



Réunion du groupe PIN (Pérennisation des Informations Numériques)

La stratégie de stockage au Centre de Calcul du CNES

Anne Jean Antoine
17 janvier 2002



17/01/2002

Séminaire du groupe PIN : La stratégie de stockage au Centre de Calcul

Le stockage est crucial ...

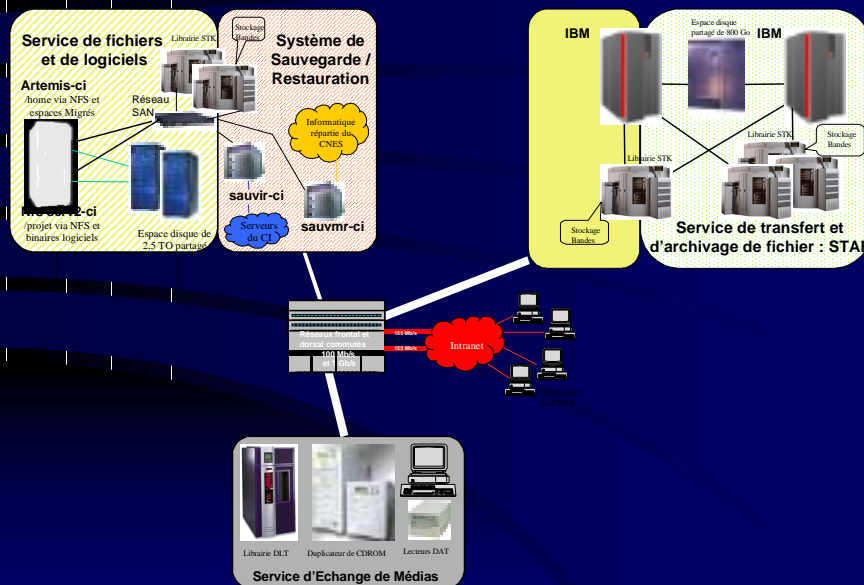
- ⇒ Les entreprises ont besoin d'un accès fiable et rapide aux données stockées
- ⇒ Elles doivent mettre en service, faire évoluer et maintenir facilement les systèmes de stockage
- ⇒ En quelques années, l'aspect stockage est devenu prépondérant et critique
- ⇒ Le stockage a maintenant un impact direct sur les budgets, le personnel technique, les infra-structures informatiques (performances réseaux, disponibilité, ...)
- ⇒ **Le résultat est que le stockage est devenu un domaine à part entière qu'il faut gérer au même titre que les autres composantes de l'informatique.**

L'organisation du Cnes autour du stockage

- ⇒ Centralisation des moyens de stockage au Centre de Calcul :
 - Service de sauvegarde des machines de traitement centrales Unix (étendu bientôt aux moyens départementaux),
 - Serveur de fichiers et archivage de contextes d'exploitation Unix,
 - Service d'archivage long terme (STAF).

- ⇒ Chacun de ces services est indépendant et dispose de ses propres moyens de stockage et logiciels ainsi que de ses propres équipes d'exploitation.

- ⇒ Principes retenus :
 - "A un besoin de stockage, une solution adaptée",
 - Mutualiser les moyens de stockage et dédier du personnel compétent.



La terminologie du stockage et de l'archivage

➔ Problématique du stockage

- temps d'accès et délai d'obtention des fichiers
- disponibilité (7/7j - 24/24h)
- fiabilité

➔ Problématique de l'archivage

- sécurité des données
- format d'accès
- coût de conservation des GigaOctets

Définitions

➔ Service de fichiers

- environnement de travail quotidien
- besoin de délai de mise à disposition très rapide
- taux d'accès très élevé
- durée de vie des fichiers courte (moins de 2 ans) => **stockage limité**
- données reproductibles donc peu sensibles => non dupliquée

Définitions (suite)

- ➔ Service de sauvegarde
 - permet de s'affranchir des erreurs utilisateurs et des pannes hardware (crash-disque)
 - permet de conserver un historique (retour arrière possible) => pseudo gestion de version
 - limitation dans le temps pour les restaurations (moins de 2 mois) => **stockage limité**
 - taux d'accès très faible
 - délai d'accès rapide

Définitions (suite)

- ➔ Service d'archivage
 - données pérennes, souvent unique et parfois non reproductibles
 - taux d'accès faible en continu
 - délai de restitution plus lent acceptable
 - durée de conservation longue (supérieure à 10 ans) => **stockage illimité**
 - nécessite une grande intégrité des données

Principes majeurs retenus dans STAF (1)

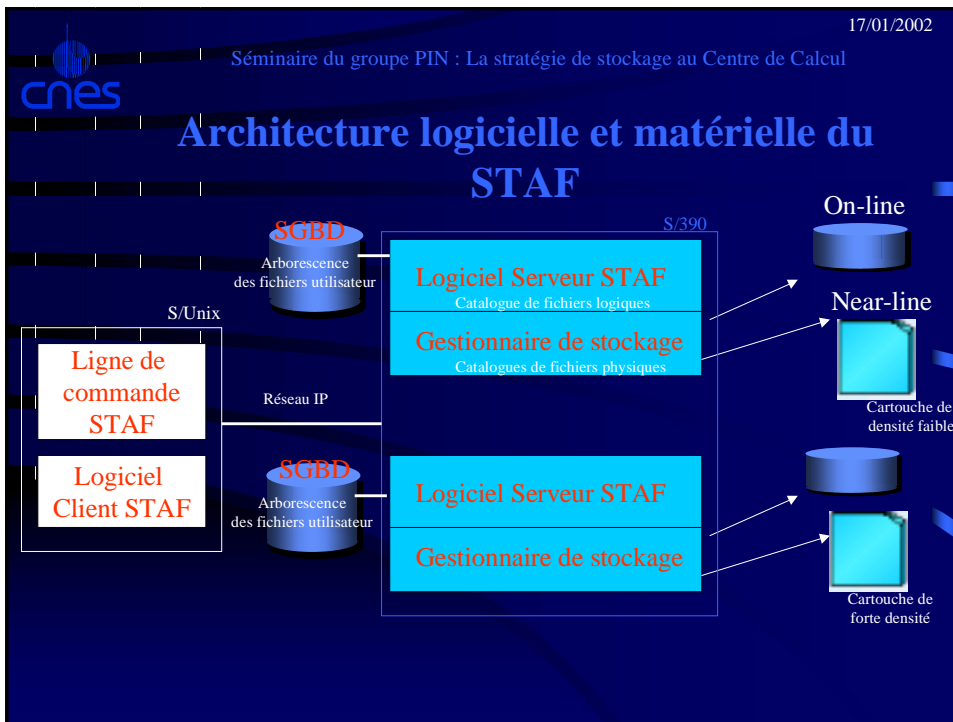
- ⇒ Moyens de stockage centralisés pour :
 - une gestion unifiée et une prise en compte à un seul endroit des besoins d'évolution (ex : migration technologique)
 - une consolidation des ressources de stockage (disque, bandes,...)
 - un centre de coût identifié.

- ⇒ Média (lecteurs et robotiques) adaptés à l'archivage long terme (selon des critères techniques et économiques) - visibilité du plan de développement du constructeur (« roadmap ») - qualité du support constructeur.

- ⇒ Migration technologique automatique tous les 5 ans (malgré la garantie du constructeur > 15 ans).

Principes majeurs retenus dans STAF (2)

- ⇒ Dans l'architecture du système :
 - privilégier un fort degré d'automatisation du système (pas de manipulation des média),
 - indépendance entre le logiciel d'accès aux données et les données stockées : modularité de l'architecture favorisant la maintenance évolutive du système - utilisation de standards,
 - extensibilité du système pour s'adapter à la demande toujours plus volumineuse du stockage (densité des média x2 tous les 18 mois),
 - adéquation de la densité des média avec la taille moyenne des fichiers stockés : garantie de délai de performance en restitution,
 - compatibilité en milieu hétérogène (transfert binaire, ligne de commande unique sur les différents systèmes client).



- 17/01/2002
- Séminaire du groupe PIN : La stratégie de stockage au Centre de Calcul
- Principaux besoins des utilisateurs STAF**
- Interface d'accès au service d'archivage unique quelque soit le moyen de traitement (système d'exploitation hétérogène) - besoin d'organiser et accéder à l'archive indépendamment des matériels sous-jacents ,
 - Plus de manipulation de média et l'utilisateur ne doit pas avoir connaissance du support physique - il définit ses besoins en terme de garantie de stockage et de délai de restitution (principe des classes de service du STAF) pour son fichier, indépendamment de sa localisation dans l'arborescence.
 - La migration technologique des supports doit être transparente pour l'utilisateur final,
 - Besoin de fiabilité et d'intégrité des données confiées au centre d'archivage, besoin de sécurité et confidentialité - disponibilité la meilleure en phase de restitution.

17/01/2002

Séminaire du groupe PIN : La stratégie de stockage au Centre de Calcul

Principe des classes de service dans STAF

- ⇒ Un attribut affecté par l'utilisateur à un fichier ou plusieurs, au moment de son archivage.
- ⇒ La Classe de Service permet à l'utilisateur d'adresser un besoin en délai de mise à disposition du fichier et un besoin ou pas de duplication de son fichier.
- ⇒ Reposant sur une hiérarchie de média à 2 niveaux :

17/01/2002

Séminaire du groupe PIN : La stratégie de stockage au Centre de Calcul

Les fonctions HSM du gestionnaire de stockage

ON LINE

ML1

ML0

HSM Migrate

HSM backup

HSM Recycle

NEAR LINE

ML2

Backup Tape

Génération N

Migration technologique des supports

Génération N+1

STAF : Le stockage ON LINE

➔ Disques magnétiques de la gamme IBM et EMC2 3390 Modèle 3 (2.8 Go) et 3390-9 (9Go) à venir

- Redondance RAID1 (disque miroir) et RAID5
- ML0 (espace tampon avant migration vers les médias near-line)
et ML1 (espace d'archivage disque -> +100 Go accumulés)
- Volume global en ligne : +500 Gbytes.

STAF : Le stockage NEAR LINE

➔ Pour la « migration » :

Cartouches (format standard IBM 3490E - canal Escon) de la gamme StorageTeK et lecteurs associés :

- 2 « Timberline » (cartouches High Tape de 800 Mo non compressés)
- 12 « 9840 » (2547 cartouches 9840 de 20 Go non compressés pour 50 To)

et de la gamme IBM - lecteurs associés :

- 8 « 4490 » (cartouches de 400 Mo non compressés) (3054 cartouches 4490 + Timberline représentent un volume global de 2.5 To)

17/01/2002

Séminaire du groupe PIN : La stratégie de stockage au Centre de Calcul

STAF : Le stockage NEAR LINE

➔ Pour la « sauvegarde » :

Cartouches (format standard IBM 3490E - canal Escon) de la gamme StorageTeK et lecteurs associés :

- 6 « Redwood » (cartouches de 50 Go non compressés)

VERS

- 9940 (cartouches de 60 Go non compressés) en cours.

17/01/2002

Séminaire du groupe PIN : La stratégie de stockage au Centre de Calcul

Retours d'expérience

- ➔ Système d'archivage adapté aux fichiers de grande taille (imposant aux utilisateurs le regroupement des fichiers de faible taille).
- ➔ Fiabilité du système (pas de perte de fichiers depuis 1995) - Sécurité des accès au service et aux données - Performances en constante amélioration.
- ➔ Intégrité référentielle : cohérence parfaite entre le catalogue des fichiers logiques et celui des fichiers physiques sous le gestionnaire de stockage.
- ➔ Limitation de DFHSM sur le mainframe IBM :
 - ne permet pas plusieurs types de médias au niveau ML2,
 - propose dans ses versions à venir le rappel de fichiers en mode « copy » pour une consultation (et non plus en mode « move » entraînant la fragmentation des médias near-line et le « recycle »)
- ➔ Ligne de commande trop riche par rapport à l'utilisation réelle des chaînes automatiques
- ➔ Taux de disponibilité du service (97.4% en fin 1998) en amélioration (arrêt obligatoire du service pour montée de niveau matérielle et logicielle).

Evolutions envisagées

- ➔ Rénovation du gestionnaire de l'application (meilleure modularité - utilisation de standards)
- ➔ Ouverture du système STAF à d'autres gestionnaires de stockage
- ➔ Passerelle entre les serveurs de fichiers sous DMF et le système STAF
- ➔ Intégration de la technologie SAN (Storage Area Network).

CONCLUSION

- ➔ « Mutualiser les infra-structures de stockage »
- ➔ « Choisir ses logiciels et supports en fonction du besoin »
- ➔ « Le stockage est un domaine à part entière de l'informatique »
- ➔ « Construire un système d'archivage souple et évolutif, basé sur des standards »
- ➔ « Préserver efficacement ses données »