Watershed

Get to zero carbon, fast.

InsightGlobal Walmart DOORDASH BlackRock. Figma BAILLIE GIFFORD

WISE airbnb Match Group Grover Sonder BLOCK

Misfits Market athenahealth ZAXBY'S monzo CSWR Central States Water Resources BEAUTYCOUNTER

EVERLANE BOMBAS TaskUs™ BOOM Klarna. shopify

data iku BOB'S DISCOUNT FURNITURE WARBY PARKER Spotify GOJO Purell GOJO, MAKERS OF PURELL™ okta

sweetgreen Oakley Capital FORTRESS General Mills stripe servicenow

sunrun THOMABRAVO SKIMS kainos BainCapital Wolt

InsightGlobal Walmart DOORDASH BlackRock Figma BAILLIE GIFFORD

WISE airbnb Grover Sonder BLOCK

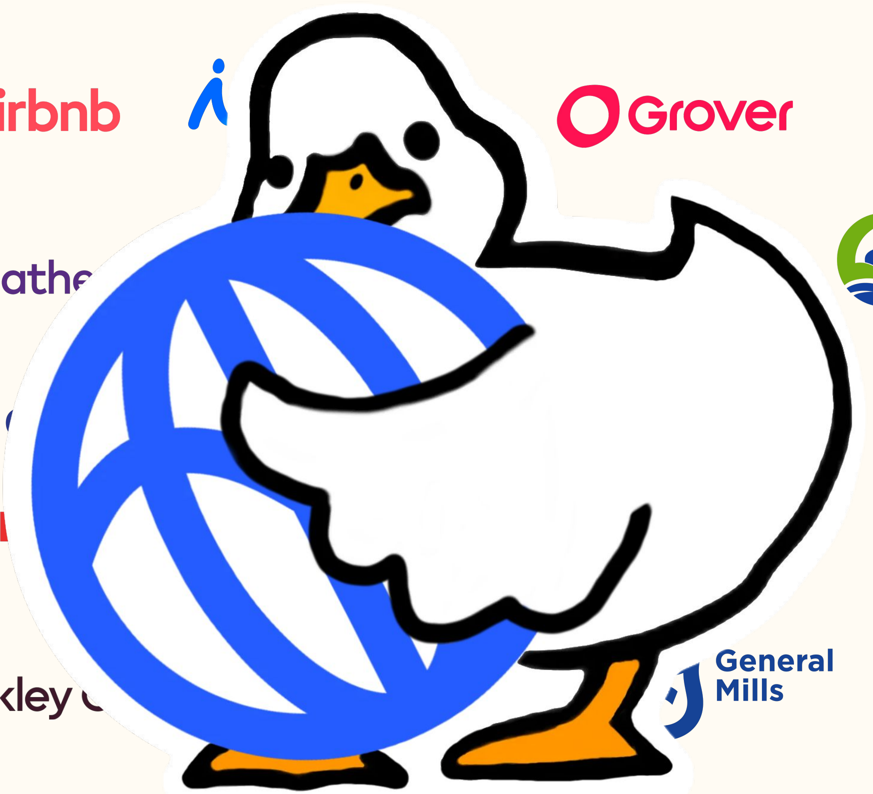
Misfits Market athe CSWR Central States Water Resources BEAUTYCOUNTER

EVERLANE B Klarna. shopify

data iku BO GOJO Purell GOJO, MAKERS OF PURELL™ okta

sweetgreen Oakley General Mills stripe servicenow

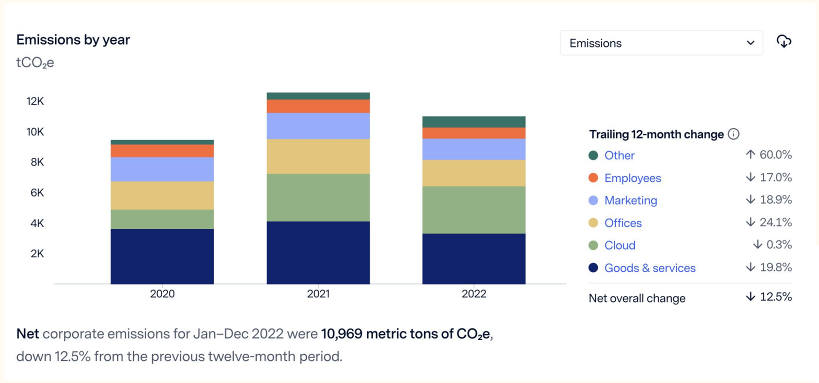
sunrun THOMABRAVO SKIMS kainos BainCapital Wolt



Watershed

Carbon footprint data

User needs: fast aggregates



Emissions by year

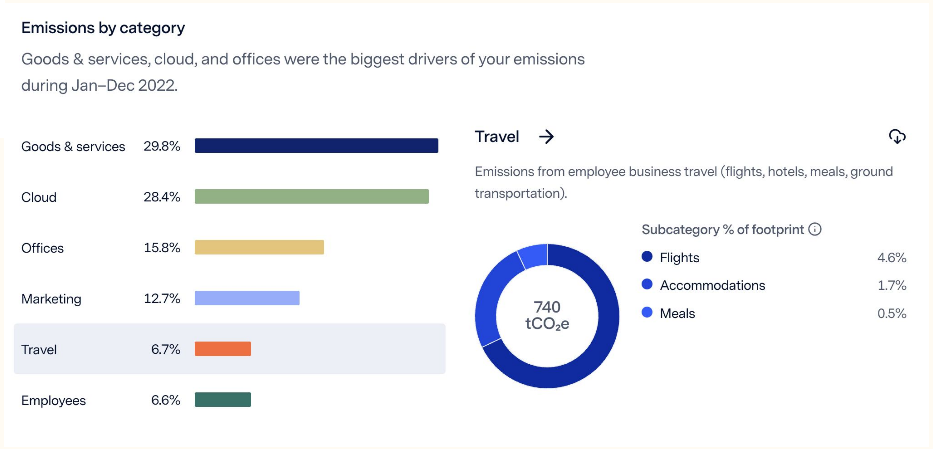
tCO_2e

Emissions

Trailing 12-month change

● Other	↑ 60.0%
● Employees	↓ 17.0%
● Marketing	↓ 18.9%
● Offices	↓ 24.1%
● Cloud	↓ 0.3%
● Goods & services	↓ 19.8%
Net overall change	↓ 12.5%

Net corporate emissions for Jan–Dec 2022 were **10,969 metric tons of CO_2e**, down 12.5% from the previous twelve-month period.



Emissions by category

Goods & services, cloud, and offices were the biggest drivers of your emissions during Jan–Dec 2022.

Goods & services	29.8%
Cloud	28.4%
Offices	15.8%
Marketing	12.7%
Travel	6.7%
Employees	6.6%

Travel →

Emissions from employee business travel (flights, hotels, meals, ground transportation).

740 tCO_2e

Subcategory % of footprint

● Flights	4.6%
● Accommodations	1.7%
● Meals	0.5%

Medium-sized data

- 12% of customers have footprints with › 1m rows
- Largest customer has ›15m rows → ~750mb parquet

Medium-sized data

- 12% of customers have footprints with › 1m rows
- Largest customer has ›15m rows → ~750mb parquet

Medium-sized data

- 12% of customers have footprints with › 1m rows
- Largest customer has ›15m rows → ~750mb parquet
- Fits on one machine, but non-trivial to query performantly

Existing solution: Postgres

Existing solution: Postgres

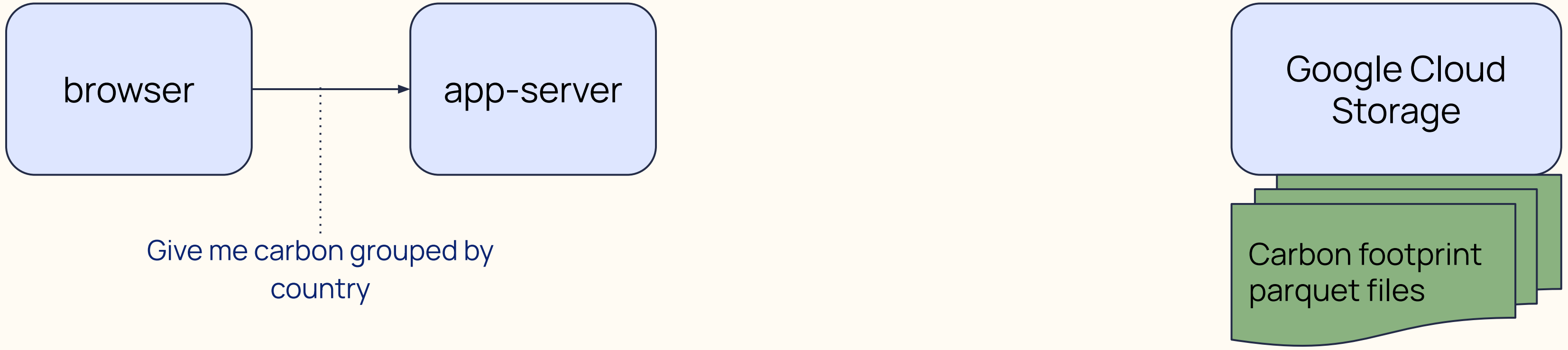
- 162GiB table / 252GiB of Postgres database size

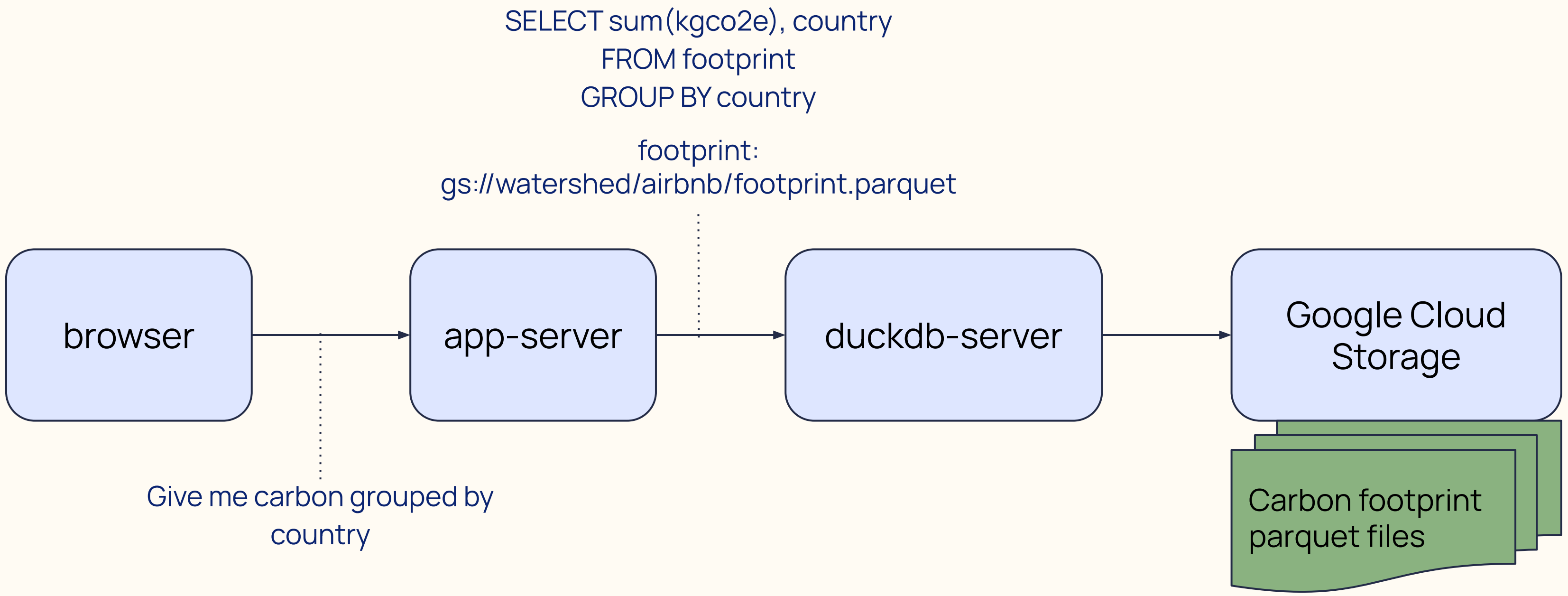
Existing solution: Postgres

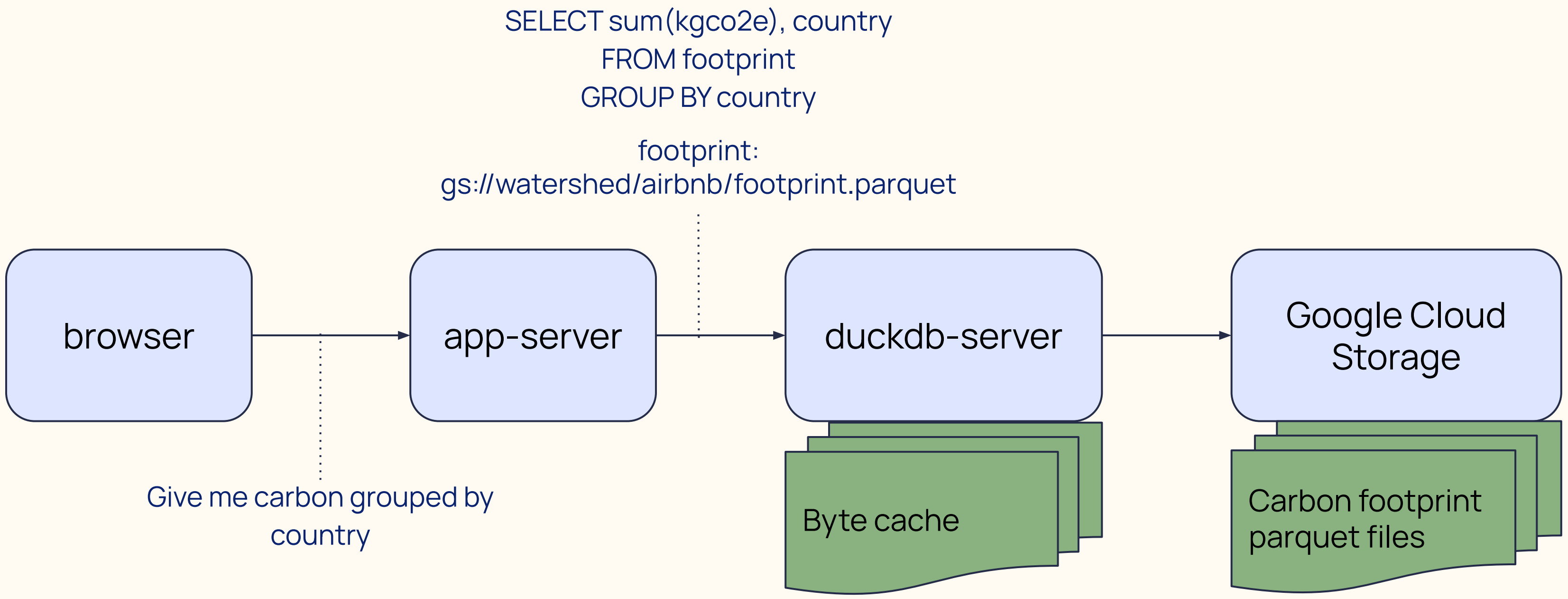
- 162GiB table / 252GiB of Postgres database size
- Painful migrations
- Arbitrary analytic queries don’t scale

Architecture

browser
Google Cloud Storage
Carbon footprint parquet files


browser
app-server
Give me carbon grouped by country
Google Cloud Storage
Carbon footprint parquet files


SELECT sum(kgco2e), country
FROM footprint
GROUP BY country
footprint:
gs://watershed/airbnb/footprint.parquet
browser
app-server
duckdb-server
Google Cloud Storage
Give me carbon grouped by country
Carbon footprint parquet files


SELECT sum(kgco2e), country
FROM footprint
GROUP BY country
footprint:
gs://watershed/airbnb/footprint.parquet
browser
app-server
duckdb-server
Google Cloud Storage
Byte cache
Carbon footprint parquet files
Give me carbon grouped by country

Why we love this

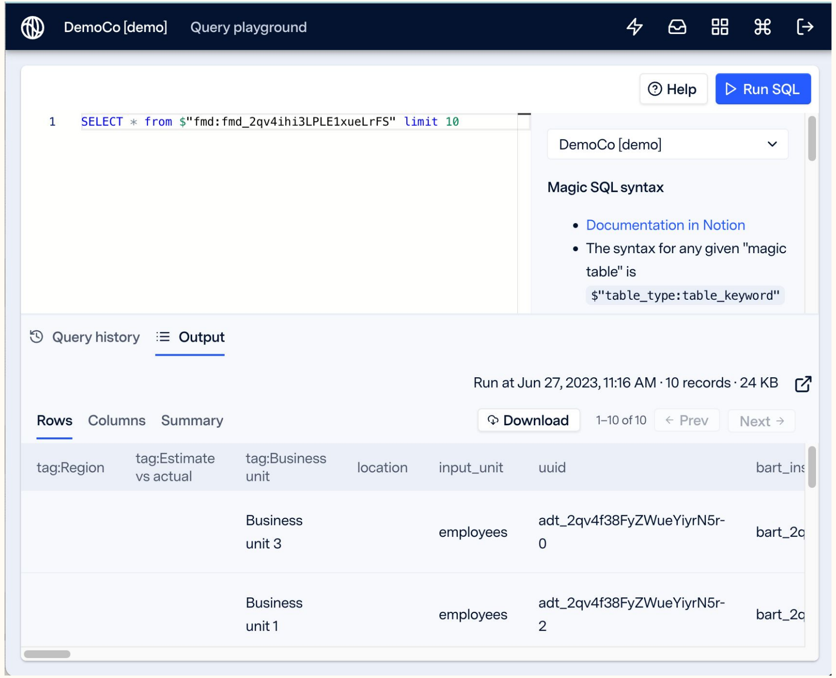
- It’s fast!
  - 10x faster than Postgres with lots of indexes
  - 100s of ms for our P99 data size
- No more giant table!
- Less adhoc caching
- Parquet 💜

Other uses

- Data pipeline: activity data → carbon footprint data
- Query any parquet file
- ~75k duckdb-server queries per day

Other uses

- Data pipeline: activity data → carbon footprint data
- Query any parquet file
- ~75k duckdb-server queries per day



```
SELECT * from $"fmd:fmd_2qv4ihi3LPLE1xueLrFS" limit 10
```

Magic SQL syntax

- Documentation in Notion
- The syntax for any given "magic table" is `$"table_type:table_keyword"`

Run at Jun 27, 2023, 11:16 AM · 10 records · 24 KB

tag:Region	tag:Estimate vs actual	tag:Business unit	location	input_unit	uuid	bart_ins
		Business unit 3		employees	adt_2qv4f38FyZWueYiyrN5r-0	bart_2q
		Business unit 1		employees	adt_2qv4f38FyZWueYiyrN5r-2	bart_2q

Thank you!

jessica@watershed.com

Watershed