MAXTA HYPER-CONVERGED SOLUTIONS OVERVIEW

July 20, 2016 Jeremy Li

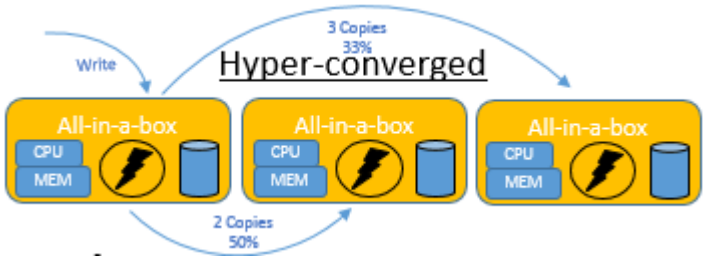
Maxta is an emerging software-only hyperconverged vendor and was named a 2015 Cool Vendor in Storage Technologies by Gartner - Recognized for providing innovative storage capabilities via new architecture and deployment method.

Hyperconvergence, or hyperconverged infrastructure, also known as hyperconverged integrated system [HCIS] by Gartner, sometimes refer to as Server SAN, provides elastic with scale up and scale out capability by integrating compute (x86) and storage resources via software-defined virtualization for both computing and storage on standard x86 platforms. The best use case scenario will be in a Virtual Desktop Infrastructure (VDI) environment.

Software defined storage (SDS) and hyperconverged infrastructure (HCI) can massively simplified data centers by eliminating an extra layer of complexity (SAN). Gartner Says Hyperconverged Integrated Systems Will Be Mainstream in Five Years

Built-in data efficiency, data protection and VM-centric management are key differentiators for next-generation hyperconvergence infrastructure.

Maxta Hyperconverged solution is based on RAIN technology, also known as “Log Structured File System”. As a result, it has 2X with resilient factor 2 and 3X with resilient factor 3 storage penalty, respectively prior to any data reduction is applied, as illustrated in the screenshot below:



Generally speaking, a usable storage is calculated as: The total raw capacity is divided by a replication factor – for example – 100TB raw storage / 3 = 33.33TB usable disk space.

Maxta Hyper-converged solutions comprise three components: (1) abstraction; (2) pooling; and (3) automation. It can simplify a traditional date center with HCI solutions for SMBs markets, as shown below:

- Eliminate the Storage Array
- Any Server Vendor
- Administer VMs, not Storage (See the following picture for details)
- Self-Healing & Self-Optimizing
- Lifetime Transferable Maxta License

(Source: http://www.maxta.com).

The below result is from a question-and-answer session:

Q1: Tell me what the reason Gartner still recommends "All Data Centers with 200 VMs or less should consider to deploy Hyperconvergence?

Below are quote from Gartner (emphasis added):

*"We believe that **highly virtualized midsize enterprises with fewer than 200 virtual machines should absolutely opt for hyperconverged infrastructure,"** wrote Gartner Analysts, Mike Cisek, Gartner Research Director and Jeffrey Hewitt, Research Vice President.*

*Gartner states on May 5, 2016 that "The market for hyperconverged integrated systems **(HCIS) will reach 24 percent of the market, by 2019.** Phase 3 represents continuous application and microservices delivery on HCIS platforms (2016 to 2025)."*

Hyperconvergence solution has a scale out limitation as Maxta has illustrated here because it can only support 12 nodes in one cluster. As a result, both Convergence solution and Hyperconvergence solution will co-exist for a long time.

This demarcation line does exist as shown in Gartner's report titled "**Simplify the Midmarket Data Center with Hyperconverged Infrastructure Solutions**." (Report published June 24, 2015)

Here's the pertinent part from the above report (emphasis added):

*"The costs associated with refreshing data center hardware, at midmarket scale (**80 to 120 production virtual machines (VMs) and 30 to 50 TBs of storage, coupled with their high levels of virtualization (80% to 90%)** are piquing midmarket I&O leaders' interest in HCIS."*

A1: [KS] We have currently tested up to 12 nodes. Architecturally we do not see any limitations to scale to the limit VMware vSphere supports. Not just in a hyperconverged environment, even with traditional storage array the average VMware cluster size on an average is around 8-12 nodes. The key reasons are

1. Isolation of Fault domain
2. Simplify VM administration and management.
3. Reduce middleware (Oracle etc.) software license costs

Q2: How many items below can Maxta meet?

Software Defined Storage: it's an approach to storage system design using the principles of distributed computing that delivers elastic, scale-out, multi-protocol, and geo-distributed capabilities all with a simplified management experience. This is how we will all store more data, and more kinds of data, than we have ever done before. **Source: EMC**

A2: [KS] Maxta support Elastic, Scale-out and Scale-up, NFS in a VMware environment, Native file system on KVM, geo-distributed capabilities and simplified management.

A slight variation in the definition from Intel:

"Software Defined Storage (SDS) is the framework for delivery of a scalable, cost-effective storage solution to serve the needs of tomorrow's Data Center. While standards are still evolving, it is a key component of a Software Defined Infrastructure."

The key aspect I would like to point out is the "cost-effective" that EMC has left out in their definition which is a critical component of software defined infrastructure.

Q3: Explain: (1) how is Maxta's hyperconvergence solution unique; (2) how does Maxta support whole system virtualization (hypervisors) and operating level virtualization (containers) technologies simultaneously; (3) Maxta solutions are in use by industry leading cloud service providers as a foundation for their client services?

A3: Maxta is the only vendor to provide software only to achieve SDS with no-lock-in vendors …

[KS] The Maxta® solution is a hypervisor-agnostic, highly resilient storage platform for the virtual data center. Maxta provides its storage solution as the software-only Maxta Storage Platform (MxSP™) and as the validated Maxta MaxDeploy™ Appliances. Maxta's software-centric, hyper-converged solution is transforming the enterprise storage market. It fully integrates with server virtualization at all levels, from the user interface to data management, and supports all possible deployments of virtual data centers, including private, public, and hybrid clouds. Maxta's software turns standard servers into a converged compute and storage solution, leveraging server-side flash

and spinning disks to optimize performance and capacity. Maxta's distributed architecture enables shared storage with enterprise-class data services and full scale-out capability without performance degradation. This results in significant cost savings, as well as dramatically simplifying IT.

The key value proposition for Maxta Storage Platform are:

- Maximize Choice o Choice of x86 server platform
- Software Agility o Lifetime license o Support for next generation hardware on day 0 o Manage VMs and not Storage
- Application Defined o Align storage policies with application on VM level granularity
- Efficiency o Eliminate Storage Arrays o Self-Healing & Self-Optimizing o Reduce Cost & Complexity

Q4: Snapshots, Replication and a VM-to-VM Replication – Built-in Data Protection, Data Efficiency and VM/data Mobility

Below are more granular questions:

a) Can Maxta actually snapshot or replicate a single VM at the storage level?
b) Can Maxta understand the busy blocks inside an actual VMDK?
c) Which block sizes does SympliVity use on the back-end for Redirect on Write (ROW) snapshots?
d) Does Maxta use asynchronous or synchronous replication?
e) Does Maxta support a VM-to-VM replication?
f) How many snapshot per VM is allowed? For example, (1) Cisco HyperFlex is limited to 30 snapshots per VM; (2) does not have a native replication capability; (3) it is unable to recover a deleted VM from a snapshot, which makes the functionality largely unsuitable for general backup; (4) as a result, a third party backup software must be used, which in turn increase the TCO; (5) also, a third party backup and replication cannot leverage the native deduplication of HyperFlex, meaning data must be rehydrated and subsequently dehydrated as it moves across cluster and site boundaries.

Source: TechTarget White Paper titled "Comparing Hyperconverged Infrastructure Options for Virtualized Environments"

Note: The smaller the block sizes (e.g., 8KB) on the back-end for Redirect on Write (ROW) snapshots, the lesser space will be used.

Some vendors use large 4MB-512MB block sizes which results in massive snapshot overhead that in turn much less snapshots can be produced and stored on storage. As a result, it will affect recovery time objective (RTO) and recovery point objective (RPO).

A4: [KS] Yes, Maxta delivers snapshot of a VM. There is no concept of taking a snapshot at a storage level. Customer manage VMs and not storage with Maxta storage platform. Maxta file system is a log structured file system. We write sequentially to MxIL which our write cache and then coalesce multiple writes and de-stage it sequentially to the HDD. From a read cache perspective, once we read a block it is pulled on to the SSD into read cache partition. Thus the traditional concept of a hot block does not apply. Maxta can assign certain parameters at a vDISK level. For example, things like page size, compression, read caching etc. Currently Maxta supports 4K through 32K. The default in our current release is 4K.
Currently we leverage 3rd party replication software. We have qualified Veeam, VMware vSphere replication, Zerto etc.

Yes, Maxta supports data-at-rest encryption. Maxta leverage Self-encrypting-drives (SED) to support data at-rest encryption. Maxta also has qualified Broadcom Safestore product to provide encryption. Specific to Fiber Channel storage arrays, QoS is only done at the FC WWN level and it is measured by IOPS or by MB/s. The only way to limit a guest VM is to limit the resources that are assigned to the individual server. Additionally, if you have QoS on a server with lower performance limits and a high performance VM is moved over to this physical node, the VM will suffer.

Q5: Is Maxta technology based on RAIN technology since it uses Striping Data Across Multiple Nodes?

A5: [KS] Although Maxta technology from a conceptual point of view is based on RAIN technology, the implementation is quite different. For example, with an RF2 when 2 drives fail on 2 different nodes, only the VMs that have data on the 2 drives will be affected and will become in-accessible. All other VMs on the cluster will not have any issues and will be accessible.

Will all subsequent snaps be also corrupted, if a corruption occurs in a parent Maxta snap?

> if corruption occurs in a parent snap, you in essence end up saving the fruit of the poisoned tree as children snaps will likely carry forward the corrupted data.

[KS] With Maxta you can always recover to any pervious snapshots. Maxta provides the ability to take 1000s of snapshots and customers can recover to any previous snapshots.

Source: TechTarget White Paper titled "Comparing Hyperconverged Infrastructure Options for Virtualized Environments"

Q6: Theoretically, Maxta can support unlimited number of nodes to existing clusters (http://www.maxta.com/solutions/cloud-service-providers/), but it only certifies 12 nodes per cluster (48 cores) at this time

Note: Most auto manufactures suggest that each tire pressure is at about 32 PSI for reasons. Yes, when 40 PSI is applied to each tire, it will still work, but a life will be greatly reduced.

A quote from Gartner report titled "Gartner Magic Quadrant for Integrated Systems" dated 11 August 2015 | ID:G00266749

"HCIS solutions offer broad scaling in theory, but are usually limited to smaller configurations."

A6: [KS] As explained in Q1, Maxta has qualified 12 nodes based on customer requirement. The architecture does not have a limit on the number of nodes. For the right opportunity Maxta will qualify higher number of nodes in the cluster.

Q7: Does Maxta support QoS (IOPS)?

A7: No. Maxta Mxinsight for VMware vSphere can view the IOPS for the entire cluster, but not each individual VM.

[KS] In the current release it is true that we cannot set a threshold for IOPS on a per VM basis. That feature is on our roadmap. Although customers cannot set IOPS thresholds on a per VM basis, we can set rebuild priorities on a per VM basis. Rebuild priorities will provide the ability to identify the VMs that has to be rebuild first when things fail.

Q8: Can Maxta Mxinsight for VMware vSphere view each VM throughput and latency?

A8: [KS] Yes

Q9: Resiliency - Can You Explain How Maxta Store Metadata with Replica Factor (RF2)?

RS2 only protects against single drive loss. If Maxta system loses any two disks or a single disk while a node is offline in a cluster because of a failure or maintenance, all data will be lost. Is it true? Note: Nutanix also has this kind of issue.

Note: If I choose a RF3, do I required a minimum of five nodes in one cluster? If so, an additional 50% physical capacity, and moves 50% more traffic over the network to store the same amount of data.

A9: [KS] Maxta metadata has 2 components.

1. Global metadata

2. Local metadata for each server

Global metadata is spread across a number of nodes within the cluster. The number of copies that are spread depends on the number of nodes in the cluster. The global metadata is also stored on a SSD partition across multiple servers in the cluster. This provides redundancy for global metadata.

Local Metadata for each server

Local metadata is also stored on a separate partition on the SSDs. Local metadata is mirrored across a second SSD within the node to provide redundancy. In case one of the SSDs fails we rebuild the local metadata by reading the local metadata form the second SSD.

We also support deployment with a single SSD. In that scenario we rebuild the metadata from scratch. We will have to rebuild the node data in that scenario.

Our best practice is to deploy a node with SSD mirroring.

Also metadata and data are not stored on the same disk. This allows Maxta to support silent data corruption. This capability is huge value that Maxta delivers and not many storage platforms support this capability.

With RF2 the key thing is that Maxta will be able to tolerate 1 node failure. When 2 disk fail only the VMs that have data on these 2 disks will be affected and all the other VMs will still be accessible.

Q10: Data Efficiency - Since Maxta does not support a Block Level dedupe, does It support Inline Compression at Ingest?

A10: [KS] Inline compression and compression at ingest are the same.

Yes, it is on our Roadmap. One of the key value that Maxta supports is very efficient capacity optimization delivering significant lower storage costs. Maxta supports

- Thin Provisioning
- Compression
- Metadata based (Deduped) Clones and Snapshots.

Also these features are enabled by default on the entire filesystem for VMs and their snapshots and clones.

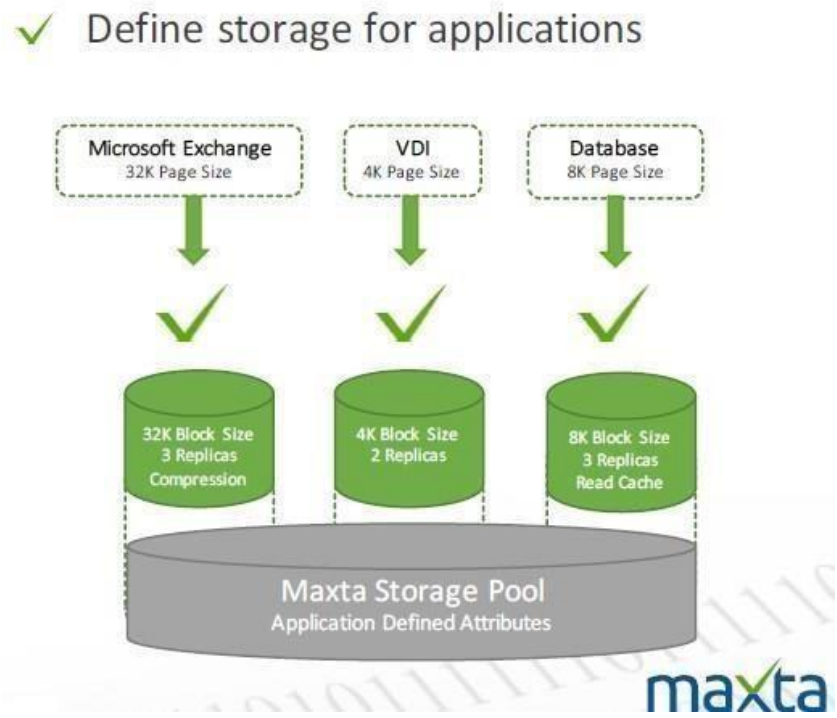
Q11: How can Maxta scale up compute or storage independently while the cluster is supporting a live production environment?

A11: [KS] Yes, Maxta support a non-disruptive way to scale-up and scale-out storage. Customers can non-disruptively add/remove/replace drive and nodes. Maxta also supports the ability to non-disruptively replace SSDs in case of failures. Maxta also supports nodes that do not have uniform storage.

Q12: With a 12-node cluster with 48 cores (max), how many VMs can one core support on average? Use three (3) SQL 2012 servers as an example!

A12: [KS] Maxta storage platform reserves 4 cores and 20GB of memory per node/host. All the remaining CPU and memory resources are available for applications to use.

Q13: Nimble storage is able to perform the following task by modifying a storage profile to match a block size for each different application, as illustrated in Maxta's define storage picture below. Can you tell me the major difference between Nimble Storage and Maxta?



A13: [KS] With Nimble customers have to set this policy at a LUN level. All the VMs deployed on that LUN will have the same policy.

With Maxta the policy is set at a VDISK level which is much more granular compared to Nimble storage. This is very helpful where an application best practice is to set a different page size for transaction logs and data files.

Additionally, Maxta also allows ability to turn on/off compression and read caching at the VDISK level. Once again this is very helpful in a database environment where enabling read caching for transaction logs does not add value or an application stores uncompressible data.

Q14: Under what kind of conditions would you recommend to disable "Striping" via a VM Creation Wizard?

A14: [KS] Maxta supports two types of striping.

Horizontal striping: In this mode the data is striped across all the nodes in the cluster. For example, if the VMDK is 900GB and there are 9 nodes in the cluster, 100GB of data will be spread across the 9 nodes in the cluster. The replica for these will also be striped across all the nodes in the cluster.

Node 1: Primary copy A = 100GB, Secondary Copy B' = 100GB

Node 2: Primary copy B = 100GB, Secondary Copy A' = 100GB

Etc.

Vertical striping: In this mode the data for the entire VMDK of 900GB will be on one node and its replica will be on the other node.

Node 1: Primary copy A = 900GB

Node 2: Secondary Copy A' = 900GB

Etc.

Data mining application are well suited for vertical striping.

Q15: Let's assume Maxta software is deployed to a cluster that comprises 12 Cisco C240 M4 nodes running VMware Hypervisor. Does the Maxta Software Defined Storage support a unified storage, namely, Block and NFS or SMB file system?

For example, NetApp FlexPod Unified management requires 3rd party software solution with its own infrastructure, software and licensing cost.

A15: [KS] Maxta delivers VM storage. Customers manage VMs and do not manage storage. In a VMware environment the only way to access storage is through VMFS or NFS. Maxta utilizes NFS to present the storage to hypervisor.

Note that this is completely different compared to a traditional NFS storage platform. Our NFS traffic under normal circumstances is local to the server and the traffic does not go on the wire like traditional NFS storage.

Maxta and ESX kernel communicate on the same vSwitch on the node and thus it is just a buffer copy within the node/server. So we should not identify Maxta as an NFS storage but just as VM storage.

Q16: How many reference architectures does Maxta have?

For example, Cisco UCS has many reference architectures such as (1) SmartStack – an integrated infrastructure solution by Cisco and Nimble; (2) FlexPod - an integrated infrastructure solution by Cisco and NetApp; (3) VersaStack - an integrated

infrastructure solution by Cisco and IBM. All those names and solutions mentioned here are often referred to as converged integrated infrastructure solution because servers and storage are a separate unit.

A16: [KS] Maxta has reference architectures for almost all the server platforms which highlights the interoperability testing. Specifically, with respect to Cisco we have completed the CVT tests.

Q17: Let's assume that all 12 4TB SATA disks are fully installed on each node with a replica factor 2, the Maxta cluster system has 144 spindles. Do you think HP 3PAR 7400 system with 144 4TB spindles will take much less time than the Maxta's when rebuilding one failed drive (4TB) with 50% data (2TB) on the disk takes place because 3PAR has a massive parallel storage architecture?

A17: [KS] Maxta treats the entire cluster as a spare cluster. There is no concept of a single spare drive, thus there is no concept of a hot disk. We distribute members of a VM across many drives in the cluster. So a disk can have members from multiple VMs. When we have a replica factor of 2 it is not a 1-1 mirror of disk drives. When a disk fails we read the replica member from many drives and write it to many drives. So it is a many-to-many read and write operations when we rebuild due to a failed drive.

The concept is very similar to the 3PAR array where a data set is divided into chunklets and that gets distributed across many drives. So when a drive fails they read from many drives and write to many drives since they have a spare space allocated across all the drives in the cluster.

Summary, both Maxta and 3PAR have a massively parallel storage architecture.

Reference:

One of the action items from the meeting was to provide you with the calculation of time it takes to perform a resynchronization when a node fails. I have included the calculations and the assumptions.

I am also including our technical whitepaper. Please let me know if you have any questions.

Assumptions

Total Amount of data to be resynchronized = 12 TB
Maxta Storage Network = 10Gbps
Rebuild priority = Average (set to 5 on a scale from 1 to 10 where 1 is the slowest)

Calculation

Average case
Resynchronization bandwidth = 200MB/sec (This means in 1 sec we can synchronize 200 MB)

In 1-Second resynchronized capacity = 200 MB
In 1-Minute resynchronize capacity = (200 x 60) = 12000 MB = 12 GB in 1-Hour resynchronize capacity = 12GB x 60 = 720 GB to Resynchronize
12TB = 12000 GB = 12000/720 = 16.6 Hours ~ 17hours

Good case

Resynchronization bandwidth = 500MB/sec (This means in 1 sec we can synchronize 500 MB)

In 1-Second resynchronized capacity = 500 MB
In 1-Minute resynchronize capacity = (500 x 60) = 30000 MB = 30 GB In 1- Hour resynchronize capacity = 30GB x 60 = 1800 GB (1.8 TB) to
Resynchronize 12TB = 12000 GB = 12000/1800 = 6.6 Hours ~ 7hours

Worst case

Resynchronization bandwidth = 100MB/sec (This means in 1 sec we can synchronize 100 MB)

In 1-Second resynchronized capacity = 100 MB
In 1-Minute resynchronize capacity = (100 x 60) = 6000 MB = 6 GB In
1-Hour resynchronize capacity = 6GB x 60 = 360 GB to Resynchronize
12TB = 12000 GB = 12000/360 = 33.3 Hours ~ 34hours

Regards -Kiran S

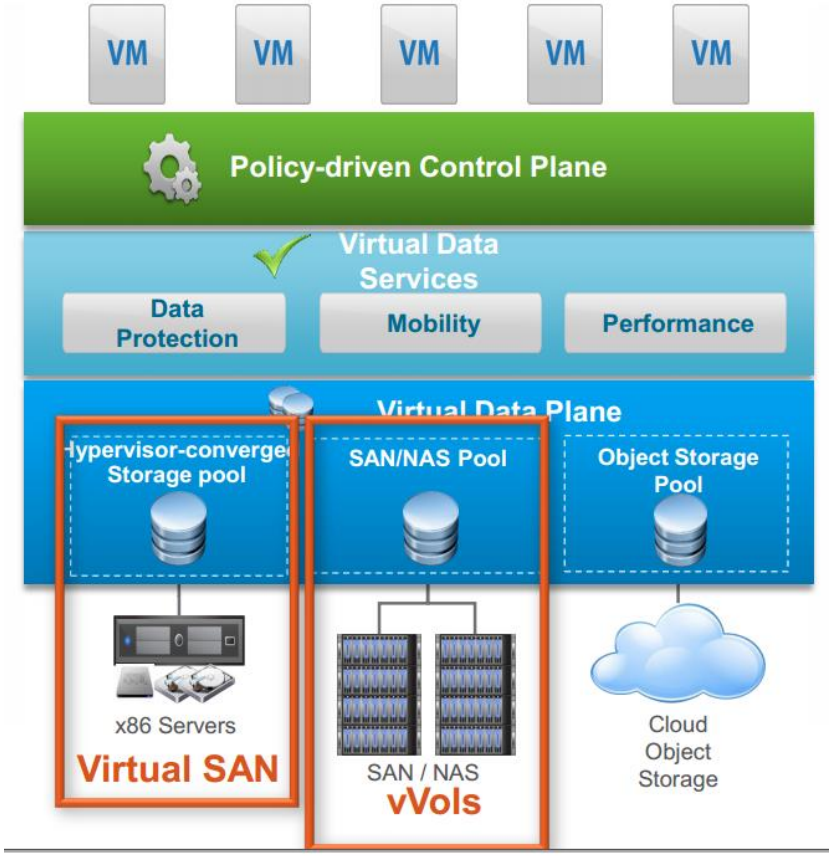
Jeremy's response:

I believe that the 3PAR rebuilt-in and passive parallel process for rebuilding a failed drive might be much faster than Maxta's because of (1) the rebuilding process for a failed drive within 3PAR system does not have to go through external switches that always introduce additional hops and latency; (2) 3PAR uses the 16Gbps Fiber Channel, while Maxta uses 10GbE (10Gbps) that has overhead from Ethernet protocol – see Carrier sense multiple access with collision detection for details.

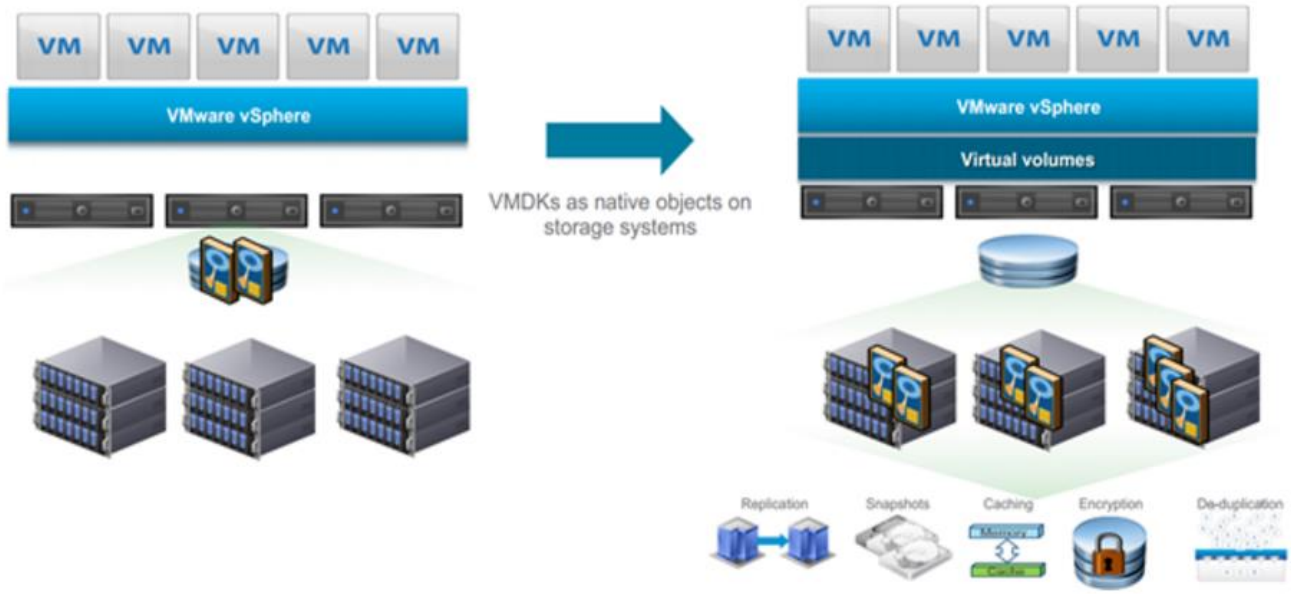
Q18: Visibility - Telemetry Data Visibility for each VM - Can you tell me how Maxta provides more statistics than vVOLs in details?

For example, Tintri's VM aware or App aware storage capability is able to display more granular storage statistics per VM level. The vVOLs technology, which only supports VMware OS and makes storage arrays aware of individual VMDK files from which more granular info can be obtained, provides a similar feature in the aforementioned VM aware storage (see the second picture below for details.)

Note: In a converged infrastructure, the vVol or Tintri's VMstore is used to see a granular visibility, throughput, IOPS, and even latency across the three segments (host, network and storage) for every virtual machine – an end-to-end visibility in a single pane of glass without needing a third party tool.



With VVol most of the data operations can be offloaded to the storage arrays. VVols goes much further and makes storage arrays aware of individual VMDK files.



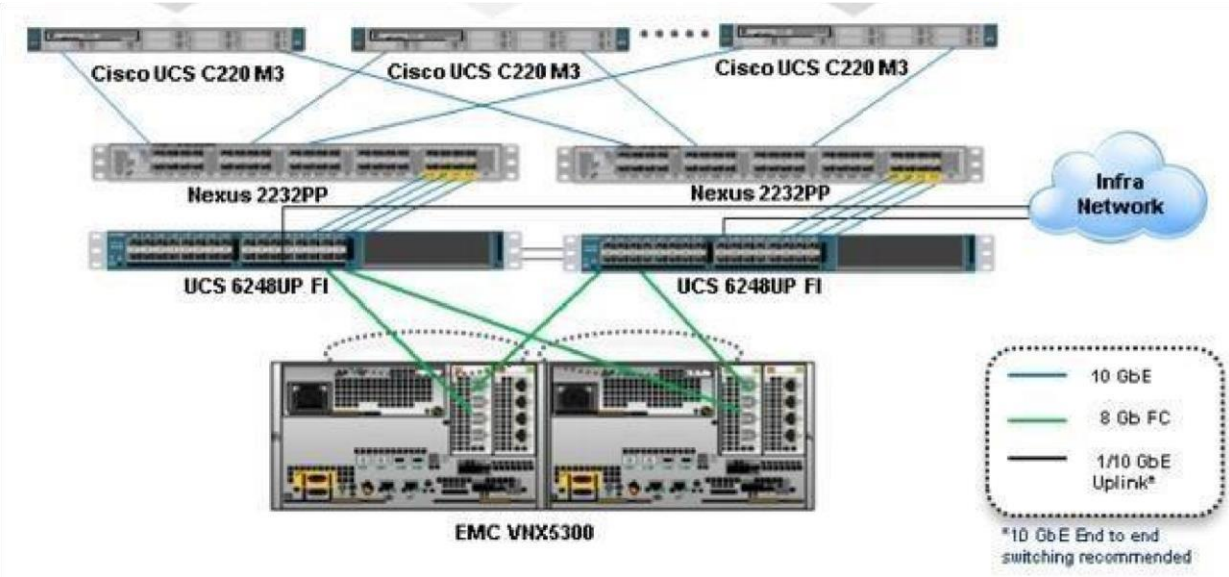
Courtesy of www.wooditwork.com | [What's New in vSphere 6.0: Virtual Volumes]

A18: [KS] The main goal of VVOL was to bring parity between NFS and block (VMFS). In block storage there was no way to get VM level granularity since VMFS was a closed filesystem. The only thing we would know is the number of VMs that were on a LUN. Even before VVOLs VMware tried to provide some level of granularity in a block storage environment by supporting VAAI protocol for block with features like “Hardware assisted locking” where the block storage devices could lock a region of a LUN and not the whole LUN.

Maxta supports VM level management and VM level stats which is much better than vVOLS.

Q19: What’s the difference between Maxta’s hyperconverged and converged infrastructure that comprises VMware vSAN, which can provide a storage choice for each customer?

A19: A converged infrastructure still comprises of complex layered components separately, as illustrated in the picture below, while the Hyperconverged infrastructure combines both commodity servers and storage into one unit. As a result, an additional layer from servers to an external storage system can be illuminated.



On the other hand, converged infrastructure can support more than 200 VMs easily, a demarcation line from Gartner’s report titled “**Simplify the Midmarket Data Center with Hyperconverged Infrastructure Solutions**.” (Report published June 24, 2015)

Below is an excerpt from the above Gartner report (emphasis added):

"The costs associated with refreshing data center hardware, at midmarket scale **(80 to 120 production virtual machines (VMs) and 30 to 50 TBs of storage, coupled with their high levels of virtualization (80% to 90%)** are piquing midmarket I&O leaders' interest in HCIS."

[KS] Maxta is a much superior solution than VMware vSAN. I am including some of key benefits that

Maxta provides compared to VMware:

vSAN ✓ No Choice in Hypervisor

— VMware vSphere only

✓ Poor implementation of storage data services

— Massive performance degradation when using their snapshot

— Compression is 2:1 or nothing ✓

Reduced resiliency

— SSD failure will require a full node rebuild ✓

No end-end data protection

— Cannot prevent silent data corruption

✓ Rebuilds due not node/disk failures cannot be prioritized

— Consumes a lot of resources, affecting production IO

Q20: Can you tell me what does the enterprise class data services mean?

A20: Enterprise class data services often refer to as (1) high availability; (2) resiliency; (3) scalability; (4) multi-tenancy; (5) replication; (6) inline deduplication; (7) multilevel data protection; (8) thin provisioning; and etc.

[KS] I would also include

Data Integrity

Call Home Capability

Auto Rebalance when drive or nodes are added/removed/replaced

Q21: How does Maxta use SSD as Read Only or both Read and Write?

A21: [KS] Maxta utilizes SSD for read/write cache and also for metadata. Maxta has introduced a new class of device called Metadev that will be used to store metadata on the SSD.

Q22: Does Maxta resolve an I/O Blender Effect in a VMware environment?

A22: [KS] Yes, Maxta addresses the IO Blender Effect. Maxta File System is a log structured file system leveraging flash for read/write caching and metadata. Maxta file systems writes sequentially to the SSDs and also de-stages data sequentially to the HDD.

Q23: Which cloud provider Maxta Cloud (MxCloudConnect) rely on?

Note: Google spend 10 billion to beef up its data centers in Google cloud in 2016, while Microsoft spend 7.5 billion in the past four quarters. Source: WSJ

Microsoft's Cloud Trip Is No Free Ride (Source: WSJ)

A23: [KS] Additionally, Maxta has developed MxCloudConnect, a cloud based proactive reporting and alerting system which is included as part of the standard MxSP installation. This service provides administrators with access to a web-based portal which has information on alerts and events across multiple Maxta clusters. Maxta's support team uses MxCloudConnect to check on the state of all MxSP deployments and respond immediately to any alerts. This quickens response time on support calls and reduces customer headaches. MxCloudConnect can also be used by managed service providers to monitor client deployments of MxSP and ensure that all clusters are functioning properly.

Q24: May you explain to me how Maxta Write I/O Path & Read I/O Path work?

For example, Nimble Storage uses Cache Accelerated Sequential Layout Architecture (CASL) with in-line compression in NVRAM to form a full write stripe, CASL writes the data to disk and will determine whether it should also write to SSD for cache for faster reads.

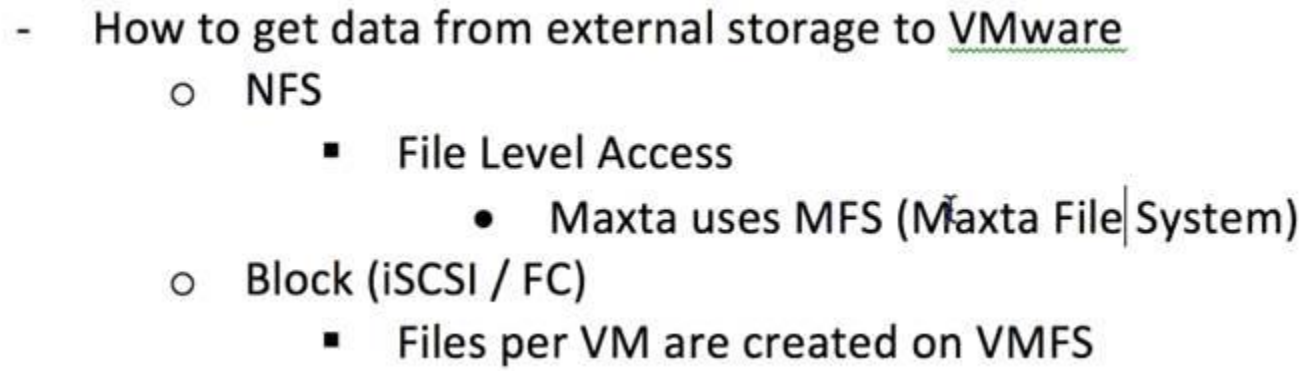
A24: [KS] A write I/O proceeds as follows:

1) A Maxta write I/O is processed through the local MxSP instance closest to the VM guest.
2) The data is striped as widely as possible and replica copies are created. The data and its replicas are written across the network to MxIL residing on flash media on at least two different nodes. 3) Each node acknowledges that data has been written to flash media.
3) The write I/O is acknowledged back to the application after all data, including replicas, has been written to flash.
4) Data is eventually de-staged onto spinning disk.

A read I/O proceeds as follows (note that if the original copy of data is marked as stale then the data is fetched from its replica copy):

1) A Maxta read I/O is processed through the local MxSP instance closest to the VM guest.
2) If the data is dirty (has not yet been moved from MxIL to spinning disk) then it is read directly from MxIL.
3) The next level of read caching is the MxSP memory.
4) If there is a cache miss from memory then the data is fetched from the read cache residing on flash media.
5) If the data is not found in flash, MxSP will read the data from HDD.
6) The read I/O is acknowledged back to the application VM after striped data has been read from all nodes.

Q25: Can you tell me a use case when enterprises must purchase a NAS (NFS/CIFS) gateway when a block-level storage (e.g., VMAX and 3PAR) is deployed as a converged infrastructure, in other words, does any enterprise still need to purchase it when Maxta HCI is deployed?

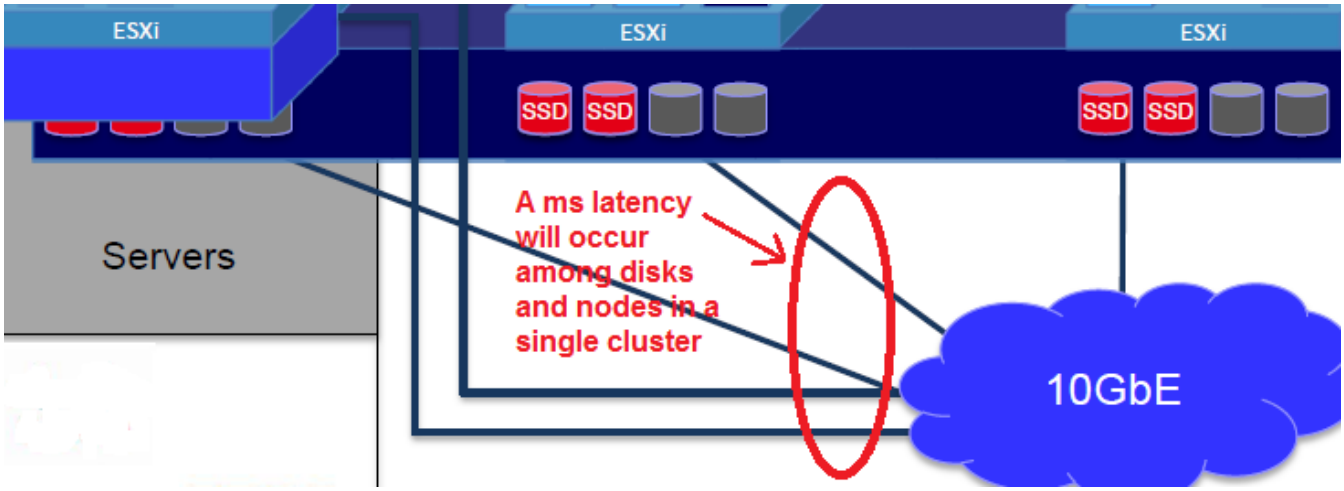


Note: Both 3PAR and Dell Complement engineers told me that an NFS/CIFS gateway must be purchased if their block-level storages are deployed when the NFS/CIFS access is required.

A25: [KS] Maxta presents a VM storage to the guest applications. There is no need to have a NFS/CIFS gateway to access Maxta storage.

Q26: Can you tell me which architecture (Maxta vs. HP 3PAR) will be able to recover a failed drive faster?

3PAR does not rely on an external Ethernet network to recover a failed drive. Instead, it relies on its internal direct-attach fiber channel connections, often known as HP Virtual Connect FlexFabric modules, to connect directly to its storage system. As a result, there will be no Ethernet collision and a shared network bandwidth issue that in turn a very fast performance will be achieved during the recovery process by using all available drives configured for the volume – **a massive parallel architecture**.



In addition, 3PAR has a built-in dedupe capability, meaning, it only sends a unique block, never sends a repeated block to a newly inserted drive, but Maxta must send duplicated blocks, while rebuilding takes place.

With the above facts, a failed drive recovery time from 3PAR will be faster than Maxta's.

The more disks installed in the system, the shorter the recovery time because the contents written to each disk will be spread to all available disks configured for the volume!

There will be a lot of I/O requests or activities sent from each VM (guest OS) among all nodes within a cluster: each VM guest OS to the **Hypervisor->Host->Network->Another VM** from another node via an external network and etc.

Infinio (www.infinio.com), a market leader in storage acceleration, claims to deliver 20X improvement in latency by keeping storage traffic within each node (or server side cache), avoiding the massive unnecessary traffic moving from the node to an external storage due to extra hops and high latency or **microseconds (us)** vs. **milliseconds (ms).** The same concept applies here.

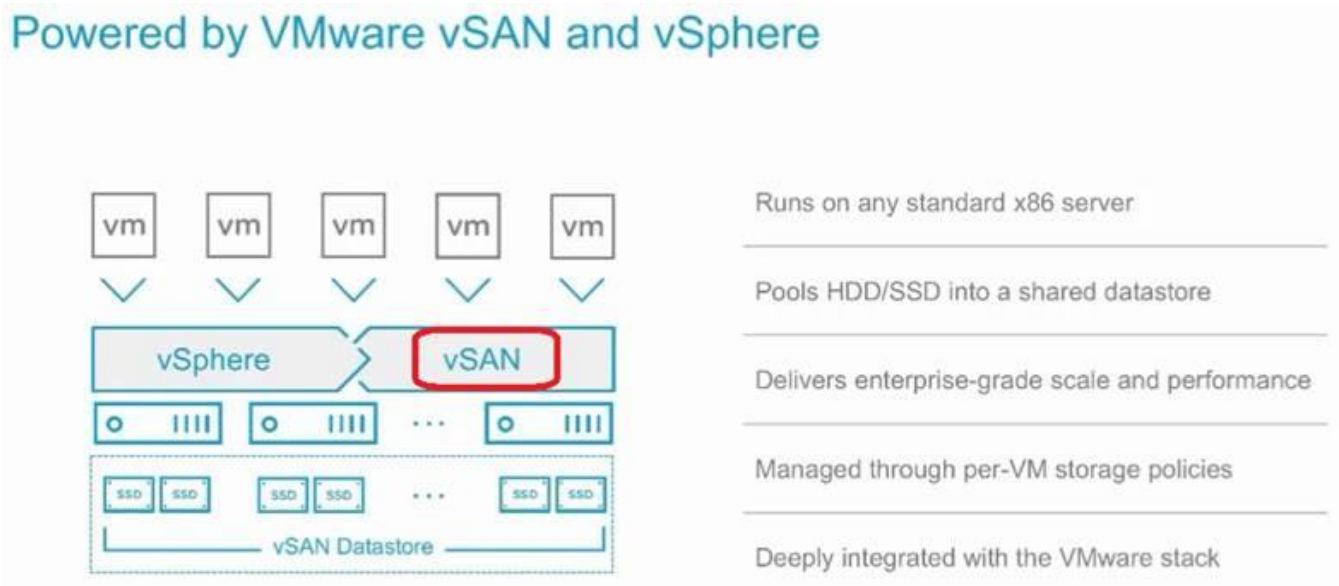
A26: 3PAR storage platform has a mesh backplane that communicates at 4GiB/sec. That is good and very beneficial between node-node communication. For the rebuild of data the bottleneck is not on the backplane. It is at the speed you can write to the disk.

See HP 3PAR REPLACING A FAILED DISK for details.

Q27: Since many on-prem clouds rely on vSAN, which works with any x86 servers and pools all local disks (either SSDs or HDDs) among all hosts (or nodes) together to form a storage pool, what's the advantage to use the Maxta technology?

See the link to the most popular ready nodes - http://vsanreadynode.vmware.com/RN/RN

Note: The vSAN is already installed along with the vSphere, as long as clicking on a check box to turn it on, as illustrated in the screenshot below:



A27: To be clear VMware VSAN is not an included feature in vSphere standard, advanced or enterprise plus. It is licensed per socket with separate support contracts.

Maxta is a superior alternative to VSAN for a number of different reasons:

1. Resiliency
    a. Maxta can withstand an SSD failure without the need to rebuild a node or a disk group across the network. VSAN cannot achieve this and sees crippling performance during SSD failure.
    b. Maxta can take unlimited snapshots without penalty to provide restore points of your VMs throughout the day. VSAN has massive performance hits when using their snapshots and they limit you at 32.

2. Effective usable capacity in a hybrid flash setup
    a. Maxta uses a powerful variable block in-line compression algorithm reducing data footprint 2 to 1 on average.
    b. VSAN does not offer any data reduction with a hybrid install. If you are looking at a 4 node VSAN setup with hybrid and you want 30TB usable, you will need to purchase roughly 90TB of capacity and associated SSD to achieve this.

3. Superior data integrity
    a. Maxta employs an end to end metadata based checksum methodology to ensure data corruption does not happen due to block rot or bad block reads.
    b. VSAN simply appends a checksum to the data block. This will not help in the event of a block read misdirect. *a very large coffee company saw massive data corruption on VSAN and is actively working on a shift to Maxta.

4. More flexibility and choice
    a. Maxta controller VM runs in user space, not in the kernel. This allows us to use the much more flexible Vsphere HCL for server options vs. the limited VSAN HCL. This is often very beneficial when customers have some existing investments that they want to use.
    b. Maxta is not limited to VMware. Maxta is not restricted to 1 hypervisor and will provide the flexibility to move to KVM for example if that becomes an option down the road.

These are just a few important items I wanted to point out. On a demo I can show you how our comprehensive UI is so easy to navigate and provides VM-centric policies that allow you to align the storage block size with the application page size for additional storage efficiencies.

Challenge:

1. Maxta is an emerging HCI software vendor. As a result, it will take time to prove its platform.

2. In the public sector here, it is almost impossible to use a white box vendor (e.g., Super Micro or Quanta). Therefore, the server vendors will be Cisco, Dell or HP, while Lenovo has a difficult time to enter into the public sector.

3. Maxta HCI does not offer a block-level dedupe. It will have a disadvantage in comparison to most HCI vendors that offer a block-level dedupe (e.g., SimpliVity). Please note that all HCI vendors that rely on RAIN technology will have **a massive storage overhead** - 2X with RF2 or 3X with RF3, respectively, including SimpliVity. For example, if 100TB raw storage is purchased, the usable disk space is only at 33.3TB with RF3 prior to any data reduction technology applied. Therefore, the data reduction technology (or software differentiation of each HCI solution) will be used to minimize its massive storage loss – one of the key differentiators –when evaluating the HCI solutions since the servers (hardware) is highly commoditized.

4. Maxta does not offer any replication technology at this time, meaning it must rely on a third party solution such as Veeam, although it offers a local snapshot.

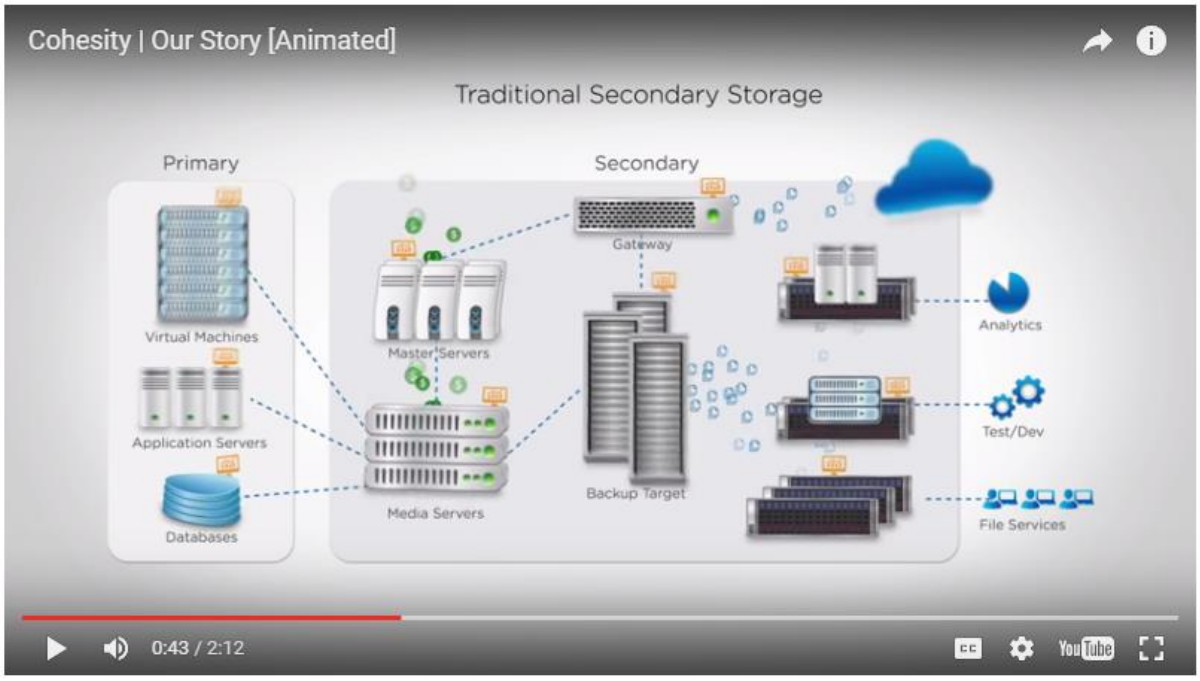
Note: Maxta claims that it is Rack awareness, meaning a cluster can be across two (2) sites as long as the latency between the sites is <=5ms. In the real world scenario, it is not practical, but problematic.

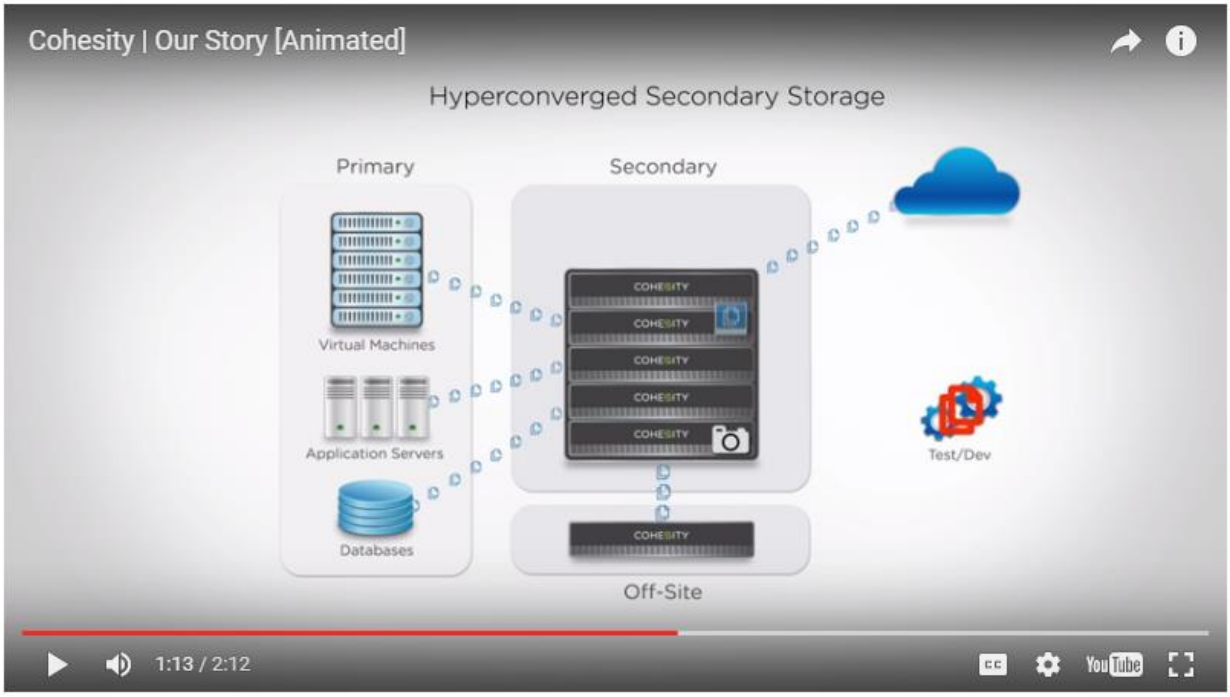
5. SimpliVity will pose a big challenge to Maxta's solution because: (1) it provides highly efficient deduplication technology or appliances, results **90% capacity savings** across all of its customers' storage including backup, whose effect is referred to as HyperEfficient, while Maxta lacks a true block-level dedupe, thus, it can only gain the best data reduction at **2:1 (50%)** on average through its inline compression;

   Data efficiency/Resiliency:
   · More than 1/3 of SimpliVity customers see >100:1 data efficiency
   · Typical ratio is at 40/50:1
   · Efficiency ratio 10:1 guaranteed.
   · RF3 requires only 1 node minimum, 2 for physical redundancy (ability to fail more than 1 drive at a time)

(2) backup software (e.g., file/folder-level as well as VM-level backup/restore capabilities); (3) replication; (4) WAN optimization technologies; (5) cloud gateways; and (6) a secondary storage that can achieve two birds with one stone, meaning providing both HCI as a primary storage as well as a secondary storage for backup and restore in a single platform. As a result, a potential traditional secondary storage either will be completely or partially eliminated, as illustrated in the two pictures below, that in turn reducing massive software licensing costs as well as infrastructure costs with a unified single pane of glass for management.

Note: 51% of SimpliVity customers choose to get rid of the traditional secondary storage at this time.





Source: www.Cohesity.com

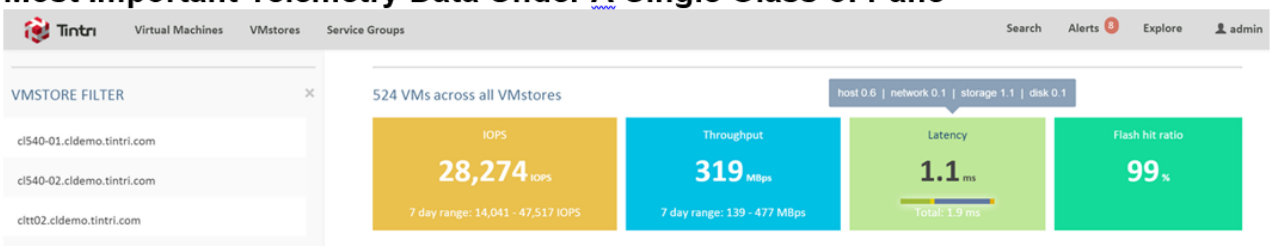
6. Other HCI vendors try to include networking in addition to computing and storage: Here's the pertinent part from the recent TechTarget report dated June, 2016 (emphasis added) - *Source: TechTarget-Comparing-Hyperconverged-Infrastructure-Options.pdf*:

"Hyperconvergence replaces the need for disparate hardware components in a legacy stack with a software-centric architecture that includes compute, storage, ***networking,*** *hypervisor and unified management in a single box."*

7. Although Maxta presents a VM storage to the guest applications and manages VM-level storage via a single management console, it lacks the capability to view a flash hit ratio.

Below is a screenshot from a single pane of glass via Tintri management console to view most important telemetry data, including a flash hit ratio:

Most Important Telemetry Data Under A Single Glass of Pane



8. Maxta lacks the cloud connection capability, for example, to move a VM to and from the cloud such as Amazon cloud, since most enterprises are looking for a hybrid cloud solution.

9. Some advantages from Maxta may not apply to most in Public Sector. For example, the choice of any server vendors are almost limited among Cisco, Dell and HP.

10. Cisco introduced its first hyperconverged product based on "RAIN" technology, also known as "Log Structured File System" and named as "HyperFlex Systems" in March 2016. Cisco markets its HCI as "**2nd Gen Hyperconvergence**" that includes:

- Unified Fabric Networking, as illustrated in a picture below or see a video for details.
- Pre-loaded HX Data Platform, a core software, which is designed for distributed storage to offer Data Services and Storage Optimization
- Dynamic Data Distribution - Elastic

Note: Cisco B200 M4 blades (or servers) are the #1 market share in the U.S. in 2016, while its UCS platform has 48,000 customers in March 2016.

- Independently scale-up and scale-out
- Security
- Call Home and Onsite 24x7 support
- Pointer-based snapshot
- Near Instant Clones
- Inline dedupe and compression
- Self-healing
- Single pane of glass for management

Special note: *Cisco acknowledges that HyperFlex is not designed for low latency apps such as databases and operational and mission critical applications. It is designed for operational simplicity – see https://www.youtube.com/watch?v=BVMpcitCQcw for reference!*

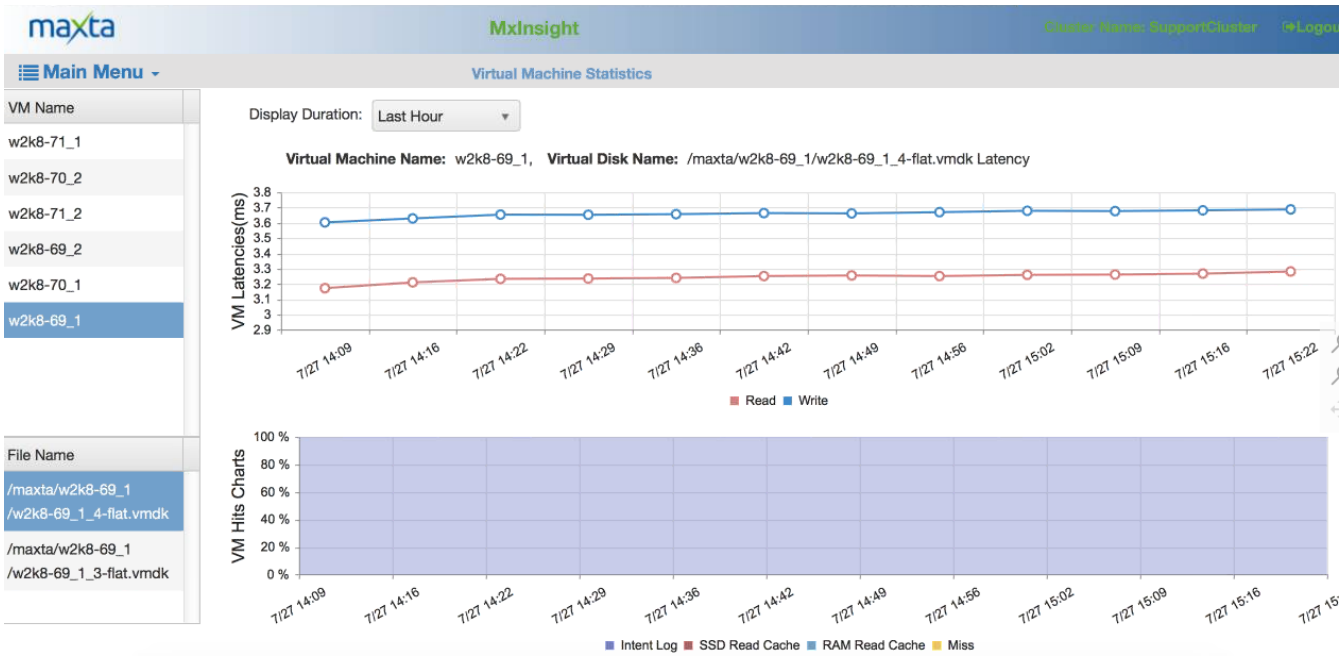
11. Most enterprises prefer to have a single point of support, meaning avoid a multi-vendor support for a quicker resolution.

In summary, Maxta hyperconverged solutions can provide:

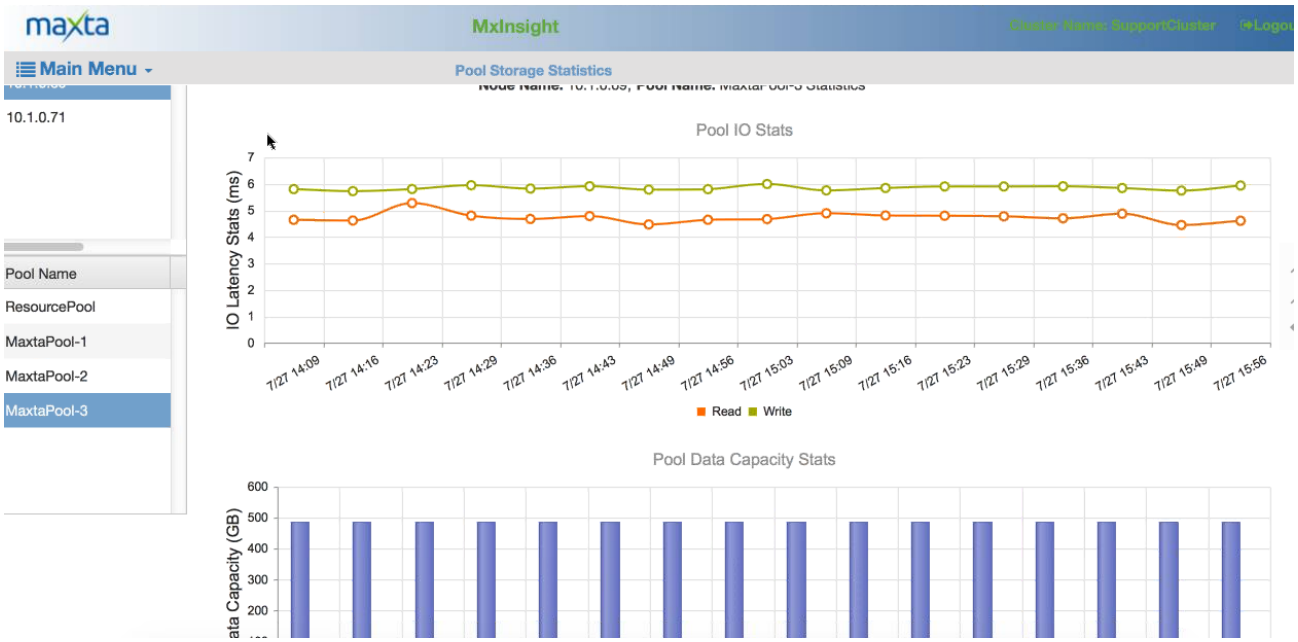
- Eliminate the Storage Array
- Any Server Vendor
- Rebalancing of data when you add nodes/drives

Note: SimpliVity does not offer a rebalancing of data when nodes/drives are added. Applications are stuck with what they are provisioned with. Customers will have a tremendous challenge in scaling beyond what they have deployed with.

- Administer VMs, not Storage (See the following pictures for details)
- Self-Healing & Self-Optimizing
- Lifetime Transferable Maxta License
- Maxta also uses a dedicated tab with a vCenter console to view a granular info, as illustrated in the screenshot below:



Viewing an individual VM and VMDK statistic



Viewing an individual disk statistic

In theory, Maxta can offer unlimited scale-out capability per RAIN technology, but it only certifies 12 nodes per cluster at this time.

For those SMBs customers, which would like to keep its existing secondary storage for their backup and restore and replication, you can consider Maxta as your potential vendor during your POC process.

It is worth noting that Maxta has an advantage to adopt a newer technology quicker than other HCI vendors due to its nature of a software-only HCI vendor. For example, it can quickly support VMware vSphere 6, while it took much longer time for other HCI vendors to support it (e.g., it took Symplivity one year to support it.)

Remember, always seek the lowest TCO and highest ROI whenever possible when evaluate your next hyperconverged solution by considering the following facts:

- vSphere licenses
- Total cost of ownership.
- Footprint /Power cooling
- Ease of management, etc.

Caution:

Maxta is an emerging software-only hyperconverged vendor and has not been in the market long enough to establish a reputation in HCIS. It will take time to prove its solutions that will appeal to enterprises, especially to the public sector, which usually always purchases the best products from the top-tier vendors such as EMC, Cisco and etc.

Here's the pertinent part from a 451 Research report dated Oct. 20, 2014 - 451_Research_Nutanix_20Oct14.pdf - (emphasis added):

"The startup and specialist space is also getting crowded: SimpliVity probably trails second behind Nutanix in terms of deployments and awareness. Although Nutanix recognizes SimpliVity as a lower-end play, it claims it will struggle to scale into larger enterprise deployments. Other startups include software-only player **Maxta**, open source specialist NIMBOXX and a couple more players (at least) **still in stealth**."

Maxta has not yet offered any HyperGuarantee similar to the SimpliVity's HyperGuarantee - the Industry's Most Complete Guarantee!

The employees turnover rate is very high in this industry. Therefore, the public sector usually looks for a more reliable vendor for a long-term support.

Recommended Reading:

- **Success –** A project delivers expected business value such as measurable improvement to revenue, profits or net income, automation to improve productivity, new product release, reduce inventory costs or some other targeted outcome.
- **Failure -** A project that did not meet or exceed expected business value.

Source: Paul Dandurand, CEO of PieMatrix and Lawrence Dillon, Practice Leader of ENKI LLC.

- TechTarget White Paper titled "Comparing Hyperconverged Infrastructure Options for Virtualized Environments", which has the following comparison chart:
- Gartner Magic Quadrant for Integrated Systems - 11 August 2015 | ID:G00266749

	Cisco HyperFlex	NetApp FlexPod	Nutanix XCP	SimpliVity OmniStack
Data Efficiency	All data deduplication and compression are done in-line with the same CPUs used for production workloads on a "best-effort" basis, which means that if the controller is busy, it may not be done at all. It is difficult to predict how much data efficiency will actually be achieved in production environments.	Deduplication is not inline. Compression and deduplication are available but recommended to run off-peak to sustain performance.	Fingerprinting of data is done inline, for sequential writes of 64KB or larger. The actual deduplication processing is largely done post-process.	All data is deduplicated compressed and optimized inline globally across all tiers once and forever, globally in 4 to 8KB chunks. Median customer data efficiency is 40:1.
VM-Centric Management	HyperFlex uses vCenter for VM-level management, the HyperFlex management interface to manage the storage layer and UCS Manager to manage the Fabric Interconnects.	Management paradigm is at the LUN level. iSCSI/ Fibre channel networking, LUN mapping and zoning are part of the standard mode of operation. Unified management requires 3rd party software solution with its own infrastructure, software and licensing cost.	VM management has an added cost due to additional licensing and requires multiple interfaces including Nutanix Prism, Prism Central and the individual hypervisor management consoles, making movement of VMs between data centers a challenge.	VM management is provided via integration with VMware vSphere and other management and orchestration software. All management is at the VM level without the complexity of LUNs and SAN concepts.
Data Protection	No built-in backup. Native snapshots are unable to recover a deleted VM and are limited to 30 snapshots/ VM. Replication requires using third-party software.	No built-in backup. Backup requires a 3rd party backup software with its own infrastructure, software and licensing cost. NetApp snaps and SnapMirror do provide local and remote data protection.	No built-in backup. Backup requires 3rd party software. Nutanix does natively offer snap shots and multi-site replication for additional license cost. File level restore requires 3rd party software.	Built-in VM backup, multi-site replication, recovery and cloning, and disaster recovery included natively. SimpliVity also includes file level restore natively.
Resiliency	RAIN-based (Redundant Array of Independent Nodes) architecture. Since data is striped across all nodes, RF3-level protection is standard to protect against the loss of every VM hosted in a cluster in the event two disks are lost, or even one disk is lost while one node is off-line. HyperFlex requires a minimum of four nodes per cluster in production.	Double-parity RAID-DP prevents data loss with double drive failure for SSD and HDD drives.	No RAID, resiliency is based on RAIN with Resiliency Factor (RF) 2 set by default. RF2 only protects against single drive loss or single node loss. Node loss plus an additional drive loss results in data corruption. RF3 is available, requiring significantly more infrastructure and cost investment.	Intra-node RAID6 can tolerate double drive failure on every node for SAS drives only. Multiple copies of data are spread evenly over several nodes for additional resiliency.

Appendix

- SimpliVity does not offer rebalancing of data when nodes/drives are added, meaning, applications are stuck with what they are provisioned with. Customers will have a tremendous challenge in scaling beyond what they have deployed with

- SimpliVity may not be able to protect against silent data corruption. This is a huge impediment to deploy a storage solution

- SimpliVity does not provide application defined storage, meaning (1) the storage cannot be customized based on the application needs; (2) cannot customize page size, compression, read caching, striping and etc. on a per VMDK basis

- SimpliVity offers HyperGuarantee , one of the five HyperGuarantee offers an innovative in-line deduplication and data compression at origin with global namespace and native VM backup

Acknowledgement

Thanks for John Hofdahl, Account Executive, and Kiran Sreenivasamurthy, Sr. engineer at Maxta to provide a web session on July 20, 2016 with a follow-up question-and-answer session via email, as well as a second live demo on July 27, 2016 to have a showcase of Maxta hyperconverged software-only solutions.