



Environnement
Canada

Environment
Canada

Canada

Le retour du “Super” EC (Super)ordinateur

EC/CIOB/ITID/NOCD

Carol Hopkins/Luc Corbeil
4 décembre 2007



Environment
Canada

Environnement
Canada

Canada

Menu du jour

- Historique
- Motivations/raisonnement
- Mise-à-jour
- Autres systèmes clés

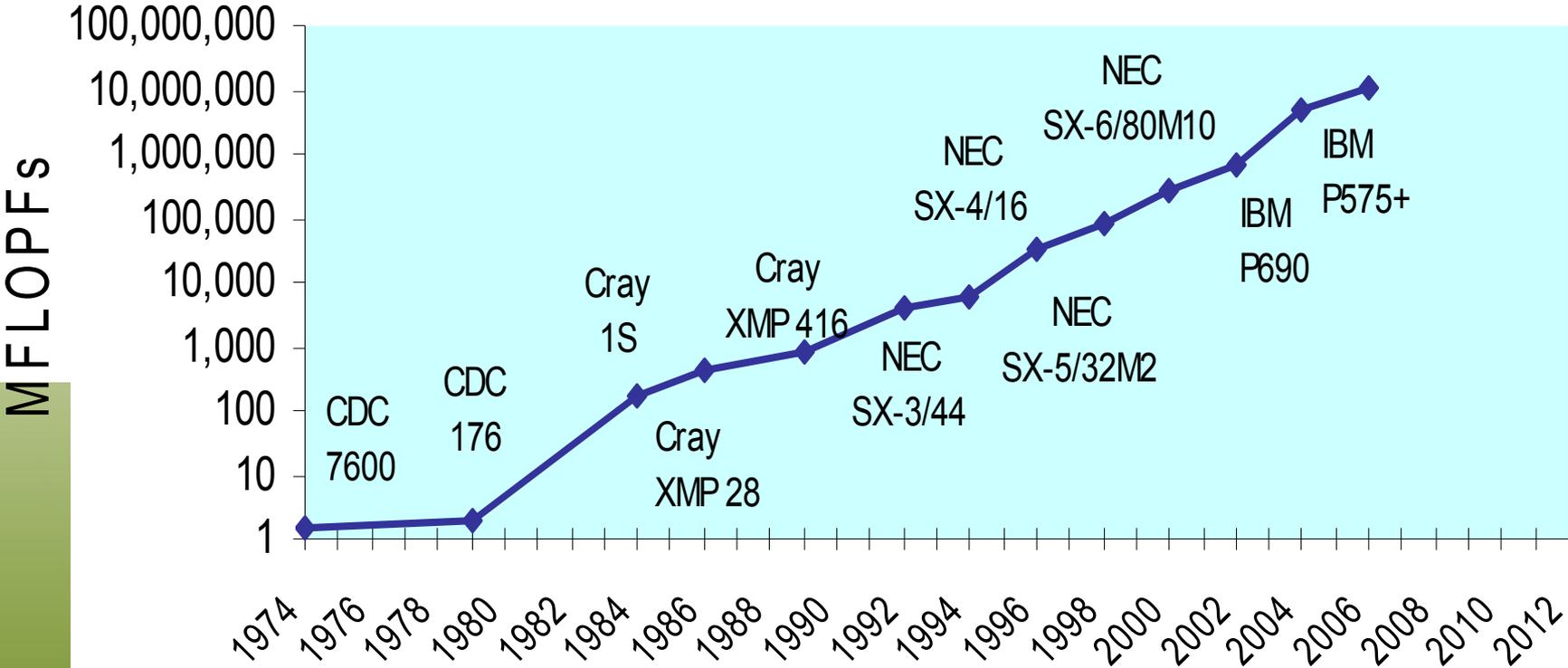


Historique

- Le contrat IBM débuta en décembre 2003
 - Azur, p690, 960 processeurs, Power4 1.3GHz
- Retard de 6 mois pour la première mise-à-jour, acceptée en décembre 2006
 - Maia/Naos, p575+, 1424 coeurs, Power5+ 2 coeurs 1.9GHz
- Seconde mise-à-jour optionelle
 - 30 mois après la première mise-à-jour, donc juin 2009
 - Irait jusqu'en décembre 2011
 - Si déclinée, appel d'offre!
- Appel d'offre requis pour 2012



Historique



Motivations/raisonnement

- Début des discussions avec IBM au printemps dernier pour la seconde mise-à-jour
- Besoins exprimés par EC:
 - Plus de puissance de calcul
 - Plus tôt que ce que le contrat prévoit
 - Pas d'interruption
 - Pas de financement supplémentaire
 - Infrastructure électrique limitée



Motivations/raisonnement (2)

- IBM suggère de rester au Power5+
 - Le Power6 est très différent
 - Refroidi à l'eau
 - Architecture du processeur
 - Performance des codes à revoir
 - Effort de conversion nil pour les usagers
 - Procédure d'installation connue et documentée
 - Pour les administrateurs systèmes, ajout de composantes connues
 - Prix du p575+ en décroissance, donc \$/flop réduit
 - Rencontre toutes les exigences contractuelles



Power6: refroidi à l'eau

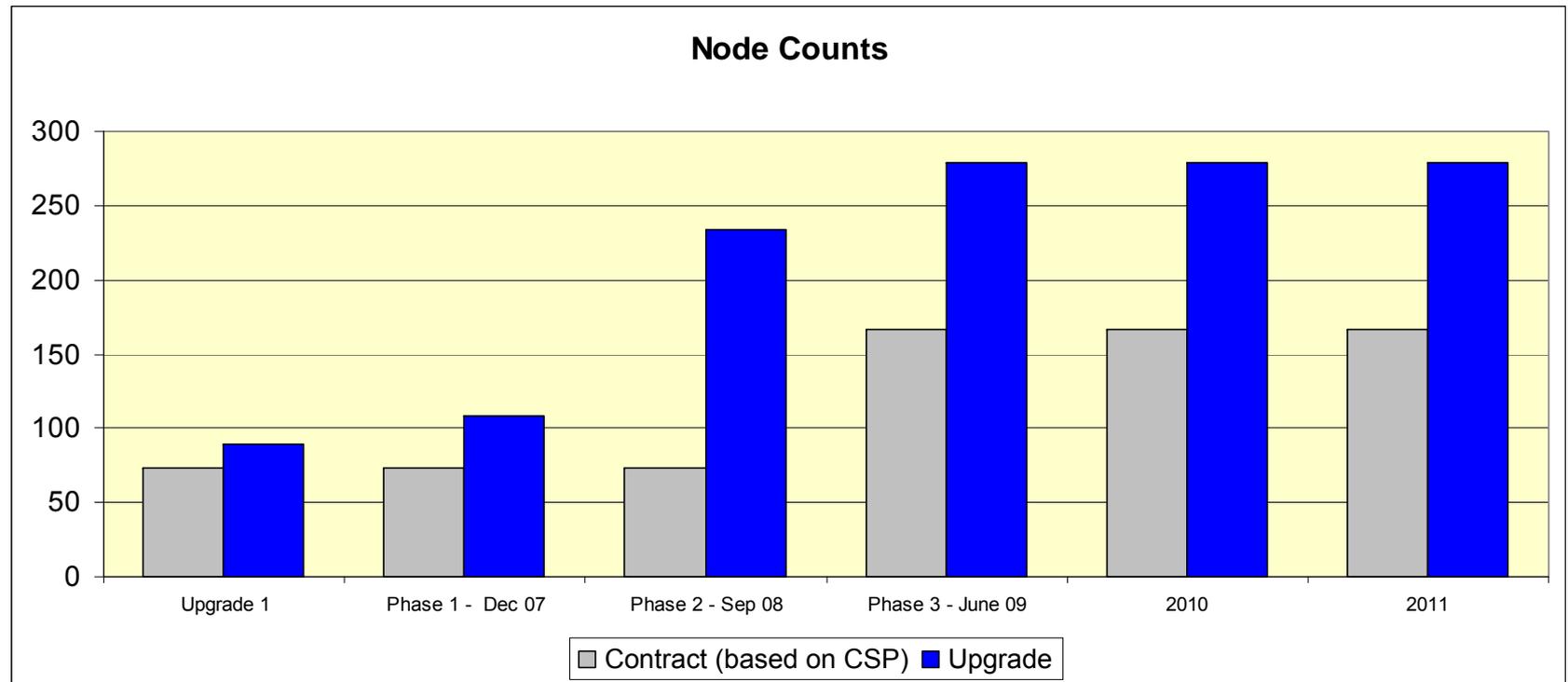


Motivations/raisonnement (3)

- Pour
 - Éviter les coûts énormes d'un appel d'offre (temps et \$\$)
 - Plus de puissance plus tôt se réalise
 - Migration très simple pour les usagers (et plus simple pour nous!)
- Contre
 - Même technologie pour 5 ans
 - Bris matériel potentiellement à la hausse
 - IBM s'engage à rencontrer 99% de disponibilité via le contrat
 - Beaucoup de redondance
 - Application sérielles ou quasi-sérielles ne verront pas de gain

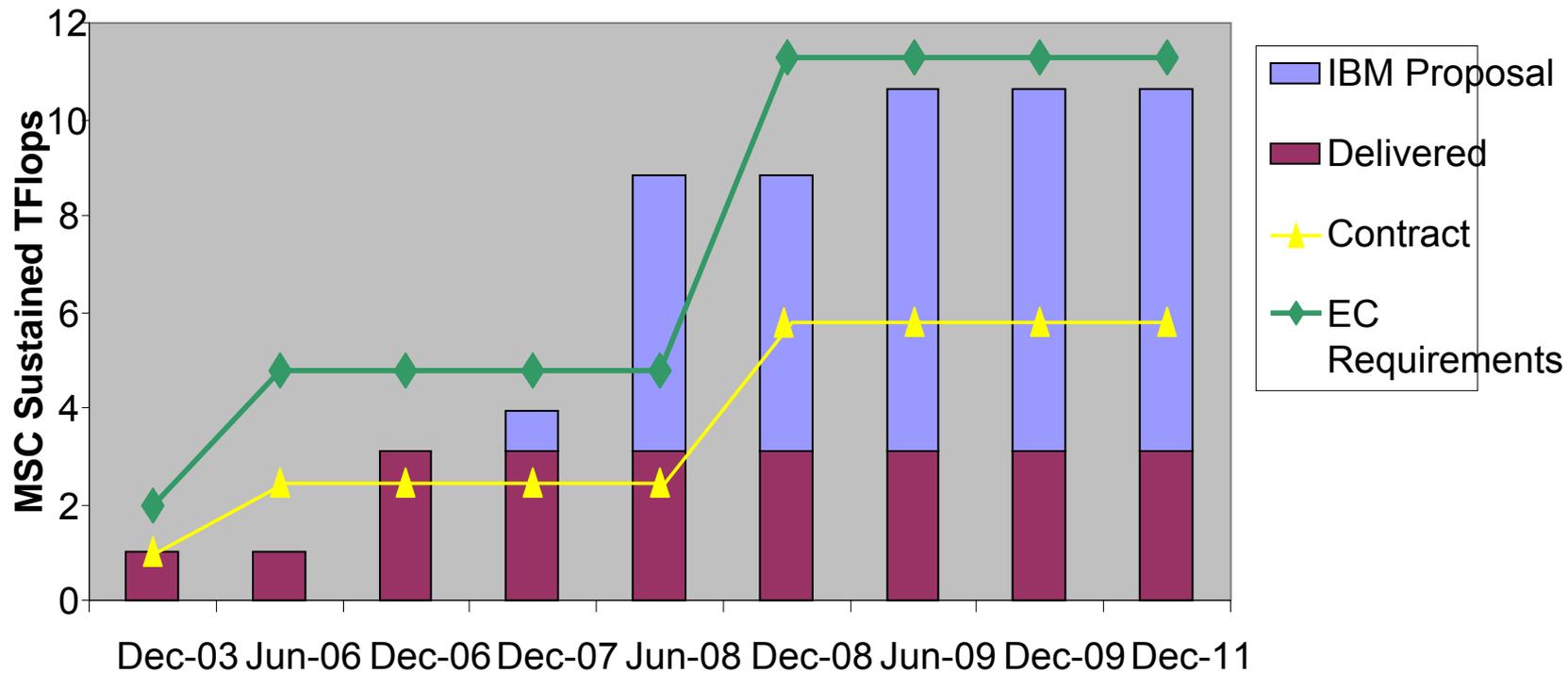


Mise-à-jour IBM



Courbe de performance

Supercomputing Performance



Nœuds par phase

Phase	Livraison	Nœuds additionnels	Total	Minimum selon le contrat
Upgrade 1 – Phase 1 et 2 (Maia / Naos)			91	73
Upgrade 1 - Phase 3	Dec 2007	19	110	73
Upgrade 1 - Phase 4	Sept 2008	127	237	73
Upgrade 2 - Phase 1	June 2009	45	282	167

Installation: le casse-tête

Date	C4 (maia)	C5 (naos)	C6 (TBD) 128 ports	C7 (TBD) 256 ports
Actuel	40	38		
Dec 2007	59	38		
Sept 2008	59	38	121	
Dec 2008	59		80	123
Jan 2009			80	182




Stockage, HPS, réseautage

- Stockage: même technologie, quantité doublée
 - De 36 à 72 TB brut, ~ 54 TB après configuration
 - Rencontre facilement les exigences contractuelle
- HPS: quantité triplée
 - De 2*64 à 128 + 256 ports.
- Réseautage: doublé
 - De 2*6513 à 4*6513 chassis Cisco
 - Ports 10 GigE supplémentaires pour autres équipements

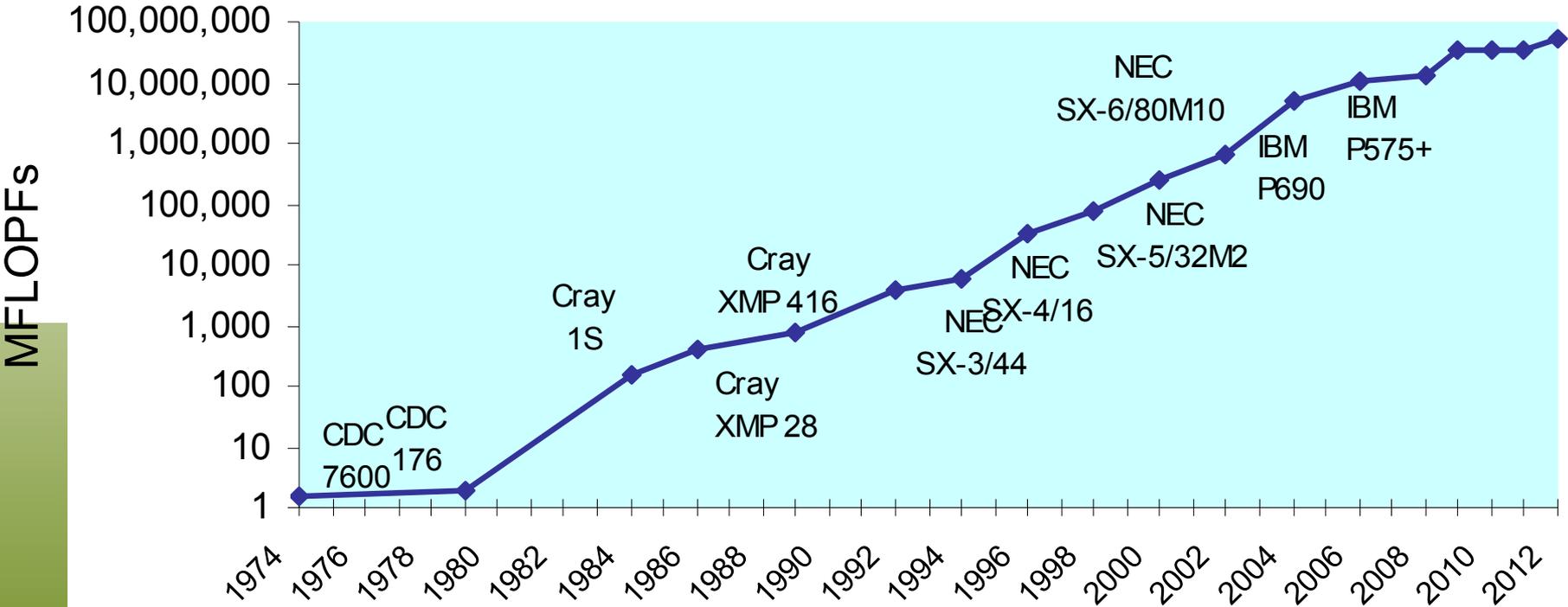


Électricité, refroidissement

- Tout est fonction de l'électricité!
 - 19 noeuds: en attente de prises, dû mi-décembre
 - 121 noeuds: idem, question de semaines
 - Certaines composantes sur Hydro
 - noeuds de calcul côté développement
- Refroidissement
 - Plomberie requise pour Power6 évitée
 - Lieberts additionnels seront installés
 - Pas seulement pour le superordinateur



Futur



Mise-à-jour IBM

- Pourquoi IBM “donne” 67% plus que le contrat?
 - Marketing: pour garder EC comme client
 - Marketing (2): reprendre le #1 sur le top500 au Canada
 - Avec les 19 noeuds supplémentaires: #400 (Nov. 2007)
 - Université de Sherbrooke: #391
 - Éviter les coûts et les efforts d’un appel d’offre
 - Assurer un revenu garanti jusqu’en 2011
 - Accéder occasionnellement un système p575+ au printemps



IBM: autres possibilités

- Grossir les grappes
 - 100 noeuds peuvent être “facilement” rajoutés
- Discussions afin de tester de nouvelles technologies
 - Blades Power6 (modèle non-HPC)
 - Blades “Cell” (processeur du Playstation)
- Partenariat avec d’autres clients similaires d’IBM



Ce qu'on vise

- Transition douce et indolore
 - Pas de migration de fichiers, espace grossira magiquement
 - Pas d'interruption significative
 - OS (AIX 6.X) et compilateurs (xlf 11.X) récents
- Donner les ressources dès que possible
- Meilleure surveillance du système, formation
 - Besoin de plus de bras!

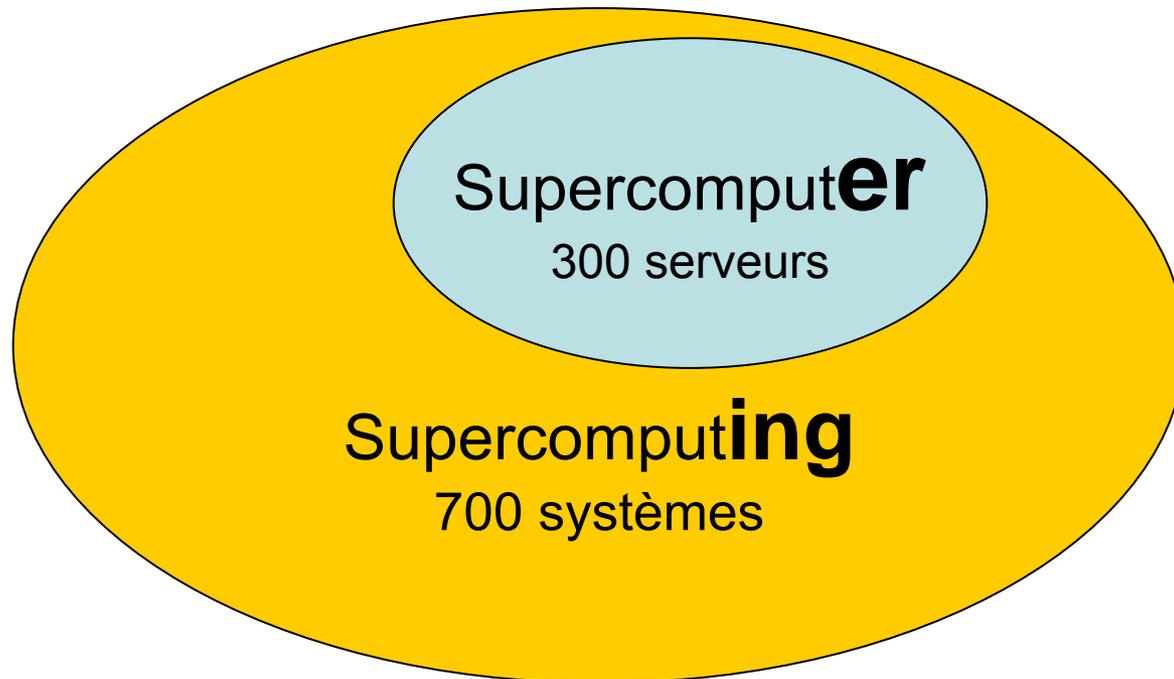


Ce qu'on attend de vous

- **Rétroaction**
 - Bonne et mauvaise
 - *DDS pour les problèmes*
 - *it_infrastructure_client_feedback@ec.gc.ca* pour commentaires
- **Patience**
 - Les opérations d'abord!
 - Pas d'électricité, peut pas livrer!
- **Volonté de tester OS et compilateurs récents**
 - Idéalement, pourrait être automatisé
- **Réflexion sur le futur**
 - Travaux MPI de $O(10k)$ ou $O(100k)$ processeurs en 2012!



Infrastructure des TI



Frontaux

- SGI O3900 (castor et pollux)
 - Pollux: 40 processeurs 600 MHz, 40 GB RAM
 - Castor: 32 processeurs 1 GHz, 32 GB RAM
 - Plus de 40 TB d'espace disque de haute-performance
 - TP9500, 2GBit
 - IS4500, 4GBit

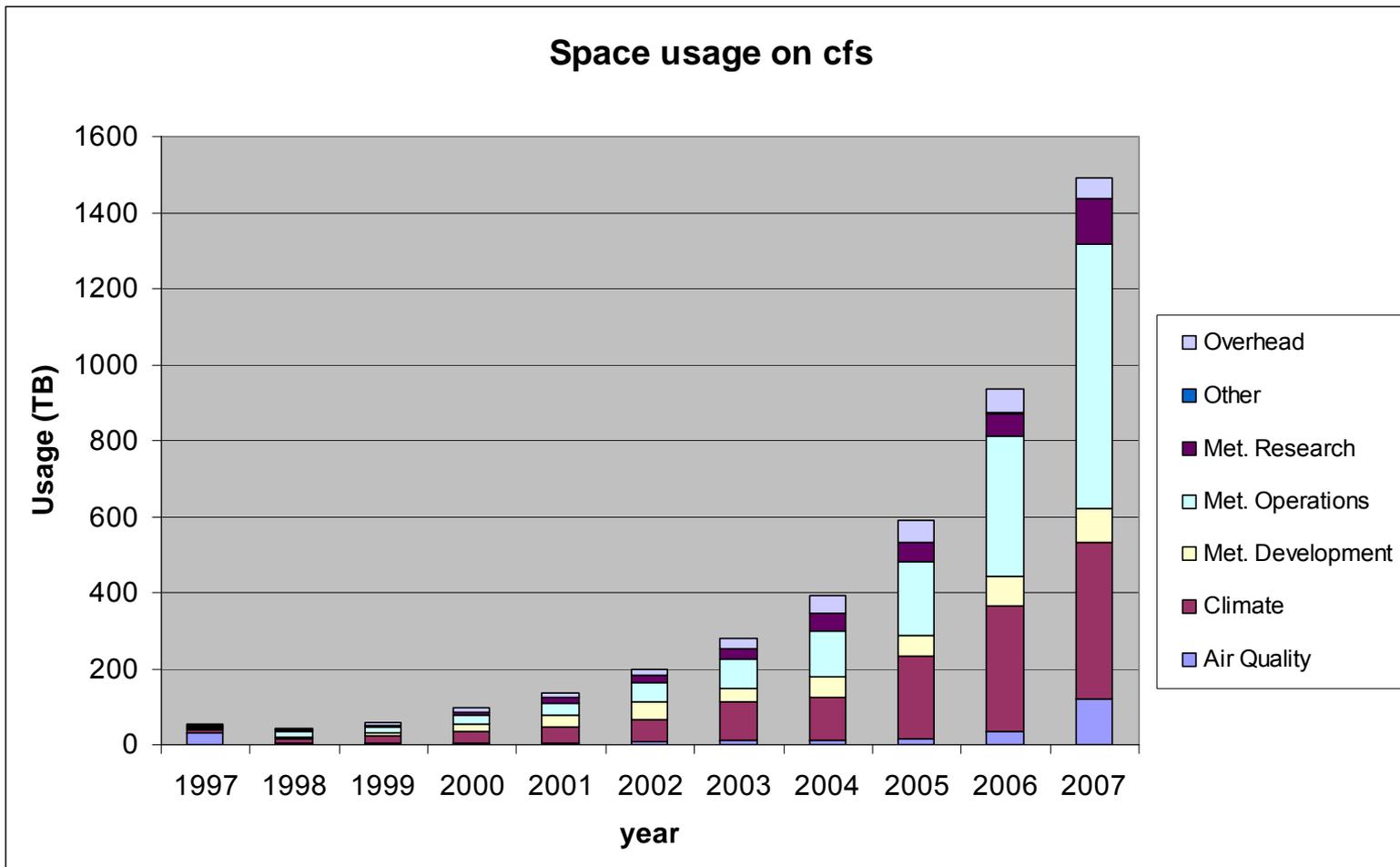


CFS: Archivage





Croissance sur le CFS



Nouveau logiciel HSM

- Provenait de ADIC/Quantum: Fileserv/Volserv
 - Support terminé par Quantum, Sept. 2007
- Maintenant Quantum StorNext
 - Basé sur Fileserv
 - Lecteurs et rubans LTO-3 (2X plus gros/plus vite)
- Migration
 - Nouveau serveur CFS: SGI Altix 350
 - Les fichiers doivent migrer de Fileserv/oldcfs à StorNext/cfs
 - Terminé à 80%



Dorval-ib

- 40 noeuds, processeurs dual Xeon
- Séparé en deux, 20 nœuds pour les opérations
- Opérationnel!
- Remplacera gfx
- Infiniband switch, 4X, 900 MB/s
- Stockage: Rackable RapidScale
 - Actuellement 4TB, 40 TB dès que nous aurons des bras libres



Prochaines étapes

- Frontaux et CFS
 - Contrats se terminent au printemps 2009 (une année d'option)
 - Projet de remplacement débutera en janvier
 - Idéalement, serait installé pour la fin 2009
- Nouveau UPS: printemps 2010
 - Beau casse-tête de logistique!
- Superordinateur: appel d'offre débutant en 2009, mise en opération automne 2011
- Contrat IBM se termine à Noël 2011



Merci!

- Questions?
- Commentaires: luc.corbeil@ec.gc.ca

