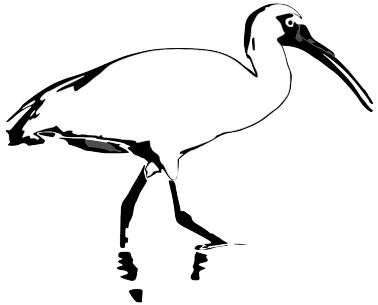
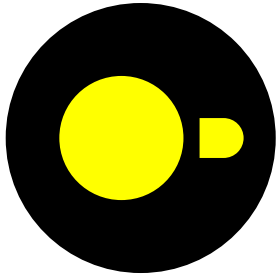
Ibis, DuckDB, and GeoParquet: Making Geospatial Analytics Fast, Simple, and Pythonic





Intro

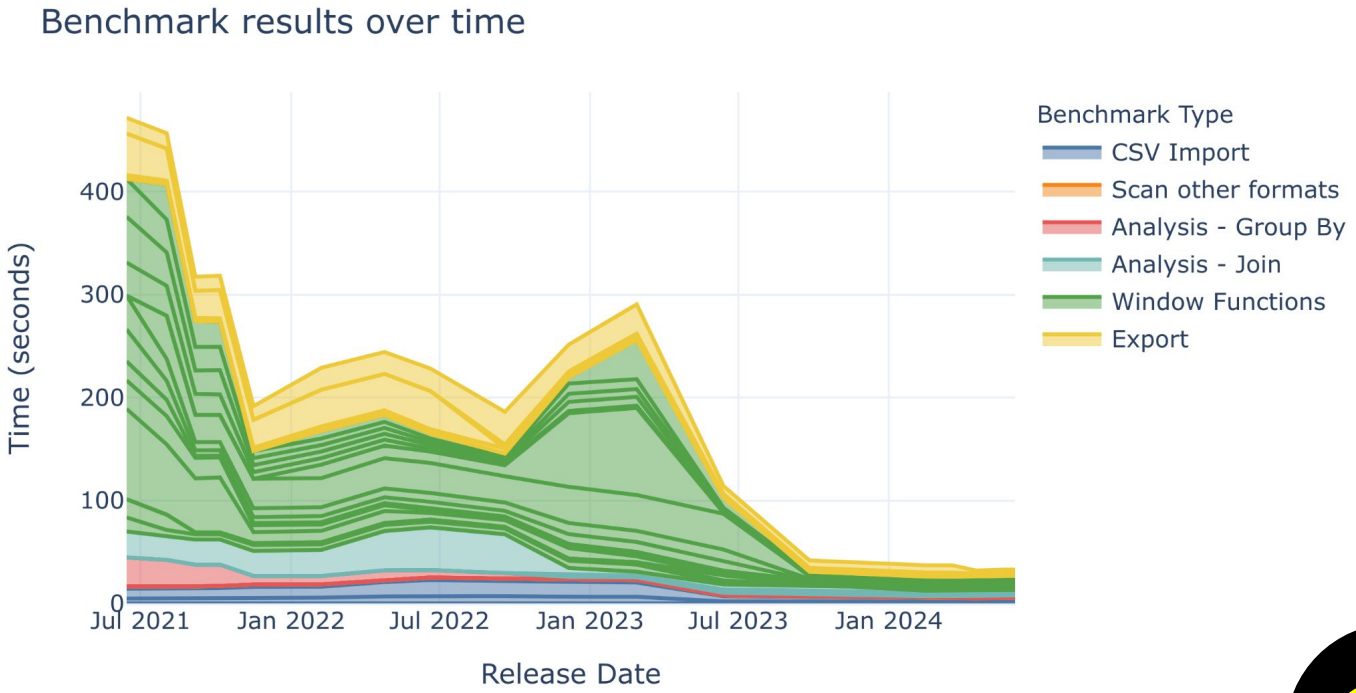
Naty Clementi, PhD
Sr. Software Engineer

ncclementi

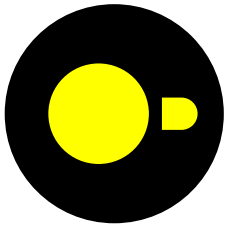
https://ibis-project.org/

Why ...

- DuckDB
- DuckDB for Geospatial
- GeoParquet
- Ibis



TL;DR: In the last 3 years, DuckDB has become **3-25x faster** and can analyze **~10x larger datasets** all on the same hardware.



https://duckdb.org/2024/06/26/benchmarks-over-time.html

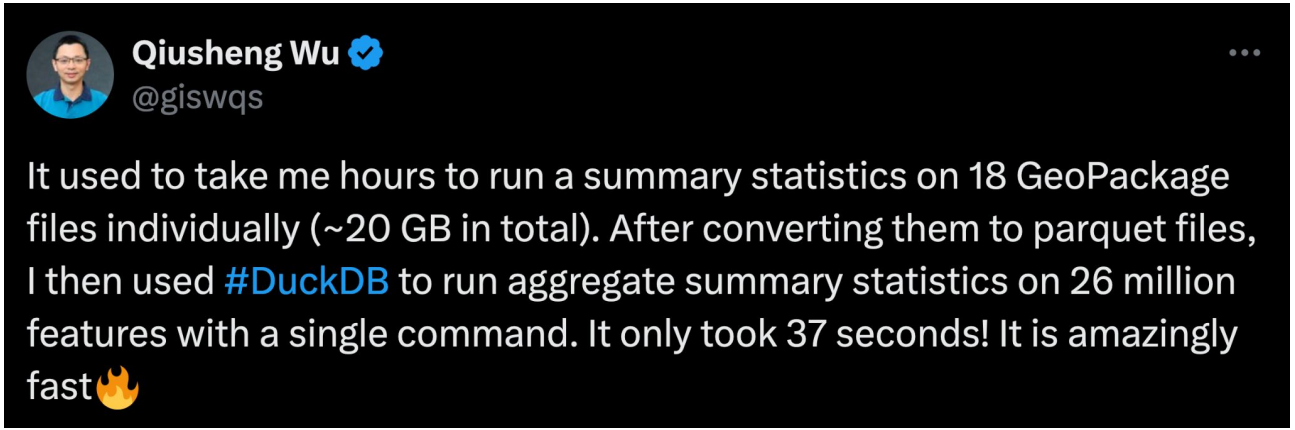
Why ...

- DuckDB ✔
- DuckDB for Geospatial
- GeoParquet
- Ibis

Why DuckDB Geospatial?

- Functions are based on GEOS, GDAL and PROJ.
- Leverage GDAL: support 50+ geospatial data formats.
- Mix and match your geospatial data with other data.
- GeoParquet support*.
- “R-Tree” spatial indexes.

```
CREATE TABLE cleaned_rides AS SELECT
    ST_Point(pickup_latitude, pickup_longitude) AS pickup_point,
    ST_Point(dropoff_latitude, dropoff_longitude) AS dropoff_point,
    dropoff_datetime::TIMESTAMP - pickup_datetime::TIMESTAMP AS time,
    trip_distance,
    ST_Distance(
        ST_Transform(pickup_point, 'EPSG:4326', 'ESRI:102718'),
        ST_Transform(dropoff_point, 'EPSG:4326', 'ESRI:102718')) / 5280
    AS aerial_distance,
    trip_distance - aerial_distance AS diff
FROM rides
WHERE diff > 0
ORDER BY diff DESC;
```



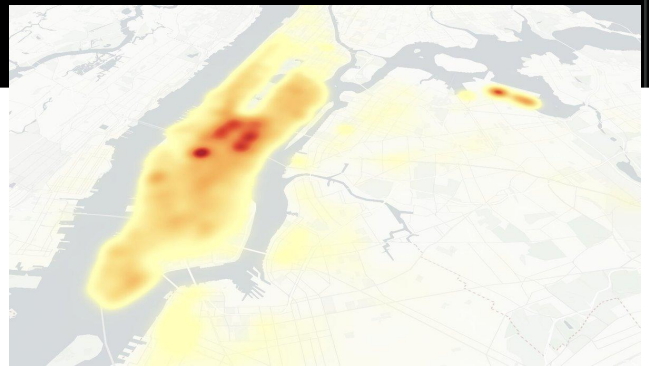
https://twitter.com/giswqs/status/1716548270191784259

https://twitter.com/giswqs/status/1729711511747096824

Kyle Barron @kylebarron@mapstodon.space
@kylebarron2

Lonboard 0.9 brings direct integration with @duckdb Spatial's Python API!

In one line of code, visualize millions of geometries from a SQL query in @ProjectJupyter. Powered by @ApacheArrow and deck.gl under the hood.



https://twitter.com/kylebarron2/status/1787535648737329248

Why ...

- DuckDB ✔
- DuckDB for Geospatial ✔
- GeoParquet
- Ibis

Why Geoparquet

- Standard Geospatial Data in Parquet / Columnar Data for Geo
- Cloud Data Warehouse Interoperability
- Benefits of compression (e.g. save money in storage)
- Integration with GeoArrow (way of representing geospatial vector data in memory)

Interview with Kyle Barron on GeoArrow and GeoParquet, and the Future of Geospatial Data Analysis

https://cloudnativegeo.org/blog/2024/12/interview-with-kyle-barron-on-geoarrow-and-geoparquet-and-the-future-of-geospatial-data-analysis/

Why ...

- DuckDB ✔
- DuckDB for Geospatial ✔
- GeoParquet ✔
- Ibis

Ibis!

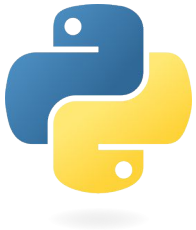
An **open source** portable Python library for data wrangling.

- A dataframe API for Python
- Interfaces to 15+ query engines
- Deferred execution model

Ibis: because SQL is everywhere and so is Python
SciPy 2024 by Gil Forsyth

Why Ibis as an interface for DuckDB geospatial?

```
SELECT * FROM tbl
```



Demo Time!

https://github.com/ncclementi/ibis-duckcon-2025

Why ...

- DuckDB ✔
- DuckDB for Geospatial ✔
- GeoParquet ✔
- Ibis ✔

```
pip install 'ibis-framework[duckdb,geospatial]'
```

ncclementi

https://ibis-project.org/

ibis-project/ibis

ibis-project.zulipchat.com/

@ibisData

Phillip in the Cloud
cpcloud

linkedin.com/company/ibis-project/

Questions

Slides & Demo