Key Considerations for Storing Data on the Cloud

Tom Augspurger

NOAA EDMW 2022

September 2022

Premise: *Everyone* has access to *all* the data

Metadata

1. Standardize metadata is essential

Premise: Everyone has access to all the data

Storage

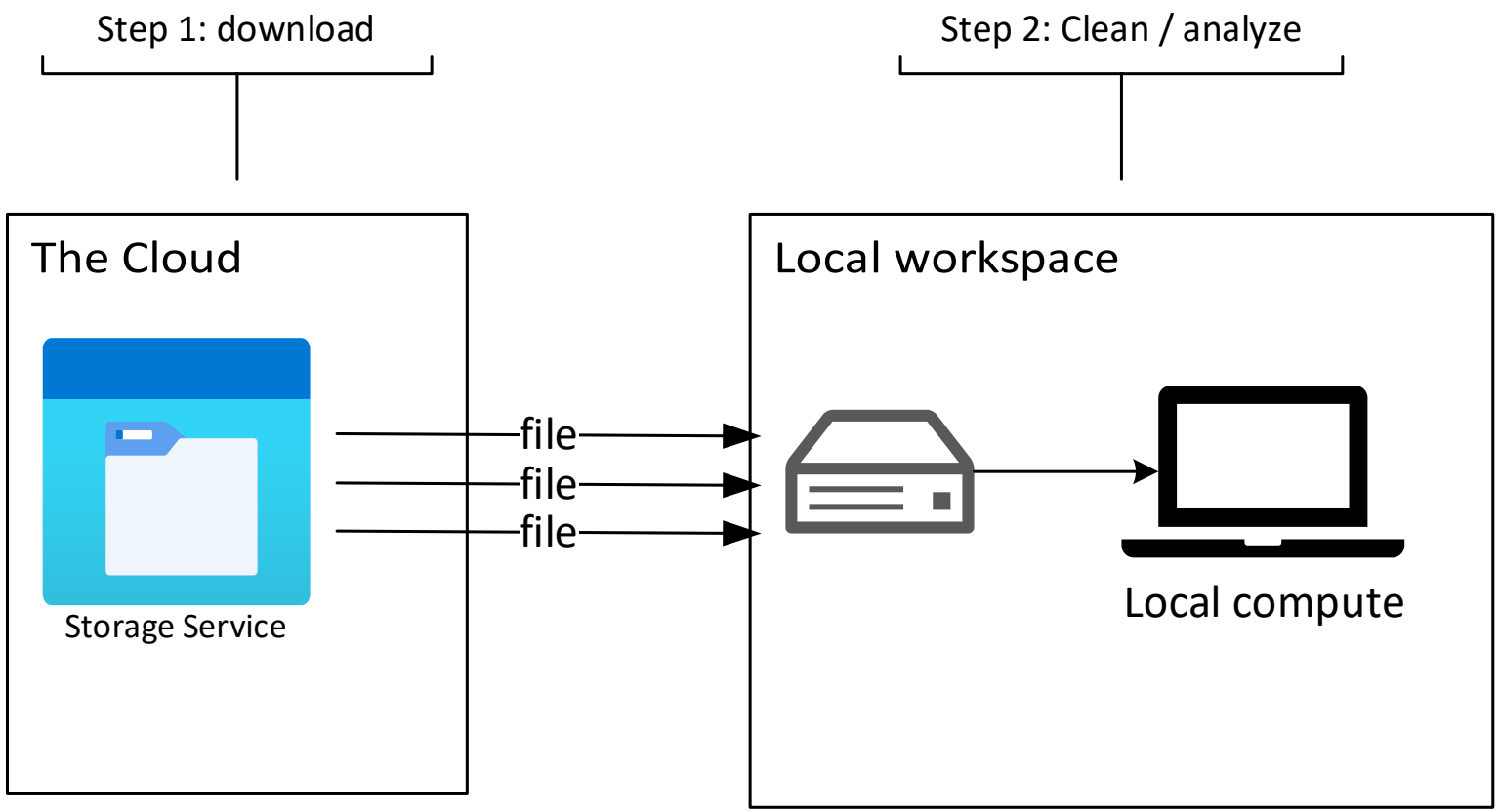
2. Data are stored in a Blob Storage service

Workflows

3. The “download model” breaks down at scale

Premise: Everyone has access to all the data

The “Download” Model

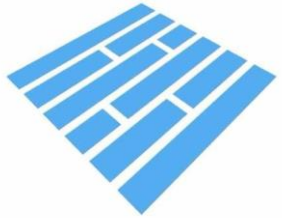


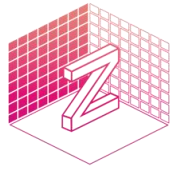
File formats

4. Cloud-friendly file formats are important

Cloud-friendly file formats

- Metadata compactly in a convenient place
- Somehow support chunking / tiling
- Combine with HTTP range requests
  - Give me bytes 100-200 from the file at https://...





Cloud-friendly *clients*

> One of the quiet secrets of the “cloud optimized” geospatial world is that, while all the attention is placed on the formats, the actual **really really hard** part is writing the clients that can efficiently make use of the carefully organized bytes.

- Paul Ramsey (http://blog.cleverelephant.ca/2022/04/coshp.html via Pete Gadomski)

Thanks!

https://planetarycomputer.microsoft.com

https://github.com/TomAugspurger/noaa-edmw-2022

taugspurger@microsoft.com