

HPC User Site Census: Storage

Christopher Willard, Ph.D.
Laura Segervall

Addison Snell

December 2010

EXECUTIVE SUMMARY

This report, part of our Site Census series, provides an examination of the storage characteristics and capacities found in a sample of HPC user sites. We surveyed a broad range of users about their current computer system installations, storage systems, networks, middleware, and supporting software.

Our goal in this analysis of storage systems is to examine storage usage within the HPC user communities and to explore how this usage varies, based on categories such as storage capacity and its location, storage supplier, and storage network.

Key findings of the survey include the following:

- Approximately 34% of the total available storage at respondents' sites resides on compute servers. Storage available to each node (referred to as node-level storage) represents 14% of the capacity and storage available to the server (referred to as system-level storage) accounts for 20% of the total available storage. The remaining 65% of storage is found at the site level, generally on Network-Attached Storage (NAS) or Storage Area Network (SAN) systems. On average, 285TB of storage resides at the site level on a storage system.
- About 75% of the 312 sites have at least one site-level storage system installed. No vendor dominates the storage system market for HPC sites. IBM has the largest share with 10.8% but is closely followed by DataDirect Networks and Sun with 9.5% each. The "in-house" and "generic" solutions combine to account for 9.8% of the storage systems installed – more than each of the number two vendors. We see this last value as reflecting the commodity nature of storage components and the availability of open software for storage systems.
- NAS and SAN systems had almost equal representation in the surveyed HPC sites, with 38% SAN and 40% NAS. Commercial sites are more likely to have a NAS storage system, while academic sites are more likely to have a SAN storage system.
- The majority of NAS storage systems are connected using 1 Gigabit Ethernet. The share of systems connected by 10 Gigabit Ethernet increased from 8% in the 2009 survey to 17% in the 2010 study.
- Data revealed almost equal usage of InfiniBand and 10 Gigabit Ethernet for site level storage system networks. About 32% of the storage systems utilized a higher performance network (InfiniBand or 10 Gigabit Ethernet) and 52% of those storage systems used InfiniBand.
- Most storage management software (36%) in use by the survey sites was provided by the storage system vendor. Very little penetration by add-on storage management suppliers was reported.

- Clustered parallel file systems are gaining share in the HPC market and represent 27% of the storage management tools. As these file systems gain greater share, the practical distinction between NAS and SAN may diminish.