Toward a petabyte-scale AFS service at CERN

Dan van der Ster, Jakub T. Mościcki, Arne Wiebalck

CERN IT-DSS-FDO

Abstract

AFS is a mature and reliable storage service at CERN, having worked for more than 20 years as the provider of Unix home directories and project areas. Recently, the AFS service has been growing at unprecedented rates (200% in the past year); this growth was unlocked thanks to innovations in both the hardware and software components of our file servers.

This work will present how AFS is used at CERN and how the service offering is evolving with the increasing storage needs of its local and remote user communities. In particular, we will demonstrate the usage patterns for home directories, workspaces and project spaces, as well as show the daily work which is required to rebalance data and maintaining stability and performance. Finally, we will highlight some recent changes and optimisations made to the AFS Service, thereby revealing how AFS can possibly operate at all while being subjected to frequent—almost DDOS-like—attacks from its users.

Status of the Service

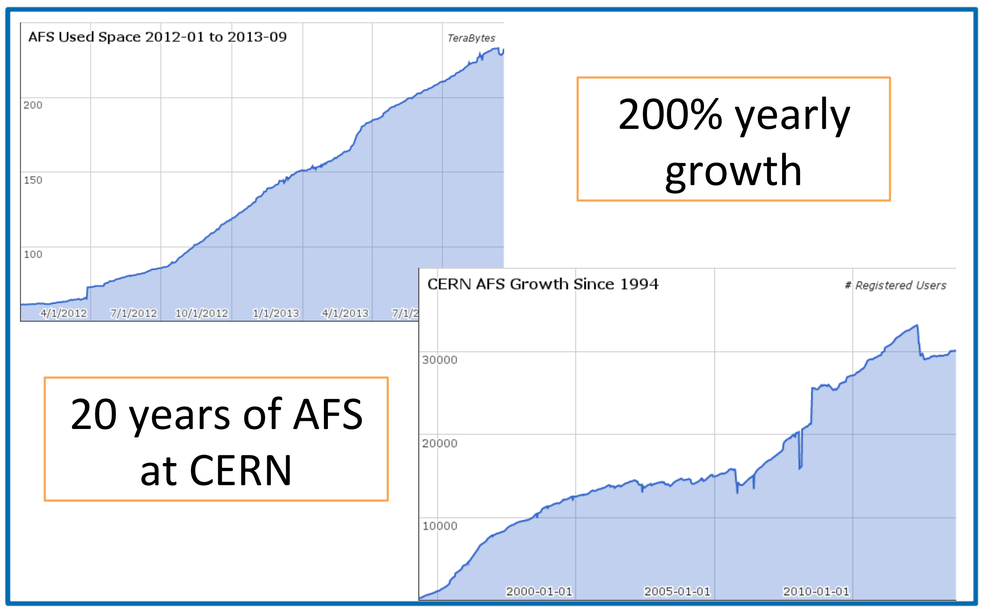
Table 1. Status of the AFS Volumes at CERN, 18 September 2013

Type	# Volumes	Quota (TiB)	Used (TiB)	# Files ($*10^6$)	Daily Accesses ($*10^6$)
Home Directories	30156	47	14	199	832
Workspaces	4627	303	113	402	2504
Project Volumes	33773	182	99	1282	3108
Archived Homes	14872	3.6	0.88	21	0
TOTAL	83428	538	228	1904	6445

Table 2. Local and Remote Accesses to /afs/cern.ch/

Date	CERN Clients	Remote Clients	TOTAL
2012	16823	19008	35830
2013 (–Sept)	16344	16979	33322
Week of 19 Sept 2013	9673	4238	13910

Growth of AFS at CERN



General Improvements to the Service

AFS Volume Management
- Transparent movement of volumes between servers to balance disk space and I/Os
- Automatic management tool was improved with increased parallelism
- Moves up to 250 volumes per hour

AFS Backup System
- At CERN we backup all data in AFS to tape for 6 months
- Growth of the service beyond 200TB was putting pressure on the 10 year old AFS backup system
- AFS Backup Service Completely rewritten to handle multiple TSM backends as well as CASTOR

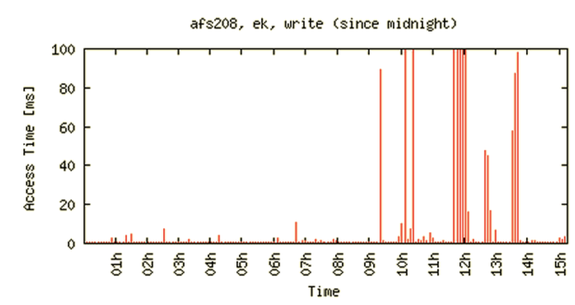
New Large Disk Servers
- Historically AFS requires fast expensive disk servers with multipath SAS hardware
- Performance improvements (see right) enabled us to move to lower cost hardware which enabled further data growth

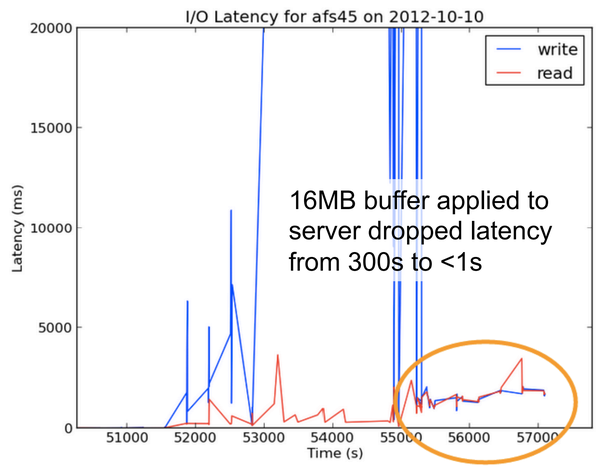
Solving the AFS Latency Issues

In 2012:
- 100GB workspaces becoming popular
- Periodic high latency incidents were becoming more and more common

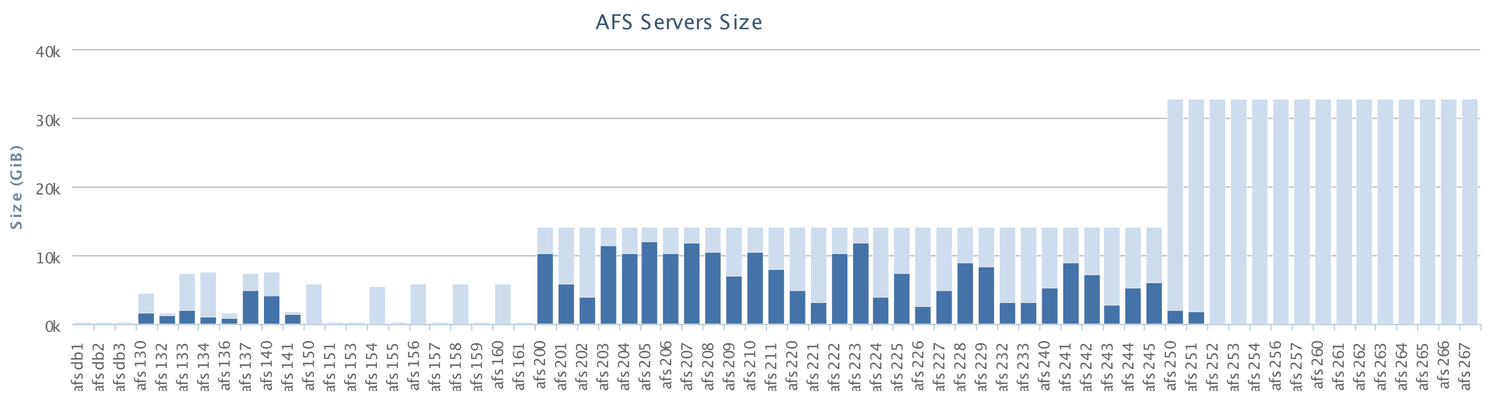
Investigating the Problem:
- LSF queue used to hammer a testing fileserver
- rxperf used to test the UDP RX protocol directly

Solution:
- UDP Buffer Size of 2MB found to be too small for AFS fileservers, leading to significant packet loss
- 16MB buffer deployed in late 2012
- Latency incidents **no longer occur**





All Deployed CERN AFS FileServers as of Mid-September 2013



Rollout of new large servers resulted in 1 petabyte available space in the AFS service

Conclusions and Future Plans

With the aforementioned improvements in place, the AFS service has grown by more than 200% in the past year and used space is expected to continue to double yearly.

The figure at the left shows the total available space in AFS at CERN – with the newly deployed 32 terabyte servers we have achieved a total 1 petabyte capacity.

Future Plans:
- Deploy OpenAFS 1.6 (performance improvements).
- Puppetize the service, run more fileservers to balance the load
- Cloudify the service: virtual AFS fileservers with Ceph-backed storage