, to be a compelling benefit. Along with the more compact footprint, energy requirements can also be reduced by deploying a smaller number of drives. Best of all, we show in this white paper that these reductions of

footprint and energy can be accomplished while also improving the response time seen by applications.

## **1.1 Audience**

This technical paper was developed to assist IBM and IBM Business Partner field sales representatives and technical specialists and IBM's clients in understanding the performance characteristics of the IBM 2107 Model 941 and the new Easy Tier feature. The IBM 2107 Model 941 is the DS8700, POWER6 model and shall be referred to as the DS8700 throughout this paper.

## 2 Introducing Easy Tier

### 2.1 A High-Level Description of Easy Tier

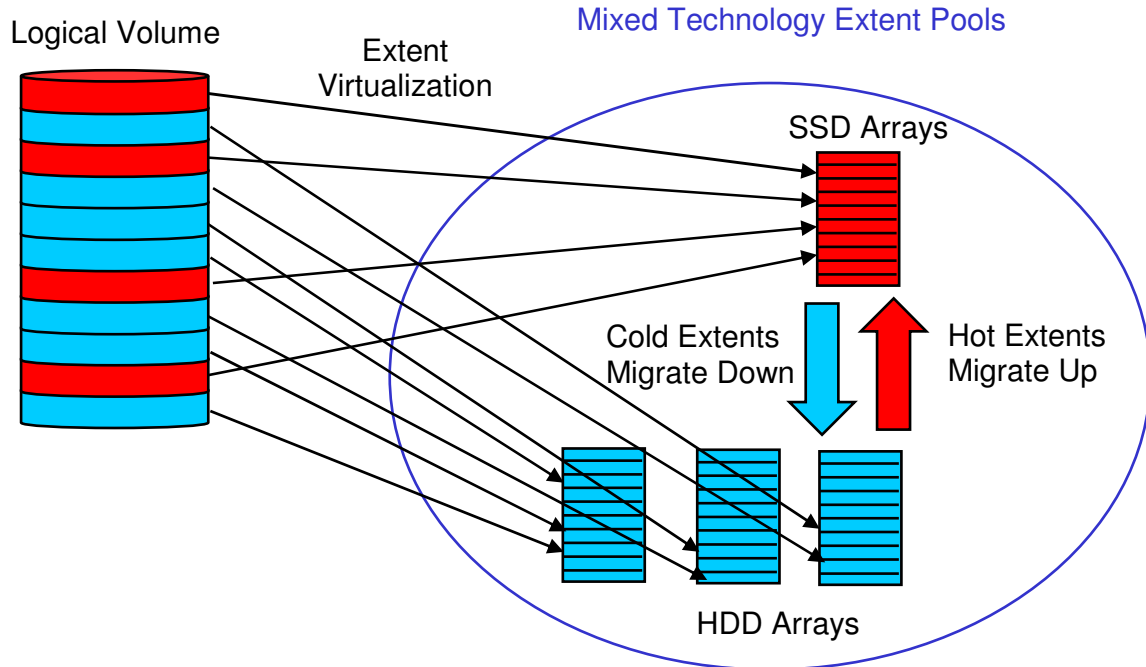
Easy Tier is a new functionality introduced for the DS8700 with Licensed Internal Code (LIC) Release 5.1. It is designed to ease storage management and improve performance for real world customer environments. All Easy Tier functions may be used concurrently without disruption to applications. Easy Tier features two modes of operation: Automatic and Manual. Automatic mode optimizes performance in a tiered storage pool while Manual mode allows movement of volumes between or within storage pools<sup>1</sup>.

Solid-State Drives (SSDs) have been available in the DS8000 product family since 2009, but deciding what data would most benefit by this technology has been largely left to storage administrators. They could rely on their experience and reports from various tools but often this was too time consuming and tedious to implement. With Easy Tier Automatic mode, the DS8700 monitors the workload in tiered storage pools and determines at a sub-volume granularity which data would most benefit from SSDs. The storage administrator need only configure the storage pools with a mix of Hard Disk Drives (HDDs) and SSDs and turn it on. Easy Tier Automatic mode will monitor continually and move data as needed between the SSD ranks and HDD ranks within the storage pool. Currently, Easy Tier Automatic mode only manages movement between SSDs and HDDs and does not distinguish classes of HDD by rotational speed or RAID (Redundant Array of Independent Disks) configuration.

Easy Tier Automatic mode is very simple to use and the primary control is to turn it on or off via the DSCLI (DS8000 Command Line Interface) or the DSGUI (DS8000 Graphical User Interface). The monitoring function is a process with very low overhead that runs on the Power 6 processors within the DS8700 and causes no perceptible effect on system performance. Easy Tier Automatic mode uses about a 24-hour window of time to determine which extents would most benefit from SSD within the tiered pool configuration. It bases its decisions on both the disk access rates and the overall latency of disk I/O operations at the sub-volume level. The Easy Tier algorithms look for high concentrations of random operations with small transfer sizes that are currently on HDDs as candidates for movement to SSDs. It also monitors the extents currently residing on SSDs to ensure that they are still active and “hot” enough compared to the extents that are on HDDs. When hotter extents are seen on the HDDs they may be swapped with cooler SSD extents. Thus, the hottest extents in the pool automatically reside on SSD over time. The backend data rate to the extents is also taken into account when placing data on SSDs to prevent overloading a single SSD rank with more I/O capacity than it can effectively provide. Refer to Figure 1 below for a visual representation of how Easy Tier Automatic mode virtualizes extents.

---

<sup>1</sup> DS8700 refers to storage pools explicitly as extent pools. For the purpose of this paper, the terms Storage Pool and Extent Pool will be used interchangeably referring to the same concept.



**Figure 1:** How Easy Tier Automatic Mode Virtualizes Extents.

Easy Tier Manual Mode provides two new functions, that give a storage administrator expanded storage pool management capabilities. **Dynamic extent pool merge** enables merging existing storage pools, and **dynamic volume relocation** facilitates the movement of volumes between storage pools or within a storage pool to facilitate re-striping the data in the pool. Pools may be merged simply by selecting the two pools to merge and executing a merge DSCLI or GUI command. If one needs to merge multiple pools, then a sequence of merges may be used. Note that the merge does not actually move any customer data so there is zero performance impact on applications. Merging pools may be useful when you have a number of existing storage pools that you wish to combine to begin using with Automatic mode.

Similarly, movement of volumes between pools may be initiated via the DSCLI or DSGUI. It will actually move customer data to the target pool but the data movement is carefully paced to minimize any impact to applications during the move. This function is most useful for managing multiple storage pools with different performance characteristics (e.g. SSD, HDD, SATA or Easy Tier managed) or balancing pools with uneven I/O workload characteristics.

Moving volumes within a storage pool is useful when changing the extent allocation method for a pool from rotating volumes to rotating extents (Storage Pool Striping). It is also a good way to grow storage capacity in a pool by adding or merging current rotate extent pools and re-striping the volumes over all of the ranks in the expanded pool.

## 2.2 The Advantage of Using Easy Tier Automatic Mode

The traditional way of doing storage tiering is to analyze volume-level performance data, decide which volumes are hot, and move these to faster storage manually. The sub-volume operation of Easy Tier is a new approach that we believe provides far greater efficiency than traditional methods.

Key advantages of Easy Tier:

- **Designed to be Easy!** The user is not required to make a lot of decisions or go through an extensive implementation process to start utilizing Easy Tier.
- **Efficient Use of SSD Capacity.** Easy Tier moves 1 gigabyte data extents between storage tiers. This enables very efficient utilization of SSD resources. Other systems may operate on a full logical volume level. Logical volumes in modern storage systems are trending towards larger and larger capacities. This makes migration of data at a volume-level of granularity all the more inefficient by:
  - 1) potentially wasting precious SSD space on portions of logical volumes that are not really hot, and
  - 2) creating more HDD contention when executing the data movement.
- **Intelligence.** Easy Tier learns about the workload over a period of time as it makes decisions about which data extents to move to SSDs. As workload patterns change, Easy Tier finds any new highly active (“hot”) extents and exchanges them with extents residing on SSDs that may have become less active (“cooled off”).
- **Negligible Performance Impact.** Easy Tier moves data gradually to avoid contention with I/O activity associated with production workloads. It will not move extents unless a measurable latency benefit would be realized by the move. The overhead associated with Easy Tier management is so small that the effect on overall system performance is nearly undetectable. This eliminates the need for storage administrators to worry about scheduling when migrations occur.

Easy Tier continues to monitor the workload over time and will migrate extents back to HDDs when changes in I/O patterns warrant. By contrast, with other implementations that are volume-oriented, the overhead of migrating data back to HDDs may be relatively high and attention may be needed to avoid impact on production I/O activity. It is likely that those systems’ management policies and settings may need ongoing updates due to common storage management activities such as adding new drives, creating new volumes, or applications changing the volumes they reside on. With Easy Tier, these worries disappear.

## 2.3 Easy Tier Tools

### 2.3.1 Storage Tier Advisor Tool

The Storage Tier Advisor Tool (advisor tool) provides a high-level summary of workload characteristics and hot spots of volumes that are monitored. It provides assistance for SSD capacity planning with Easy Tier.

The advisor tool is supported on Windows<sup>TM2</sup> and can be installed similarly as DSCLI. Input data files for the advisor tool, i.e. Easy Tier summary data, can be offloaded from the DS8700 and the output from the advisor tool can be viewed using any web browser.

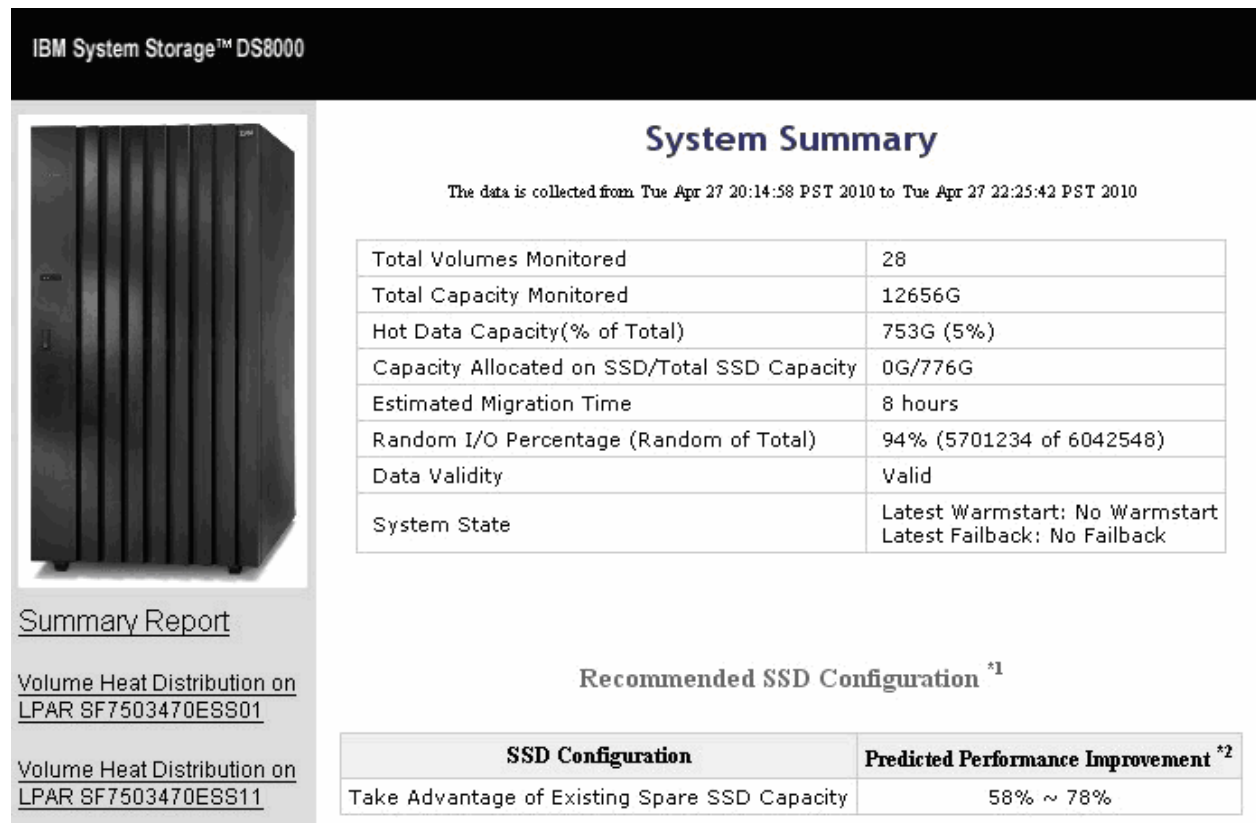
The following example was created to demonstrate how to interpret results from the advisor tool. The test utilized 224 73GB 15K RPM HDDs and 16 73GB SSDs, all configured as RAID-5 and resulting in a total capacity of about 13TB. An OLTP (Online-Transaction Processing) workload was run on the system at 20 KIOPS, which represents a typical I/O access density (1.5 IO/sec/GB).

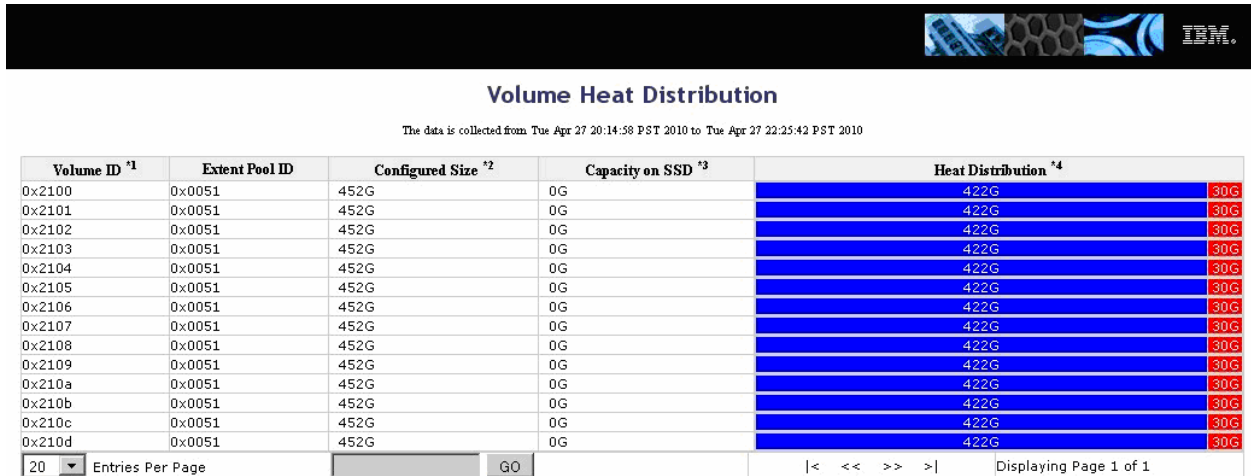
---

<sup>2</sup> Windows is a registered trademark of Microsoft Corporation.

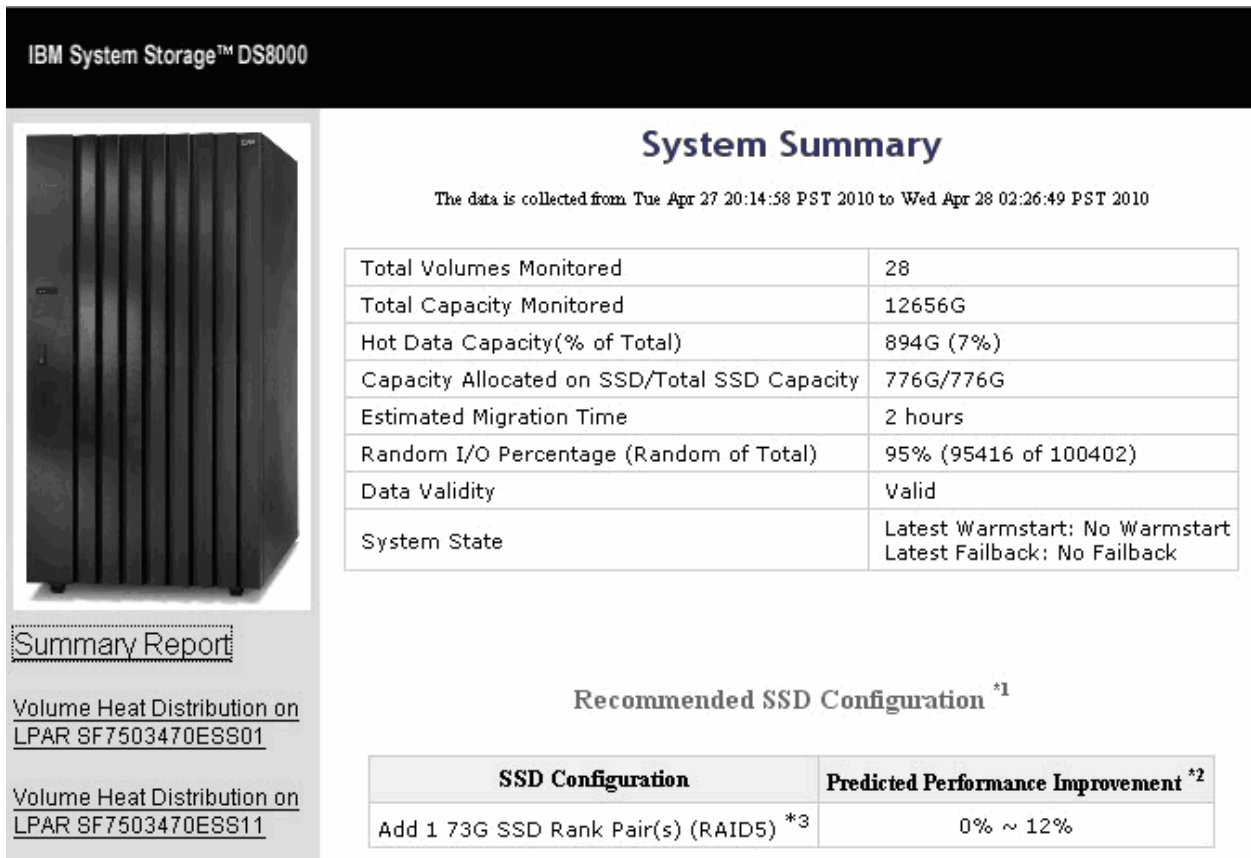
Two snapshots of data were captured: one at the end of the learning period, before any extents were moved (see Figures 2 and 3); the other after SSDs were filled with hot extents (see Figures 4 and 5). Volume Heat Distribution was generated for both Storage Servers, but to save space, the report from only one of the Storage Servers is shown here.

Figures 2 and 4 show that the OLTP workload contained about 5-7% hot data, which is consistent with the characteristics of this particular OLTP workload. The volume heat distribution in Figures 3 and 5 shows that I/O activity is evenly distributed among the volumes matching the characteristics of the workload. Figure 5 also shows the amount of extents that were moved to SSD ranks for each volume. The response time of the workload at 20 KIOPS was improved from 9 ms with HDDs to only 1 ms after most of the hot extents were migrated to SSDs, which is better than the prediction generated by the tool as shown in Figure 2.

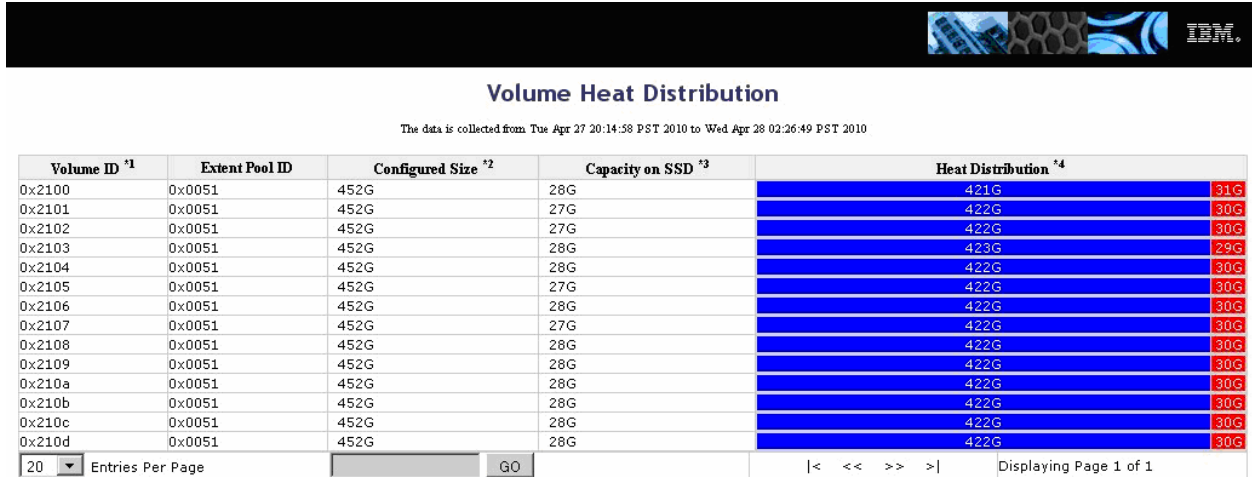




**Figure 3:** Advisor Tool Volume Heat Distribution for ESS11 after Easy Tier learning period, no extents were moved to SSD ranks yet.



**Figure 4:** Advisor Tool System Summary after Easy Tier migration fills SSD capacity.



**Figure 5:** Advisor Tool Volume Heat Distribution for ESS11 after Easy Tier migration fills SSD capacity.

### 2.3.2 Other Tools

More detailed analysis regarding Easy Tier planning can be performed by involving IBM services. Two tools available are Detailed Smart Monitoring (DSMT) and Disk Magic<sup>TM3</sup>. DSMT provides detailed analysis of workload characteristics and extent level heat map over time for each logical volume. Disk Magic is a performance modeling tool used by IBM that can help predict the expected performance of a DS8700 with a specific configuration running a specific workload and modeling of Easy Tier configurations is planned to be supported by this tool.

<sup>3</sup> Disk Magic is a registered trademark of IntelliMagic, Inc.

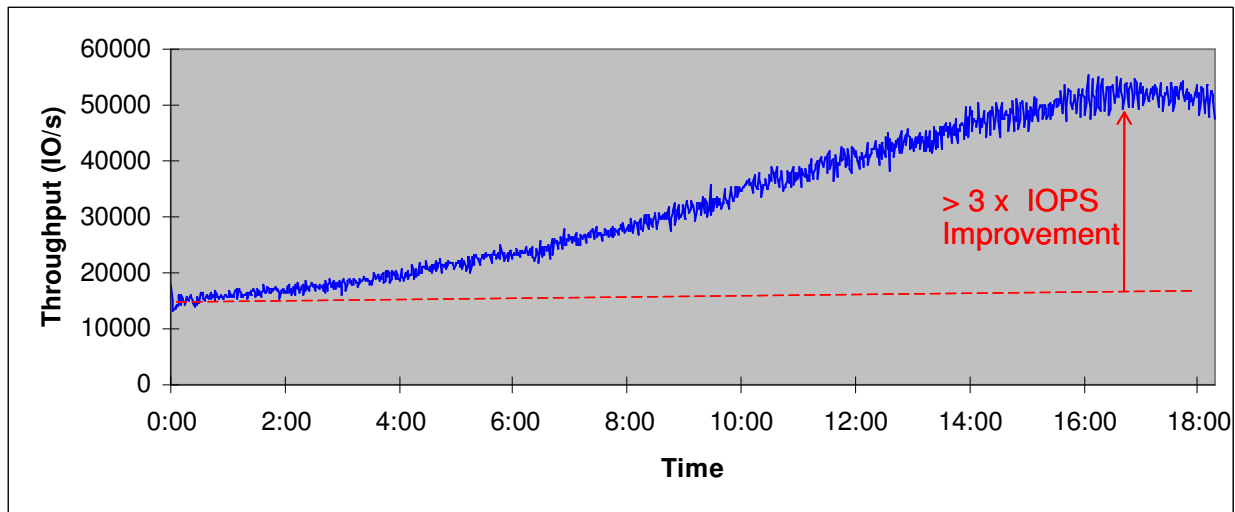
### 3 DS8700 Performance with Easy Tier Automatic Mode

#### 3.1 SPC-1 Performance

The SPC-1 (Storage Performance Council) result for Easy Tier<sup>4</sup> is unique in several ways. Perhaps the most exciting is that it is the first official SPC-1 submission to be based upon SATA drives. This says much about the potential of a storage tier that combines such drives with SSDs to augment their performance. Our SPC-1 drive configuration consisted of 96 SATAs and 16 SSDs.

Another “first” of the Easy Tier submission is that it has by far the longest warm-up period of any SPC-1 result (nearly a full day and night). The reason for this extraordinary warm up period is to show the effect of using Easy Tier for the first time, if no data is initially present on the solid state drives. During the warm-up period, approximately 4.9 percent of the SPC-1 user data was migrated to the SSD media.

This brings us to the third unique aspect of the submission, which is that the chart of events during the warm-up period is by far the most interesting and most telling data produced by the test. Figure 6 is an excerpt from the official SPC-1 submission. The figure presents the period from starting the SPC-1 benchmark until a new stable throughput capability has taken hold. The throughput capability of the system increases by more than three times during this period of just over 16 hours without manual tuning or user intervention.



**Figure 6:** SPC-1 Throughput Improvement due to Easy Tier.

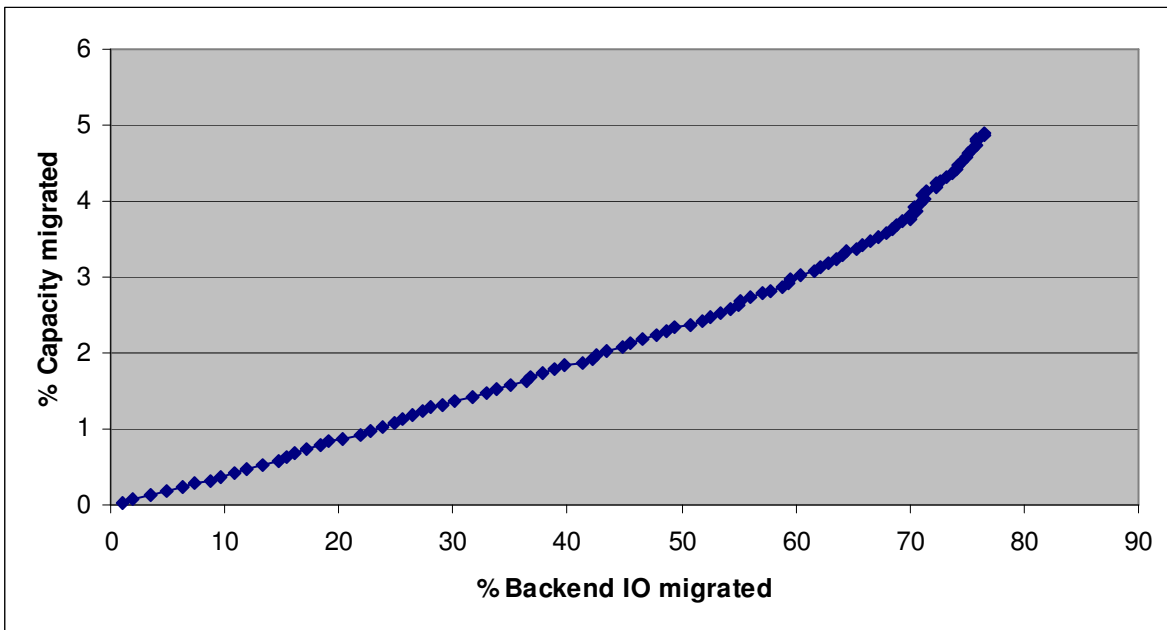
It should be noted that the absolute throughputs presented in Figure 6 are tied to the number of drives and overall size of the system. In a two-rack system, the total number of SATAs and SSD drives could be as much as three times that of the one-rack system of Figure 6, with a corresponding increase in throughput that can be achieved.

<sup>4</sup> [http://www.storageperformance.org/results/benchmark\\_results\\_spc1#a00092](http://www.storageperformance.org/results/benchmark_results_spc1#a00092)

The settings adopted for the SPC-1 test were chosen to allow the impact of Easy Tier to play out within a 24 hour period. More specifically, the minimum learning period for Easy Tier was reduced from 24 hours to one hour<sup>5</sup>. The modified setting just described can be implemented through Product Field Engineering services. However, in an ordinary production environment, 24 hours is likely to be soon enough for the desired migrations and this learning period provides a good precaution against data migrations that fail to offer a long-term performance benefit.

In our Easy Tier settings, we made no change to the maximum rate at which migrations occur. For this reason, Figure 6 does provide a realistic picture of the rate at which the system accommodates to hot spots once the minimum learning period is complete.

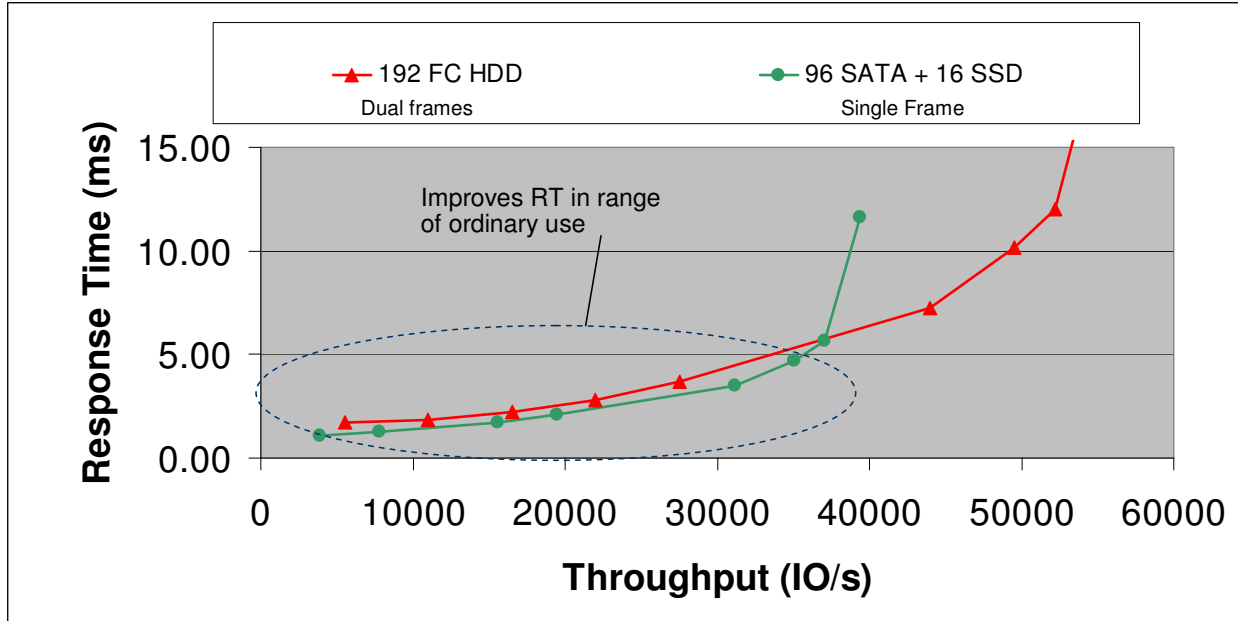
For readers interested in a more detailed view of the data migrations that occurred during the 18 hours of Figure 6, Figure 7 presents the complete migration history during that period of time and also shows the effectiveness of the data migrations in terms of offloading work from the SATA drives. During the 18 hours, approximately 4.9 percent of the SPC-1 data were migrated to SSD and this resulted in offloading about 76 percent of the total I/O.



**Figure 7:** Easy Tier Migrations during the period of Figure 6.

Figure 8 compares the response times seen with the same mix of 96 SATAs and 16 SSD drives to response times obtained with a more traditional configuration of 192 15K RPM drives with approximately the same capacity. As the figure shows, the mix of SATAs and SSDs is competitive with the more traditional 15K drives in terms of overall performance and actually improves upon the observed response times.

<sup>5</sup> As a result of this change, it was also necessary to reduce the minimum I/O threshold for extent migration (since the default for that threshold assumes a learning period of 24 hours). For the purpose of our testing, and in combination with a one hour minimum learning period, we found that a setting of 0 produced satisfactory results (but the default setting is recommended if running with a standard learning period of 24 hours).



**Figure 8:** Easy Tier can improve upon Average Response Times for OLTP.

For applications that rely upon the heavy use of small areas of data (e.g. database indexes), the use of Easy Tier may have an extraordinary impact. A specific example of the benefits of Easy Tier to DB2 transaction processing is presented in the following section.

## 3.2 DB2 Brokerage Workload Performance

A DB2 Brokerage workload was used to evaluate the effectiveness of Easy Tier for improving application performance as well as enabling higher utilization of POWER7 host server resources. This type of experiment represents a class of applications that facilitate and manage transaction-oriented business processes commonly used in a broad range of industry segments including finance, retail, and manufacturing. DB2 Brokerage workloads generate transactions related to trading, account inquiries/updates, placing orders, and market research by brokerage firms dealing with financial markets on behalf of their customers. This DB2 Brokerage workload used to evaluate Easy Tier simulates the design/redesign systems for brokerage firms.

The DB2 Brokerage Application has over 90% read hits in server memory - we observed the overall buffer pool hit ratio at ~97 % throughout the testing. Thus the workload is ideal for evaluating Easy Tier on POWER7 since the disk I/O tends to be highly random and cache unfriendly. Our lab experiments clearly demonstrated the benefit that may be gained using Easy Tier with this type of OLTP application. The tests show the potential for ease of storage management and dramatic improvement in both storage and application performance in real world customer environments without disruption to applications.

### **Workload Configuration:**

We evaluated two DB2 Brokerage workload intensities to illustrate the behavior of the system for both typical and peak I/O intensities. A default HBA queue depth value was used to run a more typical I/O intensity while a much higher queue depth value was set to evaluate peak I/O intensity workloads. In this paper, the following designations are used to identify these two different measurements:

1. Workload with Typical I/O Intensity
2. Workload with Peak I/O Intensity

In both instances, baseline tests were made using HDDs only and compared to mixes of HDDs and SSDs using Easy Tier. The difference in behavior between the typical and peak I/O intensity was solely a function of the queue depth value used in the tests.

Each base was run using a 30 minute ramp up time (RUT) period for the workload plus an added six hours of steady state run time. For the typical I/O intensity runs with Easy Tier (both with 16 and 32 SSDs) we used a 30 minute RUT and eight hours of steady state run time. Since this is a steady state benchmark, we decreased the short term learning window from 24 hours to 1 hour simply to obtain results in a shorter period of time. Likewise we changed the promotion rate from 8 GB every 5 minutes to "back-to-back" migration, which in this case increased the promotion rate to roughly 80 GB every 5 minutes. For the Peak I/O Intensity workloads, we increased the steady state run time to 12 hours for the Easy Tier 32 run while keeping the RUT to 30 minutes. This was needed because of the increased throughput and I/O intensity.

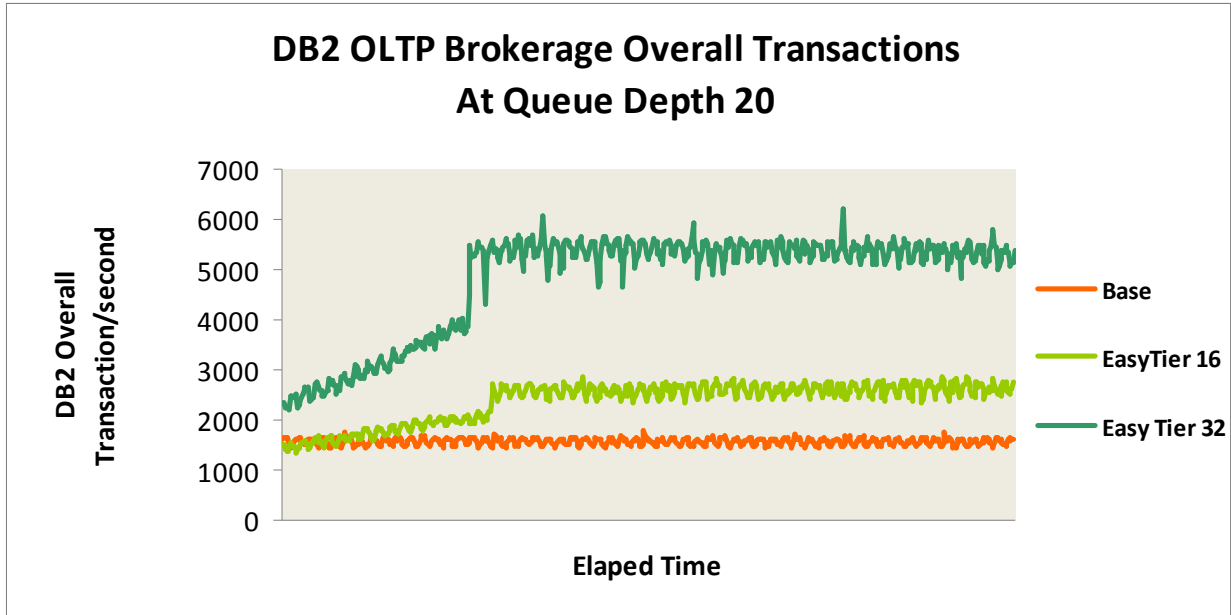
### **Disk and SSD Configuration:**

27 TB of storage capacity from 2 (Device Adapter) DA pairs on a single DS8700 frame was used to create the base: 16 RAID-5 arrays consisting of 128 300 GB 15K HDDs. The base plus 16 146 GB SSDs were used to setup the Easy Tier 16 and similarly the base plus 32 146 GB SSDs used to setup Easy Tier 32 (detail configuration available in Appendix 9.B.2).

The hot extents in the databases are moved dynamically based on both backend disk access frequency and accumulated backend disk latency. The data continues to reside on SSD after a DB2 or server shutdown (it is moved "permanently" but may be replaced later on by what is considered "hotter" data).

### **DB2 Brokerage Workload with Typical I/O Intensity**

The Overall Transaction Rates (OTR) comparison between the Base, Easy Tier 16 and 32 runs is shown in Figure 9. One can see a 241% improvement with the Easy Tier 32 and 66% increase with the Easy Tier 16 compared to the Base run.



**Figure 9:** Overall Transaction Rate at Queue Depth 20.

In addition to OTR improvement, Figure 10 also shows that simply adding Easy Tier 32 improved the backend throughput by 76% and 44% with Easy Tier 16. At the same time, the backend I/O and read response time show significant improvement.



**Figure 10:** Backend Storage Measurements at Queue Depth 20.

Easy Tier also offers a huge benefit (~70%) to the Overall Transaction Response Times (OTRT) and demonstrates the ability to optimize workload out of the box. The response times for some of the key tables – lookup, order, update, and results -- are shown in Figure 11.

Weighted Average RT (ms)					
	Lookup	Order	Update	Results	Overall Transaction
<b>Base</b>	4641	126	6373	143	556
<b>Easy Tier 16</b>	2603	70	3652	82	313
<b>Benefit (%)</b>	<b>43.91</b>	<b>44.44</b>	<b>42.7</b>	<b>42.66</b>	<b>43.71</b>
<b>Easy Tier 32</b>	1237	42	1795	47	154
<b>Benefit (%)</b>	<b>73.35</b>	<b>66.67</b>	<b>71.83</b>	<b>67.13</b>	<b>72.3</b>

**Figure 11:** Weighted Average Overall Transaction Response Times at Queue Depth 20.

So far we have seen the benefits of Easy Tier. Using a POWER7 server in our configuration also improved overall system performance. The POWER7 Server allowed us to leverage performance per core, improved processor utilization, and dynamic infrastructure. Figure 12 shows CPU usage increased by 5X with Easy Tier and POWER7 with improved throughput and response time.

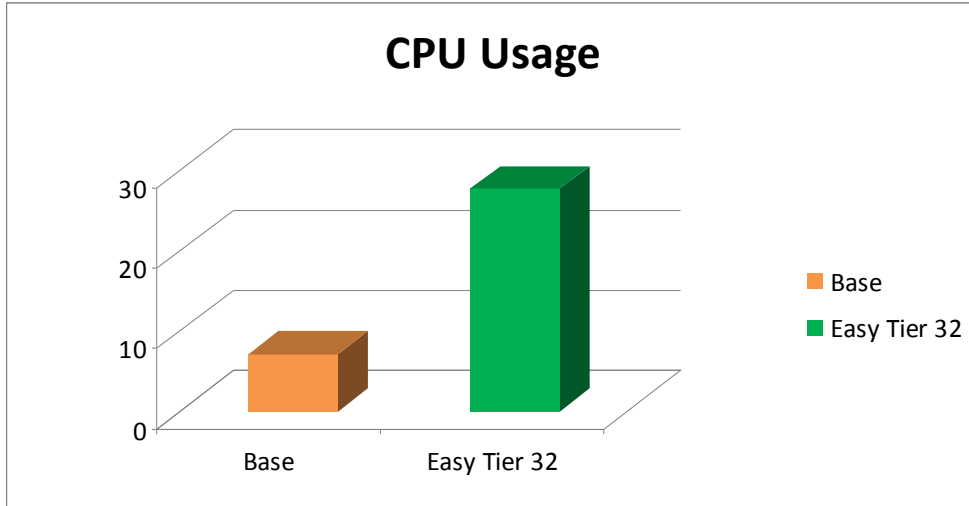


Figure 12: CPU usage at Queue Depth 20.

**DB2 Brokerage Workload with Peak I/O Intensity**

The OTR for the Base and Easy Tier 32 improved by 41% and 35% when the queue depth was set to 256. Figure 13 shows a comparison between the Base and Easy Tier 32 for the Peak I/O Intensity runs – a significant improvement of 226%.

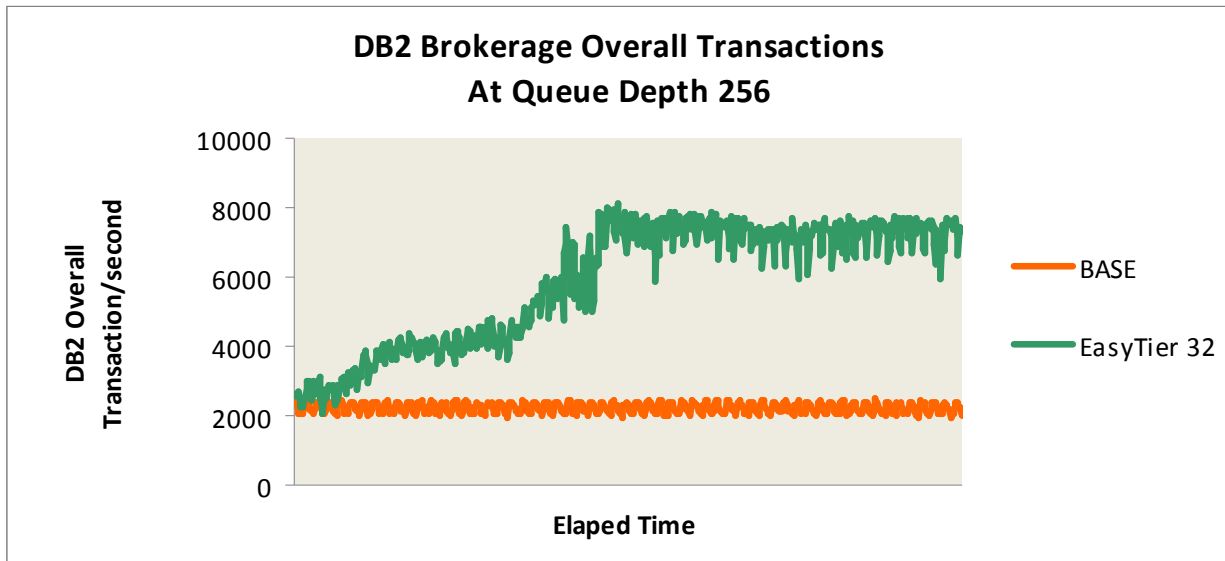
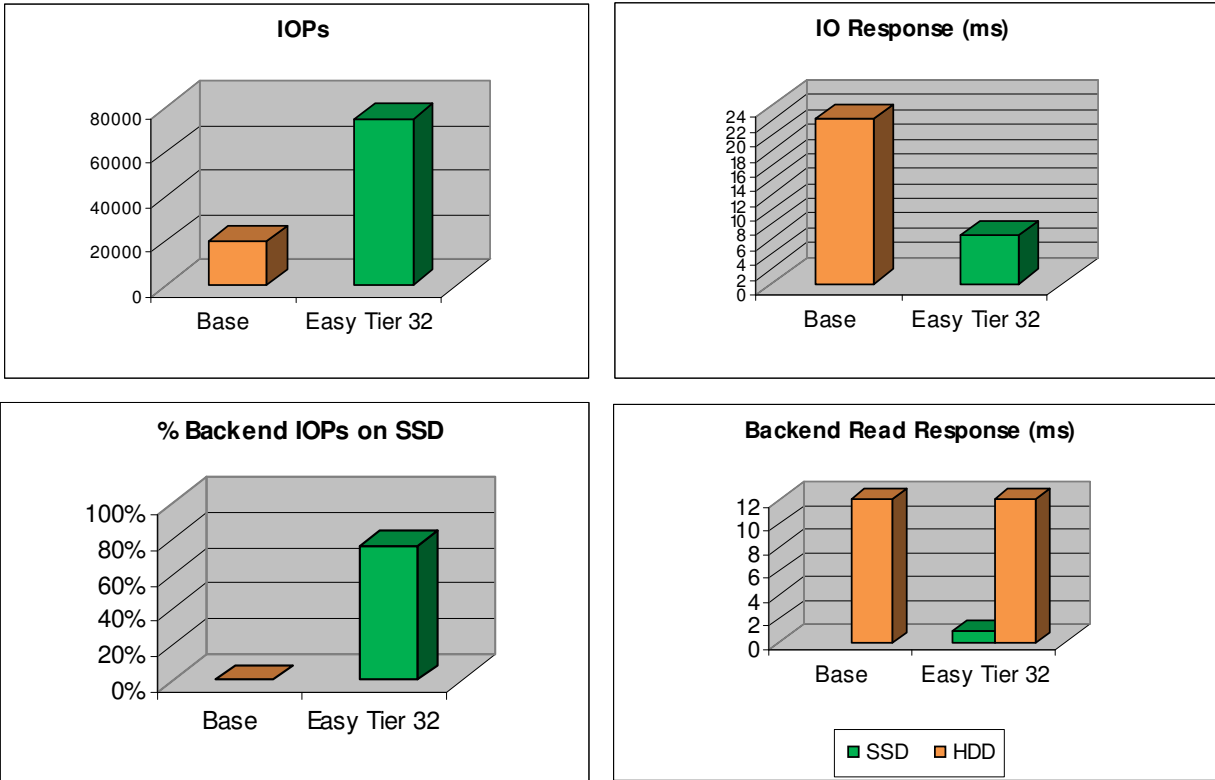


Figure 13: Overall Transaction Rate at Queue Depth 256.

For the typical I/O Intensity workloads, we saw a similar improvement in the backend throughputs. Figure 14 shows an improvement of 76% in backend throughput with Easy Tier 32 while both I/O and read response time decreased very significantly.



**Figure 14:** Backend Storage Measurements at Queue Depth 256.

Remarkable improvement was also seen in the OTRT (72%) despite increased throughputs and without additional HDDs or SSDs as seen in Figure 15.

Weighted Average RT (ms)					
	Lookup	Order	Update	Results	Overall Transaction
<b>Base</b>	3515	98	5033	115	421
<b>Easy Tier 32</b>	794.8	53	1331	52	115
<b>Benefit (%)</b>	<b>77.39</b>	<b>45.92</b>	<b>73.55</b>	<b>54.78</b>	<b>72.68</b>

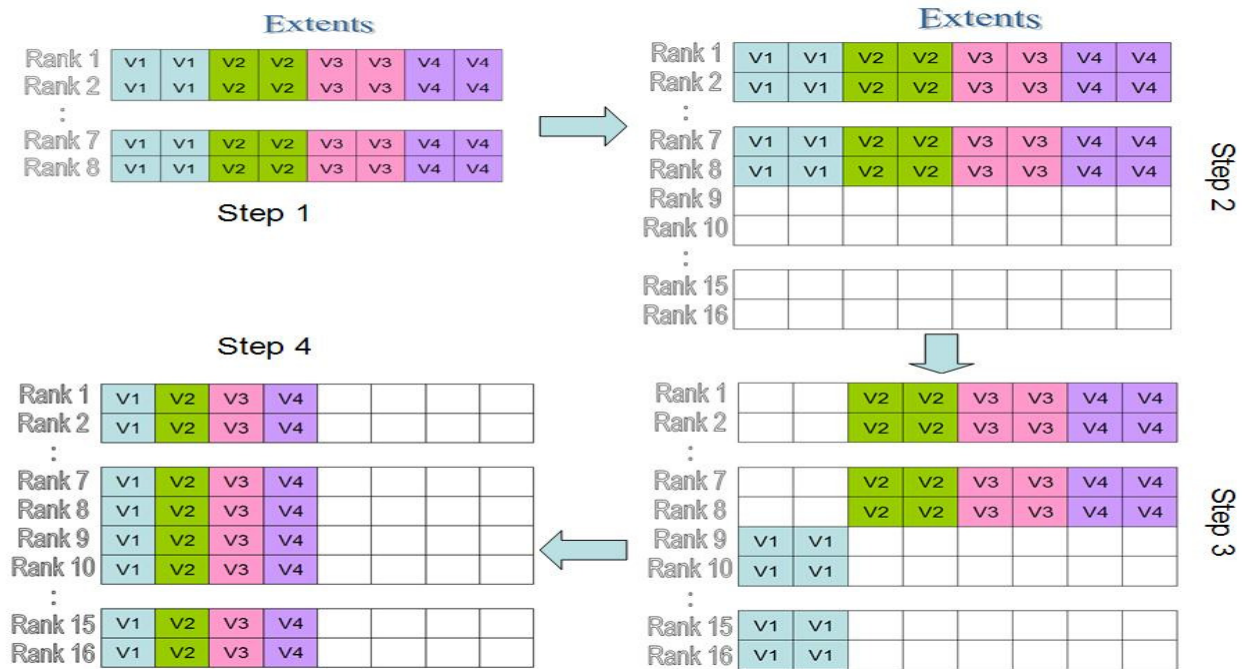
**Figure 15:** Weighted Average Overall Transaction Response Times at Queue Depth 256.

## 4 DS8700 Performance with Easy Tier Manual Mode

Easy Tier Manual Mode allows a logical volume to be migrated to the same or a different extent pool without interruption to host I/O. Similar to FlashCopy, Manual Mode generates asynchronous background activities. It is designed to move data at a rate that will have minimal impact to host I/O performance.

Performance for Easy Tier Manual Mode was evaluated for the following two use cases:

- Migrate within the same extent pool.** More rank capacity was added to an extent pool that was fully populated. Easy Tier Manual Mode was used to re-stripe existing volumes across all available ranks. There were four steps that are shown in Figure 16.
  - Four volumes were striped across eight ranks from two DA Pairs in extent pool P1 and 100% of the capacity in P1 was used.
  - An additional eight ranks from another two DAs were added to P1.
  - In order to re-stripe four volumes across sixteen ranks, one of the four volumes was relocated in P1 first, so that capacity was vacant in the first eight ranks for re-striping.
  - Then the four volumes were relocated at the same time which resulted in re-striping across sixteen ranks. Volume migration rate from this step is shown in Figure 17.
- Migrate across different extent pools.** A set of volumes were relocated from one extent pool to another extent pool. For example, this technique is relevant when there is a desire to have different applications redistributed across different resources. Two extent pools were involved, each containing eight ranks from different DAs. Eight volumes were created in extent pool P1 and four of the eight volumes were relocated to another extent pool, P2.



**Figure 16:** Steps for Migrating Volumes within the Same Extent Pool.

For both cases, the volume migration rate was measured with and without workload. The workload executed was DB Open exercised at 75% of its maximum throughput for this configuration. The impact to workload was also measured as well. Results with FlashCopy with background copy were provided for comparison. FlashCopy setup was similar to 'Migrate across different extent pools'.

Figure 17 shows that the migration rate of volume relocation using Easy Tier Manual Mode is comparable to the copy rate of FlashCopy with background copy, both with and without workload.

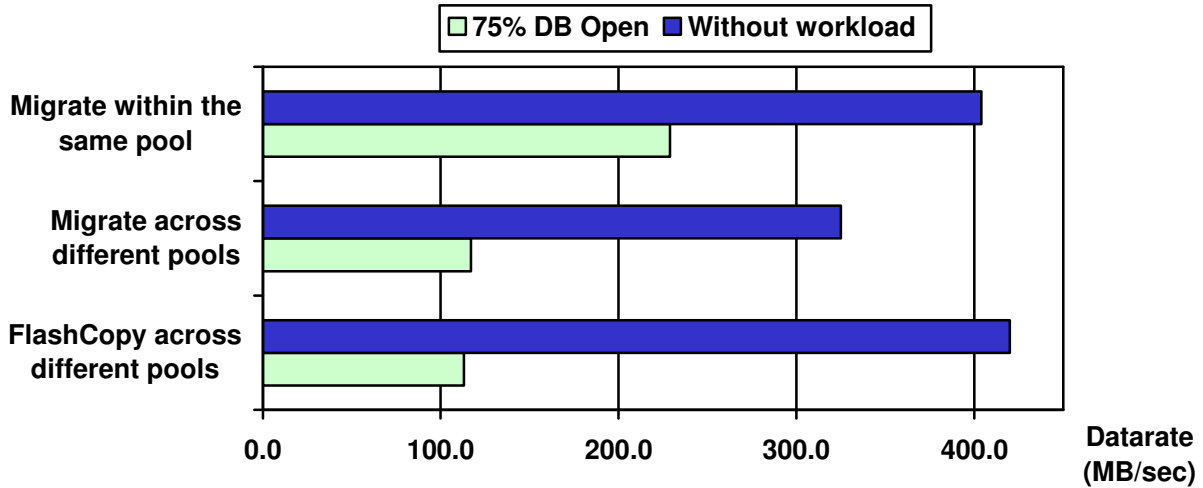


Figure 17: Data Migration and Copy Rates.

As shown in Figure 18, the impact to the host workload is 7% for Easy Tier Manual Mode, compared to 14% for FlashCopy with background copy.

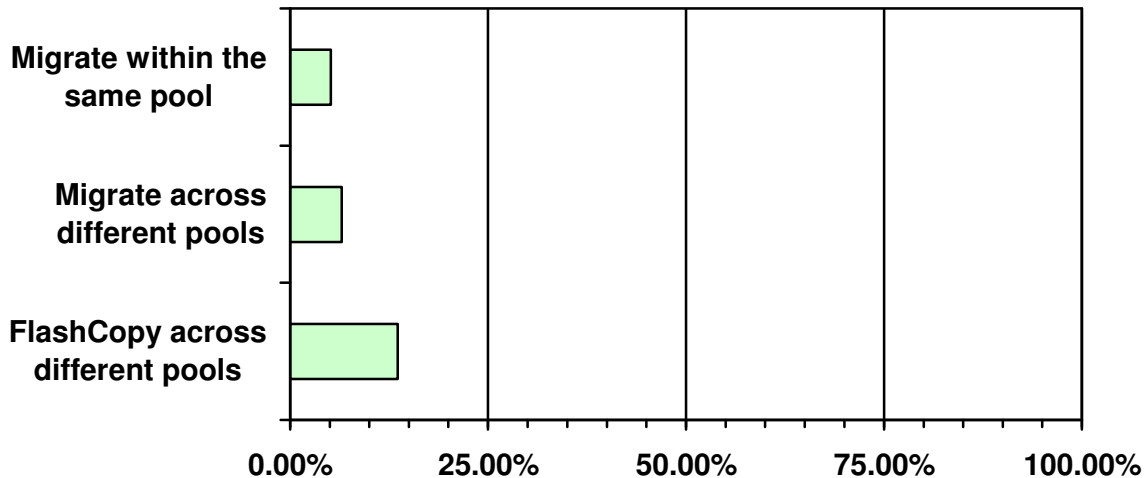


Figure 18: Impact to Host I/O during Volume Migration and FlashCopy.

## 5 Best Practices and Considerations

One of the only decisions a storage administrator needs to make when using Easy Tier Automatic Mode is how much SSD capacity is required and how the DS8700 should be configured. Once the storage pools are created with both SSD and HDDs, the rest of the storage management is done by the Easy Tier algorithms.

With the DS8700, you can exploit Easy Tier monitoring without SSDs to provide estimates of potential performance improvement. Release 5.1 LIC is required to enable this. You may use the advisor tool (described in section 2.3.1) to help with this analysis.

The SSDs should be spread one 16 DDM drive set at a time across the DA pairs. You should consider the lower capacity SSDs to spread SSDs at least across 4 DA pairs with all 8 DA pairs being optimal. Obviously, this is a lot of SSD capacity so it is fine to go with a small number of SSD drive sets and thus populating less DA pairs.

There are a few options when setting up the storage pools that you can choose from. The most obvious is two storage pools, one per DS8700 storage server, with  $\frac{1}{2}$  the SSD ranks and  $\frac{1}{2}$  the HDD ranks and capacities in each pool, with storage pool striping used for volume allocation. This configuration allows all volumes to benefit from SSD performance if one or more of their extents are hot enough to be placed on SSD. This configuration should suit the predominate number of customers.

Some customers may want two pools per storage server. One pool with SSD ranks and HDD ranks and the other with strictly HDD ranks. The HDD-only pool could be used for test volumes, to archive data where there is already in place archiving to disk, or production data that's not performance sensitive. The HDD-only pool can be of a different RAID type or HDD type (for example, 2TB SATA RAID-6 vs 600GB 15K rpm RAID-5 in the SSD+HDD pools).

Finally, there may be a very high performance workload where you want to guarantee SSD performance even if the activity is low (for example, only used at IPL time). In this case you may want to use some of your SSD ranks in SSD-only pools and do manual data allocation to these pools.

Not all workloads will benefit equally from Easy Tier. When considering what workloads to place in an Easy Tier pool, the most improvement will usually be seen on those extents with low to modest cache read-hit-ratios, high destage rates, and smaller transfer sizes that exhibit non-sequential (random) access patterns. Workloads with higher read percentages tend to show more disk response time improvement than write intensive workloads. However, Easy Tier has the capability to improve performance of any mix of reads and writes. Transactional applications such as order entry, financial services, ERP, and customer inquiry will tend to benefit most from Easy Tier. Batch oriented workloads will benefit less but mixtures of batch and transactional workloads should be well managed by the Easy Tier algorithms.

## 5.1 z/OS Considerations

When DS8700 Easy Tier is managing z/OS data in a mixed technology extent pool in Automatic Mode, sub-volume extents will be dynamically migrated between SSD and HDD according to their temperature. This activity is completely transparent to the z/OS application.

If there is a requirement to guarantee retention of selected z/OS data on SSD *regardless of its current temperature*, a separate extent pool of only SSD volumes must be defined and used to manually place this “special case” data. This is especially important when using the low latency of SSDs to optimize for mean time to recovery (MTTR). z/OS MTTR SSD candidates like the SYSRES volume (to reduce IPL time), restart and couple data sets (to reduce restart time), and paging data sets (to reduce dump time) will naturally have bursts of intense I/O activity separated by long periods of very low activity which would cause them to appear “cold” to Easy Tier and be migrated down to HDD if they were placed in a mixed technology extent pool.

If there is a requirement to place additional z/OS data on SSD for business reasons other than extent temperature as measured by Easy Tier, the Softek Data Mobility Console for z/OS (DMCzOS), FLASHDA, and other tools based on metrics provided in the SMF 42-6 and 74-5 records<sup>[5]</sup> are still available to assist in selection of appropriate SSD candidates at the dataset or volume level. Because these tools provide a system-wide view of data, they are also helpful in identifying candidates on other storage systems which could benefit from migration to SSD on the DS8700.

## 5.2 IBM Tivoli Storage Productivity Center (TPC) Considerations

Although Easy Tier appears to have a similar functionality like the TPC Storage Optimizer (i.e. the relocation of storage based on previously collected workload characteristics), the two will actually complement each other. While Easy Tier focuses on the performance-optimal placement of storage extents, the goal of the Storage Optimizer is the placement of volumes across storage pools to increase the overall utilization of the environment. Besides performance, the Storage Optimizer also takes into consideration additional aspects of storage management, such as capacity utilization. In future TPC releases, the Storage Optimizer can consider even more relevant characteristics, such as the costs of storage or the availability of copy services functions.

Future releases of TPC may also help adjust the usage of Easy Tier by tying together information about the configuration of the storage subsystem’s volumes and their respective performance. The data movement performed by Easy Tier will help the user find the optimal ratio of disks and make recommendations to add or remove faster/slower disks to a storage pool's configuration. This makes the most of the disks available in a subsystem while ensuring the performance is at least in the expected range for all storage volumes and is as optimal as possible.

## **6 Conclusions**

In the introduction, the authors predicted that Easy Tier (and the advances that follow up on it) will become as commonplace as paging or L1/L2 cache. Our evidence for this is the powerful performance improvement of Easy Tier, as presented in both the SPC-1 and DB2 Brokerage sections.

We have shown that with no need for manual tuning, Easy Tier provides over a three-fold increase in system performance. This not only allows applications to enjoy significant performance improvements, but also creates an opportunity to shift to higher capacity and less expensive storage media.

It is the shift to the use of higher capacity storage media which seems likely to make the Easy Tier capability universal within a small number of years. This shift can only lead to an acceleration of current trends toward lower storage costs and smaller footprints.

In this white paper, the authors have endeavored to present a thorough discussion of Easy Tier performance as well as planning considerations. Nevertheless, we must also emphasize, based upon our own experience, the simplicity with which this feature can be deployed and the automatic, hands-off way that it runs. This is made possible by the unique fine-granularity architecture of Easy Tier, and cannot be obtained with competitive offerings. It would be hard to imagine a better combination of high impact, combined with a low investment of effort by the storage administrator.

## **7 References**

- [1] La Frese, L., Lin, A. W., Martin, J., Williams, S. E., and Xu, Y. "IBM® System Storage™ DS8700™ Performance Whitepaper." December 2009.
- [2] La Frese, L., Sutton, L., and Whitworth, D. "IBM® System Storage DS8000® with SSDs: An In-Depth Look at SSD Performance in the DS8000." April 2009.
- [3] Roll, M. "Understanding Storage Performance: Concepts, Issues and FAQ." 2006.
- [4] Ripberger, R. and Xu, Y. "IBM System Storage, DS8000 Storage Virtualization Overview, Including Storage Pool Striping, Thin Provisioning, Easy Tier", WP101550 V2.0, May 2010.
- [5] Altman, J., Sutton, L., and Sutton, P. z/OS Hot Topics, article 20-48: z/OS Support for Solid State Drives in the DS8000. February 2009.

## **8 Frequently Asked Questions**

### **Q: Do I need SSDs installed to run the Easy Tier Storage Advisor Tool?**

A: No. any DS8700 with R5.1 code installed can do Easy Tier monitoring. Use the DSCLI or GUI to produce a file that the advisor tool (running on any Windows based workstation) will post-process to produce a report.

### **Q: How can I assure that a particular volume remains provisioned by SSDs with Easy Tier running?**

A: In a merged extent pool with both SSDs and HDDs, Easy Tier will manage which extents reside on which storage. If you have the need to manually place volumes on SSDs you will need to create an extent pool that only contains SSDs.

### **Q: Should I use Easy Tier in conjunction with SAN Volume Controller (SVC)?**

A: There is no reason not to. However, it is likely you would want to avoid mixing volumes from Easy Tier managed extent pools with other extent pools in a managed disk group.

### **Q: Are there any special considerations when using Easy Tier in an SVC environment?**

A: No. However, when introducing Easy Tier you may also introduce the use of storage pool striping at the same time. Storage pool striping always stripes at a granularity of 1 GiB. For this reason, we recommend that when SVC MDisks are storage pool striped, an SVC extent size be adopted of 256 MiB or smaller. This helps to avoid any effect where striping at one level undoes the effect of striping at another.

### **Q: Is there any reason not to order the Easy Tier feature for a DS8700?**

A: Since Easy Tier is a no-charge feature, it is likely that anyone purchasing or upgrading a DS8700 with R5.1 would want to have the Easy Tier feature, even if they currently have no SSDs. The feature would enable Easy Tier monitoring to evaluate potential future deployment of SSDs.

### **Q: Is greater cache in the DS8700 desired if you run Easy Tier?**

A: No. Easy Tier in itself places no additional demand on cache. Easy Tier may enable higher throughputs on the DS8700 via latent demand but it is unlikely that this will have any noticeable effect on cache behavior.

### **Q: What happens if I copy a volume from one address to another within an extent pool?**

A: Once completed, the copy would become managed by Easy Tier and any hot extents would be moved to SSDs over time.

### **Q: If I am running remote mirroring on DS8700, should I run Easy Tier on both the primary and secondary storage?**

A: It is recommended but not required. If you have similarly configured storage at both locations, running Easy Tier will enable balanced performance improvement. However, in the event of a site switch, there may be additional learning required due to the read activity that is introduced.

**Q: With Easy Tier Manual Mode, are there any special considerations to re-stripe a pool using rotate extents after adding new ranks to the pool?**

A: If the original pool was very full, a two-step process may be needed. First you should move a few volumes within the pool to free up some space on the full ranks. These volumes will likely be completely moved to the new ranks. Then you should move all of the volumes, including the ones moved in the first step. On the second move, the volumes will re-stripe evenly across all of the ranks in the expanded pool. This process extends the performance benefits of Storage Pool Striping after adding capacity to a pool.

**Q: The SPC-1 submission for Easy Tier shows 33K IOPS. This seems to be just an ordinary submission of a middle-of-the-road result and is way less than what we published for SVC + DS8700. What is going on here?**

A: Many SPC-1 benchmark tests use very large numbers of drives (sometimes thousands). For example the SVC+DS8700 submission used 2048 fibre channel disk drives. The objective of this test, however, was to demonstrate what could be accomplished with a moderate number of ordinary SATA drives. Our test used only 12 SATA ranks (96 drives).

This SPC-1 submission is the first based upon the use of SATA drive technology. The result shows that Easy Tier enables an effective new way to deploy a combination of SATAs and SSDs into an autonomic tiered system.

The chart on page 25 of the SPC-1 Full Disclosure Report shows the effect of using EasyTier. Over a period of about 18 hours, the throughput of the 96 SATA drives increases from about 15K (with completely unacceptable response times) to over 50K (with response times that are at least tolerable). The remainder of the run demonstrates our ability to deliver 33K of throughput at a response time of less than 4 milliseconds.

**Q: Does an Easy Tier hybrid volume look just like an SSD volume to DFSMS?**

A: Currently DFSMS has no special awareness of Easy Tier hybrid volumes and does not get notification when selected extents from a logical volume have been migrated between HDD and SSD under Easy Tier's control because of their temperature. Although the hybrid volumes may appear to be standard HDD volumes as reported by z/OS software, whenever a DFSMS Storage Group contains volumes configured from extent pools with both SSDs and HDDs, Easy Tier will automatically manage extent placement for best performance.

**Q: Are heat map analysis tools such as the advisor tool available on a DS8300?**

A: The exact tools mentioned in this paper are not available on the DS8300, but heat map evaluation can still occur. This is useful for DS8300 clients who are interested in the Easy Tier feature on the DS8700 and are curious how it would improve their current environment. Consult your local IBM account team for availability and assistance for enabling this function on an existing DS8300.

## 9 Appendix

### 9.A Appendix A: Workload Characteristics

- *DB Open*: 70% reads, 30% writes, 50% read hits. This workload is designed to be comparable to typical online transaction processing applications, also referred to as OLTP. Read/Write Ratio = 2.33, Read Hit Ratio = 0.50, Destage Rate = 17.2%, Transfer size = 4 KB.
- *OLTP workload*: simulates the workload of transaction processing systems that require small, mostly random, read and write operations (for example, database systems, OLTP systems, and mail servers). It resembles the mix I/O workload components as defined in the SPC-1 specification.

## **9.B Appendix B: DS8700 Hardware Configurations**

### **9.B.1 Configuration for SPC-1 Measurements**

- 96x1TB SATA (RAID-10), 16x146 GB SSD (RAID-5).
- 256 GB cache, 12 FC paths on 6 Host Adapters.
- SPC-1 workload generator ran on a P770 (AIX 6.1.3.0) with 12 FC ports on 11 Host Bus Adapters (8 Gbps HBAs but running at 4 Gbps).
- Measurements included an extended SPC-1 "warmup" of 21 hours to demonstrate the effect of Easy Tier data migration.

### **9.B.2 Configuration for DB2 Brokerage Measurements**

- HDD Configuration: RAID-5, 128 300GB/15K HDDs; 8 3TB volumes were allocated for database, Temp files and Data Generation 8x50GB volumes are allocated to log files.
- SSD Configuration: RAID-5, 16 or 32 146GB SSDs
- DB2 Configuration: DB2 9.7 FP1, 4 Instances, 4 DBs at 2TB each, 4 Buffer Pools at 54GB each.
- Server Configuration: P770 (AIX 6.1.3.0) , 8 Eight Core P7 (3GHz), 256 GB Memory, 16 4Gb FC Ports.

### **9.B.3 Configuration for Easy Tier Manual Mode Measurements**

- RAID-5 measurements were taken with 128 146GB 15K RPM drives.

## 9.C Appendix C: Definitions and Methodologies

- *IOPS*: Input/output operations per second.
- *RAID-5*: A popular RAID implementation that optimizes cost effective performance while emphasizing use of available capacity through data striping. RAID-5 provides fault tolerance for one failed disk drive. This scheme uses XOR parity for redundancy. Data is striped across all drives in the array and parity is distributed across all the drives.
- *RAID-10*: Combines two schemes: RAID-0 (data striping) and RAID-1 (mirroring). Volume data is striped across several disks and the first set of disk drives is mirrored to an identical set. Since redundancy is achieved through mirroring, there is no parity in RAID-10. RAID-10 optimizes high performance while maintaining fault tolerance for disk drive failures. It can tolerate at least one, and in most cases, multiple disk failures.
- *Storage Server*: Also referred to a Central Electronics Complex or CEC.
- *IPL*: Initial Program Load. The boot sequence that occurs on the CECs.
- *FlashCopy*: A Point-in-Time Copy feature. Enables one to create full volume copies of data in a storage unit.