Amazon S3 – stockage objet

1. Introduction

Amazon S3 (Simple Storage Service) est un service de stockage d'objets. Contrairement à un disque classique (système de fichiers), S3 ne stocke pas des fichiers dans des dossiers, mais des objets dans des conteneurs appelés buckets.

L'objectif de S3 est de fournir :

- un stockage massif
- hautement durable
- accessible via Internet
- sécurisé et contrôlable finement

---

2. Modèle de stockage objet

2.1 Bucket

Un bucket est un conteneur de stockage.

Caractéristiques :

- nom unique au niveau mondial
- lié à une région AWS
- point d'entrée pour accéder aux objets

2.2 Objet

Un objet est l'unité de base dans S3.

Un objet est composé de :

- données (le contenu du fichier)
- une clé (identifiant unique dans le bucket)
- des métadonnées

Exemple :

- clé : images/photo.jpg
- contenu : image
- métadonnées : type, taille, date

Important : Il n'existe pas de vraie hiérarchie de dossiers. Les chemins sont simulés via le nom de la clé.

---

3. Types de buckets et évolutions récentes

3.1 Buckets généralistes

C’est le modèle classique :

- stockage multi-AZ
- très haute durabilité
- usage général

3.2 Directory buckets (S3 Express One Zone)

Caractéristiques :

- latence très faible
- organisation en répertoires
- stockage dans une seule zone de disponibilité

Usage :

- applications nécessitant des accès très rapides
- traitements intensifs (IA, analytics)

Limite :

- résilience plus faible (une seule zone)

3.3 S3 Tables

Permet de stocker des données tabulaires optimisées pour l’analyse.

Usage :

- data lake
- moteurs analytiques

3.4 S3 Vectors

Permet de stocker des vecteurs (embeddings).

Usage :

- intelligence artificielle
- recherche sémantique

---

4. Classes de stockage

S3 propose plusieurs classes pour optimiser les coûts :

- Standard : accès fréquent

- Intelligent-Tiering : adaptation automatique
- Standard-IA : accès peu fréquent
- One Zone-IA : moins cher, moins résilient
- Glacier : archivage
- Glacier Deep Archive : archivage long terme

Principe : Plus l’accès est rare, plus le stockage est économique, mais plus l’accès est lent.

---

5. Sécurité

La sécurité dans S3 repose sur plusieurs couches.

5.1 IAM (Identity and Access Management)

Permet de définir :

- qui accède
- à quelles ressources
- avec quels droits

5.2 Bucket Policy

Règles appliquées directement au bucket.

Exemples :

- autoriser une IP
- autoriser CloudFront
- interdire l’accès public

5.3 Block Public Access

Mécanisme global permettant d’éviter toute exposition accidentelle.

Bonne pratique : Toujours activer ce mécanisme par défaut.

5.4 Principe du moindre privilège

Chaque utilisateur ne doit avoir que les droits nécessaires.

---

6. Chiffrement

6.1 Chiffrement en transit

Les données circulent via HTTPS.

6.2 Chiffrement au repos

Trois modes principaux :

- SSE-S3 : clés gérées par AWS
- SSE-KMS : clés gérées via AWS KMS (recommandé)
- SSE-C : clés fournies par le client

6.3 Bonnes pratiques

- activer le chiffrement par défaut
- utiliser KMS pour données sensibles
- contrôler l’accès aux clés

---

7. Gestion du cycle de vie des données

7.1 Versioning

Permet de conserver plusieurs versions d’un objet.

Avantages :

- récupération après suppression accidentelle
- historique des modifications

Inconvénient :

- augmentation des coûts

7.2 Lifecycle policies

Permet d’automatiser :

- transition entre classes de stockage
- suppression des objets

Exemple :

- 30 jours → Standard-IA
- 90 jours → Glacier
- 365 jours → suppression

---

8. Réplication

Deux types :

- SRR (Same Region Replication)
- CRR (Cross Region Replication)

Objectifs :

- résilience
- conformité
- performance

---

9. Logging et audit

9.1 CloudTrail

Trace les actions effectuées sur S3 :

- qui a fait quoi
- quand

9.2 S3 Access Logs

Trace les accès aux objets.

9.3 CloudWatch

Permet :

- surveillance
- alertes

---

10. Accès aux objets

10.1 Accès standard

Via :

- API REST
- CLI
- SDK

10.2 URL publiques

Un objet peut être accessible publiquement si autorisé.

10.3 Pre-signed URL

Permet de donner un accès temporaire sécurisé à un objet.

11. Événements et automatisation

S3 peut déclencher des actions lors d'événements :

- upload de fichier
- suppression

Intégrations :

- AWS Lambda
- SNS
- SQS

12. Durabilité et disponibilité

- durabilité : 99.999999999%
- stockage sur plusieurs zones

Différence :

- durabilité = ne pas perdre les données
- disponibilité = pouvoir y accéder

13. Bonnes pratiques essentielles

- activer le chiffrement
- activer le versioning
- bloquer l'accès public
- utiliser IAM correctement
- mettre en place lifecycle policies
- activer les logs
- surveiller les accès

14. Erreurs fréquentes

- bucket public involontaire
- permissions trop larges
- absence de logs
- coûts non maîtrisés
- absence de stratégie de cycle de vie

15. Conclusion

Amazon S3 est un service simple en apparence, mais puissant.

Sa maîtrise repose sur :

- la compréhension du modèle objet
- la gestion des accès
- la sécurisation des données
- l'optimisation des coûts
- l'automatisation

S3 est au cœur de la majorité des architectures Cloud modernes.