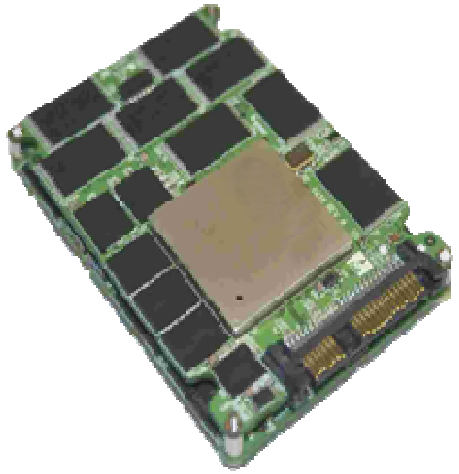
IBM Power Systems

Technologie des disques SSD et IBM i

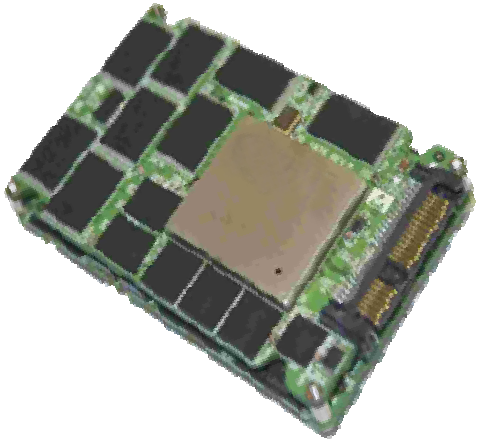
COMMON, Genève le 03 mai 2011

Agenda

- Positionnement produit
- La technologie des SSD
- L’offre IBM
- IBM i et SSD



Positionnement Produit



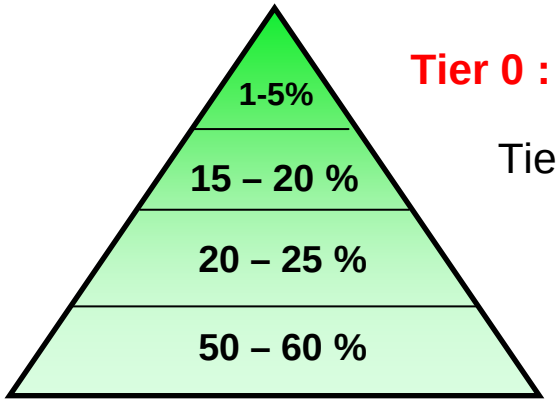
Définitions ….

✓ Unité de stockage constituée de mémoires flash qui peut être effacée électriquement et reprogrammée.

✓ Inventée par Toshiba et utilisée depuis les années 95 dans l'aérospatiale

✓ Technologie présente dans les clés USB et dans les cartes mémoire des appareils photos numériques (cartes micro SDH

Classement des moyens de stockage en Tiers:



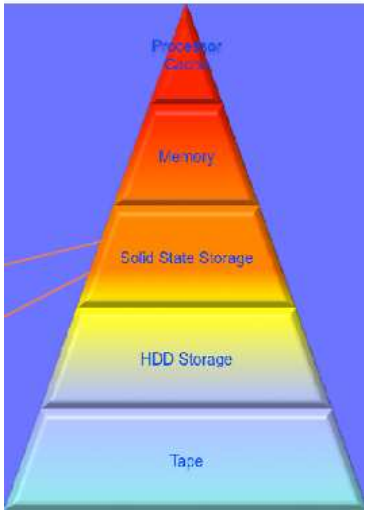
Tier 0 : Entreprise SSD (FC SSD, SAS SSD)

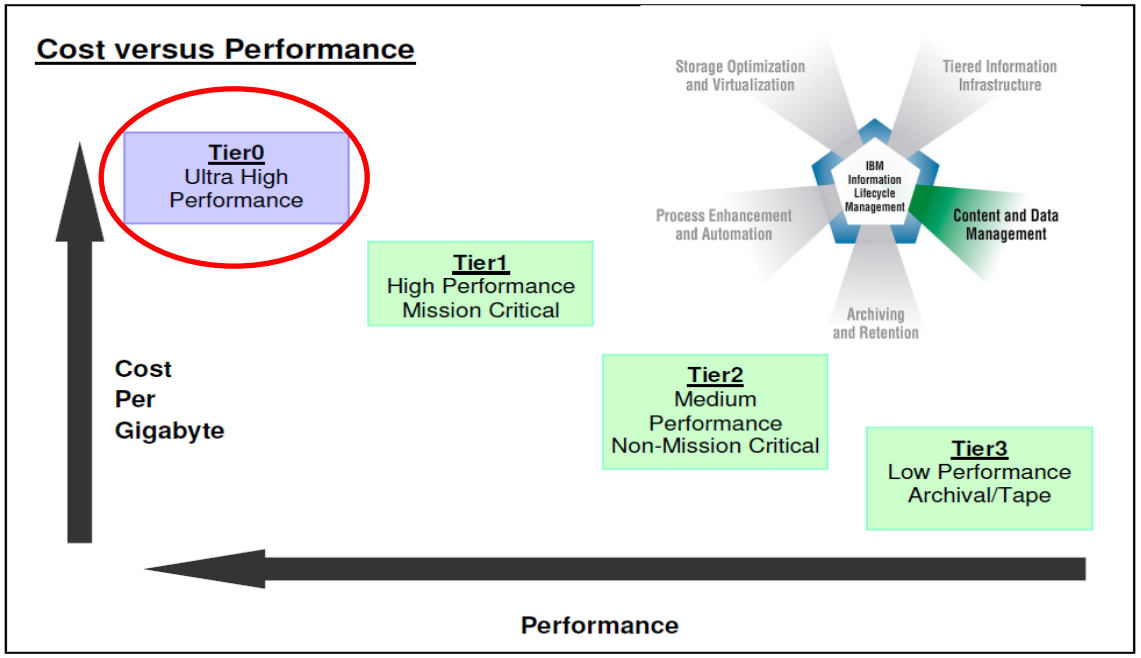
Tier 1 : HDD (FC HDD, SAS HDD)

Tier 2 : SATA SSD & SATA HDD (interface Serial ATA)

Tier 3 : Bandes et stockage optique

Les différents niveaux de stockage des serveurs d'entreprise



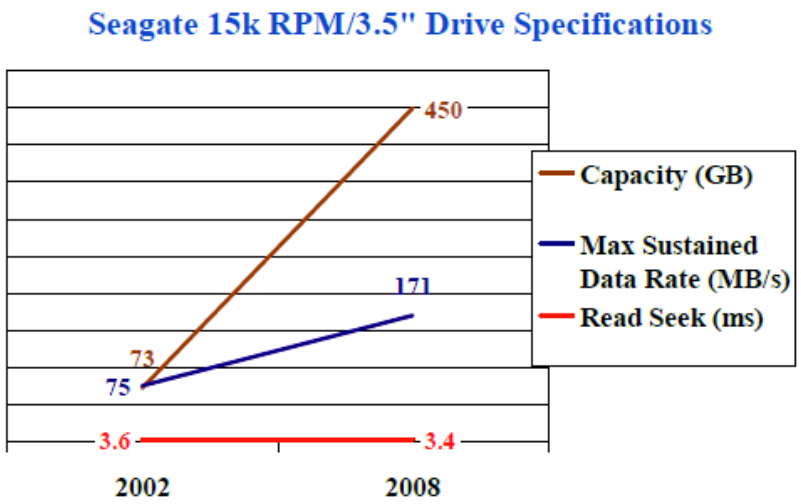


- L'emplacement des données permet de mettre en œuvre le tiering et d'optimiser la gestion des données au sein du système d'information.
- La combinaison de ces éléments constitue le socle d'une infrastructure de stockage dans le cadre de la mise en œuvre d'une infrastructure de PRA.

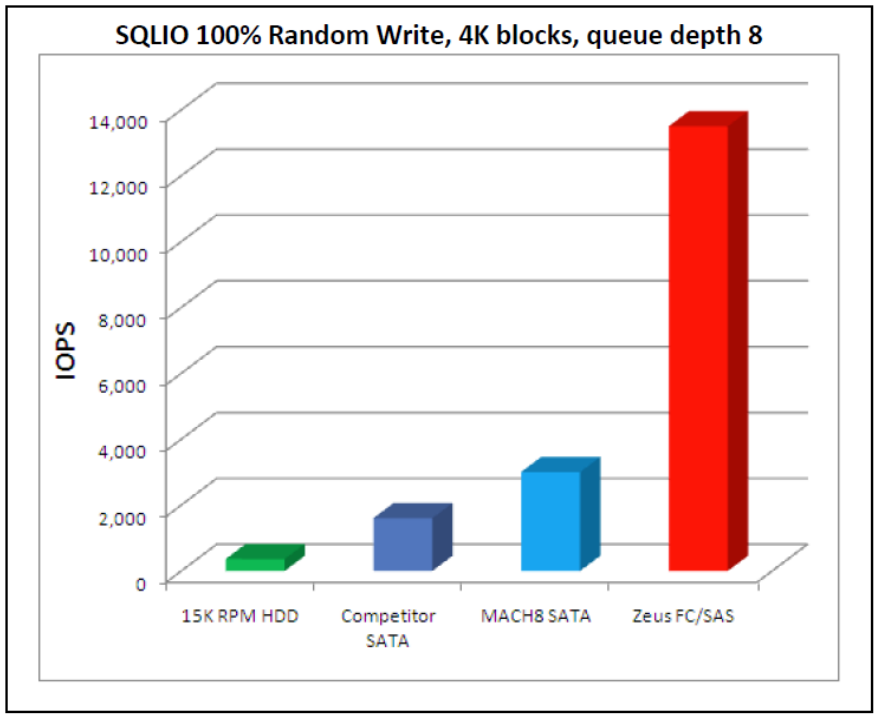
Nouveau métrique de référence

☞ Définition de nouveaux métriques pour positionner l'apport des SSD dans les serveurs d'entreprise… **prix par opération d'E/S par seconde**

(en HDD, le métrique de référence est le prix au Gigaoctet)



Les SSD de classe Entreprise sont beaucoup plus onéreux que les SSD de classe Grand Public moins performants et moins robustes.



Source STEC Inc.

Performances des SSD

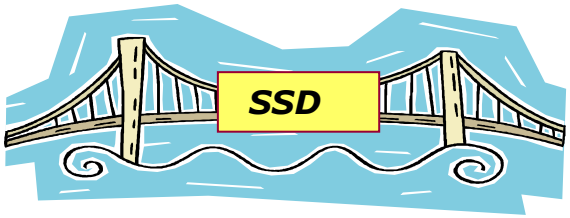
ns

Processors

Très, très, très, très, très rapide

Memory

Très, très, très rapide



Rapide

ms

Disk

Très, très lent comparativement

≤ 10 ns	~100 ns	~ 100.000 ns (100 µs)	900,000 -8,000,000 ns

Vitesse d'accès

Valeur médiane
~ 4.500.000 ns

~100 ms

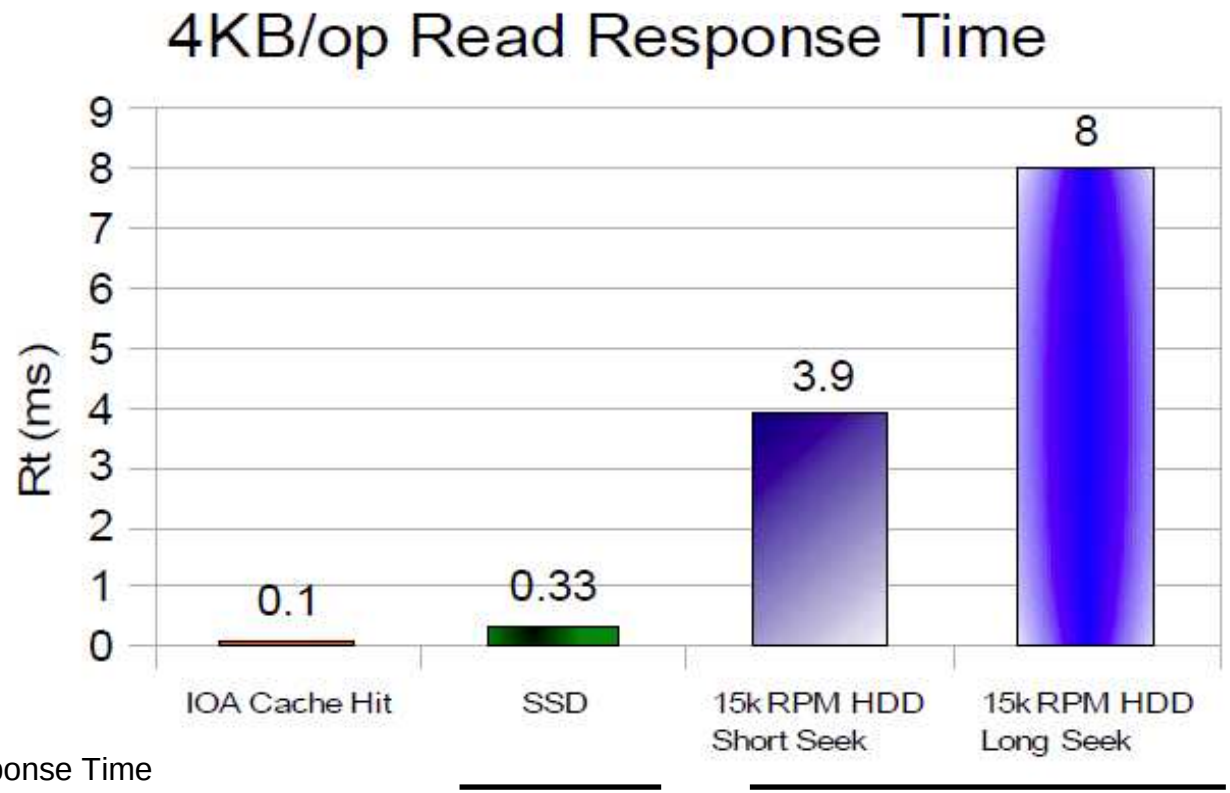
~1 seconde	~17 minutes	~ 12 heures 1/2

Contexte de temps homme

Mémoire vs Disque = x 8.000

Proc vs Disque = x 80.000

Temps de réponse des disques SSD et HDD en lecture



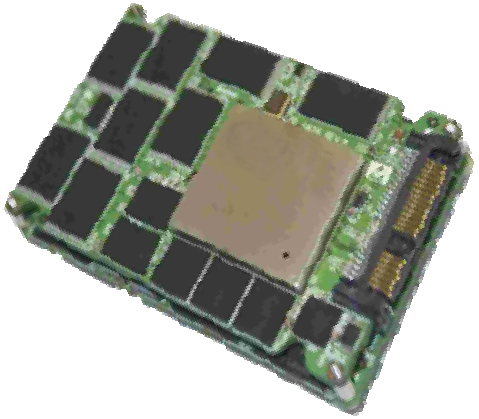
RT = Response Time

~ 28.000 op/s par unité

~ 150 à 200 op/s par bras

Le meilleur résultat est obtenu sur des travaux présentant un grand nombre de lecture aléatoires

Technologie des SSD

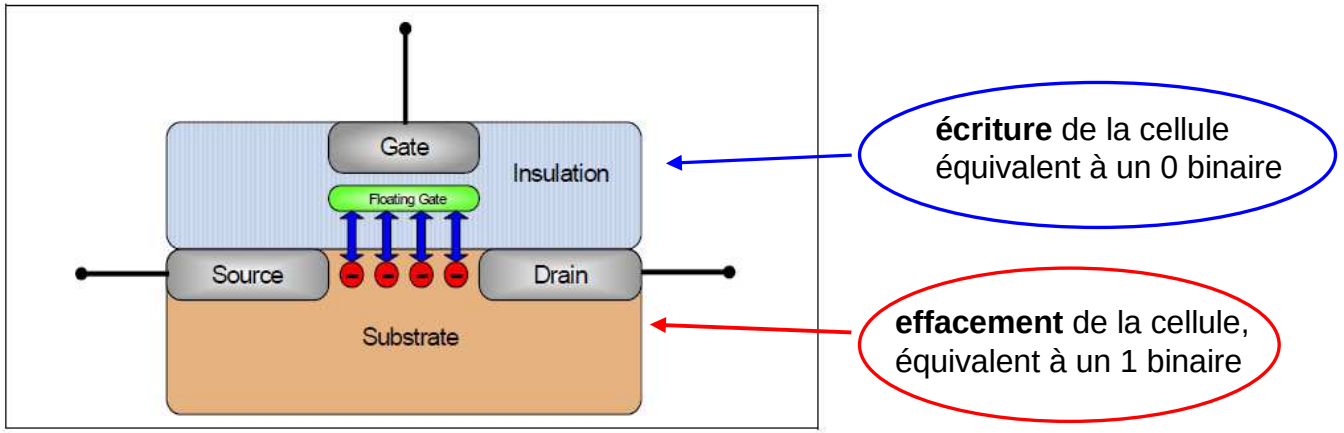


Les principes de la mémoire flash NAND (Not And)

Une mémoire flash stocke l'information dans un groupe de cellules programmables. Chaque cellule élémentaire est constituée d'un transistor à effet de champs ou effet tunnel (effet *Fowler-Nordheim),* capable de déplacer des électrons en fonction des courants électriques appliqués au semi-conducteur.

Ecriture …Tension de 12v sur la grille de contrôle et tension de 7v entre drain et source

Lecture … mesure de la résistance de la grille flottante par courant de 5v entre grille de contrôle et une des deux électrodes.



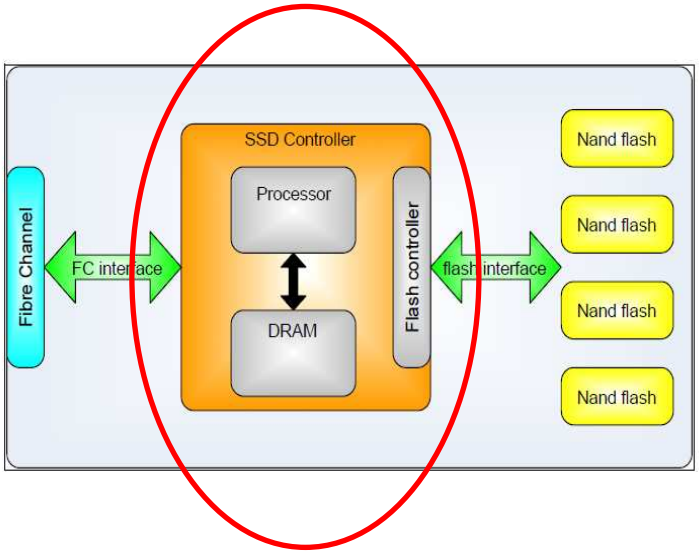
Le processus de lecture est plus rapide que celui d'écriture ou d'effacement car on ne doit pas remplir ou vider la grille flottante avec des électrons, ce qui explique la différence de performance qui existe entre lecture et écriture.

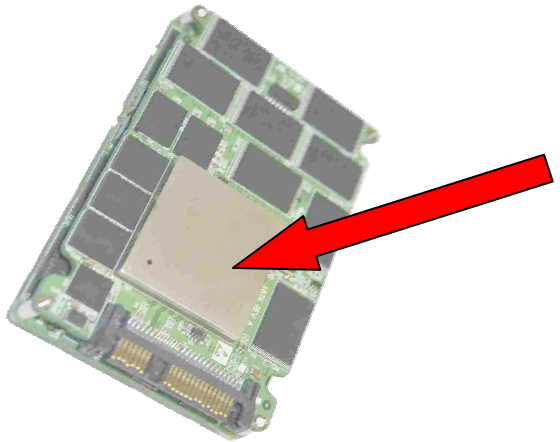
En technologie MLC (plus d'un bit par cellule), la quantité de courant qui circule est mesurée afin d'évaluer plus précisément le niveau de charge de la « floating gate » et déterminer la valeur des bits de la cellule.

La lecture des données dans une mémoire flash NAND

La mémoire flash NAND travaille avec un bus série, en accès séquentiel. Il est impossible d'accéder directement à un bit en particulier. Pour accéder à une information précise, on doit charger une partie des données (ou page) dans une petite mémoire DRAM assimilable à un cache et ensuite lire ce que l'on souhaite dans cette mémoire.

Cette mémoire contient aussi un répertoire de placement des blocs et gère l'équilibrage dynamique de l'utilisation des cellules *(Static and Dynamic Wear-Leveling)*





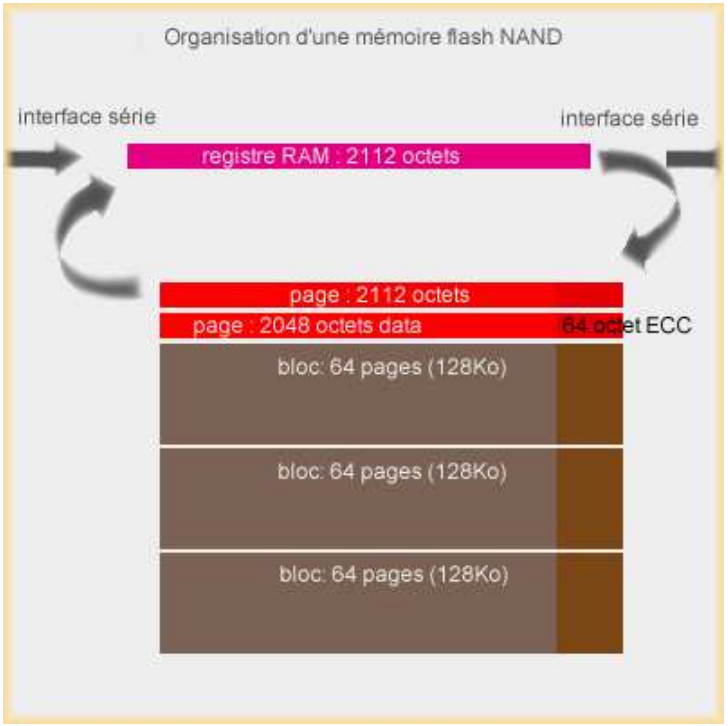
Le contrôleur s'occupe, entre autres choses, de gérer la copie de la page dans la mémoire DRAM *(shadowing)* et de la gestion des blocs et de leur intégrité.

La mémoire flash NAND et l'organisation en blocs

☞ **Lecture sur une page, écriture sur un bloc**

- la page (2048 octets de donnée + 64 octets ECC) est l'unité minimale pour la lecture
- écriture au niveau du bloc (64 pages ou 128 Ko):

  La moindre écriture oblige l'effacement du bloc de données complet (128 Ko) avant l'écriture d'une nouvelle valeur; d'où la différence de performance entre lecture et écriture.



Lecture ….

Accès à une page de données : ~ 25 µs

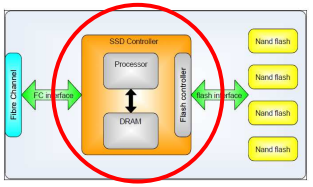
(temps de copie de la page dans la DRAM interne)

Accès aux autres pages résidentes du bloc : ~ 0.03 µs

Ecriture …

Effacement / écriture d'un bloc de 64 pages : ~ 2 ms

Le contrôleur interne



C'est la partie la plus importante de l'unité. Puce de type SoC *(System on a Chip),* il va gérer l'accès aux données mais aussi assurer la gestion de l'ensemble des cellules NAND

Les principales fonctions du contrôleur utilsé par IBM (autres que le transfert des données):

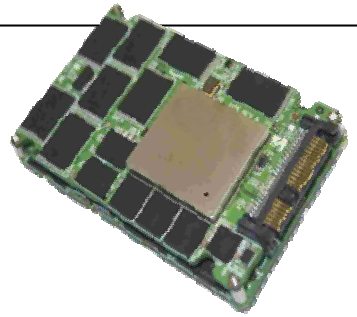
- ✓ **Algorithme de « Wear-Leveling »**
  - Dynamic Wear-Leveling
    distribution aléatoire des écritures sur les blocs libres ou effacés (usure équilibrée)
  - Static Wear-Leveling
    déplacement des données statiques en fonction de la fréquence d'utilisation des blocs
- ✓ **Algorithme du « Bad block »**

  détection des blocs défectueux lors de leur utilisation, en général dans le cycle effacement / programation et exclusion / remplacement dans la rotation des blocs
- ✓ **Algorithme de l' « Over Provisioning »**

  gestion d'une partie de la mémoire du SSD réservée et inaccessible à l'utilisateur (~75%) en combinaison avec la fonction « static wear-leveling »
- ✓ **Algorithme « Error Correction Code / Error Detection Code » *(ECC/EDC)***

  maintient l'intégrité de la donnée stockée en relisant la dernière donnée écrite avec déplacement /retrait complet du bloc si elle est altérée (EDC) et correction de la donnée en cours de lecture ou écriture avec gestion du bloc endommagé (ECC)

Mémoire SLC et mémoire MLC



Il existe deux types de mémoire NAND:

- la mémoire **SLC**, ou *Single Layer Cell* qui stocke un seul bit dans la grille flottante
  - *supporte en moyenne 100.000 cycles d'écriture*
- la mémoire **MLC** , ou *Multi Layer Cell* qui stocke plusieurs bits dans la même cellule
  - *on double la capacité de stockage en gardant la même taille physique*
  - *80% du débit d'une SLC en lecture, 50 % des performances en écriture*
  - *supporte en moyenne 10.000 cycles d'écriture*

Les données stockées dans ce type de mémoire persistent pendant une dizaine d'années sans être alimentée électriquement.

La durée de vie d'une mémoire flash

Une des particularités de la mémoire flash est sa durée de vie relativement limitée.

Les deux principales causes de ce phénomène ..

✓ l'oxyde utilisé pour séparer les grilles

Par construction, les électrons traversent cet oxyde en fonction des opérations d'écritures / effacements

De temps en temps, des électrons peuvent rester captifs de cet oxyde et perturber les lectures / écritures en provoquant un état indéterminé de la cellule

✓ la structure de la grille flottante et les tensions appliquées

Les différentes tensions appliquées à la grille (~12 volts en saturation pour écrire un 0 binaire) et 5 volts pour lire la cellule peuvent endommager la cellule avec le temps et la rendre inutilisable

Technologie NAND	Nombre d'écritures d'une cellule *
SLC	100.00 cycles
MLC	10.000 cycles

* Valeurs moyennes couramment admises

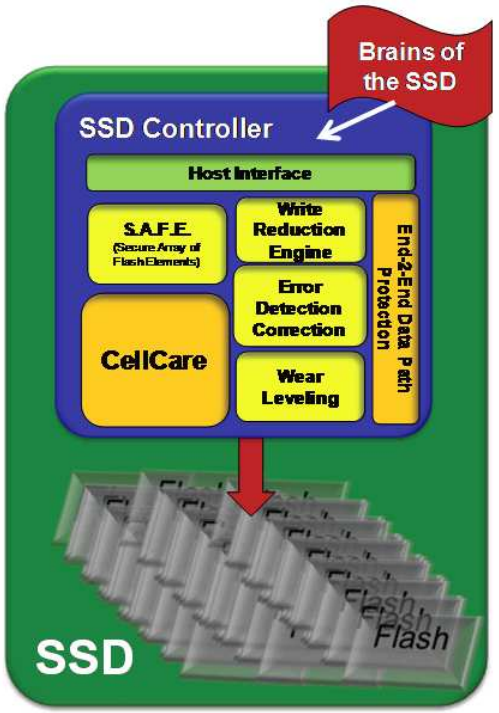
eMLC – Enterprise Multi-Level Cell - Technology

Performance et fiabilité de classe entreprise

PLUS technologie Flash MLC moins coûteuse

Comparée à l'offre 2009 des SSD 69GB sur Power Systems

- Meilleur coût au gigaoctet
- Packaging plus dense au gigaoctet
- Environ 50% de consommation électrique et calorifique par unité
- Performance comparable



eMLC – Enterprise Multi-Level Cell - Technology

eMLC notes

eMLC technology stands for "Enterprise Multi-Level Cell" Flash memory technology. IBM is the first server vendor to provide this new SSD technology option which blends enterprise class performance and reliability characteristics with the more cost effective characteristics of MLC Flash storage.

Using advances in both the SSD device controller Flash memory management plus advances in MLC technology itself, IBM can now provide much better cost on a per GB basis, much more dense physical packaging and about 50% less energy consumption and heat per drive than the IBM 69GB SLC SSD.

More impressively, eMLC does this while continuing to provide great sustained performance levels and extended endurance/reliability. For example, the new IBM eMLC SSD modules were designed to provide 24x7x365 usage even running write-intensive levels for at least five years. Typical client usage is expected to be much lower, especially regarding the average percentage of writes, and thus drive lifespan can be much longer.

Enterprise MLC Technology (eMLC) details

Prior to eMLC introduction, IBM Power Systems offered only an enterprise class SLC (Single Level Cell) technology. This SLC technology remains an important part of the current overall SSD product offering as it is the basis for SAS-bay-based SSD offerings for Power Systems.

MLC = 2 bits per cell, SLC = 1 bit per cell MLC can have more than 2, but sacrifices reliability/durability/shelf life to do so. IBM eMLC uses 2 bits per cell.

The “high levels” of writes mentioned in establishing a projected useful life were defined as around 70-80% writes and 30-20% reads. Write percentage can run higher, but at 100% writes, performance may or may not be impacted. If only a single SSD module is used per PCIe adapter instead of 2 or 4 modules then performance can be impacted if you try for 100%.

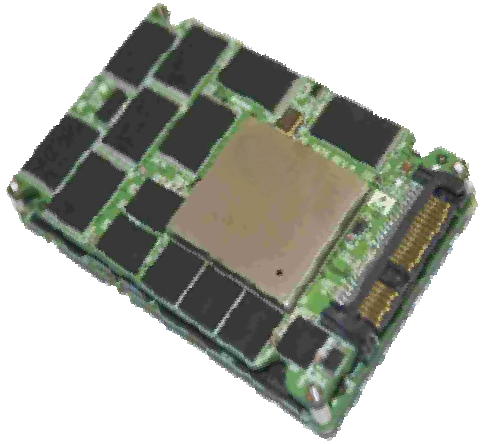
Data shelf life means you can plug a storage device which has not been used or installed in a server for some time and still retrieve the data contents. One of the ways eMLC gets its durability/speed is by providing a data shelf life of about 3 months without external power being supplied from the PCIe adapter to the SSD module. The 69GB SLC SSD module is designed with a much longer shelf life. Newer SLC products across flash industry seem to be adopting a shorter data shelf life as a good trade off in order to be more cost effective.

MLC technology seems to have more industry focus than SLC currently. Thus MLC seems to be better poised than SLC to take advantage of general technology enhancements.

eMLC SSD modules have about max of 3.5 to 4.5 W per 177GB SSD module …. about ½ that of the 69GB SSD module. The 3.5 to 4.5 W is also about ½ that of SFF disk drive.

IBM eMLC offering includes extensive ECC included in SSD module device controller. This is in addition to RAID formatting done by the PCIe adapter.

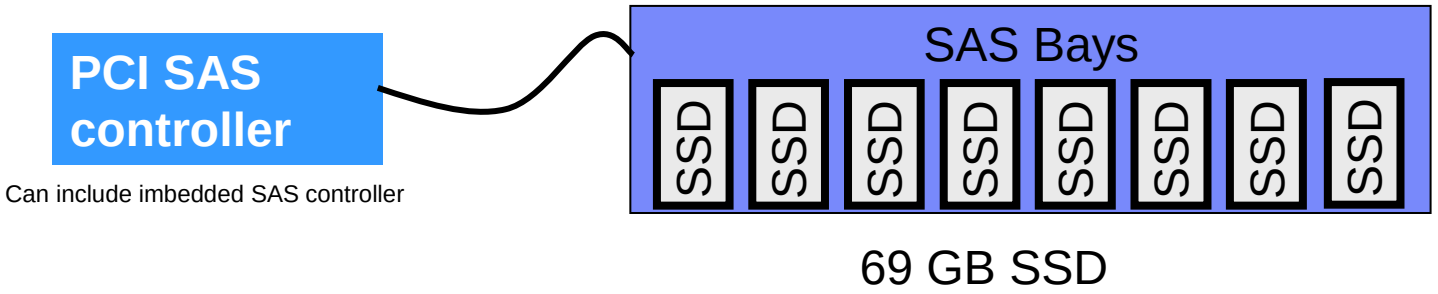
L’offre IBM



Options de configuration des disques SSD sur Power Systems

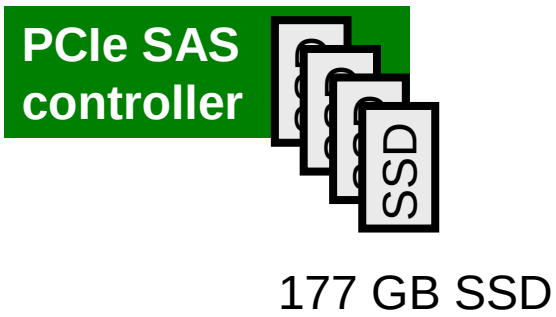
En baie SAS

– Offre 2009



En slot PCIe

– Offre 2010



Options de configuration des disques SSD sur Power Systems

En baie SAS

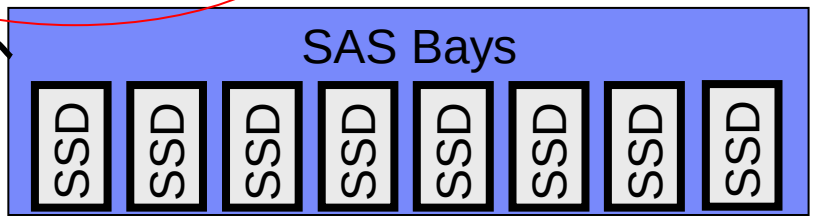
– Offre 2011

69GB SSD

177GB SSD

PCI SAS controller

Can include imbedded SAS controller

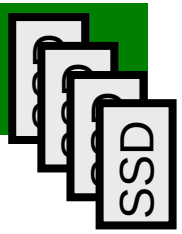


En slot PCIe

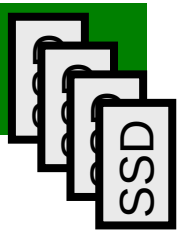
– Offre 2010

177GB SSD

PCIe SAS controller



PCIe SAS controller



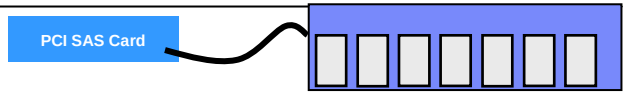
Disque SSD 177Go pour baie SAS dans tiroir #5803

- Pour le modèle POWER7 795

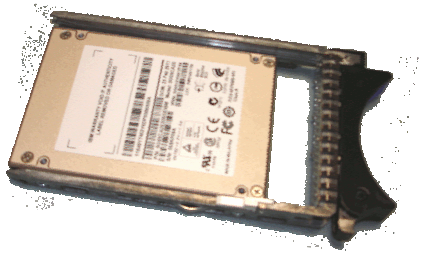
#5803 12X PCIe I/O Drawer

- Via carte contrôleur RAID PCIe #5805/5903
  - IBM i 6.1 or later (VIOS optional)
  - VIOS 2.2.0.

Disque SSD 177Go pour baie SAS



69GB SFF SSD



177GB SFF SSD

Plus Green

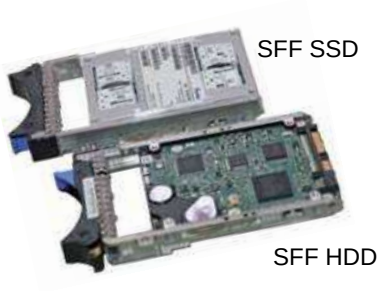
- 2,5 x Go par disque par rapport au disque SSD 69 Go

Plus économique

- 30% moins cher par baie SAS
  - $4700 contre $6882 par disque
- Presque 75% moins cher au gigaoctet
  - $26.6/Go contre $100/Go pour le disque 69 Go

Prices are USA suggested list prices as of April 2011 when ordered with the 795 server. Prices and are subject to change without notice. Reseller prices may vary.

Options de configuration des SSD en baie SAS



Les disques SSD s'intègrent à l'infrastructure HDD SAS

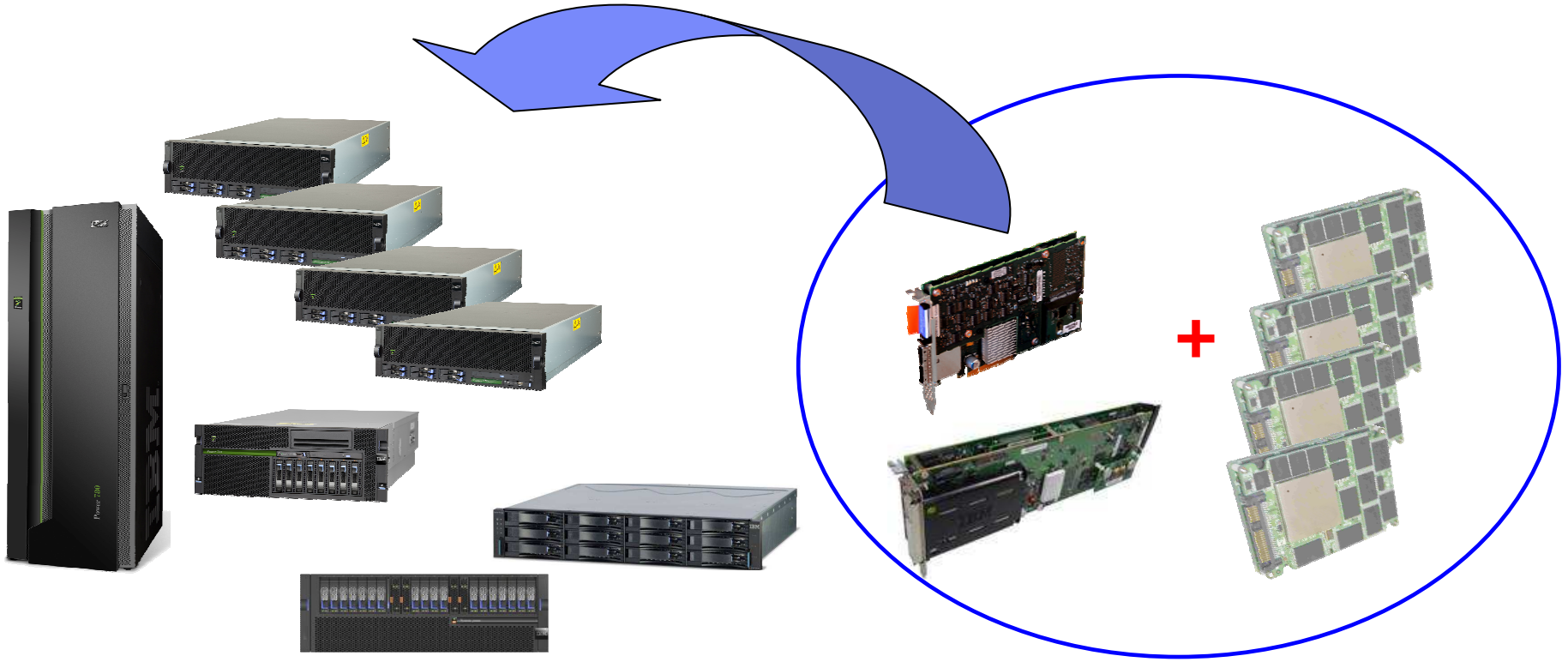
- Même logement que disque HDD
- Mêmes cartes contrôleur SAS que disque HDD
- Protection RAID ou miroir comme disque HDD
  - RAID-5, RAID-6, RAID-10, mirroring, hot spare
- Echange à chaud dans la baie SAS

Baies SAS des Power Systems

- BladeCenter JS23/JS43/PS70x
- Unités centrales POWER6 et POWER7
  - Mix SSD-HDD possible
- Tiroir disque #5886 EXP12S ,
- Tiroir 12X PCIe #5802/5803
- Pas de mix SSD et HDD dans les tiroirs I/O

Règles de configuration en Power Systems

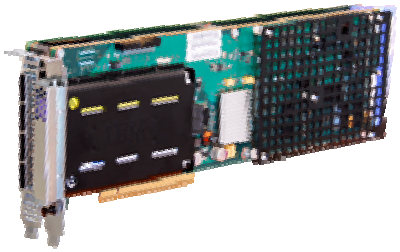
- Pour la performance en écriture, il faut associer les unités SSD à des cartes contrôleur Raid5 afin de bénéficier des caches en écriture (#5903 en paire, #5904, #5908, …)
- Limiter le nombre d'unités SSD sur les contrôleurs (de 4 à 8)



3 cartes SAS pour les disques SSD en baie SAS

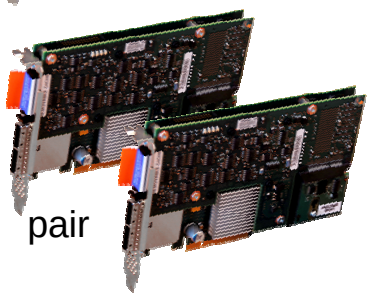
Les 3 options de carte contrôleur :

– La plus puissante : PCI-X 1.5GB Cache RAID Adapter
– Choix intermédiaire : PCIe 380MB Cache RAID Adapter
– Le moins che r: Contrôleur intégré*



PCI-X SAS RAID Adapter CCIN 572F

– Feat code #5904/5906/5908 (all same card, but 3 features indicate double-wide blind swap cassette)
– 1.5 GB effective write cache
– Read cache disabled using SSD
– Double wide adapter - uses 2 PCI-X slots



PCIe SAS RAID Adapter CCIN 574E

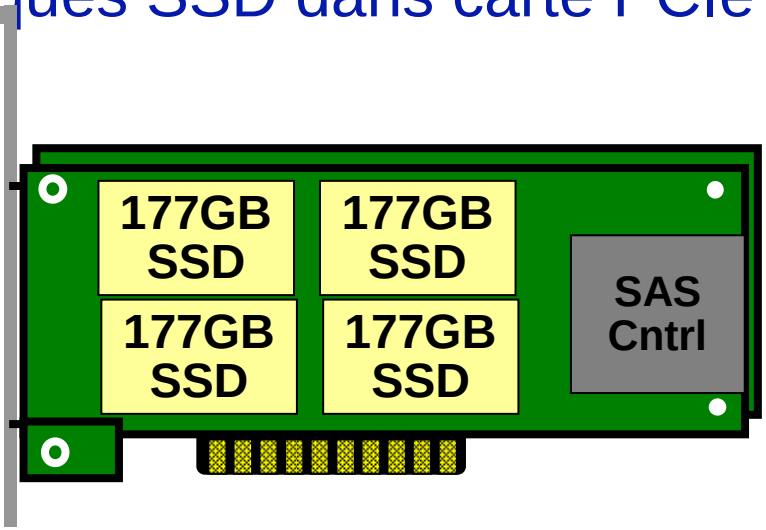
– Feat Code #5903 or #5805
– 380 MB write cache
– Used in pairs

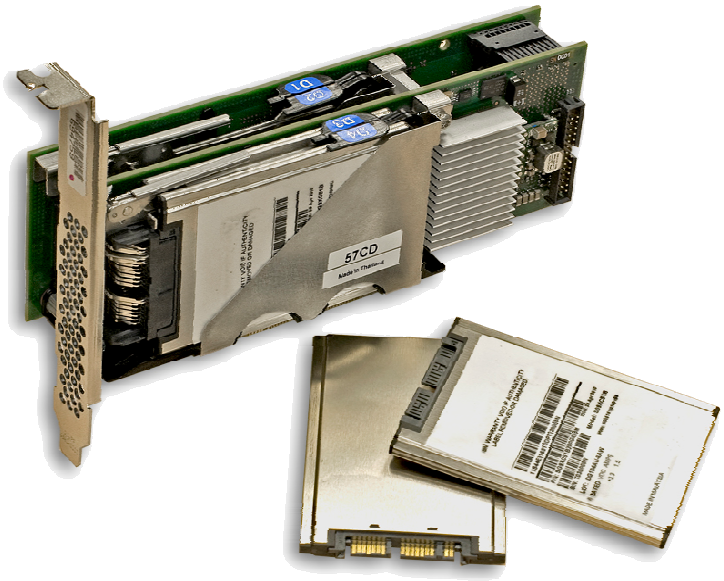
Integrated SAS Controllers

– Recommended augmented with cache
– Systems: 520 → 570 & 710 → 780

* Cache recommended

Disques SSD dans carte PCIe



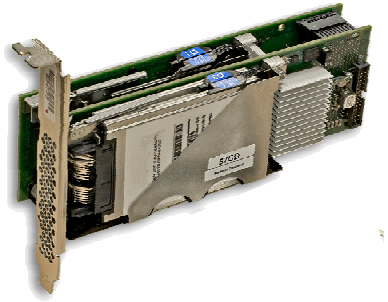


- **Carte SAS PCIe double**
- **1, 2 ou 4 modules SSD par carte**
- **177 Go par module SSD / Maximum 708 Go par carte**
- **Technologie eMLC**
- **IBM i 6.1 minimum**
- **Serveurs POWER7 (sauf 795)**

Protection des cartes PCIe-SSD

Point clé la carte PCIe doit être retirée du slot PCI pour remplacer un module SSD …

DONC SSD “hot plug” seulement si l’adaptateur PCI et ses modules sont en configuration miroir



Les modules SSD doivent être protégés au même titre que les unités de disque

- Première option de protection: Miroir via le système d’exploitation*
  - Adaptateurs redondants PLUS modules SSD redondants
    - » possibilité de hot plug … à choisir de préférence dans la plupart des situations
- Deuxième option de protection : RAID 5
  - L’adaptateur n’est pas redondant .. Pas de possibilité de modules SSD « hot plug »
  - Option Alternative: Ajout d’une unité « hot spare » (50% de capacité pour la protection)
- Troisième option de protection : RAID 6
  - L’adaptateur n’est pas redondant .. Pas de possibilité de modules SSD « hot plug »
  - 50% de la capacité des SSD affectée à la protection

*

SSD et stockage externe

DS8700/8800

DS8700 .. SSD 73Go (FC/LFF), 146Go(FC/LFF), 600Go (FC/LFF)

Protection en Raid 5, 128 SSD maxi, 16 par paire de DA

Release 5.1 pour support « Easy Tiering »

DS8800 … SSD eMLC 300Go (SAS/SFF)

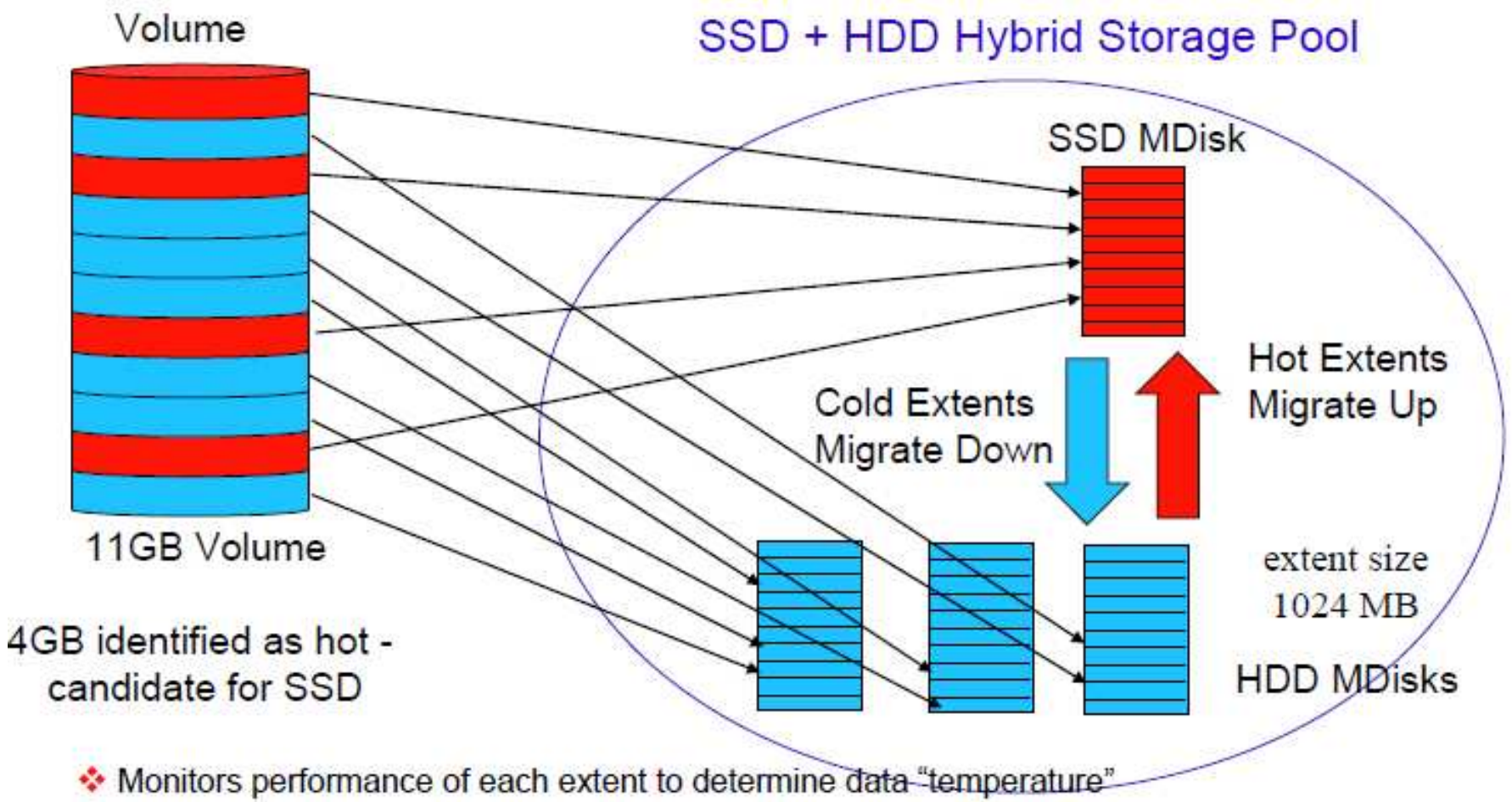
Release 6.0 et 6.1 (1H2011)

• *Utilisation de cartes #5735 de préférence et technologie P6+ ou P7*

Storwize V7000

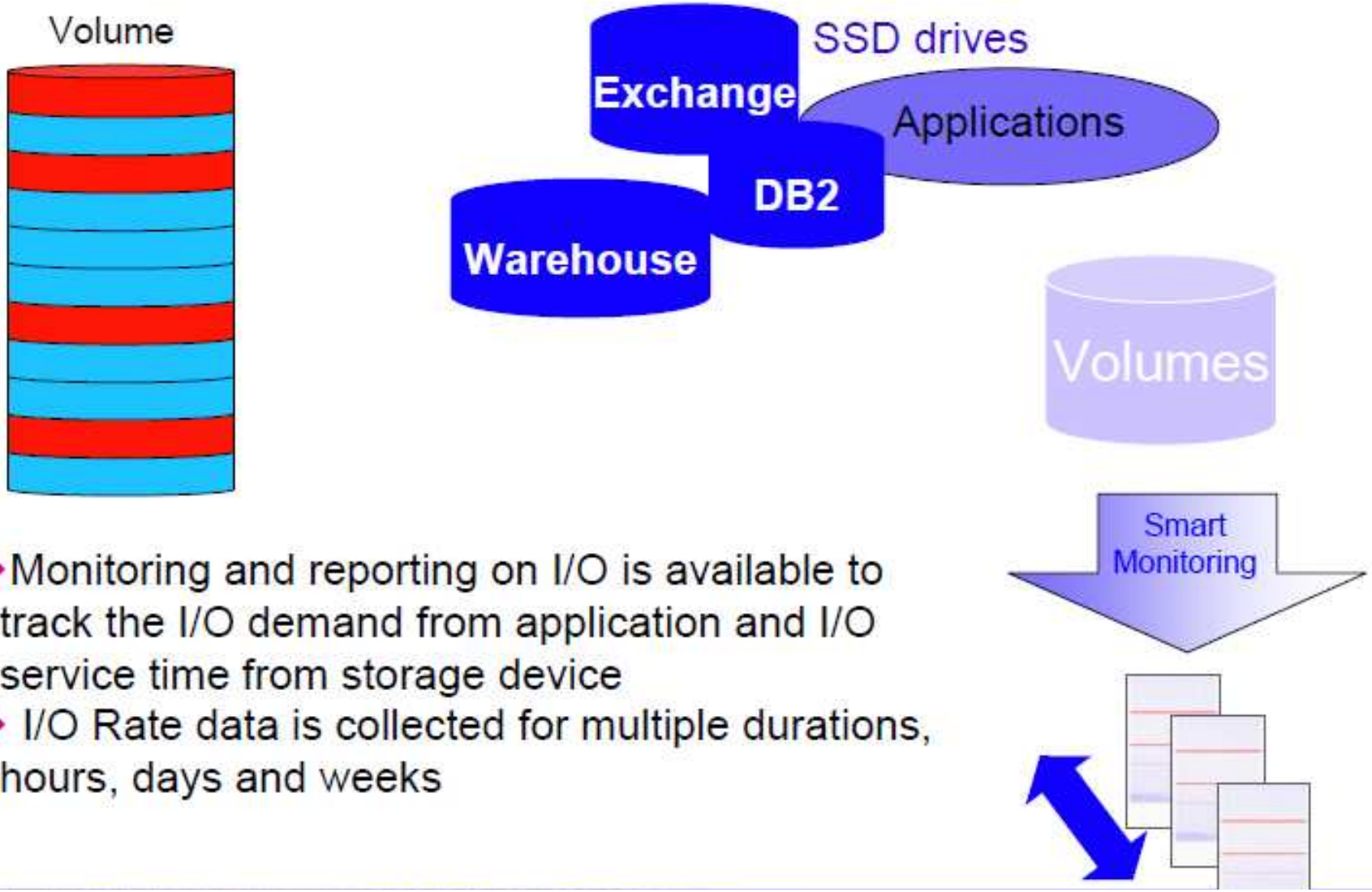
- Unités SSD eMLC 300 Go (SAS/SFF)
- Fonction « Easy Tiering » de base
- Interface graphique de gestion

Easy Tier – Automatic mode extent relocation



- Monitors performance of each extent to determine data "temperature"
- Creates migration plan for optimal extent placement every 24 hours
- Migrates extents within pool per plan over 24 hour period (limited number of extents chosen to migrate every five minute interval)

Easy Tier – I/O Rate Monitoring



- Monitoring and reporting on I/O is available to track the I/O demand from application and I/O service time from storage device
- I/O Rate data is collected for multiple durations, hours, days and weeks

Easy Tier – What's Hot/What's Not

What's Hot

Exchange

Applications

DB2

Warehouse

IOs below 64KB are considered best use
Of SSD-based MDisks. These are considered Random IOs.

Backups

User Directories

Multimedia

Large sequential data not accessed frequently. IOs above 64KB.

What's Not

IBM i et SSD

IBM i sur Power Systems … un délicat équilibre

POWER7 BUILT ON Power™

POWER6™ BUILT ON Power™

PROCESSEUR

POWER6+

POWER7

Management
Energy
Security
Availability
OS and Integration
Virtualization
Power™ Systems Software

DDR2

DDR3

MEMOIRE

Cartes PCIe

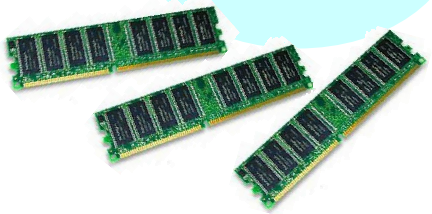
Boucles 12X DDR

Interface SAS

SFF 2.5inch

DISQUES

Batch !!!



PCI-X et PCIe

SSD !!

HDD

15 K rpm

Transitions technologiques majeures des Entrées/Sorties

1. Interface SCSI vers SAS

Disques SAS et SSD = 3.5-inch & SFF
Supports amovibles SAS & SATA

2. Adaptateurs PCI / PCI-X / PCI-X DDR vers PCIe

2008: PCIe disponible sur CEC 520/550/570 …, 2009: extension aux tiroirs d'E/S

3. Boucles RIO/HSL vers 12X

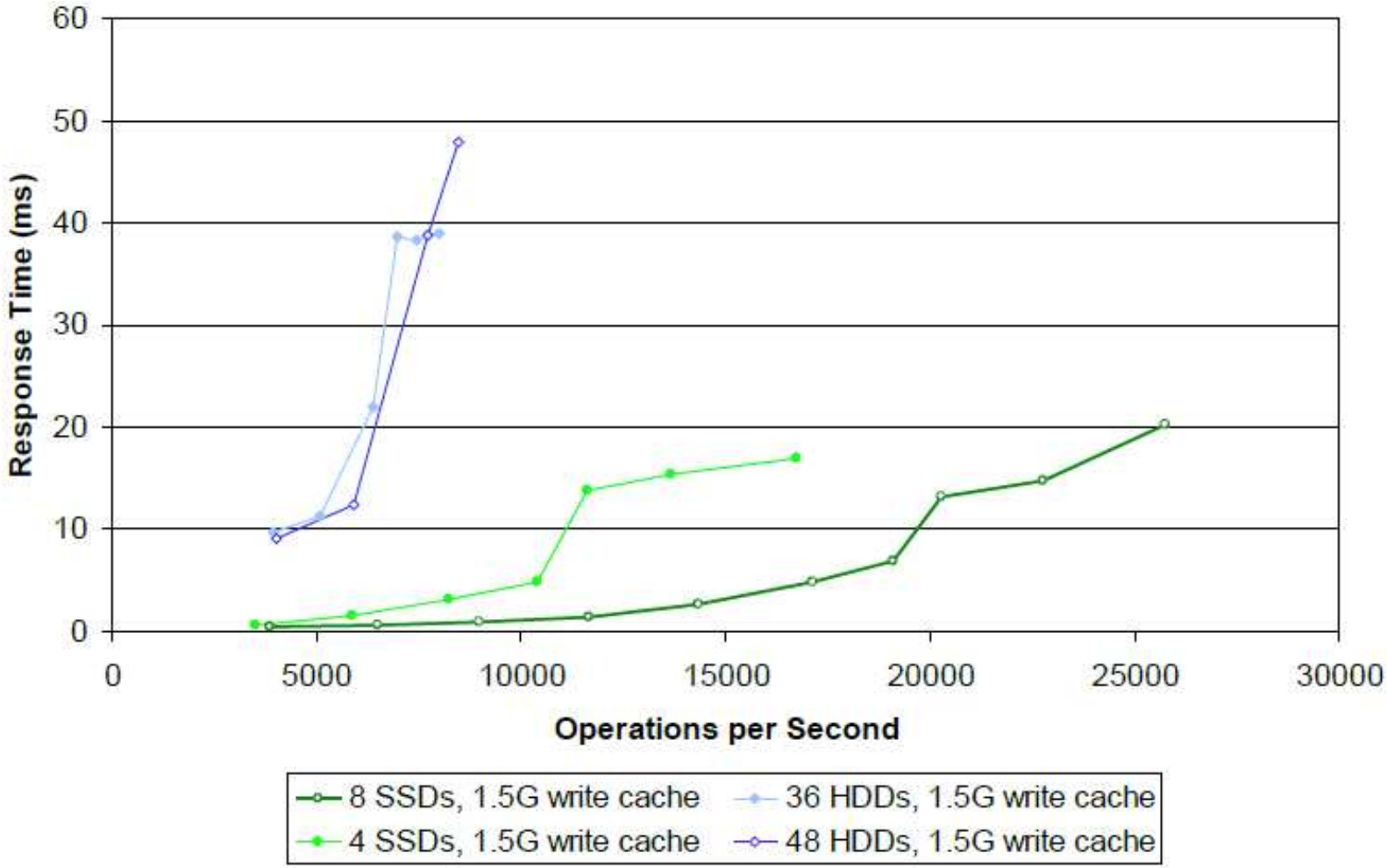
12X (SDR) et 12X DDR

4. IBM i : adaptateurs "IOP-based" vers "Smart IOA"

Plus d'IOP sur POWER 7

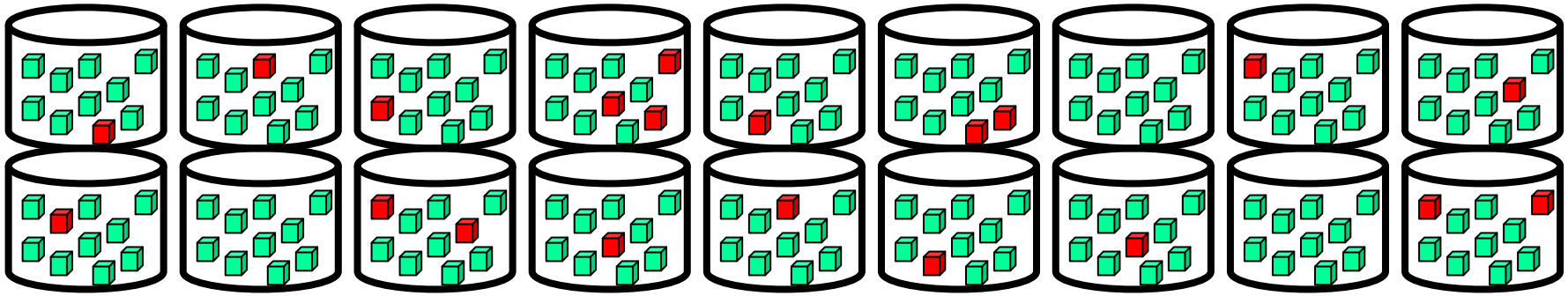
- Interface SCSI vers SAS

Comparaison SSD et HDD en environnement IBM i



Mélange SSD + HDD : une solution qui a du sens

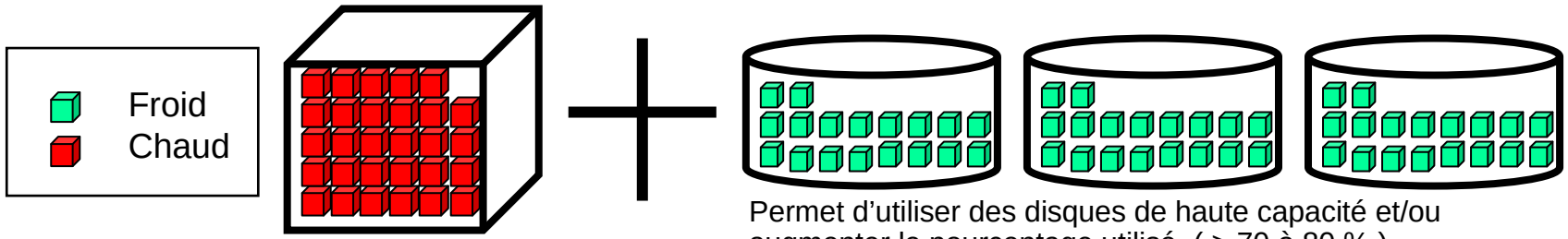
Par expérience, les bases de données ont un grand pourcentage de données qui sont utilisées à l'occasion ("froide") et un petit pourcentage de données utilisées fréquemment ("chaude")



Les données chaudes peuvent représenter 10 à 20 % de la capacité mais 80 à 90 % de l'activité

La solution SSD offre le meilleur rapport prix/performance quand elle est axée sur les données "chaudes"

La solution HDD offre le meilleur coût de stockage, axé sur la donnée "froide"…. approche de type HSM



SSD: proche de 100% de sa capacité

Permet d'utiliser des disques de haute capacité et/ou augmenter le pourcentage utilisé (> 70 à 80 %)

Placement des données “chaudes” et “froides”

IBM i

#1 La meilleure fonction automatique intégrée disponible dans l’industrie IT aujourd’hui

- La fonction “Trace and Balance” fait partie intégrante de l’IBM i
- Monitoring par partition ou ASP (Aux Storage Pool) pour déterminer les données chaudes ou froides
- A la demande, déplacement automatique des données chaudes vers les SSD et froides vers les HDD
- Processus à initier régulièrement

Les fichiers peuvent être placés automatiquement sur les SSD *(RSTLIB, RSTOBJ …)*

Certains objets spécifiques de type base de données peuvent être placés sur les SSD

Améliorations complémentaires de placement automatique intégrées dans la version IBM i 7.1

IBM i “Load Balancer”

Fonction totalement intégrée au système d’exploitation

Utilisation de la fonction “trace” pour collecter des données sur l’utilisation des partitions et ASP

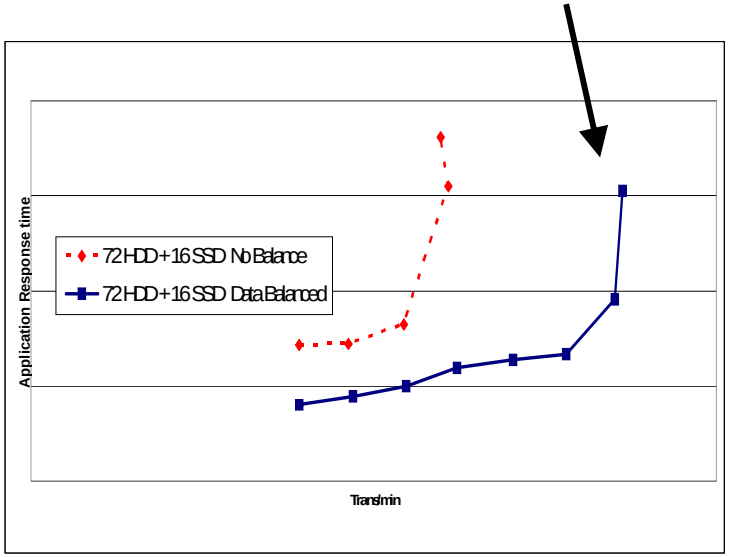
Activer la fonction trace pendant les périodes de pointe
- Désactiver cette fonction après une période de temps significative
- Impact négligeable sur la performance de la machine
- Utilisation des outils de performance pour identifier les données “chaudes”

Commande permettant automatiquement de déplacer les données « chaudes » vers les SSD et les données « froides » vers les HDD
- Peu d’impact en terme de performance, processus en arrière plan

Possibilité de “re-monitorer” et de “re-balancer” à n’importe quel moment
- Planification hebdomadaire ou mensuelle
- Probablement moins souvent si les données sont plutôt statiques

La fonction de placement des données chaudes/froides incluse dans le microcode de l’IBM i est plus performante que la distribution classique des données sur tous les disques du stockage.

Configuration: 72 HDD + 16 SSD



Si on veut analyser le % de données considérées comme “chaudes afin de déterminer le nombre de SSD nécessaire, il faut:
- Utiliser les résultats de l’outil PEX
- Interpréter les informations obtenues dans l’outil Performance Tools/400
- Utiliser l’outil (*) disponible sur le site IBM pour estimer l’intérêt d’implanter quelques disques SSD sur le serveur Power

(**) Download http://www.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/PRS3780

Amélioration du gestionnaire de stockage en IBM i pour les SSD

IBM i supporte la fonction de gestion de stockage hiérarchique

- Assure la collecte automatique des données de performance des E/S et gère le déplacement des données les plus sollicitées vers les “Solid State Drives” (SSD)
- ➢ Permet d’optimiser les investissements liés aux SSD à l’aide des amélioration de la gestion du stockage

DB2 for i supporte les SSD comme unités de prédilection

- Statistiques sur les lectures aléatoires DB2

Améliorations complémentaires pour la gestion des SSD

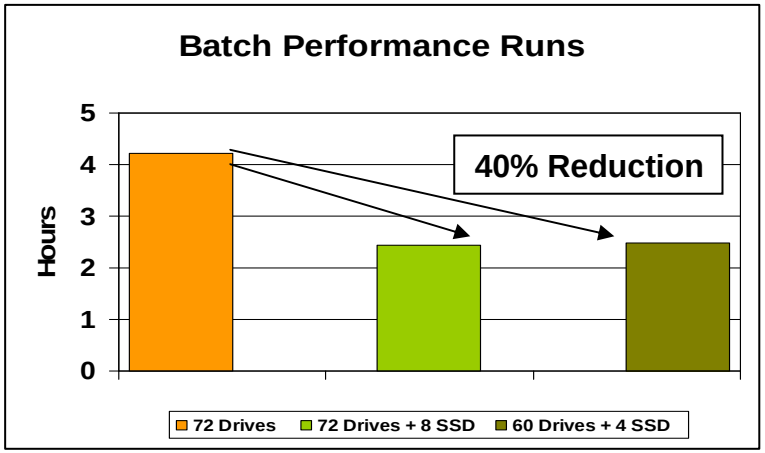
- Nouveaux utilitaires orientés SSD
- Possibilité de placer des objets de type IFS sur SSD
- Amélioration des instruments de performance
- Possibilité d’indiquer qu’une table à usage intensif ou un index soit spécifiquement positionné sur SSD:
  - **CRTPF** department **SRCFILE**(mjasrc/dds) **UNIT**(*SSD)
  - **CRTLF** departmentl **SRCFILE**(mjasrc/dds) **UNIT**(*SSD)
  - **CREATE TABLE** employee (c1 INT) **UNIT** SSD
  - **CREATE INDEX** employeeix **ON** mjatst.t2 (c1) **UNIT** SSD

« SSD Analyzer Tool »

- Utilitaire destiné à déterminer si l’utilisation de SSD permet une amélioration des performances applicatives
- Compatible i5/OS 5.4, IBM i 6.1 et 7.1 (**)

V7R1

“Associated Bank” réduit la durée de son Batch de 40% avec les SSD (*)

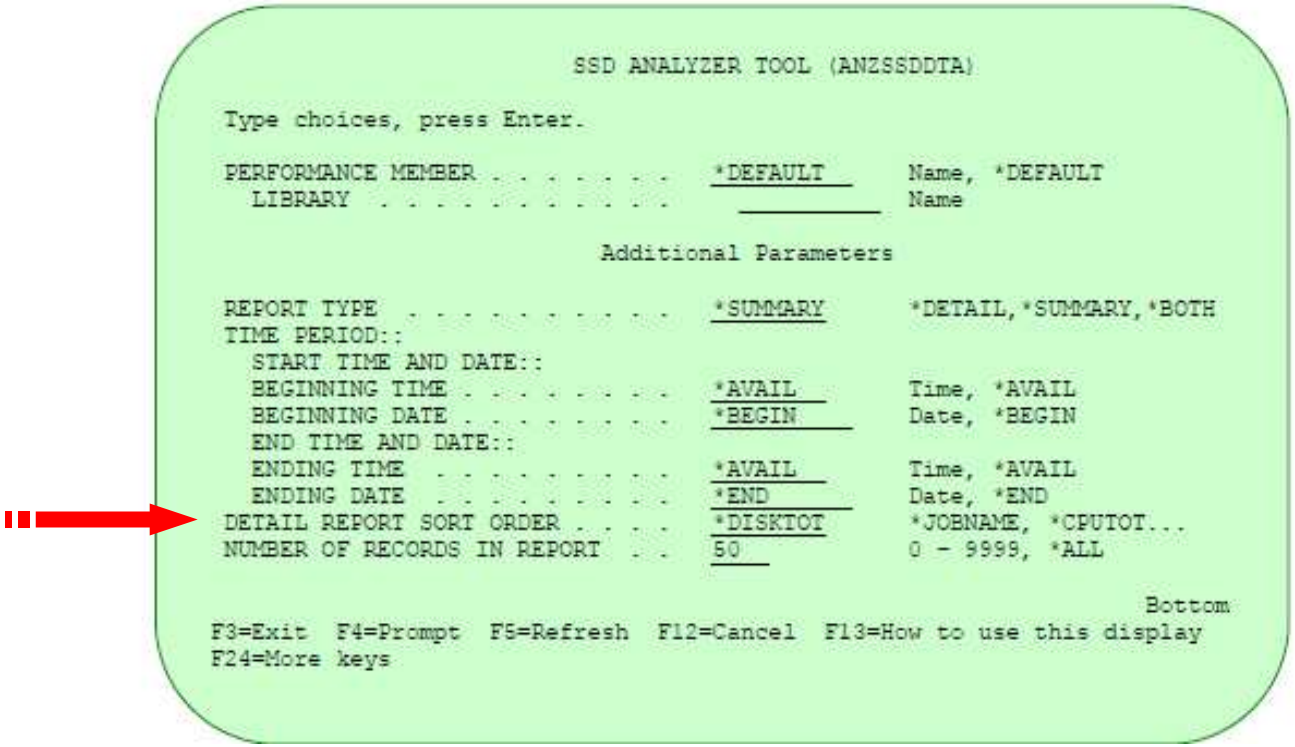


Les unités SSD peuvent améliorer les performances des longs traitements par lots et des requêtes tout en optimisant automatiquement le placement des données.

(*) http://www.ibmsystemsmagpowersystemsibmidigital.com/nxtbooks/ibmsystemsmag/ibmsystems_power_200909/index.php#/16
(**) Download http://www.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/PRS3780

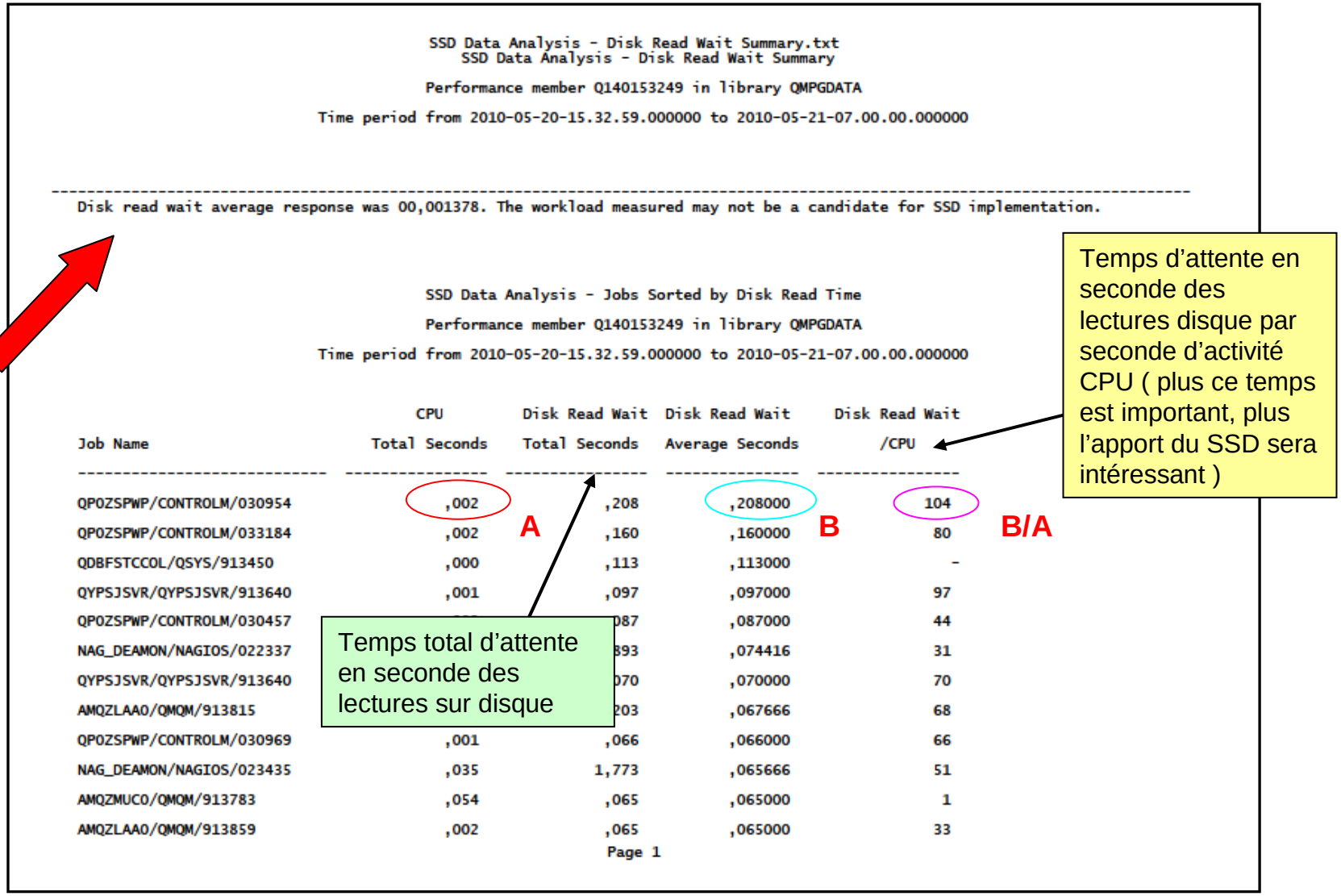
SSD Analyzer Tool for IBM i

- Rapide, simple, outil d'analyse gratuit disponible en V5R4 et plus
- Se présente comme un rapport standard de performance
- Donne une indication de type "probablement oui", "probablement non", ou "peut être »



Outil disponible sous forme de SAVF sur www.ibm.com/support/techdocs dans "Presentations & Tools". Faire une recherche sur le mot clef SSD. Attention à suivre le mode d'emploi, surtout en V5R4 *(QIBMSSD/SETUP)*

Conclusions de l'outil …



SSD Data Analysis - Disk Read Wait Summary.txt
SSD Data Analysis - Disk Read Wait Summary

Performance member Q140153249 in library QMPGDATA

Time period from 2010-05-20-15.32.59.000000 to 2010-05-21-07.00.00.000000

Disk read wait average response was 00,001378. The workload measured may not be a candidate for SSD implementation.

SSD Data Analysis - Jobs Sorted by Disk Read Time

Performance member Q140153249 in library QMPGDATA

Time period from 2010-05-20-15.32.59.000000 to 2010-05-21-07.00.00.000000

Job Name	CPU Total Seconds	Disk Read Wait Total Seconds	Disk Read Wait Average Seconds	Disk Read Wait /CPU
QP0ZSPWP/CONTROLM/030954	,002	,208	,208000	104
QP0ZSPWP/CONTROLM/033184	,002	,160	,160000	80
QDBFSTCCOL/QSYS/913450	,000	,113	,113000	-
QYPSJSVR/QYPSJSVR/913640	,001	,097	,097000	97
QP0ZSPWP/CONTROLM/030457		087	,087000	44
NAG_DEAMON/NAGIOS/022337		893	,074416	31
QYPSJSVR/QYPSJSVR/913640		070	,070000	70
AMQZLAA0/QMQM/913815		203	,067666	68
QP0ZSPWP/CONTROLM/030969	,001	,066	,066000	66
NAG_DEAMON/NAGIOS/023435	,035	1,773	,065666	51
AMQZMUC0/QMQM/913783	,054	,065	,065000	1
AMQZLAA0/QMQM/913859	,002	,065	,065000	33

Page 1

A

B

B/A

Temps d'attente en seconde des lectures disque par seconde d'activité CPU (plus ce temps est important, plus l'apport du SSD sera intéressant)

Temps total d'attente en seconde des lectures sur disque

Que choisir: SSD-SAS en tiroir ou carte PCIe ??

En 2010, les modules eMLC sur carte PCIe SAS présentent deux fois plus de capacité que les SSD SLC traditionnels, MAIS attention à bien analyser votre besoin (niveau de version, technologie des serveurs Power, protection souhaitée …)

SSD sur carte PCIe …

- Miroir: utilise 4 slots PCIe pour une capacité jusqu'à 708 Go
- RAID-5: utilise 2 slots PCIe pour une capacité de 531 Go, 304 Go si "hot spare"
- RAID-6: utilise 2 slots PCIe pour une capacité de 304 Go

SSD sur attachement SAS …

- Supporte le "Hot Plug" si baie SAS et contrôleur SAS séparés
- Paire d'adaptateurs SAS #5903, utilise 2 slots PCIe; jusqu'à 9 SSD
  - En protection RAID 5 = 557 Go dans un tiroir E/S 12X IB PCIe #5802
- Adaptateur #5904/5906/5908, utilise 2 slots PCI-X; jusqu'à 8 SSD
  - En protection RAID 5 = 388 Go dans une tiroir disque EXP12S #5886

Fonctions supportées

- Le nouvel adaptateur PCIe, de même que la solution SSD sur baie SAS, supporte la fonction d'IPL, en mode "load source" comme en mode "alternate IPL"

- Les SSD peuvent être utilisés en lieu et place des disques traditionnels dans le CEC. Ils peuvent même être une alternative (onéreuse !!) aux disques classiques sur la partition/serveur.

- Ce nouvel adaptateur PCIe, comme la solution SSD sur baie SAS, peut être utilisé par PowerVM (VIOS)

- **Fonction MIROIR:**
  - De même que les SSD classiques SAS 69 Go, les SSD 177 Go sur PCIe ne peuvent pas être en configuration miroir sur un HDD classique.
  - Le SSD sur adaptateur PCIe doit être en miroir sur un SSD de même taille lui même sur adaptateur PCIe.
  - Un SSD de 177 Go sur adaptateur PCIe ne peut pas être mis en miroir sur un SSD de 69 Go en baie SAS.

La technologie SSD en résumé

Réduit les délais: temps de réponse des transactions, durée de traitement des batchs

- *réduction des temps d'attente des E/S (de 10 à 100 fois)*

Réduit les infrastructures de stockage: diminution des coûts liés aux contrôleurs et unités de disque, énergie, refroidissement, occupation au sol ….

- *augmentation des possibilités des E/S par élément de stockage (dans un rapport de 2 à 4 fois), réduction du nombre de bras d'accès nécessaire à la performance, …*

Réduit l'infrastructure serveur: capacité des mémoires DRAM, coût et énergie

Green!

- *pagination très rapide des OS ou de l'hyperviseur (réduction dans un rapport de 2)*

Améliore la disponibilité: augmentation du MTBF et meilleure anticipation des erreurs

- *amélioration du temps d'établissement des points de synchro et des dumps, moins de composants électromécaniques, reconstructions de contexte plus rapide…*

Améliore la gestion du changement: réduit les temps d'arrêt pour maintenance et changements pour les systèmes sensibles

- *augmentation du nombre d'E/S par seconde, réduction des délais d'IPL …*

Emergence de nouvelles possibilités: nouvelles fonctions et applications possibles

- *Amélioration des performances, des coûts ….*

Glossaire …

- SCSI ... Small Computer System Interface (*technologie en fin de vie)*
- SAS … Serial Attached SCSI
- HDD …Hard Disk Drive
- SSD… Solid State Drive
- SAN … Storage Area Network
- NPIV … N_Port ID Virtualization
- VIOS … Virtual I/O Server
- SFF … Small Form Factor
- IOA … Input/Output Adapter
- IOP … Input/Output Processor
- Smart IOA … Intelligent I/O Adapter
- PCI-x … PCI eXtended (enhanced PCI card and slot)
- PCIe … PCI Express (latest and fastest enhanced PCI card and slot)
- HSL … High Speed Loop (POWER4 thru POWER6 I/O bus interconnect)
- RIO … Remote I/O -same as HSL, but called RIO when used on p system
- 12X … IBM's POWER System implementation of InfiniBand bus interconnect
- CEC … Central Electronics Complex *- refers to the processor enclosure for POWER Systems*