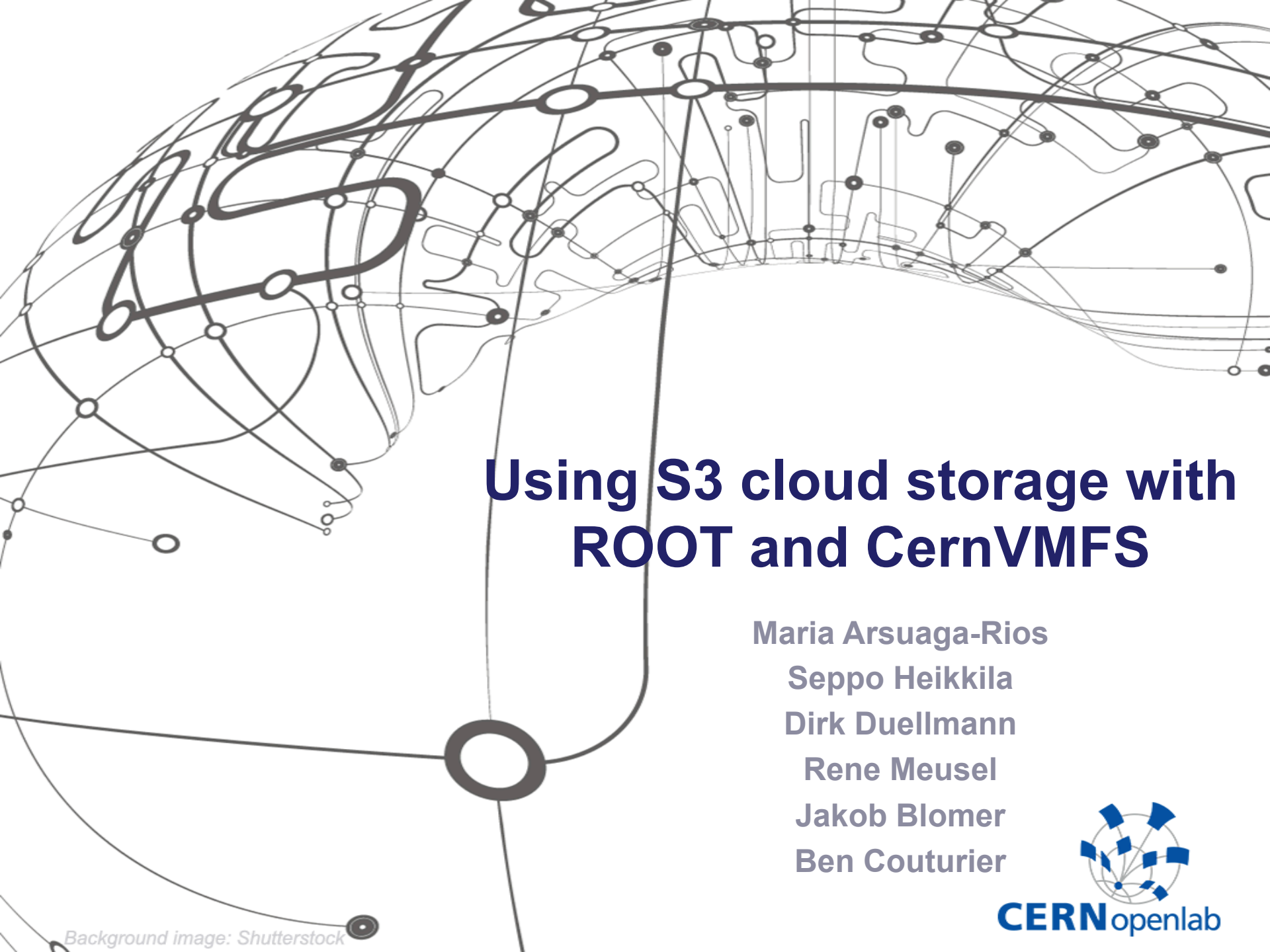


Using S3 cloud storage with ROOT and CernVMFS

Maria Arsuaga-Rios
Seppo Heikkila
Dirk Duellmann
Rene Meusel
Jakob Blomer
Ben Couturier

CERN openlab

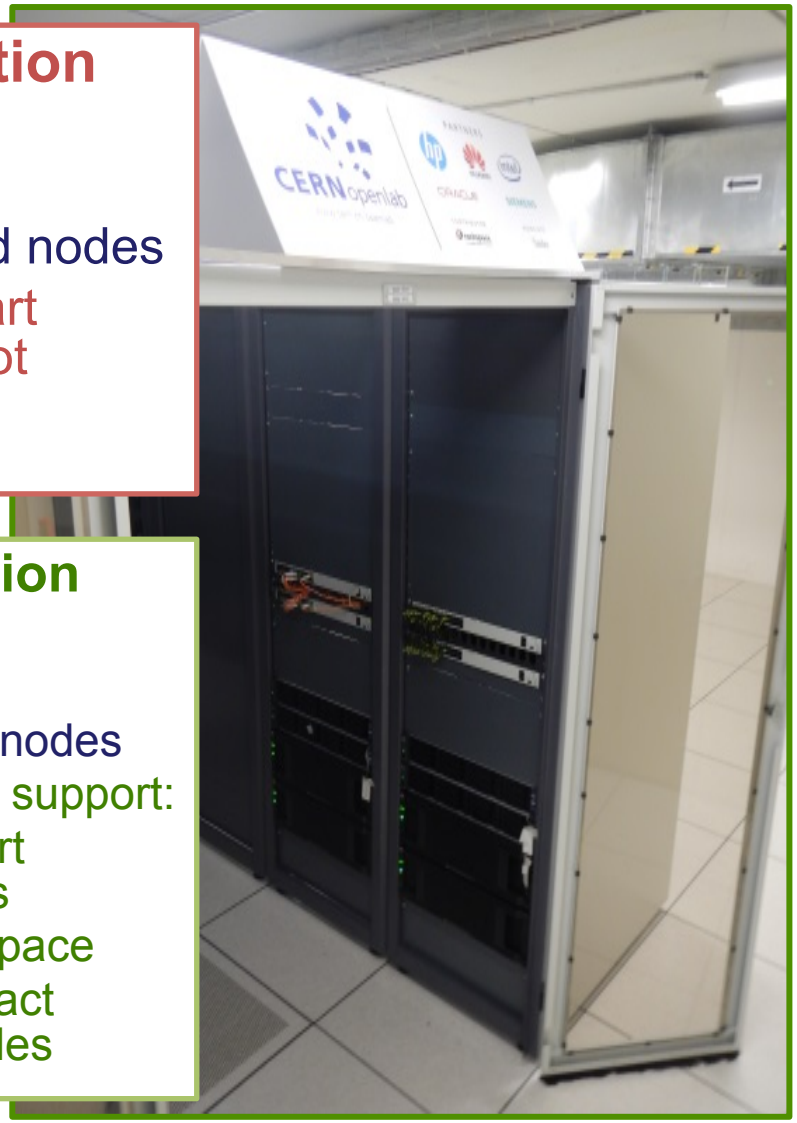
INDEX

- Huawei cloud storages at CERN
- Old vs. new Huawei UDS comparative
- Multi-client ROOT benchmarks
- Real CernVMFS application
- Summary

NEW HUAWEI UDS GENERATION

- **Old generation**
  - 384 disks
  - 768 TB
  - 7 front-end nodes
  - S3 multipart uploads not supported

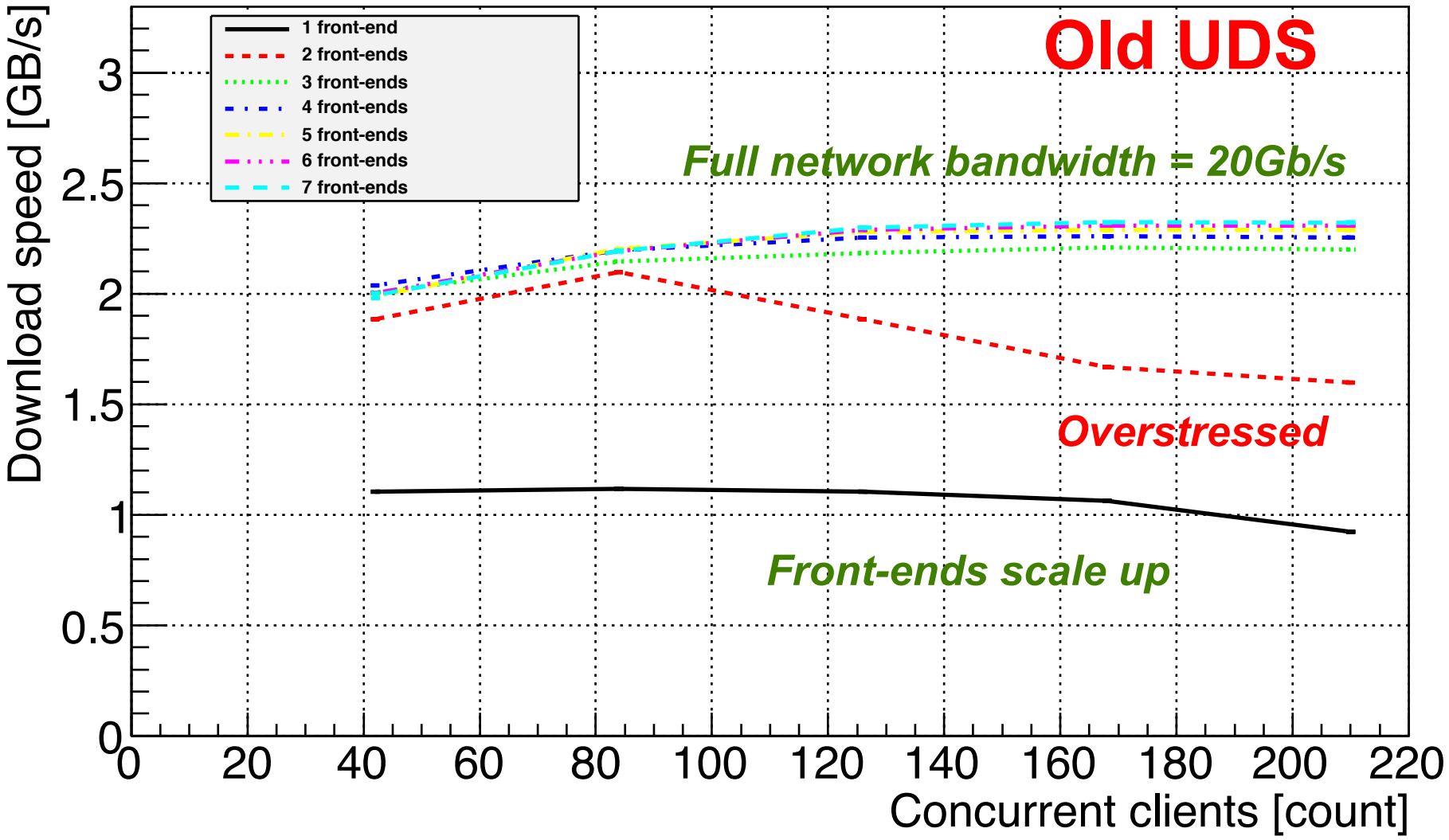
- **New generation**
  - 300 disks
  - 1.2 PB
  - 4 front-end nodes
  - Improve S3 support:
    - multipart uploads
  - Less rack space
  - More compact storage nodes



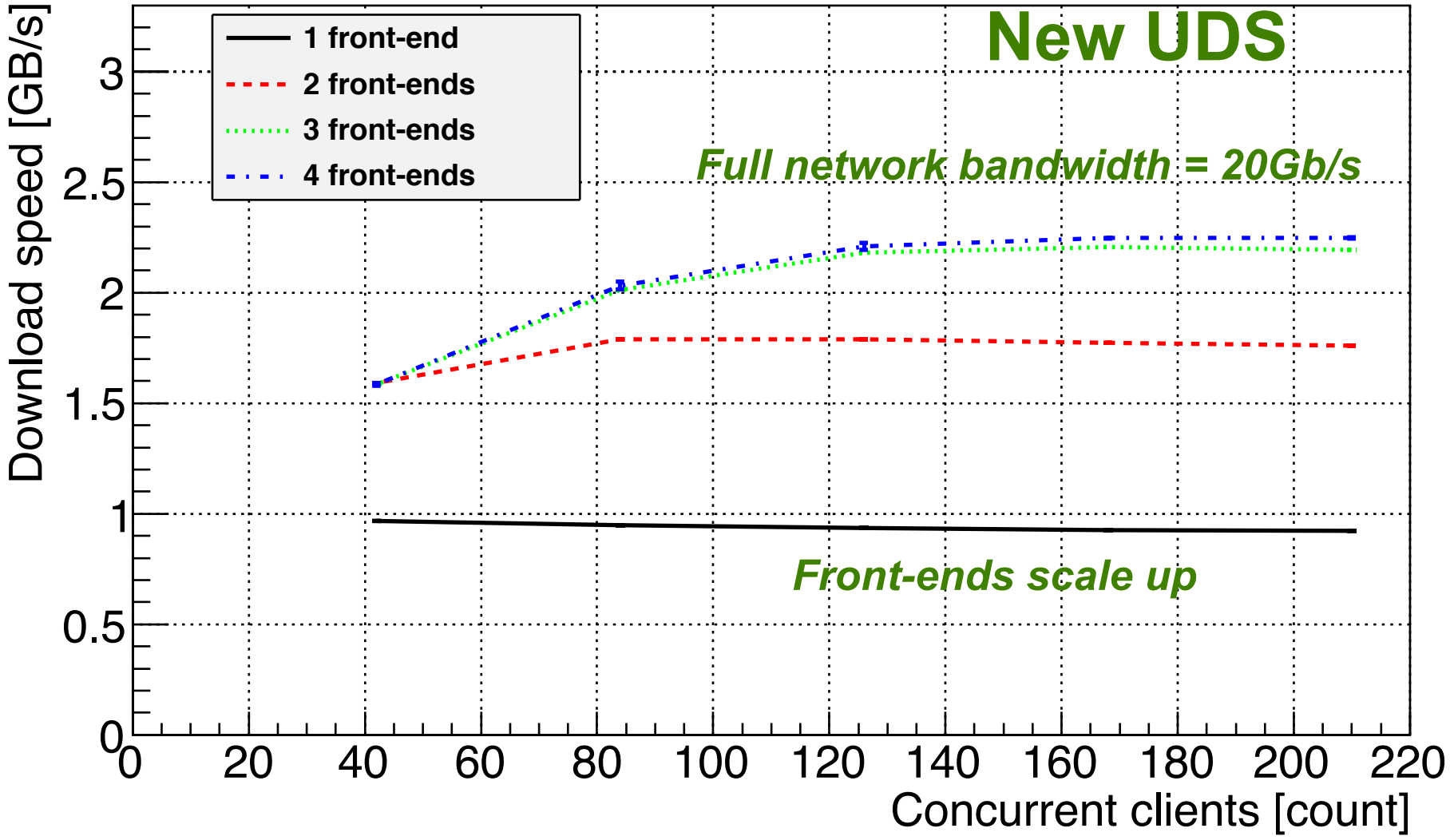
OLD vs. NEW HUAWEI UDS COMPARATIVE

- **Throughput tests**
  - 100MB downloads
  - 100MB uploads
- **Metadata tests**
  - 4kB downloads
  - 4kB uploads

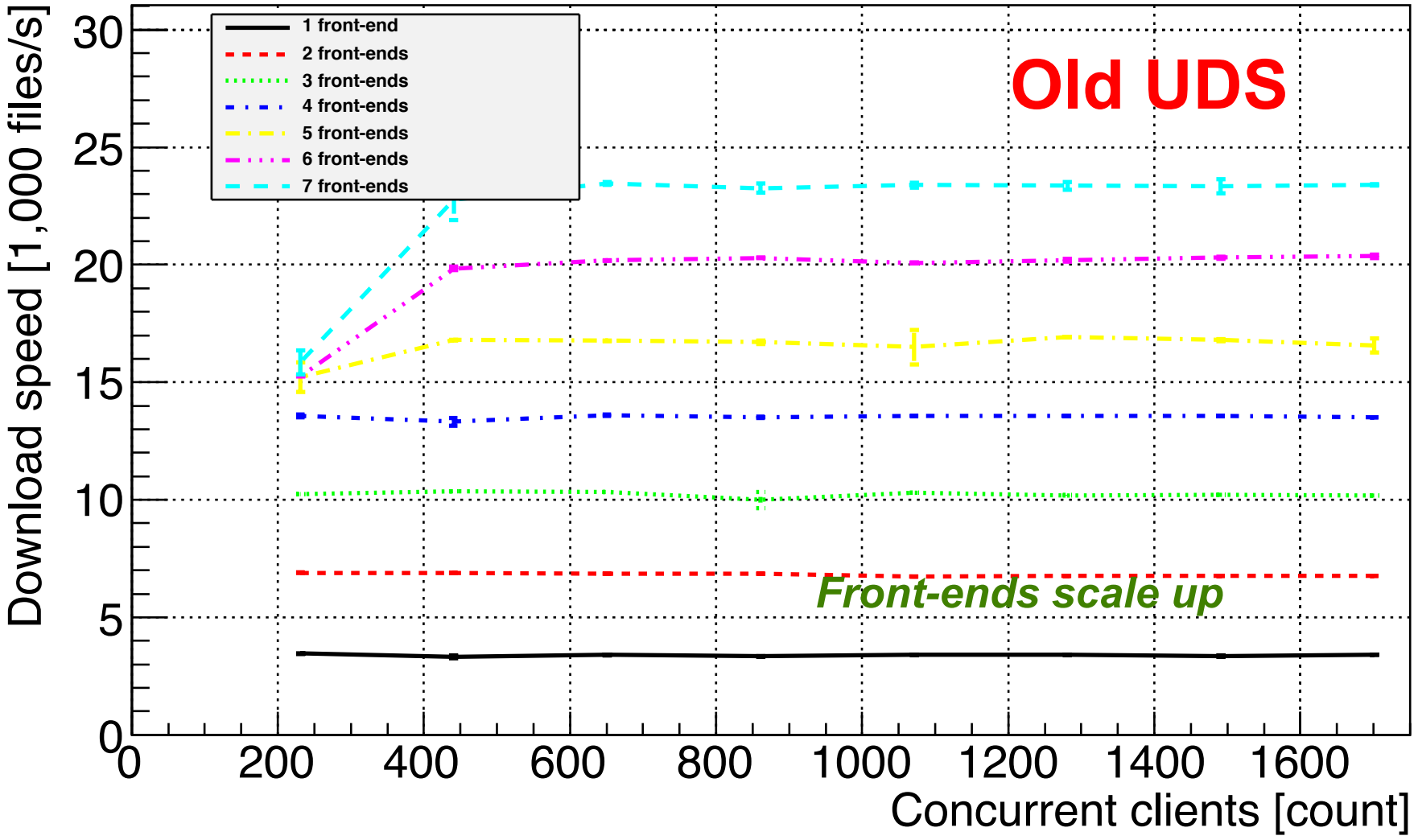
OLD vs. NEW HUAWEI UDS Throughput Download Tests



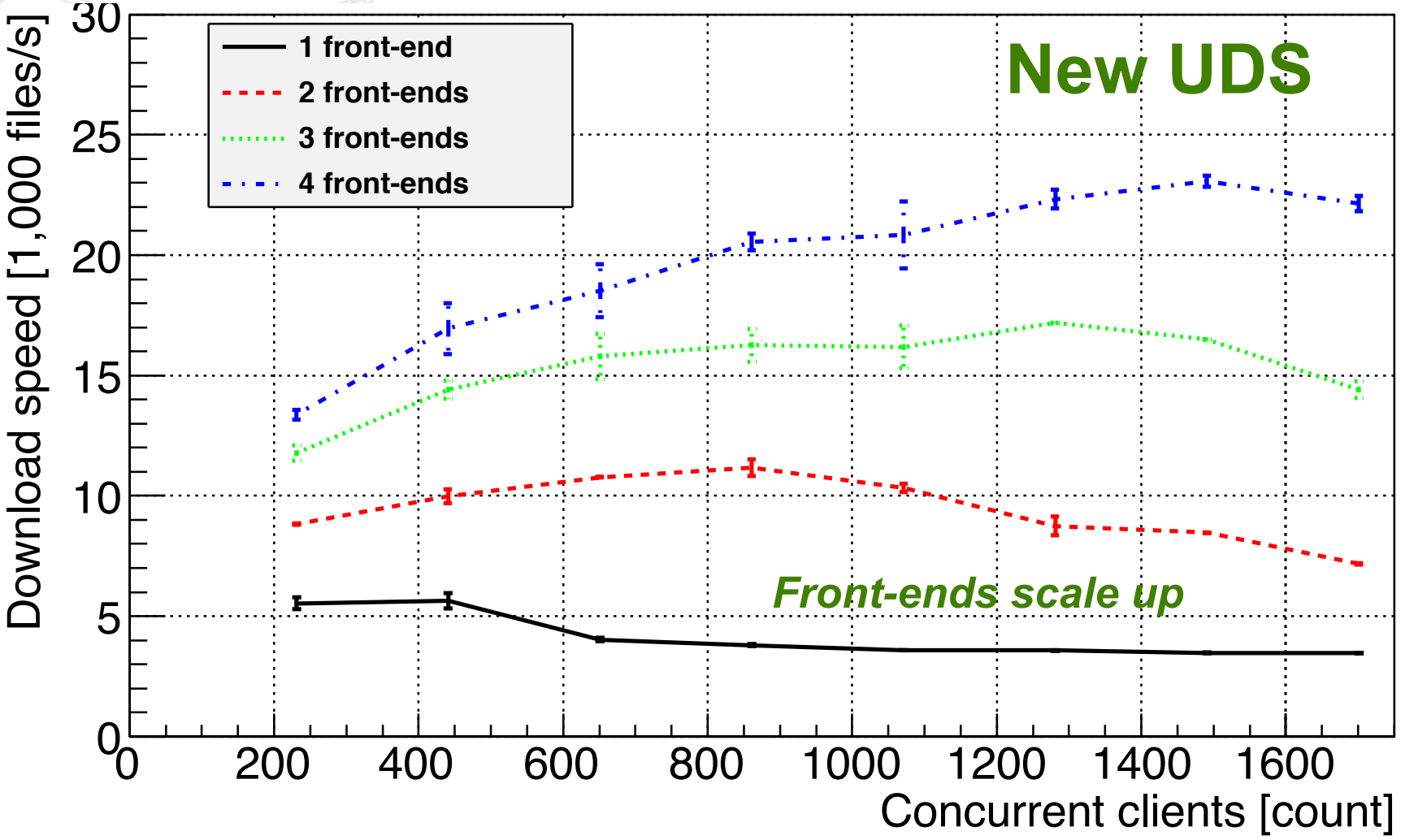
OLD vs. NEW HUAWEI UDS Throughput Download Tests



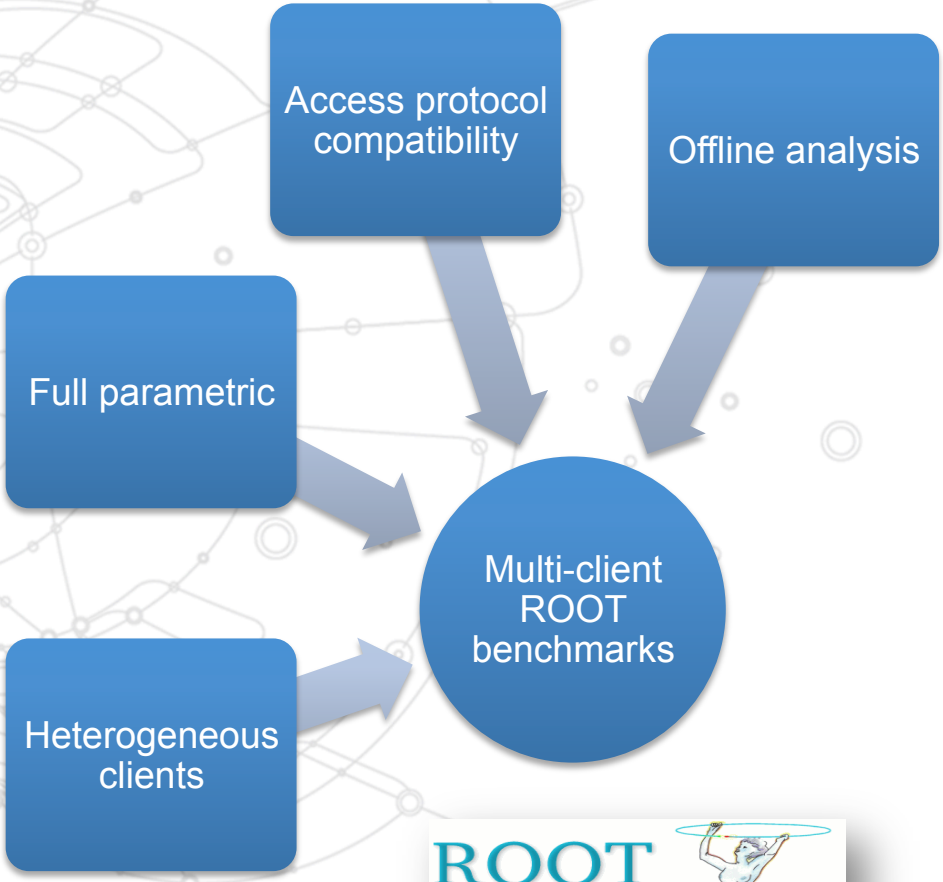
OLD vs. NEW HUAWEI UDS Metadata Download Tests



OLD vs. NEW HUAWEI UDS Metadata Download Tests

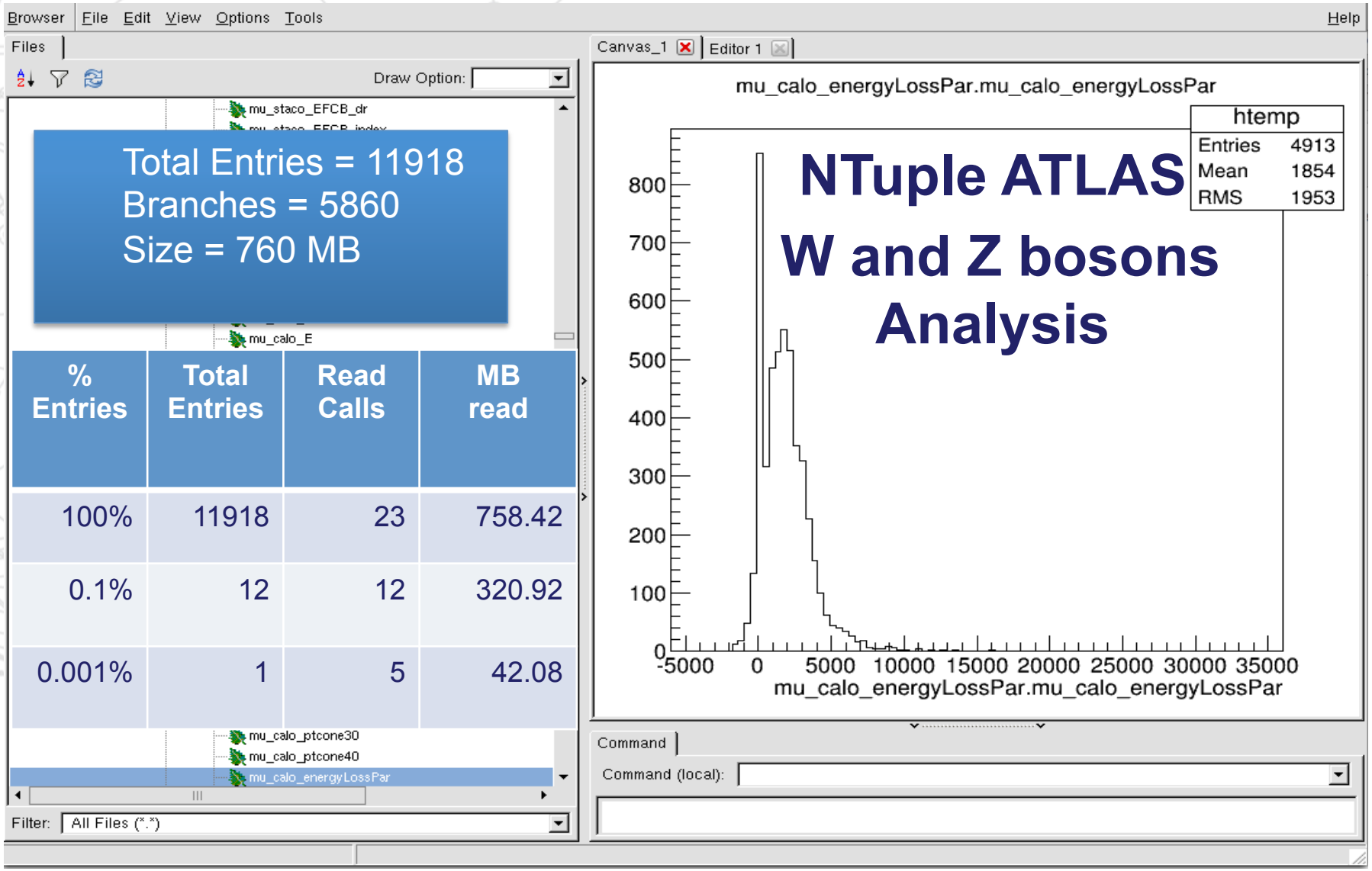


MULTI-CLIENT ROOT BENCHMARKS



- **Concurrent client accesses to real ATLAS ROOT files**
- **Sparseness, entries accessed**
  - 100%, 0.1%, 0.001%
- **2 use cases:**
  - 1 bucket
  - 64 buckets
- **Tested only in UDSs**
  - Ceph and Amazon do not support multi-range requests

MULTI-CLIENT ROOT BENCHMARKS



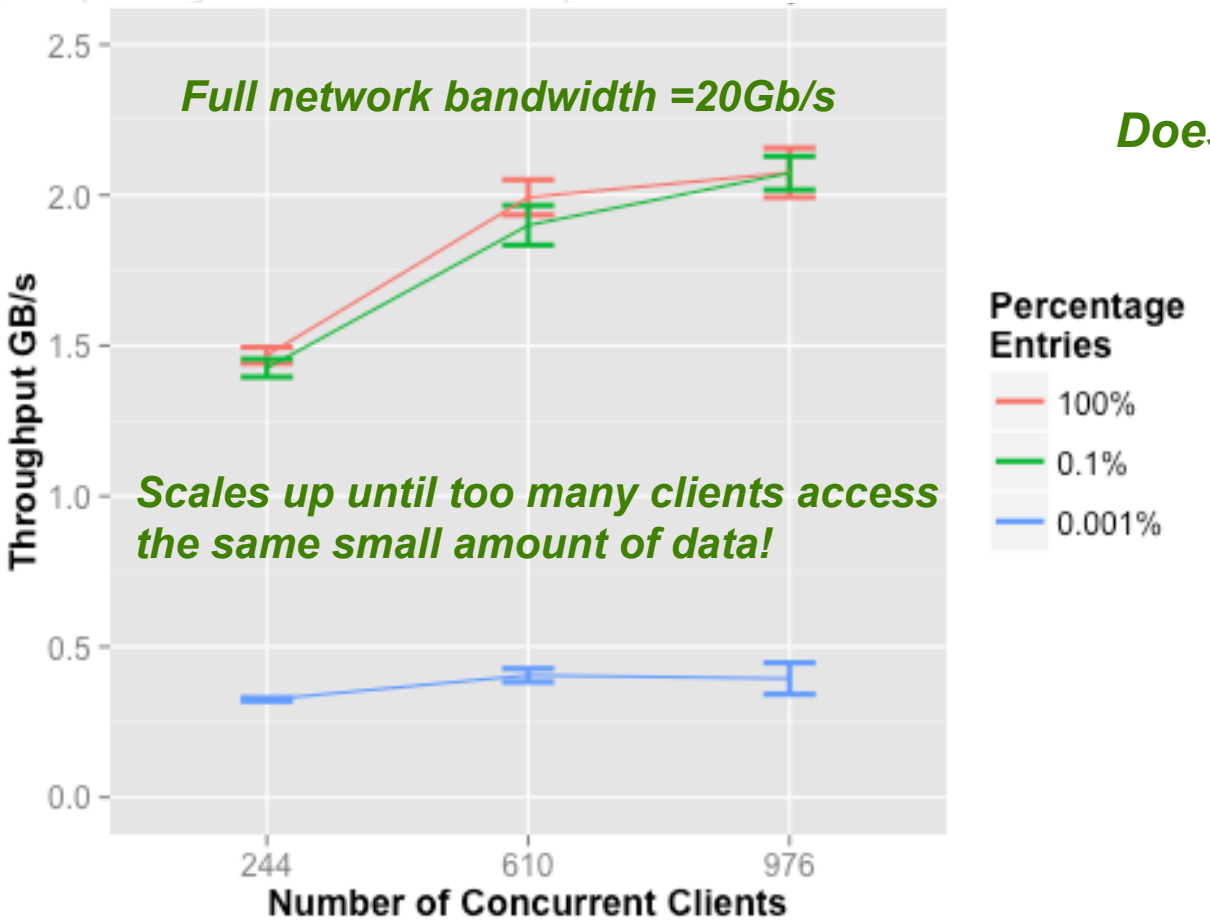
Total Entries = 11918
Branches = 5860
Size = 760 MB

% Entries	Total Entries	Read Calls	MB read
100%	11918	23	758.42
0.1%	12	12	320.92
0.001%	1	5	42.08

MULTI-CLIENT ROOT BENCHMARKS
1 Bucket

Old UDS

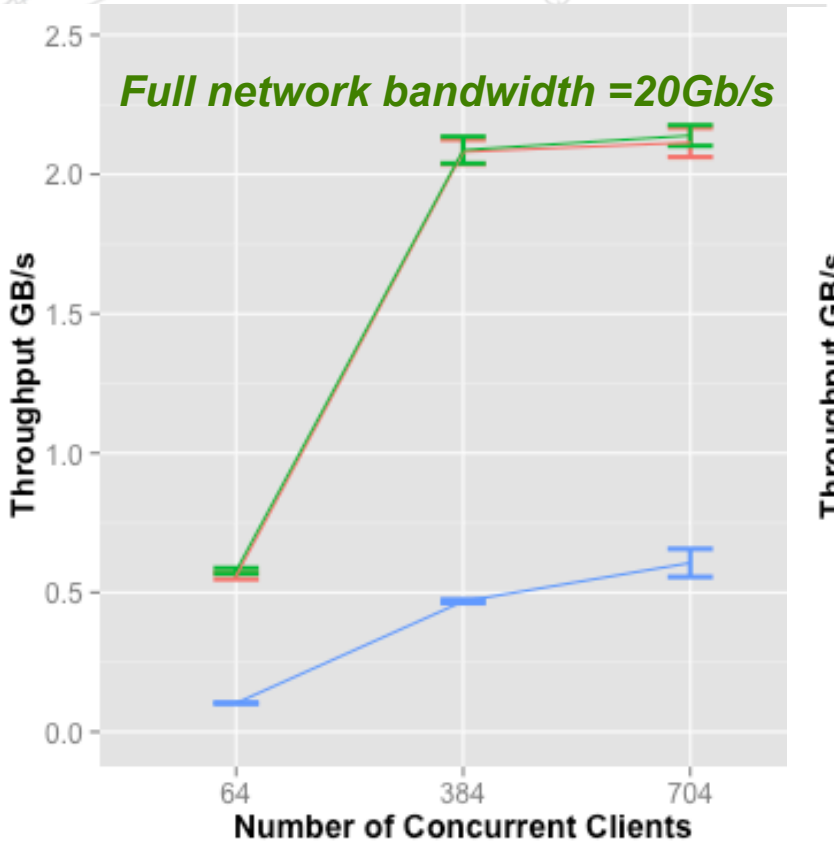
New UDS



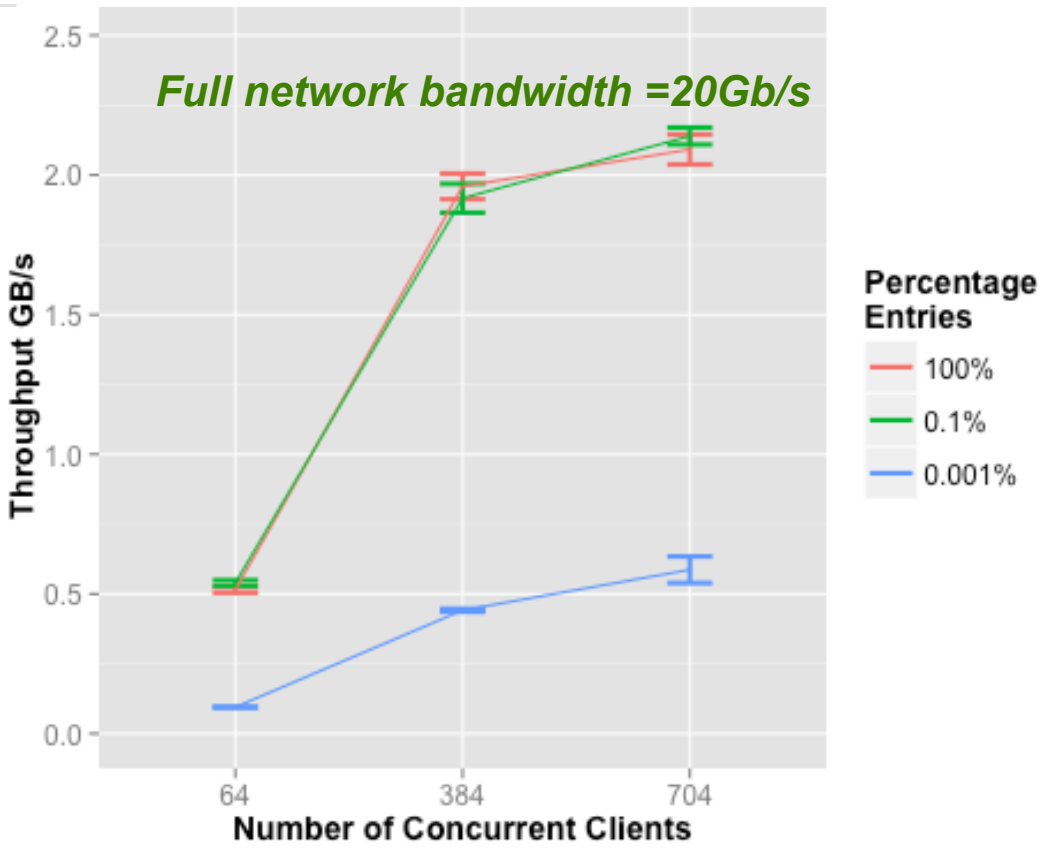
Does not scale as expected

MULTI-CLIENT ROOT BENCHMARKS
64 Buckets

Old UDS



New UDS



MULTI-CLIENT ROOT BENCHMARKS Summary

- **Full bandwidth is reached by executing sparse accesses with both UDS generations**
- **UDS supports multi-range get requests which are common in HEP analysis**
  - E.g. Ceph and Amazon S3 do not support multi-range requests

REAL CernVMFS APPLICATION Introduction

What is CernVMFS

- 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

CernVMFS back-end challenges

- Publishing new software should be as fast as possible
- Files are accessed through the HTTP

REAL CernVMFS APPLICATION S3 Backend

- **New features to CernVMFS (release 2.1.20)**
  - S3 compatible storage
  - Supports multiple buckets and accounts
- **LHCb nightly builds test case**
  - LHCb software (binary and source files)
  - Releases done daily for over 3 months
  - ~1 million files in release (~150k new files)
  - Completed successfully with UDSs and Ceph

REAL CernVMFS APPLICATION Deletion Process

- **Deletions should not affect the cloud storage functionality**
  - Old UDS does not handle deletions correctly
- **New UDS is fast enough to keep up with average LHCb nightly builds repository size growth**
  - Old versions can be deleted to make space for new ones

SUMMARY

- Scalable throughput and metadata performance verified
- Successful scalability with different ROOT access patterns
- LHCb nightly builds software stack distributed with CernVMFS using UDS backend

THANK YOU