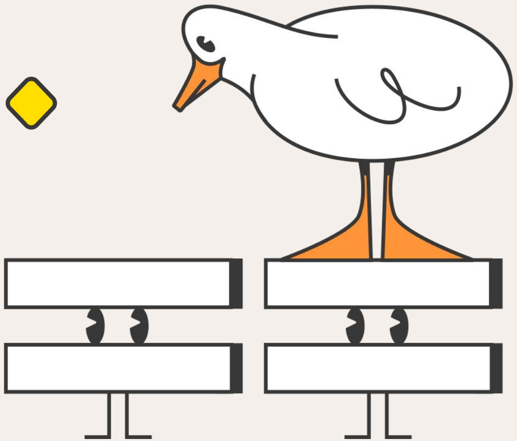
MotherDuck

Data Council

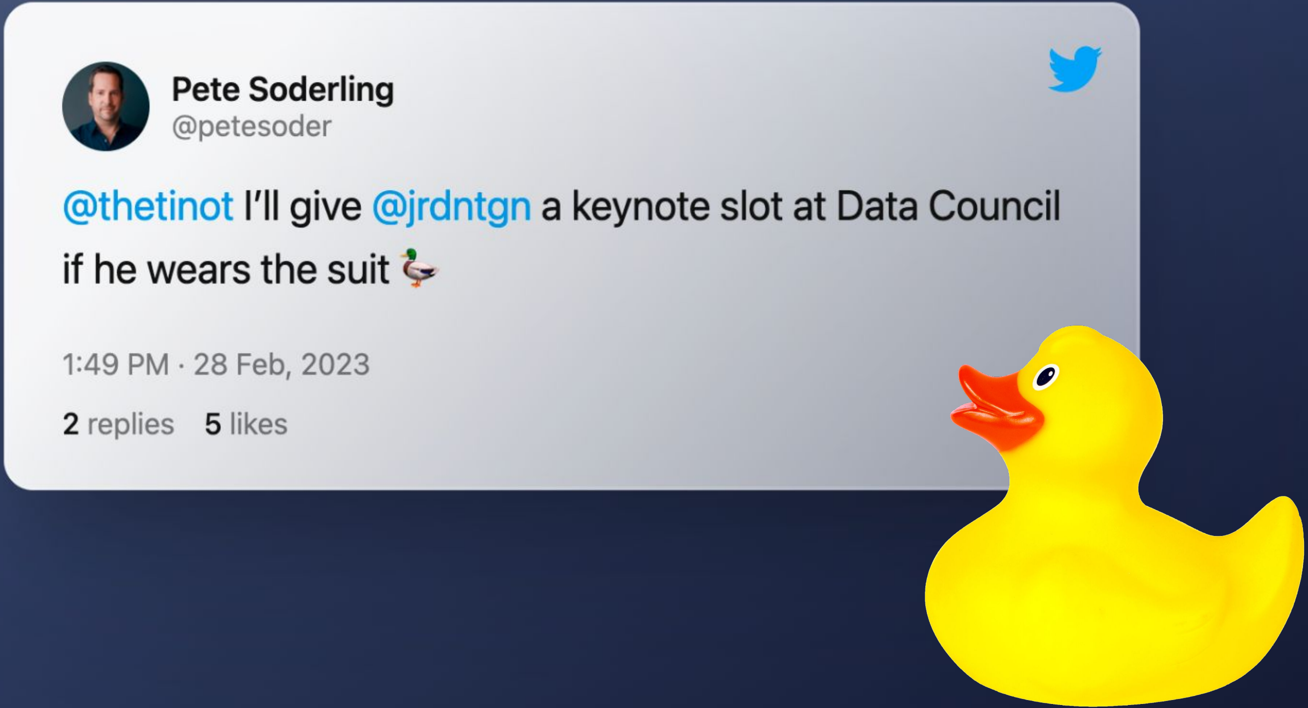
BIG DATA IS DEAD

AND WHAT THE DUCK YOU CAN DO ABOUT IT



JORDAN TIGANI

Chief Duck Herder @ MotherDuck



Pete Soderling
@petesoder

@thetinot I'll give @jrdntgn a keynote slot at Data Council if he wears the suit 🦆

1:49 PM · 28 Feb, 2023

2 replies **5** likes

WHO AM I?

- BigQuery Engineer
- BigQuery Eng Director
- BigQuery PM Director
- MemSQL/SingleStore CPO
- MotherDuck CEO

- Lots of talking about Big Data

Photo above: Me warning about the perils of big data in 2012.

MotherDuck

ABOUT THIS TALK

TELLING A STORY WITH GRAPHS

- Represent shapes rather than real data
- Based on experience with real data
- Only intended to be directionally accurate

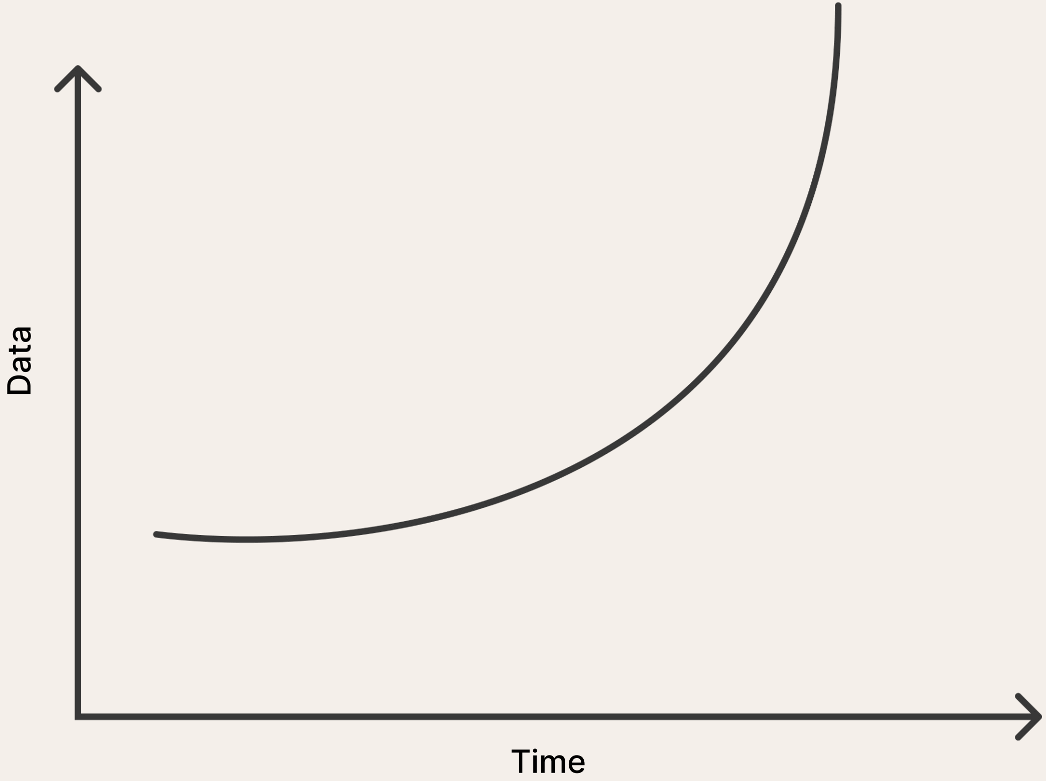
WHERE THIS DATA COMES FROM

- Query Logs
- Deal Post-Mortems
- Benchmark results
- Bugs/Support Tickets
- Customer Conversations
- Service Logs

BIG DATA IS COMING!

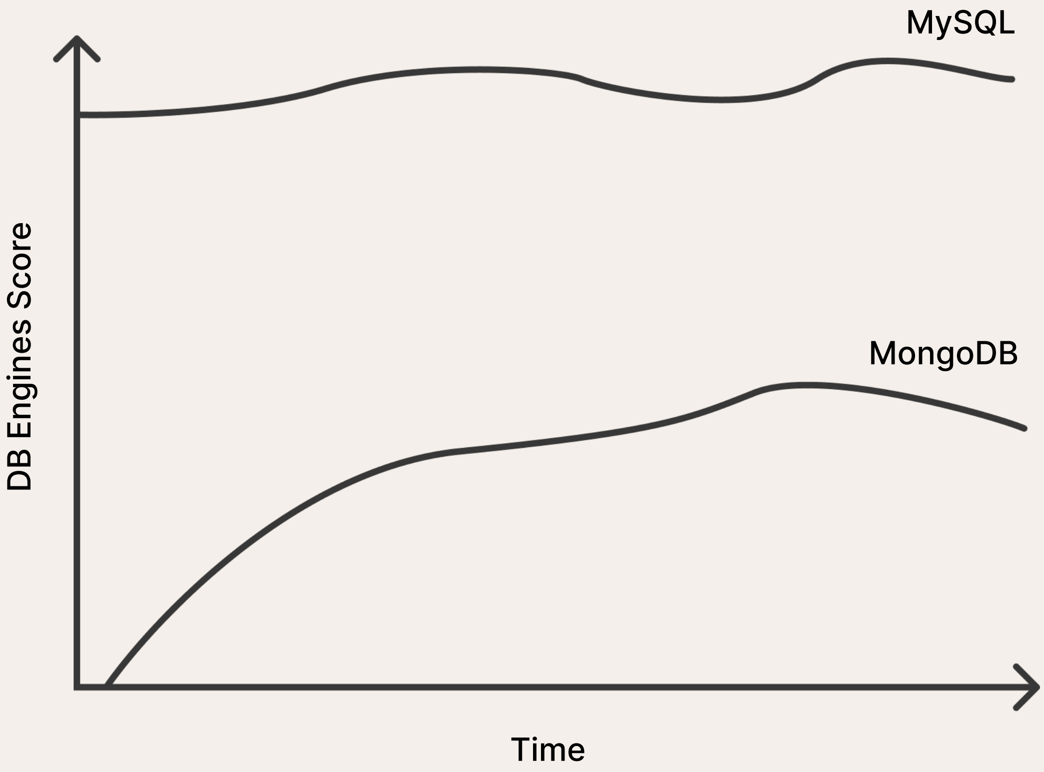
(So you’d better buy whatever I’m selling)

By year {CURRENT_YEAR() + ε} there will be {unfathomably large amount} of data generated



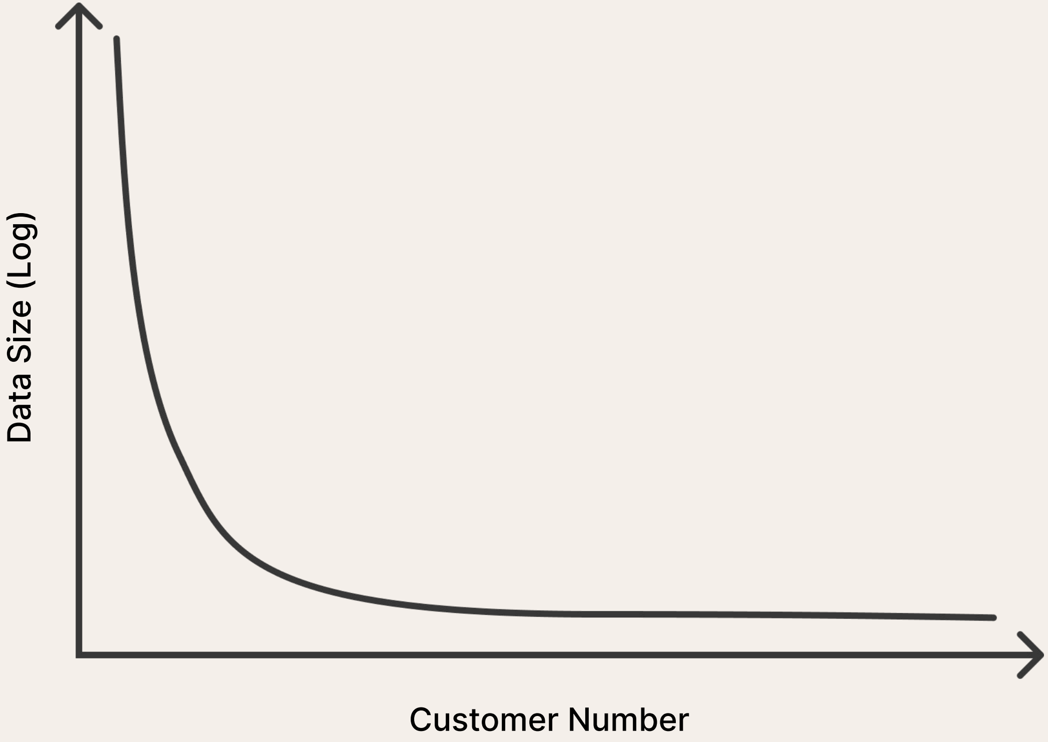
POPULARITY CONTEST:

BIG DATA IS NOT TAKING OVER

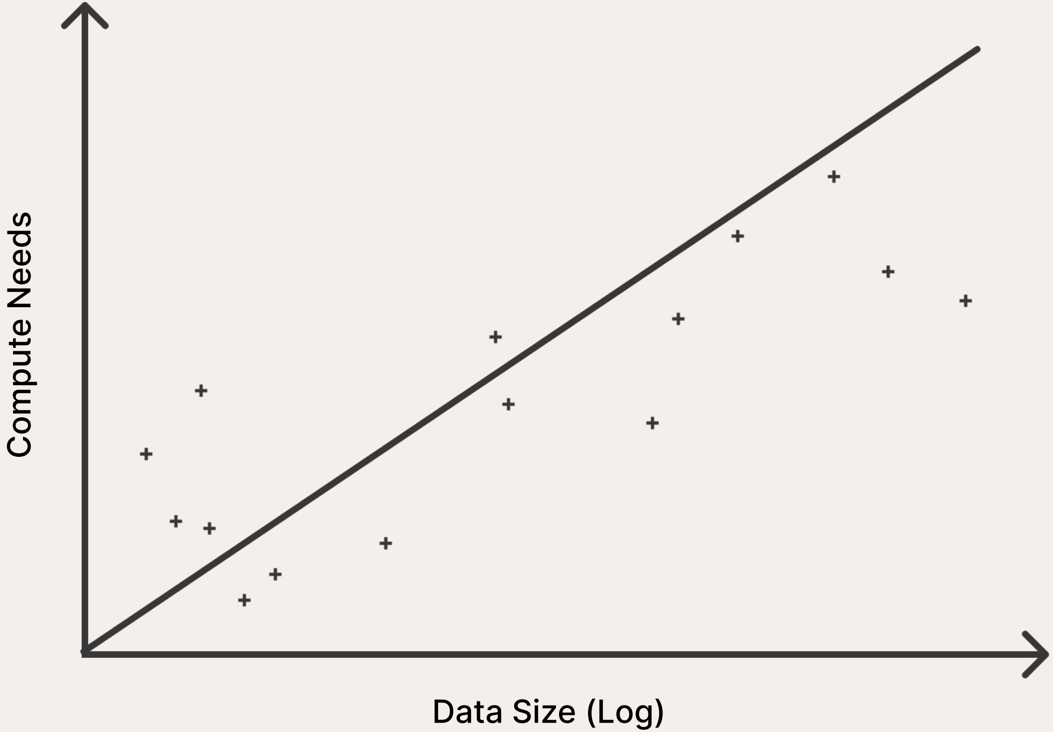


MotherDuck

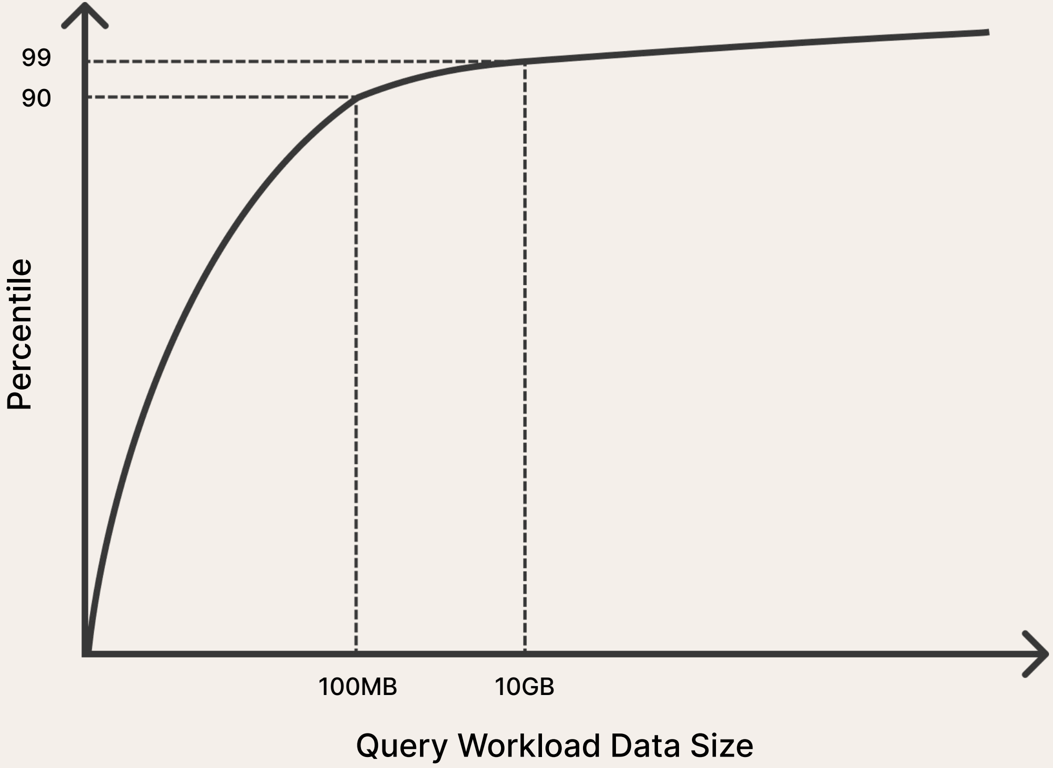
MOST PEOPLE DON'T HAVE THAT MUCH DATA



SEPARATION OF STORAGE AND COMPUTE FAVORS STORAGE

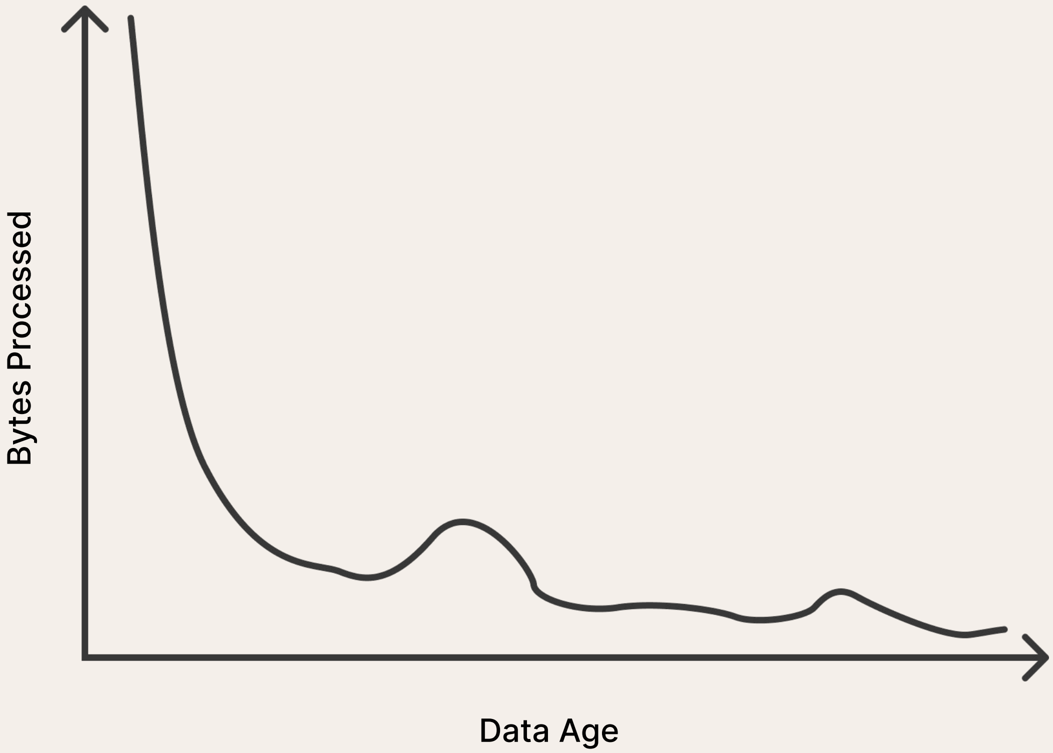


WORKLOADS USE LESS DATA THAN YOU THINK



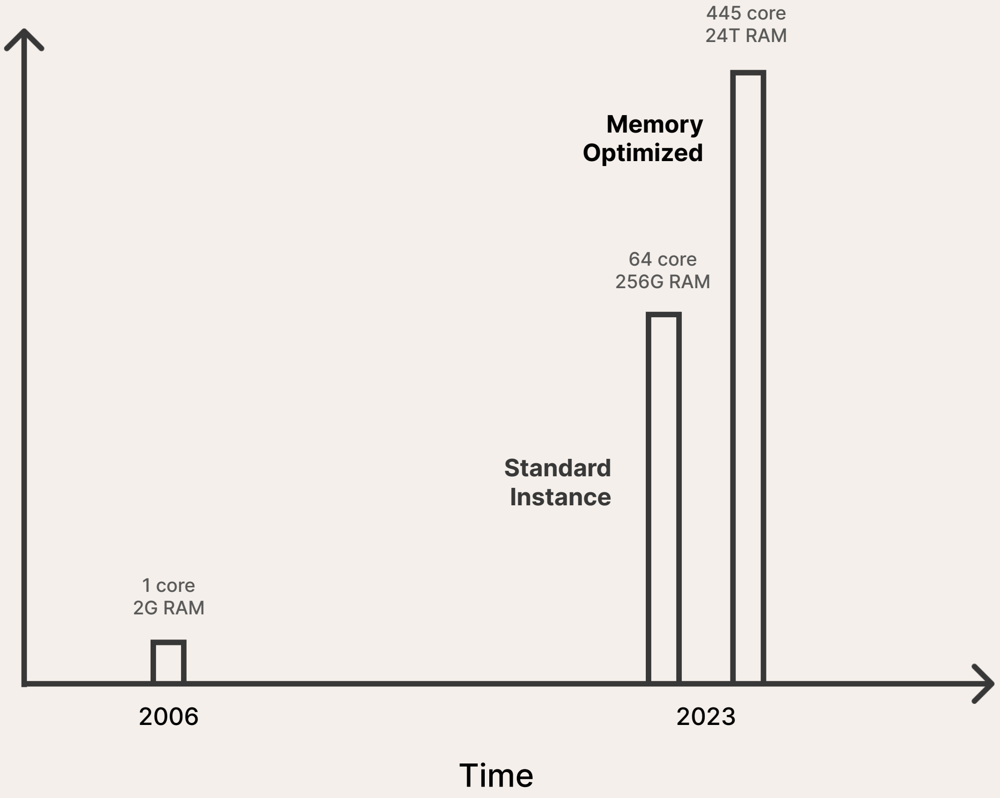
MotherDuck

MOST OF YOUR DATA JUST SITS THERE



MotherDuck

SCALE-UP IS NOT A DIRTY WORD



MotherDuck

DREMEL
IN A BOX

Dremel: Interactive Analysis of Web-Scale Datasets

Sergey Melnik, Andrey Gubarev, Jing Jing Long, Geoffrey Romer, Shiva Shivakumar, Matt Tolton, Theo Vassilakis
Google, Inc.

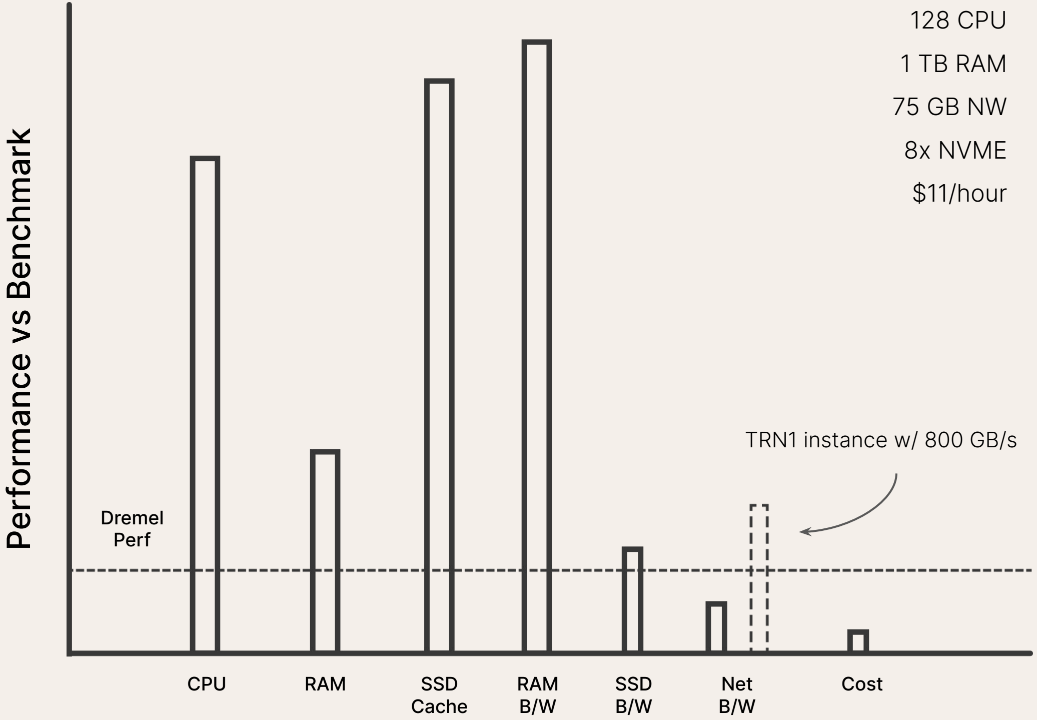
Dremel Performance:

87B Rows Scanned
512 GB processed
3000 nodes
~20 Seconds

Cached In Memory: ✓
Cached On SSD: ✓
Cold (S3): ✗ (✓?)

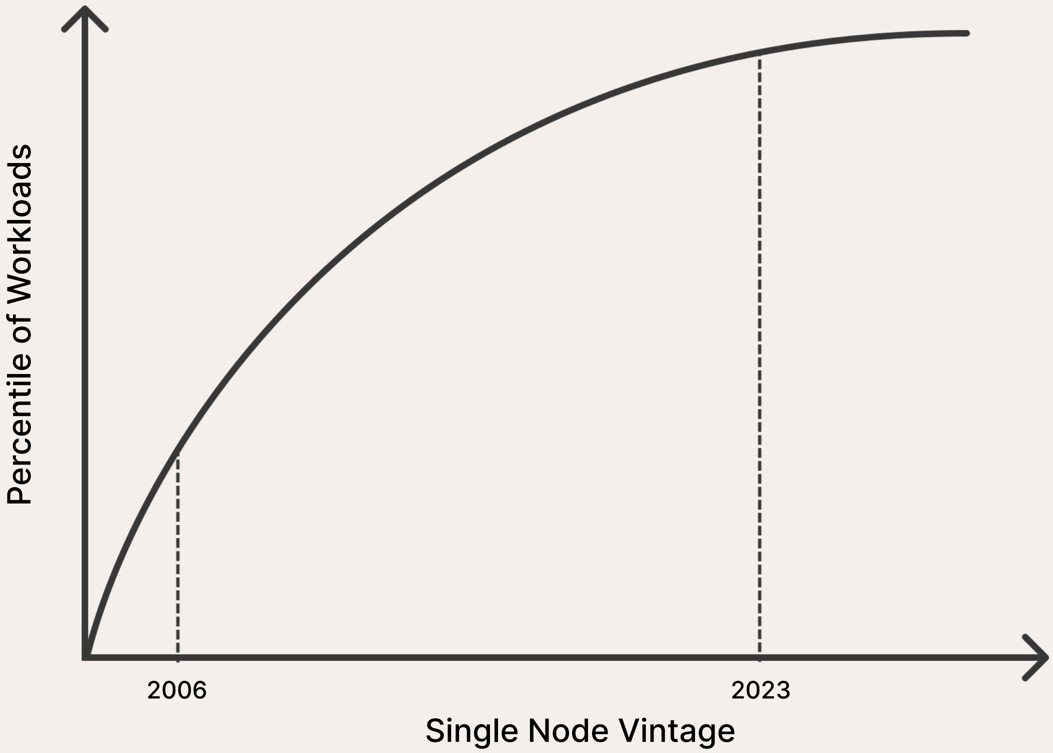
AWS I4i.metal

128 CPU
1 TB RAM
75 GB NW
8x NVME
$11/hour



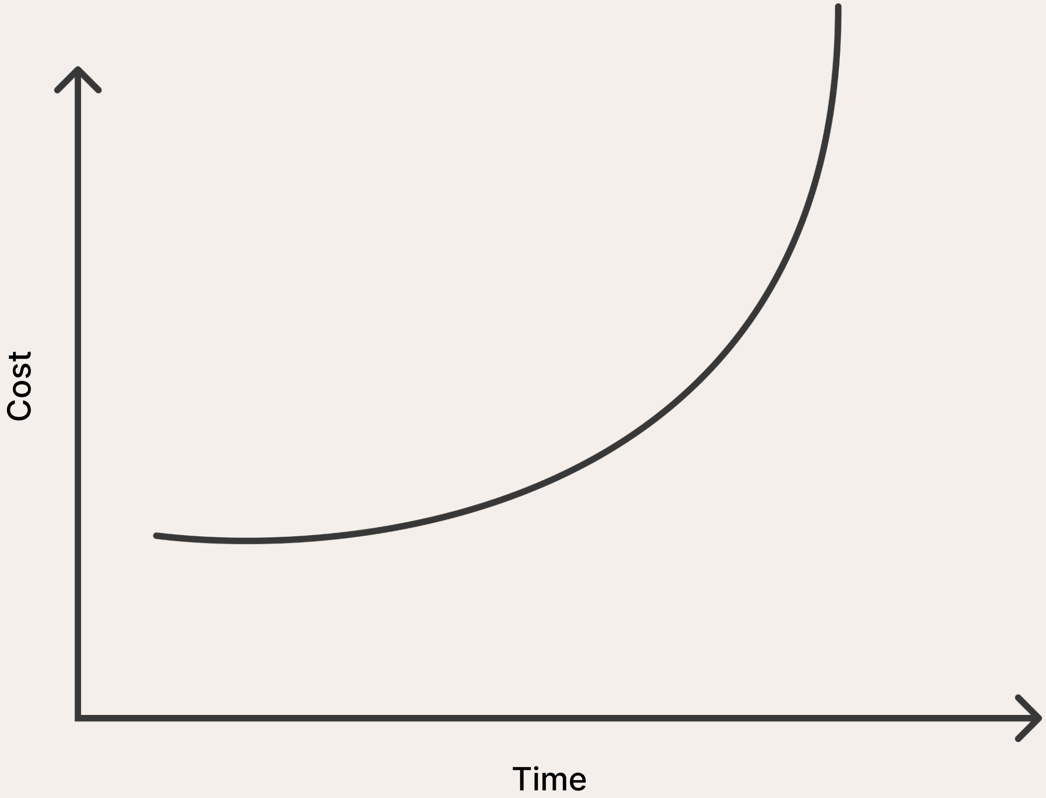
THE BIG DATA FRONTIER IS HEADING OFF INTO THE SUNSET

MotherDuck



DATA AS A LIABILITY

WHAT IS THE TRUE COST OF YOUR DATA?



MotherDuck

ARE YOU IN THE BIG DATA 1%

(and shouldn't we be building tools for the other 99%)

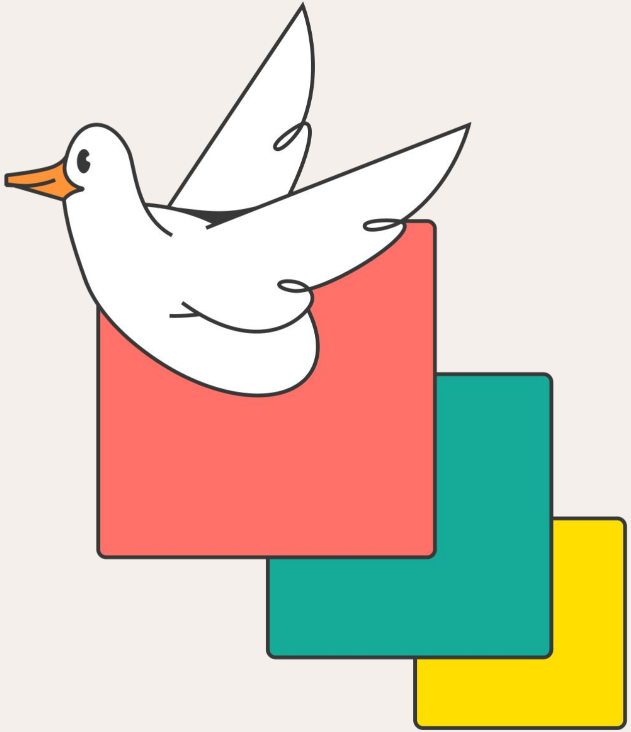
YOU HAVE BIG DATA WHEN:

- You've really got a huge amount of data
- AND you need to a lot of that data at once
- AND what you're accessing won't fit on one machine
- AND you're not hoarding your data
- AND you wouldn't be better off summarizing

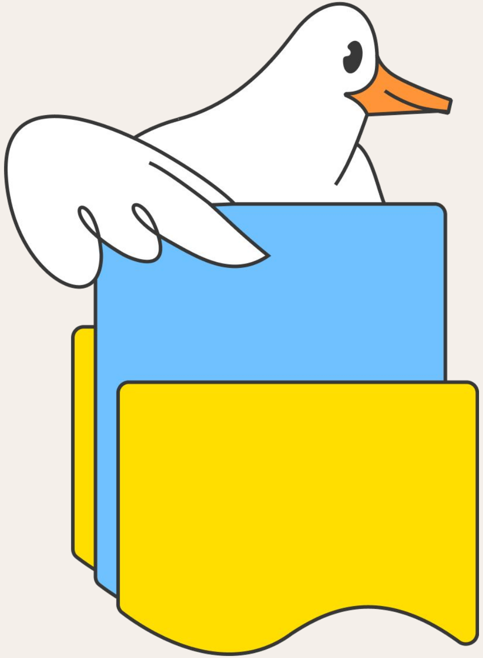
WHAT DOES THE WORLD LOOK LIKE IF SIZE IS NO LONGER THE PRIMARY DRIVER OF DATA ARCHITECTURES?

MotherDuck

DON’T BE AFRAID TO SCALE UP



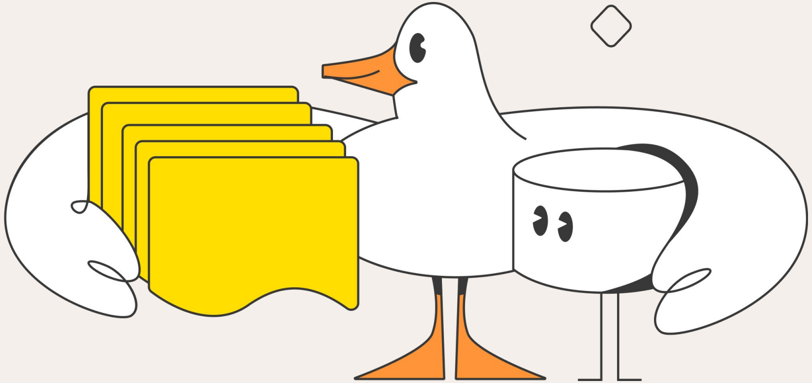
MotherDuck

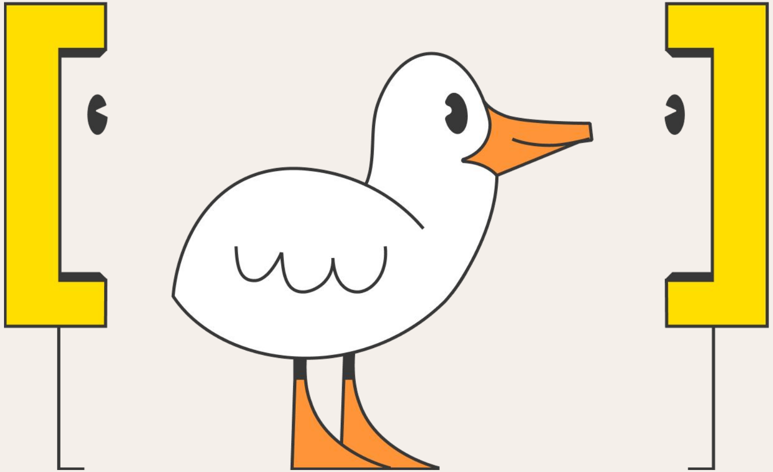


CLEAN UP YOUR DATA

MotherDuck

BRING DATA TO USERS

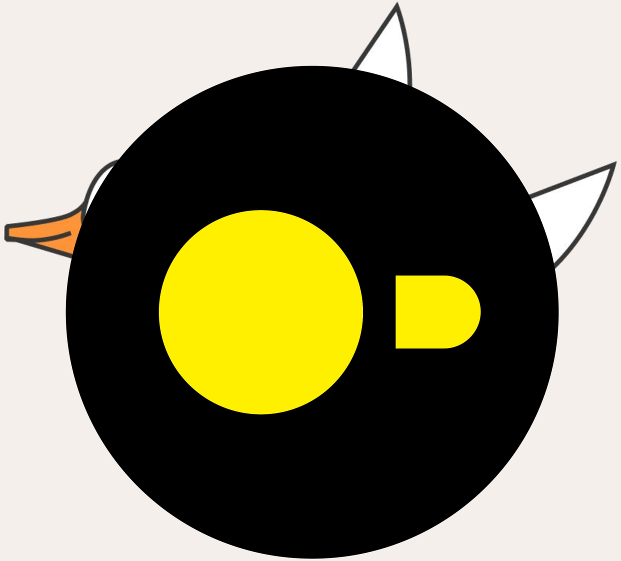




ONE SIZE
MAY NOT
FIT ALL

MotherDuck

LIFE IS BETTER WITH A DUCK



THE MODERN DUCK STACK HAPPY MEAL

Ingest
+
DuckDB

Analyze
+
DuckDB

Visualize
+
DuckDB

THANK YOU!

Don’t be a data hoarder

MotherDuck