

IBM Power Systems - IBM i

Modernisation, développement d'applications et DB2 sous IBM i Technologies, outils et nouveautés 2013-2014

13 et 14 mai 2014 – IBM Client Center Paris, Bois-Colombes

S19 - Performances IBM i : focus sur les indicateurs clé

Mercredi14 mai – 11h00-12h30

Intervenant Eric Saglier





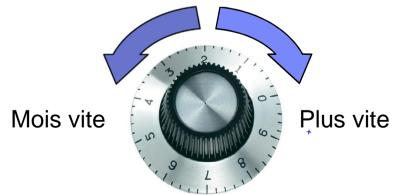
Plan de la présentation S19

- 1. Contrôle des sous-systèmes et pools mémoires
- 2. Analyse de l'activité CPU, IO, mémoires et travaux.
- 3. Vérifications des valeurs Systèmes Travaux
- 4. Pagination
- 5. Eligibilité aux disques SSD
- 6. Gestion du cache SQL.
- 7. Base de données



Tuning: Ce dont chacun aimerait disposer







Tuning: Ce dont on dispose







Nombre de cœurs





Niveau d'activité

Expert Cache



Taille des pools



Taux de défauts page



Classe



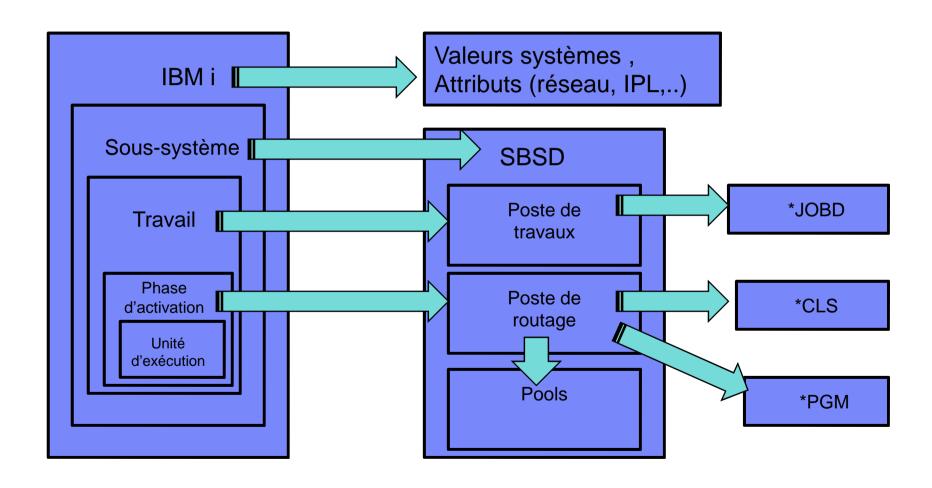
Cache SQL



Travaux à démarrage anticipés



Objets de la gestion des travaux





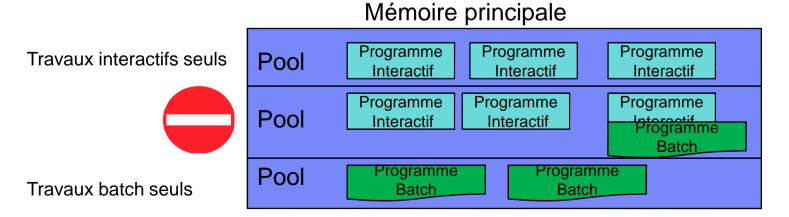
La mémoire Principale

- La RAM est découpée en pool, c'est un espace logique de la mémoire principale où s'exécutent les travaux
 - La taille d'un pool définit un montant de mémoire
 - Les pages ne sont pas obligatoirement contigües
 - La pagination d'un travail est limitée à ce pool
 - Niveau d'activité : nombre maximum de travaux "actifs" de ce pool
 - 10 pools maximum par sous-système
 - 64 pools maximum au total sur une partition.



Pourquoi des pools

- Les travaux de nature différentes (batchs, interactifs,...) ne doivent pas, en général, s'exécuter dans la même zone (pool) de mémoire principale. En effet, leurs besoins de mémoire sont la plupart du temps conflictuels. C'est une des bases du tuning.
- Les moniteurs des sous-systèmes routent chaque travail dont ils ont la charge vers un pool mémoire, et sont donc à même d'effectuer le dispatching nécessaire.





Caractéristique d'un travail

- Un TRAVAIL (JOB) est l'unité de traitement de base de l'OS.
- La PHASE D'ACTIVATION (ROUTING STEP) conduit aux attributs d'exécution du travail. A un instant donné, un travail démarré est dans une seule phase d'activation.
- La phase d'activation conduit à l'appel d'un programme.
- Chaque travail a au moins une phase d'activation.
- La plupart du temps, un travail ne comporte qu'une seule phase d'activation.
- A partir de la V4R2, la notion d'unité d'exécution (thread) apparaît dans l'OS/400. C'est le cas pour JAVA, HTTP, de certaines fonctionnalités DB2 UDB, par exemple.
- S'exécute au sein d'un pool de sous-système.



Composant d'un sous-système

- Un sous-système comprend
 - Un nom
 - Des pools mémoires
 - Un nombre maximum de travaux
 - Des entrées postes de travaux
 - Jobq
 - Écran
 - Communication
 - Démarrage automatique
 - Travaux à démarrage anticipés
 - Des postes de routage



Définition des sous-systèmes

- Proscrire le mélange de travaux de nature et de priorité différentes au sein du même pool mémoire.
- Le moniteur de sous-système s'exécutant dans le premier pool, configurer les sous-systèmes avec comme premier pool, le pool *BASE
- Définir plusieurs pools au sous-système (au moins 2)
- Définir les postes de routage en conséquence.



Architecture d'un sous-système

Postes de travaux

Description d'un sous-système	
Pools() Maxjobs()	CRTSBSD
File d'attente de travaux	ADDJOBQE
Postes écrans	ADDWSE
Travaux à démarrage automatique	ADDAJE
Entrée de communication	ADDCMNE
Travaux à démarrage anticipé	ADDPJE
Poste de routage	ADDRTGE



Les pools mémoire des sous-systèmes

- Deux types de pools
 - Privés: CRTSBSD SBSD(XYZ) POOLS((1 *BASE) (2 20000 5))
 - Un pool privé ne peut être utilisé que par le sous-système qui l'a alloué.
 - Il faut indiquer une taille de pool et un niveau d'activité pour un pool privé à la création de la description du sous-système.
 - Partagés: CRTSBSD SBSD(XYZ) POOLS((1 *BASE) (2 *INTERACT))
 - Un pool partagé peut être référencé et utilisé par plusieurs soussystèmes.
 - Son nom commence par * et peut être
 - *INTERACT
 - *BASE
 - *SPOOL
 - *SHRPOOL1 à *SHRPOOL60
 - Le niveau d'activité et la taille se modifie via la commande CHGSHRPOOL.



Les postes de travaux

- Les postes de travaux sont ajoutés à la description initiale de soussystème. Ils permettent au moniteur de sous-système d'identifier les travaux à exécuter.
- Chaque travail a toujours une JOBD associée. La JOBD utilisée est celle indiquée dans le poste origine du travail, ou dans le SBMJOB (cas du batch).

Poste de travail

*ALL *USRPRF | Nom | Jobd |

DSP* *USRPRF

File d'attente de travail

File d'attente

QBATCH

Travaux à démarrage anticipé

Prog	Jobd
PGM1	JOBD

Travaux à démarrage automatique

Travail	Jobd
STARTJ	STARTJD

Entrée de communications

Туре	Jobd
*ALL	*USRPRF



La table de routage

- La table de routage dirige les travaux vers les pools pour qu'ils s'exécutent et affecte une classe donnant la priorité d'exécution. Le poste de routage définit aussi le programme à exécuter au démarrage du travail.
- Un travail est routé grâce à la donnée de routage (RTGDTA), en provenance de la description de travail (JOBD) ou du SBMJOB. Cette donnée de routage est comparée à la donnée de comparaison (CMPVAL) de chaque poste de routage, dans l'ordre des numéros. La recherche s'arrête quand la correspondance est trouvée. La phase d'activation est alors initialisée.

SEQNBR	CMPVAL	CLASS	PGM	POOLID
10	QCMDI	QINTER	QCMD	2
30	QCMDB	QBATCH	QCMD	1
9999	*ANY	QPGMR	QCMD	1



La classe

Spécifie les paramètres suivant :

Priorité d'exécution
 RUNPTY

Tranche de temps
 TIMESLICE

Admis à sortir de mémoire
 PURGE

Temps d'attente par défaut
 DFTWAIT

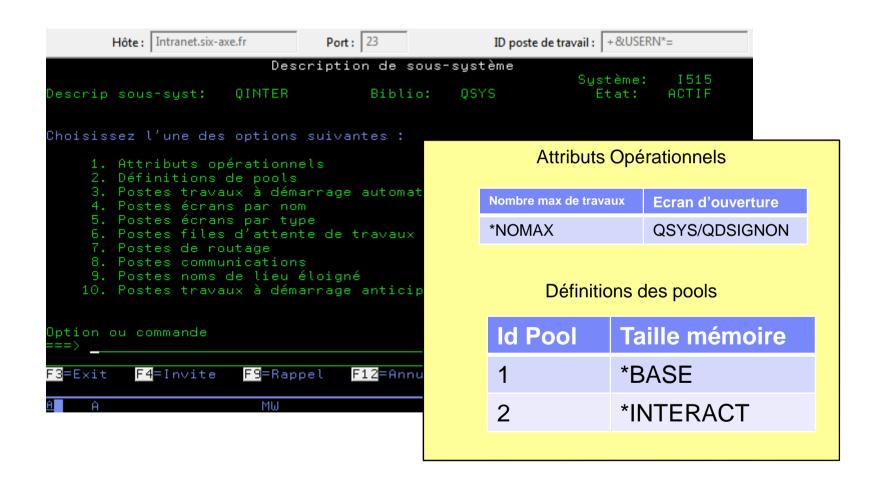
Temps maximal unité traitement CPUTIME

Mémoire temporaire maximale MAXTMPSTG

Nombre max unités exécution MAXTHD

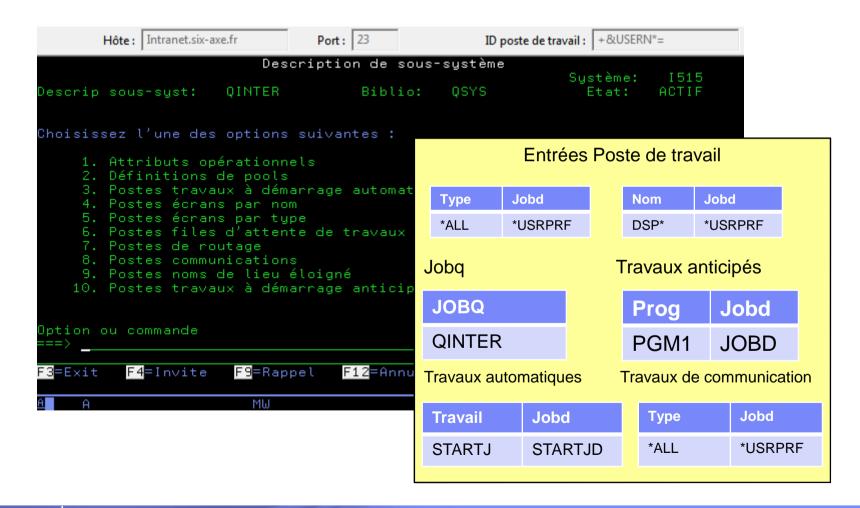


Description d'un sous-système 1/3



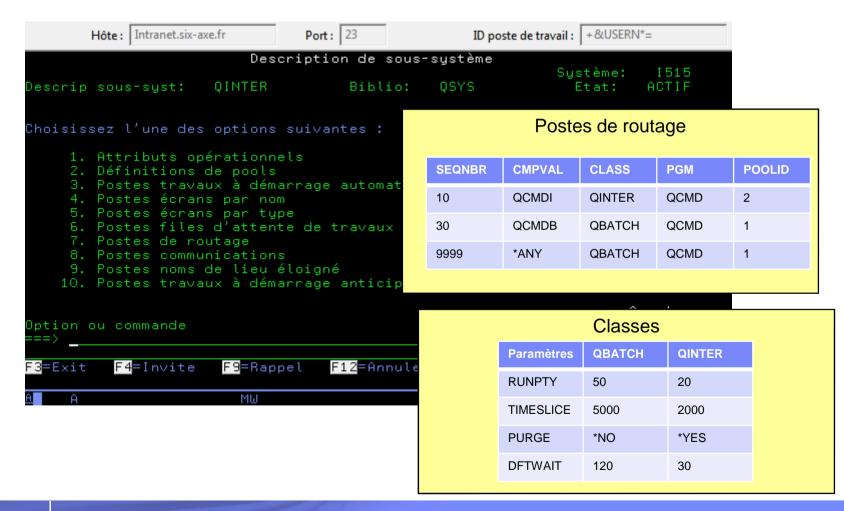


Description d'un sous-système 2/3



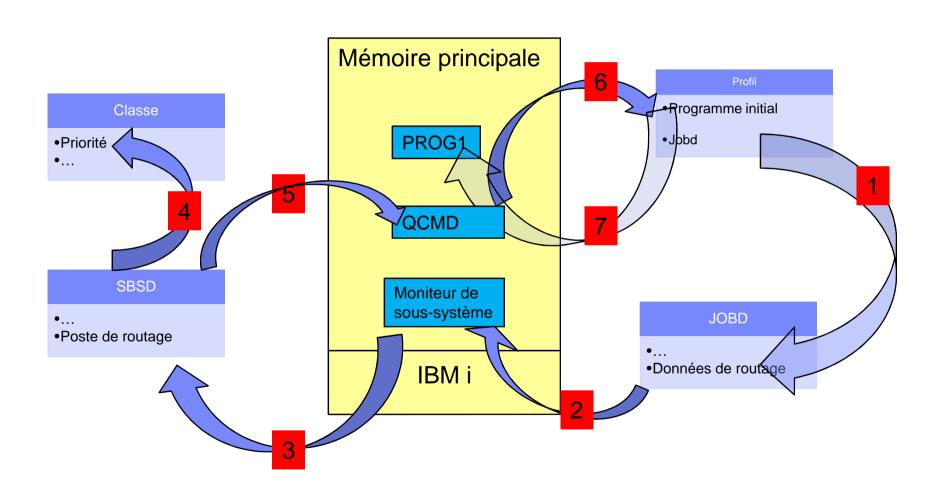


Description d'un sous-système 3/3





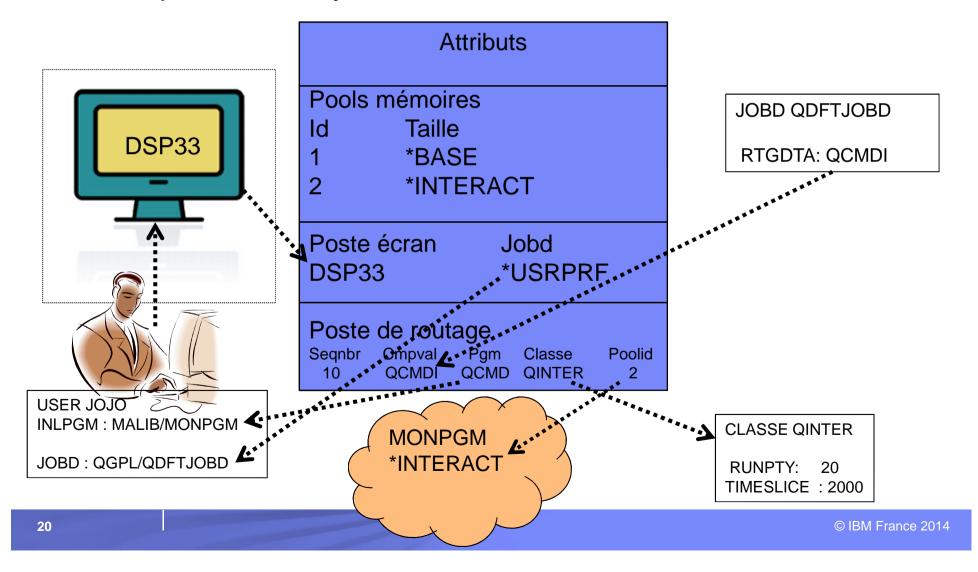
Prise en compte d'un Job





Cheminement d'un travail interactif

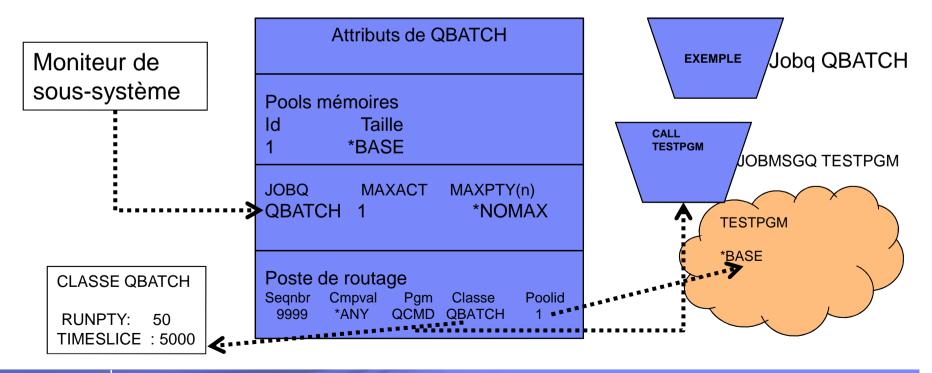
Description sous-système QINTER





Cheminement d'un travail Batch

- Le moniteur du sous-système utilise les postes de la description de sous-système pour sélectionner les travaux à exécuter.
- La donnée de routage 'QCMDB' est utilisée par défaut dans la commande SBMJOB.
- La JOBD utilisée par le travail est celle spécifiée dans le SBMJOB: la valeur par défaut est *USRPRF, ce qui fait remonter au profil du travail qui soumet. La valeur par défaut du profil est QGPL/QDFTJOBD.
- SBMJOB CMD(CALL PGM(TESTPGM)) JOB(EXEMPLE) JOBQ(QBATCH)





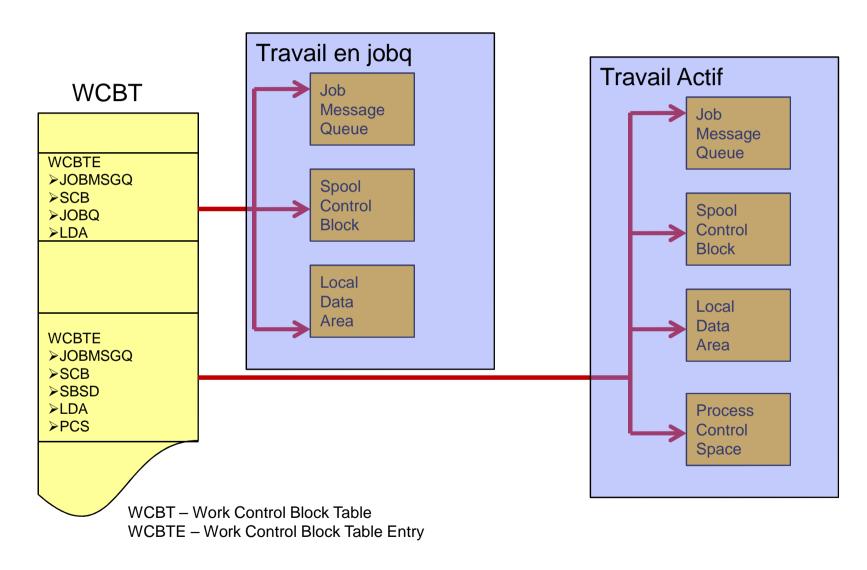
Les travaux anticipés.

- Penser à analyser l'utilisation des travaux anticipé au travers de la commande DSPACTPJ et tout particulièrement pour les travaux :
 - QZDASOINIT s'exécutant dans QUSRWRK
 - QSQSRVR s'exécutant dans QSYSWRK

```
ous-système . . . . :
                          OSYSWRK
                                           Date de réinit .
                                           Heure de réinit
rogramme . . . . :
                          OSOSRVR
                            OSYS
                                           Intervalle . . . .
 Bibliothèque . . . :
ravaux anticipés :
                                                        34,7
ravaux anticipés en cours d'utilisation :
                                                        16
                                                        15.7
                                                        16
Demandes de démarrage de programmes :
 Nombre en cours d'attente . . . . . . . .
 Mouenne en attente . . . . . . . . . . . . . . .
 Maximum en attente . . . . . . . . . . . . . .
 Temps d'attente moyen . . . . . . . . . . . .
                                                     00:00:00,0
                                                     18
 Nombre accepté . . . . . . . . . . . . . . . . .
```



Structures « permanentes » d'un travail



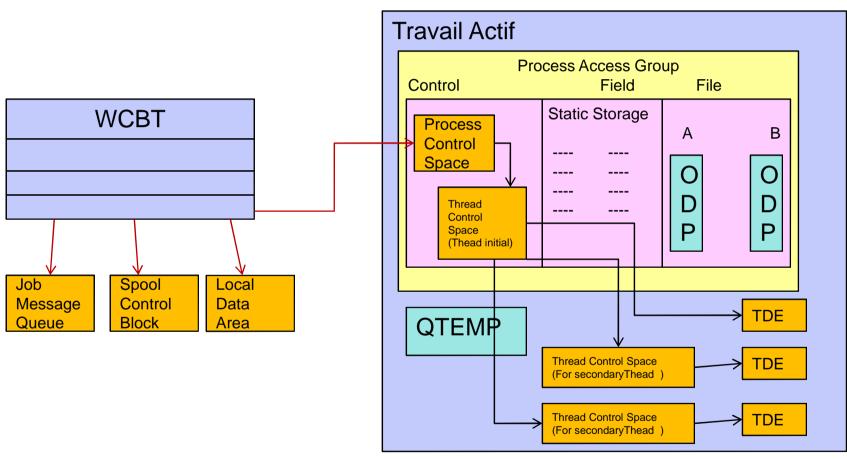


Le PAG

- C'est une structure temporaire d'un travail.
- Un par travail actif.
- Est en mémoire principale quand le travail a un niveau d'activité.
- Contient les informations sur le travail en cours.
- Le design des applications fait varier sa taille.
- Ces informations sont réparties dans des blocs physiques de 32K.
- La commande DSPJOB permet d'en visualiser l'essentiel.
- L'objectif est de diminuer la taille des PAG.



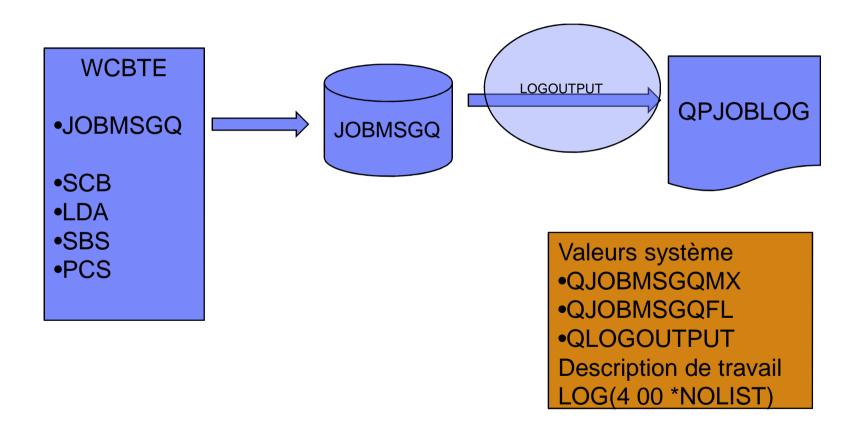
Structures « permanentes » d'un travail



WCBT – Work Control Block Table
WCBTE – Work Control Block Table Entry



File d'attente de message du travail



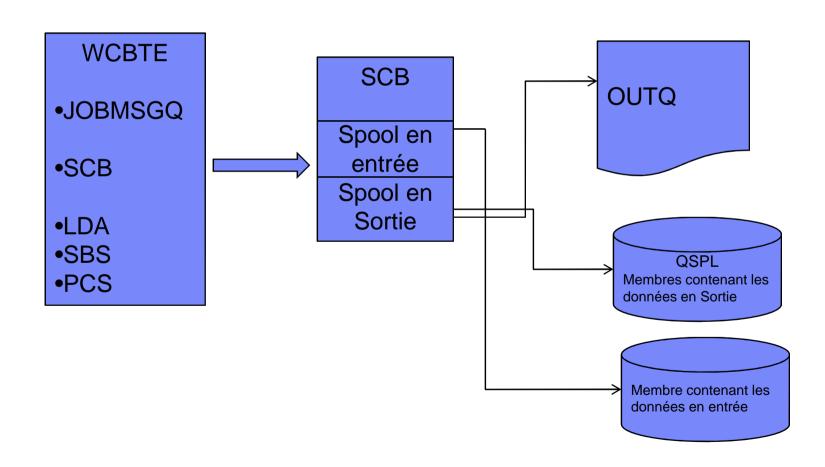


File d'attente de message du travail

- C'est un objet permanent de la structure d'un travail
- Sa taille est contrôlée par la valeur-système QJOBMSGQMX (16M par défaut)
- L'action à effectuer quand ce maximum est atteint est contrôlée par QJOBMSGQFL (*NOWRAP par défaut).
- ATTENTION, les valeurs-système QJOBMSGQSZ et QJOBMSGQTL sont périmées...
- RECOMMANDATIONS
 - Eliminer les causes de messages non nécessaires
 - Placer le niveau de l'historique à LOG(4 00 *NOLIST)
 - Changer QJOBMSGQFL à *PRTWRAP ou *WRAP.
- La production du fichier QPJOBLOG est conditionnée par le paramètre LOGOUTPUT du travail :
 - *SYSVAL Action de fichier spoule <u>*DETACH</u> *SAME, *SYSVAL, *KEEP...
 - *JOBLOGSVR
 - *JOBEND
 - *PND



Spool Control Block





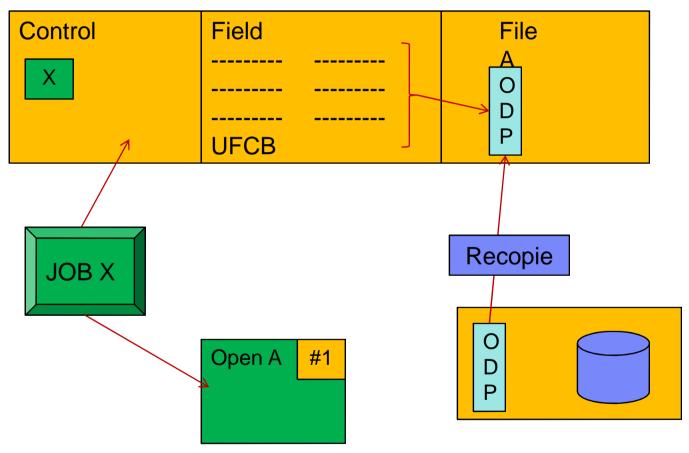
Spool Control Block

- C'est un objet permanent de la structure d'un travail.
- Sa taille initiale est contrôlée par la valeur-système QJOBSPLA (3516 octets par défaut). Correspond à 8 fichiers spools par job.
- L'élimination des membres vides est contrôlée par la valeursystème QRCLSPLSTG (8 jours par défaut). On peut aussi provoquer la récupération par la commande RCLSPLSTG.



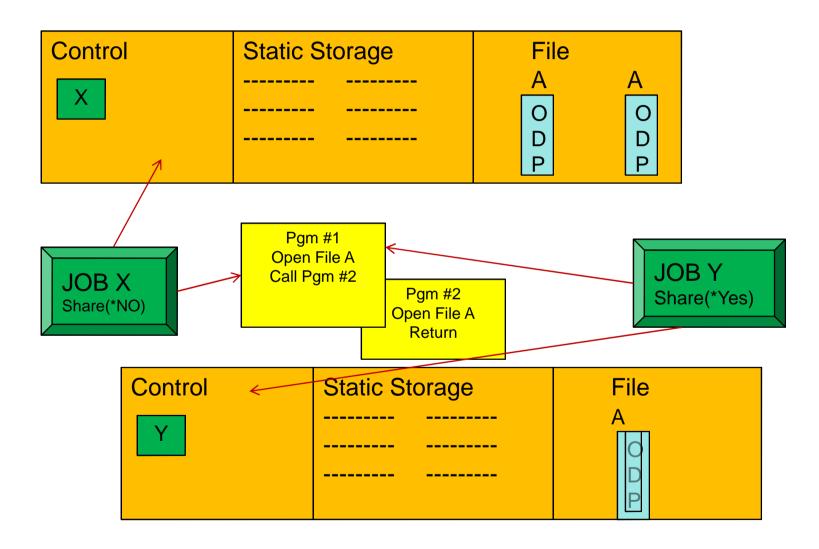
Open Data Path (ODP)

UFCB – User fiel Control Block





Partage: Open Data Path (ODP)





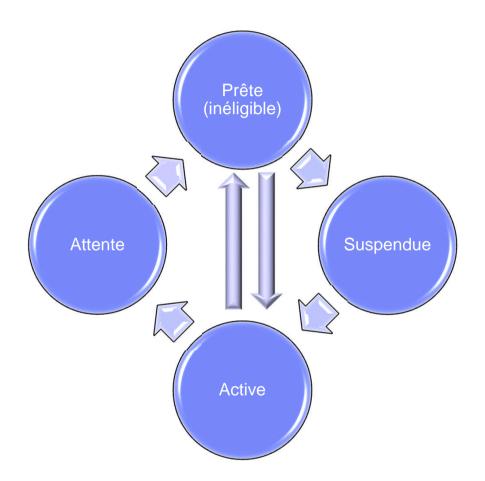
Structure des travaux et répartitions des tâches

- Les trois niveaux hiérarchiques
 - Processus
 - Objet du système MI (PCS)
 - Assemble les ressources nécessaire à l'invocation d'un programme consiste en :
 - Un programme exécutable (peut être partagé)
 - Les données de programme
 - Informations de l'Etat
 - Peut également être défini comme un programme en exécution
 - Travail
 - Objet OS/400 qui contient une structure de processus de gestion de système
 - Unité d'exécution
 - Une partie d'un processus
 - Un processus peut avoir plus d'une unité d'exécution
 - Chaque unité d'exécution est consiste en :
 - Jeu de registres représentant l'état du processeur
 - Une pile de contrôle
 - le stockage automatique de programmes (variables locales)
 - Les ressources système allouées au processus sont partagés avec d'autres unités d'exécution au sein du processus



Les états de transition d'une tâche

- Prête
- Suspendue
- Attente
- Active



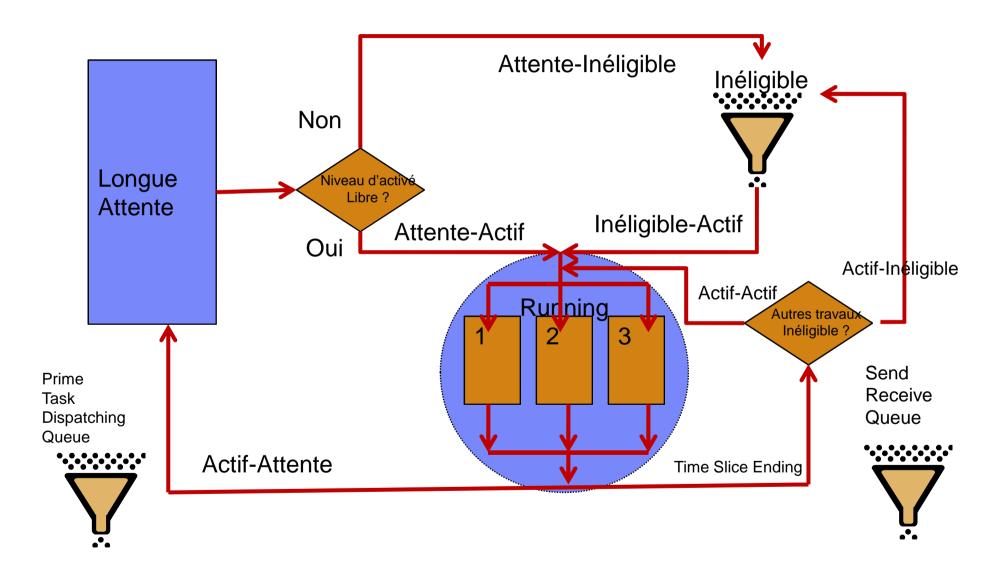


Les états d'une Thread et les transitions

- Un travail peut être dans trois états dans un sous-système: Actif en niveau d'activité, en Attente ou Inéligible.
- Les Transitions sont Act→Att (Actif vers Attente), Act→Inelg (Actif vers Inéligible), Inelg→Act (Inéligible vers Actif), TSE (Fin de tranche de temps), Att→Act (Attente vers Actif), Att→Inelg (Attente vers Inéligible).
- Les paramètres d'exécution sont définis par la CLASSE (RUNPTY, TIMESLICE).
- La PRIME TASK DISPATCHING QUEUE (TDQ) est une file d'attente unique pour le ou les processeur(s). Elle contient les travaux dans un niveau d'activité et prêts à être exécutés, ceux-ci sont ordonnés en FIFO par priorité.
- Il existe une file d'attente inéligible (INELIGIBLE QUEUE) par pool actif.
 Elle contient les travaux en attente de niveau d'activité, classé par priorité.
- Il existe plusieurs SEND/RECEIVE QUEUES spécialisées, pour les travaux non prêts à être exécutés.



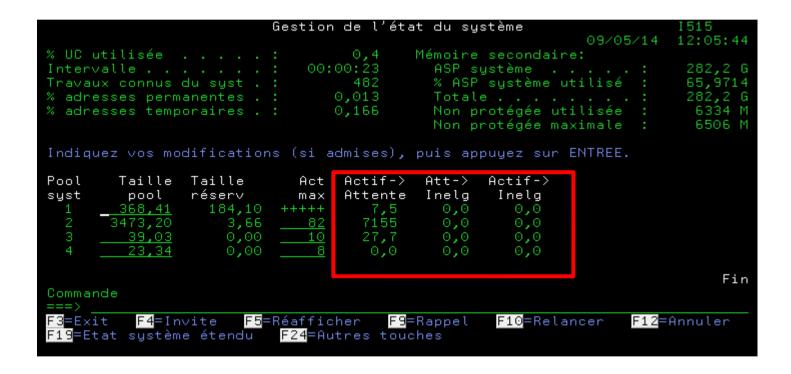
Les états d'une Thread et les transitions





Observer les transitions

WRKSYSSTS





Passage en longue Attente

LES CAUSES de LONGUE ATTENTE:

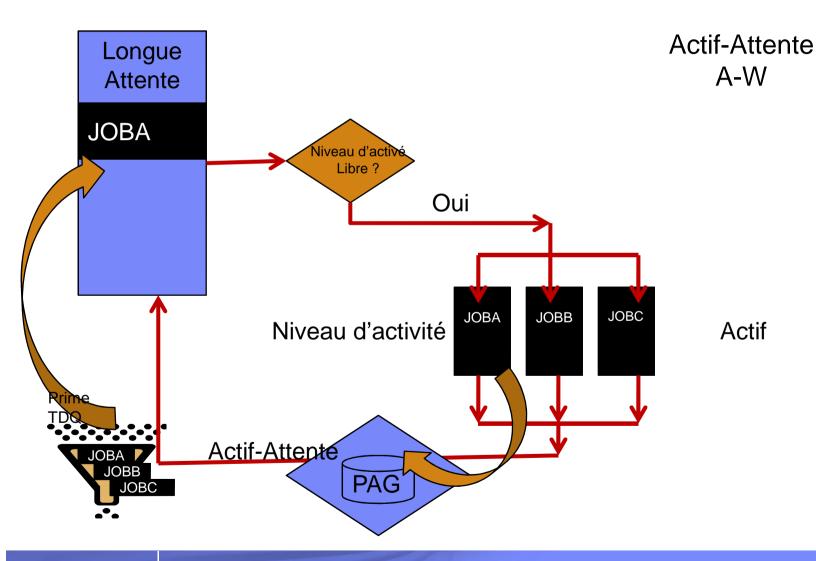
- Les travaux demandant une lecture d'un poste de travail ou d'une file d'attente de messages.
- Attente à la suite d'un verrouillage sur un fichier, une file d'attente de messages, un enregistrement base de données.
- Un programme écrivant des données sur un support à vitesse lente.

• QUE SE PASSE-T-IL ?:

- Une transition est comptabilisée dans la rubrique Act→Att, et le travail perd son niveau d'activité.
- Le compteur de la tranche de temps est remis à zéro.
- Le travail reste en longue attente jusqu'à la fin de l'événement qui cause cette longue attente (il quitte la TDQ et est placé dans une Send/Receive Queue).
- Si un travail se trouve dans la file d'attente lui affecte alors le niveau d'activité libéré



Actifs-Attente



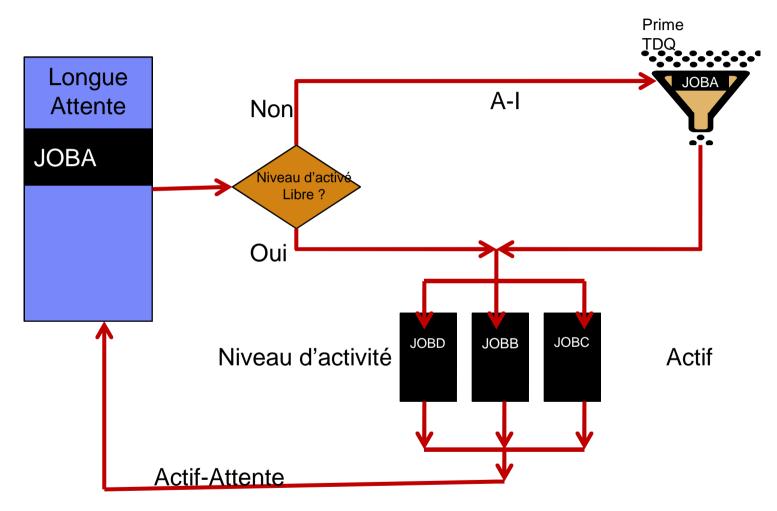


Attente - Inéligible

- L'opération qui avait mis le travail en longue attente s'est achevée.
 - Aucun niveau d'activité n'étant disponible, le travail est placé dans l'état Inéligible, et attendra qu'un niveau d'activité se libère.
 - La transition apparaît sur le panneau WRKSYSSTS dans la colonne Att→Inelg.



Attente-Inéligible



Attente -Inéligible A-I

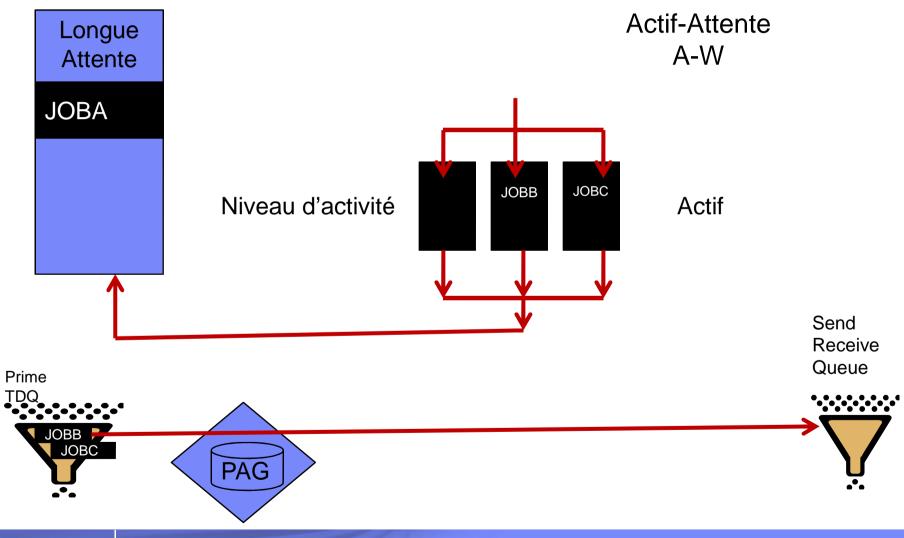


Attente dans un niveau d'activité

- Ce type d'attente s'appelle un SHORT WAIT.
- CAUSES D'ATTENTE DANS UN NIVEAU D'ACTIVITE:
 - Ouverture/fermeture de fichier
 - Gestion de message
 - Opération Base de Données
 - Défaut de page
 - Sauvegarde d'écran
 - Attente sur une DTAQ
- QUE SE PASSE-T-IL ?:
 - Le travail reste dans son niveau d'activité.
 - Pas de transition indiquée dans le panneau WRKSYSSTS.
 - Le PAG peut rester en mémoire principale.
 - Le travail quitte la Prime TDQ et est placé dans une Send/Receive Queue.
 - Si l'attente dure plus de 2 secondes, le travail passera en longue attente.



Attente dans un niveau d'activité

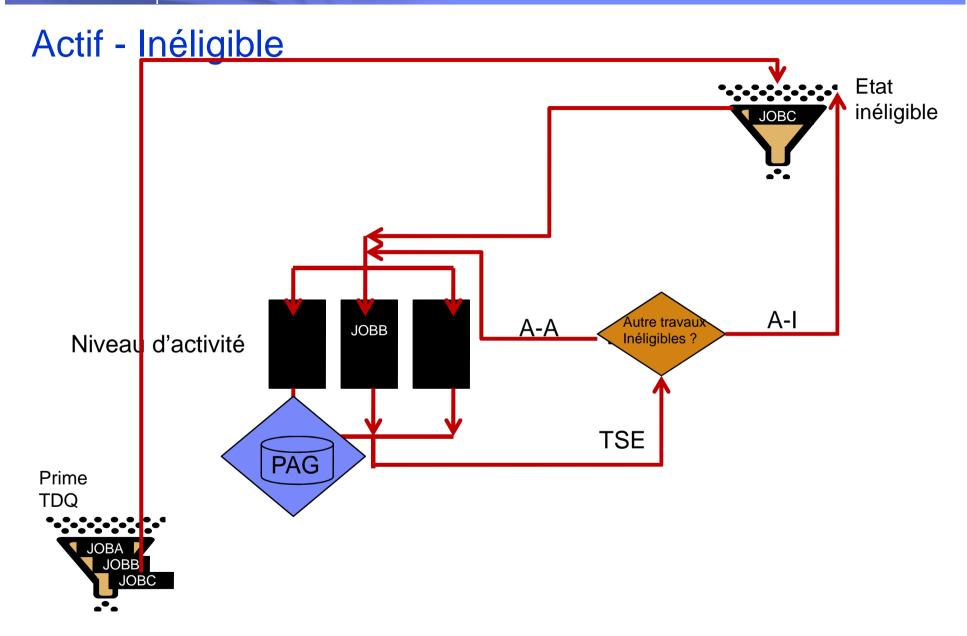




Actif - Inéligible

- La durée de la tranche de temps est indiquée dans la classe.
- A la fin de la tranche de temps, si aucun travail de même priorité ou de priorité supérieure n'est en attente dans la file d'attente inéligible, le travail obtient une nouvelle tranche de temps.
- A la fin de la tranche de temps, si des travaux de même priorité ou de priorité supérieure se trouvent dans la file d'attente inéligible, le travail est alors transféré dans la file d'attente inéligible. Une transition est alors comptabilisée (Act -> Inelg), et le travail doit libérer certains verrous.







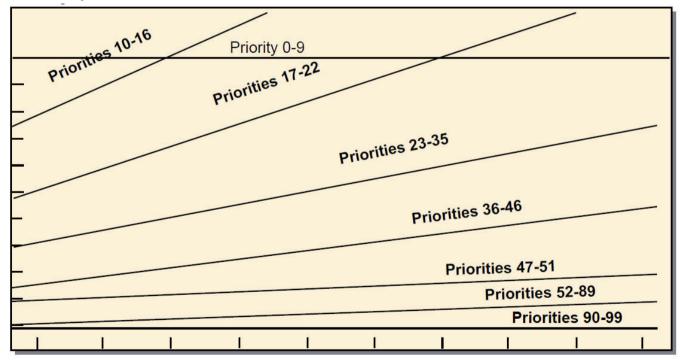
Ajustement dynamique de priorité

- Pour éviter qu'un travail monopolise le processeur
 - Sa priorité est peu à peu dégradée
- Les travaux de priorité inférieure ont plus de chance de pouvoir s'exécuter.
- Ce processus est activé via la valeur système QDYNPTYSCD
- Le processus d'ajustement ne se produit qu'aux fortes consommations du processeur (plus de 80% environ).
- La valeur-système QDYNPTYADJ, si mise à 1, permet d'éviter que l'interactif change de priorité (mais ce peut être au détriment du reste de l'activité...).



Abaques d'ajustement

- Les travaux de priorité 0-9 et 90-99 ne font l'objet d'aucun ajustement.
- Les autres ont une priorité modifiable pour tenir compte de leurs consommations respectives précédentes. Cependant, la priorité d'origine est périodiquement restaurée.



Time (Millions of cycles)



Considérations Multi-processeurs

- Le dispatching est basée sur la priorité, l'affinité du cache ainsi que sur l'admissibilité
- Trois bits dans uneTask Dispatching Entry afin de gérer l'admissibilité
 - Un bit pour chaque processeur
 - Indique que la tâche est éligible pour s'exécuter sur le processeur correspondant
 - Un bit (si Actif)
 - Un bit pour chaque processeur
 - Indique sur quel processeur la tâche est en cours d'exécution
 - Un bit d'affinité
 - Un bit pour chaque processeur
 - Indique sur quel processeur la tâche a été en cours d'exécution

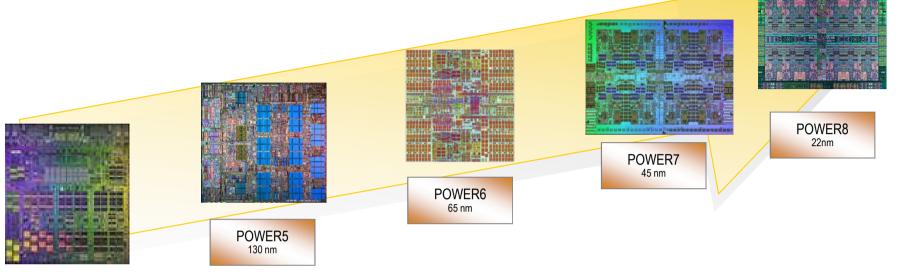


Valeur système d'allocation

- QTOTJOB (valeur par défaut 200) Spécifie le nombre initial de travaux pour lequel un espace de mémoire auxiliaire (Work Control Block Table) est alloué au démarrage de la machine, cette valeur concerne les travaux :
 - En Activité
 - En Jobq
 - En Outq
- QADLTOTJ (valeur par défaut 30) Spécifie l'augmentation de taille à effectuer dès lors que la valeur de QTOTJOB est atteinte.
- QACTJOB (valeur par défaut 200) Spécifie le nombre initial de travaux actifs pour lequel un espace de mémoire auxiliaire est alloué au démarrage de la machine.
- QADLACTJ (valeur par défaut 30) Spécifie l'augmentation de taille à effectuer dès lors que la valeur de QACTJOB est atteinte.



Les Processeurs/Les Threads



POWER4 180 nm

- Double cœurs
- Module multi-puce
- Switch distribué
- Cache L2 partagé
 Dynamic LPARs (32)

- Double cœurs
- SMT+ (2)
- Switch distribué +
- Parallélisme +
- Performance FP +
- · Amélioration de la bande passante mémoire
- Virtualisation

- Double cœurs
- SMT + (2)
 Hautes fréquences d'horloge
- Virtualisation +
- Amélioration sous-système mémoire
- Altivec
- Instruction Retry
- Gestion de l'énergie
- SMT +
- Protection Keys

- Multi Cœurs (6-8)
 SMT ++ (4)
 eDRAM sur la puce

- Performance / cœur
 Amélioration sous-système mémoire
- SMT++
- Fiabilité
- VSM & VSX (AltiVec)
 Protection Keys+

- ■Multi Cœurs (12)
- ■SMT+++ (8)
- ■Reliability ++
- ■CAPI Support
- ■Transactional Memory
- Operating System booted

2001 2004 2007 2010 2014



WCBTE – Work Control Block Table Entry

- Tout travail initialise une entrée dans la WCBT :
 - En JOBQ
 - En Activité
 - En OUTQ
- Le nombre de tables de travaux est déterminé par la valeur système QMAXJOB :
 - valeur par défaut 163520 10 tables
 - Valeur maximale 485000 (V7R1) 30 tables



Observation tables des travaux

Commande DSPJOBTBL





Comprimer les tables de travaux

- Modifier les attributs d'IPL CHGIPLA
 - Paramètre CPRJOBTBL (valeur par défaut *NONE)
 - *NEXT
 - *NORMAL
 - *ABNORMAL
 - *ALL
 - *NONE



Diminuer le nombre travaux => Réduction taille WCBT

- Diminuer le nombre travaux en outq
 - Nettoyer les outq (si possible...)
 - Détacher les états des travaux les ayants produits

Moyens

- Positionner la valeur système QSPLFACN à *DETACH
- Modifier les JOBD utilisées en positionnant le paramètre SPLFACN à *DETACH
- Changer les travaux Commande CHGJOB paramètre SPLFACN

```
Définition de valeur pour paramètre SPLFACN

Indiquez votre choix, puis appuyez sur ENTREE.

Action de fichier spoule . . . . <u>*SYSVAL</u>

*SAME
*SYSVAL
*KEEP
*DETACH
```



Techniques diverses 1/2

- Afin de ne pas avoir de surconsommation de RAM diminuer la taille des PAG
 - Partager autant que faire se peut les ODP
 - Diminuer la taille des programmes
 - Option (*NOUNREF) en RPGV en V6-V7
 - Charger les programmes de services en activation différée.

 Et surtout ne pas mélanger dans les mêmes pools mémoires des travaux de nature et des priorités différentes



Techniques diverses 1/2

- Pour limiter un processus sur une machine multi-cœurs, activé le concept de « WORKLOAD CAPPING GROUP »
 - Créer un groupe de processeur Commande ADDWLCGRP

```
Add Workload Capping Group (ADDWLCGRP)

Indiquez vos choix, puis appuyez sur ENTREE.

Workload capping group . . . . WLCGRP

Processor limit . . . . . . . PRCLMT 1
```

- Modifier les travaux au travers de la commande CHGJOB
 - Possibilité d'indiquer le « Workload Capping Group »
 - Possibilité de limité le temps CPU CPUTIME
 - Possibilité de limité la mémoire temporaire





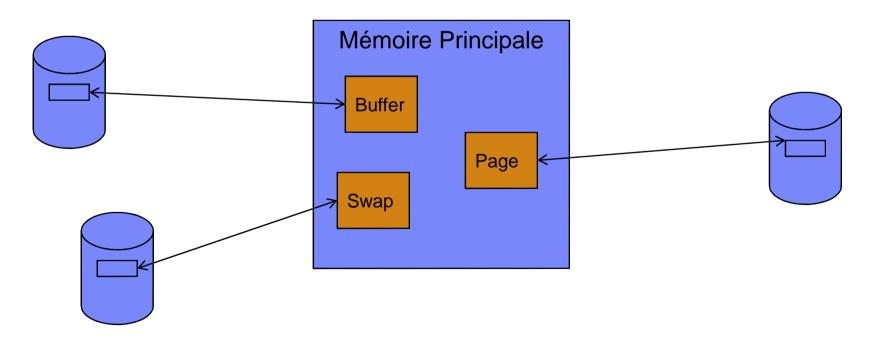
La pagination

- Le concept de l'espace adressable unique de l'OS/400 est mis en œuvre sur les 2 types de mémoires :
 - La mémoire principale
 - l'espace disque.
- Plusieurs mécanismes permettent de transférer les données entre ces 2 mémoires. Le microcode gère automatiquement les demandes.



Gestion de la mémoire

 Le système gère toutes les requêtes d'entrée/sortie avec une seule instruction machine: Perform Paging Request (PPR).
 Le nombre de pages transférées est déterminé par la fonction demandée et le contexte





Lectures Synchrones / Asynchrones

SYNCHRONE

 le programme attend que l'opération soit terminée. Une demande de lecture base de données en accès sélectif génère en général plusieurs demandes d'E/S synchrones.

ASYNCHRONE

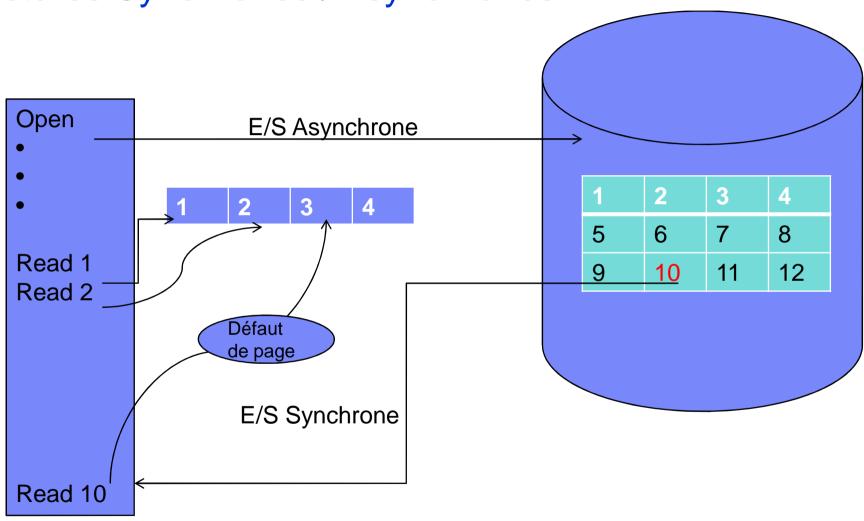
 le système exécute l'opération de lecture avant que le travail ne le demande. Des lectures asynchrones sont par exemple exécutées dans les cas suivants: ouverture de fichier, lecture base de données en accès séquentiel.

DEFAUT DE PAGE

 se produit quand la page demandée n'est pas en mémoire principale. Un défaut de page va générer une demande de pagination qui active une tâche d'E/S synchrone. Le travail attendra l'achèvement de cette tâche pour reprendre l'exécution.



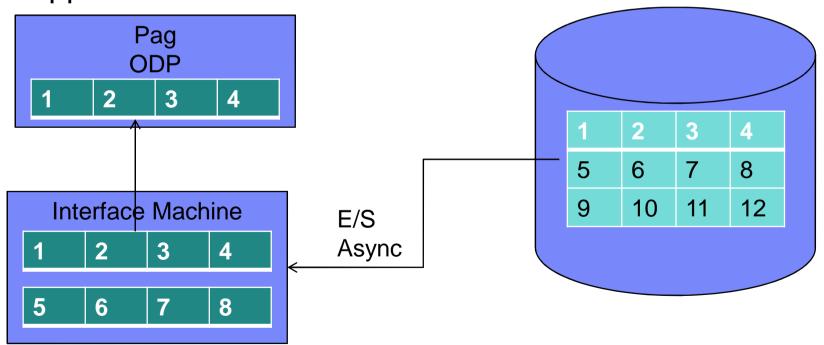
Lectures Synchrones / Asynchrones





Double buffering

 A chaque fois que le système reconnaît qu'un fichier Base de données est traité en séquentiel ou qu'un utilisateur a demandé un groupe d'enregistrements, l'opération d'E/S va automatiquement remplir un second buffer. Cette opération s'appelle un ADAPTIVE LOOK AHEAD BRING





OVRDBF SEQONLY(*YES Nb. enregistrements)

- On fournit dans le mot-clé SEQONLY le nombre d'enregistrements passés au PAG (par défaut buffer OS/400 de 4K). Les enregistrements seront traités en ordre d'arrivée. Ceci est valable pour les exécutions en lecture ou écriture seules. Cette technique peut améliorer sensiblement la performance du travail.
- Possibilité d'indiquer des tailles de buffer.

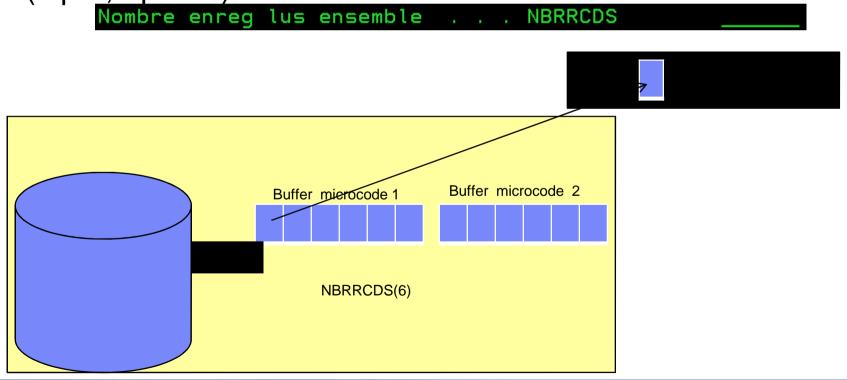
```
Type . . . . . . . . . . ENTIER
Nombre d'enregistrements . . .

*BUF32KB
*BUF64KB
*BUF128KB
*BUF256KB
```



OVRDBF NBRCDS(Nb. enregistrements)

- On fournit dans le mot-clé NBRRCDS le nombre d'enregistrements groupés dans le buffer du microcode.
- Peut être utilisé pour n'importe quelle ouverture de fichiers (input, update).



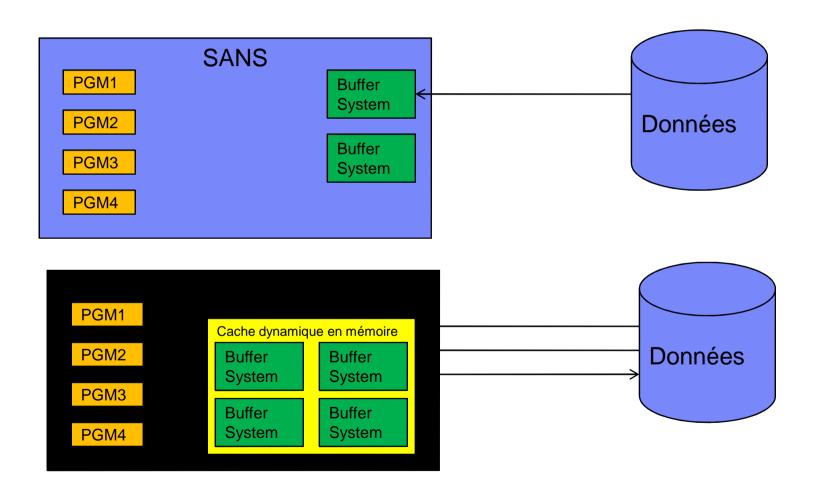


Principe de l'Expert Cache

- L'Expert Cache est un cache dynamique en mémoire principale.
- L'Expert Cache peut améliorer les performances dans les conditions suivantes:
 - La mémoire principale disponible est significative.
 - La CPU n'est pas très chargée.
 - Les E/S impactent les performances.
 - Les données sont utilisées en séquentiel ou les mêmes données sont souvent réutilisées.



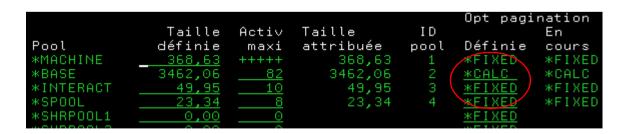
L'Expert Cache





Mise en œuvre Expert Cache

- L'Expert Cache est mis en œuvre par l'intermédiaire des commandes WRKSHRPOOL ou CHGSHRPOOL (pour les pools partagés).
- Il faut indiquer l'option de pagination *CALC à la place de *FIXED pour bénéficier de cette possibilité.
- On peut aussi utiliser WRKSYSSTS pour les pools actifs.
- N'est possible que pour les pools partagés





Ecritures Synchrones/Asynchrones

SYNCHRONE

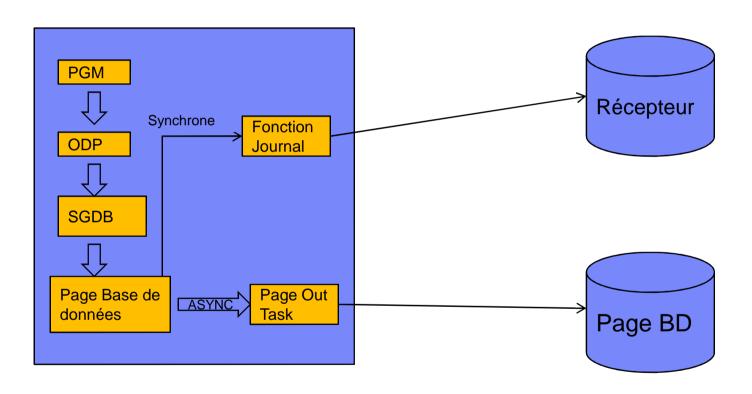
 le programme attend la fin de l'exécution de la tâche d'écriture. Ceci arrive quand une instruction de pagination WRITE est générée comme dans le cas de la journalisation ou quand le système n'a pas pu libérer des pages à temps (systèmes surchargés).

ASYNCHRONE (PURGE BEHIND)

 la demande de pagination PURGE est exécutée par les tâches de PAGE-OUT qui écrivent les pages modifiées sur disques quand le système n'est pas occupé. Le PURGE est exécuté dans les cas de MAJ, suppression et d'ajouts d'enregistrements dans la Base de données.



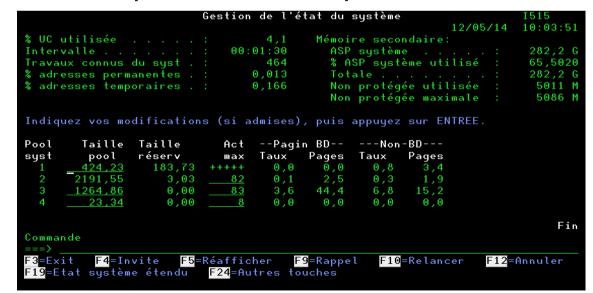
Ecritures Synchrones/Asynchrones





Observer les défaut de page

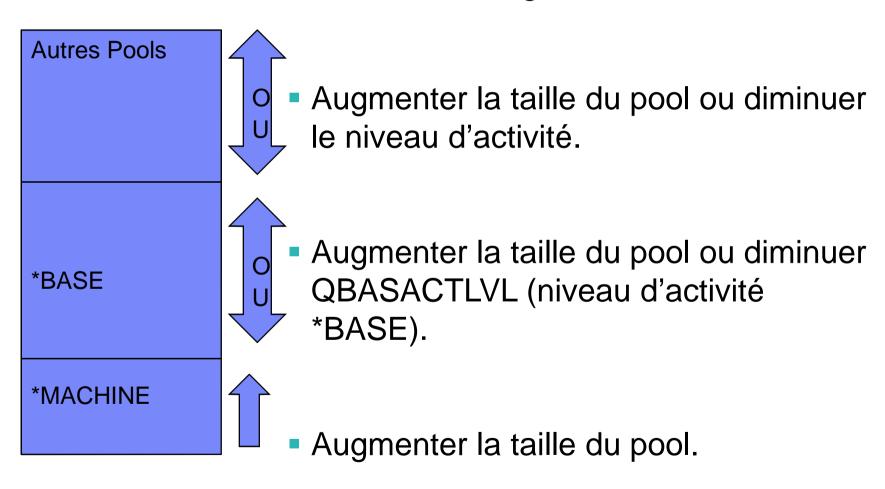
- Les taux de pagination sont indiqués par seconde.
- Un intervalle d'observation doit être de l'ordre de quelques minutes.
- Examiner les périodes et les pointes typiques.
- Prêter une attention particulière au pool *MACHINE





En cas de pagination importante

Diminuer la Pagination





Ajustement automatique

- Mise en œuvre au travers de la valeur système QPFRADJ
 - 0 : Pas d'ajustement
 - 1 : Ajustement des pools partagés uniquement à l'IPL
 - 2 : Comme 1 Avec ajustement dynamique.
 - 3 : Ajustement dynamique seul, sans recalcul à l'IPL.
- L'ajustement de la taille des pools et des niveaux d'activité ne concerne que les POOLS PARTAGE (*MACHINE, *BASE, *INTERACT, *SPOOL, *SHRPOOL1-60).
- L'ajustement dynamique échantillonne environ toutes les 60 secondes.
- Les modifications effectuées peuvent être consignées dans le journal QSYS/QPFRADJ : il suffit de créer ce journal.
- L'ajustement dynamique peut être contrôlé via une data Area : QPFRADJWT ; à créer dans QUSRSYS avec une valeur en seconde comprise entre 20 et 120.
 - CRTDTAARA DTAARA(QUSRSYS/QPFRADJWT) TYPE(*DEC) LEN(3 0) VALUE(60)



Contrôle de QPFRADJ 1/2

- La taille mémoire du pool ainsi que le niveau d'activité peut être fait via la commande CHGSHRPOOL et les paramètres :
 - MINFAULT
 - JOBFAULT
 - MAXFAULT
 - PTY
 - MINPCT
 - MAXPCT

- La priorité d'ajustement par rapport aux autres pools est contrôlée aux travers du paramètre :
 - PTY



Contrôle de QPFRADJ 2/2

- La commande WRKSHRPOOL, permet de contrôler les pools partagés.
- Ne pas hésiter à privilégier *BASE.





Identification des problèmes particuliers

- Travaux de priorité inégale dans un même pool.
- Travail présentant un pourcentage UC ou nombre d'E/S élevé.
- Temps de réponse moyen élevé.
- Verrouillages.
- Batch en interactif
- ...etc
- => Utiliser (avec parcimonie) WRKACTJOB.



WRKACTJOB

 Attention aux nombre de travaux actifs, il convient de regarder également le nombre d'unités d'exécution afin de bien régler les niveaux d'activité et la taille RAM des pools.

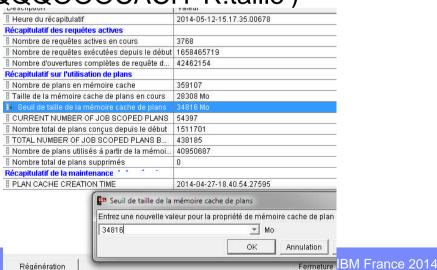
		Gest	ion des	travau	x actif	5		I515
% UC	: 1,9	Intervalle:	02:13:	55	Travau	12 x actifs:	255 255	11:44:44
						Unités		
0pt	S-syst/trav	Utilisat	Numéro	Type	% UC	exéc		
	ADMIN2	QLWISVR	820818	BČI	0,0	40		
	ADMIN3	QLWISVR	820823	BCI	0,0	27		
	ADMIN4	QWEBADMIN	820530	BCI	0,0	37		
	ADMIN5	QLWISVR	820532	BCI	0,0	28		
	IWAIAS	QLWISVR	820811	BCI	0,0	24		
	IWAIAS	QTMHHTTP	820497	BCH	0,0	1		
	IWAIAS	QTMHHTTP	820511	BCI	0,0	1		
	IWAIAS	QTMHHTTP	820526	BCI	0,0	47		
	ZENDSVR	QTMHHTTP	820498	BCH	0,0	1		
	ZENDSVR	QTMHHTTP	820512	BCI	0,0	1		
	ZENDSVR	QTMHHTTP	820513	BCI	0,0	1		
	ZENDSVR	QTMHHTTP	820518	BCI	0,0	48		
	ZENDSVR	QTMHHTTP	820557	BCI	0,1	1		
	ZENDSVR	QTMHHTTP	820558	BCI	0,0	1		
							A	suivre
===>								
F21=	Atticher inst	ructions/touc	hes					

74



Moteur SQE

- Avec l'apparition du moteur SQL (SQE) en V5R3M0 a été implémenté un cache SQL, d'une taille de 512 MO
- Ce cache est visualisable :
 - En interface Green Screen (Call QQQOOCACH)
 - Via System I Navigator (V5R4M0)
- Peut et doit être agrandi si besoin.
 - En interface Green Screen (Call QQQOOOCACH 'R:taille')
 - Via System I Navigator





Cache SQL

- Amélioration V7R1 avec Groupe PTF SF99701 niveau 26
- Redimensionnement automatique
- Indication complémentaire sur l'utilisation

Description	Valeur			
Heure du récapitulatif	2014-05-12-15.40.08.75715			
PLAN CACHE CREATION TIME	2014-04-28-12.36.55.753363			
Récapitulatif des requêtes actives				
Nombre de requêtes actives en cours	1872			
Nombre de requêtes exécutées depuis le début	844744351			
Nombre d'ouvertures complètes de requête depuis le début	1398363			
Récapitulatif sur l'utilisation de plans				
Nombre de plans en mémoire cache	8906			
Nombre total de plans conçus depuis le début	115793			
TOTAL NUMBER OF UNIQUE QUERIES SINCE START	58621			
Taille de la mémoire cache de plans en cours	948 Mo			
Seuil de taille de la mémoire cache de plans	*AUTO			
MAXIMUM PLAN CACHE SIZE Allowed FOR AUTOSIZING	*DEFAULT (18432) Mo			
CURRENT PLAN CACHE HIT RATIO	92 PERCENT			
TARGET PLAN CACHE AUTOSIZE HIT RATIO	*DEFAULT (90) PERCENT			
CURRENT NUMBER OF JOB SCOPED (QTEMP) PLANS	485			
TOTAL NUMBER OF JOB SCOPED (QTEMP) PLANS BUILT SINCE START	4022			
TOTAL NUMBER OF UNIQUE QUERIES WITH JOB SCOPED (QTEMP) REFERENCES SINCE START	307			
Nombre de plans utilisés à partir de la mémoire cache	1282721			
Nombre total de plans supprimés	67230			
CURRENT Number OF TEMPORARY RUNTIME OBJECTS STORED IN CACHE	8182			
CURRENT TOTAL SIZE OF TEMPORARY RUNTIME OBJECTS STORED IN CACHE	262665 Mo			
MAXIMUM NUMBER OF TEMPORARY RUNTIME OBJECTS STORED PER PLAN	*DEFAULT (5)			
PLAN CACHE CONFIGURATION				
🗽 Seuil de taille de la mémoire cache de plans	*AUTO			
■ MAXIMUM PLAN CACHE SIZE Allowed FOR AUTOSIZING	*DEFAULT (18432) Mo			
■ TARGET PLAN CACHE AUTOSIZE HIT RATIO	*DEFAULT (90) PERCENT			
■ MAXIMUM NUMBER OF LONGEST RUNS ALLOWED PER PLAN	*DEFAULT (3)			
■ MAXIMUM NUMBER OF TEMPORARY RUNTIME OBJECTS STORED PER PLAN	*DEFAULT (5)			
PLAN CACHE ACTIVITY THRESHOLDS				
♣ ACTIVITY THRESHOLDS START TIME	2014-04-28-12.39.50.513497			
Nombre le plus élevé de requêtes actives à un instant donné	2087 (2014-05-06-18.15.53.567735)			
HIGHEST NUMBER OF PLANS IN CACHE	36956 (2014-05-03-03.00.16.744273)			
HIGHEST NUMBER OF TEMPORARY Runtime OBJECTS STORED IN CACHE	36722 (2014-05-03-02.29.55.428591)			
LARGEST TOTAL SIZE OF TEMPORARY Runtime OBJECTS STORED IN CACHE	342378 (2014-05-07-01.54.47.043775) Mo			



Performance Base de données 1/2

- Eliminer les utilisation du moteur CQE en accès DB2 (en V5, et V6), afin d'utiliser le moteur SQE.
 - Eviter les accès via des Fichier Logiques.
- Remplacer les analyses QUERY et les OPNQRYF (CQE en V5, V6, V7R1) par des requêtes QM (SQE) ou par des vues.
- Activer les instructions de blocage d'enregistrements en lecture ou écriture
 - Commande OVRDBF avec valeur *BUF32KB, *BUF64KB, *BUF128KB, *BUF256KB
 - Procédure stockée OVERRIDE_TABLE()
 - OVRDBF : CALL QSYS2.OVERRIDE_TABLE('CORPDATA', 'EMP', '*BUF256KB');
 - DLTOVR : CALL QSYS2.OVERRIDE_TABLE('CORPDATA', 'EMP', 0);



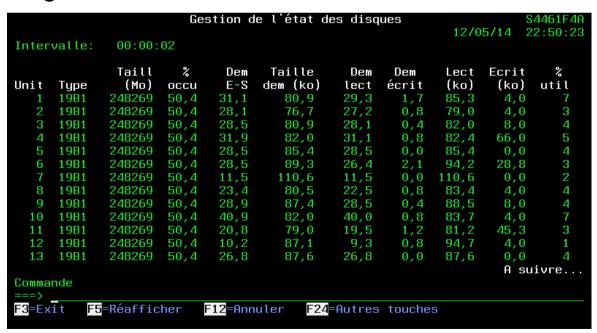
Performance Base de données 2/2

- Créer des tables en lieu et place de PF.
- Si création de logique penser à les passer à des taille de page à 64 K
 - CRTLF ... ACCPTHSIZ(*MAX1TB) PAGESIZE(64)
- Penser à la valeur système QQRYDEGREE
 - *NONE = pas de // IO pour CQE (défaut) // IO pour SQE pas de // CPU pour CQE ni SQE
 - *IO = // IO pour CQE et SQE
 - *OPTIMIZE = // IO ou // CPU pour CQE et SQE (avec DB2 SMP)
- Disposer d'un pool partagé/privé de mémoire pour les travaux serveur BD (QZDASOINIT, QSQSRVR, ...)
- Si l'ajusteur de performance est actif lui fournir des consignes pour ce pool :
 - % taille minimale de mémoire
 - Priorité
 - Si mémoire non contrainte activer l'expert cache pour ce pool



La mémoire secondaire

- Observer les disques Commande WRKDSKSTS.
- Ce panneau fournit l'occupation et l'activité des bras d'accès.
- Un intervalle de 5-10 minutes est significatif en général.
- Regardez la colonne "% util": chaque unité (mécanisme ou bras) devrait être chargée à moins de 60%.





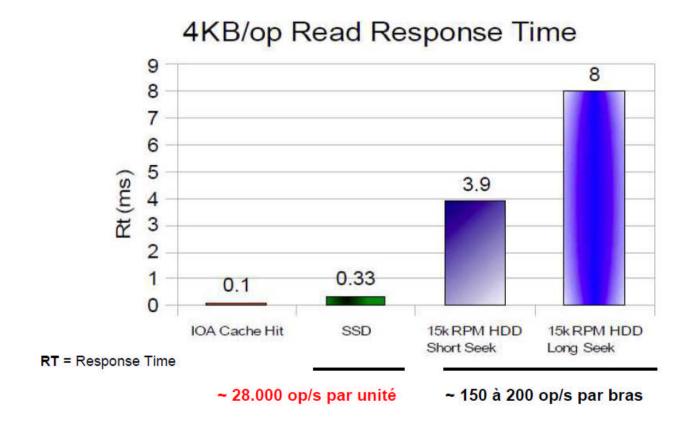
Disques SSD

- La technologie SSD: un coup de pouce pour les serveurs et le stockage !!!
- 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
 - 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



Temps de réponse des disques SSD et HDD en lecture

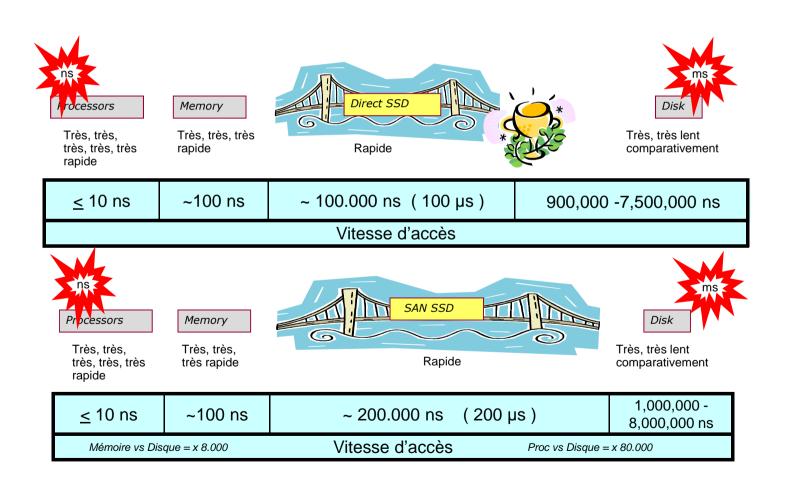
 Le meilleur résultat est obtenu sur des travaux présentant un grand nombre de. lecture aléatoires





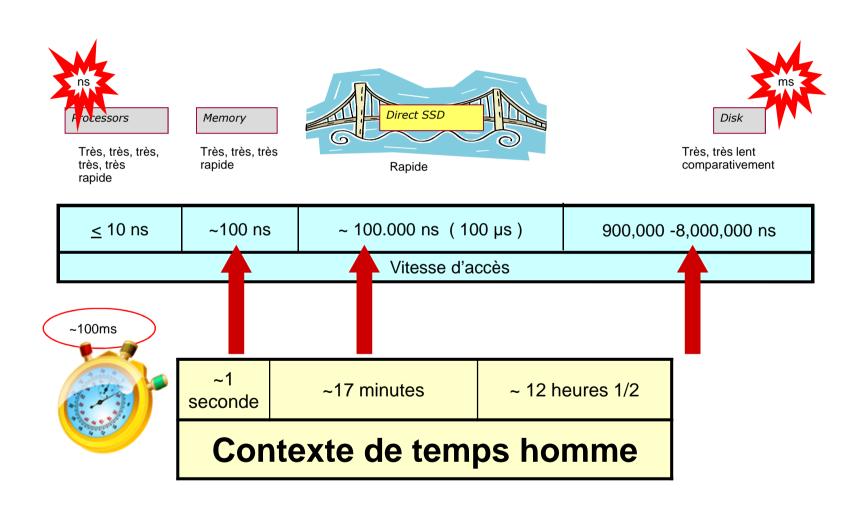
Disque SSD

Performances des SSD en fonction des configurations





Performances des SSD





Apport des disques SSD

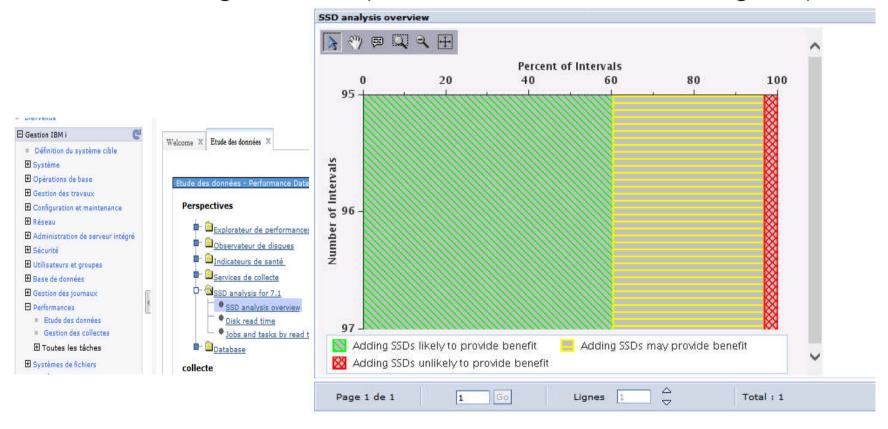
 V5R4, V6 utiliser l'utilitaire ANZSSDTA. Outil disponible sous forme de SAVF sur www.ibm.com/support/techdocs dans "Presentations & Tools". Faire une recherche via Google sur le mot clef ANZSSDDTA.

```
SSD ANALYZER TOOL (ANZSSDDTA)
Indiquez vos choix, puis appuyez sur ENTREE.
PERFORMANCE MEMBER . . . . . . .
                                                  Name, *DEFAULT
                                                  Name
                           Autres paramètres
                                                  *DETAIL, *SUMMARY, *BOTH
                                    *SUMMARY
  START TIME AND DATE::
                                                  Time, *AVAIL
  BEGINNING TIME . . . . . . .
                                    *AVAIL
  BEGINNING DATE . . . . . . .
                                    *BEGIN
                                                  Date, *BEGIN
  END TIME AND DATE::
                                                  Time, *AVAIL
                                    *AVAIL
                                    *END
                                                  Date, *END
DETAIL REPORT SORT ORDER . . . .
                                    *DISKTOT
                                                  *JOBNAME, *CPUTOT...
NUMBER OF RECORDS IN REPORT . .
                                    50
                                                  0 - 9999, *ALL
                                                                             Fin
                      F5=Réafficher
                                       F12=Annuler
                                                     F13=Mode d'emploi invite
F3 = E \times i t
          F4=Invite
F24=Autres touches
```



Apport des disques SSD

V7 IBM Navigator for i (Performance Data Investigator)





Base de donnée et disques SSD

- Possibilité de localiser des données sur disques SSD
 - CRTPF UNIT(*SSD)
 - CHGPF UNIT(*SSD)
 - CRTLF UNIT(*SSD)
 - CHGLF UNIT(*SSD)
 - ALTER TABLE STORE123.EMPLOYEE ALTER UNIT SSD



Equilibrage des ressources

- Mémoire principale : taille des pools, défauts de pages
- Processeur(s) : nombre, utilisation
- Disques : nombre, occupation, utilisation
- Réseau : attachements, débit, utilisation, gestion des erreurs
- Une sur-utilisation de l'une de ces ressources aura un effet négatif sur les performances du système. La disportibilité de chacune de ces ressources est directement liée à sa propre capacité, vitesse et utilisation.

87 © IBM France 2014

Mémoire

principale

Processeurs

Disques

Réseau



Les bonne performances dépendent...

- Du design applicatif
- De la charge d'exploitation
- De la puissance du système
- Du type de processeur
- Et surtout de votre connaissance de l'IBM i



Conclusion

Le potentiomètre du tuning c'est Vous



