

DATA MANAGEMENT FOR HIGH PERFORMANCE COMPUTING/RESEARCH



Unmatched scale, speed and security.

Scientific research increasingly utilizes high performance computing (HPC) infrastructure that carries with it some unique storage challenges. Data is produced by experiments and temporarily cached in expensive silicon storage or in proprietary parallel file system storage. These technologies are optimized to capture data at high speed. But they are either limited in capacity, or do not provide easy file access for protection or downstream workflows. So, it follows that the data collected in high performance repositories must subsequently be moved into less expensive and more scalable general purpose file or object storage kit as soon as it is practicable to do so.

Most of the movement of data today from HPC infrastructure to general purpose file storage is performed manually, making the activity prone to error. Problems such as silo'ing and misclassification make the research data difficult to access again when needed whether the result of storage silo incompatibilities or because the data simply cannot be searched for and found successfully.

Silo'ing is not a purely technical challenge. Often, in grant-driven research, storage is acquired with each project award – resulting in pockets of storage technology of different vintages from different manufacturers. Ultimately, these silos of technology may include different types of network attached storage (NAS), shared IP SAN storage, and direct-attached storage arrays without a common management layer. When data from instruments or parallel file systems are written to their designated silos, they are subsequently more difficult to access, protect or share.

Metadata is the key to finding data once it has been stored, but files are often migrated to general purpose storage silos without much attention to metadata refinement or classification. This makes files harder to locate when they are needed later on; data can become orphaned.

These issues, while challenging, are not insurmountable. When it comes to managing data in an HPC/research setting, the key to success is consistent and policy-based data management services enabled by StrongLink, the Intelligent Data Manager. StrongLink automates the annotation of files with metadata references and the migration of files from capture storage to retention storage, so a researcher doesn't need to perform these tasks manually.

StrongLink enlists intelligent data management, powered by StrongLink Autonomous Engines technology, to ensure that data has provenance and that it is placed in an optimal manner to ensure that its accessibility, preservation and protection requirements are met.

For research project leaders, StrongLink makes IT tasks much easier. It works with all data and with all storage platforms. Policies are easy to develop and simple to associate with metadata constructs

typically used by file and object systems. Policies can be extended to new data readily and actions (such as data moves) can be refined when needed or when new storage technologies are added.

StrongLink isn't artificial intelligence, it is augmented intelligence – an easy-to-use facility that storage administrators can use to help de-clutter their schedules and to increase the amount of work that they are able to accomplish each day. Deployment is a snap and organizations begin realizing benefits immediately from the creation of a Global Namespace, describing and indexing all data hosted on all storage whether on-premises or in clouds, and a Dynamic Data Mover that uses your data lifecycle policies to constantly rationalize and optimize data placement based on your priorities.

The payoff is immediate. Data value can be unleashed by enhancing the accessibility of files and objects, even in a heterogeneous storage setting, and by providing each file or object with a rich layer of inherited metadata, which automatically applies agreed-upon data classification standards defined by users or the research institution. These benefits are realized without increasing the burden on researchers, or risking data being unclassified, or misplaced; all functions are automated and performed in a way that is transparent to the user.

Now is the time to automate data management. StrongLink provides a solution that does not require the replacement of your existing storage infrastructure or the laborious and time-consuming planning and selection of storage hardware going forward. Automated Data Classification and Automated Data Migration from StrongLink enables you to streamline and expand the use of research data to solve bigger problems.

Schedule a free demo today to get answers to all of your questions and to see, first hand, why StrongLink has captured the interest of research organizations and enterprises large and small?

www.stronglink.com

With StrongLink data management automation,

- *Data can be classified automatically by applying metadata "wrappers" defined by project planners. This eliminates the need to rely on researchers to manually classify data associated with a grant, department or other references.*
- *Data provenance ensures that researchers will know when and where the data was created, and to what standards or schemas the data has been coded – not only making data easier to find when needed, but also easier to describe when the data is published.*
- *Data is managed throughout its useful life in accordance with lifecycle management policies, reducing the potential for data integrity issues or data loss.*
- *Data can be written anywhere in the storage infrastructure and managed in common via a global namespace compatible with any brand of file or object storage system.*