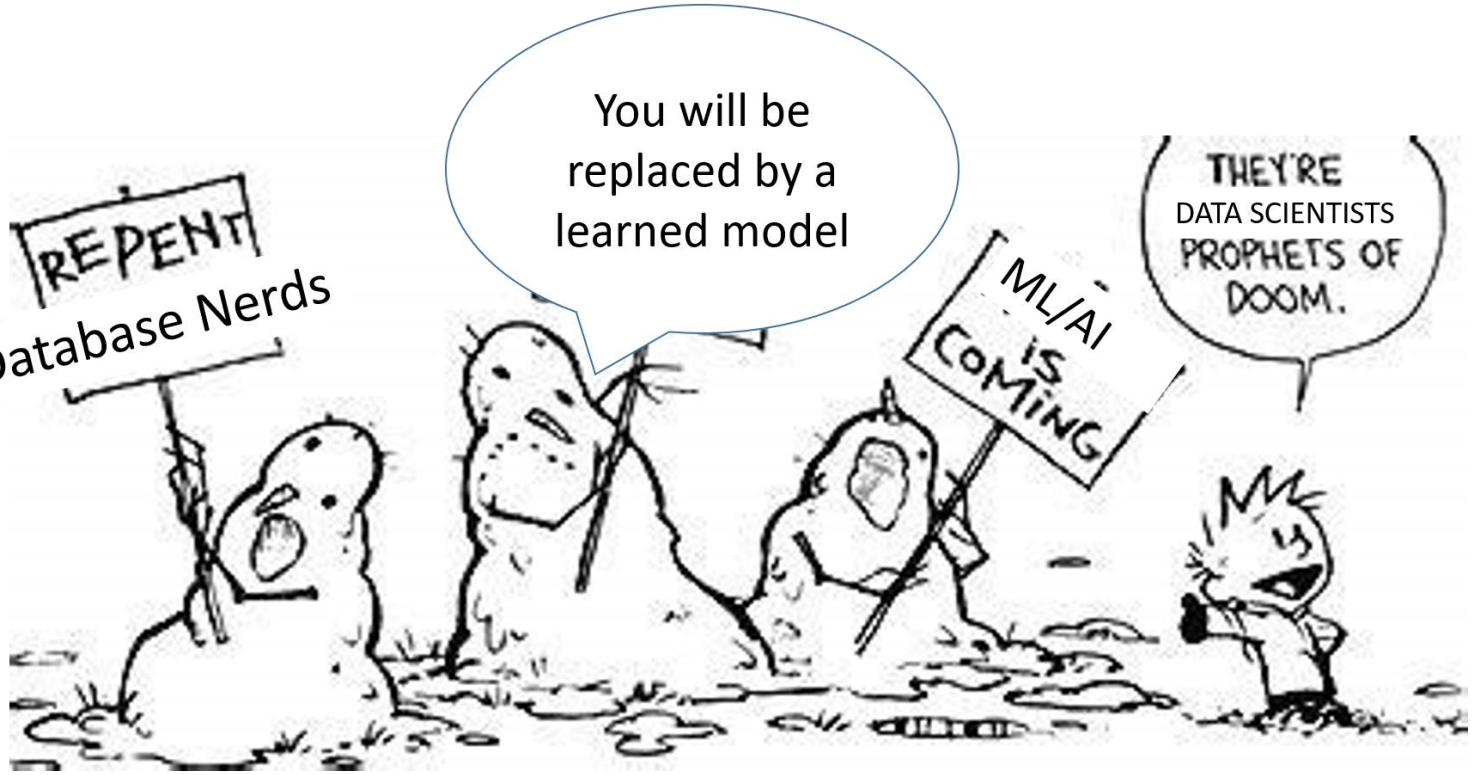
A New Golden Era for Data Management

Gustavo Alonso
Systems Group
Department of Computer Science
ETH Zurich, Switzerland

Background


REPENT
Database Nerds
You will be replaced by a learned model
ML/AI IS COMING
THEY'RE DATA SCIENTISTS PROPHETS OF DOOM.

The Database Community

- The Good
  - Great success (in research, in industry, in academia)
  - Great people (one of the best communities I know along all dimensions)
- The Bad
  - Incredibly insular and isolated
  - Too prone to fall for the latest gimmick
- The Ugly
  - Too much influence from short term interests of companies
  - Very difficult place for truly forward-looking research
  - No citations to original material, we forget our history

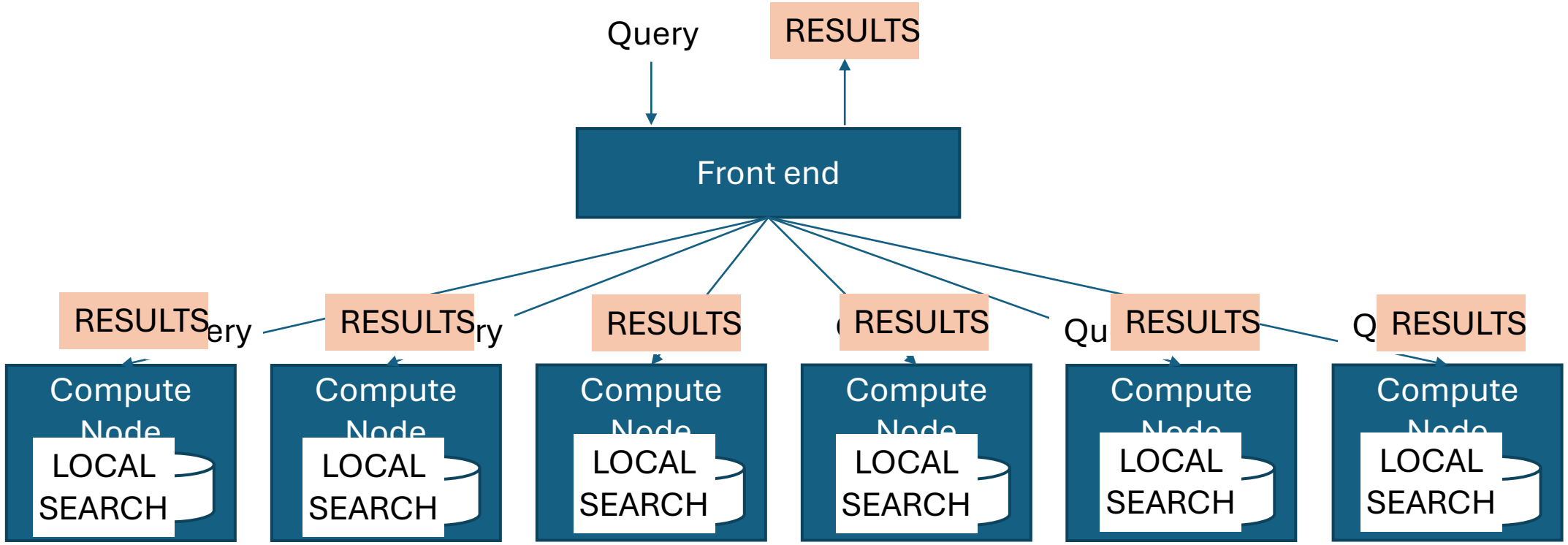
Data management research
is too gregarious, incremental,
and driven not by technology
advances but by marketing pitches.
Academic research should look
further into the future than companies

The Hardware Era

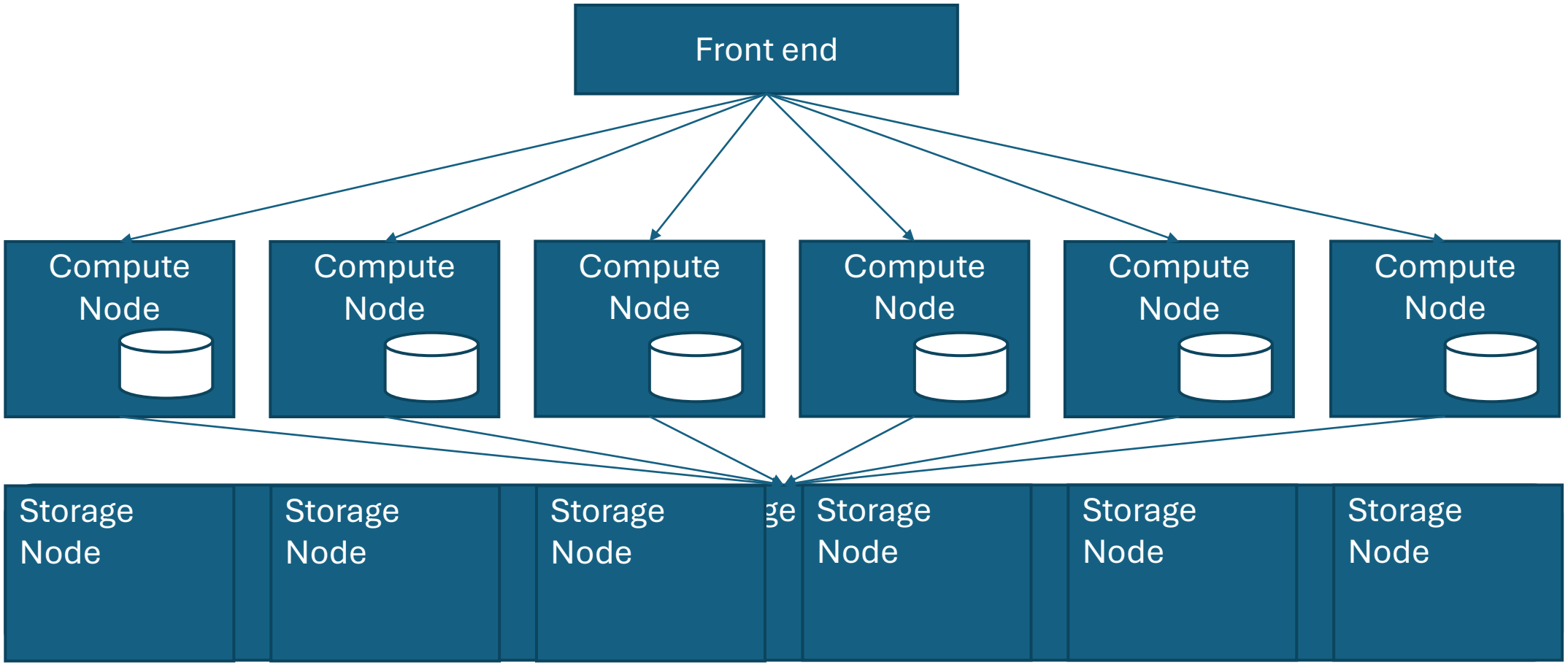
How we are ignoring one of the biggest changes in IT history

Data processing systems
are designed in a reactive manner,
they adapt to the evolving hardware,
they do not influence the hardware

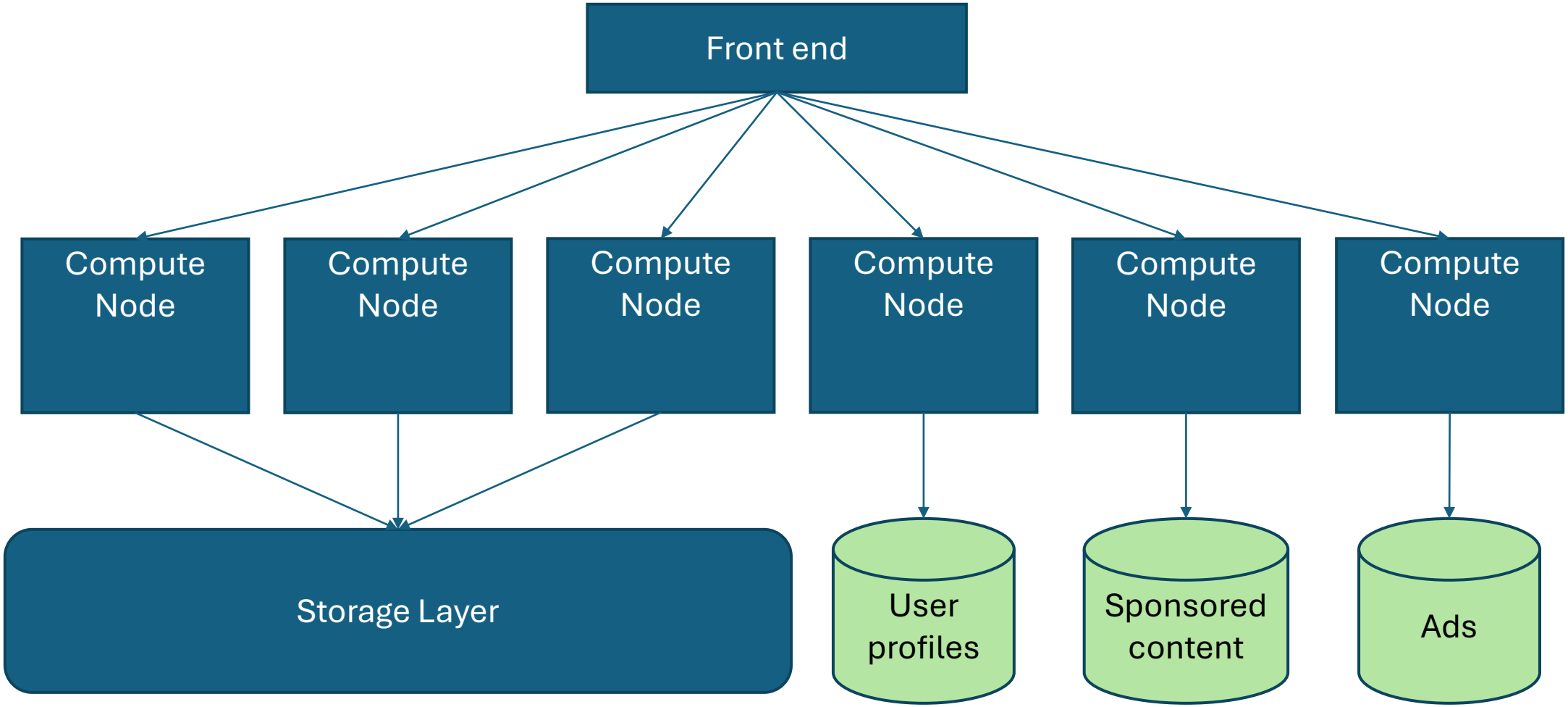
The cloud design pattern: search engine



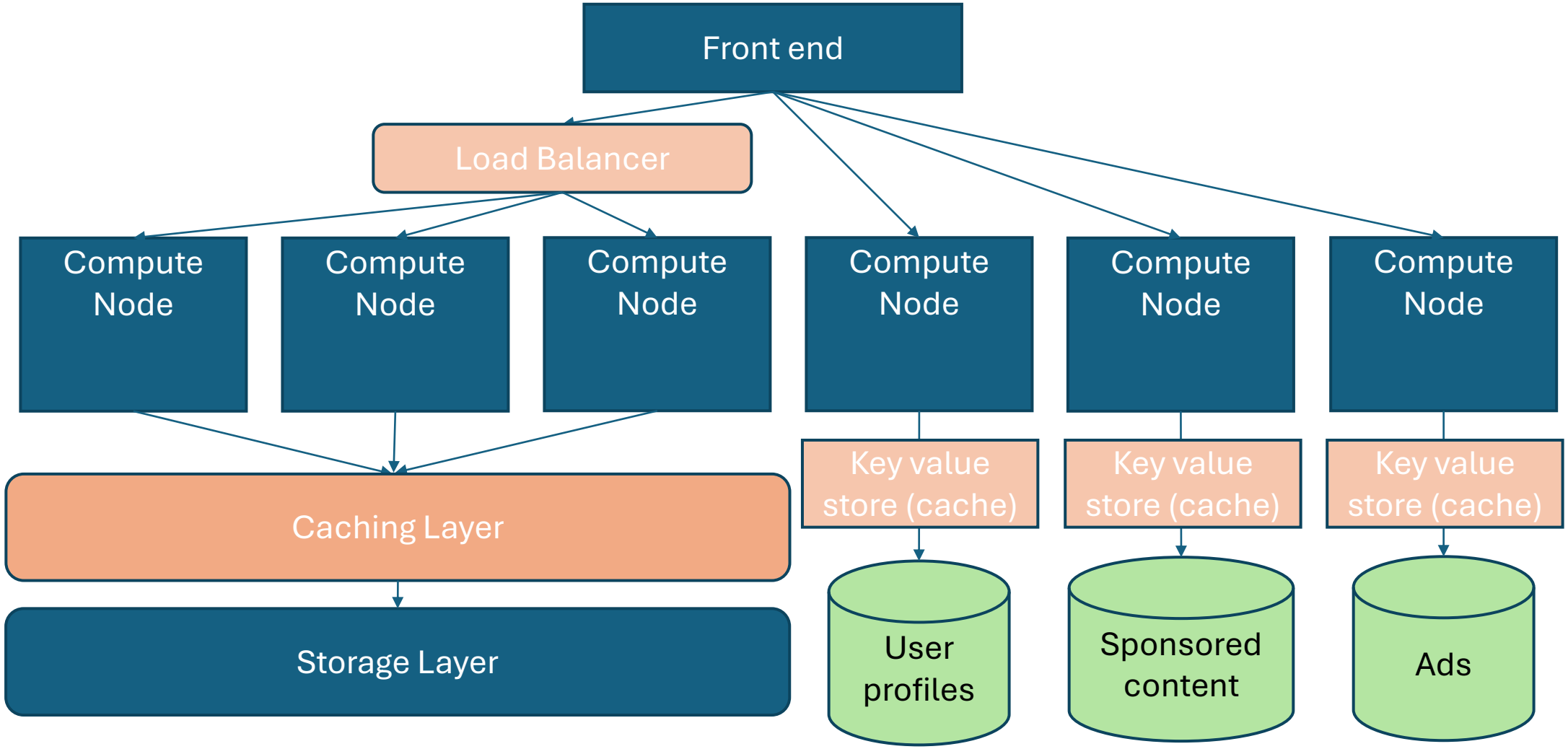
Separation of compute and storage



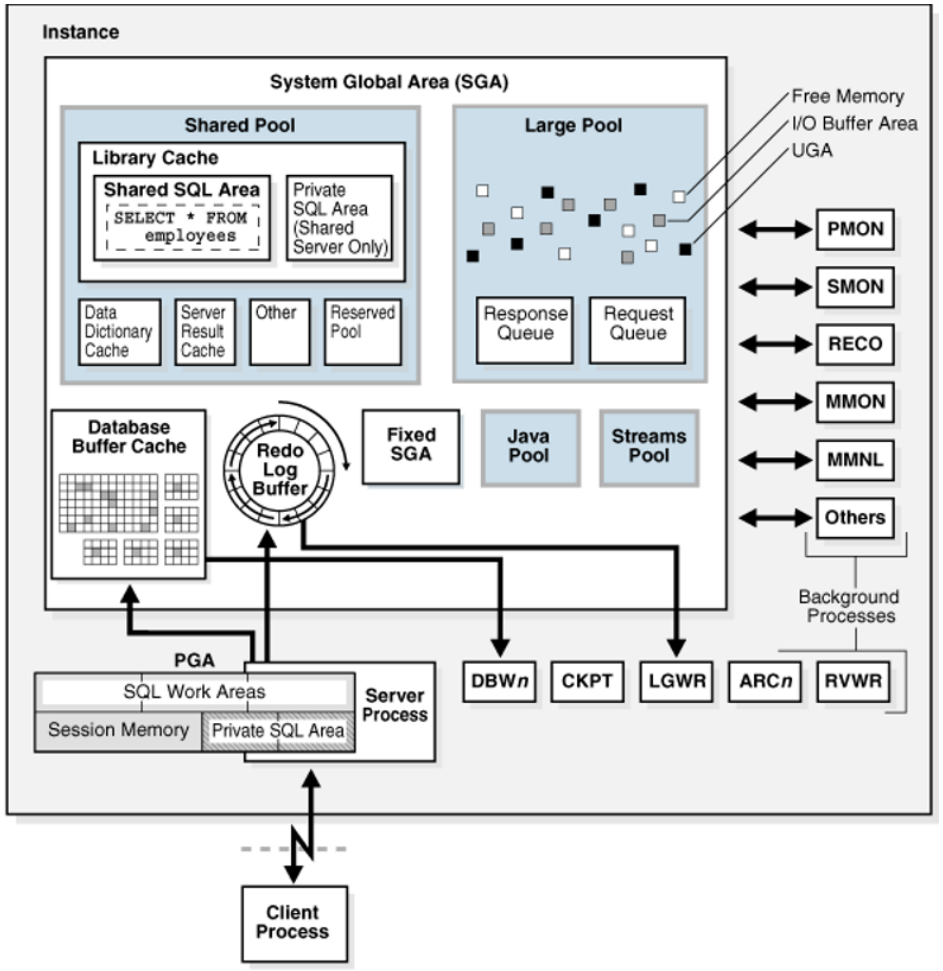
Additional functionality

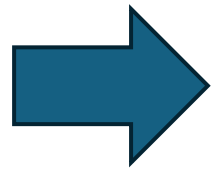


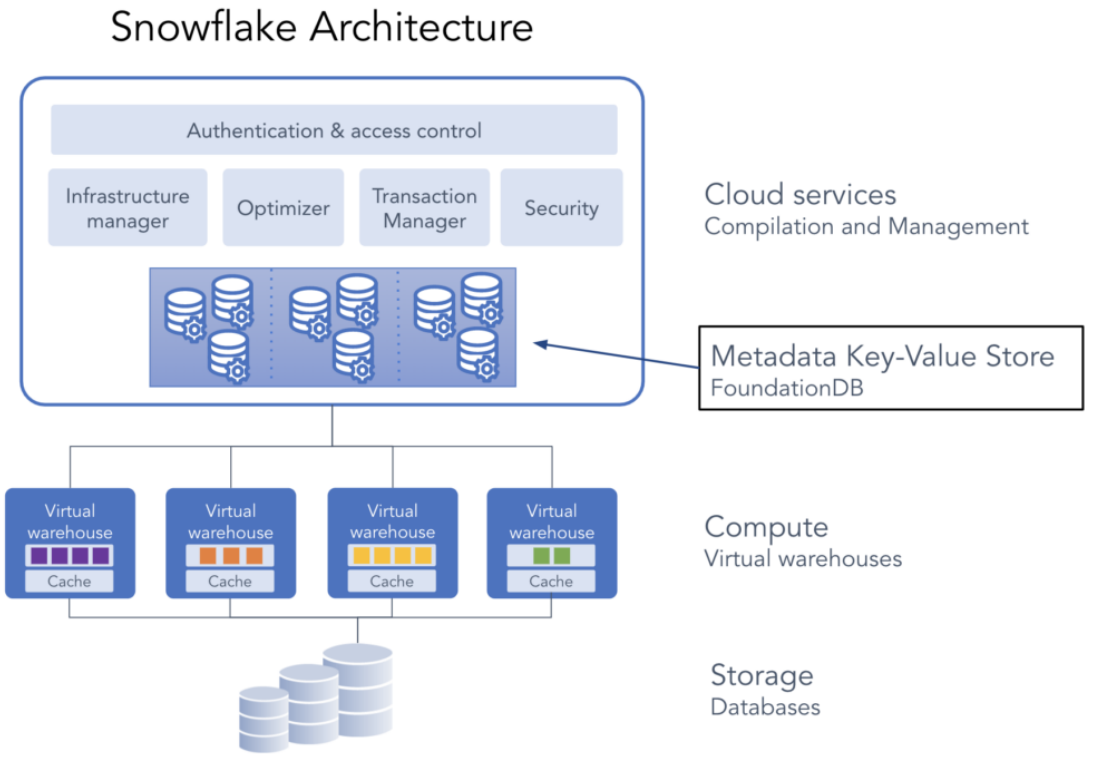
Compensate for the design problems



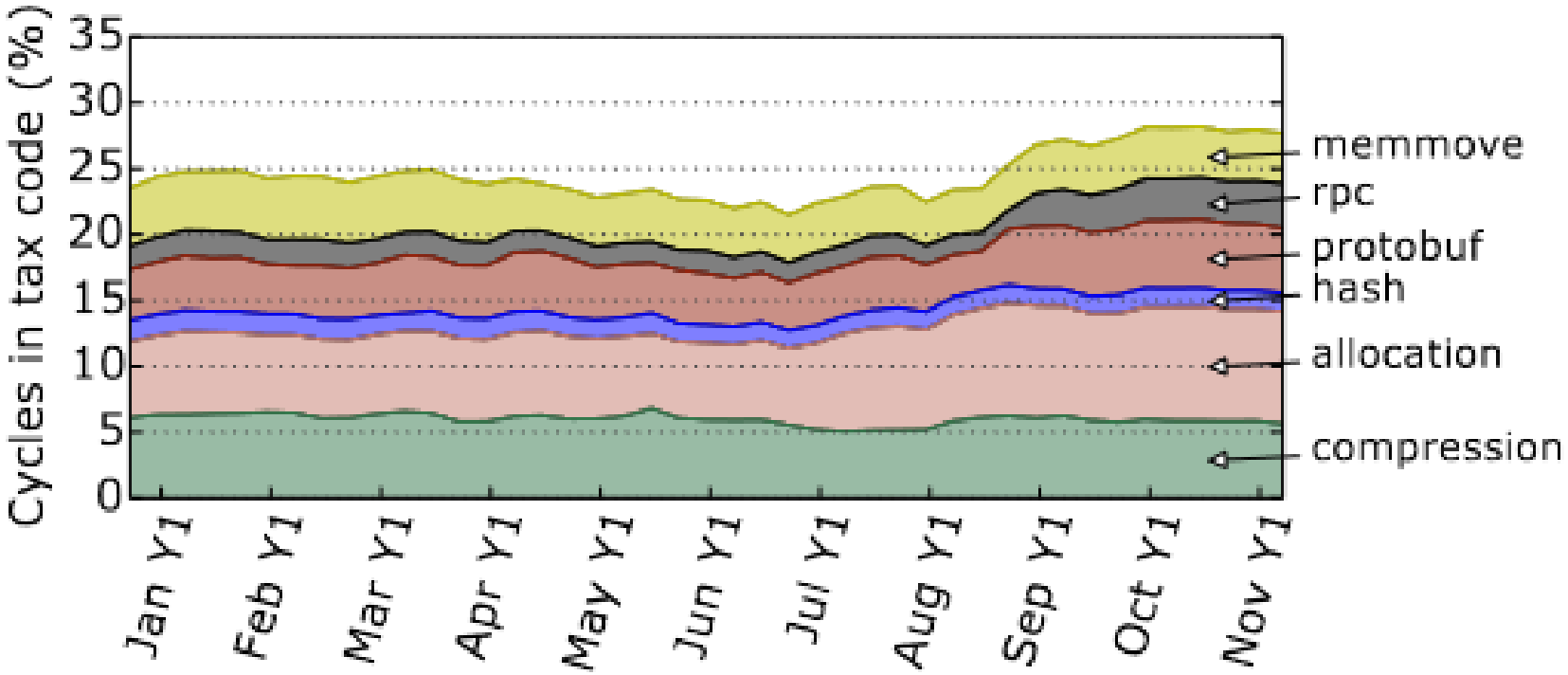
We build for the hardware that is available





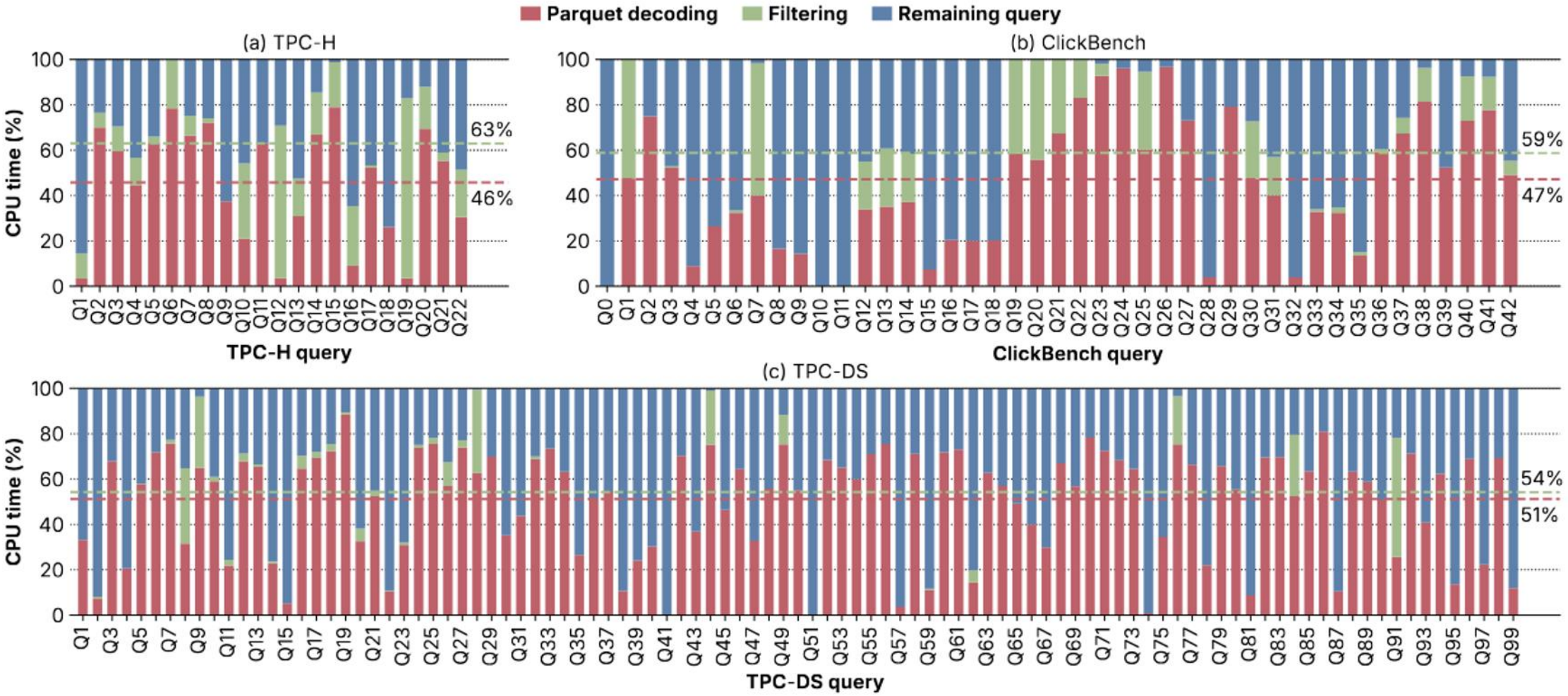


The Data Center Tax is a data movement tax



Profiling a warehouse-scale computer, ISCA 2015

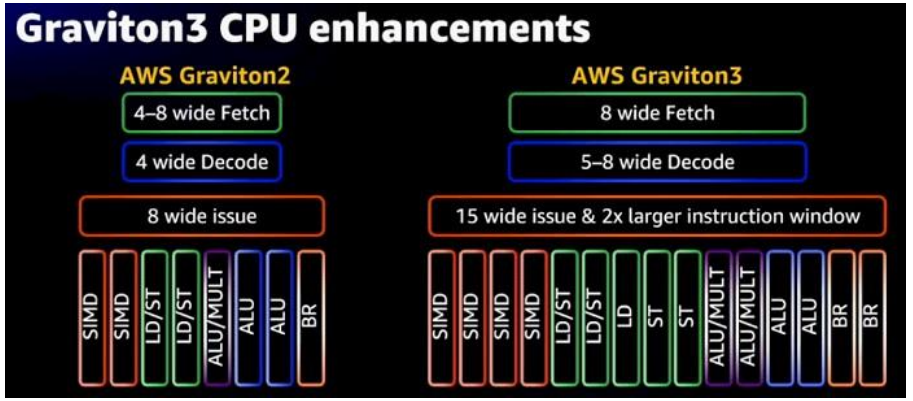
The Lakehouse tax

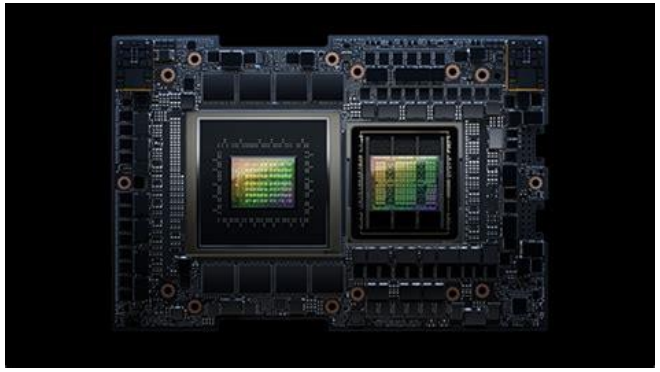


Should I hide my duck in the lake. Jonas Dann et al. DAMON 2026

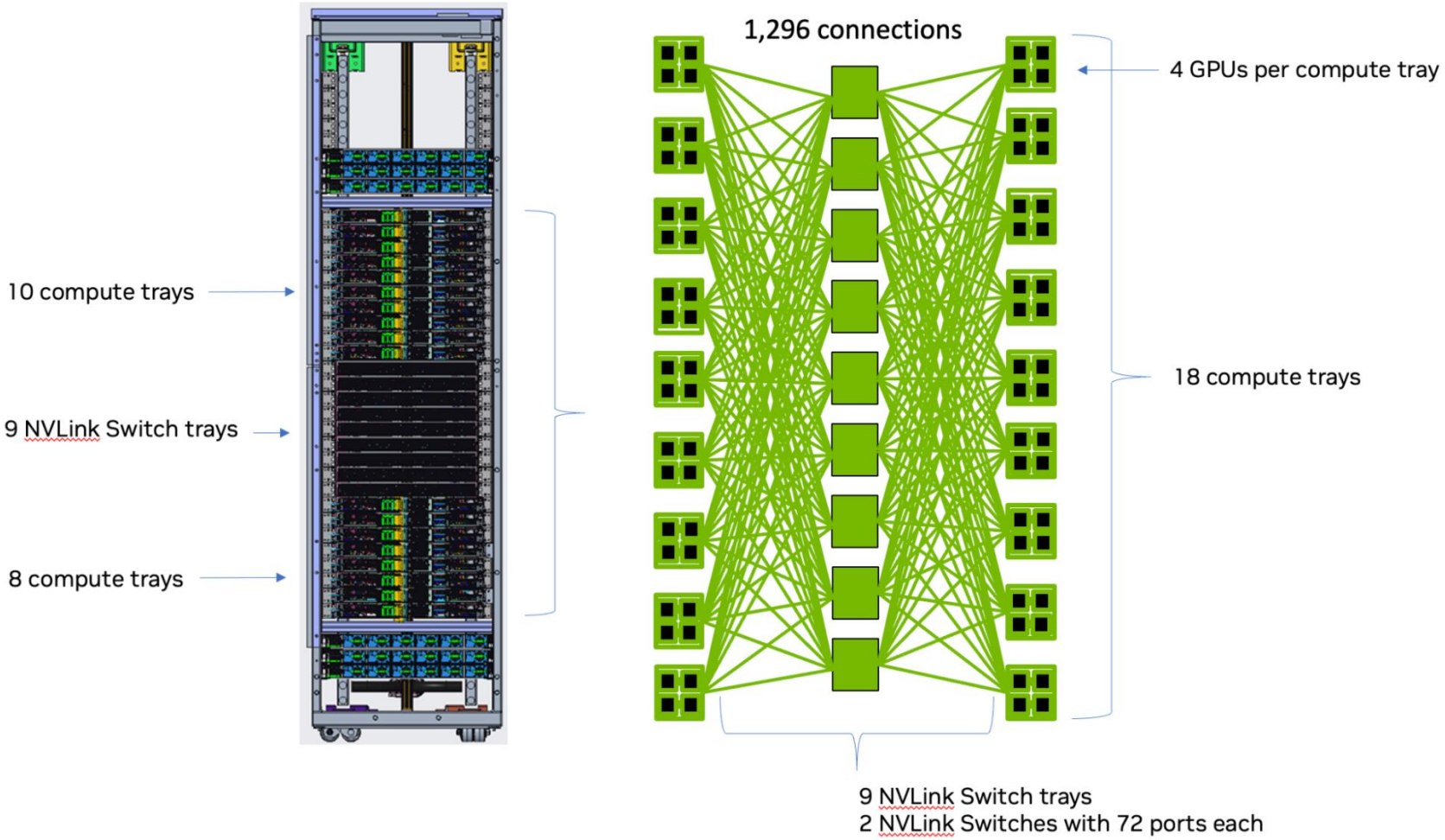
Hardware to reduce the tax

- Hardware evolving very quickly
  - New CPU architectures
  - New ISAs
  - New interconnects
  - New accelerators
  - Specialized processors
- How much of this hardware helps data processing?
- How much of this hardware has been developed for data processing?

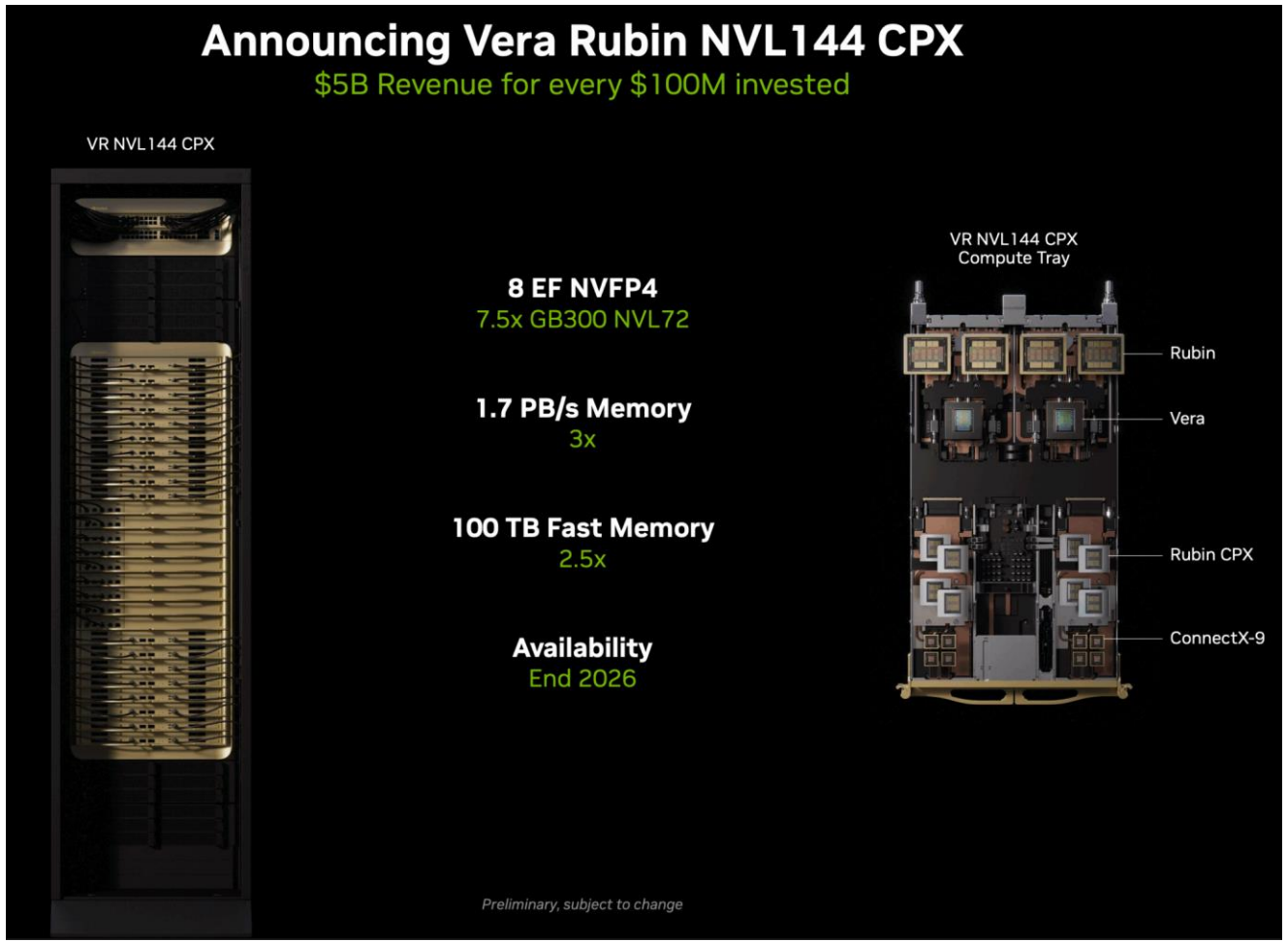


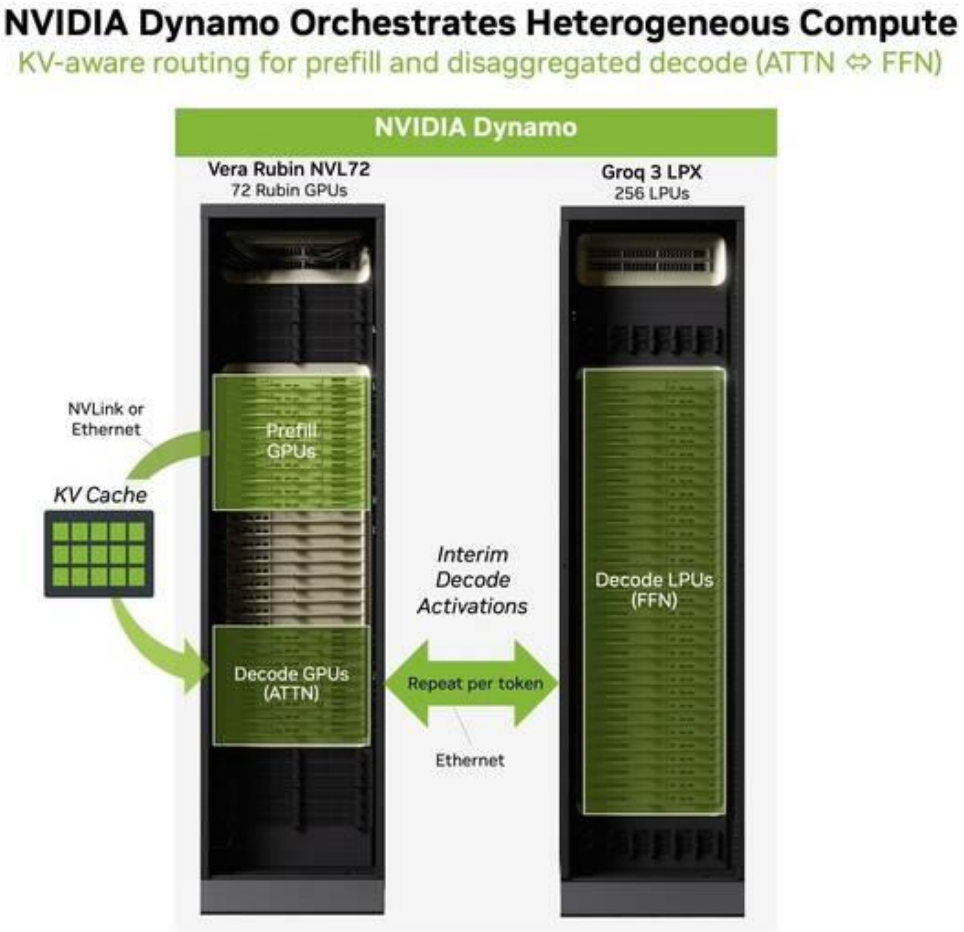


Future hardware NVIDIA NVL 72



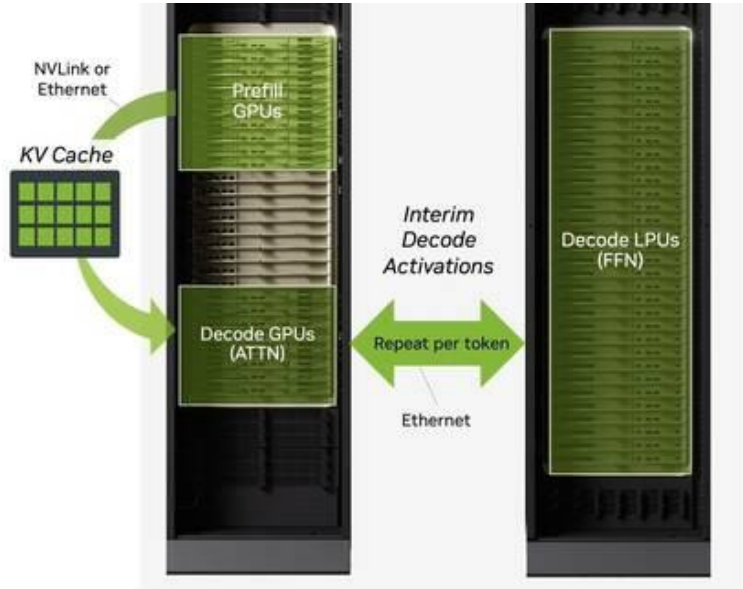
The hardware lottery





The hardware lottery

Do you think data processing will run on an NVL 144?



The advances in AI are the result of unprecedented advances in hardware and computer architecture.
There is more to gain by focusing on the advances in hardware than on the advances in AI

What is there for databases?

Outlining a research program

Raghu Ramakrishnan

Technical Fellow, CTO for Data

In 3 to 5 years, data analytics will run on GPUs and FPGAs

Ignore new hardware at
your own risk.
Everything that truly matters will
move to hardware.
If your software does not use the latest
hardware, it will be irrelevant

Why specialization is possible

- Data processing engines evolution:

- Specialization impossible in conventional engines
- Now the engine runs on the cloud
- The cloud is highly specialized …

  … see previous slides on search engines

Many possibilities yet unexplored

- Accelerators are here to stay
- New processors and ISAs are here to stay
- AI is changing computing in fundamental ways
  - New number formats and data representations
  - Specialized hardware for key operations
  - Scale out architectures instead of distributed ones
  - Shift to GPUs, TPUs, DPUs
  - Very, very large-scale systems

Many possibilities yet unexplored

- Accelerators are here to stay
- New processors and ISAs are here to stay
- AI is cha
  - New r
  - Speci
  - Scale out architectures instead of distributed ones
  - Shift to GPUs, TPUs, DPUs
  - Very, very large-scale systems

What is the matrix multiplication of data processing?

Old ideas for the old dog

- Schema integration and database federation
  - DATA LAKES, LAKEHOUSES
- Sagas and complex transactions
  - DATA LAKES, DATA FLOW SYSTEMS, AGENTIC AI
- Query optimization over heterogeneous platforms
  - FOR EVERYTHING BUT CPUs
- Storage level processing
  - SMART STORAGE, DEEP MEMORY HIERARCHIES
- System architecture
  - DATABASE APPLIANCES (as opposed to cloud)

This is the best of times for database systems research. None of our assumptions and system design principles hold anymore.

LET'S REINVENT DATA PROCESSING FOR THE XXI CENTURY