What's new in Gluster 3.2
Webinar Q&A

Craig Carl craig@gluster.com **408-829-9953 (pst)**

What's the status of BRTFS support?

Although BRTFS has been steadily moving towards maturity, they have still not released a “stable release”, and many still consider it to be experimental. It is also not included with enterprise Linux distributions such as Red Hat. Gluster will work with BRTFS but should be considered very beta and not ready for prime time.

What about FUSE, FUSE modules, tuning and performance?

One of the ways Gluster gives back to the OSS community is by maintaining FUSE in the kernel. Csaba Henk is a Gluster engineer, he maintains FUSE while also actively developing the GlusterFS. For pre 2.4.15 kernels Gluster has a small patch, all it does is change some options that were not user-tunable in those kernel revisions. For any recent kernels Gluster uses the FUSE included in the distribution, there are no known issues with any version of FUSE at this time.

If every client is connected to every server is there a performance hit when having 1000's of clients connected to 1000's of Gluster servers?

When a client mounts a Gluster volume, it only opens connections to servers that are part of that volume. In an environment where there are 1000’s of Gluster servers, it would not make sense to put 1000 servers in a single volume. In the testing that Gluster has done, increasing the number of servers in a volume has been shown to increase performance to the client systems, and we have not observed a point where performance has decreased due to a larger number of servers in a volume.

Who do I call if I want 24/7, 2 hour response time, world class support for my Gluster cluster?

Me! I'll get you connected with the sales and support teams and even send you a fancy Gluster t-shirt! Maybe even a hat if the racoons haven't run off with them again. craig@gluster.com.

What are the Gluster limits? (come on, really, no BS.) (please)

Good question. Because we have NO centralized data store, no metadata, and we operate at the file level there really, really are no theoretical limits to the size, speed, inode count, file count, number of nodes, number of bricks, number of clients, etc, etc, etc.

In reality server and IP management becomes a problem at a certain point, running an `ls` on a file system with 265 billion (with a b) files would be...something silly. Our cluster with the most nodes in production has ~250 servers participating in a single volume. Our largest volume in production is 2.5PB after hardware RAID and 1.25PB usable after Gluster replication.

We have designed but not yet implemented a 1000 node, 100PB cluster. If you've got that kind of spare hardware laying around we could use to test with please let me know. I'll get you a Gluster t-shirt *and* a water bottle!

If I have two boxes with replication. Can I add two more boxes and convert it to replication with distributed?

When you expand a Gluster volume that is replicated, Gluster will automatically change the volume from being replicated to being distributed and replicated. *(it's magic!)*

In a HPC environment, how would you achieve parallelism and how would it compare to, say, GPFS?

When using the Gluster native client to connect to your Gluster volume, it automatically opens parallel connections to the storage nodes. The Gluster client also has the Elastic Hashing Algorithm built into it which allows clients to connect directly to the storage nodes which contain the needed data in parallel without the use of metadata lookups. Gluster has no metadata to maintain or cause bottlenecks. Gluster does not do any sort of write caching also, so there are no worries regarding cache coherency. (Gluster does do write-behind operations to improve write performance)

Is there a GUI interface for Gluster 3.2?

We will be releasing a GUI later this year for Gluster commercial customers.

Is compression an option on geo-replication?

Compression is not currently used, however it is going to be added soon.

After the first sync geo-replication only copies over the changed blocks, this makes replication very efficient.

What steps are taken when geo-replication passes the ability of the pipe to pass data immediately?

Indexing processes are separate from transfers, so the index will queue up transfer requests and those requests will be fulfilled when the remote site is back online. The size of the queue is only limited by the free space in your Gluster cluster.

Is Gluster 3.2 S3 and REST compatible?

Not currently, however these items are on our near term road-map.

Is Gluster and Nexenta working on any integration?

There is no formal Gluster ↔ Nexenta relationship but plenty of people use Gluster on top of multiple Nexenta clusters for better, bigger, faster distributed NAS!

Is replication master-master supported?

Normally Gluster replicated volumes are master-master volumes as all Gluster storage nodes are all considered master servers. Data is written synchronously to both servers that are part of the mirror pair.

How can I limit bandwidth consumption used for geo-replication?

There is some loopback style *nix tricks we can use, I'll work on getting some documentation put together, send me an email if you want me to send you the details. craig@gluster.com

What is the maximum number of storage nodes recommended?

The number of storage nodes used in an environment is determined by understand performance metrics and capacity needs. You can build an environment that has very few storage nodes with a lot of storage connected to them, but the performance will be slower because the network will eventually become the bottleneck. For example, if you need 100TB of storage and 1GB/s worth of throughput, you could not build this using 2 servers with 50TB of storage each that have 1GbE network connectivity because your throughput to that 100TB of storage would only be 200MB/s. You could build this environment using 10 servers with 10TB of storage and 1GbE.

What's the fastest, easiest to deploy and manage, most cost effective distributed file system in the world?

GLUSTER! GLUSTER! GLUSTER! GLUSTER! GLUSTER! GLUSTER! GLUSTER! GLUSTER!

Are you always this subtle?

Yes. I'll grow on you.

Is there an API to Gluster?

We will be releasing a RESTful API soon. For now, all Gluster functionality is available via our fancy CLI, `gluster` :)

Is WAN replication read-write?

Currently, the geo-replication feature for Gluster is for disaster recovery so the remote end should be considered read-only. Currently the Gluster geo-replication solution is one-way, look for active-active asynchronous replication coming soon for commercial Gluster customers.

Is GlusterFS capable of Box-Level RAID?

Basically no. Gluster can protect against the loss of a server by replicating data across multiple systems. In this scenario, you can lose a server and not lose access to any data. It is still recommended to run RAID under Gluster to protect against the loss of a disk. Parity protection is not currently supported, so the only way to achieve data redundancy today is using full replication.

What about NFS locking?

We don't have support for NFS locking, we suggest using the Gluster native client if you need locking.

Do you have any 32-bit support?

Changes in the 3.1 code base required moving to a 64-bit architecture for both the Gluster client and server. There is no 32-bit versions available.

Do you have support for RDMA?

Sure do! Details here - http://gluster.com/community/documentation/index.php/Gluster_3.2:_Configuring_for_InfiniBand

Who do I call if I want 24/7, 2 hour response time, world class support for my Gluster cluster?

Me! I'll get you connected with the sales and support teams and even send you a fancy Gluster t-shirt! Maybe even a hat if the racoons haven't run off with them again. craig@gluster.com (deja vu much?)

What's the easiest way to test Gluster?

Gluster offers solutions for both physical and virtual environments. For physical machines, you can download the GlusterFS OSS or the Gluster Software Storage Appliance. Virtual appliances are available for AWS, Vmware, Xen and KVM. The easiest way to get started is probably the Amazon AMI, you can get a Gluster cluster up and running in 15 minutes, easy. More info -

http://www.gluster.com/trybuy/

http://www.gluster.org/download/

http://www.gluster.com/gluster-for-aws/

When a node within the cluster fails, what are the features of Gluster that handle this scenario? And what happens with data as well?

Assuming you have set your cluster up with replication then there is no impact at all. Your data is always on two servers so if one fails operations continue as normal. If you are running in a non-redundant configuration then any data on the failed node is unavailable until that node comes back up. Access to the oter nodes of the cluster continues normally.

How the return of a failed node within the cluster is managed?

For replicated volumes, GlusterFS has a self heal feature that will find any changes and repair as necessary. For pure distribute, files will become available again once the node has come back online.

Any plans for supporting NFS4 for clients?

We're working on it! Follow us on Twitter @Gluster for regular updates.

Are the GlusterFS native clients for Mac or Windows (or other OSes)?

No, we only have a native client for *nix.

We use Windows, what are you going to do for us! (huh!)

Gluster uses Samba and CTDB, an awesome project from the Samba team to provide highly available CIFS access with stateful fail over in case of a node failure! Just because CIFS kinda sucks it is slower than GlusterFS and NFS. The best we've ever seen using CIFS is 50MB/sec on a 1Gb interface, and as slow as 5MB/sec for some worloads. :(

Are striped volumes mentioned in documentation RAID0?

Yes, Documentation is available for striped volumes in GlusterFS 3.1.x as well as 3.2.x. A quick note about striping – you probably don't want to do it. Improvements in the Gluster distribute translator mean that stripe is no faster than distribute and

there is no way to be redundant and stripe at the same time.

http://gluster.com/community/documentation/index.php/Gluster_3.2:_Configuring_Distributed_Striped_Volumes

http://gluster.com/community/documentation/index.php/Gluster_3.1:_Configuring_Distributed_Striped_Volumes

Does Gluster support RHEL 6 and ext4?

Ext4 support is 100%, it is our default file system. Ext3 and XFS are also supported. RHEL 6 should work fine but we won't add it to our QA cycle until CentOS 6 is available.

Should I have a backend network for Gluster?

If you are using the Gluster native client then no, there is no need. If you are using NFS or CIFS clients then a backend network can help if you are network bound.

How can I detect slow performing nodes in the cluster?

Gluster has two great monitoring tools, `gluster top` and `gluster profile`. Details - http://gluster.com/community/documentation/index.php/Gluster_3.2:_Understanding_your_GlusterFS_Workload

Does Gluster support Windows or NFSv4 style ACLs?

So soon! Support for Windows ACLs is expected in June 2011, NFSv4 ACL support will be right behind that!

How easily are upgrades between Gluster versions? Can upgrades be done live?

Upgrades between Gluster versions easy! Upgrades can be done with 0 downtime for replica volumes and just a few seconds of downtime if your cluster isn't redundant.

With replicas are reads spread across both copies?

Yes. Read performance is significantly improved for replica volumes.

What about some kind of network raid 5 with distributed redundancy over multiple nodes?

No. We're working on it using erasure coding but I don't have a release date yet.

Can you talk a bit more about using RAID under GlusterFS? Do you recommend this primarily for performance? What type of RAID do you recommend?

Imagine what happens without RAID under Gluster. When a 2TB drive in a server fails then gets replaced we have to copy 2TB of data across the IP network to the replica. Slow, slow, slow and while the copy is happening all of your applications slow down too. RAID is cheap. If you have to you can use LVM or mdadm, but hardware RAID is best. Unless you need screaming performance and are using IB then RAID 5 or 6 is more that fast enough.

Do you have anyone using this as a backup target for RMAN (Oracle backups) and replicating those backups/logs?

Absolutely. Backup-to-disk is a great use case for Gluster, lots of groups use us for that. We work well with RMAN, Netbackup, Commvault, etc.

Can you change the replica count of the cluster without deleting and recreating the volume?

No, but deleting and recreating the volume doesn't require moving any data around and can be done in less than 30 seconds if you're a better typist than I am. (not hard)

Do you have any customers using GlusterFS on production XENserver or the XEN Cloud platform ?

Boy do we! Gluster won BEST IN SHOW at the Citrix Synergy conference in May 2011. You would already know that if you were following us on Twitter @Gluster!!

This space intentionally left blank so you could scribble notes about all the ways Gluster is going to make your life easier.
I know I didn't really leave enough space for that but please, think of the pretty trees.