

Saving the Planet one Query at a Time

Dr. Hannes Mühleisen



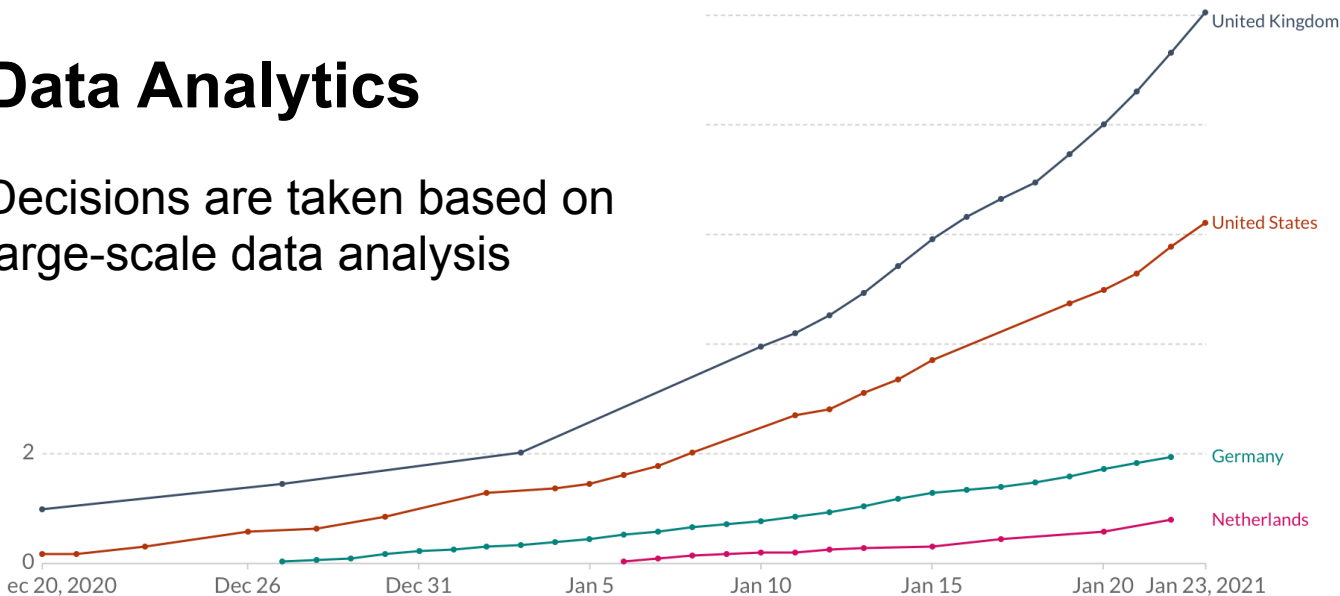
Cumulative COVID-19 vaccination doses administered per 100 people

This is counted as a single dose, and may not equal the total number of people vaccinated, depending on the specific dose regime (e.g. people receive multiple doses).

Our World
in Data

Data Analytics

Decisions are taken based on
large-scale data analysis

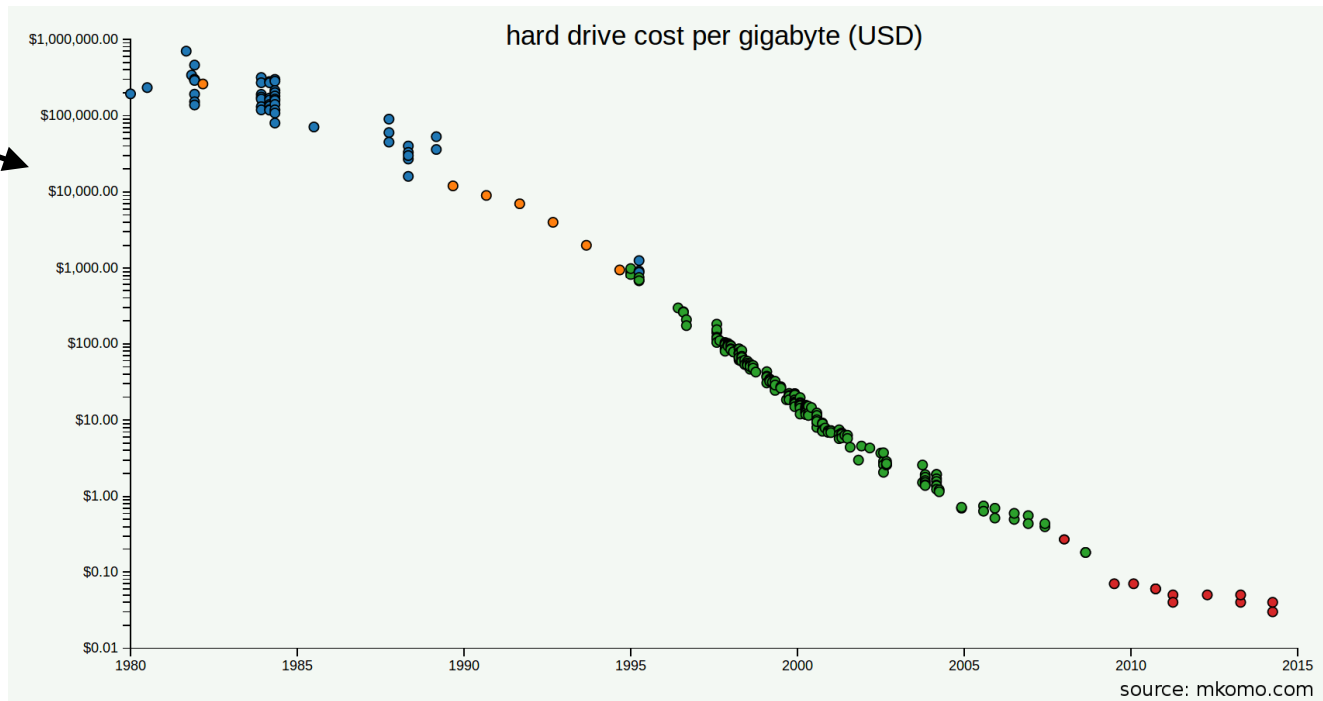
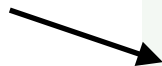


Source: Official data collated by Our World in Data - Last updated 24 January, 20:00 (London time)

CC BY

Small Price, Big Data

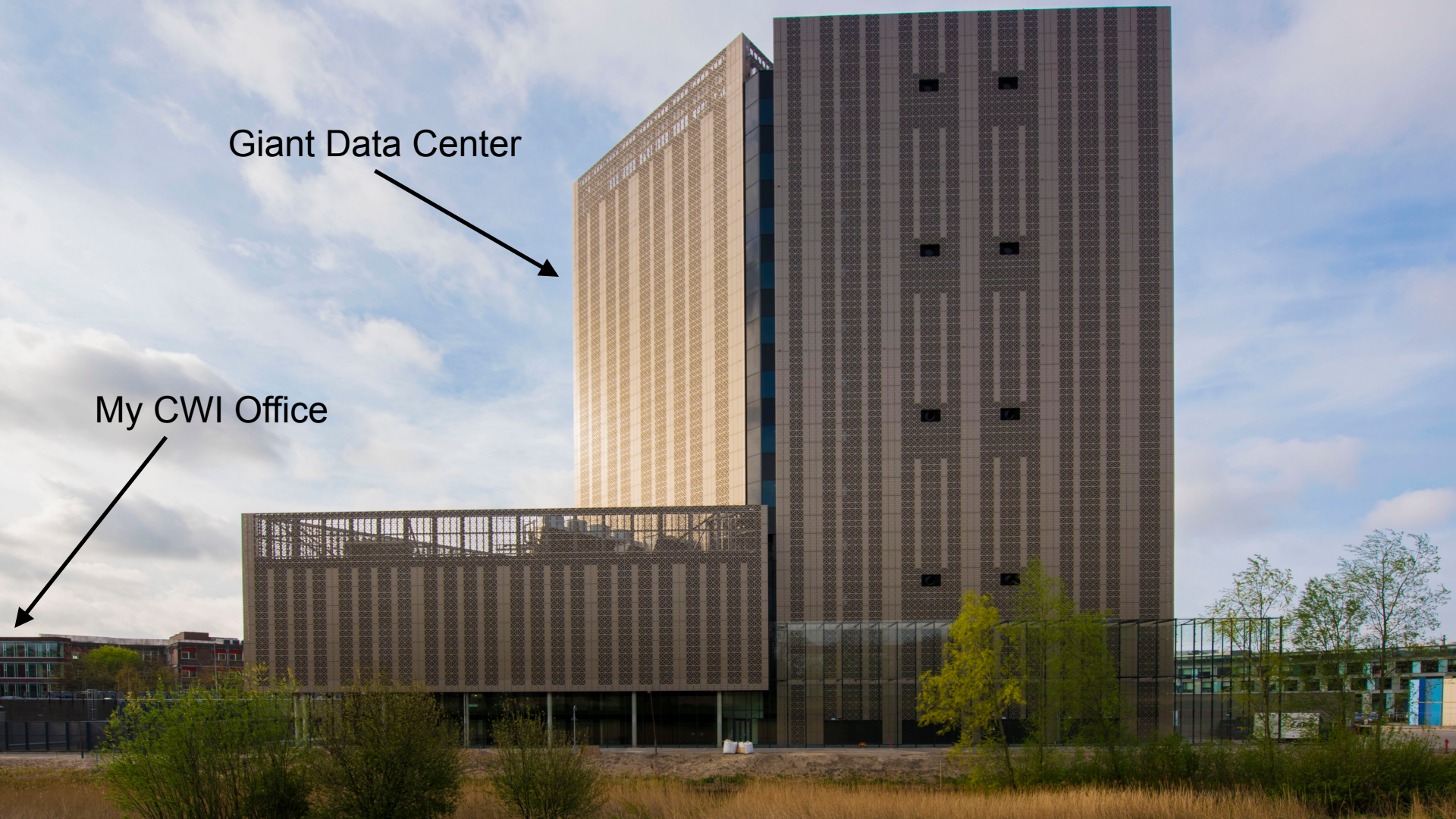
Log Scale



Giant Data Center



My CWI Office



Horizontal scaling

Standard solution, used in e.g. Spark, Flink,...

Data analysis slow? Add more computers!

Why is this problematic?

Little incentive to improve cycles/value efficiency

Enormous resource waste, Hardware, Power, CO₂, €



Efficient resource usage

Massive compute power in laptops, devices ...

90%* of “Big Data” feasible on single node

Vertical scaling by increasing efficiency

Lower cycles/value





Bringing **Big Data** back to **single node**

Feasible to analyze 10s of GB on MacBook

Available at no cost

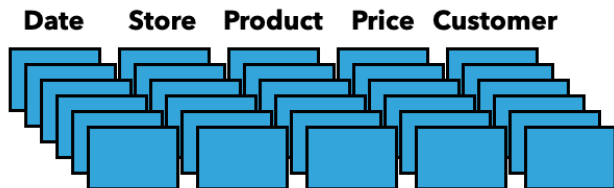
duckdb.org

Why “DuckDB”?

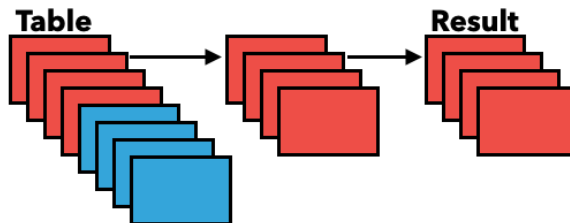


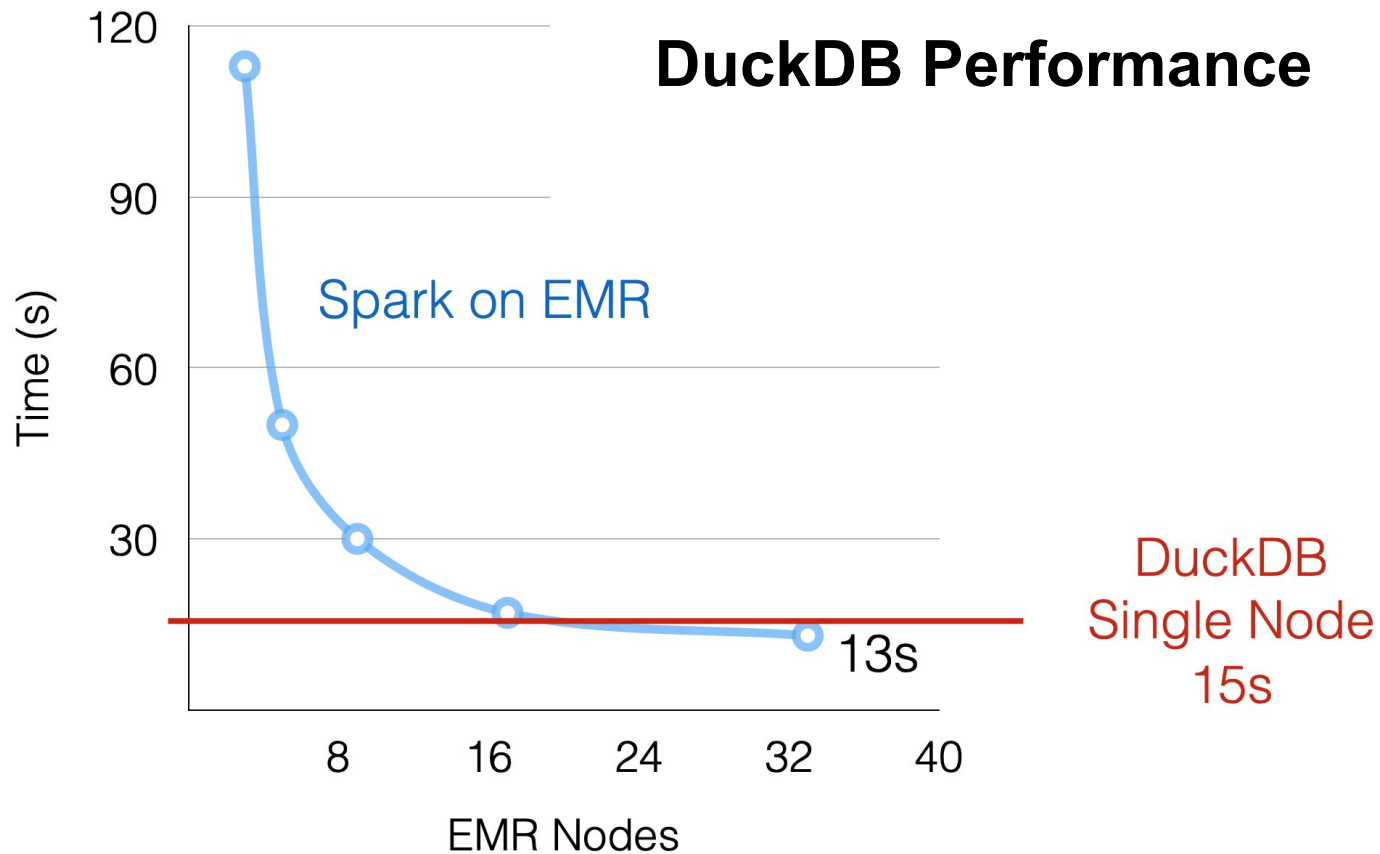
CWI Innovations in DuckDB

Column-Store



Vectorized Processing





DuckDB Engineering

No “research prototype”

Built for serious usage

Millions of tests on every change

More impact than research paper

Direct benefit to taxpayer





Hacker News new | threads | past | comments | ask | show | jobs | submit

1. DuckDB – An embeddable SQL database like SQLite, but supports Postgres f
335 points by pcr910303 8 hours ago | hide | 62 comments
2. ▲ Raspberry Pi – UASP, Trim, and Boot Performance via USB (jeffgeerling.com)
98 points by geerlingguy 5 hours ago | hide | 11 comments
3. ▲ Base 65536 (github.com)
38 points by leoh 3 hours ago | hide | 14 comments
4. ▲ Mouse found atop a 22,000-foot volcano, breaking world record (nationalgeogr
29 points by greenyoda 3 hours ago | hide | 4 comments
5. ▲ Windows Server vulnerability requires immediate attention (cisa.gov)
362 points by ohjeez 14 hours ago | hide | 80 comments

Future Data

De-centralization of data

DuckDB rules the world

But: Parkinson's Law

“Work expands so as
to fill the time available
for its completion”



Q&A



hannes@cwi.nl



[@hfmuehleisen](https://twitter.com/hfmuehleisen)
duckdb.org