SCALING AFS AT CERN

Arne Wiebalck, IT-DSS CHEP 2012, New York

Based on: OpenAFS

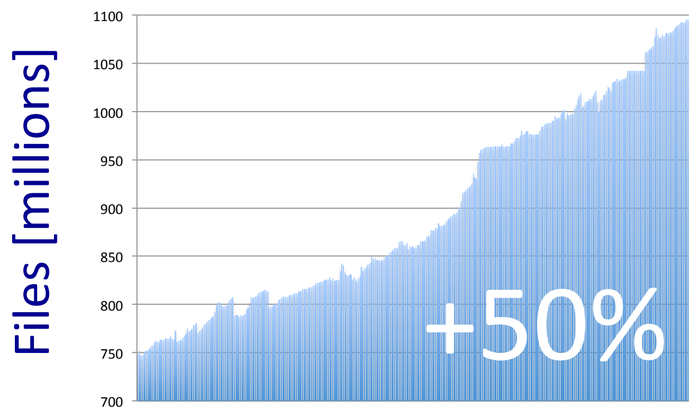
CERN

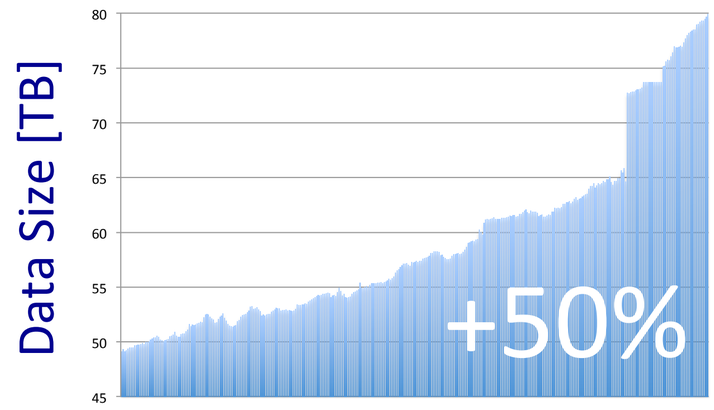
What we provide

30’000 accesses per second

- 32’000+ home directories
- 300+ project spaces
- 1.1 billion files
- 80TB of data
- 10’000 CERN clients
- 5’000 off-site clients

AFS service growth in the past year





What users demand

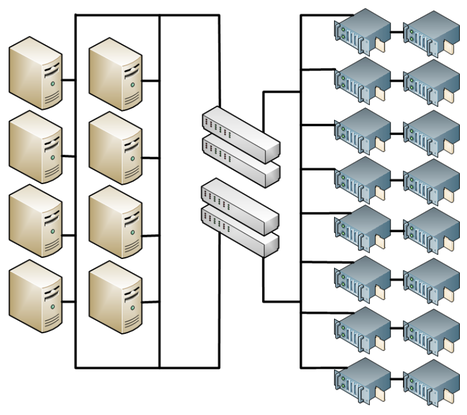
More …

- Personal space (tens of GB/user)
- I/O operations (100s/sec/user)

Features …

- High availability
- High reliability
- Daily backup
- Quota self-service
- Service simplification
- Multiple client support

What we have used so far



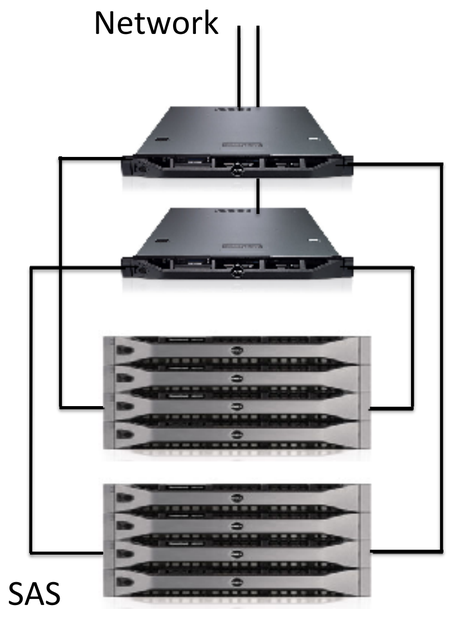
Fibre Channel Fabrics
~40 servers, ~80 disk arrays
300GB 15k SAS drives

+ Reliable setup
+ Hardware interventions do work
+ Sufficient performance in the past

- Non-standard CERN h/w
- Prone to finger trouble
- Very high price/GB

→ Prevented easy scaling

New approach: SAS Storage Units



2 servers, 2 SAS arrays form unit
All disks visible on both servers
32x 2TB 7.2k NL-SAS drives
8x 256 OCZ Vertex II SSDs

+ Lower price/GB than FC setup
+ No Single Point of Failure
+ Fast data fail-over
+ Limits the impact of failures
+ SSDs compensate for larger disks
+ Streamlined hardware

+

facebook ’s Flashcache

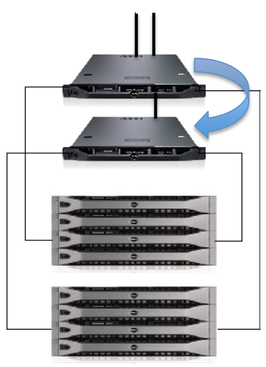
General purpose SSD block cache for the Linux kernel

developed for MySQL
integrated into device mapper
supports writeback/-through/-around
can give significant speed-up
patched to compensate for SSD drop-outs
in production at CERN since Q1/2011

https://github.com/facebook/flashcache

Option: Virtual AFS servers

Based on: KVM



Flip from physical to virtual server within a pair

+ Can increase availability (quick fail-over)
+ Can help to isolate hot data
+ Can help to guarantee service levels
+ Can help to overcome AFS limits

} Will help with further scaling

What the users get

10x more space
More I/O
Better QoS

Arne.Wiebalck@CERN.ch http://cern.ch/afs