YourFileSystem

Announcing U.S. Dept of Energy SBIR Grant Supporting Development of Next Generation OpenAFS

Jeffrey Altman, President

Your File System Inc.

26 October 2009

OpenAFS Roadmap? Or Wish List?

- At every Workshop and Conference a roadmap is presented
  - but its not a roadmap
  - No commitments
  - No delivery dates
  - How are you supposed to plan your rollout schedule?
- The problem is lack of resources
- The Gatekeepers and Elders compile a list of requests but have little influence on what people work on

HEPiX Fall 2007 was a Wake Up Call

- OpenAFS was unable to provide:
  - Commitments
  - A Delivery Schedule
  - A list of what was being worked on
- The message was received loud and clear:
  - Do not ask others to help you until you can prove that you can help yourselves

YFS Inc. Founded to Drive Demand Globally Accessible File Systems

- MobileMe, BigVault, and similar sync and access cloud storage services are far behind the capabilities that AFS provides for real-time access and collaboration.
- YFS will provide services and appliances direct to home, small business, and enterprise users and indirectly through telecommunication companies.
- With hundreds of millions of users, scalability and performance are key requirements.

The YFS Mission

- Develop, deploy, and operate “Write once, Access Manywhere™” global storage solutions
- Support the on-going development of critical path standards and open source technologies

U.S. Department of Energy Small Business Innovative Research Grant

- The U.S. DoE labs are large users of AFS:
  - support their HEPiX research
  - provide global access to home and project data
  - distribute and manage applications
- YFS Inc. applied for a grant in 2007
- In 2008, received $99,000 to fund a feasibility study
- In August 2009, awarded US$648,000 to design, standardize, and implement core protocol enhancements

YFS Requirements of AFS

- Server scalability (>60,000 clients per server vs ~1000)
- Networking
  - 10GBit networks
  - IPv6
  - TCP and/or SCTP in addition to UDP communications
- Network Traffic Optimization
  - Reducd cache trashing
  - Enhanced file change notification protocol
  - Server based virtual query volumes
- Lazy Read/write replication

YFS Requirements of AFS

- Directory improvements
  - Internationalization, Extended Attributes, Multiple Data Streams per Object
- Mandatory and byte range locks
- End-to-end Security
  - AES-256 encryption
  - Kerberos, X.509 certificates and SCRAM for authentication
  - Per Service Keys
  - Anonymous Client Access is Protected
  - Secure Callback Channels

YFS SBIR Phase I Success

- The feasibility of the project was demonstrated
- In the process, several contributions to OpenAFS were delivered
  - Documentation of existing AFS3 protocols
  - Several Rx Transport Implementation Errors were corrected
  - OpenAFS 1.4.8 provides a 9.5% improvement in throughput compared to 1.4.7

YFS Phase II First Year Road Map (August 2010 deliverables)

- Rx TCP Transport
- Rx UDP Improvements
  - Window Size Negotiation*
  - Dynamic Retransmit Calculation*
  - Path MTU Discovery
  - Large Data Buffers
  - Improved Jumbograms
  - Max Call Negotiation
- Asynchronous Rx API
- RxGK Security Class
- Protection Service
  - Machine Accounts
  - Aliases
- Client Improvements
  - Byte Range Locking
  - Direct and Synchronous I/O
  - Demand Prefetching
- Ubik enhancements

YFS Phase II Second Year Road Map (August 2011 Deliverables)

- Server Improvements
  - Event driven workflow
  - Posix Ext. Attr. Backend to replace namei
  - Service Port Independence*
  - Split Horizon Support
  - Volume Release Optimizations
  - Lazy Read Write Replication
- IPv6 Support
- Partition UUIDs
- Long Volume Names
- Metadata Improvements
  - Unicode
  - Extended Attributes
  - Alternate Data Streams
  - DOS Names/Attributes
  - Per File ACLs
- Removal of Directory Size Limitations

File System Comparison

CRITERIA	Volume Management	Filesystem snapshots	POSIX Extended Attributes	Transport	Scalability	Performance
OPENAFS	Yes	Limited	No	UDP IPv4	Yes	Moderate
OPENAFS NOTES	Transparent movement of data.	Typically one “backup”.		TCP support planned.	Thousands of clients per server in practice.	No parallel access today. Limited by transport.
LUSTRE	No	No	Yes	TCP IPv4	Yes	High
LUSTRE NOTES	Online data migration planned.	Planned for 3.0.			30000 clients per node.	Optimized; Uses object-based storage.
NFS V4	Extension	No	Yes	TCP	Yes	Varies
NFS V4 NOTES	Not always available			IPv6 not widely available.		pNFS extension, TCP allow good performance.
ZFS	Yes	Yes	Yes	N/A	N/A	High
ZFS NOTES				Local only.		Uses mirroring and striping to achieve high bandwidth.
YFS	Yes	Limited	No	UDP, TCP; IPv4, IPv6	Yes	High
YFS NOTES	Striping; Q3 2011	More than OpenAFS; Q3 2010	Q3 2011	TCP Q3 2010; IPv6 Q3 2011	Asynchronous threading model; 60,000 clients / server Q3 2010	Transport, threading, OSD; Q3 2010-11

File System Comparison (cont.)

CRITERIA	Locking	Replication	Object Storage Integration	Security	Authentication	Open Source	Commercial Support
OPENAFS	Advisory	Read-Only	No	Yes	Yes	Yes	Yes
OPENAFS NOTES	Whole file only.	Read-Write planned.	Integration to begin soon.	56 bit fcrypt. K5crypto, 2010	Kerberos 4 and Kerberos 5.	IBM Public License V1.0.	Linux Box Secure Endpoints Sine Nomine Associates
LUSTRE	Yes	Local	Yes	No	No	Yes	Yes
LUSTRE NOTES	No lockf /flock yet.	RAID, not multi-server yet.	That's largely the point!	Planned for 1.8.	Kerberos support in Lustre 1.8	GPL.	ClusterFS (now Sun).
NFS V4	Yes	Extension	Extension	Yes	Yes	Available	Yes
NFS V4 NOTES	Mandatory and Advisory.	Not widely available.	In pNFS/NFS v4.1.	GSSAPI RPC.	GSSAPI / Kerberos 5.	Citi reference implementation is GPL.	Typically from OS vendor.
ZFS	Yes	Manual	Extension	N/A	N/A	Available	Yes
ZFS NOTES	Mandatory and Advisory.	Using zfs send/receive.	Block-based ZFS.				Typically from OS vendor.
YFS	Yes	Read-Write & Read-Only	No	Yes	Yes	Yes	Yes
YFS NOTES	Q3 2011	Q3 2011	Q3 2011	RFC3961, Q3 2010	GSSAPI / Kerberos 5, X.509; Q3 2010	IBM Public License V1.0 + BSD	YFS

Open Source is a Commitment

- Open Design
- Open Standardization
- Open Implementation
- Open Contributions

- Public git and gerrit instances will be provided

- All externally funded projects will be contributed to OpenAFS upon completion under a BSD license.

There Remains Much to be Done

- User Experience Improvements
  - File System adoption is driven by demand
  - Demand is determined by usability
  - Usability is the result of user experience improvements and human computer interaction design
- Hadoop Map-Reduce Service for AFS
- AFS – ZFS integration
- Indexing Services
- Cloud Services APIs
- Automated Load Balancing

Contact Info

- Jeffrey Altman
- President
- Your File System Inc.
- jaltman@your-file-system.com
- +1 212 769-9018