



Data Science Academy

www.datascienceacademy.com.br

Arquiteto de Soluções AWS

Guia Resumo Amazon S3

Guia de Resumo do Amazon S3

Site do serviço:

<https://aws.amazon.com/pt/s3/>

Guia do Desenvolvedor:

https://docs.aws.amazon.com/pt_br/AmazonS3/latest/dev/Welcome.html

Perguntas Frequentes:

<https://aws.amazon.com/pt/s3/faqs/>

Definição de Preço do S3:

<https://aws.amazon.com/pt/s3/pricing/>

Resumo do Amazon S3:

- O S3 (Serviço de Armazenamento Simples) é o serviço padrão de armazenamento em nuvem da AWS, oferecendo armazenamento de arquivos com tamanho de 0 a 5 TB.
- Itens, ou objetos, são colocados em recipientes nomeados armazenados chamados buckets. Os nomes dos objetos são normalmente chamados de chaves. O conteúdo principal é o valor.
- Os objetos são criados, excluídos ou atualizados. Objetos grandes podem ser transmitidos, mas você não pode acessar ou modificar partes de um valor. Você precisa atualizar todo o objeto.
- Cada objeto também possui metadados, que inclui pares de chave-valor arbitrários, e é usado de forma semelhante aos cabeçalhos HTTP.
- S3 URIs: embora muitas vezes os nomes de bucket e chave sejam fornecidos nas APIs individualmente, também é uma prática comum escrever uma localização S3 na forma 's3://bucket-name/key' (onde key é o nome do arquivo).
- S3 vs Glacier, EBS e EFS: a AWS oferece muitos serviços de armazenamento, e vários além do S3 oferecem abstrações de tipo de arquivo. O Glacier é para armazenamento de arquivamento mais barato e com pouca acessibilidade. EBS, ao contrário do S3, permite o acesso aleatório ao conteúdo do arquivo através de um sistema de arquivos tradicional, mas só pode ser anexado a uma instância EC2 por vez. O EFS é um sistema de arquivos de rede em que muitas instâncias podem se conectar, mas a um custo maior. Veja a tabela de comparação:

	Padrão	Standard – Acesso ocasional	Armazenamento de redundância reduzida
Durabilidade	99,999999999%	99,999999999%	99,99%
Disponibilidade	99,99%	99,9%	99,99%
Tolerância a falhas simultânea da instalação	2	2	1
Suporte a SSL	Sim	Sim	Sim
Latência de primeiro byte	Milissegundos	Milissegundos	Milissegundos
Políticas de gerenciamento de ciclo de vida	Sim	Sim	Sim

Fonte: <https://aws.amazon.com/pt/s3/reduced-redundancy/>

Dicas de Uso do Amazon S3:

- Para fins mais práticos, você pode considerar a capacidade S3 ilimitada, tanto no tamanho total dos arquivos quanto no número de objetos. O número de objetos em um bucket é essencialmente também ilimitado. Os clientes rotineiramente têm milhões de objetos. Mas apenas 100 buckets podem ser criados em uma única conta!
- Nomeação do bucket: os buckets são escolhidos de um espaço de nomes global (em todas as regiões, mesmo que o próprio S3 armazene dados em qualquer região S3 que você selecione), e, portanto, você encontrará muitos nomes de buckets que já foram tirados. Criar um bucket significa adotar o nome até que você o exclua. Os nomes dos buckets têm algumas restrições sobre eles.
 - Os nomes dos buckets podem ser usados como parte do nome do host ao acessar o bucket ou seu conteúdo, como <bucket_name>.s3-us-east-1.amazonaws.com, desde que o nome seja compatível com DNS.
 - Uma prática comum é usar o acrônimo ou abreviatura do nome da empresa para o prefixo (ou sufixo, se você preferir a hierarquia do estilo DNS) todos os nomes do bucket (mas, por favor, não use um controle sobre isso como uma medida de segurança - isso é altamente inseguro e facilmente contornado!).

- Nomes de bucket com '.' (ponto) podem causar desajustes de certificados quando usados com SSL. Use '-' em vez disso, pois estará em conformidade com as expectativas de SSL e é compatível com DNS. Não use letras maiúsculas no nome do bucket, para evitar problemas de acesso com o DNS.
- Versionamento: S3 possui suporte de versão opcional, de modo que todas as versões de objetos sejam preservadas em um bucket. Isto é principalmente útil se você quer fazer gestão de mudanças ou restaurar arquivos devido erros.
- Durabilidade: a durabilidade do S3 é extremamente alta, pois internamente mantém várias réplicas. Se você não exclui por acidente, você pode contar que o S3 não vai perder seus dados. (AWS oferece a taxa de durabilidade aparentemente improvável de 99,999999999%, mas este é um cálculo matemático baseado em taxas de falha independentes e níveis de replicação - não uma estimativa de probabilidade verdadeira. De qualquer forma, o S3 teve um registro muito bom de durabilidade.) Observe que essa é uma durabilidade muito maior do que o EBS!
- A maioria das pessoas usa a classe de armazenamento padrão no S3, mas existem outras classes de armazenamento com menor custo:
 - Reduced Redundancy Storage (RRS) está sendo descontinuado e tem menor durabilidade (99,99%) do que o padrão S3.
 - Infrequent Access (IA) permite que você obtenha um armazenamento mais barato em troca de um acesso mais caro. Isso é ótimo para arquivos como logs que você já processou, mas pode querer ver mais tarde. Para ter uma ideia das economias de custos ao usar Infrequent Access (IA), você pode usar esta Calculadora de Acesso Infrequente S3:
<http://www.gulamshakir.com/apps/s3calc/index.html>.
 - Glacier é uma terceira alternativa discutida como um produto separado.
- Desempenho: Maximizar o desempenho do S3 significa melhorar o rendimento global em termos de largura de banda e número de operações por segundo.
- S3 é altamente escalável, então, em princípio, você pode obter arbitrariamente alta taxa de transferência.
- O throughput é, naturalmente, o mais alto entre AWS e S3, e entre instâncias EC2 e buckets S3 que estão na mesma região. A largura de banda da EC2 depende do tipo de instância.

- A transferência de muitos objetos é extremamente alta quando os dados são acessados de forma distribuída, de muitas instâncias EC2. É possível ler ou escrever objetos do S3 de centenas ou milhares de instâncias ao mesmo tempo.
- No entanto, o throughput é muito limitado quando os objetos acessados sequencialmente de uma única instância. Operações individuais levam muitos milissegundos, e a largura de banda para e de instâncias é limitada. Portanto, para realizar um grande número de operações, é necessário usar vários segmentos e conexões de trabalho em instâncias individuais e para trabalhos maiores, várias instâncias EC2 também.
- Uploads de várias partes (Multi-part Upload): para objetos grandes, você pode aproveitar as capacidades de upload de várias partes (começando com tamanhos mínimos de 5 MB). Também é possível fazer o download de pedaços de um único objeto grande em paralelo, explorando a capacidade de cabeçalho de intervalo HTTP GET.
- Para dados fora da AWS, DirectConnect e S3 Transfer Acceleration podem ajudar. Para S3 Transfer Acceleration, você paga aproximadamente o equivalente a 1-2 meses de armazenamento para a transferência em qualquer direção.
- É importante gerenciar as permissões no S3 se você tiver dados sensíveis, pois corrigir isso mais tarde pode ser uma tarefa difícil se você tiver muitos objetos e usuários internos. Crie novos buckets se você tiver diferentes níveis de sensibilidade nos dados, pois isso é muito menos propenso a erros do que as regras de permissões complexas. Se os dados são apenas para administradores, como dados de log, coloque-o em um bucket que apenas os administradores podem acessar. Limite o acesso de usuário individual (ou IAM) ao S3 ao mínimo exigido e catalogue os locais "aprovados". Caso contrário, o S3 tende a se tornar o campo de despejo onde as pessoas colocam dados em locais aleatórios que não são limpos por anos, custando-lhe muito dinheiro.
- Ao gerenciar dados, a compreensão do ciclo de vida dos dados é tão importante quanto a compreensão dos próprios dados. Ao colocar os dados em um bucket, pense em seu ciclo de vida até o fim, e não apenas em seu início. Em geral, os dados com diferentes políticas de expiração devem ser armazenados sob prefixos separados no nível superior. Por exemplo, alguns logs podem precisar ser excluídos automaticamente mensalmente, enquanto outros dados são críticos e nunca devem ser excluídos. Ter o primeiro em um bucket separado ou pelo menos uma pasta separada é uma boa prática. É muito difícil limpar grandes coleções de arquivos criados por muitos engenheiros com diferentes ciclos de vida e nenhuma organização coerente. Alternativamente, você pode definir uma política de ciclo de vida para arquivar dados antigos para o Glacier. Tenha cuidado com o arquivamento de um grande número de objetos pequenos para o Glacier, pois pode custar mais. Há também uma classe de armazenamento chamada Infrequent

Access que tem a mesma durabilidade do Standard S3, mas é descontada por GB. É adequado para objetos com acesso infrequente.

- O S3 possui um recurso para replicar um bucket entre uma região e outra, chamado Cross-region replication. Note-se que o S3 já é altamente replicado dentro de uma região, por isso geralmente isso não é necessário para durabilidade, mas pode ser útil para conformidade (armazenamento de dados distribuído geograficamente), menor latência ou como estratégia para reduzir a largura de banda de região para região custos, refletindo dados altamente usados em uma segunda região.
- Há um notório limite de 100 buckets por conta no S3, o que não poderia ser aumentado e causou muito problema nas empresas. Desde 2015, você pode solicitar aumento nesse número, o que será avaliado pela Amazon. Você pode pedir para aumentar o limite, mas ainda será limitado (geralmente abaixo de ~ 1000 por conta).
- S3 possui SLA com 99,9% de tempo de atividade. Se você usar S3 com muita frequência, inevitavelmente verá erros ocasionais de acesso ou armazenamento de dados à medida que os discos ou outra infraestrutura falharem. A disponibilidade normalmente é restaurada em segundos ou minutos. Embora a disponibilidade não seja extremamente alta, como mencionado acima, a durabilidade é excelente.
- Após o upload, qualquer alteração que você faz no objeto causa uma reescrita completa do objeto, então evite o comportamento de editar os arquivos regularmente.

Referências:

Definição de Preços do S3

<https://aws.amazon.com/pt/s3/pricing/>

Contrato de Nível de Serviço

<https://aws.amazon.com/pt/s3/sla/>

Amazon S3 Reduced Redundancy Storage (descontinuado)

<https://aws.amazon.com/pt/s3/reduced-redundancy/>

Definição do Preço do Glacier

<https://aws.amazon.com/pt/glacier/pricing/>

Amazon EBS

https://aws.amazon.com/pt/ebs/details/#Amazon_EBS_Availability_and_Durability

FAQ S3

<https://aws.amazon.com/pt/s3/faqs/>