# MIGRATION DU STAF

Service de Transfert et d'Archivage des Fichiers



#### Le STAF

#### Fonction « stockage » de l'OAIS

Dispense les projets spatiaux de la gestion des supports de stockage



#### Évolution de la volumétrie

2000 : quelques Téra-Octets

2010 : ~ 400 Téra-Octets

2020 : ~ 4 Péta-Octets

2030 : plusieurs dizaines de Péta-Octets



# **Évolution de l'architecture du STAF**

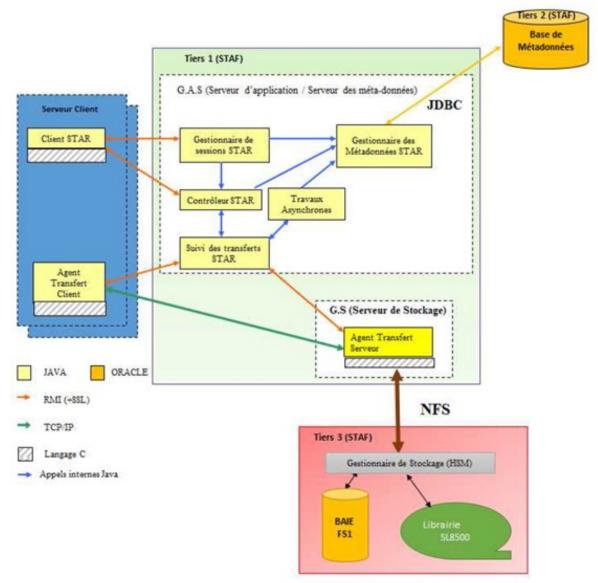
Années	1995 - 1998	2001 – 2008	2009 – 2015	2016 – 2023
Technologie bandes	T10K	T10K	T10K + LTOx	T10Kc + LTO6
Technologie de cache	DFS-HSM (IBM)	DFS-HSM	SAMFS (Oracle)	SAMFS
Serveur(s) de production	Mainframe	Mainframe	SPARC	HP(X86) + SPARC
Système d'exploitation	IBM	IBM	Solaris	Linux + Solaris
Applicatif (version)	V1	V2	V3	V3
Langage applicatif	#C/TCP	#C/TCP	JAVA/RMI	JAVA/RMI



## Évolution de l'architecture du STAF

#### STAF V3

- Application client serveur
- Abandon des technologies IBM
- Conservation de l'API STAF V1 & V2
- Séparation entre « gestionnaire d'application » et « gestionnaire de stockage »
- Couplage faible entre application et infrastructure de stockage (montage NFS)
- SAMFS (ORACLE HSM)
- Librairies SL8500





# **Problématique**

#### Obsolescence des composants du STAF V3

- Fin de maintenance des librairies SL8500 (fin 2023)
- Arrêt de production des cassettes T10K
- Obsolescence SOLARIS
- Obsolescence de l'application

#### Bilan de 25 ans d'utilisation du STAF

- Pas de perte de données
- Certaines données uniques au monde
- MAIS
- Utilisation parfois dévoyée (sauvegarde)
- Données inaccessibles (perte des catalogues des Centres de Mission) Absence de métadonnées « métier »
- Difficulté de déploiement de l'application cliente



#### Du STAF V3 au STAF V4

#### **Principes**

- Conservation de l'architecture STAF V3
  - ✓ API d'accès au STAF indépendante de la technologie de stockage
  - ✓ Couplage faible entre l'application STAF V4 et l'infrastructure de stockage
- Abandon du client déployé sur les machines des projets utilisateurs
- Présence systématique d'un catalogue des produits
- Ouverture sur l'Internet
- Recentrage du STAF sur sa fonction de pérennisation des données issues des missions spatiales
  - ✓ Profiter de la migration pour faire du ménage

#### **Briques techniques**

- Infrastructure de stockage : DATALAKE
- Applicatif : REGARDS



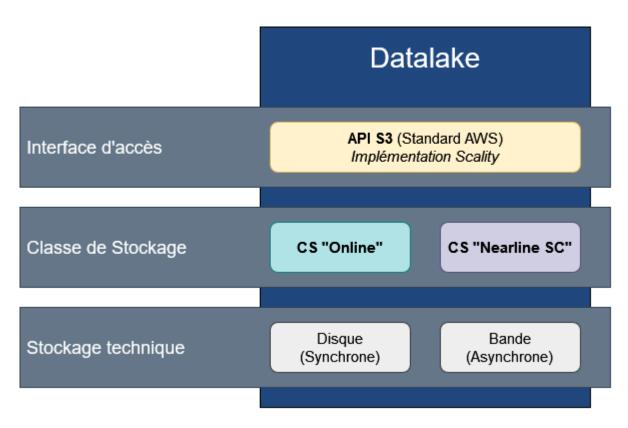
# Qu'est-ce que le DATALAKE ?

#### DATALAKE : le système de stockage du CNES pour les volumes massifs de données

- ❖ Capacité : 35 Péta-Octets sur disque à la mise en service 70 Péta-Octets à terme Évolutif > 100 Péta-Octets
- Stockage objet
- API « Simple Storage Service » (S3)
- Plusieurs classes de stockage :
  - ✓ Online
  - √ Nearline SC (simple copie)

## Une classe de stockage réservée STAF V4

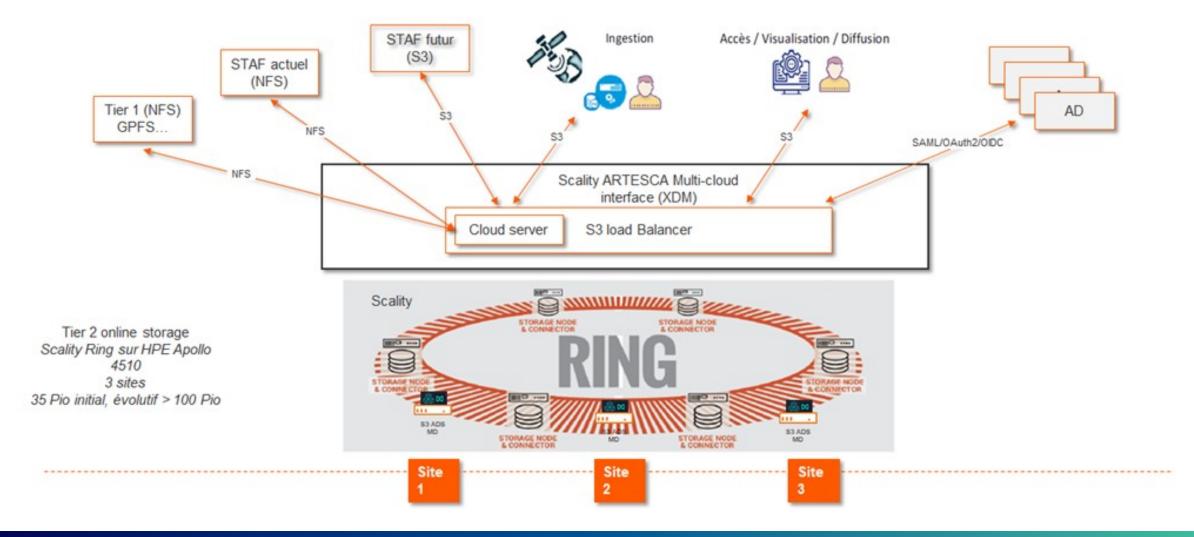
- DATALAKE « glacier »
- Tampon disque
- Double copie sur bandes (dans 2 bâtiments)





# Qu'est-ce que le DATALAKE ?

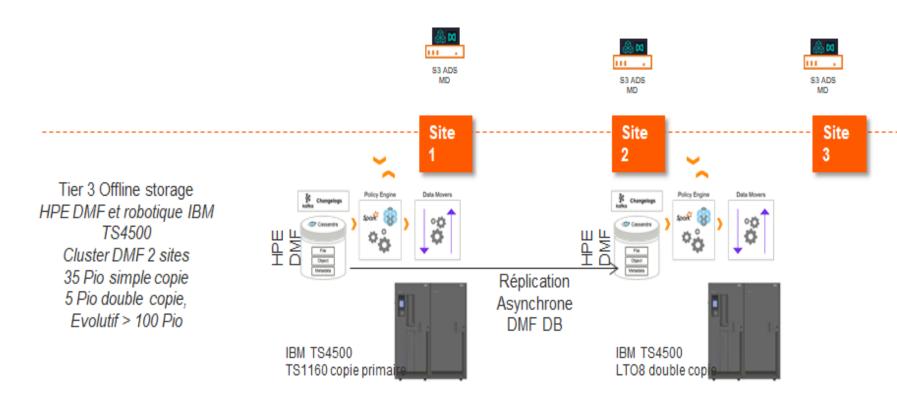
## Infrastructure disque





# Qu'est-ce que le DATALAKE ?

#### Infrastructure bandes

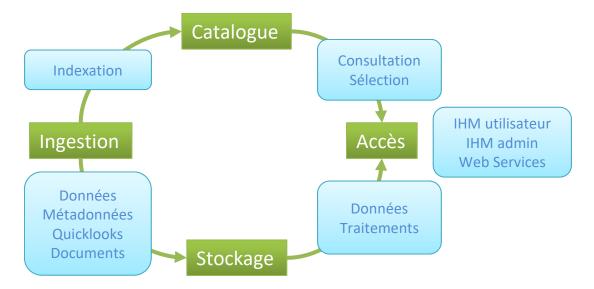




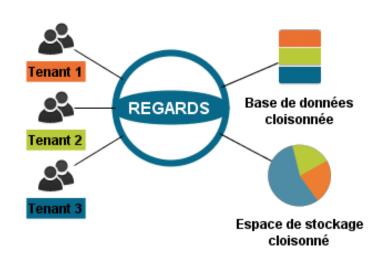
# **Qu'est-ce que REGARDS?**

#### REnouvellement des outils Génériques d'Accès et d'aRchivage pour les Données Spatiales

- Framework (cadriciel) adaptable et paramétrable
- Accès catalogue aux produits des Centres de Missions ou de Données CNES et partenaires
- Logiciel libre (licence GPL V3)
  - √ <a href="https://github.com/RegardsOss">https://github.com/RegardsOss</a>
  - √ https://regardsoss.github.io/
- Fonctions d'archivage basées sur l'OAIS

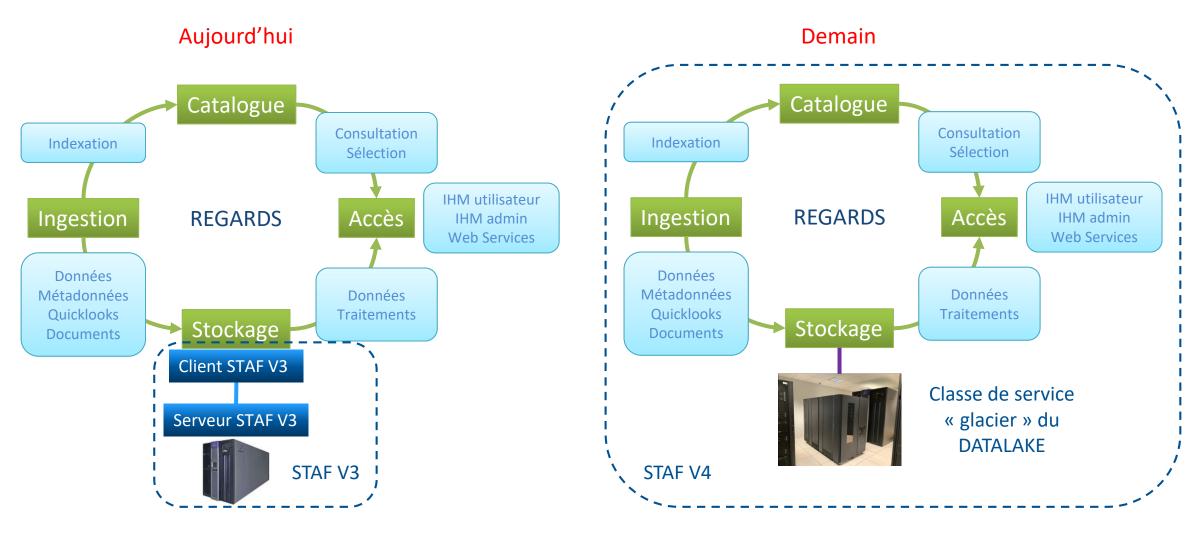


Système « multi-tenants »



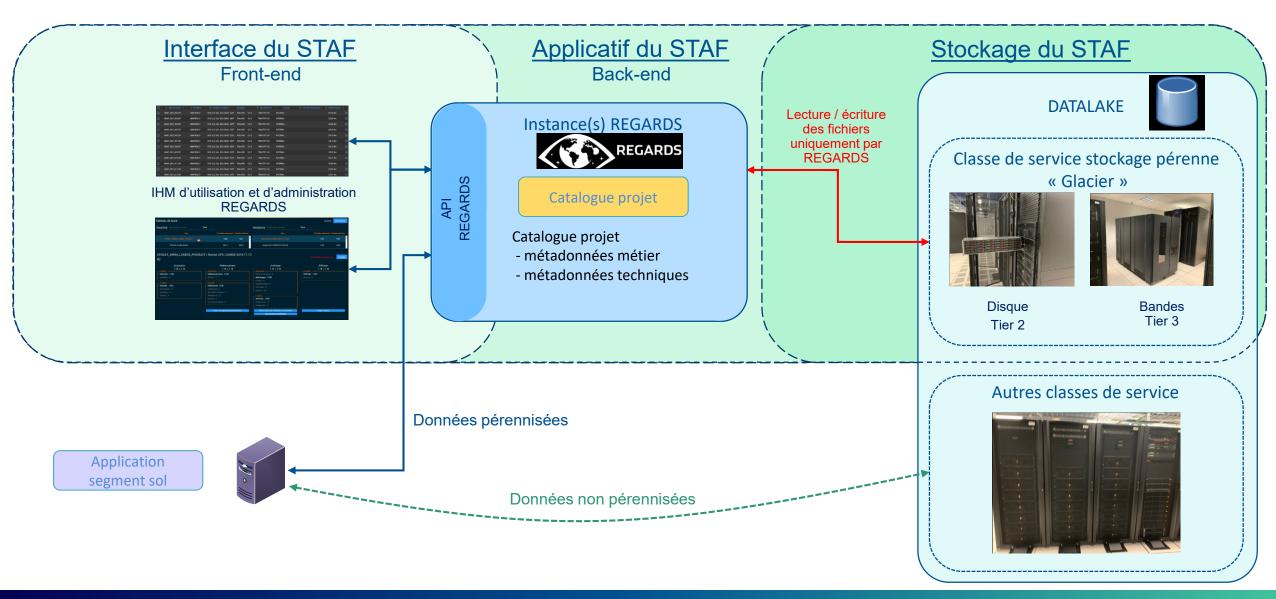


#### **Architecture du STAF V4**





# **Architecture du STAF V4**





# **Enjeux de la migration**

#### Cas 1: il existe un catalogue REGARDS

- Exemples : SWH, CDPP
- Restitution sur un espace disque des fichiers stockés dans le STAF V3 et référencés dans le catalogue
- Transfert par REGARDS des fichiers dans le DATALAKE « glacier »

### Cas 2 : il n'existe pas de catalogue REGARDS

- En créer un
- Se ramener au cas 1



# **Application à l'archive SWH**

Aujourd'hui les données L1A Spot World Heritage sont stockées sur le GPFS de façon transitoire, dans l'attente de la disponibilité du DATALAKE.

La migration SWH se fera donc directement du GPFS vers le DataLake.



Lac de Crop près de Grenoble

Etat actuel : produits L1A stockés sur l'espace GPFS

Première étape : déplacement sur le « tier 2 » du DATALAKE

Stockage objet avec forte résilience : 3 salles indépendantes

Deuxième étape : archivage pérenne

Prérequis : nouvelle infrastructure STAF et nouvelle application cliente STAF (basée sur REGARDS) disponibles

A priori, les produits L1A demeureront en ligne sur le tier 2 pour permettre un accès rapide (besoin d'y accéder pour la valorisation.)