



CENTRE DE CALCUL

# **30 ans de stockage sur bandes magnétiques au CC-IN2P3**

Pierre-Emmanuel BRINETTE

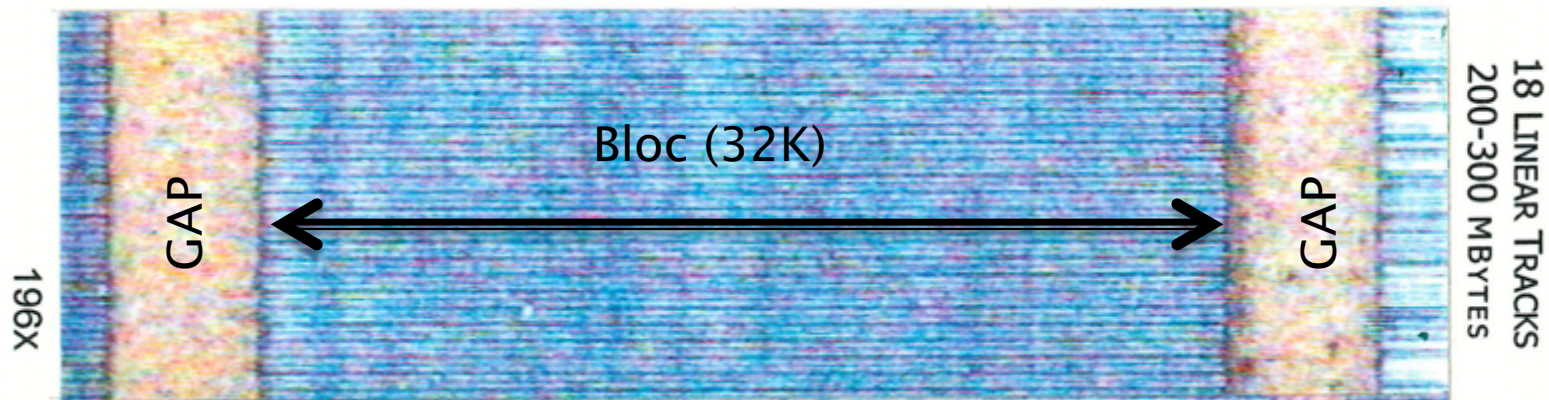
Philippe GAILLARDON

Journées Informatiques – 16 octobre 2014

- ▶ Autopsie d'une bande magnétique
- ▶ Historique
  - Les 80`s et le mainframe
  - Les 90`s et la transition vers Unix
  - Les années 2000 et HPSS
- ▶ Etat des lieux en 2014
- ▶ Tendances à venir

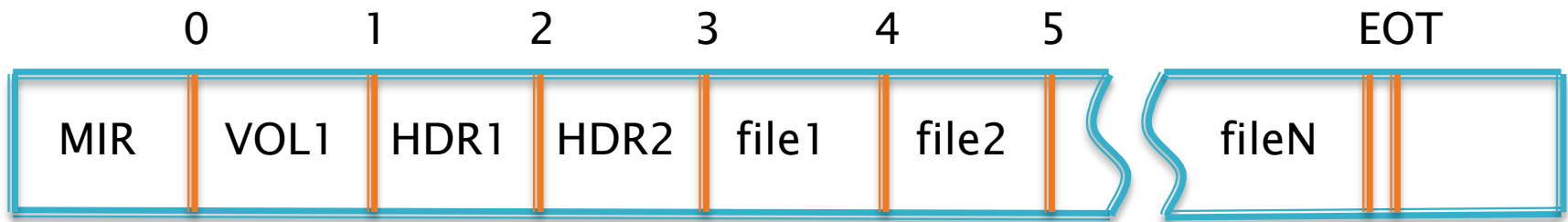
## ▶ Bande magnétique

- Film recouvert d'un substrat magnétique (MP, BeFe)
- Tête de lecture est en contact avec bande
- Information stockée numériquement
  - Codage Manchester, NRZ, ...
- Plusieurs pistes parallèles (9 pistes en 1964)
- Chaque piste est composée de blocs magnétisés entrecoupés d'espaces (GAP)
- Longueur : 100 m à 1200 m
- Capacité = longueur \* nombre de pistes \* densité



# Autopsie d'une bande magnétique

- ▶ Bande formatée (standard IBM)
- ▶ Label sur 6 caractères (VOLSER)
- ▶ Accès séquentiel en mode bloc
- ▶ Les fichiers sont écrits les uns à la suite des autres
  - Suite de blocs séparés par des Tape Mark
  - Fin de bande (EOT) : Double Tape Mark
- ▶ MIR : Media Information Record
  - Zone non accessible par l'application
  - Contient la position des fichiers d'une bande



Norme ANSI X3.27 - 1969

## ▶ Lecture d'un fichier

- Positionnement de la bande en début du 4<sup>ème</sup> fichier
  - `mt -f /dev/nst0 fsf 4`
- Lecture du fichier (READ)
  - `dd if=/dev/nst0 of=monfichier.dat`

## ▶ Ecriture d'un fichier

- Positionnement en fin de bande
  - `mt -f /dev/nst0 eod`
- Ecriture des données (WRITE)
  - `dd if=monfichier.dat of=/dev/nst0 bs=1048576`
  - Taille de bloc 1Mo

## ▶ Fermeture de la bande

- `mt -f /dev/nst0 rewind`
- `mt -f /dev/nst0 offline`

## ▶ Logique de fonctionnement inchangée depuis 60 ans !



- ▶ 1986 : Délocalisation du CCPN à Villeurbanne
  - Création du laboratoire : CC-IN2P3
  - Déménagement de l'ordinateur central
- ▶ Mainframe IBM 3081
  - 30 MIPS
  - 32 Mo de RAM
  - 38 Go de disques
  - 2 imprimantes
  - 1 lecteur de carte perforée
  - 14 lecteurs de bandes





- ▶ 14 dérouleurs IBM 3420
  - Bande ½ pouce
  - 9 pistes parallèles
  - Longueur : 730 m
  - Capacité max : 170 Mo ( à 6250 BPI )



## ► Les bandes sont stockées en salle machine



J.M. HURON

Lyon Figaro 12/06/1987

*Cent millions de francs d'investissement  
pour l'ensemble bâtiment-machine.  
(Un super ordinateur IBM d'une puissance de trente millions  
d'informations par seconde).*



- ▶ Environ 10.000 bandes  $\approx$  1 à 2 To



- ▶ Nécessite une main d'œuvre pour opérer le système
  - Equipe de 10 opérateurs en 3x8



- ▶ IBM 3480
  - Utilisé au CC depuis 1987
- ▶ Cassette au format 4"x5"x1"
  - Amorçage
  - Déroulée dans le lecteur
- ▶ Plus robuste
- ▶ Plus simple à manipuler
- ▶ Capacité 200 Mo
  - 800 Mo en version 3490
- ▶ Débit : 1-3 Mo /s





## ▶ Lecteurs 3480

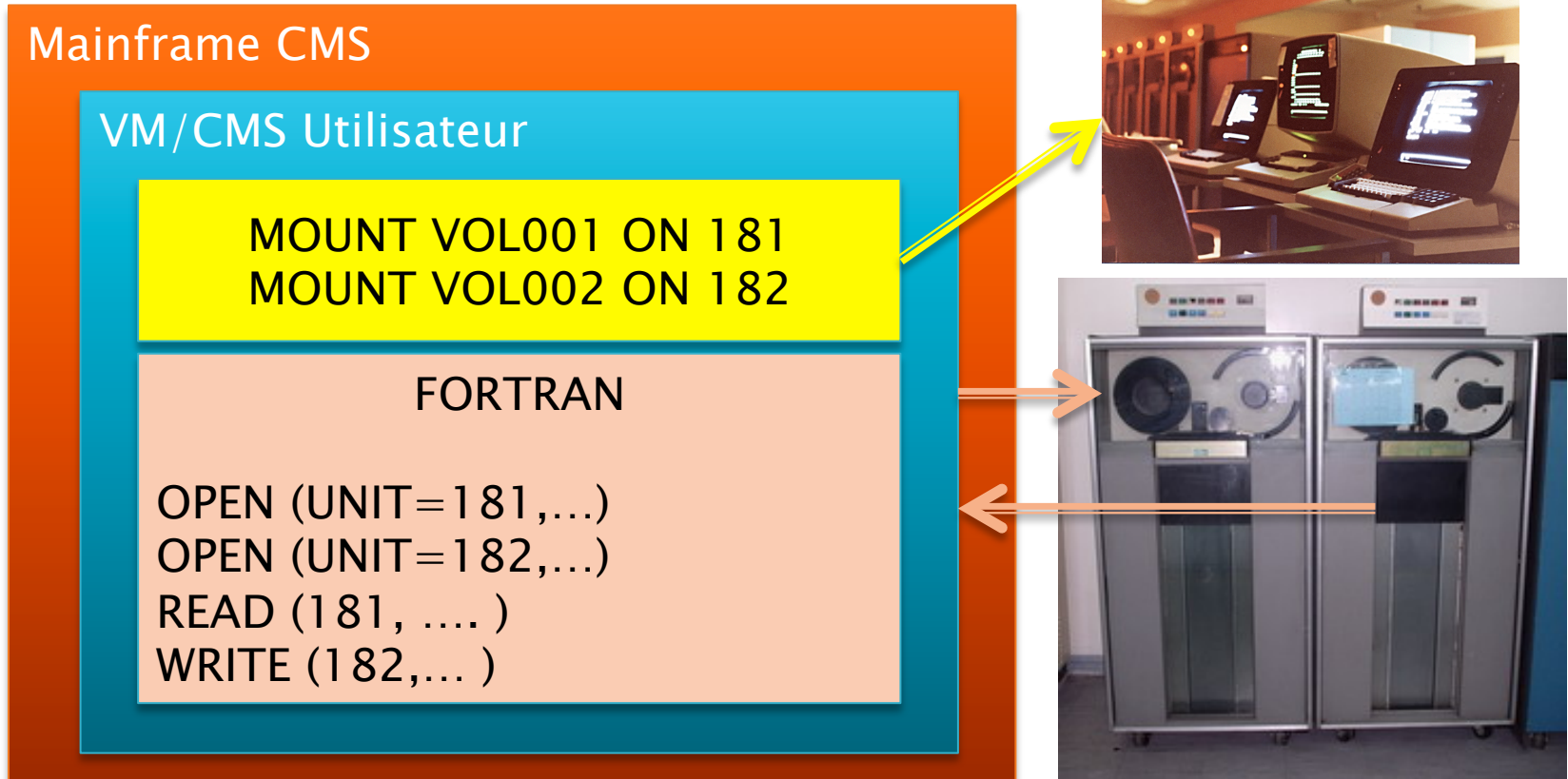




- ▶ Bandothèques
  - 40,000 cartouches



- ▶ L'utilisateur doit gérer ses volumes et ses fichiers
  - Identifiant du volume / position des fichiers sur la bande
  - Lots de bandes et liste des fichiers font partie des données des manipulations
- ▶ E/S des jobs se font directement sur bande
  - Un préambule en langage de commande permet de demander les montages
  - Les E/S se font directement au travers de périphériques virtuels

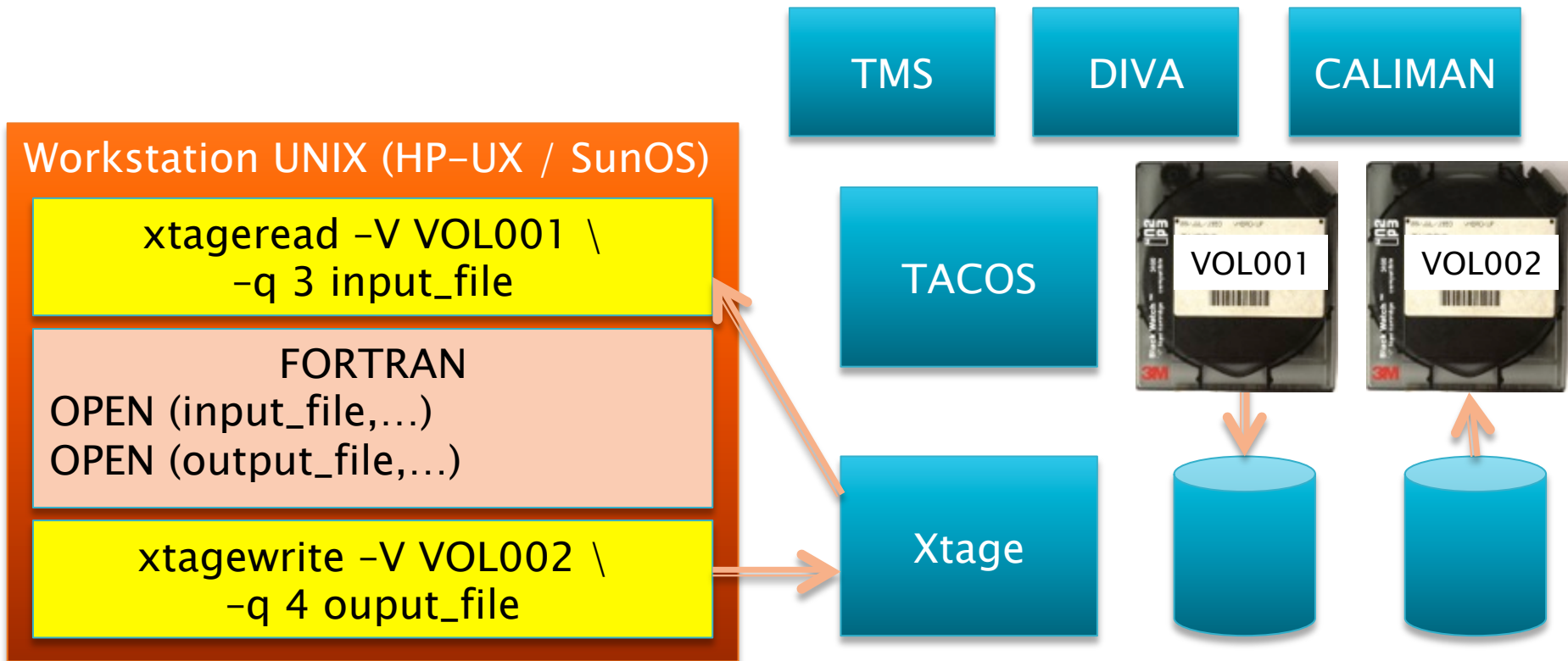




- ▶ 1993 : 6 Robots StorageTek 4400
  - Capacité : 6 x 6000 cartouches
  - Automatisation des montages depuis les jobs



- ▶ Portage des outils d'accès bandes à l'environnement UNIX
  - Les fichiers sont stockés sur des tampons disques
  - Accès aux fichiers depuis les jobs via montage NFS ou RFIO
- ▶ L'utilisateur doit toujours gérer ses bandes et ses positions





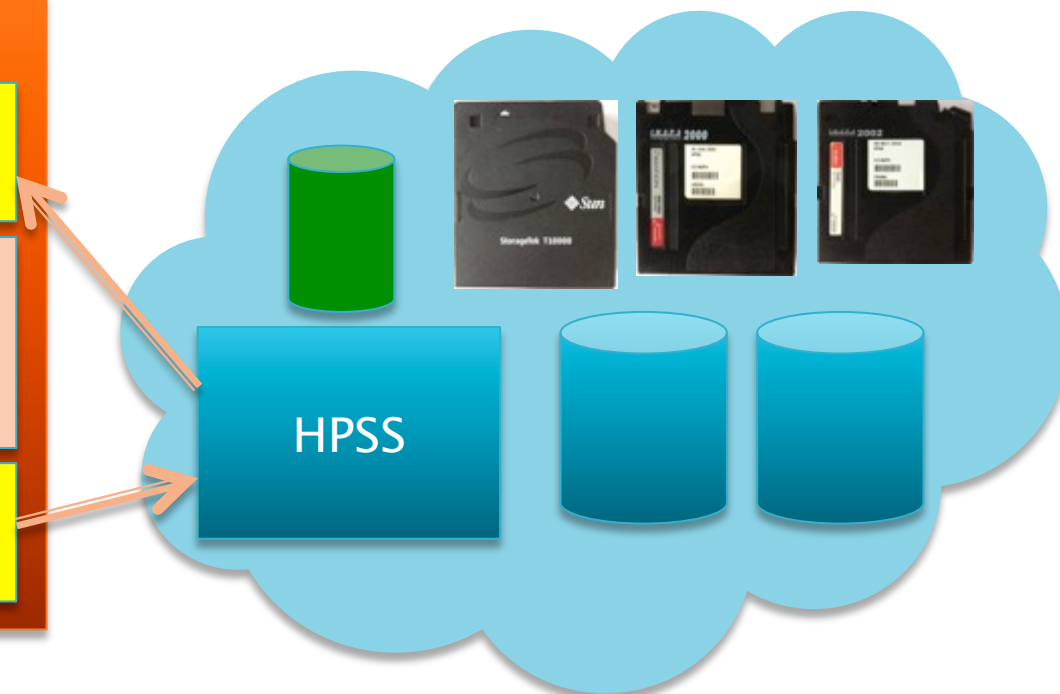
- ▶ HPSS : Système de stockage hiérarchique (HSM)
  - Gestion transparente de la localisation des données
  - 2 niveaux de hiérarchie : disque et bande
- ▶ Introduit pour Babar en 1999
- ▶ Permet de décharger l'utilisateur de la gestion des volumes
- ▶ Espace de nommage de type système de fichier POSIX
- ▶ Accès via rfiio : (rfcp , rfdir, ...)

## Worker Node Linux

```
$ rfcp /hpss/in2p3.fr/input_file  
  \ /scratch/
```

```
FORTRAN  
OPEN (input_file,...)  
OPEN (output_file,...)
```

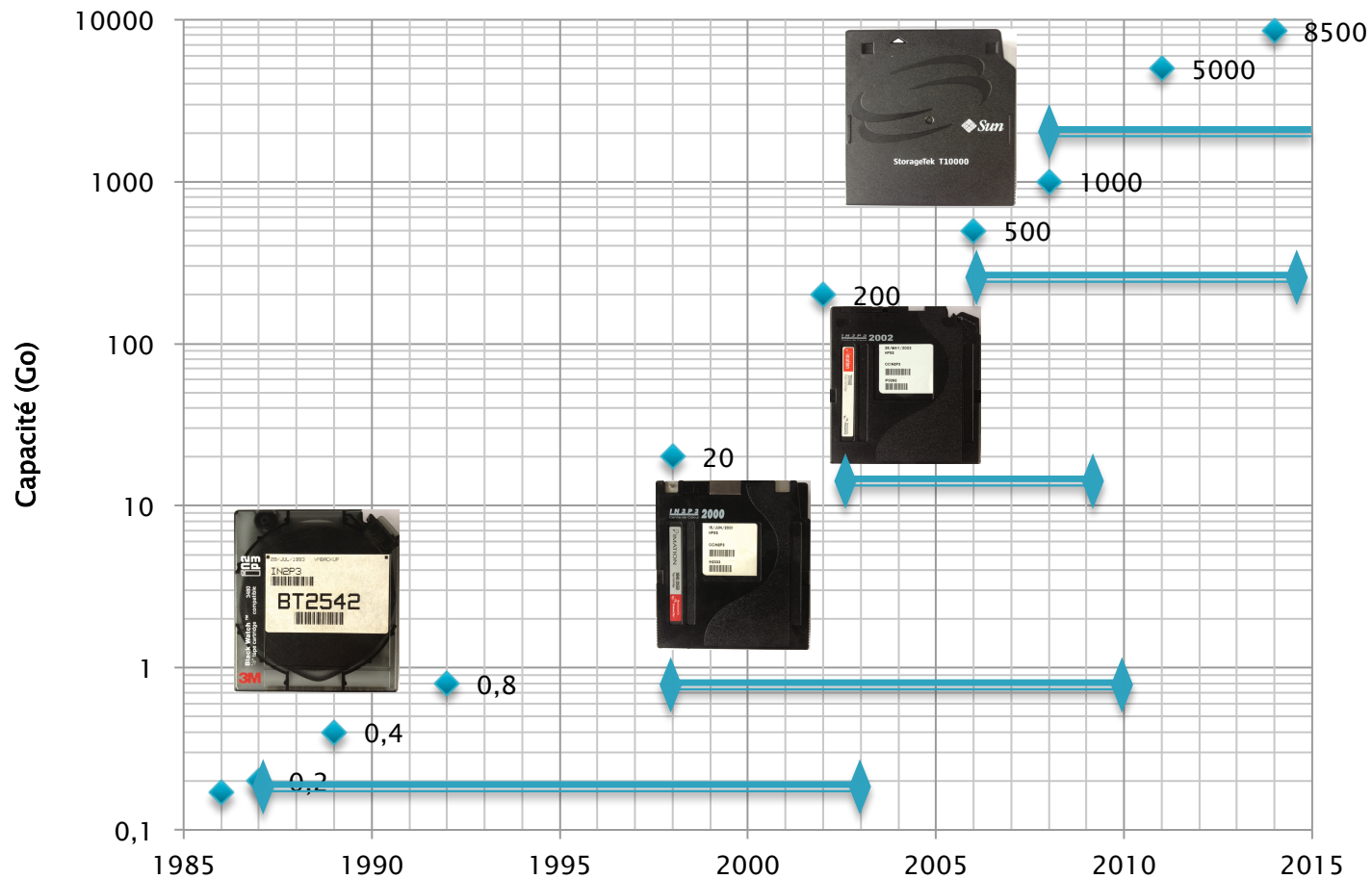
```
$ rfcp /scratch/output_file \  
  /hpss/in2p3.fr/../
```



- ▶ Les bandes : seul moyen pour transférer des données entre les sites des expériences et les centres de calculs
  - 2 navettes par semaine entre le CC et le CERN
- ▶ Babar : 1<sup>ère</sup> expérience à transférer ses données par réseau vers le CC
  - Liaison 155 Mbits entre SLAC et le CC
- ▶ En 2014 : Expérience HESS en Namibie
  - Envoi mensuel de 10 To par bande vers le CC par transporteur

## ▶ Technologies à durées de vies longues

- Lecteurs : ~ 7 ans
- Médias : ~ 10-15 ans
- Robots : ~ 15-20 ans

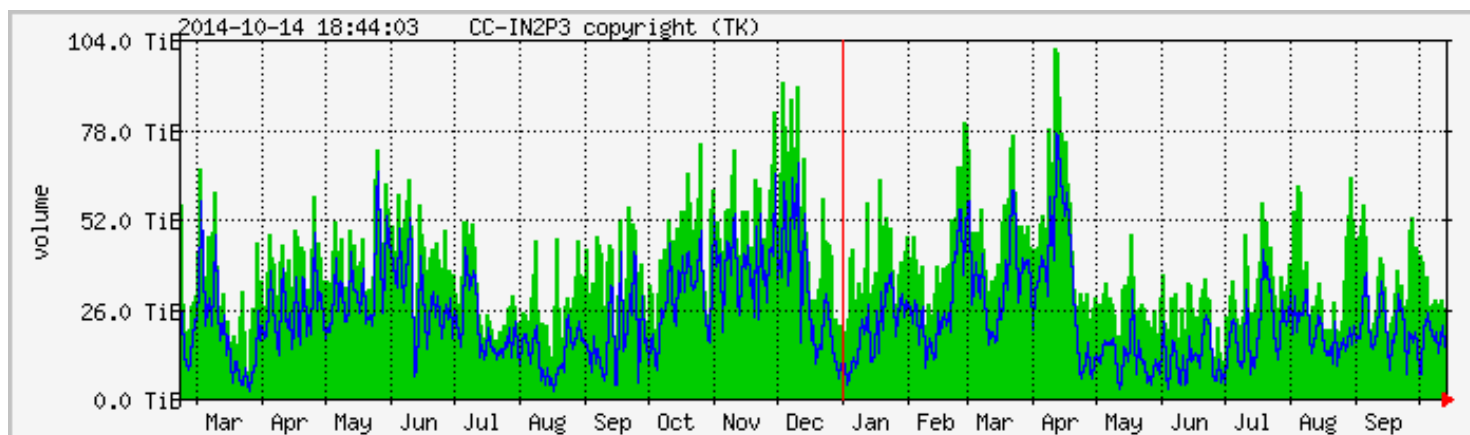
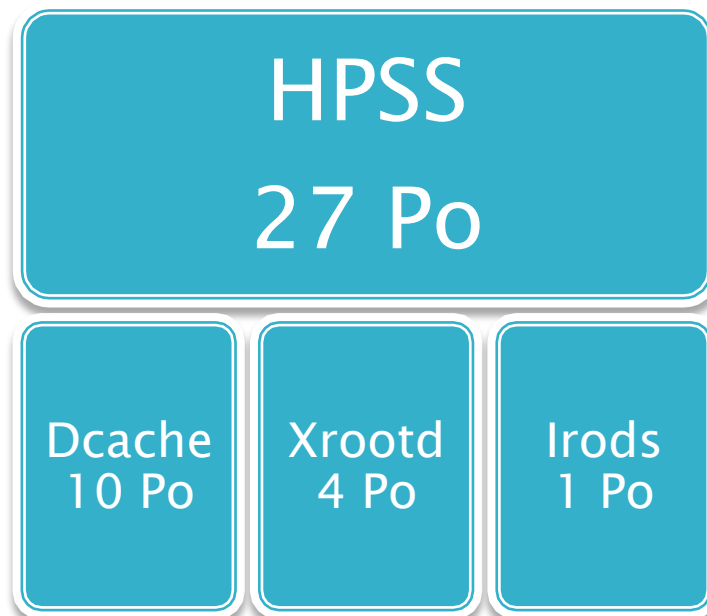


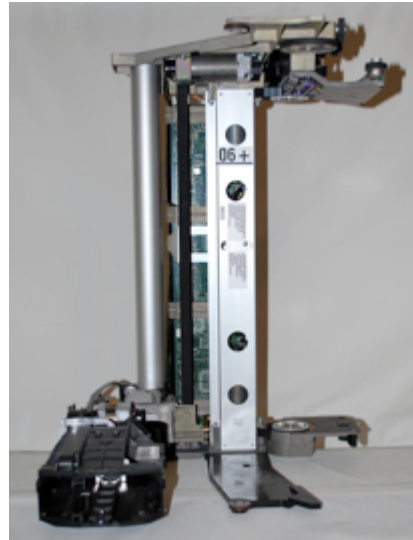
- ▶ 2005 : Arrêt du support des 3480 (au bout de 20 ans)
- ▶ 2008 : Démontage des derniers silos 4400



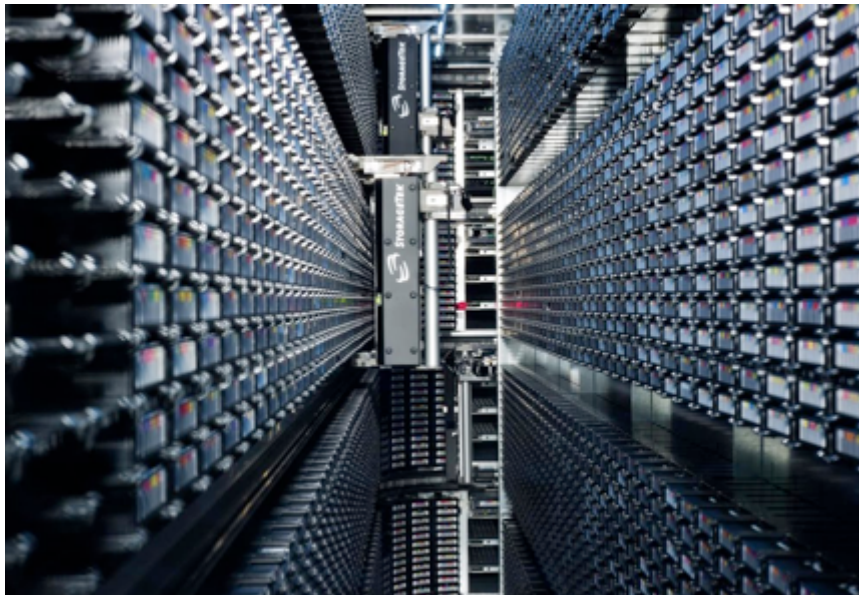


- ▶ HPSS en 2014
  - 27 Po
  - 47 Mo de fichiers
- ▶ 90 % des accès via d'autres systèmes de stockages
- ▶ Fortement sollicité
  - 37 To / jours
  - Pic à 100 To





- ▶ 4 x Oracle SL8500
  - Interconnectées
- ▶ Capacité théorique
  - 40000 bandes
  - 256 lecteurs
  - 340 Po
- ▶ En production
  - 30000 bandes
  - 6000 montages / jours
  - 130 lecteurs



## ▶ Oracle T10000-D

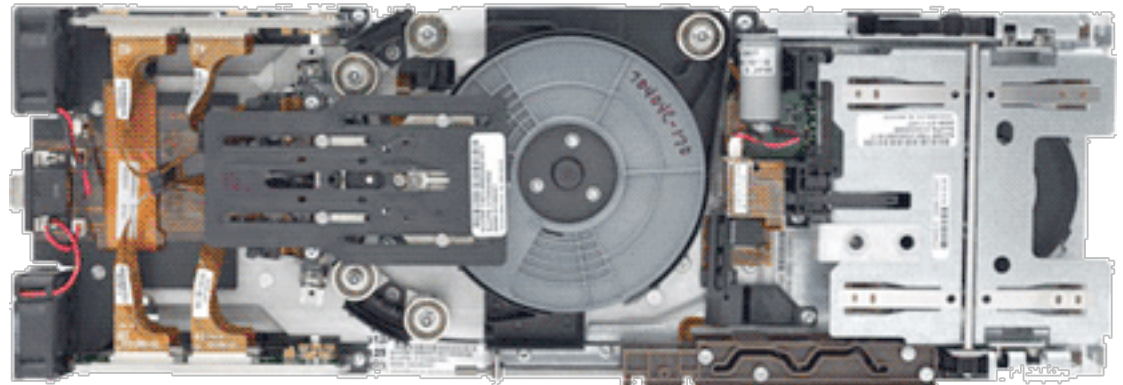
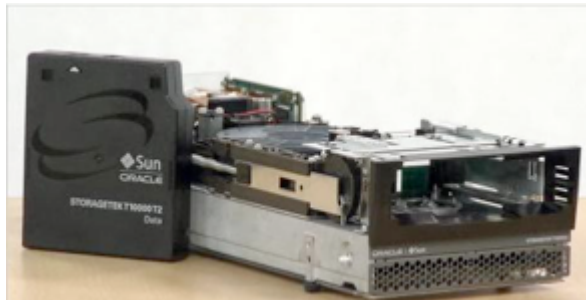
- Capacité : 8.5 To (1147 m)
- Débit : 252 Mo/s
- 4608 pistes (144 enroulements de 32 bits)
- Temps moyens de positionnement : 50 s

## ▶ Lecture/Ecriture séquentielle :

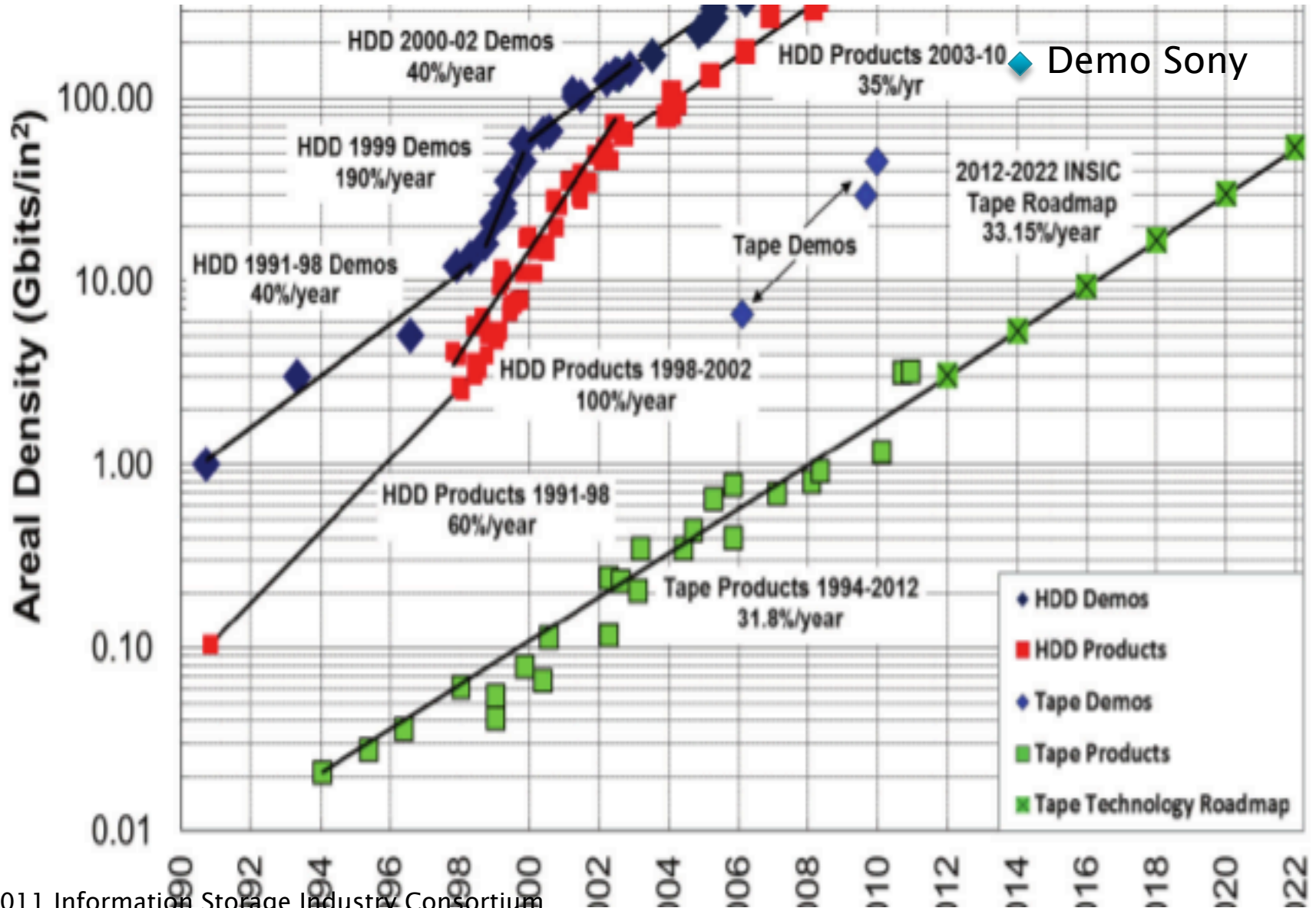
- + 9 h pour lire/remplir une bande (144 A/R)

## ▶ Lecture aléatoire :

- Temps de positionnement : 50s
- Temps de lecture du fichier (1Go) : 4s
- Durée de relecture : 127 h (+ de 5 jours)









## M Technologies

TECHNOLOGIES Jeux vidéo Hits Playtime Libertés numériques Téléphonie mobile Droit d'auteur Ré

### Une « super cassette » pour stocker des données

Le Monde.fr | 06.05.2014 à 12h46 • Mis à jour le 06.05.2014 à 14h10 |

Par Audrey Fournier

Réagir Classer

Partager

Recommander Partager 1 298 personnes recommandent ça. Soyez le premier parmi vos amis.



Les fétichistes du Walkman ont peut-être une raison de se réjouir : la bonne vieille cassette à bande magnétique est sur le point de faire un retour en fanfare. Sony, inventeur du baladeur, a dévoilé le 4 mai lors de l'International Magnetics Conference à Dresde (Allemagne) une cassette dotée d'une capacité de stockage impressionnante de 185 teraoctets, c'est-

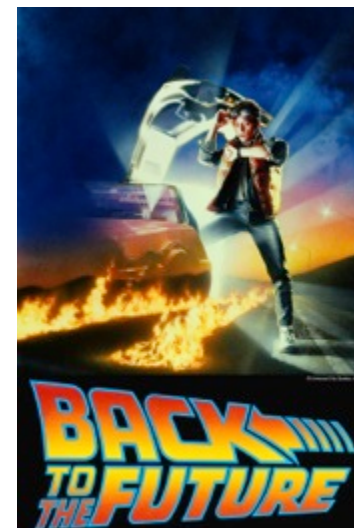
- ▶ Prévisions d'accroissement de densité :
  - Bande : +30 % par an
  - Disque : +20 % par an
- ▶ Démo Sony :
  - 185 To sur une bande
  - 148 Gbit / in<sup>2</sup>
- ▶ Disponibilité : 2025 ?
- ▶ 27 Po de HPSS sur 150 cartouches

### ▶ LTFS : Linear Tape File System

- Système de fichier auto-descriptif sur bande
- Permet de manipuler les données comme un système de fichier
- Non lié à un logiciel externe
  - Drivers spécifiques au lecteur
  - 2 partitions : index et données
- Les métadonnées dissociées des données
- Support : LTO 5/6 – IBM Jaguar – Oracle T10000D

### ▶ Cas d'utilisation possible :

- Echange massif de données entre sites
  - 1 Po transféré à 1 Gbit/s = 115 jours
  - 1 Po = 400 bandes LTO6
- Hébergement temporaire de communautés dans un CC



- ▶ La bande a encore de beaux jours devant elle :
  - Très capacitif : croissance +30% an
  - Peu onéreux : 5-10 % du CTO d'un disque
    - 17 KW pour 30 Po de données
- ▶ Malgré des inconvénients intrinsèques :
  - Média séquentiel, pas de parallélisation possible
  - Performances par TB en baisse : (sur 30 ans)
    - Augmentation de capacité : x 50,000
    - Augmentation du débit : x 250
    - Temps de montage/chargement : x 1 - 0,5
- ▶ Solution d'optimisation :
  - RAO : Recommended Access Order
  - Le lecteur donne à l'application l'ordre d'accès optimisé pour la relecture de fichiers