



CENTRE NATIONAL D'ÉTUDES SPATIALES

Un exemple de migration de la fonction "stockage" de l'OAIS : remplacement de l'infrastructure du système STAF et migration des données au CNES

PLAN DE LA PRESENTATION

■ Le contexte de la migration

- ◆ Origine du projet de refonte STAF-NG
- ◆ L'analyse des solutions possibles et les orientations choisies

■ La migration des données

- ◆ Principes généraux
- ◆ Phase de prototypage
- ◆ Phase de pré-migration
- ◆ Phase de migration / mécanismes de migration
- ◆ Phase de post-migration
- ◆ Moyens humains

■ Bilan de la migration

ORIGINE DU PROJET STAF-NG

■ Constat fin 2006 – début 2007

- ◆ Fin de vie annoncée des silos de stockage STK 9310 (5 robots, ~15 000 cartouches)
- ◆ Saturation de la capacité de traitement du mainframe
- ◆ Augmentation significative de la volumétrie

■ Une première instruction avec STK permet d'identifier les points suivants :

- ◆ une simple mise à niveau du mainframe n'est pas envisageable car
 - ce modèle est en voie d'obsolescence
 - les canaux de communications avec les lecteurs sont en voie d'obsolescence
- ◆ la nécessité de remplacer le robot de stockage implique une remise en question de la technologie lecteur actuellement utilisée

LANCEMENT DU PROJET STAF-NG

- **Début du projet : janvier 2007**
- **Périmètre du projet**
 - ◆ unités de traitement qui hébergent l'application STAF
 - ◆ moyens de stockage
 - ◆ intégration de la nouvelle version de l'application STAF V3
- **Démarche projet**
 - ◆ Phase A-B : analyse des solutions possibles et définition préliminaire
 - ◆ Phase C-D : réalisation industrielle via intégrateurs référencés
- **Principales contraintes**
 - ◆ fin de maintenance des silos en 2010
 - ◆ être capable d'accueillir le projet Pléiades sur les moyens définitifs à partir de mi-2009

PHASE A-B : ANALYSE DES SOLUTIONS POSSIBLES

■ **Planning** : février 2007 à décembre 2007

■ **Objectif** : évaluer les solutions proposées par les constructeurs et identifier les technologies viables dans le cadre de cette infrastructure du service STAF

■ **Constats** :

- ◆ Les constructeurs ont proposé, pour la partie ordinateur du projet, des solutions dans la filière technologique « mainframe » (IBM) et dans la filière technologique dite « open » (SUN, SGI),
- ◆ Le stockage tout disque n'est pas compatible avec les volumétries estimées pour le STAF en terme de budget d'investissement (2 à 2,5 fois plus cher les 3 premières années), surface au sol, consommation électrique,...
- ◆ Deux solutions viables retenues : « mainframe » (IBM) et « open » (Sun)

ANALYSE DE RISQUES – solution SUN

- **Les principaux risques identifiés :**
 - ◆ incertitude sur la faisabilité technique de la migration des données,
 - ◆ mauvaise connaissance du produit SAM-FS,
 - ◆ fonctionnement de l'application STAF non encore testé.

- **Les actions en diminution de risque :**
 - ◆ validation d'une maquette représentative de la solution Sun (fonctions, migration),
 - ◆ visite d'un site représentatif utilisant le produit SAM-FS (météo allemande).
 - la migration a duré 6 mois pour une volumétrie de 750 To,
 - une charge d'exploitation légère.

- **Risques résiduels**
 - ◆ Nécessité de former les personnels au produit SAM-FS,
 - ◆ Licences SAM-FS indexées sur le volume de données brutes,
 - ◆ Pérennité et support du produit SAM-FS,
 - ◆ Problématique de la migration des données de l'environnement « mainframe » vers l'environnement « open ».

ANALYSE DE RISQUES – solution IBM

- **Le principal risque identifié pour la solution IBM était la difficulté à stabiliser le fonctionnement de la nouvelle version de l'application STAF dans l'environnement mainframe.**

- **Les actions en diminution de risque :**
 - ◆ **une instruction de la possibilité de pérenniser l'application STAF V2**
 - Hypothèse abandonnée (IBM a évalué que le travail de pérennisation global représente environ 800 jours/hommes)

- **Risques résiduels**
 - ◆ Dépendance vis-à-vis d'une technologie propriétaire fermée,
 - ◆ Coûts de fonctionnement élevés difficiles à réduire,
 - ◆ Faible disponibilité des compétences dans l'industrie,
 - ◆ Difficulté économique à mettre en place une solution entièrement redondante,
 - ◆ Durée de la migration importante.

QUELQUES CHIFFRES DE VOLUMETRIE

■ Volumétrie en 2008

- ◆ Plus de 300 To archivés (hors copies), 7 500 000 de fichiers
- ◆ Flux de 1,5 To archivés par semaine
- ◆ Représente plus de 15000 cartouches utilisées réparties sur 5 silos

■ Volumétrie en 2010

- ◆ Plus de 650 To archivés (hors copies), 11 000 000 de fichiers
- ◆ Flux de 3 To archivés par semaine
- ◆ Représente plus de 1000 cartouches utilisées réparties sur 2 robotiques

■ Évolution des volumétries

- ◆ > 5 To archivés par semaine en 2011, volumétrie totale > 900 To (hors copie)
- ◆ > 7 To archivés par semaine en 2012 , volumétrie totale ~ 1,3 Po (hors copie)

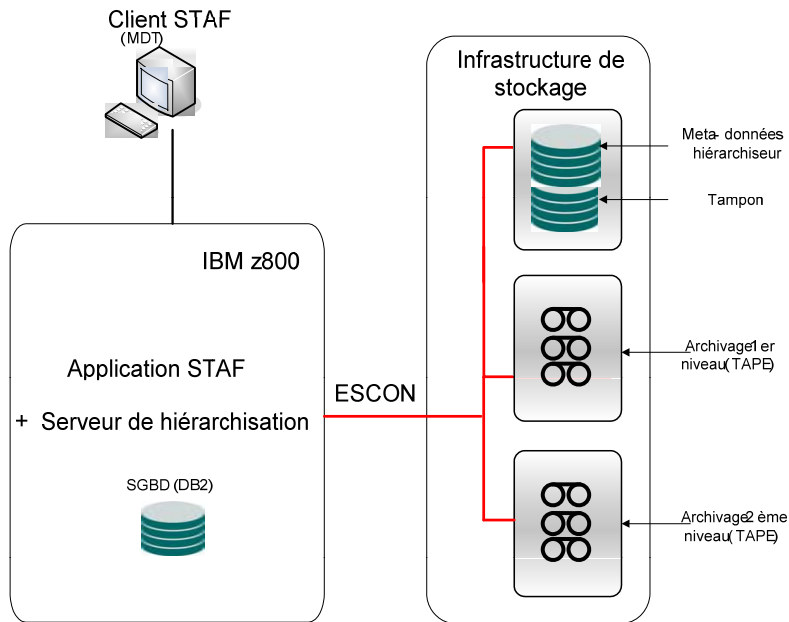
PHASE C-D : REALISATION INDUSTRIELLE

- **Planning** : mai 2008 à novembre 2009
- **Objectif** : Conception Détaillée, installation/intégration, recettes/qualification, migration
- **Consultation intégrateurs** : la proposition de l'équipe industrielle APX/SUN est retenue pour les phases C/D

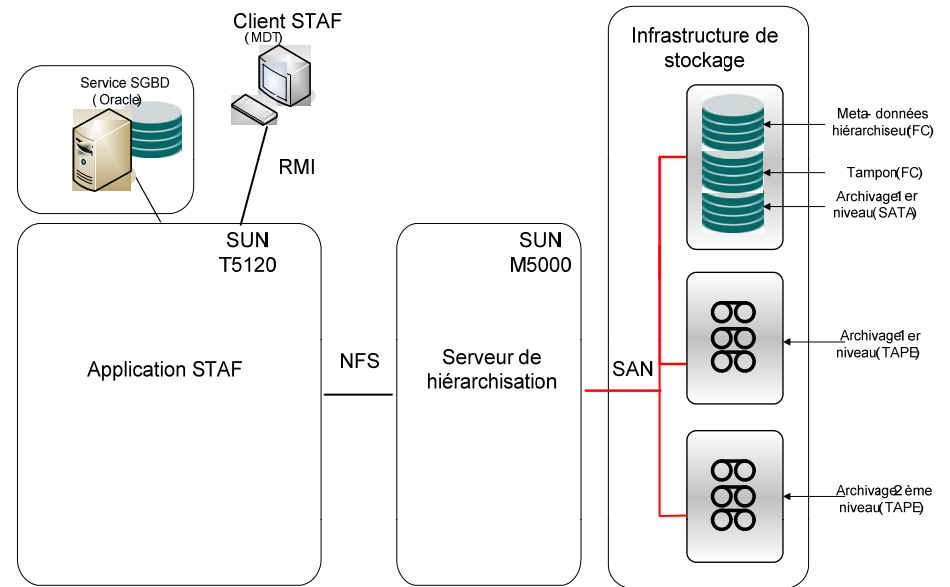


ARCHITECTURE TECHNIQUE : nouveaux principes

AVANT



APRES



Plan de Migration : Principes généraux

- **Les opérations de migrations comportent 2 périodes effectives :**
 - ◆ **Une période de pré-migration**
 - Le système de production est l'ancien système
 - Migration "continue" par bande
 - ◆ **Une période de migration consécutive à la mise en exploitation du STAF-NG**
 - Le système de production est le STAF-NG
 - Migration "aléatoire" de fichiers demandés en restitution est encore hébergés STAFV2
 - Migration "continue" par bande

- **Principales contraintes :**
 - ◆ **Pas de retour arrière possible après la mise en exploitation.**
 - ◆ **Il n'est pas possible de différencier les projets lors de la migration.**
 - ◆ **Les opérations de migration se font 7J/7 24H/24.**

Les phases de migration : phase de prototypage

■ Réalisée en juillet/août 2008

■ Objectif :

- ◆ Tester le bon fonctionnement des processus de migration utilisés
- ◆ Faire une estimation aussi précise que possible de la durée de migration

■ Principales conclusions :

- ◆ Validation fonctionnelle des processus de transfert des fichiers avec contrôle de l'intégrité
- ◆ Durée estimée de la migration de 8 à 11 mois

Les phases de migration : phase de pré-migration

■ Réalisée dès le début décembre 2008

■ Nécessaire car

- Contrainte de fin de migration à fin 2009
- Report disponibilité de la salle secours à mi-février 2009
- Durée de migration évaluée entre 8 et 11 mois

■ Objectif

- ◆ Commencer la migration des fichiers au plus tôt
- ◆ Qualifier les outils de migration testés en phase de prototypage

■ Principe

- ◆ Identification des fichiers ayant une faible probabilité d'être modifiés
 - Bandes non accédées depuis plusieurs années
- ◆ Transfert de ces archives vers la nouvelle infrastructure
- ◆ Utilisation des mêmes outils et mécanisme que pour la migration
 - Intégrité des fichiers
 - Mémorisation des informations nécessaires au post-traitement

Les phases de migration : phase de migration

■ Réalisée à partir de mars 2009

■ Objectif :

- ◆ Migrer les fichiers qui n'ont pas été pré-migrés
- ◆ Assurer la restitution des fichiers non encore migrés demandés par les utilisateurs

■ Principe :

- ◆ Un seul système actif du point de vue de l'utilisateur
- ◆ Cohabitation de 2 types de migration
 - Migration « aléatoire » des données au fil de l'eau suivant les requêtes utilisateurs
 - Migration « continue » des données par lot de bandes

Les phases de migration : Étapes de mise en exploitation STAF-NG

1. Arrêt de la production de l'ancien STAF
2. Migration des méta-données DB2 vers Oracle
3. Extraction des caractéristiques fichiers de la base Oracle
4. Mise au format SAMFS des caractéristiques fichiers
5. Création des inodes sur le tampon SAMFS
6. La migration des fichiers
 - ◆ Migration « aléatoire » générée par la production.
 - ◆ Migration « continue »
 1. construction des fichiers bandes
 2. traitement d'un fichier bande
 3. traitement des fichiers migrés sur disques (liste pseudo-bande)
 4. traitement des bandes avec erreurs en lecture (manuel)

Les phases de migration : Mécanismes de la migration

■ Garantie d'intégrité des fichiers migrés

- ◆ **Le contrôle de l'intégrité des fichiers migrés est assuré par le moteur de migration**

- ◆ **Principe du contrôle d'intégrité :**
 - Calcul et mémorisation du CRC (CRC32 de la Zlib) lors de la lecture du fichier sur l'ancien STAF.
 - Calcul du CRC du fichier sur SamFS
 - Comparaison des CRC
 - si comparaison Ok alors le fichier est renommé avec son nom définitif
 - si comparaison NOK alors suppression du fichier sur SamFS et log d'une erreur dans le journal SamFS

Les phases de migration : Mécanismes de la migration

■ Suivi de la migration

◆ Contrôle de performance

- Contrôle quotidien de l'indicateur de performance projets utilisateurs
- Contrôle deux fois par jour de l'indicateur d'ajustement du taux de migration

◆ Suivi d'avancement de la migration

- Mesure de performance pour chaque fichier et chaque bande traités
 - volume
 - durée
 - débit
 - ...
- Mesure hebdomadaire de l'avancement de la migration pour chaque partition
 - Le volume total des fichiers transférés.
 - Le temps total du transfert des fichiers
 - Le débit obtenu pour le transfert des fichiers
 - Une estimation de durée de la migration en prenant comme base le volume total et le nombre total de fichiers à traiter par rapport à ce qui est déjà réalisé.
- Tenue d'un journal de bord de la migration.

Les phases de migration : phase de post-migration

■ Réalisée en octobre/novembre 2009

■ Objectif :

- ◆ Prouver que tous les fichiers ont bien été migrés

■ Principe :

◆ Comparaison entre la liste des fichiers connus dans la base de donnée applicative et la liste des fichiers présents sur les disques SamFS

- Si sur SamFS, aucun fichier n'est manquant par rapport à la base STAF
 - c'est que tous les fichiers sont migrés
- Si sur SamFS, des fichiers sont présents mais ne sont pas répertoriés dans la base STAF
 - c'est que ces fichiers ont été supprimés de l'application StafV2 depuis la migration -> Suppression des fichiers sur SamFS
- Si sur SamFS, il manque des fichiers par rapport à la base STAF
 - c'est que des fichiers ont été oubliés dans la migration -> migration des fichiers via les mécanismes de migration "aléatoire"

Plan de Migration : Moyens humains

■ Mise en œuvre de la migration

- ◆ Une personne MOI dédiée pour toutes les phases de la migration.
- ◆ Ressources CNES
 - Équipe exploitation CNES
 - Support Infogérant
 - Support IBM
 - Support projet STAF-NG

■ Suivi de la migration

- ◆ Une personne MOI (3 jours / semaine) dès pré-migration
 - suivi optimum de la migration
 - résoudre les éventuels problèmes
 - surveiller les performances
 - respect du plan d'avancement
- ◆ Ressources MOI (MCO) dès pré-migration
 - Suivi des performances
 - réglage du taux de migration
 - résolution incidents de niveau 1
- ◆ Ressources CNES équipe d'exploitation STAF
 - Suivi des performances
 - Résoudre les éventuels problèmes liés à l'environnement de production

BILAN DE LA MIGRATION - Les problèmes rencontrés

- ◆ **Incidents de production de l'ancien système ayant conduit à un arrêt des mécanismes de migration (sur la partition concernée).**
 - Erreurs de lecture de bande (9 occurrences), les données ont pu être lues soit par « swap de lecteur » soit sur la bande de backup (2 occurrences).
 - Saturation mémoire CSA (3 occurrences)
 - Erreur sur la clôture du fichier journal du DFSHSM (3 occurrences)

- ◆ **Les pools de bandes du DFSHSM n'étaient pas complètement figés**
 - Réactualisation périodique des listes TTOC

- ◆ **Problème de performances de migration dus à la capacité CPU du z800**
 - Recompilation du moteur de migration avec des options optimisant le calcul du CRC32

- ◆ **Problèmes de performances pour le calcul des empreintes MD5**
 - Désactivation du calcul initial MD5 des fichiers

- ◆ **Le nombre d'arborescences des fichiers archivés s'est trouvé plus élevé qu'estimé initialement, ce qui génère des temps de sauvegarde des méta-données importants (restructuration en cours)**

BILAN DE LA MIGRATION

- **L'objectif de fin de migration avant fin 2009 a été atteint (2 mois de pré-migration et 8 mois de migration)**
- **Les volumes migrés sont 386 To soit 8 millions de fichiers**
- **Les processus de migration ont transféré tous les fichiers présents dans la BD applicative et sur l'ancien système**
- **L'intégrité de chaque fichier migré a été vérifiée par « checksum » et tous les fichiers migrés sont valides**
- **La post-migration a révélé que 19 fichiers avaient été perdus au cours de la production de l'ancien système (entre 1998 et 2007)**

REX : SOLUTIONS «MAINFRAME» (IBM)

■ Avantages :

- ◆ **Solution stable et éprouvée,**

■ Inconvénients :

- ◆ **coûts d'exploitation important (difficilement justifiable dans un environnement mono-application)**
- ◆ **faible disponibilité des compétences dans l'industrie**
- ◆ **difficulté économique à mettre en place une solution entièrement redondante**
- ◆ **performance limitée de la solution (limitation des débits DFSMSHSM)**
- ◆ **difficulté de qualification de l'application STAF V3**

REX : SOLUTIONS «OPEN» (APX/SUN)

■ Avantages :

- ◆ Facilité à mettre en place une solution entièrement redondante,
- ◆ Capacité à dupliquer le modèle (exp. : STAF dédié à un projet particulier),
- ◆ Coûts d'exploitation moindres que solution mainframe (< facteur 2),
- ◆ Disponibilité des compétences dans l'industrie,
- ◆ Évolutivité des architectures (virtualisation stockage, mutualisation avec d'autres infrastructures, ...).

■ Inconvénients :

- ◆ Nécessité de former les personnels
- ◆ Problématique de la migration des données de l'environnement « mainframe » vers l'environnement « open »
- ◆ HSM SAM-FS « jeune » pas aussi mature que DMF mais qui répond parfaitement au besoin

PERSPECTIVES D'EVOLUTION

- **Événement récent : rachat de Sun par Oracle**
 - ◆ Événement difficilement identifiable au début du projet qui génère un risque sur la pérennité du HSM SAM-FS
 - ◆ Aujourd'hui la pérennité de SAM-FS est assurée par Oracle (produit « open »)
 - ◆ La migration vers un autre HSM « unix » serait bien plus aisée qu'à partir d'un environnement zOS.

- **Les volumétries vont continuer à croître de façon important**
 - ◆ Fin 2012 les flux d'archivage seront au minimum ~400 To / an

- **L'utilisation des robotiques STAF-NG pour les besoins de sauvegarde des espaces de stockage hors archivage a déjà commencé**

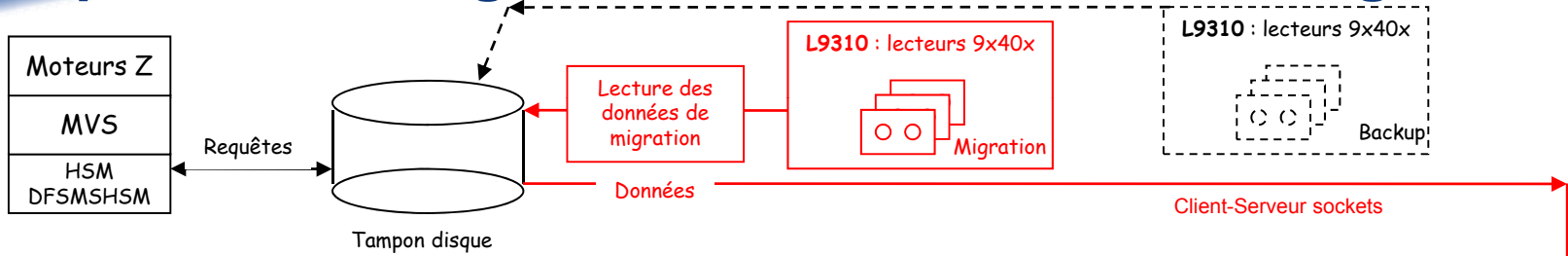
- **L'ouverture de l'infrastructure STAF-NG à des espaces de stockage de type HSM est envisagée dans le projet de refonte des infrastructures de stockage et de sauvegarde**

Questions / Réponses

ANNEXES

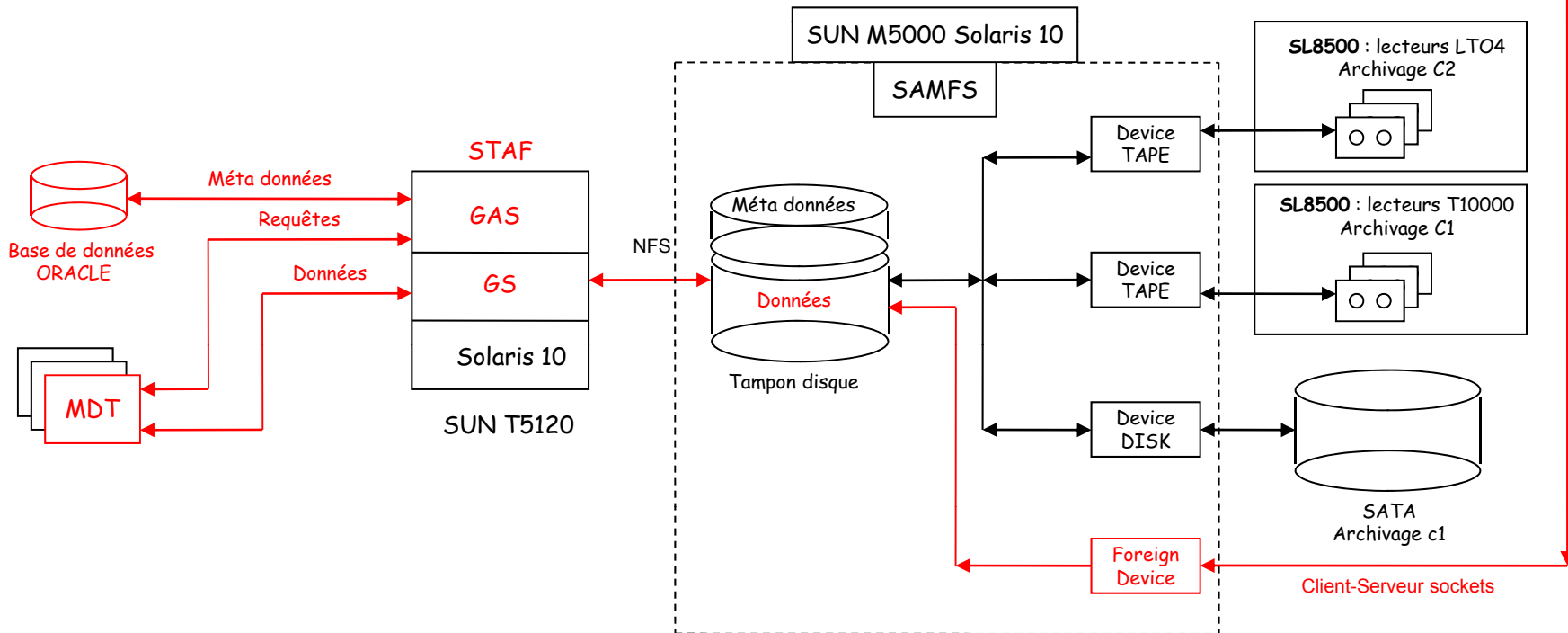
Les phases de migration : Mécanismes de la migration

STAF version 2
hors production



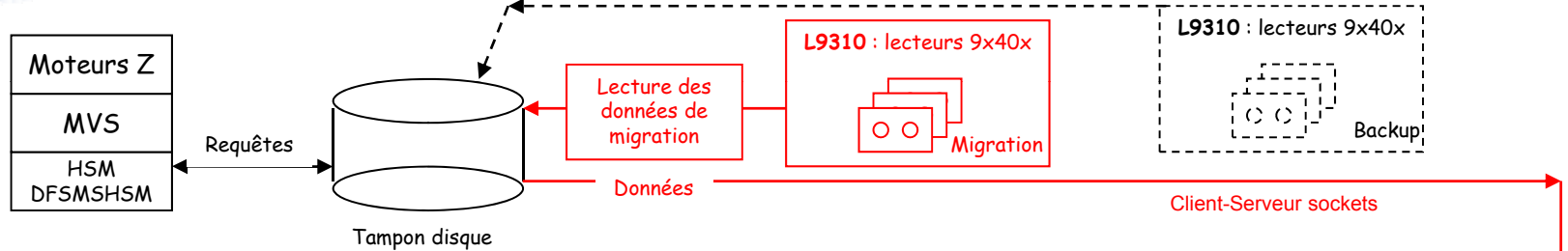
STAR version 3
en production

Migration « aléatoire » générée par la production.



Les phases de migration : Mécanismes de la migration

STAF version 2
en production



STAR version 3
en production

Migration « continue »
A partir de la liste des bandes et de la date de leur dernier accès :

