



Introduction to Flash Express

Improving Availability with Flash Express

What is Flash Express?

Flash Express is an innovative solution designed to help improve availability and performance, during workload transitions, for improved quality of service. Flash Express can help reduce latency for critical paging that might otherwise impact the availability and performance of your key workloads. For paging flexibility and efficiency, Flash Express is a compelling addition to traditional auxiliary storage.

Flash memory has been integrated into the memory hierarchy to provide even higher levels of system availability and performance. Flash Express is designed to offer exceptional performance for paging spikes by reducing paging latency. Flash Express can be especially helpful during transitional workload processing shifts, where paging might surge, such as during start of day processing, changes in loads, or collection of SVC or standalone dumps.

Flash Express is designed to improve availability through improved paging performance at times when it matters most - like periods of peak paging during morning trading sessions or retail shopping days.

Planning and Configuring Flash Express

Flash Express is an IBM z Systems™ PCIe feature containing 4 solid state drives (two card pairs) which support 1.4 TBytes of Storage Class Memory and read/write cache located in the server Hardware System Area (HSA). Flash storage is integrated on PCI Express attached RAID10 cards which fit in the PCIe I/O expansion drawer.

A maximum of four card pairs are supported providing a maximum of 5.6 TB of storage. The amount of flash cache on the IBM z13s™ (z13s) is .5 GB and on the IBM z13™ (z13) is 4 GB.

Incremental Flash memory can be brought online or offline as needed using dynamic reconfiguration capabilities. A Flash Memory allocation panel on the SE specifies the amount of flash memory initially brought online to a z/OS® partition. In addition Flash memory may be virtualized across partitions (LPARS) for further flexibility.

Sizing Flash Express

Plan to assign the same amount of memory on Flash Express as defined for paging datasets on disk. Usually one pair of Flash Express features provides enough paging space for the entire z/OS partition. There is no need to perform detailed capacity sizing to plan for Flash Express. Adding Flash Express features to your auxiliary storage can improve paging performance and availability.

Flash Express is not persistent across IPLs, and can not be used for Virtual I/O or PLPA data used in warm starts.

Resiliency

Flash Express cards are delivered as a RAID10 mirrored pair for superior resiliency and reliability. In the unlikely event of a problem, Flash Express features can also be concurrently replaced. The features are designed for superior wear leveling and have a long expected lifetime.

Security

Your Flash Express data remains protected. Data is encrypted on the Flash Express adapter with 128 bit AES encryption. Encryption keys are stored on smart cards that are plugged into the SE. Removing the smart cards renders the data on the smart card inaccessible.

Benefits from Flash Express

IBM Flash Express helps organizations improve availability and performance especially during periods of paging spikes. Those using Coupling Facilities have a cost effective, resilient solution for overflow of WebSphere® MQ® shared queues.

Improving Availability

Flash Express can improve availability by reducing significant paging delays that might otherwise slow system performance and impact mission critical workloads.

Improving Diagnostic Time

During diagnostic collection, as in SVC or standalone dumps, systems can become sluggish effectively rendering key systems unavailable. When data is transferred into main memory as part of a dump, Flash Express' fast IO rates and low latency provide decreased first failure data capture time, and faster page-ins of the critical pages needed to create the dump. The HSA cache will also help during diagnostic collection. The system is able to return to normal workload performance faster, without delays. Flash Express offers a 19% reduction in total dump time for a 36 GB standalone dump¹.

Improving Paging

z/OS uses both Flash Express and page data sets for auxiliary storage. Wherever possible the system will page first to Flash Express resulting in faster performance. Especially for data intensive applications the use of pageable large pages with Flash Express enables the transfer of large amounts of data at faster speeds, which can result in improved performance for critical workloads like DB2®. Tests demonstrated up to 28% improvement² in DB2 throughput leveraging Flash Express with pageable large pages.

Improving Performance at Transition Times

Banks and financial institutions need highly responsive start of day performance. When the workload shifts from a transactional workload, say, from prime shift to batch and back, response time delays can occur. This is due to the required page-ins of critical work needed to resume transactional processing. These delays can be dramatically reduced when data for the next shift is transferred from flash storage into memory. The large number of page-ins could otherwise delay performance at "start of day" or "market open" activities, vital to operations like trading and banking. Flash Express enables 10X faster response time and 37% increase³ in throughput compared to disk for morning transition.

Reduce CPU Cycles

Flash Express works to reduce CPU cycles associated with page translations. Typically, page translations from virtual to real memory can impact the performance of workloads like DB2 or Java®. When using small pages (4K pages), paging is less efficient than paging using fewer larger 1 MB pages.

How it works

Cache buffers are used by the operating system to reduce virtual to real address translations. Performance of this translation can be improved through the use of a greater number of 1MB page entries in cache. As a result of improved cache hits, exploiters of pageable large pages and Flash Express should experience performance improvements both in elapsed time and CPU usage.

Flash Express Benefits Many Industries

Flash Express is useful for *any* industry needing improved service levels:

- Applications requiring high availability like bank, insurance, or trading applications
- Development and test teams that collect diagnostics frequently
- Service providers competing on superior performance
- Retail applications with a web presence
- Public sector applications requiring high availability, like emergency preparedness
- Any organization that needs high SLAs

Bottom line

- Flash Express is designed to improve availability and can reduce paging latency at critical times such as during morning transitions or other periods experiencing paging spikes.
- Flash Express can improve performance with pageable large pages, for instance with DB2 and Java workloads
- Flash Express can also reduce delays from SVC or standalone dumps
- Flash Express is automatically secured by 128 bit AES encryption for your security and compliance needs
- Flash Express is easily deployed and configured

© Copyright IBM Corporation 2016

IBM, IBM logo, DB2, MQ, WebSphere, z13, z13s, z Systems and z/OS are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries, or both. Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both. Java and all Java based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates. Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both. Other company, product and service names may be trademarks or service marks of others.

IBM assumes no responsibility regarding the accuracy of the information provided herein and use of such information is at the recipient's own risk. Information herein may be changed or updated without notice. IBM may also make improvements and/or changes in the products and/or the programs described herein at any time without notice.

References in this publication to IBM products and services do not imply that IBM intends to make them available in all countries in which IBM operates.

This information is provided "as is" without warranty of any kind, express or implied, and is based on IBM's current product plans and strategy, which are subject to change by IBM without notice. IBM shall not be responsible for any damages arising out of the use of, or otherwise related to, this document. Nothing contained in this document is intended to, nor shall have the effect of, creating any warranties or representations from IBM (or its suppliers or licensors), or altering the terms and conditions of the applicable license agreement governing the use of IBM software.

1, 2 and 3: all performance information was determined in a controlled environment. Actual results may vary. Performance information is provided "AS IS" and no warranties or guarantees are expressed or implied by IBM.

ZSL03189-USEN-05

