



Instant Insight
February 12, 2010

IBM to Put Billions upon Billions of Files in the Cloud

By Clay Ryder

IBM has announced a new hardware and software offering, known as IBM Scale Out Network Attached Storage (SONAS), that assembles server and storage components into a virtual storage environment to overcome many of the scalability limitations of current NAS technology. SONAS is massively scalable with support for up to 14.4PB in a single system while offering the ability to seamlessly add storage capacity. It features automated tiering, is capable of scanning 1+ billion files in a matter of minutes, and uses policy-driven file-level tiering to enable organizations to specify where data is placed and/or how it is migrated over time. IBM states that the target market for SONAS is midsize to larger enterprises in industries that need to store, access, and manage dramatic growth in file-based data, including verticals such as financial services, insurance, banking, medical and life sciences, digital media and government.

Key business benefits outlined in the announcement include:

- ◇ **Better control of data** through consolidation of islands of data into a single network-based store that is accessible from anywhere in the world and can non-disruptively add storage capacity.
- ◇ **Smarter use of resources** achieved through policy-driven automation and tiered storage management.
- ◇ **Reduced Operational Costs** through hardware consolidation, reduced CAPEX, and maximizing of overall storage ROI as well as minimizing of administration headcount through streamlined and simplified administration, backup, and application access to data.

Net/Net

One simple IT axiom is that the number of files being created on a daily basis continues to grow, seemingly unabated. At the same time, the value of these files increases geometrically as they are made available in a policy-based fashion to an ever growing number of LoB professionals, business partners, customers, and even the world at large. Files represent information, the key component of the information-driven economy, and organizations that are cost-effectively able to capture this growing pool of data and transform it into actionable information are those which will have the competitive advantage in the marketplace.

Files almost seem to create themselves and expand to overfill the available storage space. Just ask any IT professional tasked with managing capacity just how long that “new” free storage capacity lasts after installation. Although for many organizations managing the utilization of servers has become paramount in the quest for organizational efficiency, not as many have fully embraced this reality with respect to storage. This is in part due to some additional challenges that storage management presents compared with servers. Although there are a few exceptions, notably horizontal applications such as email, or desktop productivity applications, most applications deployed in an organization are used by specific groups, departments, or other functional organizations. Thus it is possible to consolidate like applications within a subset of the corporate server resources and manage their level of service with some degree of predictability.

For the most part storage solutions have retained more of the traditional siloed approach than their server brethren, which would seem counter-intuitive given the focus on improved utilization through virtualization in most organizations. Yet part of the challenge lies in the fundamental difference between how storage is

consumed and how applications are. As we noted earlier, access to applications can be grouped by function with some degree of the physical manifestation mimicking this principle. However, file-based storage (as opposed to block-based storage) has essentially universal appeal to applications so a bevy of applications can end up making substantial demands for file space. In addition, there is the unprecedented growth in unstructured files of most any type finding their way on to network-based storage. The historic approaches that sought to deliver enterprise-wide storage, typically NAS in nature, are ultimately limited in the number of files that could be delivered within a given namespace. As space ran out, extensions to NAS required either that a potentially enormous number of files had to be moved to the new storage to keep associated files together, or users had to deal with their files being stored across multiple file namespaces and the inevitable management headache that entailed.

This is where we believe SONAS has the potential to greatly ease the management and accessibility of universal file systems and help move organizations towards a single universally accessible file store of considerable size. With SONAS, organizations can move towards Cloud-based infrastructures with the knowledge that the sheer demand for file space will not outstrip the ability of the storage infrastructure to keep up with rising demand while keeping the management expense under control. With the scale that Cloud portends, automated policy control and tiering will be absolutely essential in order for organizations to achieve the scale required while being able to match the business value of information being stored with the cost of its storage. Decisions on where data is placed upon creation, where and when it moves within the tiers of the storage hierarchy, where it is copied for disaster recovery, and when it will be deleted would be automated based upon the organization's defined policy guidelines.

While SONAS has the potential to deliver a single large file system in the Cloud, by no means would it become the sole storage approach for an organization or service provider. By virtue of its file system-based approach to storage, NAS (or now SONAS), is not the most efficient means by which to deliver block-oriented storage for applications such as large database instances or other applications requiring large quantities of raw storage that are under the application's direct management and control. Hence, even in a cloud environment, a one-size-fits-all approach will not best meet the needs of the organization.

Overall, we are heartened by this announcement as it illustrates a big-picture vision of how the vast number of files that could be found in a Cloud or very large enterprise environment could be managed in much more efficient, streamlined, and cost-effective fashion over past solutions. In line with the notion of efficiency, we are also pleased to read the very business-focused positioning of this technology, a discussion about operational efficiency and improved utilization as opposed to a singular fixation on the zillions of files that could be stored. In any large-scale deployment such as that implied by a Cloud Computing approach, the efficient management of resources, both human and capital, will be paramount for IT success. In the case of storage, this implies a high degree of automation, tiering, and policy definition and enforcement. To our way of thinking, SONAS is a logical extension to the well-understood NAS approach, but one that has the potential to allow NAS to grow far beyond its original design considerations and into its next wave of potential.