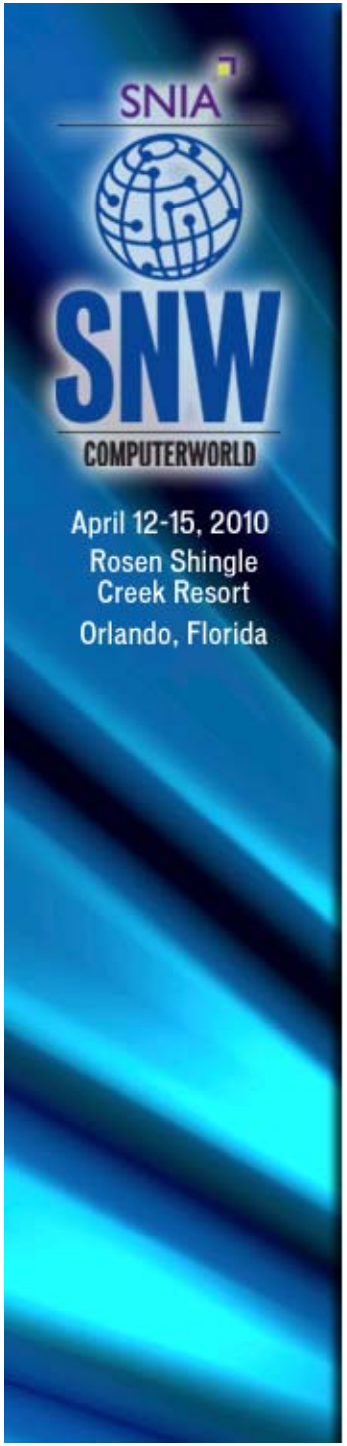
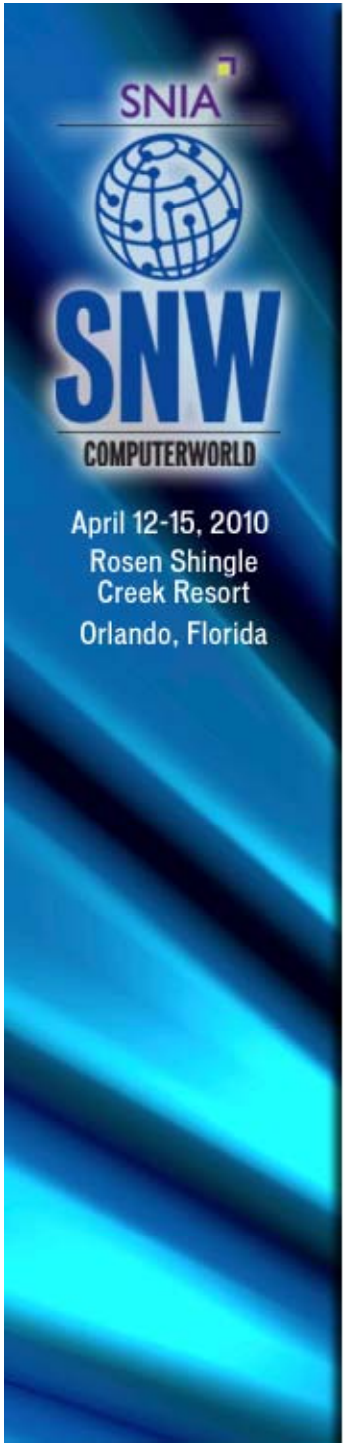


Getting Your Company on the Road to Storage and Server Efficiency

Grant Leathers

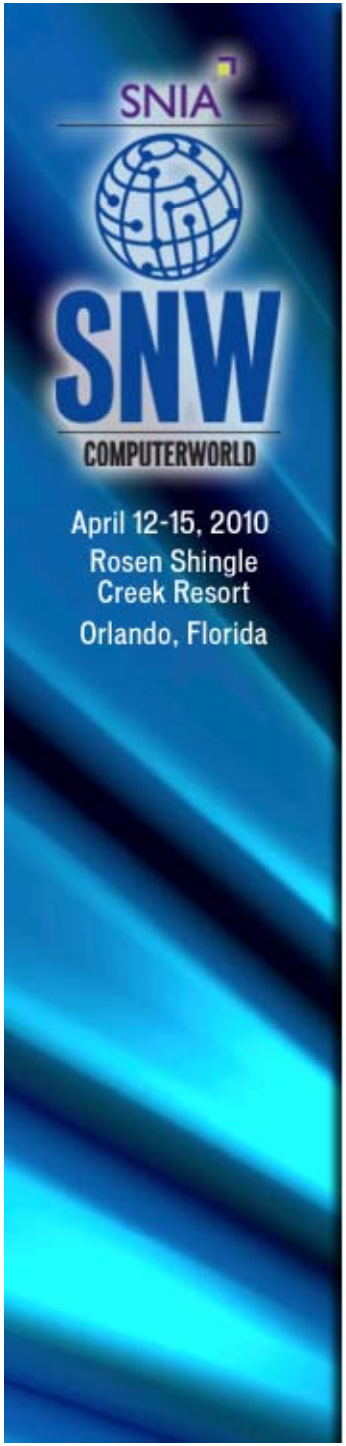
Kelley Blue Book



Company Overview

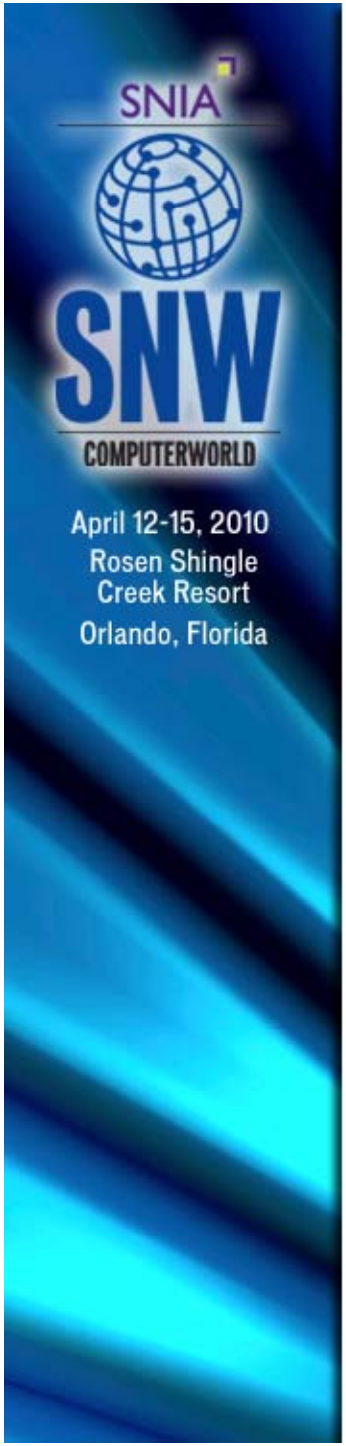
- Kelley Blue Book is "**The Trusted Resource**" for automotive values and Information
- Operate the #1 automotive website in the U.S. www.kbb.com *
- Provide SaaS products to help dealers manage and price inventory. www.cdmdata.com, www.karpower.com
- Headquarters in Irvine, CA with offices in Detroit, Scottsdale, and Beijing
- 500-600 Employees; Privately Held

* Based upon my research. Individual results may vary



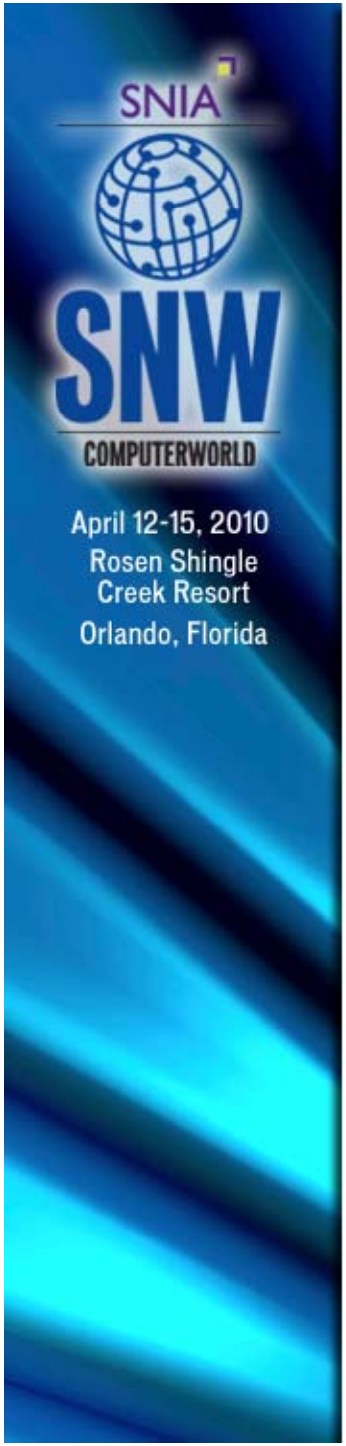
Key Metrics – Kbb.com

- 12 - 15 million monthly visitors to www.kbb.com
- 225 - 275 million page views per month in 2009
- More than 10 days of 10 million + page views in 2009
- Over 1 million images hosted online and delivered each month
- Two active data centers; Irvine, CA and Las Vegas, NV



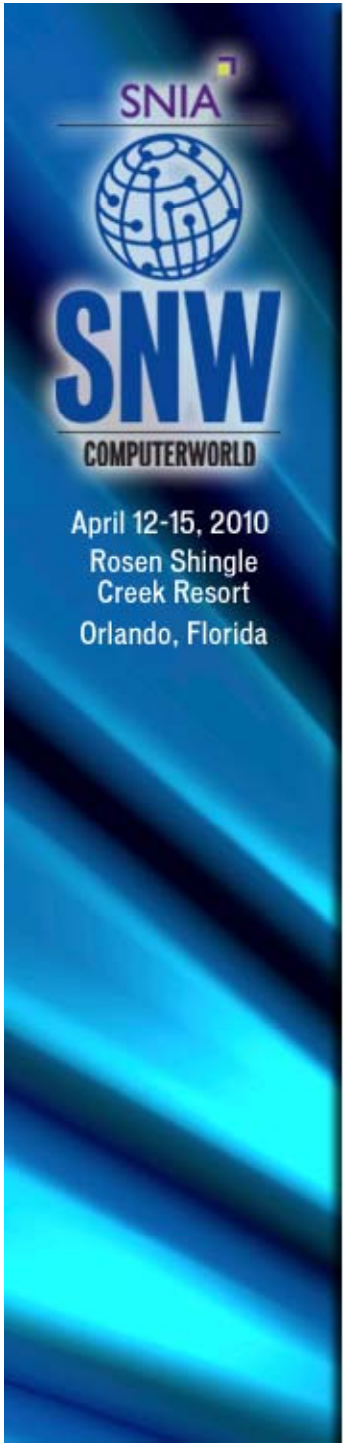
Key Metrics – SaaS Products

- 10,000 dealer customers and 30,000 users of www.cdmdata.com and www.karpower.com
- 750,000 active vehicles in inventory
- 2 million vehicles processed in 2009; 4 million vehicles in inventory
- 10 million active vehicle images hosted; over 30 million images processed in 2009



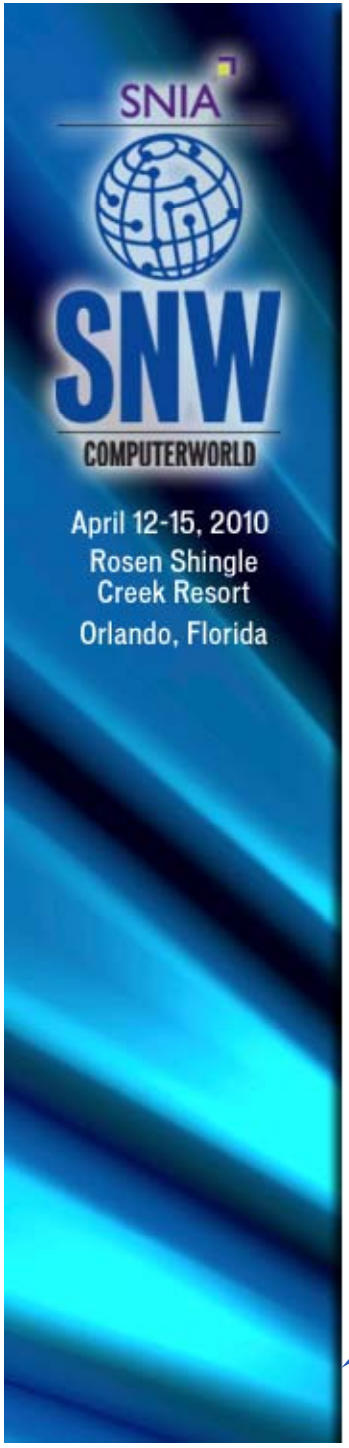
Technology Environment

- Infrastructure
  - 550 Servers – 60% Virtual
  - Storage Protocols: CiFS, NFS, iSCSI
  - 100% VSphere4; 100% NFS in VM environments
  - Upgrading to 10GB core network
  - 4 NetApp Filers; 175 TB Network Storage
  - 6 Server Engineers; 1 Network Engineer; 2 DBAs
- Key Vendors:
  - NetApp, Vmware, Dell, Cisco, Microsoft, F5
  - Savvis, Switch, Keynote, Limelight, UltraDNS, Omniture
- Key Partners: Trace3, EnPointe



Technology Environment

- Bandwidth:
  - 200 MB (compressed) to the Internet
  - 300-400 MB from servers to load balancers
  - 3-4 GB to production storage system
  - 500-700 MB on virtual servers during peak traffic
- Data:
  - 75 + production database servers
  - 4TB (compressed) database backups; 20TB uncompressed
  - 9TB (compressed) weekly backups total



Systems and Data Growth

100% Annual Server Growth from 2004-2009

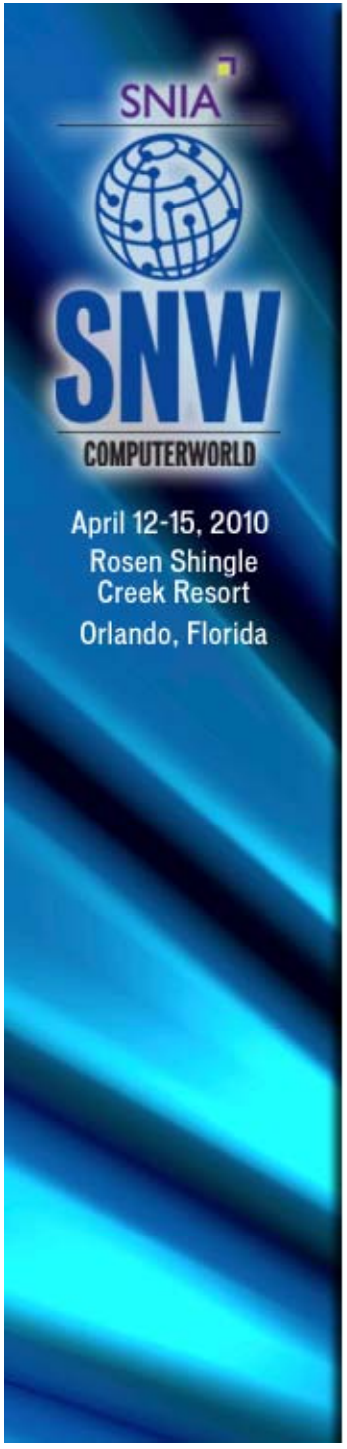
50% Annual Data Growth from 2004-2009. New data sources to drive valuation process

Offshore development partnerships 2008

Migration of dealer products from CD-ROM to SaaS in 2007

Acquisition of CDM Data in 2004

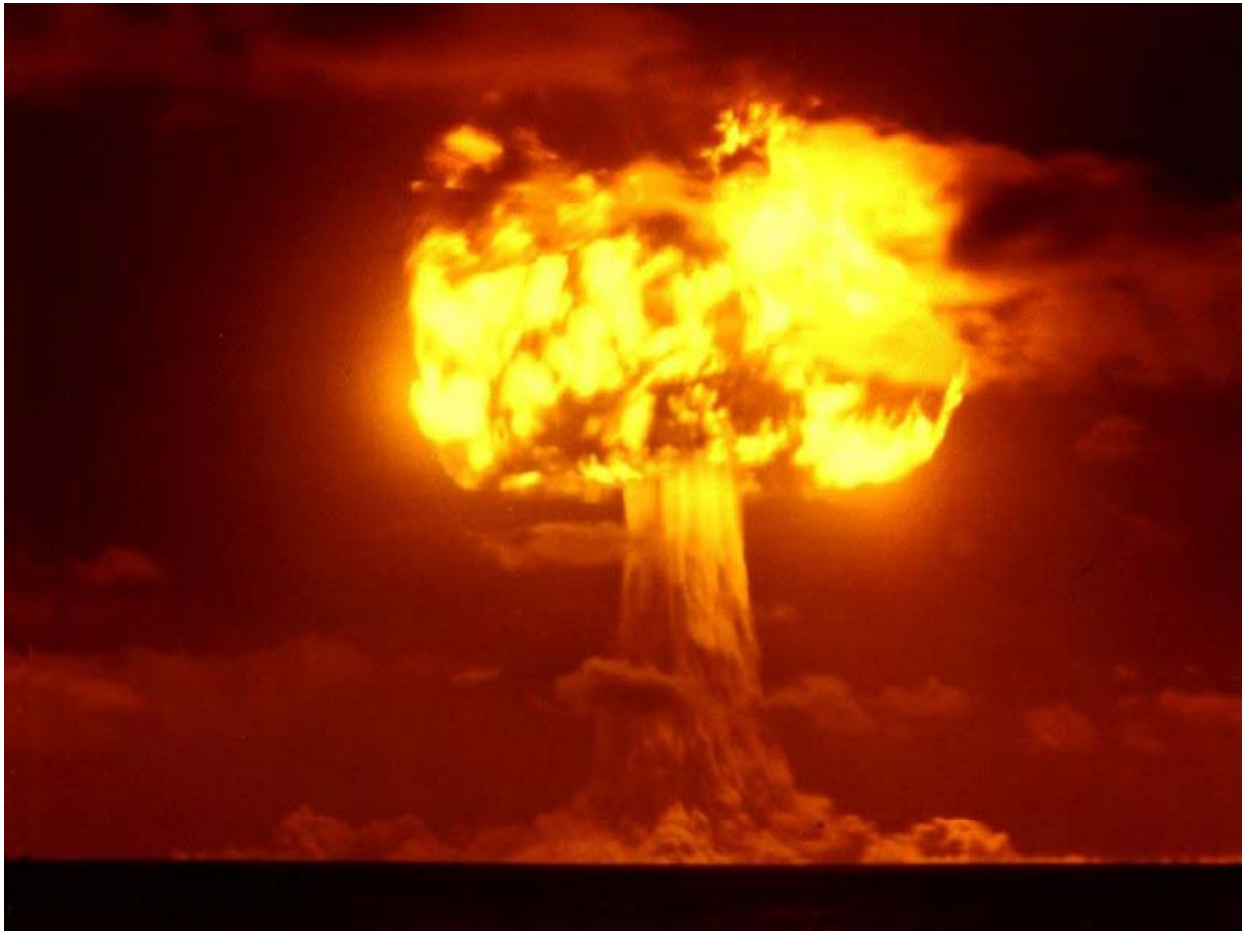
4x Employee growth from 2003-2008



Q: How was this impacting my Capex Budget?

SNIA
SNW
COMPUTERWORLD
April 12-15, 2010
Rosen Shingle Creek Resort
Orlando, Florida

A: It Was Blowing Up



SNIA

SNW

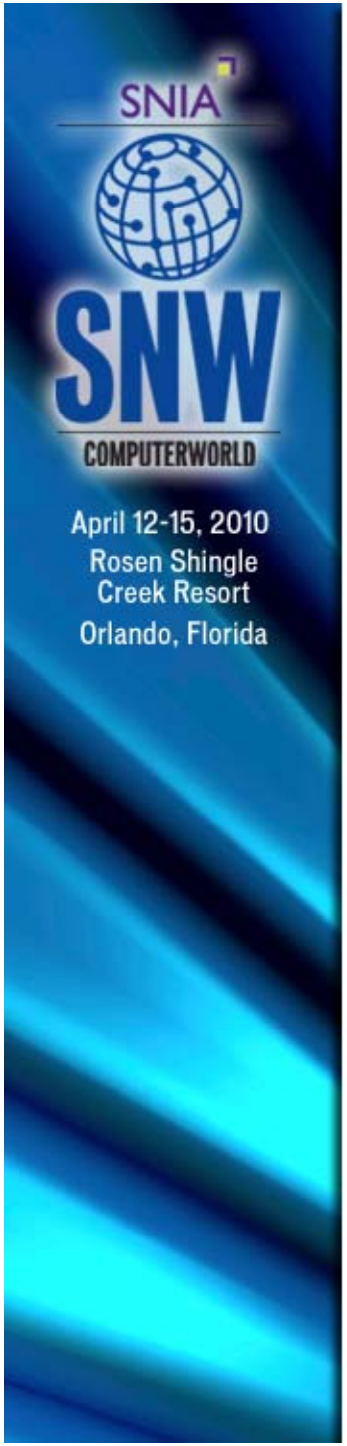
COMPUTERWORLD

April 12-15, 2010

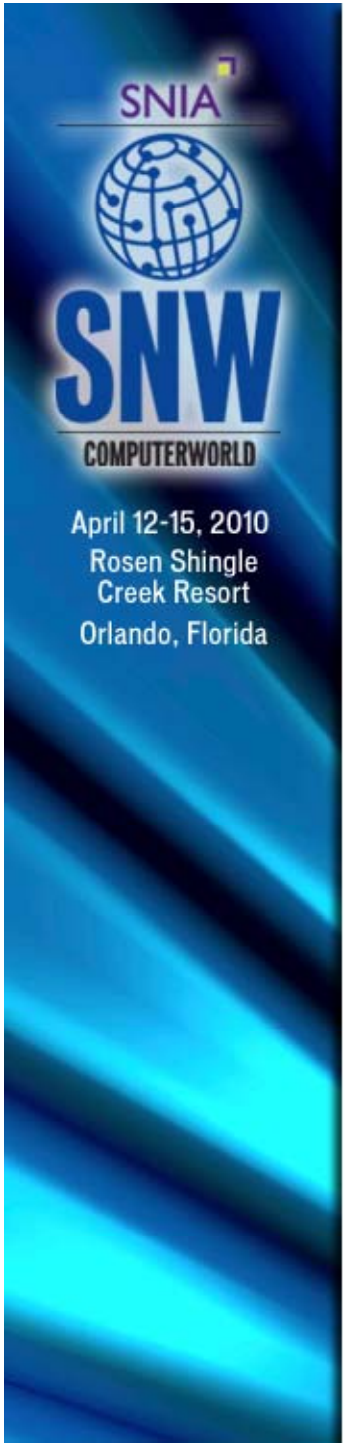
Rosen Shingle Creek Resort

Orlando, Florida

Q: How do we better (more cost effectively) accommodate growth?

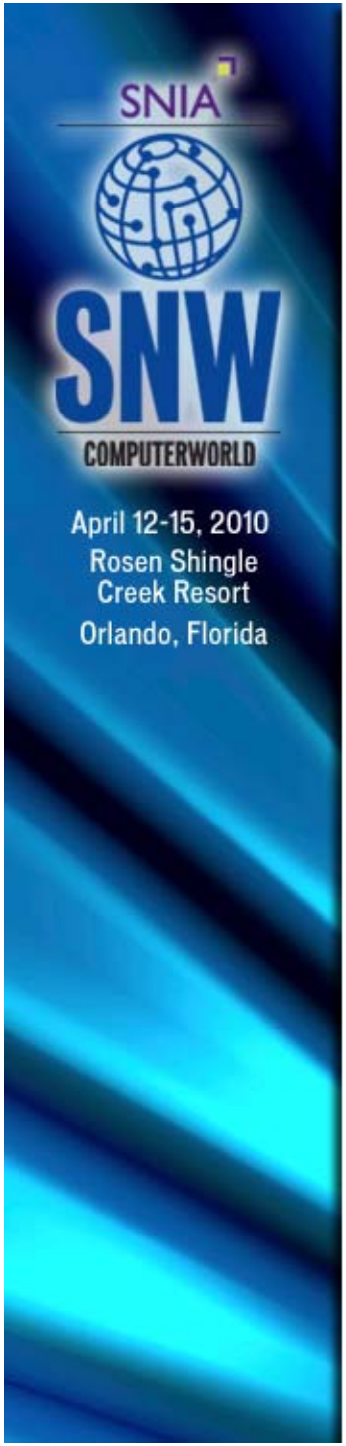


A: Virtualize everything possible



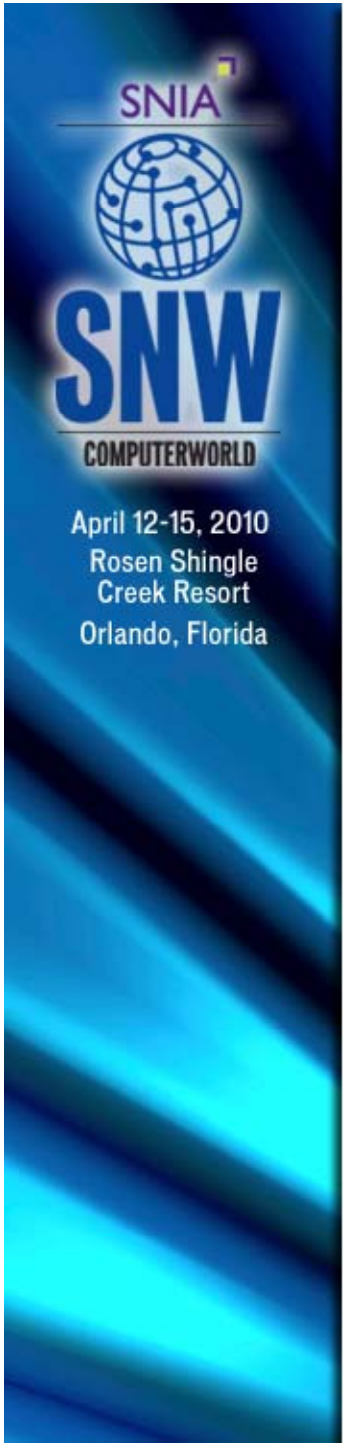
Virtualization

- In 2005 KBB had 5% of servers virtualized
- Virtualization was a 4-Letter-Word due to performance issues
- No high availability or shared storage option
- Implemented industry standard virtualization software with redundancy, high availability, etc.
- Implemented storage platform to accommodate growth of shared data.



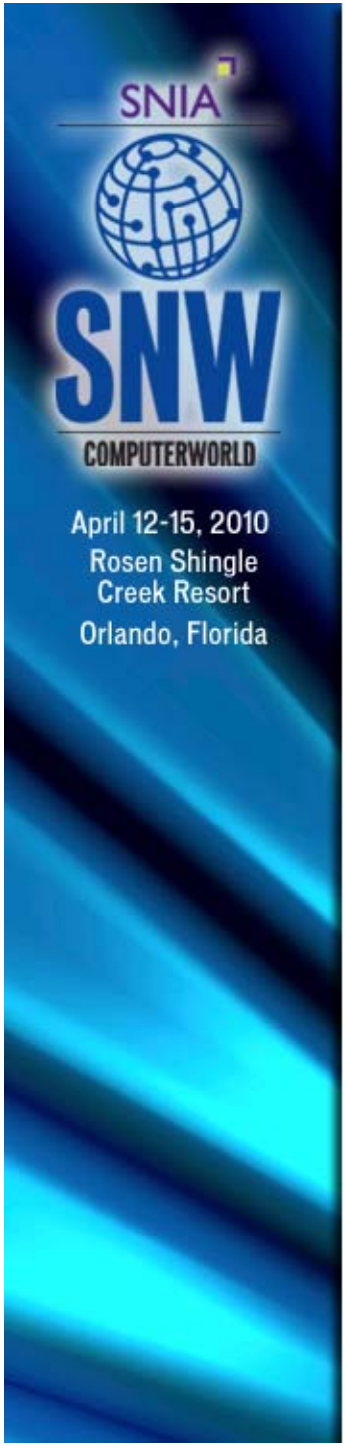
Virtualization

- Started with 4GB Fibre Channel connectivity
- Found this to be too expensive on a per port basis
- Had Questions about the technology:
  - What happens after 8GB?
  - Why do I need two different data networks?
  - What are the trade-offs in price/performance?



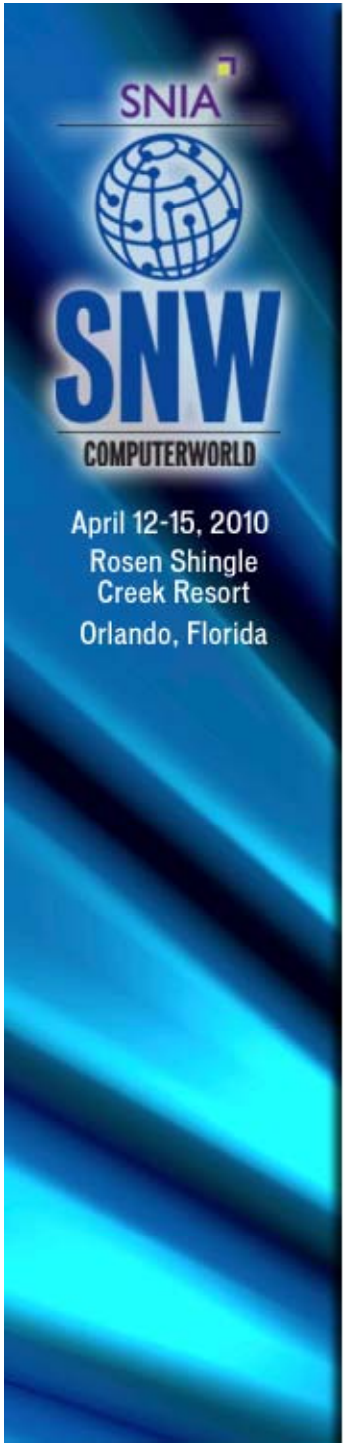
Virtualization

- Moved to iSCSI connectivity within 1 year
- Much less expensive per port
- No separate data network
- No significant degradation in performance from FC
- Much more confident in the future of the technology:
  - Industry standard
  - Clear plan for scale -10GB and beyond

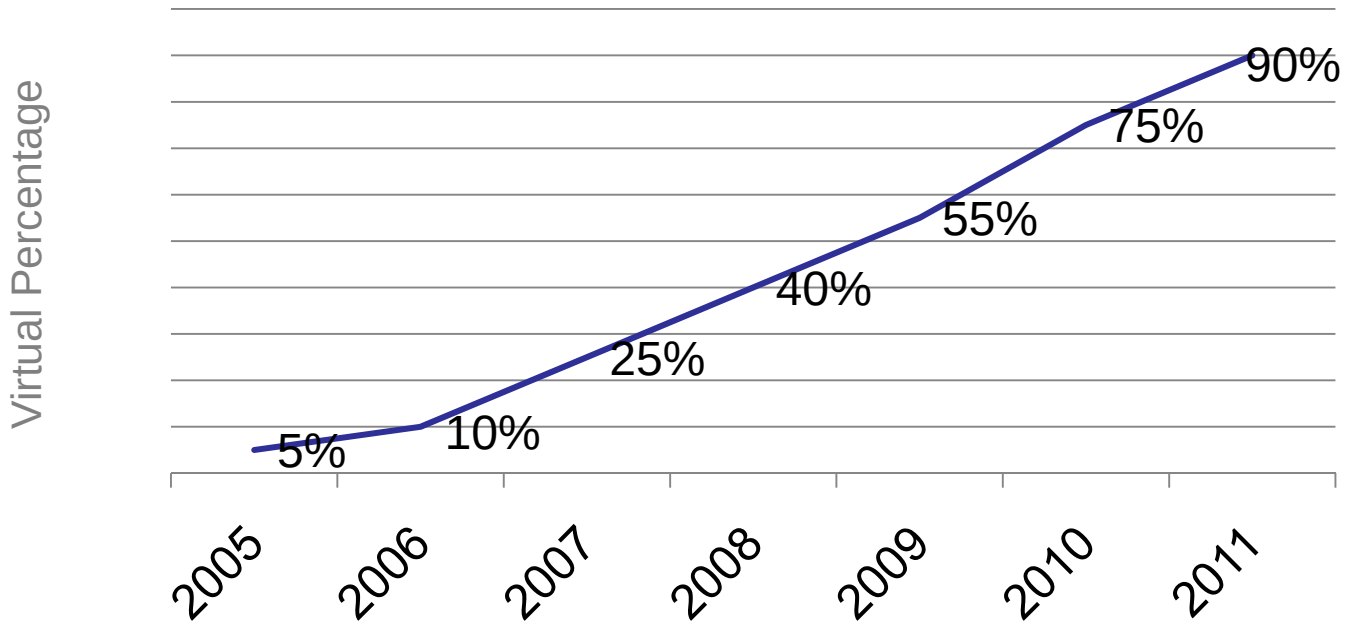


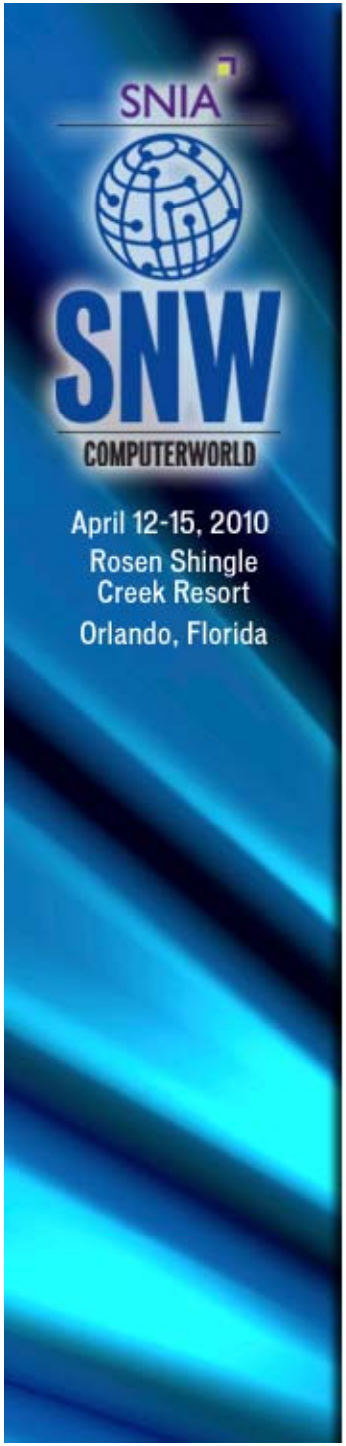
Virtualization

- Moved to NFS within 1 year
- Native protocol with our Storage Platform
- Ability to create larger volumes than iSCSI or FC without performance penalties
- No noticeable degradation in performance from iSCSI
- Same data network and benefits from using iSCSI

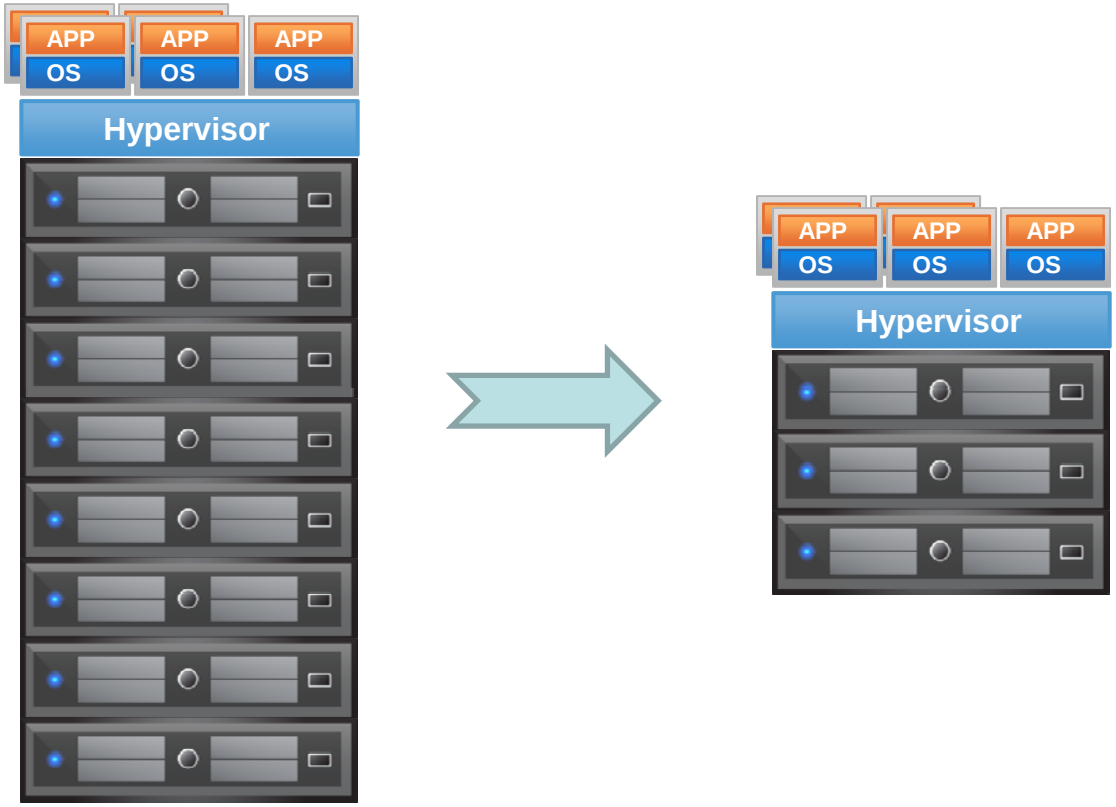


Virtualization Growth



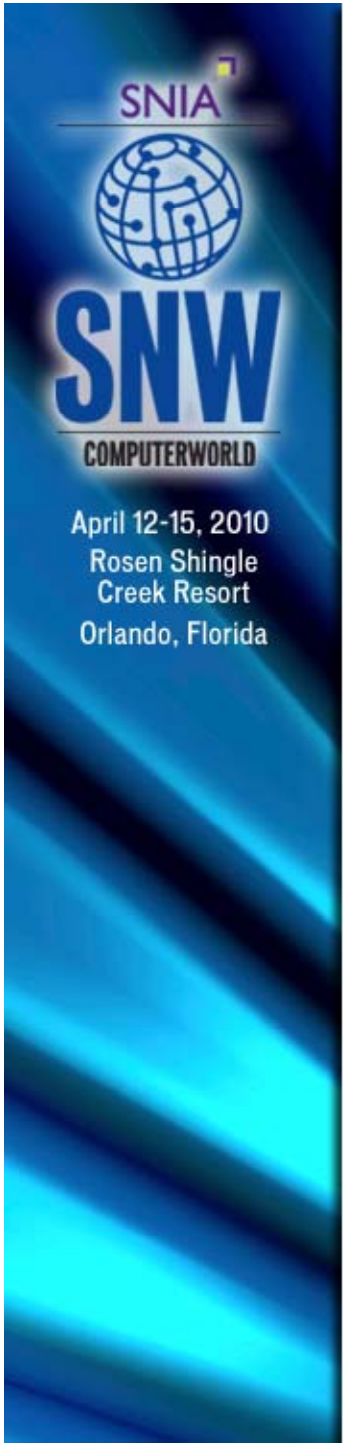


Moore's Law



2006

2009



Analyzing our hosting costs, we calculated a savings of **$1.6 *million* dollars** from our virtualization initiative since 2007

SNIA

SNW

COMPUTERWORLD

April 12-15, 2010

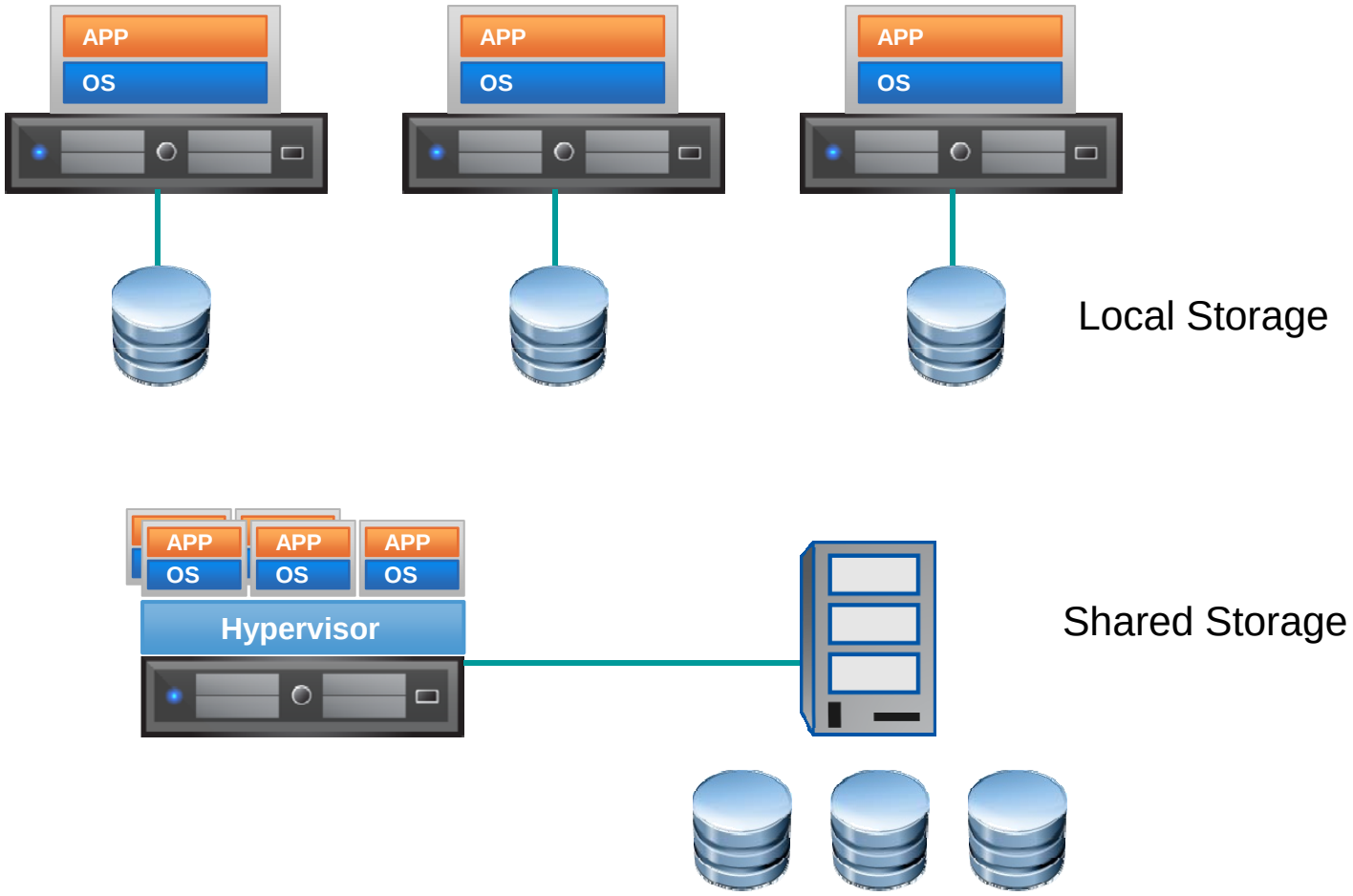
Rosen Shingle Creek Resort

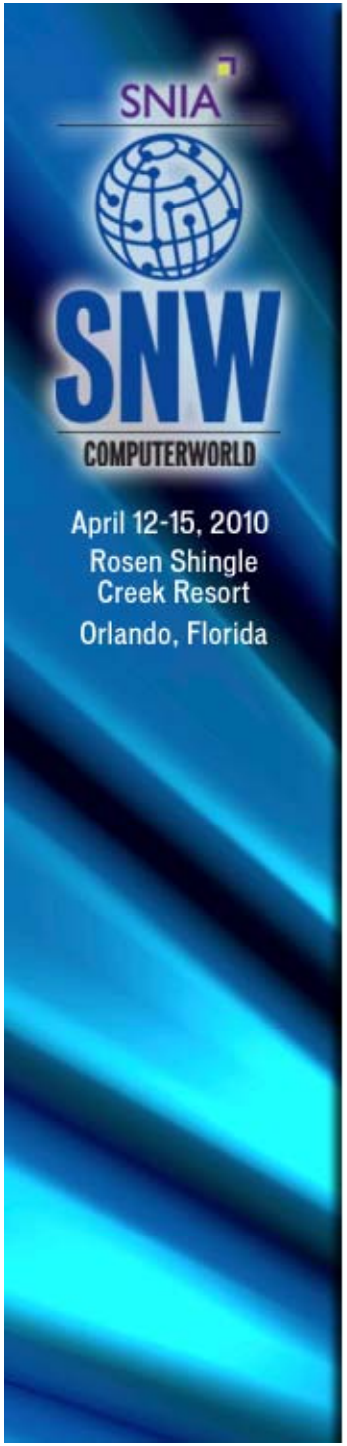
Orlando, Florida

Q: Where is all of this data going?

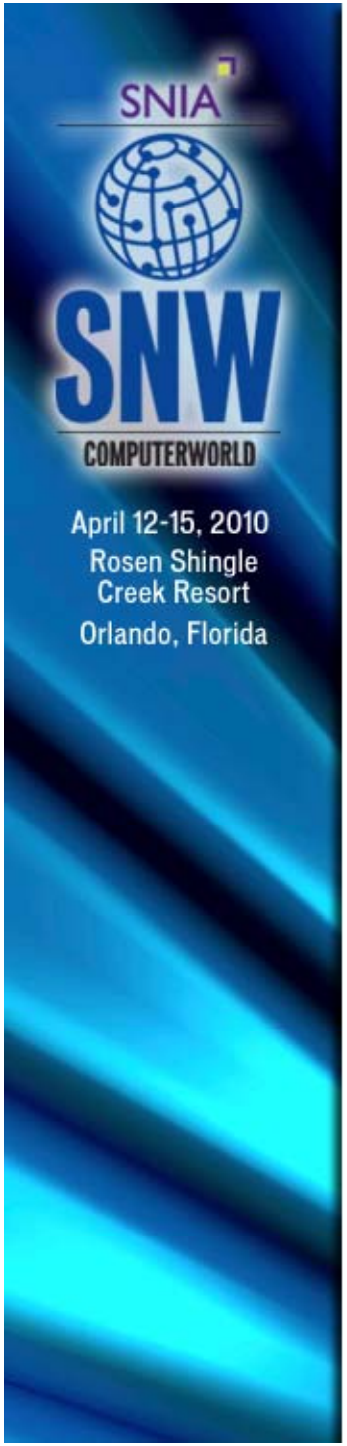
SNIA
SNW
COMPUTERWORLD
April 12-15, 2010
Rosen Shingle Creek Resort
Orlando, Florida

A: It was moving to our Shared Storage Environment

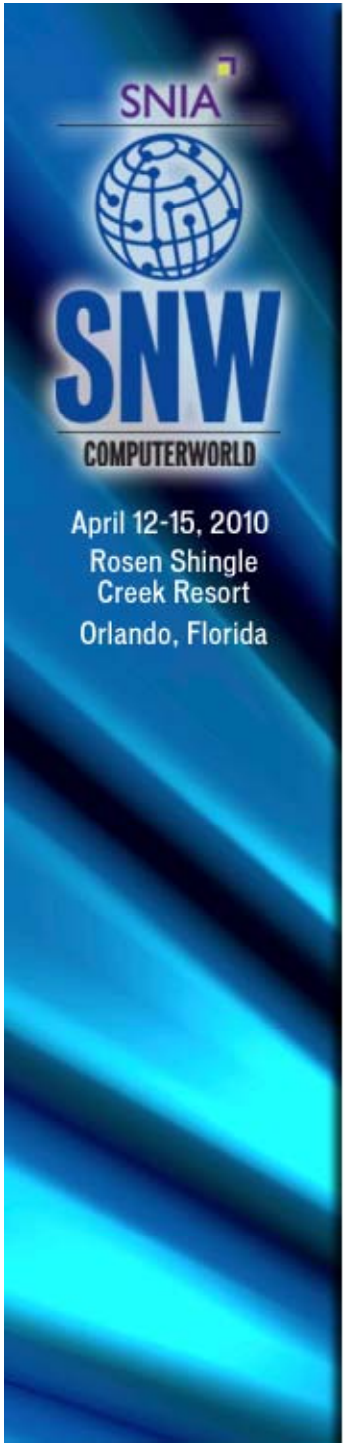




This is was a problem

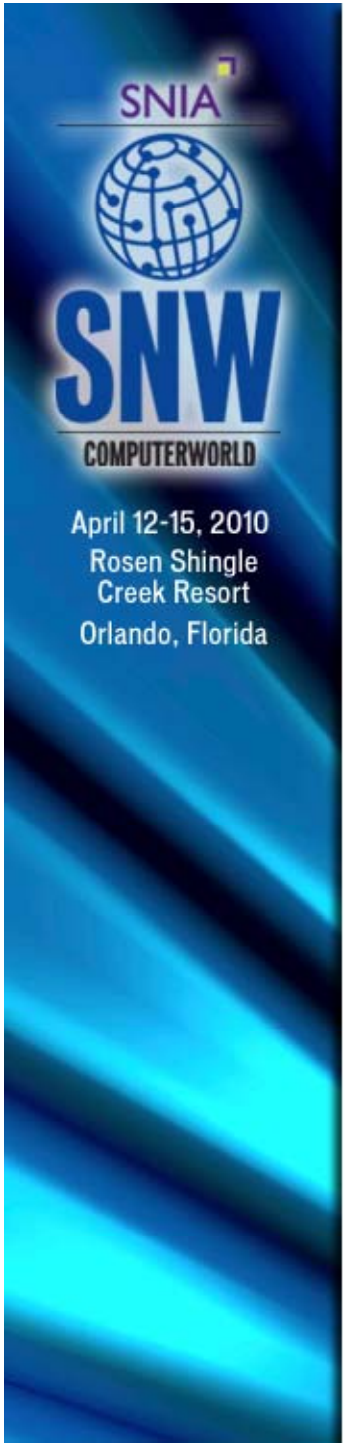


Q: What do you call a problem that was created by a solution to another problem?

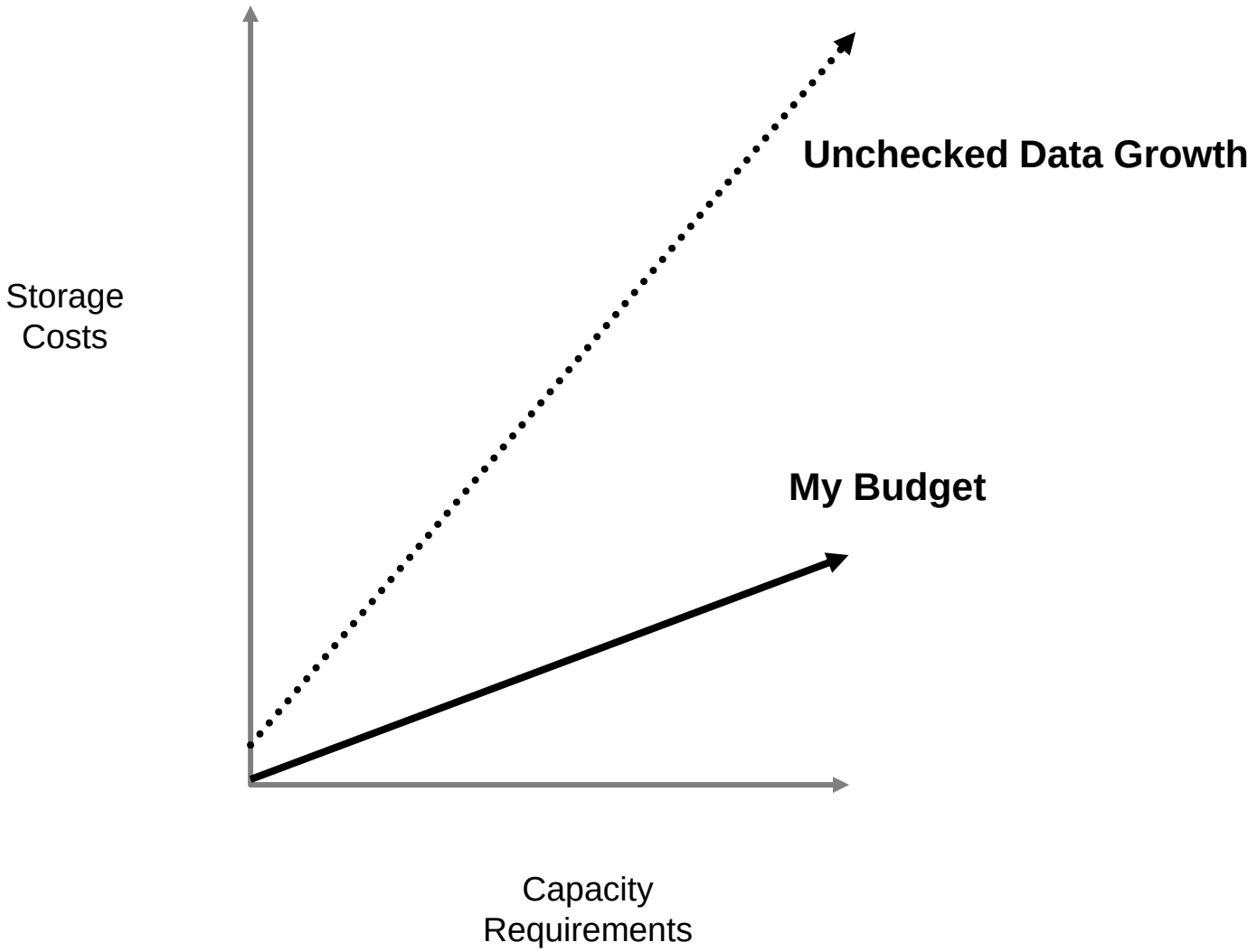


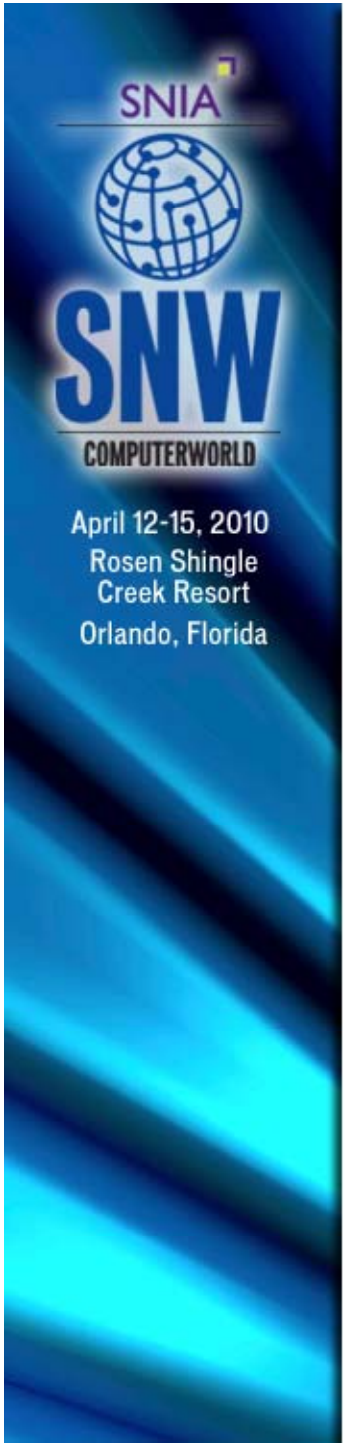
A: “Prolution” *

* According to my 5-year old son



Data Prolution



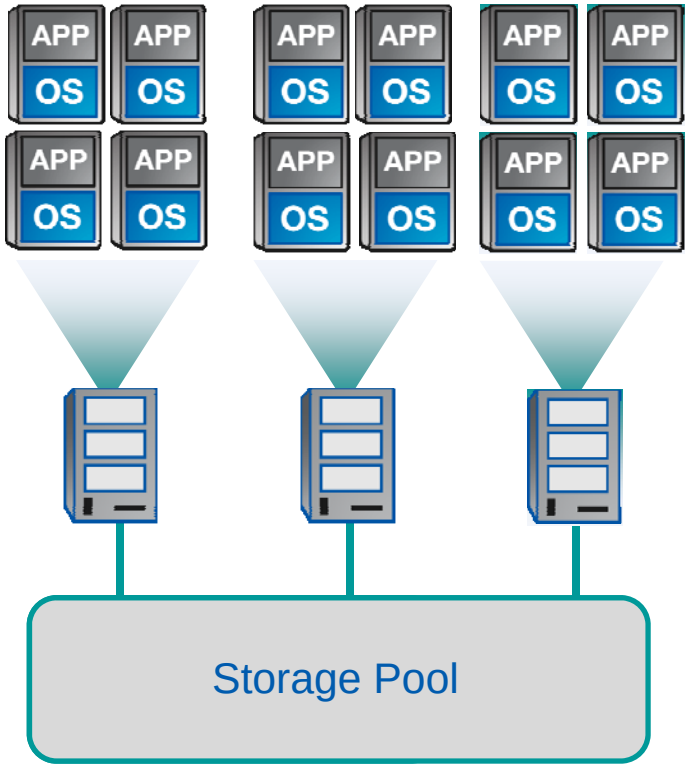


Data Prolution

- Moving data from local storage to more expensive shared storage at a rapid rate
- Unable to mirror data to other storage systems and data centers due to the amount of data being transferred.
- Although savings from virtualization more than made up for increased storage expenditures, there had to be a better way

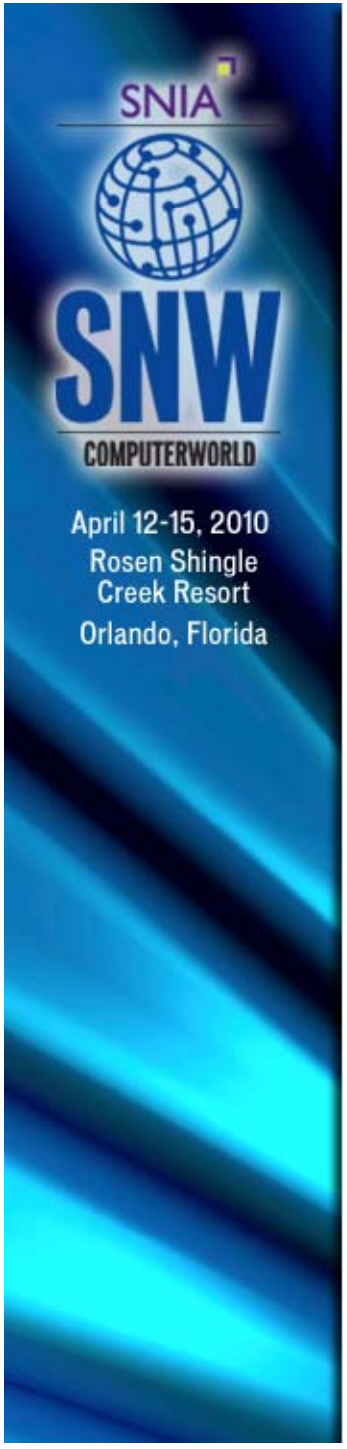
SNIA
SNW
COMPUTERWORLD
April 12-15, 2010
Rosen Shingle Creek Resort
Orlando, Florida

Data Prolution



Multiple copies of the same OS and applications are now hosted on shared storage platform

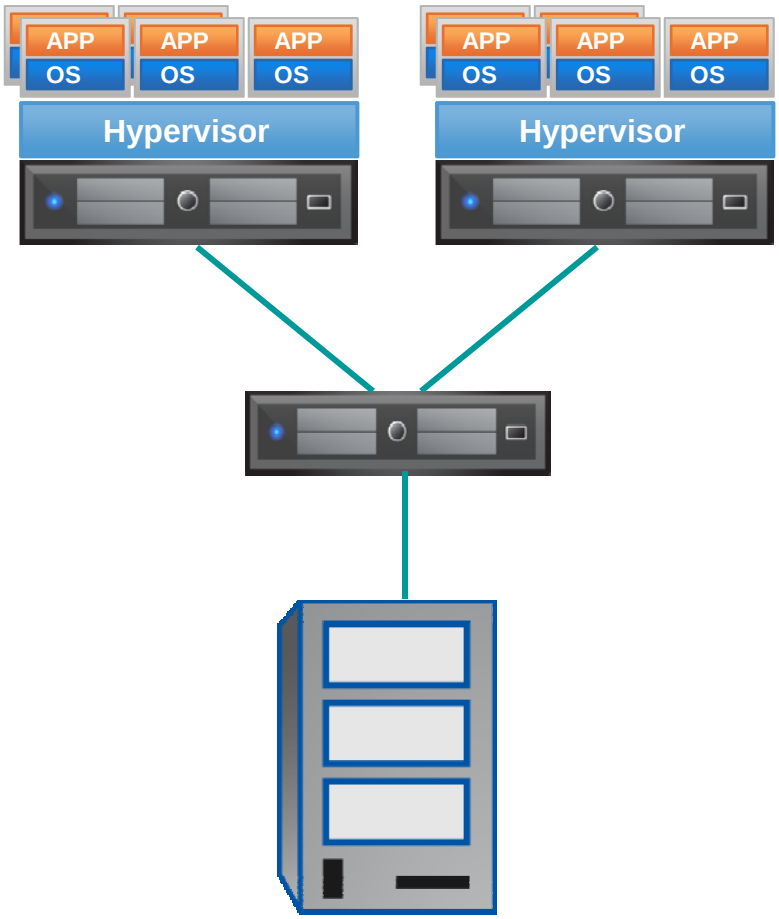
Utilization on shared storage system goes up – really fast.



We explored a few options to solve this problem

SNIA
SNW
COMPUTERWORLD
April 12-15, 2010
Rosen Shingle Creek Resort
Orlando, Florida

In-line Compression



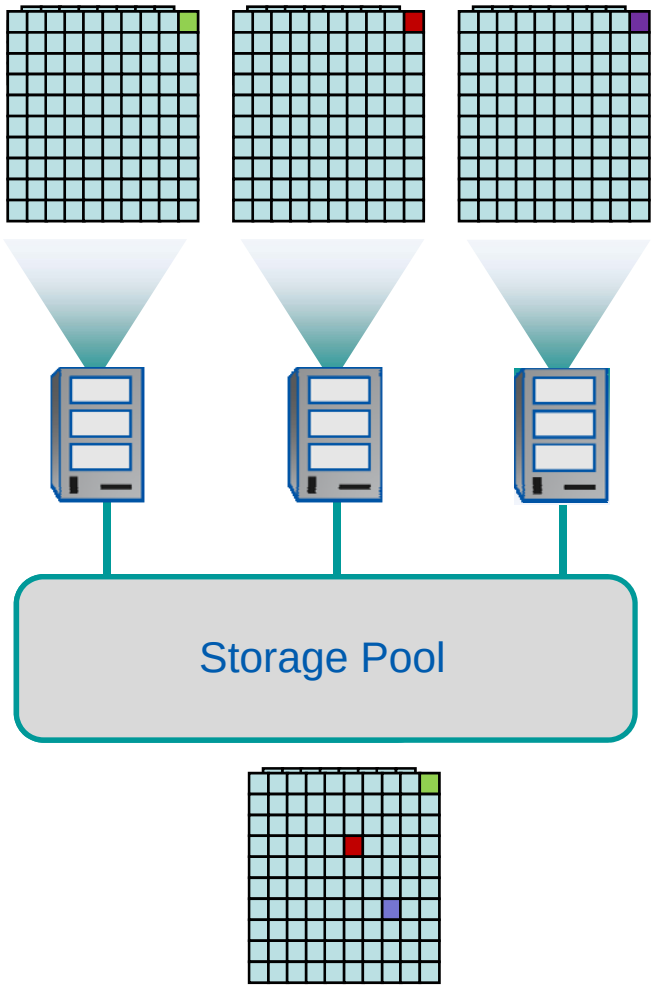
Introduce third party appliance to compress data coming from virtual environment

Questions about performance and redundancy

Questions about vendor compatibility and support

SNIA
SNW
COMPUTERWORLD
April 12-15, 2010
Rosen Shingle Creek Resort
Orlando, Florida

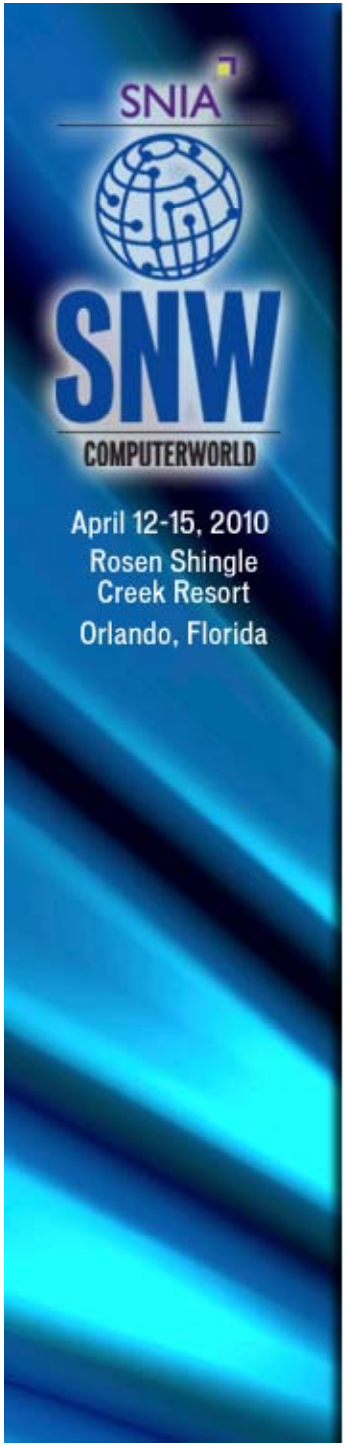
Data Deduplication



Turned on data deduplication on our storage system.

Process runs nightly – not inline

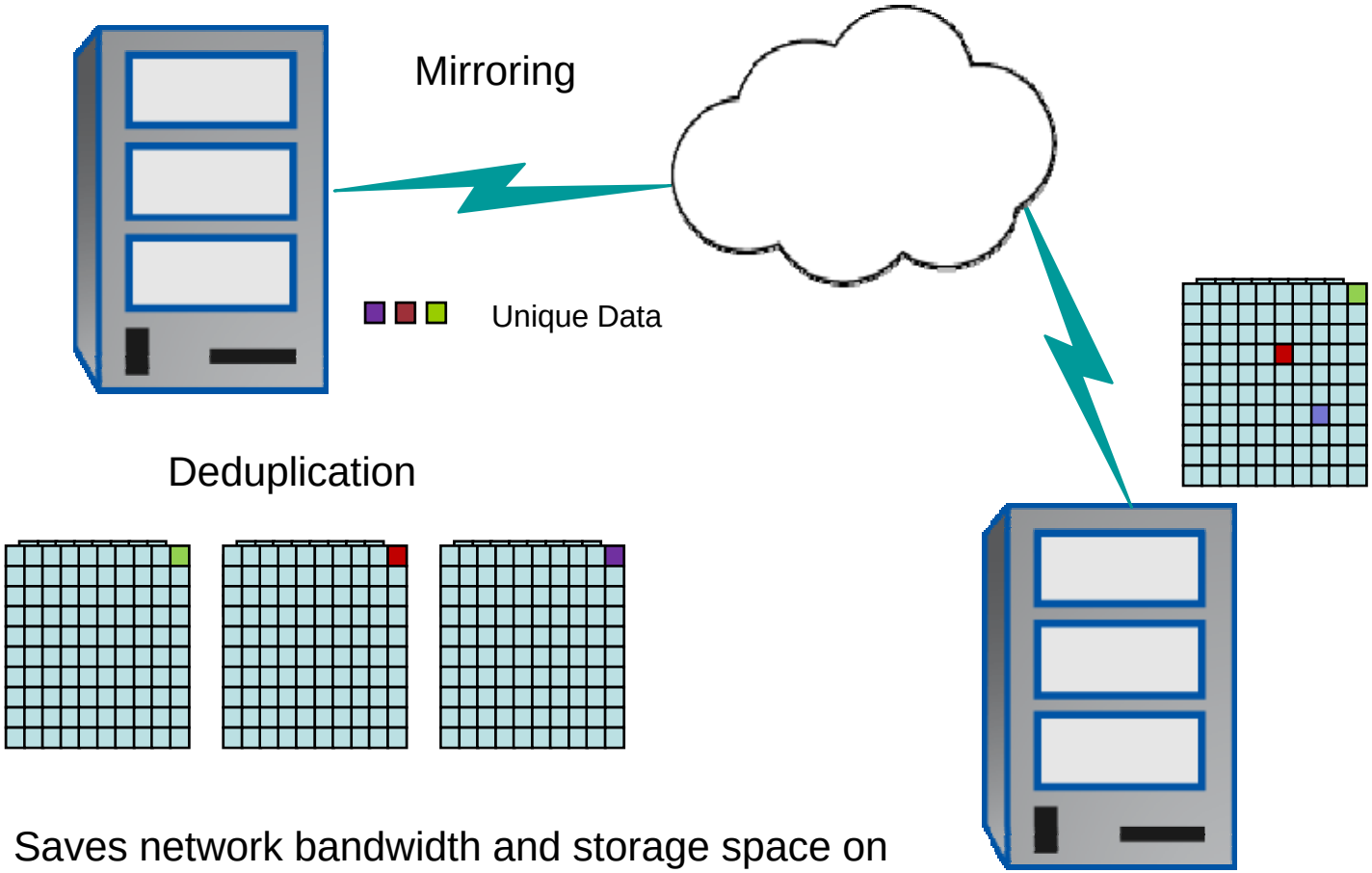
Significant reduction in storage space



Reduced space in our virtual environments by over 70%

SNIA
SNW
COMPUTERWORLD
April 12-15, 2010
Rosen Shingle Creek Resort
Orlando, Florida

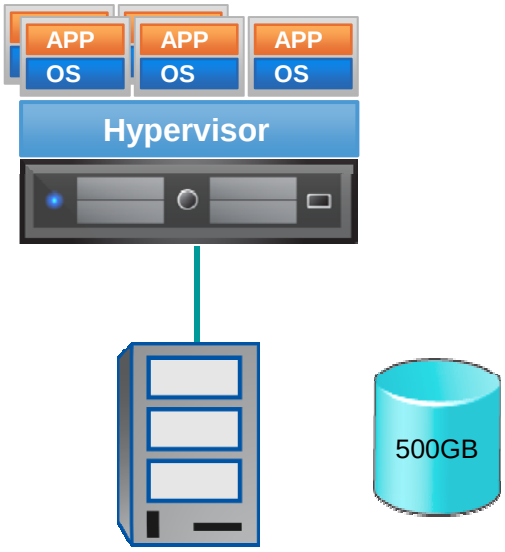
Data Mirroring

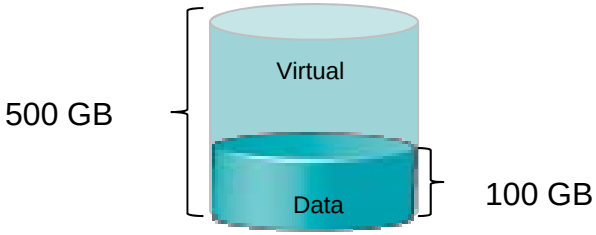


Saves network bandwidth and storage space on both units

SNIA
SNW
COMPUTERWORLD
April 12-15, 2010
Rosen Shingle Creek Resort
Orlando, Florida

Thin Provisioning



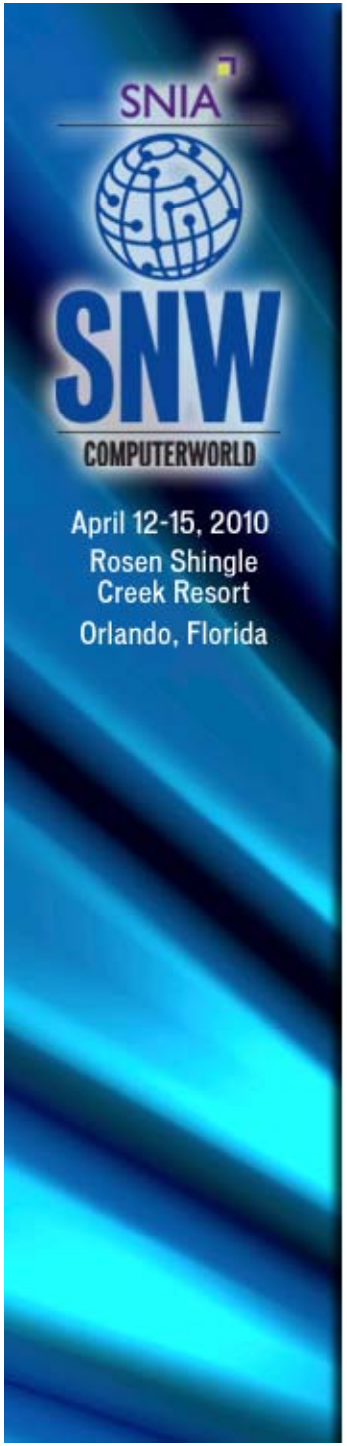


Create a virtual volume to match size of data provisioning request

Automatically allocate a smaller amount of actual physical disk space

Additional physical space allocated automatically as needed

Reduces wasteful over provisioning of storage



Thin Provisioning

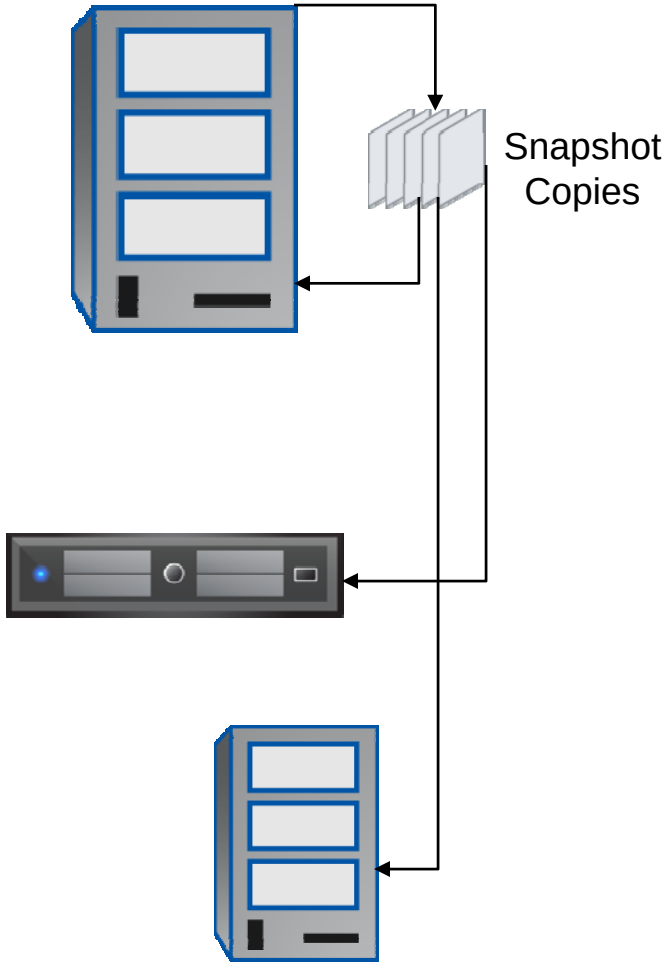
Storage over subscription at the volume level should be set to Automatically grow

Must monitor closely to make sure you don't run out of space on your physical disks

Capacity planning and trending of storage utilization is essential

SNIA
SNW
COMPUTERWORLD
April 12-15, 2010
Rosen Shingle Creek Resort
Orlando, Florida

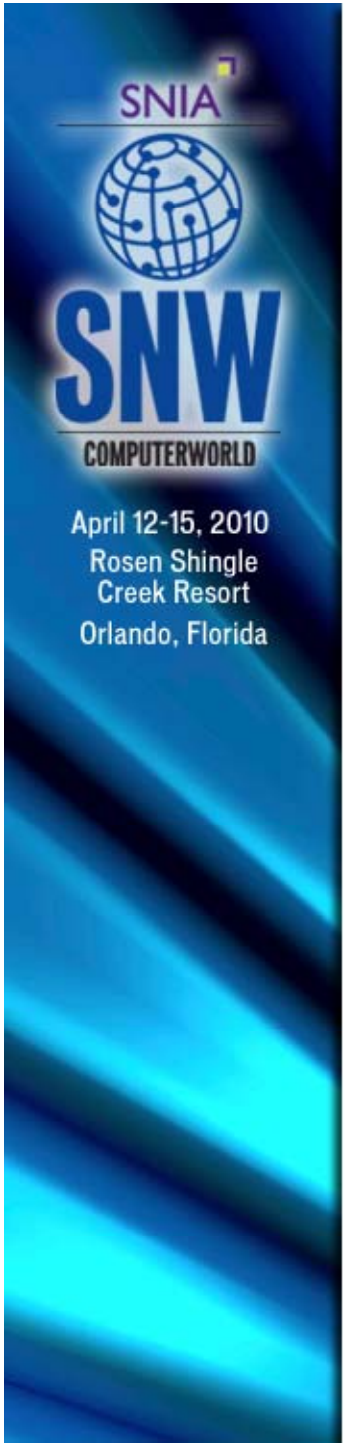
Backup Efficiency



Run snapshot copies on critical data volumes hourly

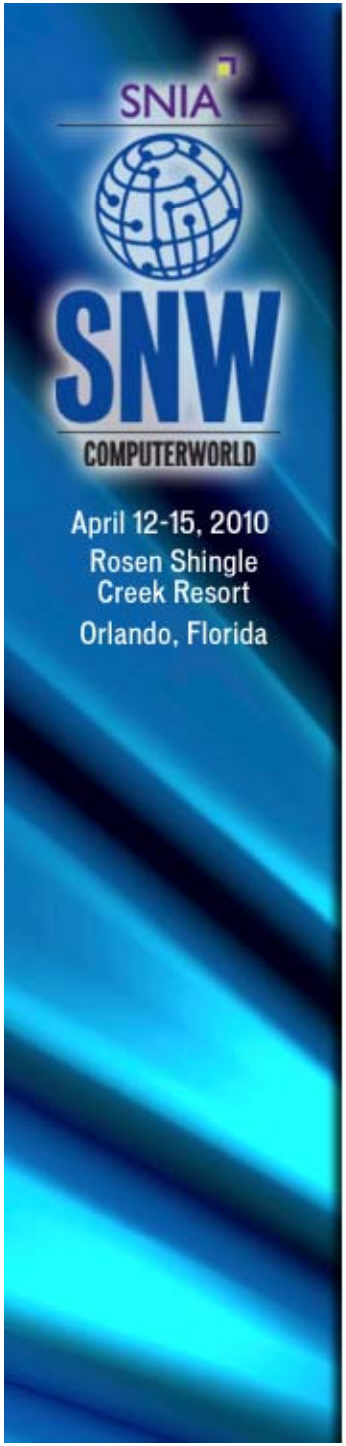
Backup snapshots to near-line disk for point and time data protection

Single mailbox restore capability with email system vs full backup and restore saves both disk space and time



Next Steps – Storage Efficiency

- Implement integrated snap shot backup solution for virtual Machines across storage system and data centers
- Implement snapshot clones of production systems for staging, qa, and development environments
- Shared source directories for web servers



Questions?

Email: gleathers@kbb.com

Linked In: http://www.linkedin.com/in/gleathers

Twitter: @grantlabt