Cloud Storage

How cloud storage differs from file storage

So what is a file?

So what is a file?



WARNING

Hardcore, low level information ahead!

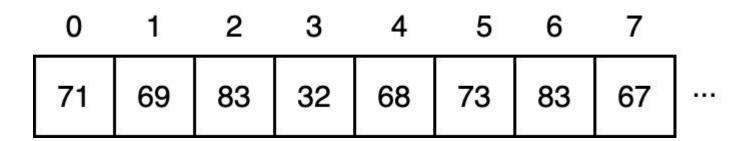


So what is a file?

Definition 1: an ordered collection of bytes that follow a structure. E.g. - ASCII, jpg, NetCDF, geotiff, ...

Definition 2: an operating system level API.

Definition 1: ordered collection of bytes

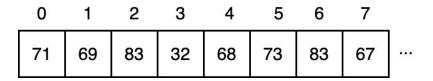


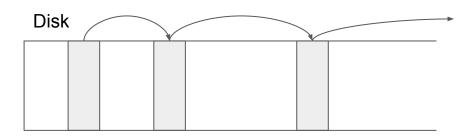
... over to notebook demo ...

Definition 2: operating system level API

- Operating systems provide a common interface (API) to storage devices
 - files
 - folders
 - file permissions
 - file locks
 - file caching
 - 0 ...
- python/c/java/ ... → OS API → hardware controllers

Definition 2: operating system level API





Definition 2: operating system level API

FAT: https://www.youtube.com/watch?v=V2Gxqv3bJCk

Inode: https://www.youtube.com/watch?v=tMVj22EWg6A

So what is s3 object storage?

Definition 1: an ordered collection of bytes that follow a structure. E.g. - ASCII, jpg, NetCDF, geotiff, ...

Definition 2: an API for network storage.

Mental model: file systems → s3 (AWS object storage)

file system (OS) concepts	s3 concepts
drive	~ bucket
file path	~ key
file/directory permissions	policies ++++
file pointer	
locking	
directories	

... over to notebook demo ...



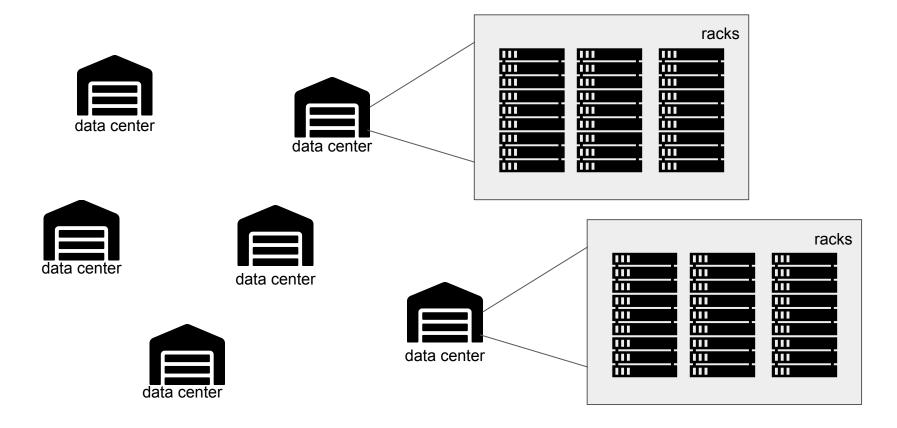


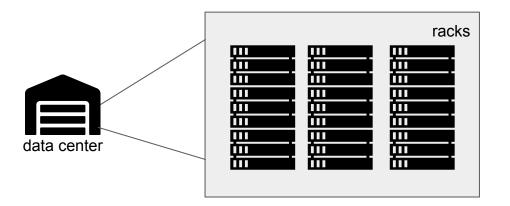


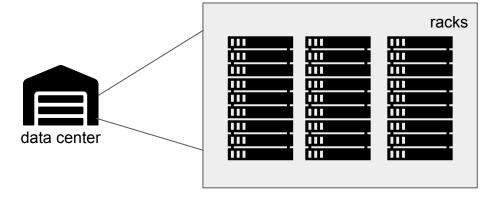






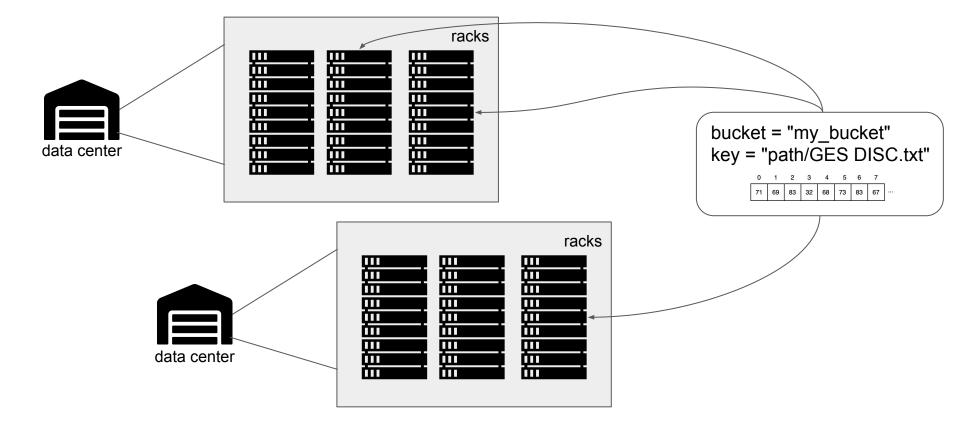


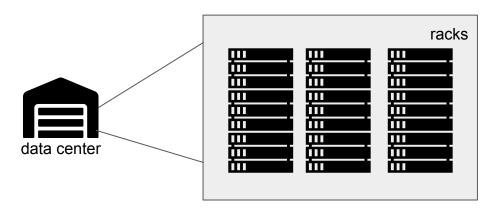


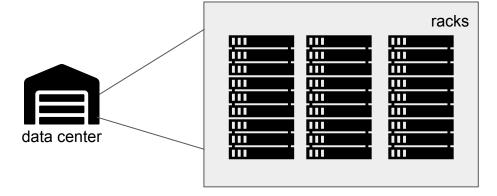


Create an object

bucket = "my_bucket"
key = "path/GES DISC.txt"

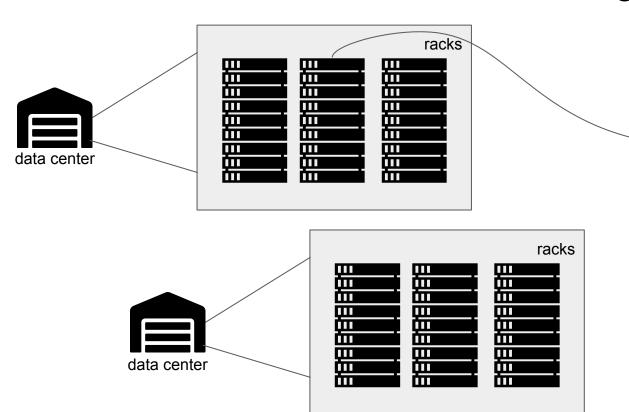






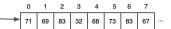
Read an object:

bucket = "my_bucket"
key = "path/GES DISC.txt"



Read an object:

bucket = "my_bucket"
key = "path/GES DISC.txt"

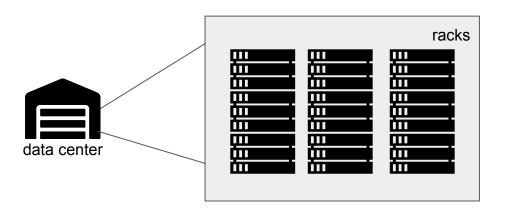


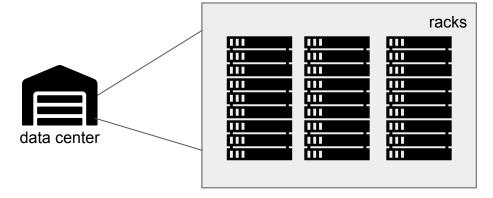


Strong read after write consistency <



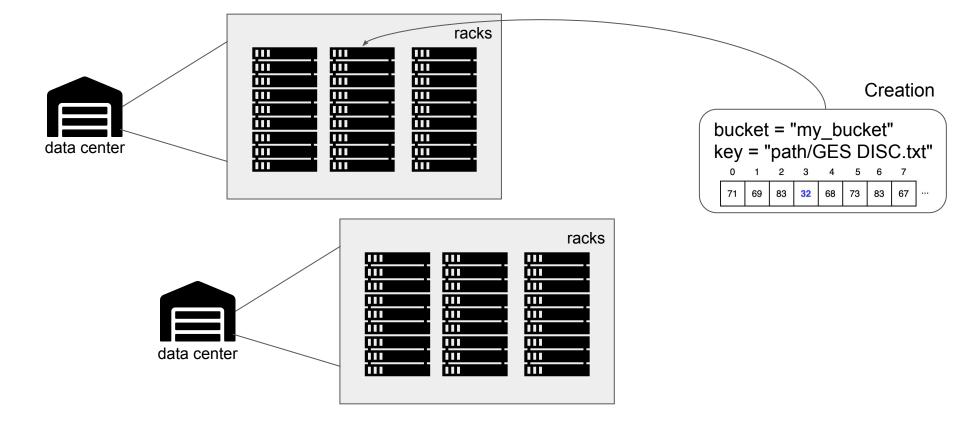
Breaking news:
announced re:Invent 2020!

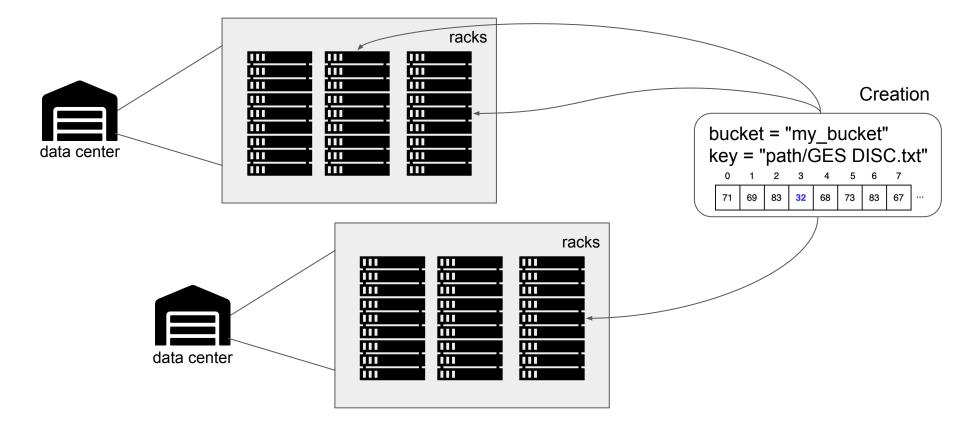


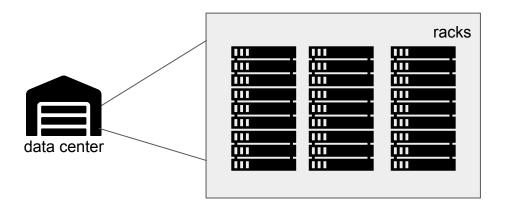


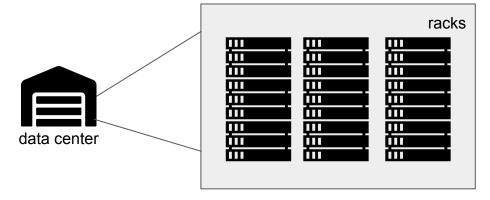
Creation

bucket = "my_bucket" key = "path/GES DISC.txt" 0 1 2 3 4 5 6 7 71 69 83 32 68 73 83 67 ...

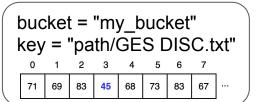


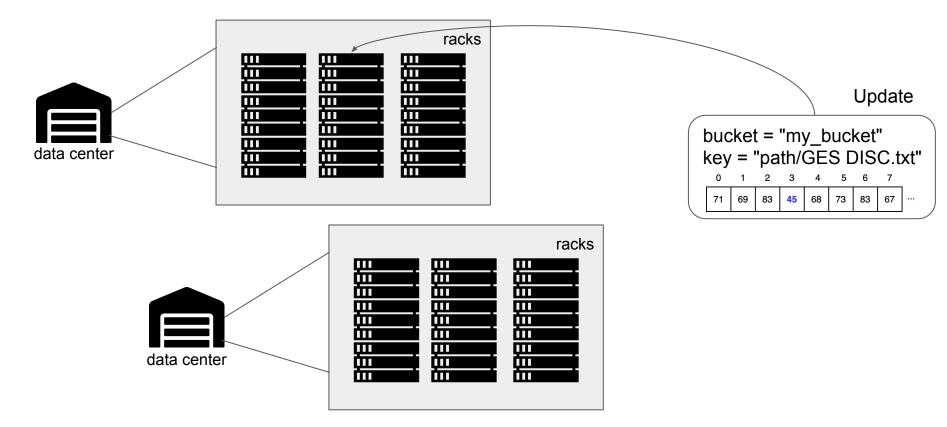


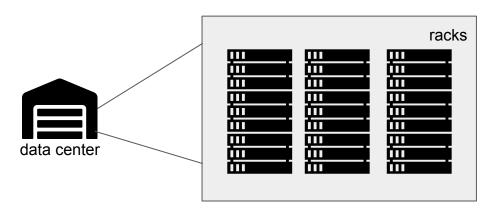




Update

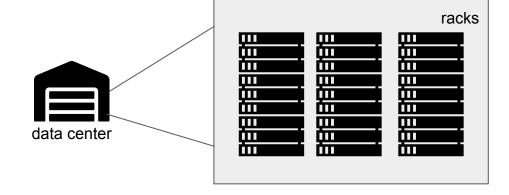


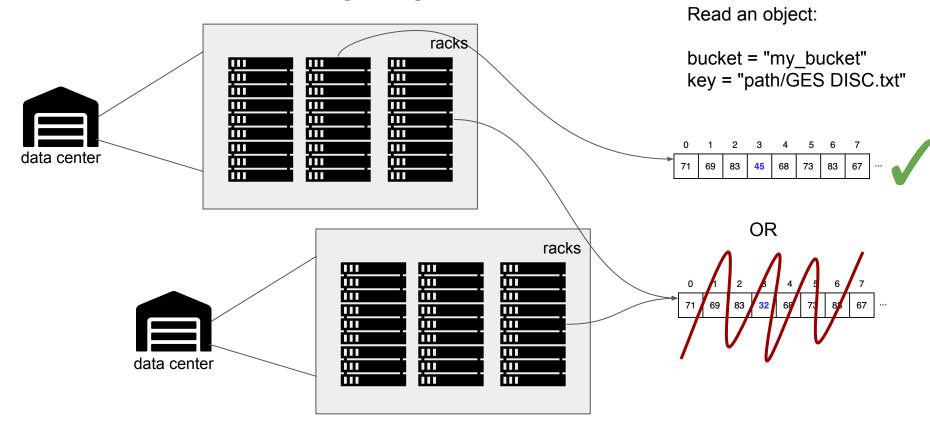




Read an object:

bucket = "my_bucket"
key = "path/GES DISC.txt"





So why use object storage?

S3 has a lot of very attractive characteristics.

- 1. cost: it's cheap
- 2. backups: happen automatically
- 3. parallel access: built in
- 4. network access: built in
- 5. versioning: built in
- 6. encryption: built in
- metadata: much more sophisticated than OS-level file metadata, including things like tagging
- 8. access control: much more sophisticated and fine-grained than file permissions

Questions?