

Open, Cloud-optimized, Analysis-ready Global Gedi Satellite Lidar Datasets For Land Surface Applications

Yu-Feng Ho (OpenGeoHub)
Johannes Heisig (University of Münster)
Milutin Milenkovic (IIASA)

29 Sep 2025



**OPEN EARTH
MONITOR**



This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No.101059548

A background image of a person with blonde hair, seen from the side, looking through a large telescope. The scene is outdoors, with palm trees and a fence visible in the background.

SAR

Optical

Spaceborne
Lidar

Scientific Community



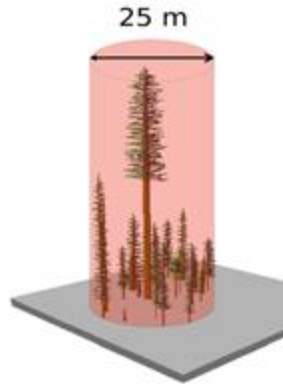
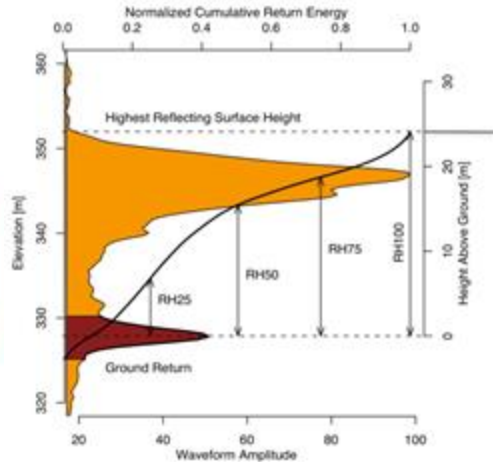


Big Data From Space - GEDI

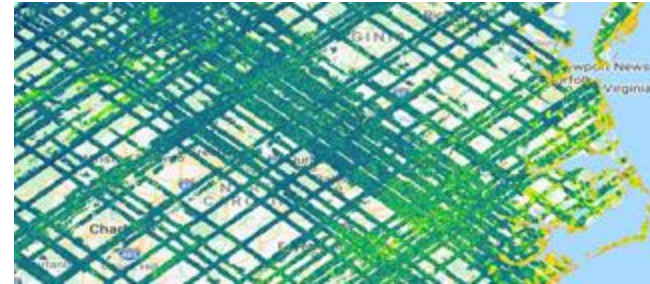


Global Ecosystem Dynamics Investigation (GEDI)

- Full-waveform LiDAR system
- Data since 2019
- N ~ 5.4 billion points
- Level 2A+B data, 94 variables
- Point geometries (lon/lat)



Sampling pattern



<https://gedi.umd.edu/mission/technology/>
<https://www.earthdata.nasa.gov/news/gridded-level-3-data-product-facilitates-use-gedi-mission-data>

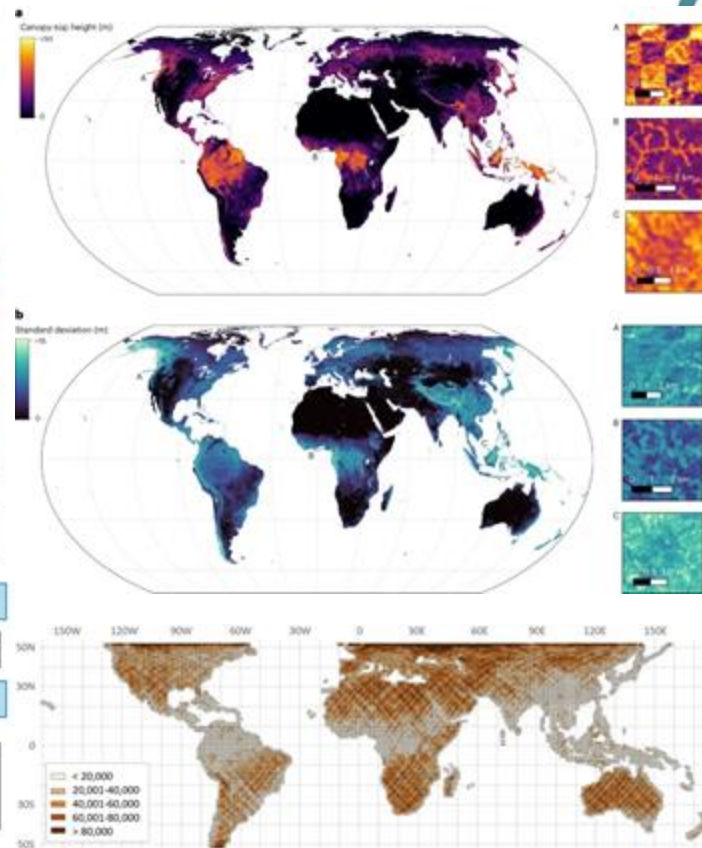
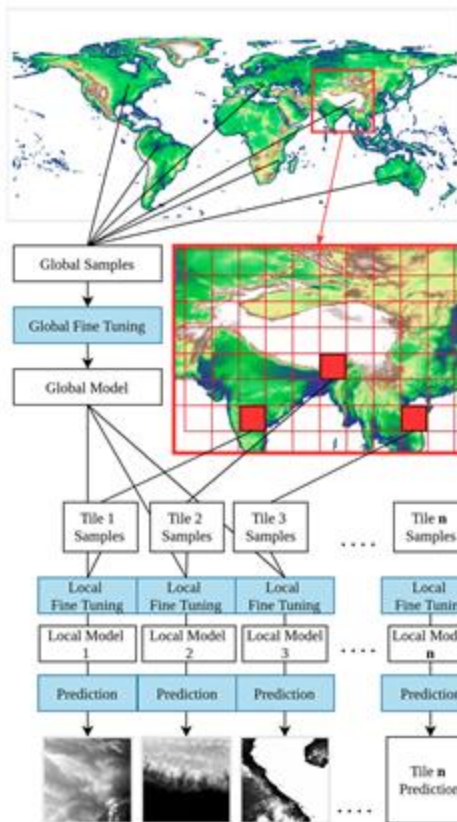


Terrain, Canopy height mapping

Ho, Y. F., Grohmann, C. H., Lindsay, J., Reuter, H. I., Parente, L., Witjes, M., & Hengl, T. (2025). GEDTM30: global ensemble digital terrain model at 30 m and derived multiscale terrain variables. *PeerJ*, 13, e19673.

<https://peerj.com/articles/19673/>

Potapov, P., Li, X., Hernandez-Serna, A., Tyukavina, A., Hansen, M. C., Kommareddy, A., ... & Hofton, M. (2021). Mapping global forest canopy height through integration of GEDI and Landsat data. *Remote Sensing of Environment*, 253, 112165. <https://doi.org/10.1016/j.rse.2020.112165>



Lang, N., Jetz, W., Schindler, K., & Wegner, J. D. (2023). A high-resolution canopy height model of the Earth. *Nature Ecology & Evolution*, 7(11), 1778-1789.

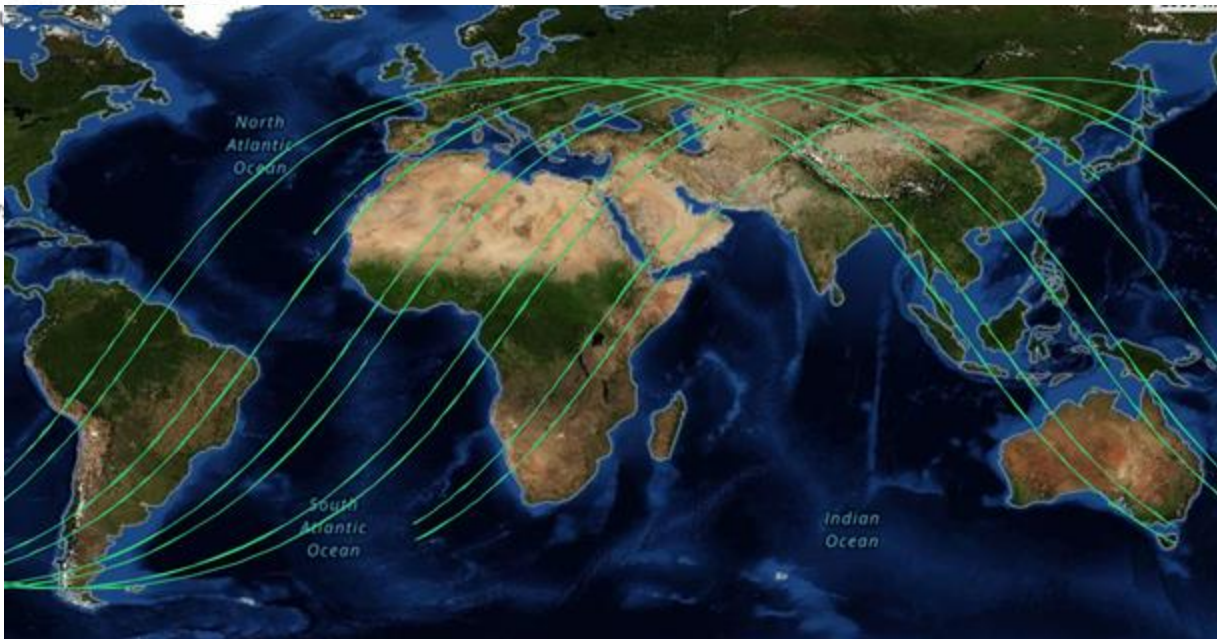
<https://www.nature.com/articles/s41559-023-02206-6>



File and Data Structure

In this example of a Version 2 Level 01B product, the filename **GEDI01_B_2019110110221_O01997_03_T03335_02_005_01_V002.h5** indicates:

- **GEDI01_B** = Product Short Name
- **2019110** = Julian Date of Acquisition in YYYYDDD
- **110221** = Hours, Minutes, and Seconds of Acquisition
- **O01997** = O = Orbit, 01997 = Orbit Number
- **03** = Sub-Orbit Granule Number
- **T03335** = T = Track, 03335 = Track Number
- **02** = Positioning and Pointing Determination System
- **005** = PGEVersion = SDPS Release Number
- **01** = Granule Production Version
- **V002** = Version Number
- **.h5** = Data Format



GEDI
ECOSYSTEM LIDAR





File and Data Structure

Search Results (0 Collections)

Showing 20 of 645 matching granules

Sort: Start Date (Newest) View: List

GED102_A_2025050180535_O350 64_02_T03388_02_004_02_V002	GED102_A_2025049032457_O350 39_03_T00012_02_004_02_V002
START 2025-02-19 18:05:35 END 2025-02-19 19:30:24	START 2025-02-18 03:24:57 END 2025-02-18 04:57:47
GED102_A_2025046194303_O350 03_02_T01353_02_004_02_V002	GED102_A_2025045050211_O349 78_03_T09361_02_004_02_V002
START 2025-02-15 19:43:03 END 2025-02-15 21:15:53	START 2025-02-14 05:02:11 END 2025-02-14 06:35:01
GED102_A_2025041063859_O349 12_03_T06209_02_004_02_V002	GED102_A_2025039220915_O348 96_02_T11375_02_004_02_V002
START 2025-02-18 06:38:59 END 2025-02-18 08:11:49	START 2025-02-08 22:09:15 END 2025-02-08 23:42:05
2025037081519_O348 02_02_T00000_02_004_02_V002	GED102_A_2025035234525_O348 35_02_T05071_02_004_02_V002
START 2025-02-06 09:15:19 END 2025-02-06 09:48:19	START 2025-02-04 23:45:25 END 2025-02-05 01:18:18
GED102_A_2025032012127_O3477 4_02_T09034_02_004_02_V002	GED102_A_2025030104131_O3474 8_03_T11197_02_004_02_V002
START 2025-02-01 12:12:12 END 2025-02-01 12:12:12	START 2025-02-01 04:13:11 END 2025-02-01 04:13:11

Download All 645

600+ files



Data are
largely outside
of Area of
Interest!



OPEN EARTH
MONITOR

Parquet



Parquet



<https://www.nasdaq.com/articles/8-types-of-flooring-2021-03-15>



<https://vertexeng.com/insights/hardwood-flooring-assessing-common-causes-of-damage/>



Parquet



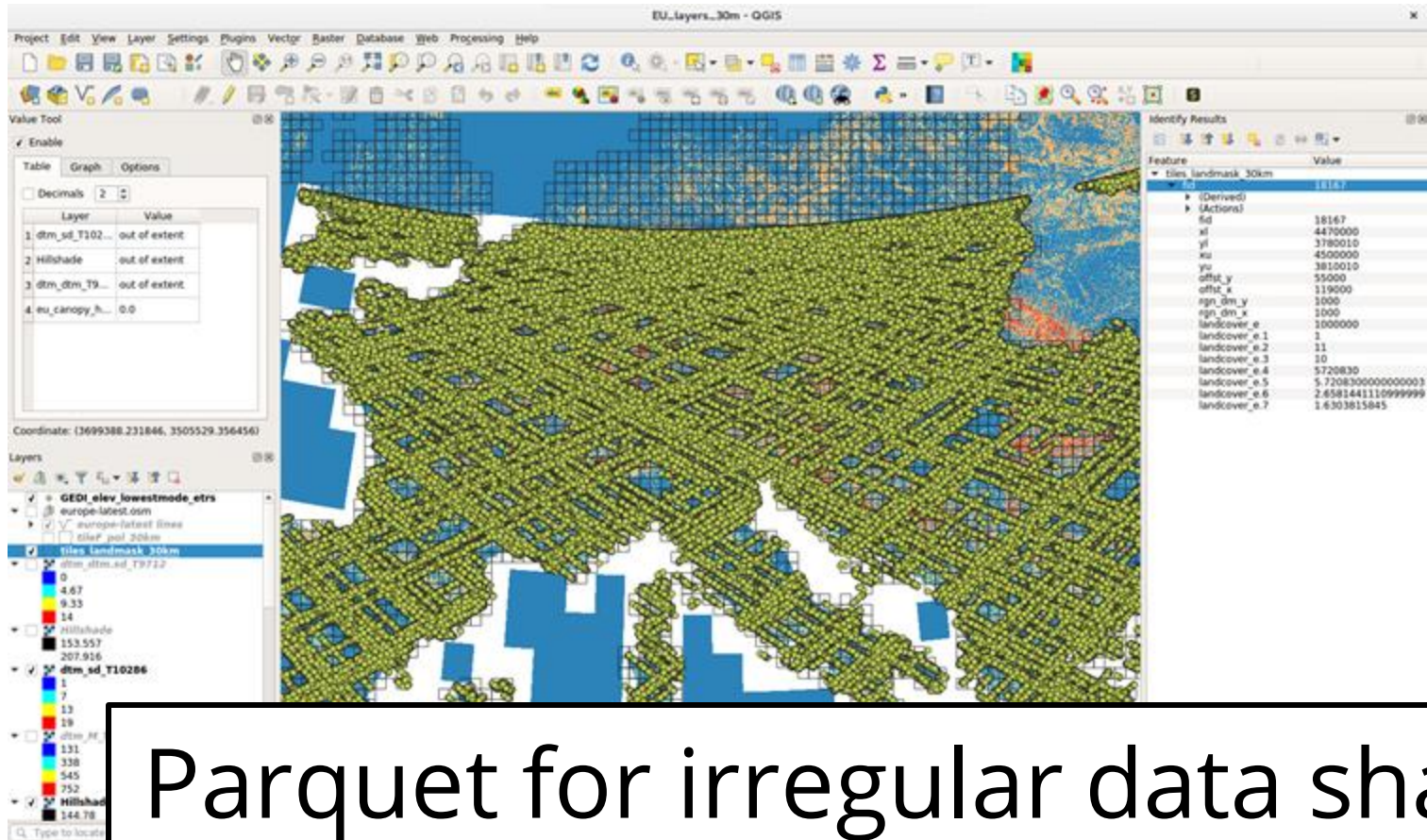
Tile

<https://www.nasdaq.com/articles/8-types-of-flooring-2021-03-15>



Parquet

<https://vertexeng.com/insights/hardwood-flooring-assessing-common-causes-of-damage/>



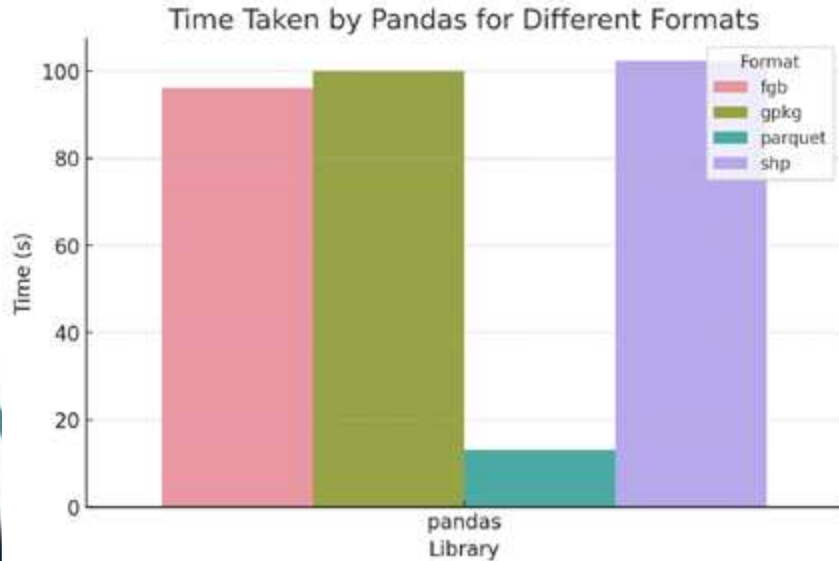
Parquet for irregular data shape



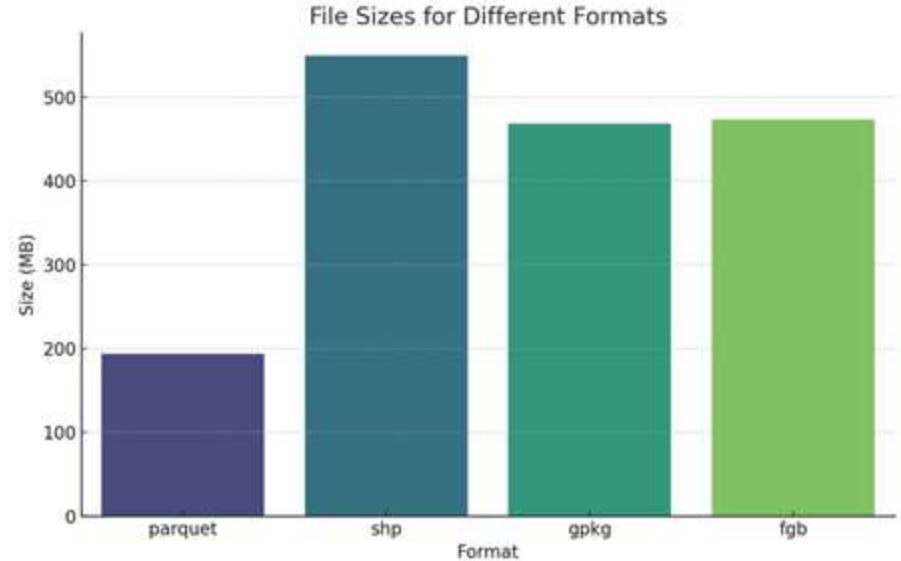
Apache Parquet is an open source, column-oriented data file format designed for efficient data storage and retrieval.

- *Feature 1: Compression*
- *Feature 2: Partitioning*
- *Feature 3: Lazy loading*

Parquet Feature: *Compression*



Fast access

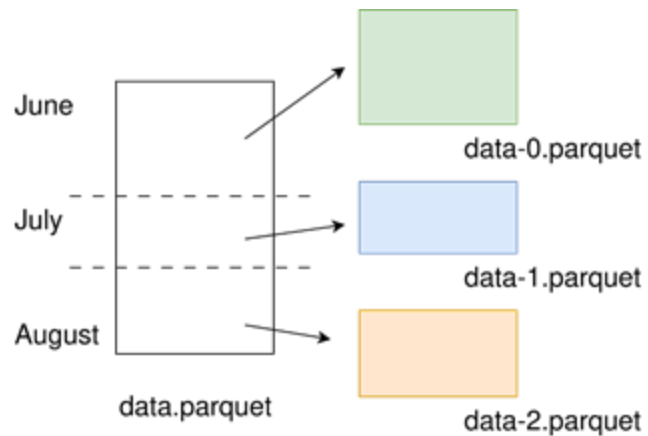


Compressed

Source: <https://cloudnativegeo.org/blog/2023/08/performance-explorations-of-geoparquet-and-duckdb/>

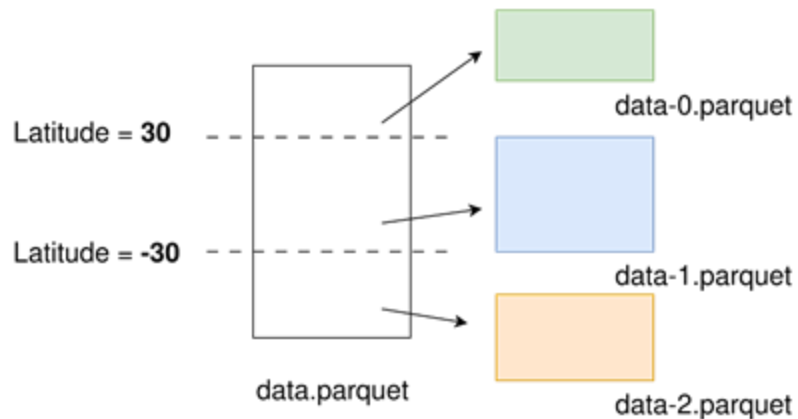
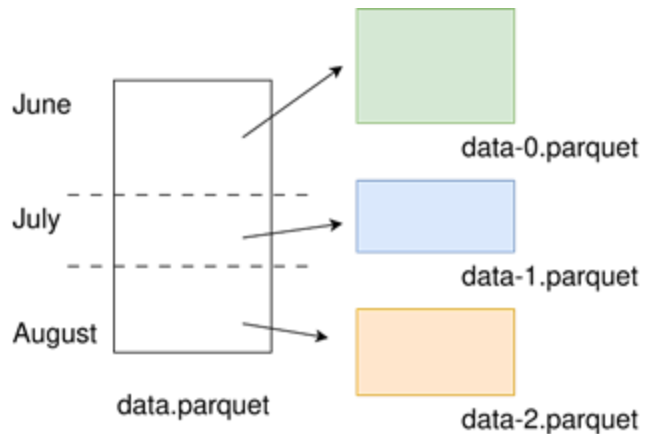


Parquet Feature: *Partitioning*



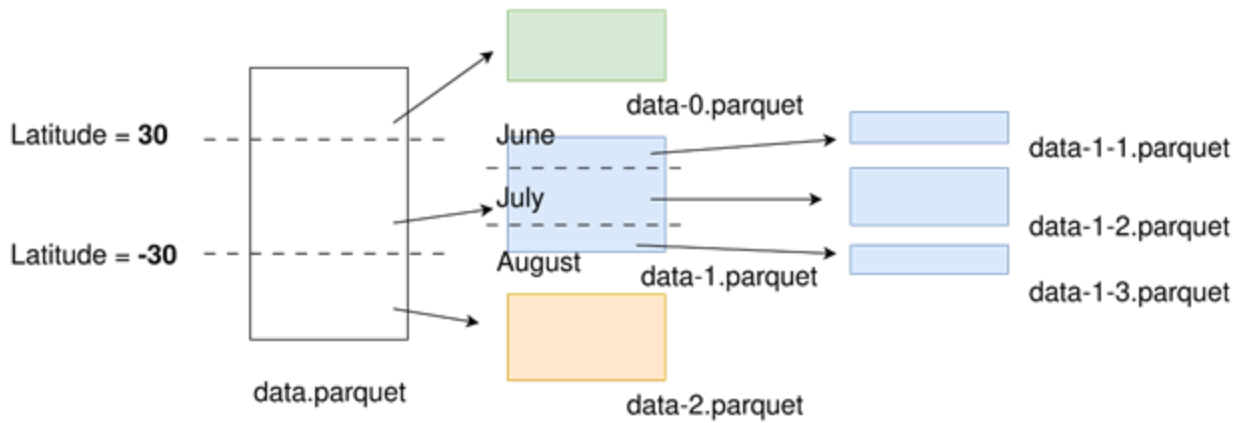
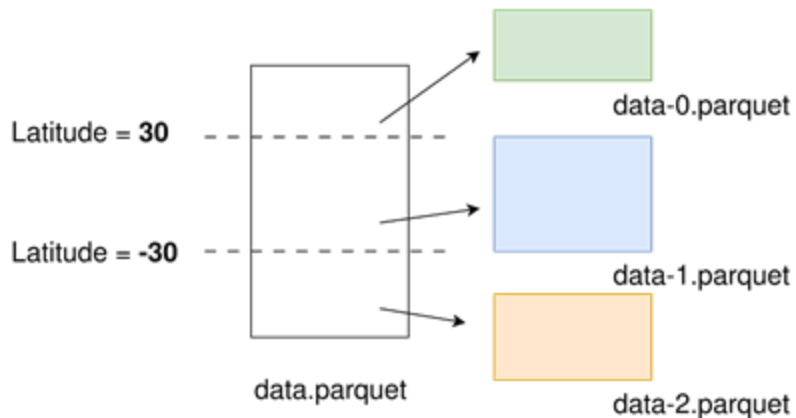
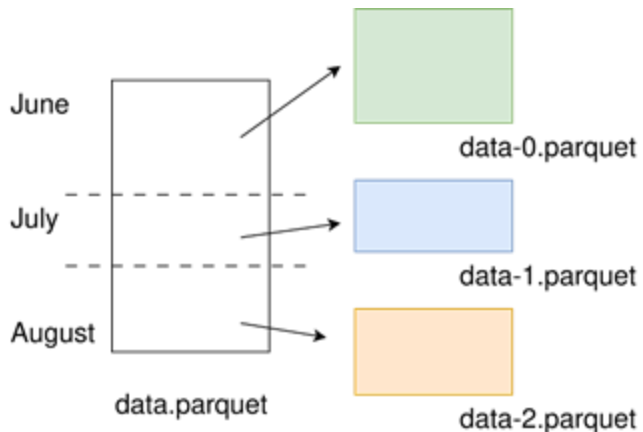


Parquet Feature: *Partitioning*





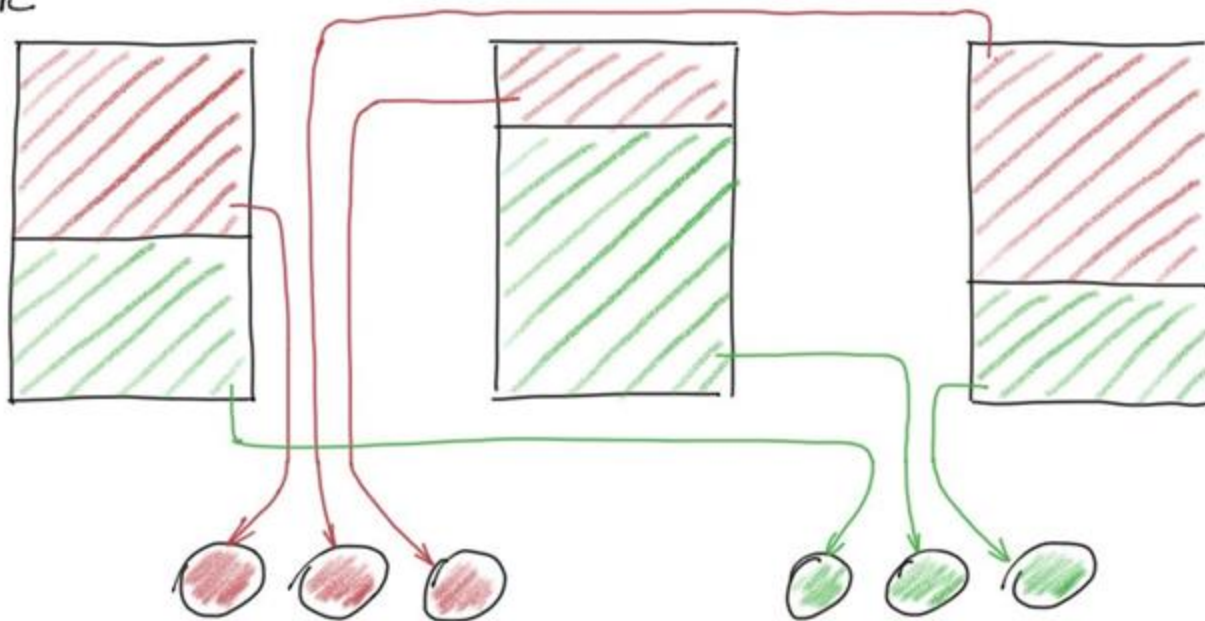
Parquet Feature: *Partitioning*





Reconstruct Global Lidar by Apache Parquet

Spark

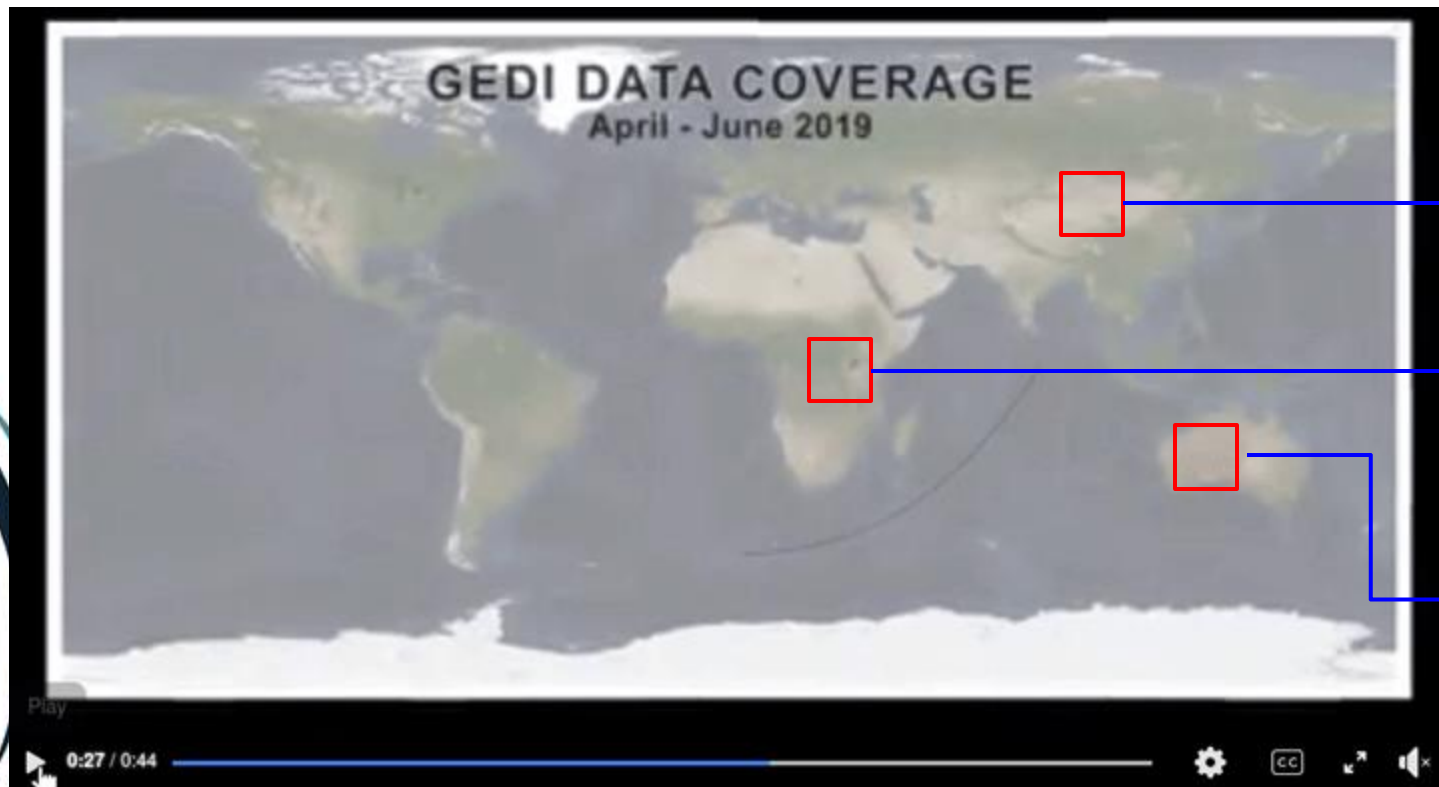


S3

Source: <https://medium.com/@vladimir.prus/spark-partitioning-the-fine-print-5ee02e7cb40b>



Reconstruct Global Lidar by Apache Parquet



partition-1

- gedi-o1a.parquet
- gedi-o2a.parquet

partition-2

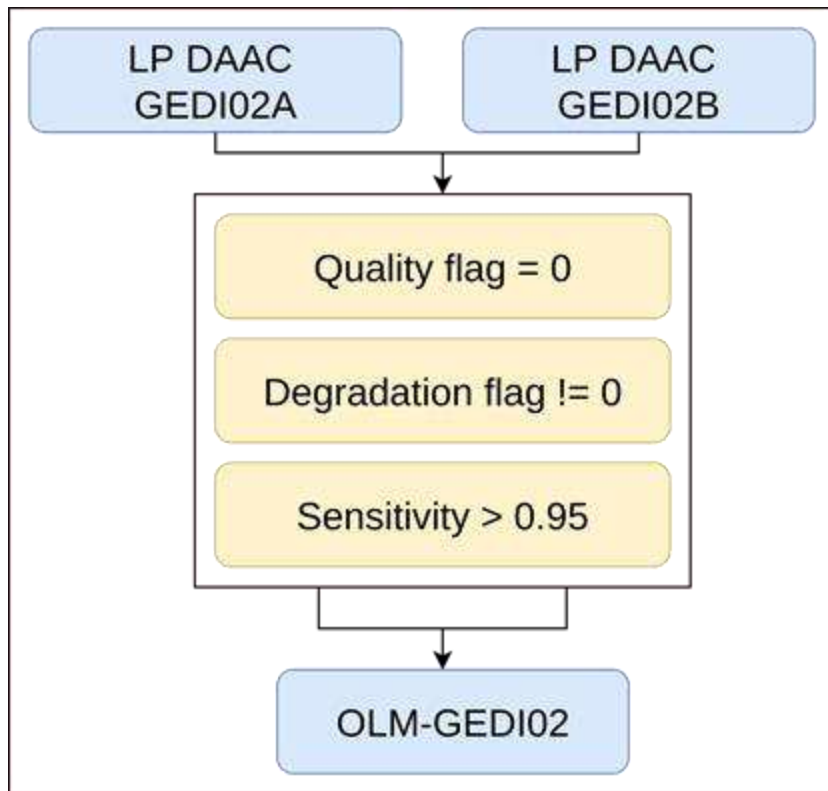
- gedi-o1b.parquet
- gedi-o2b.parquet
- gedi-o3a.parquet
- gedi-o4a.parquet

partition-3

- gedi-o3b.parquet
- gedi-o4b.parquet
- gedi-o5a.parquet



Getting closer to ARD



Processing Pipeline



HPC infrastructure

Table 1. Summary of raw GEDI and OpenLandMap GEDI.

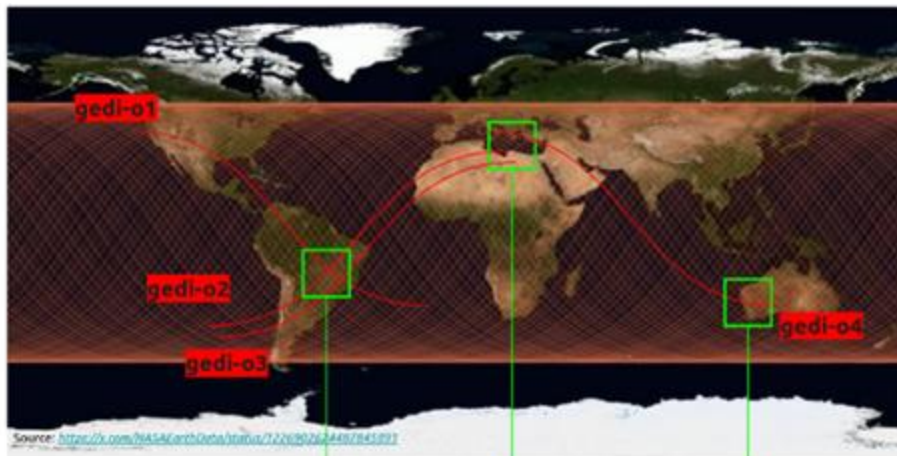
Product Name	LP DAAC GEDI02	OLM-GEDI
Time Coverage	03.25.2019 ~ 03.15.2023	
Data Size	- Level 2A: 101.8 TB - Level 2B: 22.0 TB	Level 2 A&B: 759 GB
Data Format	Hierarchical Data Format (HDF)	GeoParquet

Result



Spatio-Temporal Blocking

E.g. restructure GEDI level 2 to spatio-temporal blocks



1. Streaming

- partition-1**
 - gedi-o1a.parquet
 - gedi-o2a.parquet
 - gedi-o3a.parquet
- partition-2**
 - gedi-o2b.parquet
 - gedi-o3b.parquet
 - gedi-o4b.parquet
- partition-3**
 - gedi-o4b.parquet

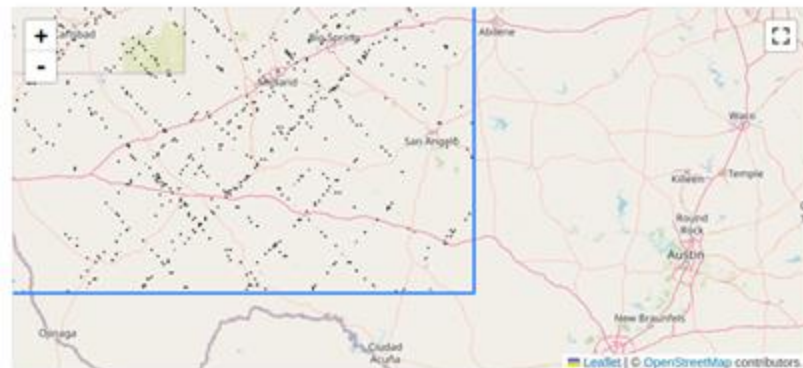
2. Aggregation

- partition-1**
 - gedi_l2ab.parquet
- partition-2**
 - gedi_l2ab.parquet
- partition-3**
 - gedi_l2ab.parquet

lon=-105_lat=30_year=2023_gedi_l2ab

in OpenLandMap STAC

[Up](#) [Collection](#) [Browse](#)



Assets

> GeoParquet

DATA

> Overview

✓ SHOWN

OVERVIEW

THUMBNAIL

COG

<https://stac.openlandmap.org/GEDI02/collection.json>

OpenLandMap GEDI level2

in OpenLandMap STAC [Up](#) [Browse](#)

[Source](#) [Share](#) [Language: English](#)

Description

Global Ecosystem Dynamics Investigation (GEDI) is a full-waveform satellite LIDAR (Light Detection And Ranging), collecting 3D measurements over land (through recorded backscattered laser energy, i.e., waveforms) near-globally, between 51.6° N and S within a laser footprint diameter of 25 m between 2019-03-25 and 2023-03-15. To foster the uptake and mask out the complexity and large data volume of GEDI products (GEDI L2, version 2, - V002) to the OEMC and broader users, we aim to host pre-filtered and high-quality GEDI points (including their relative heights, and waveform- and derived-attributes) and offer them as a cloud service.

[gedi](#) [level2](#) [in-situ data](#) [lidar](#) [canopy height](#) [canopy structure](#) [terrain height](#)

License CC-BY-4.0
Temporal Extent 3/25/2019, 12:00:00 AM UTC - 3/15/2023, 12:00:00 AM UTC



Asset

Items 3336 [Ascending](#) [Descending](#)

lon=-10_lat=0_year=2019_gedi_l2ab

4/18/2019, 6:55:14 AM UTC - 12/18/2019, 5:33:37 PM UTC

lon=-100_lat=50_year=2019_gedi_l2ab

4/19/2019, 5:13:31 PM UTC - 12/18/2019, 6:47:35 PM UTC

lon=-10_lat=0_year=2020_gedi_l2ab

1/9/2020, 9:12:51 PM UTC - 12/24/2020, 2:28:47 PM UTC

lon=-100_lat=50_year=2020_gedi_l2ab

1/8/2020, 8:22:19 AM UTC - 12/31/2020, 1:22:20 PM UTC

lon=-10_lat=0_year=2021_gedi_l2ab

1/5/2021, 9:47:02 AM UTC - 12/16/2021, 5:46:15 AM UTC

lon=-100_lat=50_year=2021_gedi_l2ab

1/1/2021, 10:57:31 AM UTC - 12/31/2021, 1:03:49 PM UTC

lon=-10_lat=0_year=2022_gedi_l2ab

1/5/2022, 9:25:53 AM UTC - 12/29/2022, 11:52:55 PM UTC

lon=-100_lat=50_year=2022_gedi_l2ab

1/1/2022, 10:38:31 AM UTC - 12/31/2022, 11:52:03 AM UTC

lon=-10_lat=0_year=2023_gedi_l2ab

1/2/2023, 10:16:58 PM UTC - 2/28/2023, 11:46:12 PM UTC

lon=-100_lat=50_year=2023_gedi_l2ab

1/1/2023, 11:04:11 AM UTC - 3/15/2023, 6:15:01 AM UTC

lon=-10_lat=10_year=2019_gedi_l2ab

4/19/2019, 5:34:59 PM UTC - 12/17/2019, 6:50:19 AM UTC

lon=-105_lat=15_year=2019_gedi_l2ab

4/20/2019, 1:03:29 PM UTC - 12/16/2019, 11:42:54 PM UTC



Parquet Feature: *Lazy loading*

Partition Pruning



a	b	c
a5	b5	c5
a6	b6	c6
a7	b7	c7
a8	b8	c8
a9	b9	c9



a	b	c
a5	b5	c5
a6	b6	c6
a7	b7	c7
a8	b8	c8
a9	b9	c9



a	b	c
a0	b0	c0
a1	b1	c1
a2	b2	c2
a3	b3	c3
a4	b4	c4



a	b	c
a5	b5	c5
a6	b6	c6
a7	b7	c7
a8	b8	c8
a9	b9	c9

+

Column selection

a	b	c
a0	b0	c0
a1	b1	c1
a2	b2	c2
a3	b3	c3
a4	b4	c4

Predicate Pushdown

a	b	c
a0	b0	c0
a1	b1	c1
a2	b2	c2
a3	b3	c3
a4	b4	c4

+

=

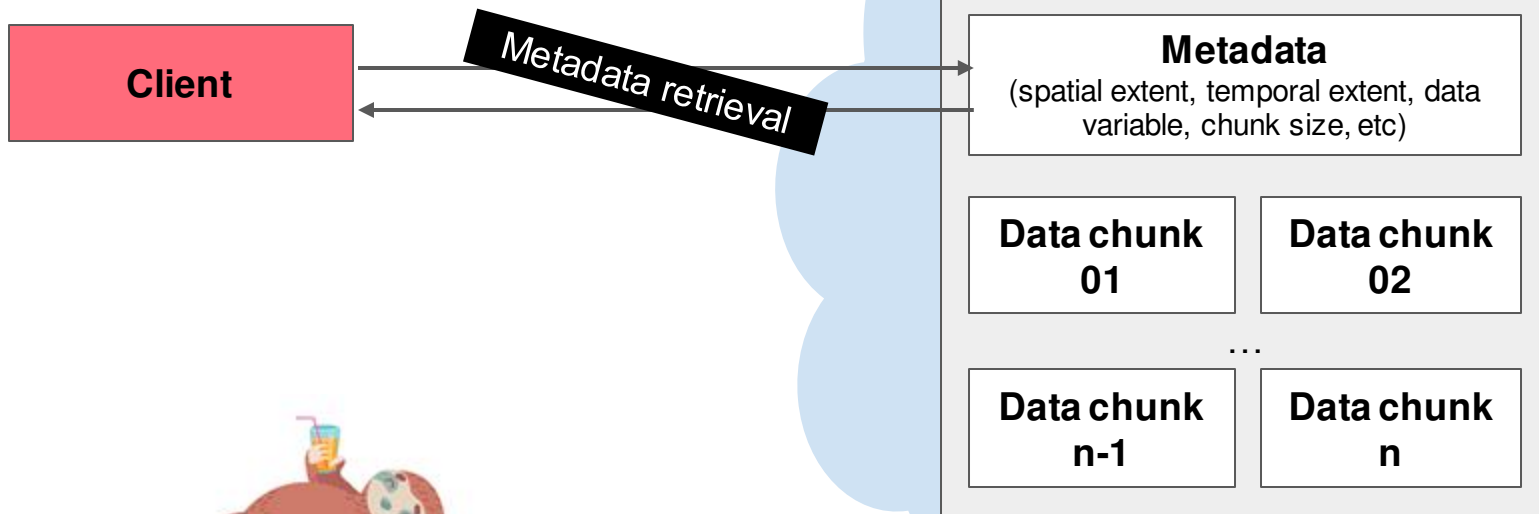
Read only the
data you need

a	b	c
a0	b0	c0
a1	b1	c1
a2	b2	c2
a3	b3	c3
a4	b4	c4

<https://docs.vertica.com/11.1.x/en/data-load/working-with-external-data/querying-external-tables/improving-query-performance/>

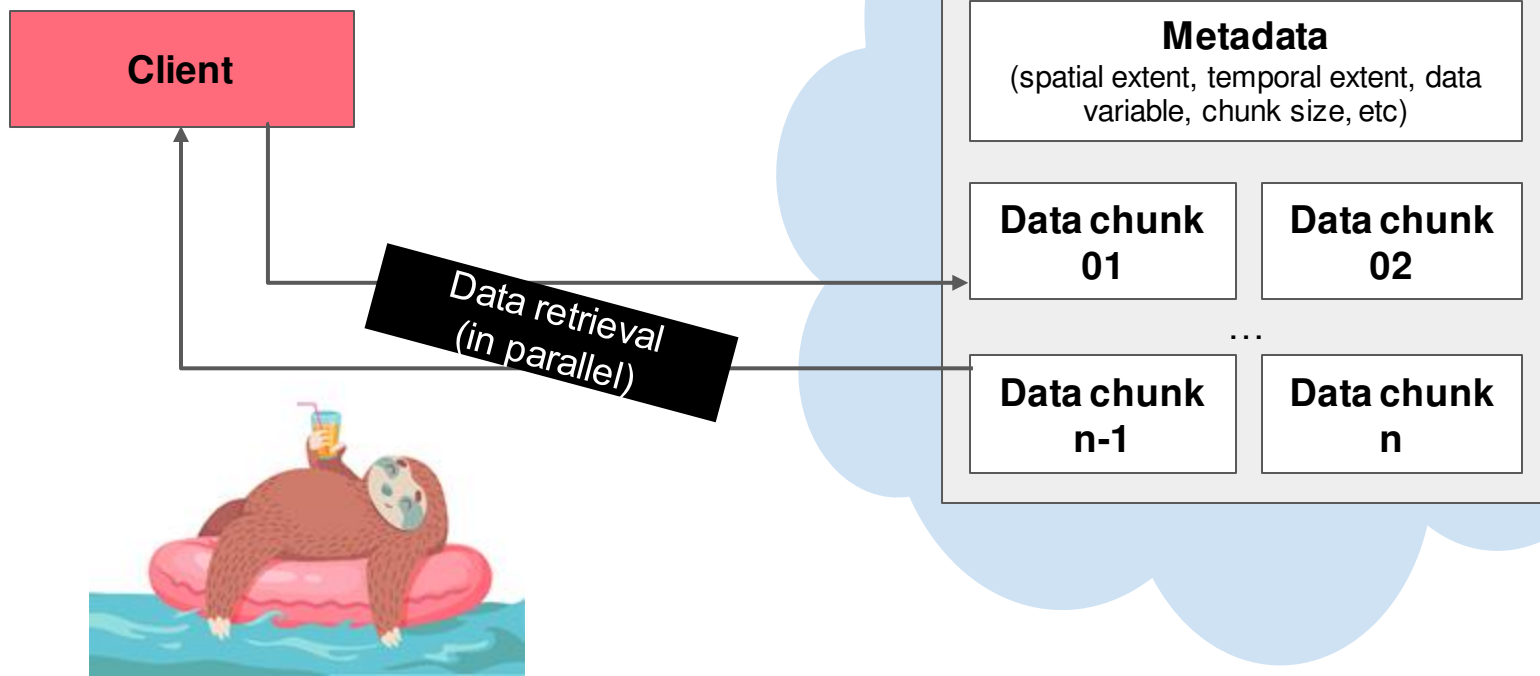
Parquet Feature: *Lazy loading*

It enables “**lazy loading**” data.



Parquet Feature: *Lazy loading*

It enables “**lazy loading**” data.



Parquet Feature: *Lazy loading*

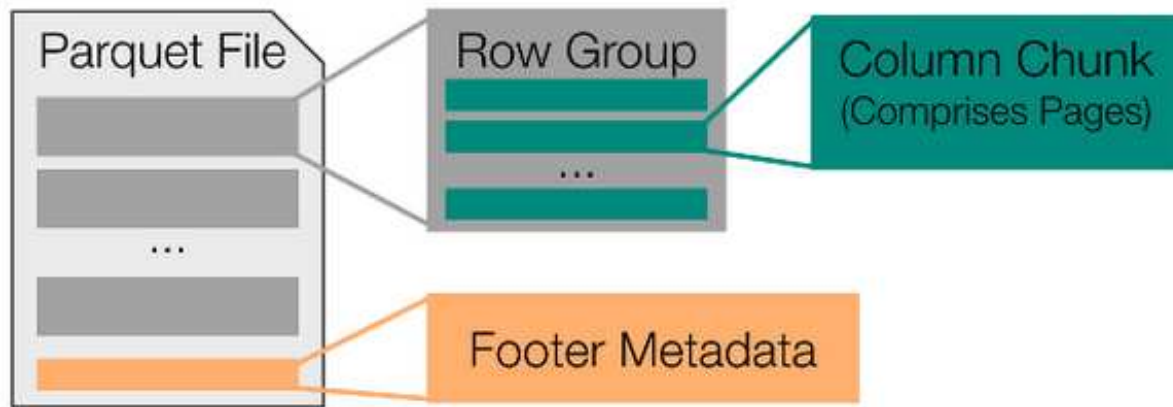


Figure 1. High-level visual representation of the Parquet file format

```
https = S3FileSystem(
    endpoint_url='https://s3.eu-central-1.wasabisys.com',
    anon=True
)
pq_dataset = pq.ParquetDataset(
    path_or_paths='gedi-ard/level2/l2v002.gedi_20190418_20230316_go_epsg.4326_v20240622',
    filesystem=https)

CPU times: user 634 ms, sys: 0 ns, total: 634 ms
Wall time: 1.35 s
```

Dataset

|--- Partition

|--- File

|--- Row group

|--- Column chunk

|--- Page

Source:

(1) [Optimizing Access to Parquet Data with fsspec](#)



<https://github.com/Open-Earth-Monitor/GlobalEarthPoint>



yu-feng-ho Update fuding statement

c40ba19 · 4 minutes ago

44 Commits



GED102

update to md format

5 days ago



ICESat-2_ATL08_v6

ICESat-2 table update

5 days ago



archive

update for Bids conference 2025

last week



convert-cloud-format

pmtile exploration

3 days ago



img

covert to jpg

7 hours ago



.gitignore

rearrange repo structure

last year



LICENSE

Initial commit

last year



Readme.html

update R function reference

last year



Readme.md

Update fuding statement

4 minutes ago



openEO_stream_GEDI_job.json

add openEO demo as json process graph

last week

Access to our
Repository for
more detail.

Thank you!

