

05. Advanced Amazon S3

Questions :

- C'est quoi Event bridge ?
- C'est quoi SNS / SQS ?
- fef

1. S3 Lifecycle Rules (Politiques de Cycle de Vie)

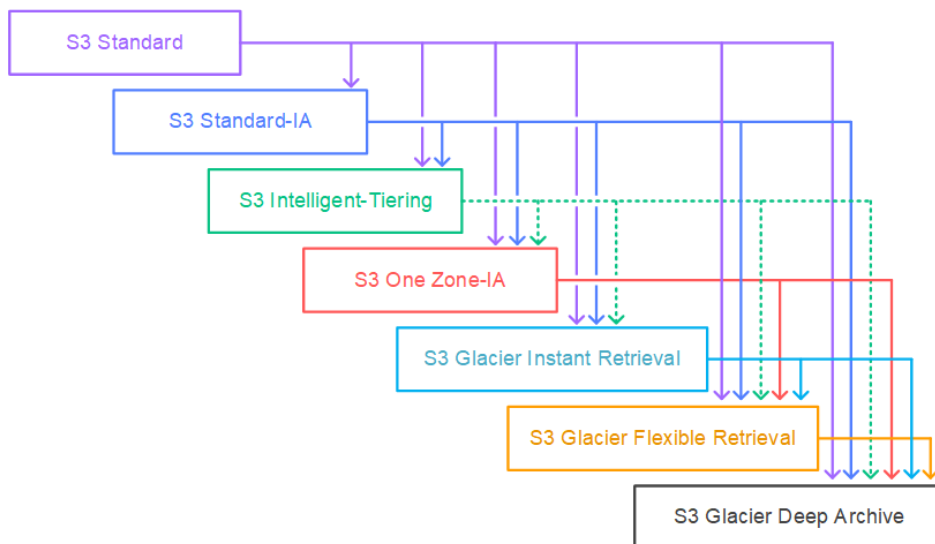
Transition Actions & The Transition Ladder (L'Échelle de Transition)

Pourquoi ? Automatiser la réduction des coûts de stockage au fil du temps sans intervention humaine. Les données vieillissent et deviennent généralement moins "chaudes".

Comment ? : Tu définis des règles pour déplacer les objets d'une classe de stockage à une autre après un nombre de jours précis. Tu ne peux pas transitionner vers le "haut" (du froid vers le chaud). Il existe un ordre strict imposé par AWS.

Remarque : AWS impose des délais minimaux avant de pouvoir passer à la classe suivante.

Exemple : Pour passer de *S3 Standard* à *S3 Standard-IA*, l'objet doit avoir **au moins 30 jours**. Si tu configures une règle à 15 jours, S3 refusera la transition.



Source : <https://docs.aws.amazon.com/AmazonS3/latest/userguide/lifecycle-transition-general-considerations.html>

Expiration Actions

- **Pourquoi ?** Pour supprimer automatiquement les données devenues inutiles (ex: logs de plus de 90 jours) ou nettoyer l'espace.
- **Cas particulier (Versioning) :** Tu peux configurer l'expiration pour supprimer définitivement les **anciennes versions** d'un objet (Noncurrent Versions) tout en gardant la version actuelle, ou supprimer les "Delete Markers" périmés.
 - **A EXPLIQUER**

Filtres : Prefix & Tags Rules

Une règle de cycle de vie ne s'applique pas forcément à tout le bucket. Tu peux cibler :

- Un **Prefix** (chemin/dossier) : ex, uniquement les objets dans `logs/2026/`.
- Des **Object Tags** (clés/valeurs) : ex, uniquement les objets marqués `Environment: Development`.

Amazon S3 Storage Class Analysis

Comment savoir s'il faut passer à Standard-IA après 30 jours ou 60 jours ?

Cet outil analytique observe tes patterns d'accès aux données pendant plusieurs semaines. Il génère des graphiques et te recommande la politique de cycle de vie idéale (le nombre de jours exact) pour maximiser tes économies.

2. S3 Requester Pays

- **Pourquoi ?** Tu es une entreprise de recherche (Compte A) qui héberge un dataset de 5 To très populaire. Si des milliers de clients (Compte B, C, D) téléchargent tes fichiers, tu vas payer une fortune en frais de transfert de données sortantes (Data Transfer Out / Egress).
- **Le fonctionnement :** En activant Requester Pays, tu restes responsable du coût du stockage des données, mais **c'est l'utilisateur qui télécharge l'objet qui paye les coûts** de transfert (Egress) et de requête (GET).
- **La condition absolue (Identité AWS) :** Pour que cela fonctionne, le demandeur doit obligatoirement être authentifié auprès d'AWS. Les accès anonymes/publics sont automatiquement rejetés sur un bucket Requester Pays. AWS doit savoir quel compte débiter !

3. Event Notifications

Filtrage des événements

- **Pourquoi ?** Déclencher une action automatique lorsqu'un événement survient dans S3 (ex: exécuter un script dès qu'une image est uploadée).

Gestion des Permissions IAM : Les Politiques de Ressources

Pour que S3 puisse envoyer une notification à un autre service, il a besoin d'une autorisation. Contrairement à d'autres services, cela ne se fait pas via un rôle IAM classique attaché à S3, mais via des **Resource-based Policies** :

Exemples :

- SNS (Simple Notification Service) : Tu dois modifier la SNS Topic Policy pour autoriser le Principal `s3.amazonaws.com` à exécuter l'action `SNS:Publish`.
- SQS (Simple Queue Service) : Tu dois modifier la SQS Queue Policy pour autoriser `s3.amazonaws.com` à faire un `SQS:SendMessage`.
- Lambda : Tu dois ajouter une Resource Policy sur la fonction Lambda (`lambda:InvokeFunction`) autorisant le bucket S3 spécifique à l'appeler.

EventBridge (Advanced Filtering)

- Pourquoi passer par EventBridge ? Les notifications S3 natives sont limitées (un seul prefix/suffix par règle, pas de filtrage sur la taille de l'objet, etc.).
- Avantages EventBridge : 1. Permet un filtrage ultra-avancé (ex: déclencher une alerte uniquement si l'objet fait plus de 1 Go ET possède le tag Confidential).
2. Permet d'envoyer l'événement vers beaucoup plus de destinations (plus de 18 cibles AWS comme Step Functions, Kinesis, API Destinations).

4. S3 Performance & Optimisation

S3 est conçu pour être massivement scalable par défaut, mais pour l'examen, tu dois connaître les limites physiques et les techniques d'optimisation.

Les limites par Préfixe (La règle d'or)

Les performances de S3 sont calculées **par seconde et par préfixe (dossier)**. Un bucket n'a pas de limite globale, chaque préfixe est indépendant :

- **3 500 requêtes** de modification (`PUT / COPY / POST / DELETE`) par seconde par préfixe.
- **5 500 requêtes** de lecture (`GET / HEAD`) par seconde par préfixe.
- *Astuce architecture* : Si tu as besoin de 11 000 GET/s, crée deux dossiers distincts (`/folder1` et `/folder2`) et réparties tes fichiers.

Multipart Upload

- **Pourquoi ?** Obligatoire pour les fichiers de **plus de 5 Go**, mais fortement recommandé dès **100 Mo**.
- **Comment ?** Découpe le fichier en plusieurs morceaux et les uploade en parallèle. Si un morceau échoue à cause du réseau, tu ne recommences que ce morceau, pas tout le fichier.

S3 Transfer Acceleration

- **Pourquoi ?** Un utilisateur en Australie doit uploader un fichier de 2 Go vers un bucket situé à Dublin. La traversée de l'internet public va être lente et instable.
- **Comment ?** Le fichier est envoyé vers l'**Edge Location** CloudFront (réseau privé mondial d'AWS) la plus proche de l'utilisateur en Australie. Le reste du voyage jusqu'à Dublin se fait sur le réseau de fibre optique ultra-rapide d'AWS.

S3 Byte-Range Fetches

- **Pourquoi ?** Tu as un fichier vidéo de 50 Go sur S3 et tu veux juste lire les 5 premières minutes, ou tu veux accélérer le téléchargement d'un gros fichier.
- **Comment ?** Tu passes une commande HTTP `Range` pour demander des segments d'octets spécifiques (ex: les premiers 10 Mo). Permet de paralléliser la lecture d'un même fichier pour doper les performances.

5. S3 Batch Operations (Opérations en Lot)

- **Pourquoi ?** Tu as un bucket avec **100 millions d'objets existants**. On te demande d'ajouter un tag sur chacun d'eux, ou de les chiffrer, ou de changer leur classe de stockage.
- **Pourquoi c'est mieux qu'un script fait maison ?**
 - *Script maison (EC2 + SDK)* : Tu vas devoir gérer les pannes de réseau, le multi-threading, le throttling (limites S3), le suivi de ce qui a réussi ou échoué, et payer l'infrastructure de calcul. Cela prendrait des jours de dev et de run.
 - *S3 Batch Operations* : C'est un service entièrement managé (Serverless). Tu lui fournis un fichier manifeste (la liste des objets à traiter via S3 Inventory), tu choisis l'action (ex: *Replace all tags*), et AWS gère la mise à l'échelle, la reprise après erreur et te fournit un rapport final complet.

6. S3 Storage Lens

- **Pourquoi ?** Comprendre, analyser et optimiser le stockage à l'échelle d'une multinationale (**toute une AWS Organization**, des dizaines de comptes, des milliers de buckets).

S3 Storage Lens fournit un tableau de bord centralisé découpé en plusieurs piliers d'analyse :

1. **Summary (Résumé)** : Vue globale sur la taille totale du stockage (en To) et le nombre d'objets.
2. **Cost-optimization (Optimisation des coûts)** : Identifie les buckets qui ont des versions obsolètes non nettoyées ou des fichiers qui devraient être migrés vers Glacier.
3. **Data protection (Protection des données)** : Te montre le pourcentage de buckets où le *Versioning*, *MFA Delete* ou la *Réplication* ne sont pas activés.
4. **Access management (Gestion des accès)** : Repère les buckets configurés avec des accès publics dangereux.
5. **Event / Performance / Activity** : Analyse la fréquence d'utilisation et les taux de requêtes.
6. **Status Code** : Met en avant les erreurs récurrentes (ex: beaucoup d'erreurs `403 Forbidden` ou `404 Not Found` sur un bucket spécifique).

Free vs. Paid (Gratuit vs Payant)

- **Version Gratuite** : Activée par défaut pour tous les comptes. Fournit 29 métriques d'activité de base avec un historique de rétention de **14 jours** (agrégé au niveau du bucket).
- **Version Payante (Advanced)** : Permet d'obtenir des métriques avancées (recommandations de coûts, codes d'erreur HTTP), d'analyser jusqu'au niveau du **Prefix**, et d'étendre la rétention des données à **15 mois** pour faire des analyses de tendance à long terme.

? Cheat Sheet Mentale pour l'Examen (Modules 3 à 6)

1. Erreur de transition de cycle de vie $\rightarrow$ Vérifier la règle des **30 jours minimum** pour Standard-IA.
2. Partager un gros dataset sans payer le trafic sortant $\rightarrow$ **S3 Requester Pays**.
3. Filtrage d'événement complexe ou cible exotique $\rightarrow$ **Amazon EventBridge**.
4. Atteinte des limites de requêtes S3 (HTTP 503) $\rightarrow$ **Créer des sous-préfixes (dossiers) supplémentaires** pour diviser la charge.
5. Action massive sur des millions d'objets *existants* $\rightarrow$ **S3 Batch Operations**.
6. Rapport de conformité/coûts de S3 sur 50 comptes AWS différents $\rightarrow$ **S3 Storage Lens**.

Tu te sens prêt à tester tes connaissances avec une mise en situation réelle d'examen sur l'ensemble de tes notes S3 ?

Revision #3

Created 2026-05-29 15:35:32 UTC by Victor

Updated 2026-05-29 15:52:07 UTC by Victor