

STORAGE DEVELOPER CONFERENCE



*BY Developers FOR Developers*

A decorative graphic on the left side of the slide, consisting of a dense cluster of small, colored dots in shades of purple, teal, and yellow, arranged in a pattern that tapers to the right.

# Beyond S3 Compatibility Claims

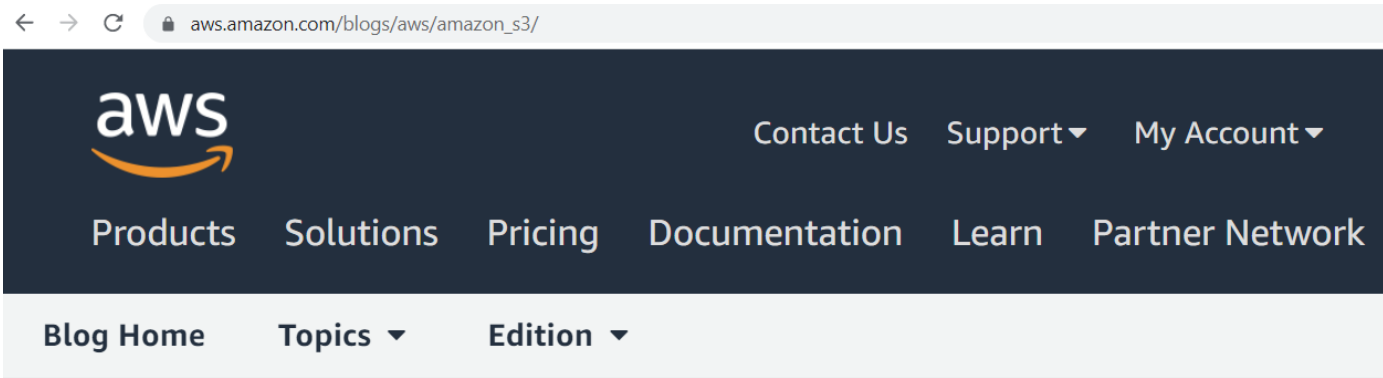
A Deep Dive into Real-World Incompatibilities

Gregory Touretsky, Seagate

# Is '100% compatibility' with Amazon S3 just a myth? Let's find out

The fine print of S3 compatibility: What vendors won't tell you

# Amazon S3



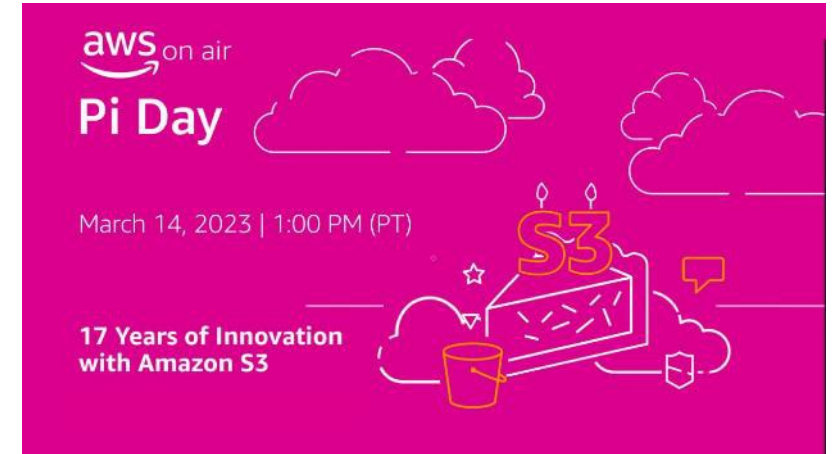
## AWS News Blog

# Amazon S3

by [Jeff Barr](#) | on 14 MAR 2006 | [Permalink](#) | [Share](#)

Earlier today we rolled out [Amazon S3](#), our reliable, highly scalable, low-latency data storage service.

Using SOAP and REST interfaces, developers can easily store any number of blocks of data in S3. Each block can be up to 5 GB in length, and is associated with a user-defined key and additional key/value metadata pairs.



- >280 Trillion objects
- >100M requests per second
- 4B checksum calculations per second

# Amazon S3 API

Amazon S3 API Reference

- Actions
  - Amazon S3
    - AbortMultipartUpload
    - CompleteMultipartUpload
    - CopyObject
    - CreateBucket
    - CreateMultipartUpload
    - DeleteBucket
    - DeleteBucketAnalyticsConfiguration
    - DeleteBucketCors
    - DeleteBucketEncryption
    - DeleteBucketIntelligentTieringConfiguration
    - DeleteBucketInventoryConfiguration
    - DeleteBucketLifecycle
    - DeleteBucketMetricsConfiguration
    - DeleteBucketOwnershipControls
    - DeleteBucketPolicy
    - DeleteBucketReplication
    - DeleteBucketTagging
    - DeleteBucketWebsite
    - DeleteObject**
    - DeleteObjects
    - DeleteObjectTagging
    - DeletePublicAccessBlock
    - GetBucketAccelerateConfiguration
    - GetBucketAcl
    - GetBucketAnalyticsConfiguration
    - GetBucketCors
    - GetBucketEncryption
    - GetBucketIntelligentTieringConfiguration
    - GetBucketInventoryConfiguration

## DeleteObject

PDF

Removes the null version (if there is one) of an object. If the object you want to delete is in a bucket with Amazon S3 Object Lock enabled, you must use the version ID of the object. If the object deleted is a delete marker, Amazon S3 sets the object's `is_delete_marker` flag to true.

If the object you want to delete is in a bucket with Amazon S3 Object Lock enabled, you must use the version ID of the object. For more information about MFA Delete, see [Using MFA Delete](#).

You can delete objects by explicitly calling DELETE them for you. If you want to block users or accounts from deleting objects, you can use the `s3:DeleteObject` and `s3:DeleteObjectVersion` actions.

The following action is related to DeleteObject:

- PutObject

### Request Syntax

```
DELETE /Key+?versionId=VersionId HTTP/1.1
Host: Bucket.s3.amazonaws.com
x-amz-mfa: MFA
x-amz-request-payer: RequestPayer
x-amz-bypass-governance-retention: BypassGovernanceRetention
x-amz-expected-bucket-owner: ExpectedBucketOwner
```

### URI Request Parameters

The request uses the following URI parameters.

Bucket

90+ Amazon S3 Actions

Amazon S3 API Reference

- Actions
  - Amazon S3
    - CreateAccessPoint
    - CreateMultiRegionAccessPoint
    - DeleteAccessPoint
    - DeleteAccessPointForObjectLambda
    - DeleteAccessPointPolicy
    - DeleteAccessPointPolicyForObjectLambda
    - DeleteBucket
    - DeleteBucketLifecycleConfiguration
    - DeleteBucketPolicy
    - DeleteBucketReplication
    - DeleteBucketTagging
    - DeleteJobTagging
    - DeleteMultiRegionAccessPoint
    - DeletePublicAccessBlock
    - DeleteStorageLensConfiguration
    - DeleteStorageLensConfigurationTagging
    - DescribeJob
    - DescribeMultiRegionAccessPointOperation
    - GetAccessPoint
    - GetAccessPointConfigurationForObjectLambda
    - GetAccessPointForObjectLambda
    - GetAccessPointPolicy
    - GetAccessPointPolicyForObjectLambda
    - GetAccessPointPolicyStatus

## CreateJob

PDF

You can use S3 Batch Operations to create jobs.

Related actions include:

- DescribeJob
- ListJobs
- UpdateJobPriority
- UpdateJobStatus
- JobOperation

### Request Syntax

```
POST /v20180820/jobs HTTP/1.1
Host: s3-control.amazonaws.com
x-amz-account-id: AccountId
<?xml version="1.0" encoding="UTF-8" ?>
<CreateJobRequest xmlns="http://s3.amazonaws.com/doc/2018-08-20/">
  <ConfirmationRequired>true</ConfirmationRequired>
  <Operation>
    <LambdaInvoke>
      <FunctionArn>string</FunctionArn>
    </LambdaInvoke>
    <S3DeleteObjectTagging>
      </S3DeleteObjectTagging>
    <S3InitiateRestoreObject>
      <ExpirationInDays>integer</ExpirationInDays>
      <GlacierJobTier>string</GlacierJobTier>
    </S3InitiateRestoreObject>
    <S3PutObjectAcl>
      <AccessControlPolicy>string</AccessControlPolicy>
    </S3PutObjectAcl>
  </Operation>
</CreateJobRequest>
```

60+ Amazon S3 Control Actions

AWS Identity and Access Management API Reference

## CreatePolicy

PDF

Welcome to the AWS Identity and Access Management (IAM) console. You can use IAM to manage your AWS accounts, managed users, and groups. You can also use IAM to manage your AWS accounts, managed users, and groups.

Actions

- AddClientIDToOpenIDConnectProvider
- AddRoleToInstanceProfile
- AddUserToGroup
- AttachGroupPolicy
- AttachRolePolicy
- AttachUserPolicy
- ChangePassword
- CreateAccessKey
- CreateAccountAlias
- CreateGroup
- CreateInstanceProfile
- CreateLoginProfile
- CreateOpenIDConnectProvider
- CreatePolicy**
- CreatePolicyVersion
- CreateRole
- CreateSAMLProvider
- CreateServiceLinkedRole
- CreateServiceSpecificCredential
- CreateUser
- CreateVirtualMFADevice
- DeactivateMFADevice
- DeleteAccessKey
- DeleteAccountAlias

### IAM, STS Actions

As a best practice, you can validate your policy version. For more information about managed policies, see [Versioning for managed policies](#) in the IAM User Guide.

### Request Parameters

For information about the parameters for this action, see [Request Parameters](#) in the IAM User Guide.

### Description

A friendly description of the policy. Typically used to store information about the policy. The policy description is immutable. Type: String Length Constraints: Maximum length: 256 characters. Required: No

### Path

The path for the policy. For more information about paths, see [Paths](#) in the IAM User Guide. This parameter is optional. If it is not provided, the path is the root path. This parameter allows (through its `regex` attribute) to end with forward slashes. In addition, it does not allow you to use double backslashes or other punctuation characters, digits, and underscores.



# S3-Compatible Storage

## Cloud Services



## Systems and Software



# Official Incompatibilities

## Unsupported S3 APIs

Table 3. Unsupported S3 APIs

FEATURE	NOTES
DELETE Bucket tagging	-
DELETE Bucket website	-
GET Bucket location	ECS is only aware of a single Virtual Data Center (VDC).



### Use metadata search queries

The metadata search feature provides a rich query language that enables objects that have indexed metadata to be searched.

Table 21. API Syntax

API Syntax	Response Body
<pre>GET /{bucket}/? query={expression} &amp;attributes={fieldname, ...} &amp;sorted={selector} &amp;include_older_version= {true false} &amp;max-keys={num_keys} &amp;marker={marker_value}</pre>	<pre>&lt;BucketQueryResult xmlns:ns2="http:// s3.amazonaws.com/doc/2006-03-01/"&gt;   &lt;Name&gt;mybucket&lt;/Name&gt;   &lt;Marker/&gt;   &lt;IsTruncated&gt;false&lt;/IsTruncated&gt;   &lt;MaxKeys&gt;0&lt;/MaxKeys&gt;   &lt;ObjectMatches&gt;     &lt;object&gt;       &lt;objectName&gt;file4&lt;/objectName&gt;       &lt;objectId&gt;09998027b1b7fbb21f50e13fabb48 1a237ba2f60f352d437c8da3c7c1c8d7589&lt;/ objectId&gt;       &lt;versionId&gt;0&lt;/versionId&gt;       &lt;queryMds&gt;         &lt;type&gt;SYSMD&lt;/type&gt;         &lt;mdMap&gt;</pre>

**NOTE:** Prefix capability is added to the metadata search. See [Prefix capability in metadata search](#).




# Protocol Compliance Tools

- Home-grown
- 3<sup>rd</sup> party applications
- <https://github.com/ceph/s3-tests>
- <https://github.com/splunk/s3-tests>
- <https://github.com/open-io/ceph-s3-tests>
- <https://github.com/minio/mint>

Solution	Failed Mint tests
Minio	0
Backblaze B2	14
Google Cloud Storage – S3	15
AWS S3	12

Is Minio more S3-compatible than Amazon S3? 😊

# What Is Behind the Endpoint?

```
aws s3 ls --debug |& grep "Response headers" | awk -F 'Server' '{print $2}'
```

Solution	Server Header
Amazon S3	AmazonS3
Google Cloud Storage (S3 compatible)	UploadServer
Ceph	-
Minio	MinIO
Wasabi	WasabiS3/7.12.1004-2023-02-17-7ff2f5bdd9 (head07)
Seagate Lyve Cloud	Seagate-LyveCloudS3
Backblaze B2	-



# What Is in the Request?

- Example: UploadPart

Component	Example
Method	PUT
Bucket	mybucket
Host	us-east-1.lyvecloud.seagate.com
Key	myprefix/myobject
partNumber	12
uploadId	2e1c42be-fc1d-4055-bbce-d10a55a0a662
authorization	AWS4-HMAC-SHA256 Credential=REDACTED/20230418/us-east-1/s3/aws4_request, SignedHeaders=content-length;host;user-agent;x-amz-content-sha256;x-amz-date, Signature=61b9391bc68984f634db8437779e76a8f609a5823b3ea0ac00a3df48e431d59c
user-agent	APN/1.0 Qumulo/1.0 S3Replication/6.0.2
...	

# What Is in the Response?

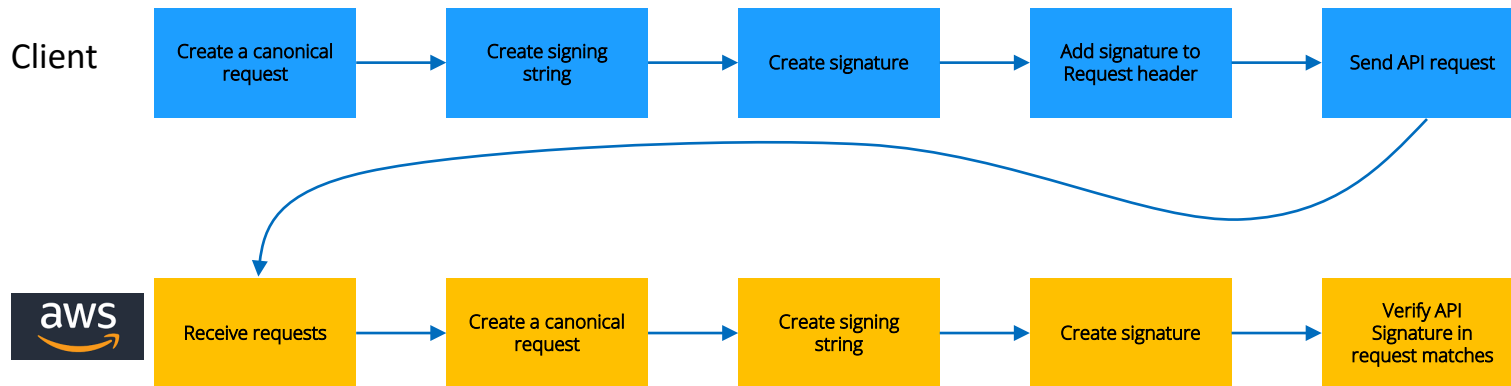
- Example: UploadPart

Component	Example
Response code	HTTP/1.1 200
Date	Tue, 18 Apr 2023 04:25:17 GMT
Etag	"3ae3ac4887f98ddefde4eeb82e37280b"
Server	Seagate-LyveCloudS3
X-Amz-Request-Id	1756ECD65A5EE9C2
...	

# Authentication / Signature



- Sig V2 – old, discontinued in 2020
  - But many client apps still use it!
- Sig V4 – current, in use since 2012:



```
#Create a Canonical Request
CanonicalRequest =
  HTTPRequestMethod + '\n' +
  CanonicalURI + '\n' +
  CanonicalQueryString + '\n' +
  CanonicalHeaders + '\n' +
  SignedHeaders + '\n' +
  HexEncode (Hash (RequestPayload))
```

Header names, sorted by lowercase character code, delimited by semi-colon

- Customer's app fails CopyObject requests: **SignatureDoesNotMatch**
- Root cause:
  - Server returns a mix of lower case and mixed-case header names, Ex: X-Amz-Server-Side-Encryption
  - Customer's app sorts headers case-sensitive → Signed headers string is not properly sorted

```
Authorization: AWS4-HMAC-SHA256
Credential=REDACTED/20230205/us-east-1/s3/aws4_request,SignedHeaders=content-type;host;user-agent;x-amz-server-side-encryption;x-amz-acl;x-amz-content-sha256;x-amz-copy-source;x-amz-date;x-amz-metadata-directive,Signature=246998b7b32681af8d6dbfdf8754da20c4633509d20fc6391bfe291d0d4caba1
```



# Object Key (Path)

- Up to 1,024 bytes long
- The following objects can coexist in a bucket:
  - mybucket/myfolder/obj
  - mybucket/myfolder/obj/
  - mybucket/myfolder/obj//
  - mybucket/myfolder/../obj/
  - mybucket/myfolder/

- Up to 1,024 bytes long
- “/” is interpreted as a directory
- Directory segments are limited to 255 bytes
- “//”, “.”, “..” are not allowed
- mybucket/myfolder and mybucket/myfolder/obj objects can't coexist



```
$ aws --profile minio s3api put-object --bucket  
bucket1 --key NameWith//Inside --body ~/empty
```

An error occurred (XMinioInvalidObjectName) when calling the PutObject operation: Object name contains unsupported characters



# Complete-Multipart-Upload Response Caching



AWS CLI Command Reference Home User Guide Forum GitHub

- Examples
- Output

## complete-multipart-upload ¶

Quick search

Description ¶

Search box

Complete Multipart Upload is an **idempotent** operation. After your first successful complete multipart upload, if you call the operation again within a short period, the operation will succeed.

```
$ aws s3api complete-multipart-upload --bucket A --key B --uploadId XX --multipart-upload file
POST 200 None
$ aws s3api complete-multipart-upload --bucket A --key B --uploadId XX --multipart-upload file
POST 200 None
```



```
$ aws s3api complete-multipart-upload --bucket A --key B --uploadId XX --multipart-upload file
POST 200 325
$ aws s3api complete-multipart-upload --bucket A --key B --uploadId XX --multipart-upload file
POST 404 449
```



# Don't Be Fooled by "Success"

## Complete-Multipart-Upload

Processing of a Complete Multipart Upload request could take several minutes to complete. After Amazon S3 begins processing the request, it sends an HTTP response header that specifies a 200 OK response. While processing is in progress, Amazon S3 periodically sends white space characters to keep the connection from timing out. **A request could fail after the initial 200 OK response has been sent.** This means that a 200 OK response can contain either a success or an error.

```
2023-03-15 21:40:29,089 - MainThread - urllib3.connectionpool - DEBUG - https://gt-test-006.s3.us-east-1.amazonaws.com:443 "POST /largeobjecttest?uploadId=szRCQ4o6dw8qjRjUjOe9WD2z2JbE5bHFuZvL27zUciZJW3um8GeIqYcPlLNu_GzUzuYCheYCYpAaWdLZF3x3I8rAdVF_7U109PBm3nd_ATIntjyYqHOcVdbS6X8vmxNI HTTP/1.1" 200 None
```

```
21:40:29,096 - MainThread - botocore.parsers - DEBUG - Response body:b'<?xml version="1.0" encoding="UTF-8"?>\n<Error><Code>EntityTooLarge</Code><Message>Your proposed upload exceeds the maximum allowed size</Message><ProposedSize>5513664266240</ProposedSize><MaxSizeAllowed>5497558138880</MaxSizeAllowed><RequestId>A1B9N8GCJ9368Z0N</RequestId><HostId>nXky785oI/qBz4qo1PO3M00bNF/SJSXiw6tLSEESNF1hVT2kEU2cKWKxbfG5iTw4KlVBNok5GoY=</HostId></Error>'
```

# Get-object-attributes



```
$ aws s3api get-object-attributes --bucket gt-test-006 --key awscliv2.zip --object-attributes "ETag"
```

```
{
  "LastModified": "2023-04-20T01:01:27+00:00",
  "VersionId": "2m1EHmC6ALD_FVJ6HZBJ9znvBEbi6pNa",
  "ETag": "75c77163c337dfd5bb5a5f9f7a6473dd-1"
}
```

```
$ aws s3api get-object-attributes --bucket gt-test-006 --key awscliv2.zip --object-attributes "ETag"
```

Unable to parse response (not well-formed (invalid token): line 1, column 2), invalid XML received. Further retries may succeed:

```
b'PK\x03\x04\x14\x00\x00\x00\x00\x00F\x8f\xf4T\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x04\x00\x00\x00aws/PK\x03\x04\x14\x00\x00\x00\x00\x00E\x8f\xf4T\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\t\x00\x00\x00aws/dist/PK\x03\x04\x14\x00\x00\x00\x08\x00\x1a\x8e\xf4TW\x92\xbe\n<\x02\x00\x00\x0b
```



# Head-bucket



```
$ aws s3api head-bucket --bucket gt-test-006 --expected-bucket-owner wronguser --debug  
|& grep HEAD  
urllib3.connectionpool - DEBUG - https://gt-test-006.s3.us-east-1.amazonaws.com:443  
"HEAD / HTTP/1.1" 400 0
```



```
$ aws s3api head-bucket --bucket gt-test-006 --expected-bucket-owner wronguser --debug  
|& grep HEAD  
urllib3.connectionpool - DEBUG - http://10.0.0.83:9000 "HEAD /gt-test-006 HTTP/1.1" 200  
0
```





# Put-object – Unsupported CRC32C

```
$ ./warp get --access-key=REDACTED --secret-key= REDACTED --bucket=gt-test-001 --
concurrent=60 --host=ENDPOINT --obj.size=16MiB --tls --duration 300s --objects=2500 --
analyze.v
warp: <ERROR> upload error: The X-Amz-Checksum-Crc32c you specified did not match what
we received.

$ aws s3api put-object --bucket gt-test-001 --key myobject --body myobject.zip --
checksum-crc32-c 8KygCQ==
An error occurred (InvalidRequest) when calling the PutObject operation: Value for x-
amz-checksum-crc32c header is invalid.
```

# Performance



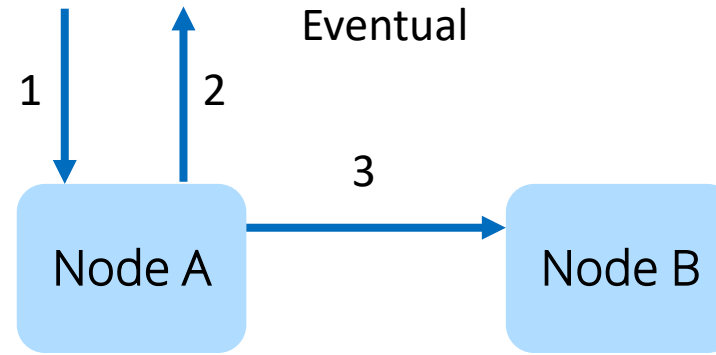
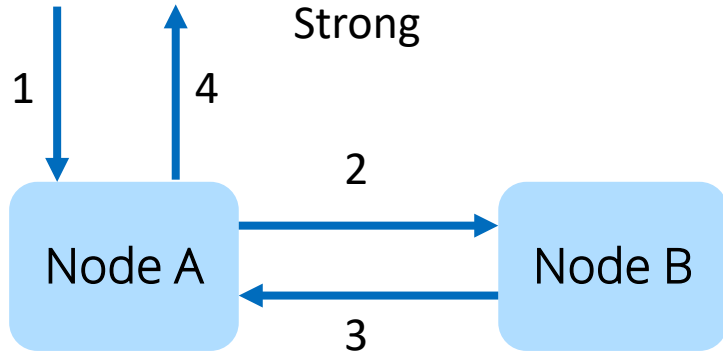
Requests/ sec

- |  |  |  |  |
|--|--|--|--|
| <ul style="list-style-type: none"> <li>• 3,500 PUTs / 5,500 GETs per sec per partitioned prefix</li> <li>• Request rates are allocated proportionally</li> </ul> | <ul style="list-style-type: none"> <li>• Up to 500 requests/sec for a single blob (object)</li> <li>• Default max request rate per storage account: 20,000 requests / sec</li> </ul> | <ul style="list-style-type: none"> <li>• Initial ~1,000 writes/sec, 5,000 reads/sec</li> <li>• Gradually autoscaling above limits, based on prefixes. Double rate every 20 min</li> <li>• Bucket                             <ul style="list-style-type: none"> <li>• Create/delete: 0.5/sec</li> <li>• Update: 1/sec</li> </ul> </li> </ul> | <ul style="list-style-type: none"> <li>• S3 API: “fair use”, depends on storage volume</li> <li>• Account Control API: GET 1000/min, PUT 100/min, DELETE 10/min</li> </ul> |
|--|--|--|--|

Throughput

- |   |  |  |  |
|---|--|--|--|
| <ul style="list-style-type: none"> <li>• single-instance ... up to 100 Gb/s</li> <li>• aggregate ... multiple Tbps</li> </ul> | <ul style="list-style-type: none"> <li>• Default max ingress per storage account: 10/25/60 Gbps (varies per region)</li> <li>• Default max egress per storage account : 50/120 Gbps</li> </ul> | <ul style="list-style-type: none"> <li>• Default egress quota: 200Gbps per region</li> </ul> | <ul style="list-style-type: none"> <li>• No details</li> </ul> |
|---|--|--|--|

# Consistency Model (\*)



Before Dec'2020



(\*) Data path. Access control, etc may vary

# Summary



Complexity + nuance



Incompatibilities



Deep understanding of the S3 API



Thorough testing



S3 compatibility = customer adoption



**HAVE YOU STRUGGLED  
WITH S3 COMPATIBILITY?**

# Driving Compatibility and Collaboration in Cloud Storage

SNIA Cloud Storage TWG Work Item

- Foster Ecosystem Collaboration
  - Facilitate collaboration and knowledge sharing among S3 developers by establishing a platform for discussions, forums, and workshops
- Enable S3 Multi-Cloud Interoperability
- Establish Compliance Certification
  - Documentation
  - List of known incompatibilities
  - Standardized compatibility tests



BoF Session: Tuesday, 9/19 @ 8pm  
[gregory.touretsky@seagate.com](mailto:gregory.touretsky@seagate.com)



# Beyond S3 Compatibility Claims

Please take a moment to rate this session.

Your feedback is important to us.