

July 1, 2014
Jeremy Li

Nutanix Integrated Systems Overview

Nutanix (www.nutanix.com) is a private company and has been shipping its disruptive VDI technology since 2011 and gained 5,200 customers in 2014. The founder came from Google and developed a new OS from Google's GFS (clustering file system) to the enterprise NDFS (Network Distributed File System) for general public use.

Gartner assessed Nutanix as a Visionary in Gartner's First-Ever Magic Quadrant for Integrated Systems dated June 2014 (ID:G00252466) and named Nutanix product as a highly innovative and scalable architecture that is generationally advanced compared with most rivals. Its solution has a potential unlimited scale out capability because its highly modular designs allow the easy addition of new server and storage resource. It can also address common denominator failure (e.g., a storage shelf is not available). Recently, it provides dedupe and compression capability.

Nutanix has an impressive reference client list across many vertical industries and geographic regions such as U.S. Army 20,000 VDI nodes deployment (ongoing) through many clustering architecture as shown in the link [Nutanix VDI: Example Architecture](#) from which Nutanix VDI Example Architecture can support 20K to 200K+ Power User Desktops, each cluster can have up to 6 blocks (e.g., [3061](#)) that leads to 24 nodes.

Click on the link <http://www.nutanix.com/the-nutanix-solution/tech-specs/#3342> for Virtual Computing Platform Specifications for Each Node, and [a Nutanix Disruptive Technology Video](#) for details.

A few highlights are listed below:

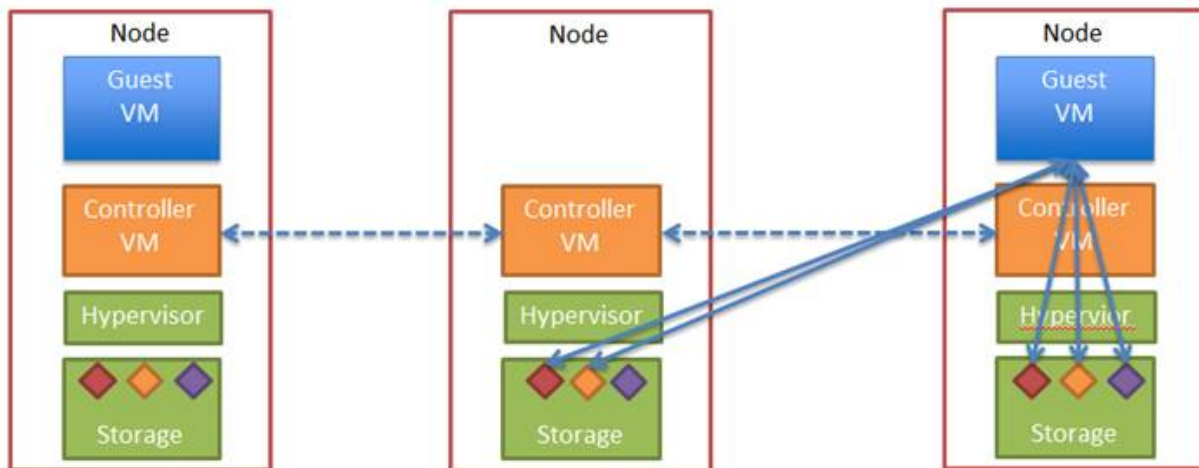
- Nutanix Virtual Storage Controller (Controller VM in the video) is loaded on each node in the cluster
- No more needs for expensive centralized storages, meaning Nutanix eliminates the need for network-based storage architecture, such as a storage area network (SAN) or network-attached storage (NAS)
- Nutanix distributed file system aggregates local storages across all nodes, creating a single storage pool that can be partitioned into one or more DATASTORES. Those DATASTORES presented to the Hypervisor, using the standard NFS protocol to provide storage to all VMs.
- Because the Hypervisor communicates to the Nutanix software exactly as what a traditional storage array, there is zero change to the virtual environment. As a result, there is no need to have to configure LUNs, Volumes and RAID groups that require any company to hire many expensive storage administrators to configure and manage them constantly
- NDFS converges all compute and storage resources into a single integrated system. This allows all the storage from all local disks to be presented as a single pool of shared storage, meaning presented to the hypervisor to run VM's

- The shared storage from the pool, any customers can have enterprise grade features such as Higher Availability (HA) and Disaster Recovery (DR)
- Additional nodes can be added seamlessly to scale out linearly. The advantage is to allow its customers to buy the capacity needed today and scale out as needed basis. This approach will reduce the total cost of ownership (TCO)
- Each Nutanix 2U appliance, also known as block, can have up to four nodes, with virtualization software pre-installed and ready to run out of the box (Turn-key solution)
- Each server node integrates a virtual storage controller to manage storage resources across the cluster

Below are from a Q & A session during the Nutanix Technology Overview via Web Conferencing:

Q1: Nutanix system design has a high storage overhead; meaning at least 2X disk storage space will be used just for replication 2X data due to its NDFS architecture.

As a result, the Nutanix's architecture has 33% storage efficiency, meaning the same data must be copied into two different nodes for redundancy purpose (HA) in addition to consuming extra IOPS between 50% to 66%, due to another extra overhead from the extra two identical copies for the HA purpose, as illustrated in a screenshot below:



Performance and availability

- Data is written locally
- Replicated on other nodes for high availability
- Replicas are spread across cluster for high performance

Figure 1 Data Replication from a node to the other

A1: This is only if you choose replication factor of 3 which we have just introduced. By default, we use a replication of factor of 2. The data is written locally, then 1 copy is stored on a remote node. To enable RF=3 you need at least 5 nodes in a Nutanix cluster. You can set the replication factor on a per-datastore basis, so you can have RF=2 containers for most of your VM's, and a RF=3 container for the VM's that must remain running in the event of a simultaneous failure (i.e. double disk failure).

It's also important to note that enabling RF=3 on a datastore will only have a 15-20% performance overhead.

Please take a look at the following blog written by one of our product managers:
<http://nutanix.blogspot.com/2014/06/nutanix-40-feature-increased-resiliency.html>

Another useful blog is the Nutanix Bible that I think you mentioned you may have read already, but I'll link it here just in case: <http://stevenpoitras.com/the-nutanix-bible/> computing resource, such as N+1 or more.

Q2: One of the 3 key features of virtualization that lead to better disaster recovery is "Hardware Independence". Can the Nutanix replication technology replicate data from a Nutanix appliance (block) in one cluster into another vendor's storage at another site?

A2: Currently our native Nutanix replication solution works only with another Nutanix block. However, you can leverage other third-party backup and replication solutions running on Nutanix, we have customers running solutions such as Veeam and [Zerto](#), and you can download the Veeam best practices doc here:
http://go.nutanix.com/NutanixandVeeambestpracticesguide_LP.html

Q3: How many snapshots typically can be created in one block?

A3: The limiting factor here is the storage capacity of the block. The first snapshot is a full clone and then incrementally from then on. So if there is a high rate of change, you could potentially use a lot of storage. The amount of storage used per snapshots is shown on the DR screen in the UI.

Q4: How long snapshots can be stored on one node or one block?

A4: Theoretically as long as you want. You define the schedule and also the retention policy of the snapshots. So if you are snapshotting every 24 hours and want to save the last 365 local snapshots, you will have a year worth of daily snapshots. Again the limiting factor is the storage capacity.

Q5: Since the Copy-on-Write (COW) consumes three I/Os while Copy-on-Redirect (COR) consumes only one I/O, can you tell me your COW process and why Nutanix

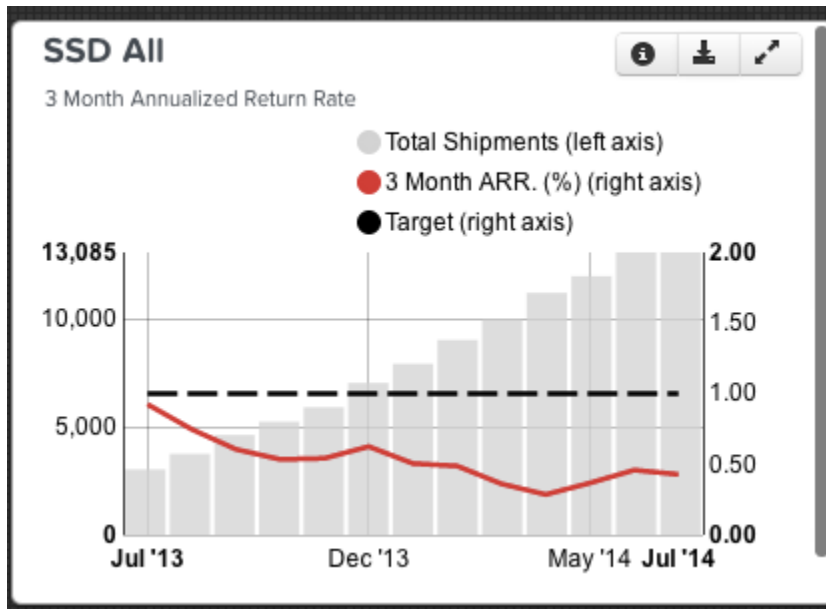
can store 365 snapshots for one year if one snapshot occurs every 24 hours without affecting the performance?

A5: I am working on getting some best practice numbers for how many snapshots we recommend storing. However we do not use LUNs so the above scenario does not apply to us. All the locally attached disks are pooled together as one shared storage resource with our filesystem on top of it.

Q6: In a Nutanix's 3-year history since the product was shipped, how often a solid state drive (SSD) must be replaced on average because Nutanix architecture writes data to SSDs directly without going through inline dedupe and inline compression?

A6: I have attached the MTBF document for the 3050 series. Please see page 4 for the MTBF info for the components. If you require this info for other models as well, let me know.

Here are some real world numbers; these are the past 3 months of SSD return rate. The annualized return rate is currently at 0.43%.



Q7: May you tell me the difference between the Nutanix's eMLC and a consumer version of the MLC available from Fry's Electronic?

A7: We are using Intel S3700 drives which uses HET MLC (which is basically eMLC). The spec sheet can be found here: <http://www.intel.com/content/www/us/en/solid-state-drives/ssd-dc-s3700-spec.html>

Discussion of HET MLC can be found at <http://www.tomshardware.com/reviews/ssd-710-enterprise-x25-e.3038-3.html>

From this website: "A consumer-grade MLC will only provide 3000 to 10000 write cycles, whereas an eMLC can offer 20,000 to 30,000."

(<http://digital.ni.com/public.nsf/allkb/2672EACD8D0A48F786257BF80035A530>)

Q8: What is the time to rebuild a failed 3TB HDD with 80% capacity full?

A8: Please see the following tech note regarding disk failures and how we compare to traditional RAID arrays:



<http://go.nutanix.com/rs/nutanix/images/nutanix-failure-analysis.pdf>





"Unlike traditional RAID arrays that must undergo a full array rebuild to restore data redundancy, including the associated performance penalty incurred, Nutanix replications happen as background processes with no impact to cluster performance. In fact, as the cluster size grows the length of time needed to recover from a disk failure decreases as every node participates in the replication. Since the data needed to rebuild a disk is distributed throughout the cluster, more disks are involved in the rebuild process. This increases the speed at which the affected extents are re-replicated."

Q9: Can you give me a quote for a 3,000 VDI nodes deployment scenario under the task and knowledge users' configuration referenced below from Nutanix website?

Real World VDI Users

With its new VDI Per Desktop Program Nutanix enables you to purchase advanced web-scale infrastructure the same way you procure virtual desktop software — per desktop. With support for multiple real-world VDI user profiles, Nutanix eliminates complex infrastructure sizing and assures one price per desktop at any scale.



	 KIOSK	 TASK	 KNOWLEDGE	 POWER
Description	Light usage with limited number of applications	Repetitive tasks with small set of applications	Complex projects with heavy internet usage	Heavy users of multi-media, variances in user habits
CPU	1 vCPU	1 vCPU	1 vCPU	2 vCPU
Memory	1.5GB RAM	2GB RAM	2GB RAM	4GB RAM
IOPS	8	30	40	60
Raw Storage	15GB	40GB	50GB	80GB
Usable Storage*	30-50GB	100-200GB	125-250GB	200-400GB

*Usable Storage based of Redundancy Factor of 2 and deduplication & compression benefits in the 5-10x range (Excluding Kiosk).

A9: Below is a chart that gives you an idea about pricing for the different bundles. If we work with a reseller, you will get better pricing. I will follow up with you in two weeks, if you have any questions before then please let us know.

	Kiosk	Task	Knowledge	Power
Configuration	1vCPU, 1.5GB RAM, 8 IOPS, 15GB disk	1vCPU, 2GB RAM, 30 IOPS, 20GB disk	1vCPU, 2GB RAM, 40 IOPS, 25GB disk	2vCPU, 4GB RAM, 60 IOPS, 40GB disk
# of desktops per pack	125	100	75	50
Minimum Purchase (3)	375	300	225	150
List Price per Pack	48k	60k	60k	60k

Q10: May you kindly verify the following list price per chart referenced in your email?

- \$5.4 million for 3,000 Task users
- \$7.2 million for 3,000 Knowledge users

Note: The above list price does not include:

- The VMware connection broker licensing fee, commonly known as VMware View licensing fee, typically at \$125 per desktop
- Microsoft VDI licensing fee such as Windows Enterprise and MDOP (Microsoft Desktop Optimization) regardless which VDI platform will be used

A10: That is correct. However this is list price, if you decide you would like to move forward we can involve a reseller of your choice who will give you a lot better pricing on this.

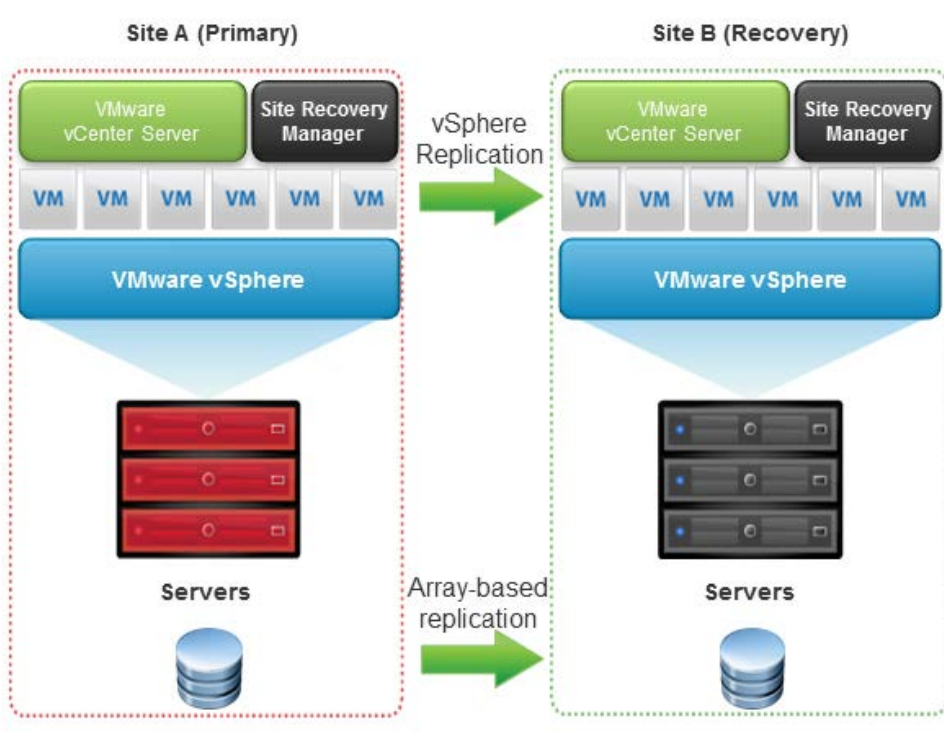
Nutanix does not come with VMware or Microsoft licensing.

Q11: Can you tell me whether Nutanix use vSphere Replication, or array-based replication?

Below is the info regarding vSphere Replication as well as array-based replication for your reference:

- In a physical environment, an array-based replication, snapshots and mirroring can reduce a workload from a server, meaning reducing IOs (data movement) from a server

- In a virtual environment, things are changed because a hypervisor maintains its own allocated storage resources that will be shared among all VMs on a host (also known as I/O blender), while arrays have no idea how many VMs on a single LUN are sharing the storage resources via the I/O blender. Therefore, the VMware VAAI was born in version 4.1 and the vSphere replication (VM-to-VM) were born in SRM 5.5 to address the issues, as illustrated in a screenshot below:



Note: One of the VAAI features resolved the locking and unlocking of small portion of storage resources, also known as region, thus, dramatically increase the storage array performance.

A SSD technology is very good for handling random IOs while the VMs are difficult to deal with random IOs because of an IO Blender issue.

See additional SRM 5.5 info in Appendix.

A11: We can use either our own native DR, or integrate with vSphere's SRM (Site Recovery Manager) using our SRA (Storage Replication Adapter). Either solution requires at least two Nutanix clusters. Since SRM is designed for block based storage, the Nutanix SRA makes the Nutanix storage abstractions look like devices to SRM & leverages our existing replication technology.

Q12: What kind of assurance Nutanix will provide after a VDI deployment?

A12: Nutanix offers [Nutanix VDI Per Desktop and Assurance](#), meaning if additional hardware is required to meet profile parameters after the deployment, Nutanix will provide it at no additional cost.

Q13: On average, what's the percentage of Microsoft Hyper-V and VMware Hypervisor deployed with Nutanix disruptive storage product?

A13: I believe it's something along the lines of 75% VMware, 25% Hyper-V.

Q14: NDFS has a big storage overhead at 100% for RAID-5 and 200% for RAID-6

For example, with 10 HDDs configured at RAID-5, a 10% disk space will be used for parity overhead in most storage vendors' configuration, while 100% penalty for a similar RAID 5 disk fault-tolerance and 200% penalty for a similar RAID 6 disk fault-tolerance in Nutanix (See Figure 1 referenced above).

A14: We have a variety of different models to scale out however you need. If you need more storage and less compute we have our NX-6020 model which has 10TB usable storage per node but less compute - it has 12 total cores and can be configured with as little as 32gb of memory. If you need less storage and more compute you can add something like a NX-3060 node which has 2TB usable storage per node, but is heavier on the compute with 20 cores and up to 512GB of memory. All of our models can be mixed and matched in the same Nutanix cluster.

If you need to add storage this typically means you need additional storage to run more VM's so including compute with storage is necessary. However as mentioned above we have a wide variety of configuration options to best fit your growth needs and we can seamlessly scale out your cluster without disruption to your existing VM's.

Q15: Does Nutanix lack a simultaneous compression and dedupe in real time?

A15: The only thing Simplivity is better than us now is the simultaneous compression which is on our roadmap for the end of this year.

Q16: Do you agree that Nutanix has Integrated System Convergence 2.0 while Simplivity has Integrated System Convergence 3.0 – see the link below for details:

http://www.simplivity.com/components/com_wordpress/wp/wp-content/uploads/Path to Hyperconvergence.jpg ?

Please note that Simplivity Integrated System Convergence 3.0 includes a full backup with a unique 8KB block, meaning its backup or replication software never repeats the same info twice. As a result, each Simplivity full backup is better than a popular snapshot. Therefore, Symplivity does not rely on third party software for snapshots or replications, as Nutanix relies on Veeam and Zerto VR for help.

In order to prove the above concept, Simplivity can show any audience a live demo to illustrate the above concept by showing only 19MB data of the original 469GB data from the U.S. data center will be copied to a Sydney, Austria data center in 10 seconds.

It is worth mentioning that the current latest VMware vSphere Replication that is only designed to protect data or vSphere Replication with SRM 5.5 (orchestration) that is designed to protect both Apps and Data will copy 469GB data from a U.S. data center to a Sydney, Australia Data Center in 8 or 9 hours depending on the WAN bandwidth because vSphere Replication lacks compression. That's why many customers must rely on Riverbed or Silver Peak for help with accelerated replication and etc.

A16: To be answered...

Challenges

1. Leading incumbencies such as VCE (Cisco, EMC and Intel company), Cisco/NetApp are still the largest Enterprise Integrated Systems vendors and largest market share in the VDI market, although incumbencies are still relying on legacy storage technology (LUN, RAID and VOLUMN). However, many incumbencies made significant changes in the latest storage technology by using PCIe based Flash (SSD) first (e.g., EMC VNX 2) and [Dell Compellent Fluid Cache](#), meaning all flash capacity installed in each node or host can be pooled together to form a big pool so that each node can borrow it for use at any time, instead of wasting high percentage flash resources in each node or host. That technique might be able to address a VMware boot storm or any burst mode issue in VDI environments.
2. Generally speaking, VCE and Cisco/NetApp (FlexPod) are leaders in Integrated Systems category in addition to having a good reputation to provide an excellent presales and postsales support.

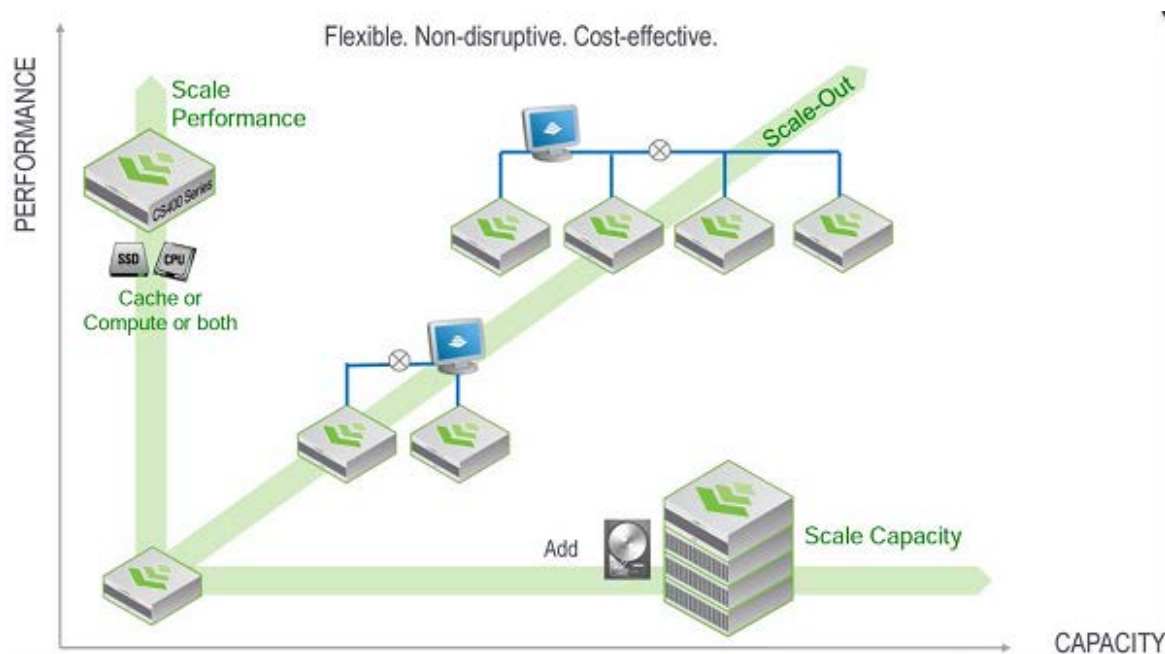
With new EMC Isilon/new OneFS 3.0 in addition to the Xtream IO 3.0 that is the fastest growth array in IT industry with 100 million and 40% scale-out and 99.999 % HA, the new Isilon can provide the followings:

- XtremIO DATA Services is always on, always Inline, not like other vendors that dump data to flash first and process data later when an array controller is idling. Therefore it redefined the IT industry in scale-out storage from which Hadoop that is based on HDFS might shine. It looks like EMC is leading in HDFS
- XtremIO Data Compression is always on and always Inline that, in turn, provide 2X to 4X performance gain for relational workspace workloads
- XtremIO provides writeable snapshots and snapshots of snapshots that are also always on and always Inline all the time.
- XtremIO snapshots achieve the same performance, meaning no performance degradation

Additional info about EMC Elastic Cloud Storage (ECS):

- EMC Elastic Cloud Storage (Object, HDFS & Block) has been in production in Vatican with 3PB objects stored on one full rack.
- ViPR 2.0 - first 50TB controller is free. ViPR SRM 3.5 is downloadable.
- Dynamic Virtual Matrix - Scale without compromise

3. As of today, Microsoft has zero VDI node deployed within the company worldwide, so as IBM. As a result, many private and even many high-security required companies such as defense companies like Northrup Grumman have not yet implemented a single VDI node within their companies due to high cost reasons such as \$2,900 per VDI node vs. a traditional desktop running at \$650.00 per node.
4. Nutanix name is never mentioned in the Gartner Magic Quadrant for General-Purpose, Midrange Storage Arrays as well as General-Purpose, High-End Storage Arrays. This is because Nutanix has 100% storage overhead as a minimum with one duplicate copy, and 200% storage overhead with two duplicate copies that are similar to a RAID-6 configuration that, in turn, can survive two disk drivers failure simultaneously.
5. Since Nutanix snapshot is based on Copy-on-write, it may not be able provide a sufficient amount snapshots most enterprises may seek. See the Q&A 5 referenced above for details.
6. The ideal scale out should be independently. Nimblestorage is able to scale out in three different direction, also known as all 3 axis that in turn can lower TCO dramatically in the Enterprise Data Centre, as illustrated in the screenshot below:



Source: [Nimblestorage](http://nimblestorage.com)

However, Nutanix scale-out solution is only in one direction. Therefore, it is challenge for Nutanix to replace all SAN and NAS in the future due to the reason referenced above.

Conclusion

In summary, Nutanix is a disruptive Integrated Systems company to challenge the incumbent with 5,200 customers in three years since 2011.

Gartner assessed Nutanix as a Visionary in Gartner's First-Ever Magic Quadrant for Integrated Systems dated June 2014 (ID:G00252466) and named Nutanix product as a highly innovative and scalable architecture that is generationally advanced compared with most rivals. Its solution has a potential unlimited scale out capability because its highly modular designs allow the easy addition of new server and storage resource. It can also address common denominator failure (e.g., a storage shelf is not available). Recently, it provides dedupe and compression capability.

As of today, Microsoft has zero VDI node deployed within its company worldwide, so as IBM as well as many other highly-security required companies, for example, defense companies who have not yet implemented a single VDI node within their companies (e.g., Northrup Grumman) because of the VDI extremely high cost reasons such as \$3,000 per VDI node vs. a traditional desktop running at \$650.00 per node, including Microsoft Windows 8.1 professional edition.

However, the VDI solutions can bring many enormous benefits such as in Healthcare and education areas.

Whether your organization requires a VDI deployment, it all depends on an environment by considering total cost ownership (TCO) and return on investment (ROI) carefully!

Therefore, any organization can start to deploy VDI by starting small and growing bigger and bigger via module architecture. Nutanix solution will fit that scenario.

Generally speaking, Nutanix solution should not be applied to General-Purpose, Mid-range or High-End Storage area such as archiving, file and print server and etc., because its NDFS architecture requires to duplicate original data to other nodes referenced in Q&A 1 and will have a 100% storage capacity penalty, and 20% or more on IOPS.

Below is some useful information:

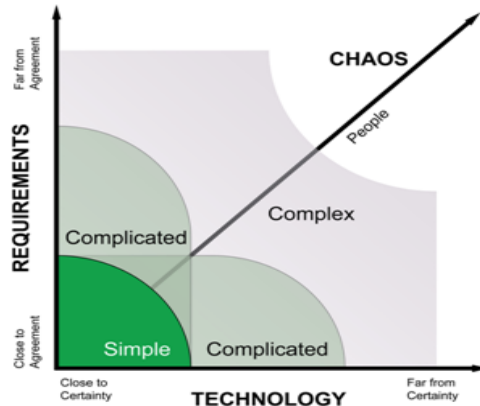
The Stacey Graph - Complexity

Simple
Everything is known

Complicated
More is known than unknown

Complex
More is unknown than known

Chaotic
Very little is known



Source: Ralph Stacey, University of Hertfordshire

Often, today's solutions will not address tomorrow's problems. The added costs associated with labor always challenge any organizations.

Recommendation

Implementation via “Agile” Approach to “Metrics”

Due to a very complicated DVI ecosystem, it is highly recommended to follow the [SCRUM](#) that is a framework for managing the development and deployment of complex products, in order to implement the entire ecosystem correctly. Agile that follows the principle of “Inspect and Adapt” and advocates team empowerment uses Scrum.

A few tips to use SCRUM for achieving your goal by Michael Vincent:

- Don't be tempted to change Scrum
- Scrum exposes inefficiency
- Fix the problem
- Don't shoot the messenger
- Don't reward a wrong person
- Scrum exposes need for change

Source: 1) [Scrum Fundamentals Do It Right](#) by Michael Vincent

2) [Case Study of a Difficult Federal Government Scrum Project](#): FBI Sentinel

Note: The FBI abandoned the VCF project in 2005 after spending \$170 million. The project went live on July 1, 2012 after spending 600 million in 12 years by switching to the Scrum.

3) <http://agile2013.agilealliance.org>

Note: The topic of SCRUM or Agile is beyond the scope of this writing.

Appendix:

Note: Currently, Nutanix only support SRM 5.0, not SRM 5.5. Therefore, any VMware customers will not be able to take many advantages from VMware SRM 5.5 listed below, while Zerto Data Protection with DRaaS should be considered – see the link below for details: <http://www.zerto.com/zerto-virtual-replication-video/> :

- The vSphere Replication provides asynchronous replication with flexible Recovery Point Objectives (RPO) that range from 15min to 24 hours
- The vSphere Replication enables simpler replication that is managed directly from vCenter Server with virtual machine granularity. Users can now replicate individual virtual machines (VM-to-VM) that in turn essentially has eliminated a third party VM-to-VM replication technology advantage in a tier-2 application environment
- Because it operates at the individual virtual machine disk (VMDK) level, it enables replication that is storage agnostic. Customers can save not only on replication software, but also on storage infrastructure by using lower end, heterogeneous arrays across sites, including Direct Attached Storage

What the new three features from SRM 5.5 do?

- SRM simplifies the setup and on-going management of recovery and migration plans. Customers can replace traditional, manual runbooks with centralized recovery plans, which reduces the time required for set up from weeks to minutes
- SRM automates the orchestration of the failover process to the secondary site, as well as the failback to the production environment. Failover and failback automation can eliminate errors with manual processes and eliminates complexity. This level of automation also enables users to test their recovery plans with non-disruptively, thus can predict the Recovery Time Objectives (RTOs) and increase the confidence level in a recovery plan
- SRM 5.5 can increase the support flexibility via different replication solutions and either leverage vSphere Replication or a wide range of storage vendors' array-based replication software

In addition, SRM 5.5 can now:

- Recover from multiple point-in-time snapshots using vSphere Replication. This features enables customers to recover from a previous known “good state” if the last restore point is corrupted

- Supports storage vMotion and storage DRS (Distributed Resource Scheduler™) for workloads moving within a consistency group at the primary site
- Adds interoperability with Virtual SAN for workloads replicated with vSphere Replication

See [VMware vCenter Site Recovery Manager 5.5 Release Notes](#) dated March, 2014 for details.

Recommended Reading

1. [Magic Quadrant for Integrated Systems](#) 16 June 2014 ID:G00252466
2. **Critical Capabilities for General-Purpose, High-End Storage Arrays** – 7 March 2014
ID:G00248908
3. **Critical Capabilities for General-Purpose, Midrange Storage Arrays** - 7 March 2014
ID:G00248904
4. [Dell Fluid Cache for SAN Frequently Asked Questions](#) – 4/1/2014

Acknowledgement

Thanks Andrew B. Mills, Territory Manager; Shaun Nelson, Account Manager for Los Angeles; Laura Jordana, System Engineer and Thomas A. Gustaveson, Sales Development Representative at Nutanix for providing me a webcast and a follow-up Q&A session via email and web conferencing.

Thanks Andrew B. Mills for allowing me to use a few graphics in my notes for clarification purpose.