Streamlining Scientific Workflows:

Computational Storage Strategies for HPC

Dominic Manno
Los Alamos National Laboratory
HPC-Design

SNIA COMPUTE, MEMORY, AND STORAGE SUMMIT

Solutions, Architectures, and Community
VIRTUAL EVENT, MAY 21-22, 2024

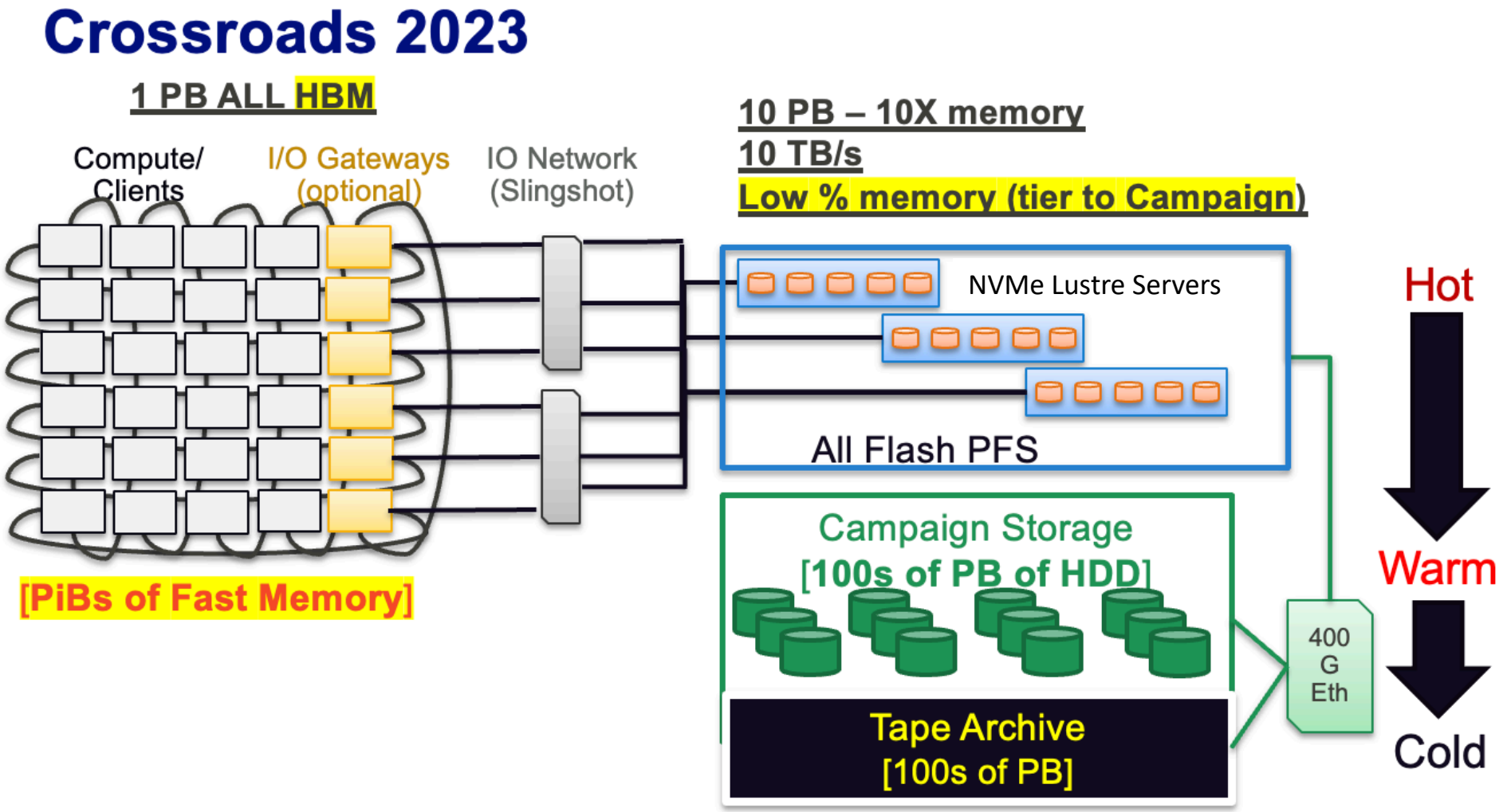
Scientific Workflows and HPC Storage

- **Weapons Science Workflow**
  - Dismantle Weapon, find issue
  - Experiment to learn physical properties of materials (accelerators, gas guns, hydro tests, lasers, etc.)
  - CAD drawings and experimental set up as input
  - Physics simulation that obeys material properties at extreme conditions
- **Computation Workflow**
  - Run, Checkpoint, run … analyze/viz maybe backup, smooth, run, checkpoint, run…. Analyze
  - Plan for refurb/dismantle/etc.
  - Prepare for report out to President/Congress

Checkpoint, what is that?

Exploration 1: Data Management

HPC Storage



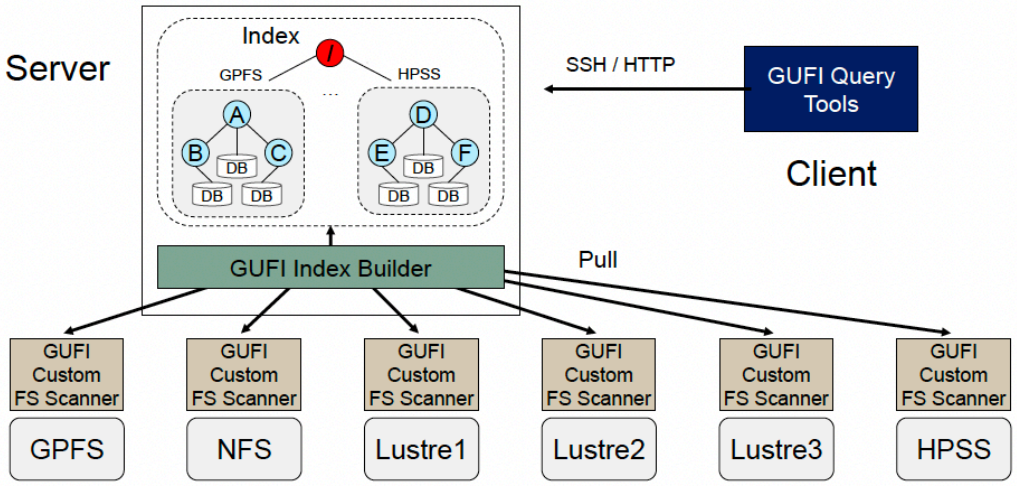
User Workflow

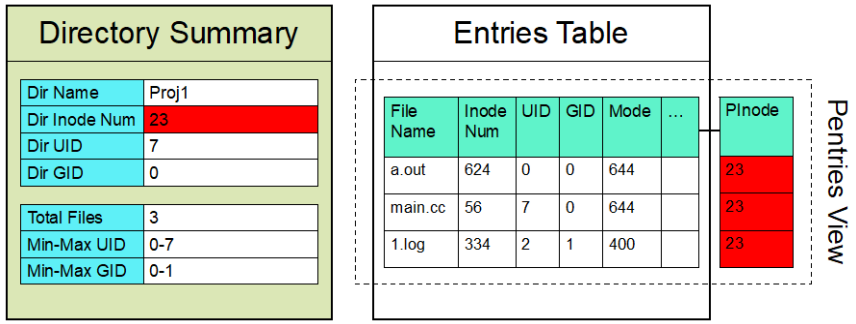
- Determine required input data
  - Input decks
    - (initial state, tabular constants, mesh configuration, …)
    - Re-start information (checkpoint)
  - Data location (scratch, campaign, archive)
  - **Find, stat, user driven data organization**
- Run sim, checkpoint, migrate, run, chkpt, …
- Analyze, smooth/modify, run sim, viz, archive
- Requires users to *know* their data and org
  - Metadata search slow, limited to each FS capabilities
  - Data movement parallel, slightly complex, separate cluster

Filesystem	Type	Dirs	Files
/users	NFS	6.1M	43M
/proj	NFS	35.7M	263M
/scratch1	Lustre	7.4M	102M
/scratch2	Lustre	16.5M	225M
/archive	HPSS	5.7M	193M

Key Technology

- Metadata indexing – Grand Unified File Index (GUFI) (link paper)
- Data movement scheduling – CONDUIT
- User defined metadata – GUFI + Data Science Infrastructure (DSI) (link)
- Computational storage – (file system offloads)

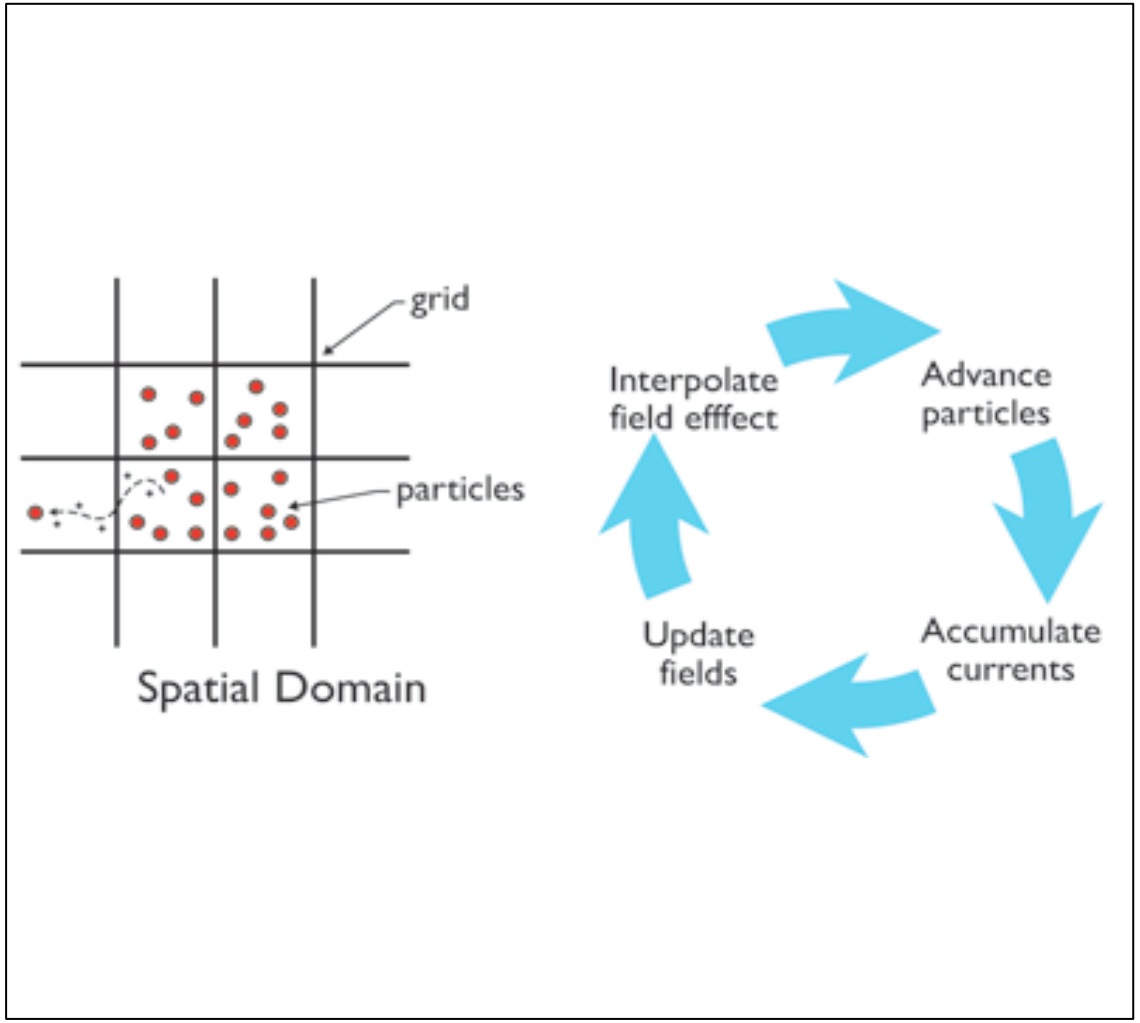




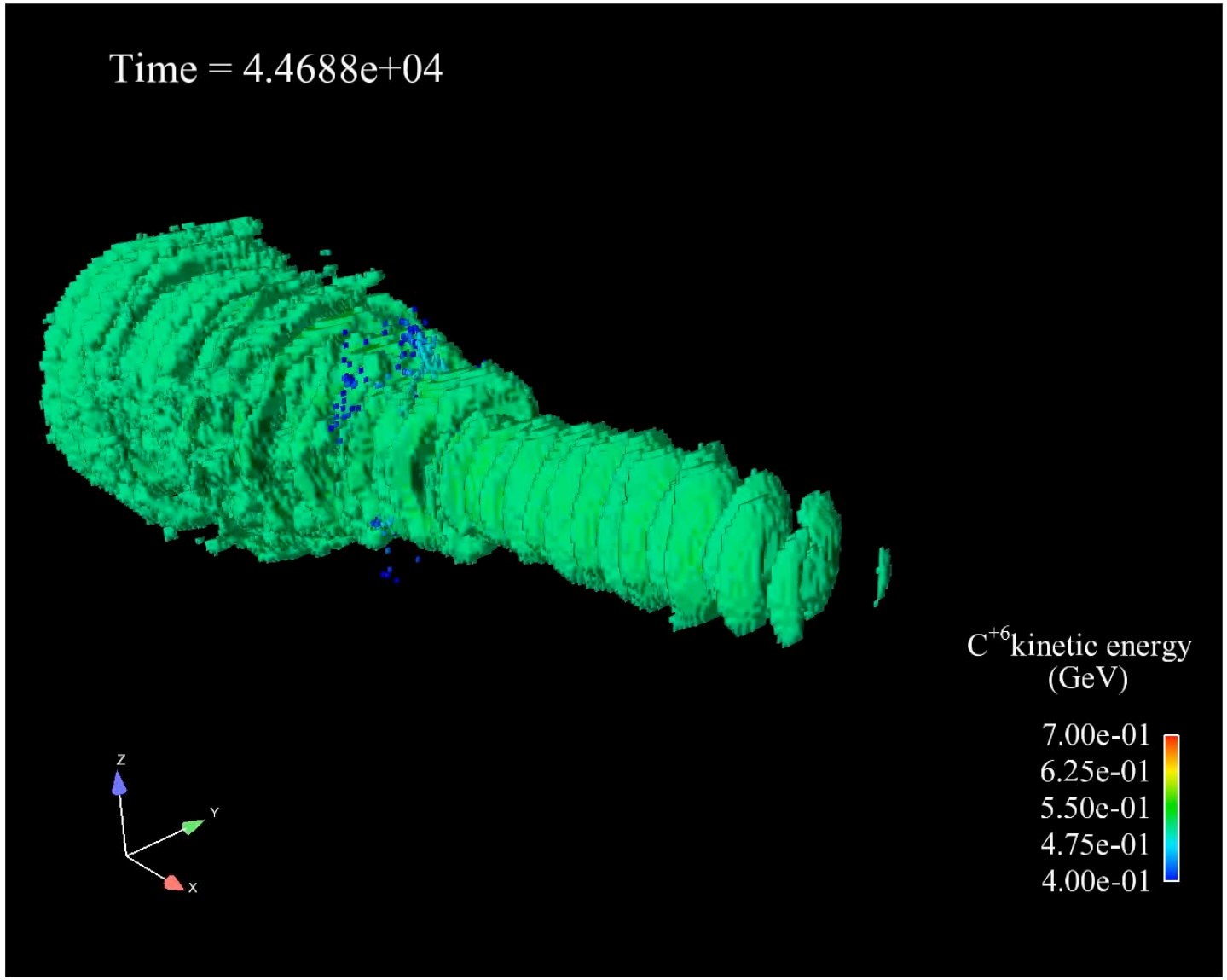
Exploration 2: Plasma Simulation

User Workflow

- I want to understand how magnetic reconnection occurs:
  - Need to sim plasma flowing over one another
  - Need to understand which particles are high and define high
  - Data is “un-ordered” by application
  - Can move the data, sort, viz – but we can do better
  - Often file per n-processes or particles – particle id: related values



Plasma Particle Sim

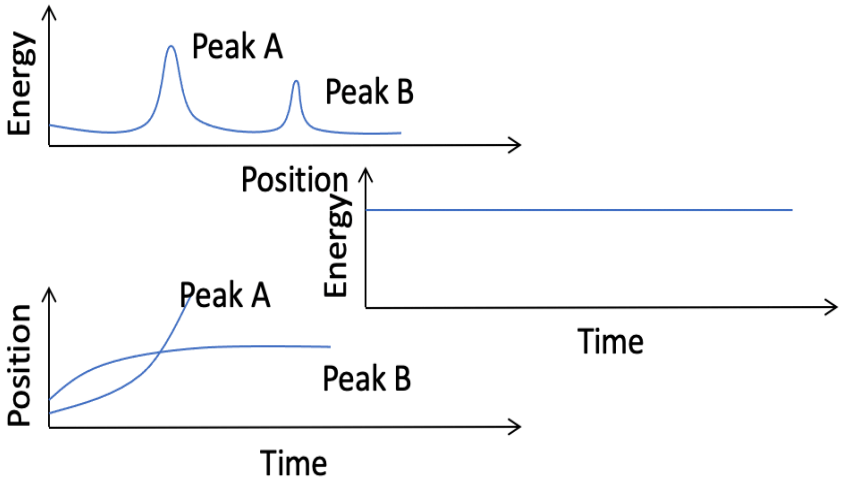


Key Technology

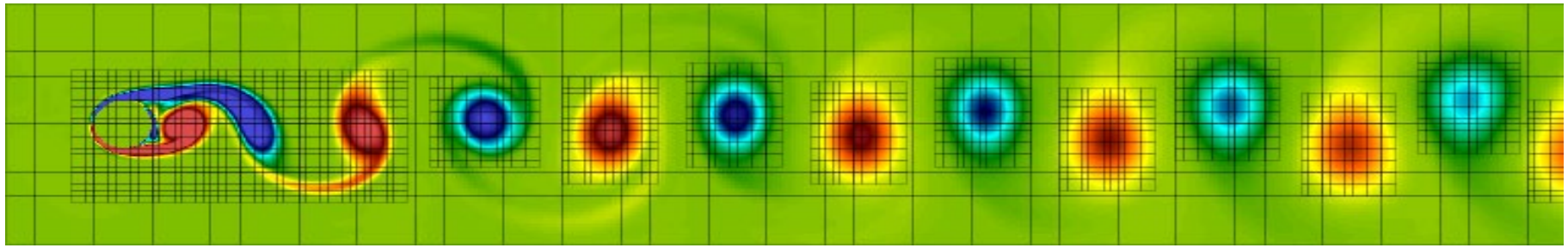
- Key-value API
- SSTables (levelDB, rocksDB)
- Secondary indexing to support range query
- Offloading function to device (partner: SK hynix)
  - Filtering, compaction
- Histogram generation (partner: Airmettle)
- **Significant speedups to end to end workflow!**

Exploration 3: Tracking a Shockwave

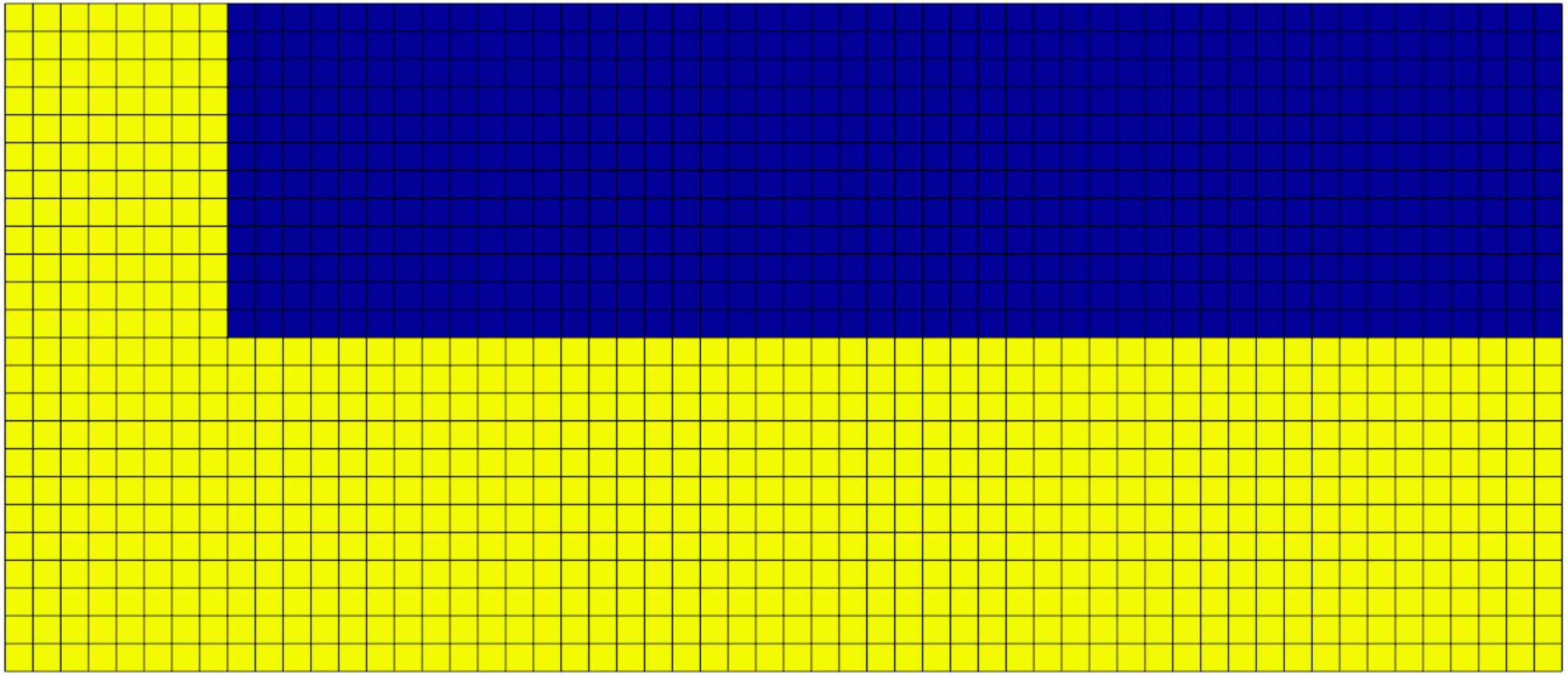
User Workflow



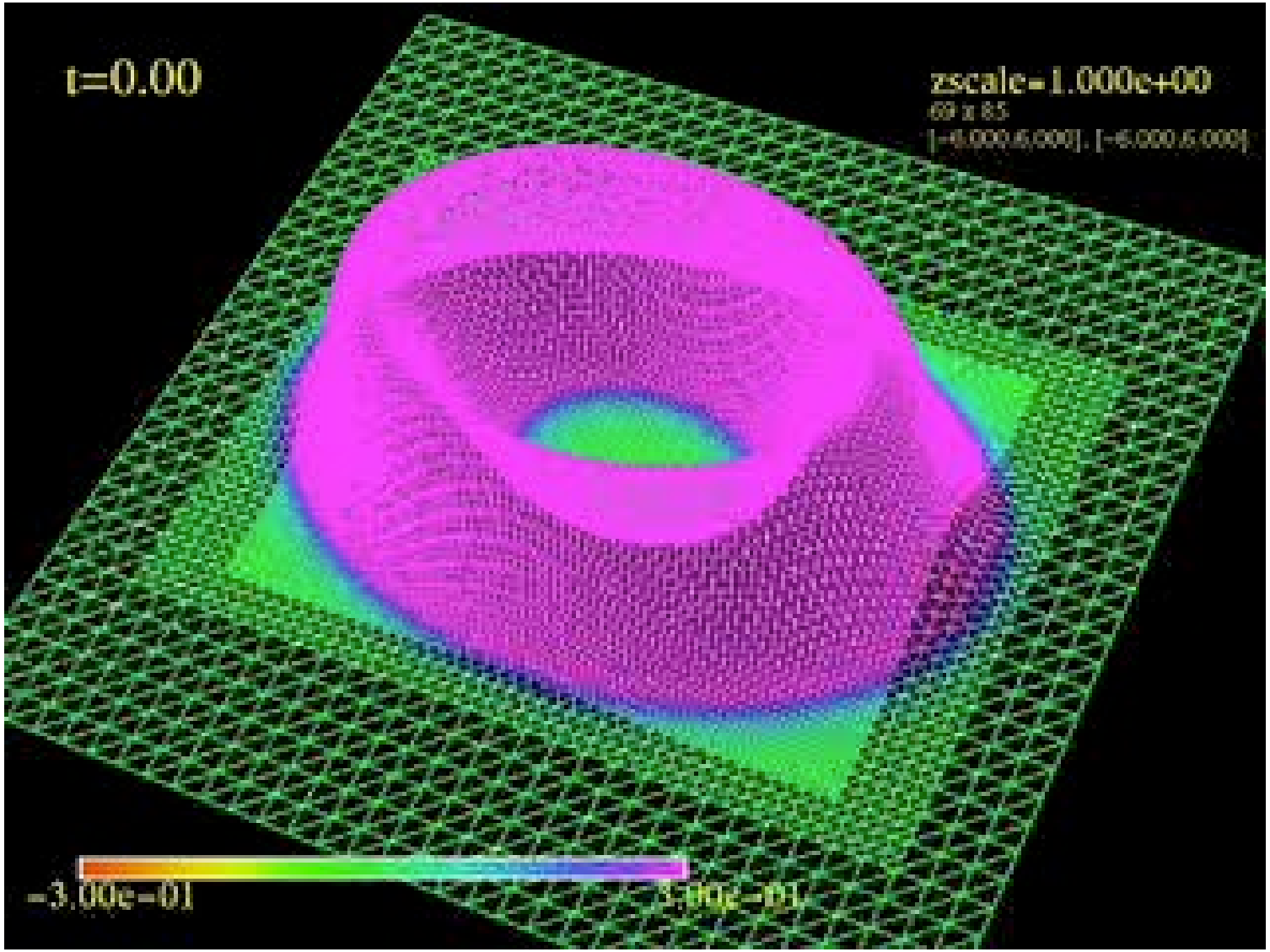
- Need to track shockwave edge
  - Determine prior simulations (2d, 3d)
  - Leverage data management tools
- Run simulation, checkpoint, run, analyze, smooth, …
- Understand how the shockwave moves through time
  - Where is energy high, what is high, what's happening near high energy
- More complex data organization – columnar or "arrays", often single file



Wave Tracking – 2D



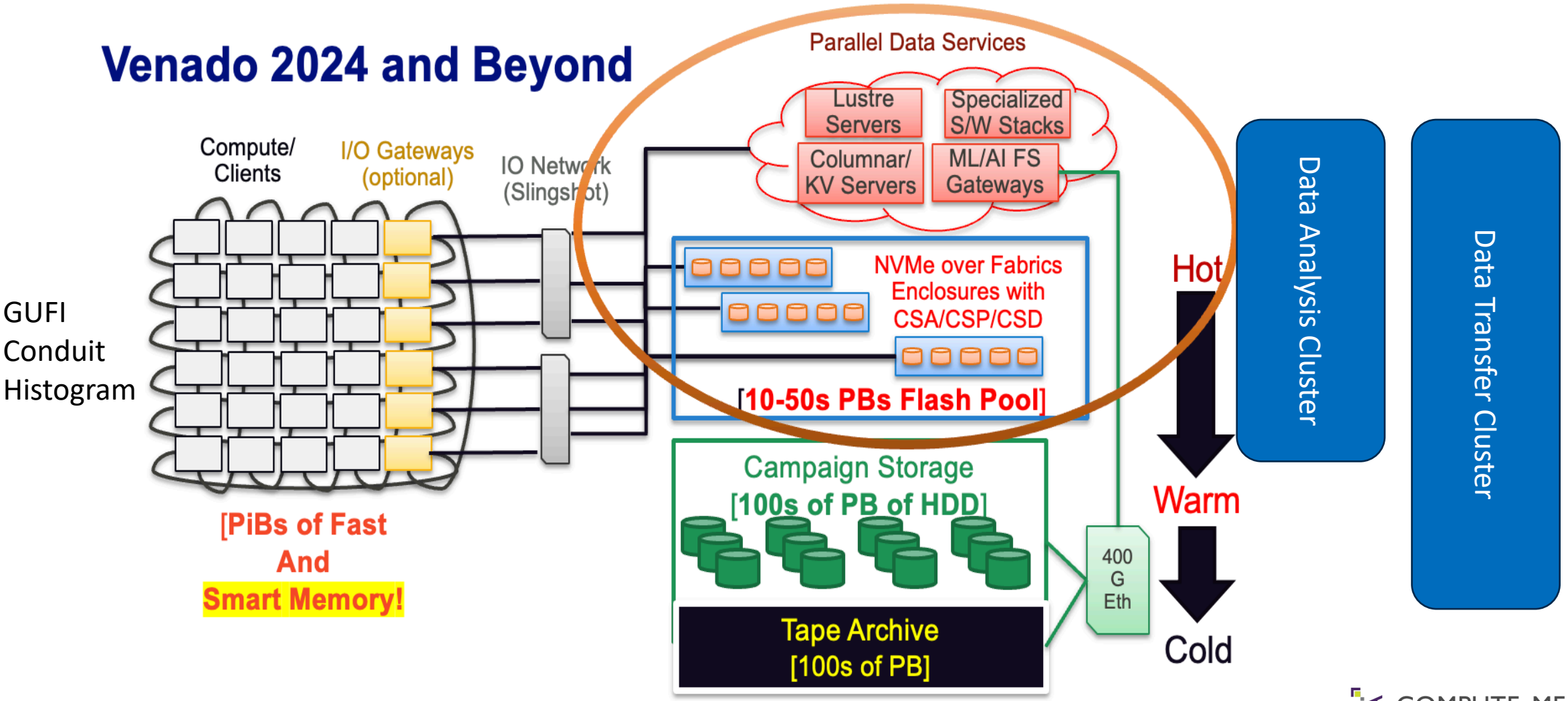
Wave Tracking – 3D



Key Technology

- Leverage open ecosystem and existing data analysis tools
- Columnar file-format with index capabilities (Parquet)
- Pushdown mechanics – substrait, computational storage
  - Filter near storage! Retrieve/transfer orders of magnitude less data!
- Alternative pushdown “medium” (block, file, object, kv)
- Histogram
- Require significantly fewer resources for analysis phase
- Middleware (histogram, viz, pushdown)

Putting it All Together



Please take a moment to rate this session.

Your feedback is important to us.