Université du Mainframe 2013

4-5 avril

Flash Express - Une nouvelle mémoire intelligente

Alain Maneville

Senior Certified I/T Specialist – zChampion

Vendredi 5 avril 2013 – 13H45-15H05

Université du Mainframe 2013

4-5 avril

Trademarks

The following are trademarks of the International Business Machines Corporation in the United States, other countries, or both.

Not all common law marks used by IBM are listed on this page. Failure of a mark to appear does not mean that IBM does not use the mark nor does it mean that the product is not actively marketed or is not significant within its relevant market.

Those trademarks followed by ® are registered trademarks of IBM in the United States; all others are trademarks or common law marks of IBM in the United States.

For a complete list of IBM Trademarks, see www.ibm.com/legal/copytrade.shtml:

*BladeCenter®, DB2®, e business(logo)®, DataPower®, ESCON, eServer, FICON, IBM®, IBM (logo)®, MVS, OS/390®, POWER6®, POWER6+, POWER7®, Power Architecture®, PowerVM®, S/390®, System p®, System p5, System x®, System z®, System z9®, System z10®, WebSphere®, X-Architecture®, zEnterprise, z9®, z10, z/Architecture®, z/OS®, z/VM®, z/VSE®, zSeries®

The following are trademarks or registered trademarks of other companies.

Adobe, the Adobe logo, PostScript, and the PostScript logo are either registered trademarks or trademarks of Adobe Systems Incorporated in the United States, and/or other countries.
Cell Broadband Engine is a trademark of Sony Computer Entertainment, Inc. in the United States, other countries, or both and is used under license therefrom.
Java and all Java-based trademarks are trademarks of Sun Microsystems, Inc. in the United States, other countries, or both.
Microsoft, Windows, Windows NT, and the Windows logo are registered trademarks of Microsoft Corporation in the United States, other countries, or both.
Intel, Intel logo, Intel Inside, Intel Inside logo, Intel Centrino, Intel Centrino logo, Celeron, Intel Xeon, Intel SpeedStep, Itanium, and Pentium are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.
UNIX is a registered trademark of The Open Group in the United States and other countries.
Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both.
ITIL is a registered trademark, and a registered community trademark of the Office of Government Commerce, and is registered in the U.S. Patent and Trademark Office.
IT Infrastructure Library is a registered trademark of the Central Computer and Telecommunications Agency, which is now part of the Office of Government Commerce.

* All other products may be trademarks or registered trademarks of their respective companies.

Notes:

Performance is in Internal Throughput Rate (ITR) ratio based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput that any user will experience will vary depending upon considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve throughput improvements equivalent to the performance ratios stated here.

IBM hardware products are manufactured from new parts, or new and serviceable used parts. Regardless, our warranty terms apply.

All customer examples cited or described in this presentation are presented as illustrations of the manner in which some customers have used IBM products and the results they may have achieved. Actual environmental costs and performance characteristics will vary depending on individual customer configurations and conditions.

This publication was produced in the United States. IBM may not offer the products, services or features discussed in this document in other countries, and the information may be subject to change without notice. Consult your local IBM business contact for information on the product or services available in your area.

All statements regarding IBM's future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only.

Information about non-IBM products is obtained from the manufacturers of those products or their published announcements. IBM has not tested those products and cannot confirm the 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.

Prices subject to change without notice. Contact your IBM representative or Business Partner for the most current pricing in your geography.

AGENDA

- **Introduction sur la mémoire FLASH**
- **Introduction sur FLASH EXPRESS**
- **Utilisation de FLASH EXPRESS**
- **FLASH EXPRESS en action**
- **Paramétrage z/OS**
- **Les états RMF**
- **Les chiffres de performance (2012)**

Introduction sur la mémoire FLASH

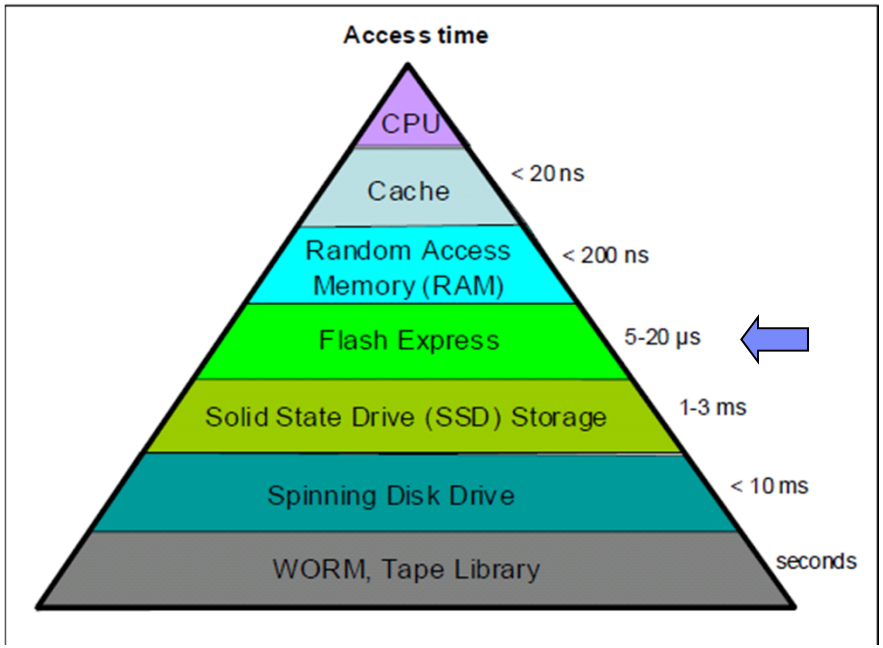
Introduction sur la mémoire FLASH

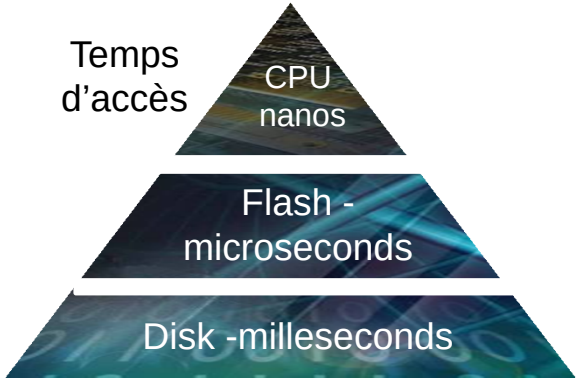
- **Technologie de mémoire non volatile**
- **Introduite il y a des dizaines d'année en informatique**
- **Utilisée aujourd'hui dans (exemples):**
  - Clés USB
  - Solid State Disks (SSD)
- **Des amélioration récentes en font une option performante (coût/vitesse) pour une utilisation au niveau de l'entreprise**
- **Un SSD**
  - Unité de stockage de données qui utilise des assemblages de circuits intégrés pour stocker des données de manière permanente.
  - Utilisation d'interfaces I/O compatibles avec les disques traditionnels
  - Un SSD n'utilise pas de mécanisme de mouvement de données (c'est la différence avec les Hard Disk Drives)
  - Pas de temps de « seek » ou de latence rotationnelle, donc meilleure performance.
    - Les temps de latences sont de 10 à 50 fois inférieurs aux meilleurs disques HDD.

Introduction sur FLASH EXPRESS

Introduction sur FLASH EXPRESS

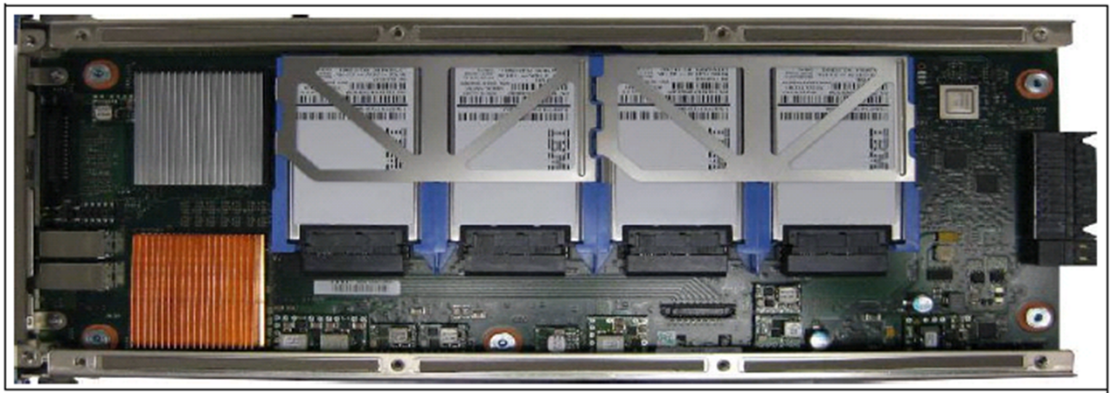
- **Flash Express introduit la technologie Solid State Drive (SSD) sur l'IBM zEnterprise EC12 en implémentant des Flash SSDs montés dans les cartes PCIe Flash Express feature.**
- **Intérêt de cette mise en œuvre:**
  - Amélioration de la disponibilité et de la performance
  - Meilleur qualité de service (et garantie de celle-ci) pour les travaux critiques qui ne peuvent pas supporter du « Paging »
  - Améliore aussi la vitesse de diagnostic (DUMPs) et en minimise ses effets.
- **Nouvelle hiérarchie de mémoire sur le zEnterprise EC12**

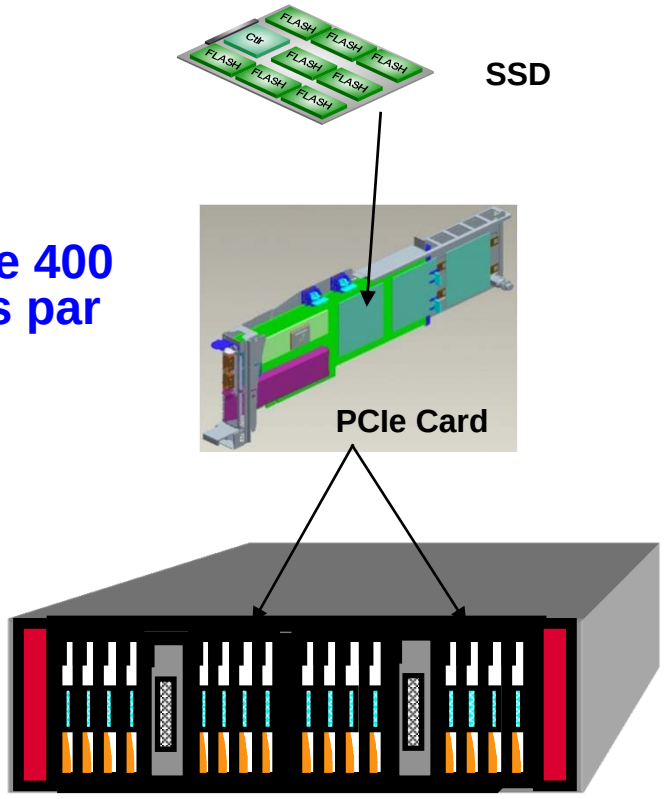




Introduction sur FLASH EXPRESS

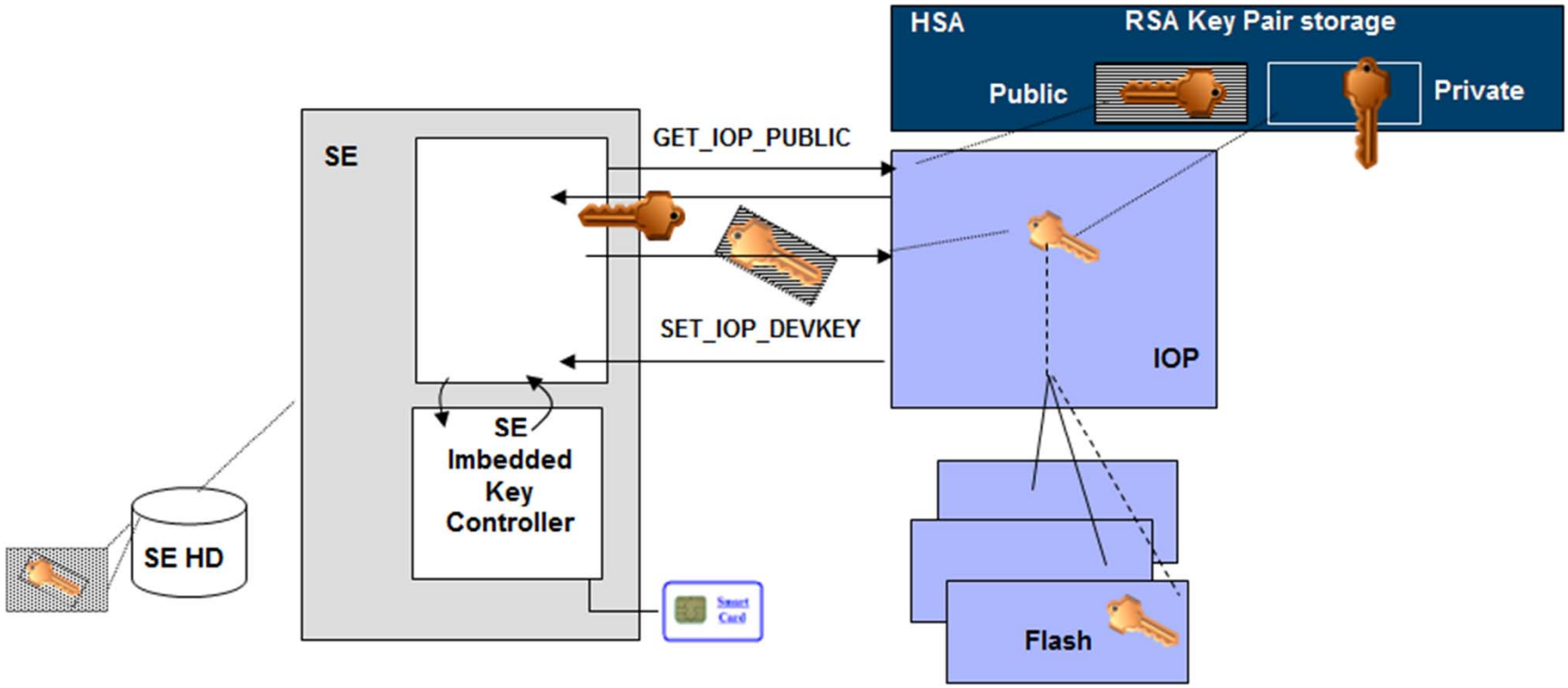
- **Flash Express est un dispositif optionnel**
- **Carte dans le PCIe I/O Drawer (en mix avec les autres cartes)**
  - Slots 1 et 14 réservés
- **Deux dispositifs minimum (FC#0402) et 8 maximum**
  - RAID 10 Mirrored Pair
- **Affectée à un PCHID (même s'il n'y a pas de ports)**
- **Pas de définition HCD**
  - Utilisation de SubChannel réservés dans les SS(0)
- **Une carte Flash Express PCIe adapter intègre 4 cartes de 400 GB pour un total utilisable de 1.6TO – toujours installées par paire – (1.4TB utilisable soit 1424GB)**





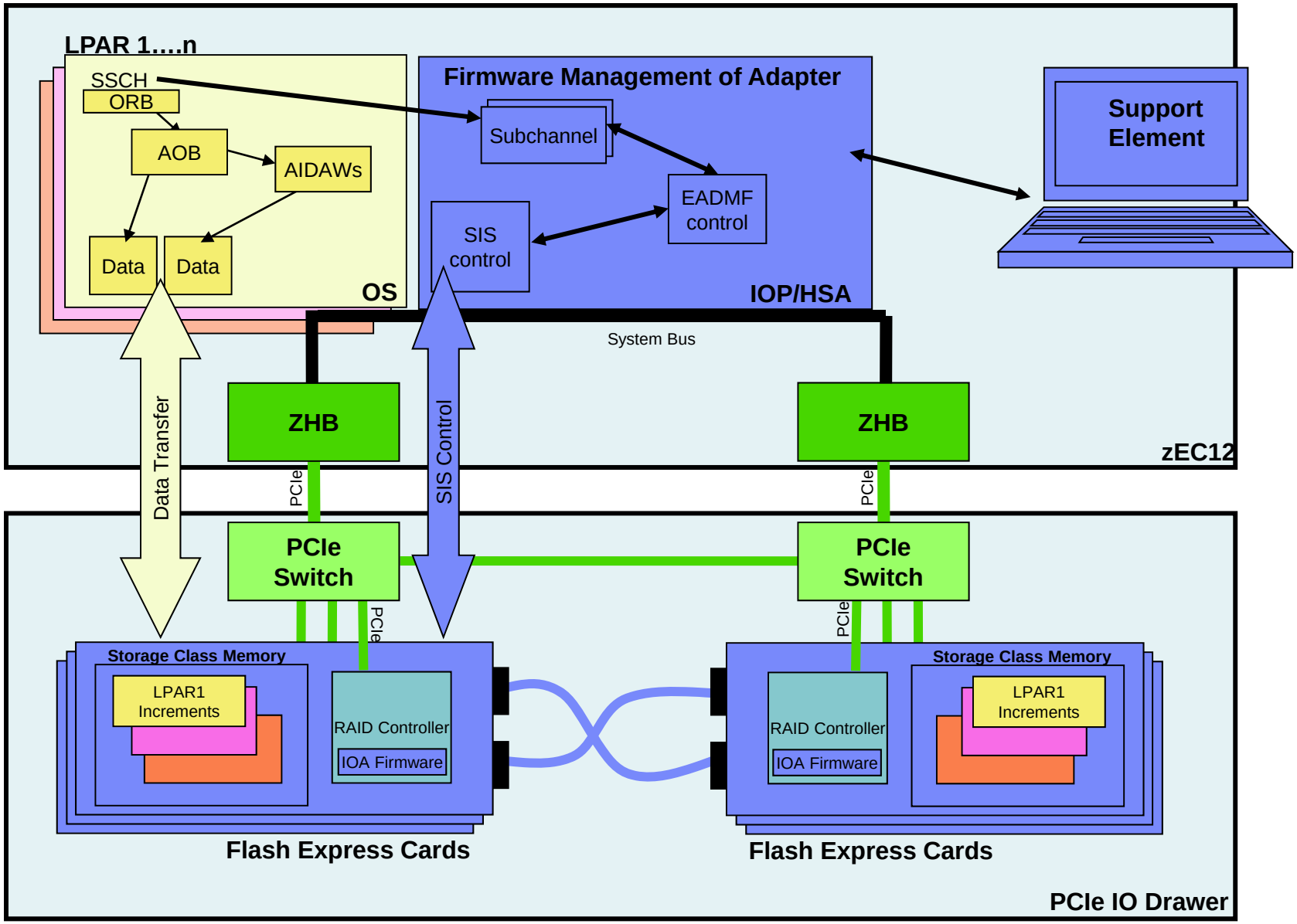
Securité des données avec Flash Express

- **Cette carte interne peut être utilisée pour la Pagination, les Dumps et….**
  - Elle peut contenir des données, incluant des données personnelles etc.
- **Les données Client sur cette carte sont protégées par du 128 bit AES Encryption**
  - Réalisé à partir du Hardware Encryption comme IBM's Disk and Tape encryption
- **La gestion de la clé est basée sur une Smart Card**
  - Le lecteur Smart Card Reader est installé sur le Support Element



Introduction sur FLASH EXPRESS

- **Diagramme général de mise en œuvre**



Introduction sur FLASH EXPRESS

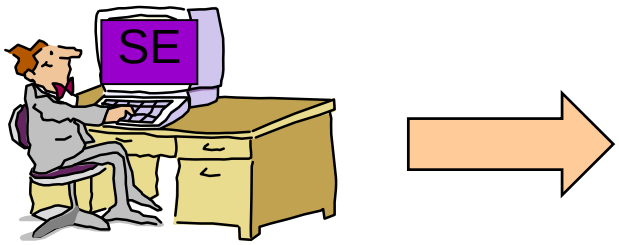
- **Valeur**

Client Value		Flash Capability
▪ Clients get business value from Continuous Availability ▪ Clients have max times for unavailability (few seconds) that are critical for the success of their enterprise	 ***Reliability***	▪ Improved z/OS recovery and diagnostic times ▪ Handling of workload shifts ▪ Coping with Dynamic Environments
▪ Clients see high per-MIPS costs mainly through software pricing ▪ Clients have elapsed time objectives for CPU bound workloads	 ***CPU Performance***	▪ Enables use of pagable large pages yielding 1-5% CPU performance benefit
▪ Clients over-configure IO infrastructure to support large dynamic capacity changes ▪ Clients are constrained by performance rules for paging IO	 ***IO Performance***	▪ Offloads GBs per sec of random IOs from IO Fabric ▪ Block & Predictive Paging
▪ Clients depend on IBM to keep their investment in System z transparently current over technology changes	 ***Technological Trends***	▪ Flash is an enabling technology for future middleware capability like cooperative memory management

Utilisation de FLASH EXPRESS

Utilisation de FLASH EXPRESS

- **Allocation de Flash Express à une partition (Initial)**
  - Attention à la taille réelle disponible pour la LPAR
  - Dans ce cas, le paramètre PAGESCM n'a pas été précisé (ALL par défaut)
    - Le total est de 1424GB
  - 64 GB on déjà été alloués
  - Il reste 1424-64= 1360



HMCZ12B: Manage Flash Allocation - Mozilla Firefox

172.26.2.182 https://172.26.2.182/hmc/content?taskId=18&refresh=84

Manage Flash Allocation - HPIB1

Summary

Allocated:	64 GB	Storage increment:	16 GB
Available:	1360 GB	Rebuild complete:	0 %
Uninitialized:	0 GB		
Unavailable:	0 GB		
Total:	1424 GB		

Partitions

--- Select Action ---

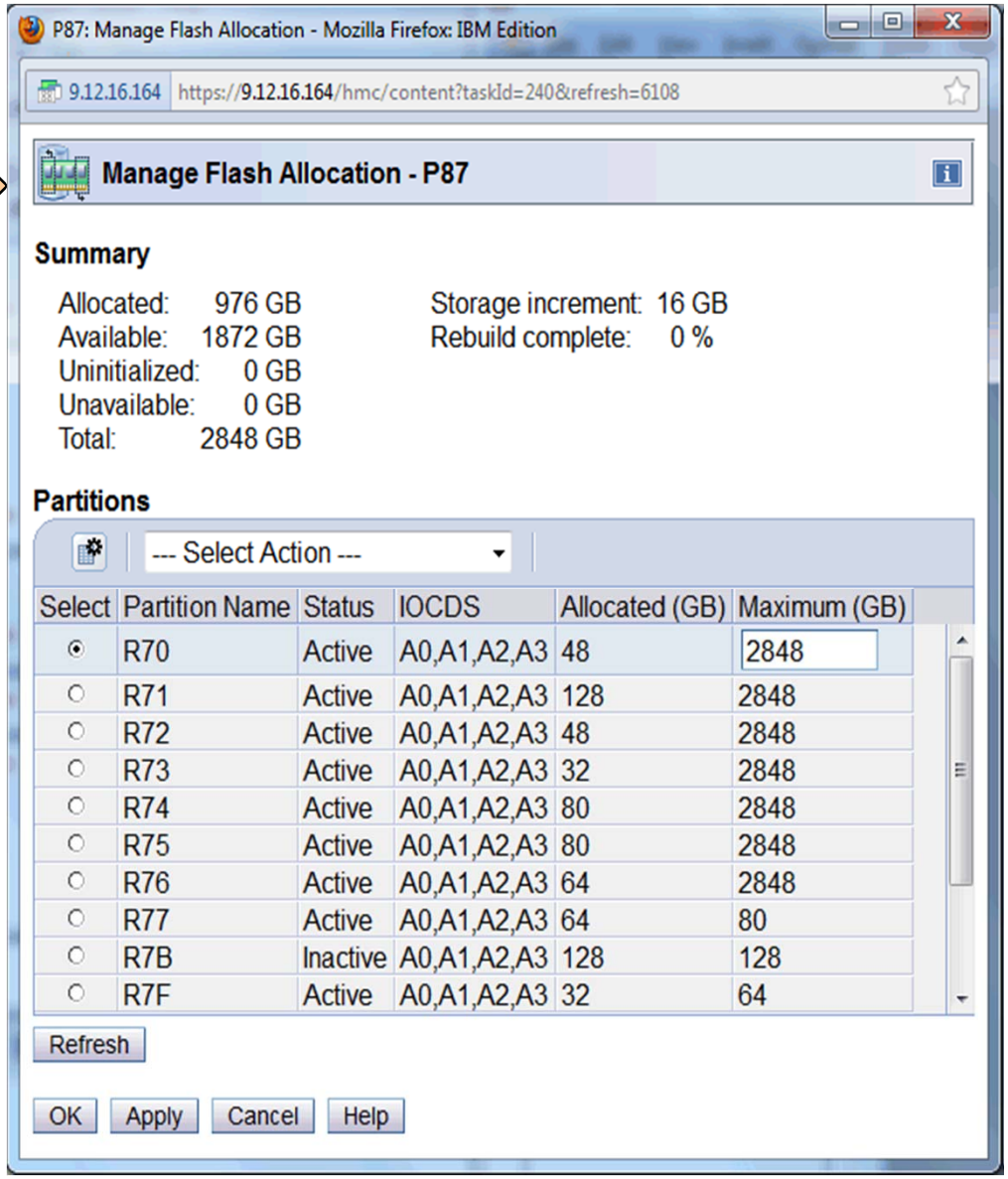
Select	Partition Name	Status	IOCDS	Allocated (GB)	Maximum (GB)
◉	B1C2P1OS	Active	A0,A1,A2,A3	32	32
○	B1C2P2OS	Active	A0,A1,A2,A3	32	32
○	B1C2P3OS	Inactive	A0,A1,A2,A3	32	16
○	B1C2P4OS	Active	A0,A1,A2,A3	32	16

Refresh

OK Apply Cancel Help

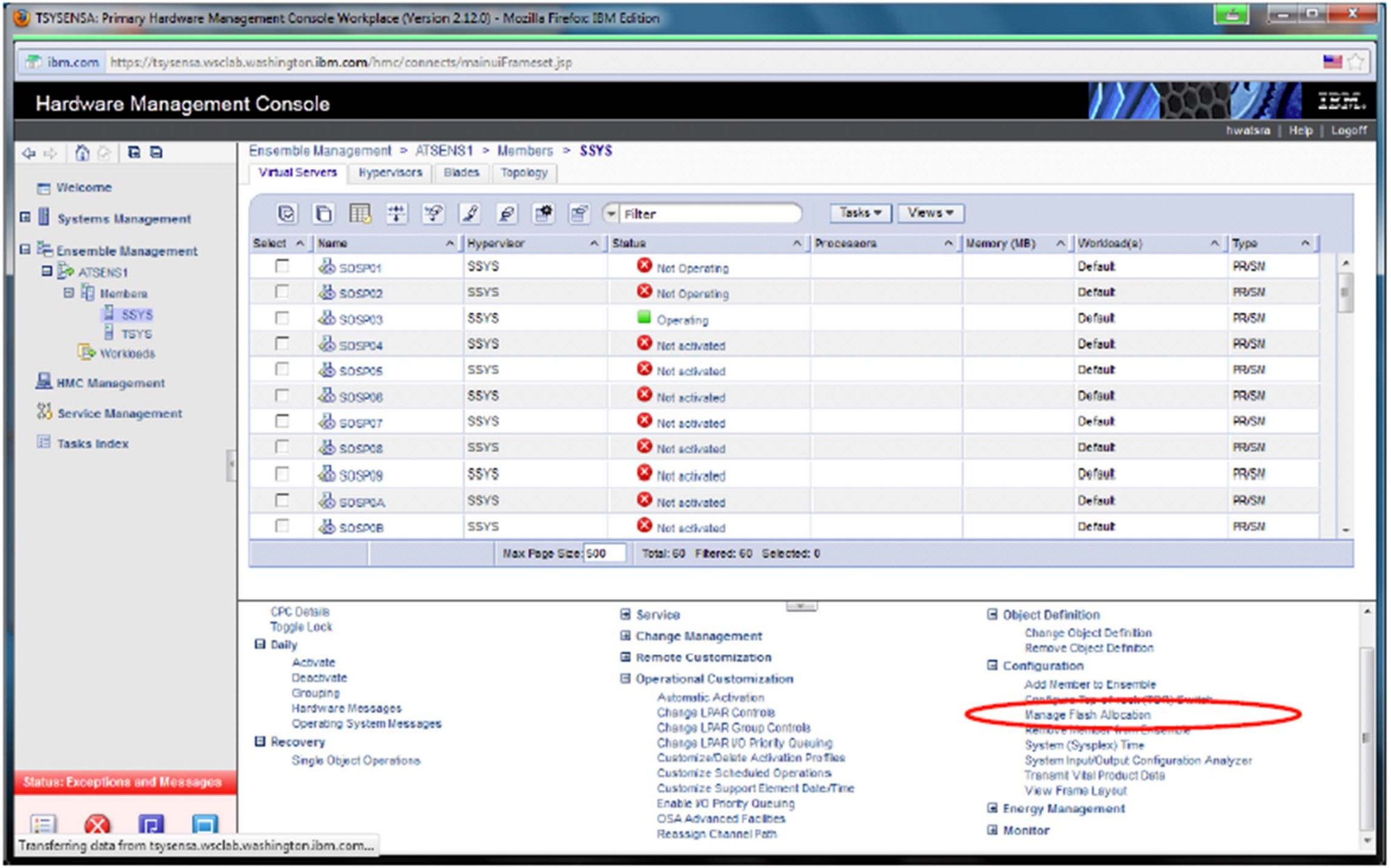
Utilisation de FLASH EXPRESS

- **Allocation de Flash Express à une partition**

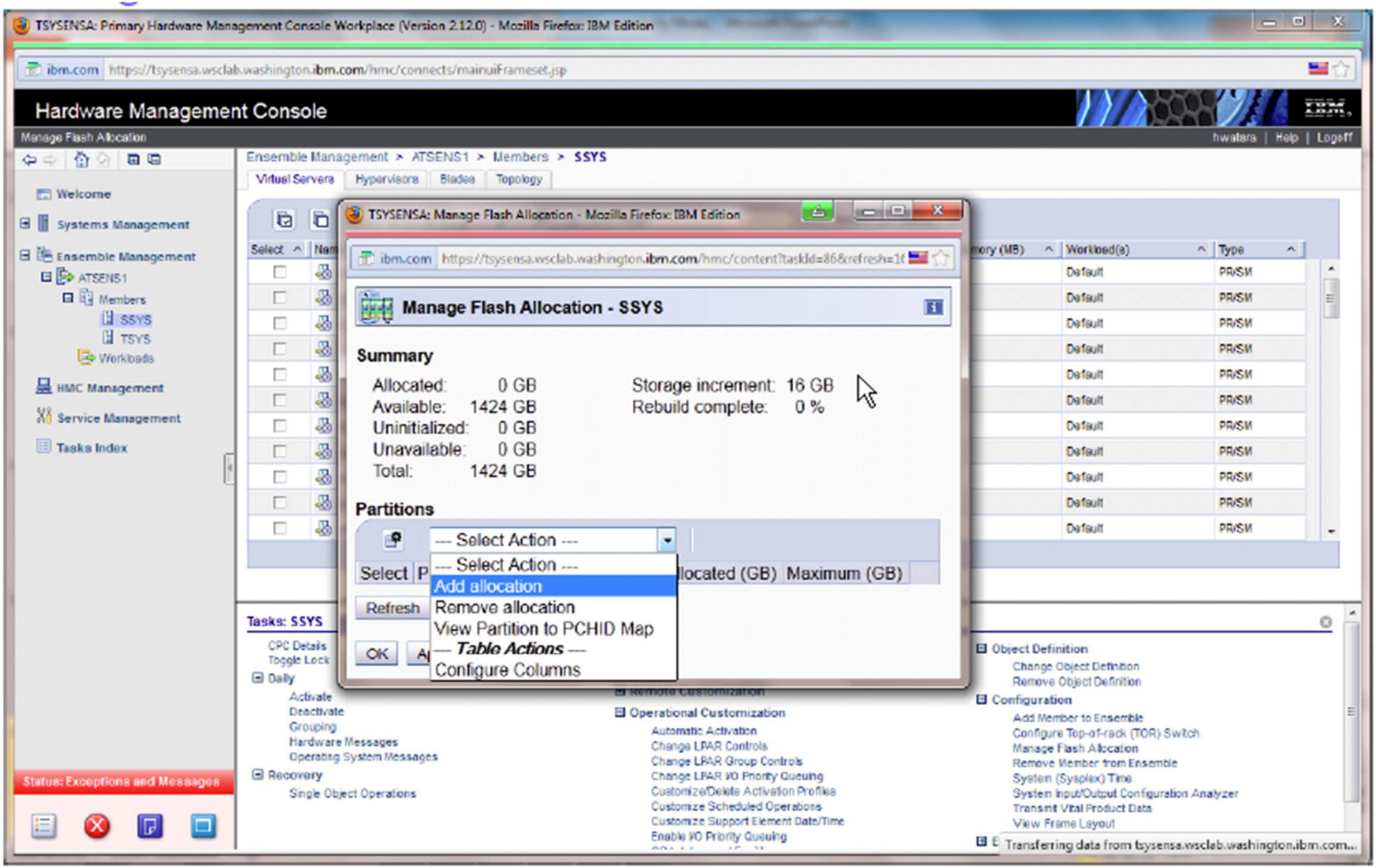


  - La quantité initiale et maximum de mémoire Flash disponible pour une LPAR est indiqué par le SE ou HMC par un nouveau Flash Memory Allocation panel
  - Quantité modifiable dynamiquement pour une LPAR.
  - Plus de Flash Memory (jusqu'au maximum permis) peut être configuré ONLINE pour une LPAR via SE ou HMC
    - Pour z/OS ceci peut être fait par une commande operateur
  - Peut dynamiquement configurer Flash Memory OFFLINE pour une LPAR via SE ou HMC
    - Pour z/OS ceci peut être fait par une commande operateur
  - Subchannels prédéfinis – pas d'IOCDS

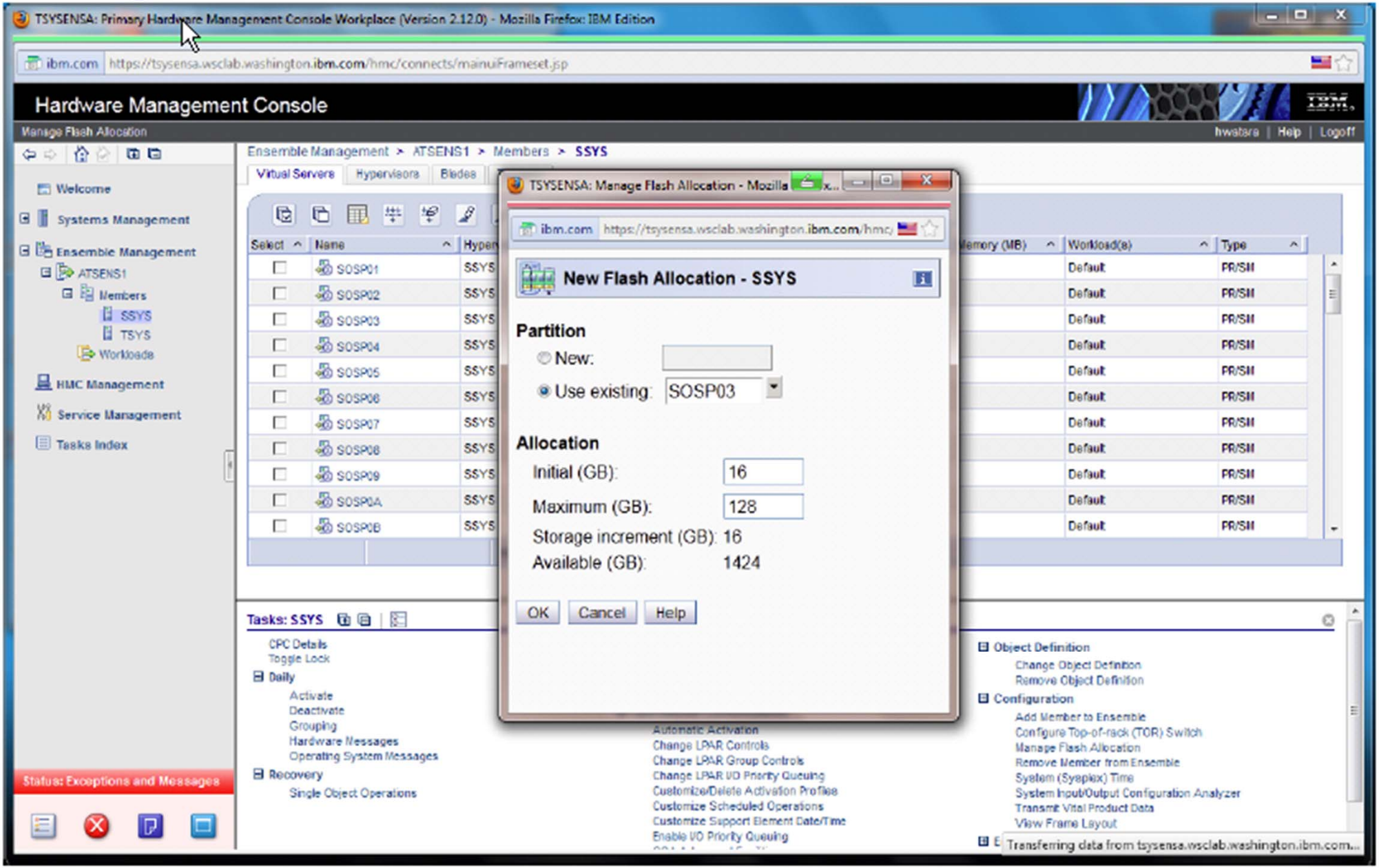
Utilisation de FLASH EXPRESS



Utilisation de FLASH EXPRESS

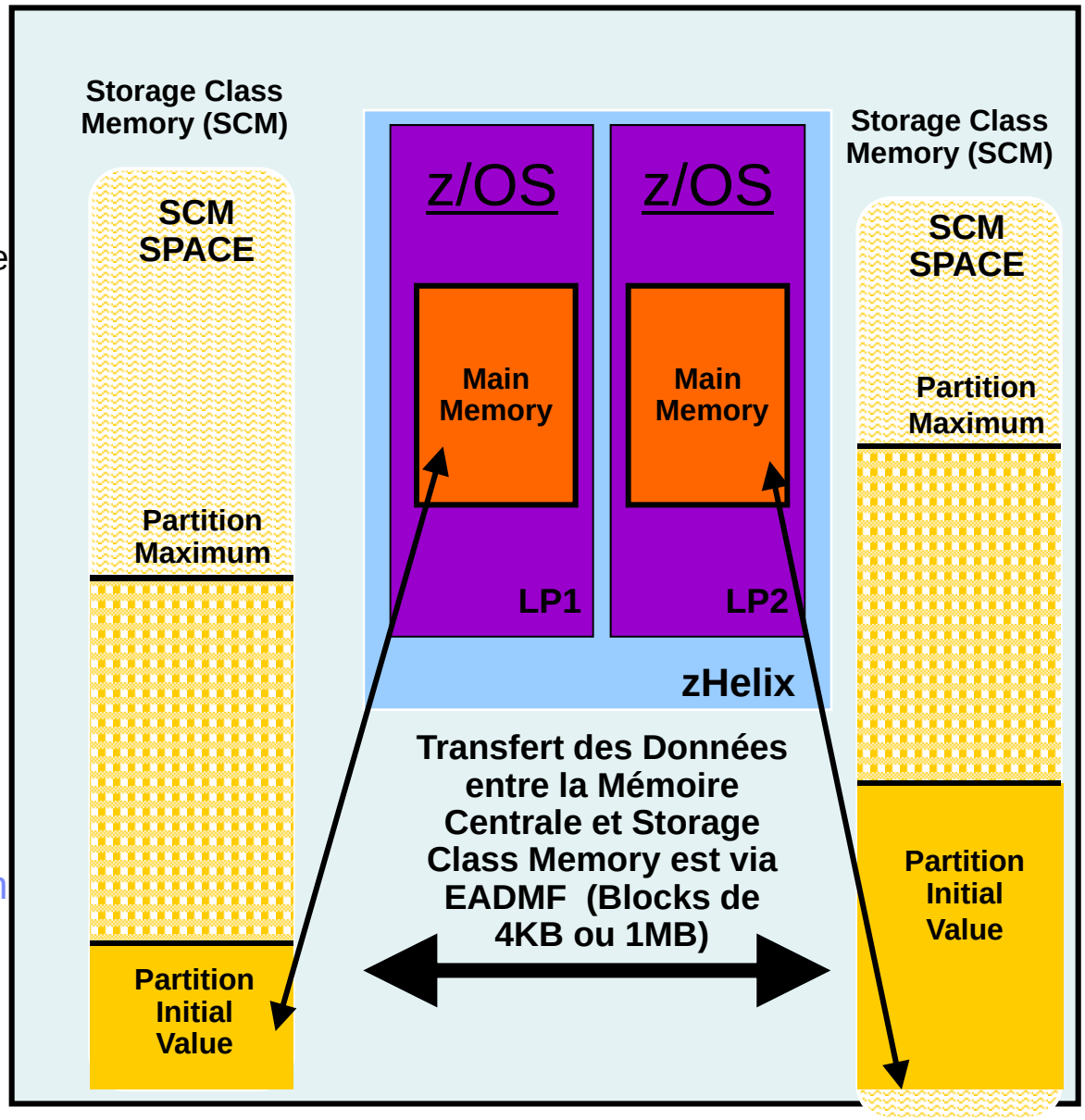


Utilisation de FLASH EXPRESS



Utilisation de FLASH EXPRESS

- **Virtualisation**
  - Chaque z/OS ne voit « qu'un espace Flash Express ».
    - Permet à chaque LPAR d'être configurée avec son propre SCM (Storage Class Memory) Address Space
    - Allocation de Flash aux LPARs par quantité (pas par taille de cartes).
- **Toute la gestion du matériel est faite de manière transparente pour le z/OS**
  - Erreur
  - Diagnostic
  - Etc.
- **EADM facility**
  - Extended Asynchronous Data Mover
  - EADM est initialisé avec une instruction
    - Start Subchannel.



Utilisation de FLASH EXPRESS

- **Pré requis**
  - Matériel – zEC12
  - Logiciel
    - z/OS V1.13 Flash Web Deliverable – GA December 14, 2012
    - z/OS V1.13 enabling PTFs for RSM enhancements – 1Q2013
      - Flash Dynamic Reconfiguration
      - Optional PLPA and COMMON Page data sets - (PLPA/COMMON (*NONE*)
    - A l'IPL z/OS détecte si Flash Memory est défini à cette LPAR et utilisera automatiquement Flash pour le Paging sauf si PAGESCM=NONE est indiqué dans IEASYSxx.
  - Futurs exploitants
    - DB2 for z/OS V10 and JAVA SDK7 SR3 will support pageable Large Pages -
      - Availability targeted for 4Q12
    - IBM is working with it's Linux Distribution partners to include support in future Linux on System z distribution releases
    - IMS* CQS will use pageable Large Pages when IMS runs on zEC12
      - Availability targeted for end of 2013

Utilisation de FLASH EXPRESS - RSM

RSM Enhancements will be delivered via RSM Enablement Offering Web Deliverable (FMID JBB778H) for z/OS V1.13

- Exploit Storage Class Memory (SCM) technology for z/OS paging and SVC dump
  - Is expected to yield substantial improvements in SVC dump data capture time
  - Removes the requirement for PLPA and Common page data sets when used for cold start (CLPA) IPLs.
  - It can also be used to remove the requirement for non-VIO local page data sets when the configuration includes enough SCM to meet peak demands
  - However, local page data sets remain required for VIO, and when needed to support peak paging demands that require more capacity than provided by the amount of configured SCM
- Pageable 1MB Large Page Support
- Dynamic reconfiguration support for Storage Class Memory (SCM) - target 1Q2013*
- Optional PLPA and COMMON page data set support – target 1Q2013*
- 2GB Large Page Support – target 1Q2013*

PLPA/COMMON (*NONE*)

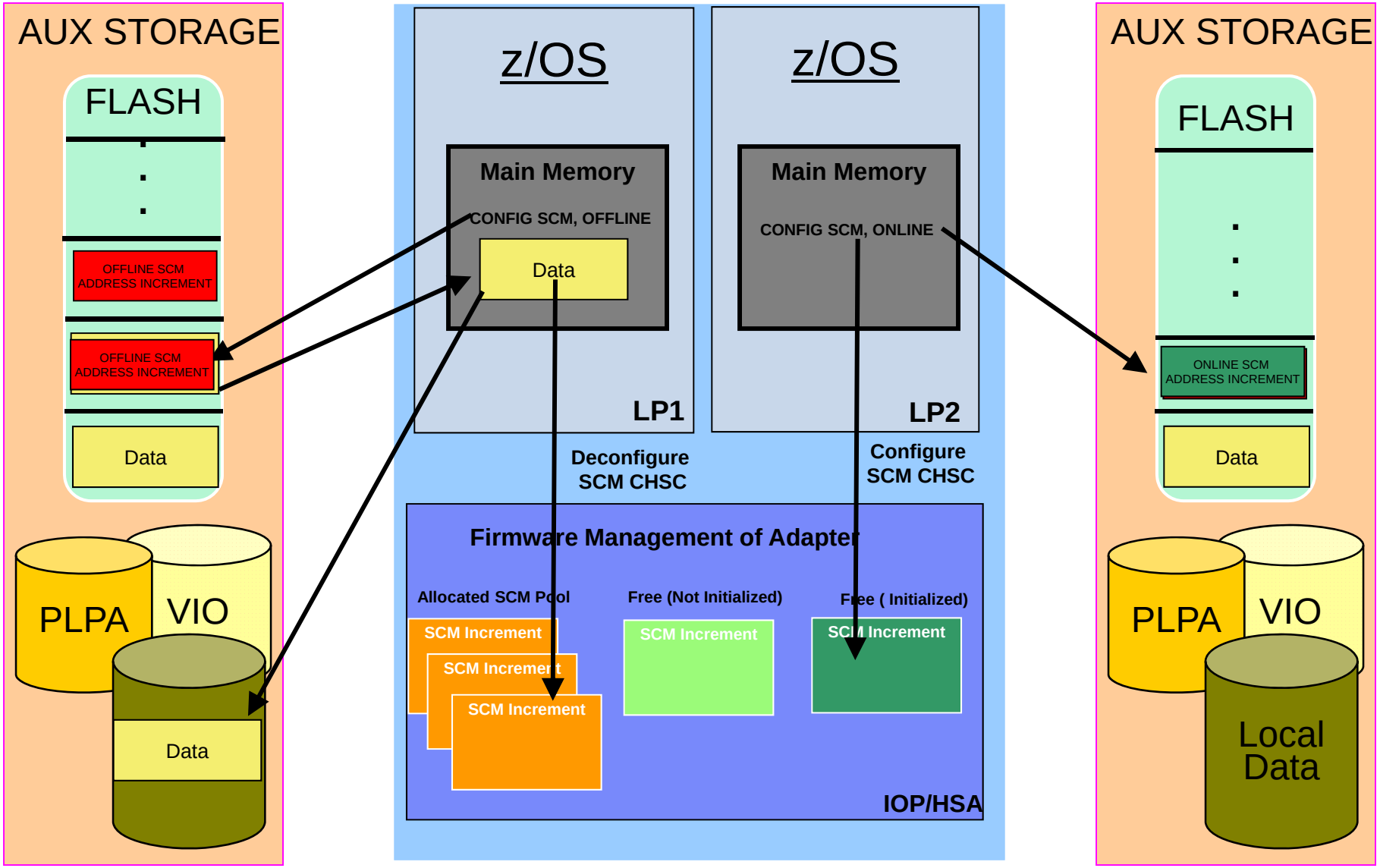
Installation of the z/OS V1R13 RSM Enablement Offering Web Deliverable (JBB778H) will:

- Increase the size of the Nucleus by roughly 280K
  - You may need to analyze your private storage usage
- Add 24K bytes of ESQA per CPU
- New memory pool (Pageable Large Page) is automatically carved out (approximately 1/8 of above the bar real storage)
  - Converted to Pageable 4K Pages if needed by the system

FLASH EXPRESS en action

FLASH EXPRESS en action

- **Exemple**

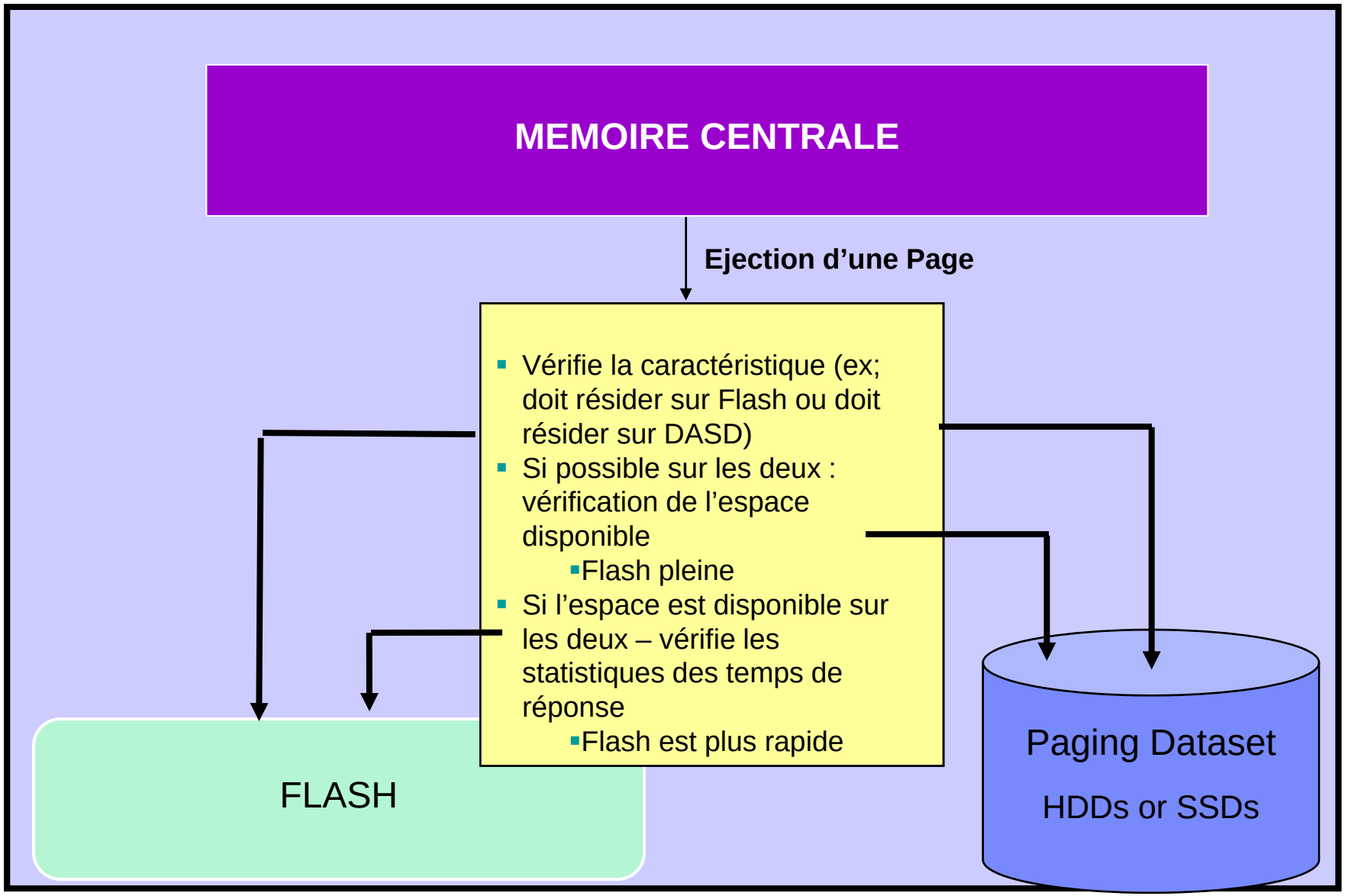


FLASH EXPRESS en action – La pagination de z/OS

- **Le z/OS paging subsystem utilisera un « mix » de Flash et DASD**
  - Self Tuning basé sur les performances mesurées
  - Amélioration des performances de Pagination , Simplification de la configuration
- **Début de la Pagination des 1 MB Large Pages uniquement sur Flash**
  - Exploitation du « Flash's random I/O read rate » pour gagner des cycles CPU en mettant en œuvre des utilisations plus grande des « Large Pages ».
  - Actuellement les Large Page ne sont pas « paginables ».
- **Les DUMPS**
  - Minimise la durée des SVC Dump – donc de l'impact sur le système
    - Flash performance pendant le SDUMP
    - Flash performance après le SDUMP
- **Prefetching sur les PAGE FAULTs**
  - z/OS profite du « Flash's fast random reads » pour amener plus de pages « locales »
  - Etude en cours pour ramener d'autres pages potentiellement référençables.

FLASH EXPRESS en action – La pagination de z/OS

- **Le choix de z/OS pour la pagination**



FLASH EXPRESS en action – La pagination de z/OS

- **Les critères de placement**

Data Type	Data Page Placement
PLPA	**A L'IPL/NIP les pages PLPA seront écrites sur Flash et DASD**
VIO	**Les Fichiers VIO seront écrits sur DASD**
Hyperswap Critical Address Space data	**Si de l'espace est disponible sur FLASH, les pages virtuelles appartenant à un Hyperswap Critical Address Space seront écrites sur Flash.** **Sinon, ces pages seront gardées en mémoire réelle et uniquement paginées sur DASD quand le système sera en contrainte de mémoire réelle et qu'aucune autre alternative n'existe**
Pageable Large Pages	**Si de l'espace contigu existe sur Flash, les Pageables Large Pages y seront de préférence écrites.** **Sinon, les pages seront "rétrogradées" et transformées en pages de 4K**
All other data	**Si de l'espace existe à la fois sur Flash et DASD, alors sélection du média en fonction du temps de service du média**

Paramétrage z/OS

Paramétrage z/OS

- **IEASYSxx**
  - PAGESCM= indique la taille minimum de Flash Express à réserver pour la pagination.
    - Valeur en unité de M, G, ou T
    - NONE indique de ne pas utiliser Flash pour la pagination
    - ALL (défaut) indique que toute la quantité de Flash définie à la LPAR est disponible pour la pagination.
- **Message à l'IPL**
  - `IAR031I USE OF STORAGE-CLASS MEMORY FOR PAGING IS ENABLED -PAGESCM=ALL, ONLINE=00001536M`
  - `ou`
  - `IAR032I USE OF STORAGE-CLASS MEMORY FOR PAGING IS NOT ENABLED – PAGESCM=NONE`

Paramétrage z/OS

- **COMMANDES - DISPLAY**

```
COMMAND INPUT ===> /D ASM_
RESPONSE=J90
 IEE200I 12.07.24 DISPLAY ASM 613
 TYPE      FULL STAT    DEV  DATASET NAME
 PLPA       18%   OK   8571  SYS1.J90.PLPA
 COMMON      0%   OK   8570  SYS1.J90.COMMON
 LOCAL       0%   OK   8570  SYS1.J90.LOCAL
 LOCAL       0%   OK   8570  SYS1.J90.LOCAL2
 LOCAL       0%   OK   8571  SYS1.J90.LOCAL3
 LOCAL       0%   OK   8571  SYS1.J90.LOCAL4
 LOCAL       0%   OK   8572  SYS1.J90.LOCAL5
 SCM         1%   OK    N/A   N/A
 PAGEDEL COMMAND IS NOT ACTIVE
```

```
COMMAND INPUT ===> /d m=scm
RESPONSE=J90
 IEE174I 12.04.16 DISPLAY M 340
 STORAGE-CLASS MEMORY STATUS
 384G DEFINED
 ONLINE
  0G-80G
 304G OFFLINE-AVAILABLE
 2% IN USE
 SCM INCREMENT SIZE IS 16G
```

```
COMMAND INPUT ===> /d asm,scm                                  SCROLL ==
RESPONSE=J90
 IEE207I 12.05.28 DISPLAY ASM 494
 STATUS      FULL              SIZE            USED        IN-ERROR
 IN-USE       1%          20,971,520         210,210              0
```

```
COMMAND INPUT ===> /D M=SCM(DETAIL)
RESPONSE=J90
 IEE174I 12.06.13 DISPLAY M 557
 STORAGE-CLASS MEMORY STATUS - INCREMENT DETAIL
 384G DEFINED
 ADDRESS   IN USE   STATUS
       0G      10%  ONLINE
     16G        0%  ONLINE
     32G        0%  ONLINE
     48G        0%  ONLINE
     64G        0%  ONLINE
 ONLINE: 80G  OFFLINE-AVAILABLE: 304G  PENDING OFFLINE: 0G
 2% IN USE
 SCM INCREMENT SIZE IS 16G
```

  - CF SCM(quantité),ONLINE ou OFFLINE

Les états RMF

RMF

- **Le REPORT FRAME AND SLOTS COUNT**

FRAME AND SLOT COUNTS

(181 SAMPLES)

CENTRAL STORAGE FRAMES	TOTAL	AVAILABLE	SQA	LPA	CSA	LSQA	REGIONS+SWA	HV SHARED	HV COMMON
MIN	34019328	26255728	37,820	17,276	38,587	314,895	6,851,620	249,699	78,909
MAX	34019328	26436096	38,631	17,276	38,671	316,023	7,032,986	253,953	78,909
AVG	34019328	26359600	37,871	17,276	38,638	315,349	6,928,872	251,582	78,909

FIXED FRAMES	TOTAL	NUCLEUS	SQA	LPA	CSA	LSQA	REGIONS+SWA	<16 MB	16MB-2GB
MIN	955,576	2,930	33,839	84	11,706	76,085	830,629	92	62,216
MAX	980,148	2,930	34,649	84	11,713	77,032	855,001	297	64,103
AVG	958,948	2,930	33,891	84	11,706	76,463	833,873	95	63,006

SHARED FRAMES / SLOTS	TOTAL	CENTRAL STORAGE	FIXED TOT	FIXED BEL	AUX DASD	AUX SCM
MIN	76,781	67,383	300	0	0	0
MAX	76,869	67,471	301	0	0	0
AVG	76,823	67,425	300	0	0	0

LOCAL PAGE DATA SET SLOTS	TOTAL	AVAILABLE	BAD	NON-VIO	VIO
MIN	2,879,995	2,879,995	0	0	0
MAX	2,879,995	2,879,995	0	0	0
AVG	2,879,995	2,879,995	0	0	0

SCM PAGING BLOCKS	TOTAL	AVAILABLE	BAD	IN-USE
MIN	20971520	20956624	0	14,492
MAX	20971520	20957024	0	14,889
AVG	20971520	20956912	0	14,605

RMF

- **Le REPORT P A G I N G A C T I V I T Y**

```
                                                P A G I N G  A C T I V I T Y

            z/OS V1R13                    SYSTEM ID J90              START 09/04/2012-06.00.00  INTERVAL 000.30.00
                                          RPT VERSION V1R13 RMF      END   09/04/2012-06.30.00  CYCLE 1.000 SECONDS
-OPT = IEAOPTHD   LFAREA SIZE =  2147484K   MEMORY OBJECTS AND HIGH VIRTUAL STORAGE FRAMES
-------------------------------------------------------------------------------------------------------------------
MEMORY OBJECTS          COMMON      SHARED       1 MB
------------------   ---------   ---------   ---------
      MIN                   64           6       1,915
      MAX                   64           6       1,915
      AVG                   64           6       1,915
1 MB FRAMES          ------------ FIXED ------------   ----------- PAGEABLE -----------
------------------        TOTAL   AVAILABLE      IN-USE       TOTAL   AVAILABLE      IN-USE
      MIN                 2,048           0       2,048      14,336      14,329           7
      MAX                 2,048           0       2,048      14,336      14,329           7
      AVG                 2,048           0       2,048      14,336      14,329           7
HIGH SHARED FRAMES        TOTAL       CENTRAL STORAGE                  AUX DASD     AUX SCM
------------------   ---------   ---------------                  ---------   ---------
      MIN             136902.1M                 249,699                    0           0
      MAX             136902.1M                 253,953                    0           0
      AVG             136902.1M                 251,582                    0           0
HIGH COMMON FRAMES        TOTAL       CENTRAL STORAGE    FIXED 4K    AUX DASD     AUX SCM
------------------   ---------   ---------------   ---------   ---------   ---------
      MIN              17301504                  78,909       4,195           0           0
      MAX              17301504                  78,909       4,195           0           0
      AVG              17301504                  78,909       4,195           0           0
```

RMF

- **Le REPORT PAGE DATA SET AND SCM USAGE**

```
                                    P A G E   D A T A   S E T   A C T I V I T Y

            z/OS V1R13              SYSTEM ID J90            START 09/04/2012-06.00.00  INTERVAL 000.30.00
                                    RPT VERSION V1R13 RMF    END   09/04/2012-06.30.00  CYCLE 1.000 SECONDS
-
 NUMBER OF SAMPLES =   1,800                  PAGE DATA SET AND SCM USAGE
                                              --------------------------
-PAGE                                                                    %    PAGE                        V
 SPACE    VOLUME   DEV  DEVICE    SLOTS  ---- SLOTS USED ---  BAD       IN   TRANS  NUMBER  PAGES  I
 TYPE     SERIAL   NUM  TYPE      ALLOC    MIN    MAX   AVG   SLOTS     USE  TIME   IO REQ  XFER'D O  DATA SET NAME
-LOCAL    PAG901   8571 33909    575999      0      0     0       0    0.00 0.000        0       0 Y  SYS1.J90.LOCAL4
 LOCAL    PAG902   8572 33909    575999      0      0     0       0    0.00 0.000        0       0 Y  SYS1.J90.LOCAL5
 SCM      N/A      N/A  N/A      20972K  14492  14893 14600       0    0.94 0.000   97,965       0    N/A
 LOCAL    PAG900   8570 33909    575999      0      0     0       0    0.00 0.000        0       0 Y  SYS1.J90.LOCAL
 LOCAL    PAG900   8570 33909    575999      0      0     0       0    0.00 0.000        0       0 Y  SYS1.J90.LOCAL2
 LOCAL    PAG901   8571 33909    575999      0      0     0       0    0.00 0.000        0       0 Y  SYS1.J90.LOCAL3
 COMMON   PAG900   8570 33909     71999    298    298   298       0    0.00 0.000        0       0    SYS1.J90.COMMON
 PLPA     PAG901   8571 33909     71999  13077  13077 13077       0    0.00 0.000        0       0    SYS1.J90.PLPA
```

Les chiffres de performance

Environnement IBM contrôlé

Les chiffres de performance

Flash Performance Test Setup

- z/OS Tests were designed to demonstrate flash performance under paging workloads that are typically encountered in a z/OS enterprise environment
  - SSD performance is not only about the number of IOPS but about steady performance over time and consistent latency
    - Preconditioned SSDs with random-write IO engage the device's wear leveling, error handling, and flash management algorithms
  - Comparison DASD Characteristics used current device configurations
    - DS8800 model 2107-951
    - 60 GB cache, cache hit rates of 95-100% were observed during the tests
    - DASD was not shared with any other systems and did not have any I/O traffic other than the paging traffic used for these tests
    - Configured 16 local page datasets spread across 8 LCUs

Bottom line: The tests were set up to similate real customer environments

Les chiffres de performance

Flash Performance Benefits

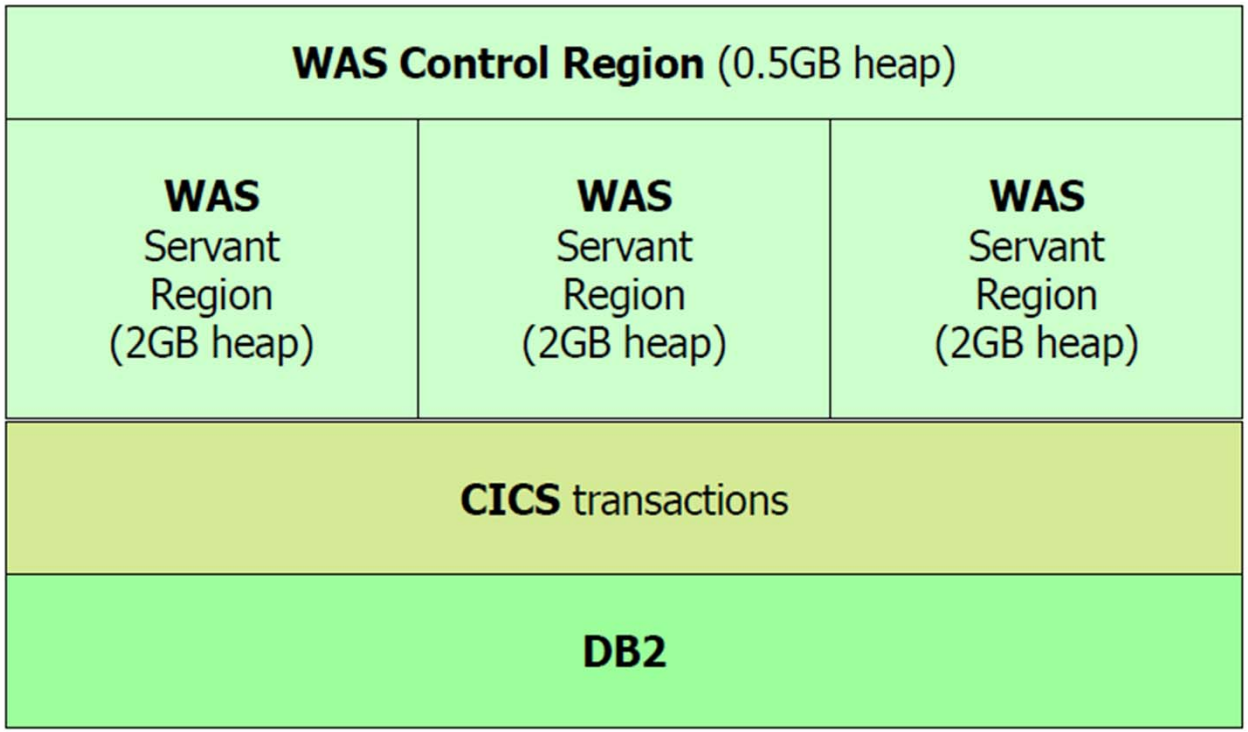
Test Results

- FLASH paging benefits
  - Improved availability through faster paging at critical times
    - Faster workload transitions (e.g.; morning startup)
      - *meaning less time to reach peak transaction rates*
    - Faster SVC dumps (reduced **non-dispatchable** time)
      - *meaning higher availability*
- Pageable Large Page benefit
  - Java realizes performance benefits from use of large 1MB pageable pages
    - Large pages benefits for JIT Code Cache, 31 bit Java applications
    - No authorization needed to access fixed large pages
    - Approximately 5-8% CPU improvement from PLP

Les chiffres de performance

Workload Configuration Block Diagram
Building block – A WAS instance accessing CICS and DB2

Each WAS instance has a WAS Control Region and 3 WAS Servant Regions.
Each WAS Control Region has a 0.5GB heap plus a JIT Code cache.
Each WAS Servant Region has a 2GB heap plus a JIT Code Cache.



Test Configuration

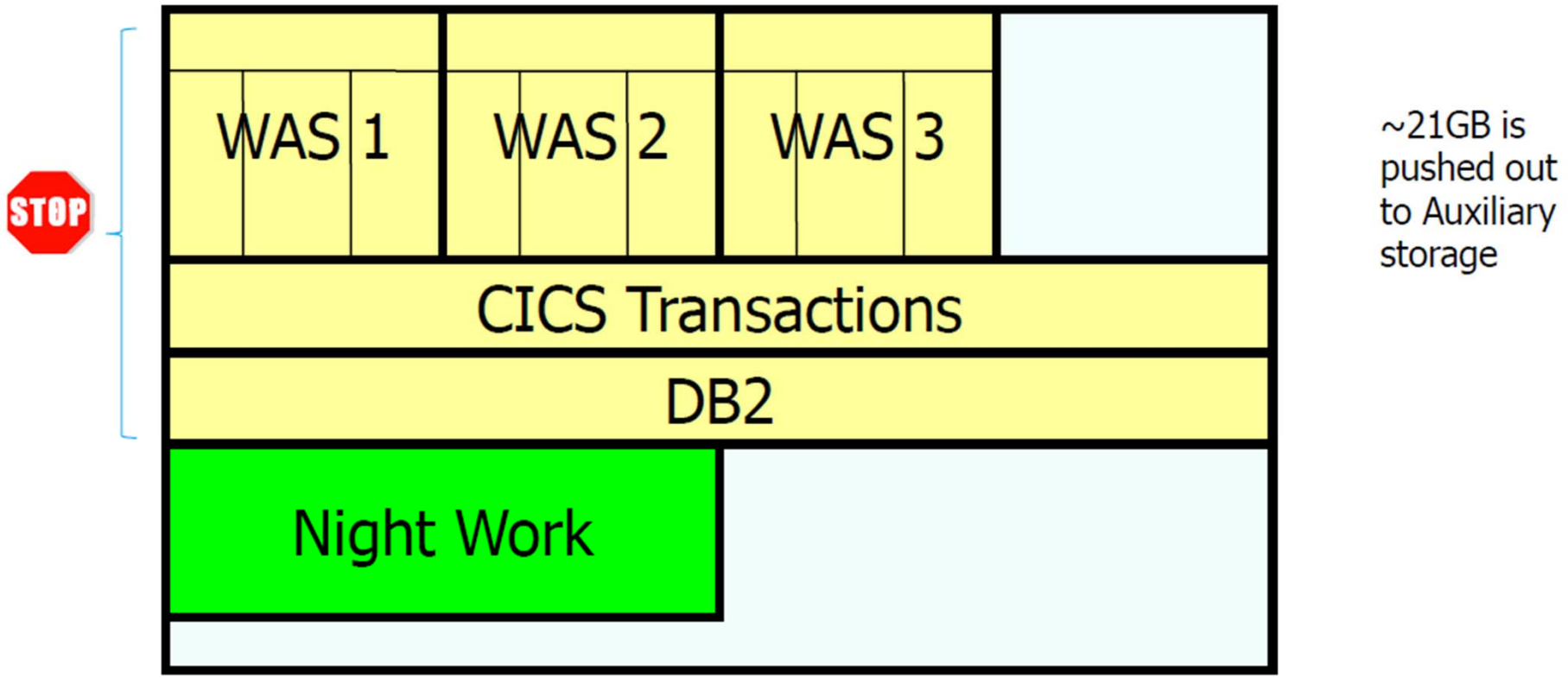
- WAS 7 (3 servants each with two GB heap) + 1 control region (.5 GB Heap)
- CICS V4.2, DB210 on a zEC12
- Storage: DS8800 2107-951 with 60GB cache, very fast device
- Tests simulated morning transition time typical of trading or call center work
- SVC dump measurements were taken for an 18 GB dump.

Les chiffres de performance

I. Workload Transition

Transition from night batch to OLTP

WAS workload to CICS and DB2 represents OLTP work which is then stopped
Simulated overnight work consumes real storage pushing other pages out



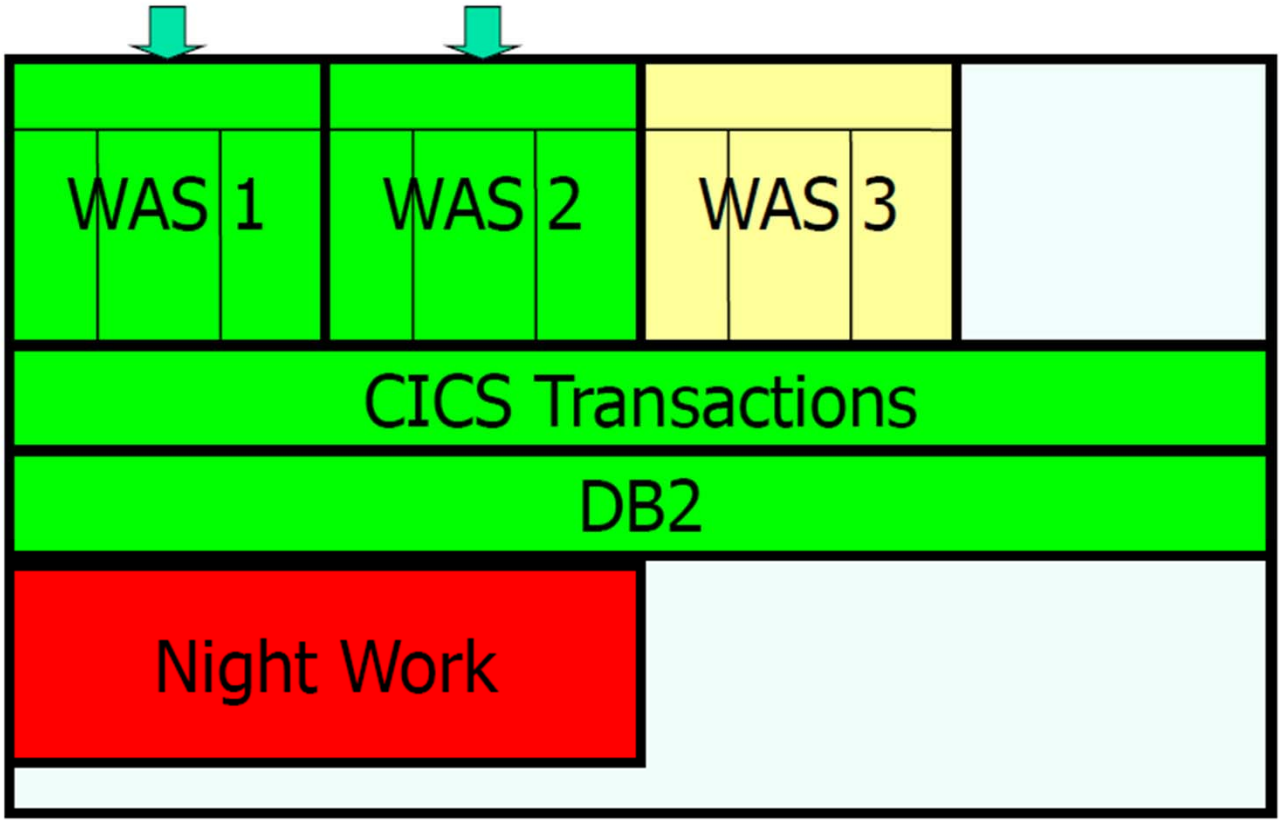
~21GB is pushed out to Auxiliary storage

Les chiffres de performance

Workload Transition

Transition from night batch to OLTP

The "Night Work" is then stopped and OLTP work is started (WAS 1 and WAS 2)
Measure the time needed to bring the OLTP work to full speed.

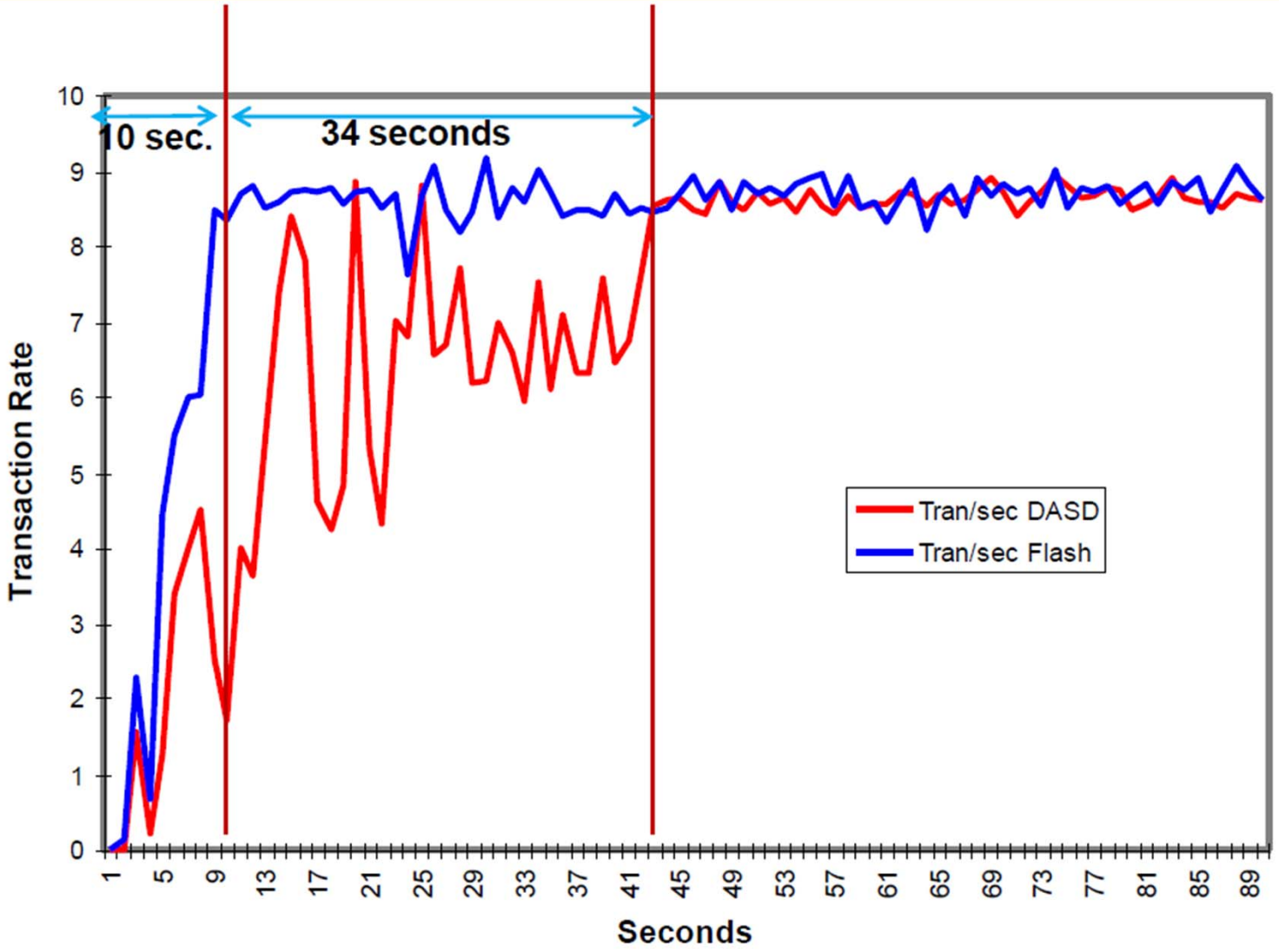


~14GB is paged back in from Auxiliary storage

Les chiffres de performance

Workload Morning Transition - Results

- **During morning transition, workloads using Flash Express reached peak throughput in less than a quarter of the time required when using DASD**



Paging to **DASD** required about **44 seconds** for the workload to reach steady state

Paging to **Flash** required only **10 seconds** for the workload to reach steady state

Les chiffres de performance

Workload Morning Transition – Results Apparent in First 45 Seconds

Transaction completion & response time	DASD	Flash	Improvement
Total Transactions within first **45** seconds	251	343	**37% increase**
Average response time within first **45** seconds	0.62	0.06	**90% reduction**

- **Paging to Flash generated a 90% reduction in response time…**
  And…
- **a 37% increase in transaction throughput (1)**

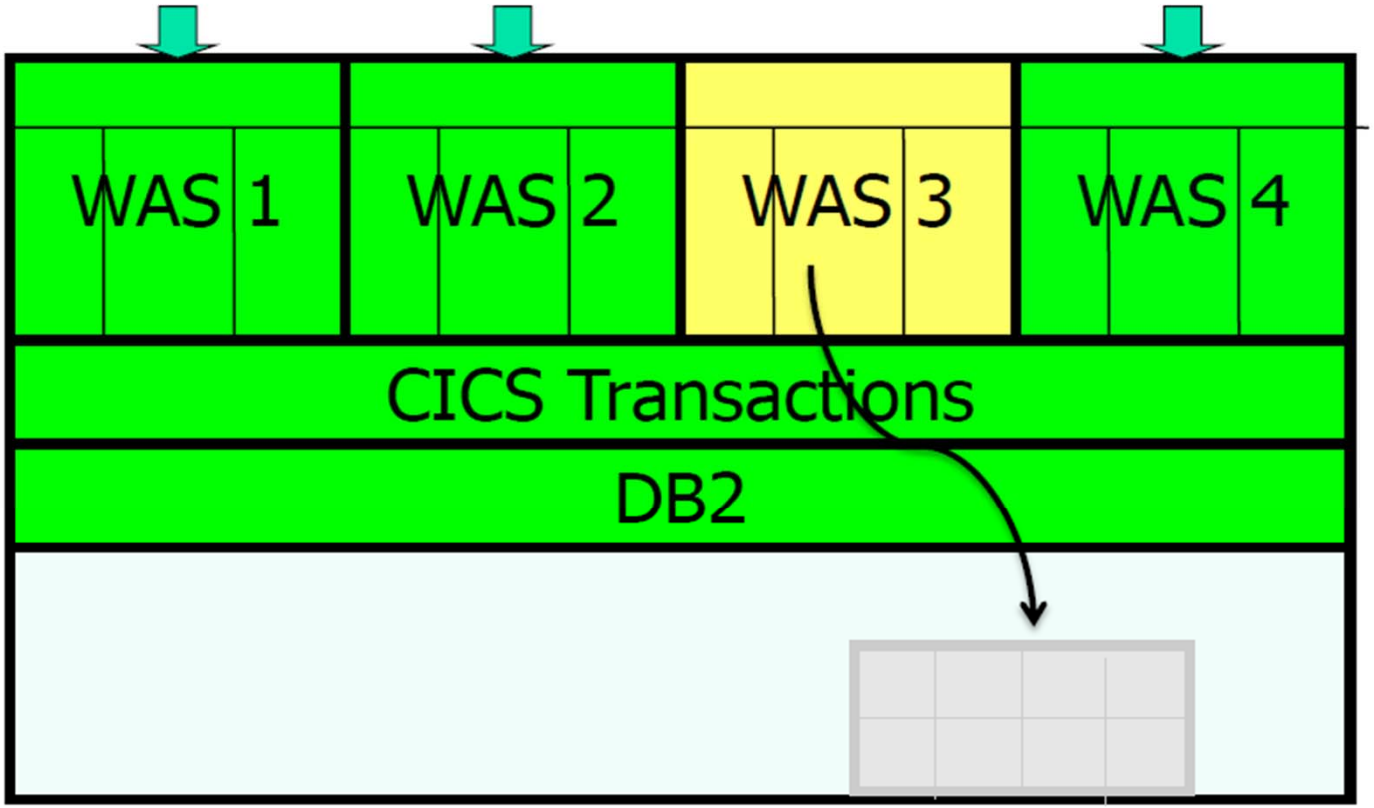
(1) Test was for the first 45 seconds of morning transition time

Les chiffres de performance

II. SVC Dump

SVC dump with pages out

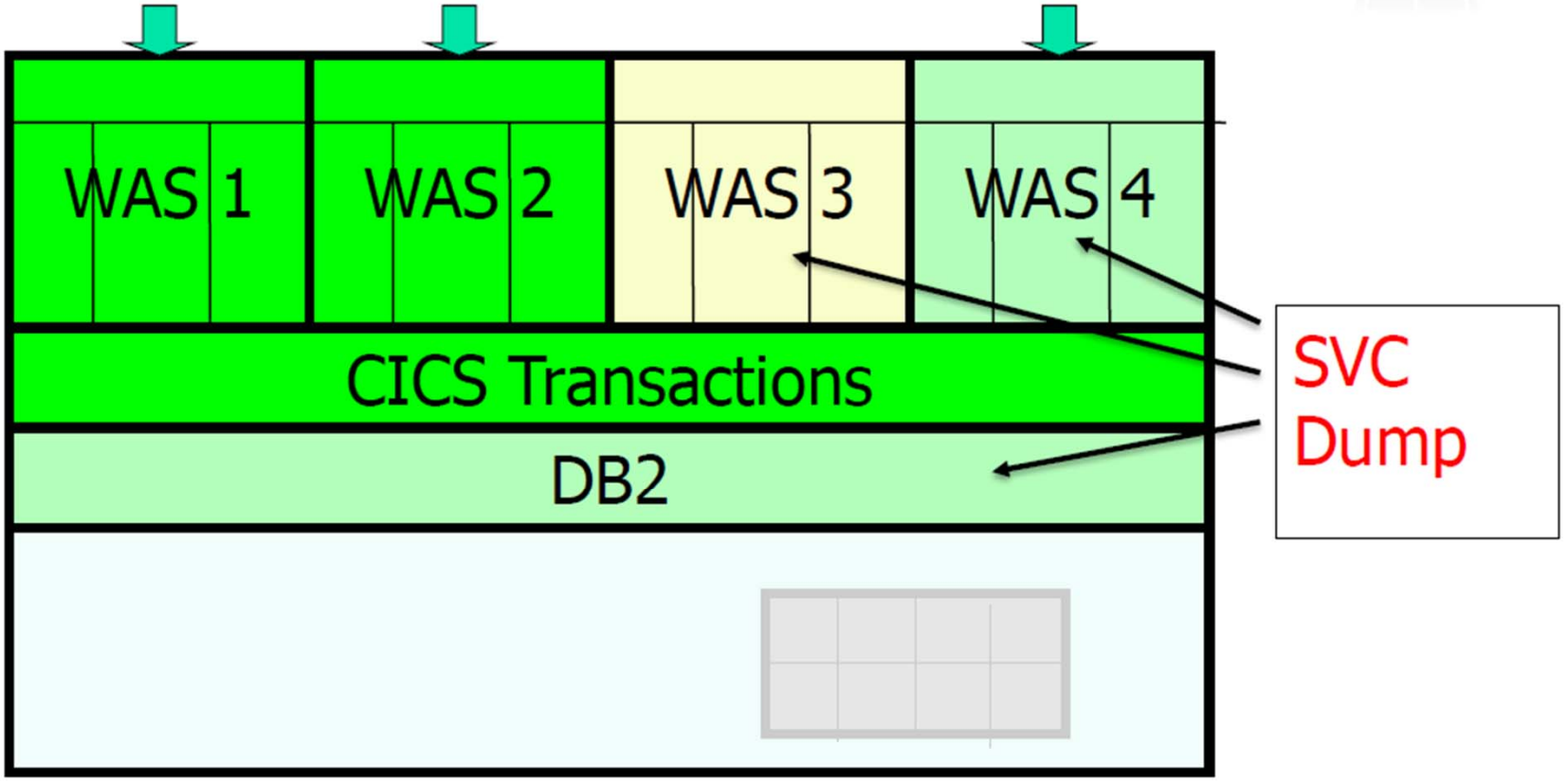
Three of four WAS instances were active.
One WAS instance was stopped and most pages were paged out.



Les chiffres de performance

SVC Diagnostics capture

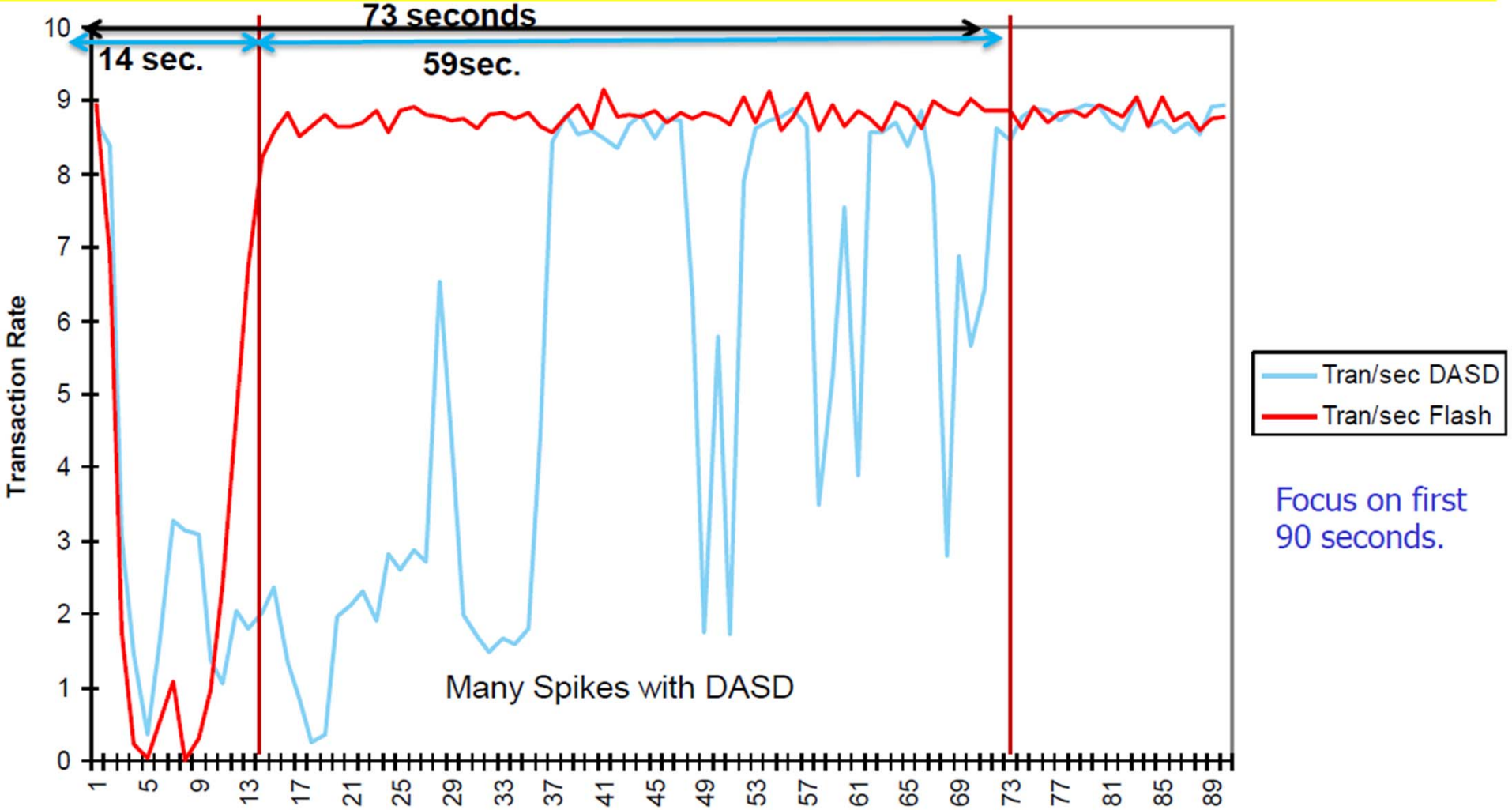
Capture an SVC dump of WAS instance 3 and 4, and DB2.
Measure the capture time for the SVC dump.



Les chiffres de performance

SVC Dump - Results

❖**Steady state is reached in 14 seconds with Flash Express, vs 73 seconds DASD - an 80% reduction. Another way to say this is steady state was achieved 5 times faster with Flash Express**



Les chiffres de performance

SVC Dump - Results

❖ **With Flash, SVC dump elapsed time was almost 25% shorter than when using DASD**

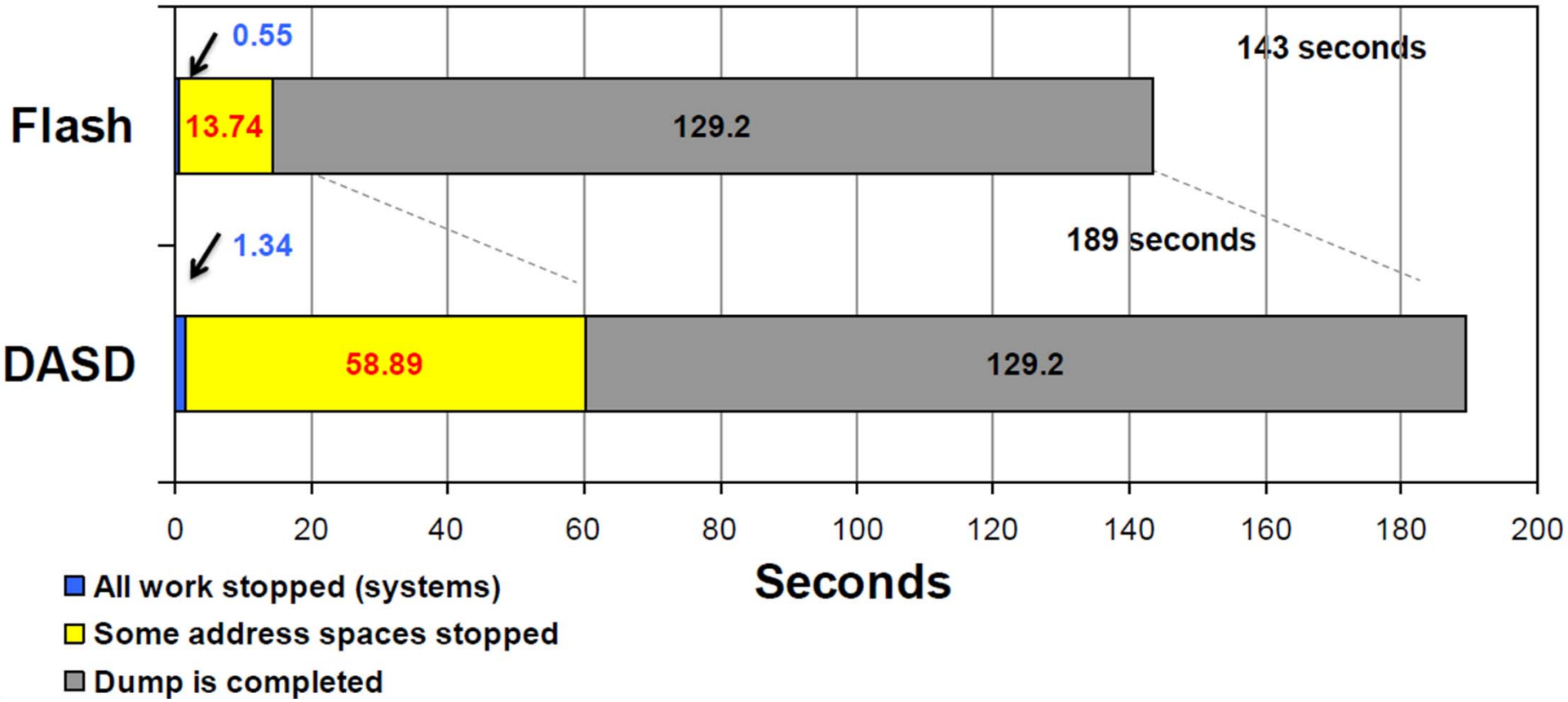
SVC Dump Metrics	DASD	Flash
SVC Dump size (in bytes):	18GB	18GB
% of pages from Aux storage:	50%	53%
DUMP Elapsed time:	189	143
Max address space non-dispatchable seconds	58.89	13.74
System non-dispatchable seconds	1.34	0.55

Let's graph these results….

Les chiffres de performance

SVC Dump –Another View

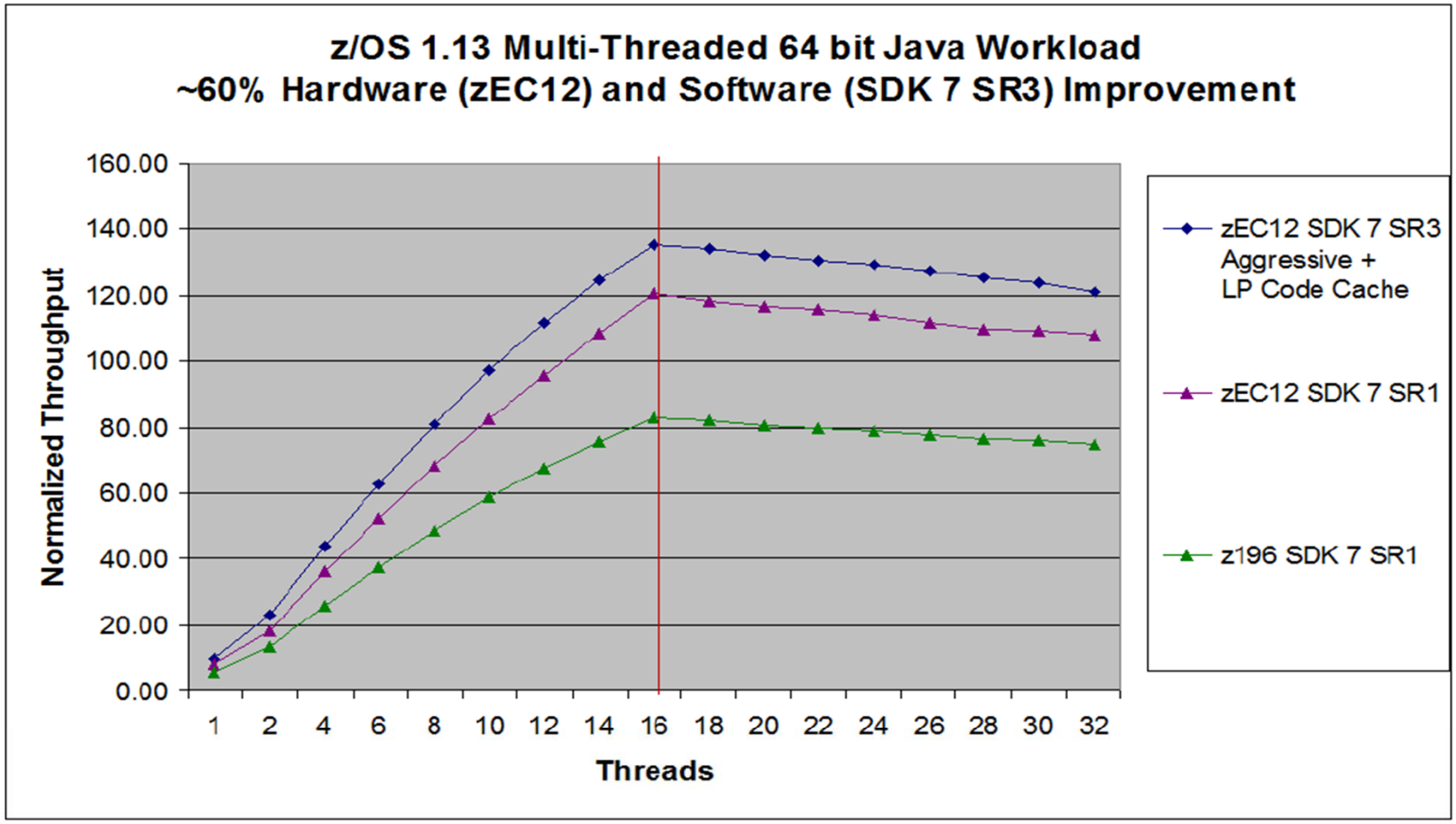
- Using Flash, Address Spaces were stopped 77% less time and Systems were stopped 59% less time than when using DASD in the SVC dump test.
- This means availability was 4 times better for Address spaces. And twice as good for systems.



Les chiffres de performance

z/OS Java SDK 7:16-Way Performance Shows up to 60% Improvement

64-bit Java Multi-threaded Benchmark on 16-Way



Aggregate 60% improvement from zEC12 and Java7SR3

- zEC12 offers a ~45% improvement over z196 running the Java Multi-Threaded Benchmark
- Java7SR3 offers an additional ~13% improvement (-Xaggressive + Flash Express pageable 1Meg large pages)

Les chiffres de performance

z/OS Performance for Pageable Large Pages

❖ **WAS* Workload showed up to an 8% performance improvement using Flash**

Java 7 SR3	JIT	Java Heap	Multi Threaded	WAS Day Trader 2.0
31 bit	yes	yes	4%	
64 bit	yes		1%	3%
64 bit		yes	4%	5%

* WAS Day Trader 64-bit Java 7 SR3 with JIT code cache & Java Heap

Details

- **64-bit Java heap (**1M fixed large pages (FLPs) or 1M Pageable (PLPs)) versus 4k pages
  - Java heap 1M PLPs improve performance by about
    - **4%** for Multi-Threaded workload
    - **5 %** for WAS Day Trader 2.0
- **64-bit** Java 7 SR3 with **JIT code cache** 1M PLPs vs without Flash
  - **3** % improvement for traditional WAS Day Trader 2.0*
  - **1** % improvement for Java Multi-Threaded workload
- ***31-bit** Java 7 SR3 with **JIT code cache and Java heap** 1M PLPs vs without Flash
  - **4** % improvement for Java Multi-Threaded workload

Note: Used WAS Day Trader app that supports PLP, Earlier version of 31-bit Java did not allocate 1M large pages

Les chiffres de performance

Summary of Benefits

- FLASH paging benefits
  - Improved availability through paging to Flash Express
    - Faster workload transitions (morning startup)
    - Faster SVC dumps (less non-dispatchable time)
- Pageable Large Pages can provide the **same benefit** as fixed large pages.
  - Java heap
    - Uses PLP **without configuring security and LFAREA**
    - With Java 7 SR3 and Java 6.0.1 SR4, users running Java experienced
      - 64-bit applications show 4%-5% performance benefit from Java heap alone
        And an additional 1%-3% improvement for the JIT Code Cache
      - 31-bit applications were able to use PLP for both Java heap *and* JIT Code Cache

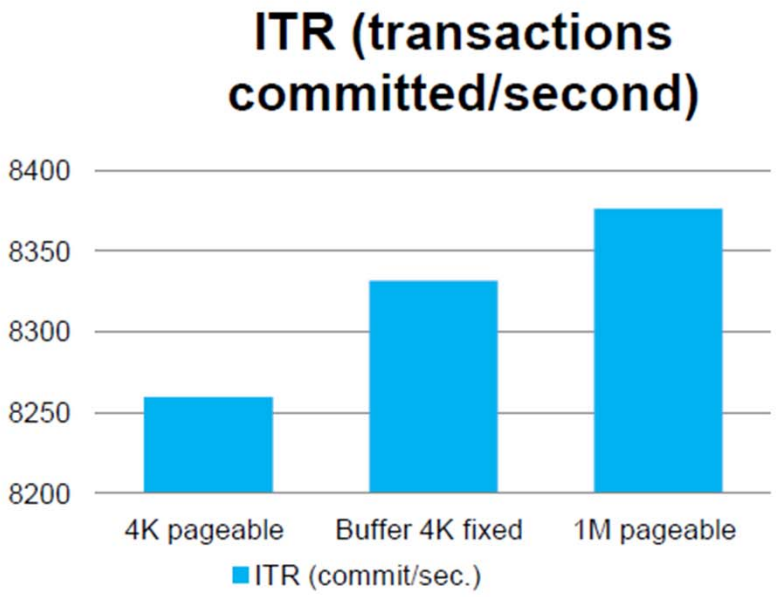
- **Java Multi-threaded benchmark with 31-bit Java, shows a 4% benefit from the Java heap and JIT Code cache**
- **Java WebSphere Day Trader 2.0 with 64-bit Java, shows a 8% benefit using PLP for Java heap and JIT Code cache**

- *NOTE Tests use Java 7 SR3 or Java 6.0.1 SR4. Earlier Java version did not support large pages for 31-bit Java*

Les chiffres de performance

DB2 and zEnterprise EC12 Team to Bolster Capacity and Performance

- **Faster CPU – 1.25X compared to z196**
  - **20-28%** CPU reduction measured with DB2 OLTP workloads
  - **25%** reduction measured with DB2 query and utilities workloads
  - Less compression overhead with DB2 data **(1-15%)**
- **50% More System Capacity**
- **New Features DB2 plans to exploit**
  - FLASH Express and 1MB Pageable Large Pages
    - Larger DB2 Buffer Pools
  - 2GB frame support
    - Larger frames expected to provide additional CPU savings, especially for very large memory
  - DB2 code backed by large frames for CPU reductions
- Transactional Execution opportunities for performance gains
- Initial support planned for DB2 10 with APARS



***DB2 transaction throughput** improvement stems from **reduced CPU** needed for buffer pool management*

❖ **Initial results: Pageable Large Pages for DB2 helps DB2 achieve up to a 3% transaction throughput improvement from CPU savings. The savings are due to reduced buffer pool management.**

Les chiffres de performance

- **Synthèse**

WORKLOAD TRANSITION

- **During morning transition, workloads using Flash Express reached peak throughput in under one-fouth (23%) of the time compared to using DASD.**
- **In the first 45 seconds of Morning transition, Paging to Flash showed a 10 times faster response time and a 37% increase in throughput as compared when using DASD.**

WAS

- **The WAS Day Trader 64-bit showed 8% performance improvement using Flash. The test used Java 7 SR3 with JIT code cache & Java Heap leveraging Flash and PLP.**

Startup morning transition time is much faster and steady state performance achieved rapidly which can mean more consistent performance for customers. It also means companies are better prepared to meet their performance SLAs.

With faster response time and huge throughput improvements, more transactions can be processed -- which can mean more business for your customer.

WAS day trader benchmarks, indicate a significant performance boost from pageable large pages and can serve as a proxy for similar customer workloads.

Les chiffres de performance

- **Synthèse**

DIAGNOSTICS

- In the SVC dump scenario, Address Spaces were stopped 77% less time and Systems were stopped 59% less time when using Flash. This means higher availability for both address spaces and overall systems when taking diagnostics.
- Flash steady state performance was 5 times faster with Flash Express in SVC Dump tests; transaction steady state was reached in 14 seconds with Flash Express, vs 73 seconds DASD.
- Flash Express SVC dump elapsed time was almost 25% shorter than with DASD.

Without Flash and pageable large pages, dumps can disrupt operations not only of the workload involved directly in a dump, but for other workloads as well. Flash therefore improves availability for *many* workloads; this is ideal for both development and production use cases.

Les chiffres de performance

Synthèse générale

WORKLOAD TRANSITION

- During morning transition, workloads using Flash Express reached peak throughput in under 1/4th the time
- Paging to Flash Express during morning transition showed up to a 10 times faster response time and up to a 37% increase in throughput within the first 45 seconds

WAS JAVA PERFORMANCE BENCHMARKS

- The WAS Day Trader benchmarks showed up to an 8% performance improvement using Flash Express.(2)

** This test used 64-bit Java 7 SR3 with JIT code cache & Java Heap leveraging Flash and pageable large pages.*

Improved Availability During Diagnostics

- In SVC dumps, availability was up to 4 times higher for workloads and up to twice as high for systems*
- In SVC dump tests, steady state performance was achieved up to 4 times faster *
- Flash Express SVC dump elapsed time was up to 25% shorter

** Transaction steady state was reached in 14 seconds with Flash Express, vs. 60 seconds DASD.*

DB2

- Up to 28% improvement in DB2™ throughput due to faster CPU and leveraging Flash Express with Pageable Large Pages (PLP)*
- Workloads leveraging Flash Express with PLP can see up to a 8%** price performance improvement over the z196.

* *PLP for DB2 helps DB2 to achieve* ***"additional" up to 3% additional performance on top of zEC12 CPU expected throughput improvements of 25%.***

** *based on average 5% discount for zEC12 workloads under the AWLC pricing plus up to 3% more performance per MSU with Flash Express.*

(1) All tests are comparing the use of Flash Express as compared to using DASD (DS8800)

(2) System non dispatchability and address space non dispatchability time were dramatically reduced enabling work to be processed that would otherwise have been stopped

Flash Express - Une nouvelle mémoire intelligente

FIN DU DOCUMENT

Vendredi 5 avril 2013 – 13H45-15H05

Université du Mainframe 2013

4-5 avril