DSS

Data & Storage Services

CERN IT Department

Huawei Cloud Storage

Seppo S. Heikkila
Maitane Zotes Resines
CERN IT

Openlab Major Review Meeting
26th of September 2013
CERN, Geneva

CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

HUAWEI

CERN openlab

Outline

- Timeline
- Huawei and benchmark setup
- Past phase results
- New S3 library Davix
- Front-end activity monitoring
- Testing infrastructure updates
- File system with Huawei back-end
- Long-term stability tests
- Conclusions and future plans

Timeline

1.5 years of Huawei...

Major Review

Major Review

Major Review

Major Review

Minor Review

Board of Sponsors

Minor Review

Board of Sponsors

01/2012

01/2013

09/2013

Project starts

First tests

Upgrade of the system

Stress testing

File-system integration

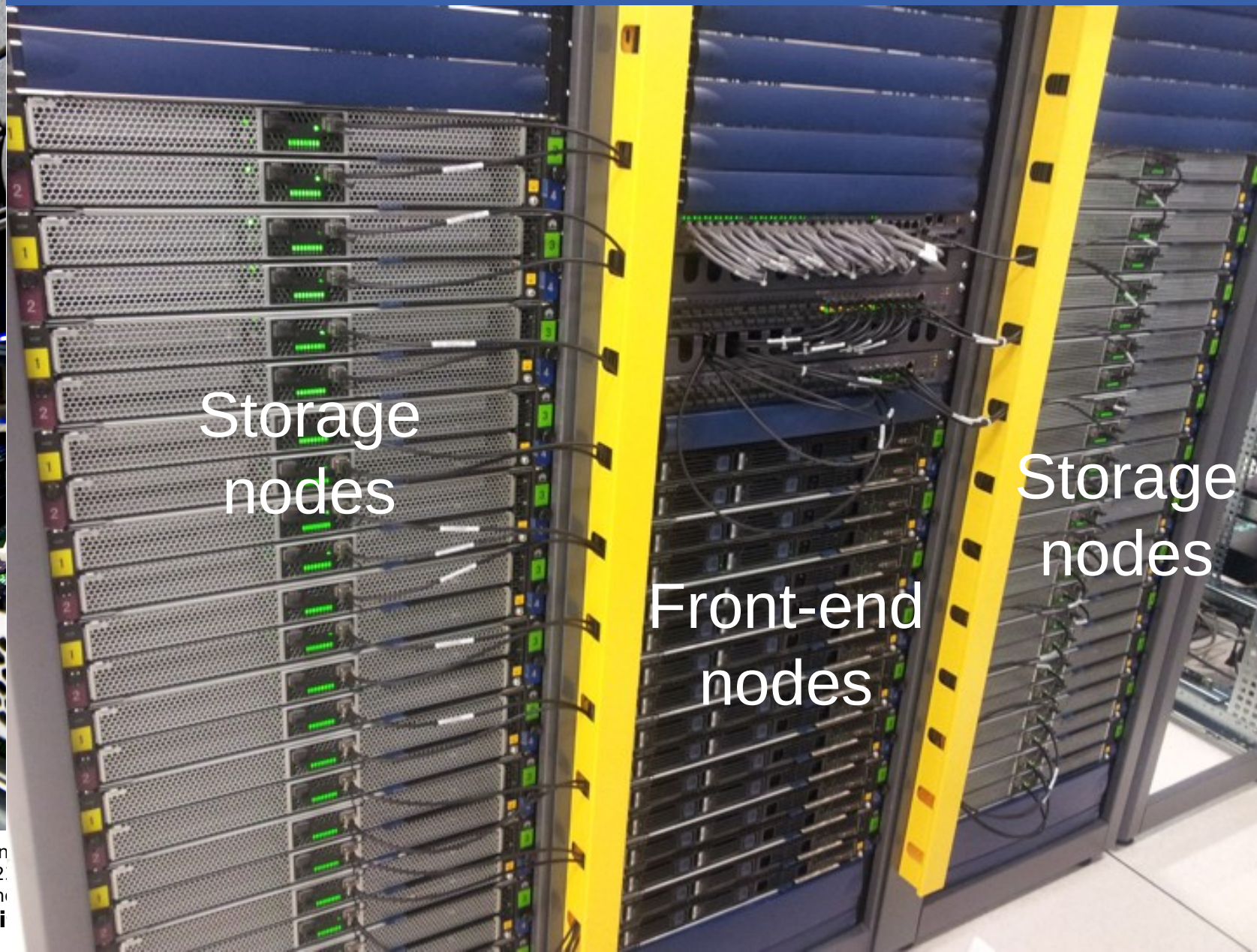
Commissioning of the system

Failure recovery testing

Full-scale stress testing

CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

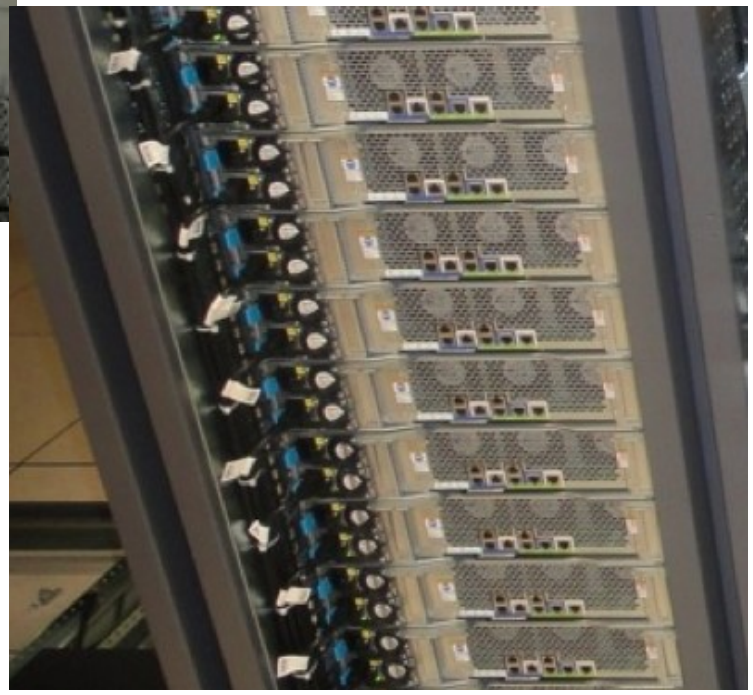
Current Huawei setup



CERN IT Departmen
CH-1211 Genève 2
Switzerlan
www.cern.ch/i

Storage nodes

One chassis has two blades



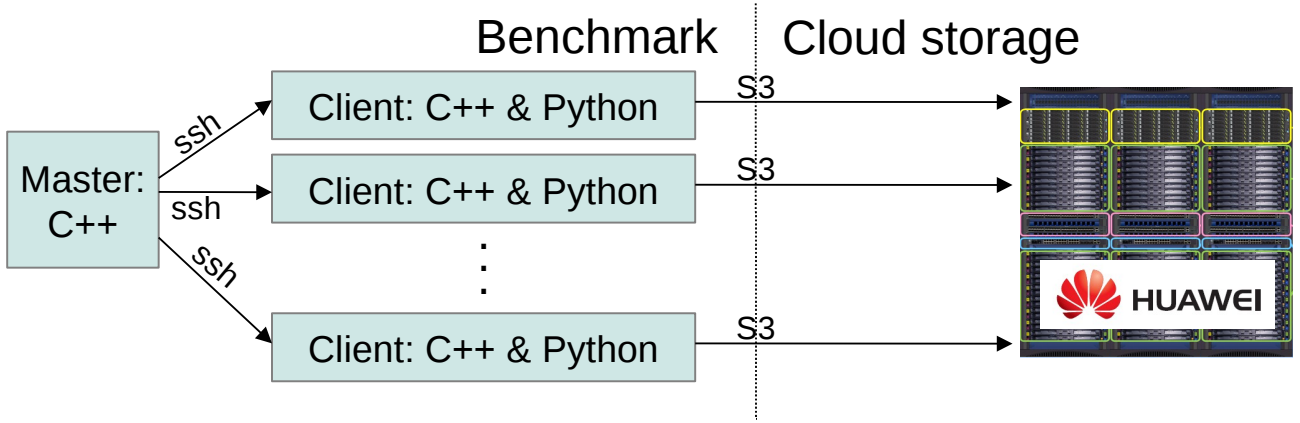
Each blade has eight storage nodes

CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

Benchmark setup

Distributed C++ benchmark

- Integrated with ROOT
- Client nodes connected with ssh
- S3 Python library to read and write files
- Histograms about specific metrics
  - Operation time, read/write speed, CPU/memory utilisation



Previous phase results

- Upload performance
  - 1400 files/second metadata (4kB files)
  - 2000 MB/second throughput (100MB files)

- Download performance
  - 18000 files/second metadata (4kB files)
  - 2300 MB/second throughput (100MB files)

- Recovery after powering off a chassis
  - Transparent disk failure recovery proven

CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

Davix - new S3 library

- Problem: tested C++ S3 libraries
  - Not able to detect all upload failures

?

Davix - new S3 library

- Problem: tested C++ S3 libraries
  - Not able to detect all upload failures

?

- Solution: Davix HTTP library
  - Adapted to use together with CERN grid storage developers (IT-SDC group)
  - Targeted for high performance file access
  - Provides S3 support
  - Performed as expected (after few minor updates)

!

Front-end log analysis

- Problem: detailed cloud storage monitoring
  - Monitor S3 requests (GET, PUT, etc.)
  - Monitor internal state (overloading, etc.)

?

CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

Front-end log analysis

- Problem: detailed cloud storage monitoring
  - Monitor S3 requests (GET, PUT, etc.)
  - Monitor internal state (overloading, etc.)

?

- Solution: front-end log analysis tool
  - Collect and archive logs from the front-ends
  - Summarise events from requested time range
  - Plot number of events vs time

!

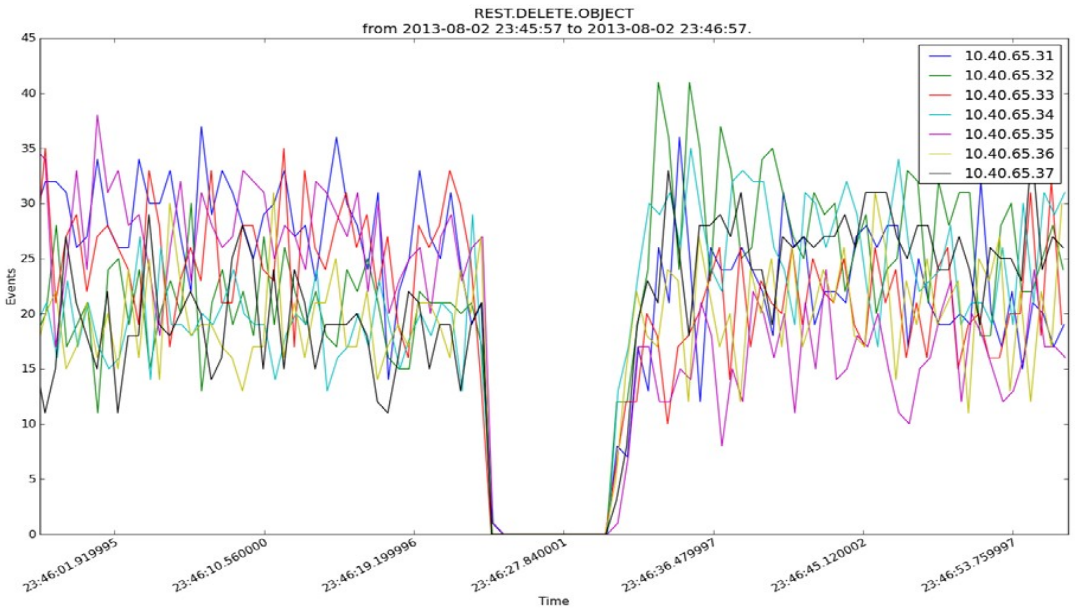
CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

Front-end log analysis

Summary of front-end activity

```
Req range: '03/Aug/2013 01:47:39' and '03/Aug/2013 01:47:40' (1 seconds).
                              Front end node...
Number of...                  #1      #2      #3      #4      #5      #6      #7      =SUM
       Total in range:        70      54      79      89      78      95      79      =544
       REST.DELETE.OBJECT     70      54      79      89      78      95      79      =544
       Unclassified:          0       0       0       0       0       0       0       =0
       Old entries:           3029652 1545741 13120   17494   20314   4881    24763   =24763
```

Plotting of selected parameters



The plotting feature was part of the contribution of Openlab summer student Carolina Lindqvist (2013).

CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

Updates to testing infrastructure

- Practical problems
  1) Verify test results with another framework
  2) Remove millions of files between test runs
  3) Change access rights for millions of files
  4) Create hundreds or thousands of buckets

CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

Updates to testing infrastructure

- Practical problems
  1) Verify test results with another framework
  2) Remove millions of files between test runs
  3) Change access rights for millions of files
  4) Create hundreds or thousands of buckets

?

- Solutions
  1) Second test framework with C++11 threads
  2) Multithreaded file deleting (200 files/s)
  3) Scripts to modify permissions of multiple files
  4) Scripts to create multiple buckets to multiple accounts

!

CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

CVMFS introduction

- What is CVMFS (CernVM File System)
  - Read only cached file system to deliver software
  - Widely used in WLCG (Worldwide LHC Computing Grid)
  - Mounted by clients and files are downloaded on demand

CVMFS introduction

- What is CVMFS (CernVM File System)
  - Read only cached file system to deliver software
  - Widely used in WLCG (Worldwide LHC Computing Grid)
  - Mounted by clients and files are downloaded on demand

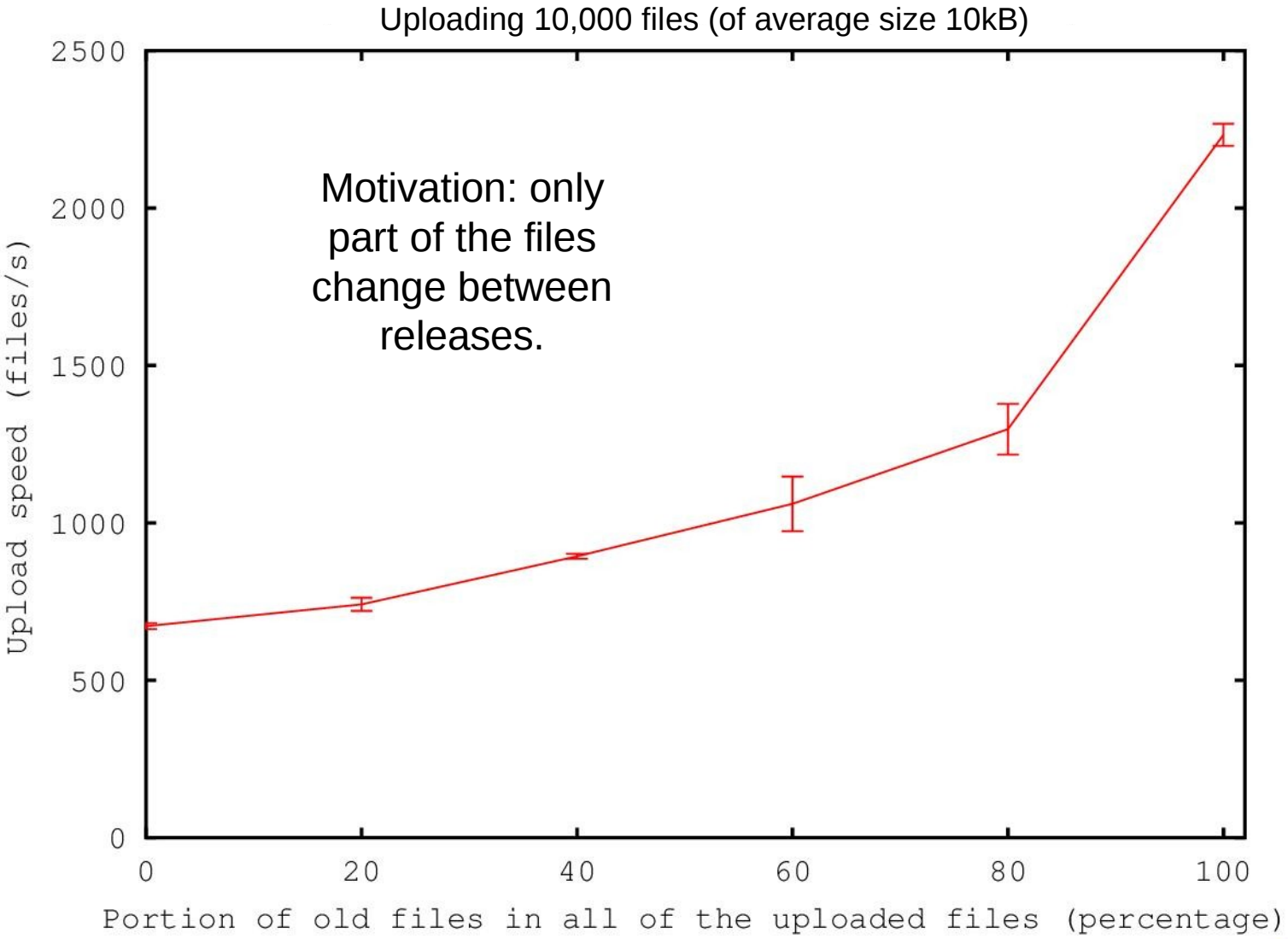
CernVM File system

- CVMFS challenges
  - Publishing new software should be fast (upload tens of thousands of files)
  - Files should be accessed with HTTP protocol

File system with Huawei back-end

- Implementation
  - Files are uploaded to multiple accounts and buckets in Huawei cloud storage
  - Files are downloaded with a flat namespace, i.e. no bucket names in the addresses (mapping to correct buckets is done by Squid proxy servers)
- Result
  - Full publish procedure tested to work using 30,000 small files
  - Upload speed 600 files/second (with 300 threads)

CVMFS partial uploads

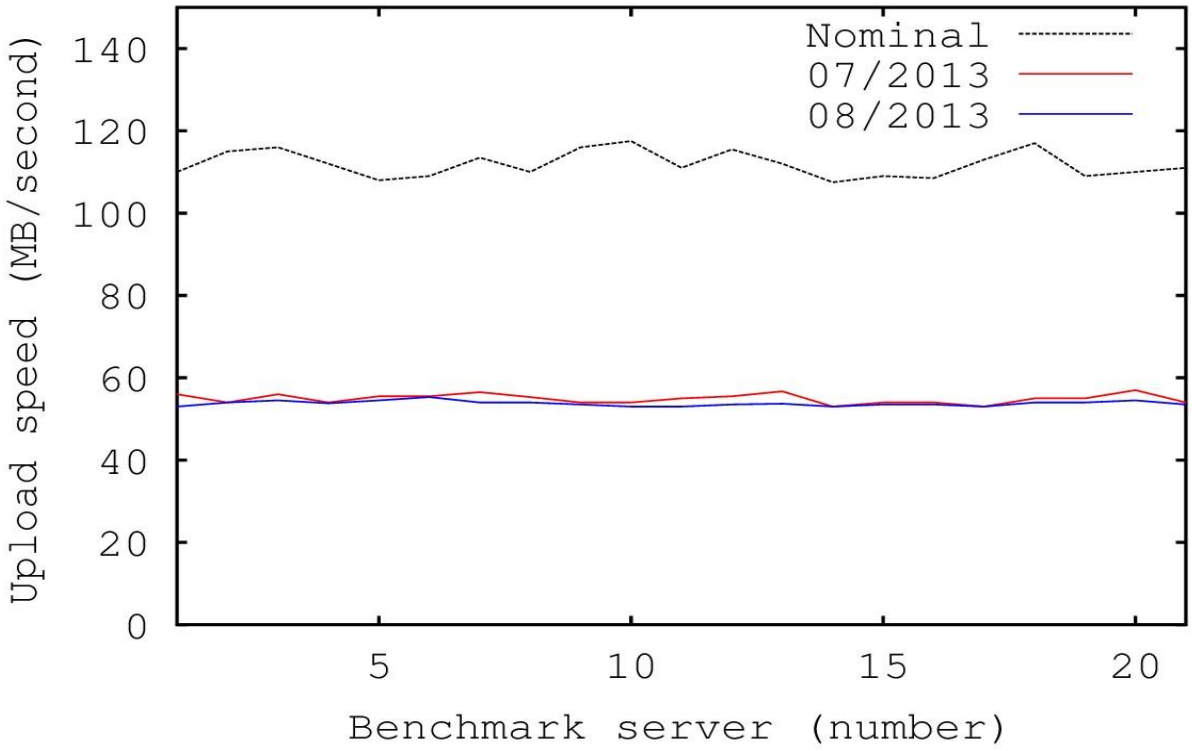


Motivation: only part of the files change between releases.

CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

Long-term upload stability

Upload throughput performance



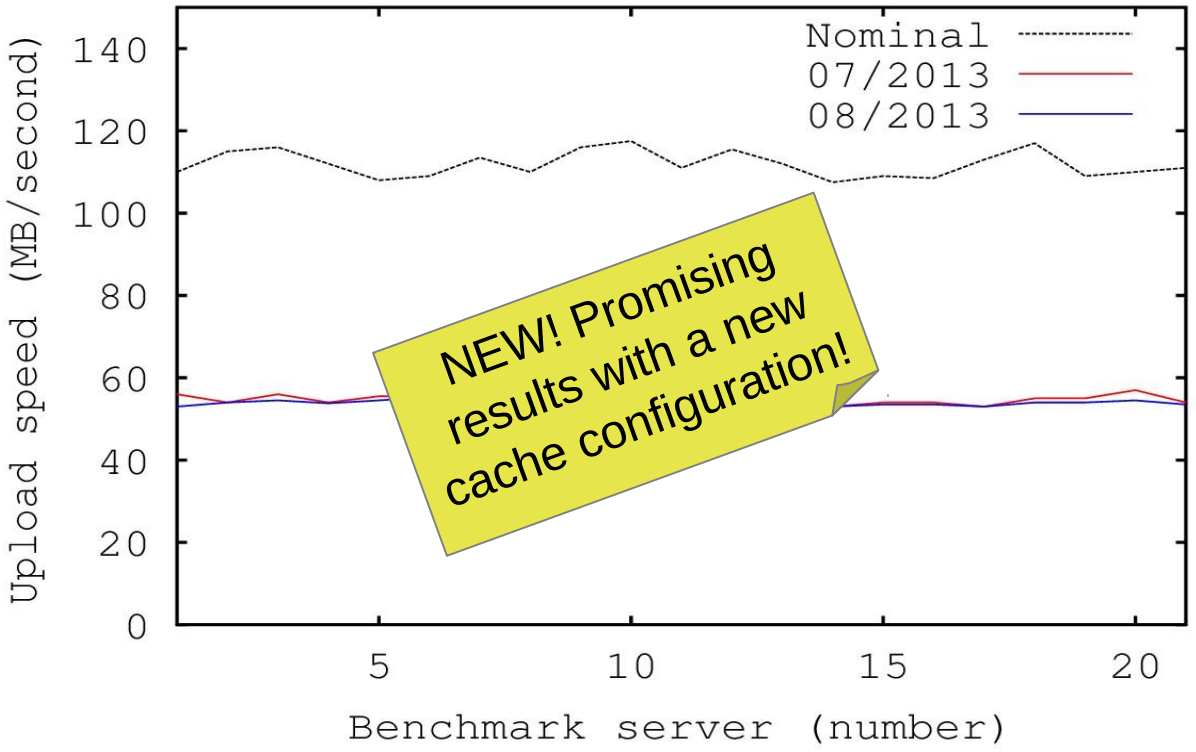
Uploading metadata (4kB) performance decreased from 1300 files/s to 800 files/s .

Emerged issues are being investigated.

CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

Long-term upload stability

Upload throughput performance



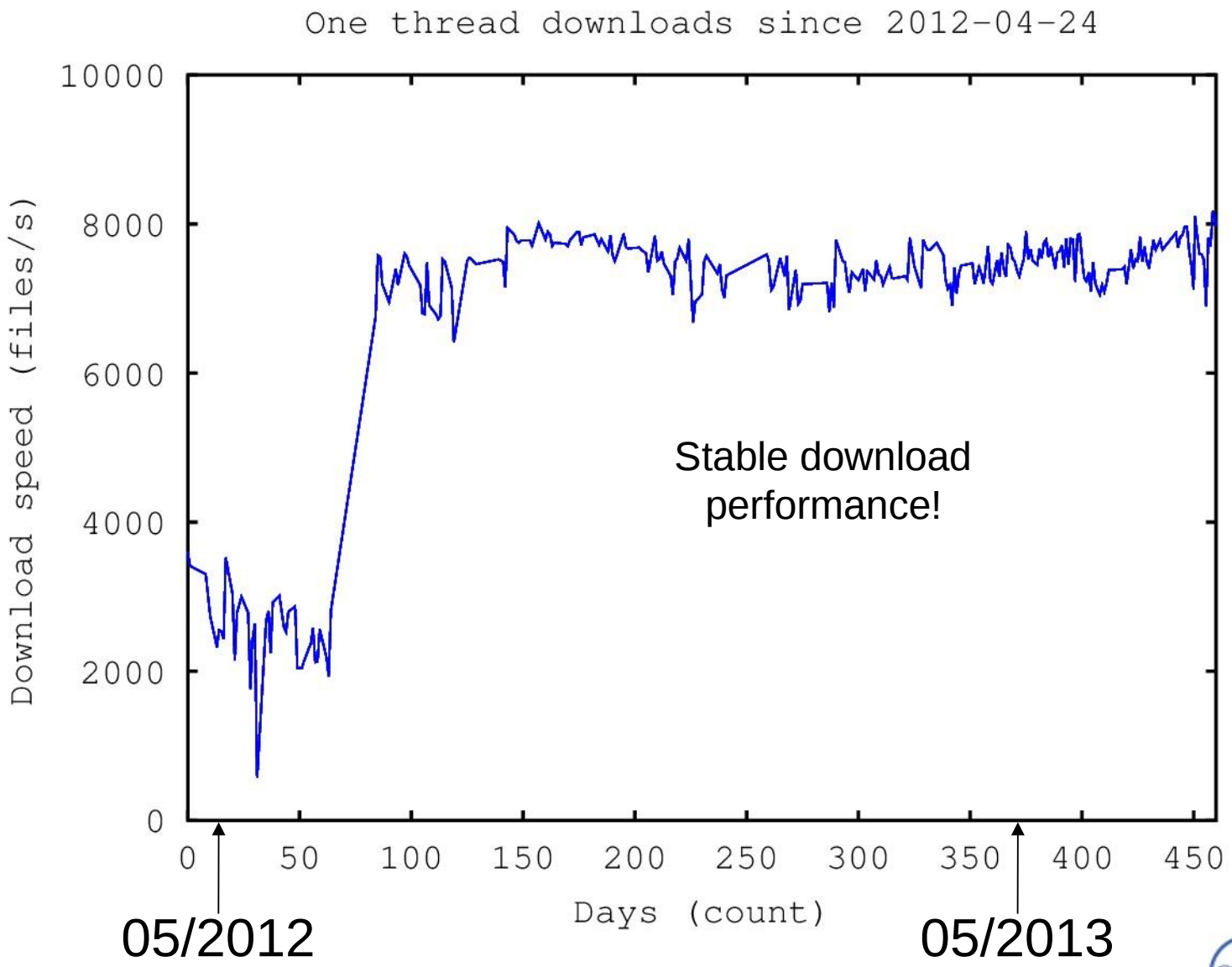
Uploading metadata (4kB) performance decreased from 1300 files/s to 800 files/s .

Emerged issues are being investigated.

CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

DSS Long term download stability

CERN IT Department



CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

Conclusions

- Testing infrastructure updates
  - Cloud storage activity analysis
  - Davix S3 library adopted in use
  - Another C++ framework used to verify results
  - Concurrent file deleting, permission modification scripts
- File system (CVMFS) with Huawei back-end
  - Full publish procedure tested (download and upload)
  - Uploading of only new files feature tested (speedup)
  - Publish speed of 600 files/second

Future plans

- Short term
  - Benchmark CVMFS with real release data
  - Investigate upload performance issues
- Long term
  - Second petabyte system with enterprise disks expected to arrive soon
  - Upgrade old Huawei cloud storage software version
  - Replication tests between two cloud storages
  - Erasure code impact on performance and space overhead
  - Prove total cost of ownership (TCO) gains of the system as part of a production service

DSS

Data & Storage Services

CERN IT Department

Huawei Cloud Storage

Seppo S. Heikkila
Maitane Zotes Resines
CERN IT

Openlab Major Review Meeting
26th of September 2013
CERN, Geneva

CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

HUAWEI

CERN openlab