

# Dimensionnement des systèmes :

## Quelques règles

*Mai 2000*

**René J. Chevance**

- **Remarques préliminaires :**
  - *Cette présentation est dérivée d'un article récent de Jim Gray et Prashant Senoy [GRA00]. Dans cet article, les auteurs proposent différentes règles empiriques (c'est-à-dire fondées sur l'observation) pour le dimensionnement des systèmes.*
  - *Des règles avaient été formulées par Gene Amdahl il y a 35 ans. Ces règles n'ont pas fondamentalement changé si ce n'est que tous les ratios concernant la mémoire doivent être révisés.*
- **Stockage des données : performance et coût**
- **Règles d'Amdahl concernant l'équilibrage des systèmes**
- **Règle de Gilder sur les réseaux**
- **Quand utiliser les techniques de cache?**
- **Conclusion**

# Stockage des données

- **Loi de Moore : la densité des circuits intégrés double tous les 18 mois. Conséquences :**
  - En dix ans, la capacité de stockage est multipliée par 100
  - Il est nécessaire de disposer d'un bit d'adressage supplémentaire tous les 18 mois
    - 16 bits d'adresse physique en 1970
    - 40 bits en 2000
    - 64 bits en 2020/2030?
- **Évolution des caractéristiques des disques (en 2000) :**
  - Capacité : x 100 entre 1990 et 2000 (70 Go)
  - Densité : de 20 Mbitsi (Mégabits par pouce carré) en 1985 à 35 Gbitsi à fin 1999 (Gigabits par pouce carré)
  - Rotation : x 3 en 15 ans, débit : 30 x en 15 ans, facteur de forme : / 5 en 15 ans
  - Temps d'accès : ~10 ms (100 accès/seconde par disque)
  - Débit : 25 Mo/s
  - Temps d'exploration total du disque : ~45 minutes
  - Prix : ~42 K\$/To

# Stockage des données(2)

## ■ Évolution des caractéristiques des disques [GRA00]

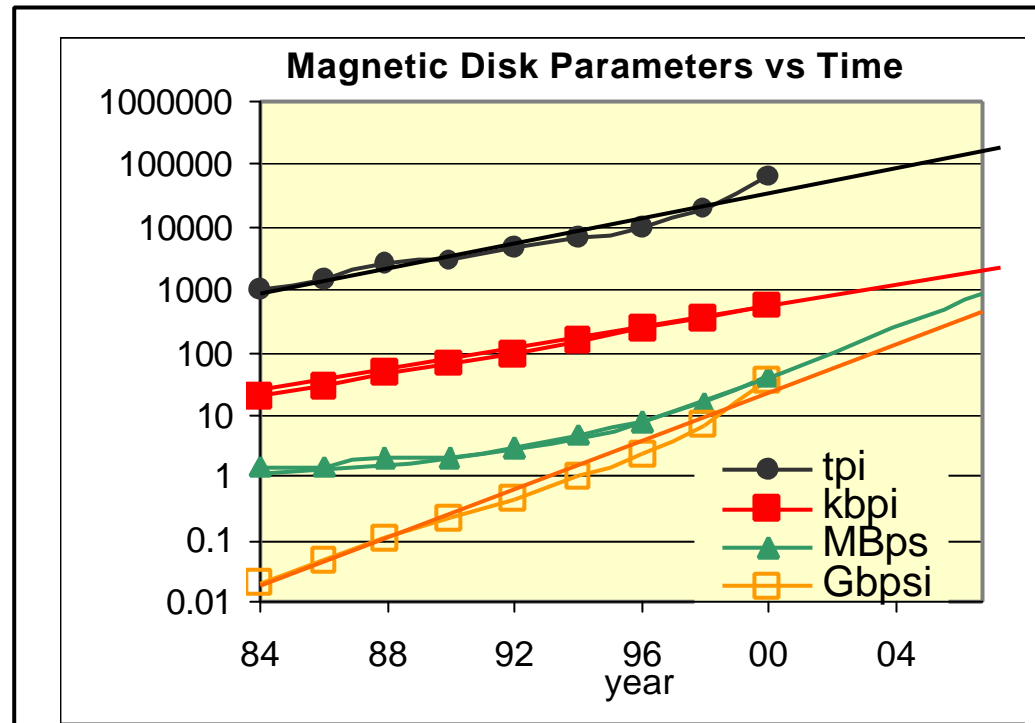


Figure 1: Disk capacity has improved 1,000 fold in the last 15 years, consistent with Moore's law, but the transfer rate MBps has improved only 40x in the same time. The metrics are tracks per inch (tpi), thousands of bits per linear inch of track (kbpi), megabits per second as the media spins (MBps), and

## Stockage des données(3)

- **Évolution des disques :**
  - le ratio entre la capacité des disques et le nombre d'accès par seconde augmente d'un facteur 10 tous les 10 ans
  - le ratio entre la capacité des disques et le débit des disques augmente d'un facteur 10 tous les 10 ans
- **Implication : les accès aux disques deviennent des « denrées » précieuses**
- **Actions pour réduire les accès aux disques**
  - Utilisation intensive des techniques de cache
  - Utiliser des grands transferts séquentiels de données e.g. grands blocs de données (e.g. plusieurs Ko), système de fichiers de type « journal »
  - Utiliser la technique du miroir (RAID 1) plutôt que le RAID 5 pour économiser les accès aux disques

# Stockage des données(4)

- **Évolution des coûts des différents niveaux de stockage bandes, disques et mémoire RAM :**
  - Historiquement, les rapports de prix (par unité de capacité) se sont maintenus à 1/10/1000
  - Avec l'utilisation de robots pour les bandes, les rapports de prix sont maintenant de l'ordre de 1/3/300
  - Sur 10 ans, le prix des RAM est divisé par 100
- **Conséquences :**
  - Déplacement de l'endroit de stockage préférentiel au cours du temps : RAM au lieu du disque (mais problème de permanence du stockage), disque au lieu de bande.
  - Utilisation de la technique du miroir (éventuellement à distance) au lieu des bandes
  - Le coût de l'administration du stockage tend à excéder le coût du matériel (même phénomène qu'avec les PC). En 1980, on considérait comme raisonnable d'avoir une personne pour administrer 1 Go de données (1 M\$ de prix d'achat). En 2000, il conviendrait d'avoir un administrateur pour 30 To de données (1 M\$ de prix d'achat).

# Règles du stockage

- **Loi de Moore : la technologie produit une amélioration par un facteur 4 tous les 3 ans**
- **Nécessité d'un bit supplémentaire d'adressage tous les 18 mois**
- **En 10 ans :**
  - **Les capacités de stockage s'améliorent d'un facteur 100**
  - **Le débit du stockage s'améliore d'un facteur 10**
  - **La fréquence d'accès aux données sur disque diminue d'un rapport 10**
  - **La taille des blocs de données sur disque est multipliée par 5**
- **Le coût du stockage sur bande (robot), disque et RAM est dans un rapport 1/3/300**
- **Dans 10 ans, le stockage RAM coûtera ce que coûte le stockage sur disque aujourd'hui**
- **En termes économiques, il est possible de confier à une personne l'administration d'un stockage de 1 M\$ (soit 30 To aujourd'hui).**

# Règles d'équilibrage des systèmes

- **Règles dérivées de celles de Gene Amdahl (formulées il y a 35 ans)**
  - **Parallélisme** : Si un traitement a une partie séquentielle  $S$  et une partie susceptible d'être parallélisée  $P$ , l'accélération maximale est  $(S+P)/S$
  - **Équilibre** : Un système nécessite 8 mips (million d'instructions par seconde) pour chaque Mo/s d'entrée-sortie, mais les débits en termes d'instructions et d'entrées-sorties doivent être mesurés sur les charges de travail appropriées : CPI (Clock Per Instruction) = 2.1 pour un profil aléatoire (TPC-C) et CPI = 1.2 pour un profil séquentiel (TPC-H)
  - **Mémoire/Traitement** : le nombre de Mo de mémoire par mips (appelé  $a$ ) est passé de 1 à 4 (4 Mo par mips), cette tendance devrait continuer
  - **Entrées-sorties** : un programme demande une entrée-sortie toutes les 50 000 instructions pour un profil « aléatoire » et de l'ordre de 200 000 pour un profil séquentiel

## Règles d'équilibrage des systèmes(2)

### ■ Chiffres dérivés de l'analyse de bancs d'essai TPC [GRA00]

**Table 2:** Amdahl's balanced system law and the parameters of two recent TPC benchmarks ([www.tpc.org](http://www.tpc.org)). The CPI varies among the workloads, and the IO sizes also vary, still, the instructions/byte are similar to Amdahl's prediction of eight instructions per byte (a bit of IO per instruction).

	MHz/ cpu	CPI	mips	KB/ IO	IO/s/ disk	Disks	Disks/ cpu	MB/s/ cpu	Ins/ IO Byte
Amdahl	1	1	1	6					<b>8</b>
TPC-C = random	550	2.1	262	8	100	397	50	40	<b>7</b>
TPC-H = sequential	550	1.2	458	64	100	176	22	141	<b>3</b>

## ■ Loi de Gilder (1995)

- La bande passante installée triple tous les ans
- Le débit des liaisons quadruple tous les 3 ans

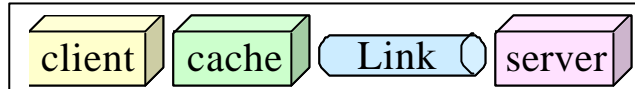
## ■ Entrées-Sorties

- Un message réseau coûte 10 000 instructions et 10 instructions par octet
- Une entrée-sortie disque coûte 5 000 instructions et 0,1 instruction par octet
- Le coût de traitement (processeur) associé à un message SAN (Storage Area Network) est de 3 000 tops d'horloge et 1 top par octet.

# Quand utiliser les techniques de cache?

- **Analyse de retour sur investissement fondée sur le coût du stockage, le prix de revient de l'heure de travail humain**
  - **Règle des 5 minutes pour l'aléatoire : cacher les pages pages du disque accédées de façon aléatoire et qui sont ré-utilisées dans un intervalle de 5 minutes**
  - **Règle de la minute pour le séquentiel : cacher les pages de disque accédées de façon séquentielle et qui sont ré-utilisées dans la minute**
  - **Règle pour le web : cacher les pages web si il y a une chance pour qu'elles soient accédées de nouveau durant leur durée de vie**

# Quand utiliser les techniques de cache?(2)



**Figure 2.** The client-side or proxy web cache improves response time by eliminate link transmission times and server times.

**Table 3:** Shows the benefits of browser and proxy caching (pennies saved) assuming people's time is worth 20\$/hr.

connection	cache	R_remote seconds	R_local seconds	H hit rate	People Savings ¢/page
LAN	proxy	3	0.3	.4	0.6
LAN	browser	3	0.1	.2	0.3
Modem	proxy	5	2	.4	0.7
Modem	browser	5	0.1	.2	0.5
Mobile	proxy	13	10	.4	0.7
Mobile	browser	13	0.1	.2	1.4

**Table 4:** Caching is a very good deal: cache web pages if they will be re-used within the few years.

	A \$/10 KB download network cost	B \$/10 KB storage/mo	Time =A/B Break-even cache storage time	C People Cost Of download \$ (table 3)	Time= (A+C)/B Break Even
Internet/LAN	1e-4	8.E-06	13 months	0.02	184 years
Modem	2E-4	8.E-06	27 months	0.03	307 years
Wireless	1E-2	2.E-04	44 months	0.07	30 years

# Article et sites de référence

**[GRA00] Jim Gray, Prashant Shenoy « Rules of Thumb in Data Engineering »  
IEEE International Conference on Data Engineering  
San Diego April 2000**

**Sites de référence pour la performance des systèmes :**

- **System Performance Evaluation Corporation**
  - <http://www.spec.org>
- **Transaction Processing Council**
  - <http://www.tpc.org>