DSS

Data & Storage Services

CERN IT Department

Cloud storage performance and first experience from prototype services at CERN

Maitane Zotes Resines, Seppo S. Heikkila, Dirk Duellmann, Geoffray Adde, Rainer Toebbicke, CERN

James Hughes, Huawei

Lu Wang, IHEP

20th International Conference on Computing in High Energy and Nuclear Physics (CHEP)

14th of October 2013

Amsterdam, The Netherlands

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

CERN

Introduction

Motivation

- Cloud storage market is growing fast
- CERN uses custom made storage solutions

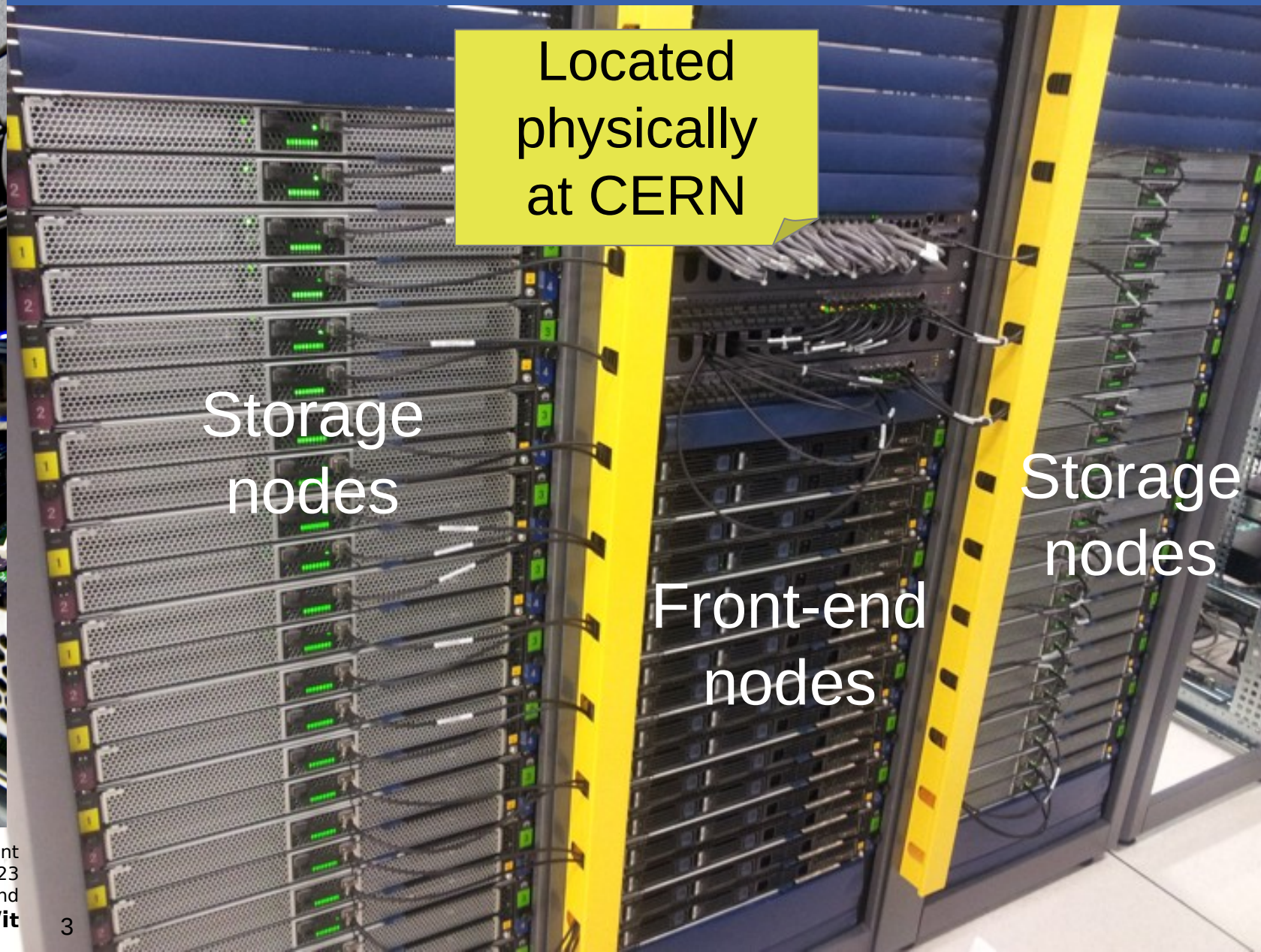
Question

"Are cloud storages able to meet the High Energy Physics (HEP) data storage requirements?"

Method

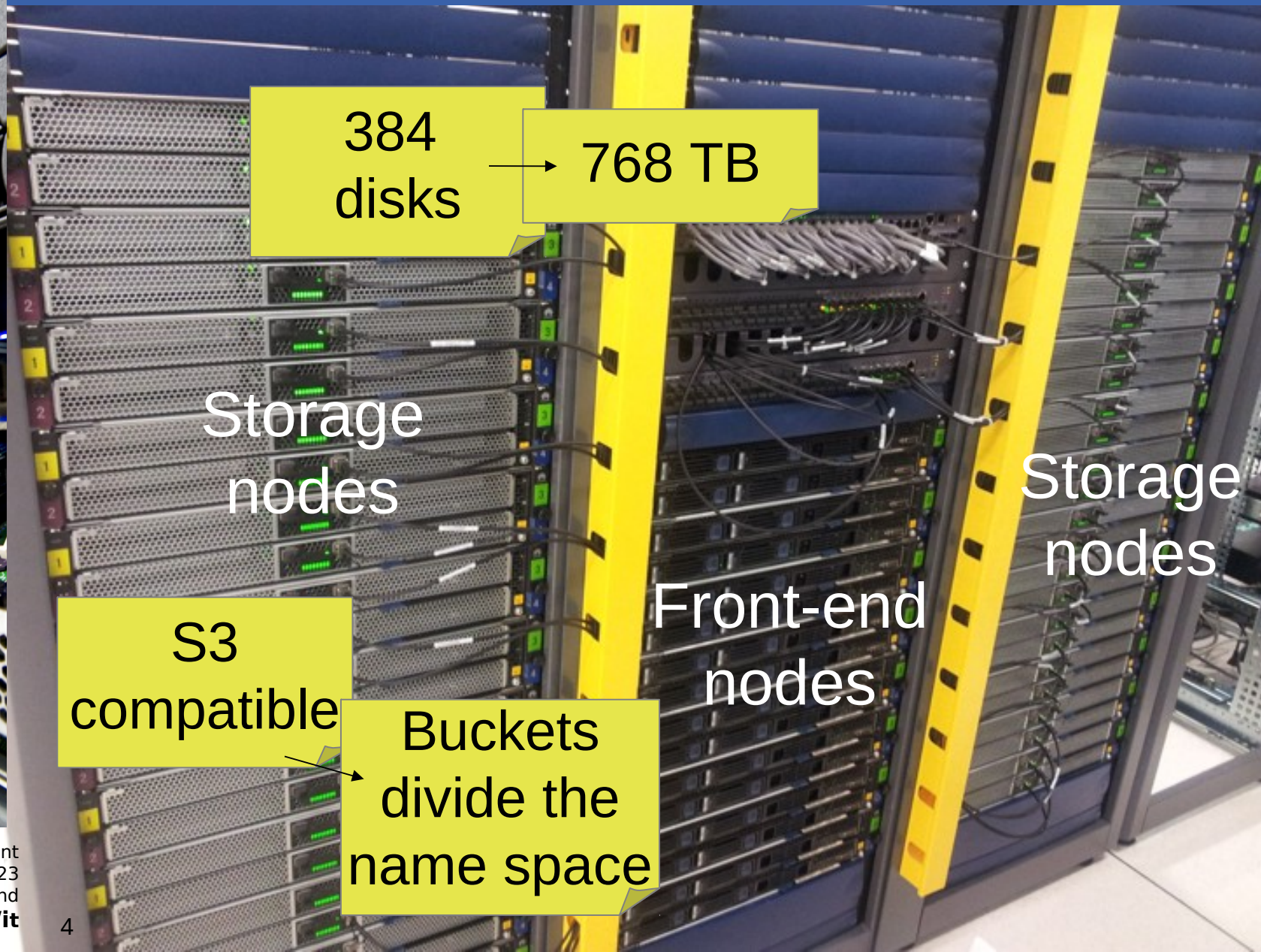
- Evaluate scalability and fault-tolerance
- Test with real applications

Huawei cloud storage setup



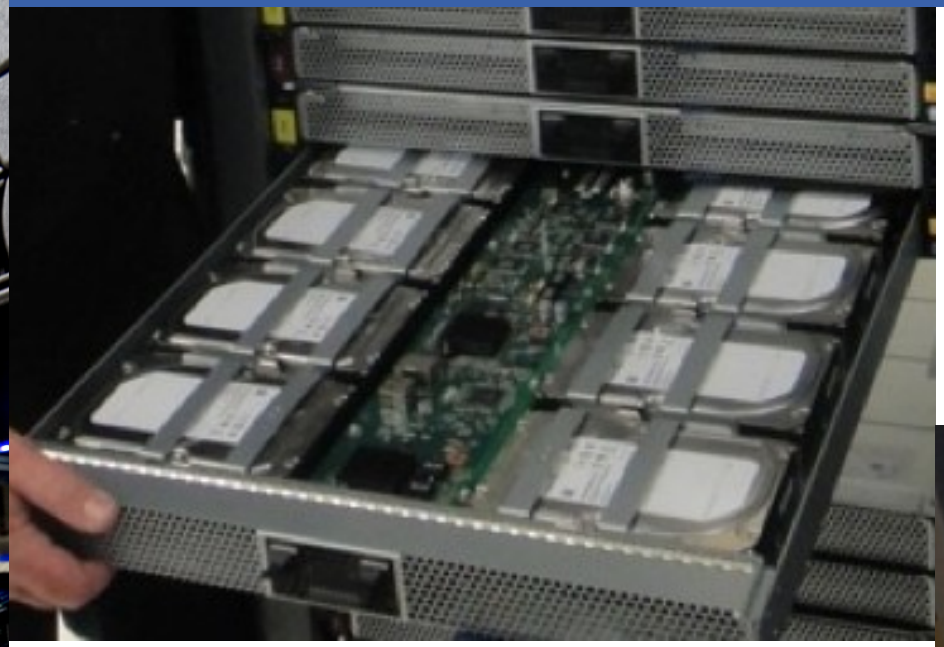
CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

Huawei cloud storage setup

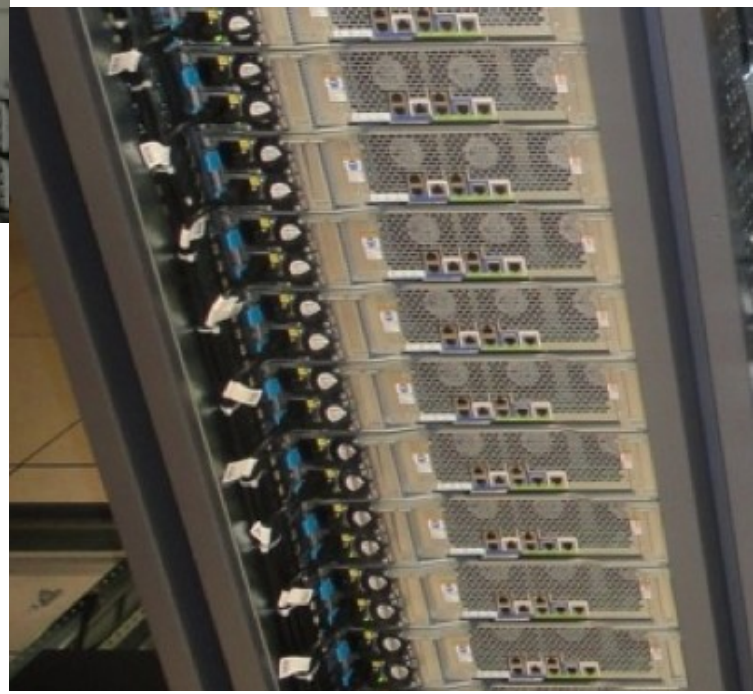


CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

Storage nodes



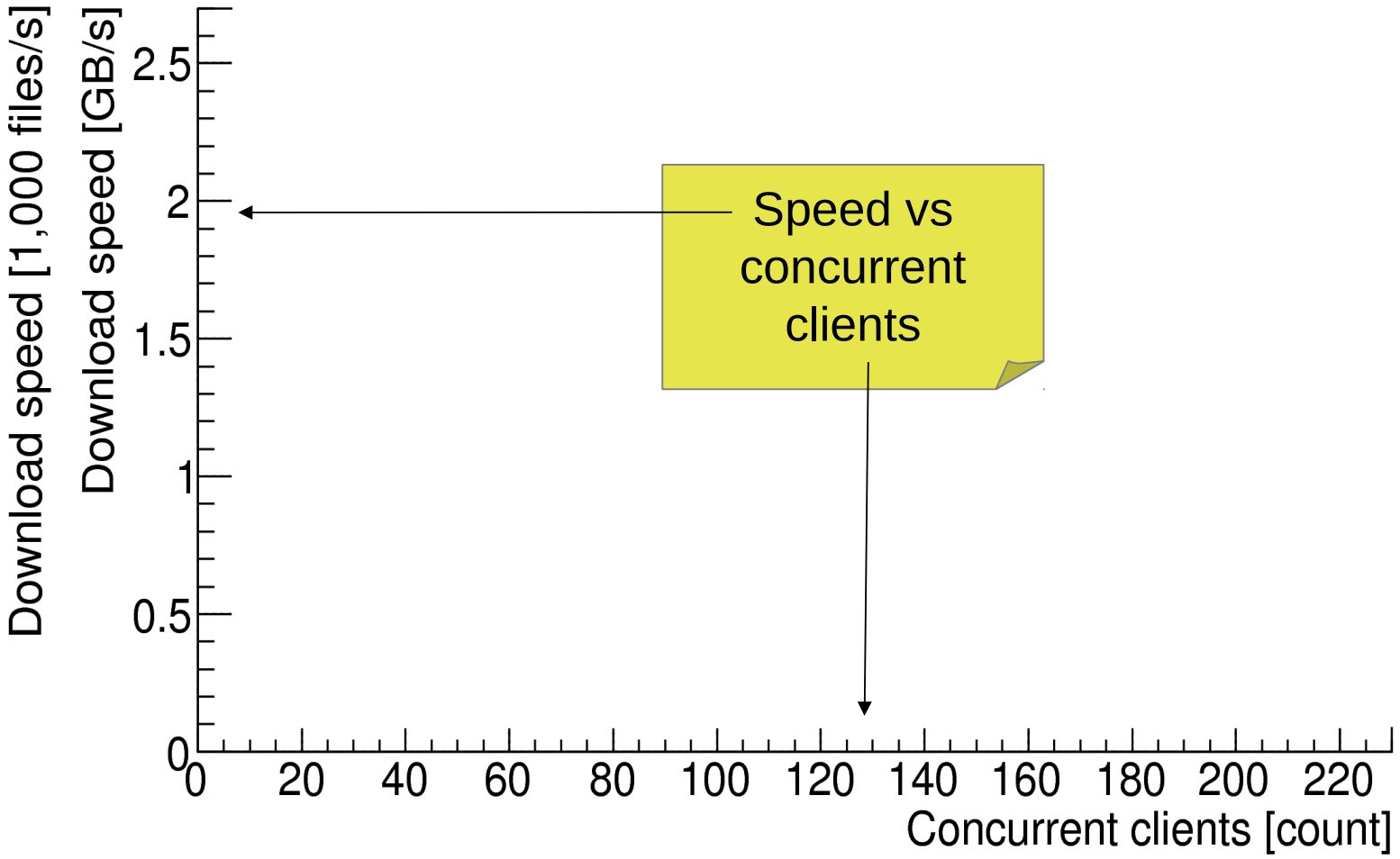
Each blade has eight storage nodes



One chassis has two blades (16 disks)

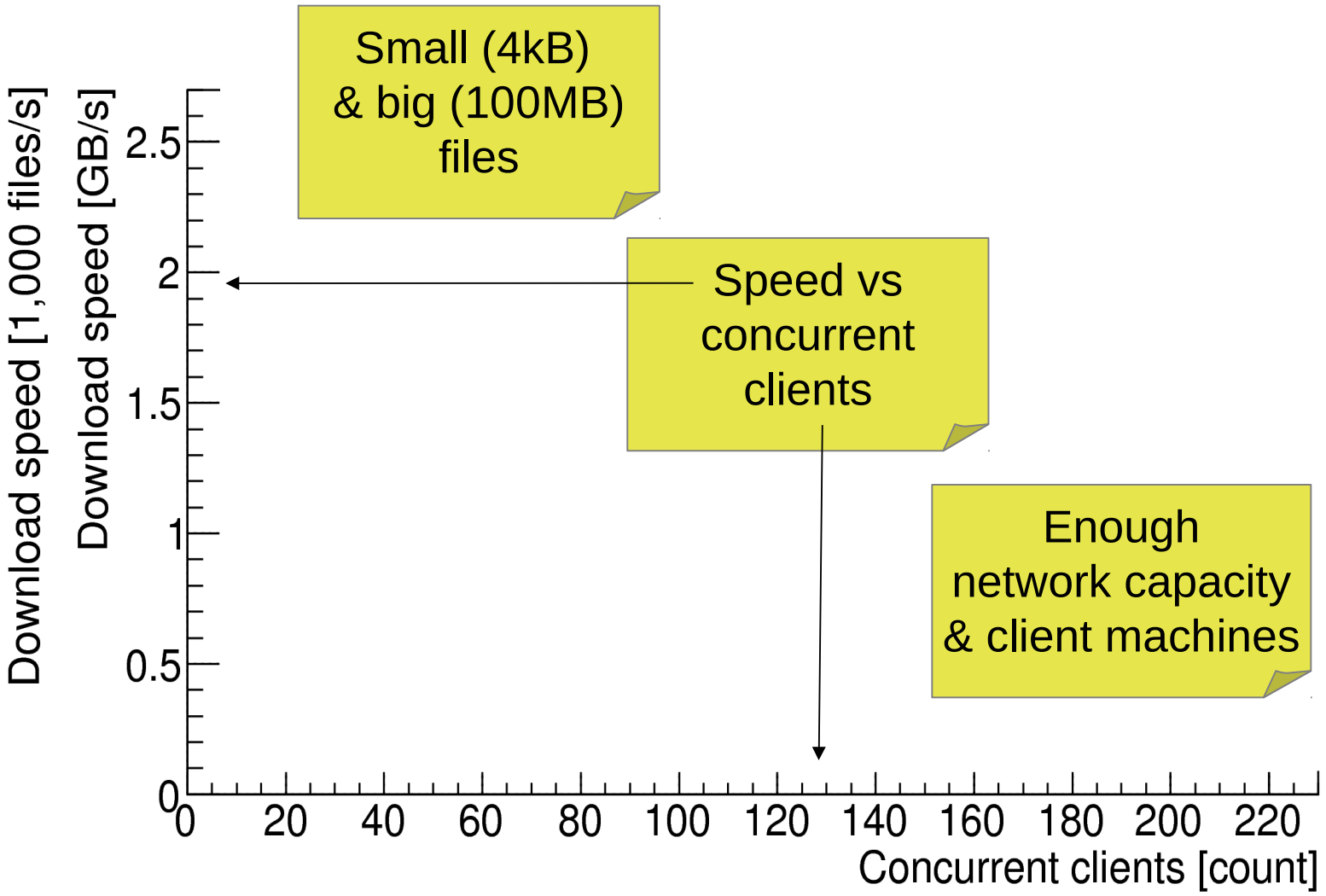
CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

Benchmarking scalability



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

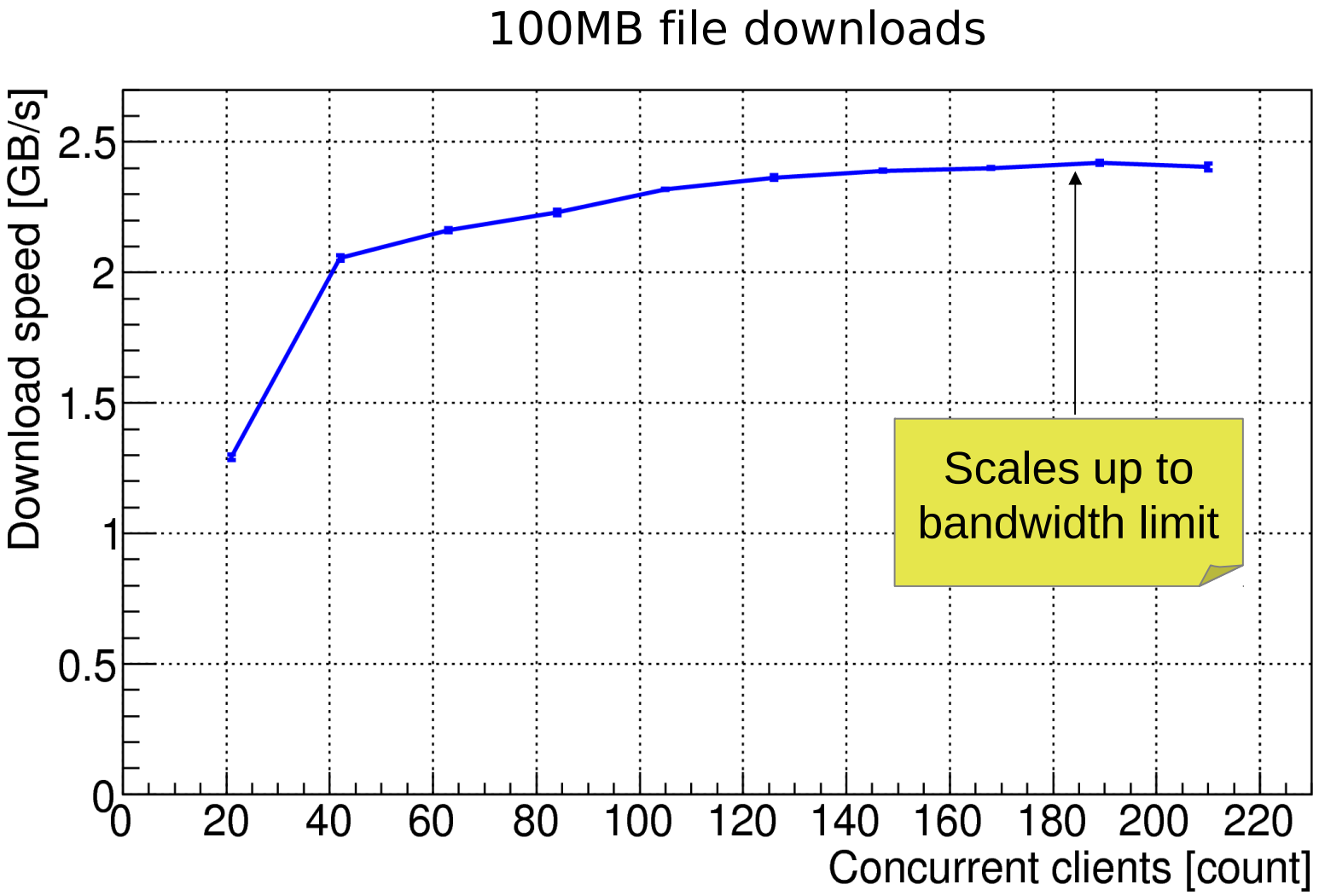
Benchmarking scalability



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

DSS Download throughput scalability

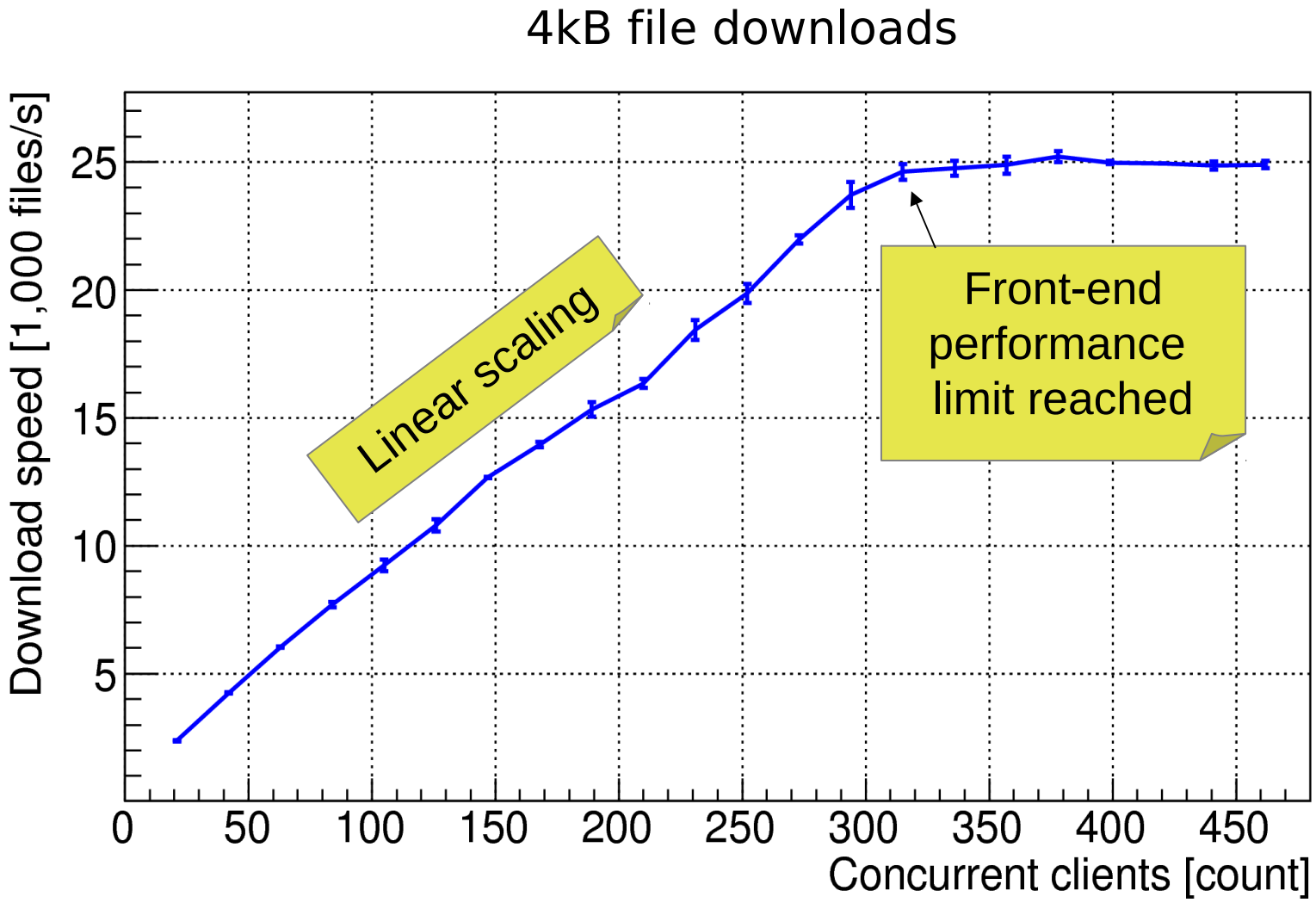
CERN IT Department



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

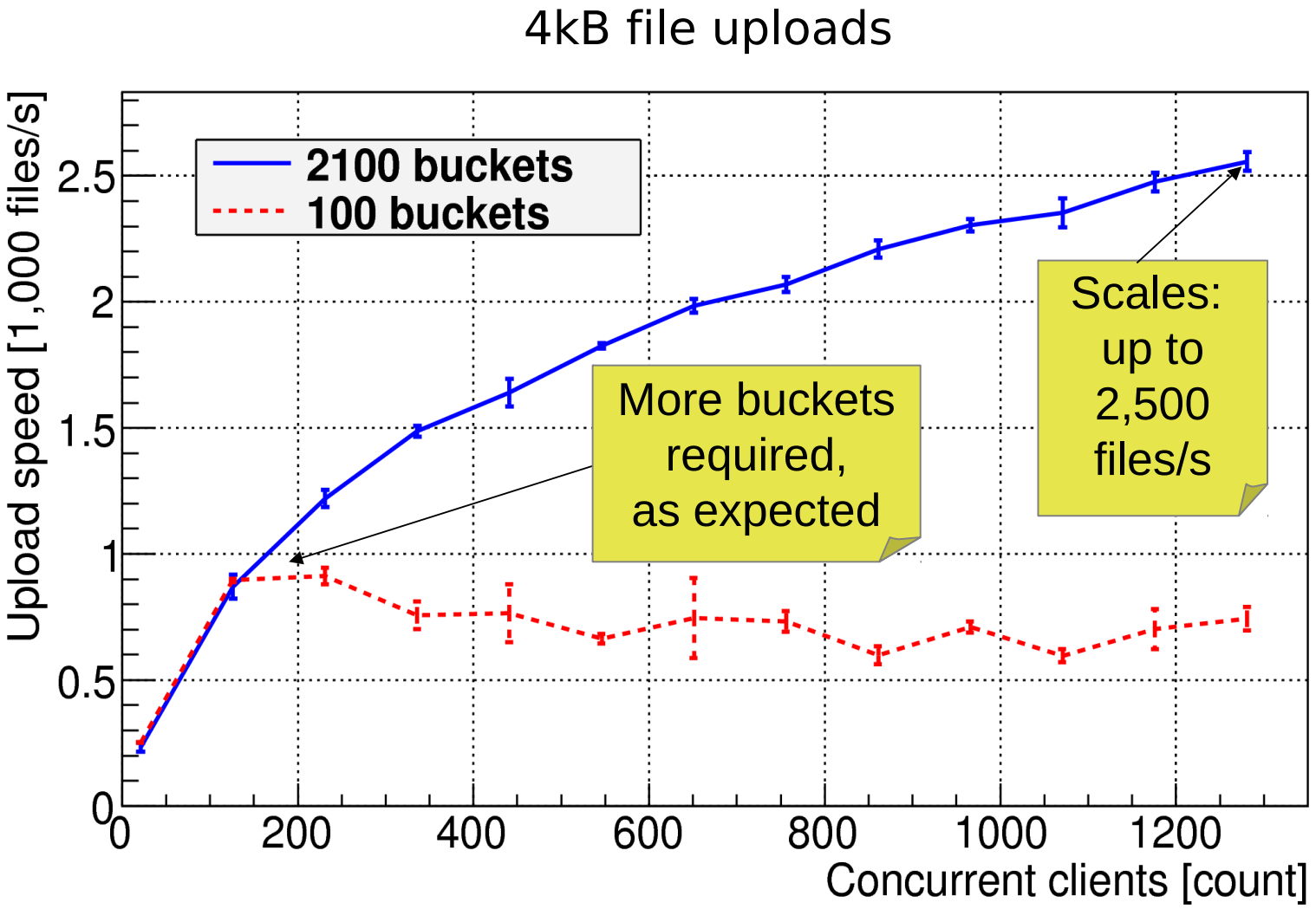
DSS Download metadata scalability

CERN IT Department



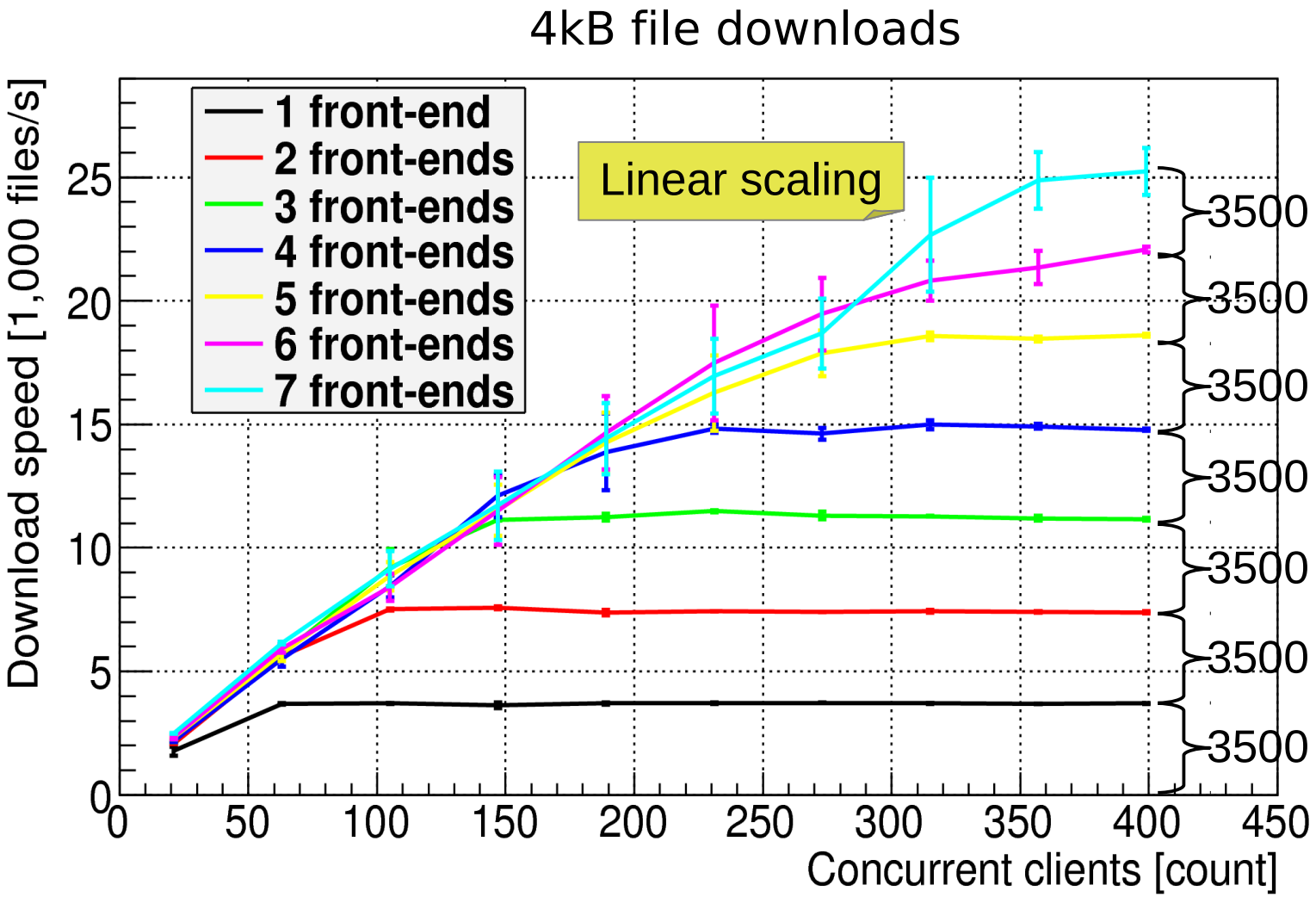
CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

Upload metadata scalability



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

Front-end scalability



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

Raw performance summary

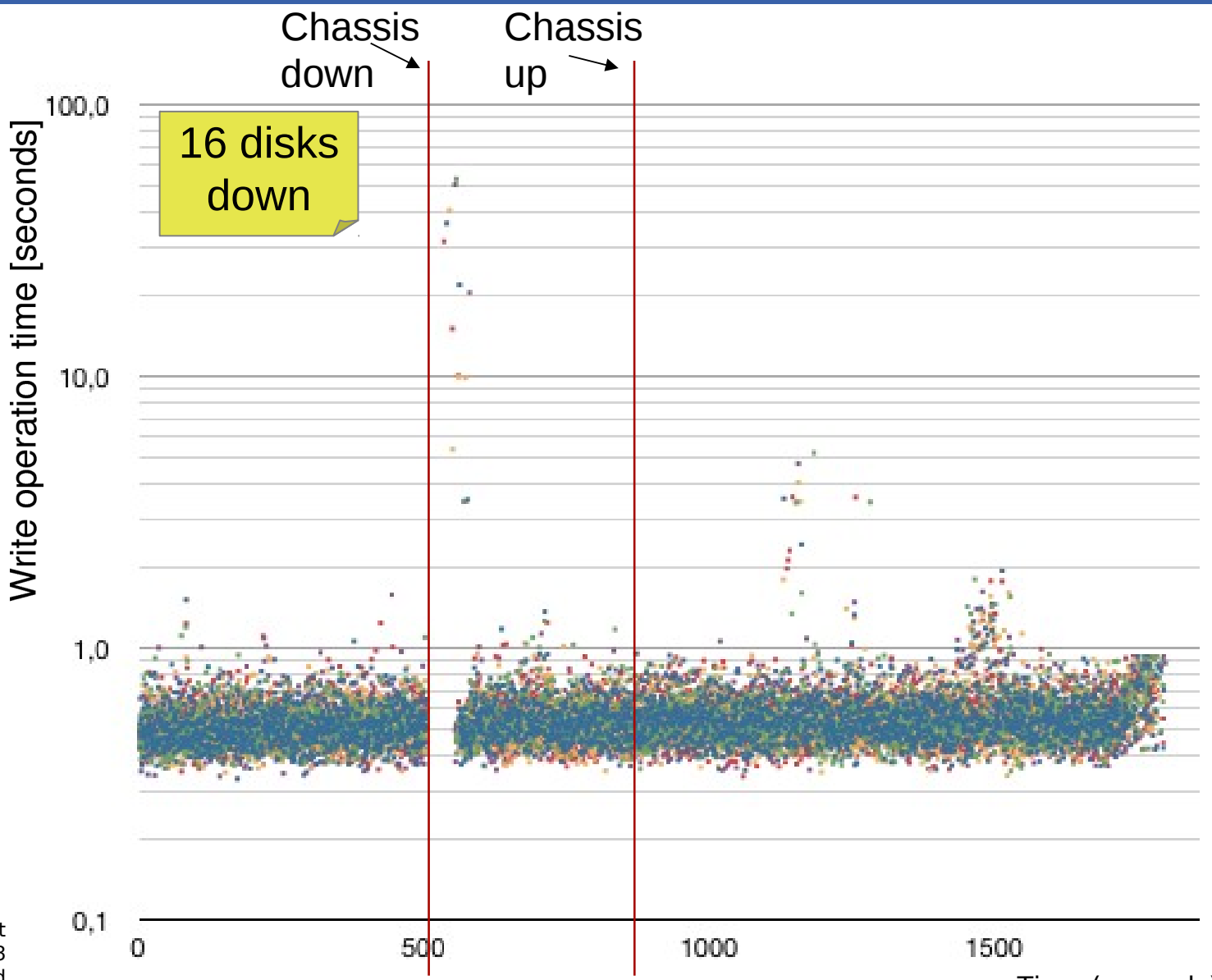
- Metadata (4kB) performance
  - 2,500 files/second upload
  - 25,000 files/second download
- Throughput (100MB) performance
  - 20Gbit network fully utilized
- Front-end scalability
  - Each front-end downloads 3500 files/s

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

DSS

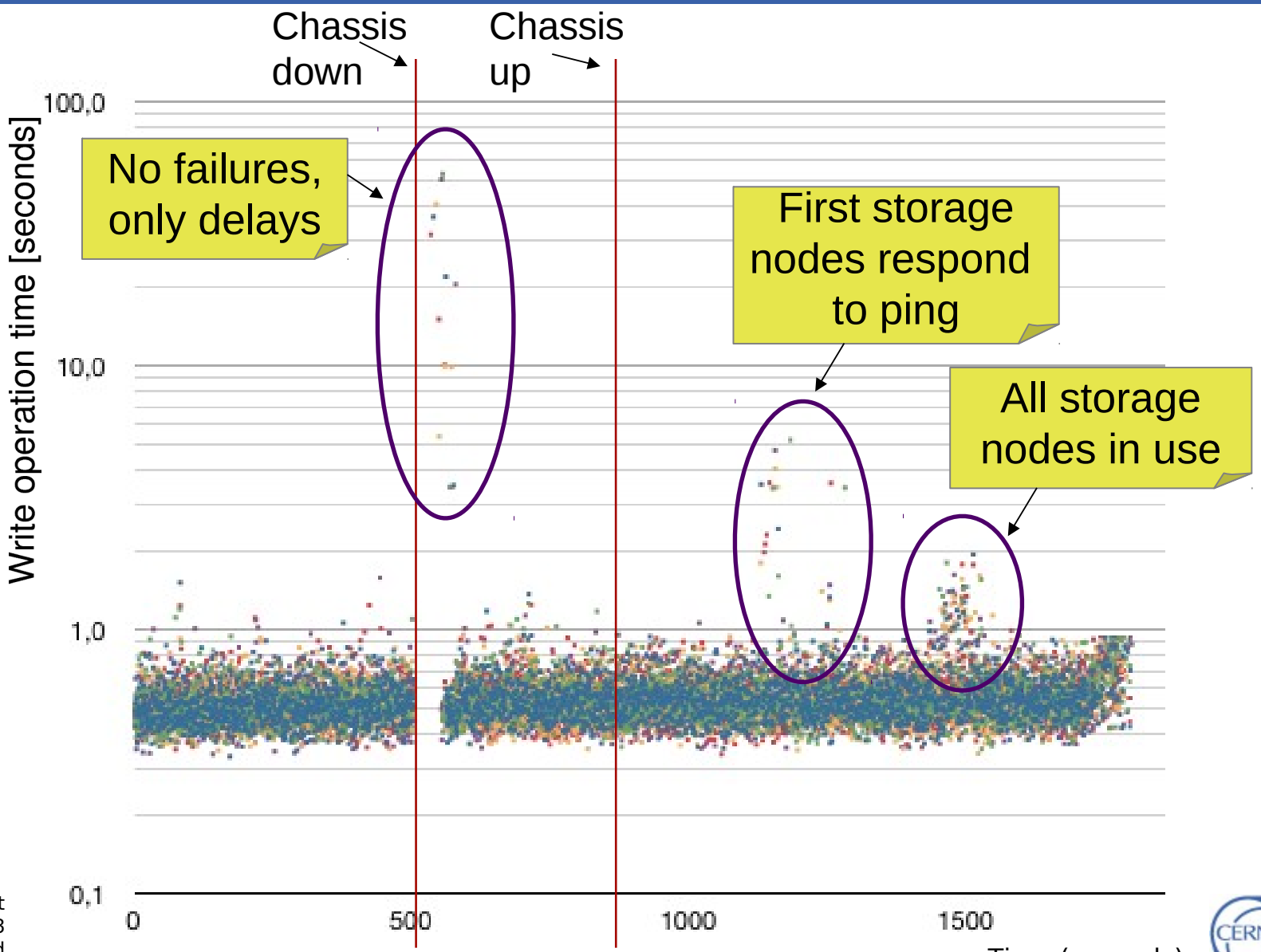
Recovery after powering off a chassis

CERN IT Department



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

Recovery after powering off a chassis



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

DSS CVMFS introduction

CERN IT Department

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

CernVM File system

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 users 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

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

File system with Huawei back-end

- Implementation

CernVM File system

  - Files are uploaded to multiple buckets in the cloud storage
  - Files are downloaded with unified name space
    http://cloud.cern.ch/~~bucket-42~~/file001.bin

    http://cloud.cern.ch/file001.bin

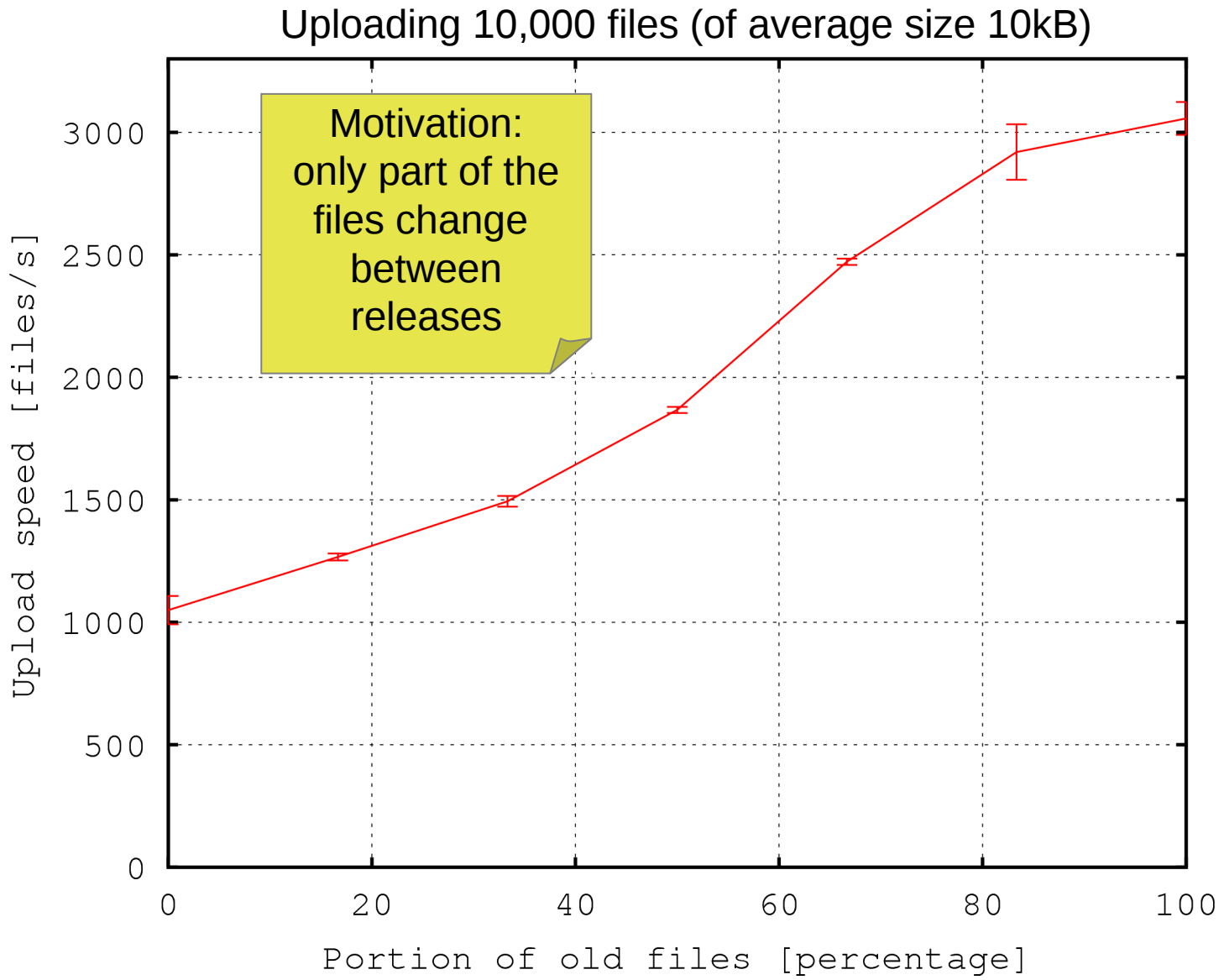
File system with Huawei back-end

- Implementation
  - Files are uploaded to multiple buckets in the cloud storage
  - Files are downloaded with unified name space
    http://cloud.cern.ch/~~bucket-42~~/file001.bin

    http://cloud.cern.ch/file001.bin
- Result
  - Full publish procedure tested to work using 30,000 small files
  - Upload speed 1200 files/second (with 240 threads)

CernVM File system

CVMFS partial uploads



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

Conclusion

- Raw performance
  - Upload and download scalability demonstrated
  - Additional front-end nodes increased linearly the performance
- Fault tolerance: powering off a chassis
  - Transparent disk failure recovery demonstrated
- File system with cloud storage back-end
  - Full publishing procedure tested
  - Uploading of only new files feature tested

Conclusion

- Raw performance
  - Upload and download scalability demonstrated
  - Additional front-end nodes increased linearly the performance
- Fault tolerance: powering off a chassis
  - Transparent disk failure recovery demonstrated
- File system with cloud storage back-end
  - Full publishing procedure tested
  - Uploading of only new files feature tested

Dank je wel. Questions?

seppo.heikkila@cern.ch