Modernizing the geo-data stack

CARTO

Matthew Forrest

Field CTO

CARTO

Overview

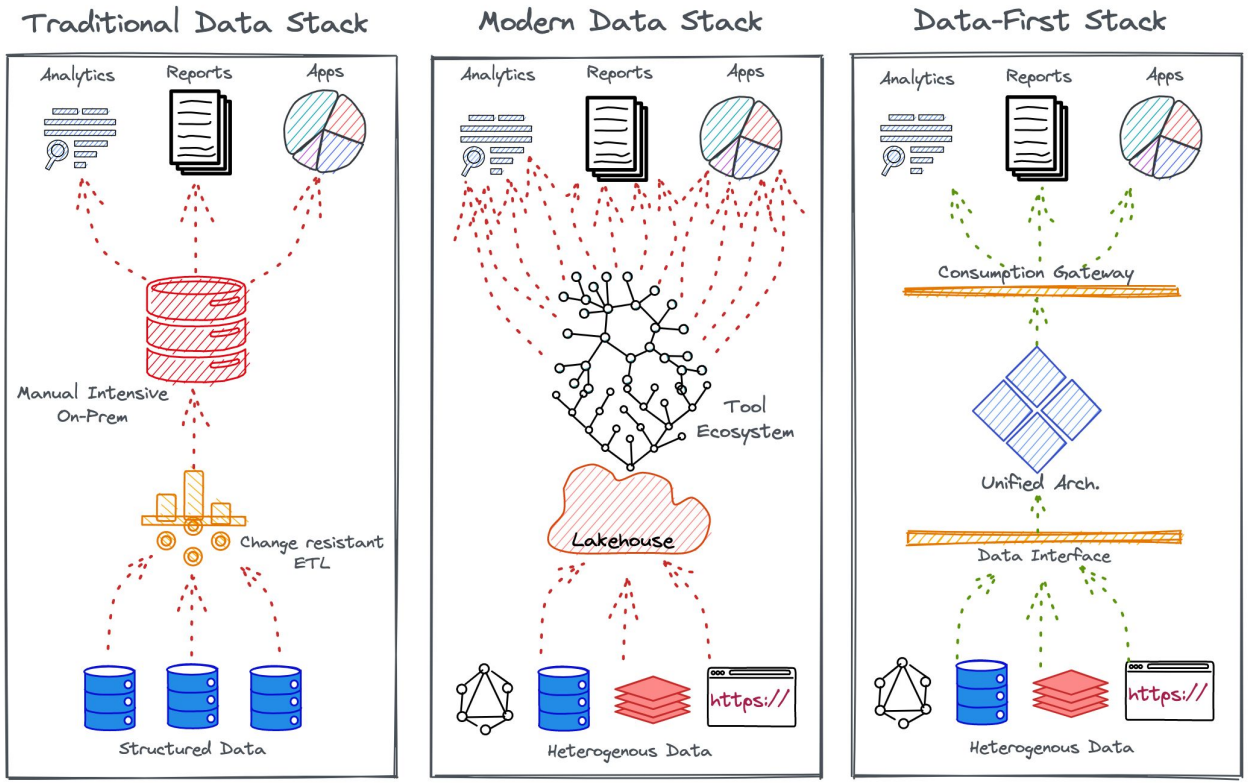
Our agenda

The modern data stack	What the modern data stack means in the wider world of data
Modernizing two ways	How to modernize locally and within the cloud (and when to move to the cloud)
The modern geo-data stack	What does the modern data stack look like for geospatial?
In the cloud	When and why to use the cloud
Where we see the future	Why we believe SQL, data warehouses, and cloud can be a part of the future

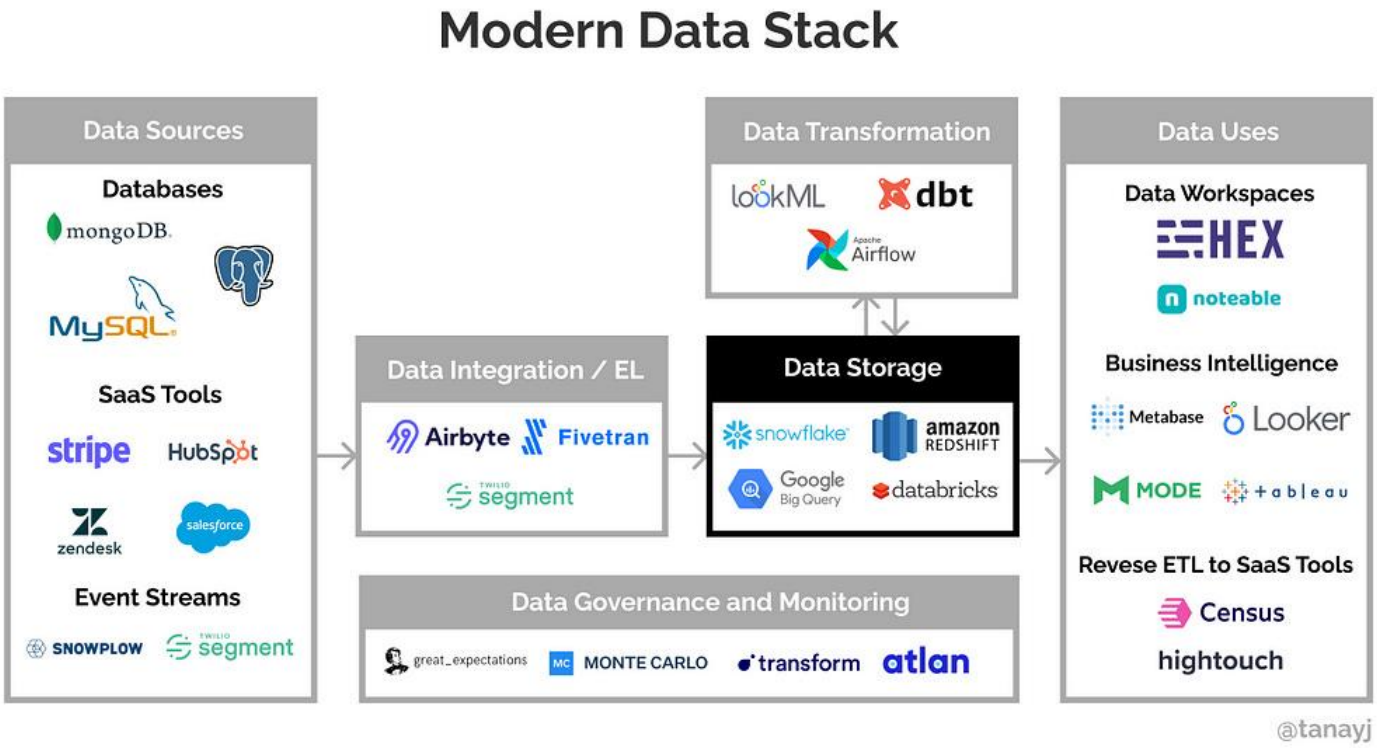
Understanding the modern data stack

Why?

- Growth in the OLAP database
- Distributed computing and regular releases
- SQL as the *lingua franca*
- Ecosystem tools to support



https://moderndata101.substack.com/p/evolution-of-the-data-stack-the-story



https://tanay.substack.com/p/understanding-the-modern-data-stack

The modern geo-data stack

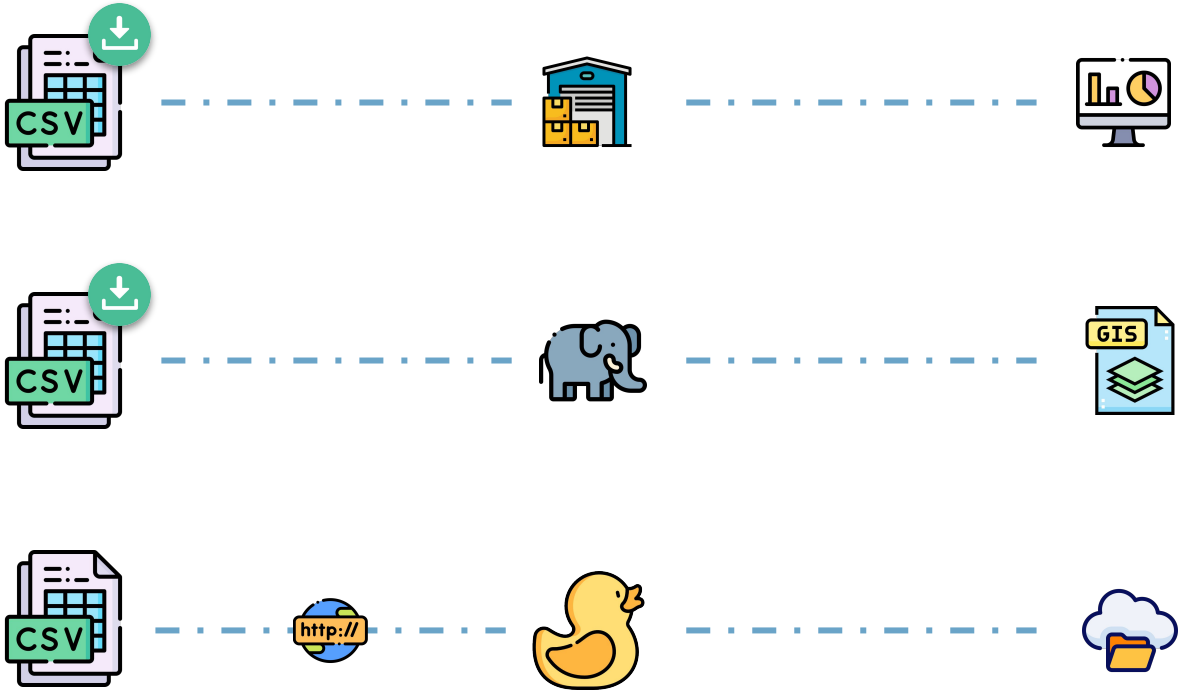
Modernizing in many ways

Processing storms data from NOAA

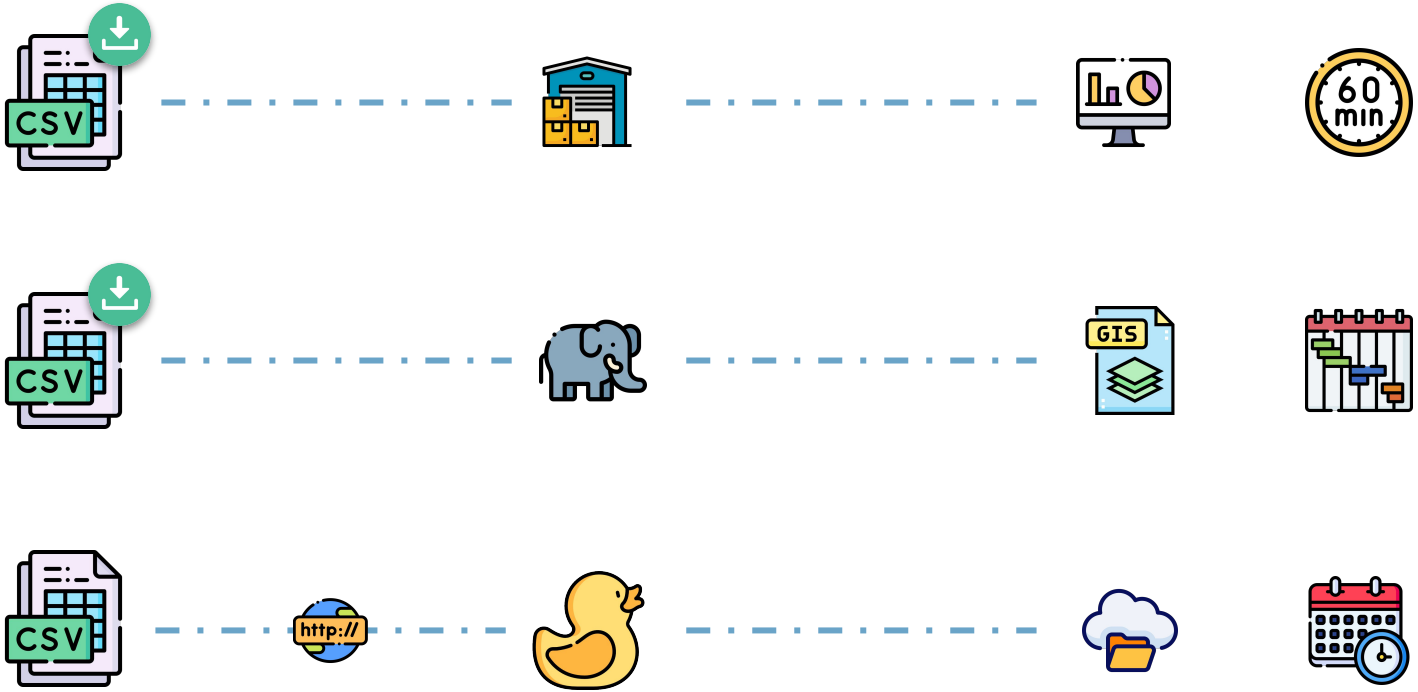
Index of /pub/data/swdi/stormevents/csvfiles

Name	Last modified	Size	Description
Parent Directory		-	
Storm-Data-Bulk-csv-Format.pdf	2020-07-17 13:10	161K	
Storm-Data-Export-Format.pdf	2020-07-17 09:17	163K	
StormEvents_details-ftp_v1.0_d1950_c20210803.csv.gz	2021-08-05 09:53	10K	
StormEvents_details-ftp_v1.0_d1951_c20210803.csv.gz	2021-08-05 09:56	12K	
StormEvents_details-ftp_v1.0_d1952_c20210803.csv.gz	2021-08-05 09:56	12K	
StormEvents_details-ftp_v1.0_d1953_c20210803.csv.gz	2021-08-05 09:56	21K	
StormEvents_details-ftp_v1.0_d1954_c20210803.csv.gz	2021-08-05 09:56	26K	
StormEvents_details-ftp_v1.0_d1955_c20210803.csv.gz	2021-08-05 09:56	52K	
StormEvents_details-ftp_v1.0_d1956_c20210803.csv.gz	2021-08-05 09:56	62K	
StormEvents_details-ftp_v1.0_d1957_c20210803.csv.gz	2021-08-05 09:56	80K	
StormEvents_details-ftp_v1.0_d1958_c20210803.csv.gz	2021-08-05 09:56	69K	
StormEvents_details-ftp_v1.0_d1959_c20210803.csv.gz	2021-08-05 09:56	66K	
StormEvents_details-ftp_v1.0_d1960_c20210803.csv.gz	2021-08-05 09:56	70K	
StormEvents_details-ftp_v1.0_d1961_c20210803.csv.gz	2021-08-05 09:56	81K	
StormEvents_details-ftp_v1.0_d1962_c20210803.csv.gz	2021-08-05 09:56	83K	
StormEvents_details-ftp_v1.0_d1963_c20210803.csv.gz	2021-08-05 09:56	70K	
StormEvents_details-ftp_v1.0_d1964_c20210803.csv.gz	2021-08-05 09:56	84K	
StormEvents_details-ftp_v1.0_d1965_c20210803.csv.gz	2021-08-05 09:56	102K	
StormEvents_details-ftp_v1.0_d1966_c20210803.csv.gz	2021-08-05 09:56	81K	
StormEvents_details-ftp_v1.0_d1967_c20210803.csv.gz	2021-08-05 09:56	95K	
StormEvents_details-ftp_v1.0_d1968_c20210803.csv.gz	2021-08-05 09:56	112K	

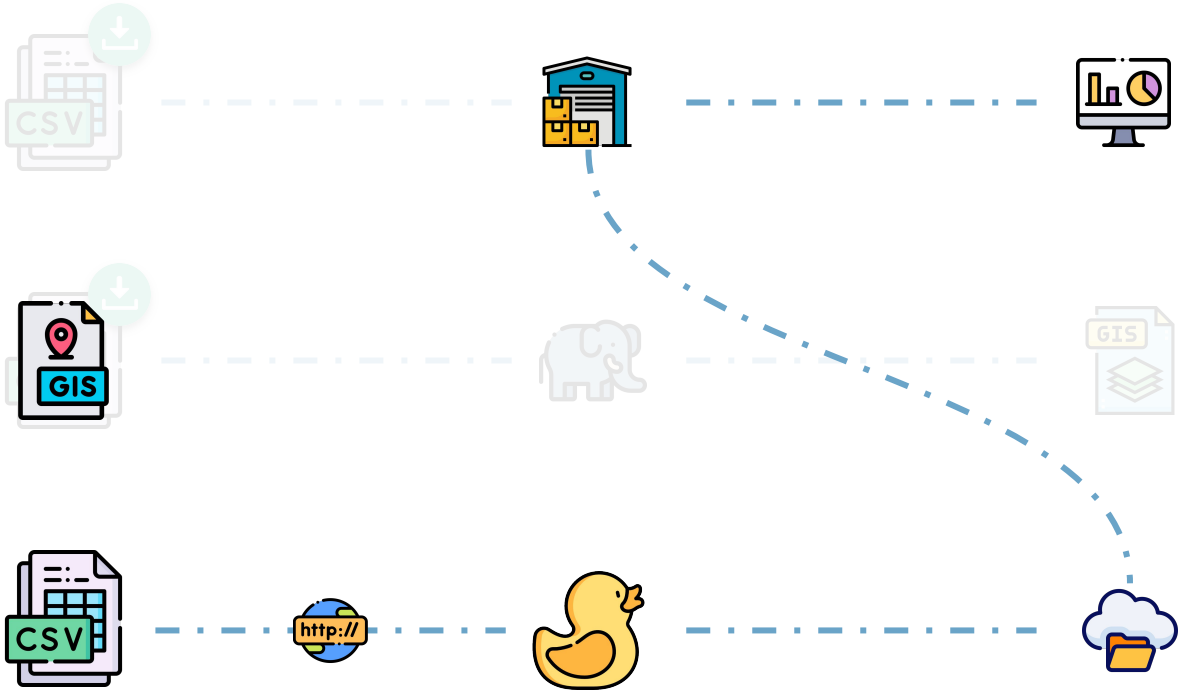
Three different paths



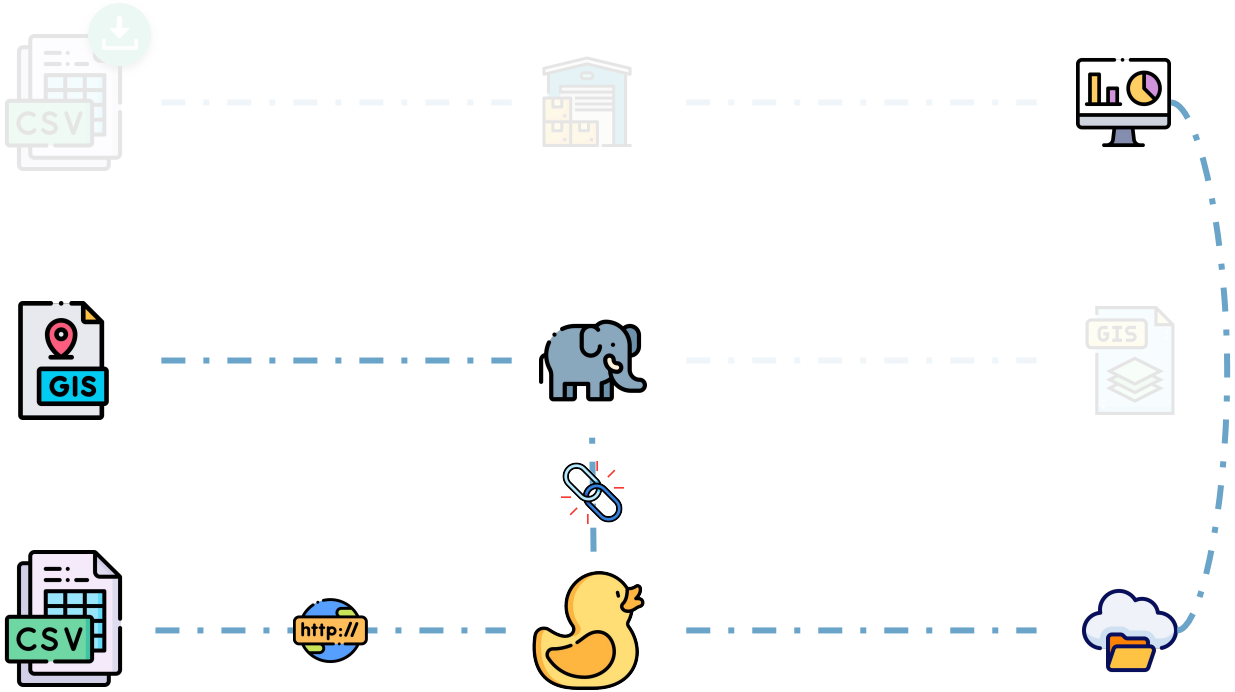
Three different paths



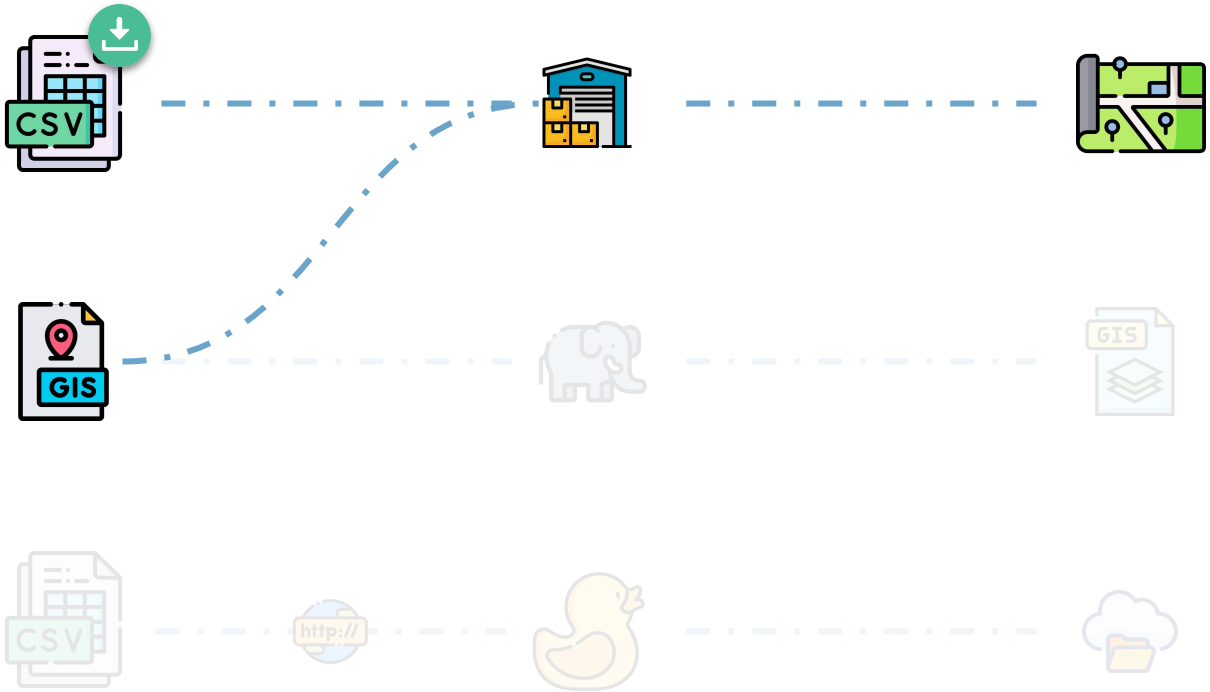
Three different paths



Three different paths



Three different paths



Looks the same, but there are many ways to use the modern geo-data stack

The modern geo-data stack

The modern geo-data stack

Data Sources

Ingestion

Storage

Analytics

Data Science

Reverse ETL

Transformation

Applications

When and why to use the cloud

The cloud does a lot more than just getting your work online

Three use cases for the cloud

Your analysis needs to be repeated or available regularly

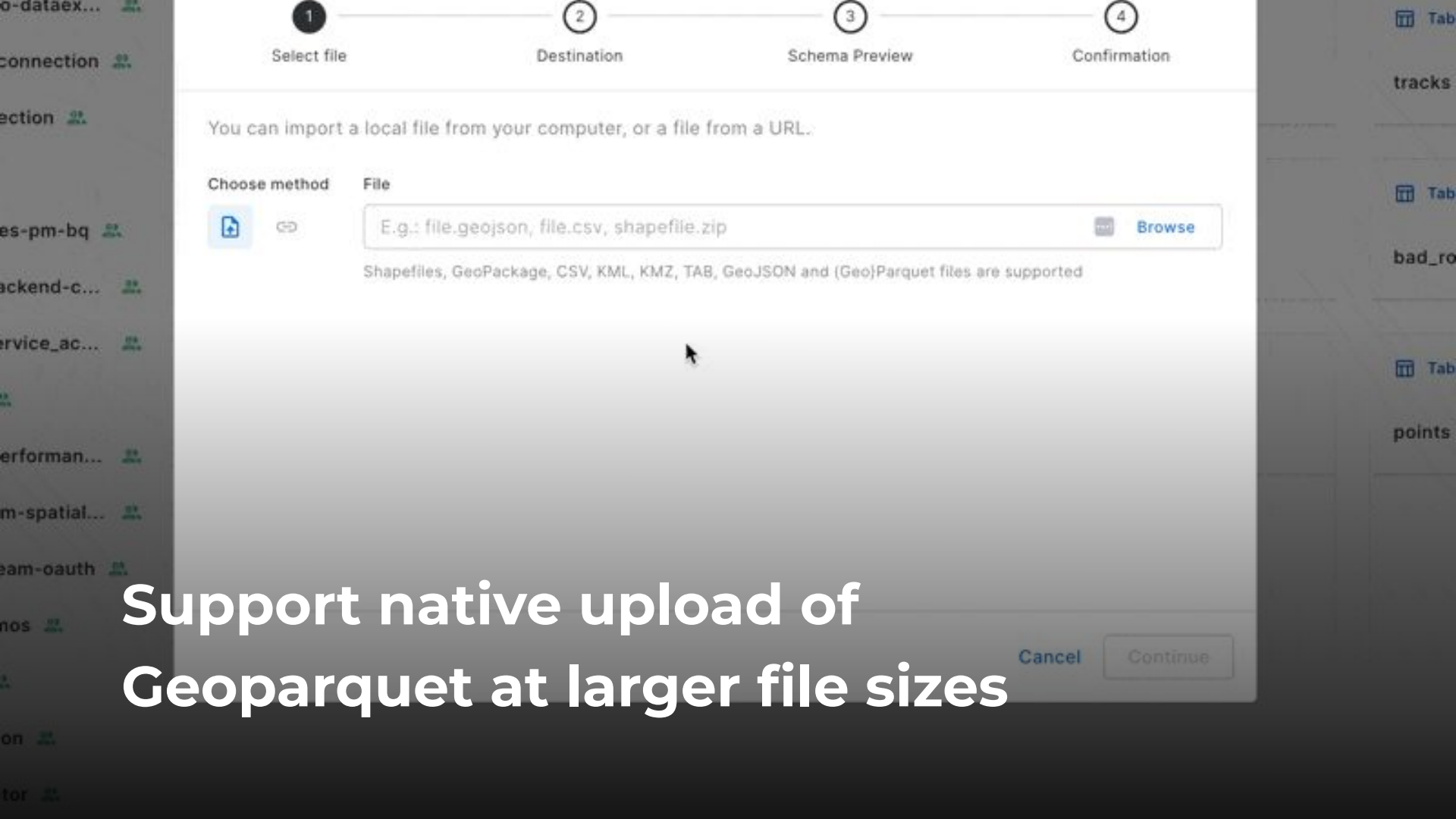
You need to run your analysis with large scale data

The geospatial need isn’t just a one off

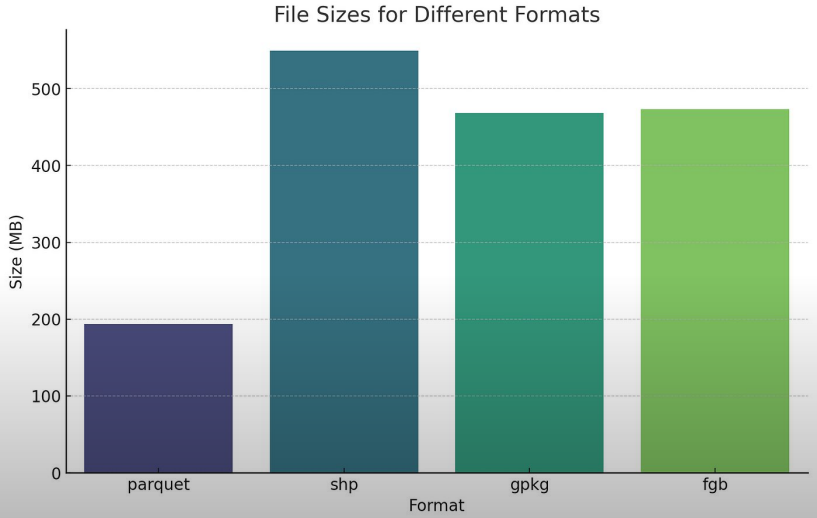
FUTURE RADAR
SAT 7 AM
PENSACOLA
TALLAHASSEE
JACKSONVILLE
PANAMA CITY
GAINESVILLE
OCALA
DAYTONA
ORLANDO
MELBOURNE
TAMPA
FORT PIERCE
SARASOTA
FORT MYERS
WEST PALM BEACH
NAPLES
MIAMI
KEY WEST
Understanding exposure at any point in time

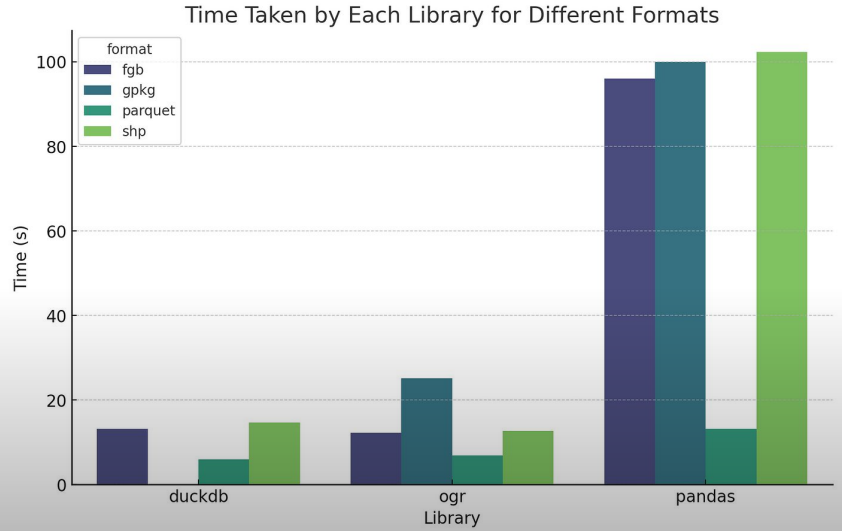
Where CARTO is focusing

Geoparquet & Raster	Improving data at the source, making raster ingestion easy
Workflows	Orchestrating ingestion, transformations, and analytics
Storage	Improving connections to the cloud data warehouses
Analytics	Builder makes analytical reporting dead simple
Data Science	Embedding pure spatial data science into the Analytics Toolbox
Applications	Integrating with Deck GL, tiling at scale like never before



Support native upload of Geoparquet at larger file sizes

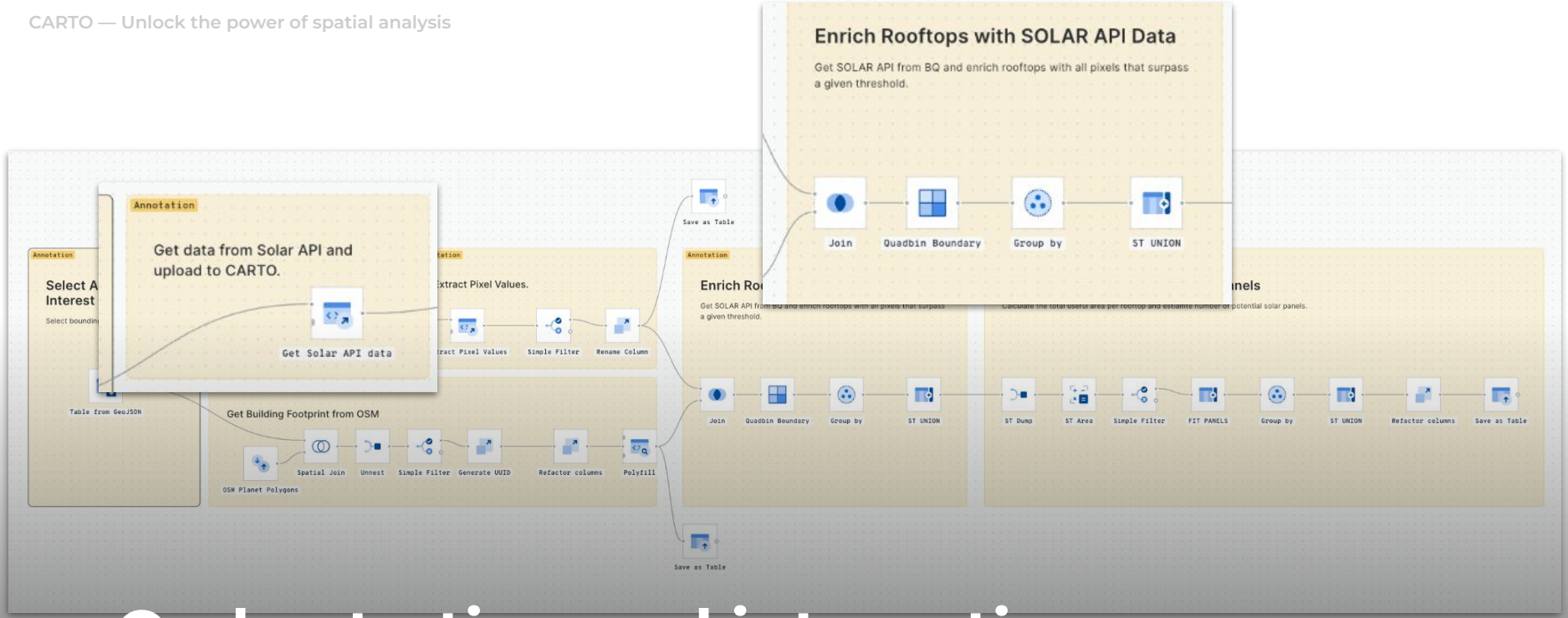




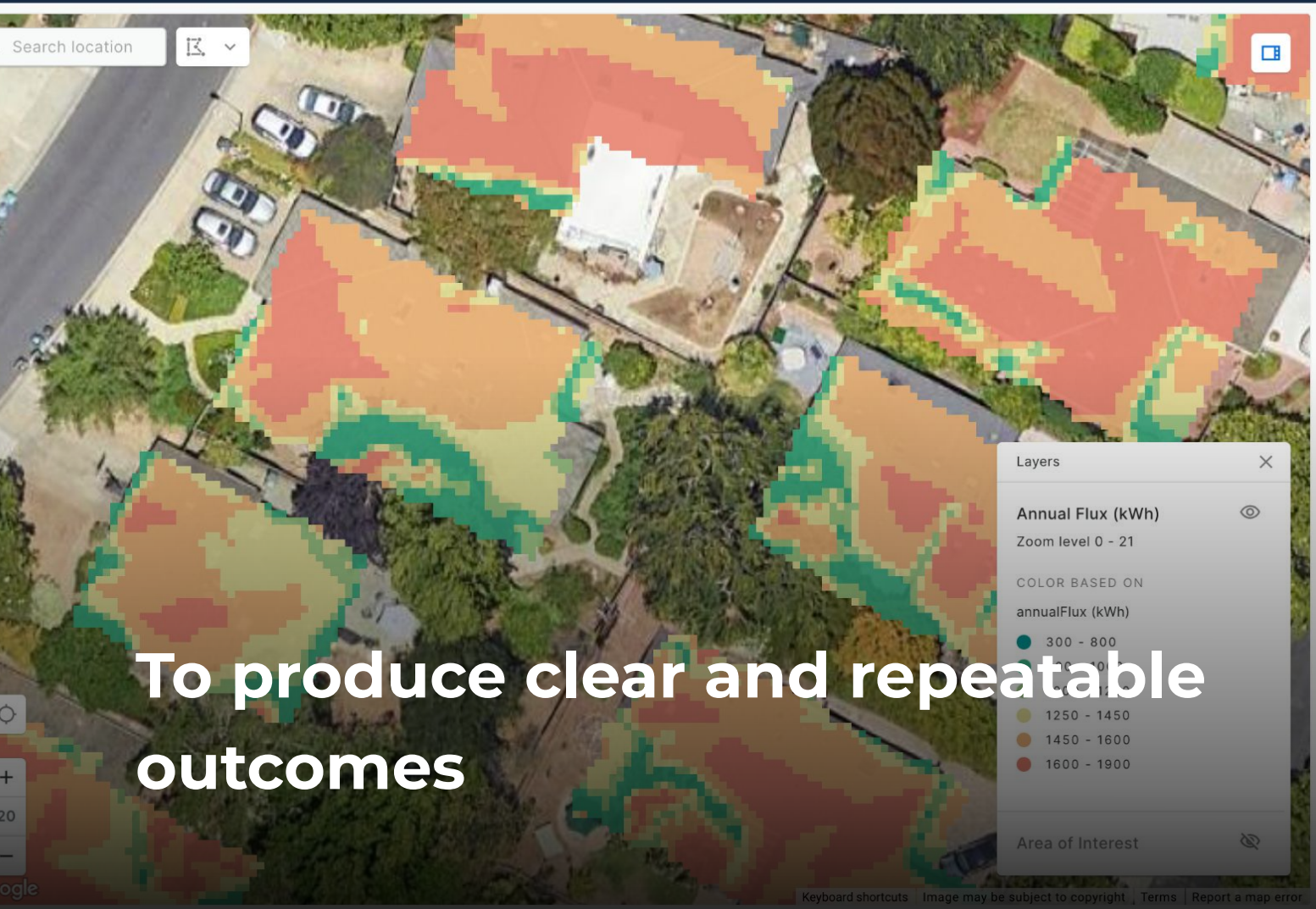
Improved size and performance across the board

```
CALL
  `carto-un.carto.RASTER_ST_GETVALUE`(
      "my-bigquery-project.my-bigquery-dataset.UKv1p1",
      ST_GEOGFROMTEXT('POLYGON ((-3.100891 51.436889, -3.100891 51.538221, -3.308258 51.538221,
      NULL,
      "my-bigquery-project.my-bigquery-dataset.UKv1p1_cardiff"
  )
```

Raster analysis and ingest in the data warehouse



Orchestrating and integrating geospatial analysis tools



To produce clear and repeatable outcomes

Widgets

Annual Flux Histogram

All selected

1k

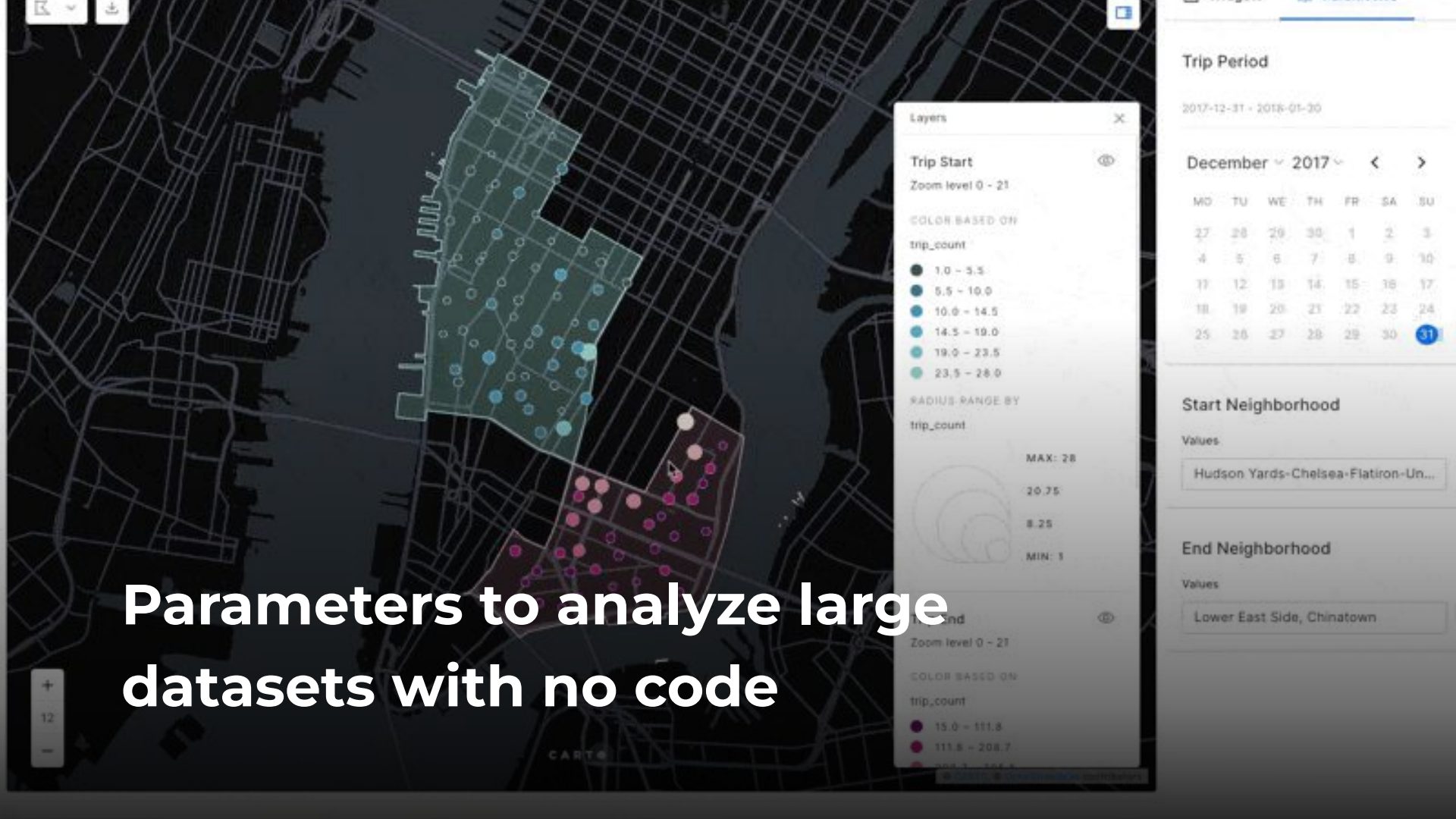
241 474 707 940 1.17k 1.41k 1.64k

Is rooftop

1 selected Lock

False

True



Parameters to analyze large datasets with no code

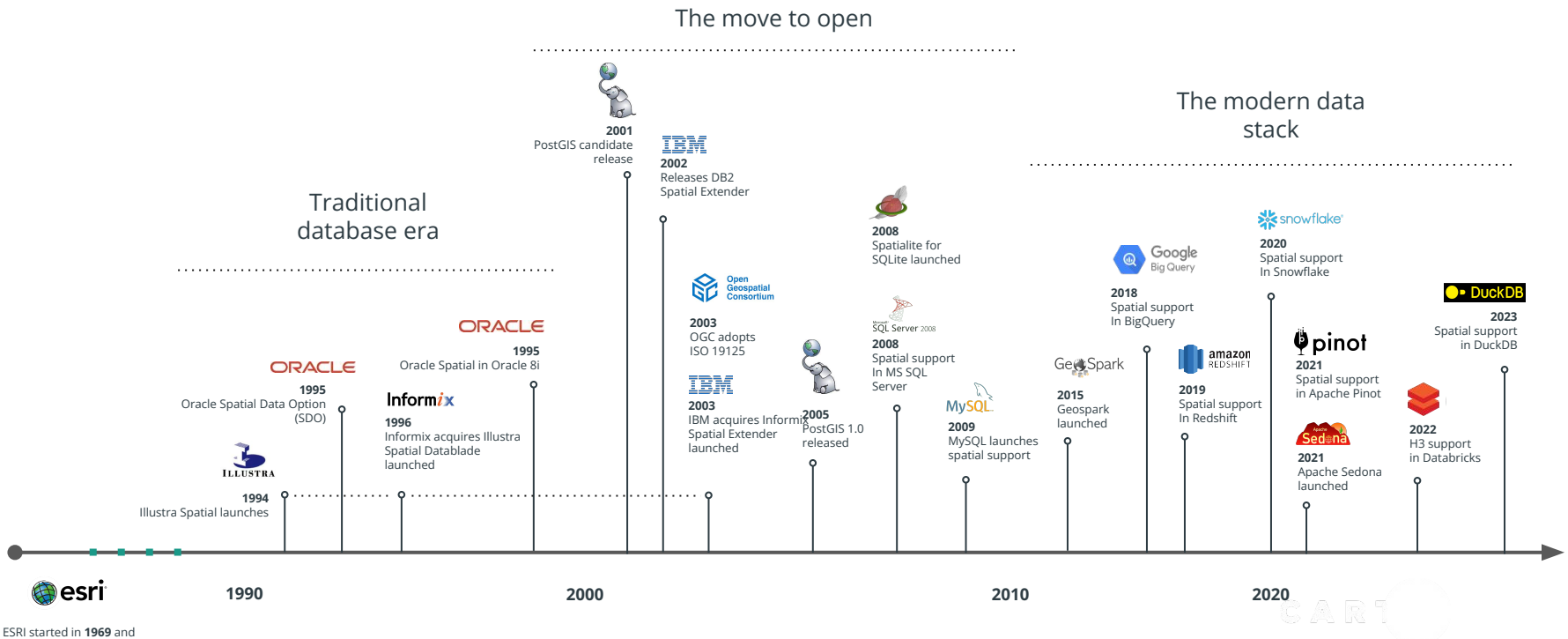
Spatial data science at scale in the data warehouse

Tiling millions and billions of features, without map tiles

Pregenerated tiles for even more scale

Why do we believe in data first approach?

Spatial SQL has always been at the center



ESRI started in **1969** and created commercial GIS