

# Performance système IBM i

**28 mars 2023**

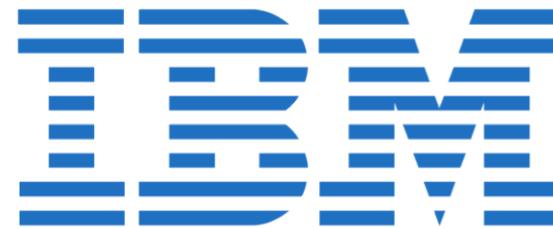
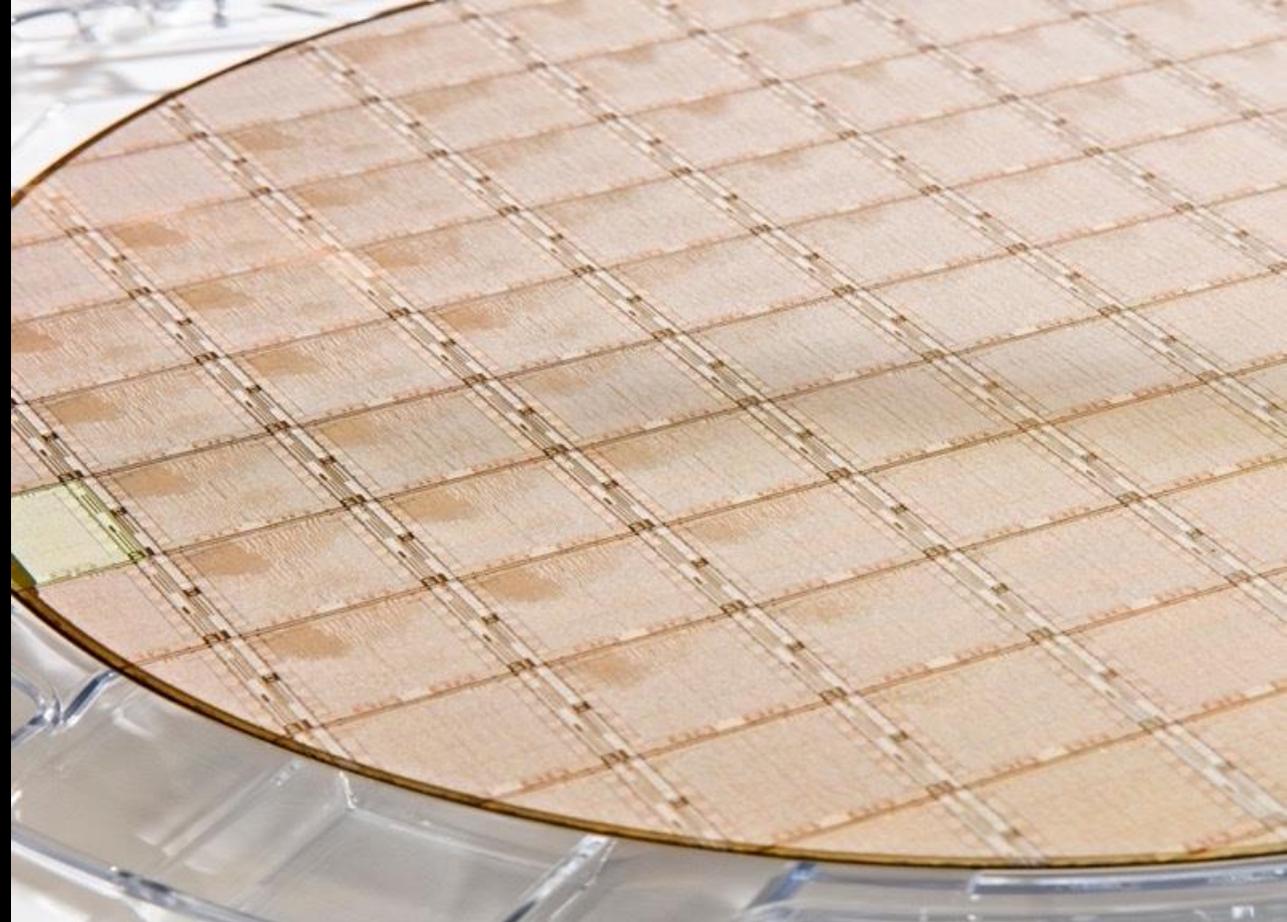
Common france

Comprendre la performance pour faire  
évoluer son système

Ce n'est pas parce que c'est rouge, ça brille et fait du  
bruit que c'est une Ferrari,  
cela peut être un Massey-Ferguson ....

Laurent Mermet

[laurent.mermet@ibm.com](mailto:laurent.mermet@ibm.com)



## Rappel de base sur les concepts

La CPU

La mémoire

Les disques

Stockage externe VS stockage interne

La virtualisation

Les métriques

Analyse de performances



## Connaître les mots clés :

- **VIRTUALISATION:** pouvoir faire fonctionner plusieurs systèmes d'exploitation en même temps sur un seul serveur physique, en faisant « croire » au système d'exploitation qu'il possède son propre matériel, elle peut être matérielle ou logicielle
- **PROCESSEURS** Un processeur est un dispositif qui traite les instructions programmées. Plus le nombre de processeurs affectés à une partition logique est élevé, plus le nombre d'opérations simultanées que la partition logique peut exécuter à tout moment est important.
- **Processeurs dédiés (DP):** Les processeurs dédiés sont des processeurs entiers affectés à une seule partition logique.
- **Processeurs partagés (SP):** Les processeurs partagés sont des processeurs physiques dont la capacité de traitement est partagée entre plusieurs partitions logiques. La possibilité de diviser les processeurs physiques et de les partager entre plusieurs partitions logiques est connue sous le nom de technologie Micro-Partitioning®.
- **Processeurs virtuels (VP):** Un processeur virtuel est une représentation d'un cœur de processeur physique pour le système d'exploitation d'une partition logique qui utilise des processeurs partagés.
- **SMT:** Le multithreading simultané permet de partager les processus afin d'exécuter deux applications ou deux threads de la même application en même temps.
- **CPU Queuing:** La mise en file d'attente de l'unité centrale se produit lorsque vous exécutez plus de threads que le nombre de VP multiplié par le niveau de SMTx. C'est le nombre de VP multiplié par le niveau SMTx. Si vous avez 4 VP sur un Power9 (SMT8 par défaut), vous pouvez au maximum exécuter  $4 * 8 = 32$  threads en même temps sans avoir de file d'attente CPU..

## Connaître les mots clés :

- **MEMOIRE** Les processeurs utilisent la mémoire pour conserver temporairement les informations. Les besoins en mémoire des partitions logiques dépendent de la configuration de la partition logique, des ressources d'E/S attribuées et des applications utilisées.
- **PAGING:** La pagination est le terme général pour désigner le mouvement des données dans et hors de la mémoire, à la fois de manière synchrone et asynchrone. Les pannes(Fault) provoquent la pagination sur le serveur..
- **FAULTING:** La défaillance se produit lorsqu'une page référencée, ou un élément de données, n'est pas en mémoire. Les programmes s'arrêtent alors car ils doivent attendre que les données soient mises en page. La défaillance est un comportement normal sur IBM i.
- **Pools de mémoire:** Les pools de mémoire sont des subdivisions logiques de la mémoire physique. Utilisés avec les sous-systèmes pour isoler l'utilisation de la mémoire par différentes applications. Il en existe deux types : les pools partagés et les pools privés
- **Sous-systèmes:** Environnements d'exécution prédéfinis au sein d'un seul IBM i. Permet la séparation de la charge de travail. Contrôle du flux de travail avec les pools de mémoire, permet un certain contrôle de l'utilisation de la mémoire, et de regrouper l'utilisation des ressources dans les données de performance

## Connaître les mots clés :

- **Espace adressable unique:** Définition du stockage IBM i, où les pages de mémoire peuvent se trouver dans le stockage primaire (RAM) ou dans le stockage secondaire (disque), et que l'emplacement physique d'une page n'a pas d'importance pour un processus. Il faut considérer la mémoire comme un cache géant pour le disque
- **Stockage:** Ensembles des équipements nécessaires à la conservation des données informatiques.
- **Stockage interne:** Ensemble de disques pilotés par un contrôleur de disque attaché directement au serveur (carte Raid) et fournissant différents niveau de protection (RAID 0, 1, 5, 6, 10)
- **Stockage externe:** Ensemble de disques pilotés et contrôlés par un système autonome ayant sa propre intelligence et fournissant différents niveaux de protection communiquant soit par FC ou IP (baie de disque, NAS, stockage Cloud)
- **ASP System (Auxiliary Storage Pool):** Espace de base de l'environnement I contenant le microcode, L'OS et les LPP, et données utilisateurs (ID 1)
- **ASP Base :** espace contenant des données utilisateurs séparées de l'ASP system (ID de 2 à 32)
- **IASP Independent Auxiliary Storage Pool:** permet de partager le stockage entre des systèmes, de mettre en ligne du stockage supplémentaire et, à l'inverse, de mettre offline du stockage non nécessaire, et indépendamment de la structure ASP(ID de 33 à 255)

## Connaître les mots clés :

- **JOB:** Représente un "processus" sur l'IBM i .Orienté vers l'utilisateur, contrôle par l'utilisateur. Contient au moins un thread (initial) et beaucoup contiennent des threads secondaires supplémentaires
- **THREAD:** Unité de travail indépendante exécutée dans le cadre d'un travail et utilisant les mêmes ressources que ce dernier. Les threads ont certains de leurs propres attributs, tels que la pile d'appels. Les travaux "multithreadés" sont un modèle typique des applications Java ou Domino.
- **TACHES:** Les tâches du système s'exécutent en arrière-plan pour le compte d'un travail demandeur. Pas de contrôle par l'utilisateur. Généralement non visibles, sauf via WRKSYSACT, les outils IBM i Performance ou les outils de service.
- **USERS:** Un utilisateur est une personne, pas un profil, ni un programme, ni une session ou autre qui accède à l'IBM i et échange des informations d'identification directement avec l'IBM i ou avec une application s'exécutant sur IBM i
- **LICENCE:** La « licence » est également appelée « droit » d'usage (« entitlement »)
- **Processor Group:** Plusieurs niveaux de facturation en fonction de la taille du serveur **Petit** : P05/P10/P20 **Moyen/Large** : P30
- **LPP: Licensed Program Products** La plupart des LPPs sont commandés séparément du système d'exploitation IBM i avec un code produit. Les LPPs sont disponibles en fonction de la Version/Release de l'IBM i. Le mode de licence dépend du LPP: par système, par utilisateur, par cœur

Rappel de base sur les concepts

## La CPU

La mémoire

Les disques

Stockage externe VS stockage interne

La virtualisation

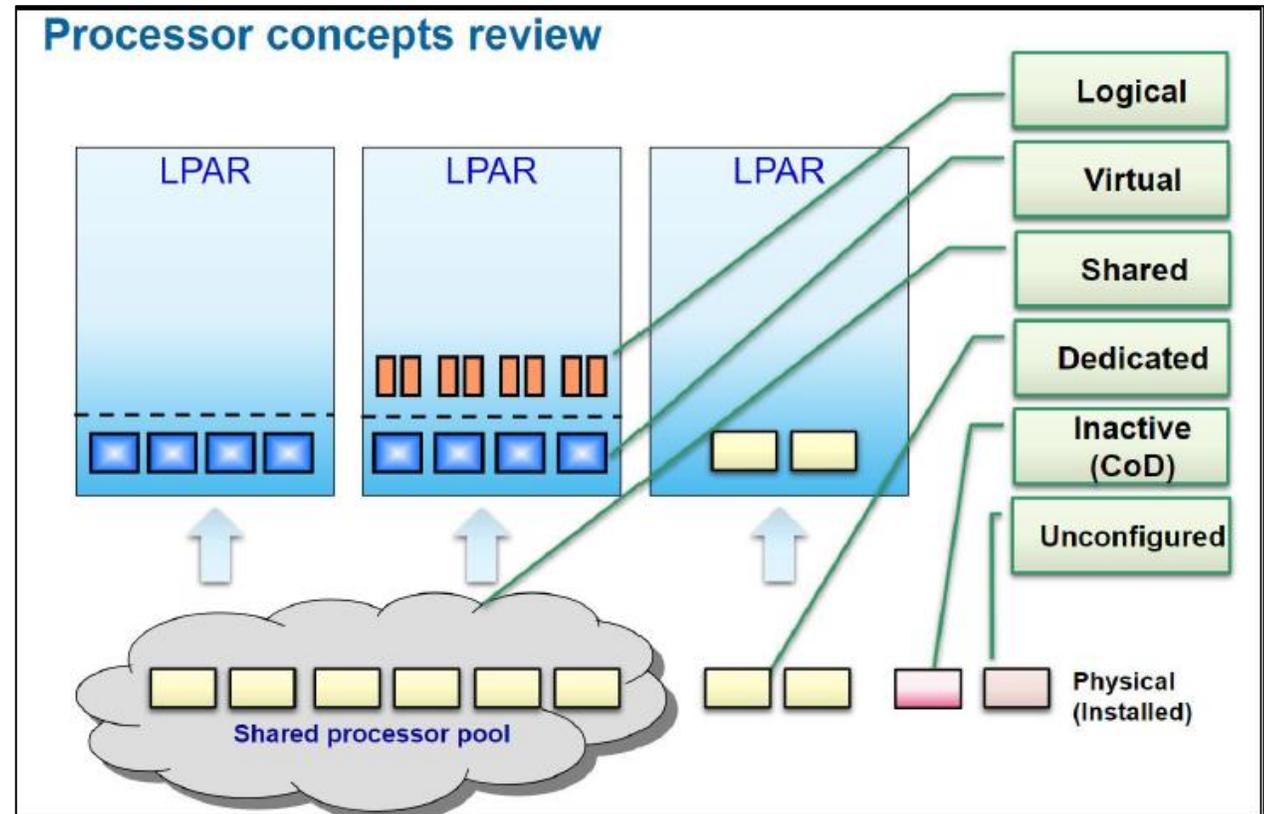
Les métriques

Analyse de performances



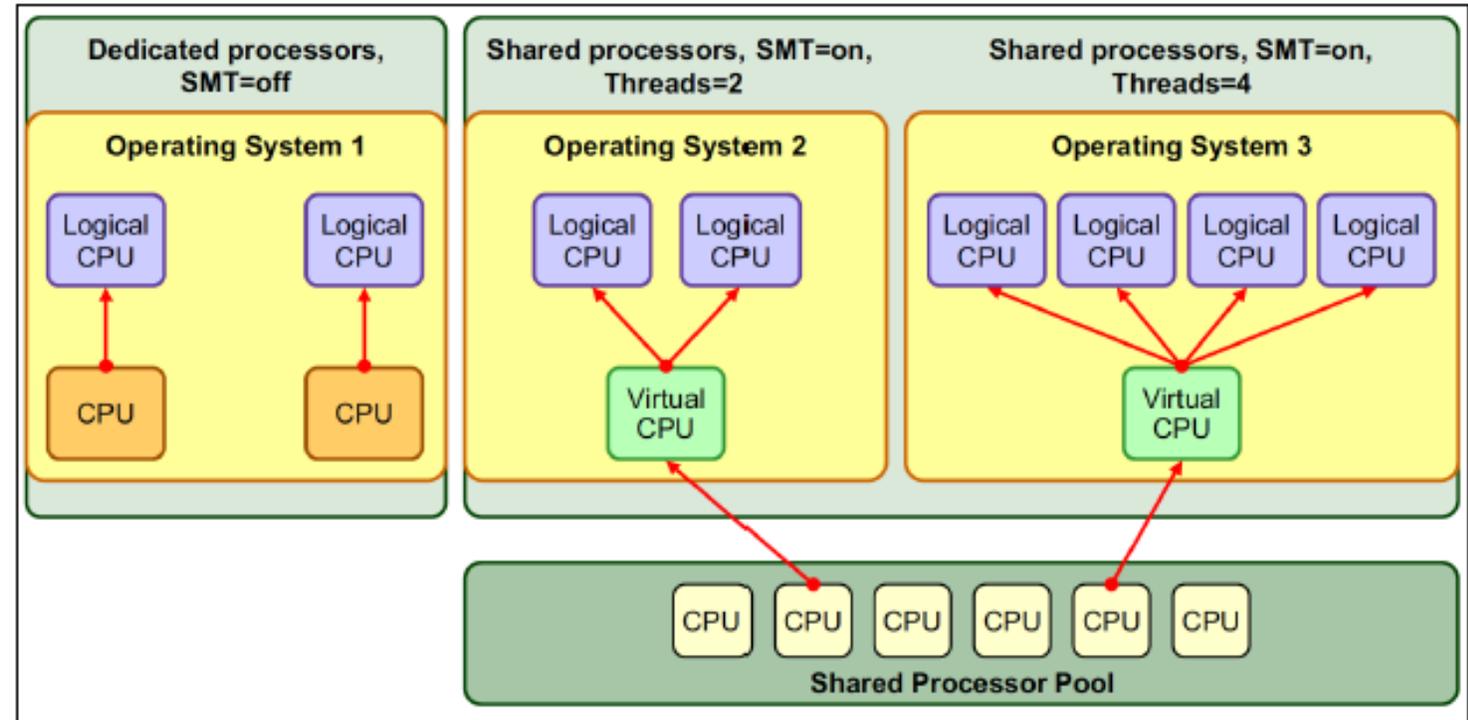
# MICRO-PARTITIONNEMENT : PROCESSEURS PARTAGÉS

- La capacité du processeur est attribuée en unités de traitement à partir du pool de processeurs partagé :
  - La quantité garantie de la partition est sa capacité autorisée (EC entitled capacity).
- Chaque partition est configurée avec un pourcentage de temps d'exécution pour chaque tranche de temps de 10 ms (fenêtre d'exécution).
- Chaque processeur virtuel donne accès à un seul processeur physique du pool.
  
- Un LPAR peut avoir des processeurs dédiés ou partagés.
  
- Les processeurs dédiés sont réservés au LPAR et ne peuvent pas être utilisés par d'autres LPAR, à moins que les paramètres ne soient reversés au pool de processeurs partagés.
  
- Un LPAR peut avoir des processeurs plafonnés (Capped) ou non plafonnés (Uncapped)
  - Les processeurs plafonnés sont limités par l'EC (Entitled Capacity)
  - Les processeurs non plafonnés sont limités par la VP (Virtual Processors and Shared Processor Pool).



# MULTITHREADING SIMULTANÉ ET MICRO-PARTITIONNEMENT

- Le multithreading simultané peut être utilisé avec les micro-partitions.
- Avec le multithreading simultané, chaque processeur virtuel exécute deux threads (POWER6), quatre threads (POWER7) ou huit threads (POWER8 POWER9 Power10).
- Chaque thread est appelé processeur logique.
- Même la plus petite partition aura toujours accès à un processeur complet.
- La EC minimale est de 0,05 = 0,5ms par 10ms.



- Exemple LPAR1 :1,6 EC sur Power10
- LPAR UNCAPPED
- Deux processeurs virtuels
- Multithreading simultané activé SMT8
- 16 processeurs logiques

- Exemple LPAR2 :0,9 EC sur Power10
- LPAR CAPPED
- Deux processeurs virtuels
- Multithreading simultané OFF
- 1 processeur logique

# SHARE PROCESSOR POOL ET PROFIL DE PARTITION

Pool Name ^	Pool ID ^	Reserved Processing Units ^	Maximum Processing Units ^
DefaultPool	0	N/A	N/A
IBM i	1	0	4

Processor Mode :

Shared

Shared Processor Pool :

IBM i

Available Processing Units in Pool :

4.0

Capped ◀ ▶ Uncapped

Weight :

128

## Virtual Processors

Maximum 4

Allocated 1

Minimum 1

## Processing Units

Maximum 4

Allocated 0.1

Minimum 0.1

## Advanced Settings

Processor Compatibility Mode

Effective :

POWER9\_Base

Pending :

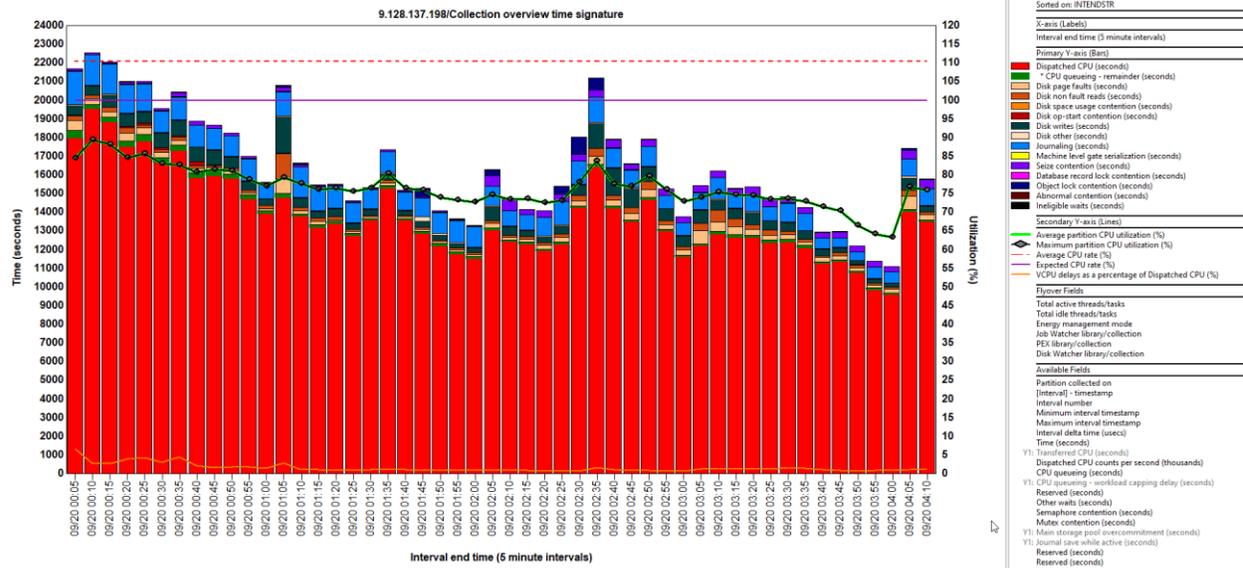
default

# UTILISATION ÉLEVÉE DU CPU

Lorsqu'une utilisation élevée du CPU est observée. Observez-vous :

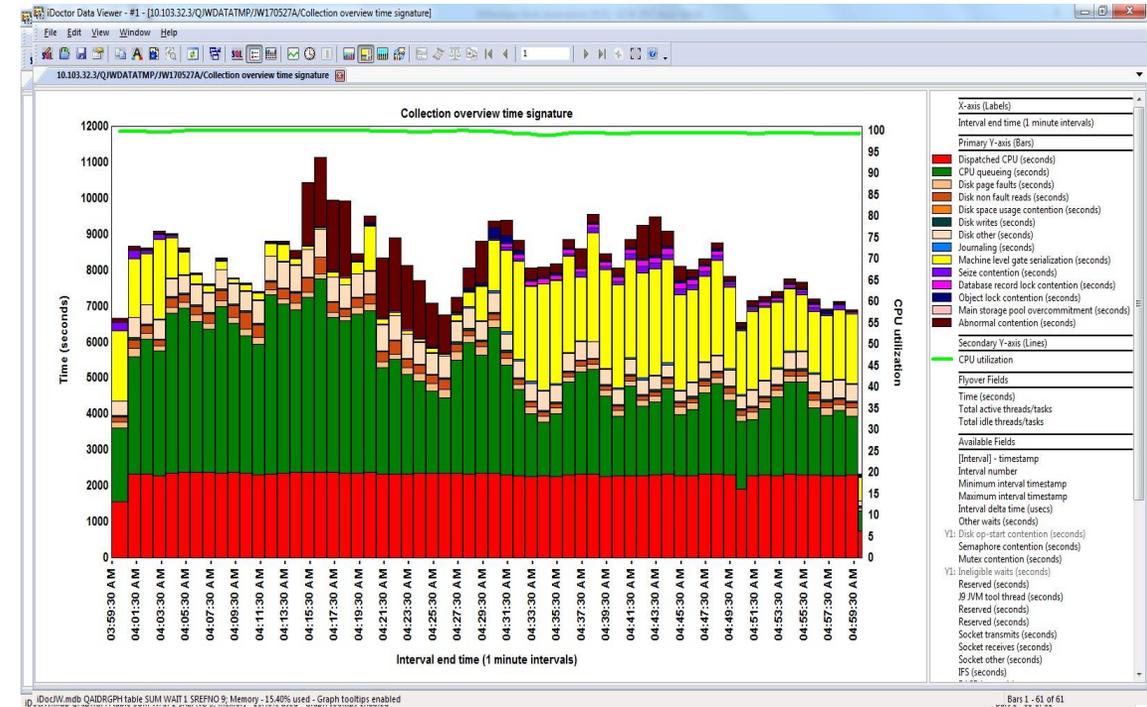
- CPU queuing?
- Abnormal contention
- Seizes and locks
- Machine level gate serialization
- D'autres attentes colorées

Si ce n'est pas le cas, l'utilisation élevée de l'unité centrale peut être justifiée.



Exemple d'un trop grand nombre de tâches en cours d'exécution ??

P8 7.1, SMT4, EC 10, VP 10, dedicated, capped



Exemple d'une CPU bien utilisée mais d'autres problèmes ???

P9 7.1, SMT8, EC 16, VP 20, shared, uncapped

# SYSTEM VALUE QQRYDEGREE

- **La valeur par défaut de QQRYDEGREE est \*NONE**
  - Le programme de licence SS1 option 26, DB2 Symmetric Multiprocessing permet l'utilisation du traitement parallèle lors de l'exécution de SQL/Query.
  - Fréquemment observé pour être changé en \*OPTIMIZE
  - De nombreuses installations n'ont jamais eu beaucoup de SQL. Au fil des ans, ce dernier s'est développé, de sorte que l'option \*OPTIMIZE au niveau du système peut s'avérer agressive.
- **Apprendre de \*OPTIMIZE au niveau du système**
  - Toutes les requêtes peuvent obtenir beaucoup de ressources et s'exécuter au mieux.
  - Cela peut influencer le reste de la partition
  - Examinez si \*OPTIMIZE est le bon paramètre pour votre partition.
  - Vous pouvez également utiliser QAQQINI et avoir un paramètre de degré de requête modéré.
  - Laissez les requêtes importantes pour la production bénéficier des bons paramètres.

# PARAMETRAGE DU PROCESSOR MULTITASKING

Changer la valeur système QPRCMLTTSK à 0 SMT OFF - **NON RECOMMANDÉ**

Changer la valeur système QPRCMLTTSK à 2                      System controlled, default

```
CALL PGM(QWCCHGPR) PARM(X'00000002') /* SMT2 context */
```

```
CALL PGM(QWCCHGPR) PARM(X'00000004') /* SMT4 context */
```

```
CALL PGM(QWCCHGPR) PARM(X'00000008') /* SMT8 context */
```

```
CALL PGM(QWCCHGPR) PARM(X'00000000') /* SMT OFF context */
```

Le système est réglé sur le niveau par défaut

- SMT4 for P8 7.1,
- SMT8 for P8 7.2 and 7.3,
- SMT8 for P9, SMT8 P10

Récupérer la valeur actuelle du SMT via l'API QWCRTVPR. Exemple de programme CL :

[https://www.ibm.com/docs/en/i/7.4?topic=ssw\\_ibm\\_i\\_74/apis/qwcrtvpr.html](https://www.ibm.com/docs/en/i/7.4?topic=ssw_ibm_i_74/apis/qwcrtvpr.html)

# CHANGEMENTS D'APPLICATION NÉCESSAIRES POUR UTILISER P9 OU P10? IBM

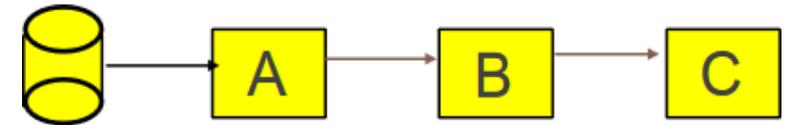
- Les programmes RPG à thread unique doivent-ils être modifiés ? → **NON**
- Faut-il des programmes multithreads ? → **NON**
- Comment faire fonctionner une application multithread sur IBM Power 9

**vous devez exécuter plusieurs tâches simultanément**

Un cycle CPU en SMT8

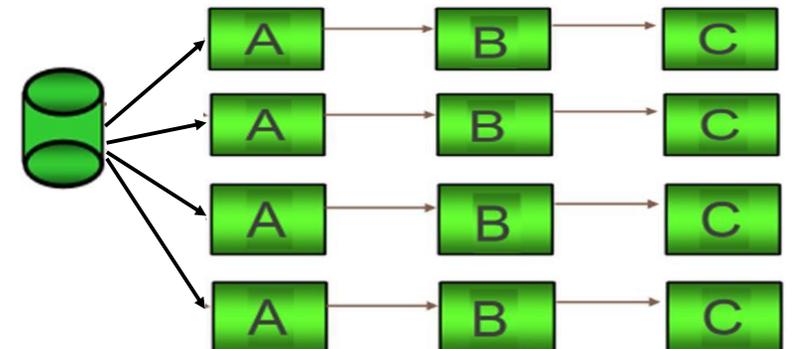
Instruction provenant de		Thread
Job xxx,	RPG à un seul thread	0
Job yyy,	un thread de la JVM multithreadée	1
Job yyy,	un thread de la JVM multithreadée	2
Job yyy,	un thread de la JVM multithreadée	3
Job zzz,	un thread de la JVM multithreadée	4
Job aaa,	RPG à un seul thread	5
Job bbb,	un thread d'un RPG multithreadé	6
Job bbb,	un thread d'un RPG multithreadé	7

Mono thread:



Diviser le flux d'entrée en plusieurs tâches simultanées de manière à ce que chaque tâche travaille sur des parties distinctes des données d'entrée.

multi threads:



Rappel de base sur les concepts

La CPU

## La mémoire

Les disques

Stockage externe VS stockage interne

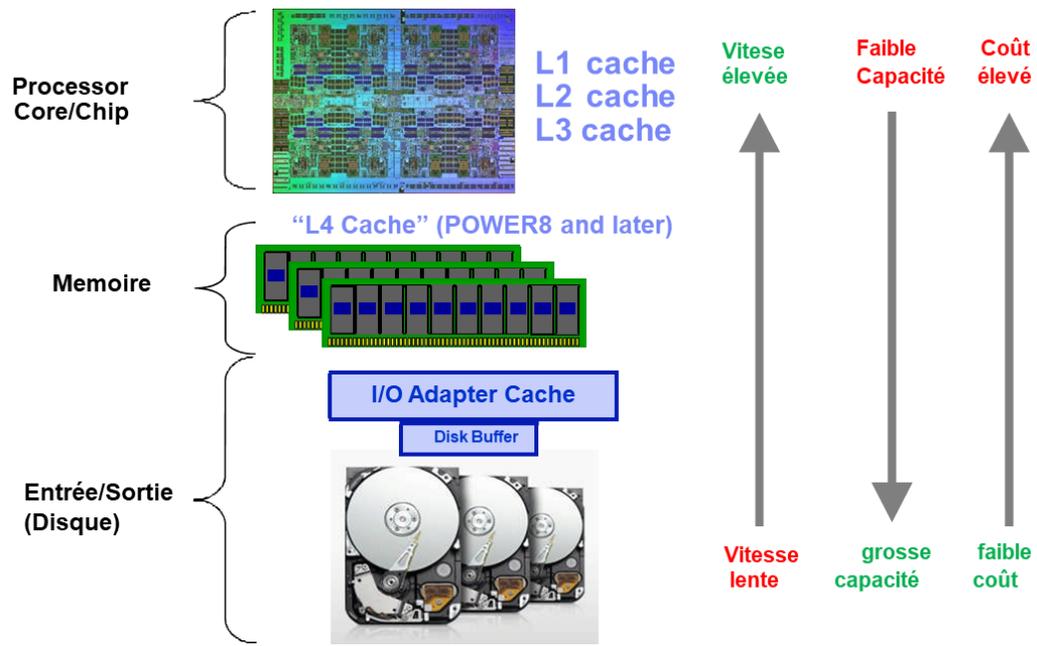
La virtualisation

Les métriques

Analyse de performances



# HIÉRARCHIE DE LA MÉMOIRE ET DU STOCKAGE POWER



- Combien de fois est-il plus lent d'accéder à quelque chose en mémoire que dans le cache d'un cœur de processeur ?
- Combien de fois est-il plus lent d'accéder à quelque chose sur un disque que dans la mémoire principale ?

	Size	Speed / cycles
POWER9 L1 cache	32K instr + 32K data per core	~3 cycles
POWER9 L2 cache	512 KB per core	~10 cycles
POWER9 L3 cache	120 MB shared per chip	~30 cycles
POWER9 "L4 cache"	16 MB per memory chip*	~200 cycles
Internal Memory	10s of GB per core	~300-800 cycles (~100ns)
Solid State Drives	100s of GB per drive	<1 ms
Hard Disk Drives	Can be TBs per drive	1-5 ms

Range due to "distance" of access.

1 ms = 1000 μs = 1,000,000 ns  
400 cycles is about 0.1 μs if 4.0 GHz

Pour introduire des données dans la mémoire, il existe deux types de transferts de mémoire

## • IMPLICIT

- les "défauts de page" et les "défauts de page en attente d'E/S".
- Le système d'exploitation ou l'application tente d'accéder à une adresse qui ne se trouve pas dans la mémoire principale.
- Elle doit attendre que les données soient introduites dans la mémoire à partir du disque.

## • EXPLICIT

- "Lecture sans erreur »
- Le système d'exploitation récupère les données avant que l'application n'en ait besoin (Db2 "read-ahead" ou "asynch brings").
- La commande SETOBJACC place un objet dans un pool de mémoire.
- Lecture de la partie des opérations CRTDUPOBJ et CPYF

# HIÉRARCHIE DE LA MÉMOIRE ET DU STOCKAGE POWER

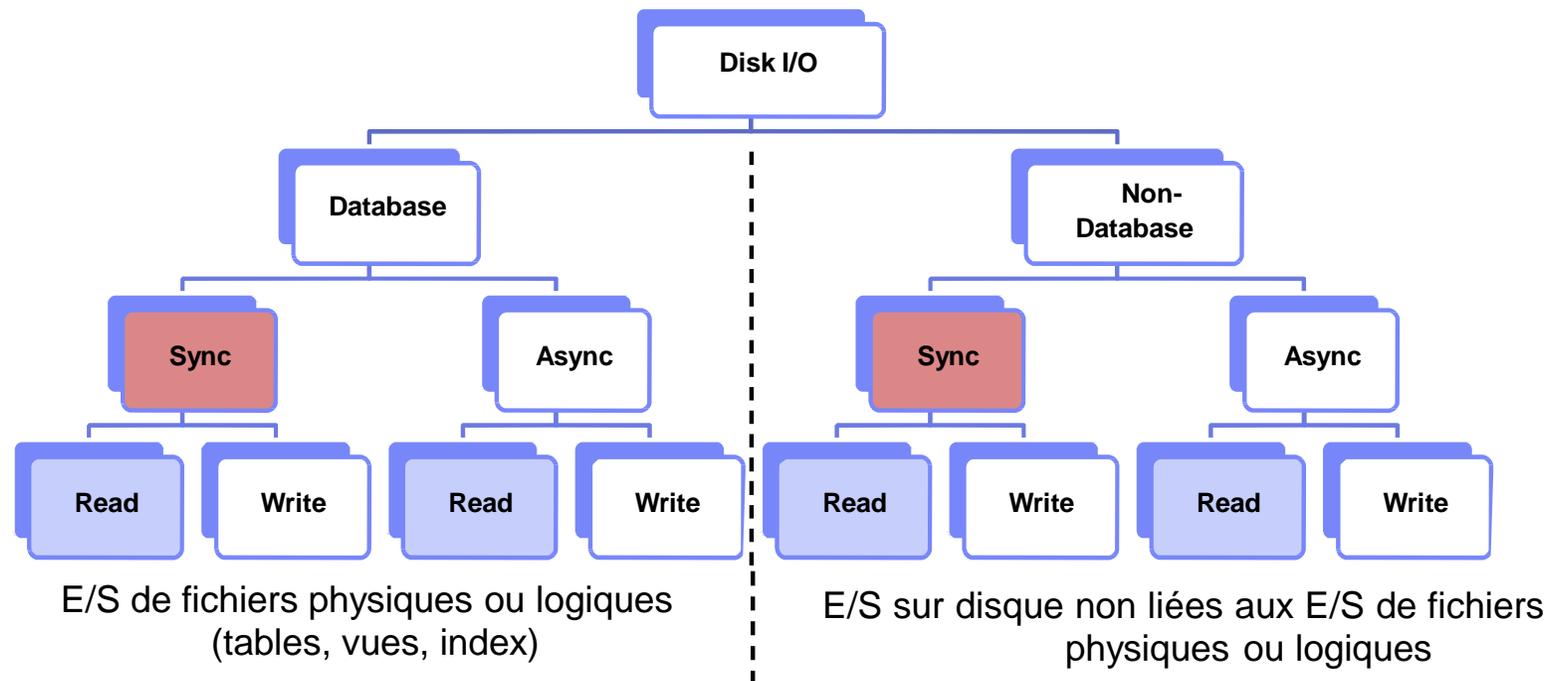
## Toutes les demandes d'E/S ne sont pas égales

- **E/S synchrones**
  - Le traitement attend que les E/S soient terminées
  - Contribue à la réponse/au temps d'exécution
  - Exemples (lectures) : SETOBJACC, CRTDUPOBJ, CPYF, Défauts
- **E/S asynchrones**
  - Traitement simultané avec les E/S
  - Peut devenir synchrone
  - Exemples (lecture) : Db2 asynchrone apporte



## Opérations sur les disques

- Classé par Gestion du stockage principal
  - E/S de base de données
    - E/S de fichiers physiques ou logiques
  - E/S non liées à la base de données
    - E/S de disque non liées aux E/S de fichiers physiques ou logiques
- Classées selon la concurrence avec le traitement
  - E/S synchrones
    - Le traitement attend que l'E/S soit terminée
  - E/S asynchrones
    - Traitement simultané avec les E/S



# PAGINATION DE LA MÉMOIRE ET DÉFAILLANCE

## Notions de base sur les erreurs de page

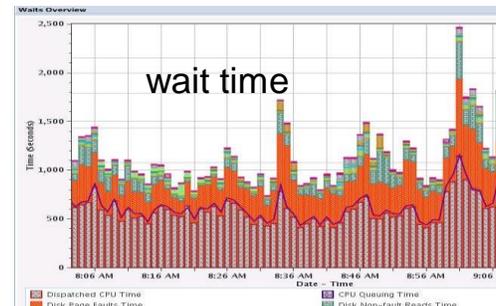
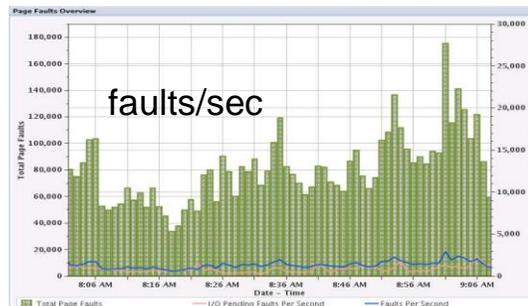
L'objectif principal de l'analyse de la mémoire est de réduire le temps d'attente des erreurs de disque.

- Les erreurs sont normales et attendues.
  - C'est le mécanisme utilisé pour amener la plupart des éléments en mémoire. Certains éléments doivent être mis en mémoire pour être nettoyés ou supprimés.
- Une seule erreur peut amener plusieurs pages en mémoire
  - Défauts par seconde < pages par seconde
  - Nous ne nous préoccupons généralement pas des pages par seconde.
- Les pages demandées sont placées dans le pool de mémoire dans lequel le travail s'exécute.
  - Les pages peuvent être partagées entre les tâches
  - Un travail peut accéder à une page dans un pool différent.

## Défauts de page

Défauts par seconde vs. temps d'attente quel est le plus révélateur?

- Se concentrer sur le temps d'attente des erreurs, et non sur le taux d'erreurs.
  - Les grands pools de mémoire peuvent avoir un nombre élevé d'erreurs, mais peu d'impact sur les temps d'attente.
- Cependant, les taux peuvent être utiles lorsqu'il est difficile d'obtenir le temps d'attente.
  - Tableau de bord, moniteurs, WRKSYSSTS, etc.
- Quel est le graphique le plus utile pour déterminer si la défaillance peut être un problème ?



# GESTION DE LA MÉMOIRE : TROP OU PAS ASSEZ ??

## Suppression des données de la mémoire

- **Contrôlé par le système**
  - Deux tâches sont chargées d'écrire les pages modifiées de la mémoire sur le disque pour faire de la place aux données mises en page dans la mémoire.
    - SMPOL001** : tâche de sortie de page de faible priorité
    - SMPO0001** : tâche de sortie de page de haute priorité
      - devient active si la tâche de faible priorité ne peut pas répondre à la demande.
  - Une activité élevée dans ces tâches, en particulier dans la tâche de haute priorité, peut indiquer un besoin de mémoire supplémentaire.
- **Contrôlées par l'utilisateur/le programme**
  - Les pages modifiées sont écrites à la fin d'un travail.

## Peut-on avoir trop de mémoire ?

### Il est presque toujours préférable d'avoir plus de mémoire, mais...

- L'ajout de mémoire peut ne pas réduire les erreurs  
Données sur le disque >> espace en mémoire, Activité E/S aléatoire, Chemins d'accès Db2
- Peut indirectement entraîner des problèmes lorsque les programmes se terminent et qu'il faut forcer de grandes quantités de données sur le disque → Provoque des pics d'activité sur le disque
- Compromis coût/bénéfice du retour sur investissement pour l'ajout de plus en plus de mémoire
- Souvent, le problème est que la mémoire existante n'est pas utilisée efficacement, et non pas qu'il n'y a pas assez de mémoire.
  - L'optimisation des applications peut réduire les besoins en mémoire
- Peut utiliser les disques SSD pour améliorer les performances pour des objets spécifiques → Peut agir comme un très grand cache

# POOLS MÉMOIRE ET SOUS SYSTEMES

## Pools de mémoire

- Les pools de mémoire sont des subdivisions logiques de la mémoire physique
  - Utilisés avec les sous-systèmes pour isoler l'utilisation de la mémoire par différentes applications
  - Il en existe deux types : les pools partagés et les pools privés ( ne profitent pas de la répartition de charge)

PARTAGE		PRIVE	Machine Pool
Sous-systèmes	simple ou multiple	simple	User Pools
QPFRAJ	OUI	NON	*INTERACT
Cache expert	OUI	NON	*SPOOL
			*SHRPOOL1-n
			*BASE

## Niveau d'activité maximum du pool

- Le nombre maximum de threads dans le pool qui peuvent utiliser le(s) processeur(s) simultanément.
  - Les threads qui n'ont pas de niveau d'activité ne peuvent pas être exécutés.
- Ne s'applique pas au pool de machines
  - Aucun travail n'est exécuté dans le pool de machines
- Peut être ajusté + ou - par l'Ajusteur de performance
- Considérations
  - En général, il faut que le niveau soit suffisamment élevé pour éviter les transitions vers l'inéligibilité.
  - Un réglage trop bas peut entraîner de graves problèmes de performance
  - Un réglage trop élevé peut entraîner davantage de défaillances

# POOLS MÉMOIRE ET SOUS SYSTEMES

## Optimisation des Pools

- Automatic Performance Adjustment (QPFRADJ)
  - 0 ou 3 on évite 1 et 2
- Expert Cache (\*CALC)
  - Gère la taille des pools de mémoire partagée pour vous.- Il ajustera le niveau d'activité maximal.

Pool	Taille définie	Activ maxi	Taille attribuée	ID pool	Opt pagination	
					Définie	En cours
*MACHINE	705,56	+++++	705,56	1	*FIXED	*FIXED
*BASE	2907,40	104	2907,40	2	*FIXED	*FIXED
*INTERACT	405,93	102	405,93	3	*FIXED	*FIXED
*SPOOL	40,59	5	40,59	4	*FIXED	*FIXED

ExpertCache  
Activation en passant de \*FIXED à \*CALC

## Memory Pools

Actions Total Memory: 3.96 GB

System Pool Identifier	Status	Pool	Description	Type	Current Size (MB)	Current Threads	Maximum Eligible Threads	Total Faults
1	ACTIVE	Machine	Utilisé par les fonctions internes du pool machine	SHARED	705,56	133	2147483647	
2	ACTIVE	Base	Pool système par défaut	SHARED	2907,4	896	104	
3	ACTIVE	Interactive	Travail interactif	SHARED	405,93	0	102	
4	ACTIVE	Spool	Impression	SHARED	40,59	0	5	

# POOLS MÉMOIRE ET SOUS SYSTEMES

## Recommandations QPFRADJ – Paramètres

- Il est recommandé d'utiliser 0 (=off) ou 3
  - Il est possible de régler certains pools sur min=max pour les "désactiver".
  - Si l'option est désactivée, il faut ajuster manuellement les pools de temps en temps.
- Les considérations ci-dessous s'appliquent lorsque QPFRADJ est activé
  - Ajuster les priorités des pools
  - Ne mettez pas tout sur 1 ou 2

## Définir les tailles minimale et maximale des pools (lorsque QPFRADJ est activé)

- Pools Java/WebSphere
  - La taille minimale doit toujours être suffisante pour contenir toutes les JVM du pool.
  - Il ne faut pas qu'il y ait de défaillance dans ce pool en temps normal.
- Fixer un maximum pour les pools à forte intensité de mémoire afin de limiter l'impact sur les autres travaux de la partition.
- Des changements importants dans la taille d'un pool de mémoire peuvent amener l'optimiseur de requêtes à reconstruire les plans d'accès, ce qui peut contribuer à de mauvaises performances.
- Si vous constatez que les pools restent souvent à la taille minimale, envisagez de diminuer la taille minimale.
- Si des pools atteignent la taille maximale, il faut envisager d'augmenter la taille maximale.

**Maintenir le total des tailles minimales < 70 % de la mémoire pour permettre à QPFRADJ d'avoir une certaine flexibilité.**

# POOLS MÉMOIRE ET SOUS SYSTÈMES (NAVIGATOR FOR I)

## Memory Pools

Actions Total Memory: 3.96 GB

System Pool Identifier ↑↓	Status ↑↓	Pool ↑↓	Description ↑↓	Type ↑↓	Current Size (MB) ↑↓	Current Threads ↑↓	Maximum Eligible Threads ↑↓	Total Faults ↑↓
1	ACTIVE	Machine	Utilisé par les fonctions internes du pool machine	SHARED	705,56	133	2147483647	
2	ACTIVE	Base	Pool système par défaut	SHARED	2907,4	893	104	
3	ACTIVE	Interactive	Travail interactif	SHARED	405,93	1	102	
4	ACTIVE	Spool	Impression	SHARED	40,59	0	5	

### Base Properties

General

**Configuration**

Performance

Tuning

**Size**

Minimum: 0 MB

Current: 2907.4 MB

Reserved: 7.28 MB

**Activity level**

Maximum eligible threads: 104

Current threads: 893

Current ineligible threads: 0

Page option: Fixed

### Base Properties

General

Configuration

Performance

**Tuning**

Automatically adjust memory pools and activity levels :

At system restart

Periodically after restart

**Tuning values**

Priority (1-14): 1 1-14

**Size:**

Minimum (0-100): 40.32 %

Maximum (0-100): 100 %

**Page faults per second:**

Minimum: 12

Additional minimum per thread: 5

Maximum: 200

Reset to defaults

# POOLS MÉMOIRE ET SOUS SYSTÈMES (NAVIGATOR FOR I)

Actions Total Memory: 3.96 GB

System Pool Identifier ↑↓	Status ↑↓	Pool ↑↓	Description ↑↓	Type ↑↓	Current Size (MB) ↑↓	Current Threads ↑↓	Maximum Eligible Threads ↑↓	Total Faults ↑↓
Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter
1	ACTIVE	Machine	Utilisé par les fonctions internes du pool machine	SHARED	705,56	133	2147483647	
2	ACTIVE	Base	Pool système par défaut	SHARED	2907,4	893	50	
3	ACTIVE	Interactive	Travail interactif	SHARED	405,93	1	102	
4	ACTIVE	Spool	Impression	SHARED	40,59	0	5	

\*BASE Pool Subsystem

Actions

Subsystem ↑↓	Status ↑↓	Active Jobs ↑↓	Description ↑↓	Library ↑↓
Filter	Filter	Filter	Filter	Filter
Lpar2rrd	Active	Lpar2rrd/Lpar2rrd Jobs		
Qbatch	Active			
Qcmn	Active			
Qctl	Active			

Lpar2rrd/Lpar2rrd Jobs

Actions

Job Name ↑↓	Detailed Status ↑↓	Current User ↑↓	Type ↑↓	CPU % ↑↓	Run Priority ↑↓	Thread Count ↑↓	Temporary Storage Used(MB) ↑↓
Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter
RTV_SYSSTS	Waiting for dequeue	LPAR2RRD	Batch	-1	50	1	-1
SND_SYSSTS	Waiting for dequeue	LPAR2RRD	Batch	-1	50	1	-1
LPAR2RRD	Waiting for dequeue	QSYS	Subsystem	-1	0	2	-1



# POOLS MÉMOIRE ET SOUS SYSTEMES

## Sous-systèmes

- Environnements d'exécution prédéfinis au sein d'un seul IBM i
  - Permet la séparation de la charge de travail
  - Contrôle du flux de travail
  - Avec les pools de mémoire, permet un certain contrôle de l'utilisation de la mémoire
  - Permet de regrouper l'utilisation des ressources dans les données de performance
- Exemple de sous-systèmes :QINTER; QBATCH; QSYSWRK; QUSRWRK; Défini par l'utilisateur.
- Les sous-systèmes peuvent être utilisés pour contrôler:
  - le pool de mémoire dans lequel un travail peut s'exécuter
  - le nombre de travaux exécutés simultanément et donc indirectement la quantité de mémoire dont dispose chaque travail.
    - par le biais des files d'attente de travaux
    - Par les paramètres de pré-démarrage des travaux
    - Le niveau d'activité maximum dans le pool
  - La quantité de CPU que les tâches peuvent consommer (avec les groupes de charge de travail)
  - La priorité d'exécution, la tranche de temps et les capacités maximales de CPU et de stockage temporaire des tâches en cours d'exécution par le biais des définitions de classes.

# POOLS MÉMOIRE ET SOUS SYSTEMES

## Quand créer des sous-systèmes distincts (Performance Focus)

- Gestion de la mémoire, exemples :
  - Limiter l'impact des travaux à forte intensité de mémoire
  - Veiller à ce que les JVM disposent toujours d'une mémoire suffisante
- Paramètres de performance (par exemple, niveau de priorité)
- Groupes de charge de travail
  - Pour limiter l'impact du CPU sur l'ensemble de la partition ou pour des raisons de licence
- Gestion du travail
  - Il est plus facile de diviser l'utilisation des ressources dans les données de performance.
  - Permet de limiter l'impact des **requêtes ad hoc**

### Qu'est-ce qu'une requête ad hoc ?

- Requêtes qui ne sont pas prédéfinies dites « ponctuelles »
- Souvent exécutées par un utilisateur interactif
- Elles peuvent être destinées à l'établissement de rapports.

### Quel est le problème ?

- Souvent, les index ne sont pas corrects pour un bon fonctionnement
  - Peuvent accéder à des fichiers volumineux qui ne sont pas actuellement en mémoire
  - Peuvent causer des problèmes de performance intermittents

### Comment les traiter ?

- Ils sont souvent exécutés à l'aide de tâches de pré-démarrage QZDASOINIT.
- Ne pas vouloir impacter tous les jobs QZDA

### Deux méthodes pour isoler - par adresse IP et par identifiant d'utilisateur

Rappel de base sur les concepts

La CPU

La mémoire

## Les disques

Stockage externe VS stockage interne

La virtualisation

Les métriques

Analyse de performances



# SOUS SYSTEMES DISQUES

## Importance du sous-système disque

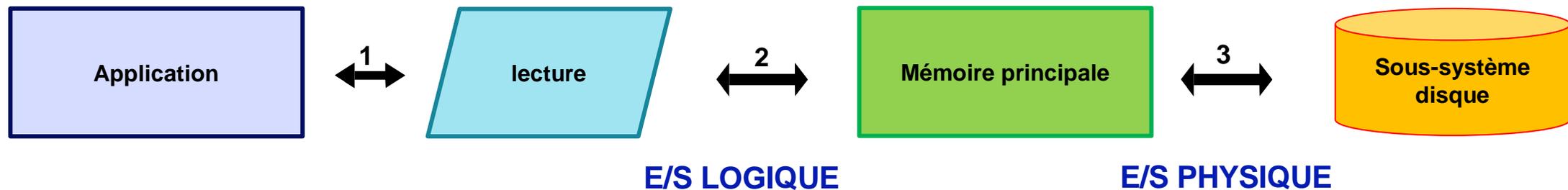
- Le fait de devoir aller sur le disque est vraiment lent, en termes relatifs.
- Une mauvaise configuration du sous-système d'E/S peut réduire les performances globales du système.

## Comment améliorer les performances :

- Veiller à la bonne configuration du sous-système de disque
- Les adaptateurs d'E/S dotés de caches plus importants et les nouvelles technologies de stockage (par exemple, SSD / FlashSystem) peuvent réduire les temps d'attente des E/S
- Les changements d'application peuvent réduire considérablement le temps d'attente de vos tâches pour les E/S sur disque.

## E/S logiques et physiques des disques

1. L'application effectue une opération de lecture (READ)
2. Les données demandées sont trouvées en mémoire- Il en résulte une E/S logique
3. Si les données demandées ne sont pas en mémoire, alors les données sont lues du disque vers la mémoire. Il en résulte une entrée/sortie physique.



Une opération de lecture entraîne **toujours** une entrée/sortie logique et **peut** également entraîner une ou plusieurs entrées/sorties physiques.

# SOUS SYSTEMES DISQUES

## Composants du sous-système d'E/S

- Types d'unités de disques (lecteurs)
- Contrôleurs de disque (également appelés adaptateurs d'E/S ou IOA) : ils relient plusieurs disques entre eux.
  - Importance de la mémoire cache
  - Considérations relatives à la protection des disques
- Bus / "I/O Fabric" - relie plusieurs contrôleurs entre eux
  - Concepts de pool de stockage auxiliaire (ASP)

## Types de disques durs

- **Hard Disk Drive (HDD)**
  - Plateau magnétique « tournant »
  - Délais de recherche et de rotation (latence)
  - Des centaines d'E/S par seconde
- **Solid State Drives (SSDs)**
  - Pas de pièces mobiles / utilisation de mémoire flash non volatile
  - Chemin de données contrôlé par logiciel
  - Des dizaines de milliers d'E/S par seconde, faible latence
- **FlashSystem Storage/ NVMe**
  - Pas de pièces mobiles / utilisation de mémoire flash non volatile
  - Chemin de données contrôlé uniquement par le matériel
  - Le plus grand nombre d'E/S par seconde, le temps de latence le plus faible



# SOUS SYSTEMES DISQUES

## Mesures de performance clés de l'unité de disque

- Temps de service du disque
  - La mise en cache signifie que le temps de service rapporté est inférieur au temps d'E/S physique, en particulier pour les écritures.
- Temps d'attente du disque
  - Temps d'attente dans la file d'attente du pilote du périphérique d'E/S
- Temps de réponse du disque (attente + service)
  - Attention au calcul de la moyenne des lectures et des écritures
- Pourcentage d'occupation du disque
  - Minimiser les files d'attente en conservant moins de 40 % (préférer < 20 %)

Le service de collecte à intervalles de 5 ou 15 minutes peut "**cacher**" des performances médiocres intermittentes. Job Watcher, Disk Watcher et PEX peuvent vous aider.

## Contrôleur de disque Cache

- **Cache de lecture**
  - Utilise des algorithmes pour compléter la mémoire de stockage à niveau unique et le cache de lecture des unités de disque.
  - Favorise la conservation des données provenant d'accès aléatoires au disque
- **Cache d'écriture**
  - Peut combiner et éliminer les écritures physiques sur le disque
  - La notification "écriture terminée" est signalée en une fraction de milliseconde.
    - L'écriture réelle se produit de manière asynchrone
  - Vouloir éviter les dépassements de cache d'écriture en raison d'un cache plein
    - Les nouvelles écritures doivent attendre que les écritures existantes soient forcées sur le disque.

Le cache se trouve dans l'IOA pour les systèmes de stockage à connexion directe et au niveau du processeur pour les systèmes de stockage en connexion.(c'est-à-dire "au-dessus" du niveau IOA)

# SOUS SYSTEMES DISQUES

## Efficacité de la mémoire cache en lecture/écriture du contrôleur de disque

- **Efficacité élevée (près de 100 %)**
  - Performances optimales en lecture/écriture sur le disque → Prévoir une efficacité d'écriture élevée
- **Faible efficacité (proche de zéro)**
  - Mauvaise relation entre :
    - l'accès aux données
    - Organisation des données- Les fichiers ne sont pas organisés en fonction des clés les plus fréquemment utilisées
  - Nombreux enregistrements supprimés
  - Taille des blocs de lecture trop importante
    - Travail entraînant la lecture d'une plus grande quantité de données sur le disque que nécessaire
    - Le cache expert (\*CALC) peut réduire les accès au disque.
  - Fragmentation du disque
  - Trop de bras sur une IOA (pas de bras pas de chocolat, trop de bras pas assez de chocolat) **≠ en stockage externe**
- Taille du cache - lecture et écriture - plus c'est gros, mieux c'est !
- Bras par contrôleur - moins il y en a, mieux c'est pour les performances
- Une batterie de cache d'écriture défectueuse entraîne une dégradation des performances.
- Protection des disques
  - Les bras défectueux dans les jeux RAID entraînent une dégradation des performances jusqu'à ce qu'ils soient reconstruits.
  - Veillez à ce que les nouvelles unités ne soient pas ajoutées aux jeux de parité existants.
- Lorsque plusieurs contrôleurs sont impliqués :
  - Nombre constant de bras par contrôleur
  - Taille de disque cohérente pour chaque contrôleur

Rappel de base sur les concepts

La CPU

La mémoire

Les disques

## Stockage externe VS stockage interne

La virtualisation

Les métriques

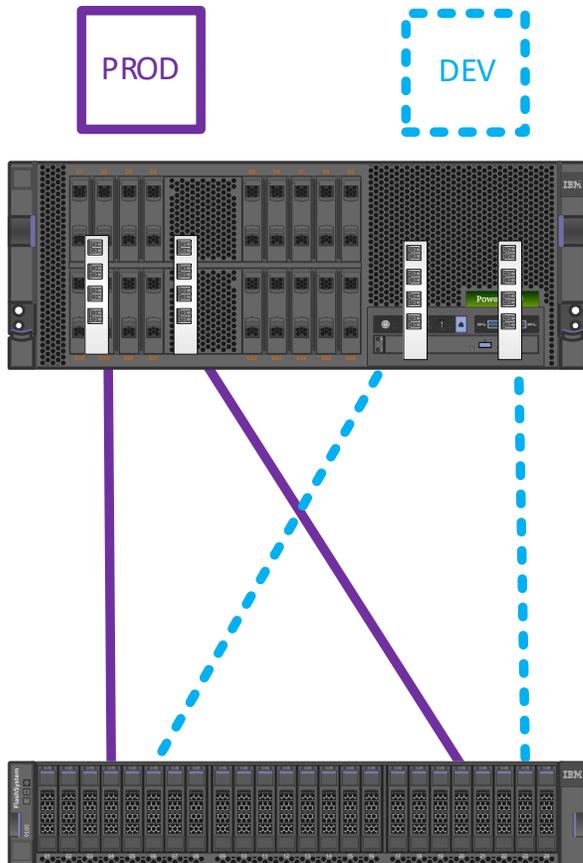
Analyse de performances



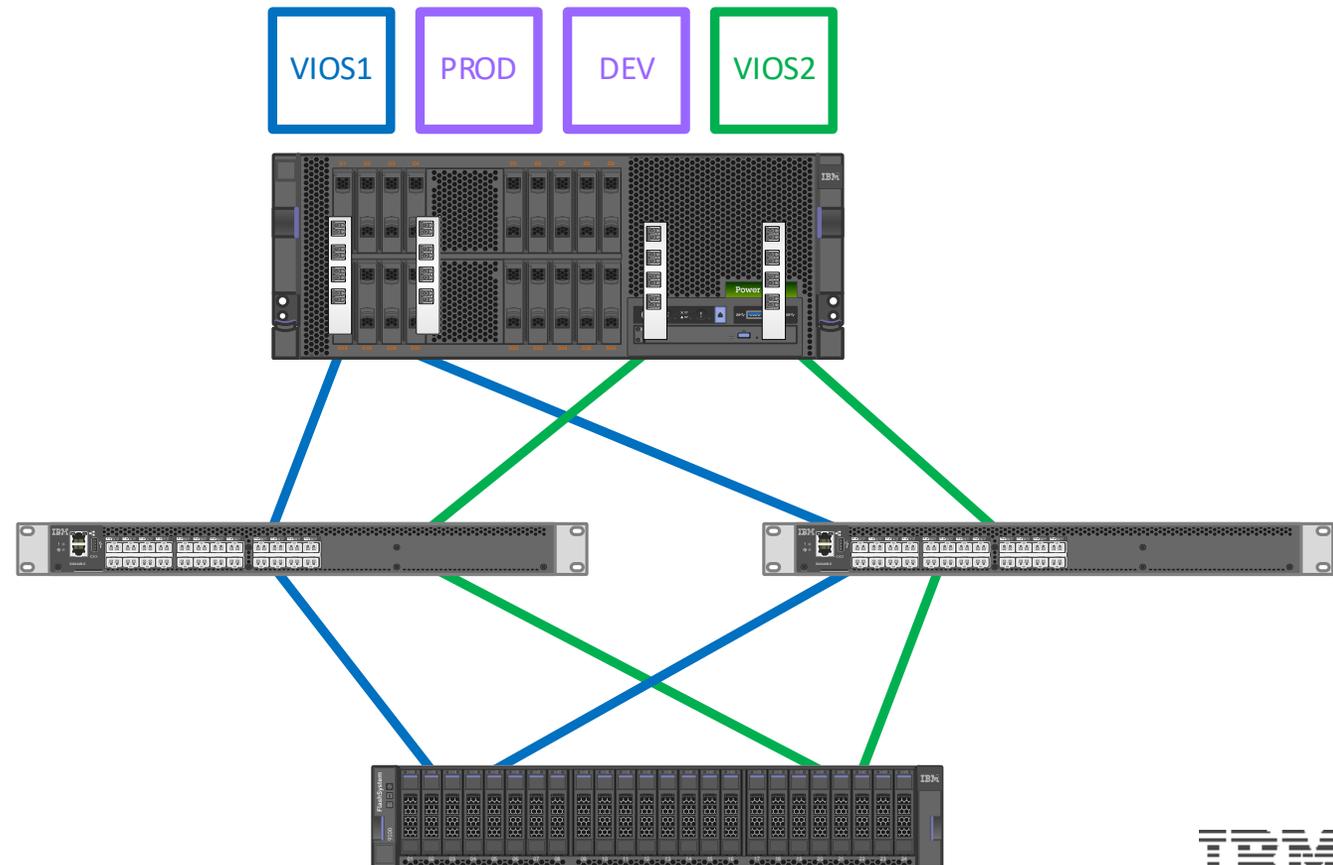
# ATTACHEMENT DIRECT VS FABRIC

Il existe 2 modes d'attachement pour du stockage externe

- Attachement direct



- Attachement par Fabric



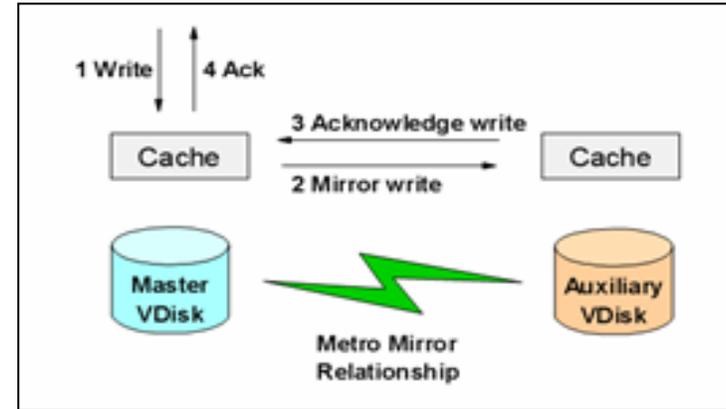
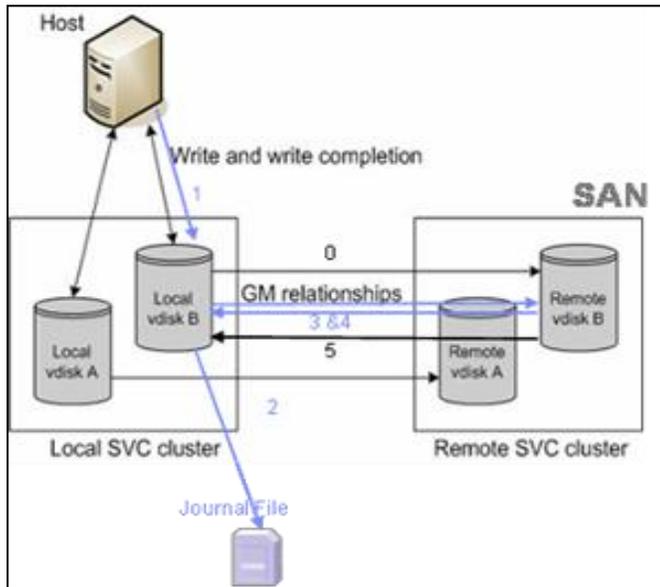
# CONSIDERATIONS POUR LE STOCKAGE EXTERNE

Les points d'attention pour une bascule en stockage externe sources d'échec

- Pas d'analyses de performances du système source → pas beaucoup d'IOPS ça passe facile....
- Pas de prise en compte de la configuration disque source
  - Carte Raid
  - Type de protection
  - Nombre de disques
  - Types de disques
  - Nombre d'iops
- Sous-dimensionnement de la baie de stockage
  - Taille des disques → attention à la règle des 8/9
  - Taille du cache → attention le cache est partagé sur la baie en totalité et pas dédié à une partition
  - Nombre de ports d'attacheements fibre
- Mauvais choix de virtualisation
  - VSCSI/NPIV
  - Nombre d'adaptateur insuffisant
  - Taille de LUNs inadaptée
  - Nombre de LUNs inadapté
- Mauvaise analyse des charges applicatives
- Mauvaises analyse des utilisations disques ( SAVF, save while active, ...)

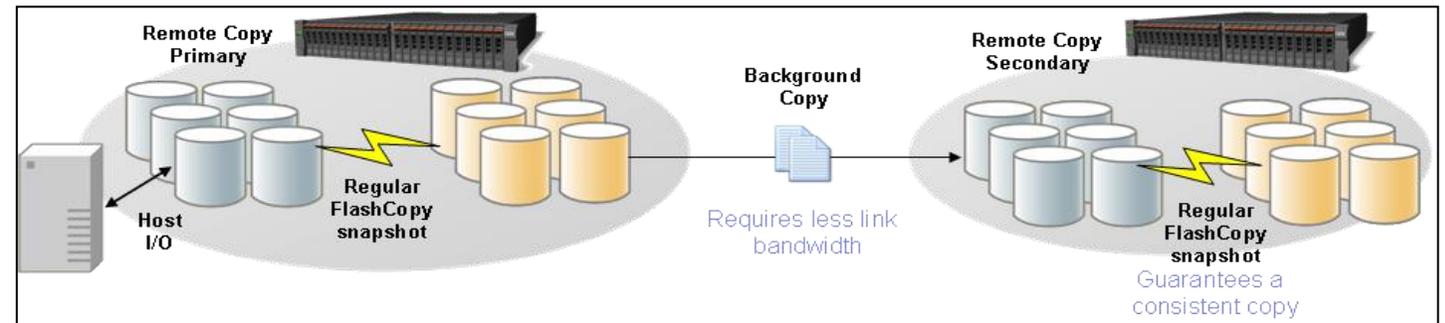
Trois modes de réplication sont proposés sous la même licence (il est possible de passer d'un mode à un autre).

- Metro Mirror (MM): Réplication synchrone



- Global Mirror (GM): Réplication asynchrone « au fil de l'eau ».

- Global Mirror (GMCV) Réplication asynchrone avec Flashcopy



Rappel de base sur les concepts

La CPU

La mémoire

Les disques

Stockage externe VS stockage interne

## La virtualisation

Les métriques

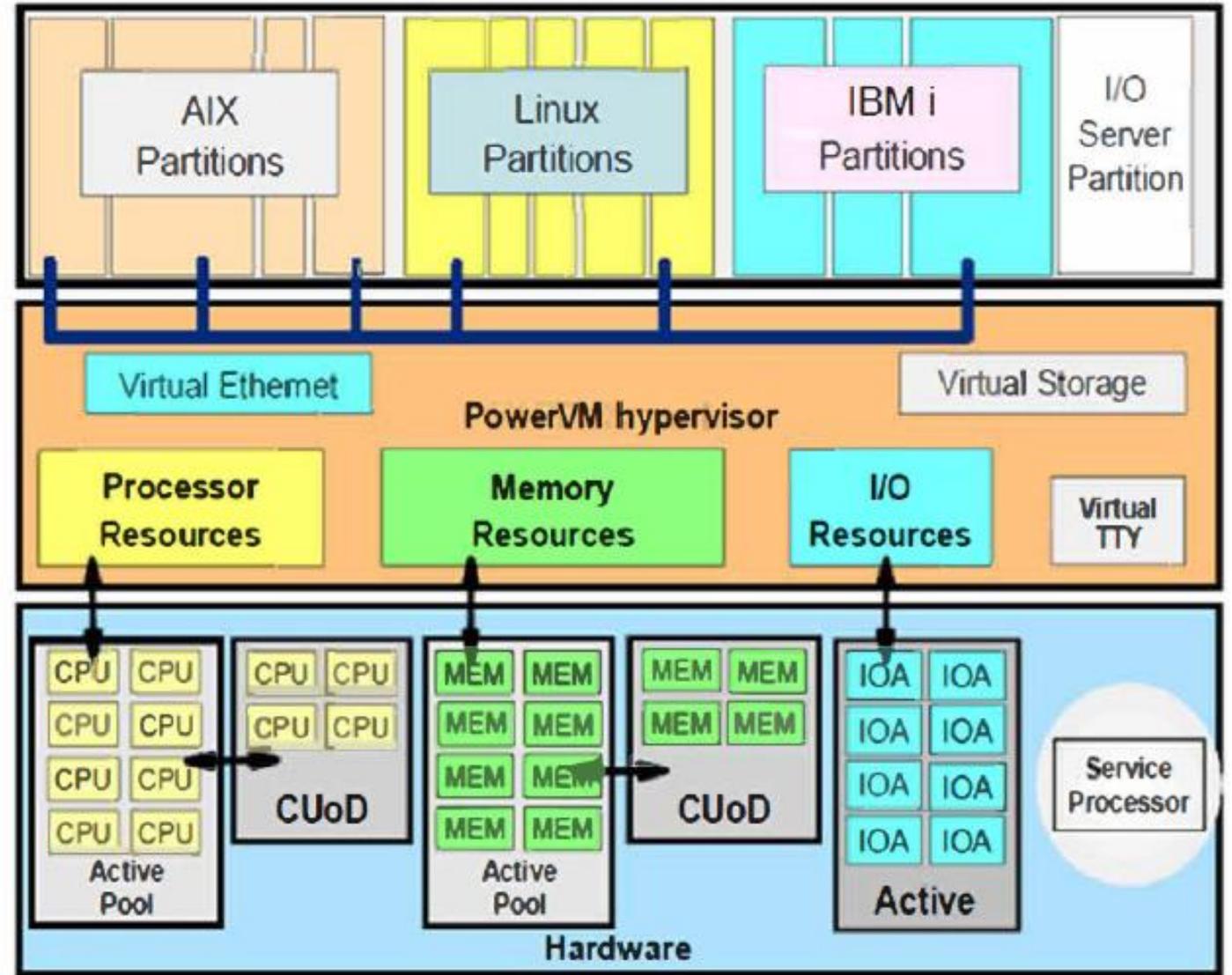
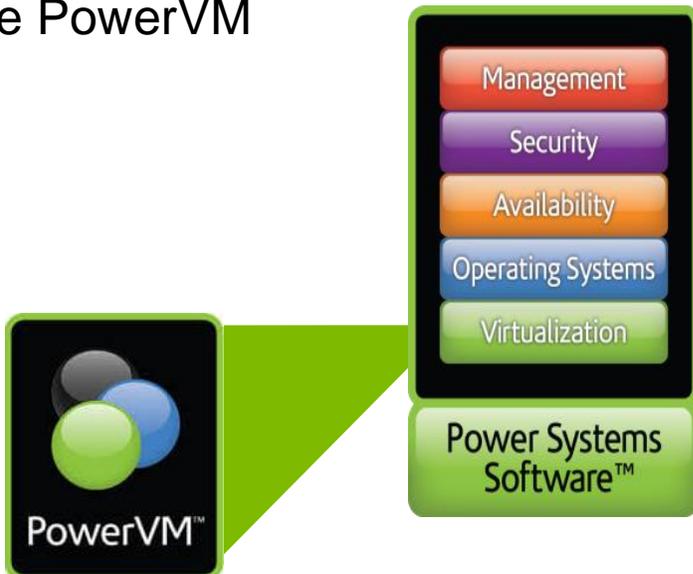
Analyse de performances



# VIRTUALISATION 1/3

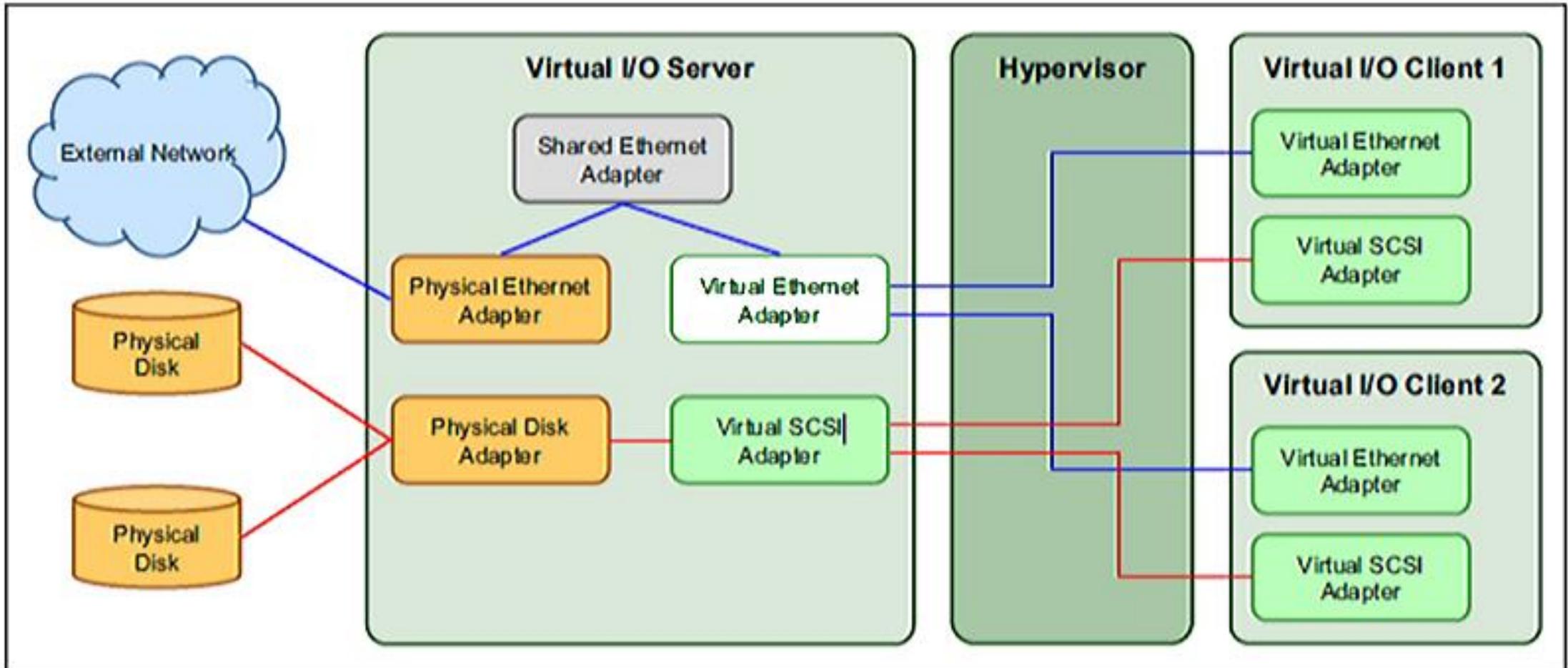
L'ensemble des ressources physiques processeur, mémoire, sont virtualisées par la partie Hypervisor pour être présentées aux VIOS et aux partitions clientes

La virtualisation sur Power est assurée au travers de la couche PowerVM



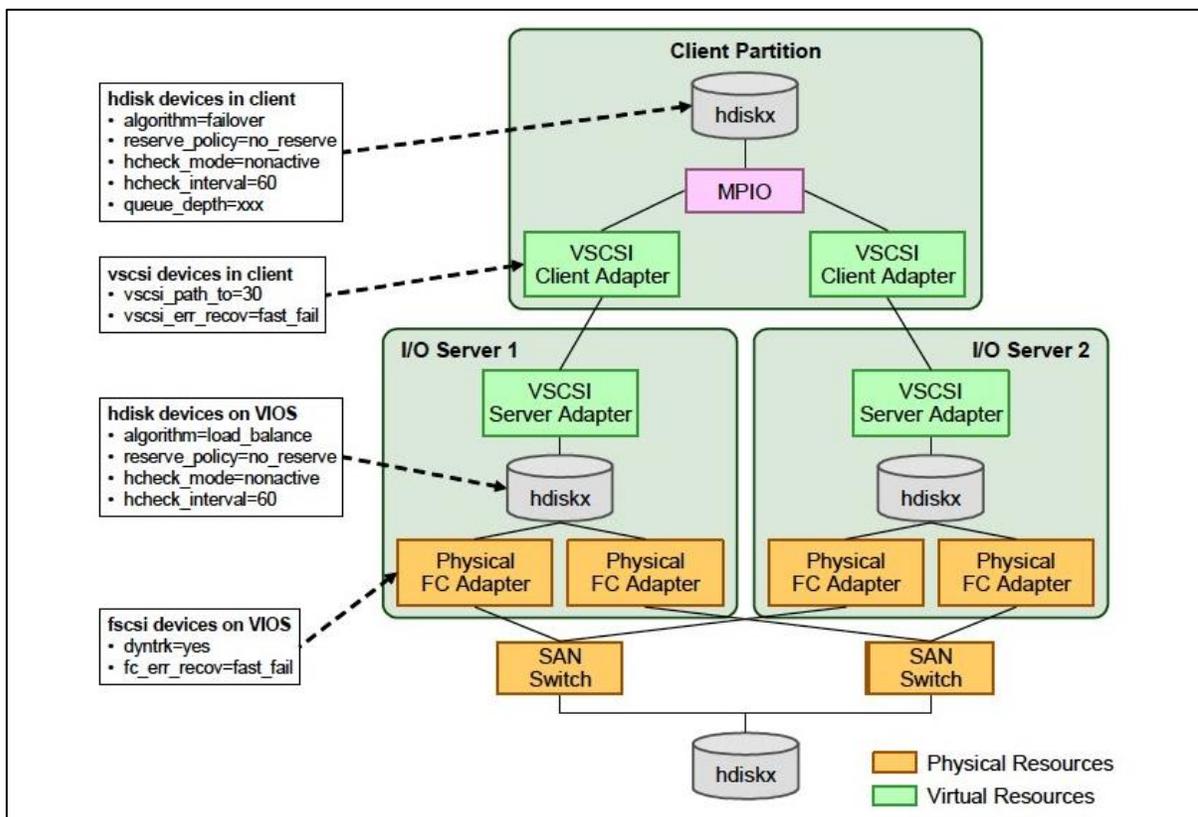
# VIRTUALISATION 2/3

Le VIOS (Virtual I/O Server) va permettre de virtualiser les ressources physiques de la machine pour les partitions clientes (SEA, NPIV, VSCSI)



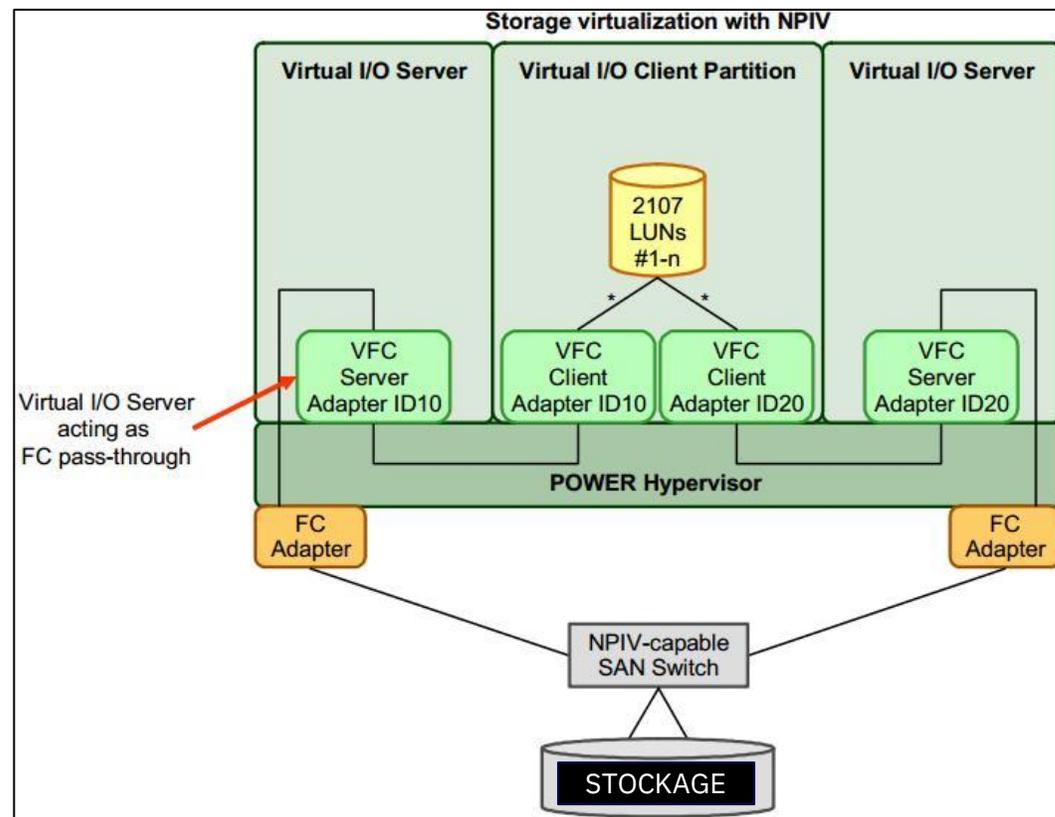
## VSCSI avec/sans Switch SAN

- LUNS → VIOS → LPAR
- Adaptateur Server/Client VSCSI
- Zoning uniquement VIOS/Stockage



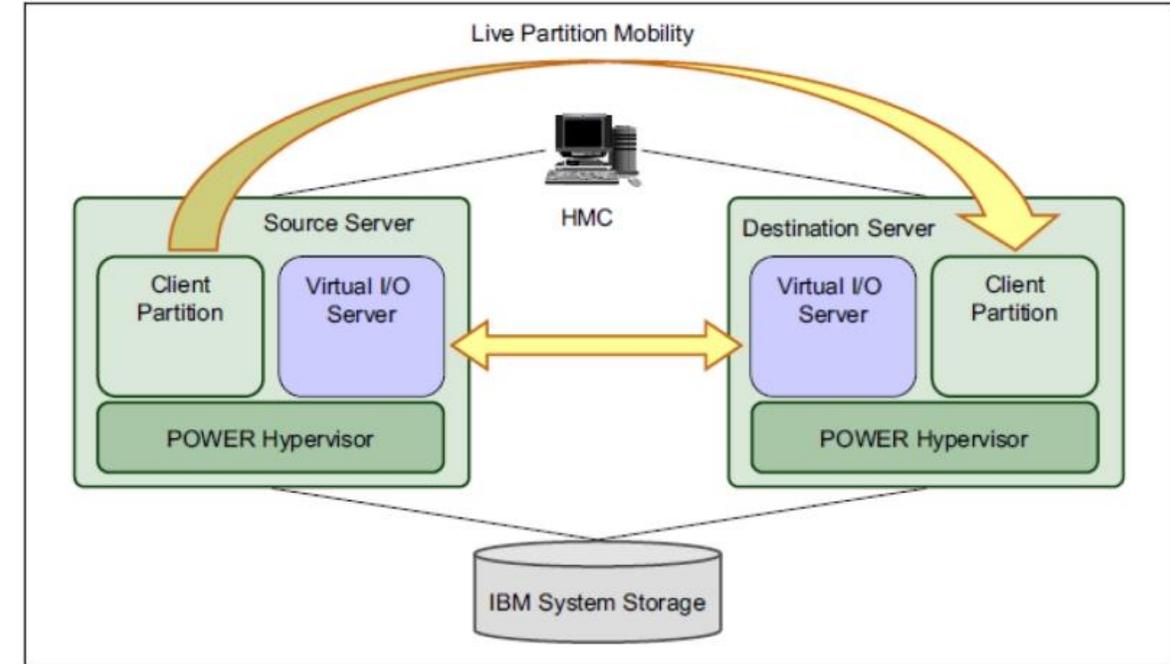
## NPIV → switch SAN OBLIGATOIRE

- LUNS → LPAR
- Adaptateur VFCHOST (wwn)
- Zoning CLIENT/STOCKAGE



Live Partition Mobility est une fonctionnalité de PowerVM Enterprise Edition permettant de déplacer une partition en cours d'exécution d'un système à un autre.

- Cette fonctionnalité est disponible pour Linux et AIX depuis les Power6 et jusqu'au Power10.
- Partition Mobility (en direct ou inactif) et Partition Migration (Active ou Inactif) se réfèrent à la même fonction.



## ATTENTION

- les systèmes **doivent impérativement** partager un stockage commun
- Les ressources sont **obligatoirement** virtualisées
- Les réseaux Ethernet **doivent être identiques** sur les systèmes

Rappel de base sur les concepts

La CPU

La mémoire

Les disques

Stockage externe VS stockage interne

La virtualisation

## Les métriques

Analyse de performances



- Outils **libres** de monitoring et de reporting des performances pour les serveurs, le stockage, le SAN et le LAN
- Open source sous License GPL v3
- Service de support professionnel optionnel
- Le support inclus les éditions entreprise des outils qui ont des fonctions complémentaires
- Les données sont collectées sur les périphériques et stockées dans les outils
- Fonctionnent sur UNIX/Linux ou bien sur une appliance VMWare (Linux)

# LPAR2RRD TUNING 1/4

```

Gestion de l'état du système                                PMROD001
                                                           10/04/19 14:46:53
% UC utilisée . . . . . :      108,6   ASP système . . . . . :    20433 G
Intervalle . . . . . :    00:22:16   % ASP système utilisé . :    41,9533
Travaux connus du syst . :    31684   Mémoire secondaire totale:    20433 G
% adresses permanentes . :    1,441   Mémoire temporaire
% adresses temporaires . :    74,188   utilisée . . . . . :    578122 M
                                       Mém max temp utilisée . :    589756 M

```

Pool syst	Taille pool	Taille réserv	Act Max	-Paging Taux	BD-- Pages	--Non-BD--- Taux	Pages	Act-> Att	Att- Inelg	Act-> Inelg
1	<u>16896,0</u>	6745,0	+++++	0,0	0,0	1,2	4,1	13088	0,0	0,0
2	<u>177408,0</u>	28,4	<u>2540</u>	63,3	5332	559,7	2369	85105	0,0	0,0
3	<u>5632,0</u>	0,0	<u>1408</u>	4,0	32,0	7,8	14,3	481,3	0,0	0,0
4	<u>28160,0</u>	<.1	<u>5011</u>	967,9	6735	3609	8325	19849	0,0	0,0
5	<u>5632,0</u>	0,0	<u>350</u>	0,3	62,4	0,3	0,4	+++++	0,0	0,0
6	<u>2816,0</u>	0,0	<u>224</u>	0,7	13,4	12,8	24,5	4265	0,0	0,0
7	<u>14080,0</u>	0,0	<u>679</u>	66,3	145,7	47,9	64,2	3244	0,0	0,0
8	<u>2816,0</u>	0,0	<u>160</u>	224,9	1956	90,5	141,8	377,0	0,0	0,0
9	<u>28160,0</u>	1,8	<u>5000</u>	8,1	93,2	7,4	15,1	19733	0,0	0,0

Fin

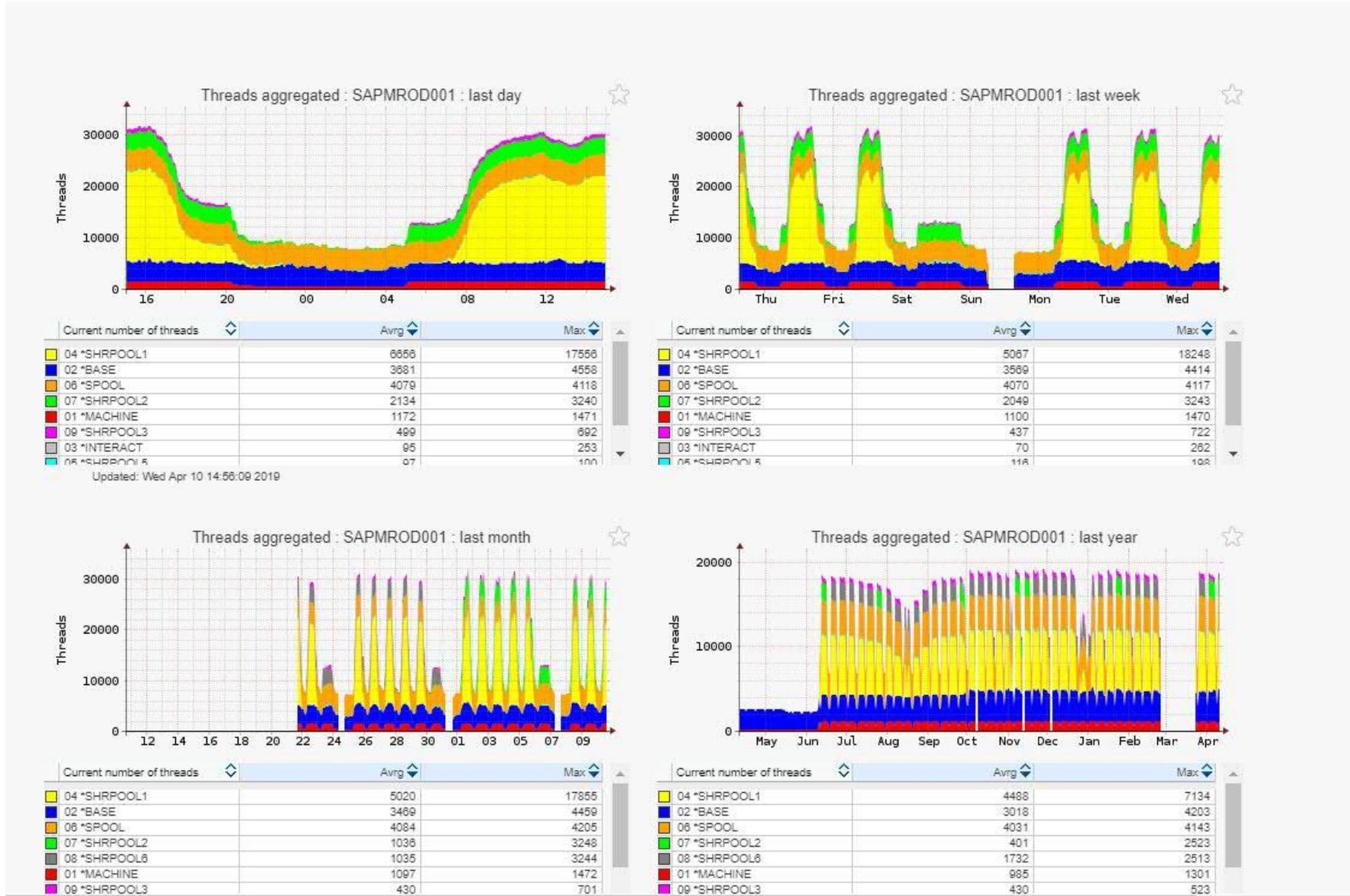
```

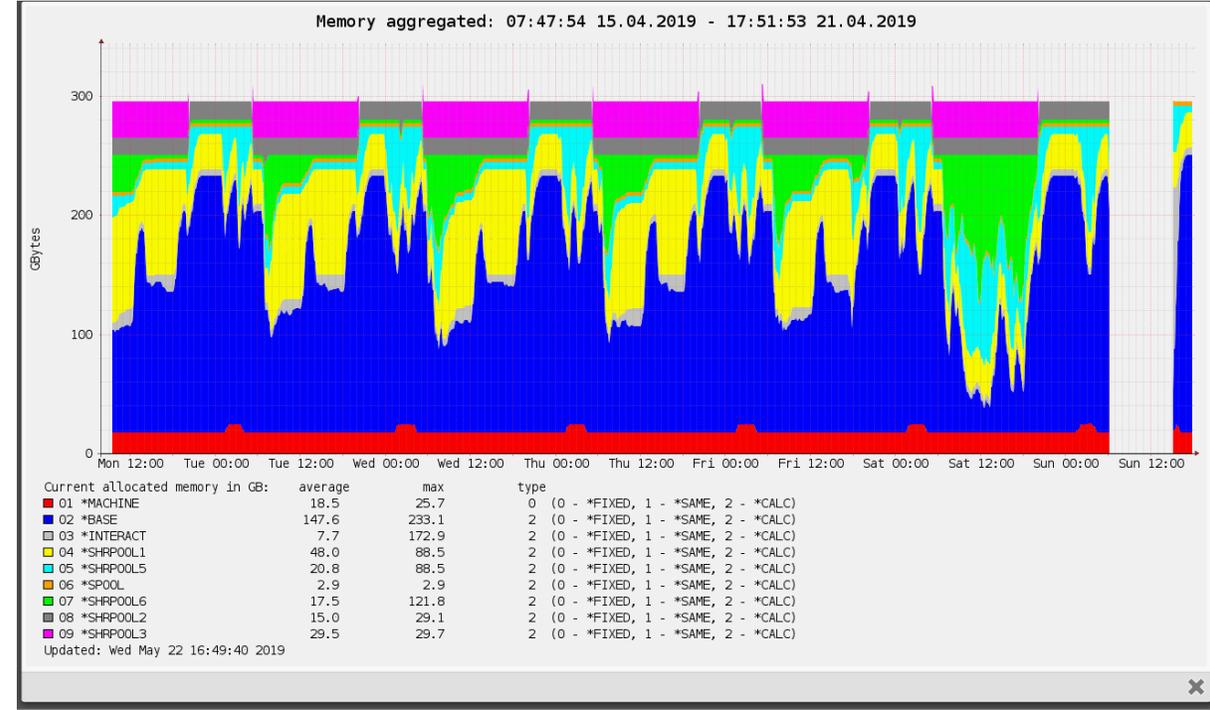
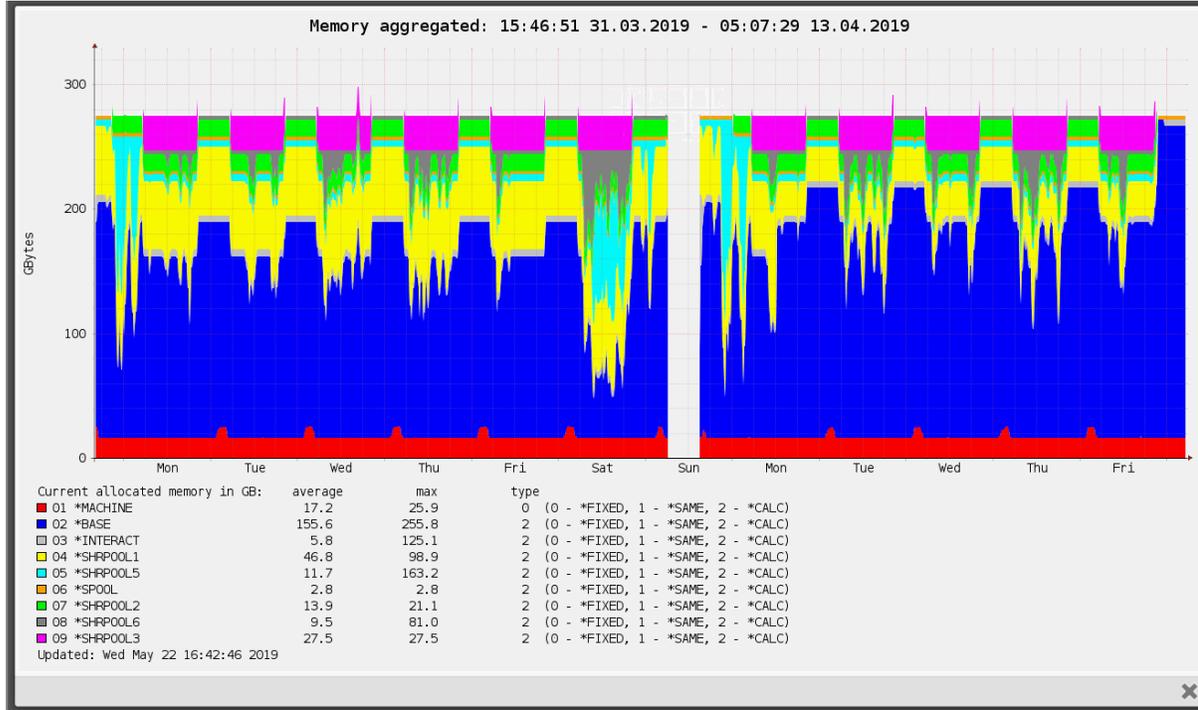
Gestion des pools partagés                                Système: PMROD001
Taille de la mémoire principale (Mo) . . . . . :    281600,00
Indiquez vos modifications (si admises), puis appuyez sur ENTREE.

```

Pool	Taille définie	Activ maxi	Taille attribuée	ID pool	Opt pagination Définie	En cours
*MACHINE	<u>16896,00</u>	+++++	16896,00	1	*FIXED	*FIXED
*BASE	<u>177408,00</u>	<u>2540</u>	177408,00	2	*CALC	*CALC
*INTERACT	<u>5632,00</u>	<u>1408</u>	5632,00	3	*CALC	*CALC
*SPOOL	<u>2816,00</u>	<u>224</u>	2816,00	6	*CALC	*CALC
*SHRPOOL1	<u>28160,00</u>	<u>5011</u>	28160,00	4	*CALC	*CALC
*SHRPOOL2	<u>14080,00</u>	<u>679</u>	14080,00	7	*CALC	*CALC
*SHRPOOL3	<u>28160,00</u>	<u>5000</u>	28160,00	9	*CALC	*CALC
*SHRPOOL4	<u>1,00</u>	<u>1</u>			*CALC	
*SHRPOOL5	<u>5632,00</u>	<u>350</u>	5632,00	5	*CALC	*CALC

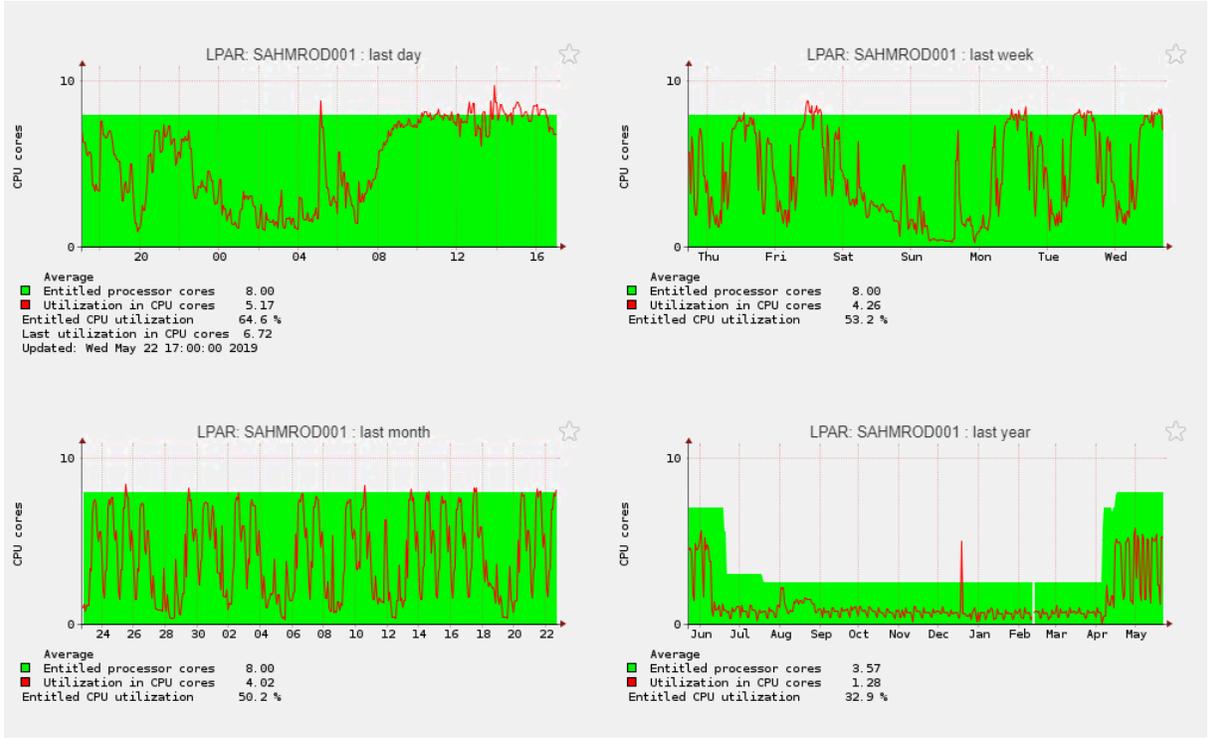
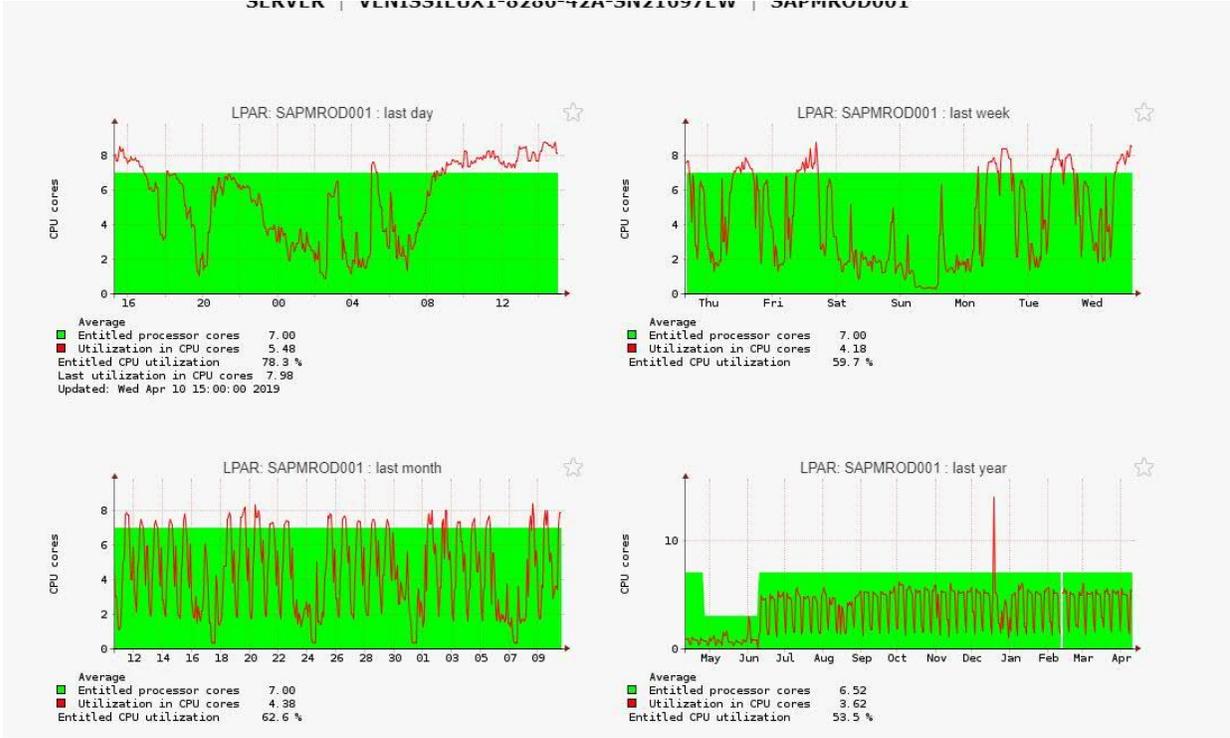
A suivre...



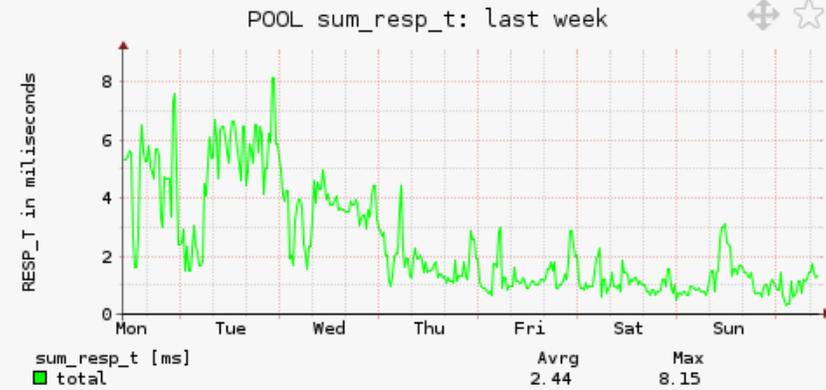
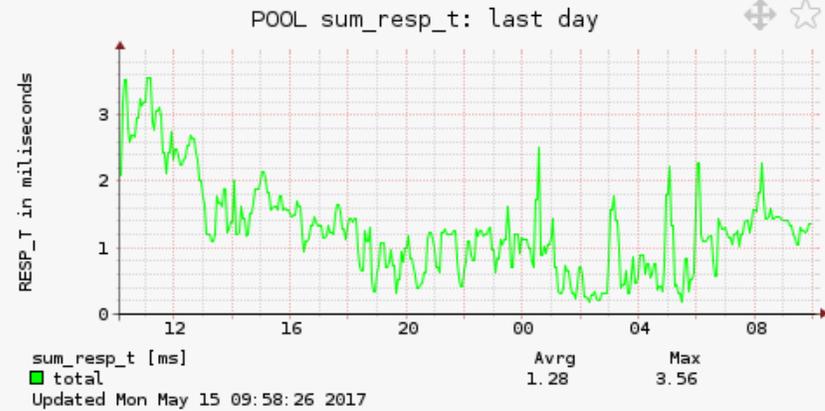


# LPAR2RRD TUNING 4/4

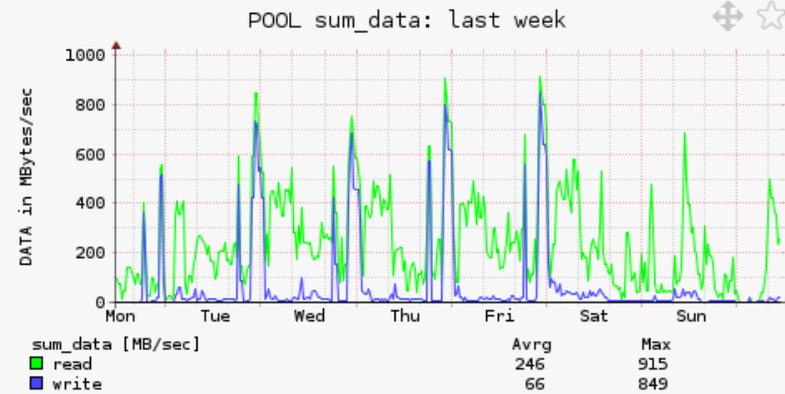
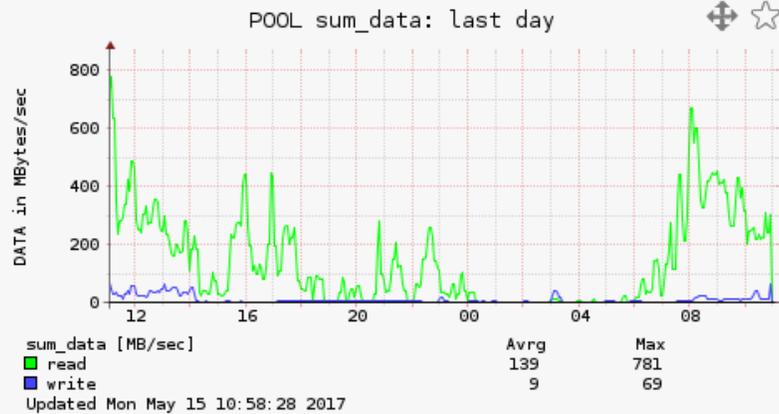
SERVER | VLN1551L0X1-0200-74A-5N2109/LW | SAPHMROD001



## C12SAN01 | POOL | Response time



## C12SAN01 | POOL | Data



Rappel de base sur les concepts

La CPU

La mémoire

Les disques

Stockage externe VS stockage interne

La virtualisation

Les métriques

## Analyse de performances



```
MG                               Must Gather Data Collector
                                (C) COPYRIGHT IBM CORP. 2009, 2012
Select one of the following:

 1. System Snapshot
 2. HA (High Availability)
 3. Performance/Misc collection
 4. Client/Server
 5. Communications menu
 6. Database menu
 7. CTA/EWS (JAVA/HTTP/DCM/WAS)
 8. Save/Restore menu
 9. Misc tools
10. FTP data to IBM
11. View FTP to IBM statuses
12. Display build date
13. Check IBM for updated QMGTOOLS

14. External Storage
15. Work Management
16. Internals
17. Electronic Services
18.
19. Hardware data collection
20. IBM -----
```

```
Performance Menu                QPERF Menu
Select one of the following:

 1. Combined PEX/JobWatcher/Collection Service trace
 2.
 3. Collection Services
 4.
 5. Monitors
 6. PEX clean up
 7.
 8. PEX status (PEX started by QMGTOOLS)
 9.
10. Compare PEX/JW PTFs from IBM public FTP site
11. POOL Info Monitor
12. Gather Disk Magic sizing data
13. Work with JW statuses
14. Remote Command Exit Program
15. Collect disk metrics from CS data
16. Performance Menu help
Selection or command
```

<https://www.ibm.com/support/pages/mustgather-how-obtain-and-install-qmgtools-and-keep-it-current>

## Collector Data Characteristics

### Collection Services

- Sample data
- Designed to collect data 24x7
- Support for small intervals
- No information concerning specific I/O operations
- Wait time information

### Disk Watcher

- Statistics and Trace data
- Detailed information focused on I/O operations to disk units

### Job Watcher

- Sample data
- Support for very small intervals
- Focus on job data
  - Call Stacks
  - SQL statements
- Wait time information

### Performance Explorer

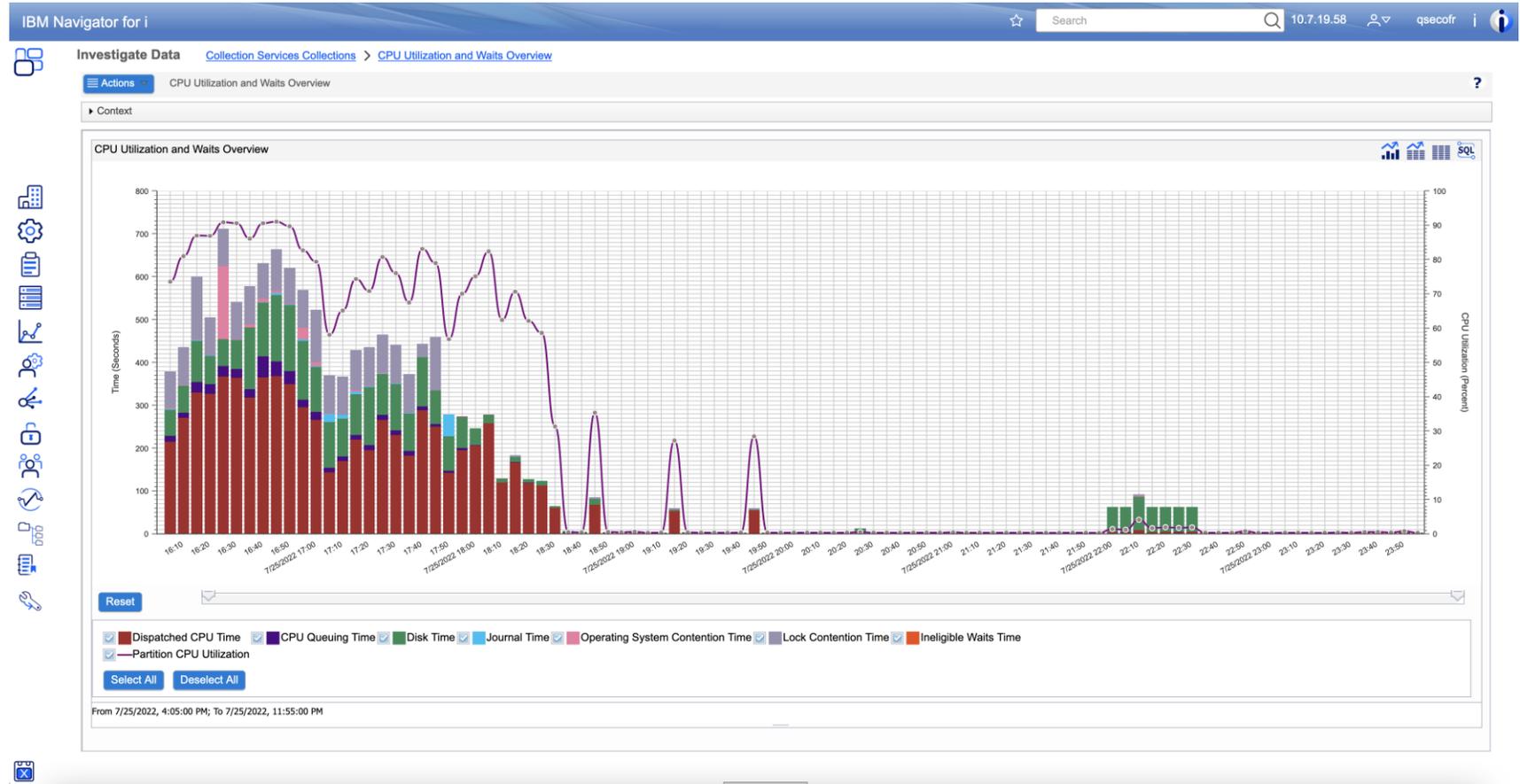
- Statistics, Profile, or Trace data
- Information collection for every I/O event
- Collection and analysis can be complex

Welcome to IBM Navigator for i

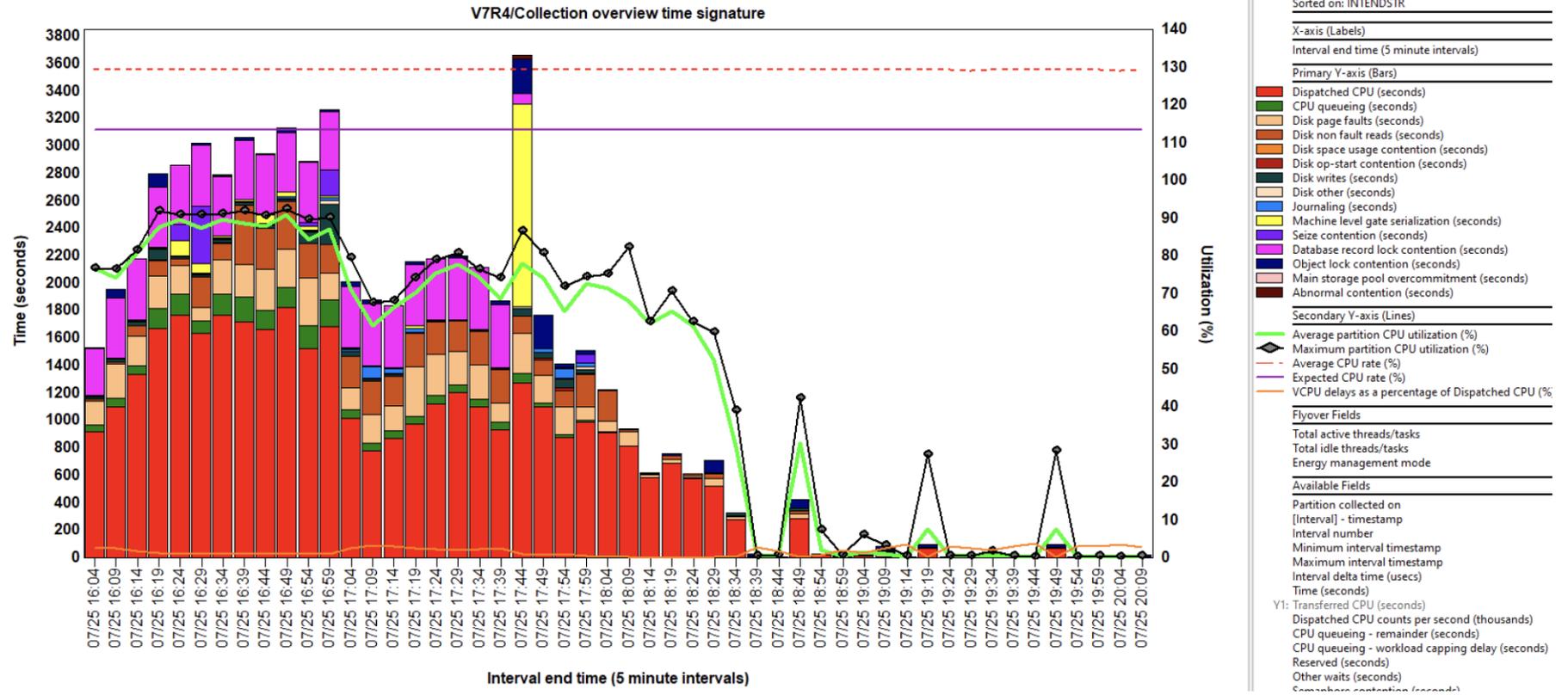
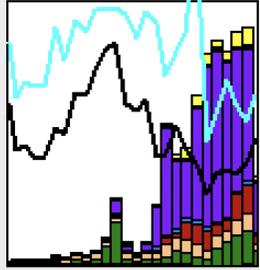
User Name:

Password:

**Log in**



## iDoctor for IBM i





Search or jump to... Pull requests Issues Marketplace Explore

IBM / **ibmi-oss-examples** Public

<> Code Issues 12 Pull requests 26 Actions Projects Wiki Security Insights

master ibmi-oss-examples / nodejs / grafana-backend / README.md

dmabupt [nodejs-grafana] update README.md

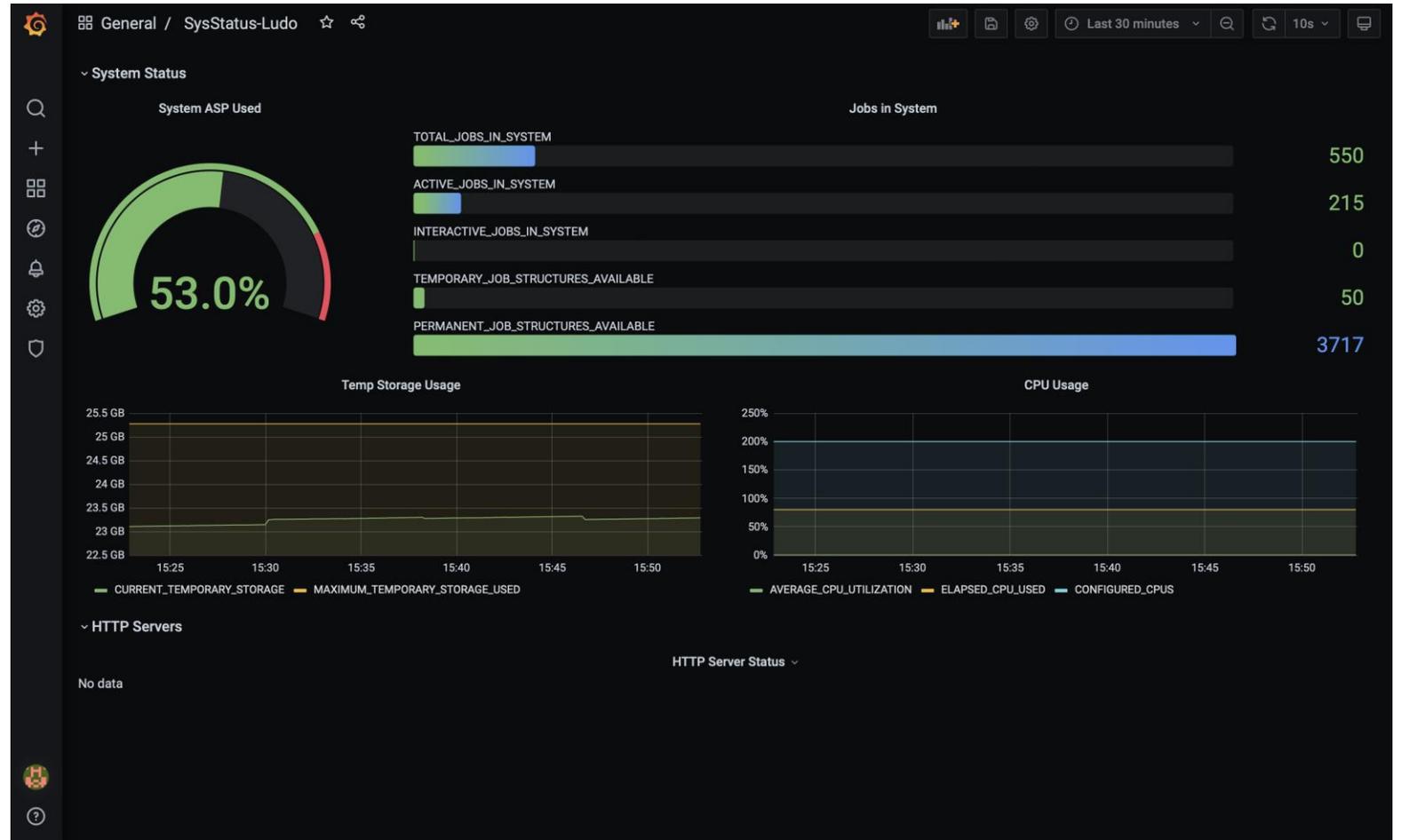
1 contributor

28 lines (24 sloc) | 838 Bytes

### Grafana Backend Example

#### On IBM i

```
git clone git@github.com:IBM/ibmi-oss-examples.git
cd ibmi-oss-examples/nodejs/grafana-backend/
npm i
node index.js
Server is listening to port 3333
```



<https://github.com/IBM/ibmi-oss-examples/blob/master/nodejs/grafana-backend/README.md>

## Sur la partition IBM i

```
git clone git@github.com:IBM/ibmi-oss-examples.git
cd ibmi-oss-examples/nodejs/grafana-backend/
npm i
node index.js
Server is listening to port 3333
```

## Sur le serveur Grafana

1. Install the SimpleJson plugin -> <https://grafana.com/grafana/plugins/grafana-simple-json-datasource/>

Note: If you are using Docker, use this to run Grafana with the SimpleJson plugin

```
docker run -d \
  -p 3000:3000 \
  --name=grafana \
  -e "GF_INSTALL_PLUGINS=grafana-simple-json-datasource" \
  grafana/grafana
```

<https://github.com/IBM/ibmi-oss-examples/blob/master/nodejs/grafana-backend/README.md>

[IBM i on Power - Performance FAQ](#)

[IBM i 7.5 performance](#)

[IBM i Technology Updates](#)

[Contact IBM Systems Lab Services](#)

## ➤ **La prestation de service :**

- ✓ Analyse des performances de votre partition IBM i
  - Par la collecte de performances de la journée / weekend
  - Par la collecte d'un plan cache
  - Utilisation des outils QMGTOOLS pour la collecte d'informations de l'environnement de l'IBM i
  - Restitution de l'analyse
  - Discussion / recommandations par une session de WebConf à distance

## ➤ **Gratuit si vous disposez d'un voucher IBM i –**

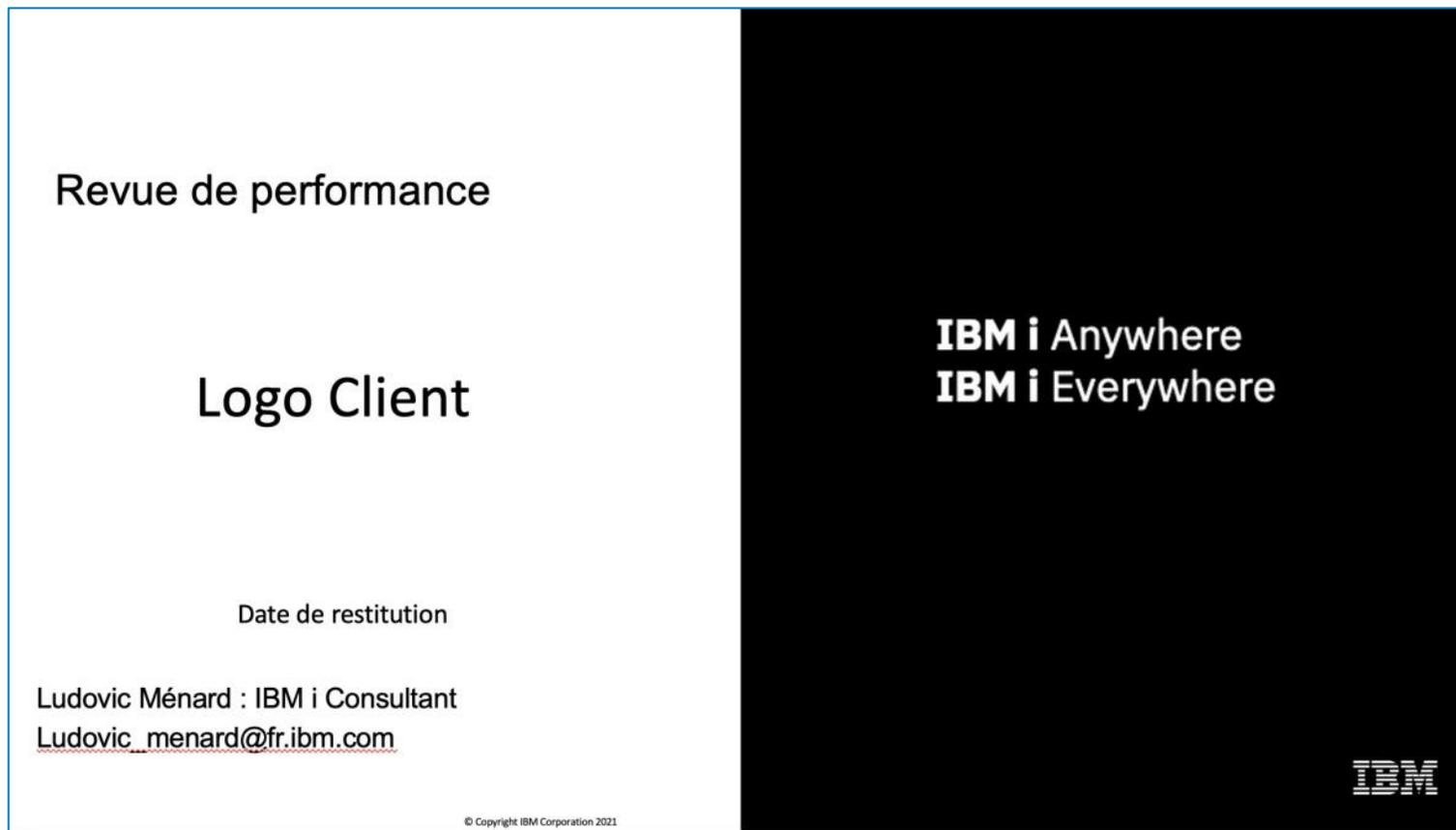
➤ contacter : Ludovic Ménard → [ludovic\\_menard@fr.ibm.com](mailto:ludovic_menard@fr.ibm.com)

- ✓ avec le(s) numéro(s) de série de votre(vos) serveur(s) pour connaître votre éligibilité aux vouchers
- ✓ ou pour connaître les conditions financières si vous ne disposez pas de vouchers

- Cette analyse de performances **n'inclut pas** :
  - ✓ Utilisation de Job Watcher pour analyser un travail spécifique
  - ✓ Utilisation de l'outil PEX pour analyser en profondeur un travail spécifique
  - ✓ Analyse de programmes
  - ✓ Analyse de requêtes SQL spécifiques
  
- Ces demandes peuvent faire l'objet de prestations complémentaires

- CPU
- Mémoire
- Disques
- Waits – Locks
- Journalisation
- SQL (vue d'ensemble)
- Plan cache
- Configuration de l'infrastructure
  - Configuration de la partition
  - VIOS / switch / stockage

Rapport (environ 30 pages) + restitution d'1 heure en WebConf



Explications et recommandations  
Support PDF en français



# Questions / Réponses



# Merci

