RED HAT SUMMIT
10 YEARS and counting
SAN FRANCISCO | APRIL 14-17, 2014

Red Hat Enterprise 7 Beta File Systems

New Scale, Speed & Features

Ric Wheeler
Director
~~Red Hat Kernel File & Storage Team~~
Red Hat Storage Engineering

Agenda

- Red Hat Enterprise Linux 7 Storage Features
- Red Hat Enterprise Linux 7 Storage Management Features
- Red Hat Enterprise Linux 7 File Systems
- What is Parallel NFS?
- Red Hat Enterprise Linux 7 NFS

10 YEARS *and counting*

SAN FRANCISCO | APRIL 14-17, 2014

Red Hat Enterprise Linux 7

Storage Foundations

redhat.

Red Hat Enterprise Linux 6 File & Storage Foundations

- Red Hat Enterprise Linux 6 provides key foundations for Red Hat Enterprise Linux 7
  - LVM Support for Scalable Snapshots
  - Device Mapper Thin Provisioned Storage
  - Expanded options for file systems
- Large investment in performance enhancements
- First in industry support for Parallel NFS (pNFS)

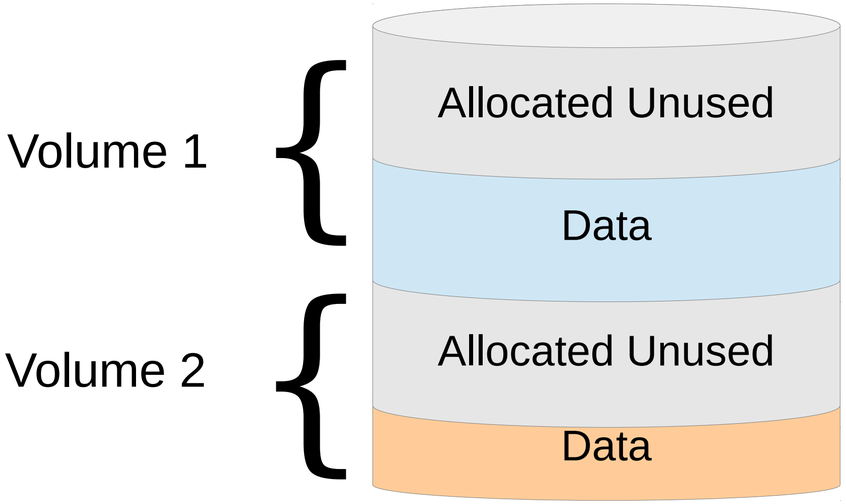
LVM Thinp and Snapshot Redesign

- LVM thin provisioned LV (logical volume)
  - Eliminates the need to pre-allocate space
    - Logical Volume space allocated from shared pool as needed
  - Typically a high end, enterprise storage array feature
  - Makes re-sizing a file system almost obsolete!
- New snapshot design, based on thinp
  - Space efficient and much more scalable
    - Blocks are allocated from the shared pool for COW operations
    - Multiple snapshots can have references to same COW data
  - Scales to many snapshots, snapshots of snapshots

RHEL Storage Provisioning
Improved Resource Utilization & Ease of Use

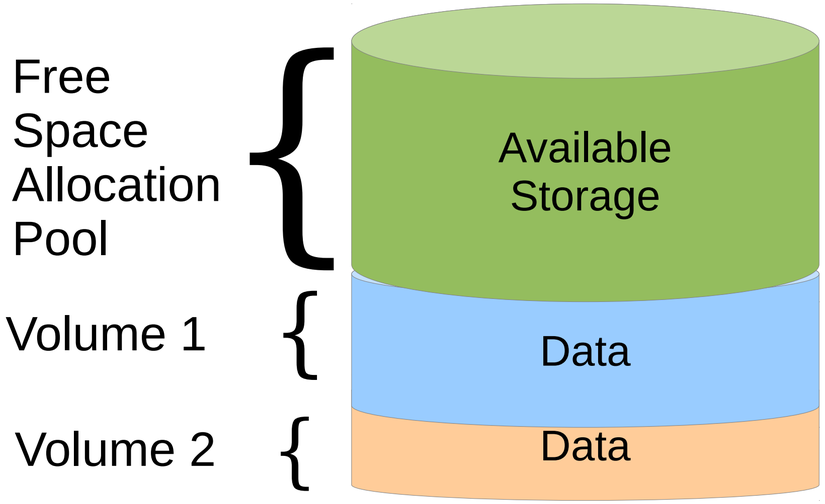
Avoids wasting space
Less administration
* Lower costs *

Traditional Provisioning



Inefficient resource utilization
Manually manage unused storage

RHEL Thin Provisioning



Efficient resource utilization
Automatically managed

Even More LVM Features

- LVM provides additional RAID features through the use of existing software RAID (md) code
  - RHEL 6.4 added LVM Support for RAID10
- Caching of metadata provides major performance gains on systems with many disks
  - Prevents large delays while by preventing rescanning of SAN fabrics

Tiered Storage Caching Schemes

- SSD's are fast but expensive
  - Cost makes it difficult to have a large system with only SSD storage
  - Using traditional drives for capacity and SSD's for cache is a cost effective performance enhancement
- Red Hat Enterprise Linux 7 block caching
  - Device mapper's dm-cache target is a block level cache
- Red Hat Enterprise Linux 7 file caching
  - Fs-cache caches data for network file system clients like NFS
- Both systems are flexible about the class of storage used

10 YEARS *and counting*

SAN FRANCISCO | APRIL 14-17, 2014

Red Hat Enterprise Linux 7

Storage and File System Management

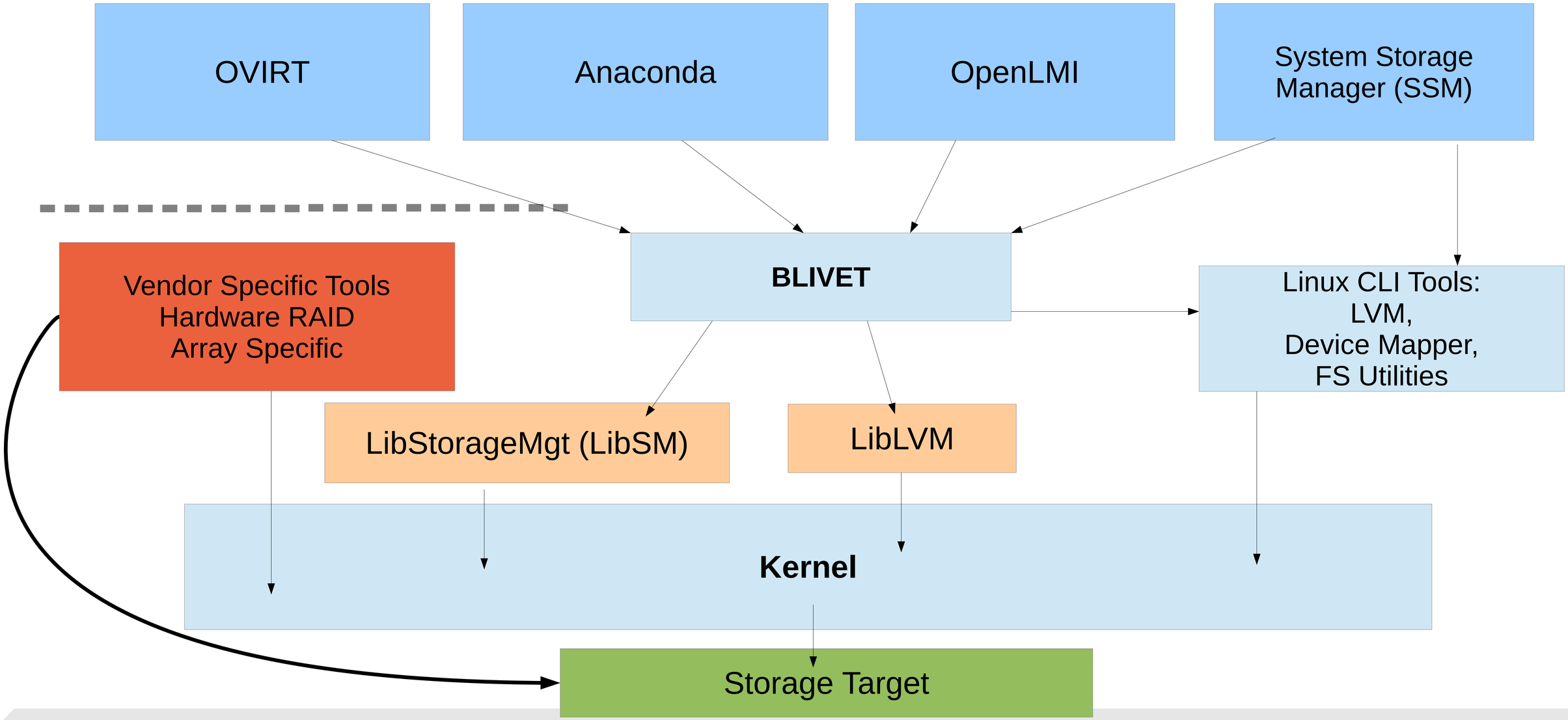
redhat.

Storage Management APIs and CLI

- libstoragemgt manages SAN and NAS
- liblvm is the API equivalent of LVM user commands
- Blivet is a new high level storage and file system library that will be used by anaconda and OpenLMI
- System Storage Manager provides an easy to use command line interface which integrates volume and file system management

Unification of storage management code

Future Red Hat Stack Overview



Thinly Provisioned Storage & Alerts

- Thinly provisioned storage “lies” to users
  - Similar to DRAM versus virtual address space
  - Each user sees a large, virtual device
  - Pools of user devices backed by a shared pool of physical storage
- Alerts provided when the physical pool hits a watermark
  - Supported for storage arrays and for device mapper dm-thinp storage

RHEL 7 Storage Summary

- SSD’s
  - Hierarchical/Tiered storage (Device mapper cache)
  - MultiQueue block layer support for high speed SSD's
- Interconnect support for NVMe, SOP, SAS-3 devices
- Linux-IO (LIO) SCSI Target
- Asynchronous SCSI events
- Software RAID enhancements

10 YEARS *and counting*

SAN FRANCISCO | APRIL 14-17, 2014

Red Hat Enterprise Linux 7

File Systems

redhat.

Red Hat Enterprise Linux 7 - Areas of Focus

- Expanded file system choices
- Support for new hardware
  - Focus on very high performance, low latency devices
  - Support for higher capacities across the range of file and storage options
  - Working with our storage partners to enable their latest devices
- Ease of use and management

Red Hat Enterprise Linux 7 - More Choices

- RHEL 7 will support ext4 and XFS
  - All can be used for boot, system & data partitions
- Ext2 and Ext3 are fully supported
  - Uses the ext4 driver, which is invisible to users
- BTRFS is still under intensive development
  - Working on finalizing multi-disk, RAID and btrfs user space tools
  - Technology preview in Red Hat Enterprise Linux 7
- Additional support for new pNFS client layouts

Red Hat Enterprise Linux 7 - Default File System

- In RHEL 7, Red Hat made XFS the new default file system
  - XFS is the default for boot, root and user data partitions
  - Included as part of the base on all supported architectures
- Red Hat worked with partners and customers during this selection process to test and validate XFS

XFS Strengths

- XFS is the reigning champion of larger servers and high end storage devices
  - Tends to extract the most from the hardware
  - Well tuned to multi-socket and multi-core servers
- XFS has a proven track record at scale
  - RHEL 6 certified partners run XFS up to 300TB
  - Maximum RHEL 7 XFS file system size is 500TB
- Popular base for enterprise NAS servers including Red Hat Storage

EXT4 Strengths

- Ext4 is very well know to system administrators and users
  - Default file system in RHEL 6
  - Closely related to ext3 our RHEL 5 default
  - Base file system for Android and Google File System
- Can outperform XFS in some specific workloads
  - Single threaded, single disk workload with synchronous updates
- Increased maximum file system size in RHEL 7 is 50TB
  - RHEL 5 and RHEL 6 maximum for ext4 is still 16TB

BTRFS – A New File System Choice

- Integrates many IO functions into the file system layer
  - Logical volume management functions like snapshots
  - Can do several versions of RAID
- Designed around ease of use
- Back references map IO errors to file system objects
- Great choice for systems with lots of independent disks and no hardware RAID
- Maximum supported file system size is 50TB
- Target is for full support in a minor RHEL 7.x release

Why Is BTRFS Still Technology Preview?

- File systems take a ***long*** time to develop to enterprise quality
  - btrfs as a project is over five years old
  - Now used in some distributions for system partitions only
  - btrfs repair tools still very basic
  - Lots of feature and code churn in the upstream project
- Red Hat is working with partners in upstream on stability
  - Pushing back on feature development
  - Focused on data integrity and specific use cases

How to Choose a Local File System?

- The best way is to test each file system with your specific workload on a well tuned system
  - Use the RHEL tuned profile for your storage
- The default file system will just work for most applications and servers
  - Many applications are not file or storage constrained
  - If you are not waiting on the file system, changing it probably will not make your application faster!
- Start with the default mkfs and mount options as well!

Red Hat Enterprise Linux 7 - GFS2 New Features

- Streamlined journaling code
  - Less memory overhead
  - Less prone to pauses during low memory conditions
- Faster fsck
- RAID stripe aware mkfs
- New cluster stack interface (no gfs_controld)
- Performance co-pilot (PCP) support for glock statistics

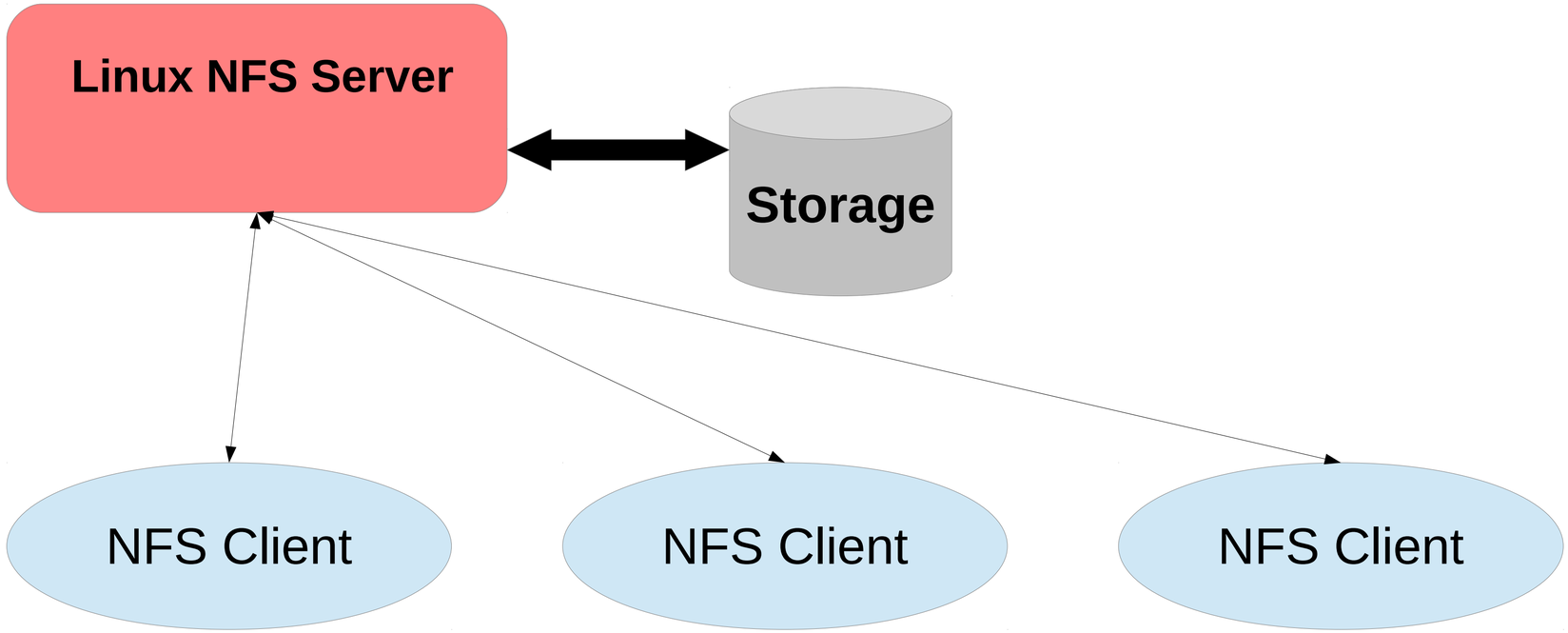
RED HAT SUMMIT
10 YEARS *and counting*
SAN FRANCISCO | APRIL 14-17, 2014

Red Hat Enterprise Linux 7

What is Parallel NFS?

Traditional NFS

One Server for Multiple Clients = Limited Scalability

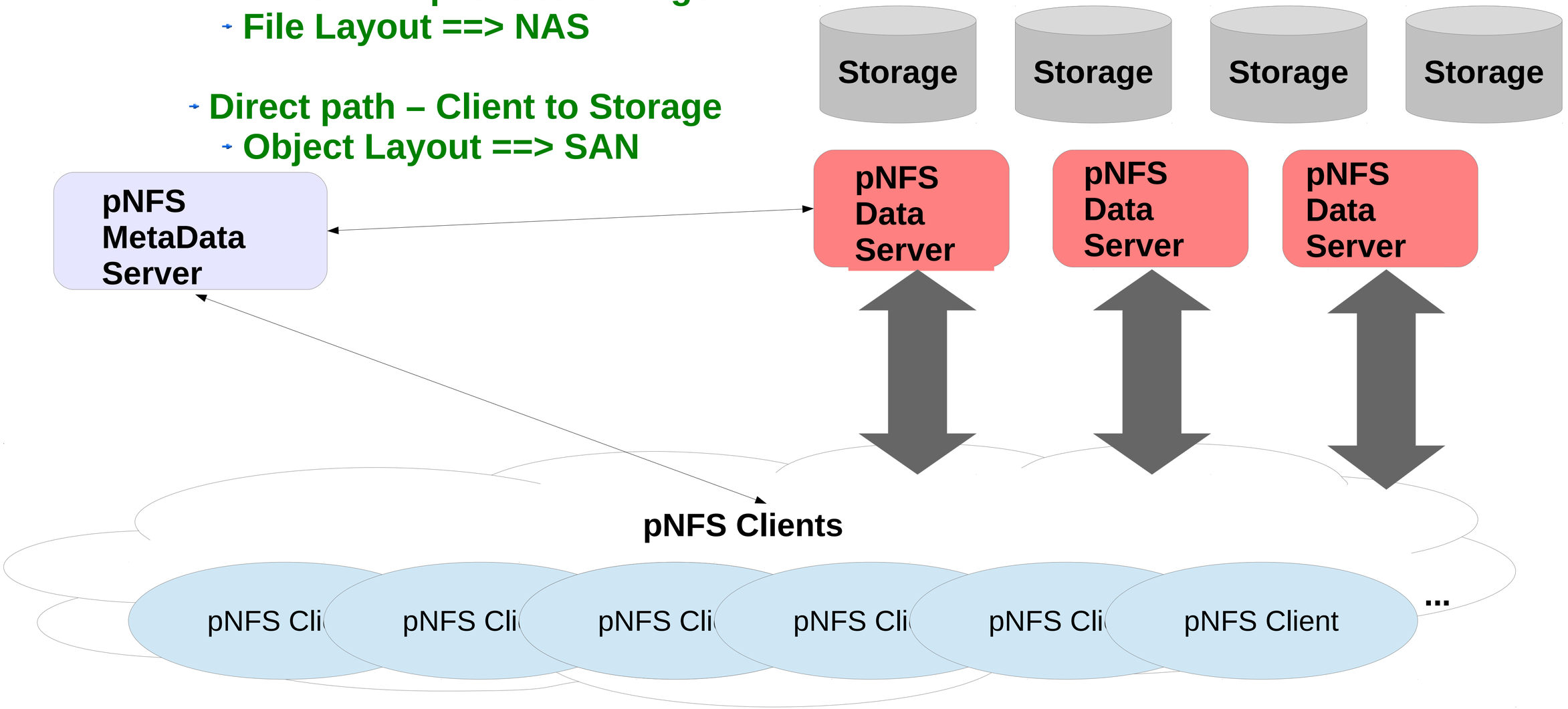


Parallel NFS (pNFS)

- Architecture
  - Metadata Server (MDS) – Handles all non-Data Traffic
  - Data Server (DS) – Direct I/O access to clients
  - Shared Storage Between Servers
- Layout Define server Architecture
  - File Layout (NAS Environment)
  - Block Layout (SAN Environment)
  - Object Layout (High Performance Environment)

Parallel NFS = Scalability

- Parallel data paths to Storage
  - File Layout ==> NAS
- Direct path – Client to Storage
  - Object Layout ==> SAN



pNFS Introduction in RHEL 6.4

- First to market with pNFS client support
  - File layout popular with leading many enterprise NFS array vendors
  - Active work done with the Linux community and Red Hat partners
- Easy to use
  - **mount -o v4.1 server:/export /mnt/export**

RED HAT SUMMIT
10 YEARS *and counting*
SAN FRANCISCO | APRIL 14-17, 2014

Red Hat Enterprise Linux 7

NFS Features

Parallel NFS Updates

- Parallel NFS has three layout types
  - File layout is in full support in RHEL 6.5 and RHEL 7
  - Object layouts to an object backend is technology preview in RHEL 7
  - Block layouts for a SAN backend is not supported
- Feedback welcome on your vendor layout type!

Red Hat Enterprise Linux 7 NFS Server Updates

- Red Hat Enterprise Linux 7.0 completes the server side support for NFS 4.1
  - Support for only-once semantics
  - Callbacks use port 2049
- No server side support for Parallel NFS
  - Red Hat Storage is working on a technology preview of a user space pNFS enabled Ganesha server

Support for SELinux over NFS

- Labeled NFS enable fine grained SELinux contexts
  - Part of the NFS 4.2 specification
- Labeled NFS does not give full support for xattrs over NFS
  - Still debating that case with upstream
- Use cases include
  - Secure virtual machines stored on NFS server
  - Restricted home directory access

Learn more about File and Storage

- *The new world of NFS*
  Tuesday 2:30 pm
- *Red Hat Storage Server: Roadmap & Integration with OpenStack*
  Tuesday 2:30 pm
- *Fundamentals of LVM with Red Hat Enterprise Linux 7 beta (**Lab**)*
  Tuesday, April 15 3:50 pm
- Demonstration (**Partner Pavilion**) of System Storage Manager (SSM)
  Tuesday 10 am-noon & Wednesday 1-2 pm
- Engage the community:
  - http://lwn.net
  - Mailing lists: linux-lvm, linux-ext4, linux-btrfs, linux-nfs, linux-xfs, ...