This is an except from a personal yearly retrospective of 2013. When I joined the organization the infrastructure layer was plagued by instability and performance issues. We completely rebuilt the platform layer on vSphere, going 100% virtual, with massive improvements to application performance verified with extensive infrastructure telemetry, application instrumentation and custom reporting.

”Jupiter” was the ‘As-Is’ codename, because it’s ‘gassy and bloated’.

While “Mercury” was the ’To-Be’ because it’s ‘hot and spins fast’

… the space theme was chosen by the team and served as an inspiration for everyone to come together on the massive organization required.

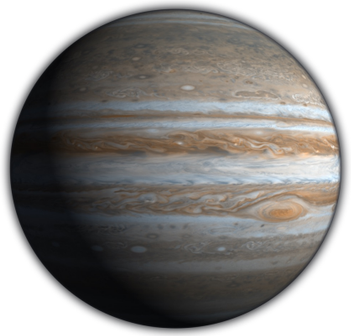
This presentation is part of a larger experiment in embracing public, visual personal brand building, available at:

https://mattschneider-visualcv.github.io/

VisualCV started as a Pathfinder’s project at Dell Technologies while I was mentoring engineers through our career ladder into roles requiring panel & packet review, Principals & Distinguished.

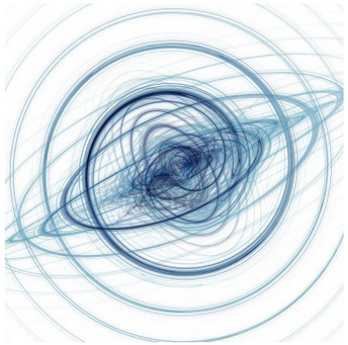
2013 Year-in-Review for Platform

Agenda



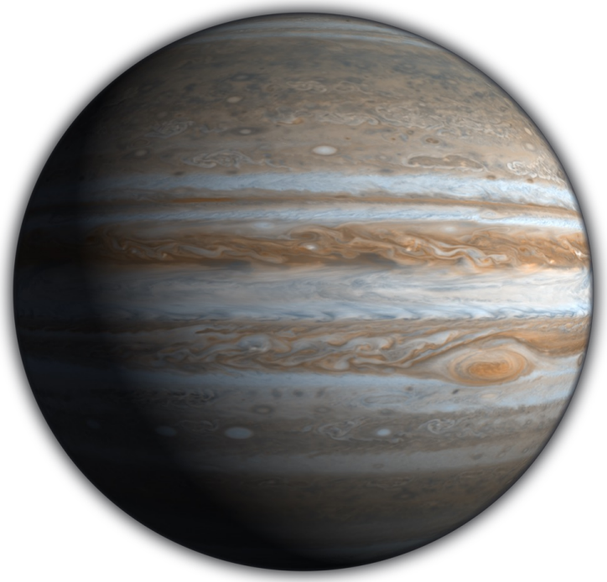
Before – challenges since November 2012

After – accomplishments made since then



Beyond - What’s next in 2014

VMware and Storage – November 2012



"Jupiter" environment

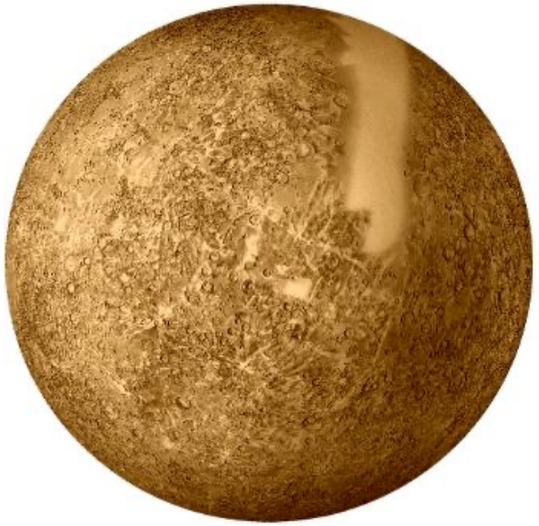
VMware challenges

- Development VMs live on same hosts as production VMs
- A rogue VM in one app can impact performance to other apps
- No resource management to adjust for changing VM workloads
- Development running out of support Lab Manager environment
- Production running on out of support hardware
- No redundancy of management network on any hosts
- No dedicated management or vMotion network; lives on SQL vlan
- 1Gb network architecture introduced risk performing routine tasks
- VMware hosts not patched with no product lifecycle management
- Almost out of cabinet space in 4120 datacenter for additional hosts
- Almost out of physical ports running directly to the core for more VMware hosts
- Inefficient resource utilization on host and VMs leads to less VM density per host
- VMware hosts straddle both customer facing and internal production networks
- vCenter stability issues
- No standardization of settings across each host
- No SRM protection makes manually recovering in a DR event a manual process
- Lax permission delegation granting more access than required

Storage challenges

- Development running on two nearly out of support arrays
- Almost out of fibre channel ports running directly to the core for more VMware hosts
- Dated 4gb fibre channel infrastructure
- 4gb Qlogic cards don’t have removable gbics; need to take server down to replace card
- No dial home on the EMC arrays to proactively monitor and alert on issues
- Improperly architected Compellent storage tiers leads to painful peaks
- Not enough write cache on Compellents to withstand additional workload from peak
- 33 Compellent arrays leads to high management overhead
- Multiple storage arrays across vendors presented to hosts
- Datastores presented across multiple clusters
- Thick VMs eating space, with no additional room to add more shelves on existing arrays
- Storage multi-pathing, queue depth, not properly tuned for VMware
- Close to maxing the bandwidth available for replication between 4120 and Data Bank
- Cloud’s entire infrastructure runs on iSCSI that hasn’t been properly setup
- Cloud has production VMs running in both 4120 and Data Bank
- Cloud is spread out between 4 arrays, with not enough physical space to add more disks

VMware and Storage – November 2013



"Mercury" environment

VMware accomplishments

- Development broken out into its own vCenter for Dev, Test, and vCloud Director
- Resource pools used to isolate apps so no VM can impact another apps' performance
- DRS used to dynamically vMotion VMs to balance CPU and memory workload
- Lab Manager 3 migrated to Lab Manager 4 with vCloud Director coming online currently
- Production only runs on in-support hardware
- Management networks of VMware hosts now fully redundant
- Dedicated management and vMotion network created to better manage network traffic
- Each VMware host now runs with dual 10Gb NICs via top-of-rack, solving port problems
- Product lifecycle in place to properly test patches in Dev before pushing to prod
- More dense VMware hosts and only virtualizing going forward solves the space problem
- VMware clusters laid out based on OS, application workload, and licensing
- Hosts are now fully allocated on RAM, allowing for higher VM to host density
- Better segmentation on VMware hosts prevents VMs straddling both networks
- Mercury built on a fresh install of 5.1 with 5.5 currently being tested in Dev
- VMware host settings standardized and documented
- Every VM built in Mercury is protected via SRM for more automated DR failover
- More granular permissions implemented to better lock-down access to appropriate teams

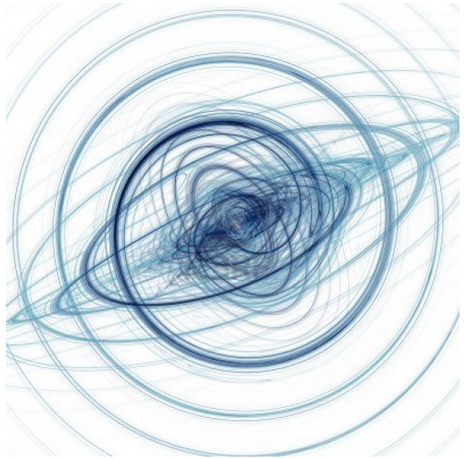
Storage accomplishments

- 5 storage arrays have been decommissioned, all data migrated to VMAX
- New 16Gb switches installed in top-of-rack, future proofing and solving port problems
- Brocade 16Gb fibre channel cards installed in new systems going forward
- VMAX arrays setup with ESRS to properly phone home to EMC Support in case of alerts
- VMAX array has more available IO per tier in addition to real-time tiering of data
- VMAX has more than 30x more cache than our Compellent arrays combined
- One interface needed to manage the 3 VMAX arrays in production, dev, and DR
- Only one array presenting storage to Mercury, simplifying management
- Converted VMs from thick to thin, freeing up multiple TBs of unallocated space
- Datastores are broken up by application and cluster
- Storage is properly tuned on VMware, with proper queue depth and PowerPath installed
- Replication bandwidth from 4120 to Data Bank upgraded from 1gb to 10gb
- Installed Isilon as future solution for NFS, backups, Cloud DropBox and future projects
- Cloud's shared environment converted from iSCSI to fibre channel
- Cloud is currently migrating their production Data Bank back over to 4120
- Cloud's storage has been consolidated to two arrays adding 100+TB of capacity

Stats – Before and After

	Prod		Dev/Test	
	Nov. 2012	Nov. 2013	Nov. 2012	Nov. 2013
Hosts	153	184 **(+20%)**	36	58 **(+61%)**
VMs	1596	2382 **(+49%)**	512	2041 **(+298%)**
Physical Cores	2262	3000 **(+33%)**	320	644 **(+100%)**
Physical Memory (TB)	24.5	48.3 **(+97%)**	3.8	9.1 **(+140%)**
Virtual Cores Assigned	3136	5922 **(+89%)**	830	2203 **(+165%)**
Virtual Memory Assigned (TB)	14.4	30 **(+110%)**	2.3	7.5 **(+220%)**
Active Memory Usage (TB)	--	2.2	--	.529
Total CPU (GHz)	--	3059	--	1487
Peak CPU Usage (GHz)	--	1102	--	407
Total Storage Capacity (TB)	728.23	823**(+13%)**	125.48	235.1 **(+53%)**
Total Subscribed Space (TB)	406.5	1041 **(+156%)**	111.13	323**(+191%)**
Total Allocated Space (TB)	--	554.26	--	173.1
Total Fibre Channel Ports	672	1200 **(+78%)**	80	256 **(+220%)**
Activated Fibre Channel Ports	598	868 **(+45%)**	70	128 **(+83%)**

2014 Major Initiatives



Backup System Refresh

The current backup system is plagued with issues. We have regular audit issues due to inability to see backups end-to-end. Insufficient on-site retention to handle frequent restore requests. Multiple teams involvement for restores, plus multi-step process means long restore delays. No backups of operating system means no short-term recovery of failed machines. All backups are only by requests, and not by profile, meaning lots of systems are not being backed up that likely should be. In many cases, backup are not even possible on existing system (26 hours a day backup times).

A POC is being started in 2013Q4 to test upgrading EMC Networker to the latest version, deploying the agent to all machines, backing up the virtual machines directly for non-database, and leveraging Data Domain storage for client side de-duplication for faster backups and reduced space consumption. The goal is automated backups through one tool, 30 day on-site retention at both Prod and DR, with tape retention at both sites.

Deploying this solution will require all teams involvement. The agent installation on database servers, configuring direct backups, training all teams on how to perform ad-hoc backups and restores, plus code changes in some cases.

Estimated cost – 1MM per quarter, financed deal.

Non-Production DR

Currently, there is no protection in our non-production environment from any major failure. There is no backups, nor DR plan. Should we need to execute DR for production systems, we would quickly need to perform code changes.

Completely replicating with like-for-like in non-production would not be a cost effective solution, nor would it solve the need for backup and retention for non disaster situations.

The approach for this situation from an infrastructure design, will be to leverage Networker and DataDomain to backup all virtual machines directly, replicating this to Databank, where restores of Dev/Test environment will be possible as we trickle down servers from DR production clusters into DR testing.

Additionally if we acquire adjacent space to our cage at DR, we could spread the dev/test workload down to DR through the vCloud Director deployment, allowing targeting both locations for non-production environments.

Cost for backup system included in previous slide.

Databank expansion in another slide.

Databank Expansion

The space at DR is insufficient to handle internally facing or dev/test systems. There is adjacent space available that will provide adequate space for these additional DR systems, plus completely separating the cages for Co-Lo customers, reducing issues with security.

Q1 Budgeted Forecast: 15K

- Badge Readers
- New cage walls

Annual cost estimate: 30K

- 2500 square feet expansion

Mercury Migration - continued

In 2013, a new vSphere platform was created; integrating Vmware best practices, 10Gb networking, multi-tenant controls, expanded memory, SRM configuration, etc. From March to September, over 1,100 virtual machines were moved or re-built in Mercury in production.

About 1,200 virtual machines remain in Jupiter, including a mix of Customer Facing, Internally Facing and Non-Production. In 2014 this migration must continue and finish. With the other investments in this deck, 2014Q1 should mark the completion of all customer facing machines. Q2/Q3/Q4 will follow the same plan for internally facing as 2013. Provide seed capacity to begin migrations, moving/rebuilding hosts along the way.

Large involvement from all teams. Project Managed.

<u>Q2/Q3/Q4 Budget Forecast: 500K per quarter</u>

- 2x cabinet upgrades (4x FC switches, 4x ETH switches)
- 30x server RAM upgrades
- 30x server HBA/NIC upgrades
- Cabling & Power

Storage Fabric Upgrade

The current storage fabric is 4Gb functional level due to the core switches all being 4Gb. Additionally these switches were out of support in 2013, though extended for free through the VMAX deal to 2014Q1. The plan is to replace the current core switches (Cisco 9505) with Brocade DCX. These will be 16Gb, coupled with the Brocade 6510 TOR switches, we'll upgrade the functional level to 8Gb end-to-end with 16Gb TOR/Core backbone.

This infrastructure upgrade will increase the speed of our storage, as well improve the visibility of the fabric through more advanced monitoring.

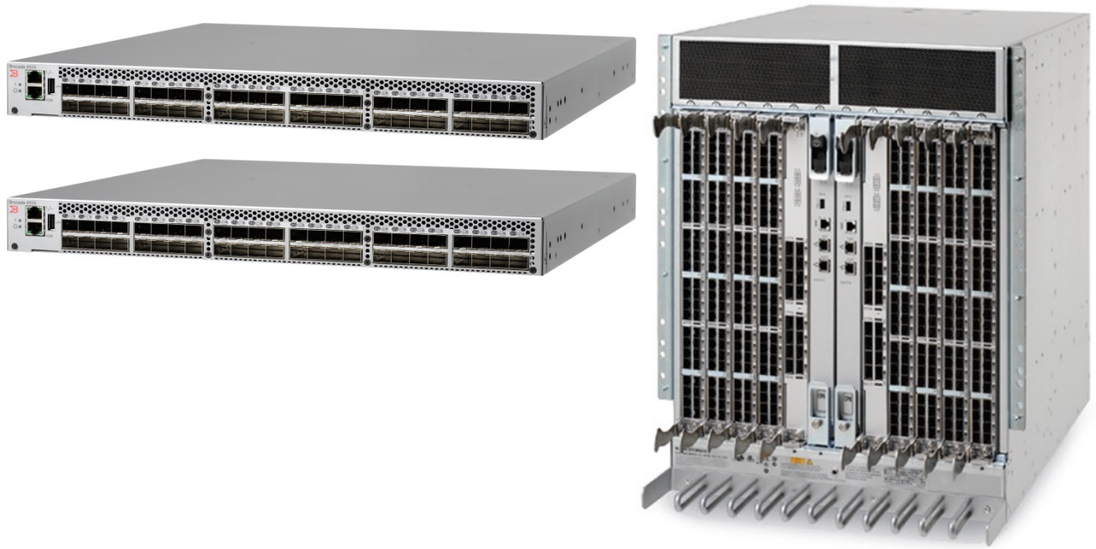
Minimal interaction from other teams needed.

<u>Q1 Budget Forecast: 1MM</u>

- 4x DCX 8510 switches (~125K each)
- 14x 6510 TOR switches (~15K each)
- Cabling

<u>Q2 budget Forecast: 500K</u>

- 250x host upgrades to 8Gb HBAs (2K per host)



Dev/Test Infrastructure Expansion

In 2013, a lifecycle program for servers was implemented. Moving release -2 generation servers out of production into Dev/Test. This has dramatically expanded the hosts available for our development and QA use.

There are currently over 20 hosts ready for implementation in Dev/Test, with more sitting in 4120 ready to be moved. When once the constraint was storage, then servers, now it is network connectivity (both ethernet and fibre channel).

in 2014 we need to plan for ongoing growth in 4000, until we can reach trickle down of networking.

Q1/Q2/Q3/Q4 Budget Forecast: 200K (per quarter)

- 2x 48port 10Gb ETH (80K)
- 2x 48port 8Gb FC (80K)
- Cabling/Power

vSphere 5.5

VMWare released in 2013Q3, we will run in test 2013Q4, with a planned implementation of 2014Q1. vSphere 5.5 provides numerous performance enhancements specifically targeted at latency sensitive applications, advanced DRS guidance controls , application aware high availability and server side flash caching.

Upgrades will roll through our normal cadence. vCenter and Management upgrades and early January, then upgrading individual hosts with a slow roll-out to monitor.

The upgrades themselves will be transparent with no impact and minimal team involvement, however in Q2 VMTools and Hardware Versions will need to be upgraded on all guests to take advantage of new features.

Based on preliminary testing, the vFlash caching feature it highly promising, we'll look towards a Q2 implementation. vFlash is server side SSD read cache, offloading this activity from the array and speeding up read latency while supporting DRS.

Q2 Budget Forecast: 375K

- Upgrade 250 hosts with 200GB SAS SSD @ $1,500 each

vSphere 6.0 Alpha Test

We have been selected as one of the few companies to alpha test vSphere 6.0. This version of vSphere will be a major release, with vCloud Director functionality being moved into the core vSphere product.

Testing will all be performed off-site on VMWare hosted environment. The desire is to include a cross section of RP engineers to work on this test and provide feedback to VMWare.

vCloud Director Rollout

With the effort beginning 2013Q4 (after the ELA uplift), all non-production systems will be moving to vCloud Director. This provides a self-service portal, as well as advanced virtualization around networking and security, allowing easy replication of dev/test environment plus segmentation between products and SDLC phases.

vCloud Automation Center Roll-out

Part of the vCloud suite purchased in 2013, is the vCloud Automation Center. This provides a blue-print based, policy driven deployment engine. The core advantages to this product included simplified deployment, with high visibility to deployments by providing approval workflows. While also ensuring deployed machines are following standards set in the blue-print.

The goal will be to stand-up and configure vCAC in Q1, with test deployments leveraging the tool in Q2, and production deployments in Q3.

vCloud Operations Manager

vCOPS is a powerful, analytics based monitoring tool for the vSphere platform. It includes anomaly detection, forecasting and powerful visulizations to show the health of systems from an infrastructure perspective.

A major enhancement is expected early 2014, which should allow sizing guidance based on configurable peak values. The system will need to be updated.

As well a major goal for 2014 is to provide training across the organization on using the tool to provide deeper insight into our systems.

System Center 2012 Roll-out

In 2013, System Center Operations Manager and Configuration Manager 2012 were installed, and deployed for limited use. The 2014 goals is complete roll-out and leverage.

Q1

- Catapult SCOM Engagement – 90K
  - Custom Management pack conversion
  - Nagios/SCOM integration testing (single pane for NOC, automated incidents to ServiceNow)
- SCOM – ServiceNow Integration Setup

Q2

- SCOM 2007 Decomission
- Non-Prod SCOM/Nagios integration

Q3

- Prod SCOM/Nagios integration
- SCCM Server rollout for software inventory/usage
  - Possible consulting engagement – 50K

(application) Web Dedicated Cluster

In 2013 Linux and Windows systems were split into separate clusters. Inside the Windows cluster, logical segmentation was implemented. Monitoring has shown application needs physical segmentation for future growth. As well the single-url project needs a 2x temporary capacity and additional capacity for the POD initiative is needed.

In 2014 a dedicated vSphere cluster for (application) Windows will be created. 30 Dell r720 release 2 servers will be needed to reach the 1:1 virtual/physical ratio of 150 OneSite web servers (4 cores and 32GB ram each). The existing hosts will trickle down to the main windows cluster for organic growth.

This project will build the platform, VMs will be built by Windows Engineering.

<u>Q1 Budget Forecast: 1.4MM</u>

- 60x Dell R720 R2, 30x per site (16K per server)
- 8x TOR FC switches, 4x per site (15K per switch)
- 8x TOR ETH switches, 4x per site (15K per switch)
- 12TB of Storage, 6TB per site (10K per TB)
- Cabling/Power

Xtreme SQL Cluster Upgrade

In 2013, selected MSSQL servers were promoted into the ‘Xtreme Cluster’. This cluster is the highest-end hardware with EMC Xtreme flash cache, provisioning is all manual with attention to tuning alignments like NUMA. Currently the hardware in this cluster does not have a Intel QPI cross bridge, this lowers the potential performance of application needs on 4 sockets. Today, there is no Xtreme cluster in Databank either.

In 2014Q2, Intel/Dell will release the Ivy Bridge EX chip, re-introducing the cross bridge. This chip will support 12 cores @ 3Ghz, which is significantly expand the processing power we can provide to guest workloads.

The plan is to build a 5 nodes (4+HS) cluster in both 4120 and DR, moving the existing guests and providing space for more. The existing hosts will trickle down to SQL cluster for organic growth. 2 additional nodes for testing are needed.

<u>Q2 Budget Forecast: 900K</u>

- 12x R920 hosts, 5x per site, 2x test (50K per sever)
- 12x Xtreme cards, 5x per site, 2x test (25K per card)

(application) DB Cluster Upgrade

In 2013Q4 the goal is to split the (application) DB servers out of the Linux cluster into a dedicated cluster. This will be done with the current servers, though additional VMAX Directors will be dedicated to this cluster (pending VMAX expansion purchases). This includes the database for (application) , (application) , (application) , (application) , & (application) .

In 2014, this cluster will be upgraded to Ivy Bridge, boosting database performance of all apps. The existing hosts will be moved into the Linux cluster for organic growth. Sizing will be based on the 279 vCPUs currently in use, plus PW server plus 30% growth. Expansion will be done over Q2/Q3/Q4 to spread out spend.

Q2/Q3/Q4 Budget Forecast: 350K/192K/192K

- 36x Dell r720, 18x per site (16K per server)
  - 12x per quarter, 6x per site/quarter)
- 4x FC TOR switches, 2x per site (15K per switch)
- 4x ETH TOR switches, 2x per site (15K per switch)
- Cabling

VMAX Expansion

Using OpenScale, the 2013Q4 plan is to buy 50TB of capacity on the general layout for CDS, and build a two engine configuration on the FDA layout; freeing ~50TB in the general layout for expansion. This should provide adequate capacity for 2014Q1.

As the year progresses, expand both the general layout and FDA layout for growth, leveraging OpenScale to only pay for allocated capacity.

Q1 Budget Forecast: 2MM

- Buy-out 2 engines, 32x 200GB EFD, 128x 900GB SAS

Q2 Budget Forecast: 1MM

- Buy-out 50TB general layout per site
- Stage additional 2 engines, 32x 200GB EDF, 128x 900GB SAS under OpenScale

Q3 Budget Forecast: 2MM

- Buy-out OpenScale 2 engines, 32x 200GB EDF, 128x 900GB SAS

Q4 Budget Forecast: 1MM

- Buy-out 50TB general layout per site (organic growth)

Cloud Backup Storage Lifecycle Replacement

The Cloud division currently uses lower end Dell NAS appliances. To handle their on-site backup capacity they stripe multiple appliance together created a huge risk of data loss with large scalability problems. Currently this space is 180TB, and has no Like-For-Like configuration, meaning if the NAS is lost, 30 days worth of backups are permanently lost, plus recent restores would not be possible during a DR event.

In 2014Q1, replace the entire storage solution for Cloud backups with Isilon. This cuts the per TB cost roughly in half, provides a DR solution and a much more scalable model. Expand the space in 2014Q3 for growth.

<u>Q1 Budget Forecast: 300K</u>

- 4x 100TB nodes, 2x per site, 200TB usable per site (55K per node)
- Add backup accelerator for disk-to-tape, long term retention (50K per site)

<u>Q3 Budget Forecast: 110K</u>

- 2x 100TB nodes, 1x per site, 100TB usable per site (55K per node)

NAS Migrations

In 2013 RP invested in the Isilon scale-out NAS solution. This provides a highly scalable, highly available, file-based storage solution, with snap-shot retention and replication at 1/10th the cost of Tier3 VMAX (1/100th the cost of tier1). This platform should be leveraged to replace all Product based file servers (CIFS and NAS) currently on virtual machines.

The Platform team will expand the array, plus work with Linux and Windows team to provision CIFS or NFS shares and ensure proper snapshot, replication and backup policies. The Windows & Linux team will need to configure security, migrate data and adjust application paths. The current Isilon has enough capacity to begin this effort, with Q2 & Q4 expansions budgeted. Migrations off VMAX storage will free up block storage for other efforts.

Q2 Budget Forecast: 200K

- 2x 100TB nodes, 1x per side, 100TB usable per side (55K per node)
- 2x Performance accelerators, price TBD

Q4 Budget Forecast: 110K

- 2x 100TB nodes, 1x per site, 100TB usable per site (55K per node)

Compellent Collapse

Because of our enterprise agreement with Dell, we have unlimited licensing on space, we pay for the controllers and disks. Because of this, the maintenance is only tied to the controller itself, meaning we can move disks off an expiring controller to one under maintenance, there-by extending the maintenance of the disks.

This method is how we have doubled space for Cloud this year, collapsing expiring arrays into the Cloud arrays.

In 2014 this effort will continue, with the end goal being reconfigured arrays for all internally facing systems with completely automated DR.

All teams should prepare for similar efforts as 2013, leverage storage vmotion to move virtual machines between Compellent arrays.