Python-Blosc2
Compress Better, Compute Bigger!

Francesc Alted / francesc@ironarray.io

Luke Shaw / luke.shaw@ironarray.io

ironArray

PyData Global

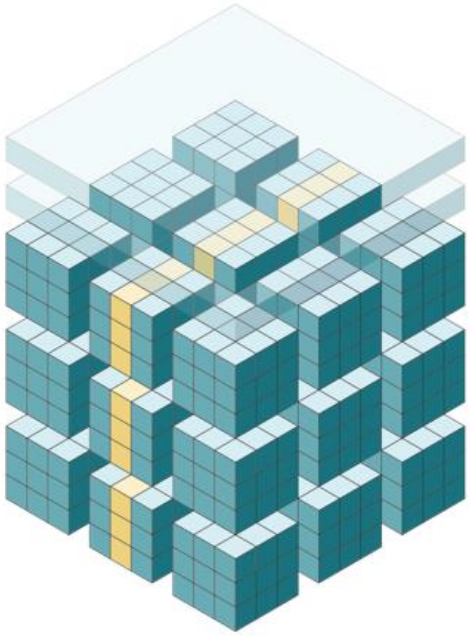
December 10th 2025

Who is ironArray SLU?

- We are the developers of PyTables, numexpr and Blosc libraries.
- Team of experts empowering you to harness the full potential of compression for big data: we are here to help!

ironArray

https://ironarray.io



Blosc2

Better compression for multidimensional, binary data

https://www.blosc.org/

What Is Blosc2?

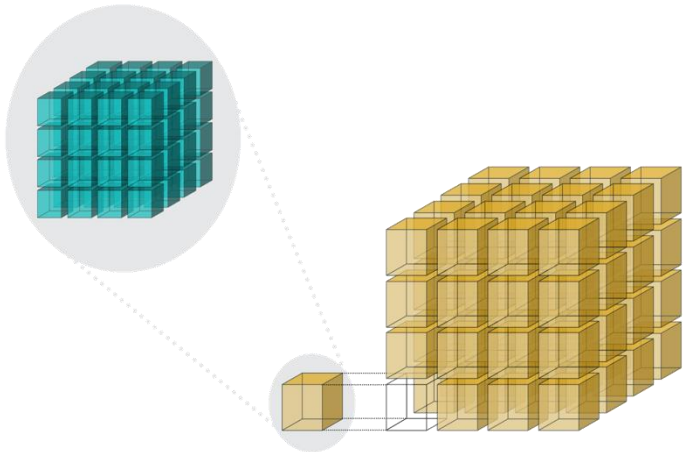
A set of codecs and filters to compress data, orchestrated in a way that **leverages modern computer architecture**:

- Support for **multithreading**: use the full cores in your CPU.
- Automatically use of **modern SIMD instructions** (SSE2, AVX2, AVX512, NEON, ALTIVEC) for performance.
- **Double partitioning**, mimicking the multi-level caches in modern CPUs, and adapted to their sizes automatically.
- **In-memory** or **on-disk storage**.

https://github.com/Blosc/c-blosc2

Blosc2 Architecture

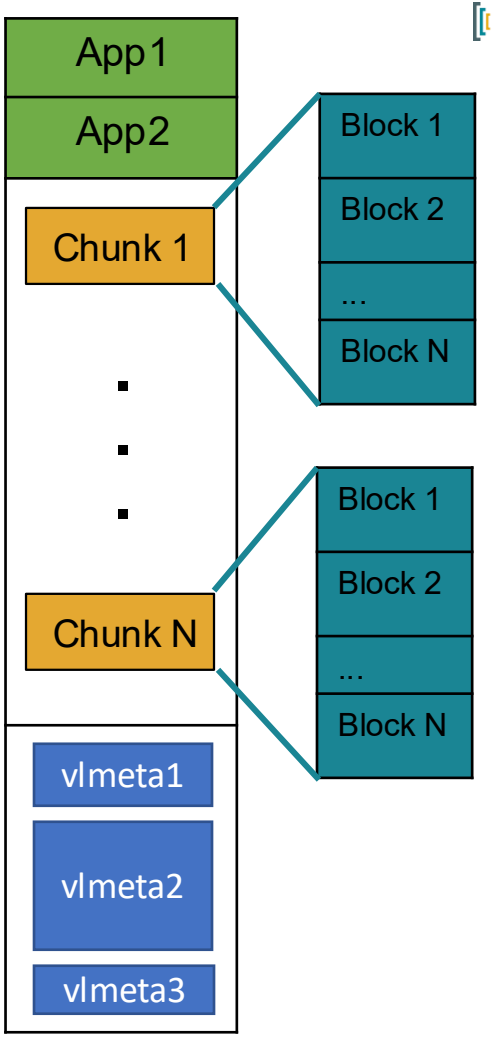
- ✓ 64-bit containers
- ✓ Metalayers for adding info for apps and users
- ✓ Blosc2 NDim: Multi-dim blocks and chunks



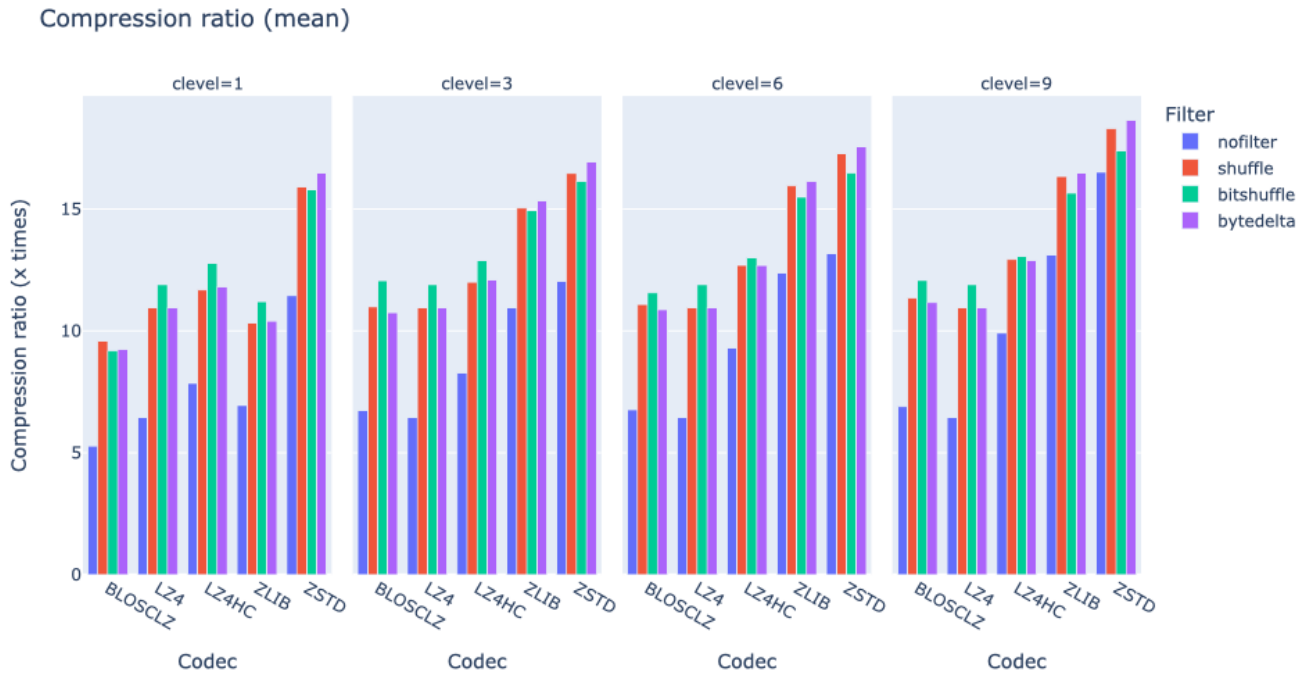
Header:
Fixed Length
Metalayers

Data:
Super-Chunk

Trailer:
Var Length
Metalayers
(up to 2 GB)



Different Codecs and Filters



How to predict the best combination?
https://ironarray.io/btune

Blosc2: Compute Engine

Compute with your big compressed arrays, fast!

Blosc2 Compute Engine: Computing With Compressed Data, Transparently

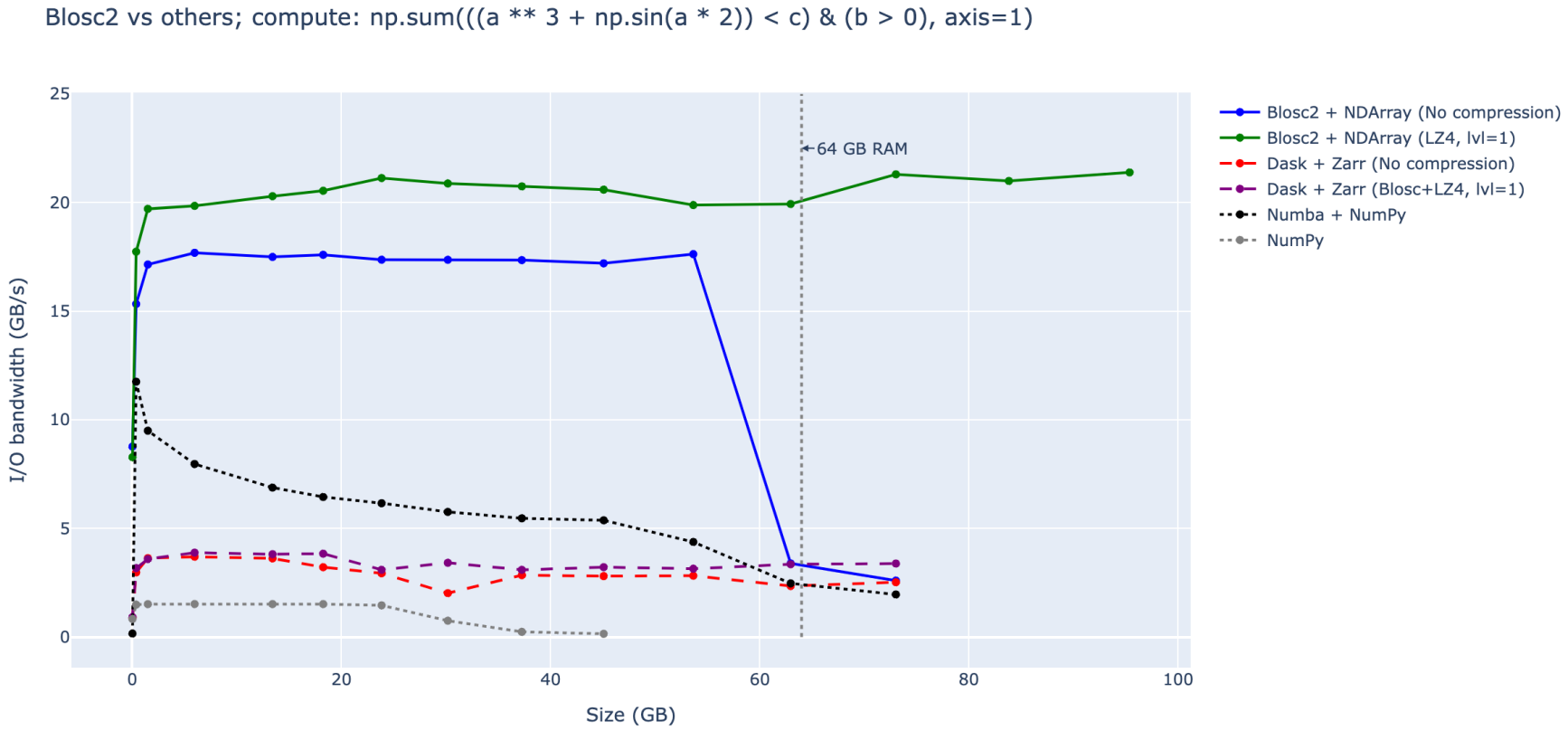
For optimal speed, it's crucial to **understand** and **utilize** modern CPU capabilities:

- Multicore processing
- SIMD
- Cache hierarchies

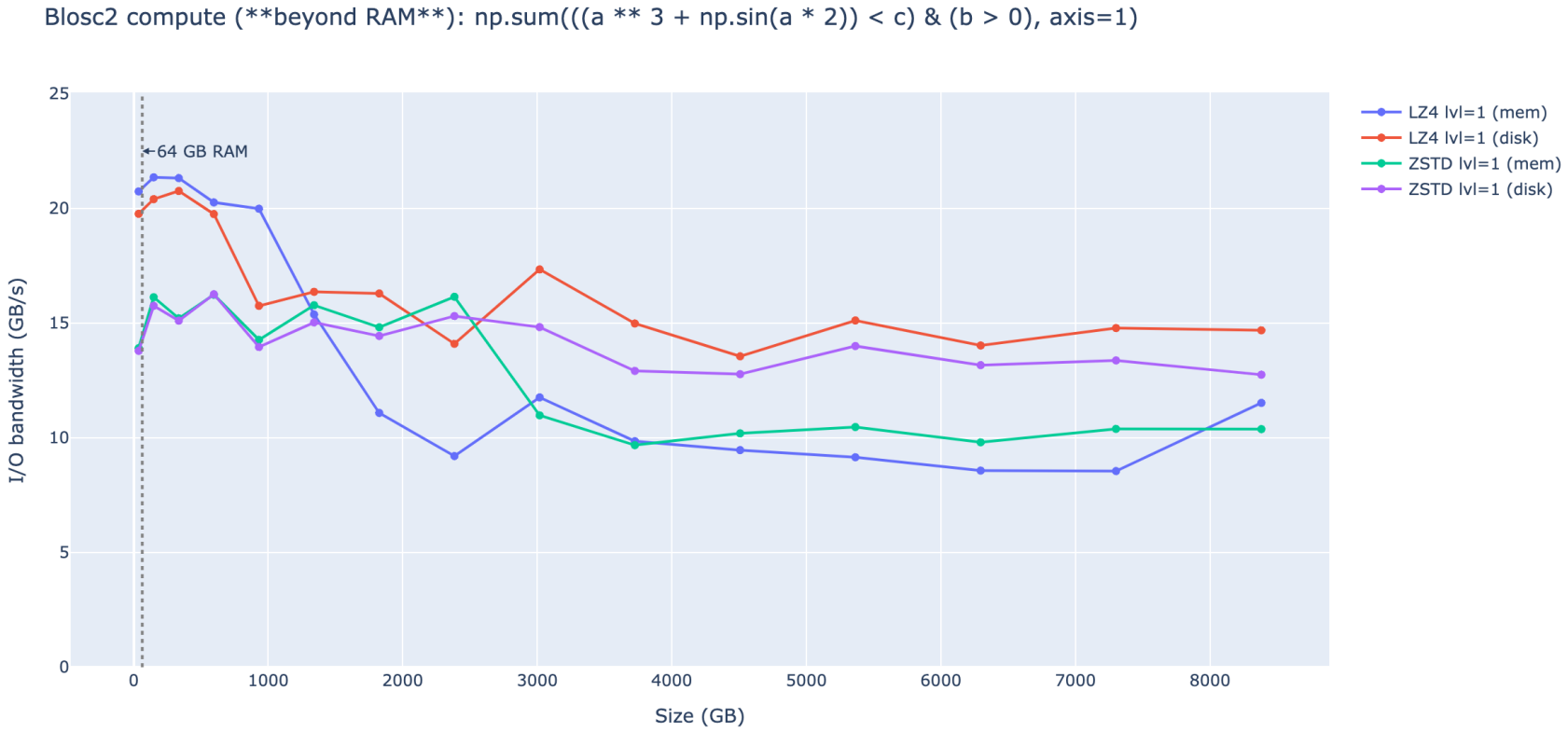
```
N = 1000_000
a = blosc2.linspace(0, 1, N * N, shape=(N, N))
b = blosc2.linspace(1, 2, N * N, shape=(N, N))
c = blosc2.linspace(0, 1, N, shape=(N,)) # 1D; broadcasting supported
# Blosc2 NDArrays override NumPy's universal functions (ufuncs)
out = np.sum(((a ** 3 + np.sin(a * 2)) < c) & (b > 0), axis=1)
```

https://www.blosc.org/python-blosc2/getting_started/overview.html#computing-with-ndarrays

Efficient Compressed Computing



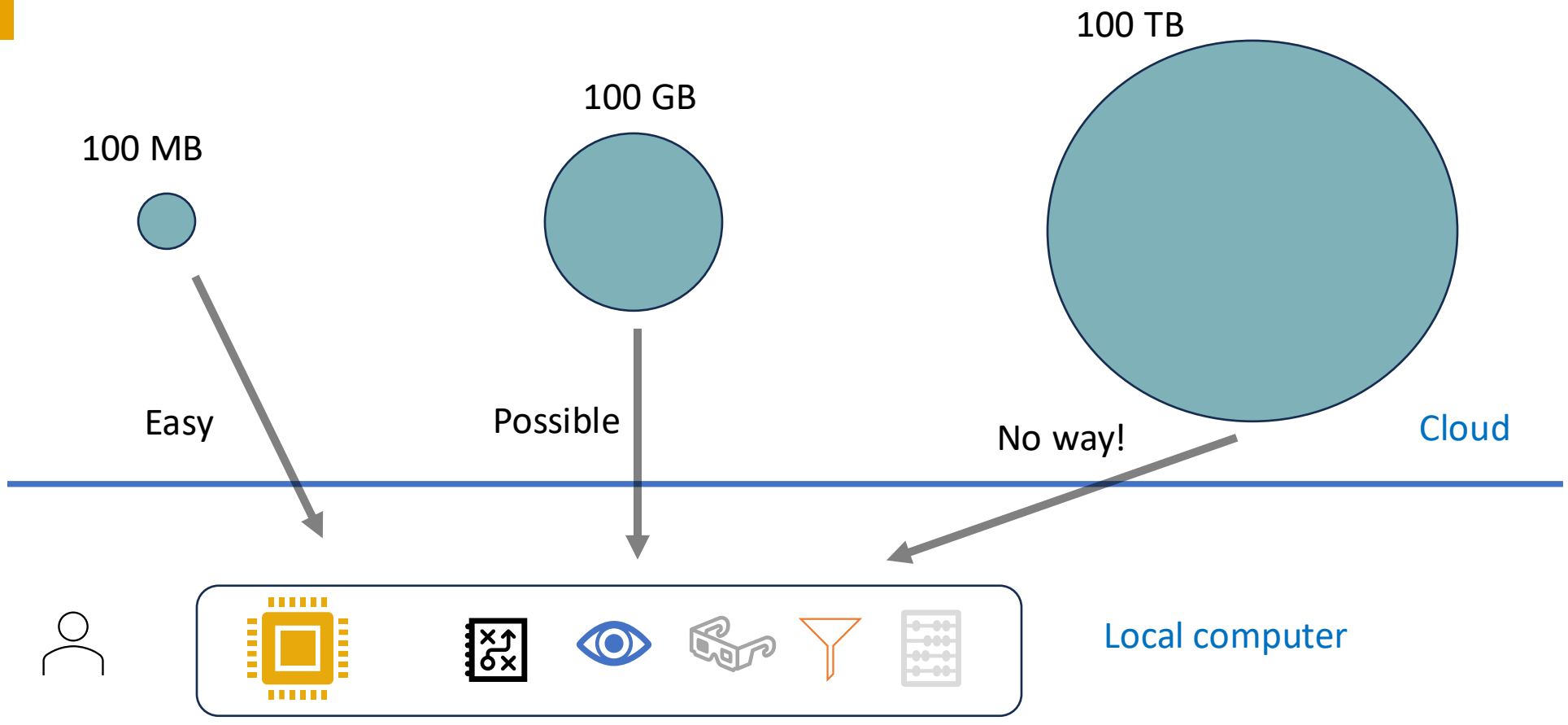
Going Bigger: Computing Beyond RAM



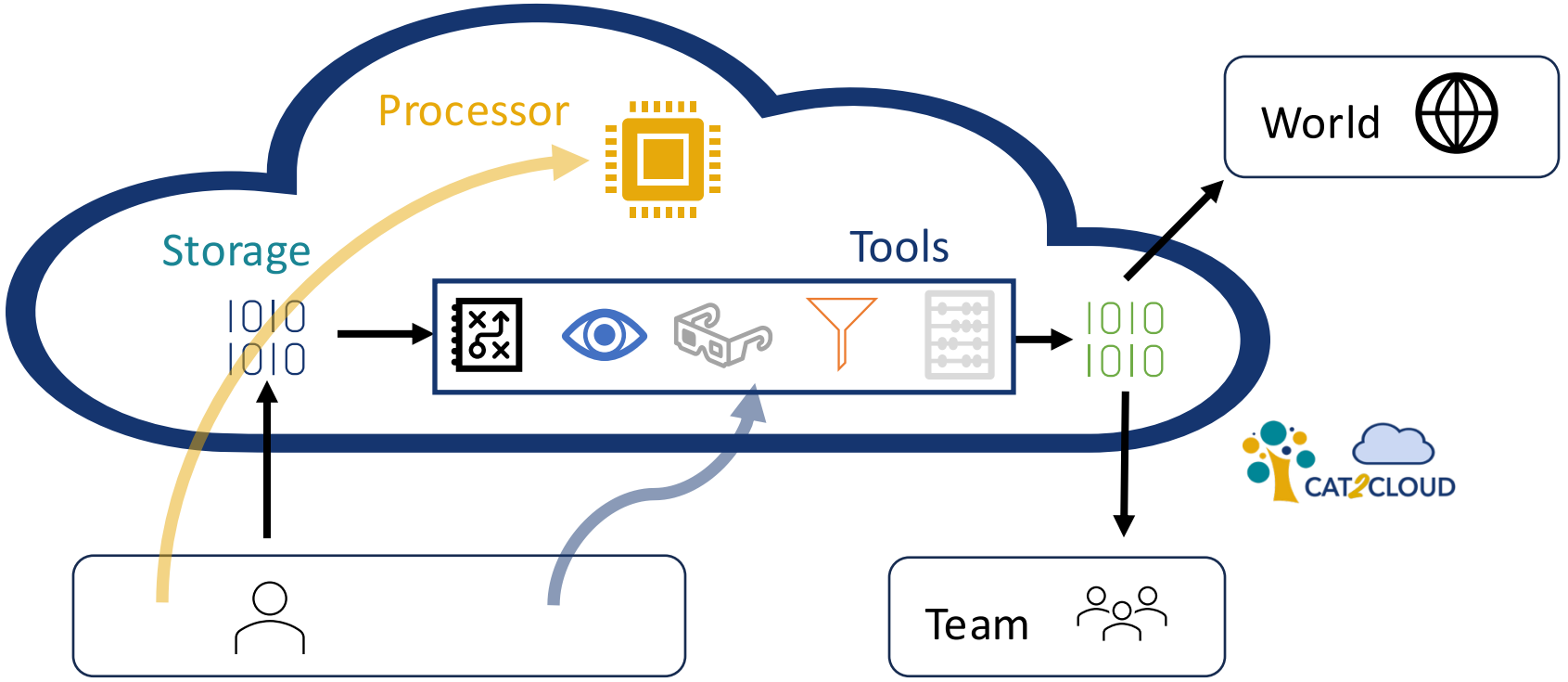
https://ironarray.io/blog/compute-bigger

Bring Computation Closer To Where Data Is Stored

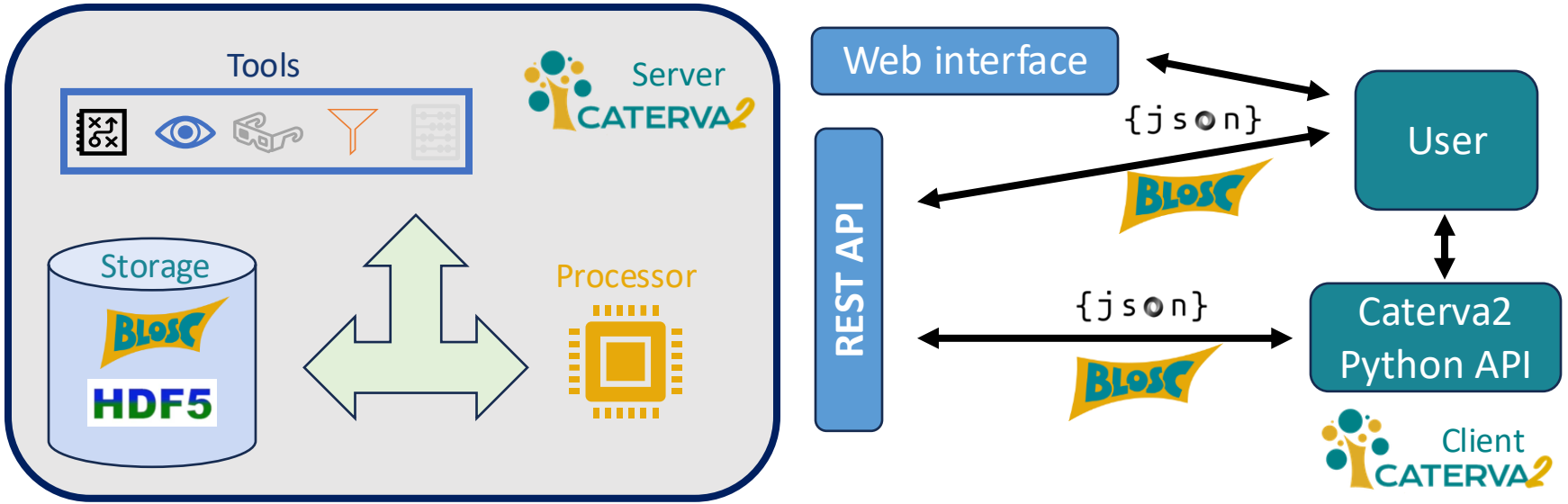
Data Is Affected By Physical Laws!



Computation Needs To Be Closer To Were Data Is Stored



Serving data through Internet



https://ironarray.io/caterva2

https://github.com/ironArray/Caterva2

Time for hands-on