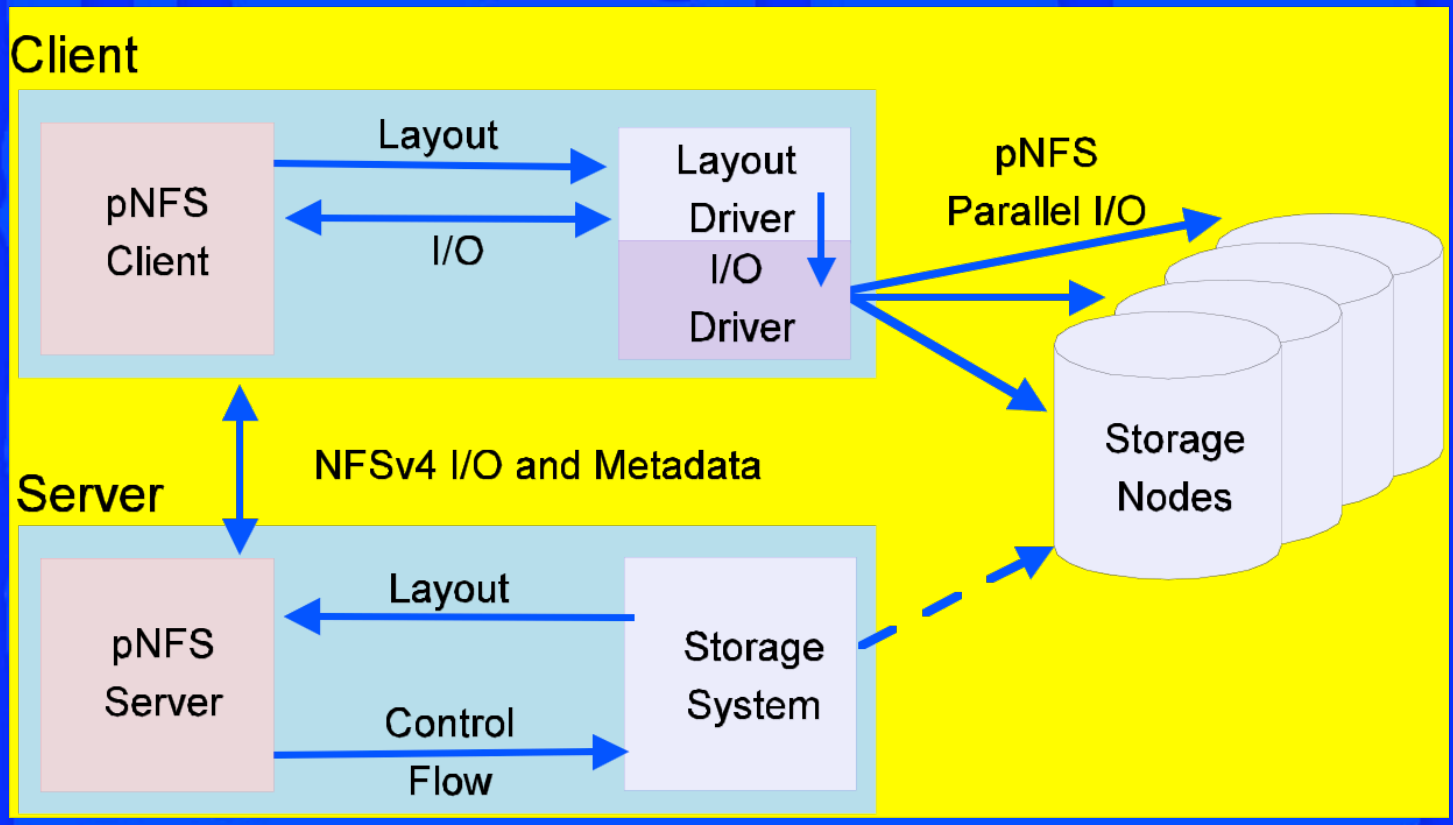
parallel NFS in NFSv4.1

* peter honeyman
center for information technology integration
university of michigan, ann arbor

pNFS schematic

- client I/O bypasses NFS server to scale with storage



NFSv4.1 addresses HPC problems

- scalability
- performance
- standard security mechanisms
- heterogeneity
  - operating systems and hardware platforms
- transparent file system access
  - predictable semantics/security
- file system independence
- leverage layout driver development and support

where NFSv4.1 is @

- active development of pNFS spec
  - currently on version 8
  - draft-ietf-nfsv4-minorversion1-08.txt
- object and block specs also in draft
- prototype interoperability began in 2006
  - san jose connect-a-thon march '06
  - ann arbor NFS bake-a-thon september '06

where NFSv4.1 is @

- linux pNFS code freely available
  - file and PVFS2 layout drivers
    - www.citi.umich.edu/projects/asci/pnfs/linux
  - block driver under development
    - anticipate 2/07 release
- sun file layout code under development
- everyone is at “early prototype” state
  - successfully interoperated sun client with linux server

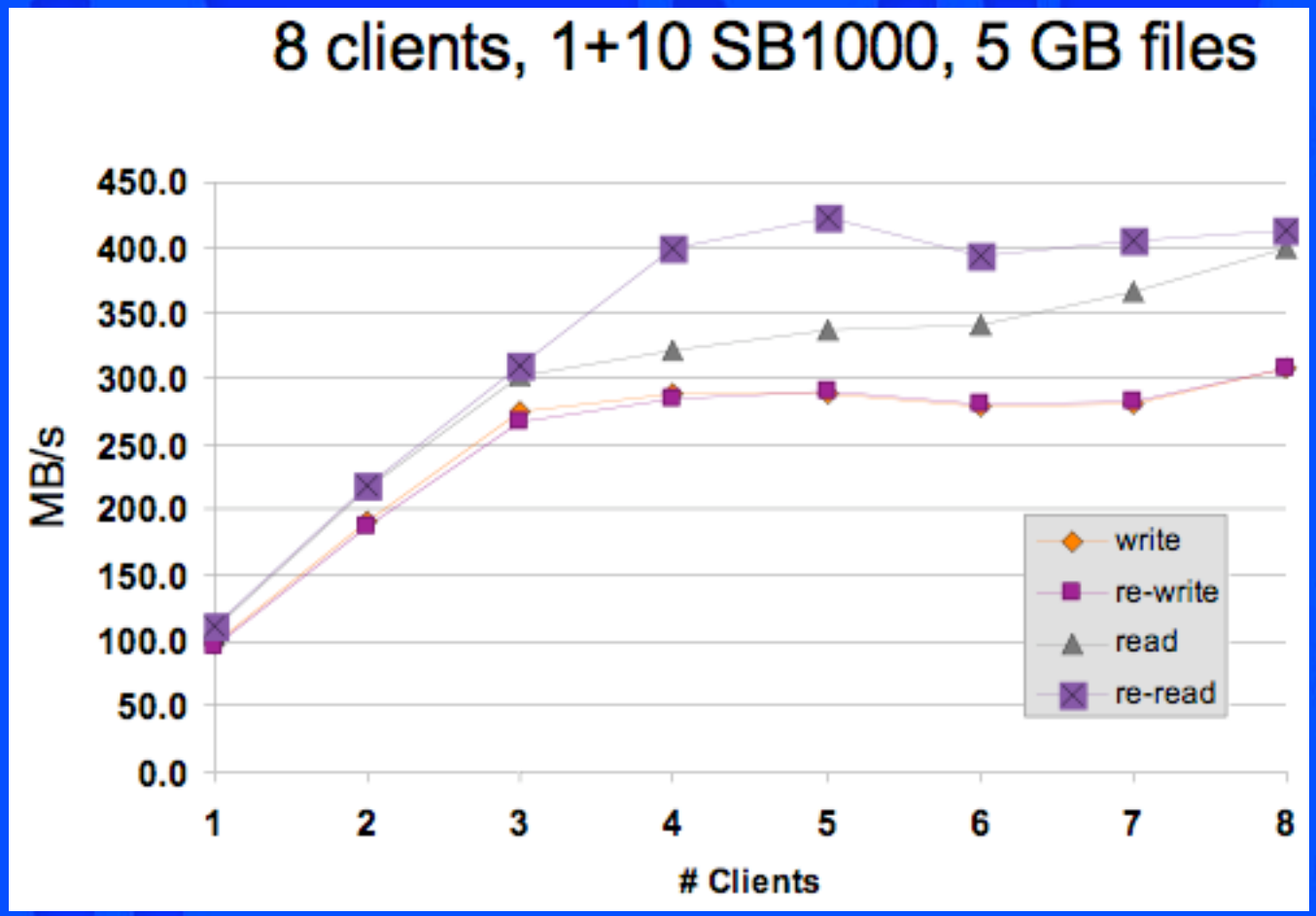
where NFSv4.1 is @

- linux pNFS weekly conference call
  - CITI (files over PVFS2, files over NFSv4)
  - netapp (files over NFSv4)
  - IBM (files, based on GPFS)
  - EMC (blocks, based on highroad)
  - sun (files over NFSv4, objects based on OSDv1)
  - panasas (objects, based on panasas activescale storage cluster OSDs)
  - CMU (performance and correctness testing)
- wiki.linux-nfs.org

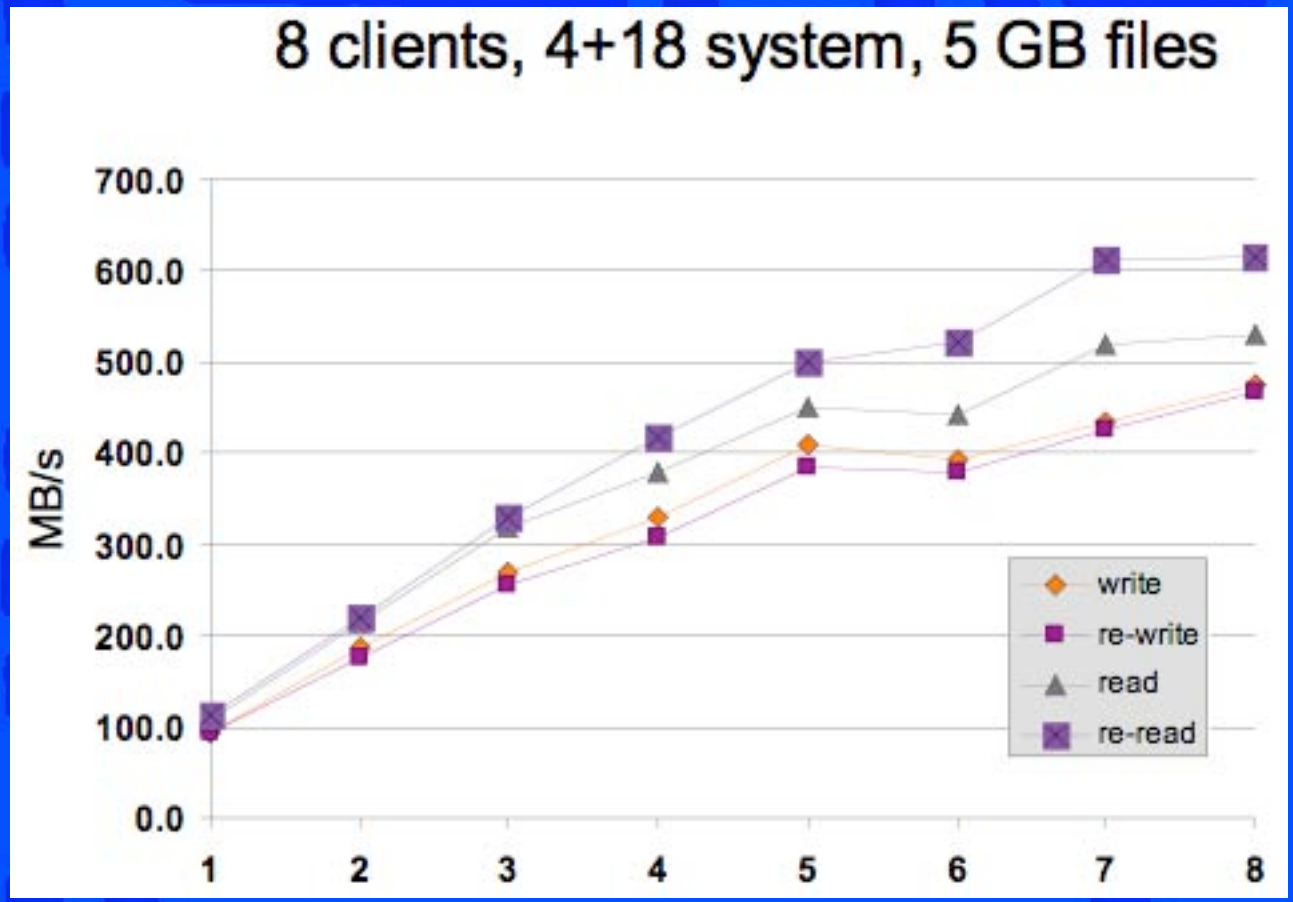
linux pNFS milestones

- pluggable layout driver client architecture using I/O and policy software interfaces
- file and PVFS2 layout drivers for PVFS2
- demoed complex data copies
  - single client
  - two layout drivers (file, PVFS2)
  - gpfs, dcache, PVFS2, and netapp
- CITI has issued four pNFS research papers
  - and three related papers

pNFS iozone write throughput



pNFS iozone read throughput



how others can contribute

- linux pNFS
  - MPI-IO support (ROMIO module)
  - NFSv4 listio support
  - linux NFSv4 server request gathering
- pNFS protocol/prototype
  - scalable metadata management (optimize interaction of NFSv4 and parallel file system metadata management)

resources

- linux NFS wiki
  - wiki.linux-nfs.org
- file and PVFS2 layout drivers
  - www.citi.umich.edu/projects/asci/pnfs/linux
- CITI technical report series
  - www.citi.umich.edu/techreports