The Scalable Data Management, Analysis, and Visualization Institute   http://sdav-scidac.org

In-Memory Data Management for Coupled Simulation Workflows using Staging-as-a-Service with DataSpaces and ADIOS

Tong Jin, Fan Zhang, Qian Sun, Melissa Romaus, Hoang Bui, Manish Parashar, **Rutgers University**
Scott Klasky , Qing (Gary) Liu, Norbert Podhorszki, **Oak Ridge National Lab**

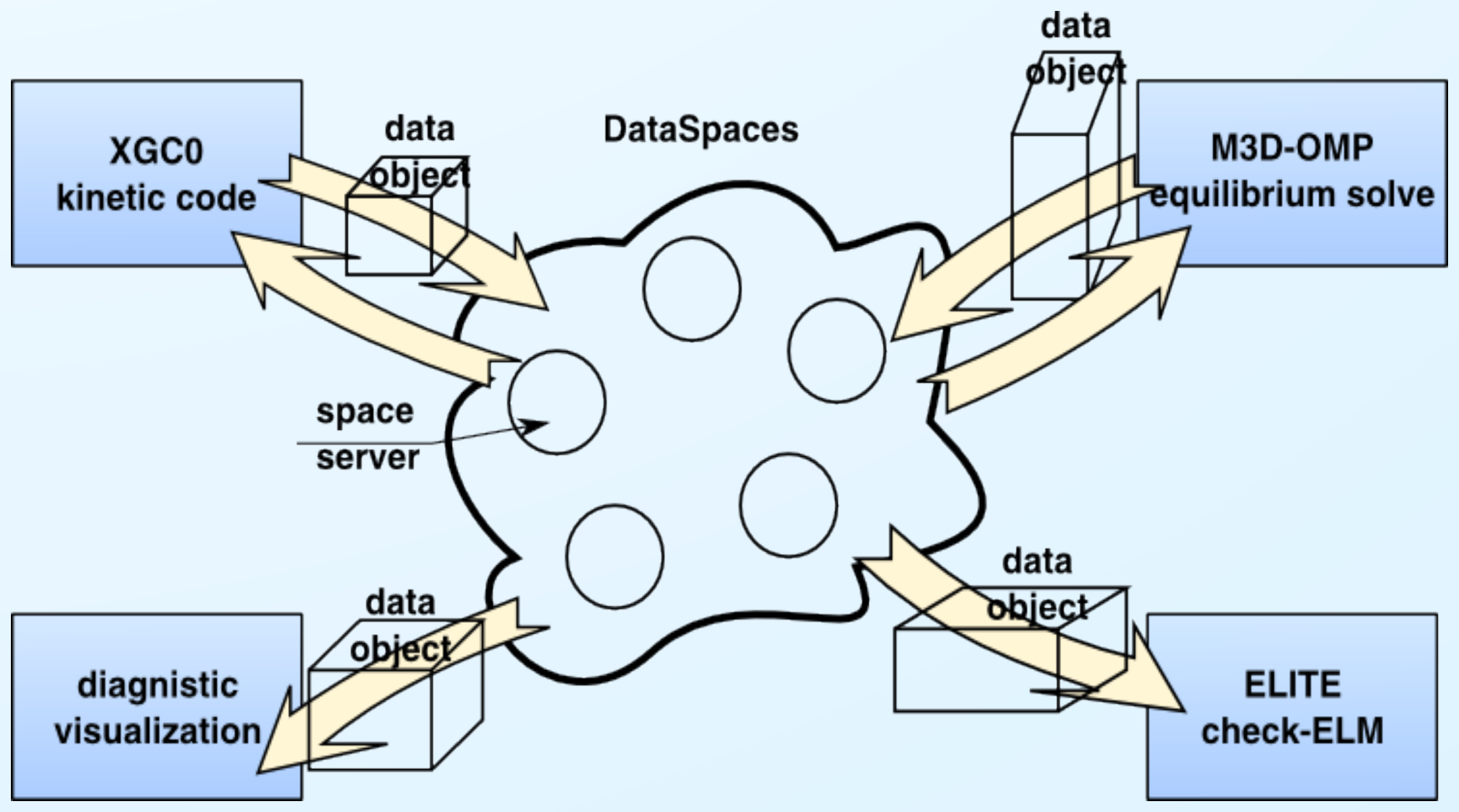
Motivation

Emerging high-end computing systems are enabling data-intensive coupled simulation workflows involving many interacting services. A challenge is enabling the efficient and scalable execution of these **in-situ** workflows, and managing the orchestration and data exchange required.

Overview of DataSpaces

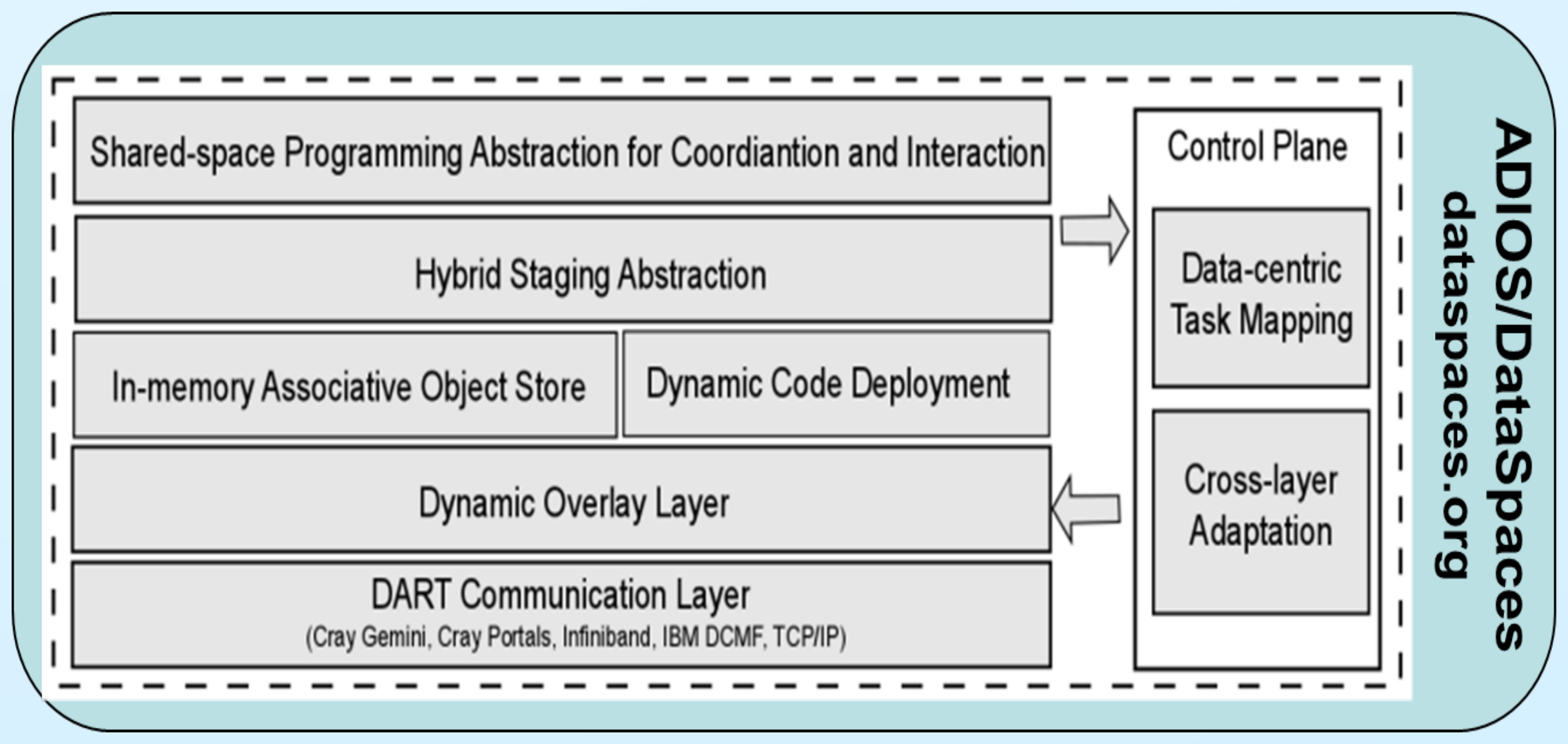
Introduction

- **DataSpaces** is a programming system designed to support in-situ workflows on current and emerging high-end systems. It efficiently and scalably enables dynamic online interaction, coordination and data exchange patterns between coupled applications and services.



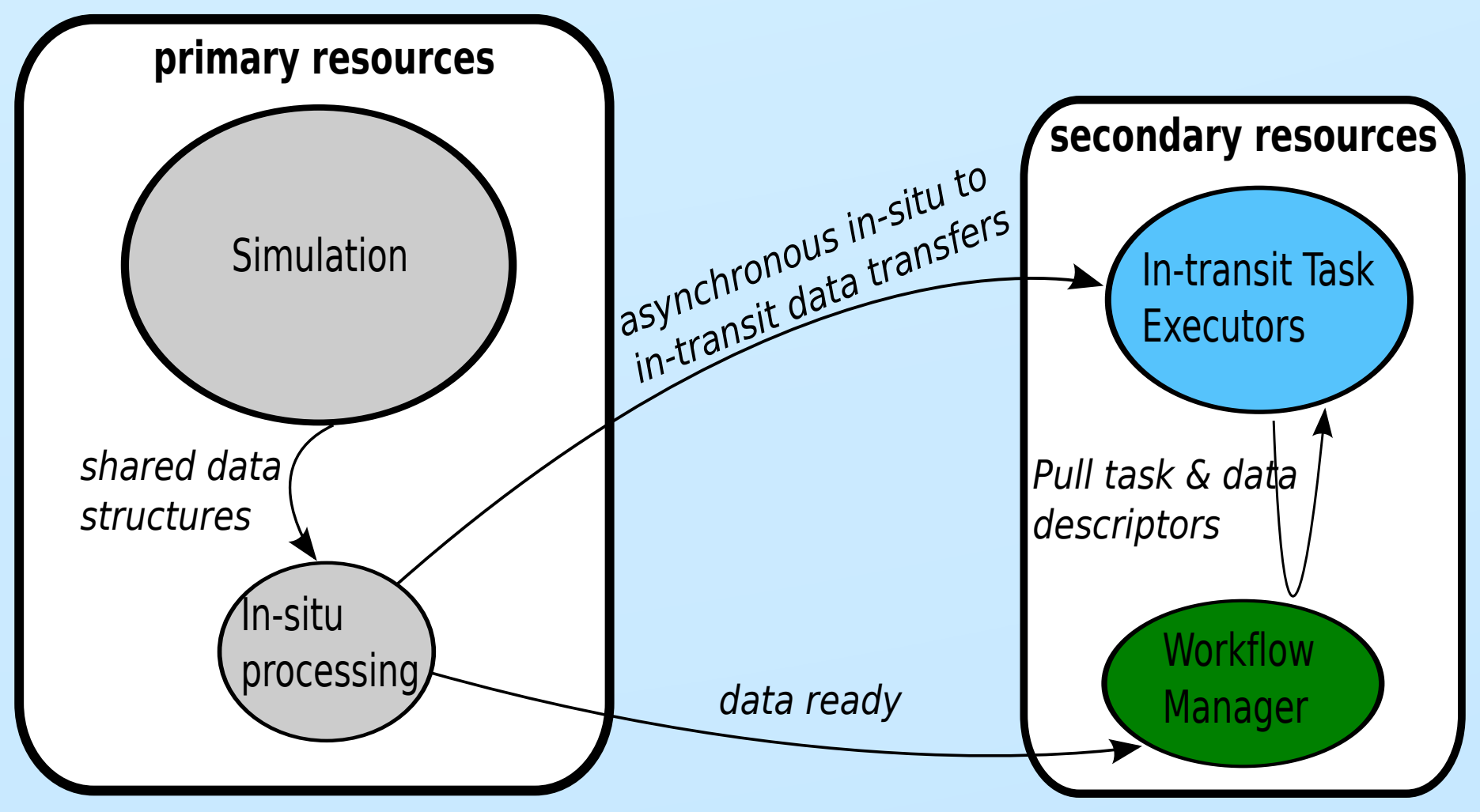
DataSpaces provides the abstraction of virtual semantically specialized shared space that can be associatively and asynchronously accessed by the different simulations and services that are part of the in situ workflow. The services can use the space to coordinate their execution and to share data.

Architecture and Features


Shared-space Programming Abstraction for Coordiantion and Interaction
Hybrid Staging Abstraction
In-memory Associative Object Store
Dynamic Code Deployment
Dynamic Overlay Layer
DART Communication Layer
(Cray Gemini, Cray Portals, Infiniband, IBM DCMF, TCP/IP)
Control Plane
Data-centric Task Mapping
Cross-layer Adaptation
ADIOS/DataSpaces
dataspaces.org

DataSpaces has a layered architecture, which includes a communication layer on top of the underlying network, an overlay layer, an object storage layer, a service layer, and the programming abstraction layer.

- Shared-space programming abstraction over hybrid staging
  - Simple API for coordination, interaction and messaging
  - Provides a global-view programming abstraction
  - Distributed, associative, in-deep-memory object store
  - Online data indexing, flexible querying
- Exposed as a persistent service
- Autonomic cross-layer runtime management
- High-throughput/low-latency asynchronous data transport



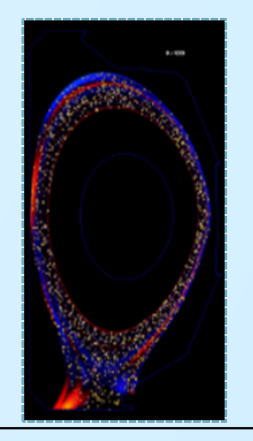
Overview of the in-situ/in-transit data analysis framework enabled by DataSpaces.

2013 R&D 100 Award Winner
ADIOS named one of top 100 tech products by R&D Magazine
ADIOS
easy-to-use, fast, scalable, and portable I/O

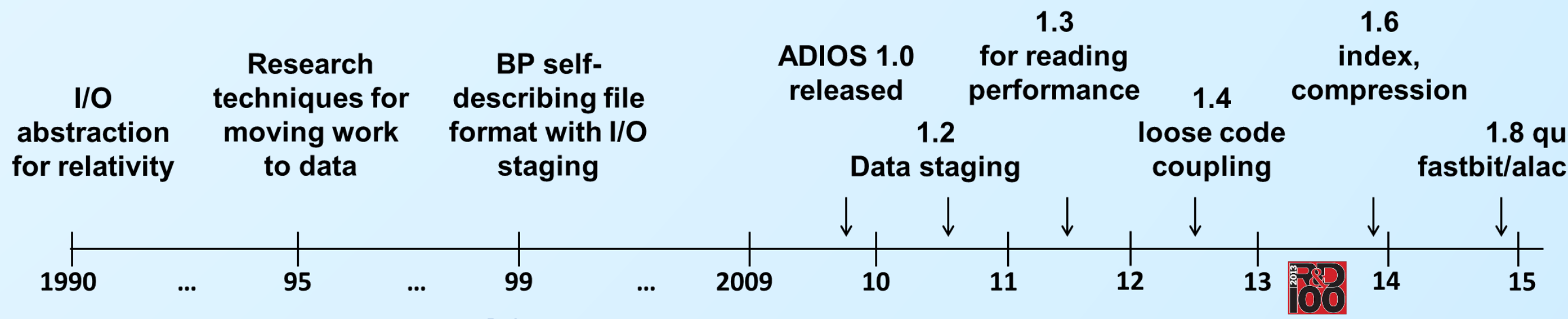
ADIOS Timeline

How did we come about this approach

- Problem
  - Before ADIOS, application writers had trouble achieving high-performance I/O for self-describing data
- Solution
  - Working with many leading DOE applications we developed a new framework for I/O with an API to abstract the implementation from the API
  - Burst Buffers, a simplified version of I/O-staging, becoming the de-facto standard for exascale I/O was created as part of the ADIOS framework
  - The first fully developed DOE I/O framework developed for sustainable I/O on LCFs
- Impact
  - Applications using ADIOS demonstrated input/output results more than 10 X faster than previous implementations
  - Now used by more than 30 LCF applications, totaling over 1B hours on the LCFs, ADIOS won a the R&D 100 Award in 2013

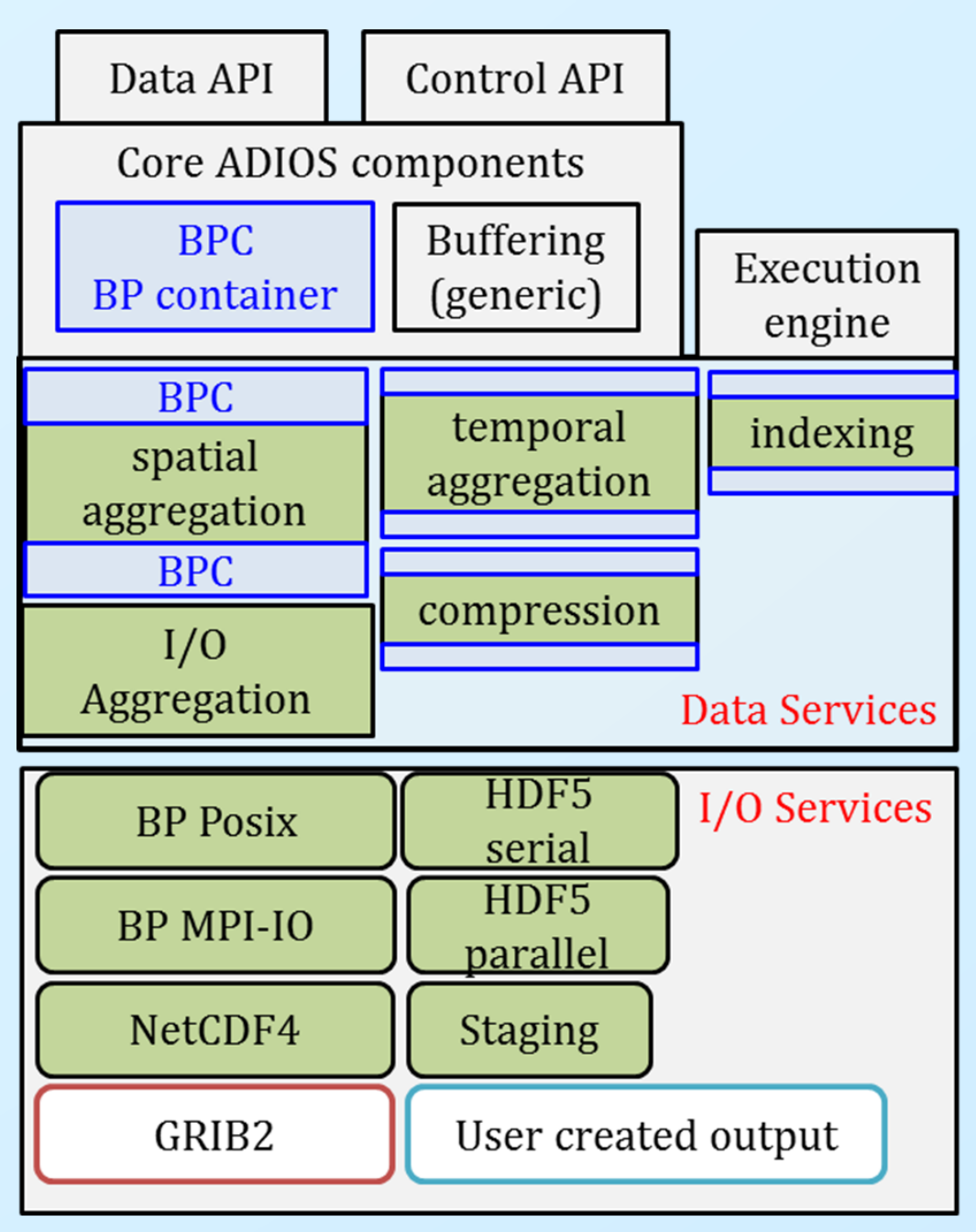


Using ADIOS, I/O for the XGC code went from 4,000 secs/hour using HDF5 to 252 secs/hour on Titan on 24K nodes


I/O abstraction for relativity
Research techniques for moving work to data
BP self-describing file format with I/O staging
ADIOS 1.0 released
1.2 Data staging
1.3 for reading performance
1.4 loose code coupling
1.6 index, compression
1.8 query fastbit/alacrity
1990
...
95
...
99
...
2009
10
11
12
13
14
15

I/O MiniApps

ADIOS framework



- An I/O abstraction framework: API is abstracted away from the method
- I/O componentization framework for Data-at-Rest and Data-in-Motion
- Provides portable, fast, scalable, easy-to-use, metadata rich output
- Change I/O method on-the-fly
- http://www.nccs.gov/user-support/center-projects/adios/
- Need to provide solutions for “90% of the applications”
- Q. Liu, J. Logan, Y. Tian, H. Abbasi, N. Podhorszki, J. Choi, S. Klasky, R. Tchoua, J. Lofstead, R. Oldfield, M. Parashar, N. Samatova, K. Schwan, A. Shoshani, M. Wolf, K. Wu, W. Yu, “Hello ADIOS: the challenges and lessons of developing leadership class I/O frameworks”, Concurrency and Computation: Practice and Experience, 2013

ADIOS Features 1.7-1.8 *new*

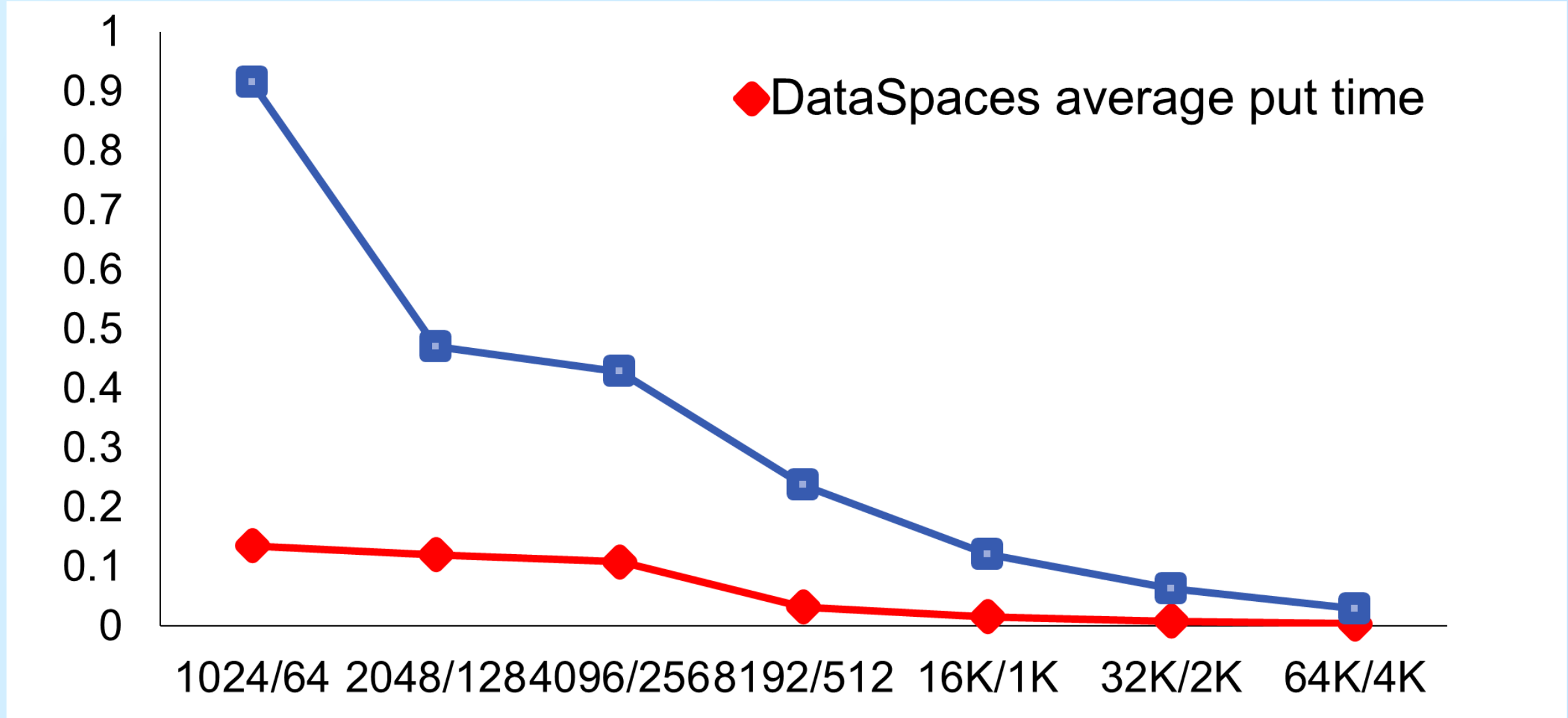
- Topology aware writing on Titan and BG/Q
- Dataspaces:
  64 bit, n-dimensions, BG/Q
  Can run as a service
- Improved Usability (cmake, etc.)
- skeldump, skel replay
- Indexing and Queries
- WAN Staging
- GRIB2 file format
- Automatic profiling saved with data

ADIOS roadmap

- User experience
  - Better buffering
  - "Best" I/O method that works good enough
  - C++ interface
- Usability/functionality
  - Encryption as transformation
  - Read method for hdf5 datasets
  - Improved PHDF5 output method
  - Support for “undef” value in datasets
- Parallelize metadata output
- Improve query read performance
- Staging
  - Make it more robust
  - Improve WAN staging
- Developer’s experience
  - Make ADIOS extensible to add new info
    - Characteristics, Attributes
    - Add/modify variables by methods
  - Links, internal/external
- Integration with VTK-M
  - Visualization plugins for data staging

Performance

- Strong scaling performance on Titan Cray XK7 at ORNL
- 1.16TB/s write throughput and 0.2TB/s read throughput

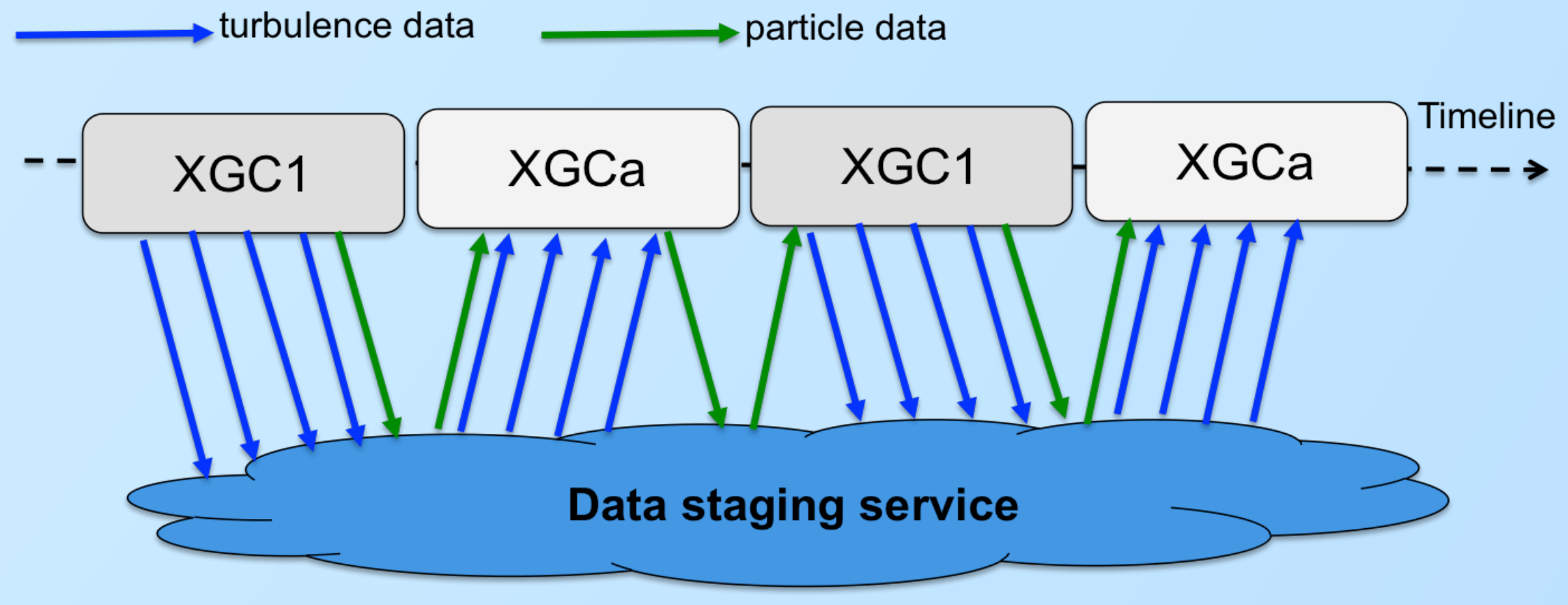


DataSpaces average put and get query time. X-axis shows the number of writer processes/reader processes in the testing workflow. Y-axis shows the query time.

Application 2: EPSI Coupled Fusion Workflow

Workflow Overview:

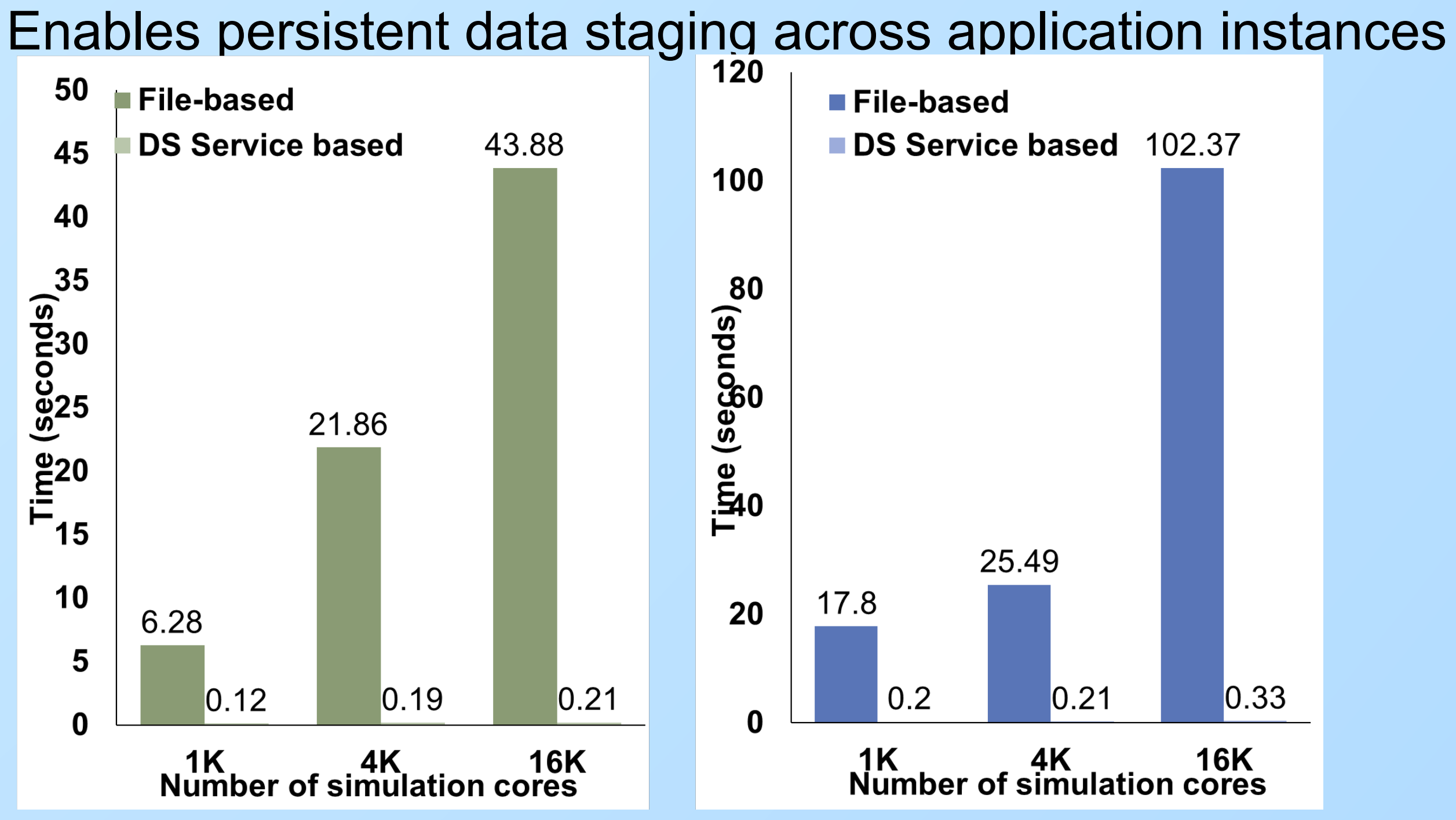
- Plasma fusion simulation workflow coupling XGC1 and XGCa
- One-way data exchange - particle data (large size & single iteration) and turbulence data (small size & multiple iterations)
- DataSpaces uses node-local shared memory segments as part of an in-memory hybrid data staging service



Execution sequence and data flow for the XGC1-XGCa coupled fusion simulation workflow using an in-memory hybrid staging service provided by DataSpaces.

Impact of using DataSpaces :

- Enables tightly coupled simulations at very large scales
- Results in significant performance improvement over traditional file-based approaches
- Enables persistent data staging across application instances



Particle data read: reduced by ~98% compared to file-based (left);
Turbulence data read time reduced by ~99% compared to file-based (right).

Summary

DataSpaces:

- Enables large scale in-situ coupled simulation workflows in different domains
- Provides in-memory hybrid staging as a persistent service
- Enables efficient data sharing and coordination with low overhead at very large scales

SDAV
Scalable Data Management, Analysis, and Visualization
U.S. DEPARTMENT OF ENERGY
Office of Science