# Retour d'expérience, dimensionnement des scénarios d'évolution

Loredana FOCSA
Jean-Marc NICOLAS
Henri MEURDESOIF
Sylvain NEUT

UMS ICARE

# SOMMAIRE

Analyse des besoins et retour d'expérience

Solutions techniques et évolutions

# Analyse des besoins

**CAPACITE**

*A un besoin de stockage une solution adaptée*

- *Faire évoluer facilement et maintenir le système de stockage*
- *Temps d'accès et délai d'obtention des fichiers*
- *Contrainte de disponibilité*
- *Croissance des données avec une augmentation annuelle*
- *Migration technologique*

**PUISSANCE DE CALCUL**

- *Traitement massif de données*
- *Moyens de calcul pour les utilisateurs (cluster)*
- *Moyens de calcul web (traitements, post-traitements, interopérabilité)*

**PROTECTION / RECUPERATION**

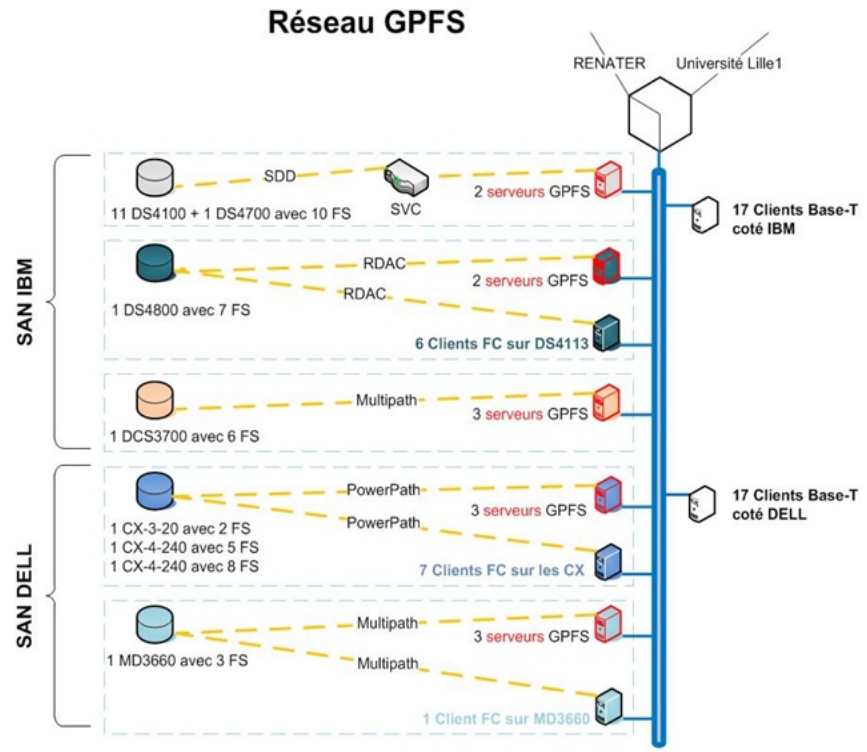
*Il suffit d'une fois ...*

- *Fiabilité de la sauvegarde et des médias pour réduire les risques de pertes de données*
- *Adéquation des ressources : s'assurer que tout est sauvegardé et peut être restauré dans le temps imparti*
- *Reporting et contrôle des ressources*

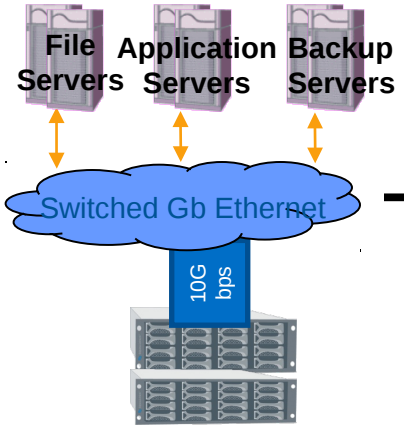
**Qualité des services**

- *Améliorer la qualité et la réactivité du service*
- *Déclencher, analyser et traiter les alertes en cas de problème*
- *Surveillance et monitoring*

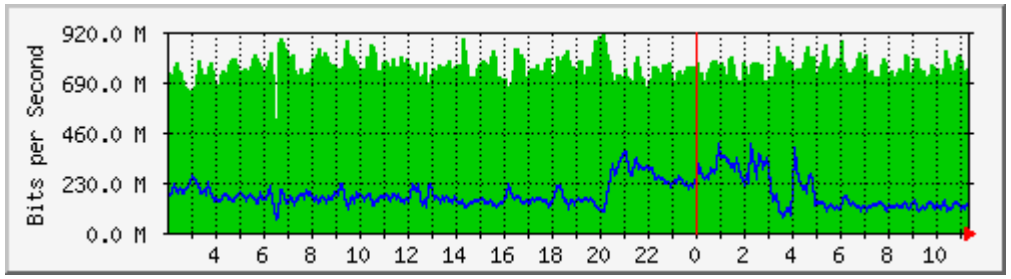
# Évolution du réseau Ethernet




- ***Étape obligatoire pour améliorer les vitesses d'accès malgré l'augmentation du volume, des serveurs, des traitements et des utilisateurs***
- ***Pour répondre à des nouveaux besoins***
- ***Stratégie de répartition de charge et de flux***




***Switchs 10Gb Ethernet pour la consolidation du réseau de stockage***




- ***Saturation des serveurs GPFS en émission***
- ***Architecture gourmande en bande passante***

- ***Multiplier les serveurs*** (coût licences) et ***les passer en 10G***
- 2013 - *Upgrader le coeur du réseau* avec des switch stackables et passer les 6 serveurs GPFS + 2 chassis blade au 10Gb
- 2014 – Passer en 10Gb les autres serveurs

# Alimentation électrique et climatisation

***Disponibilité des services***
***Résistance aux pannes***
***Mesurer et maîtriser la consommation énergétique***

- ***Surveillance de l'environnement physique***
- ***Disposition des matériels dans la pièce***
- ***Tolérance aux petits « pics » de chaleur***

***Alimentation électrique***

→ ***Pas de changement prévu à court ou moyen terme***

Pour améliorer la disponibilité des services par rapport aux coupures de courant :

- *Amélioration / évolution* ***des onduleurs :*** Les *deux onduleurs* 100kVa du CRI doivent être remplacés par un onduleur 400kVA
- *Acquisition d'un* ***groupe électrogène*** : qui maintiendrait l'alimentation des serveurs, des baies et de la climatisation pendant des coupures de courant supérieures à 15 min (coûts des travaux électriques correspondants + coût de la location du groupe électrogène). Cette solution n'a pas été retenue pour l'instant.

***Climatisation***

- ***La température de la salle ICARE reste un peu supérieure à la température des salles de calcul du CRI***

- Améliorer la régulation des clims installées en 2009

• Un couloir chaud sera mis en place en 2014 afin de diriger l'air chaud provenant des baies et des serveurs vers le faux-plafond (<avant-arriere> couloir chaud <arriere-avant>)

***Bénéfices en termes économiques et éco-responsables***

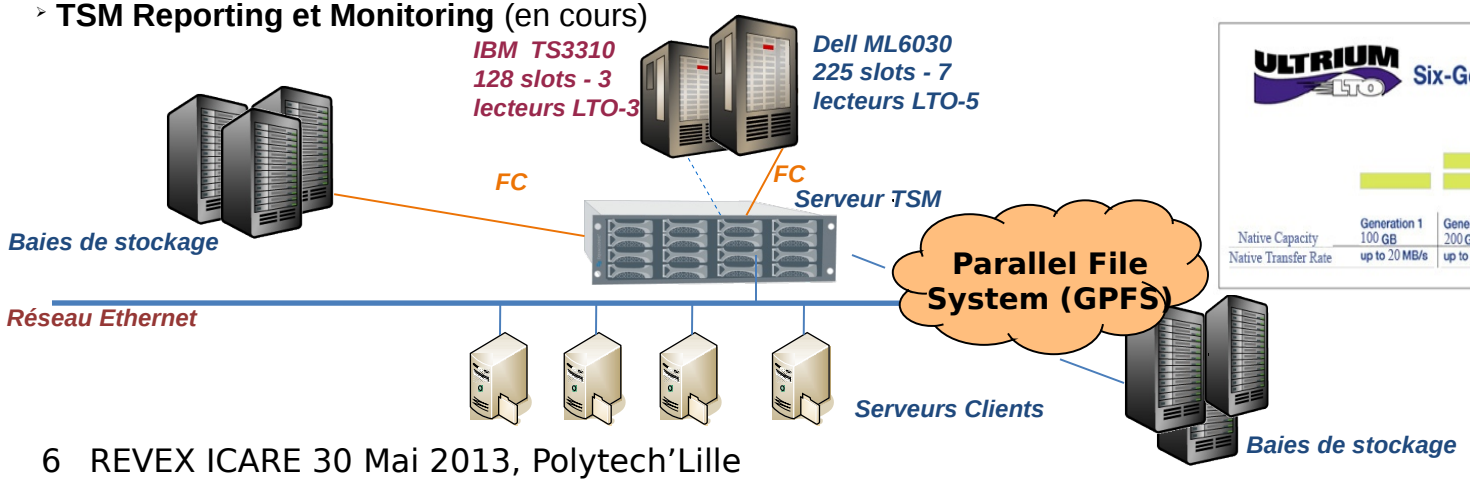
• Baisse de la consommation électrique
• Le refroidissement représente entre 4% et 6% de la consommation des serveurs !
L'installation d'un couloir froid au CRI a permis une réduction de la consommation électrique du Centre de l'ordre de 6%.

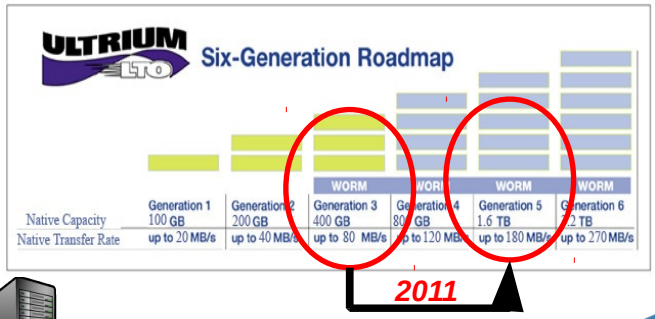
# Évolution du système de sauvegarde

***Fiabilité***
***Respect de performances***

- ***Le volume accumulé et l'explosion des volumes à sauvegarder dans quelques années***
- ***Contrôle et confirmation de la bonne exécution de sauvegardes (tableau de bord, reporting)***
- ***Amélioration et prédictibilité des durées de sauvegarde et restauration***
- ***Évaluation et identification des données à sauvegarder***
- ***Complexité des opérations***

- ***Adéquation des ressources*** (nombre de lecteurs et de la capacité sauvegarde/stockage) - extension de deux librairies *337,5 To → 1.2 Po* ***pour la fiabilité de la sauvegarde et des médias***. Actuellement 656 LTO5 (984 To sauvegardés) avec une *rétention infinie.*
- ***Mise à jour de la configuration TSM***
- ***Amélioration des débits de sauvegarde*** pouvant réduire les fenêtres de sauvegarde ou permettre d'accroître la charge dans les fenêtres existantes (25-30 MB/s actuellement → ~ 6h pour sauvegarder 500 Go/jour). Ce taux peut varier fortement suivant la charge sur le réseau !!
- Les données sauvegardées vont aussi influencer les temps de sauvegarde. Par ex, pour un même volume donné, un seul fichier sera sauvegardé plus rapidement que des dizaines de milliers.
- **Utilisation des commandes optimisées** GPFS "mmbackup" et hiérarchisation des données (l'importance et la criticité)
- **TSM Reporting et Monitoring** (en cours)




***50 To → 337,5 To***

# Moyens de stockage et de calcul dédiés aux utilisateurs

***Disponibilité***
***Fiabilité***
***Réactivité***

- ***Des accès sécurisés web, ftp et ssh***
- ***Des ressources partagé : puissance de calcul, de stockage, réseau***
- ***Des applications et du middleware : logiciels et outils qui évoluent à la demande***
- ***Hosting data and services***
- ***Sauvegarde journalière avec une rétention de 100 jours***

- Upgrade système de la machine access (passage en Redhat 6.3) prévu en 2013
- Ouverture du cluster basé sur Torque / Maui pour soulager le serveur access ( 2 machines 32 bits ==> 24 GB RAM + 20 cœurs et 2 en 64 bits ==> 64 GB RAM + 32 cœurs)
- Possibilité de mettre en place des machines supplémentaires en fonction de la charge observée

→ équilibrage des ressources RAM et CPU
→ gestion des files d'attente

- Ce cluster est déjà en test au CGTD + quelques utilisateurs externes
- Mise à disposition des utilisateurs d'un espace de ***12 To*** (début juin) ==> ***50 To*** (fin 2013) avec une organisation à définir (proportion, quotas, etc...)
- Engagement d'une ouverture des accès ftp/ssh en 3 jours ouvrés

➔ Environnement logiciel centralisé qui peut évoluer à la demande :

Compilateurs : C, C++, Fortran,suites Intel et Pgi
Bibliothèques pour la manipulation des données : hdf et netcdf
Logiciels propriétaire : IDL, MATLAB
Logiciel open source de calcul numérique : Scilab, Octave etc...
Logiciels graphiques : Ferret, Opengrads, Gmt, Ncar graphics etc...

# L'évolution du stockage et des ressources de calcul

***Fiabilité***
***Performances***
***Puissance de calcul***

→

- ***Évoluer au rythme de l'activité***
- ***Efficace et souple***
- ***Prêts pour les technologies de demain***

- Possibilité ***d'augmenter la capacité de stockage et de calcul*** et d'ajouter des éléments physiques supplémentaires
- ***Croissance stable pour les deux dernières années -*** incrément ***de ~ 300 To/an***
- ***Un flux de données de 500 Go/ jour actuellement évoluant vers 1To / jour en 2015 avec l'arrivée des nouvelles missions***
- ***En cas de retraitement le flux peut augmenter jusqu'à 3To / jour***
- ***Pour les prochaines années - acquisition de 20 serveurs de type X86, bi-processeurs multi-coeurs supplémentaires par an (144 cœurs/an au minimum)***
- ***Nécessité de remplacer les matériels vieillissants acquis entre 2006 -2009 (120 To à migrer en 2013 et 150 To à partir du 2014)***
- ***Des technologies concurrentes FC vs. Ethernet***

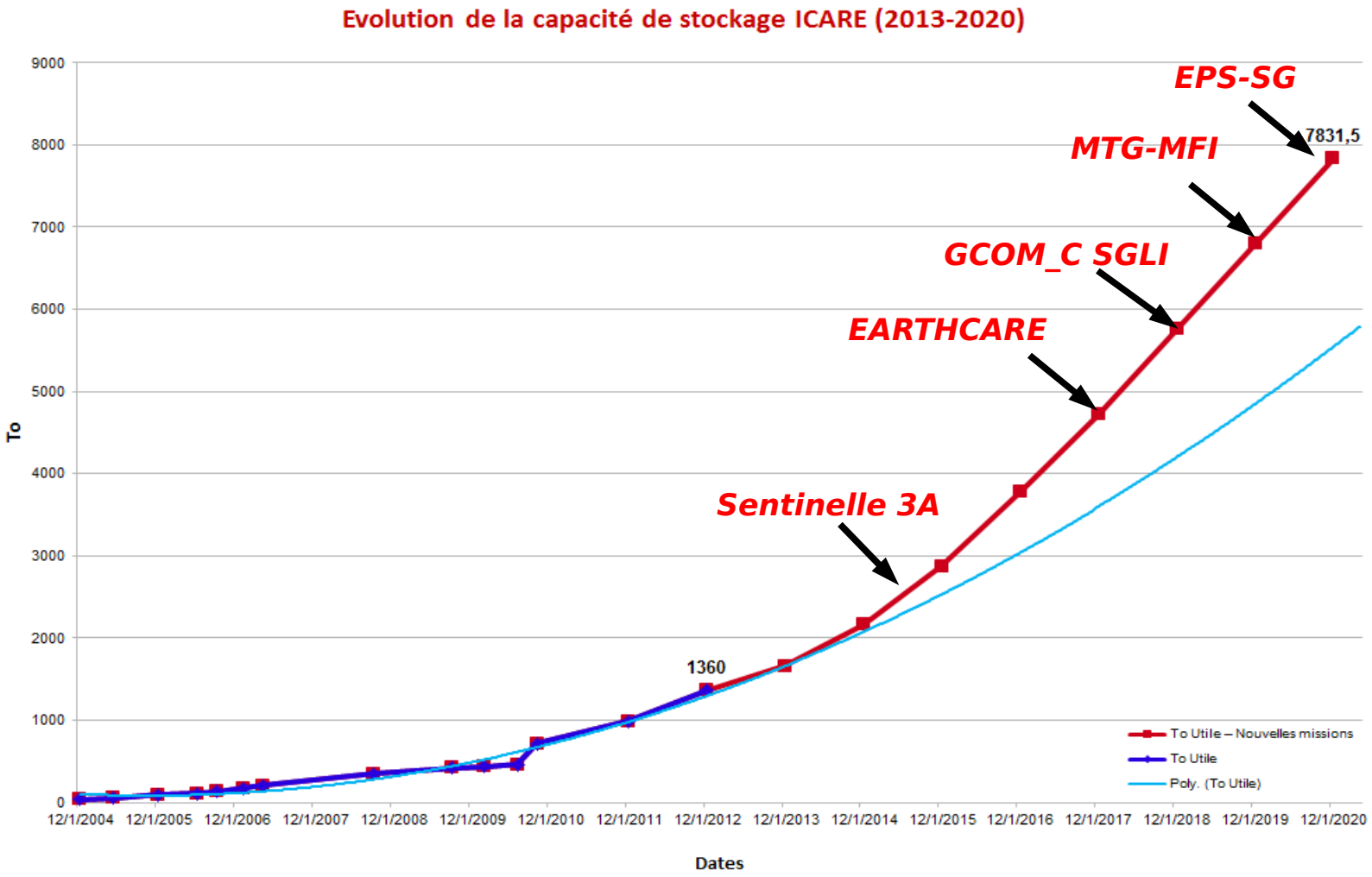
  ***FC*** (*Fibre Channel*) (2, 4, 8 Gb/s) → **16 Gb/s**

  ***FCoE*** (*Fibre Channel over Ethernet*) (10 Gb/s) (simplification de l'architecture, et réduction des coûts (une seule infrastructure d'interconnexion au lieu des deux existantes).

  ***iSCSI*** (SCSI over Internet Protocol) (10 Gb/s) (protocole qui autorise le transfert de blocs de données sur des réseaux Ethernet via des adresses IP - nécessite un réseau ethernet dédié)
- ***Évolution au rythme de l'activité***

**Stockage : ~6000 To ? (prévision 2020)**
**Puissance de calcul: ~520 cœurs (fin 2013)**
**~2500 cœurs ? (prévision 2020)**

# Une capacité de stockage en progression



# Prêts pour accompagner la croissance

## *Upgrade réseau - étape obligatoire pour assurer la croissance*

**Assurer les débits importants, la qualité de la bande passante et les temps de latence faible c'est investir dans le réseau** !!

- Dépense supplémentaire non budgétée pour 2013 mais nécessaire (~30-40K€ ) et (~40K€ ) en 2014

## *Stockage: un saut technologique ?*

**D'ici 2015 nous devrons multiplier par deux notre capacité de stockage et d'augmenter également notre puissance de calcul**

- Le modelé actuel ICARE (espace de stockage physique hétérogène, reposant sur le filesystem partagé GPFS - quelques problèmes de jeunesse, aujourd'hui oubliés !) permettrait cette croissance sans changement technologique
- Le changement d'échelle prévu sur la période 2016-2020 (multiplier par 6), pourrait introduire des problématiques (hétérogénéité de ressources, tolérance aux pannes etc.) ==> ***un saut technologique est peut-être à prévoir dans deux ans***
- Choix technologique en mettant en jeu les acteurs majeurs dans le domaine

## *Sauvegarde*

- Concilier l'augmentation du volume et la fiabilité du système de sauvegarde
- Extension de deux librairies te révision du système de sauvegarde ==> **priorité 2013**
- Dépense déjà budgétée pour 2013 (~60K€ )

## *Climatisation*

**Concevoir des techniques de refroidissement appropriés ==> objectif 2014**

- Une meilleure régulation de la température ambiante dans la salle ICARE
- Baisse de la consommation électrique
- Mesurer en permanence la température (points chauds, etc.) pour adapter la puissance de la climatisation