

# The Armada framework for parallel I/O on computational grids

Ron Oldfield and David Kotz  
Department of Computer Science  
Dartmouth College



© Copyright 2002 by the authors

January 2002

FAST 2002 WIP Presentation

# Introduction

**Computational Grids: geographically distributed networks of heterogeneous computer systems and devices.**

## Data-intensive grid applications

- Must use large remote datasets
- Often computationally intensive
- Datasets often need pre- and/or post-processing
- Examples
  - Climate modeling (EOS-DAS)
  - Astronomy (Digital Sky Surveys)
  - Comp. Biology (Computed MicroTomography)
  - Computational physics

## I/O system requirements

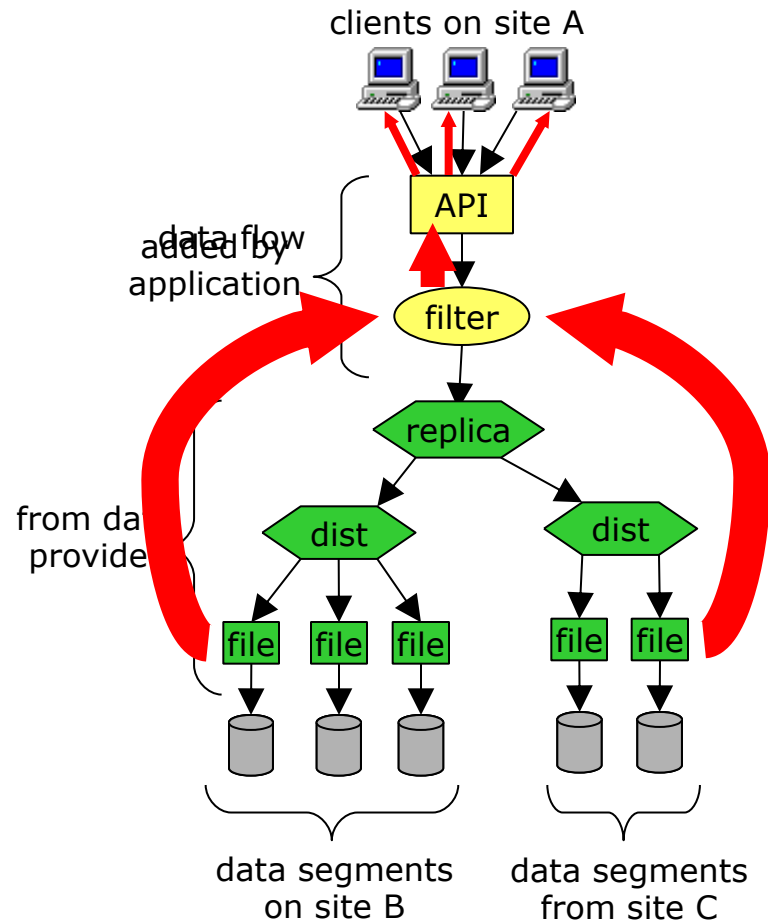
- Flexibility
  - Application control of the interface
  - Application control of system policies (caching, data-dist., ...)
- Performance
  - Parallel data transfers.
  - Remote execution of user code (e.g., filtering, transforms, compression, encryption)



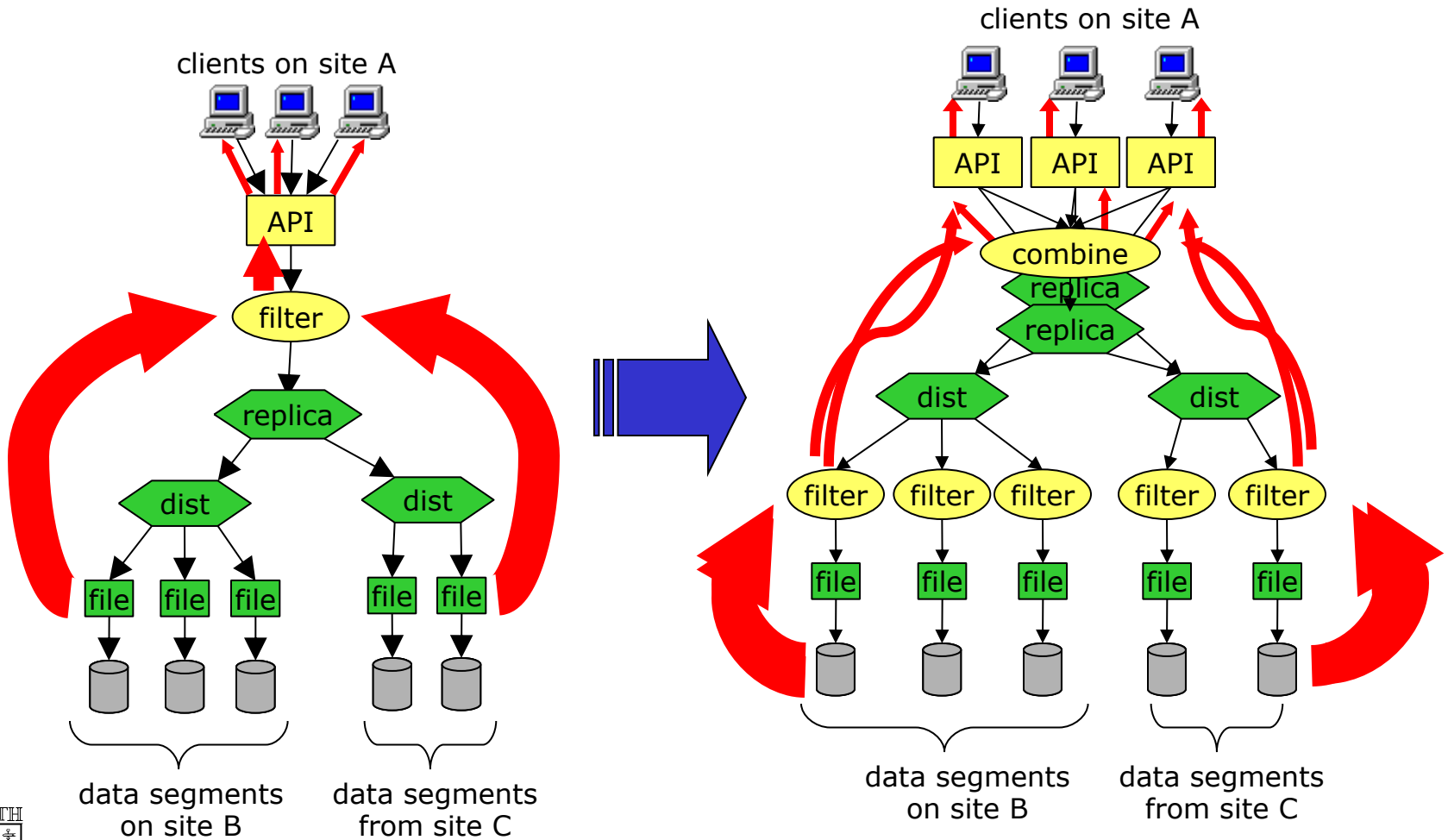
# Armada

## An I/O framework for data-intensive grid applications

- Flexible design, based on stackable file systems.
- Applications access data through a graph of “ships” called an “armada”.
- Requests travel toward data servers.
- Data is pushed toward clients for reads, pulled toward servers for writes.
- The armada abstracts details of the I/O system
  - Caching, filtering, data distribution



# Improving Performance



# In progress...

- Automate graph restructuring.
    - Formalize rules and algorithms
  - Develop placement algorithms.
    - Requires detailed information on ship requirements and available resources.
  - Performance monitoring and analysis.
- 

## Contact Information

Ron Oldfield (raoldfi@cs.dartmouth.edu)

David Kotz (dfk@cs.dartmouth.edu)

<http://www.cs.dartmouth.edu/~dfk/armada>

