S3

O que é?

- Serviço para armazenamento de arquivos
- Arquivos podem pesar de 0 bytes a 5 teras
- Armazenamento ilimitado
- Os arquivos são guardados em Buckets (como a AWS chama uma pasta de arquivos)
- Cada conta pode ter até 100 buckets por padrão.
- O S3 possui um namespace global, ou seja, o nome do seu S3 deve ser o único no mundo (tipo um username)
- O motivo do seu S3 precisar possuir um nome único é que ele terá uma URL própria.
- Quando um upload de arquivo é feito com sucesso, o código HTTP retornado é o 200. **(Costuma cair na prova).**

Features:

- Armazenamento em camadas (tiers)
- Gerenciamento do ciclo de vida dos arquivos
- Versionamento
- Criptografia de dados
- Pedir MFA para a deleção (opcional)
- Proteção de daqdos utilizando **Access Control List** e **Bucket Policies.**

Consistência de Dados

- Todos os objetos possuem uma consistência Read after Write **(Leitura disponível imediatamente após a escritura do arquivo)**

Em suma:

- Se você criar, atualizar ou deletar um novo arquivo, você poderá ver essas mudanças imediatamente, sem tempo de propagação.

Garantia:

- A Amazon garante 99.99999999999% de durabiidade para informações no S3. **(11 números 9 - costuma cair no exame)**

https://aws.amazon.com/pt/s3/faqs/ - FAQ do S3, MUITAS PERGUNTAS DAQUI CAEM NA PROVA

Lifecycle Rule

- Através da opção de lifecycle rule, é possível ajustar a "data de validade" dos arquivos e suas versões, de maneira que se automatize o tier em que o arquvio está de acordo com a sua idade

Object-based

- Todos os arquivos são armazenados em formato de objeto
  - ***Isso faz com que não seja possível instalar um SO ou hospedar uma base de dados no S3, já que o tipo de armazenamento não é Block-based***
- Chave e valor
  - A chave será o nome do arquivo
  - O valor será a sequência de bytes do arquivo
- ID da versão (O versionamento do arquivo é rastreado desta forma)
- Metadados (Dados sobre os dados armazenados)
- Subrecursos:
  - Lista de controle de acesso aos dados
  - Download como torrent

Cobrança

A cobrança pelo S3 ocorre baseada nos seguintes fatores:

- Armazenamento
- Utilização
- Preço da camada
- Preço da transferência de dados
- Aceleração da transferência (Utilização do backbone da Amazon)
- Replicação cross-region

Camadas do S3

1. **Standard ⇒** 99.99% de disponibilidade, 99 vírgula onze 9s de durabilidade, armazenado de maneira redundante em diversas facilities.
2. **Infrequently Acessed (IA) ⇒** Para os dados que são acessados com menos frequência, mas que exigem acesso rápido quando precisarem ser acessados. (Será cobrada uma taxa de recuperação do arquivo)
3. **Reduced Redundancy Storage (RRS) - IA ⇒** Uma opção de custo mais baixo para os dados IA, que não oferece a resiliência em múltiplas AZs.
4. **Intelligent Tiering ⇒** Feito para a otimização de custo, esta opção move arquivos entre camadas de acordo com o uso pelo usuário, de maneira que não ofereça impacto na performance.
5. **Glacier Flexible Retrieval ⇒** Classe de armazenamento segura, durável e de baixo custo, você pode guardar qualquer quantia de dados a um custo mais barato do que soluções on-premise. O tempo de recuperação dos arquivos pode levar de minutos a horas.
6. **Glacier Deep Archive ⇒** Essa é a classe mais barato do S3, tempo para recuperação de arquivos levará mais de 12 horas.

	S3 Standard	S3 Intelligent-Tiering*	S3 Standard-IA	S3 One Zone-IA†	S3 Glacier	S3 Glacier Deep Archive**
Designed for durability	99.999999999% (11 9's)	99.999999999% (11 9's)	99.999999999% (11 9's)	99.999999999% (11 9's)	99.999999999% (11 9's)	99.999999999% (11 9's)
Designed for availability	99.99%	99.9%	99.9%	99.5%	N/A	N/A
Availability SLA	99.9%	99%	99%	99%	N/A	N/A
Availability Zones	≥3	≥3	≥3	1	≥3	≥3
Minimum capacity charge per object	N/A	N/A	128KB	128KB	40KB	40KB
Minimum storage duration charge	N/A	30 days	30 days	30 days	90 days	180 days
Retrieval fee	N/A	N/A	per GB retrieved	per GB retrieved	per GB retrieved	per GB retrieved
First byte latency	milliseconds	millseconds	milliseconds	milliseconds	select minutes	select hours

Acesso ao Bucket

- Políticas do Bucket ⇒ São aplicadas no bucket inteiro
- Políticas do Objeto ⇒ Aplicadas somente em um arquivo individual
- Políticas IAM para usuários e grupos ⇒ São aplicadas para usuários e grupos específicos.

Versionamento

- Armazena todas as versões já existentes do objeto (todas as escritas, até quando esta foi deletada)
- Uma vez habilitado, o versionamento não pode ser desabilitado!
- Permite deleção com MFA
- Cada uma das versões do arquivo tem suas próprias permissões de acesso, por isso, você deve tornar cada um dos objetos públicos individualmente.
- Neste modo, quando um objeto for deletado, ele ainda ficará no S3 como um “delete marker”, e se você deletá-lo como “delete marker”, então ele, o

objeto, será restaurado.

- Para deletar o objeto de fato, você deve ir na aba “propriedades” e pausar o versionamento.

Lock Policies

Object Lock:

- **O que é:** O *S3 Object Lock* permite que você bloqueie objetos de um bucket, garantindo que eles não possam ser deletados ou alterados por um período de tempo específico ou indefinidamente (retention periods). Seguindo o modelo **WORM (Write Once, Read Many),**

- **Modos de Retenção:**
  - **Governance Mode:** Nesse modo, os usuários com permissões específicas (como administradores) ainda podem modificar ou deletar os objetos bloqueados, m**as** a maioria dos usuários não pode.
  - **Compliance Mode:** Nesse modo, NENHUM usuário, nem mesmo o root, pode modificar ou deletar os objetos durante o período de retenção. Isso é crucial para cumprir requisitos regulatórios onde a imutabilidade dos dados é obrigatória.
- **Legal Holds:** Além de modos de retenção, você pode aplicar *Legal Holds* (retenções legais) aos objetos, que funcionam como uma suspensão de qualquer retenção de tempo, tornando o objeto imutável até que o *Legal Hold* seja removido.
  - Qualquer usuário com a permissão `s3:PutObjectLegalHold` pode definir um legal hold.

- Aplicabilidade pode ser feita no bucket inteiro ou em um objeto individual.

- **Bucket Policy para Imutabilidade de Objetos**:

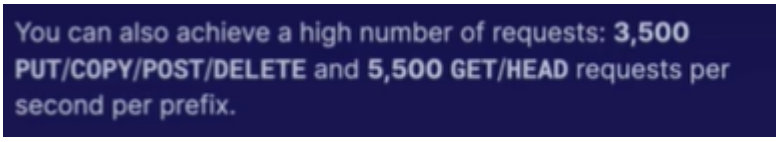
- **O que é:** As *Bucket Policies* permitem definir regras sobre o acesso e o comportamento dos objetos dentro do bucket. Com essas políticas, você pode configurar regras que evitem modificações ou exclusões não autorizadas, protegendo assim os dados de alterações indesejadas.
- **Como Funciona:** Você pode aplicar políticas baseadas em condições que, por exemplo, proíbem a exclusão de objetos ou permitem apenas operações de leitura. Isso é útil para cenários onde você quer garantir que os dados sejam somente leitura.

Glacier Vault Lock

Com o GVL é possível definir políticas de retenção no Glacier, de maneira semelhante ao Object Lock, podendo optar por bloqueio de leitura temporário ou até a desativação.

Performance & Prefixos

Ganho de Performance



Prefixos?

- **Prefixos em Nomes de Objetos:** O prefixo é a parte do nome de um objeto que precede qualquer caractere especial como uma barra (`/`). Por exemplo, em `folder1/subfolder2/file.txt`, `folder1/subfolder2/` seria considerado o prefixo.
- **Particionamento de Dados:** O S3 utiliza esses prefixos para particionar os dados em diferentes segmentos lógicos. Quanto mais diversificados forem os prefixos dos objetos armazenados, maior será a capacidade do S3 de distribuir as requisições entre diferentes partições.
- **Aumento de Performance:** Se você tiver uma grande quantidade de dados que recebe um volume muito alto de requisições, diversificar os prefixos

pode aumentar o throughput das operações, permitindo que mais requisições sejam atendidas simultaneamente.

Limitações do KMS

A criptografia de objetos do KMS pode trazer uma limitação para a quantidade de requests, a limitação varia de acordo com a região, mas está entre 5,500 e 30,000. Portanto, pense nisso na hora de ativar a criptografia.

Multipart Uploads

A técnica de multipart uploads aumentar a velocidade do upload de arquivos, Quando você faz o upload de um arquivo maior, como um de 5 GB, o Amazon S3 o divide em várias partes menores que são enviadas simultaneamente, em paralelo, ao bucket S3.

- **Envio Paralelo:**
  - As partes são carregadas simultaneamente, o que acelera significativamente o tempo de upload. Isso é especialmente útil em redes com largura de banda alta, onde várias conexões podem ser usadas ao mesmo tempo.
- **Resiliência a Falhas:**
  - Se o upload de uma parte falhar, apenas aquela parte específica precisa ser reenviada, em vez de reenviar todo o arquivo. Isso melhora a confiabilidade do upload, especialmente para arquivos muito grandes.
- **Montagem Final:**
  - Uma vez que todas as partes tenham sido carregadas, o S3 as une em um único objeto. Esse processo é transparente para o usuário, e o arquivo completo fica disponível como se tivesse sido enviado de uma só vez.
- Essa opção é facultativa para arquivos maiores que 100MB e OBRIGATÓRIA para arquivos com mais de 5GB.

Cross Region Replication

- Este é o processo realizado para transferir objetos do bucket de uma região X para um bucket da região Y. Para realizar a replicação **É OBRIGATÓRIO QUE OS BUCKETS DE ORIGEM E DE DESTINO TENHAM VERSIONAMENTO HABILITADO**.
- Só serão replicados os objetos criados a partir do momento que você definiu a regra de replicação. Ou seja, os objetos que já existem no bucket não serão replicados!

Transfer Aceleration

Recurso que melhora a velocidade de upload e download de arquivos para e de buckets S3 ao otimizar a transferência de dados através da rede global de edge locations da Amazon CloudFront. Em vez de enviar diretamente para o bucket S3, os dados são roteados para o ponto de presença mais próximo do usuário, que então encaminha os dados pela rede otimizada da AWS até o destino final, reduzindo a latência e acelerando o processo.

Este recurso é especialmente útil para usuários que precisam transferir grandes volumes de dados de regiões geograficamente distantes do S3, proporcionando uma maneira mais rápida e eficiente de mover dados na AWS.

Datasync

serviço que facilita a transferência rápida e segura de grandes volumes de dados entre sistemas de armazenamento locais e serviços de armazenamento da AWS, como Amazon S3, Amazon EFS (Elastic File System) e Amazon FSx. Ele é projetado para automatizar e otimizar a movimentação de dados, reduzindo a complexidade e o tempo necessário para transferências em massa. Ele suporta, especificamente sistemas de arquivos compatíveis com NFS (Network File System) e SMB (Server Message Block).

Storage Gateway & Snowball

Ambos são serviços destinados a migração de dados para a plataforma de armazento da AWS.

- O **Storage Gateway** um serviço que cria uma ponte virtual entre seu ambiente local e o armazenamento na nuvem da AWS. Ele permite que você use o armazenamento da AWS como uma extensão transparente do seu ambiente local.
    - **Backup e Arquivamento:** Ideal para empresas que desejam integrar seus sistemas de backup locais com o armazenamento da AWS.
    - **Extensão de Armazenamento:** Para ambientes que precisam expandir rapidamente a capacidade de armazenamento sem grandes investimentos em hardware local.
    - **Migração para a Nuvem:** Facilitando a migração gradual de dados locais para a nuvem, mantendo a compatibilidade com os sistemas locais existentes.

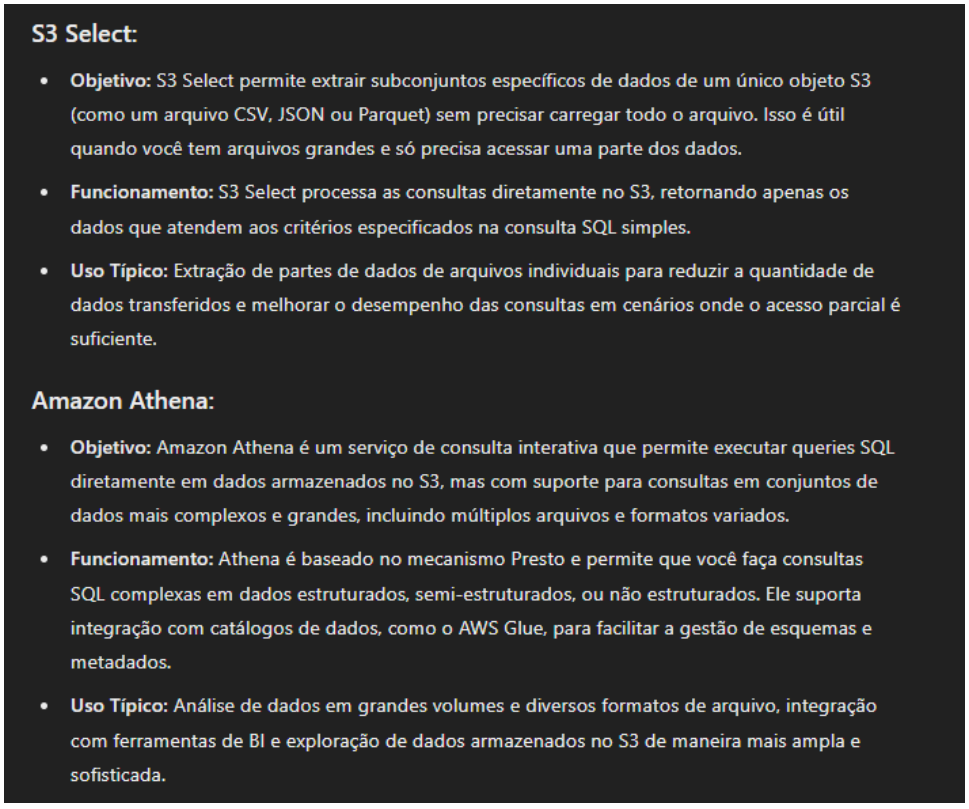
- Já o Snowball é um dispositivo físico usado para transferir grandes volumes de dados de e para a AWS. Ele é enviado para o local do cliente, carregado com dados, e depois devolvido à AWS para upload, doidera.
    - **Transferência de Grandes Volumes de Dados:** Ideal para mover petabytes de dados para a AWS de maneira econômica e segura, especialmente em locais com largura de banda de rede limitada.
    - **Migração em Massa:** Para empresas que precisam transferir grandes quantidades de dados para a nuvem em um período curto.
    - **Coleta de Dados em Campo:** Para ambientes remotos ou desconectados, onde é necessário coletar e processar dados localmente antes de enviá-los para a AWS.

Athena vs Macie

- **Athena** é um serviço de queries interativas, que permite buscar dados no S3 utilizando linguagem SQL
    - Serverless, nada de pay per query / per TB.

- Não é necessário estabelecer um ambiente ETL completo para análise de dados.
- Funciona com dados armazenados diretamente no S3.

▼ Sim, possui diferenças com o S3 Select, em especial no ponto onde o S3 Select busca somente extrair partes de um arquivo maior; um ZIP, por exemplo.



S3 Select:

- **Objetivo:** S3 Select permite extrair subconjuntos específicos de dados de um único objeto S3 (como um arquivo CSV, JSON ou Parquet) sem precisar carregar todo o arquivo. Isso é útil quando você tem arquivos grandes e só precisa acessar uma parte dos dados.
- **Funcionamento:** S3 Select processa as consultas diretamente no S3, retornando apenas os dados que atendem aos critérios especificados na consulta SQL simples.
- **Uso Típico:** Extração de partes de dados de arquivos individuais para reduzir a quantidade de dados transferidos e melhorar o desempenho das consultas em cenários onde o acesso parcial é suficiente.

Amazon Athena:

- **Objetivo:** Amazon Athena é um serviço de consulta interativa que permite executar queries SQL diretamente em dados armazenados no S3, mas com suporte para consultas em conjuntos de dados mais complexos e grandes, incluindo múltiplos arquivos e formatos variados.
- **Funcionamento:** Athena é baseado no mecanismo Presto e permite que você faça consultas SQL complexas em dados estruturados, semi-estruturados, ou não estruturados. Ele suporta integração com catálogos de dados, como o AWS Glue, para facilitar a gestão de esquemas e metadados.
- **Uso Típico:** Análise de dados em grandes volumes e diversos formatos de arquivo, integração com ferramentas de BI e exploração de dados armazenados no S3 de maneira mais ampla e sofisticada.

- Macie é um serviço de segurança que usa Machine Learning e NLP para descobrir, classificar e proteger informações sensíveis (PII - Personal Identifiable Informations) armazenadas no S3.
  - Oferece relatórios, dashboards e alertas sobre os dados
  - Trabalha diretamente com os dados no S3
  - Também pode analisar logs da CloudTrail para verificar atividades suspeitas da API.
  - Excelente para sites que trabalham com pagamento via cartão de crédito (PCI-DSS)