

JoSy Stockage S3

Rapport sur les contributions

ID de Contribution: 1

Type: **Non spécifié**

Introduction de la JoSY S3

mardi 7 octobre 2025 14:00 (15 minutes)

ID de Contribution: 2

Type: **Non spécifié**

NFS pour calculer, S3 pour partager : moderniser la diffusion des résultats scientifiques

mardi 7 octobre 2025 14:15 (45 minutes)

Notre infrastructure informatique mutualisée, au service des plateformes de recherche en génomique (séquençage NGS) et métabolomique, repose largement sur une baie PowerScale (NFS) pour la gestion des workflows de calcul et de traitement des données.

Ces traitements produisent de gros volumes de résultats de plusieurs To, destinés à être partagés avec des partenaires externes (collaborateurs, cliniciens, partenaires industriels).

Historiquement, ce besoin était couvert par un serveur Fex, qui permettait le dépôt temporaire de fichiers volumineux via des liens de téléchargement. Ce service vieillissant et non sécurisé ne répondait plus aux attentes des utilisateurs, ni aux exigences actuelles (RGPD, ergonomie, traçabilité).

Nous avons donc choisi de le remplacer par un service transverse S3 basé sur MinIO, en complément de notre infrastructure NFS existante.

Ce nouveau service permet aux plateformes de générer des liens de diffusion temporaires et sécurisés, sans exposer directement le stockage NFS, tout en s'intégrant dans les workflows de traitement.

Notre retour d'expérience portera sur :

- Le besoin initial : pourquoi et comment nous avons remplacé Fex
- Pourquoi NFS seul ne suffisait pas pour la diffusion externe
- L'architecture déployée : VM Proxmox + MinIO + reverse proxy HTTPS, stockage local provisionné via PowerScale
- L'intégration dans les workflows : automatisation du push vers S3, génération de liens, gestion du cycle de vie
- Les bénéfices observés : ergonomie, maîtrise locale, sécurité et conformité RGPD, interopérabilité avec les partenaires
- Les limites et perspectives : performances, politique de cycle de vie, scénarios d'évolution

Ce projet illustre comment un couplage NFS pour le calcul + S3 pour le partage permet de moderniser simplement et efficacement la diffusion de résultats scientifiques dans un environnement de recherche mutualisé.

Orateur: Mme LEDOUX, Anne-Sophie (Head of IT Infrastructure and data security - UMR1283 - 8199)

ID de Contribution: 3

Type: **Non spécifié**

Ceph S3 : Du Placement Dynamique à la Rétention Optimisée.

mardi 7 octobre 2025 15:00 (45 minutes)

Au cours de cette présentation (éventuellement accompagnée d'une démo), nous verrons comment optimiser les performances et l'efficacité de stockage des objets S3 dans Ceph, en tirant partie des classes de stockage S3, de manière synchrone dès l'accueil des données (LUA) et de manière asynchrone pour traiter les données les plus anciennes (LifeCycle Policies).

Type : Il s'agit d'une présentation technique.

Biographie: Frédéric est ambassadeur du projet communautaire Ceph pour la France et Ingénieur Ceph chez Clyso GmbH. Il opère du conseil, du support et du dépannage sur les clusters Ceph de clients privés et institutionnels à travers le monde. Ingénieur civil de formation, il s'est reconverti dans l'informatique et a accumulé 20 ans d'expérience dans le secteur académique français, se spécialisant sur les technologies de la virtualisation, du stockage et la conteneurisation. Frédéric possède 10 années d'expérience sur Ceph, gérant des pétaoctets de données diverses, incluant des jeux de données scientifiques, des fichiers utilisateurs, des e-mails, des machines virtuelles et des volumes de conteneurs. Frédéric intervient lors de conférences nationales et internationales, comme au Red Hat Summit en 2018. En plus de son rôle principal, Frédéric dispense des formations Ceph et Kubernetes aux étudiants en tant qu'enseignant universitaire à temps partiel.

Orateur: M. NASS, Frédéric (Université de Lorraine)

ID de Contribution: 4

Type: **Non spécifié**

Artesca + Veeam : La combinaison gagnante 2 en 1 pour des backups immuables et un budget maîtrisé”

mardi 7 octobre 2025 16:00 (45 minutes)

Scality, leader en solutions de stockage cyber-résilientes, lance ARTESCA+ Veeam, une nouvelle appliance logicielle unifiée issue de sa collaboration avec Veeam® Software. Cette innovation intègre les logiciels Veeam Backup & Replication™ et Scality ARTESCA –un stockage objet cyber-résilient –sur un unique serveur.

Cette solution simplifie drastiquement les infrastructures en éliminant le besoin d’une infrastructure Veeam séparée, qu’elle soit physique ou virtuelle. Cela se traduit par une réduction allant jusqu’à 30 % des coûts, du temps et de la complexité de déploiement. Les organisations peuvent choisir de la déployer sur leurs serveurs existants (HPE, Supermicro, Lenovo). ARTESCA+ Veeam vise à renforcer la cyber-résilience des entreprises et à optimiser leurs sauvegardes face aux menaces actuelles, tout en s’intégrant dans la gamme polyvalente de Scality. Elle combine la sécurité et la simplicité d’ARTESCA avec la résilience des données de Veeam pour des défenses robustes sans compromettre la performance.

L’offre ARTESCA+ Veeam présente des bénéfices significatifs pour les institutions académiques et les centres de recherche :

Déploiement et gestion simplifiés : Les équipes informatiques peuvent rapidement et facilement configurer la sauvegarde et le stockage, assurant une cyber-résilience de bout en bout pour les données sensibles de la recherche et de l’administration. Cela libère du temps pour des tâches à plus forte valeur ajoutée.

Sécurité accrue des données de recherche : L’hébergement de Veeam et Scality ARTESCA sur une appliance “hardened” basée sur les principes “zero trust” garantit une protection optimale contre les cyberattaques, essentielle pour la propriété intellectuelle et les données de recherche confidentielles.

Protection renforcée contre les menaces : La réduction des vecteurs d’attaque grâce à l’exposition limitée des identifiants et l’intégration d’un pare-feu sécurisant les composants Veeam protègent efficacement les bases de données étudiantes et les projets de recherche contre les intrusions. L’accès sécurisé via MFA pour l’administration Windows renforce encore cette sécurité.

Intégrité des données scientifiques garantie : Les contrôles d’intégrité non disruptifs (“SureBackup Lite”) permettent d’analyser le contenu des sauvegardes indépendamment, assurant la validité et la fiabilité des données sans impacter les systèmes de production, crucial pour la continuité des projets de recherche.

Optimisation des coûts d’infrastructure : En éliminant le besoin d’une infrastructure serveur et de stockage séparée, cette solution réduit significativement les coûts d’acquisition, de support et d’exploitation. C’est un atout majeur pour les budgets souvent contraints de l’enseignement supérieur et de la recherche.

Gestion centralisée et simplifiée : Un tableau de bord unique pour superviser ARTESCA et Veeam simplifie les opérations quotidiennes, permettant aux équipes informatiques de maintenir efficacement les environnements sans complexité inutile.

Portrait Peter LONG :

Peter Long est un expert en architecture système et stockage, notamment en stockage objet, avec plus de vingt ans d’expérience. Actuellement Senior System Architect chez Scality, il y conseille les clients ayant les projets les plus critiques et complexes en France et EMEA. Auparavant, chez

Cloudian Inc., il a été Principal Architect et Technical Sales Engineer, gérant les relations techniques et partenaires. Il a également passé plus de six ans chez Dell en tant qu'Architecte Principal et architecte infrastructure de stockage. Son parcours révèle une expertise constante dans la conception d'infrastructures complexes et une solide capacité à bâtir des relations client, avec un focus croissant sur le stockage objet.

Orateur: M. LONG, Peter (SCALITY)

ID de Contribution: 5

Type: **Non spécifié**

Table ronde "se faire accompagner dans son projet de stockage S3 et sa gestion du quotidien"

mardi 7 octobre 2025 16:45 (45 minutes)

Orateurs: M. SAILLARD, Christophe (IRMA Strasbourg); M. DUPONT, Yann (Université de Nantes)

ID de Contribution: 6

Type: **Non spécifié**

Retour d'expérience : héberger une application Vue.js via des buckets S3 configurés en mode « website »

mercredi 8 octobre 2025 08:45 (45 minutes)

Auteure : Pauline Gilg

Université de Strasbourg, Direction du numérique, Services métiers, Développeuse front end au sein du pôle Développement, intégration et paramétrage (DIP)

Pauline Gilg est développeuse front end, avec une formation en linguistique, spécialisée en études japonaises et en localisation. Elle a évolué dans le domaine de la communication numérique en tant

que webmaster avant de se spécialiser dans le développement web. Ses intérêts portent principalement sur l'expérience utilisateur et l'accessibilité numérique, avec une attention particulière aux standards du web et à la conception d'interfaces inclusives. Elle travaille également

sur des problématiques de déploiement, en mettant en place des pipelines d'intégration continue (CI) et des processus d'automatisation pensés pour être facilement utilisables, notamment dans des

contextes multi-locataires ou institutionnels. Son parcours croise ainsi des compétences en sciences

humaines et en technologies du numérique, qu'elle mobilise dans le cadre de projets orientés utilisateur.

Résumé

Campulse est une plateforme web développée par l'Université de Strasbourg à destination des associations étudiantes et des Services de la Vie Universitaire (SVU), centralisant annuaire, signature de chartes, demandes de subventions et suivi administratif. Financé dans le cadre du programme Services numériques aux étudiants (France Relance), le projet est en cours de déploiement dans 11 universités et suscite un intérêt national.

L'architecture repose sur un découplage strict entre front end et back end. Le front, développé en Vue.js, consomme une API REST exposée par un back end Django. Ce découplage a permis de faire un choix d'hébergement spécifique pour l'interface : le mode static website hosting de S3, optimisé pour la diffusion de fichiers HTML, CSS et JavaScript.

L'objectif principal : permettre à chaque université d'adapter l'interface à son identité (images, couleurs, textes, polices), de manière totalement autonome, tout en conservant une solution automatisée, scalable et maintenable. Cette personnalisation est injectée à la compilation via des scripts de parsing et des variables, puis buildée en une instance front end autonome.

Une CI (intégration continue) basée sur GitLab CI orchestre le processus :

- Build de l'application en fonction des personnalisations,
- Création à la volée du bucket S3 si besoin (un par environnement : préproduction et production),
- Déploiement automatisé des fichiers statiques.

Chaque bucket S3 est configuré en mode « website » pour servir directement les fichiers, avec :

- Définition explicite des MIME types pour assurer le bon rendu dans les navigateurs,
- Paramétrage des Access Control Lists (ACL) pour exposer uniquement les fichiers requis.

L'accès public aux interfaces est exposé via des objets VirtualService gérés par Istio et déployés sur

l'infrastructure cible.

Ce choix technique présente plusieurs avantages :

- Scalabilité horizontale : chaque établissement dispose de son propre bucket déployé automatiquement, sans incidence sur les autres.
- Sécurité : aucun accès SSH nécessaire ; seules des clés d'accès S3 sont utilisées, réduisant la surface d'attaque.
- Simplicité d'administration : une console S3 permet un accès rapide à la liste des buckets et à leur contenu.
- Pertinence technique : les fichiers buildés sont statiques et non sensibles.

Ce retour d'expérience propose donc un modèle d'hébergement statique, découplé, sécurisé et multi-locataire, particulièrement adapté aux interfaces front end dans un contexte public ou institutionnel. L'approche allie automatisation, personnalisation et robustesse, tout en s'intégrant dans des pipelines DevOps modernes.

Orateur: Mme GILG, Pauline (Université de Strasbourg)

ID de Contribution: 7

Type: **Non spécifié**

Le stockage S3: Un exemple pour la visualisation de cartes interactives

mercredi 8 octobre 2025 09:30 (45 minutes)

L'imagerie satellitaire est un gros producteur de données utilisé dans de nombreux domaines d'activités. Un satellite comme Pléiades Neo, lancée en 2021, peut produire des images de la Terre à 30 centimètres de résolution, de quoi crée une mosaïque globale de 1.5 peta-pixels sur l'ensemble du globe; qu'il faut régulièrement mettre à jour pour suivre l'évolution de notre planète. Manipuler de tel volume de fichier (~15 Po/annuel) impose d'utiliser des technologies de stockage particulièrement efficace pour héberger l'archive historique et la mettre à disposition des utilisateurs. Pour faciliter la manipulation de ces cartes, tout particulièrement pour les appareils mobiles, les mosaïques globales doivent être découpées en petites tuiles de quelque kilo-octet pour n'envoyer à l'utilisateur que la sous-partie de l'image qui lui est pertinente. Historiquement, cette étape de tuilage était effectuée soit, à la demande par un serveur de carte (WMS) sous la forme d'un middleware, qui devait absorber en direct la charge de tous les utilisateurs; soit en amont et de manière systématique en hébergeant des milliards de tuiles de manière statiques (WMTS), typiquement sur un stockage objet de type S3. En 2019, l'Open Geospatial Consortium (OGC) a introduit un nouveau standard, le Cloud-Optimized Geotiff (COG) permettant de combiner le meilleur des deux mondes. Moyennant une structuration préalable du fichier, il est désormais possible d'utiliser les "HTTP range requests" sur une gateway S3 pour s'affranchir d'un middleware et ne plus avoir besoin de pré-tuiler les mosaïques en milliards de fichiers. Nous verrons dans cet exposé comment les COG permettent de grandement simplifier la génération de cartes interactives.

Orateur: M. SEIGNOVERT, Benoit (Université de Nantes)

ID de Contribution: 8

Type: **Non spécifié**

Garage : un backend S3 aussi robuste que possible

mercredi 8 octobre 2025 10:30 (45 minutes)

Quentin Dufour est co-fondateur de l'association Deuxfleurs et il travaille pour Despite Glitches pour fournir du support sur les logiciels Deuxfleurs (y compris Garage), et plus généralement accompagner les organisations ayant des problèmes de fiabilité (scaling, bug, déploiements, etc.).

Garage a été conçu pour les besoins de l'association Deuxfleurs qui opère dans des conditions complexes : hors datacenter, géo-distribué, avec du matériel peu puissant et non spécialisé. Je présenterai d'abord les choix de conception qui soutiennent ces propriétés (limitation des requêtes consécutives, topologie flexible, maintien des performances en situation dégradée, synchronisation minimisée, single binary, cross-OS et cross-CPU, etc.), je présenterai ensuite des exemples de déploiement de Garage en production ainsi que des retours utilisateurs, enfin je présenterai l'actualité de développement du logiciel, ses fonctionnalités et limitations actuelles.

Orateur: M. DUFOUR, Quentin (DeuxFleurs)

ID de Contribution: 9

Type: Non spécifié

Le stockage objet DataCore Swarm à l'IBMP et plus généralement : comment faire tenir un océan dans une bouteille...

mercredi 8 octobre 2025 11:15 (45 minutes)

La rétention et la sécurisation des données de la recherche sur le long terme nécessitent des systèmes de plus en plus capacitifs, évolutifs et résilients. Dans cette optique, l'IBMP a investi dans la solution commerciale DataCore Swarm, qui a été déployée dans le laboratoire. Nous avons choisi cette solution du fait de sa structure objet, de sa facilité de déploiement et d'administration, ainsi que pour des raisons de coûts. Les premiers usages à ce jour concernent le stockage des grands jeux de données de bio-informatique, mais aussi le stockage des données intégrées au sein du cahier de laboratoire électronique ElabFTW déployé à l'IBMP. Dans une vue plus prospective, nous aborderons les développements logiciels à l'étude au sein du laboratoire afin de donner toute sa puissance à l'architecture objet du stockage, indépendamment de la solution S3, qui est notoirement sous-utilisée et constitue un frein à son adoption par les utilisateurs finaux.

Nous explorerons avec DataCore la problématique de la production de données qui dépasse la capacité de stockage généralement présente sur nos campus. En effet, les jeux de données grandissent exponentiellement, mais les budgets IT sont au mieux stables, s'ils ne baissent pas. Entre financements de recherche contraints et explosion des volumes, l'équation est complexe : comment maximiser la valeur scientifique tout en optimisant les investissements IT ? Comment équilibrer les besoins de performance HPC pour l'analyse active et les stratégies de stockage objet économiques pour l'archivage à long terme ? Enfin, comment éviter que la recherche d'un jeu de données spécifique ne devienne elle-même un projet de recherche ? Nous allons examiner, dans le contexte particulier de la recherche, des architectures HPC haute performance couplées avec le stockage parallèle, les stratégies de tiering intelligent vers du S3 et stockage objet, les modèles économiques hybrides on-premise, les systèmes d'indexation intelligente et l'enrichissement automatique des métadonnées (parce que "grep" a ses limites). Enfin, nous aborderons l'optimisation des coûts via la migration automatique HPC vers objet pour une rétention à long terme rentable afin de laisser l'essentiel des ressources pour la découverte scientifique.

Orateurs: M. EVRARD, Jean-Luc (IBMP); M. LBARS, Laurent (DataCore)