SNIA Cloud Storage Summit

Sajai Krishnan
CEO

January 2009
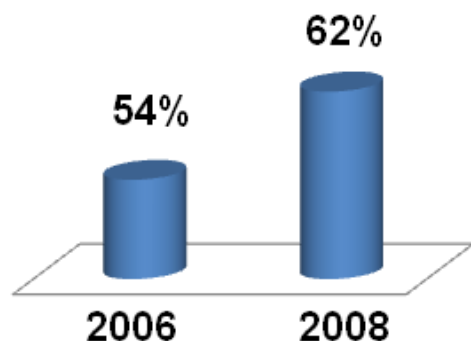
parascale
Powering Cloud Storage

Agenda

- Cloud storage 2009 & beyond
- Challenges SNIA can help solve
- Definition of Cloud Storage
- Use cases ParaScale is seeing
- About ParaScale
- SNIA & Cloud Storage
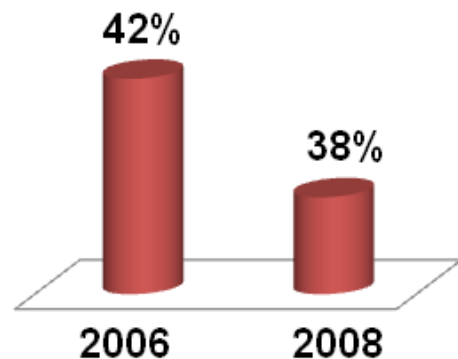 - Problems needed solving
  - Suggested APIs

parascale
Powering Cloud Storage

State of the Industry

Unstructured Data
- 54% — 2006
- 62% — 2008

versus

Structured Data
- 42% — 2006
- 38% — 2008

Content Depot % of Unstructured Capacity
- 4.3 % — 2005
- 17.4% — 2008
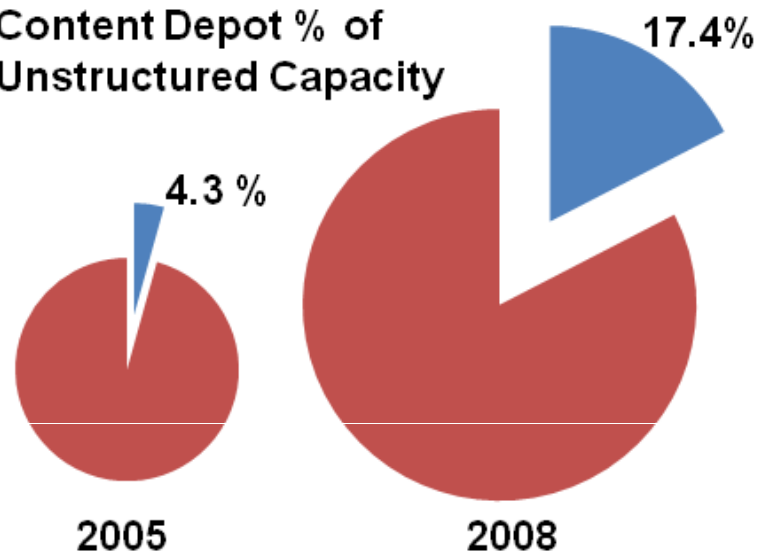
- Content depots of unstructured data are growing rapidly
- Content depots costs represent 29% per GB of average enterprise costs

Cloud Storage addresses the "different"

- **Explosion in use of storage and content**
  - Many examples
    - 530 TBs – YouTube content
      - Don't forget about - Tivo, Netflix, Hulu.com, etc.
    - 460 TBs – Digital weather data complied by National Climatic Data Center
    - 600 TBs – Ancestry.com's Genealogy Database
    - 1 PB – amount of data processed by Google's Servers every 72 Minutes

- **It's not about more, it's about different**
  - Legacy storage platforms were not built to satisfy digital content requirements

parascale
Powering Cloud Storage

2009 and Beyond

- Transformational wave
- Early adoption stage now
  - 2008: cloud compute, and public cloud storage
  - 2009: private cloud storage
  - 2010: broad adoption
- Explosion in public providers
- Huge interest in private clouds
- But
  - Grid? Utility? next "shiny thing"?
  - Roach motel? "Cloud Lock-in"?
  - Security, privacy, SLAs

parascale
Powering Cloud Storage

Cloud Storage 2009 and Beyond

- **The economic downturn helps drive adoption**

  – The addition of private cloud solutions creates an environment that enables incremental adoption of cloud storage on a very broad scale

- **Building storage clouds is becoming as simple as installing a new application on your laptop**

  – Custom service offerings become self-install activities

- **Cloud storage starts small and scales-up as needed.**

  – Organizations are no longer over-building to address the potential for rapid growth

parascale
Powering Cloud Storage

Cloud Storage 2009 and Beyond

- The Enterprise recognizes that clouds are designed to be self-managing and don't require heavy IT manpower

  – Storage tiering, provisioning, and data movement are time consuming tasks that are automated in cloud storage

- Specialty clouds are created

  – Storage clouds can be tuned for specific uses or applications

  – For example, clouds can be tuned for archival very cost-effectively, or for streaming media performance

parascale

Powering Cloud Storage

Challenges SNIA Can Help Solve

- **Customers continue to ask about "Cloud Lock-in"**
  - Once deployed, how do I get out?
  - Can I choose different providers based on cost?
  - How do I need to change my application to speak different cloud languages?

- **Application providers are struggling as well**
  - Each cloud API / protocol is different, which do they implement?
  - Why do they need to duplicate work for different clouds?

- **Everybody wants a standard, SNIA can help**

Confidential

parascale
Powering Cloud Storage

Cloud Storage Defined

- **Defined by 3 characteristics**
  - Storage service delivered over a network
    - internet or intranet
  - Easy to scale (w.r.t to effort, economy)
  - Easy to manage

parascale
Powering Cloud Storage

What About Clustered NAS and HPC?

- **High Performance Computing**
  - For bleeding-edge single file performance
  - Use a custom protocol or client ⊗ for access
  - Custom high-speed intra-cluster interconnects, as every node is aware of every action
  - Deployed and managed by specialists/scientists
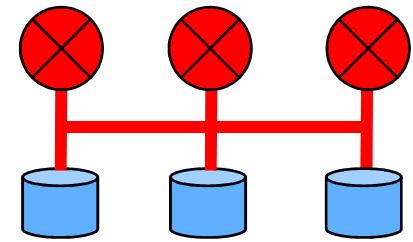
- **Clustered NAS is tightly coupled**
  - HPC heritage …more general purpose access, but design is still around optimized single file I/ODifficulty scaling beyond 10's of nodes
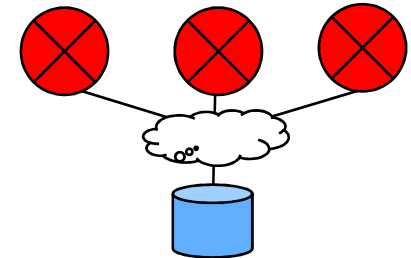 - Managed by expert storage admins

parascale
Powering Cloud Storage
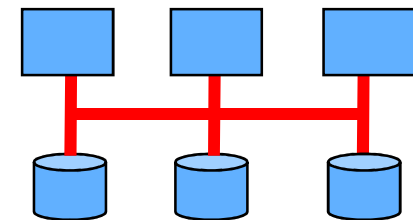
Public Versus Private Cloud Storage

- ## Public Clouds
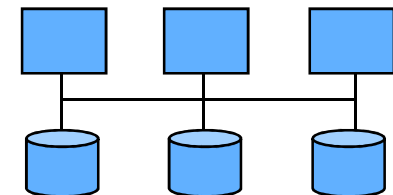
  – Storage service … delivered over the internet

  – Typically uses a custom protocol or client ⊗ for access

  – $/gig/month ++ Bandwidth fees

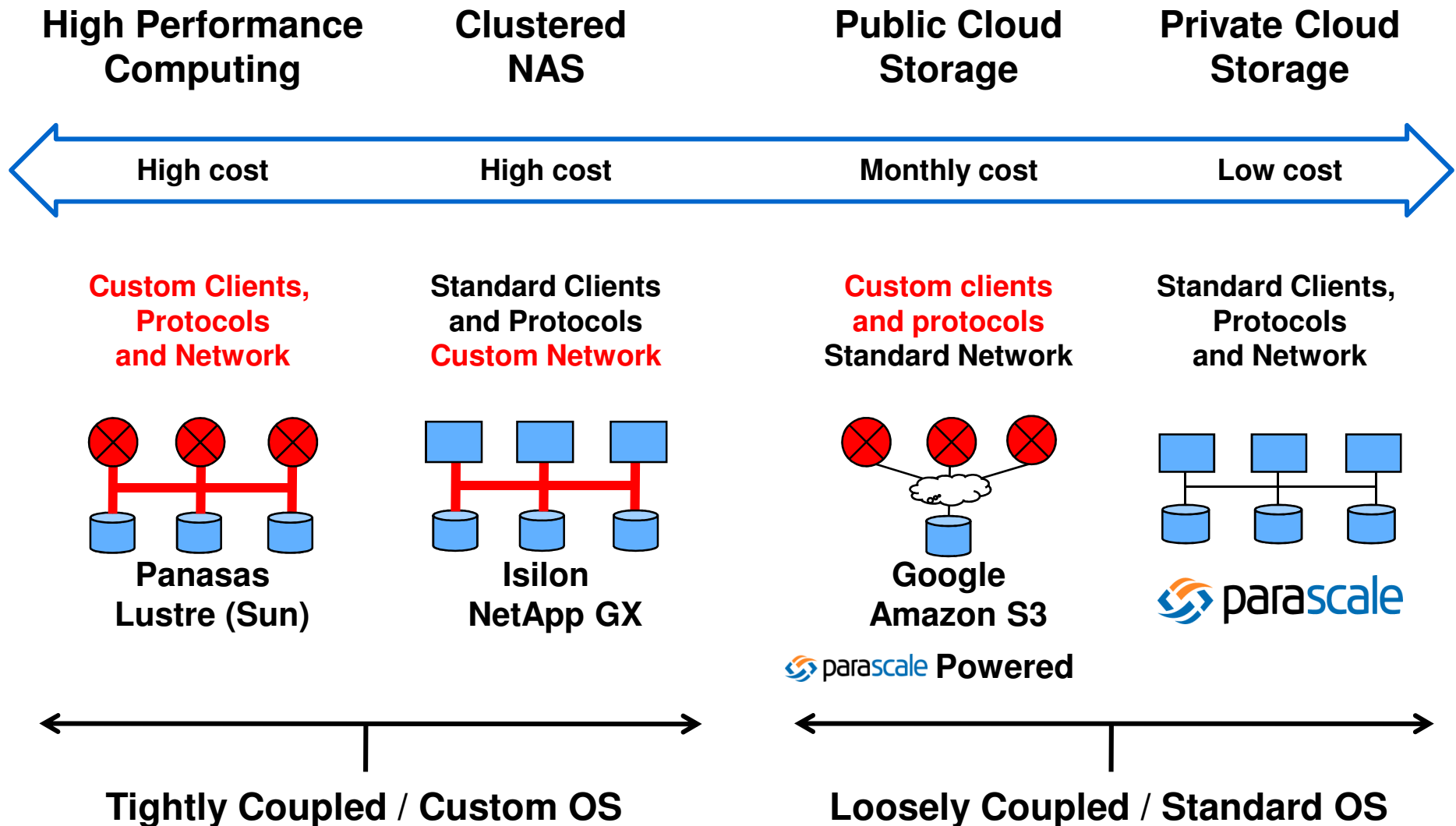- ## Private Clouds

  – A service, but delivered over intranet by IT org

  – Standard protocols for access

  – Inside the firewall, managed by own IT team

  – Start small, scale easily with commodity hardware

parascale
Powering Cloud Storage

Scalable File Storage Technology Landscape

High Performance Computing	Clustered NAS	Public Cloud Storage	Private Cloud Storage
High cost	High cost	Monthly cost	Low cost
Custom Clients, Protocols and Network	**Standard Clients and Protocols** **Custom Network**	**Custom clients and protocols** **Standard Network**	**Standard Clients, Protocols and Network**
Panasas Lustre (Sun)	**Isilon NetApp GX**	**Google Amazon S3** parascale **Powered**	parascale

Tightly Coupled / Custom OS

Loosely Coupled / Standard OS

Data Types for Cloud Storage

- **Larger files with lots of read access**
  - Digital content

- **Parallel streaming writes**
  - video surveillance (private clouds)

- **Long-term storage files**
  - Backup and archival files (private clouds)
  - Medical images, Energy exploration, Genomics

- **Geographically shared files**
  - Access from different geographies (public clouds)
  - Movie trailers, training videos

parascale
Powering Cloud Storage
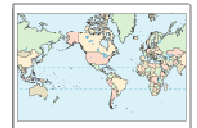
Where is Cloud Storage a Poor Fit?

- **Active Corporate Data**
  - Advanced data protection schemes
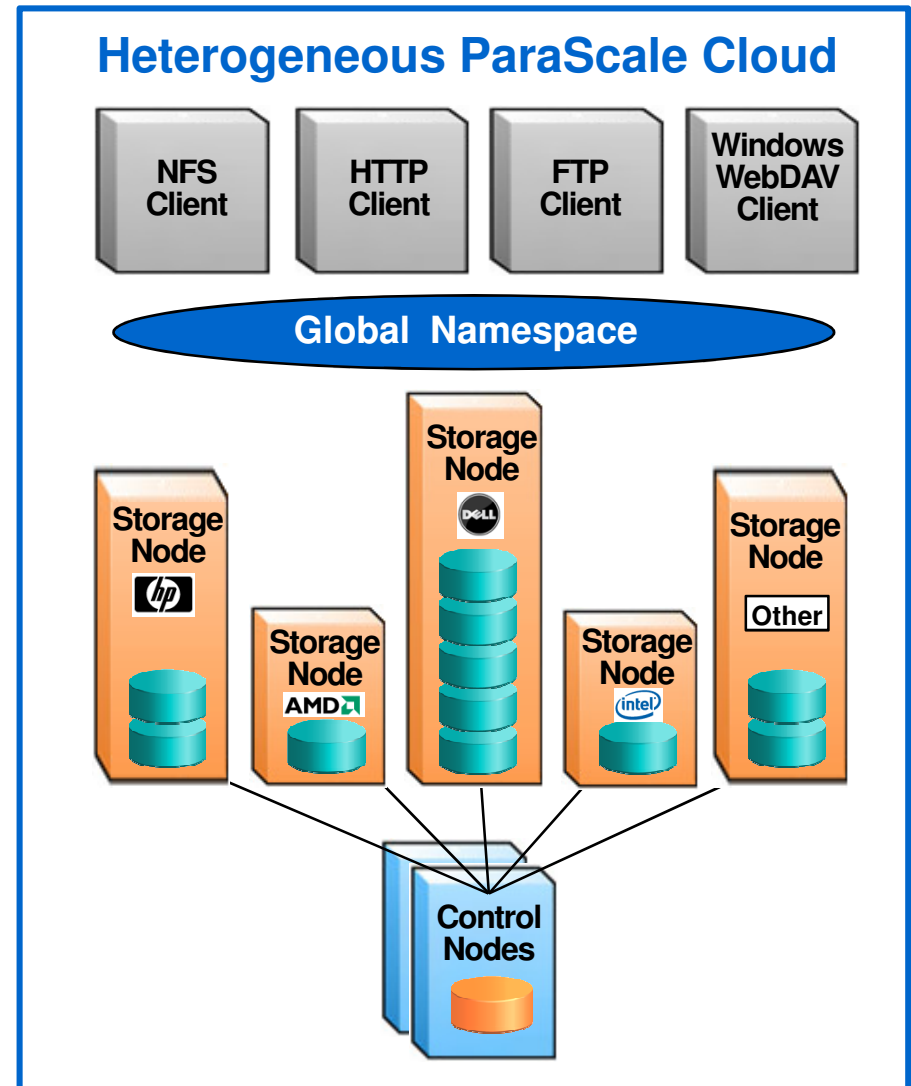 - Office Documents, Spreadsheets
  - Source-code

- **Transactional Data**
  - Frequent read and write access
  - Massive I/O requirements
  - Database, source code, Active VMware images

parascale
Powering Cloud Storage

ParaScale: commodity servers, standard protocols

- ## Software-only on standard hardware
  - Standard clients and protocols
  - No agents
  - Commodity hardware and networks
  - Single file performance limited to performance of single node

- ## Benefits
  - Transparent data migration
  - Massive scalability, start small
  - Economical
  - Your choice of hardware

Heterogeneous ParaScale Cloud

NFS Client | HTTP Client | FTP Client | Windows WebDAV Client

Global Namespace

Storage Node (DELL) | Storage Node (hp) | Storage Node (AMD) | Storage Node (intel) | Storage Node (Other)

Control Nodes

Confidential

parascale
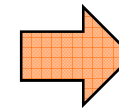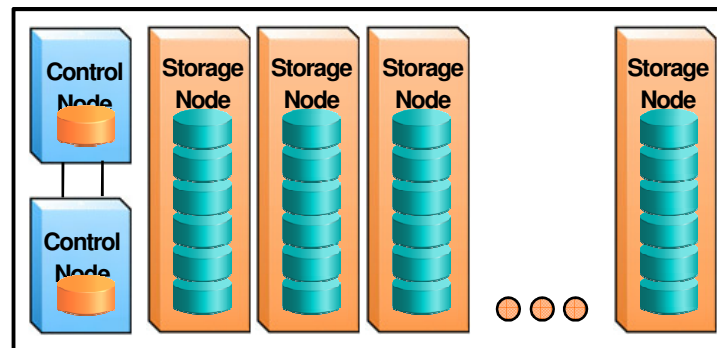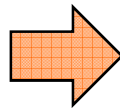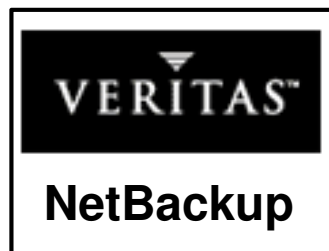Powering Cloud Storage

Blue Coat - Disk-to-Cloud-to-Tape Backup

Accelerate file backup and recovery

- VERITAS Netbackup writes multiple system backups to many ParaScale storage nodes in parallel to shorten the backup window

- Backups are retained on disk for 20 days and then spooled to tape

- Netbackup chunk sizes can be tuned to maximize performance when writing to the cloud

Wants Cloud Storage to:

Avoid the cost of extending Tier1 NAS for D2D2T

"ParaScale gives us a rock-solid platform for maintaining backups on-line. Helps me meet my back-up window affordably"

NetBackup

Control Node

Control Node

Storage Node   Storage Node   Storage Node   Storage Node

parascale
Powering Cloud Storage

Confidential

Other Use Cases

- ## Video Streaming

  - Increased per-node performance based on faster hardware

- ## Genomics Archives

  - Leverage larger disk drives and legacy servers

Confidential

SNIA Cloud Storage Summit

Cameron Bahar
CTO

January 2009

A Pitch to Vendors: There will be many winners

- ## Cloud is not about a single vendor owning the space

  - This is a transformational process in IT and there will me many winners
  - Proprietary APIs and lock-in only preserve early customers and limit the market for everyone

- ## A standard should address both local and remote access

  - The private versus public debate may rage on but customers are leveraging both
  - We need to build a standard that addresses both

- ## Speed is the key

  - A standard 2 years from now only makes the market smaller for everyone

Confidential

parascale
Powering Cloud Storage

Problems Needed Solving

- **Accessing data in the cloud**
  - A common and simple API

- **Managing the Cloud**
  - Web services and ISV hooks

- **Security**
  - Everybody's number one concern

- **Extended Metadata**
  - Manage information, not files

Confidential

Accessing data in the cloud

- **Common data access protocol**
  - HTTP based

- **Need for a simple universal API**

- **WAN optimized**

- **Open source and commercial clients emerge…**

Confidential

Managing the cloud

- **Common set of API's**

- **Use web services**

- **SPs tie managing clouds to existing OSS**
  - OSS = operational support system

- **ISVs extend existing products to support these API's and extend offerings to cloud platforms**

- **Ecosystem builds**

Confidential

parascale
Powering Cloud Storage

Security

- **Need for common WAN friendly security model**

- **Currently use NFS and HTTP access control**
  - Doesn't scale to WAN and public clouds

- **Need for trusted third party authentication**
  - OpenID?

- **Enables a multi-vendor cloud deployment model**
  - Backup one cloud to another vendors cloud!
  - Synchronously write to 2 clouds for DR!

Confidential

parascale
Powering Cloud Storage

Extended Metadata

- **API's to extend current metadata**
  - Extend current inode style metadata
  - Manage information; not files or objects

- **Allow tagging**

- **Allow external search and discovery**

Confidential

parascale
Powering Cloud Storage

parascale
Powering Cloud Storage

For more information:

www.parascale.com
ps_sales@parascale.com