

# Cloud-Optimized HDF5 Files

SC22, Dallas, TX  
November 16, 2022



Dana Robinson  
The HDF Group



# EOSDIS

NASA'S EARTH OBSERVING SYSTEM  
DATA AND INFORMATION SYSTEM



# Creating Cloud-Optimized HDF5 Files

2022 ESIP Summer Meeting

Aleksandar Jelenak

NASA EED-3/HDF Group

*[ajelenak@hdfgroup.org](mailto:ajelenak@hdfgroup.org)*

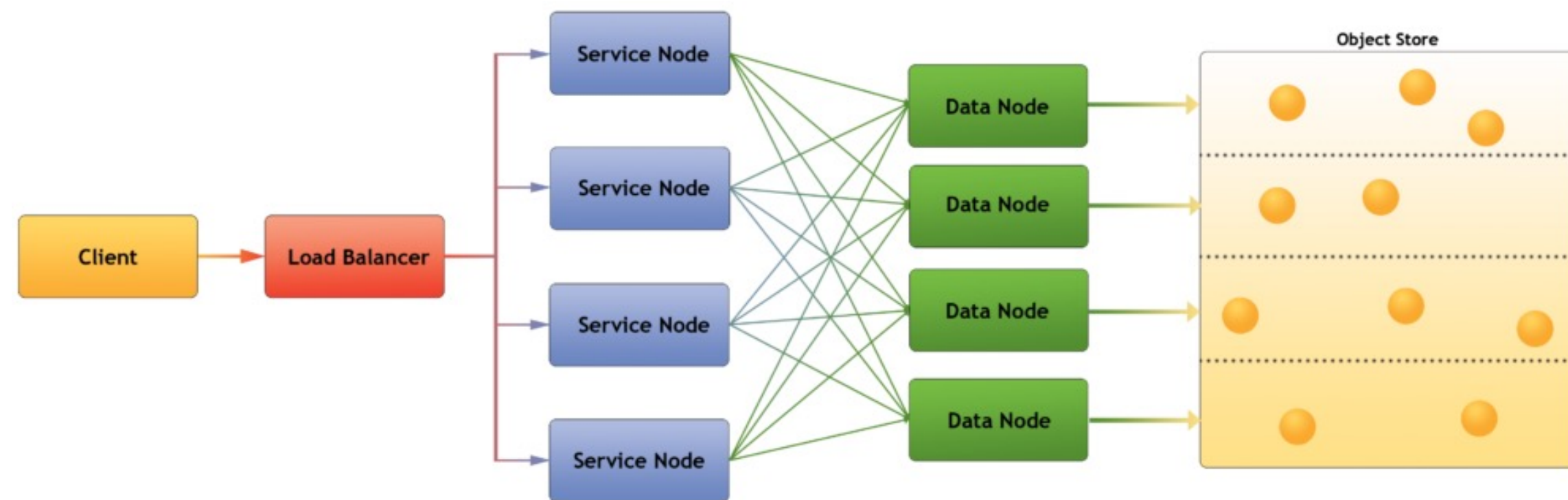
This work was supported by NASA/GSFC under Raytheon Technologies contract number 80GSFC21CA001.  
This document does not contain technology or Technical Data controlled under either the U.S. International  
Traffic in Arms Regulations or the U.S. Export Administration Regulations.

# Overview

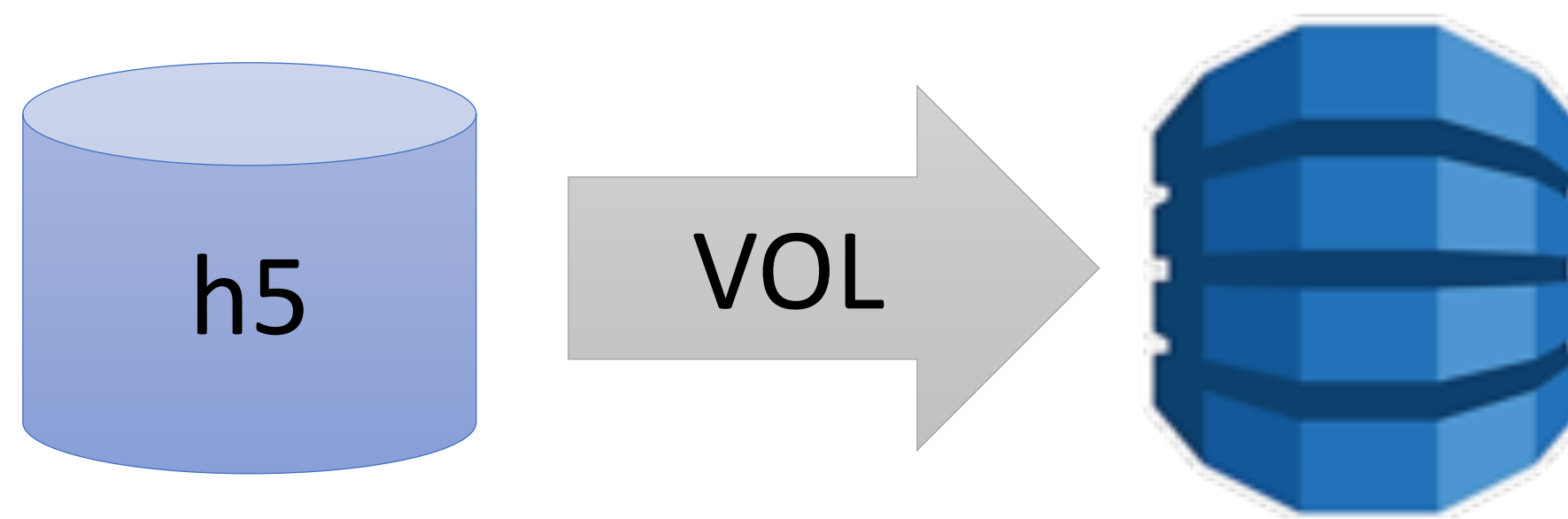


- What does HDF5 in the cloud mean?
- Cloud-optimized HDF5 Files

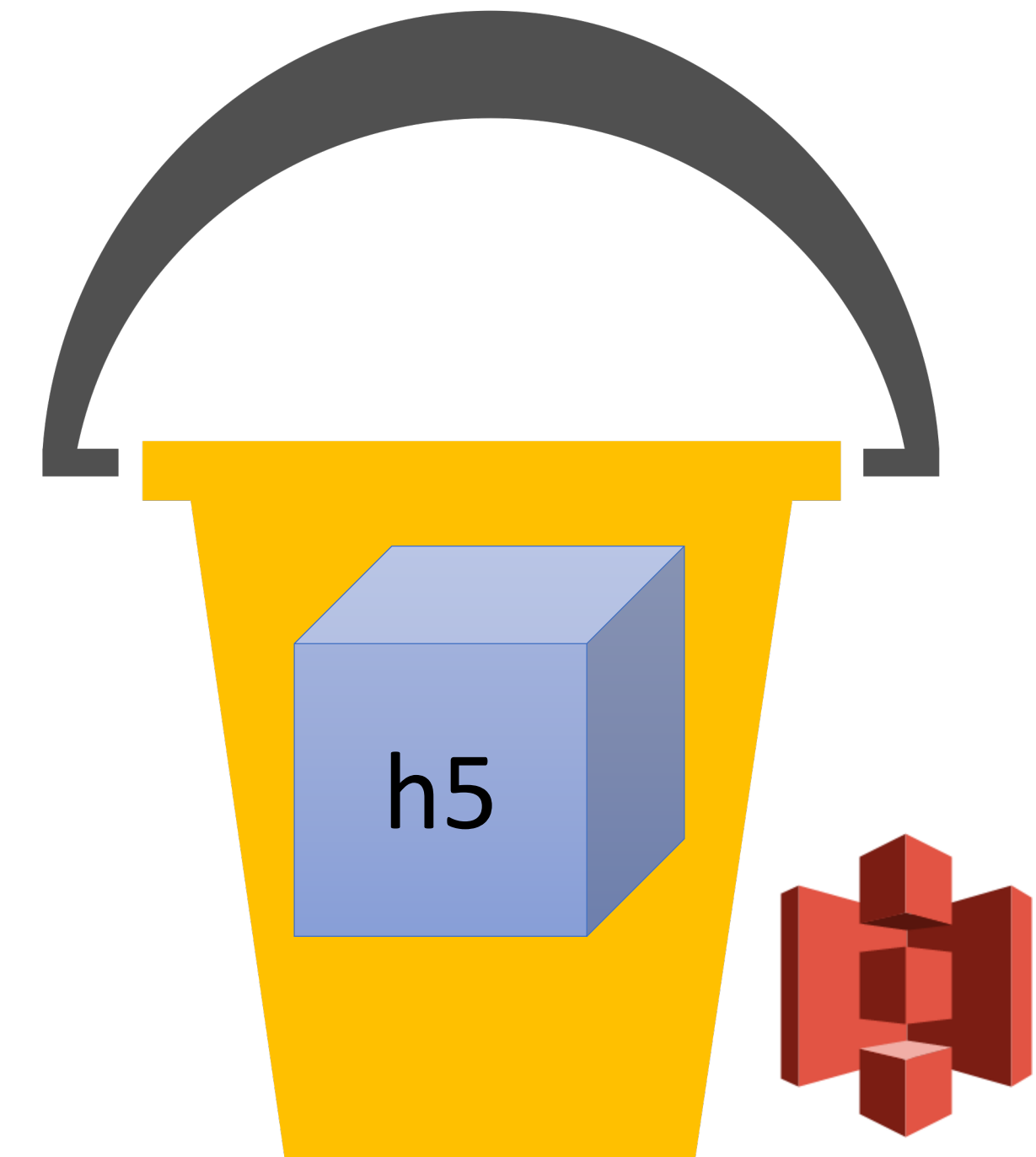
# What Does "Cloud-Optimized" HDF5 Mean?



Deconstructed into an object store like HSDS?

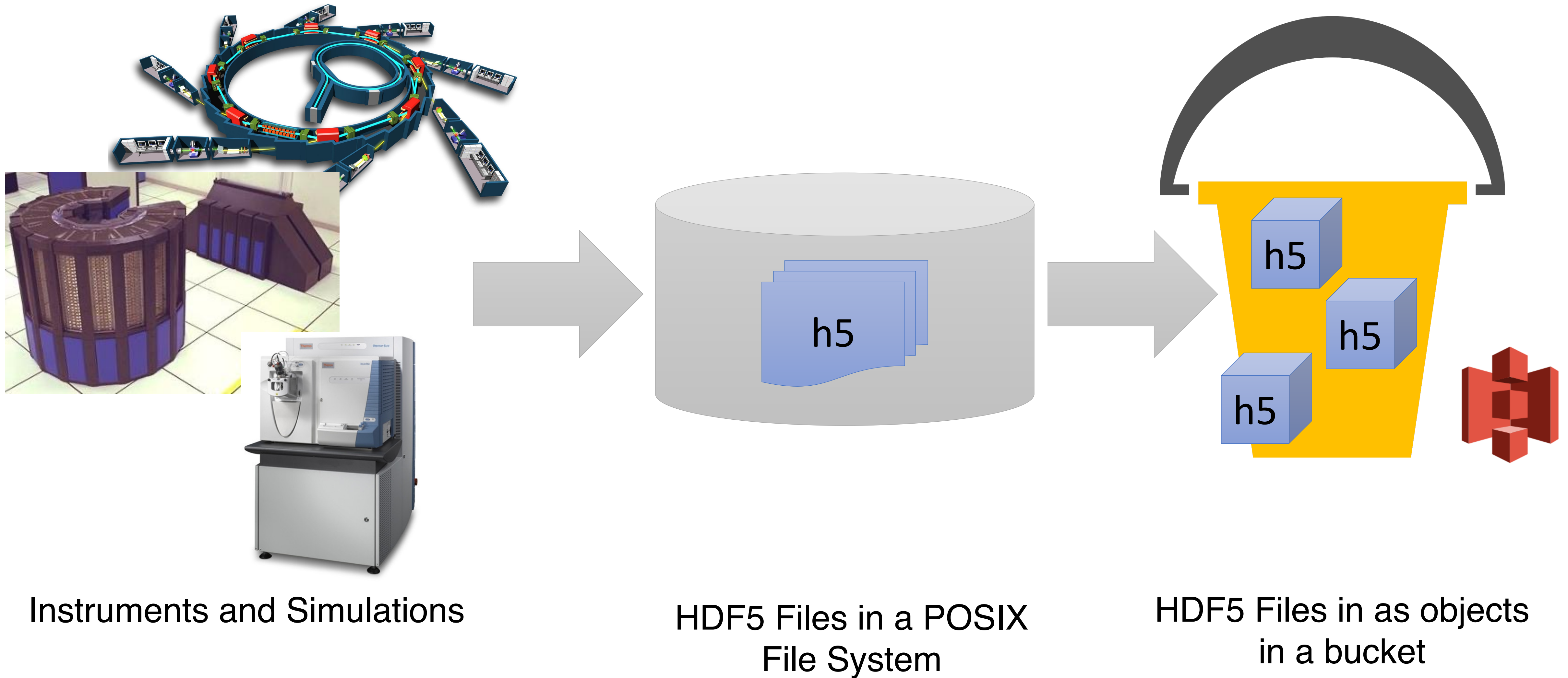


Something else?

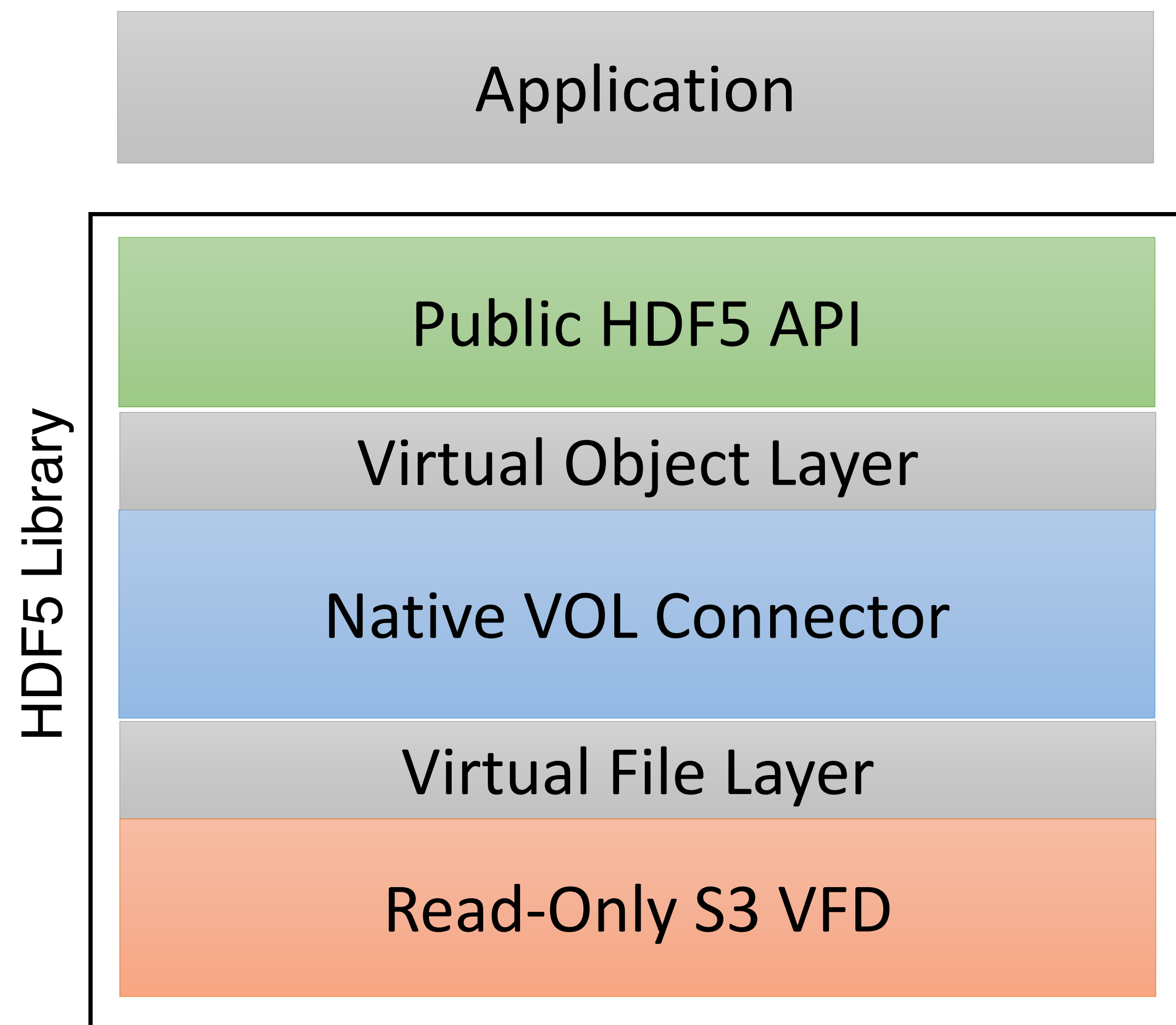


HDF5 file as an object?

# Focus: Storing Native Files In an Object Store

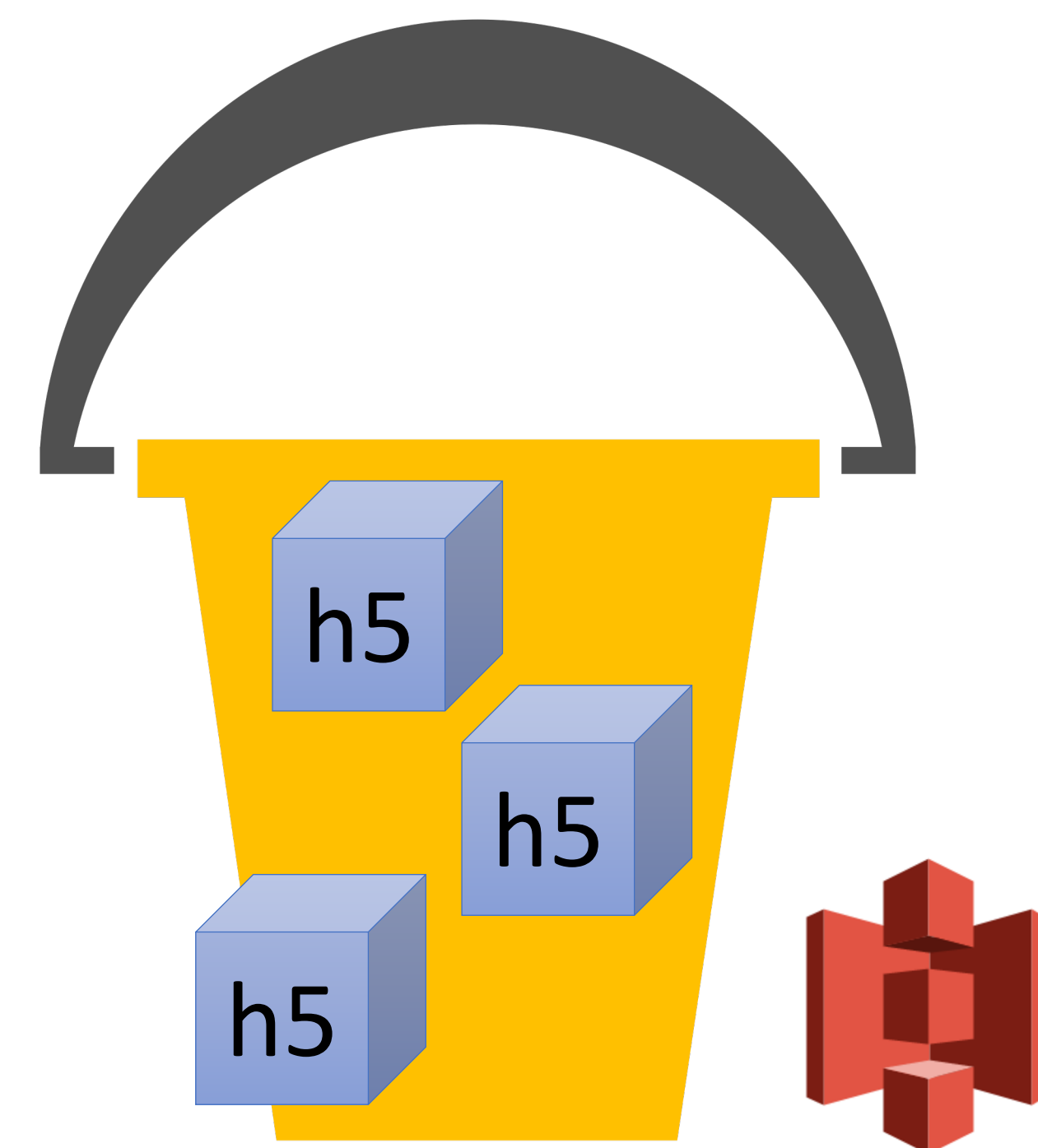


# How Do We Access Such a File?



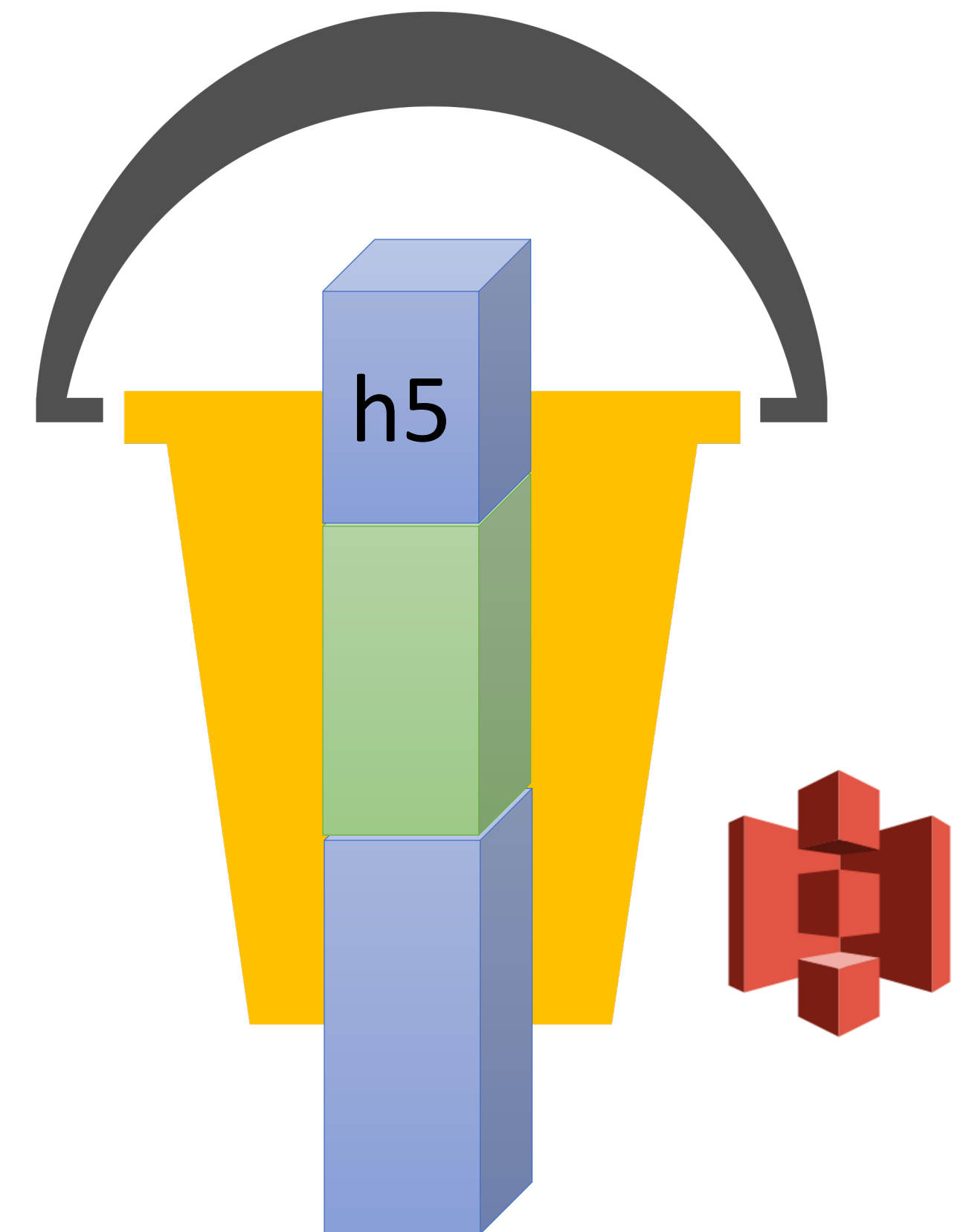
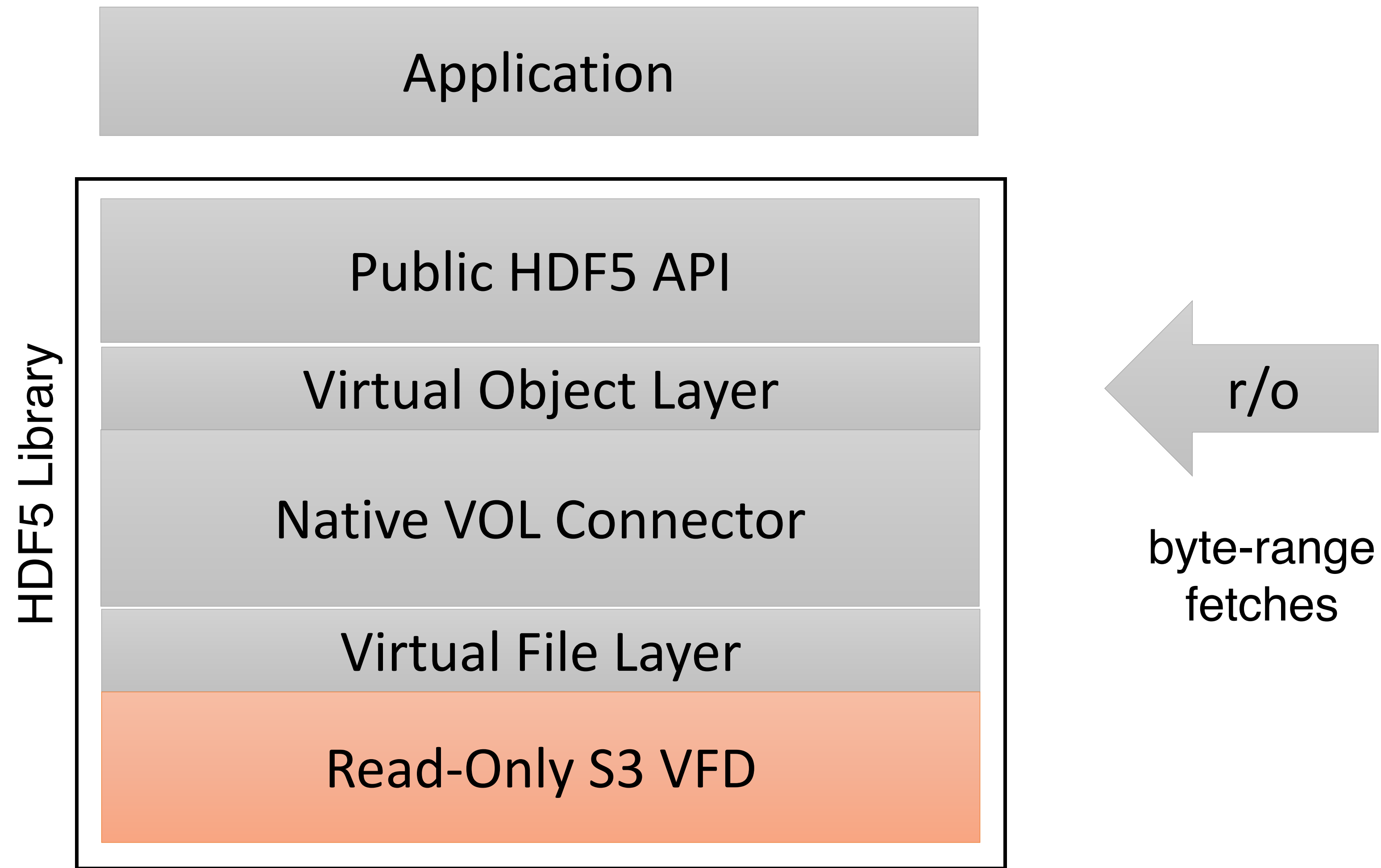
← r/o

(have to upload  
using s3 tools)



HDF5 Files in as objects  
in a bucket

# How Do We Access Such a File?



# Cloud-Optimized HDF5: Properties

- Two streams: data and metadata. Want to optimize both!
- Avoid lots of little I/O operations
- Avoid variable-length types
- If you have no control over the original file, repack with h5repack

**Metadata:** Use paged aggregation and page buffering

**Data:** Use larger chunk sizes\*, tune chunk cache & compression

\* Several MiB is a good place to start

# An Example HDF5 File



- ICESat-2 ATL03 product.
  - File size: 2,458,294,886 bytes (2.29 GiB)
  - 171 HDF5 groups
  - 1,001 HDF5 datasets
  - HDF5 file metadata size: **7,713,028 bytes**
- Repacked original file with two file space page sizes: 4 and 8 MiB.
  - 4MiB version larger by 4,345,472 bytes (+0.18%)
  - 8MiB version larger by 8,539,776 bytes (+0.35%)

# List Contents and Chunk Locations (AWS S3)

File Version	Page Buffer Size	Total Runtime	Page Buffer Stats
Original	N/A	<b>19 min 18.71 sec</b>	N/A
4 MiB Page	4 MiB	<b>22 min 23.86 sec</b>	accesses=6634, <b>hits=5691,</b> <b>misses=943,</b> <b>evictions=942</b> , bypasses=0
8 MiB Page	8 MiB	37.1 sec	accesses=6636, <b>hits=6635,</b> <b>misses=1, evictions=0,</b> bypasses=0
4 MiB Page	8 MiB	37.979 sec	accesses=6634, <b>hits=6632,</b> <b>misses=2, evictions=0,</b> bypasses=0
8 MiB Page	16 MiB	41.042 sec	accesses=6636, <b>hits=6635,</b> <b>misses=1, evictions=0,</b> bypasses=0
4 MiB Page	16 MiB	45.977 sec	accesses=6634, <b>hits=6632,</b> <b>misses=2, evictions=0,</b> bypasses=0

# List Contents and Chunk Locations (POSIX)



File Version	Page Buffer Size	Total Runtime	Page Buffer Stats
Original	N/A	43.945 sec	N/A
4 MiB Page	4 MiB	44.934 sec	accesses=6634, hits=5691, misses=943, evictions=942, bypasses=0
8 MiB Page	8 MiB	34.885 sec	accesses=6636, hits=6635, misses=1, evictions=0, bypasses=0
4 MiB Page	8 MiB	33.523 sec	accesses=6634, hits=6632, misses=2, evictions=0, bypasses=0
8 MiB Page	16 MiB	36.089 sec	accesses=6636, hits=6635, misses=1, evictions=0, bypasses=0
4 MiB Page	16 MiB	32.423 sec	accesses=6634, hits=6632, misses=2, evictions=0, bypasses=0

# What To Take Away From This



- The cloud is not a local POSIX file system
- You may have to adjust your best practices to make things work well in the cloud

# THANK YOU!

Questions & Comments?