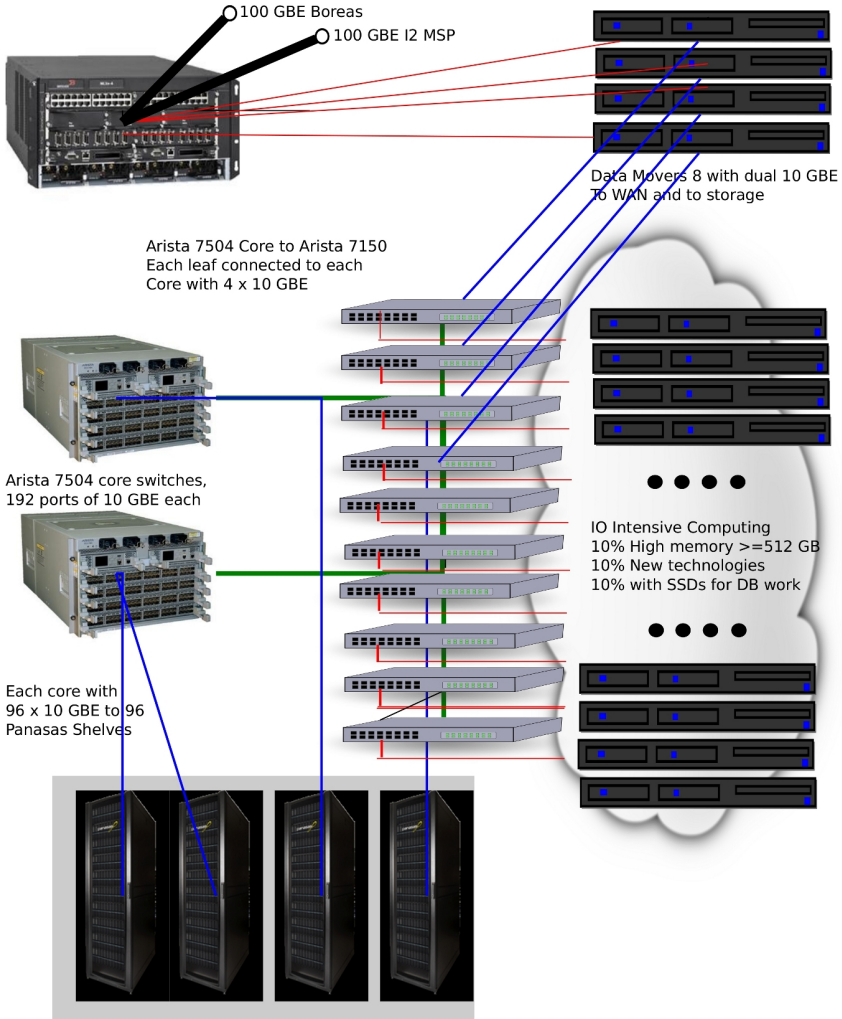
A FEW SMALL THOUGHTS ON BIG DATA

Jorge Viñals

School of Physics and Astronomy and
Minnesota Supercomputing Institute

University of Minnesota

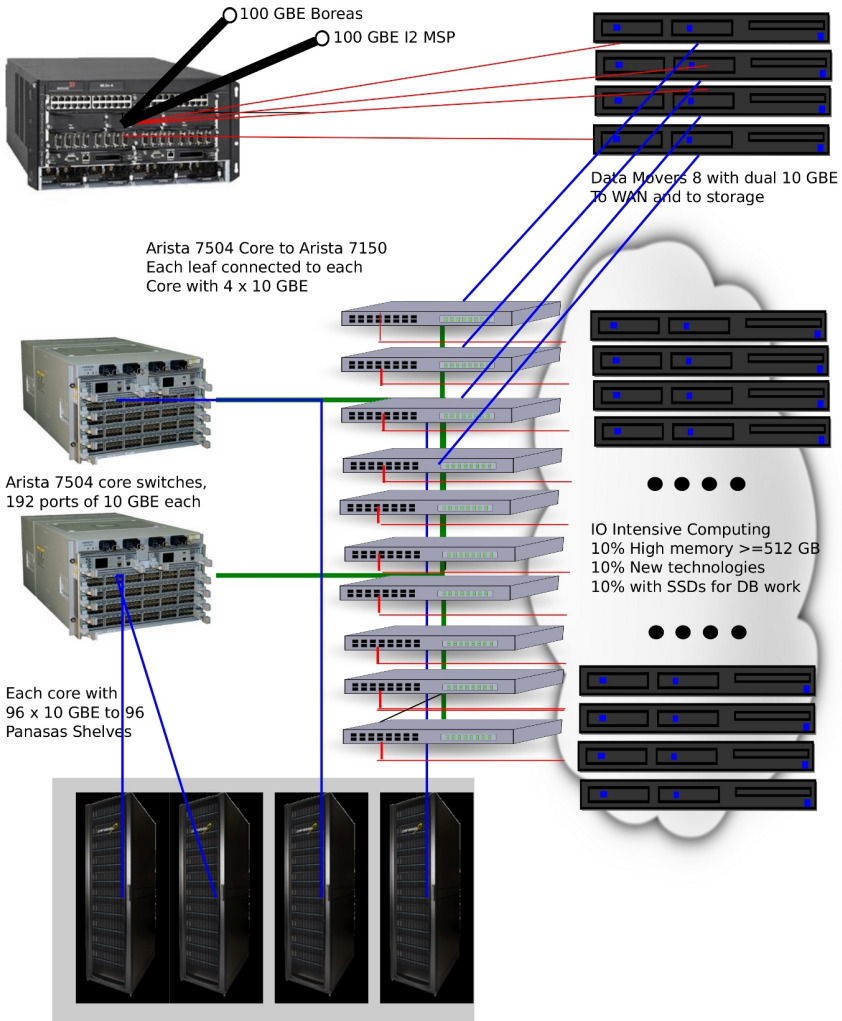
BIG DATA HARDWARE FOR SCIENTIFIC APPLICATIONS



A few design considerations

- The bulk of the budget into storage, not computing. Not an appliance, not a number cruncher.
- Balance - what to balance ? Memory, latency, bandwidth, capacity ... Likely a heterogeneous system.
- Benchmarks - what benchmarks ?

BIG DATA HARDWARE FOR SCIENTIFIC APPLICATIONS



A few design considerations

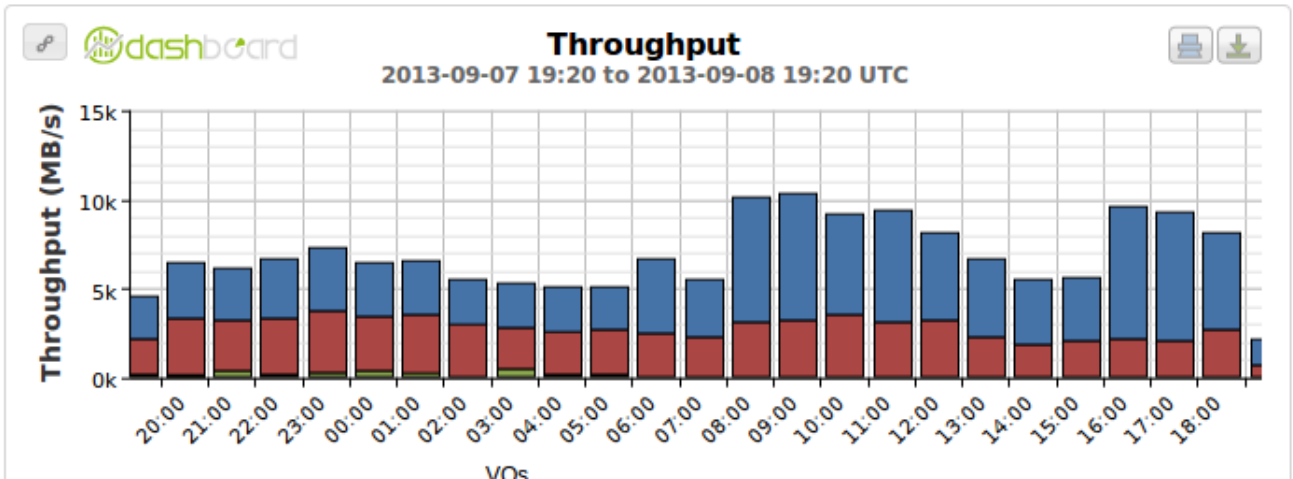
- The bulk of the budget into storage, not computing. Not an appliance, not a number cruncher.
- Balance - what to balance ? Memory, latency, bandwidth, capacity ... Likely a heterogeneous system.
- Benchmarks - what benchmarks ?

One example configuration

- 85% of budget on storage (12-15 PB), 225 TB of SSD, IP switching (10/40 GbE, not InfiniBand)
- 15% of budget on computing: 200 compute nodes, 140 for computation, 10 data movers, 20 visualization nodes, 10 large database (3TB SSD each), 10 large memory nodes, 5 GPGPU, 5 Phi.
- I/O specs: 125 GB/s, 4M IOPS.

A FEW HIGH LEVEL CHOICES

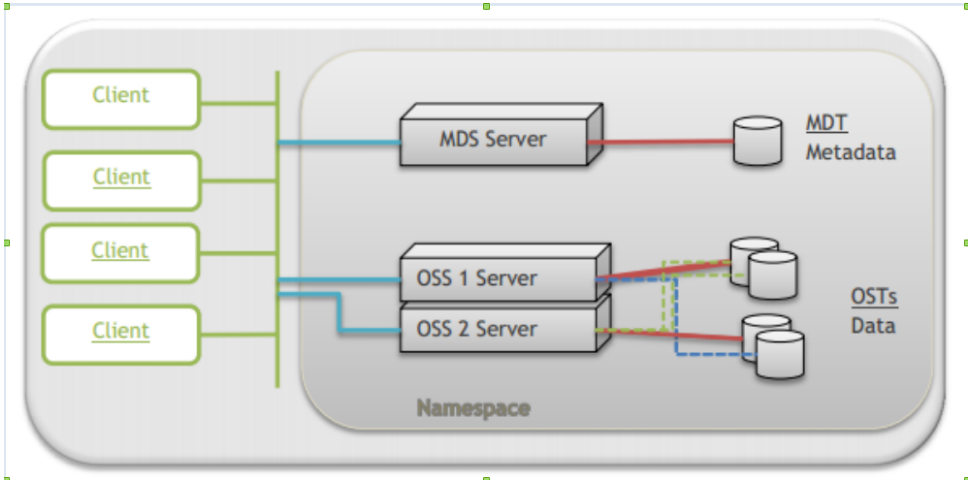
1. Computing to the data, or data to computing ?
   - Physical colocation of disparate repositories.
   - New hardware technologies for rule based (not file system based) distributed storage (e.g., Web Object Storage, WOS)
   - Data generation capacity growing faster than network bandwidth.
2. Fast large block transfers or high IOPS count ?
   - Scientific systems and Business Intelligence systems have different optimal criteria. No such thing as an "end to end appliance".



GENERAL STORAGE CONSIDERATIONS

Generally, storage falls into one of two categories: low-IOPS, high bandwidth (many large block scientific datasets), or high-IOPS and low bandwidth (many small files, or many transactions; database work).

Parallel File Systems (Lustre, gpfs, etc.)



- Key insight: separate metadata and data streams. Good data scaling with additional OSS servers.
- No parallel metadata yet. MDS becomes a bottleneck for high IOPS workloads.
- Data integrity (RAID) also a bottleneck.

BENCHMARKS, WHO NEEDS THEM ?

1. Synthetic I/O benchmarks well established: IOR, iozone, bonnie++.
2. No application class specific benchmarks.
3. Only one standarized benchmark of server throughput and latency: SPECsfs2008. Measures response time as a function of throughput, effetive measure of number of operations per second.
   - Requires NFS filesystem (a NFSv3 server).
   - Scaling studies not allowed.

BIG DATA MIDDLEWARE