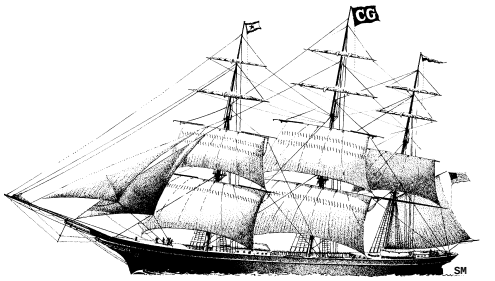
THE CLIPPER GROUP
Navigator™



Navigating Information Technology Horizons℠

Published Since 1993 Report #TCG2005022 April 19, 2005

The Data Side of Grid — The Roles of Containers and a Single Name Space

Analyst: Anne MacFarland

Management Summary

The buzz about grid computing has engendered an echo of muttering about what such opportunistic and flexible computing architectures will demand of the data structures they use. Data has traditionally been stored in application-specific spaces assigned to particular servers. What changes await us as this data is used by applications that move, and by multiple applications?

In fact, many of the changes that grid demands have been brought to the table already by the flexible ways that businesses are using their enterprise data. The need to reuse operational data in different contexts such as analysis, and the increasing need to *find* data across the enterprise has changed what is expected of data storage and structures like file systems. The pervasiveness of global competition brings new attention to business process efficiencies, and the "real-time" data synchronization that those efficiencies has changed what is expected of data access. Moreover, the combination of constrained budgets with rapidly growing data has changed how data is managed. All data cannot be treated alike. Dynamic data will require more frequent documentation than static data. Data that is used broadly will need a more robust infrastructure than that with a limited relevance.

These expectations cannot be met indirectly by management of the media on which the data sits. Data must be characterized, and classified for management by a logical level construct, such as a container, not simply tracked by physical location. Access, security, and privileges must be based on logical- rather than physical-level constructs. File systems must be unified across the scope of an organization.

Every copy of data must have a reason to exist, and must be managed. The dumps of *just-in-case* copies, particularly those of questionable currency, are redundancies of little value. Their improper use can be a considerable risk to an organization. Regardless of how these demands are met, they must be transparent to users, and must make administration easier, not more complex. It should be built into the IT environment, and built to open standards, so that it can work with other network and management elements.

Network Appliance, of Sunnyvale, CA, is evolving its products to provide storage for the new paradigm needed for efficient use of networked enterprise data assets, and for the support of grid architectures. This is, after all, what the vision of NAS is all about. Read on for more details.

IN THIS ISSUE

The Reformation of Storage

Business always has needed agility, operational resilience, and cost management. Now that much of business process is done with a technology assist, the technology needs to meet these same business criteria. *Grid computing* meets these requirements by decoupling workload and/or applications from a fixed location on a particular server, and deploying it opportunistically on an appropriate processor. This gets more use out of enterprise servers. It also allows an enterprise to use its servers more flexibly, throttling low-priority workloads to free up resources to address unexpected, high-priority needs. *Service-oriented architectures* are a more subtle incarnation of grid computing – focused more on building links between applications with XML and Web Services – to allow business processes to be more completely and seamlessly supported by IT infrastructure. **The efficiency of both architectures is improved if the data is characterized and secured as part of the data structure and not merely by the application that generated it.**

As businesses seek to use operational data for modeling and for real-time inventory sharing with partners and suppliers, **data strategies must be developed to keep *real-time data* safe and secure yet quickly available to all of those authorized to use it.** Traditional (often console-based) management of disk and tape infrastructures to this business end is inadequately automated and overly cumbersome, and cannot meet the business criteria for IT agility, resilience and cost control.

The vision of NAS pioneered the idea of access to data not based by ownership of physical media zones, but by a file system, and constrained by access permissions. This was more than just a change in the size of data chunks. It was a change of perspective from the physical perspective of the storage hardware to the logical perspective of the application, and through that shift, to the business process. Storage management, for the most part, has been preoccupied with physical concerns, like preventing data overwriting and the pre-assignment of space to applications. Network Appliance was able to simplify storage administration with the introduction of WAFL[1], which, by its nature, solves the problem of overwriting.

Containers

Containers are a concept familiar from other parts of the IT infrastructure as partitions of server processors and switch ports, and as virtual machines, and other application containers[2]. Like an envelope, these containers *contain* and *isolat*e their contents, both to protect them from external faults and to allow multi-tenancy (of applications on a single processor or of application data streams on a single switch port). The storage equivalent of this kind of container is the zones on a SAN moderated by a switch, or partitions created on internal or direct attached storage (DAS). Many of these physically-oriented containers are somewhat inflexible[3]. In addition, while they can isolate and protect, they cannot support the well-arbitrated sharing that is the heritage of NAS.

A logical container, on the other hand, can be easily expanded – or shrunk. It gives a logical-level degree of containment and the isolation, but, in addition, it provides an opportunity to characterize the data in the container so that it can be more rationally managed – like data objects but at a larger scale with less aggregate bulk. Data is still accessed by application as blocks or files, as heretofore. The container is invisible to users.

Containers can be used to make routine functions, like back up, much more effective, and can make granular restores quicker. Searches can be better targeted, for a quicker find and a high proportion of relevant results. ILM becomes easier to do. Basically, logical-level data containers turn heaps of data into multi-dimensionally characterized folders with evolvable indexes. They turn the obscurity of an enterprise's unstructured data environment into a more completely navigable space. And, in terms of rationalizing your knowledge of, access to and use of enterprise data, this is a very good thing.

[1] WAFL stands for Write Anywhere File Layout.

[2] For more information on one kind of application container see **The Clipper Group Navigator** dated May 11, 2004, entitled *Civilizing the Unruly Application with Softricity*, at http://www.clipper.com/research/TCG2004042.pdf.

[3] *Thin provisioning* of storage, introduced by 3Par Data does allow the administrator to configure generously and only assigns space as it is written to.

Copyright © 2005 by The Clipper Group, Inc. Reproduction prohibited without advance written permission. All rights reserved.

To do this, Network Appliance uses *FlexVols*[4]. These containers can rationalize both structured and unstructured data. They can organize files by subject, or provenance, or use. For database data, organizing tables, binaries, and logs in separate containers with separate backup schedules makes it possible to back up dynamic data more frequently and quickly. It can make sharing data with partners easier.

Flex Clones, NetApp's new, read/write copies of data, could also be used for testing and analytics where the changes will not be synchronized. These clones only require space as they deviate from the original. Their use in parallel processing over a grid could create time efficiencies for the project as a whole.

NetApp has also added the ability to split these containers easily – something extremely difficult with old, physical-location-based volumes. This, allows data to be sorted (for instance, by freshness), and then to be segregated, allowing improved performance in analysis of truly active data.

NetApp also permits clone hierarchies[5]. Altogether, this set of container capabilities offers many ways to rationalize the enterprise information environment. And, of course, all of these containers preserve volume semantics to satisfy the expectations of applications.

There is one caveat: These containers cannot be used with NetApp *SnapLock* for compliance purposes. Data subject to regulations should be copied when written and the separate compliance copy secured with SnapLock.

Single Name Space

For all files that are named and arranged in hierarchies, there must be a wholeness that can be searched to assure that all items of relevance are included in a search. NetApp's Spinnaker acquisition gives this capability. It will be integrated into the next release of NetApp's operating system, *Data ONTAP*.

[4] For more information about Network Appliance's FlexVols and FlexClones, see **The Clipper Group Navigator** dated November 15, 2004, entitled *NetApp's FlexVols and FlexClones Raise the Rate of Return on Storage Resources* at http://www.clipper.com/research/TCG2004094.pdf.

[5] Clone Hierarchies allow the creation of logical-level subsets within a FlexClone.

The Benefits of this approach

This NetApp reformation of how information is managed is part of a move to *data management*, as contrasted with *storage management*. The need to configure LUNs to meet the expectations of applications will persist, but as the repeatable particulars are automated, the focus can switch from managing the media to managing the data as it is used by the enterprise and, more frequently, by multiple applications.

Today, NetApp offers containers by which to characterize and manage tan enterprise's data. **In the longer term, these containers, together with the pervasive name space to navigate that data, provide the basis for using enterprise data in more natural and rational ways, without having to mind all the limitations set up by a media-focused management structure.**

Conclusion

Due to rampant digitization, businesses have both *more* and *more kinds of data* than they did a few years ago. Spurred by compliance to keep data handy, and goaded by the need to survive, they are using their data differently. Network Appliance has a clear vision of the environment needed to support more extensive use and reuse of enterprise data in a way that is safe, auditable, and efficient. This is not a matter of features but of the architectures and infrastructures that will make the features effective.

Think of how your enterprise is changing, and what you have to do to get your business processes evolved to where you want them to be. Think of the role Web Services and Grids have in getting more out of your infrastructure. Managing your data by well-characterized containers and a single name space complements these efforts, and makes evolving your business a more rational process.

Copyright © 2005 by The Clipper Group, Inc. Reproduction prohibited without advance written permission. All rights reserved.

About The Clipper Group, Inc.

The Clipper Group, Inc., is an independent consulting firm specializing in acquisition decisions and strategic advice regarding complex, enterprise-class information technologies. Our team of industry professionals averages more than 25 years of real-world experience. A team of staff consultants augments our capabilities, with significant experience across a broad spectrum of applications and environments.

- ***The Clipper Group can be reached at 781-235-0085 and found on the web at www.clipper.com.***

About the Author

Anne MacFarland **is Director of Infrastructure Architectures and Solutions for The Clipper Group.** Ms. MacFarland specializes in strategic business solutions offered by enterprise systems, software, and storage vendors, in trends in enterprise systems and networks, and in explaining these trends and the underlying technologies in simple business terms. She joined The Clipper Group after a long career in library systems, business archives, consulting, research, and freelance writing. Ms. MacFarland earned a Bachelor of Arts degree from Cornell University, where she was a College Scholar, and a Masters of Library Science from Southern Connecticut State University.

- ***Reach Anne MacFarland via e-mail at Anne.MacFarland@clipper.com or at 781-235-0085 Ext. 128. (Please dial "128" when you hear the automated attendant.)***

Regarding Trademarks and Service Marks

The Clipper Group Navigator, **The Clipper Group Explorer**, **The Clipper Group Observer**, **The Clipper Group *Captain's Log*, The Clipper Group Voyager**, and *"clipper.com"* are trademarks of The Clipper Group, Inc., and the clipper ship drawings, "*Navigating Information Technology Horizons*", and "*teraproductivity*" are service marks of The Clipper Group, Inc. The Clipper Group, Inc., reserves all rights regarding its trademarks and service marks. All other trademarks, etc., belong to their respective owners.

Disclosure

Officers and/or employees of The Clipper Group may own as individuals, directly or indirectly, shares in one or more companies discussed in this bulletin. Company policy prohibits any officer or employee from holding more than one percent of the outstanding shares of any company covered by The Clipper Group. The Clipper Group, Inc., has no such equity holdings.

Regarding the Information in this Issue

The Clipper Group believes the information included in this report to be accurate. Data has been received from a variety of sources, which we believe to be reliable, including manufacturers, distributors, or users of the products discussed herein. The Clipper Group, Inc., cannot be held responsible for any consequential damages resulting from the application of information or opinions contained in this report.

Copyright © 2005 by The Clipper Group, Inc. Reproduction prohibited without advance written permission. All rights reserved.