Active Storage Networks

John A. Chandy
Department of Electrical and Computer Engineering

HECIWG File Systems and I/O Research and Development Planning Workshop
August 7, 2007

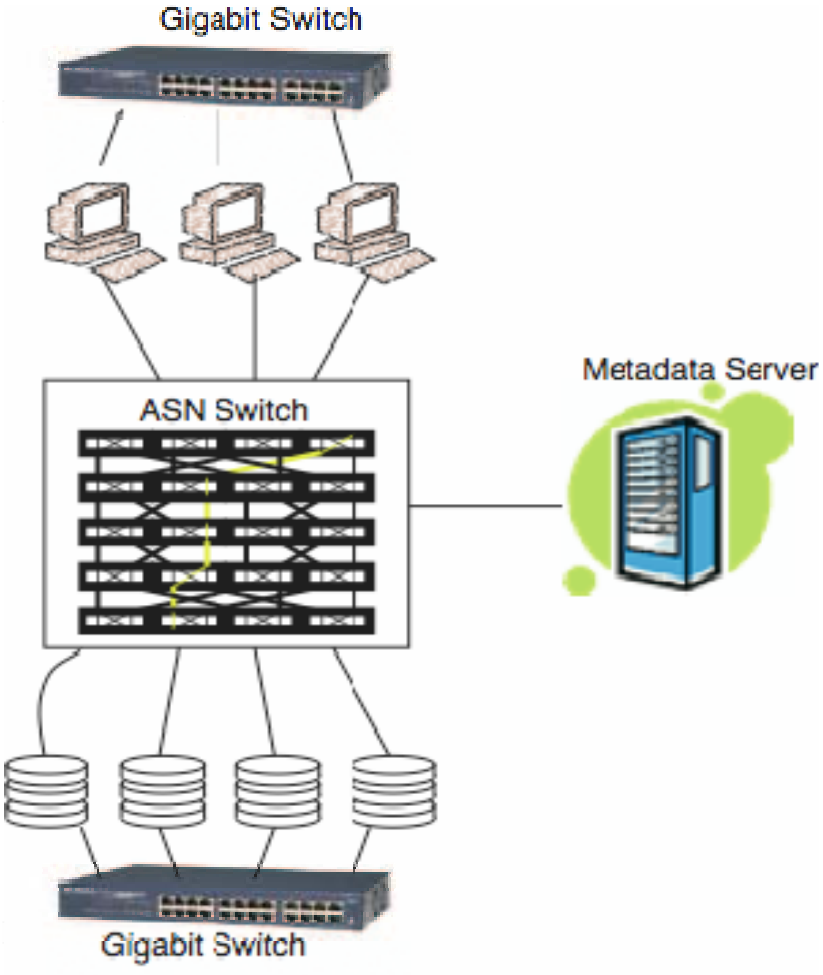
Active Storage Networks

- Can intelligence in a storage network
  - Accelerate computation?
  - Accelerate storage access?
- How do you build an active storage network?
  - Switch design
  - File system design

Motivation

- Active Disks
  - Intelligence at the disk can distribute computation to parallel disks
  - Process data in streams
  - Disks only have local view of data
- Active Storage Network
  - Network has a global view of data
  - Distributed caching of file system metadata and data
  - Redundancy optimizations

Active Storage Network



Active Storage Networks

- Application operations
  - Reduction operations
    - Database queries
      - SELECT … ORDER BY .. LIMIT k
      - m disks n items per disk
      - $O(nm)$ for normal disks
      - $O(n + km)$ for active disks
      - $O(n + k \log m)$ for active disks with ASN
    - Scientific applications
      - MPI_SUM, MPI_MAX, etc.

Active Storage Networks

- Application operations
  - Transformational operations
    - Sorting
      - $O(mn \log mn)$ for normal disks
      - $O(n \log n) + O(mn \log m)$ for active disks
      - $O(n \log n) + O(n \log m)$ for active disks with ASN
    - Scientific applications
      - Matrix transformations
    - Stream-based
      - Video editing

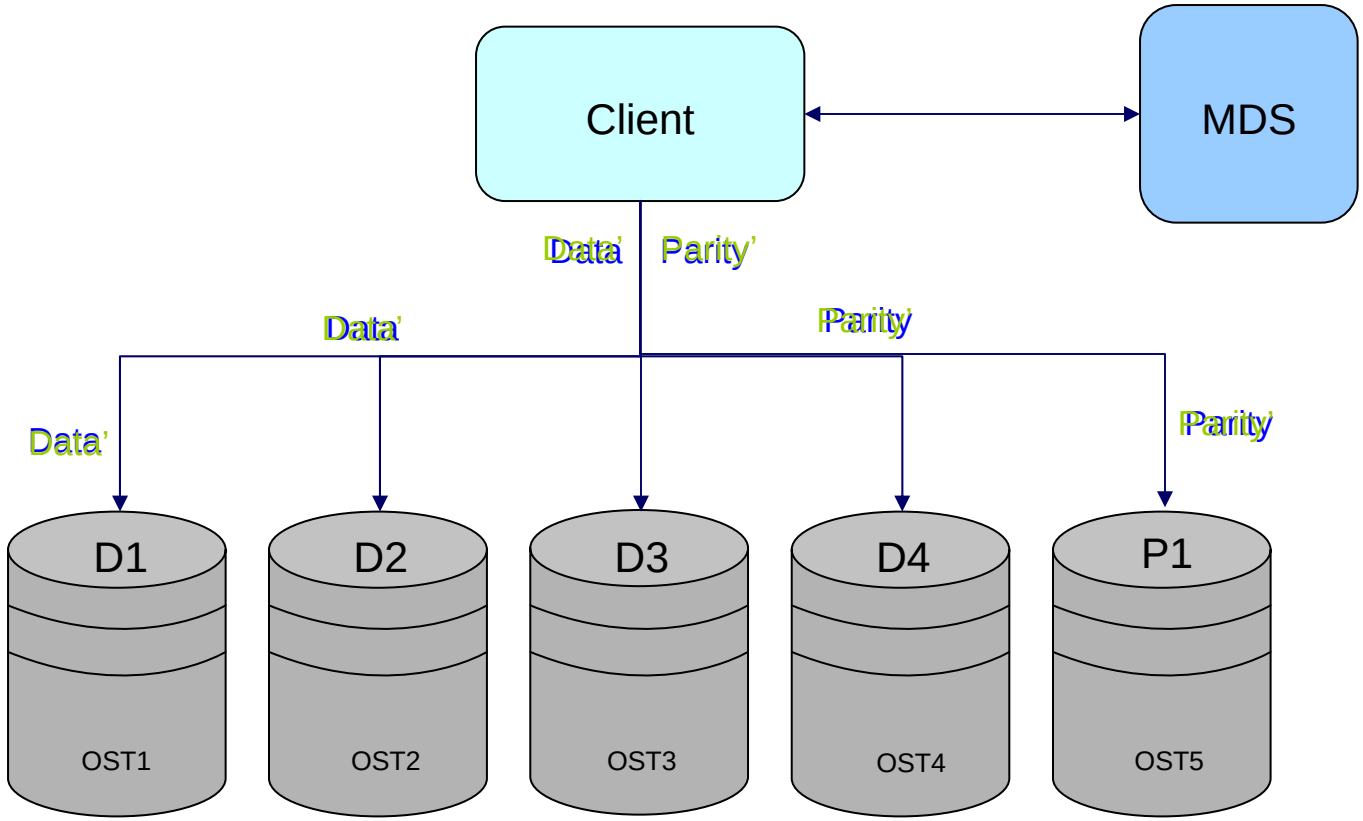
File System Caching

- Centralized cache frees up memory at the clients
- Metadata caching reduces access to metadata server
  - File layouts
  - Directory lookups
  - File access attributes

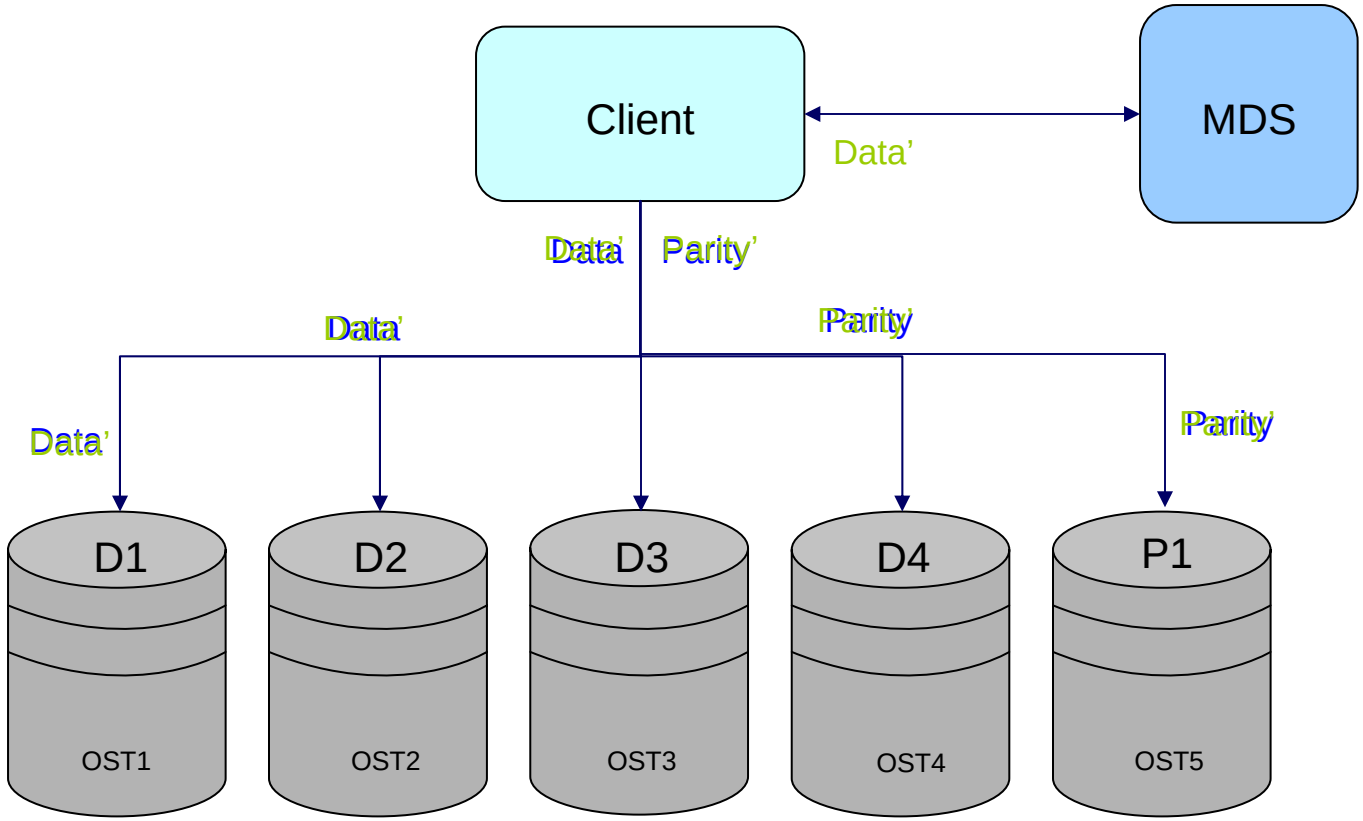
Redundancy Optimizations

- Attempt to apply RAID mechanism at target level for object-based storage file systems
  - Pros
    - Provides protection to data on loss of a target (network disruption, system failure, crash..)
  - Cons
    - Processing time, memory

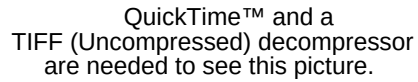
RAID in Lustre Targets



RAID in Lustre Targets



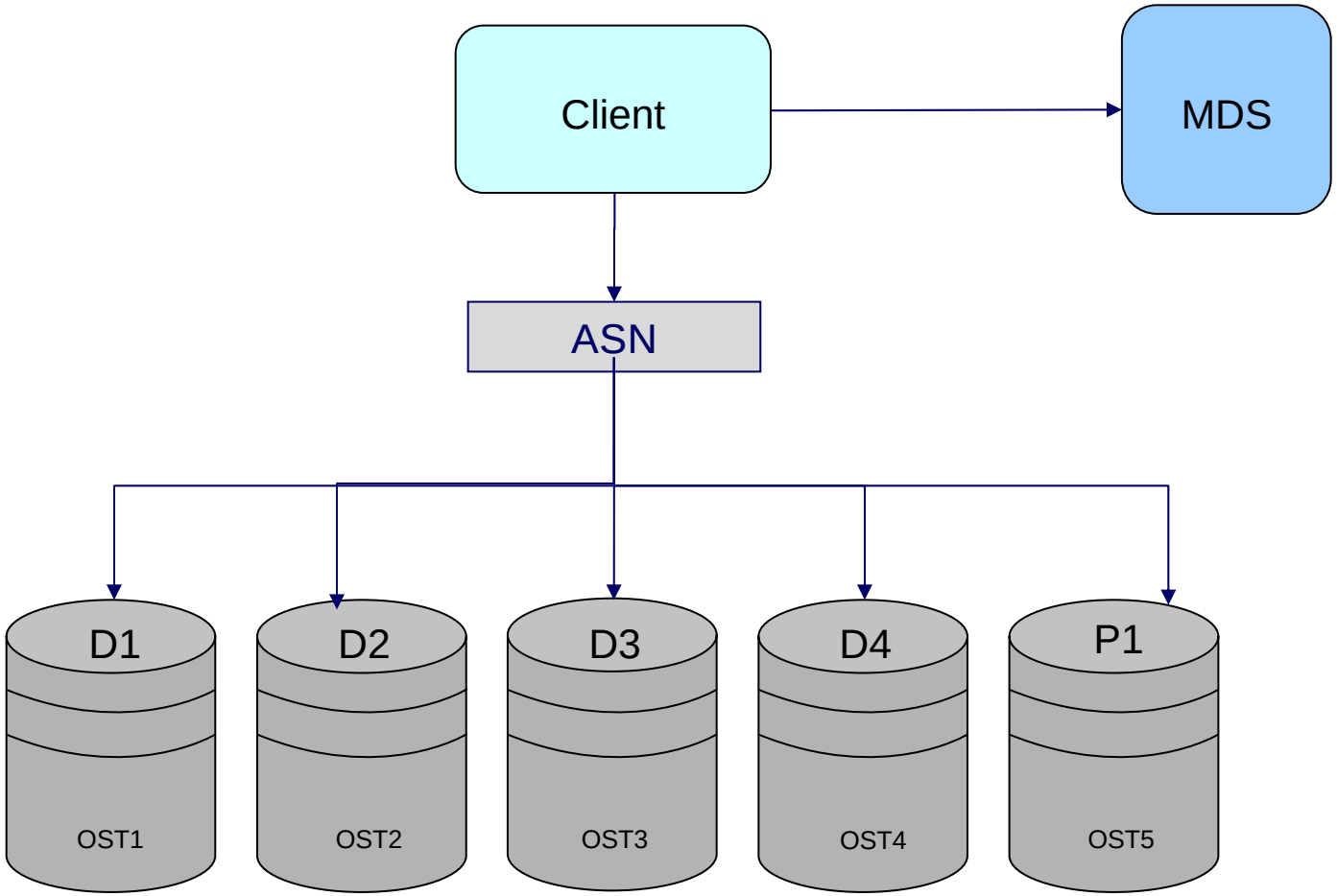
RAID in Lustre



Benefits with Active Storage Network

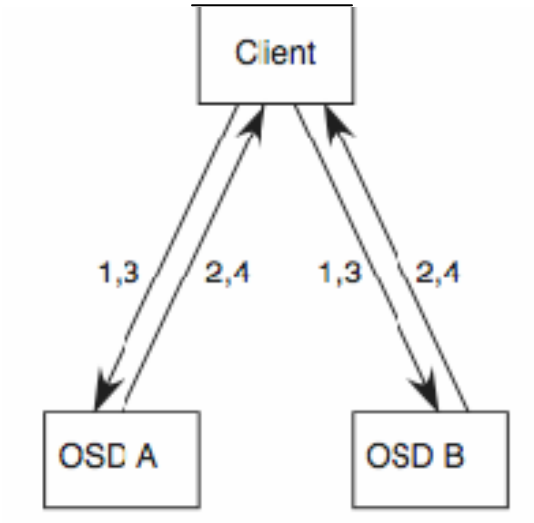
- Intelligence embedded within network.
- Intermediate level between client and target storage – used to assist performance in caching, redundancy management, locking.
- ASN can be used to buffer & generate parity (delayed parity) – independently handling read-generate-write operations.

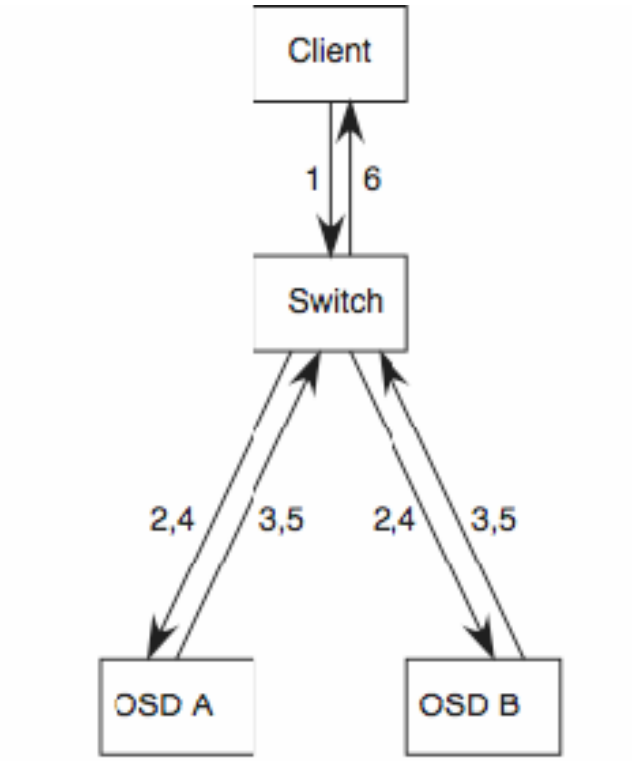
RAID in Lustre Targets



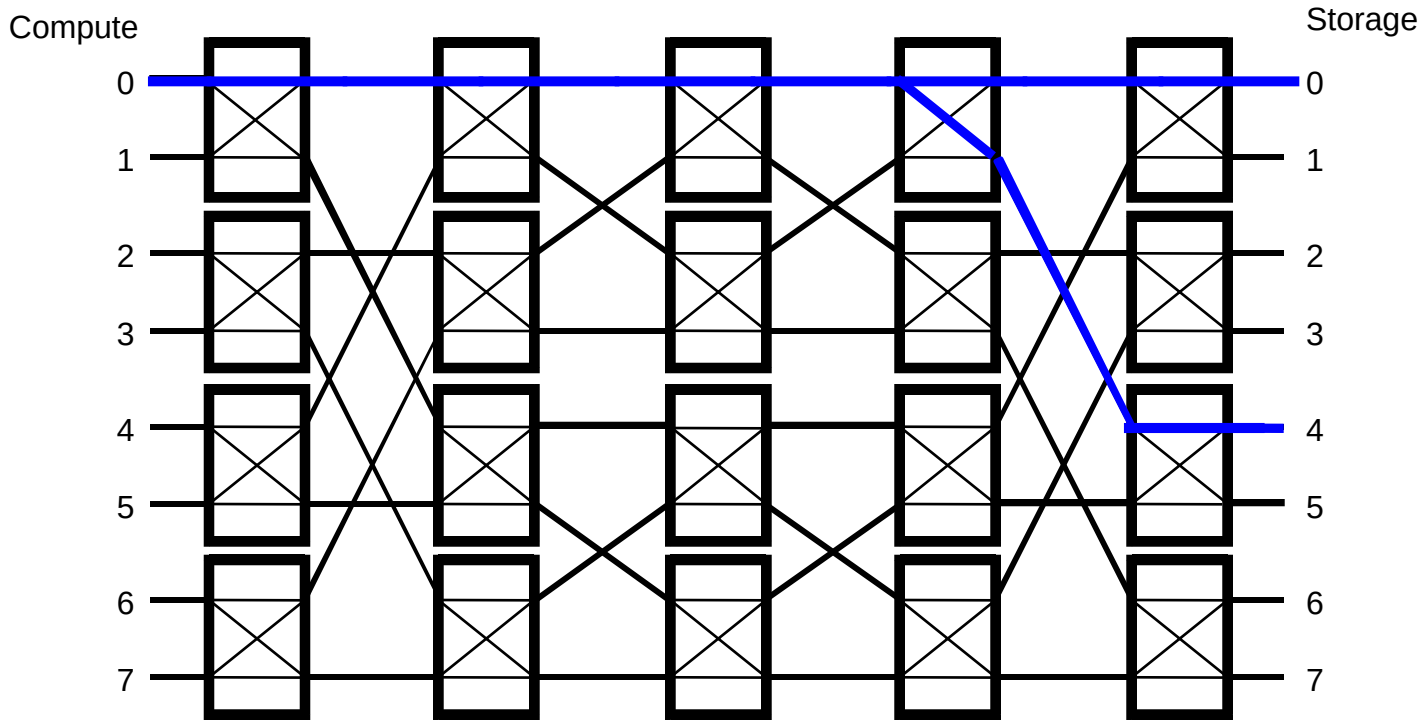
Redundancy optimizations

Parity calculations





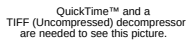
Network Structure



ASN Switch Implementation

- Reconfigurable elements (FPGA)
  - Faster design and lower cost than ASIC
  - Fast enough to handle gigabit speeds
- Downloadable functions (netlets)
  - Software functions on embedded processors
  - Hardware functions on FPGAs

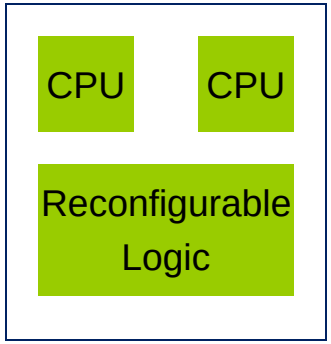
Hardware Implementation

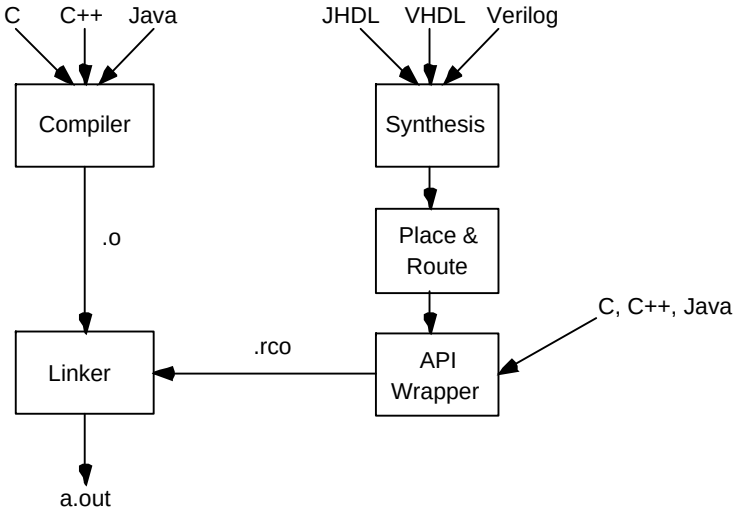


- NetFPGA board from Stanford
  - Nick McKeown and John Lockwood
- 4 GigE connects
- 2 SATA connectors for node to node communication
- PCI bus for node to node communication

Adaptive Computing

- Hybrid Computing Tools





- Applications: Embedded Systems, High-Performance Computing, Compute-intensive algorithms
- Algorithms: String matching, Numerical Methods

Summary

- Storage system optimizations for ASNs
- Investigation of ASN topologies and architectures
- Creation of ASN switch from reconfigurable components
- HEC applications for ASNs
- Programmable functions for ASNs
- Students
  - Sumit Narayan, Ajithkumar Thamarakuzhi, Mike Kapralos, Janardhan Singaraju