# 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



# **ADIOS Timeline**

How did we come about this approach

Problem

Data API

BPC

**BP** container

spatial

aggregation

I/0

Aggregation

**BP** Posix

**BP MPI-IO** 

NetCDF4

GRIB2

Core ADIOS components

Control API

Buffering

(generic)

temporal

aggregation

compression

HDF5

serial

HDF5

parallel

Staging

User created output

engine

#### Performance

0.9

0.8

0.7

0.6

the XGC code went

secs/hour on Titan

secs/hour using

from 4,000

HDF5 to 252

on 24K nodes

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

DataSpaces average put time

**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

- 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 Using ADIOS, I/O for
- 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



### **ADIOS framework**

- An I/O abstraction framework: API is abstracted away from the method
- I/O componentization framework for Data-at-Rest and Execution Data-in-Motion
- Provides portable, fast, scalable, easy-to-use, metadata rich output



1024/64 2048/1284096/2568192/512 16K/1K 32K/2K 64K/4K

## **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

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. urbulence data particle data





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

Change I/O method on-the-fly http://www.nccs.gov/user-support/center-projects/adios/ Data Service: Need to provide solutions for "90% of the applications" /O Service • 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

Staging

Topology aware writing on Titan and BG/Q

- Dataspaces:
- 64 bit, n-dimensions, BG/Q
- Can run as a service

**ADIOS roadmap** 

User experience

- Improved Usability (cmake, etc.)
- skeldump, skel replay Indexing and Queries
  - WAN Staging
  - GRIB2 file format
  - Automatic profiling saved with data

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



#### • High-throughput/low-latency asynchronous data transport



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



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

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

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

