NFS: A Customer Perspective

Bridging the Gap Between Research & Production

Joyent

Joyent Highlights

- Web 2.0 Infrastructure & Application Company
- Provides 5 core products:
  - Shared & Business Hosting
  - Solaris Container (VPS) Hosting
  - BingoDisk WebDav Storage
  - StrongSpace Secure SFTP Storage
  - Connector Collaboration Application

Joyent

Providing NAS to Web 2.0

- Clients running a highly diverse range of applications with diverse needs
- File sizes vary widely, from IMAP to Oracle Databases
- Storage demands vary widely, from serving movies to OLTP databases
- No direct control over how users utilize storage
- One storage infrastructure must accommodate every possible customer need without customization and co-exist with everyone else.

Joyent

Block & File: Yin & Yang

- iSCSI & NFS must co-exist in harmony
- … and be managed similarly.
- … on one storage device.
- … and it should be easy.

Joyent

iSCSI & NFS: The Zones Issue

- Solaris Containers can not (in the foreseeable future) live on NFS
- Multiple Containers need to access shared storage
- Binding a Container to a physical system (hunk of metal) is impractical in a diverse environment
- Solution: Store Containers on iSCSI to provide physical independence from the metal and provide shared storage via NFS.

Joyent

OpenSolaris Makes It Possible

- OpenSolaris contains a native iSCSI Target Implementation
- OpenSolaris contains a native iSCSI Initiator Implementation
- OpenSolaris contains an excellent NFSv3/4 client/server implementation
- OpenSolaris contains the worlds most advanced integrated Volume Manager & File System that binds everything seamlessly together: ZFS

Joyent

Dreams Become Reality: ZFS

- ZFS provides a hub for ultimate storage integration
- Disk Management via zpool
- Snapshot Management & Replication
- iSCSI: zfs set shareiscsi=on pool/dataset
- NFS: zfs set sharenfs=rw pool/dataset
- Check-summing and Integrity Assurance
- Quotas, Reservations, Compression, Thin Provisioning, Encryption (Soon), ...

Joyent

Pull Out the Propeller Caps

- OpenSolaris provides the best research platform on the planet
- Observability is, um, important!
- Man made it, man should be able to see what its doing
- Best tools available for research: mdb, Dtrace, ptools, resource control, etc.
- Complete Open Source codebase and development model

Joyent

OpenSolaris End-to-End

- Observability by tools like DTrace are revolutionary on the server
- … they’re unheard of on your storage subsystem.
- Non-heterogeneous environment provides complete compatibility and removes countless variables from problem analysis and design.
- Building blocks can be used and re-used on both server and storage.

Joyent

ZFS on ZFS

- ZFS has the unique ability to host itself: ZFS with a ZFS (Zpool within a ZVol)
- We can have control and give control.
- Example: ZFS providing ZVol's as iSCSI Targets to customer Containers within which are ZFS filesystems

Joyent

NFSv3 Environment

- Well understood.
- Well trusted.
- Mature.
- We’re continuously evaluating NFSv4.
- Unknowns in any technology can add up quickly.

Joyent

Replication & pNFS

- ZFS Replication is currently limited in environments with thousands of datasets
- Each dataset is snapshotted and then replicated individually
- Sun StorageTek Availability Suite provides synchronous replication, but secondary copy is unusable.
- pNFS provide a real replication scheme, but OpenSolaris development has a ways to go.

Joyent

Storage Architects Can Architect

- Pulling all the components of OpenSolaris provides us with a building blocks to build new an innovative solutions... today.
- Example: A replication scheme would be achieved by using a frontend NFS server which forges multiple backend iSCSI Targets (zvol's) into a single scalable zpool and provide NFS and iSCSI storage in a single unified pool.

Joyent

Thumper: Freedom to Dream

- 24TB of storage means flexability
- ZFS gives us unparalleled control with minimal effort
- More power than any existing storage hardware on the market
- A full enterprise grade OS brings features that are unconventional on a storage solution, such as C4 Auditing, Advanced Security, Traffic Shaping, Resource Control, etc...

Joyent

Customer Result: Unparalleled Flexibility

- If a customer wants 1TB of storage, we don't worry, we give it to them in the time that it takes to type: 'zfs set quota=1t pool/dataset'
- If a customer needs a mix of block and file storage, its trivial to implement.
- Reduced support issues due to a unified architecture
- Its easier to scale our internal support staff.

Joyent

But We Still Have Room To Improve

- Small file locking (IMAP) can overwhelm NFS-on-ZFS when done on large scale. Work in progress.
- NFS-on-ZFS bulk file creation is impacted by the ZFS Intent Log (ZIL). Currently needs to be disabled in production. Work in progress.
- Recursive replication is still a work in progress.
- Associated technologies are still in progress, example: iSNS Server.

Joyent

Community is Important

- Bring together observability with a connection to the development community means problem analysis is done in the field, in production, not held up for reproduction in a lab
- Ability to drive the future of the technologies
- Ability to contribute to the future of the technologies
- Ability to test, provide feedback on, and implement technology while its fresh.

Joyent

NFSv4 Provides More Possibility

- We continue to evaluate NFSv4
- We’re anxiously awaiting a pNFS server prototype
- NFS RDMA on 10g Ethernet?

Joyent

Questions?

Ben Rockwood
Director of Systems
Joyent, Inc

Company: www.joyent.com
Personal: www.cuddletech.com

Joyent