Abstract

This paper outlines the recommended approach to plan an Isilon storage environment for a VMware vSphere virtual data center that satisfies storage capacity and performance requirements. Steps to gather and document capacity and performance requirements are outlined, along with a recommended process to determine the appropriate Isilon storage configuration.

February 2012
# Table of Contents

**Executive summary** ................................................................. 5  
Audience .................................................................................. 5  
Additional Resources .................................................................. 5  
Revision Summary ..................................................................... 6  

**Introduction** ........................................................................ 7  

**Capacity Planning Guidelines** .................................................. 7  
Understanding and Measuring Workloads ................................. 8  
Performance Basics .................................................................. 8  
Data Access Patterns .................................................................. 9  
Data Operation Types and Management ..................................... 9  
Sizing Methodologies and Risks .................................................. 10  
Performance Factors for Virtualized Workloads ....................... 10  
Risk Factors for Virtualized Workloads ..................................... 11  

**Advantages of EMC Isilon Scale-Out Storage** ....................... 11  
Scalability .................................................................................. 11  
Simplicity ................................................................................... 11  
Automatic Balancing and Bottlenecks ...................................... 12  
Data Protection .......................................................................... 12  
Efficiency to Drive Down Costs ................................................ 12  

**Workload Planning, Capture and Analysis** ............................. 13  
Gathering Performance Data ..................................................... 13  
Planning for Data Collection .................................................... 13  
Scope of Capture ....................................................................... 13  
Duration .................................................................................... 14  
Metrics Gathering ....................................................................... 14  
Recommended Metrics for Capture ......................................... 15  
Capture Interval(s) ..................................................................... 15  
Performance Capture Tools ...................................................... 16  

**Metrics Analysis** ................................................................. 17  
Capture and Analysis Risks and Issues ...................................... 17  
Compiling and Assessing Results .............................................. 18  
Metrics Interpretation ............................................................... 18  
Analysis Methods ....................................................................... 19  
Additional Performance Considerations .................................. 20  
Data Access Patterns ............................................................... 20  
Backups ..................................................................................... 20  
Integration with Existing Workloads ......................................... 20
Executive summary

This Sizing Guide provides technical information to consider when planning a VMware vSphere™ infrastructure built on a foundation of Isilon™ scale-out NAS storage. Its fundamental intent is to help ensure that the Isilon storage configuration implemented will meet the overall capacity, availability and performance requirements of your combined virtualized workloads. This Sizing Guide also provides a recommended methodology to gather, compile, and analyze relevant storage-related metrics for all applications and services in the environment that are targeted for either migration to—or deployment on—the vSphere infrastructure.

While no two environments are alike and no single approach is appropriate for every organization, this Sizing Guide provides a useful capacity and performance planning tool in typical vSphere environments. For more individualized planning requirements, in addition to using this Guide, EMC recommends enlisting the services of an appropriately-qualified external party, such as an authorized EMC reseller, your EMC account team, or an EMC Isilon Services engineer. Doing so will help ensure that your Isilon cluster is optimally configured to match your specific virtualized workloads.

Audience

This Sizing Guide is intended for IT executives, managers, architects, and engineers who are either considering, or are currently engaged in, a virtualization initiative that will utilize Isilon storage as the underlying platform for your virtualized infrastructure.

It is assumed that readers of this Guide have a working knowledge of the following:

- Consolidated storage technologies, including Storage Area Network (SAN) and Network-Attached Storage (NAS)
- Data integrity protection in consolidated storage environments
- Isilon scale-out storage, and the OneFS™ operating system from EMC
- Enterprise Capacity Planning
- VMware vSphere virtual infrastructure components (either vSphere 4.x or vSphere 5.0) appropriate to your organization and initiative
- Spreadsheet applications, including chart creation using raw data sets, for performance and capacity analysis

Additional Resources

Before using this Guide, it is recommended that you review the following materials:

- The VMware vSphere 4.1 Documentation Portal, or the VMware vSphere 5 Documentation Portal, as appropriate for your organization
- Isilon Scale-out Storage for Virtualization
• Solution Brief: EMC Isilon Virtualization at Scale
• Insight from Isilon: Understanding the Role Storage Plays in Virtual Environments

Revision Summary

<table>
<thead>
<tr>
<th>Date</th>
<th>Version</th>
<th>Author</th>
<th>Change Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td>February 2012</td>
<td>1.0</td>
<td>James Walkenhorst</td>
<td>Initial Document</td>
</tr>
</tbody>
</table>
Introduction

In the past, most enterprise IT applications had specific, distinct performance requirements that did not change significantly over time. To meet the demands of varying applications and workloads, a wide variety of storage products and platforms, grouped into three major categories, materialized on the market:

- Traditional, direct-attached-storage (DAS)
- Network-attached storage (NAS)
- Storage-area network (SAN) systems

All three approaches have typically used a scale-up model, characterized as monolithic systems controlled by a single processing unit, with fixed performance and capacity limits.

With the application landscape transforming into distributed environments, and with the tremendous cost savings offered by virtualizing workloads onto consolidated platforms (servers, network, and storage in particular), the old, static model of resource planning no longer applies. One application in your environment may experience abrupt growth at a much higher rate than other workloads, but due to competing business demands, the growth must be planned and managed in a manner that minimizes the adverse impact on other applications and services that share the same infrastructure.

In a centralized, virtualized environment, the ability to effectively provision those consolidated resources to a number of applications and workloads concurrently requires that the underlying infrastructure be sized to meet the performance demands of multiple workloads simultaneously.

This guide is intended for planning an Isilon storage infrastructure appropriate to your vSphere platform.

Note: Planning for the other necessary vSphere-related components—host configuration, vCenter server, network infrastructure—is not included within the scope of this white paper.

Capacity Planning Guidelines

Capacity planning, at a high level, entails scaling your infrastructure resources—including network, storage, and processing capacity—to accommodate the multiple, competing demands of the workload(s) you will place on them. In the case of a vSphere infrastructure that utilizes an Isilon storage cluster from EMC, you will need to ensure that you have sized your storage to meet the necessary performance objectives for both disk throughput and network throughput. Most virtual environments also have stringent availability requirements, both to protect the integrity of the underlying data and the ability of the infrastructure components to
provide continuous access to that data in the event of a hardware or connection failure.

To determine your specific disk performance and availability requirements, a thorough analysis of your expected workload should be performed, with the resulting data used to determine the size and specific configuration of your Isilon cluster, as well as to assist in planning out and scaling your network and vSphere host infrastructure to support that workload.

Since every organization will have different requirements, and since most Isilon clusters will have been customized to meet those unique requirements, there is no single rule for optimizing your Isilon cluster for availability and performance to satisfy all vSphere environments. Instead, this guide provides a useful framework for gathering and analyzing your specific needs, and then working with your Isilon storage specialists to design a specific solution to satisfy your particular requirements.

**Understanding and Measuring Workloads**

Planning and measuring the capacity characteristics of storage systems can be challenging across many fronts. In addition to the simple question of how much raw disk space is needed to satisfy the capacity requirements of the intended mix of workloads, planning a storage environment to support a virtualized infrastructure requires knowledge of the performance needs of the multiple virtualized workloads that will compete for, and depend on, that storage. It also requires an understanding of the inherent abilities and limitations of enterprise storage technology to satisfy those needs.

Different aspects of storage architecture and data access should be considered when evaluating storage system performance. Because different application workloads have very different data access patterns, different storage systems may vary widely in their ability to handle concurrent operations: latency-bound operations, sequential-access operations, and random-access operations.

**Performance Basics**

Storage performance is typically measured in terms of either *throughput* or *latency*. Throughput is measured by calculating the average rate of data transfer, (e.g. megabytes per second, or MB/s), or by quantifying the number of operations performed (measured in Input/Output Operations per Second, or IOPS). Latency is measured by the time interval between initiating a single I/O request and receiving a response/acknowledgment to that particular request.

Throughput requirements can be calculated by either analyzing or predicting how data is accessed during both read and write operations. Latency can be addressed by minimizing the round-trip time for each data access operation, either by using low-latency network equipment and storage, or by using data buffering to minimize the overhead of latency-heavy transactions within the data path. It follows, then, that different storage configurations produce different latency and throughput performance results.
Data Access Patterns

Random access is characterized by small-block I/O operations across the dataset. It is typically expressed in terms of IOPS. Since random operations cannot be predicted ahead of time, the current data location (either in memory or on disk) has a direct impact on the speed in which an I/O operation is completed. When data is cached in memory, I/O operations are much faster than when data is stored on disk. It follows, therefore, that the amount of cache memory and the speed of the underlying disk drives are an important consideration when optimizing for IOPS throughput. The CPU power of the storage controller(s) is also critical. As the number of data transactions and the associated metadata calculations increase, an underpowered storage controller can be constrained by CPU bottlenecks that limit the number of I/O operations a storage system can sustain.

Sequential access patterns are characterized by large data transfers to contiguous sections of a file over a single connection. Application workflows requiring single-stream sequential-data access rely on fixed-bandwidth performance to guarantee data can be transmitted at an appropriate rate for the application. Applications that generate many concurrent connections and access many small files, such as are often found in virtualized environments, are another example of sequential-access-pattern workloads.

Applications producing unbound, sequential I/O operations need to maximize the bandwidth between the storage system and the application in order to move data as fast as possible. Due to their need for high-bandwidth data access, these applications benefit from having high-bandwidth, low-latency network links—such as dedicated 10Gb Ethernet connectivity—and dedicated storage resources.

Data Operation Types and Management

Fundamentally, there are two primary types of data operations on file-based storage systems: file-data operations, and file-system-metadata operations. File data operations refer to accessing or modifying the physical blocks that comprise the data stored in a file. Metadata operations refer to access to filesystem information such as file location, filename, create time, modify time, ownership, access permissions, etc., that are managed by the filesystem.

In most organizations, storage-based workflows are a mix of data-access patterns, file-type operations, and file sizes. Because traditional storage systems differ greatly in the way they support these distinct access patterns and data types, storage configuration and sizing becomes a complex and error-prone process.

Traditional scale-up storage systems typically consist of a multi-layer architecture with filesystem, volume, and RAID management layers. The filesystem, managed by the client or by a dedicated server, is bound by the CPU processing power and filesystem size limit. The RAID management layer is bound by the RAID controller’s processing power, throughput capacity, and by the number of spindles in each RAID group. The volume-management layer is mainly an abstraction layer that resides between the other two.
Sizing Methodologies and Risks

The traditional approach to sizing a storage array required either upfront knowledge—or a reasonably accurate compilation of estimates and assumptions—of the workload profiles, the quantity of data needed for each workload, and the overall performance requirements for all dependent applications and services—over the expected 3-to-5-year lifecycle of the array.

If the estimated performance requirements were overstated initially, then the resulting storage array would very likely be overbuilt. The net effect is a potentially significant waste of scarce IT funds to meet requirements that never materialize.

On the other hand, underestimating any of the components of storage-based workloads—capacity, access-pattern settings, throughput and latency requirements—over the lifetime of the array could produce even worse results: an undersized or underperforming storage array that requires a forklift upgrade to remedy, at significant cost and disruption to the organization.

Performance Factors for Virtualized Workloads

When planning a storage configuration that will satisfy the demands of a virtualized environment, an enterprise will have its own unique variables that should be identified and evaluated. Applications that are candidates for virtualization will have different workflow profiles: some applications demand high performance for random-access-pattern activity, while other applications will need sequential-access-pattern support from the underlying storage array.

In situations where existing workloads are being migrated to virtual hardware, it is possible to compile and evaluate the applications’ workload profiles (including peak-load times and the CPU, memory, disk, and network footprints necessary to sustain those workloads), and then begin to plan a virtual infrastructure that will support that migration. In other circumstances, however, such as a scenario wherein a target workload has not yet been deployed in production for the organization, an accurate assessment of the application’s performance data—and its overall infrastructure requirements—cannot be known with certainty. In those cases, a reasonable estimate may be the only available solution.

A common use case involves virtualizing a combination of existing applications and deploying new applications on the same vSphere infrastructure. In this scenario, EMC highly recommends a phased approach to migrating services to the new vSphere infrastructure, including a proof-of-concept phase to validate the fundamental design and performance characteristics of the vSphere solution.

Every environment’s needs and challenges will be different, so while this guide attempts to identify the most common factors, it should not be considered a fully comprehensive list. Where possible, EMC recommends engaging the services of either your EMC/VMware/hardware reseller, or another trusted consulting-services organization with a proven record of planning and deploying vSphere environments, when planning your virtualized infrastructure.
Risk Factors for Virtualized Workloads

In virtualized environments, the stakes for availability and performance are higher than many traditional workloads. If a server or device fails, the applications, services and data sets that depend on that component can be either lost or corrupted. Moreover, since a key goal of virtualization is to improve system efficiency, a principal design objective for most organizations deploying virtual infrastructures is to maximize system utilization without compromising system availability.

Advantages of EMC Isilon Scale-Out Storage

Built of modular nodes that offer varying degrees of performance-to-capacity ratios, and powered by an intelligent, distributed filesystem that pools all available resources together to maximize availability and performance, an EMC Isilon cluster is designed to address the storage needs for almost any kind of application workload. Specific advantages of Isilon storage are described below.

Scalability

An Isilon cluster consists of between 3 and 144 independent storage nodes. Each self-contained node comes with its own internal disk drives, multi-core CPUs, memory, and network connectivity. As additional nodes are added to a cluster, all components of the cluster scale symmetrically and predictably. Every node adds its collective power to a single unified storage system, one that can tolerate the failure of any of its constituent components: disks, network interfaces, even entire nodes.

Additionally, the Isilon storage architecture simplifies the hardware-refresh cycle by allowing you to easily add and retire nodes from the cluster while keeping the cluster itself online, and the data available, throughout the refresh process.

Simplicity

Organizations with virtualized infrastructures can simplify their storage infrastructure while at the same time positioning their storage environment to scale out to levels not supported in other storage architectures. Up to 15PB\(^1\) can be hosted in a single filesystem, providing better alignment with the dynamic nature of virtualization.

At the same time, pay-as-you-go scalability, combined with the enhancements to storage utilization, allow IT organizations to consider a simpler, more cost-effective approach to purchasing, installing, and provisioning storage. The capability of the storage system allows it to scale for capacity and performance simultaneously, or independently, resulting in an agile storage environment that can expand quickly to changing application workloads.

For example, if a given workload is constrained by inadequate disk I/O throughput, more platform nodes can be dynamically added to the cluster. This has the effect of

---

\(^1\) Based on OneFS 6.5.x.
increasing both the number of available spindles and the overall global cache of the cluster, and improves both performance and capacity almost immediately.

**Automatic Balancing and Bottlenecks**

A new storage node can be added to an Isilon cluster in less than 60 seconds, eliminating the need to over-provision storage for future requirements. As soon as the node is turned on and network cables are connected, Isilon’s AutoBalance™ feature immediately begins migrating data from the existing storage nodes to the newly added node, quickly re-balancing all of the data across all available nodes in the cluster.

This automatic rebalancing ensures the new storage node will not become a bottleneck for new data, and that existing data will gain the benefits of the increased resources now available to the storage system. AutoBalance is completely transparent to both end user and storage clients, and can be adjusted to minimize impact on high-performance workloads. Scaling the file system up, even to its limits, requires no management time from the administrator.

**Data Protection**

An Isilon cluster is designed to tolerate one or more simultaneous component failures without any interruption of service or loss of data. Data protection is applied at the file level, rather than relying on either a global setting or specific RAID groups as found in other storage platforms. In the event of a component failure, the file system is able to focus on recovering only the impacted data rather than having to repair and rebuild the entire dataset.

Because all data, metadata, and parity information is distributed across all nodes in the cluster, a dedicated parity node or drive is not needed. This eliminates any possibility of a single point of failure, since all nodes share equally in the workload and function as peers in a complete symmetrical storage environment.

Also, unlike other storage platforms, changes in data-protection settings for files and directories can be modified at any time without the need to take either the cluster or the impacted files offline. Service outages, which are often required in traditional storage environments when re-balancing workloads, are also eliminated with Isilon scale-out NAS.

**Efficiency to Drive Down Costs**

Isilon storage systems provide highly efficient data storage at scale for virtualized environments. For clusters larger than five nodes, utilization rates of over 80%—without degrading performance—become routine for many organizations using Isilon storage. Isilon’s globally coherent cache and SmartConnect™ client load-balancing provide high performance with industry-leading levels of data protection, high availability and system resiliency. Complex storage management tasks, including troubleshooting disk or network bottlenecks, or migrating VMs between datastores to rebalance capacity, are also eliminated with Isilon storage.
Workload Planning, Capture and Analysis

At its most basic level, sizing a vSphere storage environment for performance is simply a matter of following these steps:

1. Gather and analyze the necessary metrics from your in-scope applications and systems
2. Identify both peak and average sustained workload levels, and times for each application
3. Compile and document the net requirements from all workloads that will run concurrently on the vSphere virtual data center
4. Design your vSphere storage infrastructure to support those levels.

This section will review the options and decisions to assist you in determining an appropriate storage solution for your vSphere environment.

Gathering Performance Data

Broadly speaking, the purpose of data gathering is to determine the level of performance that your workloads will require if they are to be successfully virtualized. Miscalculating any of the constituent resource requirements in a vSphere infrastructure—hosts, network, storage, etc.—can lead to widespread adverse consequences for the entire virtualized environment, as well as the applications, business processes and end users dependent on that environment.

Determining performance requirements in advance, then, is not just an exercise in due diligence. It is also a critical first step toward ensuring that whatever virtualized infrastructure is deployed is effective, both in terms of the availability and performance it delivers and in terms of its cost to your organization.

Planning for Data Collection

A successful data-gathering phase requires that several decisions be made beforehand, particularly relating to the size of the collection phase, its duration, the level of detail, and the tool(s) appropriate to gathering and analyzing the results.

Scope of Capture

Determining how broad an assessment is necessary will be one of the first steps in creating a plan for gathering data.

Access to Target Workload

The size of the assessment may be determined in large part by how much of the target workload is actually available for review before the virtualized infrastructure has been built. Ideally, all workloads that are targeted for virtualization will be already online and available for analysis. In practice, however, having all target workloads available for gathering and analyzing performance metrics is rare. In some cases, one or more workloads that are in-scope for virtualization have not yet been fully deployed in your environment, so an accurate performance data sample is difficult to
obtain. Other applications and services may already be running in production but have so many nodes necessary for providing the service—e.g. end-user desktops—that gathering a comprehensive data set for every node is not feasible in the available timeframe.

Workload Components
Each workload may be composed of multiple constituent components, such as one or more database servers, middleware components, web servers, application-layer servers, and possibly even client services. For a more comprehensive dataset, and a better idea of the overall performance requirements for each workload, the scope of assessment should include all of its component systems—even if one or more of those components are not under consideration for virtualization. The resulting data will provide a useful idea of not just the I/O requirements for each individual target system, but also the overall I/O data access patterns and storage-performance requirements for the entire workload.

If there are multiple instances of a particular application component to be analyzed—web servers would be a useful example—then it may be feasible to collect metrics on a subset of those servers, rather than the entire group. In many cases, the results can be reasonably extrapolated to the larger group to determine overall requirements.

The overarching objective of the assessment exercise is to gain an adequate understanding of the collective performance requirements for all in-scope applications and their dependent business processes. There are no universal rules for ensuring that your sample is of sufficient size to provide adequate data. Generally, however, larger datasets are preferable to smaller ones, and a final infrastructure plan is likelier to be successful if it is based on comprehensive information from your targeted workloads.

Duration
The available timeframe for data collection is another significant factor in planning the assessment. If possible, the duration of the data-gathering phase should correspond to the business cycles of the in-scope applications and services. If, for example, an application runs on a quarterly business cycle, then an accurate determination of the level of performance required from your vSphere infrastructure by that application, particularly under peak loads, would require that the assessment phase run for at least one business quarter. Shorter business cycles may allow for shorter data-collection windows, but once again the aim is to compile a comprehensive picture of the virtualized applications’ long-term performance needs.

Metrics Gathering
Since the focus of this document is to assist in designing the appropriate Isilon cluster configuration for your vSphere environment, this section will focus on capturing storage-specific performance data. Additional information about other metrics is available from other sources.
Recommended Metrics for Capture

Preferably, the following storage metrics should be collected for every application that is in-scope for deployment on your virtualized infrastructure:

- Logical Disk: % Free Space
- Physical Disk: Avg. Disk sec/Read
- Physical Disk: Avg. Disk sec/Write
- Physical Disk: Disk Read Bytes/sec
- Physical Disk: Disk Reads/sec
- Physical Disk: Disk Transfers/sec
- Physical Disk: Disk Write Bytes/sec
- Physical Disk: Disk Writes/sec

Note: the interpretation of the above counters is provided in the Metrics Analysis section below.

If the server(s)/workstation(s) whose metrics are being captured are currently configured with multiple physical disks, then the above metrics should be captured per disk in order to accurately gauge the total required disk footprint. The % Free Space metric should be included for every logical disk on every system in the capture to assess the necessary collective storage capacity for all systems in the capture.

Note: The above disk-performance counters assume that the target system is hosted on a Windows-based platform. The list of metrics that can be collected in your environment may be determined at least in part by the tool(s) used to gather data, and by whether the metrics are available within each in-scope application or its underlying operating system. Regardless of the tool and platform, the above disk-related metrics should have equivalent counters that can be captured and analyzed on other operating systems and/or application suites.

Capture Interval(s)

Since the act of monitoring and collecting performance metrics can generate significant overhead on the target systems, care should be taken regarding the frequency with which performance counters will be captured. If statistics are gathered too frequently, then the resulting data sample will grow large and unwieldy very quickly. Additionally, depending on how the performance capture process is configured, the target systems may experience performance degradation as a result.

The target capture interval should be large enough that the overall impact on disk space and performance is not unduly large, but frequent enough to capture both performance trends and spikes. Setting the capture interval at 15-30 minutes should be sufficient for most workloads.
Performance Capture Tools

At the same time, determining which performance-capture and analysis tool(s) are appropriate depends on a number of factors. Some tools require a license and maintenance purchase, or specific subject matter expertise, or—as with VMware’s Capacity Planner™ tool—a solution-provider partnership for which your organization may need to plan and budget as part of its virtualization initiative costs.

Enterprise Monitoring Applications

Many organizations already have enterprise availability and monitoring tools in place. If available within your organization, these tools offer the advantages of proven compatibility with your existing environment and in-house subject matter expertise to support the monitoring application and to manage the resulting datasets after collection. Depending on how long the enterprise-monitoring application has been installed and with what features, it may also provide the benefit of historical performance-trending data for one or more of the in-scope applications slated to run on your vSphere infrastructure.

A caveat to relying exclusively on existing data from an established application, however, is that the specific performance counters relative to storage may not have been captured historically. Additionally, some applications may not retain historical data for a sufficiently long period of time to validate against an entire business cycle for the workload, e.g., they may be able to provide only 30-60 days’ worth of performance data for an application that runs on a quarterly business cycle.

If you plan to use existing data from your organization’s enterprise monitoring application(s), then for best results you should verify beforehand that the data exists over a sufficiently long timeframe, and that the dataset includes the necessary disk-level metrics for determining your storage performance requirements. If both of these conditions are not satisfied, then your enterprise monitoring data may not be sufficiently comprehensive for planning your virtual infrastructure requirements, and you may need to find an alternative method.

Having an enterprise monitoring platform already in place may enable you to simply add the workstations, servers and metrics you want to gather to the enterprise application’s management configuration for the duration of the capture window. This may require additional licensing costs, however, so be sure to check your existing license agreement before proceeding.

Native Monitoring Tools

Different applications and operating systems often install with native performance monitoring tools. SQL and Exchange are bundled with application-specific counters that integrate with Performance Monitor in Windows. Most Linux distributions use iostat. Many of these tools are capable of monitoring and capturing the recommended disk-performance and capacity counters.

Some of those same tools are also capable of simultaneously capturing relevant performance metrics for the CPU, memory and network. Since a successful
virtualization infrastructure design will depend on being able to accurately analyze all of the various performance requirements, the ability to combine and cross-reference counters and data from all sources is essential.

It is possible that, for reasons of compatibility, licensing, subject matter expertise, etc., your organization will need to deploy multiple tools to capture and compile the necessary metrics to gain a full picture of your performance requirements. EMC recommends choosing tools that will allow you to export their datasets into a common format, which you can then use to cross-correlate performance requirements across all in-scope applications.

To compile the bigger picture of what will be required of your virtual infrastructure under periods of peak workloads, you should be able to correlate performance data for all in-scope applications within the same timeframe.

**Metrics Analysis**

Once the capture phase is complete, the next step is to assemble the resulting data sets from each captured component of each targeted workload. This will allow you to build an accurate assessment of the total resources required for your vSphere infrastructure.

With regard to reviewing storage metrics, using a commonly accessible format for compiling the results, such as a database or comma-separated-values (CSV) file, should enable the creation of performance charts for the various counters and metrics from each collected dataset. By combining the complete set of results, you should be able to gauge the storage capacity and the level of throughput required to satisfy the collective demands of your virtualized workload.

**Capture and Analysis Risks and Issues**

In practice, impediments to collecting the relevant metrics can make it difficult to perform a complete and accurate collection.

- In many cases, the application(s) or system(s) to be virtualized have not yet been deployed in the organization. While the identical application may exist elsewhere, and may even have already been successfully virtualized elsewhere, variables between organizations negate the possibility of a direct extrapolation of performance characteristics and requirements from one environment to the next.

- The scope of the virtualization effort may be so large, or the timeline available for gathering and analyzing performance metrics so small, as to make it impossible to gather even a meaningful sample. An example would be an application that runs on a quarterly business cycle, but there are only a few weeks available in which to gather the necessary data.

- The tools available for gathering data may have security or compatibility issues with the in-scope workloads.
Even with accurate reports of disk-read-vs.-disk-write operations, a reliable measure of whether disk access is based on random-access requests or sequential-access requests does not currently exist.

Even in circumstances where the workload performance characteristics and metrics can be successfully gathered, if the resulting data is collected in isolation from the rest of the environment, then an accurate assessment of how the application(s) or system(s) will perform under load in a mixed environment can be very difficult.

There is no universal solution to resolve these issues. By anticipating them, however, you may be in a better position to select capture tools that minimize the risks, or you may be able to adjust for them, or to plan sufficient time into the collection window to troubleshoot and resolve these issues while leaving enough time to conduct an adequate performance survey.

**Compiling and Assessing Results**

Depending on the tool(s) used for the data collection, the analysis of the results can take one of several paths. If an enterprise monitoring tool was the sole method of data collection, then the compilation and interpretation of the data can likely be accomplished through the tool’s interface.

If the tool used doesn’t include the ability to aggregate the performance data from all captured systems, or if multiple tools were used with different interfaces, then the ability to export the datasets to a CSV file will still allow you to plot out the performance metrics over time.

Whichever tool(s) you use to compile and interpret the results, analyzing and interpreting the data will require a similar approach. Remember the target objective is to create a full picture of the expected load that will be placed on your Isilon storage by your virtual infrastructure.

**Metrics Interpretation**

Having collected the appropriate storage metrics, you can use the following guidelines for interpreting the results.

- **% Disk Free Space**: By comparing this counter against the rated capacity for each monitored volume in the collection, the disk-space requirements for each VM can be calculated. This counter measures useable space, not raw space, so in a centralized storage environment, additional capacity will be required to satisfy the data-protection requirements for each volume.

- **Avg. Disk sec/Read, Avg. Avg. Disk sec/Write**: Taken together, these counters indicate the overall level of latency in each system. This information will be necessary for calculating the number and type of Isilon nodes necessary for optimal performance.

- **Disk Read Bytes/sec, Disk Write Bytes/sec**: These counters provide the throughput requirements of each system’s disks. By dividing the numbers in each
counter by 1,024, the total throughput requirements (in MB/s) for both read and write operations can be determined.

- **Disk Reads/sec, Disk Writes/sec:** When compared against one another, these counters indicate the overall ratio of read-to-write activity that each system's individual disks experience under operational workloads. This ratio is likely to vary over time, so look for both average values and peak values when sizing your storage cluster to meet this demand.

**Analysis Methods**

The following steps for interpreting the data may be useful in accurately gauging the performance requirements for your workloads:

**Mathematical Breakdown**

Analyzing the data mathematically involves looking at each workload’s performance counters and determining that workload’s:

- Peak utilization levels for each counter
- Lowest utilization levels for each counter
- Duration of peak workload for each counter
- Average performance levels for each counter

This information—gathered and assembled for each target component in the workload—can be useful both for assigning each application to a particular pre-designated performance tier for later performance planning, and for compiling aggregate workloads for each tier’s application.

**Graphical Analysis**

While the mathematical analysis will give a clear indication of peaks and valleys, creating charts from the datasets—either from an exported CSV file or from a database built using your performance metrics—will give a clear, visual representation of each application’s performance characteristics.

Performance charts are also valuable in identifying overall patterns of application behavior, and for identifying performance spikes in addition to the peaks that can be identified mathematically. If there are external workloads that are known to have run against the in-scope applications, such as backup or batch jobs that were not captured as part of the assessment, then comparing the times in which those jobs ran against the performance charts may give an idea of the relationship between the workloads, and the combined impact those workloads can be expected to produce. Knowing this information may in turn influence the design of the virtualized infrastructure.

Note: Most spreadsheet applications have the capability to create multiple types of charts besides the straightforward, line-style graphs. By experimenting with different
chart types while analyzing the performance data, you may find that a chart type that works in analyzing one metric value is less valuable with another.

Additional Performance Considerations
In addition to the performance metrics that can be captured and measured, you may also need to consider other factors and services before a complete picture of the necessary resources to support your new or changed virtual workloads can be assembled.

Data Access Patterns
While the counters recommended above will provide a picture of the overall disk-activity levels, the ratio of read operations to writes, and the latency requirements for each application, they may not tell the whole story. A standardized, reliable measure of an application’s overall data-access-pattern profile (e.g. sequential access vs. random access) is difficult to find, for example.

In such cases, consultation with one or more appropriate subject-matter expert(s) for a particular application or workload may be necessary in order to determine your applications’ actual access patterns, and to plan your storage infrastructure accordingly.

Backups
If the virtual systems being introduced into your organization require regular backups, then in some cases the additional data they will introduce will need to be factored into your existing backup window. In other cases, either the backup window will need to be adjusted, or the backup approach you currently use will need to be reassessed in light of the additional volume of data and the time available for your enterprise backups to complete.

Integration with Existing Workloads
Another consideration to remember is that, in many consolidated storage environments, the new workload that will be created by your virtualized infrastructure may be adding to the workload already on your existing storage. If your Isilon cluster will provide services to applications, services and processes beyond that of a new vSphere infrastructure, then the storage array will need to be sufficiently sized to meet all these simultaneous capacity and performance demands.

Performance and Capacity Planning
Once you have reviewed the performance metrics for your applications, then by comparing the performance and resource requirements required for each of them, you can begin to assemble the full requirements list for your virtualized workloads. In order to simplify this planning process, you may choose to classify the capacity and performance of your consolidated workloads into discrete performance tiers.
Application Tiering

For the purposes of evaluating enterprise workloads for consolidation, and planning the infrastructure platforms and systems that will host them, it can be useful to sort applications into different tiers based on the performance requirements of each application, then to match those application tiers to the appropriate resource pools—sized with hardware, network, storage, etc., as applicable—for their particular performance needs.

Since different infrastructure platforms will have different price-to-performance ratios, matching application tiers to their corresponding infrastructure tiers can serve two objectives: it ensures that each application’s capacity and performance needs are consistently being met; and that your most valuable infrastructure resources are being used appropriately by your most valuable workloads.

Tiering Parameters

Defining and assigning application-performance tiers could include any or all of the following high-level factors:

- **CPU** – architecture and utilization
- **Memory** – architecture and utilization
- **Network** – throughput (IOPS and/or Gb/s), latency (response times), link-redundancy requirements.
- **Storage** – capacity, throughput (IOPS), access-pattern requirements (e.g., read vs. write, random vs. streaming, etc.), queue-length, latency, data-protection, and replication requirements.

Performance Tiers - Examples

Overall performance requirements vary by organization. The following tier criteria are given as examples to serve as a starting point to plan for application-performance-requirement sorting. Specific sorting criteria is likely to vary based on the needs of your organization.

**Tier-I**

Tier-I applications typically require consistently high levels of CPU, memory, network and disk performance. Data access, whether file-based or transaction-based, is frequently high-volume, with a high rate of changes.

**Tier-II**

Applications in a mid-level tier have more moderate performance requirements from CPU, memory, network, and disk. Data tends to be more persistent, such as transformed transactional data, and is seldom modified once it has been created.
Tier-III
Relative to the higher application tiers, Tier-III applications are those that require relatively low levels of CPU, memory, network and disk performance. Once created, data in Tier-III applications, such as live-access archive applications and e-discovery services, tends to be static.

Tier-IV
Tier-IV may not be necessary in all environments. If used, it is typically reserved for those applications with the lowest overall performance requirements from all components. Examples of Tier-IV applications would be certain types of disaster-recovery, business-continuity, and compliance archiving applications and services.

Note: These guidelines are intended to be a suggestion only. Please review and consider their suitability within your organization, and make adjustments as necessary for meeting your unique requirements.

Workload Tiering – Examples
Consider the following examples of application tiering:

Enterprise Database Application
A database-based application that hosts one or more of your organization’s core business function(s) is likely to have stringent capacity and performance support from all of the above components. Such an application would likely receive a Tier-I designation within your organization, and its hardware, storage and overall performance requirements would then be used as the benchmark for determining the infrastructure performance requirements for all similar applications in your environment.

Development Platform
At the other end of the performance spectrum, you may have a single workstation-grade system within your enterprise that hosts multiple components for testing code currently under development. The end-user community consists of the developer, and the performance experience delivered by the workstation hardware is more than sufficient for what the end-user requires. This application would probably warrant the lowest tier in your rating system (e.g. Tier III/IV). In order to maximize the return on your infrastructure investment, such an application will most likely be consolidated onto a lower tier of hardware and storage than the Tier-I app(s) on the high-end infrastructure.

Virtualization Tiering
Establishing a single performance-tier designation for your virtualization platform may not be the best approach in many environments. Most virtualization projects entail virtualizing a heterogeneous mix of workloads, including applications whose performance characteristics and requirements span multiple tiers.
Examples

Even the concept of virtual servers vs. virtual desktops may not be an indication of the overall level of infrastructure resources necessary to support even a single virtual machine. Consider the following examples of virtual workloads:

Web Services

For intranet web servers whose locally stored data is largely static, and who typically encounter low levels of web traffic, the collective burden on the virtualized infrastructure will probably be quite low. Web servers with dynamic or transaction-based content, however, will likely require a higher degree of performance capability from the virtual infrastructure, as will web servers with higher rates of traffic.

Virtual Desktop Infrastructure (VDI)

As with web servers, the aggregated performance demand of a collection of virtual desktops is highly dependent on the number and type of workloads that will be run on those desktops. The overall performance requirement of a VDI platform is driven by a number of factors, such as:

- The desktop virtualization architecture and approach. A one-user-per-desktop virtualization model will have significantly different storage capacity and performance needs than would pools of non-persistent desktops based on linked clones.

- Workload profiles of individual users.
  - Desktop operating system platform(s), since different operating systems can have different I/O characteristics.
  - Type of end-user applications in use. Some commonly-used desktop applications have significant IOPS requirements. These should be factored in when tallying up the aggregate network and storage figures.
  - Number of simultaneous open applications per user.
  - Number of desktop screens per user.
  - Desktop anti-virus software architecture and configuration

- Business cycles, which drive workload volumes, and may run in daily, weekly, monthly, quarterly or annual timeframes.

- Average and peak volumes for concurrent end-user connections, including peak login times, login volumes, and login settings.

Just as there is no one universal model of end-user, there is no single paradigm for virtualizing either desktops or servers that will be applicable to all environments.

Workload Tier Classification and Aggregation

In order to translate gathered performance data into concrete storage infrastructure design plans, you may find it beneficial to take the following approach:
5. Complete an inventory of all the constituent systems for each application, e.g. web server(s), middleware components, database services, client systems, etc.

6. Conduct a mathematical and graphical analysis of each application’s component disk counters, and compile the relevant performance and capacity data.

7. Compile the overall results for each application, including all its constituent components, to calculate the overall capacity and IOPS requirements for the application.

8. Compare the resulting capacity and IOPS requirements for each application against the others.

9. Sort the results into high-, medium-, and lower-performance tiers as appropriate for your environment.

Once you have sorted your applications into their designated tiers, you can combine the capacity and IOPS requirements of all the applications assigned to each tier. This process should provide you with a general idea of the total storage capacity, availability, and performance levels that each tier will need for a successful deployment.

**Single Resource Pool Model**

If you have opted not to use a tiering model for capacity and performance planning, you may choose instead to plan out overall resource requirements simply by combining the total IOPS requirements of all targeted systems and workloads to arrive at the total numbers. Such an approach may be appropriate if your virtualized workloads will all have comparable performance requirements overall, or if your targeted virtualized workloads are all of the same comparable value to your organization.

Whichever approach you choose, once you have the fullest picture available of the number and type of applications to be virtualized, and of the capacity, availability and performance requirements for each application, you can begin to compare those requirements against the capacity and performance capabilities of the various Isilon storage nodes.

Note: While many of the principles and concepts enumerated in this Sizing Guide can be successfully applied toward assessing other vSphere infrastructure (such as the number and type of vSphere hosts necessary, or the implementation and use of such vSphere features as: Resource Pools, High Availability, Dynamic Resource Scheduling, Site Recovery Manager, etc.), the following section on planning for capacity and performance will focus on storage infrastructure. For information on planning your vSphere host configuration, please contact your VMware reseller or consult the available online resources.
Planning for Storage Needs

When planning for a virtualized environment, determining the number and type of nodes necessary can be a complicated effort.

Node Types

EMC offers several different types of Isilon storage nodes, each designed to meet specific performance-to-capacity requirements. Choosing the appropriate node type—or appropriate mix of node types—for your virtualized needs requires both an understanding of the capacity and performance characteristics of each type of node, and the aggregated performance and capacity requirements of your virtualized workloads.

Despite the architectural differences between the different node types, the OneFS operating system manages the aggregation of multiple node types into one namespace, and the combination of all available resources from every node in the cluster. An internal, InfiniBand-based network functions as a virtual backplane between the nodes, and enables any node in the cluster to service data for any client in the cluster, regardless of where the data actually resides.

S-Series Nodes

The Isilon S-Series product line is optimized for high-performance enterprise workloads for mission-critical, transaction-based, or random-access-file-based applications and workloads. Each node can be configured with a mix of 24 Serial-Attached SCSI (SAS) hard disk drives (HDD) and/or solid-state drives (SSD) to deliver very high throughput while minimizing latency. Each S-Series node can be configured with up to 96GB of cache memory, as well as 2 x 1Gb Ethernet and 2 x 10Gb Ethernet network interfaces.

If you have mapped out a system of application-tier designations, the S-Series may be appropriate for your Tier-I or Tier-II workloads.

X-Series Nodes

The Isilon X-Series product line is designed to provide a cost-effective balance of performance and capacity. These nodes function well in high-concurrent and sequential-throughput workloads such as most virtualized environments. X-Series nodes can be configured with 12 Serial ATA (SATA) disks (SSDs can be added as well), up to 48GB of cache memory, as well as 2 x 1Gb Ethernet and 2 x 10Gb Ethernet network interfaces.

Depending on your overall application-performance requirements, the X-Series may be appropriate either for your Tier-I, Tier-II, or Tier-III workloads.

NL-Series Nodes

Designed and optimized for reliability and economy, the NL-Series typically functions as an archiving repository, a disk-to-disk backup solution and disaster recovery platform. Each NL-Series node can be configured with either 24 or 36 SATA HDD
spindles—for a capacity of up to 108TB per node—as well as up to 16GB of cache, and 2 x 1Gb Ethernet and 2 x 10Gb Ethernet network interfaces.

The NL-Series of storage nodes typically provide optimal value for your Tier-III and Tier-IV workloads.

Even in a heterogeneous Isilon cluster, with a mixture of S-, X-, and/or NL-Series nodes, you can still manage and provision vSphere datastores across a single namespace.

The general performance capabilities of a given cluster are determined by the aggregated characteristics of all the constituent components and settings of the cluster:

- Total number of nodes of each type
- Number and types of disk in each node
- Total memory per node
- Number and type of available network connections
- Number and type of connected storage clients
- Number and distribution of SmartConnect network pools
- Protection levels and optimization settings
- Protocols used (CIFS, NFS, iSCSI) and number of clients

**Performance and Protection Tiering**

**Performance Tier Assignments**

Assigning datastores to a particular storage tier is as simple as either manually choosing the appropriate disk pool for an entire datastore, or using [SmartPools](#) policies to automate storage-tier assignments based on a number of different management criteria. Either option can be implemented at any time. There is no need to take either a datastore or a VM offline to change its settings.

**Protection Levels**

The Isilon system provides several levels of configurable data protection settings, which you can modify at any time without taking the cluster or file system offline or needing to reboot.

**Isilon FlexProtect**

The Isilon storage cluster provides a proprietary data-integrity system called FlexProtect™, which detects and repairs files and directories that are in a degraded state. FlexProtect protects data in the cluster based on the configured data-protection policy, rebuilding failed disks, utilizing free storage space across the cluster to prevent further data loss, and monitoring and preemptively migrating data off of at-risk components.
FlexProtect distributes all data and error-correction information across the entire Isilon cluster, ensuring that all data remains intact and accessible even in the event of simultaneous component failures.

The supported FlexProtect data protection levels are:

- **N+1**: The cluster can absorb the failure of any single drive or the unscheduled shutdown of any single node without causing any loss in stored data.
- **N+2:1**: The cluster can recover from two simultaneous drive failures or one node failure without sustaining any data loss.
- **N+2**: The cluster can recover from two simultaneous drive or node failures without sustaining any data loss.
- **N+3:1**: The cluster can recover from three simultaneous drive failures or one node failure without sustaining any data loss.
- **N+3**: The cluster can recover from three simultaneous drive or node failures without sustaining any data loss.
- **N+4**: The cluster can recover from four simultaneous drive or node failures without sustaining any data loss.

OneFS enables you to modify the protection policy in real time, while clients are attached, and while they are reading and writing data.

Note: Increasing the protection level of a cluster will increase the amount of space consumed by the data on the cluster.

**FlexProtect Overhead – Disk Capacity and Node Footprints**

For each protection level, the cluster must contain a minimum number of nodes, as shown in Table 1 below:

<table>
<thead>
<tr>
<th>Data Protection Level</th>
<th>Minimum Number of Nodes Required</th>
</tr>
</thead>
<tbody>
<tr>
<td>N+1</td>
<td>3 nodes</td>
</tr>
<tr>
<td>N+2:1</td>
<td>3 nodes</td>
</tr>
<tr>
<td>N+2</td>
<td>5 nodes</td>
</tr>
<tr>
<td>N+3:1</td>
<td>3 nodes</td>
</tr>
<tr>
<td>N+3</td>
<td>7 nodes</td>
</tr>
<tr>
<td>N+4</td>
<td>9 nodes</td>
</tr>
</tbody>
</table>

*Table 1: FlexProtect Levels and Corresponding Node Requirements*

The parity overhead for each protection level depends on the file size and the number of nodes in the cluster. As shown in Table 2 below, the percentage of parity overhead
declines as the cluster gets larger. In general, +1-level protection has a parity overhead equal to one node's worth of capacity, +2-level protection has a parity overhead equal to that of two nodes, +3 is equal to three nodes' capacity, and so on.

<table>
<thead>
<tr>
<th>Number of nodes</th>
<th>+1 (overhead)</th>
<th>+2:1 (overhead)</th>
<th>+2 (overhead)</th>
<th>+3:1 (overhead)</th>
<th>+3 (overhead)</th>
<th>+4 (overhead)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 nodes</td>
<td>2+1 (33%)</td>
<td>4+2 (33%)</td>
<td>N/A²</td>
<td>3+3 (50%)</td>
<td>N/A²</td>
<td>N/A²</td>
</tr>
<tr>
<td>4 nodes</td>
<td>3+1 (25%)</td>
<td>6+2 (25%)</td>
<td>2+2 (50%)</td>
<td>9+3 (25%)</td>
<td>N/A²</td>
<td>N/A²</td>
</tr>
<tr>
<td>5 nodes</td>
<td>4+1 (20%)</td>
<td>8+2 (20%)</td>
<td>3+2 (60%)</td>
<td>12+3 (20%)</td>
<td>N/A²</td>
<td>N/A²</td>
</tr>
<tr>
<td>6 nodes</td>
<td>5+1 (17%)</td>
<td>10+2 (17%)</td>
<td>4+2 (34%)</td>
<td>15+3 (17%)</td>
<td>3+3 (50%)</td>
<td>N/A²</td>
</tr>
<tr>
<td>7 nodes</td>
<td>6+1 (14%)</td>
<td>12+2 (14%)</td>
<td>5+2 (28%)</td>
<td>16+3 (15%)</td>
<td>4+3 (43%)</td>
<td>N/A²</td>
</tr>
<tr>
<td>8 nodes</td>
<td>7+1 (12.5%)</td>
<td>14+2 (12.5%)</td>
<td>6+2 (25%)</td>
<td>16+3 (15%)</td>
<td>5+3 (38%)</td>
<td>4+4 (50%)</td>
</tr>
<tr>
<td>9 nodes</td>
<td>8+1 (11%)</td>
<td>16+2 (11%)</td>
<td>7+2 (22%)</td>
<td>16+3 (15%)</td>
<td>6+3 (33%)</td>
<td>5+4 (44%)</td>
</tr>
<tr>
<td>10 nodes</td>
<td>10+1 (10%)</td>
<td>16+2 (11%)</td>
<td>8+2 (20%)</td>
<td>16+3 (15%)</td>
<td>7+3 (30%)</td>
<td>6+4 (40%)</td>
</tr>
<tr>
<td>12 nodes</td>
<td>11+1 (9%)</td>
<td>16+2 (11%)</td>
<td>10+2 (17%)</td>
<td>16+3 (15%)</td>
<td>9+3 (25%)</td>
<td>8+4 (33%)</td>
</tr>
<tr>
<td>14 nodes</td>
<td>13+1 (8%)</td>
<td>16+2 (11%)</td>
<td>12+2 (15%)</td>
<td>16+3 (15%)</td>
<td>11+3 (21%)</td>
<td>10+4 (29%)</td>
</tr>
<tr>
<td>16 nodes</td>
<td>15+1 (6%)</td>
<td>16+2 (11%)</td>
<td>14+2 (13%)</td>
<td>16+3 (15%)</td>
<td>13+3 (19%)</td>
<td>12+4 (25%)</td>
</tr>
<tr>
<td>18 nodes</td>
<td>16+1 (5%)</td>
<td>16+2 (11%)</td>
<td>16+2 (11%)</td>
<td>16+3 (15%)</td>
<td>15+3 (17%)</td>
<td>14+4 (22%)</td>
</tr>
<tr>
<td>20 nodes</td>
<td>16+1 (5%)</td>
<td>16+2 (11%)</td>
<td>16+2 (11%)</td>
<td>16+3 (15%)</td>
<td>16+3 (15%)</td>
<td>16+4 (20%)</td>
</tr>
<tr>
<td>30 nodes</td>
<td>16+1 (5%)</td>
<td>16+2 (11%)</td>
<td>16+2 (11%)</td>
<td>16+3 (15%)</td>
<td>16+3 (15%)</td>
<td>16+4 (20%)</td>
</tr>
</tbody>
</table>

Table 2: Node Counts, Protection Levels and Parity Overhead

You can apply protection settings globally across the cluster, or at the file or directory level. This flexibility enables you to protect disparate sets of data at different levels. Any child object inherits its policy from the parent, so it is important to ensure that any parent object has an equal or greater level of protection than any of its subordinate objects. For example, if a child directory were protected at the N+2 level, it could tolerate a double device failure; however, if the parent directory were protected at the N+1 level, it could sustain only one loss.

Snapshots

An additional factor to consider when planning out vSphere storage capacity needs and node types is the expected use of snapshots. While snapshots are useful for point-in-time backup and recovery purposes, snapshots consume disk space over and above that used by the VMs themselves. The amount of disk capacity needed for your vSphere environment’s snapshots is directly related to:

---

2 OneFS will allow you to specify a protection level that the cluster is currently incapable of matching. If you specify an unmatchable protection level, the cluster will continue trying to match the requested protection level until a match is possible. For example, in a four-node cluster, you might specify a 5x protection level; in this example, OneFS would protect the data at 4x until you added a fifth node to the cluster, at which point OneFS would apply the 5x level of protection to the data.
• Whether snapshots will be taken at the VM or datastore level
• The size of the dataset for which snapshots are captured. This includes questions such as whether VM memory state information will be included in the snapshot.
• The frequency with which snapshots will be taken and purged
• How many snapshots will be maintained at any given point

While you have the ability to manage where snapshot data will be stored, the disk capacity associated with creating and maintaining VM snapshots should still be factored in when calculating total storage capacity requirements.

**Network Connectivity**

All Isilon nodes can be configured to use either 1Gb or 10Gb Ethernet connections, so determining which type of connection to plan for is largely a function of what type of connections your vSphere hosts and network equipment will support. Many virtualized workloads see acceptable performance over 1Gb storage links, but since every environment will see different performance characteristics, if the intent is to use 1Gb connectivity for the storage-to-vSphere network, it is important to validate that the resulting performance will be acceptable at the VM and end-user levels.

**Network Path Redundancy**

Each Isilon node is configured with at least 2 x 1Gb Ethernet NICs. They may also come with either two more 1Gb NICs or two 10Gb links. If path redundancy is desired for the vSphere environment, then each vSphere host should have two fully independent network paths—separate NICs connected to separate switches, which in turn connect to separate ports on the Isilon storage nodes—to ensure continued connectivity in the event of a network-path failure.

Additional redundancy can be configured at the per-datastore level, whether using NFS datastores or iSCSI LUNs to host VM data.

**Network Throughput Performance**

In order to minimize latency while making the most of available network bandwidth between the vSphere hosts and the Isilon storage cluster, the following steps should be taken when planning the storage network fabric:

• Network traffic should be segmented on the vSphere hosts, so that VM network traffic does not share bandwidth with either management traffic or storage (VMkernel) traffic. If the vSphere hosts will have few available network ports (e.g. blade servers), use high-speed NICs with VLAN segmentation to ensure bandwidth is allocated and used appropriately.

• vSphere hosts and Isilon storage nodes should be local to one another on the same subnet. Additionally, the host(s) and the storage nodes should be connected to the same network switch(es) to minimize the number of hops between hosts and storage. This, in turn, will reduce the latency in the environment and improve overall storage performance.
- Ensure the network switching equipment is sufficiently robust, with enough backplane bandwidth, port buffering and port throughput, to support the expected network storage traffic for the vSphere infrastructure.

- If the network equipment and vSphere host hardware support it, consider enabling Jumbo frames from host(s) to storage to reduce packet overhead and network latency.

**vSphere Storage Planning**

Most of the enterprise features provided by VMware in all versions of vSphere are dependent on a shared-storage model in which all vSphere hosts in a vSphere cluster have concurrent access to the same data pool(s). VMware’s approach to implementing that model is to abstract the physical storage layer from the virtual machines using a “datastore”, or a logical shared-storage repository, to enable simultaneous access from multiple vSphere hosts.

**Datastore Types**

Isilon supports shared vSphere datastores based on either Internet SCSI (iSCSI) or Network-Attached Storage (NAS) protocols. While either type of datastore architecture can satisfy the sustained demands of most virtual workloads, there are differences in both their setup and in their ongoing management that may directly impact the capacity and performance of your Isilon cluster and your vSphere environment.

Table 3 below highlights the storage capabilities of each datastore protocol.

<table>
<thead>
<tr>
<th>Capability</th>
<th>iSCSI ³</th>
<th>NFS</th>
</tr>
</thead>
<tbody>
<tr>
<td>File System</td>
<td>VMFS/RDM</td>
<td>OneFS</td>
</tr>
<tr>
<td>Maximum Number of vSphere Datastores</td>
<td>256</td>
<td>256</td>
</tr>
<tr>
<td>Max vSphere Datastore Size</td>
<td>60TB ⁴</td>
<td>15PB</td>
</tr>
<tr>
<td>Max LUN/File System Size</td>
<td>2TB/60TB ⁴</td>
<td>15PB</td>
</tr>
<tr>
<td>Recommended Number of VMs per LUN</td>
<td>File System</td>
<td>20</td>
</tr>
<tr>
<td>Network Bandwidth</td>
<td>1GbE/10GbE</td>
<td>1GbE/10GbE</td>
</tr>
</tbody>
</table>

Table 3: VMware storage capabilities for NFS and iSCSI

³ Isilon currently supports vSphere Software iSCSI initiators only
⁴ VMFS5 (requires vSphere 5) only. VMFS3 (available with previous versions of vSphere) are restricted to 2TB in size.
NFS

Fundamentally, file-based VM datastores are presented to the vSphere hosts by the NAS array via a Network File System (NFS) export. Each vSphere host accesses the datastore(s) through an NFS mount using standard network interface cards (NICs).

NFS Advantages

Among the advantages of NAS-based datastores:

- **Simplified provisioning and management** – once an NFS datastore has been provisioned in OneFS and mounted on the vSphere hosts, it is immediately available for hosting VM data. No further configuration is necessary on the hosts. Additionally, because iSCSI datastores are limited in both their optimal size and the number of VMs each datastore can host, NFS-based datastores can be much larger, limited in practice only by the amount of network bandwidth available for each datastore.

- **Load-balancing configuration** – By creating multiple NFS exports to the same mount point within the Isilon cluster, additional bandwidth can be added between the vSphere hosts and the Isilon cluster. Additionally, VMs can be migrated between NFS datastores (to balance network traffic across all available network paths) without actually having to move the underlying VMDK files.

NFS Considerations

Although NFS datastores provide sufficient scalability and performance to satisfy most vSphere workloads, some additional factors should be taken into account when choosing which type of datastore to present to your vSphere environment:

- While both NFS and iSCSI can be configured for path failover in the event of a network-link failure event, the mechanics of the configuration are different. NFS datastore high availability is configured through OneFS, using either SmartConnect Advanced™ or Link Aggregation. High availability for iSCSI datastores is managed through vSphere using Multi-Path Input/Output (MPIO).

- Each NFS datastore requires two TCP connections per vSphere host. One connection is for control data, and consumes minimal bandwidth. The other is for the actual NFS data and carries nearly all TCP traffic between the host and the NFS datastore. The throughput rate between vSphere and the datastore is therefore restricted to one network path, and is limited by the available network bandwidth. It is not possible to add a second TCP connection from the vSphere host to the same datastore. Using OneFS, however, it is possible to create multiple datastores to the same NFS mount point, thereby providing multiple paths to the same data.

- The default limit on NFS datastores in vSphere 4.x is set to eight, although it can easily be increased through the vSphere client to 64. vSphere 5 allows a maximum of 256 datastores of either type.
• Some VMs have specialized disk-access requirements that may preclude the use of NFS datastores in specific scenarios.

iSCSI

The iSCSI protocol packages traditional SCSI commands into IP packets and transmits them across an Ethernet network. In this manner, industry-standard NICs act as host bus adapter (HBA) cards, without the level of complexity or cost often associated with Fibre-Channel storage area networking.

Note: A separate license is required to enable iSCSI within OneFS.

Advantages to iSCSI-based storage include:

• vSphere can leverage iSCSI multipathing to increase aggregate bandwidth to, and provide network redundancy for, a single iSCSI Logical Unit (LUN). This improves performance and availability to support the aggregate I/O stream from a large mix of virtual machines across multiple vSphere hosts.

• For specific use cases, vSphere can provide VMs direct access to iSCSI LUNs, including: physical-to-virtual-machine host-based clustering; Microsoft Cluster Server (MSCS); and/or applications that need direct access to a block storage device. The feature that enables direct LUN access to a VM is referred to as Raw Device Mapping (RDM) over iSCSI. In this design, the vSphere hypervisor acts as a connection proxy between the VM and the storage array.

• A guest OS can use an internal iSCSI initiator to gain exclusive access to an iSCSI LUN. In this mode, only higher-level networking calls are managed by the virtualization layer while SCSI- and/or iSCSI-level commands are handled by the guest OS.

• The ability to clone LUNs either as full clones, shadow clones, or read-only-snapshot clones.

iSCSI Considerations

As with NFS datastores, there are additional considerations to evaluate whether to use iSCSI datastores in your vSphere environment.

• Path failover when using iSCSI datastores is managed in vSphere via Multi-Path Input/Output (MPIO) configuration, rather than through OneFS using SmartConnect Advanced or Link Aggregation.

• If using MPIO for iSCSI path redundancy, [VMware requires that the vSphere hosts be on the same subnets as the storage nodes](#).

• Generally, because EMC recommends configuring all vSphere hosts to use the same preferred path to a given datastore, care should be taken to ensure that the total network bandwidth for all vSphere hosts using the datastore should be less than the throughput rating of the network path. A datastore that will be accessed concurrently by five vSphere hosts over a dedicated 1Gb network link should be sized to ensure that each vSphere host uses no more than about 100-200Mb/s of...
that bandwidth. The number and size of iSCSI datastores necessary should be driven by the performance requirements of the hosted VMs, taking into account both the network connectivity to each datastore and the underlying IOPs capabilities of the supporting Isilon storage nodes.

- While multiple NFS datastores, which are in reality different paths to the same data, can be presented to the vSphere host(s), each iSCSI datastore contains data unique to that LUN. Migration of a VM from one iSCSI datastore to another requires the underlying data to be migrated.
- While iSCSI LUNs can be thin-provisioned, and can be dynamically expanded, each datastore’s capacity is restricted by the allocated capacity of the underlying iSCSI LUN.
- Optimizing iSCSI performance in OneFS often requires mirroring the iSCSI LUNs, rather than choosing to inherit the protection settings from the parent folder. Choosing iSCSI datastores may reduce available capacity on the OneFS cluster as a result.
- If upgrading your vSphere environment from version 4.x to 5, the datastores’ file systems will have to be upgraded separately from VMFS3 to VMFS5. Consult with VMware’s existing documentation for more detailed information on this process.
- The Isilon iSCSI module does not support Internet Protocol Security (IPSec), two-way Challenge Handshake Authentication Protocol (CHAP), multiple connections per session, or hardware-based iSCSI initiators/host-bus adapters.
- Because iSCSI LUNs are presented to the vSphere host(s) as block-level storage, VM disk alignment must be managed in the guest VMs on your iSCSI datastores.

In addition to the differences in configuration and manageability between NFS and iSCSI datastores, there are also differences in how each type of datastore will respond to your specific workloads’ I/O profiles. Since NFS and iSCSI are each optimized for different functionality and performance requirements, EMC recommends evaluating your environment’s specific needs when choosing between deploying NFS- or iSCSI-based datastores in your vSphere environment.

Storage Planning Assistance and Validation

Every organization will have a unique set of applications, performance tiers, and coalesced workload profiles on their Isilon cluster(s). Different organizations may also have different levels of expertise with Isilon products and features.

Because of the number of variables involved in planning and building an Isilon cluster to suit your organization's particular application requirements, EMC recommends engaging the appropriate external resources to assist you in reviewing your organization’s distinctive needs. Suggested resources include your EMC reseller, your EMC account team, or Isilon’s Global Systems professional services team.
Advanced Isilon Features for Optimal Management

In addition to the native Isilon storage elements described above, when planning your vSphere infrastructure, consider utilizing the Isilon storage features described below to further optimize your storage systems. They can be very valuable tools to ensure that your Isilon cluster is configured—and that your vSphere data is optimally placed—for maximum availability and performance.

Using SmartConnect Advanced to Provide NFS High Availability

EMC Isilon SmartConnect Advanced implements NFS failover by assigning one or more dynamic IP addresses to each node’s member interface in the cluster from a configured range of addresses. If a single interface or an entire node experiences a failure, SmartConnect reassigns the dynamic IP addresses to the remaining pool member interfaces. All datastore I/O is immediately re-routed to the newly assigned member interface and VM virtual disk access continues without interruption.

Using SmartPools to Automate Datastore Management

EMC Isilon SmartPools™ enable you to automate both the performance management and the protection levels of your datastores, even across multiple storage tiers.

The SmartPools licensed module includes the following features:

1. **Disk pools**: Dynamic groups of disks associated in a single pool of storage, for example, all disks of all S-series nodes on the cluster. Disk pool membership changes through the addition or removal of nodes and drives.

2. **Disk pool provisioning**: Rules to automatically allocate newly-added capacity to the appropriate disk pool(s) within minutes of bringing the new capacity online.

3. **Virtual hot spares**: Reserved space in a disk pool (up to four full drives) which can be used for data re-protection in the event of a drive failure.

4. **File pools**: Logical collections of files and associated policies governing attributes such as file size, file type, location, and file creation, change, modification, and access times.

5. **Disk pool spillover management**: Rules governing handling of write operations to a full disk pool.

By using Isilon SmartPools, it is possible to tier virtual machine files, datastores and RDMs—dynamically and automatically—to correspond to your applications’ designated performance and availability tiers, thereby optimizing your storage resources. These can then be automatically migrated to the appropriate class of storage (SSD, SAS or SATA) via disk pools, using policies that you define and manage.

Finally, with SmartPools, protection (parity or mirroring) and I/O optimization can be applied per-file, per-VM, or per-datastore.
Operational Monitoring and Management Using InsightIQ

The use of EMC Isilon InsightIQ™ enables monitoring of the long-term usage statistics of all components of your storage. InsightIQ can quickly identify and correct potential performance bottlenecks and limitations within your cluster. It also allows you to track the storage and network overhead associated with each hosted workload on the cluster to ensure that all applications are optimally placed and protected within the cluster.

InsightIQ enables the detailed collection of the following metrics:

- Network traffic on a per-interface, per-node, per-client, and per-protocol basis
- OneFS operation rates
- Protocol operation rates and latencies on a per-protocol, per-client and per-operation class
- Per-node CPU utilization
- Cluster events
- Per-node disk throughput

At the OneFS file-system level, InsightIQ also allows you to monitor data growth at any of a number of levels, including per-directory, per-client, and per-file-type. This in turn enables you to monitor your overall usage per workload and adjust your capacity and performance expansion plans accordingly.

Note: The use of SmartConnect Advanced, SmartPools, and InsightIQ requires separate license purchases along with your Isilon cluster.

Conclusion

A comprehensive picture of the workloads targeted for virtualization should include the following information, gathered and compiled using the appropriate tool(s) for your organization:

1. The overall size (including VM count and storage capacity) and performance needs for each workload
2. The classification of each workload to the appropriate performance tier
3. The capabilities and considerations of the different Isilon storage node families relative to your designated performance tiers
4. The capabilities and considerations of the different vSphere datastore types
5. Isilon features and toolsets for managing and optimizing your VM data on the Isilon cluster

Because of the level of complexity involved in each organization’s workloads, scope of implementation, and unique availability and performance requirements, there is no one approach that will be successful in any organization. EMC instead recommends
enlisting the services of an appropriately-qualified external party, such as an authorized EMC reseller, your EMC account team, or an EMC Isilon Services engineer. Doing so will help ensure that your Isilon cluster is optimally configured to match your specific virtualized workloads.
References

EMC OneFS Technical Whitepaper
Working with Datastores in vSphere 5
Isilon S-Series Product Brochure
Isilon X-Series Product Brochure
Isilon NL-Series Product Brochure
About EMC Isilon

EMC Isilon is the global leader in scale-out NAS. We provide powerful yet simple solutions for enterprises that want to manage their data, not their storage. Isilon products are simple to install, manage and scale, at any size and, unlike traditional enterprise storage, Isilon stays simple no matter how much storage is added, how much performance is required, or how business needs change in the future. We’re challenging enterprises to think differently about their storage, because when they do, they’ll recognize there’s a better, simpler way. Learn what we mean at www.isilon.com.


Contact EMC Isilon

www.isilon.com

505 1st Avenue South, Seattle, WA 98104

Toll-free: 877-2-ISILON | Phone: +1-206-315-7602

Fax: +1-206-315-7501 | Email: sales@isilon.com

© 2012 EMC. All rights reserved. Isilon, Isilon Systems, OneFS, and SyncIQ are registered trademarks of Isilon Systems LLC. Isilon IQ, SmartConnect, AutoBalance, FlexProtect, SmartCache, SmartPools, InsightIQ, and the Isilon logo are trademarks of EMC.