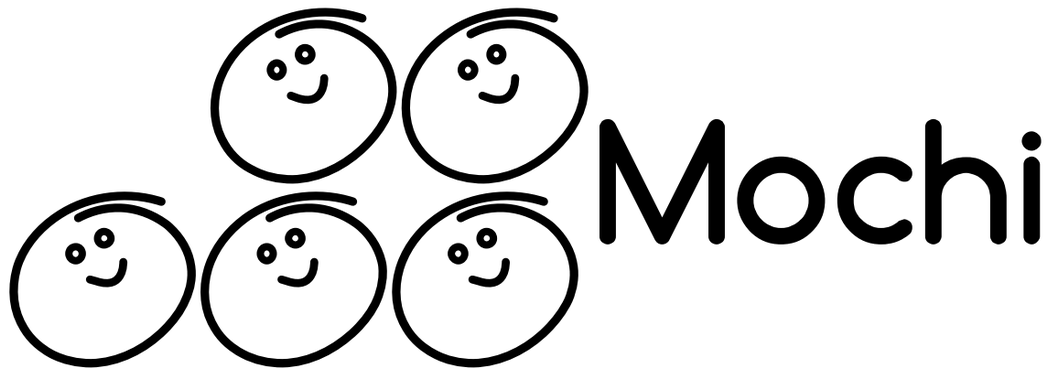
6TH JLESC WORKSHOP

Argonne NATIONAL LABORATORY

MOCHI: COMPOSABLE LIGHTWEIGHT DATA SERVICES FOR HPC

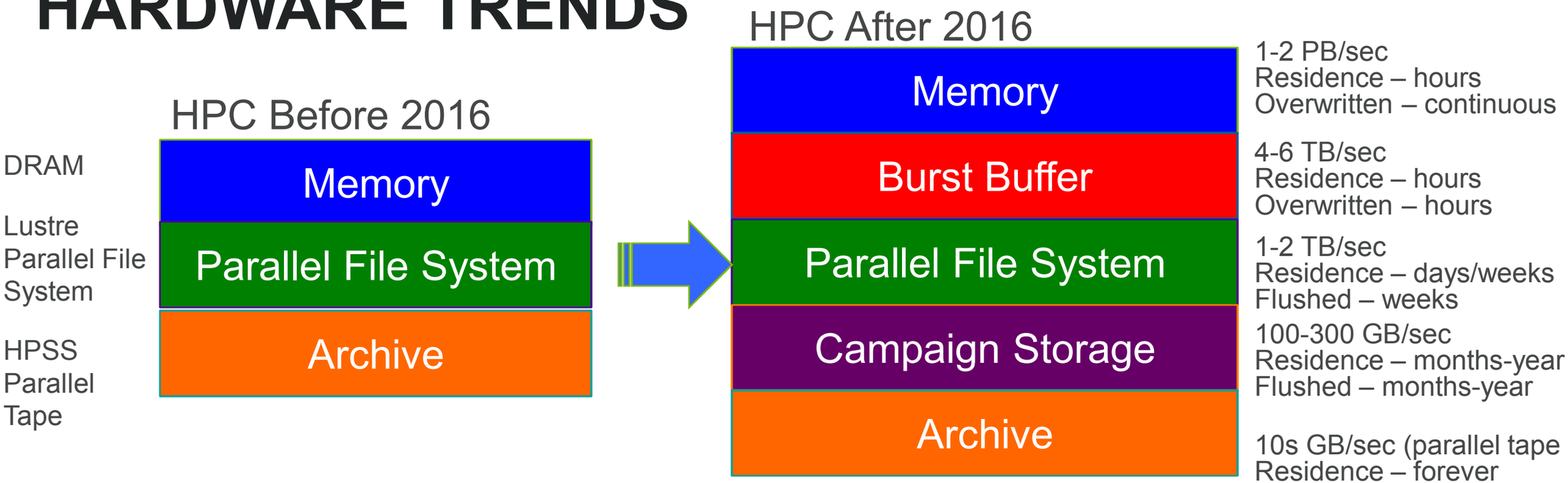


PHIL CARNS
Mathematics and
Computer Science Division
Argonne National Laboratory

November 30, 2016
Kobe, Japan

MOTIVATING TRENDS IN HPC DATA AND STORAGE

HARDWARE TRENDS



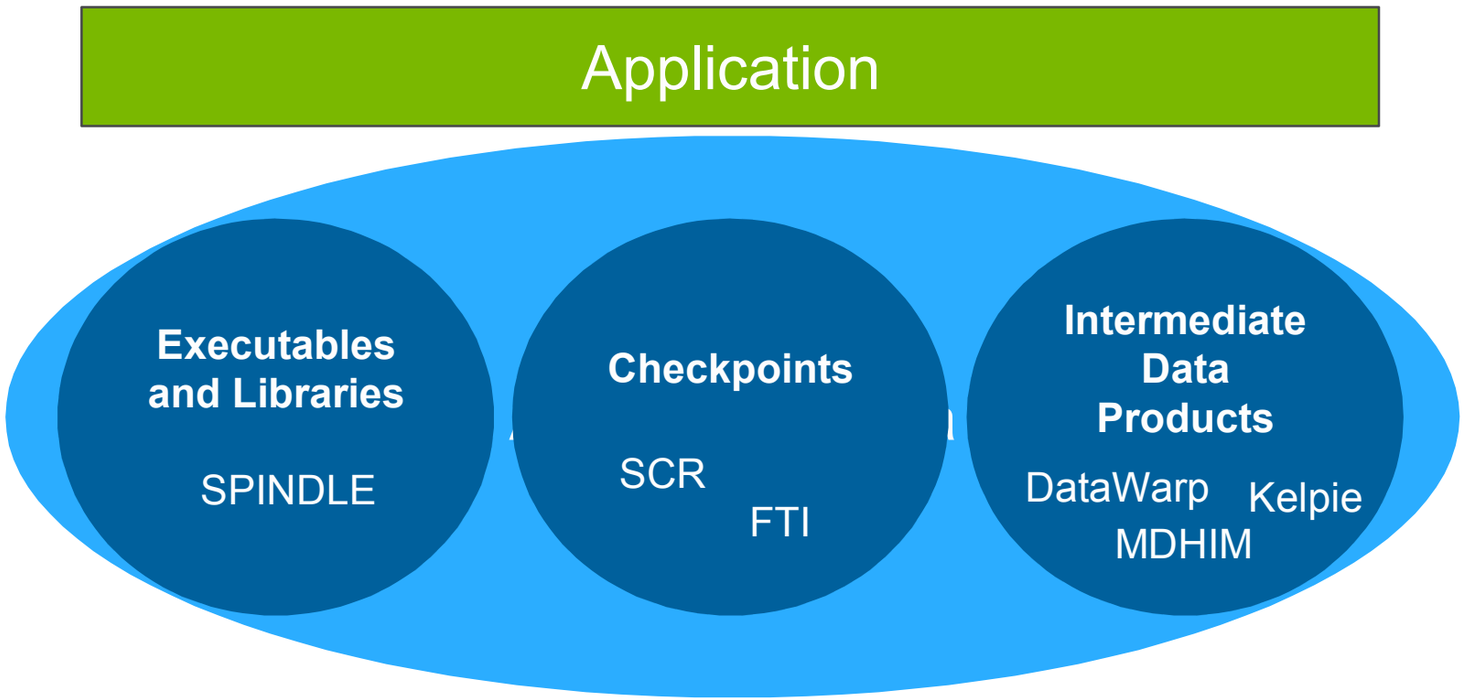
More layers are being added to the storage hierarchy

Tradeoffs in price, maturity, capacity, longevity, performance, and density

Slide adapted from
Gary Grider (LANL)

SOFTWARE TRENDS

Data services are specializing to provide application and domain-specific functionality



- Alternative to one-size-fits-all “parallel file system for data management” model
- Data services are provisioned on demand
- Co-designed with applications in some cases

WHAT'S NEXT?

Enabling the evolution of data-intensive scientific services

- Support emerging hardware (e.g., NVM), storage hierarchies, and deployment scenarios
- **Lower the barrier to entry for new specialized services**
  - Avoid building new services from the ground up
  - Provide a reusable toolkit of components and "microservices" for critical functionality
  - Compose to as needed for the task at hand

- Example microservices: key/value storage, group membership, replication
- Example composed scientific service: "Genomics Query Service"

ENABLING DATA SERVICES

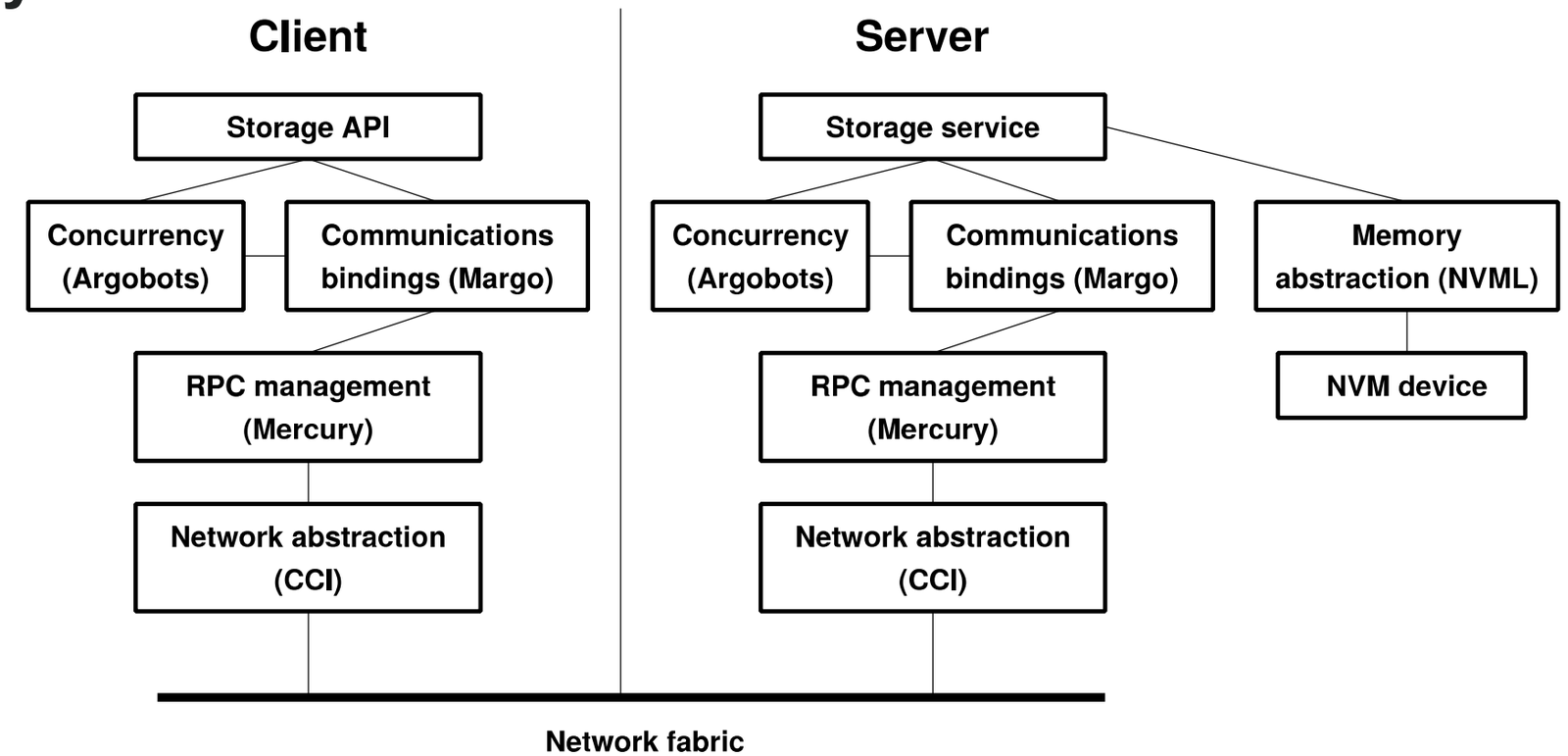
ROB ROSS, PHILIP CARNS, KEVIN HARMS, JOHN JENKINS, AND SHANE SNYDER	Argonne National Laboratory
GARTH GIBSON, CHUCK CRANOR, QING ZHENG, AND GEORGE AMVROSIADIS	Carnegie Mellon University
JEROME SOUMAGNE AND JOE LEE	The HDF Group
GALEN SHIPMAN AND BRAD SETTLEMYER	Los Alamos National Laboratory

Argonne NATIONAL LABORATORY

EXAMPLE DATA SERVICE

Early proof of concept: lightweight object storage service

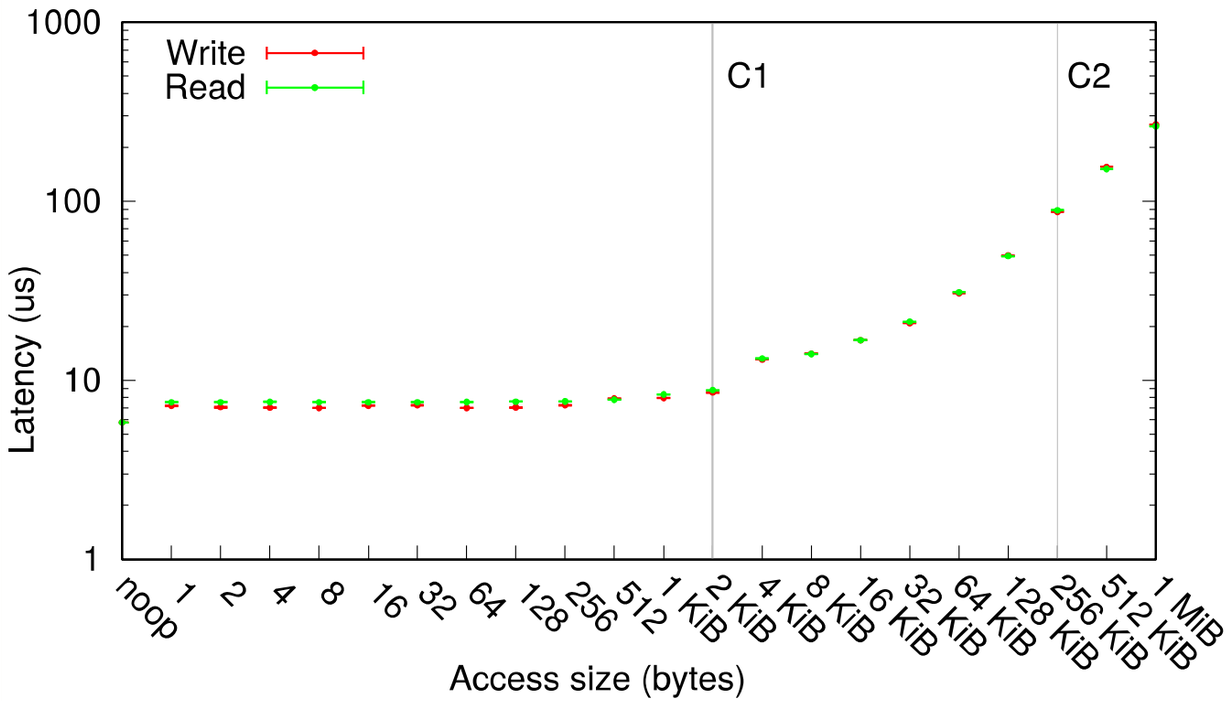
- RPC framework: **Mercury**
- Concurrency: **Argobots**
- Network layer: **CCI**
- Storage access: **NVML/libpmem**



ACCESS LATENCY

Evaluation on InfiniBand network with RAM storage

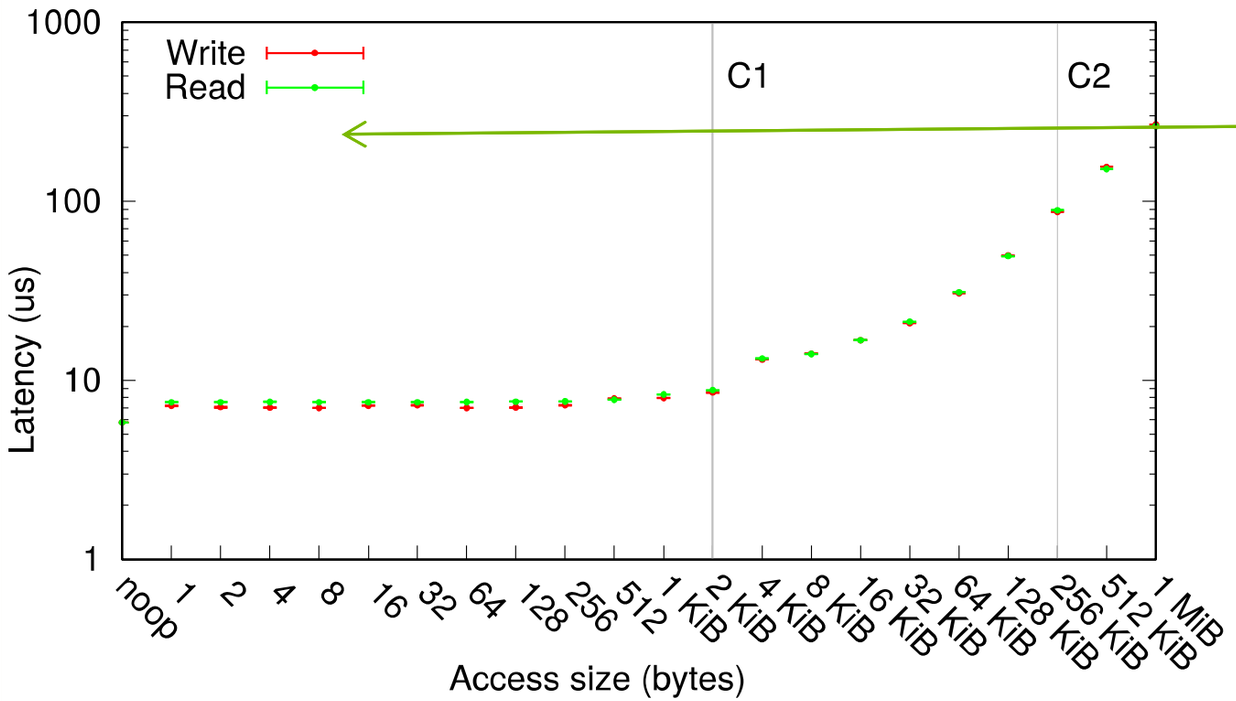
- Can we retain the performance characteristics of NVM across remote/distributed services?
- Must also be resource-friendly (no busy-polling on network, limited core usage)



ACCESS LATENCY

Evaluation on InfiniBand network with RAM storage

- Can we retain the performance characteristics of NVM across remote/distributed services?
- Must also be resource-friendly (no busy-polling on network, limited core usage)



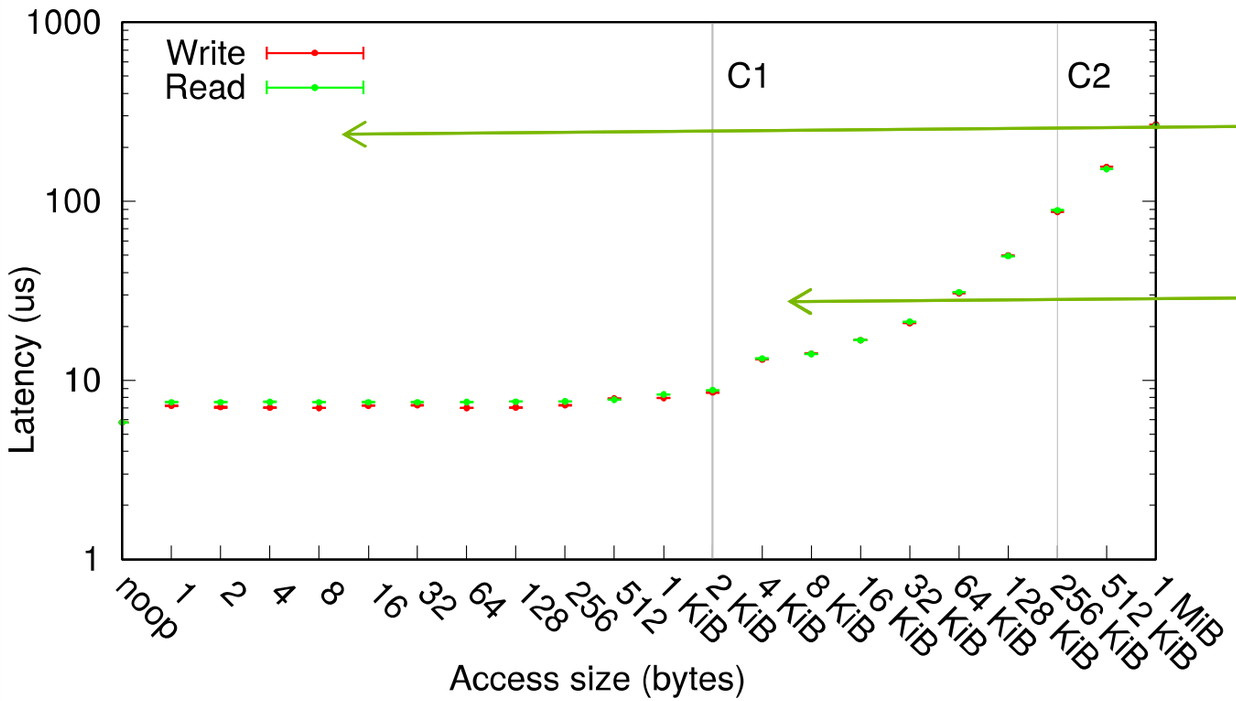
Protocol modes:

- Eager mode, data is packed into RPC msg

ACCESS LATENCY

Evaluation on InfiniBand network with RAM storage

- Can we retain the performance characteristics of NVM across remote/distributed services?
- Must also be resource-friendly (no busy-polling on network, limited core usage)



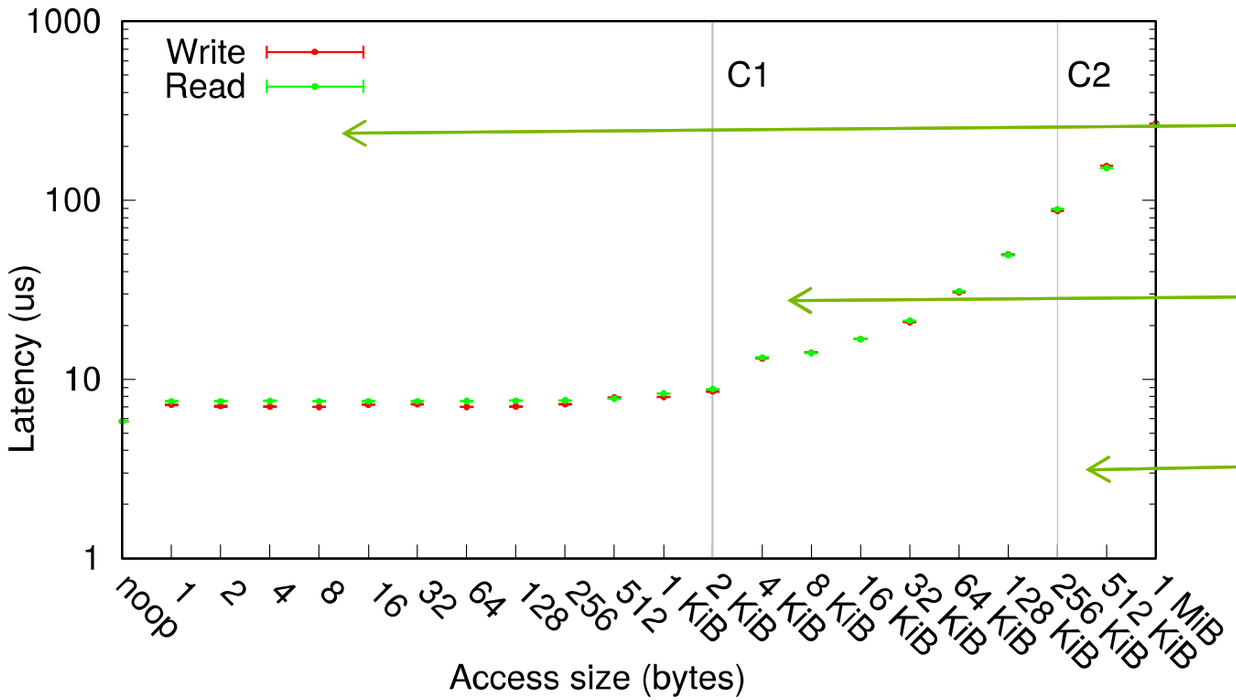
Protocol modes:

- Eager mode, data is packed into RPC msg
- Data is copied to/from pre-registered RDMA buffers

ACCESS LATENCY

Evaluation on InfiniBand network with RAM storage

- Can we retain the performance characteristics of NVM across remote/distributed services?
- Must also be resource-friendly (no busy-polling on network, limited core usage)



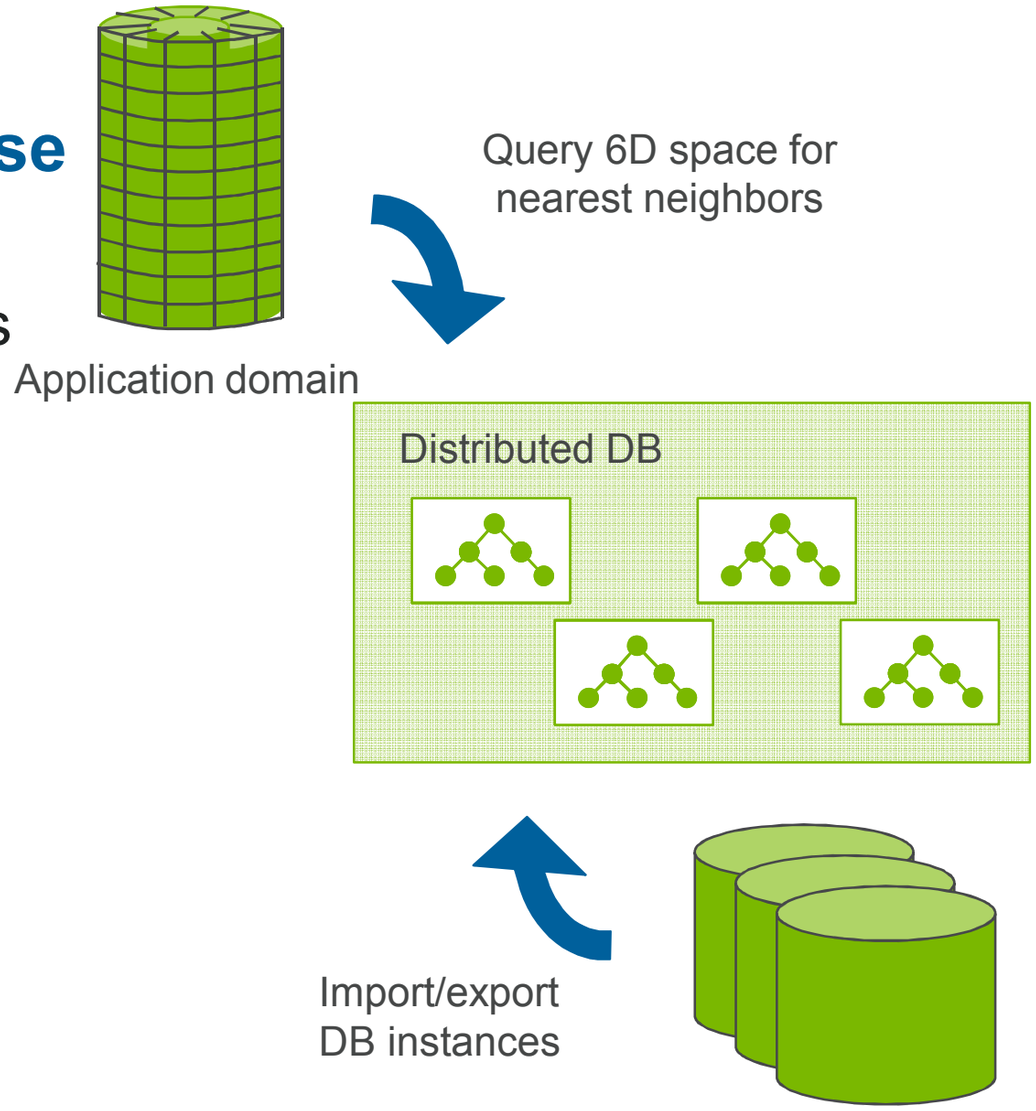
Protocol modes:

- Eager mode, data is packed into RPC msg
- Data is copied to/from pre-registered RDMA buffers
- RDMA “in place” by registering memory on demand

EXAMPLE DATA SERVICE

Co-designing a fine-scale model database

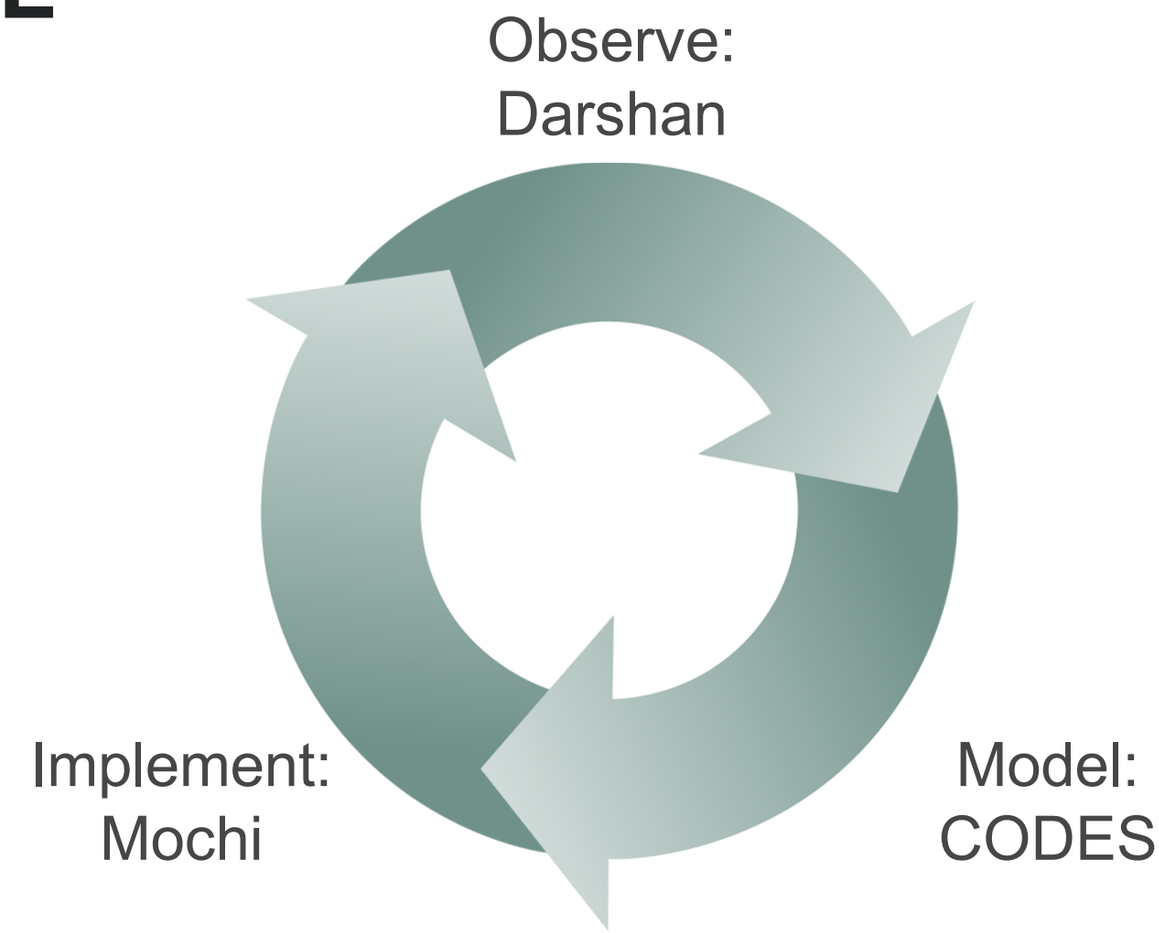
- Objective: Provide computation cache as a service to minimize fine-scale model executions
- Approach:
  - Start with a key/value store
  - Distributed approximate nearest-neighbor query for interpolation candidates
- Status:
  - Mercury-based in-memory DB service
  - Investigating distributed, incremental nearest-neighbor indexing
  - Applying technique to multiple applications



DATA RESEARCH AT ANL

Mochi is just one component

- Mochi:
  - Specialized data services
  - http://www.mcs.anl.gov/research/projects/mochi/
- Darshan:
  - Large scale I/O characterization
  - http://www.mcs.anl.gov/research/projects/darshan/
- CODES:
  - Discrete event simulation of HPC storage and networks
  - http://www.mcs.anl.gov/research/projects/codes/



THANK YOU!

THIS WORK WAS SUPPORTED BY THE U.S. DEPARTMENT OF ENERGY, OFFICE OF SCIENCE, ADVANCED SCIENTIFIC COMPUTING RESEARCH, UNDER CONTRACT DE-AC02-06CH11357.

WE ARE HIRING! THE MATHEMATICS AND COMPUTER SCIENCE DIVISION OF ARGONNE NATIONAL LABORATORY IS SEEKING OUTSTANDING PEOPLE AT MULTIPLE CAREER LEVELS.

HTTP://BIT.LY/2G8MOPV
(COMPUTER SCIENCE RESEARCH)

HTTP://BIT.LY/2FSPJW5
(SOFTWARE DEVELOPMENT)