Redpaper

Tim Bohn
Dave Pierson
Brian Haan

The Benefits of Optimizing OLTP Databases Using IBM eXFlash Solid-State Drives

Ensuring that business-critical data is available when needed is an ever-growing need in IT. Your systems must store massive amounts of data quickly and retrieve it efficiently. Simultaneously, you must use new technologies that can improve efficiency and take advantage of these technologies within limited budgets.

One measure of growing efficiency in recent years is CPU processing power, which far exceeds growth in disk input/output (I/O). For this reason, disk I/O is often the reason for bottlenecks in high-performance applications. One technology that addresses this issue by increasing I/O is solid-state drives (SSDs).

In this IBM® Redpaper™ publication, we describe tests that the IBM Software Group Competitive Project Office performed to determine the costs and benefits of using IBM eXFlash SSDs compared to using hard disk drives (HDDs). For this testing, we compared the 200 GB eXFlash SSD with the 15k RPM serial-attached SCSI (SAS) HDD.

Results of these tests indicate that eXFlash SSDs offer a dramatic increase in I/O throughput. The eXFlash SSD can provide more than 50 times the throughput of an HDD so that you can eliminate I/O bottlenecks in a database system at a much lower cost while maintaining performance levels.

This publication is directed at IT professionals and IT decision-makers, such as CEOs, CIOs, CFOs, practitioners, information architects, and current and prospective IBM clients and Business Partners.

This Redpaper provides details about the following topics:

Executive summary

The mission of the Competitive Project Office was to compare the relative performance and cost of the 200 GB eXFlash SSD with the 15k RPM SAS HDD.

For the database server, we used the System x3690 X5 with two Intel Xeon x7650 processors with eight cores/processor. The System x3690 X5 contained 128 GB of RAM, but only 64 GB were used for the database instance, and 20 GB were used for the database buffer pool.

Testing focused on a 100 GB online transaction processing (OLTP) database to determine how many HDDs are needed to match the performance of a single eXFlash SSD. This entire scenario was designed to test the required performance levels. We did not consider the extra storage capacity of the HDDs.

The team tested the HDDs by filling the internal drive bays with eight IBM 146 GB, 15k RPM, 6 GBps SAS HDDs attached to a ServeRAID M5015 SAS/Serial Advanced Technology Attachment (SAS/SATA) controller. For the SSD test, we used two HDDs and one eXFlash 200 GB SSD. In both cases, one HDD was used for the operating system and applications, and one HDD was used for the database logs.

As expected, we found the measured input/output operations per second (IOPS) and throughput MB per second (MBps) to be much higher with the SSD than with the HDD. The performance difference between the SSD and the HDD for the OLTP test was not as dramatic; however, the cost to use the SSD is lower than the cost of HDDs.

Balanced systems design 101

For database systems, a delicate balance exists between CPU processing power and the I/O throughput needed from the disk. Other factors are involved, such as memory. However, when you add more CPU processing power, you limit I/O because the processor waits for I/O throughput from the disk to proceed to the next instruction. As you add more to your I/O system, the system becomes CPU starved, as depicted in Figure 1.

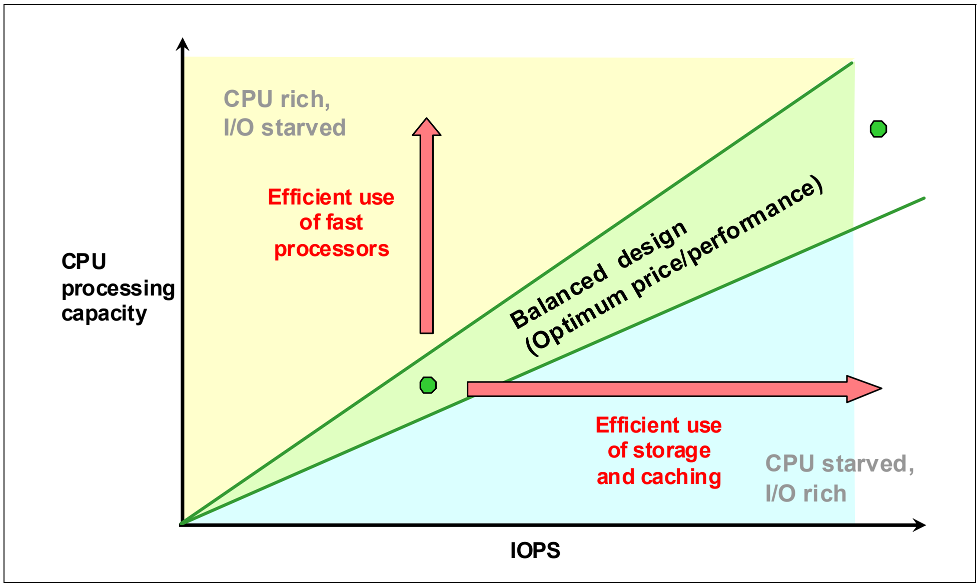

Figure 1 Correlation between CPU processing capacity and I/O throughput

Even a well-tuned database can still experience a substantial I/O wait time. During the time slice depicted in Figure 2, the processor waits at least half of the time for the I/O operations to be processed[1]. If the I/O response time decreases, minimizing processor wait time, the application can process many more operations. Furthermore, if the operations being measured are revenue-generating transactions, minimizing the wait time also helps you generate more revenue. In such cases, the processors in your expensive servers sit idle most of the time.

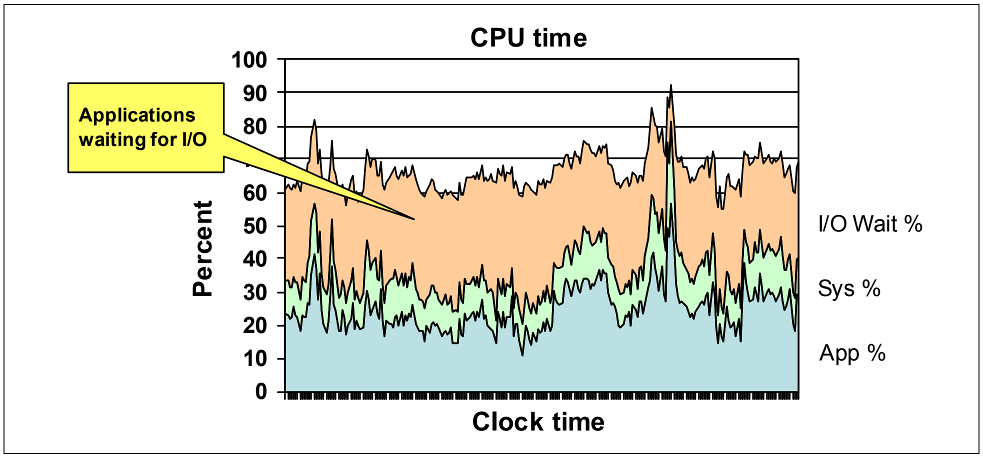

Figure 2 Percent of CPU time spent in I/O Wait compared with Sys % and App % times

Part of the problem with attaining a balance between CPU processing power and I/O throughput, is that growth in CPU processing power is far greater than the growth in I/O rates. The Intel 8088 had 29,000 transistors in 1979[2]. The Intel Xeon 7500 series processors had 2,300,000,000 transistors in 2009[3]. This rise in processing power provided a growth factor (increase in transistors) of nearly 80,000 times.

During this same time, disk I/O rates went from 0.9375 MBps, with the run length limited (RLL) attached drive technology of 1979, to 800 MBps, with modern 8 GB Fibre Channel-attached drives. This improvement resulted in a growth factor (increase in disk I/O rate) of approximately 850 times. The gap between CPU processing power and I/O rates rapidly increased over the last 10 years. This gap is a problem for fast processors that must wait for data to be moved off disks and into memory, and the gap continues to widen.

SSDs can help fill that gap. They can help ensure that critical data is moved to the processor or memory much more quickly. SSDs reduce I/O wait time by initiating and completing data operations much more quickly than spinning hard disks. For the highest level of performance, the goal is to keep the processor busy by reducing wait time and spending more time in running operations.

An SSD stores data uses flash memory, known as solid-state Negated AND, NOT AND (NAND) technology, instead of spinning disk within HDDs. NAND-based memory retains its memory even when the power is cut off. Additionally, NAND-based memory provides much faster access time, latency, and throughput. SSDs use the same interfaces that HDDs use, so you can easily replace HDDs with SSDs.

1 Here, the I/O Wait % is the time spent waiting for data from the storage system. App % is the time spent running user instructions in a program. System % is time spent managing database locks, shared memory, context switches in memory, and other elements, in support of user programs. See *IBM Information on Demand 2010* by Mike Barton and David Lebutsch, May 2010.

2 http://www.intel.com/technology/timeline.pdf

3 http://download.intel.com/pressroom/pdf/nehalem-ex.pdf

Currently, two key SSD technologies exist, as follows:

- Single-Level Cell (SLC) SSD

  Contains a binary state in each cell (therefore 1 bit per cell) and is faster, more reliable, and much more durable than the multi-level cell (MLC). SLC SSDs are also more expensive.

- Multi-Level Cell (MLC) SSD

  Doubles the capacity of the SSD by interpreting four states in each cell (therefore affecting 2 bits per cell), but typically has a considerably shorter life span. The internal design structure of the SSD can mitigate this limited time span by interleaving, changing how algorithms are written, and higher over-provisioning (adding more excess capacity) with which the wear-leveling algorithms can work.

MLC SSDs target the notebook computer market, and SLC SSDs target the enterprise market. IBM released a newer version of MLC, known as enterprise MLC (eMLC[4]), which offers increased performance and endurance in an MLC SSD.

Compared with a standard hard disk drive, both SLC and MLC SSDs have a lower write endurance, but that gap is narrowing as technology continues to improve in this area. Because SSDs are expensive compared with standard HDDs, they tend to be used when performance is the gating factor, rather than size, although SSDs are closing that gap. From 2000 to 2009, HDD capacity grew five-fold, from 180 GB to 2 TB. During that same time, SSD capacity grew 14-fold, from 14 GB to 200 GB, with capacities continuing to grow.[5]

Testing

Figure 3 shows the hardware and software used for the Competitive Project Office tests. We used an IBM System x3690 X5 with two Intel x7560 processors, 16 total cores, and 128 GB of RAM. We used 146 GB, 15k RPM SAS HDDs, and eXFlash 200 GB SSDs.

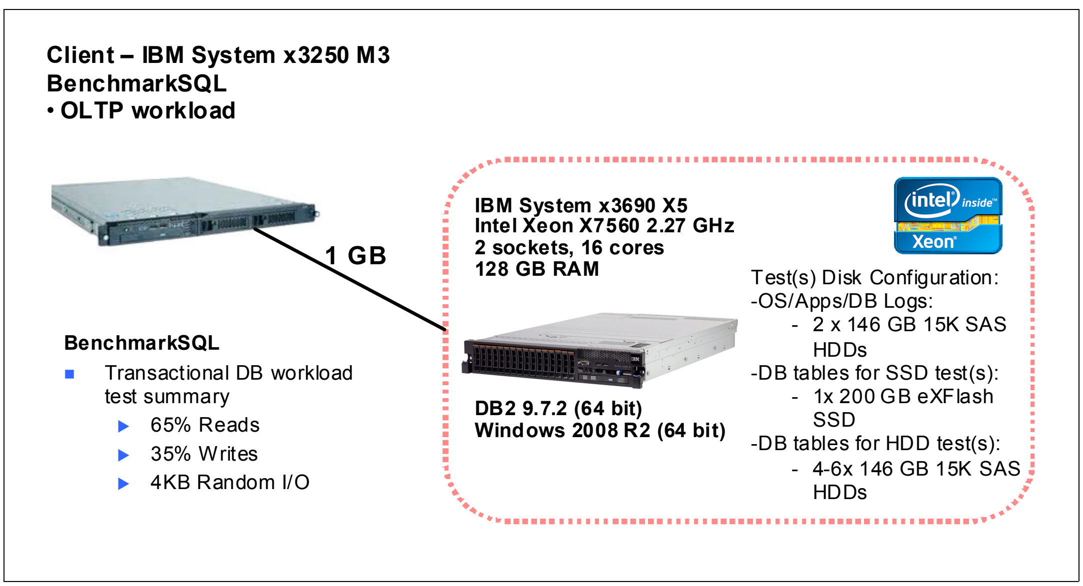

Figure 3 Hardware and software for testing

4 http://www.redbooks.ibm.com/abstracts/tips0792.html
5 http://www.storagesearch.com/bitmicro-art3.html

First, we tested to measure the input/output per second (IOPS) of the SSDs and HDDs to determine the expected difference. This test isolates the drives on the server. We used the IOMeter utility, running at 4k blocks, an 80/20 ratio of reads/writes, and 80% random/20% sequential.

Second, we conducted an OLTP workload test using BenchmarkSQL to simulate an OLTP application. This OLTP workload test ran as the driver on a client server against the System x3690 as the database server. The database application on the System x3690 X5 was the IBM DB2® Server Enterprise Edition, version 9.7.2.

Test results

We realize that measuring the different drives for IOPS is not a predictor of OLTP performance; however, we wanted to start by understanding what to expect from the HDDs and SSDs for IOPS and throughput. Running this OLTP workload test on the System x3850 using both types of drives provides a good measure of the relative difference. Many parameters can be configured, such as the transfer request size, percent random/sequential distribution, and percent read/write distribution. The team configured IOMeter with 4k blocks, an 80%/20% ratio of reads/writes, and 80%/20% random/sequential reads/writes for the tests to mimic the characteristics of a simple OLTP workload.

Comparing the results of a single HDD to a single SSD, the team saw a 50.7 times relative difference between the two. The maximum numbers for the HDD and the eXFlash SSD are shown in Figure 4.

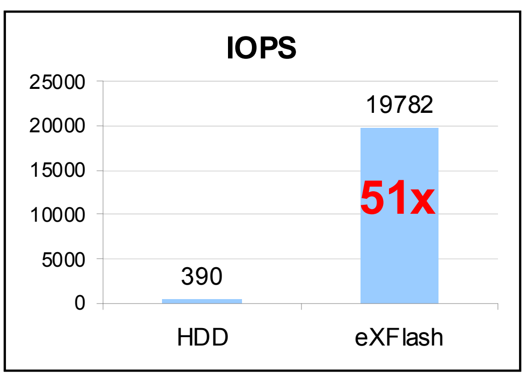

Figure 4 Maximum IOPS values for the HDD and eXFlash SSD

Measuring the data throughput in terms of MBps, you can see a similar difference. In Figure 5, the results show how the team achieved 51.5 times greater throughput with a single SSD than with a single HDD. Using either method of throughput measurement, we saw a tremendous boost with SSDs. Although these results do not hold constant for database throughput, they reflect how we can limit the I/O wait time for database processing using SSDs, thus increasing the transactions per second (TPS) that we want from OLTP processing.

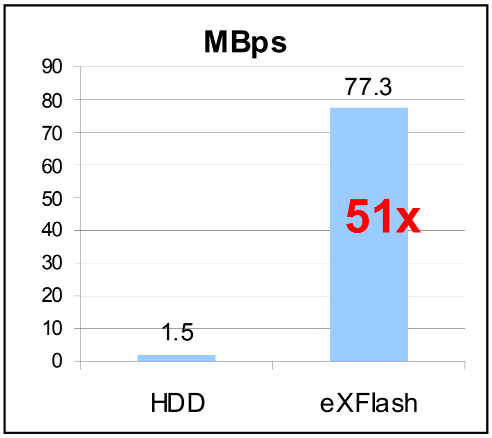

Figure 5 Maximum throughput values, in MBps, for the HDD and eXFlash SSD

Next, the team ran the OLTP test to compare the eXFlash SSD with multiple HDDs to come up with the same TPS.

For an I/O constrained OLTP database, many companies add HDDs to separate the database across many drives to alleviate the I/O bottleneck. The Competitive Project Office team wanted to see how many HDDs are required to lessen this bottleneck.

First, the team ran the OLTP test with the database on a single eXFlash SSD. The database was approximately 100 GB large, and the logs were placed on a separate HDD. With this setup, we achieved 1366 TPS.

Second, we removed the eXFlash drive and filled the server with six HDDs. The OS and logs were still on separate HDDs. We ran the same test, this time spreading the database across four of the HDDs, achieving 325 TPS. Then, we spread the database across the six HDDs and reran the test, reaching a TPS of 457.

This result means that, theoretically, if we continued adding internal HDDs to the System x3690 X5, and the TPS value scaled linearly as more drives were added, we need 20 HDDs to produce 1381 TPS. In other words, in the best case scenario, we need 20 HDDs to reach the same performance level noted with one eXFlash SSD, but can easily need more.

SSDs cost considerably more than HDD, so we calculated what the cost of using one eXFlash SSD compared with 20 HDDs. Table 1 presents the cost of 20 HDDs, but for the total cost, we must consider the real cost of the required storage enclosure and extra controller.

Table 1 Total cost of eXFlash SSD compared with HDDs[6]

Drive	eXFlash SSD[a]	HDD[a]
Drives	$3,199	20 x $439 ($8,780)
Controller		$249
Enclosure		$4,495
Total	$3,199	$13,524

a. Prices stated were retail prices as of February 2011 and are in US dollars. Prices can vary by country and will likely change over time as technology matures.

As shown in Table 1, the cost for the HDD solution can be 4.2 times greater than the cost of the eXFlash SSD. If we normalize the costs by the actual TPS for each medium, the eXFlash SSD is $2.234 per TPS, and the HDDs cost $9.79 per TPS. This data shows a 76% lower cost for the eXFlash SSD compared with the TPS using HDDs. These cost comparisons are depicted graphically in Figure 6.

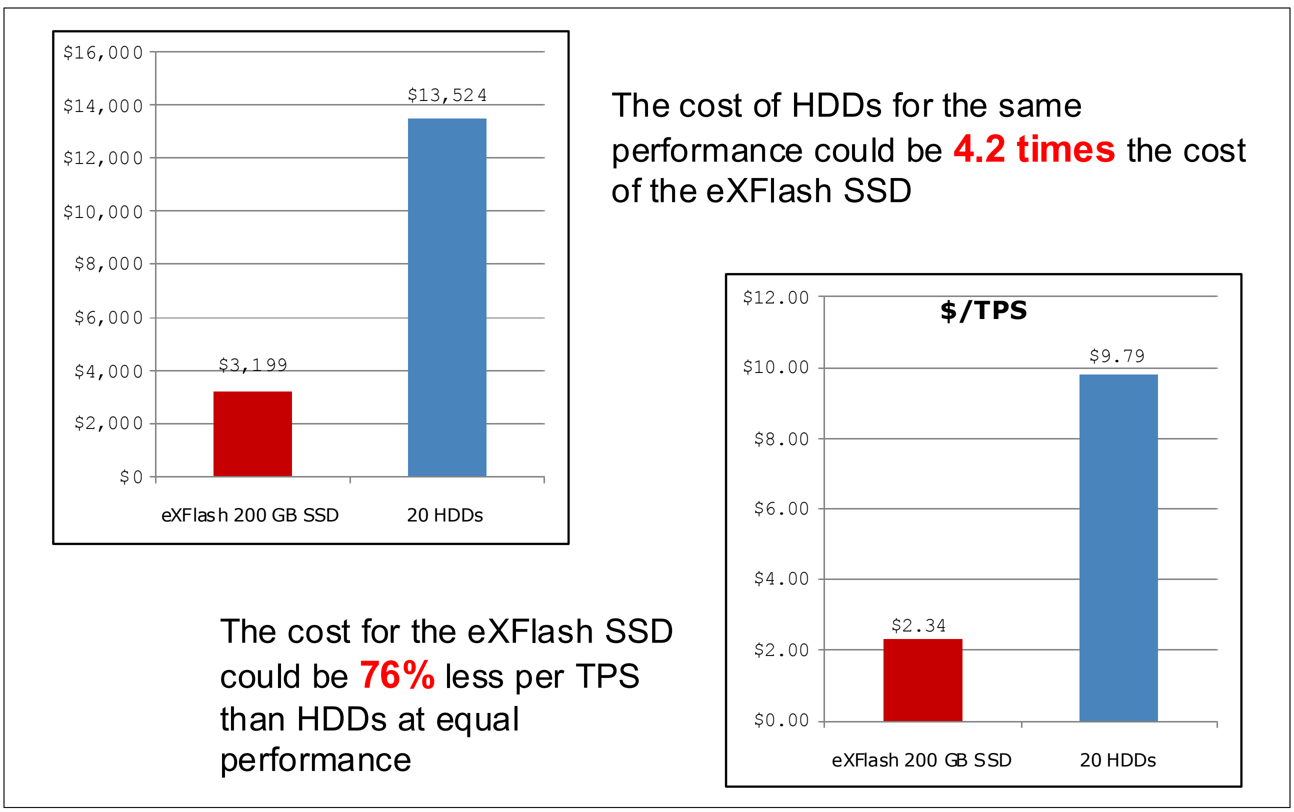

Figure 6 Cost comparison of SSDs and HDDs, by performance and cost per TPS

[6] Prices are retail as of February 2011. Current retail pricing is available using the online IBM shopping site at `https://www-304.ibm.com/shop/americas/content/home/store_IBMPublicUSA/en_US/evp/x3690X5.html` (for the System x3690 X5) and at `https://www-304.ibm.com/shop/americas/content/home/store_IBMPublicUSA/en_US/evp/x3850X5.html` (for the System x3850 X5). From these websites, you can view retail pricing, or build your own IBM System x® model to determine pricing for your environment.

Consider too that the costs in Table 1 and Figure 6 on page 7 are for the drives and storage system only. Over time, the SSD solution will save even more because SSDs do not take up extra rack space and use less power than the HDD solution.

Summary

The growth in CPU processing power far exceeds the growth in disk I/O. For this reason, disk I/O is the culprit in major bottlenecks in many high-performance applications. The eXFlash SSD can provide up to 51 times more throughput, making it possible for you to eliminate I/O bottlenecks in a database system. Compared with HDDs, the eXFlash SSD eliminates I/O bottlenecks at a much lower cost while maintaining the same level of performance, thus making it a more cost-effective workload-optimized system.

Specifically, For an OLTP database application, you need a balance between CPU processing power and I/O throughput. eXFlash SSDs offer a dramatic increase in I/O throughput. For more powerful servers, such as the System x3690 X5, adding eXFlash SSDs is a cost-effective way to add needed extra I/O throughput, rather than using multiple hard disks.

The team who wrote this paper

This paper was produced by a team of specialists from around the world.

Tim Bohn is an IBM Senior Certified IT Specialist and Open Group Master Certified IT Specialist with the IBM Software Group, Competitive Project Office (CPO). His current focus is the IBM System x platform hardware and software. Before joining the CPO team in 2010, Tim was the Global Community of Practice leader for the IBM Rational® Systems Development Community, where he helped customers with systems and software development practices. Before joining IBM Rational Software, Tim worked as a software and systems engineer for 15 years. Tim holds Bachelor's and Master's degrees in Computer Science from USC. He is a yearly guest lecturer at USC and UCLA.

Dave Pierson is a Software Engineer and member of the Competitive Project Office. His areas of expertise include the IBM System x product hardware portfolio and Enterprise Security. Dave recently worked with relational databases and optimized hardware to improve database performance. In his former roles, Dave worked on hardware/software integration test teams and worked extensively with IBM Business Partners around the world.

Brian Haan is a Consulting Software Architect in the IBM Competitive Technology Laboratory. He is the Competitive Project Office team lead, responsible for understanding IBM Information and Integration management products and their directions. Before joining the Competitive Technology team, Brian spent 10 years in IBM development, specializing in database systems. In this capacity, he worked with IBM WebSphere® Commerce Suite, Content Manager OnDemand, Printing Systems Integrations Services, and other development teams. Before joining IBM, Brian worked as a database and data communications consultant and also spent nine years working at Time Inc. in research and development.

Thanks to the following people for their contributions to this project:

- Randall W. (Randy) Lundin, Product Marketing, High End System x
- David Watts, IBM Redbooks® (ITSO), System x Specialist

Now you can become a published author, too!

Here's an opportunity to spotlight your skills, grow your career, and become a published author—all at the same time! Join an ITSO residency project and help write a book in your area of expertise, while honing your experience using leading-edge technologies. Your efforts will help to increase product acceptance and customer satisfaction, as you expand your network of technical contacts and relationships. Residencies run from two to six weeks in length, and you can participate either in person or as a remote resident working from your home base.

Find out more about the residency program, browse the residency index, and apply online at:

ibm.com/redbooks/residencies.html

Stay connected to IBM Redbooks

- Find us on Facebook:

  http://www.facebook.com/IBMRedbooks

- Follow us on Twitter:

  http://twitter.com/ibmredbooks

- Look for us on LinkedIn:

  http://www.linkedin.com/groups?home=&gid=2130806

- Explore new IBM Redbooks publications, residencies, and workshops with the IBM Redbooks weekly newsletter:

  https://www.redbooks.ibm.com/Redbooks.nsf/subscribe?OpenForm

- Stay current on recent Redbooks publications with RSS Feeds:

  http://www.redbooks.ibm.com/rss.html

Notices

This information was developed for products and services offered in the U.S.A.

IBM may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not give you any license to these patents. You can send license inquiries, in writing, to:
IBM Director of Licensing, IBM Corporation, North Castle Drive, Armonk, NY 10504-1785 U.S.A.

The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law: INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

Any performance data contained herein was determined in a controlled environment. Therefore, the results obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems, and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurements may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM websites are provided for convenience only and do not in any manner serve as an endorsement of those websites. The materials at those websites are not part of the materials for this IBM product and use of those websites is at your own risk.

IBM may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

COPYRIGHT LICENSE:

This information contains sample application programs in source language, which illustrate programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to IBM, for the purposes of developing, using, marketing or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. IBM, therefore, cannot guarantee or imply reliability, serviceability, or function of these programs.

This document REDP-4849-00 was created or updated on May 3, 2012.

Send us your comments in one of the following ways:
- Use the online **Contact us** review Redbooks form found at:
  ibm.com/redbooks
- Send your comments in an email to:
  redbooks@us.ibm.com
- Mail your comments to:
  IBM Corporation, International Technical Support Organization
  Dept. HYTD Mail Station P099
  2455 South Road
  Poughkeepsie, NY 12601-5400 U.S.A.

Redpaper™

Trademarks