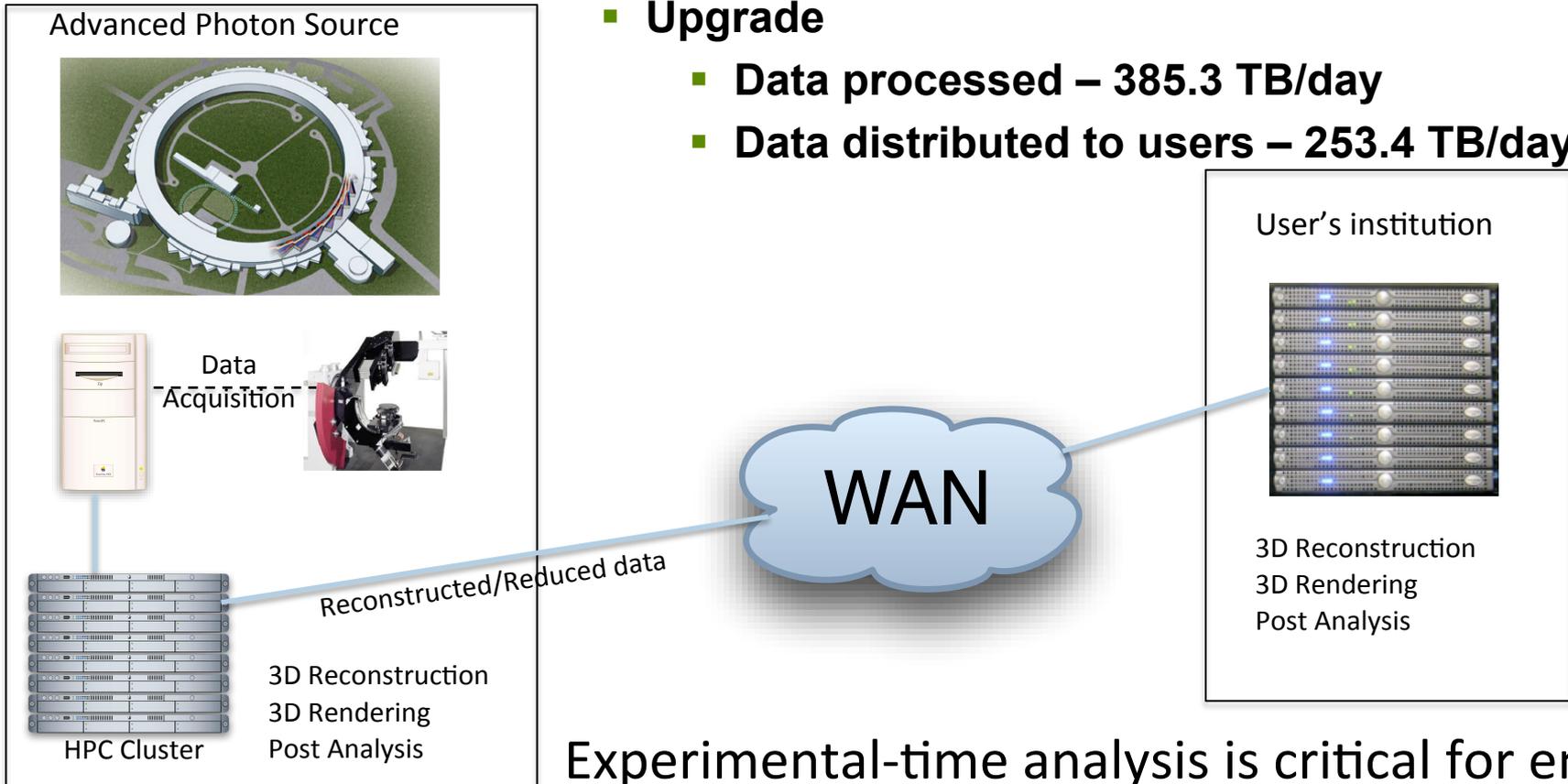


Parallel Networks and Storage for Predictable End-to-End Data Movement

Eun-Sung Jung, Raj Kettimuthu, Venkat Vishwanath

Tomography at APS - Experimental time Data Analysis

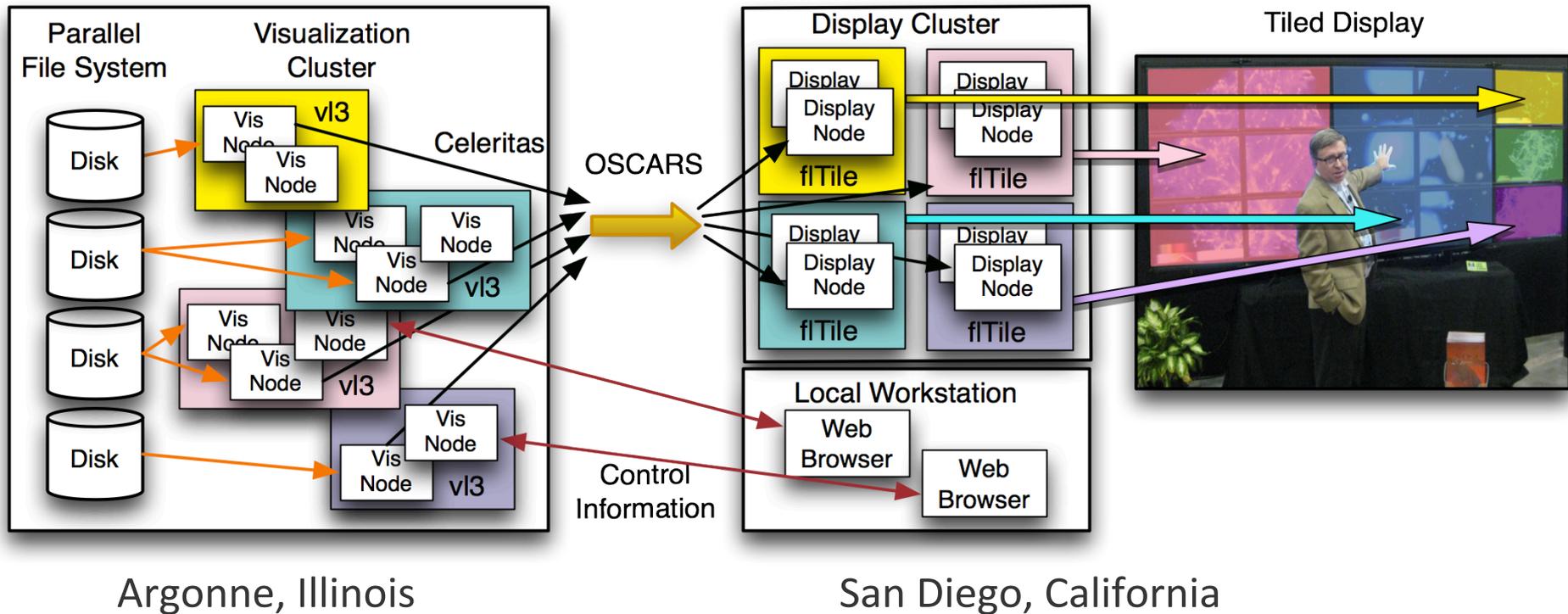
- **Current**
 - Data processed – 5.6 TB/day
 - Data distributed to users – 3.3 TB/day
- **Upgrade**
 - Data processed – 385.3 TB/day
 - Data distributed to users – 253.4 TB/day



Experimental-time analysis is critical for enabling interactive changes to experiment parameters



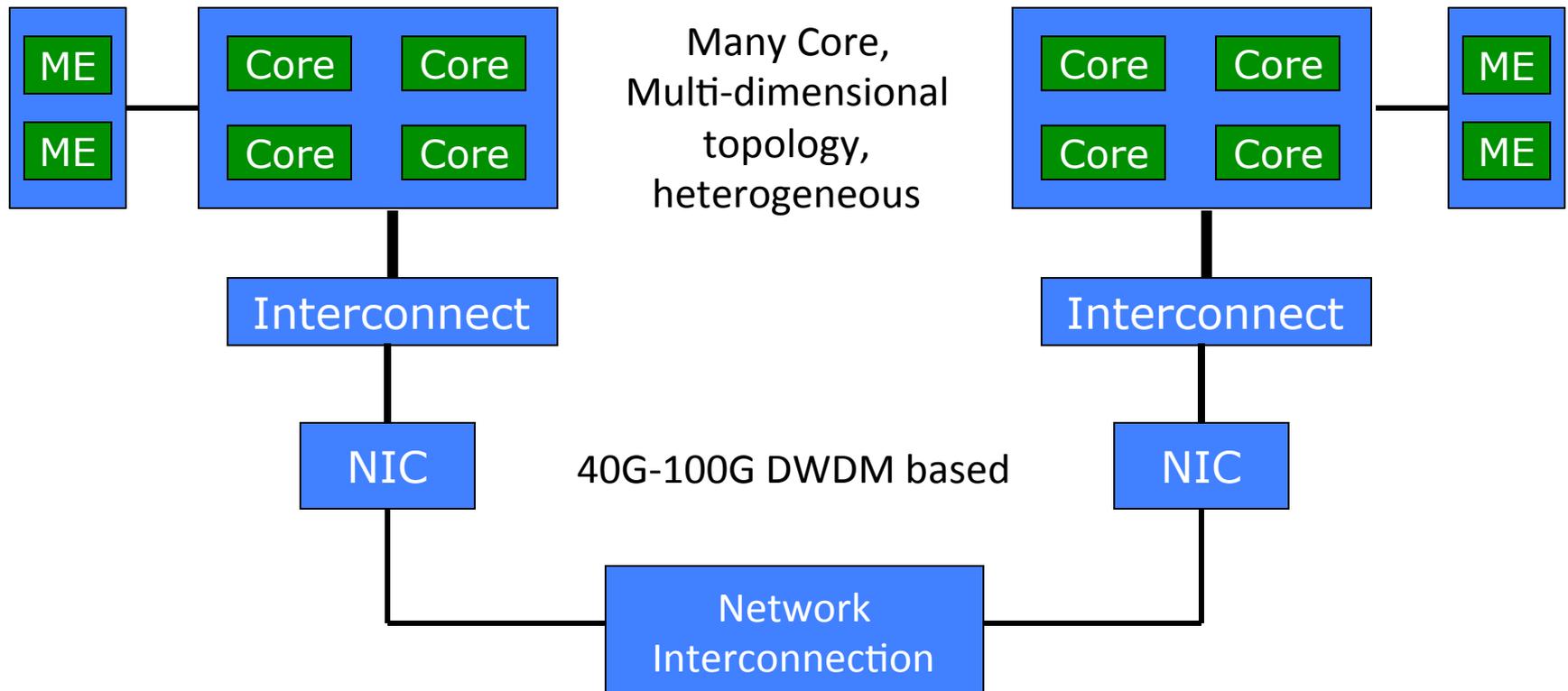
Moving beyond File Transfers: Interactive Remote Visualization of ENZO Simulations



Parallel data movement paradigms are of increasing importance to large-scale science



End Systems



Applications need to contend with the deep and complex system hierarchies and take advantage of parallelism in the various sub-systems



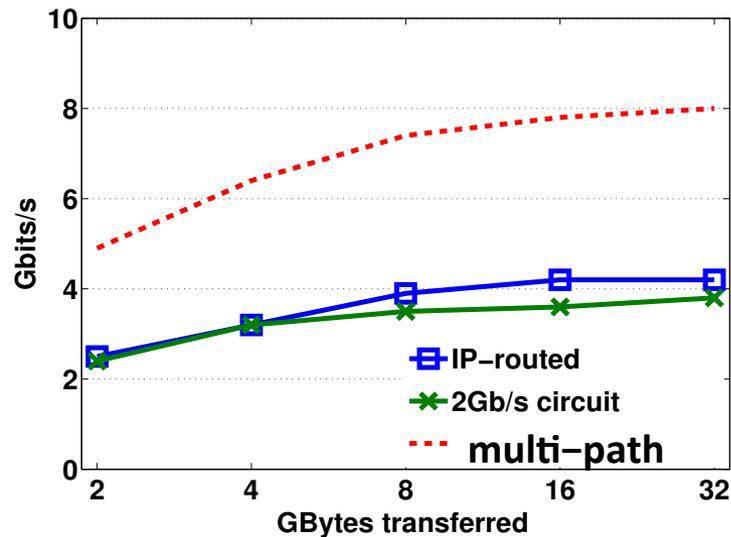
ESnet

ESnet5 Routed Network November 2012
DRAFT

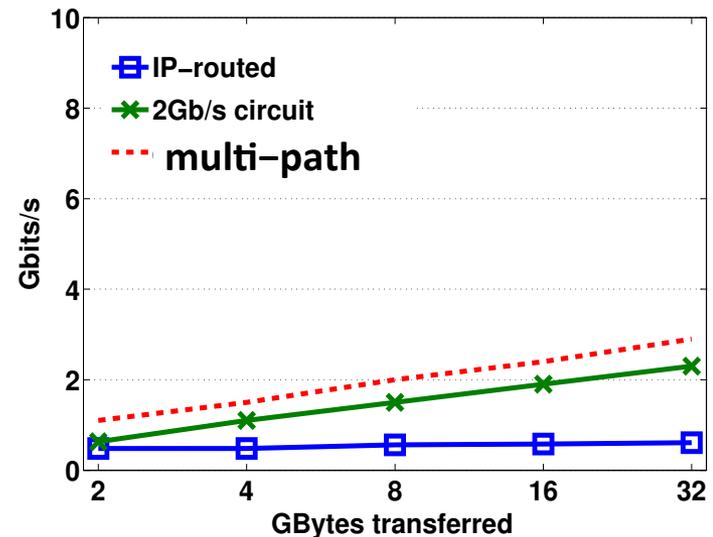


Network Parallelism

- Multiple stream data movement given a single path
 - Single stream is not enough to saturate the high bandwidth network links (e.g. 100g)
 - Multiple TCP streams are commonly used to improve throughputs
- Multipathing



ANL and NERSC



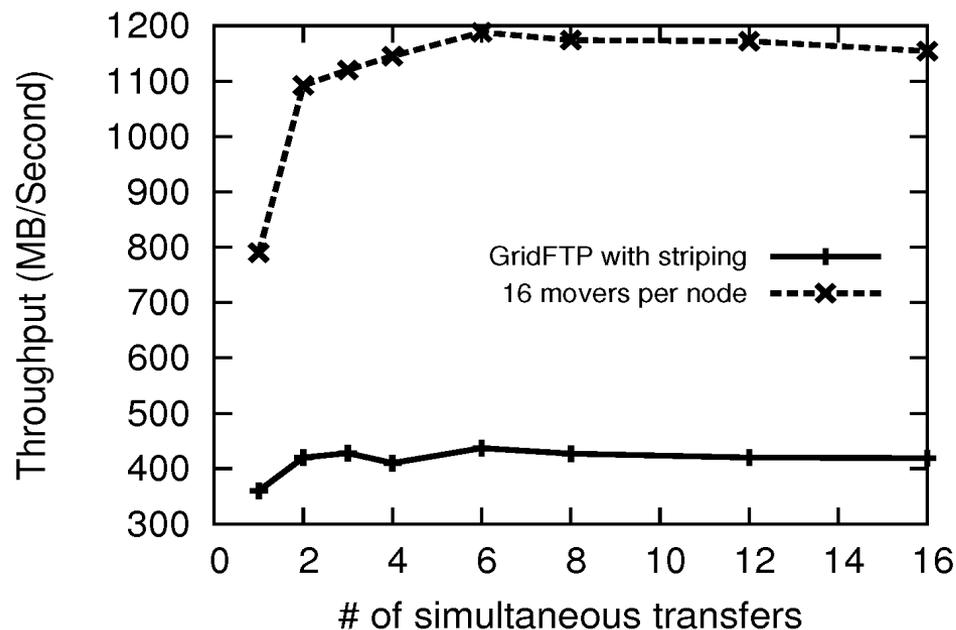
UMichigan and Caltech

Take advantage of both dedicated and best-effort paths



Network Parallelism: Heterogeneous Networks/ End Systems

- Paths may have different bandwidth characteristics.
 - Different network link bandwidth/NIC capability: 1Gbps ~ 100Gbps
- In GridFTP, the total data to be transferred is split statically and evenly among data transfer nodes. → Bottleneck in the slowest link
- End system aware parallel data transfer

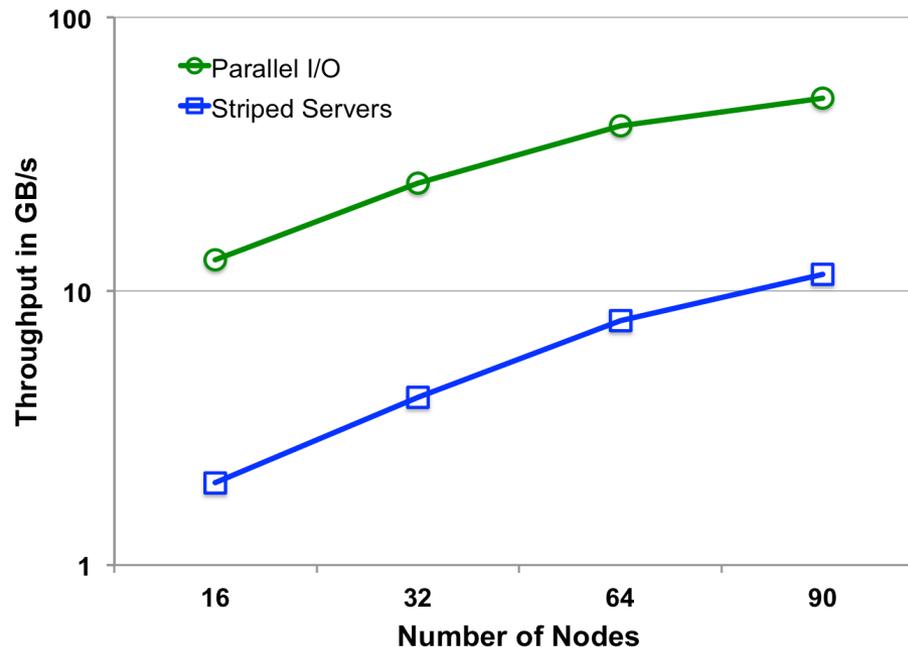


Adaptive file transfers over heterogeneous WAN



Parallel I/O

- Parallel filesystems such as Lustre, GPFS, and PVFS are being deployed for higher I/O requirements of applications.
- MPI-IO interface or higher-level libraries such as HDF5 are used to achieve scalable I/O from parallel filesystems.
- GridFTP provides an open-standard Data Storage Interface (DSI), but does not supports parallel I/O interfaces now.

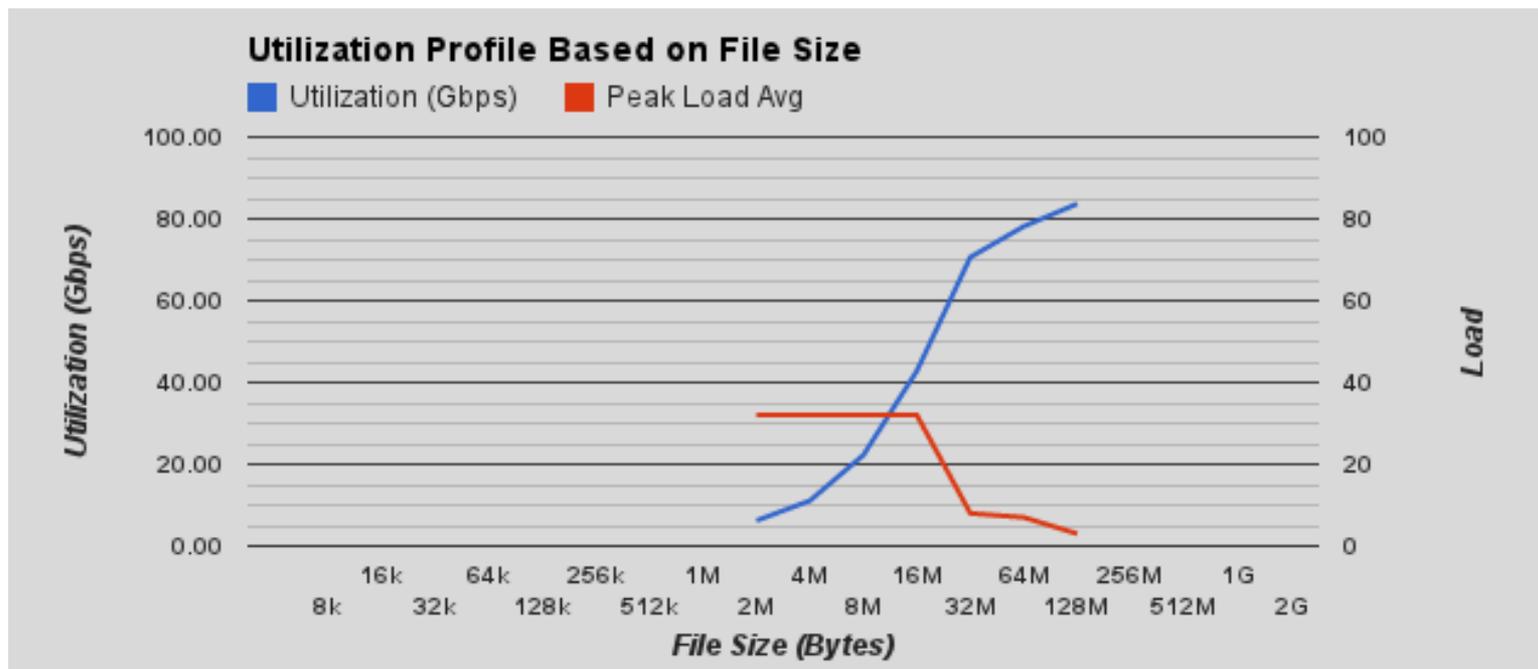


Accelerating I/O of ENZO data



Parallel I/O: Lots of Small Files (LOSF)

- Even though application data sets continue to increase, these may consist of lots of small files.
- Climate datasets are characterized by tens of thousands of small files.



Reliable End-to-End Data Transfer

- Bottleneck of a data transfer varies depending on the configuration of end hosts, storage systems, and networks.
- Possible approaches
 - Data buffering or staging to hide capability of storage systems
 - Overlapping storage and network transfers
 - Throttling network transfers based on the achievable storage performance



Thank you!