Backups Using Storage Clusters

Joshua T. A. Davies
Garrett W. Ransom
Nicole M. Shaw

Mentors: David Kennel, Sonny Rosemond, Cindy Valdez, Timothy Hemphill
(DCS-CSD)
Overview

- The Project
- The Cluster
- Software
- Issues
- Conclusions
- Future Work

―"We back up our data on sticky notes because sticky notes never crash."

Introduction

• Los Alamos National Laboratories generates petabytes of data
• Estimates for the unclassified network suggest the amount of data needing backup may easily exceed 2.5 PB
• The options available now are non-ideal
  – Traditional tapes may be too slow to restore from in the event of a large scale disaster
  – The amount of data exceeds the capabilities of most commercial solutions
  – Disk based storage tends to be prohibitively expensive
The Project

- Goal – construct and test a new design of commodity storage cluster
- Consisted of two tiers and a single control (head) node
  - Head Node: ownCloud server and tier management
  - Tier 1: Primary ownCloud Storage
  - Tier 2: Subdivided into two groups, each serving as a redundant copy of Tier 1
The Cluster

- 11 nodes
  - One head node
  - Ten compute nodes divided into two tiers

- Centos 6.5 Operating System
- Warewulf Administration
  - Stateless nodes
- IPMI
ownCloud

- Open source cloud server
- Can upload via desktop client app or web interface
- Server configuration installed on the head node
- Version 6.0.4-8.1
Gluster

- Open source distributed file system
- Version 3.5.1
- Aggregates node storage into single volumes
- Makes use of geo-replication feature
  - copies data between different volumes
Node Control and Tier

- Node control (nodectl) gives access to individual nodes
- Provides information on power state, tier membership, Gluster volume name
- Toggles power state

- Tier script controls each tier as a unit
- Brings tiers up (nodes must be on): creates Gluster volume, mounts as needed
- Synchronizes Tier 1 with given Tier 2 by starting geo-replication
- Readies tiers for safe shutdown
Switch

Tier 1

Old geo-replication session

Tier 2A

Power Switch

Tier 2B

New geo-replication session
Restore

- Halts geo-replication with active Tier 2 volume, and powers down nodes.
- Powers on initially inactive Tier 2 nodes.
- Creates Gluster volume on newly booted Tier 2 nodes.
- Starts geo-replication from Tier 2 to Tier 1
- Waits for separate command to stop replication, shut down nodes, and resume normal behavior
Issues

• Original file permissions were not preserved by ownCloud
  – ownCloud uses a global mask that will set all permissions to a default
  – At present, the preservation of such permissions does not seem to be a supported feature
Issues

• Discovered an ownCloud corruption issue occurring with files of sizes 2GB or greater
  – We confirmed this by comparing hex dumps of the original file and the downloaded file. The differences began at the 0x7fffffff byte of the file, which defines the 2GB limit.
  – This corruption was confirmed to appear across Mac, Linux and Windows clients
Conclusions

• The system showed promise in its basic functionality
  – Providing service to clients of varying operating systems
  – Storing data into GlusterFS volumes, aggregated across nodes
  – Utilizing geo-replication to duplicate data between tiers
  – Conducting automated tier switches

• The issues of file permissions and corrupted files makes this prototype unreliable until ownCloud bugs are addressed
Future Work

• Collaborate with ownCloud developers to fix the current file permissions and corruption issues
• Investigate the scalability of both ownCloud and GlusterFS
• Test the use of multiple ownCloud servers, handling large numbers of clients
• Test whether Gluster can support the use of Infiniband interconnects for geo-replication
Summary

• Measures need to be in place to prevent data loss and provide a means of recovery from large scale failures

• Our project focused on a new design for a storage cluster system integrating ownCloud and GlusterFS to provide reliable and low cost backup services

• Overall, the prototype showed promise, yet file permission and corruption issues prevent the use of the design in its current state
Special Thanks

Instructor: Dane Gardner
TA: Christopher Moore

Mentors: David Kennel, Sonny Rosemond, Cindy Valdez, Timothy Hemphill

Josephine Olivas
Carol Hogsett
Carolyn Connor
QUESTIONS?