A Highly Available Cluster Storage System Using Scavenging

High Availability and Performance Computing Workshop (HAPCW'04), held with 2004 Los Alamos Computer Science Institute Symposium (LACSI'2004)

Xubin (Ben) He, Li Ou
{hexb, lou21}@tntech.edu
Storage Technology & Architecture Research (STAR) Lab
Electrical and Computer Engineering Department

Stephen L. Scott, Christian Engelmann
{scottsl, engelmanncc}@ornl.gov
Computer Science and Mathematics Division

Outline

Introduction
Highly Available Metadata Management
Metadata Management using Bloom Filters
Conclusions and Future Work
Motivations

- Data intensive scientific applications → high performance computing.
- Distributed storage systems: high performance storage, wide area mass storage, cluster storage.
- Excellent performance, parallel support ↔ administration costs, central points of failure and control

Existing work:

- HPSS
- Peer-to-Peer Storage
- Scavenging [Vazhkudai]
Introduction

✓ Highly Available Metadata Management
   Metadata Management using Bloom Filters
   Conclusions and Future Work

User Data vs. Metadata
Distributed Metadata Management

Metadata Management Layer

Scavenge Manager Unified Interface

Client Access

Cluster Storage Management

Scavenged Storage Morsels

High Availability: Scheme #1 (Pure P2P)

Peer-To-Peer Network

Metadata

Manager
High Availability: Scheme #2 (active/hot-standby)

Peer-To-Peer Network

Metadata    Active Manager    Hot-Standby Manager

High Availability: Scheme #3 (Partitioning)

Master/Load Redirector

Network

Metadata    Manager
High Availability: Scheme #4 (leader/helper)

Group Communications

- Peer-to-Peer distributed control
- Reliable broadcast, Atomic Broadcast
- Atomic transactions guarantee metadata integrity
Bloom Filter

- A Bloom filter is a fast and efficient method for representing a set $A = \{a_1, a_2, \ldots, a_n\}$ of $n$ elements to support membership queries.
Metadata Management using Bloom Filters

File queries

- Scavenge Manager Unified Interface
  - Hashing and LRU Cache
    - Bloom Filter Array
      - Multicast the queries
        - hit
          - Update
          - hit
          - miss
          - Update

Introduction

Highly Available Metadata Management

Metadata Management using Bloom Filters

✓ Conclusions and Future Work
Conclusions

- Investigate Availability issues Scavenged Storage Systems, propose 4 solutions for maintaining multiple metadata managers:
  - P2P
  - Active/hot-standby
  - Partitioning
  - Leader/Helper

- Speed up the metadata searching:
  - Bloom Filters

Future Work

- Comparing the proposed 4 schemes
- Scalability
- Metadata Cache
Acknowledgements

• Research Office and Center for Manufacturing Research, Tennessee Technological University

• Ralph E. Powe Junior Faculty Enhancement Award by Oak Ridge Associated Universities (ORAU).

A Highly Available Cluster Storage System Using Scavenging

High Availability and Performance Computing Workshop (HAPCW’04), held with 2004 Los Alamos Computer Science Institute Symposium (LACSI’2004)

Xubin (Ben) He, Li Ou
{hexb, lou21}@tntech.edu

Storage Technology & Architecture
Research (STAR) Lab
Electrical and Computer Engineering
Department

Stephen Scott, Christian Engelmann
{scottsl, engelmannc}@ornl.gov

Computer Science and Mathematics Division

Tennessee Tech UNIVERSITY

Oak Ridge National Laboratory
Computer Science Group

NOW Handout Page 11