# HEC IWG File Systems and Storage Workshop

Peter Corbett

Technical Director, Network Appliance

August 15, 2005

# What is Netapp Currently Funding?

- Open source efforts
  - Linux NFS client and server
  - FreeBSD
  - Xen
- Several university research groups
- Internal R+D efforts
  - pNFS
  - Indexing
  - Parallel file systems
  - High performance file systems
  - RDMA
  - NFS-RDMA

# Problems to Solve

- All the big problems arise from scaling
- Exponential growth rates of all interesting system performance and capacity numbers
  - Disk capacity growing faster than anything else
  - System capability: CPU speed, MP, clusters
  - Federations, Wide Area file systems, Storage Grids
- Total accessible storage is growing at a phenomenal rate
- Number of spindles per system must grow faster than CPU count to maintain balance of I/O and processing

# Scaling Issues

- Four key problem areas:
  - How to utilize commodity hardware effectively to very large scale
  - How to manage vast amount of data and storage
  - How to increase reliability, integrity and security
  - How to extract information from data

# Utilizing Commodity Hardware

- Speeds and Feeds
  - Like air and water
- Traditionally, most of the research effort has gone here
- Four areas of rapid development:
  - Parallel file systems
  - NFSv4
  - Clusters and Federations
  - Low-cost high-performance hardware

# Parallel File Systems

- Parallel file systems will eventually mature
  - Become well-integrated into system
  - Become ubiquitous
  - Become reliable and high-performing under a variety of workloads
  - Present a standardized interface to the clients
- There is still plenty of work to do here
  - Much of it will be done by system vendors

# NFSv4

- Three rules:
  - Standards, standards and standards
- V4 can (and should) become the standard upon which HPC deployments take place
  - pNFS
  - NFS RDMA
  - Sessions
  - Directory delegations
  - Byte-range delegations
  - Security
  - Redirection
- Standards leverage the whole community and level the playing field

# Clusters and Federations

- Huge array of interesting problems to solve
- How to connect, manage, balance, recover, secure
- There is room to define standards for interoperability
  - Data migration
  - Remote caching
  - Mirroring and DR

# Low-cost Hardware

- IB
- SAS
- SATA
- PCI express
- Ethernet

# How to manage vast amounts of data

- Human admin does not scale
  - Limited cognitive budget per byte
  - Must reduce management cost per byte by approximately the rate that accessible capacity scales
- Boundaryless storage
- Virtualization at all levels of system
- Transparent data migration
- Robust systems

# Four reasons to migrate data

- Protection
- Load balancing
- Cost of storage
- Proximity to user
- All of these can drive automated data migration

# Four stages of automation

- Baseline: System reports all events, admin filters and acts
- System filters information, presents outliers only, admin acts
- System automates activity, reports to admin, admin sets policy, admin adjusts if needed
- System performs autonomously, admin can query status and adjust, but otherwise can safely ignore, admin sets policy
- This progression can be applied at various tiers of the storage hierarchy

# Reliability, Integrity, Security

- As systems become more complex, they become more vulnerable
  - To failures
  - To attack
- Cheap scale is both a challenge and an opportunity
  - Lots of failures
  - Lots of built-in redundancy

# Extracting Information from Data

- Data becoming more semantically rich
  - XML, embedded schema, self-describing
- File systems have under-utilized capabilities to annotate data
  - V4 supports named attributes
  - Additional attribute namespace below each file
- Indexing is a huge and very interesting problem area