SDC
STORAGE DEVELOPER CONFERENCE
SNIA SANTA CLARA, 2016

Efficient Data Tiering in GlusterFS

RAFI

Original authors:
Joseph Fernandes
Dan Lambright

About me

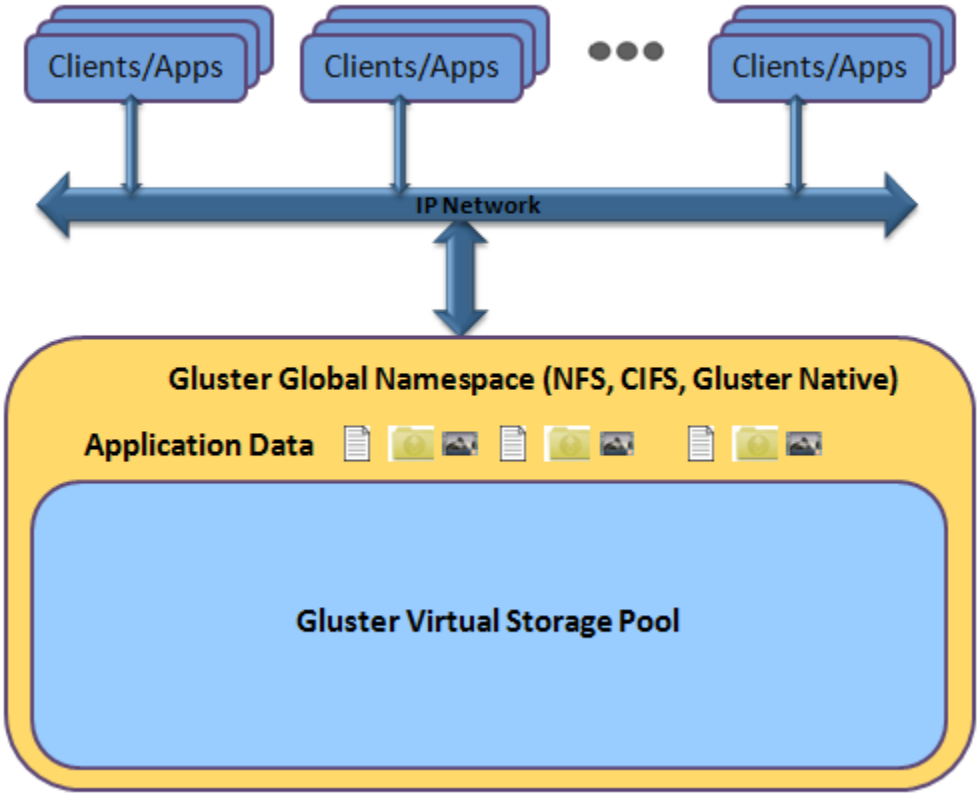
Rafi KC (Software Engineer, Red Hat)

- Gluster-rdma, Gluster-snapshot, Gluster-tiering.

Agenda

- Quick GlusterFS Overview
- Data Tiering
- GlusterFS - Data Tiering
- Detailed Implementation
- Interesting problems solved
- Indexing techniques in other Tier cache

What is GlusterFS



Distributed File System

Software Define NAS

TCP/IP or RDMA

Native Client, SMB, NFS

Automated Data Tiering

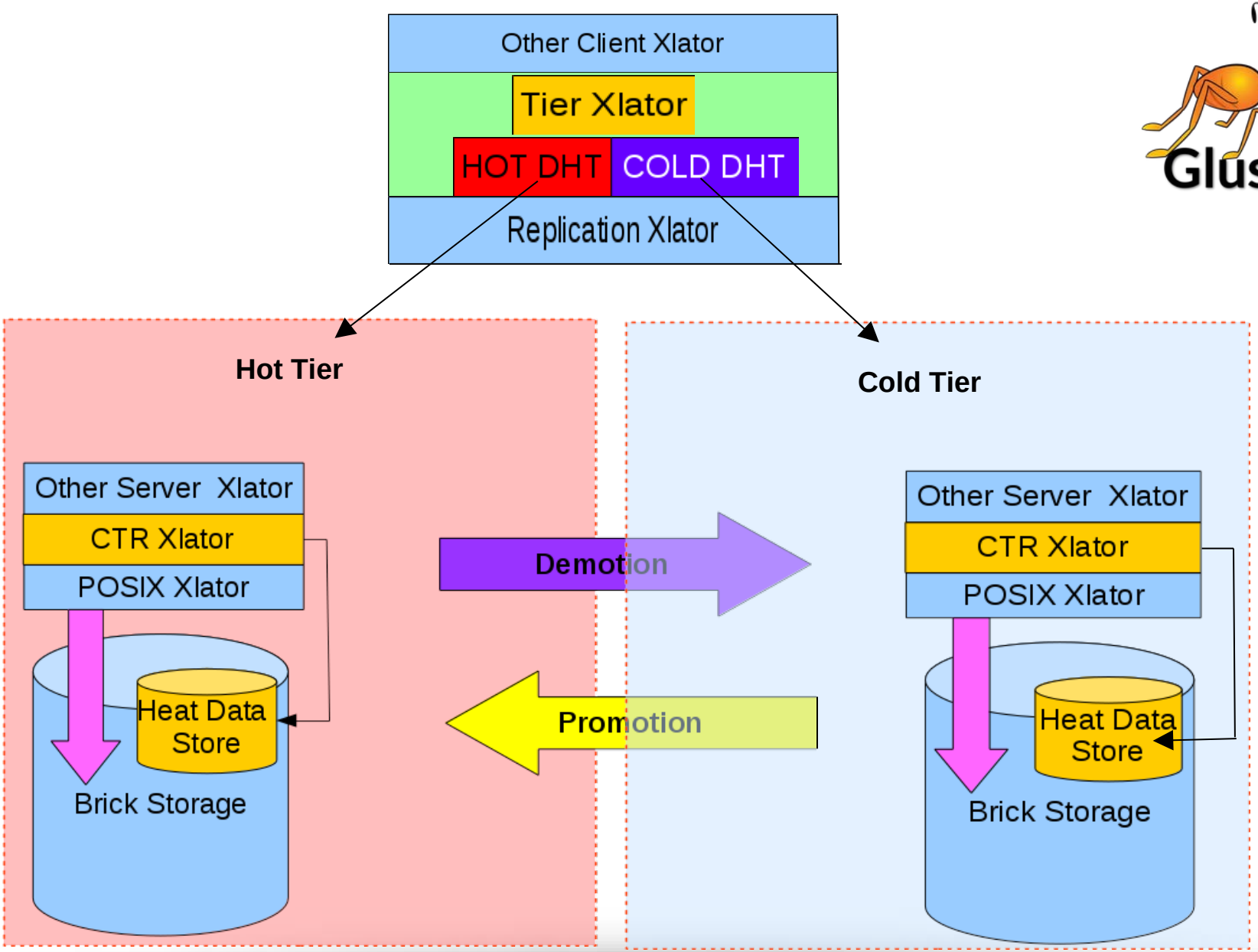
- A logical volume composed of diverse storage units
  - Fast / slow
  - Secure / nonsecure
  - Expired hold time / expired
  - compressed / uncompressed,
  - Cloud expensive elastic storage / cheap
  - Etc.
- Data moves automatically across tieres
- Efficient use of different storage tieres

GlusterFS – Data Tiering

- Two Tiers
  - HOT and COLD
  - Fa$t SSD, slow HDD
- Fast 2X replicated, slow erasure coded
- Files will be moved across HOT and COLD tiers

Policies for Smart Migration

- File size
- Access rate
- Migration frequency
- Water mark
- Break files into chunks
  - Gluster “sharding” feature

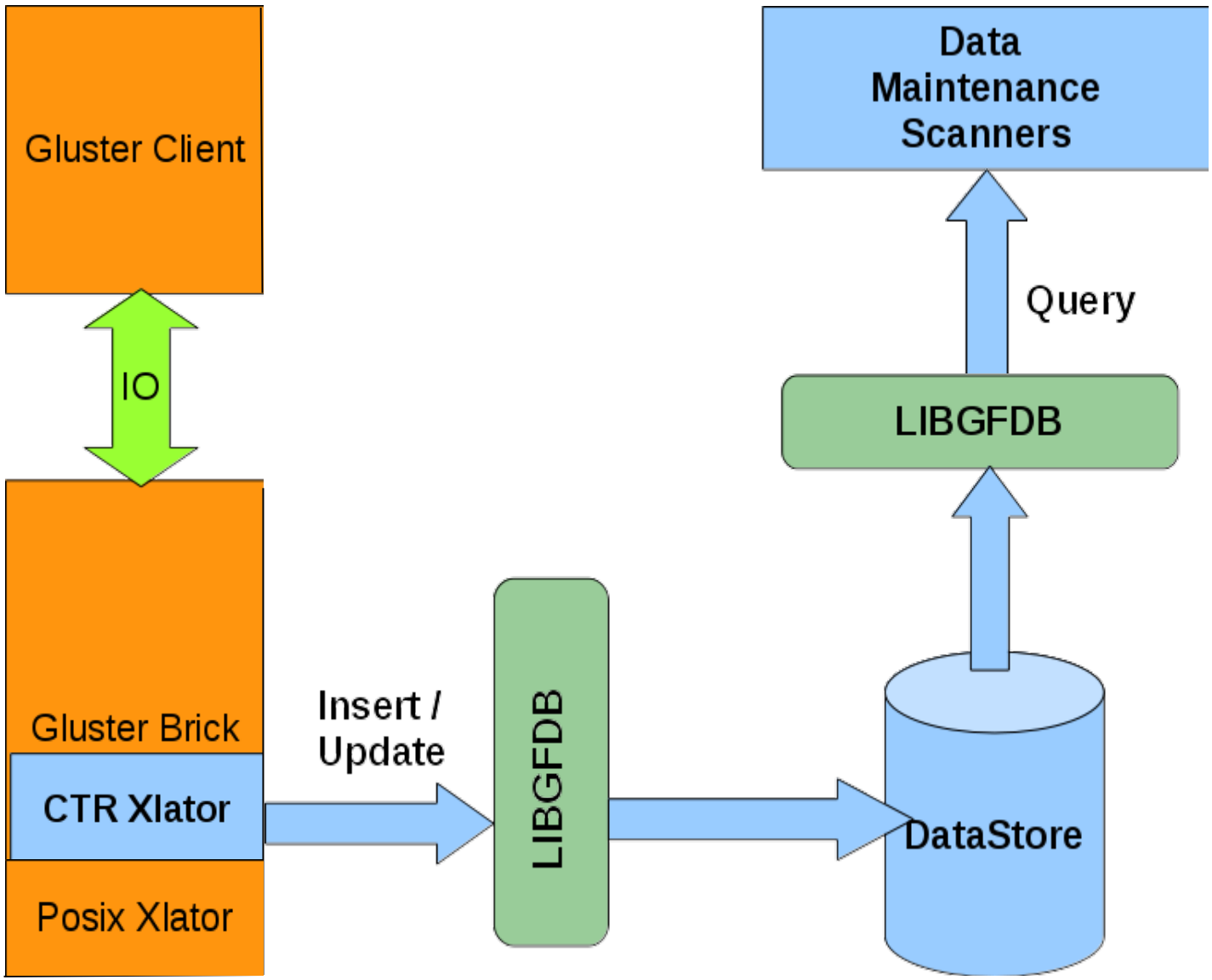


Challenges in Data maintenance

Data Maintenance has a overhead on CPU, Memory, Storage, Network.. Therefore..

Implementation

- Change Time Recorder
- Libgfdb
- Migration Daemon



LibgfDB

- API Abstraction
- Rich Search Filters
- Performance optimization options

Optimized DB for GlusterFS

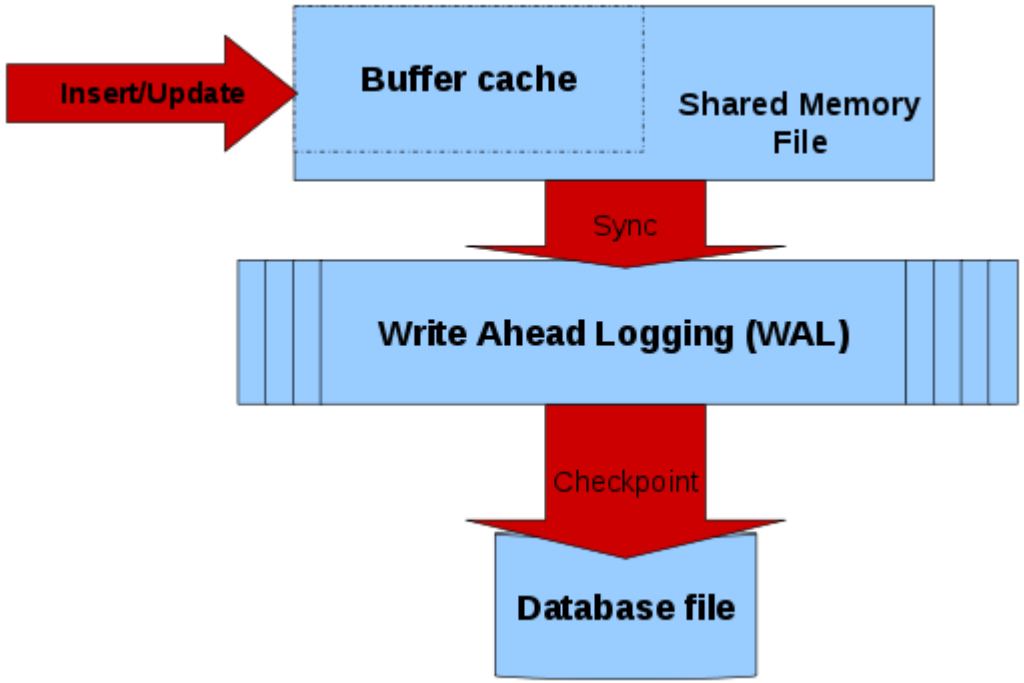
" Record now , consume later"

- Database optimized to record fast
- Good Querying Capabilities
- Embedded Database

Datastore Optimization: Sqlite3

- PRAGMA page_size: Align page size
- PRAGMA cache_size: Increased cache size
- PRAGMA journal_mode: Change to WAL
- PRAGMA wal_autocheckpoint : Less often autocheck
- PRAGMA auto_vacuum : Set to NONE

DataStore Optimization: Sqlite3



Problems we solved

- DB updates can be expensive
- DB query may have scalability problems
- Durability (ACID semantics) is expensive

Indexes for Tiered Cache

	Advantages	Disadvantages
Hash (DM cache)	Predictable	Grow/shrink
Database (Gluster)	Easier implementation, Flexible, Rich Metadata, Precision	Opaque, space inefficient, Performance optimizations
Bloom Filter (Ceph)	Space efficient	No metadata. Collisions, Counters

Credit : Dan Lambright

Resources

- Feature Page
http://blog.gluster.org/2016/03/automated-tiering-in-gluster/

- Gluster Github:
https://github.com/gluster/glusterfs

THANK YOU

Q&A