IDC Analyze the Future

IDC ANALYST CONNECTION

Noemi Greyzdorf
Research Manager, Storage Software

Success Strategies for Storage Data Migration

January 2011

There are numerous reasons why organizations periodically need to perform data migration. These migrations can often be costly, time-consuming, and disruptive. Without the proper tools, they can also require service outages and manual reconfiguration of client systems and applications. Therefore, it's important to consider intelligent file virtualization solutions that enable file-based data migration without downtime and without complex reconfiguration.

The following questions were posed by F5 Networks to Noemi Greyzdorf, research manager with IDC's Storage Software program, on behalf of F5's customers.

Q. What are some typical scenarios that involve data migration?

A. According to the Storage Networking Industry Association (SNIA), "migration or data migration" is defined as "a movement of data or information between information systems, formats, or media." The following are typical scenarios that may require data migration:

- **Decay of storage media.** A typical enterprise-class hard disk drive (HDD) has a life expectancy measured in mean time between failure (MTBF). Some of the drives on the market boast an MTBF of 2 million hours. A lower-class drive, just as a SATA drive, may have a lower MTBF of closer to 1 million hours. Drives of all types may fail for a number of reasons, such as manufacturing defects, mechanism wear and tear, overheating, or damage from rotational vibration. When a drive fails, data residing on the drive must be moved to another drive.

- **Hardware refreshes or vendor migrations.** This is one of the most common reasons for data migration. Typical storage systems have a life expectancy of three to five years on average. Once the hardware has reached its end of life, the data residing on such systems has to be moved to the newer system that is taking its place.

- **Changing performance requirements.** Recent budget constraints have driven organizations to look at how they may gain further efficiencies in their storage infrastructure. The idea that static data, which represents over 70% of all data (based on general industry consensus), can be moved off faster, more expensive disk drives to slower, less expensive drives has gained mindshare very quickly. The movement of data to a different tier of storage has become a desirable feature in datacenters large and small.

- **Consolidation of assets.** Server virtualization has gained momentum as a result of many organizations looking to consolidate assets and drive greater utilization rates. On the storage side, consolidation initiatives require data to be moved from many systems to a few systems. Such migrations are complicated because storage resources have to be allocated differently, giving little room for error.

Q. Why is it important to effectively manage data migrations?

A. Many organizations are struggling to keep up with data growth. Data is continuing to grow at over 50%, driven by the digitization of content, instrumentation data, and reliance on business communication and operation through the creation and sharing of data. Managing storage as infrastructure to support business processes that rely on this data is becoming expensive and complex. The ability to move data becomes a critical function of managing the infrastructure in support of business requirements.

A company's reliance on data to stay competitive while continuously operating puts stringent requirements on the availability of data and the reliability of storage. In such environments, data migration must be achieved seamlessly with minimal impact on operations. Downtime caused by data migration may result in the accrual of unnecessary costs and loss of revenue. To ensure that migration is accomplished without impact to the business, organizations must take the time to do a significant amount of planning prior to the actual movement of data. This is true for all major migrations.

In addition to major migrations that typically occur between storage systems, there may be minor migrations to achieve greater storage efficiencies and lower storage costs. For example, moving data from fast media to slower media for static data may result in significant savings in storage and overhead costs. For organizations to realize the benefit of such migrations, the act of moving data must not be intrusive or disruptive to the application, and the decision to move must be directed by policies automatically executed.

Q. What are the costs and complexities associated with data migration?

A. A typical process for migrating data between systems starts with the discovery of the current environment. Before any data can be moved, it is important to understand what data is out there, how it is being used or accessed, what storage resources have been allocated to each type of data or workload, and what the storage characteristics are of the current system such as performance in IOPS and throughput. Once there is a clear understanding of the current configuration, there need to be processes by which workloads and assets are reconciled. This implies determining which workloads continue to need the highest-performance storage versus those that might be migrated to other media.

The next step in the process is to configure the new system in such a way that it accommodates all the workloads being moved over. Once the plan for how data and workloads will be handled on the new system is determined, there are some organizational processes that must commence. The business units and IT must agree on when such migration will occur, typically a short window when the business unit may handle some downtime. Since the workloads on a single storage system often belong to more than one business unit, there is a lot of coordination that must occur.

Once the data has been copied over to the new system, tests must be performed to ensure that everything has migrated properly before the production applications are cut over to the new environment. The data is copied rather than moved over to allow for managers to roll back or go back to the existing environment if something goes wrong.

For many organizations, especially those with large shared storage systems, it may take three months or longer to complete the whole migration process so that the old system can be turned off. In addition to extensive resources that are typically dedicated to the planning process, there is a cost of having two systems run in parallel. Three months off the common 36-month life span of a system is a significant cost to pay.

Q. How can organizations mitigate some of the costs and complexities?

A. There are two ways the costs and complexities associated with the migration of storage systems can be mitigated: storage virtualization (block and file) and clustered or scale-out storage, often referred to as grid based with hot upgrades.

Storage virtualization is a way that numerous storage systems, block or file, can be abstracted from the hosts. The host is mounted to a volume or a file system and doesn't know or care how the system manages physical storage resources on the back end. The virtualization engine knows and tracks all data stored on the systems it represents. When data must be migrated off one system to another, the virtualization engine handles the process of moving data to the new system without any downtime, ensuring the new system is operating as intended, and automatically redirects users to the new system once it is ready. This abstraction of the physical systems expedites the movement of data to a new system, eliminates potential problems caused by human error, and minimizes impact to the business by accomplishing the migration without interruptions to the business.

Scale-out storage architectures use commodity components to scale both the performance and capacity of a storage system. These components are aggregated into a system through an intelligent software layer. In such systems, data is distributed across many nodes in a cluster of nodes. When a node needs to be decommissioned or a new node needs to be added, the software redistributes data across all available nodes and drives. The movement of data is done in the background without impact to the application or user.

There is also value in having a scale-out system deployed behind a storage virtualization layer. In this scenario, storage virtualization serves as the central point of management for all storage assets — a sort of storage broker — thus simplifying the overall management of the infrastructure.

Q. When does it make sense to use solutions such as storage virtualization and scale-out storage?

A. Both storage virtualization and scale-out storage are viable options for mitigating storage migration pains, but they are not mutually exclusive. The following are criteria to help determine when to leverage storage virtualization:

- If there are a number of workloads with different characteristics that require different storage systems, then storage virtualization is a more appropriate solution to broker the right resources for the need.
- If the environment consists of systems from a variety of vendors, then storage virtualization will help centralize management more effectively.

ABOUT THIS ANALYST

Noemi Greyzdorf is a research manager with IDC's Storage Software program. Ms. Greyzdorf's primary focus is to analyze and track the file systems and storage infrastructure management software markets. The key objective of her research is to provide insights to clients in the areas of market direction, end-user demands, positioning, messaging, and go-to-market strategies.