Container-Native Storage & Red Hat Gluster Storage Roadmap

Sayan Saha
Director, Product Management, Storage Business

Sudhir Prasad
Product Management, Storage Business

Date: 08-May-2018
AGENDA

- Overview
- Red Hat Gluster Storage roadmap
- Container-Native Storage roadmap
- Red Hat Storage One
RED HAT STORAGE PORTFOLIO FOCUS

**RED HAT STORAGE**

**PHYSICAL**
- RED HAT CEPH STORAGE
- RED HAT GLUSTER STORAGE
- RED HAT ENTERPRISE LINUX

**VIRTUAL**
- RED HAT CEPH STORAGE
- RED HAT GLUSTER STORAGE
- RED HAT ENTERPRISE LINUX
- RED HAT VIRTUALIZATION

**PRIVATE CLOUD**
- RED HAT CEPH STORAGE
- RED HAT GLUSTER STORAGE
- RED HAT ENTERPRISE LINUX
- RED HAT OPENSTACK PLATFORM

**CONTAINERS**
- RED HAT CEPH STORAGE
- RED HAT GLUSTER STORAGE
- RED HAT ENTERPRISE LINUX

**PUBLIC CLOUD**
- RED HAT CEPH STORAGE
- RED HAT GLUSTER STORAGE

#redhat #rhsummit
Gluster’s Journey @ Red Hat

Stability & Performance
2011-13

Enterprise Class Features
2014 - Present

Deep integration with OpenShift & Red Hat Virtualization
Ongoing & future
Gluster Architecture

Aggregates systems to one cohesive unit and presents using common protocols.
Key Features: Red Hat Gluster Storage

**SCALABILITY**
- Multi-petabyte support across hundreds of nodes
- Elastic hashing algorithm
- No single point of failure (metadata server)
- Pro-active self-heal and rebalance

**PROTOCOLS**
- Active/Active NFSv4
- SMB3 (protocol negotiation, in-flight encryption, server-side copy)

**EFFICIENCY**
- Standard servers and disks
- Erasure coding - reduced footprint
- Tiering

**PERFORMANCE**
- Rebalance Performance
- Self-heal
- Client-side caching (upcoming)
- Quick-read, write-behind

**SECURITY & DATA INTEGRITY**
- Bit rot detection
- SSL-based Network Encryption
- At-rest encryption using dm-crypt
- SELinux Enforcing Mode

**DATA SERVICES**
- Async Geo replication
- Snapshots & cloning
- SNMP monitoring and Nagios integration
- Backup hooks
- Compression and de-dup (via partner)
Flexible Deployment Options

- **BARE METAL**
  - Red Hat Enterprise Linux
  - Red Hat Gluster Storage

- **VIRTUAL MACHINES**
  - VM
  - Red Hat Gluster Storage

- **CONTAINERS**
  - Red Hat OpenShift Container Platform

- **PUBLIC CLOUD**
  - Red Hat Gluster Storage
Red Hat Storage - A visionary


Red Hat Storage ranked as a Visionary two years in a row

Red Hat Storage positioned farthest and highest in both Completeness of Vision and Ability to Execute in the Visionaries quadrant.

**Highlights**

- Strong customer traction across geographies/verticals
- Solid product strategy and road map
- Leadership in container-native storage and storage for the hybrid cloud, Private Cloud
- Object Storage, Hyperconverged Infrastructure

Gartner does not endorse any vendor, product or service depicted in its research publications, and does not advise technology users to select only those vendors with the highest ratings or other designation. Gartner research publications consist of the opinions of Gartner’s research organization and should not be construed as statements of fact. Gartner disclaims all warranties, expressed or implied, with respect to this research, including any warranties of merchantability or fitness for a particular purpose.

This graphic was published by Gartner, Inc. as part of a larger research document and should be evaluated in the context of the entire document. The Gartner document is available upon request from [https://engage.redhat.com/gartnermagic-quadrant-storage-s-201610121525](https://engage.redhat.com/gartnermagic-quadrant-storage-s-201610121525).
Red Hat Gluster Storage Product Family

- Software-only RHGS
- Container-Native Storage (for OpenShift) CNS
- Red Hat Hyper Converged Infrastructure (with RHV) RHHI
- Red Hat Storage One (plug & play storage servers)
PAST: Red Hat Gluster Storage Roadmap
Dynamic Volume Life Cycle Management (**Heketi**)

**What**
- RESTful service endpoint for **automated** volume lifecycle management
- Allocate/delete storage volumes dynamically and on-demand
- Provide Day-2 management capabilities including device, node removal and replacement

**Why**
- Enable integration with Kubernetes & OpenShift
- Key enabler for dynamic provisioning for OpenShift storage

**When**
- Late CY-2016
Example dynamic provisioning workflow with Heketi
Block Storage Support via iSCSI

What

- Provide block storage based on Linux iSCSI target stack (TCMU)
- iSCSI target is backed by a file in a gluster volume
- Management capabilities are provided by gluster-block

Why

- **Complete** storage solution for OpenShift
- Provide block storage for popular OpenShift workloads like MySQL, Cassandra, MongoDB and PostgreSQL.

When

- October CY-2017
- Automatically provisioned for OpenShift logging and metrics
Block Storage for Container-Native Storage

OPENSIFT NODE 1

Application Pod

APP CONTAINER

File I/O

OpenShift PersistentVolume

XFS

ISCSI LUN via gluster-block

ISCSI LUN via gluster-block

OPENSIFT NODE 2

GlusterFS pod

ISCSI

TCM core

TCM user space

Sparse File

GlusterFS Volume Mount

Kernel Space

User Space

File I/O

Block I/O
**Brick Multiplexing**

**What**
- Brick Mux allows handling of multiple bricks by one process, instead of the older brick per process model
- Reduces resource consumption, utilizing a single port, lower memory footprint
- Improves ability to manage resource contention

**Why**
- Increase PV density per 3-node cluster for OpenShift (from 300 to ~1000)
- Support high density storage servers for deep archival workloads

**When**
- October CY-2017
- Currently supported for container storage use-cases
- Generally available in summer 2018
Small-file Performance Enhancements

**What**
- Client-side caching (SMB, Fuse)
- Faster self-heal
- Faster rebalance
- Negative Lookup Cache (SMB, Fuse)
- Parallel Readdir (SMB)
- Lookup Optimize
- dm-cache support in Gluster bricks
  - Block based tiering
- Client io threads (dispersed volumes)

**Why**
- Make Gluster perform better for small file workloads (sub 64kb)
- Make Gluster interactive at scale when storing small files
- 8 - 10x improvement for small file performance

**When**
- Last 2 years
Small File Performance Improvements

Source: Red Hat Storage Architecture Team
Sharding Translator

What

- Traditionally VM images would map to exactly one file in a Gluster volume
- Fragments ("shards") a large file (like a VM image) into chunks and distributes them across the cluster
- Enables more granular healing and geo-replication, thus improving performance

Why

- Foundational feature for RHHI product
- Better back-end storage for VM images - prevents lock ups & pauses
- Support large files that would exceed the size of a brick
- Better disk bandwidth utilization

When

- Summer CY-2016
Ansible - **gdeploy**

**What**
- Ansible based tool for deploying Gluster
- Day-1 & Day-2 capabilities supported
- Provide Life-cycle management for
  - clusters
  - backend
  - client mounts
  - volumes
  - shares (Fuse, NFS, SMB)
  - SSL encryption

**Why**
- Red Hat Gluster Storage
- RHS One
- RHHI

**When**
- Last 1.5 years
# Red Hat Gluster Storage Web Admin (Monitoring)

<table>
<thead>
<tr>
<th>Overview</th>
<th>Hosts</th>
<th>Volumes</th>
<th>Bricks</th>
</tr>
</thead>
<tbody>
<tr>
<td>At-a-glance</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

## Health

**Healthy**

## IOPS

0

## Capacity Utilization

15.3%

## Capacity Available

31.7 GiB

## Weekly Growth Rate

6.0 MB

## Weeks Remaining

5.6
File/NAS Roadmap

RHGS 3.3.1
NOV 2017
- SMB performance improvements (NL cache - create & rename)
- Parallel readdir (dir listing)
- Rebalance progress monitoring
- Expanded EC coverage (8+2, 16+4)
- Monitoring + metrics
  - Capacity, Health & Performance
  - SNMP, SMTP
- Subdir exports using Fuse (Preview)

RHGS 3.3
Sep 2017
- SMB performance improvements (NL cache - create & rename)
- Parallel readdir (dir listing)
- Rebalance progress monitoring
- Expanded EC coverage (8+2, 16+4)

RHGS 3.4
Jul 2018
- Support upgrade to RHEL 7
- macOS client support
- NFS-Ganesha - directory chunking
- SMB Multi-channel (full support in async)
- Subdir exports using Fuse (full support)
- Punch Hole Support
- Brick Mux support generally available
PRESENT & FUTURE: Red Hat Gluster Storage Roadmap
GlusterD2

What

- Highly scalable and programmable management plane
- Dynamic provisioning (Heketi functionality) built-in
- More consistent and performant (etcd based)
- Full functional parity with current shipping GlusterD

Why

- Larger cluster sizes, higher volume densities
- Autonomous management - low admin intervention for DevOps, containerized workloads

When

- RHGS 4
Plus-1 Scaling

What

- Traditionally Gluster scales in multiples of server “replica counts” (2, 3, 6 etc.)
- Plus-1 scaling allows Gluster to scale in one brick (disk) increments
- Add one or more hosts with a bunch of disks, gluster will consume and serve it up
- No rebalance needed after storage addition

Why

- Flexible and economical storage expansion
- Dynamic storage addition

When

- RHGS 4
Dedupe & Compression

**What**
- Leverage VDO technology (from Permabit acquisition)
- Gluster bricks will use VDO backed LVM volumes
- Transparent to end user
- Eliminate zero blocks, dedupe, compress inline

**Why**
- Storage efficiency
- Data reduction
- Dedupe container registries

**When**
- Already available in RHEL 7.5
- RHGS will incorporate this in RHGS 4 for File Storage
- RHHI will leverage this in summer 2018
Example savings with VDO

- 259 GB reduced to 46 GB
- 5X space savings

Source: Red Hat Engineering
Active-Active multi-master replication

What
● Stretch a Gluster volume across multiple data-centers including public clouds
● Write/update a single Gluster volume from multiple Geographical sites
● Policy based reconciliation of conflicting writes
● All reads are local, writes are propagated via Gluster self-heal

Why
● Enable storage federation for OpenShift/kubernetes
● Multi-master replication/data management for NAS use-cases

When
● RHGS 4.x
RED HAT GLUSTER STORAGE 4
EARLY PROJECTIONS

- When: 2018-19
- RHEL 7 only (no RHEL 6 support)
- Key Features
  - GlusterD2 (mgmt plane) - RESTful, etcd based config store
  - Dedupe + Compression
  - Plus-1 scaling - scale by 1 node, instead of 6, 12 etc.
  - Management (Native ansible for Day-1, Day-2)
  - NFS-Ganesha only (no Gluster-NFS)
  - Multi-site, active-active replication (tentative)
  - Rio (Tech Preview)

These capabilities are planned based on active development in upstream development communities, and will only become available once they reach the necessary level of maturity. Timelines are subject to change.
Container-Native Storage Roadmap
OPENShift Storage Needs

OCP Infrastructure
- Registry
- Metrics
- Logging

OCP Application
- Service 1
- Service 2

Local/Ephemeral Storage

Container-Native Storage Focus
RWX/RWO backed by File, Block, S3
CONSISTENT STORAGE ACROSS THE HYBRID CLOUD
APPLICATION PORTABILITY AND LOWER COSTS

RED HAT CONTAINER NATIVE STORAGE

BARE METAL | VIRTUAL | CONTAINERS | PUBLIC CLOUD | LEGACY

Does not matter where you deploy - now or in future, no change Flexibility to extend across cloud DR, Stretch, Separate Cluster
CLOUD NATIVE APPS

One Cluster
Storage like any other service

Value Proposition of Container & Kubernetes in one
Simplify Container Availability

AVAILABILITY ZONE A

AVAILABILITY ZONE B

AVAILABILITY ZONE C

AWS (AZ), On Prem, Hybrid Cloud, Multi-Cloud
Leverage legacy storage...container-ready!

Create a Seamless Fabric
CONTAINER PLATFORM - COMPLETE FROM Red Hat

UNIFIED CLUSTER | COMPLETE PLATFORM UNIFIED | INTEGRATED & TESTED
SINGLE VENDOR | SINGLE POINT OF SUPPORT

#redhat #rhsummit
RED HAT CONTAINER-NATIVE STORAGE

Be the de facto storage for OpenShift Platform

✓ Technology & Capability Leadership
✓ Operational & Management Simplicity
✓ Invisible but integral part of OpenShift
## What's New?

<table>
<thead>
<tr>
<th>OCP Application &amp; Infra Support</th>
<th>Technology Leadership</th>
</tr>
</thead>
<tbody>
<tr>
<td>✓ RWX (File) &amp; RWO (Block), S3</td>
<td>✓ Resize, Volume Metrics</td>
</tr>
<tr>
<td>✓ Registry, Logging, Metrics</td>
<td>✓ Custom Volume Naming</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Hybrid Deployment Options</th>
<th>Operational Simplicity with OCP</th>
</tr>
</thead>
<tbody>
<tr>
<td>✓ On Top or Decoupled from OCP</td>
<td>✓ Integrated Installation</td>
</tr>
<tr>
<td>✓ On-Prem, Cloud, Bare Metal, VM</td>
<td>✓ Integrated Mgmt &amp; Monitoring</td>
</tr>
</tbody>
</table>
Roadmap Focus

Technology Innovation
- Contribute & Enhance K8s Storage
- Block, Operator, Data Services

Ease of Consumption
- Storage as a service
- Service Level Objective Svcs

Enterprise SLA Enablement
- Arbiter Vol, SLA Driven RPO,RTO
- Multi-Cluster, Multi-Cloud

Ease of Operational & Mgmt
- One Control & Management Plane
- Integrated Monitoring
Container-Native Storage roadmap 2018-2019

CNS 3.10
June 2018
- Arbiter Volume Support
- Stateless Heketi
- Block Enhancements
- Raw Block*
- Service Catalog S3 Support

Fall 2018
- Manage using OCP Web Console
- Snapshot & Restore
- Stateful Set Support
- Raw Block & Local PV
- CSI Adapter

Early 2019
- Operator Managed CNS
- Monitor using Prometheus
- Storage as Service
- Multi-Cloud Support
- SLO Based Deploy, Manage
Red Hat Storage One
## WHAT IS RED HAT STORAGE ONE?

A hardware/software/support offering pre-configured for a target workload

<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>4-24 servers</td>
<td>pre-configured for a workload personality</td>
</tr>
<tr>
<td>30 minutes or less</td>
<td>to get up-and-running</td>
</tr>
<tr>
<td>Sold by Supermicro</td>
<td>or accredited reseller</td>
</tr>
<tr>
<td>Pre-loaded Red Hat® Gluster Storage®</td>
<td>and a workload-specific quick-deploy utility</td>
</tr>
<tr>
<td>Shipped and supported (L1/L2)</td>
<td>by Supermicro</td>
</tr>
</tbody>
</table>

#redhat #rhsummit
MAKES IT EASY TO STEP INTO SOFTWARE-DEFINED STORAGE

- Easy to design
- Easy to purchase
- Easy to install
- Single-vendor support
WHILE PROVIDING ENTERPRISE DATA MANAGEMENT FEATURES YOU EXPECT

Such as...

- Quotas
- Snapshots
- Geo-replication

#redhat #rhsummit
CURRENT WORKLOAD IDENTITIES

General NAS and content repositories

**General NAS**
User directories, mix of small and large files in NFS / GlusterFS-native folders

**Content repositories**
Photos, rich images, and videos at large scale
For RHHI

Red Hat Hyperconverged Infrastructure: Your open hyperconverged solution

When: Today at 4:30 PM
Where: 2003