DE LA RECHERCHE À L'INDUSTRIE

Évolutions du Stockage pour les centres HPC de demain

Jacques-Charles Lafoucrière

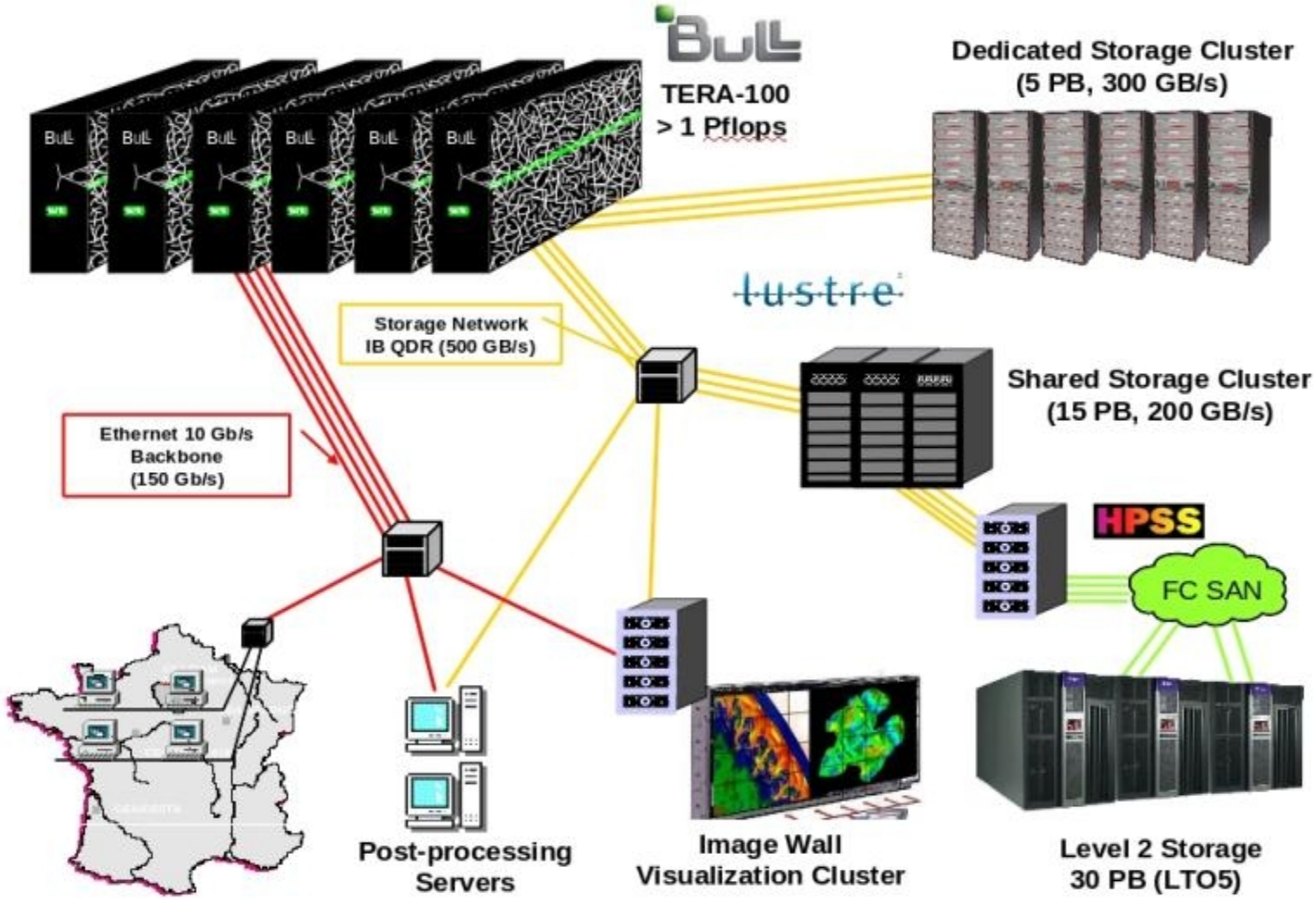
www.cea.fr

Centres de calcul Petaflopique

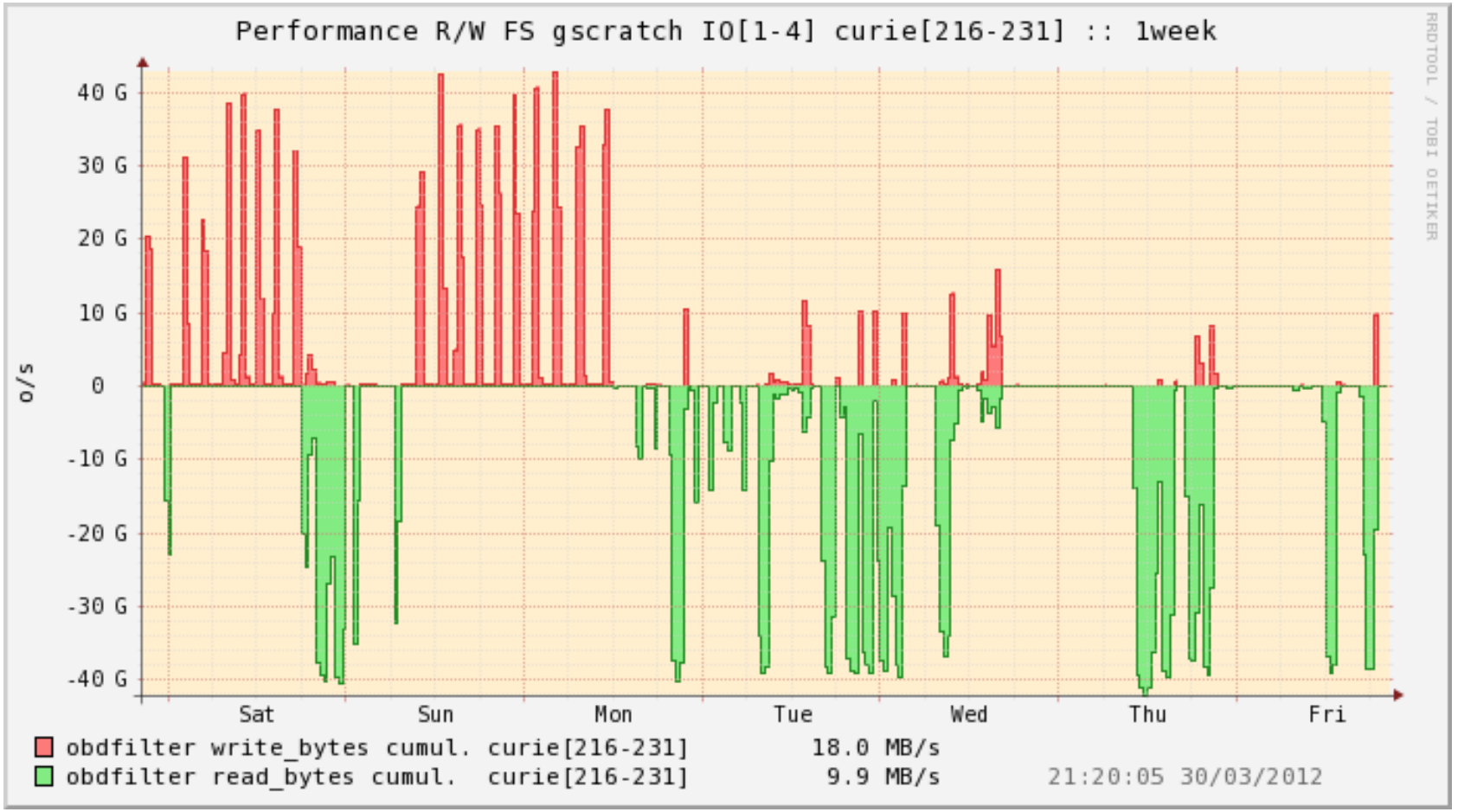
Tera
1.25 Pflop/s
Mem : 290 To
FS : 500 Go/s

TGCC/Curie
2 Pflop/s
Mem : 340 To
FS : 250 Go/s

Architecture Globale



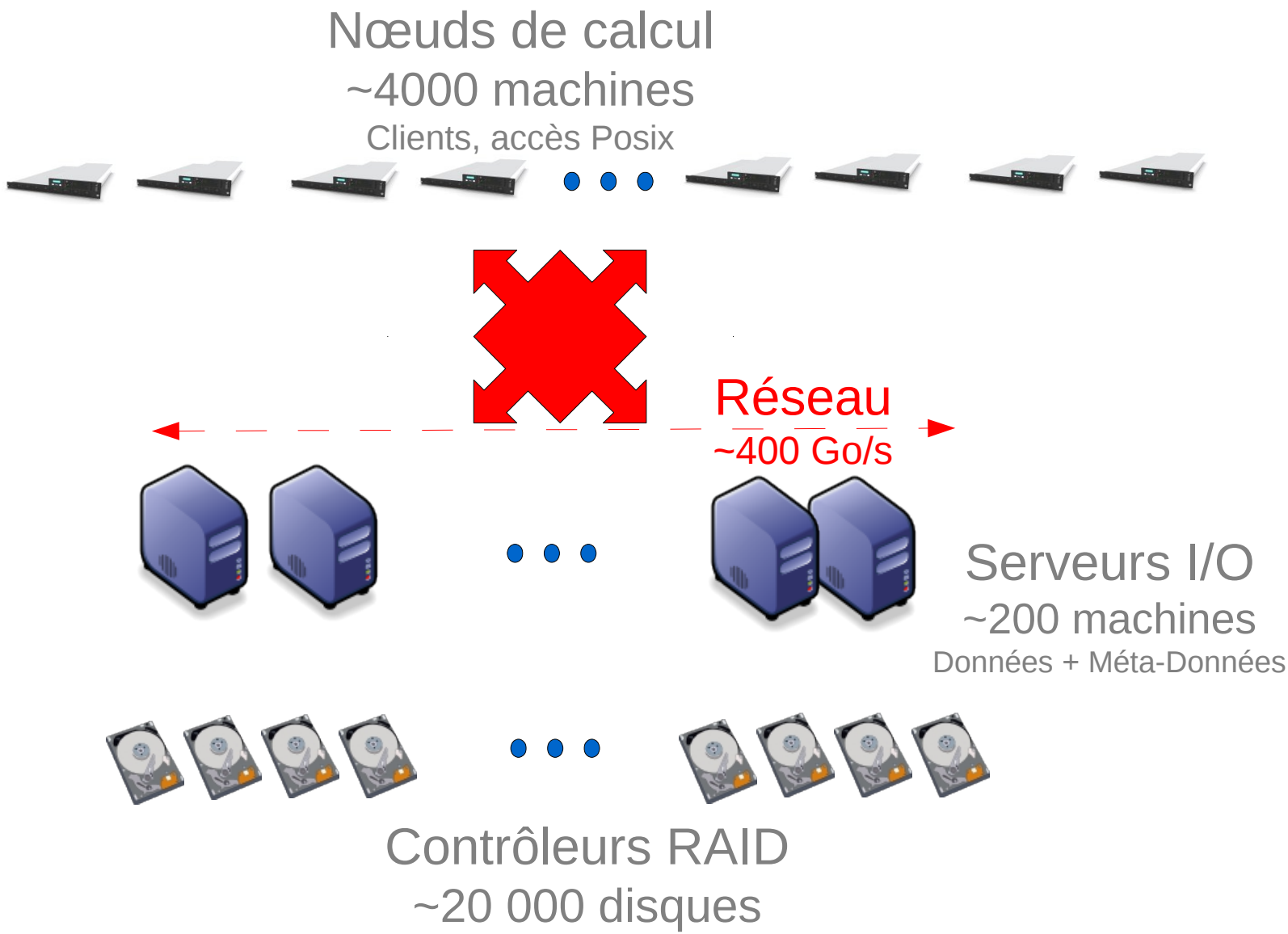
Exemple d'utilisation d'un FS Petaflopique



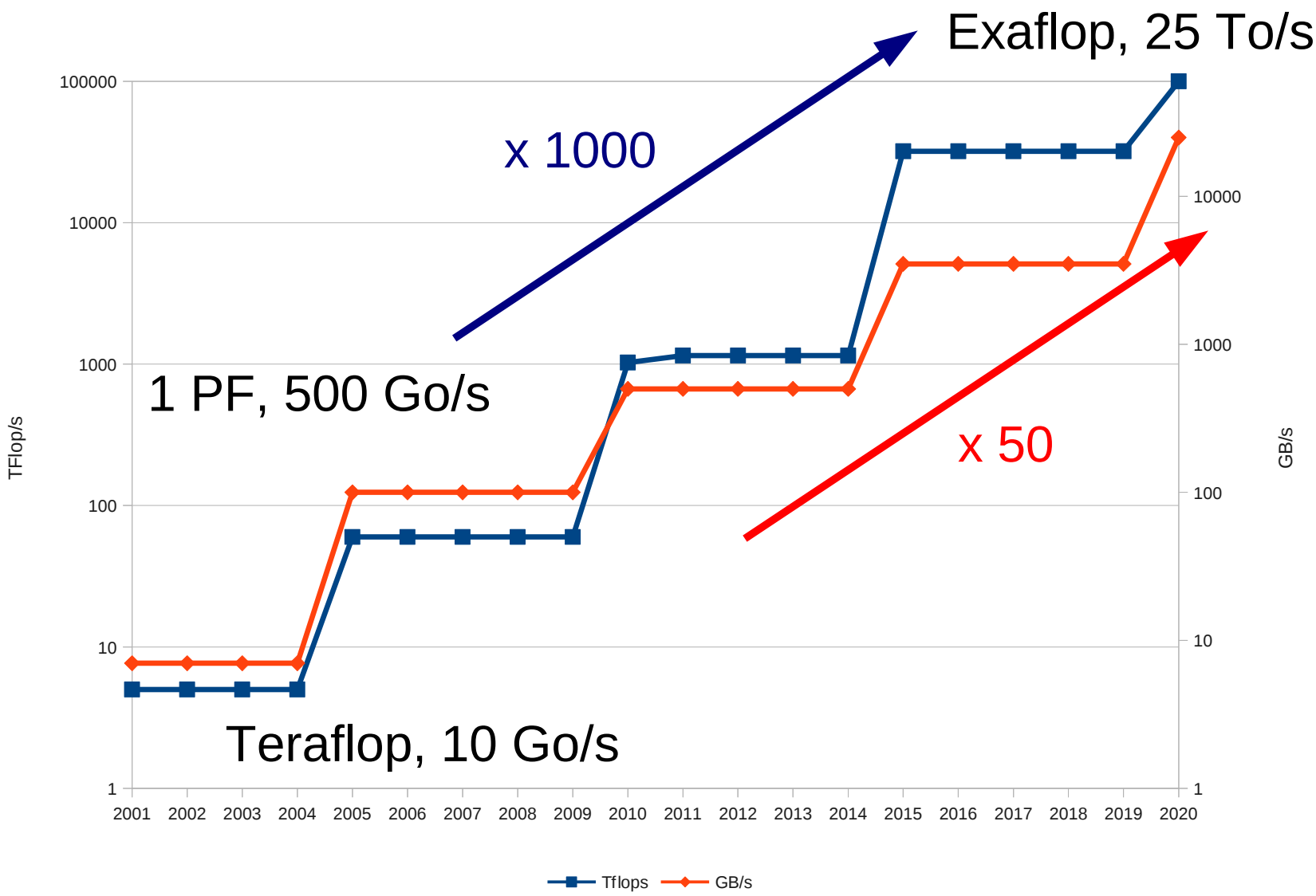
Grand Challenge DEUS

1Po produit en qlq jours

Le Stockage d'un centre de calcul



Évolution d'un centre de calcul



Feuille de route pour satisfaire un besoin Exaflop en 2020

2015

- 30 Pflops
- 3.5 To/s
- 10 000 nœuds
- > 20 000 disques

2020

- ~1 Exaflops
- 25 To/s
- > 50 000 nœuds
- > 30 000 disques

Impossible de garder le même ratio en bande passante IO qu'en puissance de calcul

Amplification de la pénurie mémoire sur les nœuds de calcul

- Due à l'utilisation d'architecture ManyCore pour les nœuds de calcul (ratio Go/Thread va en diminuant)

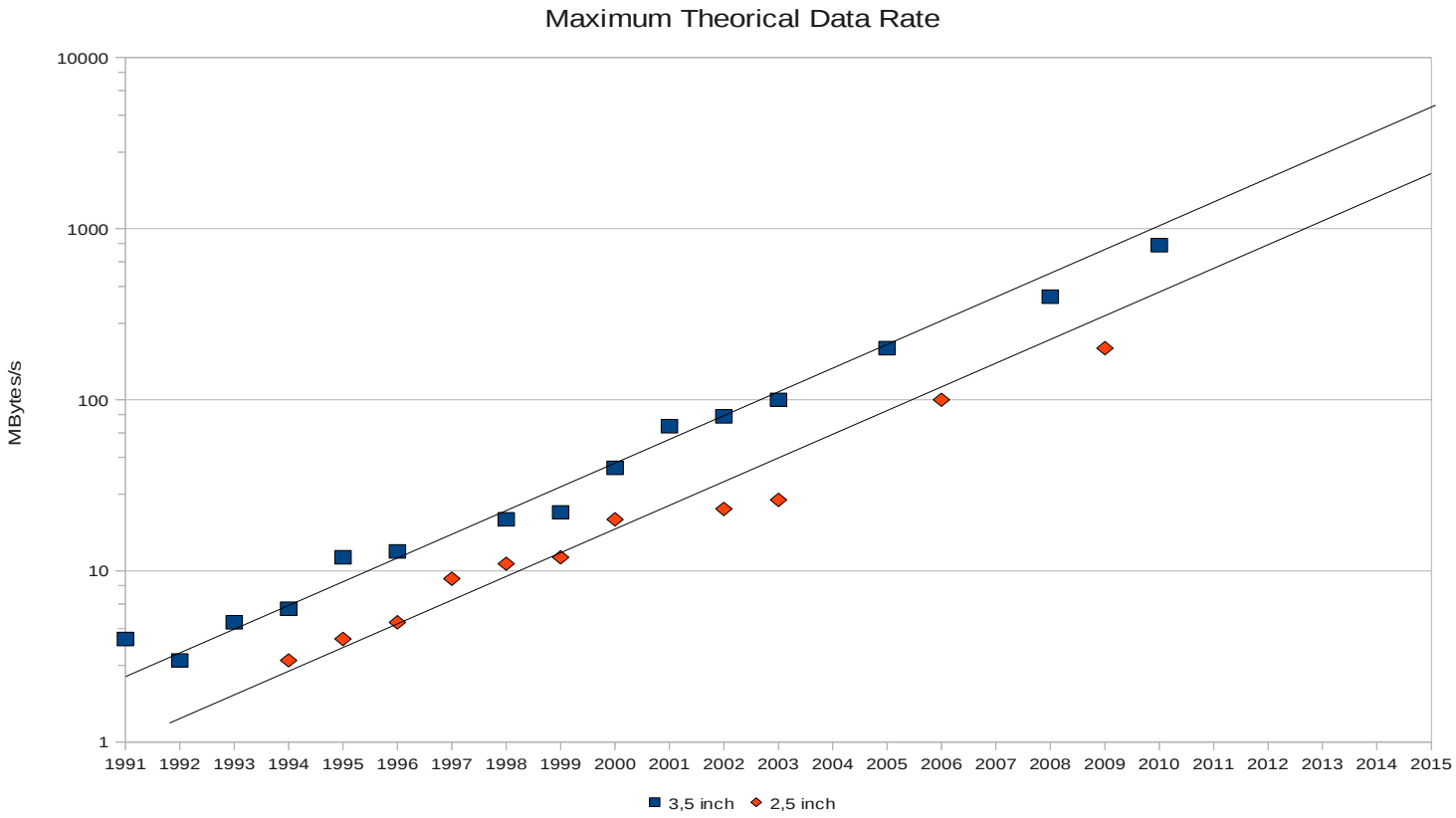
DE LA RECHERCHE À L'INDUSTRIE

cea

www.cea.fr

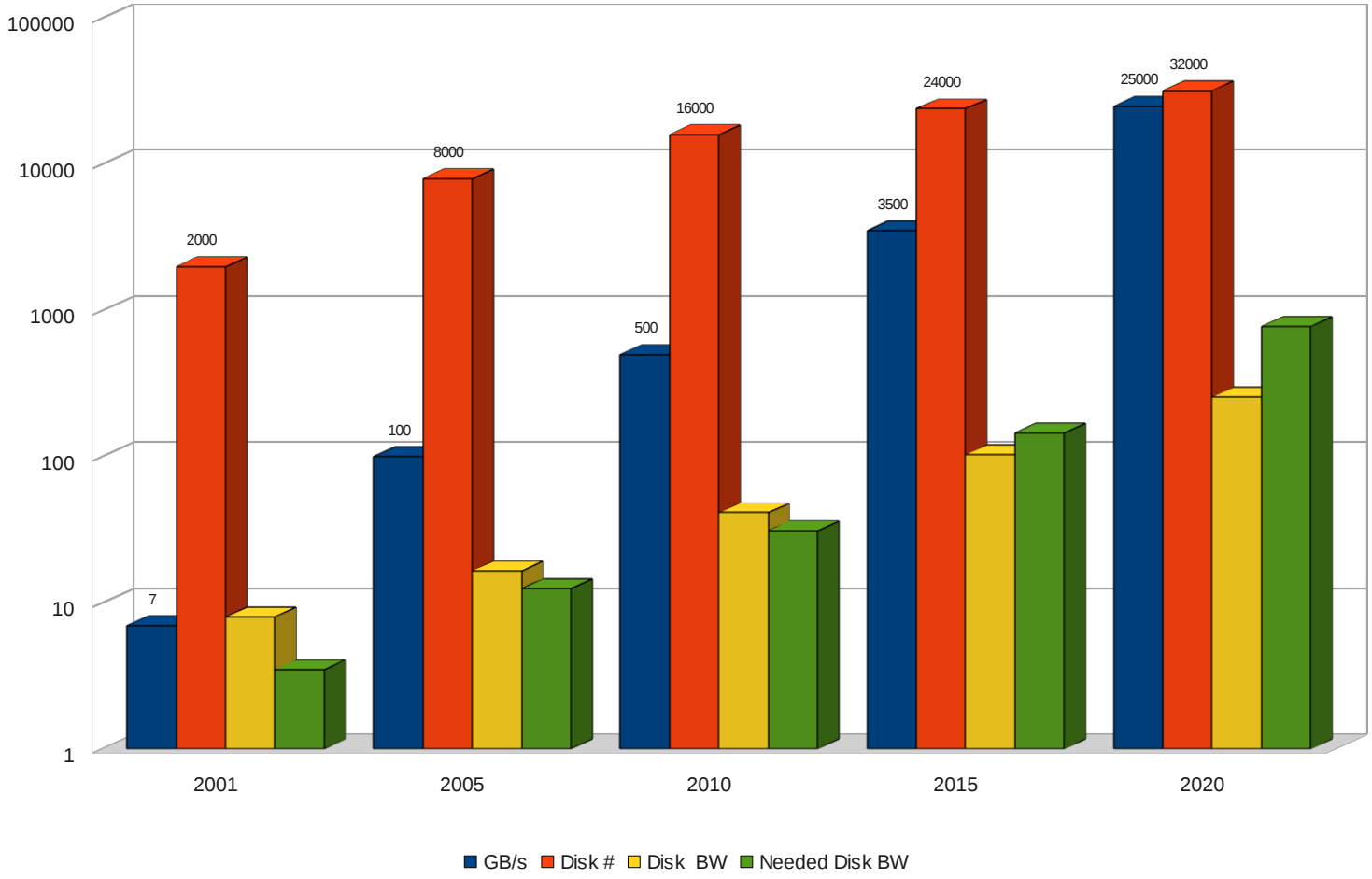
Comment atteindre le débit ?

Bande Passante Maximale d'un disque magnétique



Max Data Rate = F(Densité Linéaire, Vitesse Rotation, Diamètre)

Dimensionnement pour l'exaflop



A partir de 2015

La bande passante d'un système à base disques rotatifs ne suffit plus

- Utilisation de massive de la technologie Flash (500Mo/s en 2015)

Les défis pour les systèmes de fichiers

- Support de technologies hétérogènes
  - Du Flash aux Disques
- Explosion du nombre de clients
  - Scalabilité des meta-données
- Évoluer vers le modèle embarqué (Stockage Actif)
  - Les serveurs des systèmes de fichiers tournent dans les contrôleurs de disques
  - Réduire le coût du aux serveurs
  - Augmenter l'efficacité des serveurs

DE LA RECHERCHE À L'INDUSTRIE

Quelle Architecture pour le Système de Fichiers ?

www.cea.fr

Système Exaflop : les clients

Le nœud de calcul

- Sera massivement multithread
- Aura un ratio mémoire/thread peu favorable

Le client du système de fichier

- Supportera un fort parallélisme au sein d'un nœud
- N'aura pas de mémoire pour des caches locaux

Nécessité d'introduire un mécanisme de délégation des entrées sorties vers les systèmes de fichiers

- Allocation dynamique de serveurs de délégation proches des nœuds de calcul
- Utilisation de copie directe distante

Système Exaflop : les serveurs

L'architecture classique du stockage (accès blocs) ne passe pas à l'échelle

- Besoin d'une nouvelle architecture
- Modèle d'objets réseaux
  - Le serveurs de fichiers devient un serveur d'objets
  - Parité réseau entre serveurs

Scalabilité des Méta données

- Les Méta Données doivent être hébergées par plusieurs serveurs
- La contrainte due à la logique Posix doit être relâchée
  - Plus possible de garantir une cohérence « gratuite » et performante sur toute la machine
  - Aide fournie par les applications et le gestionnaire de ressource
    - Topologies, mode d'accès, ...

Conclusions

De nombreux défis doivent être relevés pour atteindre des systèmes de stockage pour des machines Exaflopiques

- Gestion des données, des Meta-données
- Aide nécessaire de la part des « sachants »
- Nouvelle architecture de serveurs I/O
- Mise en place de délégations d'I/O

Renforce le besoin de solutions OpenSource

- Pour adapter au besoins spécifiques et débugger sur site
- Comme base pour la recherche académique en Europe
  - Besoin fort de personnes compétentes en stockage
  - Besoin de former les utilisateurs

Un FS Exascale ne sera pas « naturel » mais des solutions commencent à voir le jour pour faire évoluer les systèmes de fichier Petascale vers l'Exascale

Merci