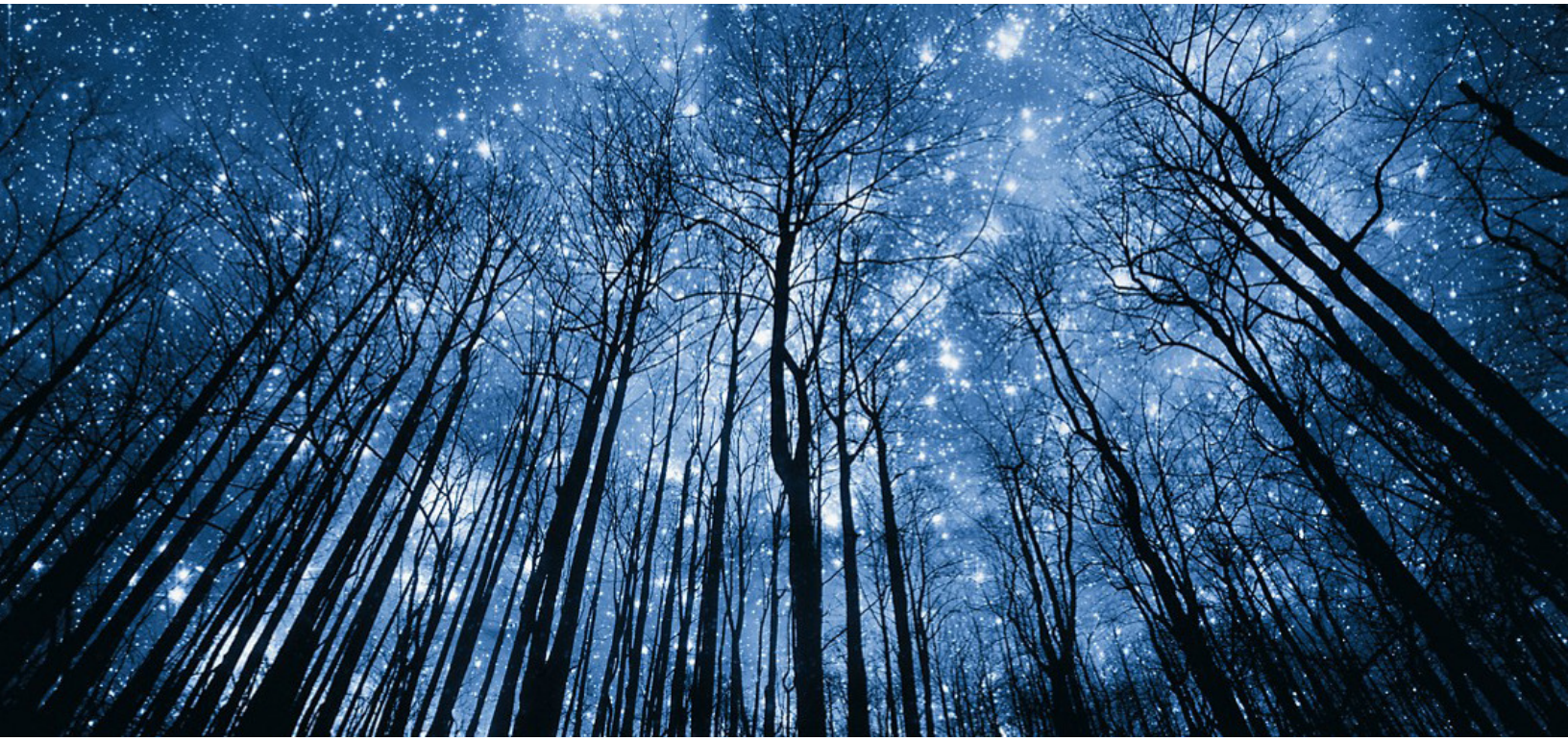
DELL Technologies

# STARTING A STORAGE CONVERSATION




## Hok Pui CHAN

Sr. Systems Engineer
Storage Platforms and Solutions
Dell Technologies
Hokpui.chan@dell.com


Knowledge Sharing Article



DELL Technologies Proven Professional


The Dell Technologies Proven Professional Certification program validates a wide range of skills and competencies across multiple technologies and products.

From Associate, entry-level courses to Expert-level, experience-based exams, all professionals in or looking to begin a career in IT benefit from industry-leading training and certification paths from one of the world's most trusted technology partners.

Proven Professional certifications include:

- Cloud
- Converged/Hyperconverged Infrastructure
- Data Protection
- Data Science
- Networking
- Security
- Servers
- Storage
- Enterprise Architect

Courses are offered to meet different learning styles and schedules, including self-paced On Demand, remote-based Virtual Instructor-Led and in-person Classrooms.

Whether you are an experienced IT professional or just getting started, Dell Technologies Proven Professional certifications are designed to clearly signal proficiency to colleagues and employers.

Learn more at www.dell.com/certification

# Table of Contents

## Introduction

This is aimed at infrastructure architects, pre-sales consultants, and storage administrators. We will discuss how to start and carry on a storage conversation to understand the customers' storage infrastructure and requirements. Asking the right questions is an effective way to understand the storage requirement, current pain points, and customers' plans. This article provides discussion points and insights for follow-up questions in a storage discussion. In the end, after gaining a holistic view of what is going on, we can architect the correct storage solutions for our customers.

Sometimes, customers do not know what can be included in their requirement – they may only know how much capacity they need. They may not have information about the latest technologies and the value they provide to their businesses. Our job is to help our customers uncover more requirements, identify their improvement opportunities, propose storage solutions that are relevant to them, and let them realize the actual values provided by the storage solutions.

We are going to discuss some common storage topics in this article. It is time to "go back to basics". For example, many customers have multiple storage arrays and want storage consolidations. So, what questions shall we ask? What information do we need to know? Well, we could probably ask questions like, "Why do you want to consolidate your storage?", "How are you going to handle applications with different Service-Level Agreement (SLA) on the same storage?", "Why do you have multiple storage arrays in the first place?" etc. Another example is when a customer tells you that they are going for cloud and don't need any on-premises storage array; is that the end of storage discussions? How are we going to respond?

There are no right or wrong answers to these questions. However, expanding these storage conversations will avoid talking too much about products to which the customers may not relate or how the proposed products are relevant to them. We can even extend our scope to cover other infrastructure components which come together with storage. Our customers will truly appreciate our advice, expertise, and dedication – they will see us as trusted advisors and the "go-to" persons for their problems.

We selected a few common storage topics and will focus more on primary storage, i.e. traditional storage arrays and Storage Area Network (SAN).

## Exploring Existing Storage Environment

When we meet with a new customer, we only know a little about their storage environment. There are so many things that we need to know, some technical, some not.

We can start the storage conversation by asking the first question: “What storage environment(s) do your most critical applications run on?”. Critical workloads could have the following characteristics:

- Running 7x24x365
- Databases
- Workloads have functions like executing transactions, billing, providing services to external users, or any other process involving money
- Running on physical servers
- Require DR/storage replication
- Demanding performance requirement

The customer could tell you that they have a mix of multiple SAN storage arrays, Network Attached Storage (NAS), Hyper-converged Infrastructure (HCI), etc. We can then ask how they distribute their workloads onto these platforms and if there are multiple storage tiers, how they allocate storage resources to different applications.

Many customers also have business continuity requirements for the applications. We can gather insight to these requirements from the storage side by understanding if the customer has multiple data centers with storage-based replication enabled. Depending on how demanding the SLA is, replication can be synchronous, asynchronous, or even active-active.

It will be a good start to know the customers' most critical storage requirements and how their existing storage infrastructure caters to them. There can be a lot of follow-up questions to be further discussed.

At the end of the meeting, if it is possible, we need to ask the customers to share a storage inventory and a storage architecture diagram. We need to know the kind of storage array (e.g. brand, model, configurations) and how things are interconnected (e.g. SAN, inter-site connection, etc.).

## Exploring Customer's Existing/Future Storage Requirements

After we know something about the customer's storage environment, we need to ask questions to make sense of it, e.g. the reasons for having multiple storage arrays, the reasons for not using storage replication, etc. Also, more importantly, how we can help the customer improve their storage infrastructure to support their future business needs.

We can start by asking the customer whether they have new projects/initiatives in the foreseeable future. For example:

- launching new critical applications
- improving business continuity
- active-active data center
- data center relocation
- storage consolidations

All the above can be further explored and new storage requirements can be formulated. New technologies such as advanced storage replication, low latency storage, scaling-out storage can be introduced to the customer.

Another important question is whether the customer plans to improve their storage environment. We want to stay positive when asking these questions, so we should avoid wordings like "pain-points", "past storage failure", "not able to meet SLA", etc. We can start asking leading questions like:

- Do you want to improve storage performance?
- Do you want to improve storage availability?
- Do you want to better cater for growth?
- Do you want to consolidate storage consolidation and overcome the challenge of managing many storage arrays?
- Do you want to enable storage replication to replicate data to a secondary data center?
- Do you want to have active-active storage across the data centers to enable workload mobility and improve availability?

Storage requirements keep changing. In recent years, customers have asked more about security; things like encryption, immutable snapshots, access controls and so forth. Moreover, given recent natural disasters and social unrest events, more and more customers ask about how the storage environment supports a third data center, probably in another city, to protect their data in case the primary data centers are compromised.

## Discussions on Performance

People usually focus on IOPS (Input/Output Operations Per Second) when they talk about storage performance. I am sure that you encountered customers saying that "Vendor A's array can achieve XXX IOPS while your array can only achieve YYY IOPS – Vendor A's array got better performance." Yes, IOPS is a critical measurement for storage performance, but that's not the whole story. The analogy is that Car A has 200 horsepower while Car B has 180 horsepower; does that mean Car A must be faster? Not necessarily.

When we talk about IOPS, we need to also speak about IO sizes and Read/Write ratios. Just like we will look at the weight of the car, torque and aerodynamics to evaluate whether is car is fast. We need to guide customers to look at the fine print when looking at a vendor's specifications. Usually, that includes assumptions such as "8K IO size", "100% read", etc. These IOPS figures can be impractical for the customer to evaluate whether they are buying a correctly sized storage array. Just like we will not take the straight line acceleration figure in car specification very seriously unless we are buying a racing car which runs on a race track.

Over the past many years, we focused on IOPS since IOPS was the bottleneck – customers bought a lot of spinning disks to boost IOPS, and IOPS is easy to understand. From the applications point of view, they do not care about maximum IOPS that the storage arrays can perform and a single application/server usually cannot exhaust the performance capability of the whole storage array. Now, in era of All-Flash, it is time to also look at IO response time.

It is our job to guide the customer through the storage performance conversations. The aim is to identify the bottleneck of the existing storage infrastructure and try to make improvements. We should focus on fixing/improving the bottleneck and prioritize investments accordingly. Investing in improving the bottleneck yields the most significant returns. In the past, it was the disks; now, the bottleneck shifts to other places, such as storage controllers.

Moreover, CPU speeds and memory are still much faster than storage. There is always a demand for faster storage (lower response time). We can expect more performance discussions with customers – the customers want to know how the latest storage technologies, e.g. NVMe (NVM Express), can help boost the performance of their applications.

# All-Flash or Hybrid

## A Bit of History and Introduction

We have been talking about flash or Solid-State Drive (SSD) for years. When flash media first came out, it was costly. An array full of flash drives was just too expensive. So, Hybrid-Flash Arrays (HFA), which mix flash drives and traditional spinning hard disks, appeared in the market to balance performance and costs. When different media types are mixed, there are mechanisms distributing the data among these media types within a storage array – we called that storage tiering.

All-Flash Arrays (AFA) refer to arrays with only flash media(s), e.g., NAND SSD, Storage Class Memory (SCM).

This section will talk about what we should tell the customer, their concerns about using All-Flash, and how to address their concerns.

## Hybrid-Flash Array

Everybody wants AFA – but it was just expensive. HFAs are cheaper ($ per TB) because it uses cheaper storage media, i.e. spinning disks. Most HFAs are designed to have "tiering" in which the arrays automatically put frequently accessed data to faster (and more expensive) media while putting less frequently accessed data to slower (and cheaper) media. Some HFAs also have cache extensions using flash medias. However, how does the array know which block of data is hot and which block of data is cold? Usually, it's by the performance history of that data block. In theory, if everything works fine, the applications can achieve the performance they need from the mix of flash and spinning hard disks. However, customers could face several difficulties:

- **Inconsistent performance** – data is sometimes handled by flash, sometimes by spinning drives. There is no guarantee that flash drives are serving every piece of data. The applications could be sometimes faster and sometimes slower.
- **Large footprint** – to satisfy the IOPS requirement, we need to put a large number of spinning disks into the arrays. The arrays could quickly end up consuming multiple racks in the data center. Hybrid arrays also consume more power and cooling than an AFA providing similar capacity and performance.
- **Storage management** – there are different tiers of storage media and customers need to consider how capacity upgrades are to be done, e.g. capacity required for each storage tier, how to allocate limited budget on different types of media, etc.

## Values of All-Flash

- **Better performance** – Flash media provide much more IOPS per device and more importantly, much lower response time for every single IO. Even if the customer needs very little IOPS, AFA can still provide immediate improvements for applications by significantly lowering response time of the IO operations. A customer might say that they do not need an AFA since their IOPS requirement is low – if so, remember to discuss response time with them. While even an HFA could satisfy the IOPS requirement, there is no way an HFA can have the same consistent low response time as an AFA.
- **Save money** – What? How? We can guide the customer to look at "$ per IO" instead of "$ per TB" which makes flash media much more attractive. More importantly, we need to look at the total cost of ownership (TCO) instead of only the initial purchase price of the AFAs. For example, AFAs use only a fraction of rack space compared to HFAs. Also, AFAs use much less power and cooling. These savings need to be considered as well by the customers. Customers typically use a storage array for 3, 5, or even 7 years. We need to include these savings in operation costs over so many years. Keep in mind that environmental friendliness is priceless, as well.
- **Higher availability** – Enterprise-class flash media are more reliable than spinning drives; they are less likely to fail, they live much longer. Also, flash drives are much more predictable than spinning drives; potential failures are easier to detect and drives can be replaced proactively before they fail. Even if a flash drive fails, the data can be rebuilt much faster and shorten the windows of data loss exposure (double faults).

## Garbage Collection

NAND SSD requires garbage collection. It is how the drives write/re-write data onto the storage media. However, there are different ways to perform garbage collection. Some arrays use the controllers to manage/initiate the garbage collection process – the drawback is that there will be performance impacts. The processing powers of the storage controllers are being used to handle garbage collection instead of processing IO. It is not ideal especially when the storage controllers are busy.

Some other arrays leverage the functions on the drive itself to handle garbage collection such that valuable storage processors' resources are not diverted. This approach is more scalable and allows more predictable performance, especially for demanding workloads or when the storage array is approaching high capacity utilization.

## Endurance

SSDs are designed to handle a finite number of writes. When some customers hear that their flash drives will die after writing a certain amount of data to them, they are scared. Yes, it can be scary, just like when we go to a fortune teller and he tells you about your "deadline" – you will then be scared and counting down your life every day, preparing your probate, making a bucket list, etc. But in fact, you still could have more than 100 years ahead of you. So, the concern should be how long the SSD would last. It is measured by Drive Writes Per Day (DWPD) or Terabyte Written (TBW) – the higher the number, the longer the SSD would last. We should see things from the bright side – this shows that SSDs are very predictable, and proactive actions can be taken if they are running low on endurance. There can be alerts from a storage array if the drives are running out of endurance such that actions, e.g. drive replacements, can be done proactively. In most cases, SSDs are still not running low on endurance when the storage array is decommissioned after using it for many years. It should not be a big concern.

In All-Flash arrays, there can also be mechanisms such as caching, write coalescing and wear leveling to maximize the life of the SSDs.

## Data Reduction

SSD is still more expensive than spinning disks. We can leverage data reduction technologies, i.e. compression and deduplication, to reduce the size of the data so that less physical storage capacity to store the same amount of written data is needed. Customers often worry that the desired data reduction ratio cannot be achieved and runs out of storage capacity unexpectedly before they have a new budget to spend. They also have concerns about potential performance degradation after turning on data reduction.

For performance, make sure the customer knows that all storage sizing from every vendor factors-in any performance impact for data reduction; even the customer is not going to turn that on in day one. Some arrays have hardware-assisted data reduction which introduces minimal impact to performance.

On the other hand, determining the data reduction assumption is not straightforward. We need to know the workloads. Some workloads are compression friendly, e.g. databases, while others can be very deduplication friendly, e.g. VDI. Unlike backup appliances, there is usually only one copy of the data for primary storage, so deduplication of multiple copies of data is generally not significant. Also, if the data is already compressed or encrypted before storing on the storage array, we can assume there will be no data reduction on the storage side at all.

If the customer is very conservative on the uncertainties of data reduction, we can suggest that they buy more usable capacity on day one while still turning on data reduction. If the data reduction is efficient, they can delay or even eliminate the need for capacity upgrades in the future. Also, there are storage data reduction guarantee programs from different vendors to ease customers' minds on this. We will have more discussions on these guarantee programs later in the article.

## Scaling-Out

In this article, we consider "Scale-Out" as increasing the performance and system limitations of a storage array. In simpler terms, it is about adding storage controllers to enhance performance capability of a storage array.

There are different "types" of "Scale-Out". We consider two types here.

The first type is that storage controllers can be added to the same storage array so that the same storage resources (e.g. LUNs) can be accessed from any of the controllers simultaneously – this is usually the way scale-out works in high-end arrays. This type of scale-out is sometimes considered the "Ultimate" or "True" scale-out. We do not want to categorize which type is superior to the other, we just call this "Type 1".

The second type is what we sometimes call "clustering" or "federation" – multiple storage arrays (or controller pairs) can form a relationship with each other. Storage resources can only access from one of the storage arrays (or controller pairs). This is "Type 2".

We do not want to conclude which one is better. We will discuss the characteristics of the two types of Scale-Out. We can then tell which type can satisfy the customer's requirements.

### Type 1

For this type of Scale-out, after storage controller pairs (storage controllers are usually added in pairs) are added to the storage array, hosts can access their storage resources (LUNs) from every storage controller. Here are some of the characteristics of this type of Scale-Out.

- Usually, minimal manual rebalancing needs to be done after Scaling-Out. However, hosts may need to establish new connections to the new controllers to fully utilize all controllers. Hosts can then enjoy the performance capability from all the controllers.
- Typically required to have very similar hardware configurations for the additional controllers. Sometimes, it can be less flexible than the other type of Scale-Out.
- Since the hosts can connect to all the controllers to access all their storage resources, this type of Scale-Out can enhance availability. In some scenarios, the storage array can still provide service after two controllers fail simultaneously, although it is not guaranteed.
- When there are more than two controllers, after one of the controllers failed, there is only less than 50% impact on the processing power of the storage array. Say, when the storage array got four controllers, losing one only means losing 25% of the total processing power – the hosts can utilize the remaining three controllers to access their storage resources.

### Type 2

We can treat these clustered storage arrays to be a single storage array as long as a single management UI (User Interface) manages them, i.e. from the single management UI the storage administrator can perform all the storage administrative tasks and monitor all the hardware components within the cluster. If a separate UI is required for each storage array, even clustered, it is difficult to say that this cluster can be considered a single storage array. Although this scale-out method does not directly enhance resilience, it provides a lot of other flexibilities.

Let's consider this scenario with two storage arrays (two controller pairs) clustered together so that we can illustrate the benefits of this type of scale-out:

- There are two controller pair which each pair own some of the LUNs. It can be considered as a 4-controller storage array and hosts can access a particular LUN through one of the controller pairs.
- We can physically separate workloads so that workloads on different controller pairs do not affect each other.
- Clustering can allow LUNs to be moved around different controller pairs for resources balancing.
- LUNs can be moved around based on the decision of the storage administrator.

- The clustered storage arrays can have some intelligence to make recommendations on LUNs placements.
- Customer can further expand their cluster to cater to additional performance and capacity workloads.
- Different models of storage arrays can be clustered together – some can have a higher performance to cater to more demanding workloads in the future.
- Different storage arrays (controller pairs) can be purchased at different times can have different maintenance periods.
- Through clustering, older storage arrays can be retired after seamlessly migrating all data to a newer storage array in the same cluster. There is no application downtime for such a tech refresh method.

## Storage Consolidation

Many customers want to consolidate their storage infrastructures, e.g. consolidating multiple SAN, consolidating multiple storage arrays. They want to manage as few arrays as possible. When a customer talks about storage consolidation, we need to ask several questions. We will focus on the consolidation of storage arrays in this section.

Before we talk about storage consolidation, we need to understand why the customer came to have multiple storage arrays in the first place. Is it the case that a single storage array could not satisfy the performance requirement from the applications? Are they used in different landscapes, e.g. production, UAT, etc.? There could be many non-technical reasons for having multiple storage arrays or even separated storage environments (e.g. SAN, storage arrays, management portals, etc.).

### Typical Storage Challenges Indicating a need for Storage Consolidations

When a customer experiences the following challenges on their existing storage environment, they may need storage consolidations:

- **Unbalanced storage utilization** – When a customer finds that some of the storage arrays are highly utilized (for both capacity and performance) while others are idle enjoying the cool air in the data center, it is time to discuss storage consolidation to share the spare capacities among different applications.
- **Complex management** – Too many storage arrays to manage. If you notice the storage administrators have a large spreadsheet recording the IP addresses and authentication details for each storage management interface, it is time to discuss storage consolidations with your customers. There are different tools to manage different storage arrays, especially for different product lines. Customers might think It's challenging to maintain the skill sets to manage different storage arrays. Also, it is error prone.
- **Storage sprawl** – Some customers will buy a new storage array for a new project. When a customer acquires storage arrays in such a way, we need to look at the trend of the number of storage arrays in the past few years.

### Typical Considerations of Storage Consolidation

When customers start considering storage consolidation, there can be some obstacles, such as:

### Noisy Neighbors

Storage consolidation means sharing storage resources among different applications. Different workloads can indeed affect each other. In the old days, we used many techniques to allocate resources, e.g. separating disks into different pools, so different LUNs would not compete for the sacred IOPS from the disks. At present, we can still do things the old way. However, when it comes to All-Flash Arrays, the bottleneck is usually the controllers which are being shared no matter how we separate the physical storage medias into different pools. Separating resources physically does not always help.

We need to leverage QoS (Quality of Service) or IO limitation functionalities on the arrays. For example, we can set a limit to some LUNs so that the LUNs will not go over a certain IOPS to reserve resources for other LUNs.

Setting the limit can be tricky. It can be both art and science. We need to understand the behavior of the applications to develop a meaningful limit. For example, an application's most IO-demanding period can be when the data is backed up and could affect other applications. A limit can be set during this backup period if we know about this behavior. The limit needs to be low enough to protect other workloads but high enough to maintain backup speed. We need to look at the historical performance data and the nature of the workloads to come up with a meaningful setting. Settings should be reviewed from time to time and fine adjustments made.

## Cater for Growth

When workloads are consolidated to a single storage array, we need to cater for growth (capacity and performance). Usage should be projected carefully to see if limits of some arrays will be reached in the future, identifying storage arrays which can be Scaled-Up as well as Scaled-Out.

## Times when Consolidation is not Practical

There can be many reasons that workloads cannot be consolidated to a single storage array:

- **Different application SLAs** – Application criticality varies. Placing less critical applications to a high-end storage array may not be cost-effective. The customer could still have multiple tiers of storage arrays that balance the costs and the benefits of having storage consolidations.
- **Different landscape** – Customers might want to separate their production, testing, and development environment. Another example is that some customers might want to separate the storage arrays for their Mainframe and Open Systems environments.
- **Security** – Some application owners are concerned about security of their applications when sharing resources with other workloads. Technically speaking, there are different mechanisms (e.g. logical separation) to address this. However, their concern can be non-technical, e.g. human error, internal regulations, etc.

## The Requirement of an Ideal Consolidation Platform

Suppose the customer is considering storage consolidation and choosing a storage array to achieve their goal. In that case, we need to remind them that a few characteristics and functionalities are necessary:

- **Exceptional and predictable performance** – Many different applications will run on the same storage platform after consolidation. Suppose the storage array cannot deliver the necessary performance consistently. In that case, application users will likely blame the storage for the poor performance of their applications – no matter whether the poor performance is related to storage or not. Ensuring the applications have the same or better performance is the first critical success factor for storage consolidation.
- **High availability** – This one is obvious. We need ultra-high availability since the storage platform will be hosting the most business-critical applications. If the storage is not available, it will affect all applications.
- **Scale-out** – There should be a way to increase performance capability of the storage array by Scaling-Out. (Please refer to the previous section for more discussions.) Otherwise, the customer could end up having to acquire an additional array to cater to the performance growth of their applications, defeating the purpose of storage consolidation.
- **Rich data services** – Data services include data reduction, data replication, etc. Different applications have different requirements for data services. This consolidation platform should provide all the necessary data services.
- **Easy to use UI** – Storage administrators now manage a larger storage environment with many different hosts connected to it. A single pane of glass is essential to manage the storage arrays. The storage administrator should be able to quickly review utilization and trends from the storage management UI.

## Fiber Channel (FC) or Ethernet?

FC and Ethernet are networks that connect hosts to storage resources in a SAN. Lots of discussions, blogs, and presentations are out there and we are not going deep down to discuss the technical comparisons of the two in this article. In the end, there are some non-technical aspects that a customer would consider. Sometimes, discussions would be non-technical rather than technical.

### Why Use FC?

FC was built and optimized for storage networks.

FC is usually a separate network for storage traffic, and customers love that. It is considered more secure since it is a different type of network that does not interconnect with the Internet. FC also has other access control features such as zoning – unless we specify which devices can communicate with each other, the devices cannot "see" each other.

Since the devices on SAN usually have hardware resilience, i.e. multiple physical paths from multiple initiators to multiple targets, FC SAN is designed to have physically separated networks ("fabrics") which maximize availability of the SAN. When changes are made to one of the FC SAN fabrics, the change will not affect the other fabric – there is always half of SAN unaffected. This protects against human error and physically segregates failures. On the other hand, all Ethernet equipment is usually designed to be interconnected somehow.

When a customer gets a storage team, the storage team usually controls everything related to storage, e.g. storage arrays, management tools, SAN switches, etc. It does not mean that the storage team wants to dictate and protect their "territory", but there are practical perspectives. For example, ports on switches need to be assigned when there is a new array – it would be easier for the storage team to provide the storage to the host with end-to-end control. Also, it would be easier to troubleshoot storage-related issues when the storage team gets the end-to-end view, especially when the FC switches are usually purchased from storage vendors who understand storage well. If it is an Ethernet SAN, the storage team might need to work with the network team and network equipment vendors for troubleshooting – it could delay problem resolutions and could lead to finger-pointing.

Some customers also consider an FC SAN easier and simpler to manage than an Ethernet SAN. FC SAN is usually not as big as the Ethernet and most things can be up and running by plugging in power and cables with simple configurations.

### Why Use Ethernet?

Customers usually consider costs when evaluating whether they should use Ethernet for SAN. They might also consider leveraging/sharing existing network equipment for SAN.

Customers also consider Ethernet more flexible. As long as the storage equipment or host is connected to the network, additional access control on the network is not needed – the host can "see" everything on the network with IP addresses.

Nowadays, as Ethernet bandwidth can now exceed FC, more and more customers are starting to consider Ethernet as a SAN networking option, especially as newer protocols like NVMe/TCP (NVMe over TCP) are currently supported by many storage vendors.

## Guarantee Programs

There are a few kinds of guarantee programs for storage. Below are tips to guide our customers on interpreting these programs.

### Data Reduction Guarantees

These Data Reduction Guarantee Programs do not guarantee a pre-defined data reduction ratio (the magnitude of data being shirked). Instead, they tell the customers what will happen if the storage array cannot achieve a certain data reduction ratio. Every vendor has its own Terms and Conditions (T&Cs) for the guarantee. To compare them, the customer should consider the followings:

- **The guaranteed data reduction ratio** – Different products, even from the same vendor, can have different guaranteed data reduction ratios.
- **Compensation** – Usually, this is about shipping extra capacity to the customer when the agreed data reduction ratio cannot be met. The customer should be aware of the limitation on the amount of additional capacity for compensation.
- **Conditions for triggering the compensation process** – Conditions can include the utilization of the storage capacity needs to reach a certain percentage, the storage array needs to be running the latest code, etc.
- **Duration of the guarantee** – Customers typically use a storage array for years. If the guarantee period is too short, it could expire even before data migration is completed. It is better to look for a longer guarantee period.
- **Pre-Assessment** – Some vendors require an assessment of current workloads to provide a guarantee program. The guaranteed data reduction ratio could depend on the assessment results. However, after the storage array is deployed, the customer may decide to put different workloads to the array. Customers need to determine what would happen if storage usage changed, e.g. a new application not included in the pre-assessment. Would the vendor still honor the guarantee agreement? If not, this agreement may not be flexible enough.
- **Exclusion clauses** – Some data types/applications are excluded, e.g. compressed or encrypted data, etc.

### Availability Guarantees

Availability Guarantees do not guarantee high availability. Availability Guarantees outline how customers will be compensated if a specific availability cannot be met, e.g. data unavailability events.

The customer needs to look at how they will be compensated if something happens. Usually, the compensation can be free maintenance, a discount towards a new purchase, etc. The customer needs to evaluate whether these options can really compensate for the loss to their business.

It is more important to evaluate whether the storage array to be purchased can deliver the availability the customer wants. An Availability Guarantee does not make a storage array more reliable.

## Cloud Discussions

We have been talking about “Cloud” for a least a decade. Most customers consider Cloud the alternative for their on-premises IT environment. We have spoken to customers saying that they are moving things to the Cloud and want to minimize spending on IT infrastructure. This article does not discuss whether it is more economical to put data, applications, and so forth on the cloud. Instead, as infrastructure and storage architects, we should aim to understand more about customers’ cloud initiatives and develop the best solution for them.

We need to understand why the customer wants to use Cloud. Some might say they want to minimize costs or consider CapEx vs. OpEx. Other customers might want to have a Cloud consumption model, electing to pay for what they consumed. There are a lot of vendors offering different consumption models for their products and can immediately address customers’ wishes without moving/migrating workloads to Cloud. We can expand our discussions beyond storage to how an IT infrastructure consumption model helps. This enables the customer to enjoy financial flexibility while maintaining specific control over the infrastructure and the corresponding SLA.

After we understand why the customer wants Cloud, we also need to know how they plan to execute their cloud vision. How are they going to move/migrate existing workloads to Cloud? Migrating to Cloud is not straightforward, and the customer needs time (years?) to shift things to Cloud. On the other hand, regulations can restrict data/workloads on Cloud. The customer could end up required to put their infrastructure on-premises. We need to keep reminding the customer that they still need on-premises IT infrastructure for some years to come. If the customer has a migration plan to Cloud, their on-premises infrastructure demands would diminish. However, it can be frustrating for the customer to know that while there will be fewer and fewer workloads on-premises, they still need to invest significantly at the beginning before migrating everything to Cloud. Again, we can leverage the different infrastructure consumption models to cater to such dynamic and diminishing on-premises requirements. Some consumption models can let the customer pay fewer infrastructure costs after workloads are migrated to Cloud in the years to come.

Other customers might want some “Cloud” elements in their environment while their peers and competitors talk about “Cloud”. They could ask questions like, “Does your storage array support Cloud?”, “Can we move data from the storage array to Cloud?”. Some storage arrays have the capability to somehow connect to Cloud – some storage arrays can copy a snapshot to Cloud, some can leverage a Cloud platform for monitoring and reporting. In the end, we need to help the customer build a proper use case for leveraging Cloud for their storage. For example, if the customer wants to copy a snapshot to Cloud, we need to understand what the Cloud snapshot is for – if the customer wants to look at the snapshot data on the Cloud for data analysis, it could be a sound use case. However, suppose the customer wants to recover production data from the Cloud snapshot. In that case, it might not be an efficient solution since the recovery takes time and might not be realistic for their recovery objectives. There are many other better solutions for this.

## Conclusion

Every customer and company have a lot of data. When there is data, there will be demands for storage. Consider our own storage requirements and you'll find that our circumstances are similar to those of our customers.

For example, we have more and more digital photos to be stored and we want to keep our photos safe. Some photos, e.g. photos taken at your kid's birthday party, are more important than the others. We will consider where to store our photos – somewhere reliable, fast and cost-effective. Storage capacity requirements for photos is constantly accelerating as the photos' resolution is getting better and better and we are taking more and more photos. There are also many photo-taking devices within a family and you need to manage the photos taken from these devices. We need to think about the scalability of storage and how to handle more photos, e.g. are we going to buy a larger hard drive and migrate all the photos over to it? Or simply buy another storage device/hard drive to store the latest photos. Can we retrieve our photos from different devices and play them back instantly on our TV when friends visit you? Shall we put the photos to the cloud directly after they were taken? How about security concerns? How to avoid the photos being accidentally deleted? There are also other questions to ask ourselves and we are now actually establishing a storage conversation with ourselves.

The thinking process for storing digital photos is very similar to our customers for their applications. Together with the ideas we talked about in this article, we hope that our customers can enjoy the true values of the storage products and consider that as suitable investments for their companies.