

NetApp Services Viewpoint

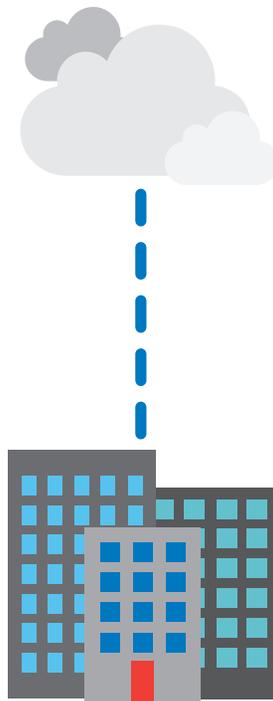
How to Design Storage Services Like a Service Provider





The Challenge for IT

Innovative technologies and IT delivery modes are creating a wave of data center transformation. To meet performance, cost, and agility requirements, IT organizations are adopting flash and cloud solutions in record numbers. Architectures built on these new technologies frequently generate new data silos—prisons into which data goes but never leaves. Such silos hinder digital transformation and restrict the benefits that a business can realize from its data.

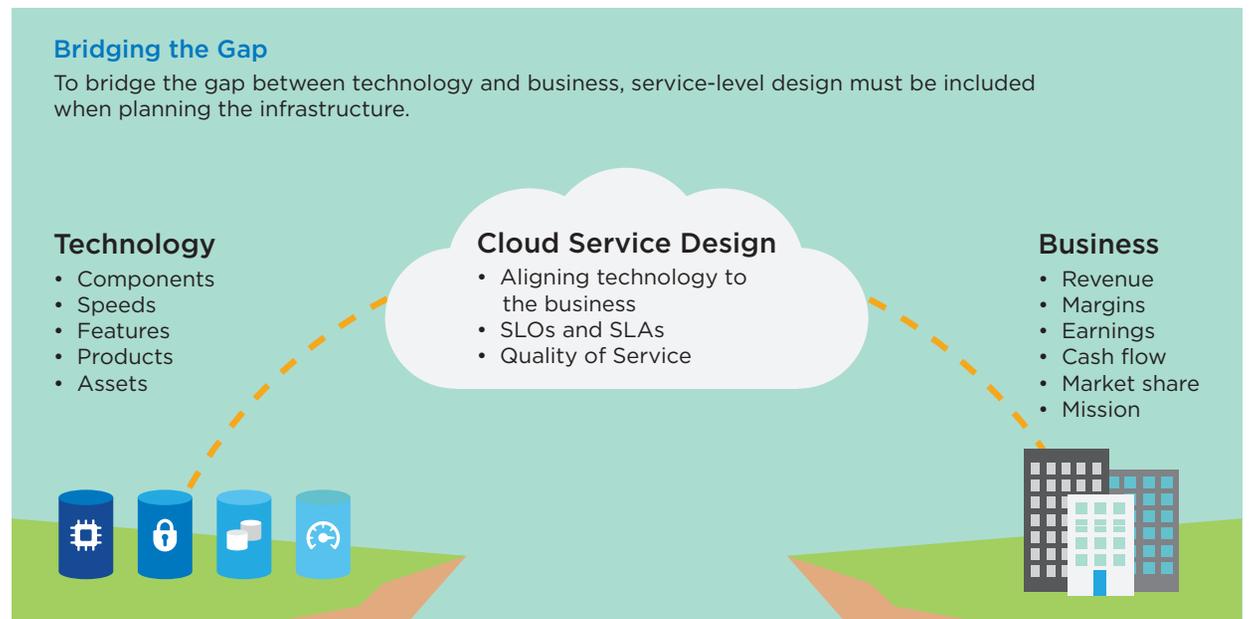


At the same time that complexity is rising, cloud services are raising the bar for IT delivery. Business units expect IT to deliver high performance at low cost with on-demand agility. IT consumers expect secure delivery of storage and data management services that can be consumed wherever and whenever they are needed. It is a new world for IT, and it requires a new approach. Today's successful IT organizations overcome these challenges by thinking outside the box and thinking like service providers.

A New Approach for On-Premises Storage Operations

Traditionally, IT organizations have gone directly from strategy to architecture design. To be successful in today's transition to the digital economy—whether deploying flash solutions or cloud technology—IT organizations must connect technology to the business.

Service design bridges the gap between technology and business. It enables IT to function like a service provider and operate under a delivery model that reduces costs while improving performance and agility.



Five Steps to Achieving a Service Provider Delivery Model

1. Define service metrics.
2. Assess application demands.
3. Define storage service levels.
4. Architect a shared infrastructure.
5. Automate delivery.

Are You Operating Like an Asset Manager or a Service Provider?

Traditional Approach

Asset Manager

On an application-by-application basis

- Assure performance
- Maximize asset utilization
- Reduce asset cost



Outcome

A custom infrastructure for every application:

- ① Increased costs
- ② Added complexity and rigidity
- ⚠ Elevated risk

Modern Approach

Service Provider

- Automate and orchestrate a shared infrastructure
- Deliver measurable service levels
- Align service levels to application cost and performance needs



Outcome

Streamlined storage operations:

- ① Simplification and consolidation
- ② Predictable storage cost
- ✓ Reliable performance

Operate Like a Service Provider

Cloud service providers use a business model that delivers predictable services with measurable cost and performance per unit of service. Like a restaurant, they provide customers with a menu of items that are priced per meal. Careful management of value and cost per unit makes this service model successful.

When IT acts only as an asset manager, it's acting like a restaurant without a menu. There's no understanding of the cost or the value of each meal, making it difficult to manage and meet expectations.

By seeking cloud services, IT consumers are signaling their preference for predictable value and cost. They like knowing exactly what they will get for their money.

Defining a menu of IT service levels is the first step in breaking down silos and transitioning to this new way of doing business. Organizations gain the ability to leverage all that the cloud model has to offer, both internally and externally.

All successful services share three attributes:

Service metrics



CPU



RPO

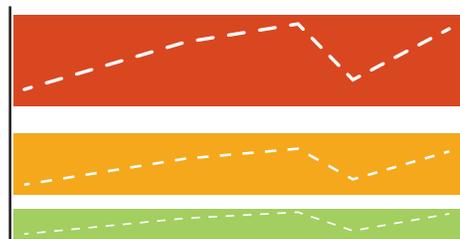


IOPS/TB

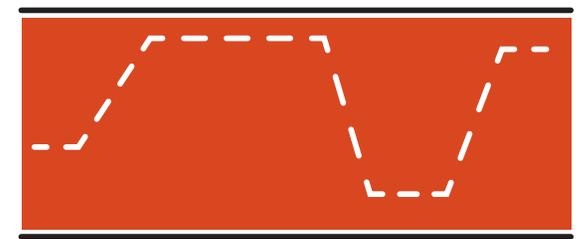


Users

Service levels



Demand regulation



Step 1: Define Service Metrics

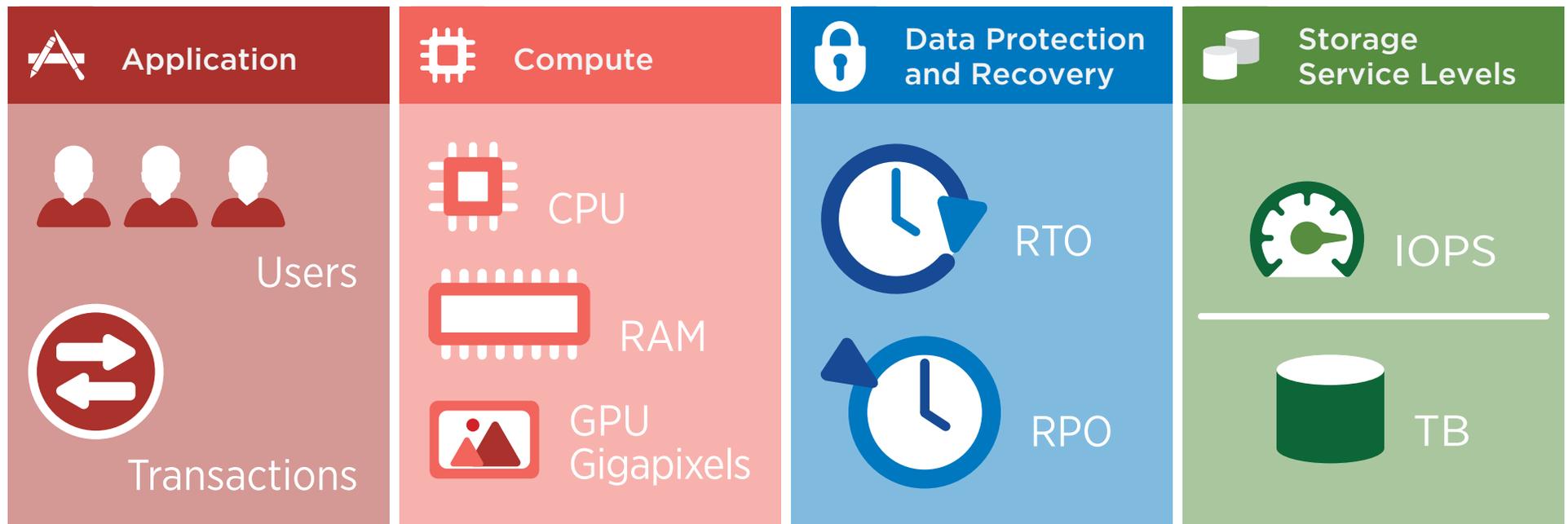
Shifting from the mindset of an operator of storage assets to the mindset of a service provider begins with defining your service metrics. Your applications and users don't consume hardware. They don't even consume latency. Storage hardware delivers input and output operations measured in IOPS, along with storage capacity measured in bits.

You will use the relationship between IOPS and capacity to define service levels. This relationship is expressed

as IOPS divided by stored capacity in terabytes: IOPS/TB. This single metric is critical when defining the design of the underlying architecture—all the way down to the storage hardware. The industry term for describing this storage metric is I/O density: the density of IOPS for a given amount of capacity stored.

Latency is a secondary metric that measures how quickly the service is delivered.

Service Consumption Metrics for IT Services



Step 2: Assess Application Demands

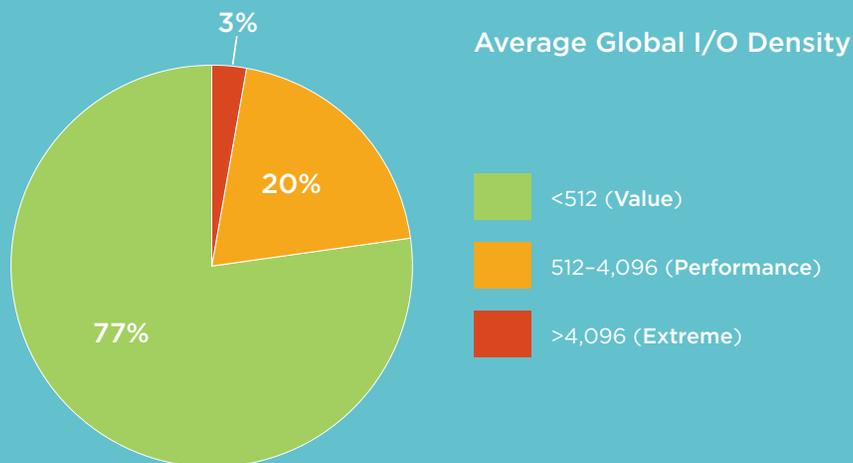
Start by assessing the I/O density that is delivered by your storage technologies to the applications that are requesting I/O.

Next assess the distribution of peak latency by capacity stored. This distribution illustrates how well your services are being delivered by the current infrastructure.

Average Global I/O Density

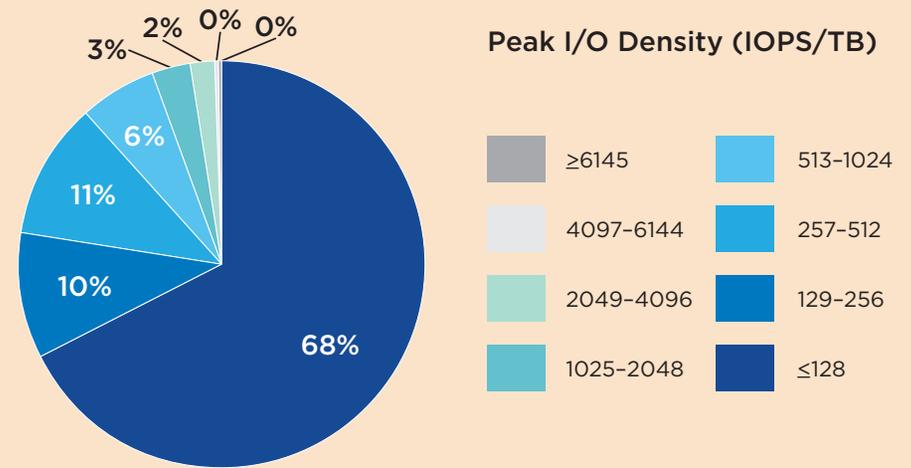
NetApp Services measured the average global distribution of I/O density across roughly 200PB of data from nearly 100 large enterprises and service providers.

77% of storage has an I/O density of less than 512 IOPS/TB.

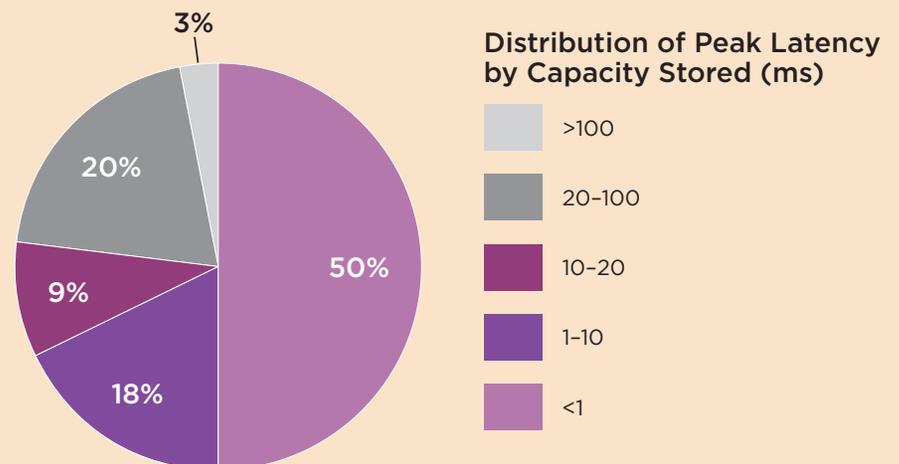


Example Distribution of Density and Latency Across Current Infrastructures

This data represents the speed of data services across the storage infrastructure. The data is independent of IT mode or platform type.



68% of workloads operate at a peak I/O density of 128 IOPS/TB or less.



50% of workloads experience a peak latency of less than 1ms.

Step 3: Design Storage Service Levels

In data centers that have evolved application by application, you will find hundreds or even thousands of unique configurations. Limiting your service offerings is the first step in simplifying IT infrastructure to provide predictable performance, cost, and agility.

A limited set of service levels aligns to the way that people make choices:

- Slow, medium, or fast?
- Cheap, midrange, or expensive?

With a shared infrastructure that delivers a limited range of service levels, your consumers can choose the service level that best fits their performance and cost requirements. Your storage consumers can rely on storage architects and operators to design and deliver the service that predictably meets their needs.

Application-Aligned Storage Service Levels

| |  Value |  Performance |  Extreme |
|-----------------------------|---|---|---|
| Costs | \$ | \$\$ | \$\$\$ |
| Workload | E-mail, web, file shares, backup | Database and virtualized applications | Latency-sensitive applications |
| Minimum SLA | 128 IOPS/TB | 2,048 IOPS/TB | 6,144 IOPS/TB |
| Maximum SLO* | 512 IOPS/TB | 4,096 IOPS/TB | 12,288 IOPS/TB |
| Latency Disk + Flash | 17ms | 2ms | 1ms |

*Maximum SLOs in the table are derived from proprietary assessments of customer workloads made over several years by NetApp Services and validated by NetApp partners.

Application owners choose service levels based on the minimum number of IOPS/TB required to support their applications during peak loads. Latency is generally a secondary consideration.

Regulate Demand

Overdelivery is the root cause of nearly every performance problem in a storage environment. The only way to control costs and meet minimum performance expectations without incident is to limit overdelivery.

Many providers don't publish the maximum speed of a service level (maximum SLO) because it's not an important consideration for the application owner. However, it's critical for the service provider to limit maximum performance in order to guarantee the minimum to all users.

Step 4: Architect a Shared Infrastructure

IT is constantly changing, and data has a lifecycle. The worst thing you could do in this dynamic environment would be to design your infrastructure on an application-by-application basis. Only a shared infrastructure can provide nonstop storage services on a broad range of service levels. Eliminating dedicated infrastructure by consolidating many point solutions into shared infrastructure also lowers complexity and labor costs.

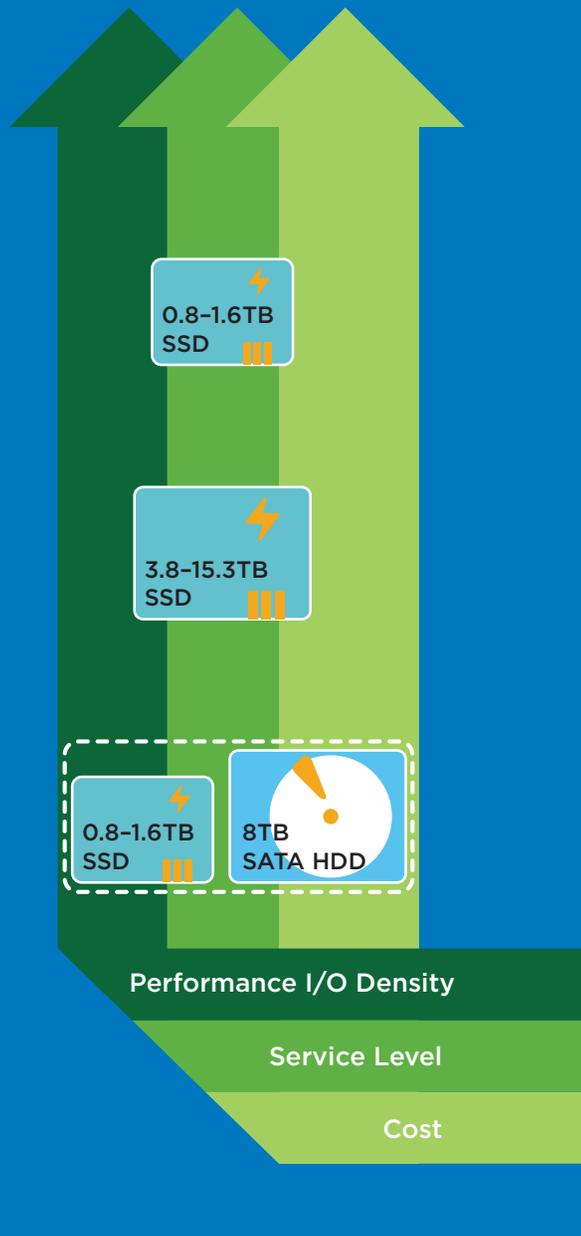
The performance and cost requirements of each service level should determine your architecture, for example, whether storage is all flash or flash accelerated.

Infrastructure Capabilities Required for Efficiency

- Support for all-flash or hybrid flash-accelerated disk
- Support for all protocols required by applications
- Constant availability without data loss
- Movement between service levels without disrupting applications
- Ability to start small and scale to very large capacities without disruption



Cost, Service Level, and Performance I/O Density by Drive Type



Incorporate Flash Everywhere to Align Service Levels to Architecture

Today, all-flash solutions are dominating the middle and highest service levels. They dominate not only because of speed and latency, but also because physical density allows the consolidation of whole racks of spinning disk drives into a few rack units of solid-state drives (SSDs).

Smaller SSDs, such as 800GB and 1.6TB, provide the highest-performance I/O density (IOPS/TB) for the highest service level. Large SSDs, such as 3.8TB and 15.3TB, are the lowest-cost choice for the middle service level because of the lower physical density and because storage operating systems can more efficiently store data on SSDs. Large SATA drives, such as 8TB SATA HDDs, accelerated by 800GB or 1.6TB SSDs currently provide the lowest-cost media for the lowest-cost service

level, where peak latency of up to 17ms is acceptable. The SSDs cache hot data, and the less-expensive but slower SATA drives store the cold data.

Maximize Storage Efficiency

Modern storage operating systems on shared storage platforms can store several times more application data than the amount of raw storage. This capability reduces your solution and labor costs to far below what direct-attached storage (DAS) can support. Your storage system should optimize efficiency techniques for SSDs.

Efficiency Techniques for SSDs

- Deduplication
- Compression
- Compaction
- NetApp® Snapshot® copies and clones
- Thin provisioning

With these techniques, SSDs can offer significantly lower cost per gigabyte than DAS.

Manage Quality of Service

Quality of service (QoS) addresses many problems simultaneously. It enables a predictable cost per GB and provides a performance commitment to applications and storage consumers. Nearly every storage

performance underdelivery

problem is caused by an overdelivery somewhere else.

Simply overbuying infrastructure doesn't solve the problem because any one application can consume all

the available IOPS from the allocated storage resources.

Without QoS, the performance cost of any volume in the system is completely random, regardless of the underlying media.

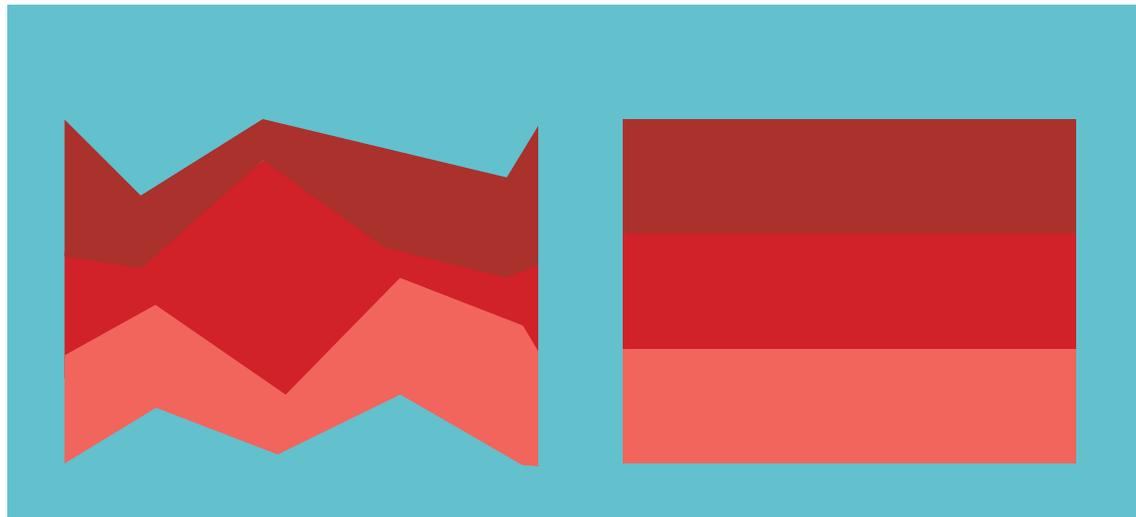
SSDs are creating a problem for shared infrastructure: The drives are faster than the components above them. Just a small amount of storage can overwhelm the controller resources. QoS solves this problem by evenly distributing

performance across all volumes, allowing a greater amount of SSD capacity to be attached to controllers, without stranding storage or causing unacceptable latency.

Some organizations don't implement

QoS because of the complexity and cost of managing individual QoS settings for hundreds or thousands of volumes. A QoS policy manager automates the task of managing QoS at the volume level by translating service-level policies into the QoS settings for individual volumes.

Performance Before and After Quality of Service



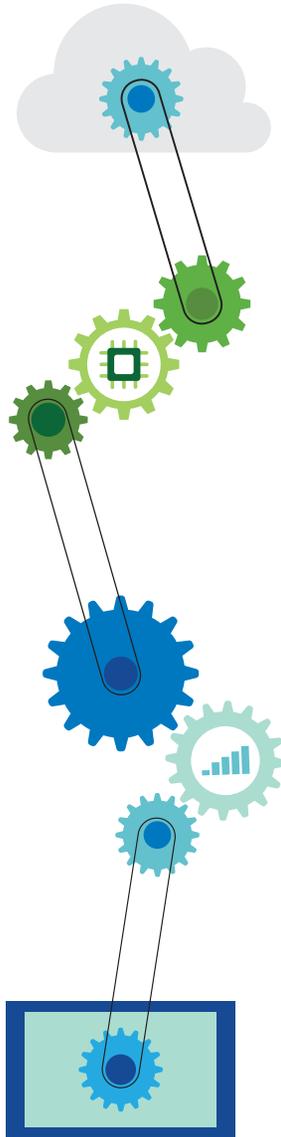
Step 5: Automate Delivery

Service providers are able to provide and monitor a performance SLA with rapid service provisioning on a large shared infrastructure using:

Automated QoS Policy Management by Service Level

It is impossible to deliver the minimum performance expectation without a QoS limit on storage volumes at each service level. Use a QoS policy manager to automate QoS management at the volume level and translate service-level policies into QoS settings for individual volumes.

By having a small number of service levels that are governed by QoS, you can help prevent performance incidents and make sure that the cost for every volume is predictable.



Automated Service-Level Monitoring and Reporting

Every service consumer wants to know the answer to three questions:

1. Am I getting the service level that I expect from all volumes?
2. Should I move a volume to a lower-cost service level?
3. Should I move a volume to a faster service level?

Automated Service-Level Provisioning and Orchestration

Automated provisioning and orchestration can reduce labor costs, enhance agility, and increase the value of your services.

To automate provisioning and orchestration, you must integrate these components:

- Customer-facing service-ordering portal
- Storage volume provisioning workflow
- Orchestration workflow of other services
- Configuration management
- Reporting

Although this integration might seem complex, it is less expensive and easier overall than delayed and error-prone human processes.



Get Started

Today's successful IT organizations are taking a new approach to meeting the expectations for more predictable storage costs, performance, and agility for IT services. They are moving from managing assets to managing services and operating their IT like a service provider does. Connecting technology to the business is key to a successful transition. NetApp can help you get started.

A NetApp Storage Service Design Workshop helps bridge the gap between technology and business. It creates a strategy for enabling IT to function like a service

provider and operate under a delivery model that offers predictable storage cost, application performance, and business agility. The workshop provides key service delivery metrics and recommendations for delivering consistent storage service levels using all flash or a combination of flash and high-density disks.

For more information about how this workshop can help you build the right strategy for aligning service levels to your business needs, contact your local NetApp sales representative.

Contact Us

To learn more about how NetApp products, services, and solutions can help solve your business and IT challenges, contact your local sales representative or visit www.netapp.com.

About the Author

Evan Miller

NetApp Senior Manager
and Executive Architect

Evan brings more than 25 years of experience in helping customers deploy global IT services to accelerate their companies' agility. As the developer of the NetApp Service Design methodology and the lead of the NetApp Service Design team, he is focused on helping enterprises seeking to improve their IT performance and efficiency by running their IT like a service provider.

