DATA MIGRATION BEST PRACTICES

NetApp Global Services

January 2006

Introduction

For IT managers, data migration has become one of the most routine—and challenging—facts of life. Workload balancing, technology refresh, server and storage consolidation, data center relocation, data classification, and mergers/acquisitions are all drivers to move data from one storage device to another on a regular basis. You'd think that any operation performed routinely would become easy. But not so. Migration—moving data from one device to another (especially as the capacity of storage devices continues to grow) and then redirecting all I/O to the new device—is an inherently disruptive process. With the increase in the percentage of mission-critical data and the proportionate increase in data availability demands, downtime—with its huge impact on a company's financial bottom line—becomes unacceptable. In addition, business, technical and operational requirements impose challenging restrictions on the migration process itself. Resource demands—staff, CPU cycles, and bandwidth—and risks—application downtime, performance impact to production environments, technical incompatibilities, and data corruption/loss—make migration one of IT's biggest challenges.

Since the majority of storage systems purchased by customers is used to store existing—rather than new—data, getting these new systems production-ready requires that data be copied/moved from the old system to be replaced to the new system being deployed. As the services arm of one of the industry's leading storage companies, NetApp Professional Services has therefore developed a core competency in data migration, fine-tuning its migration methodology over time, with each migration services engagement delivered. Whether the migration is a same vendor system upgrade (for example, NetApp FAS940 to NetApp FAS3050) or a vendor take-out (for example, EMC Celerra to NetApp FAS3020), whether the migration is performed by internal IT or an external services provider, the migration methodology is the same.

Migration Methodology

Although the term *data migration* is often used to refer to several different operations, this paper will focus on data migration as the process of moving data from one storage device (*source*) to another (*destination*), as in a technology refresh or an archiving ("HSM") operation.

All migrations follow the same basic methodology illustrated in Figure 1.

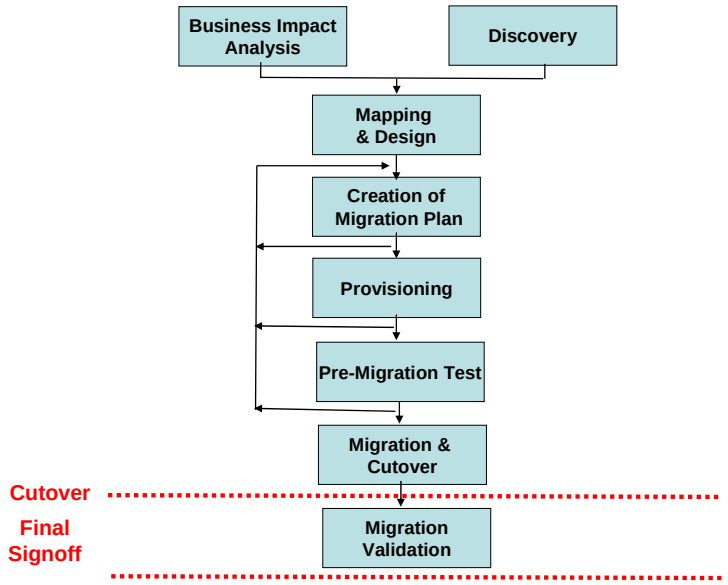

Figure 1) Migration methodology.

Although the ultimate goal of a data migration is to move data, a lot of upfront planning needs to happen prior to the move in order to ensure a successful migration. In fact, planning is the number-one success factor for any migration project, independent of the complexity. Not only does upfront planning help shorten the duration of the migration process, but also it reduces business impact and risk—for example, application downtime, performance degradation, technical incompatibilities, and data corruption/loss. The migration plan—the end result of the planning—defines what data is moved, where it is moved, how it is moved, when it is moved, and approximately how long the move will take.

There are a few key environmental factors that influence the migration methodology selected—how the data is moved.

The first factor is the number of tiers within the storage hierarchy that are included in the destination.

Single-tier migration is a migration in which all source data is migrated onto one or more devices within a single tier—that is, primary storage. Within a single-tier migration, three other factors are important in determining the methodology:

- Storage vendor:

  Homogeneous: one vendor

  Heterogeneous: multiple vendors

- Data layout:
  - One-to-one mapping: source and destination have the same storage layout

    - Relayout: source and destination (one or both of which can be one or more devices) have different storage layouts

- Access protocol:
    - NAS (file)
    - SAN (block)

Multitier migration is a migration in which data is migrated to different tiers of the hierarchy based on a best fit between data/application requirements and storage tier attributes. There are two types of multitier migration:

- One-time migration: data placement recommendations, based on the results of an analysis of the data and the storage tiers in the hierarchy (primary filer, ATA filer, and NearStore®), are provided
- Continuous and policy-driven migration: unlike all other data migrations, this is a continuous process (requiring the installation of an archival application) that uses a set of user-defined policies to determine, in real time, when data needs to be moved where

Figure 2 illustrates the basic migration environments defined by these key factors.

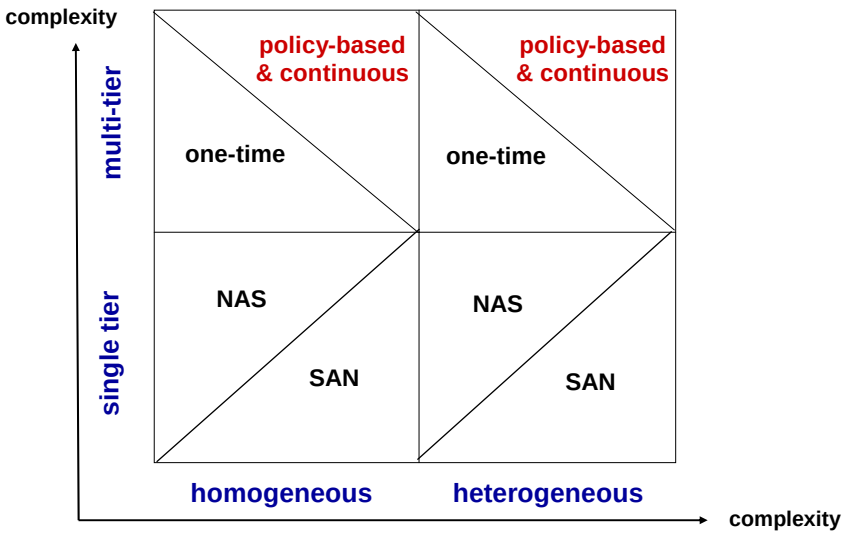

Figure 2) Migration environments.

All migrations fall into one of the above migration environments.

Business Impact Analysis

The objective of the business impact analysis is to identify the business and operational requirements that impact the migration process. Various stakeholders within the company need to be consulted to ensure that their requirements are factored into the migration planning.

- IT staff defines available network bandwidth and CPU cycle load, allowable downtime, and migration schedule

- Database and system administrators define application and database requirements
- Business owners define the importance of, and requirements for, specific applications and types of data
- Security and compliance groups define compliance requirements

Discovery

Details about the migration hardware and software environment are collected during the discovery step. The migration environment—whether a heterogeneous SAN migration or a homogeneous NAS migration—determines the specific data that has to be collected. For example, in a NAS migration environment, current I/O workload (IOPS) on the servers, number of files/folders/volumes involved in the migration, share permissions (NFS, CIFS, or both), network configuration, DNS and AD domain information, and so on are important. A SAN migration environment has different discovery requirements. Accurate and complete discovery is critical to the success of the migration, so although discovery can be manual (via command line) or automated (via a discovery tool), using an automated discovery process is the preferable option. Not only does automated discovery reduce discovery time, but also it minimizes the chance of error. Unfortunately, no one discovery tool provides all the required data, so often multiple tools must be used.

Mapping and Design

During discovery, the migration source—what data is moved—is identified. During mapping and design, the second half of the equation, the migration destination—where is the data moved—is identified. There are two basic mapping layouts: one-to-one, where the source and destination layout are the same, and relayout, where the source and destination layouts are different. Although a one-to-one mapping enables a much simpler migration, migration is often seen as an optimal opportunity to consolidate and/or optimize performance and/or capacity utilization, so relayout is a very common scenario. A combination of the migration goals and best practices drives the layout of the new (destination) storage environment.

Creation of Migration Plan

The migration plan, which is the end deliverable of the planning phase, functions as the blueprint for the migration implementation, specifying customer expectations, defining project deliverables, and identifying migration methodologies to be used. There are four major inputs into the migration plan:

- Business and operational requirements, which provide the constraints
- Data to be migrated, with all associated attributes
- Available migration tools
- Storage and application best practices

Creating a workable migration plan is often quite challenging. Different types of data may require different migration tools and strategies, and business and operational requirements—the downtime window—may require creative ways of moving the data.

The migration plan components should include, but are not limited to:

- Migration strategy and key activities
- Dependencies
- Required equipment and migration tools
- Customer expectations (customer business, technical and operational requirements)
- Test plan

- Verification procedures
- Risks and contingency plans
- Change control procedures
- Project schedule
- Postimplementation activities/responsibilities
- Migration completion criteria

Although the migration plan is the end deliverable of the planning phase, it is really a living, rather than static, document. Variables in the environment can change, or execution can lead to unexpected results, impacting the migration plan as documented.

Provisioning

During provisioning, the destination storage environment is prepared for the data move. LUNs, volumes, directories, and so on are allocated, security attributes are set, and shares/exports are created. Provisioning for a one-to-one mapping is simple; for a relayout, it is a more complex task. However, by using information generated from the mapping and design step, it is possible to automate many of the provisioning tasks.

Premigration Test

Before any data is moved, it is important that some portion of the migration plan—with the scope depending on the specific situation and the agreement with the customer—be tested and validated. Results of the migration test determine whether modification of the migration plan—for example, timeline, migration tools used, amount of data migrated per session, and so on—is required. For example, if testing shows that allowable downtime would probably be exceeded, the migration methodology needs to be revisited.

Migration and Cutover

This is the point within the migration process when the source data can be moved to the destination devices.

There are many options for moving data, each with its own advantages and disadvantages:

- Migration utilities can run on the host, storage device, or network appliance/switch
- Data can be moved in two basic ways:
    - Out-of-band (or out of the data path), by creating a baseline copy, moving the baseline copy via media transfer or electronically, and then applying incremental changes, redo logs, and so on
    - Inband (or in the data path), by creating a mirror in real time and then breaking the mirror

The optimum migration methodologies vary significantly, based on many factors, including:

- Source and destination storage systems
- Network storage topology (NAS or SAN)
- Equipment locations
- Specific applications and data usage
- Customer business, technical, and operational requirements
- Available/supported migration utilities

Many vendors offer products that play in the migration space, so there are many choices.

Basically, there's no one right way; it's all about making trade-offs based on knowledge of the environment, business and operational requirements, and migration experience, which provides a proven set of best practices. In fact, the entire planning phase focuses on developing the best-in-class migration plan for the specific migration project at hand. By migration and cutover, all the upfront planning, analysis, and trade-offs have been made; the migration plan serves as the implementation guide.

Once the data has been moved, all clients must be redirected to the destination devices.

Migration Validation

Before the migration can be considered a success, one critical step remains: to validate the postmigration environment and confirm that all expectations have been met prior to committing. At a minimum, network access, file permissions, directory structure, and database/applications need to be validated, which is often done via nonproduction testing.

Internal or External?

The disadvantage of doing a task over and over is that it can get pretty boring; the advantage is that you fine-tune your procedures, learn to recognize the pitfalls and develop workarounds, and become proficient—in other words, you become an expert. Although data migration has become a routine operation within IT, when migrations are complex, new technology is being introduced into the environment (for example, NetApp storage or FlexVol™ technology), or migration skills are in short supply, bringing in an external services provider makes a lot of business sense.

NetApp Professional Services complements your internal capabilities, working with all the key stakeholders to understand goals and requirements and providing

- Migration expertise and field-proven methodology and tools
- Incorporation of best practices
- Risk mitigation
- Professional project management throughout the migration process lifecycle, including:
    - Creating a statement of work (SOW) and obtaining customer approval
    - Identifying the various customer stakeholders
    - Establishing a migration team, composed of both customer and service provider representatives
    - Creating a migration plan and obtaining customer approval
    - Providing migration progress reports
    - Communicating issues and their impact and recommending alternatives as required
    - Obtaining customer project acceptance based on predefined completion criteria

Data migration is—and will continue to be—a way of life for IT as workload balancing, technology refresh, server and storage consolidation, data center relocation, data classification, and mergers/acquisitions drive the need to move data from one storage device to another.