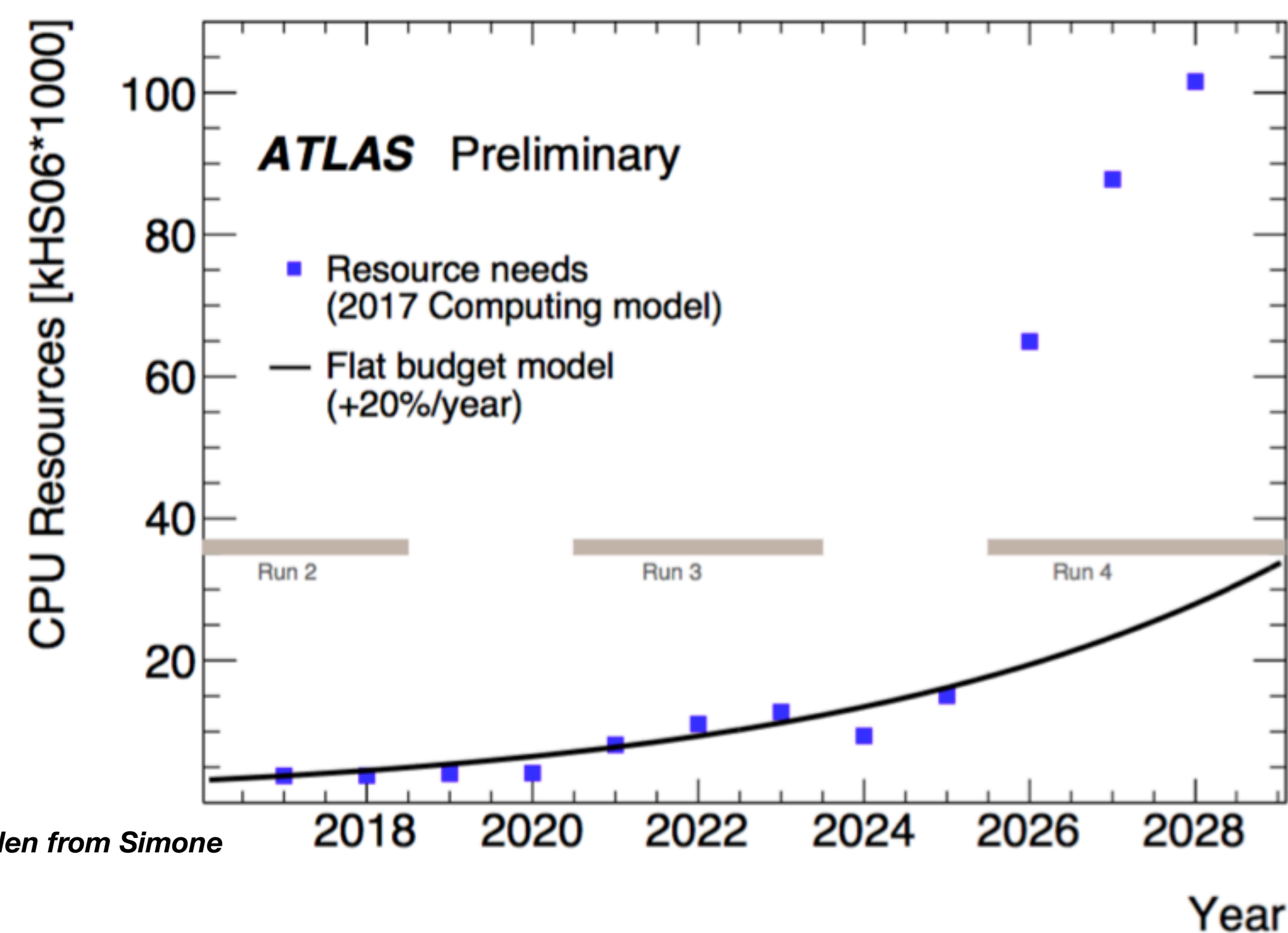
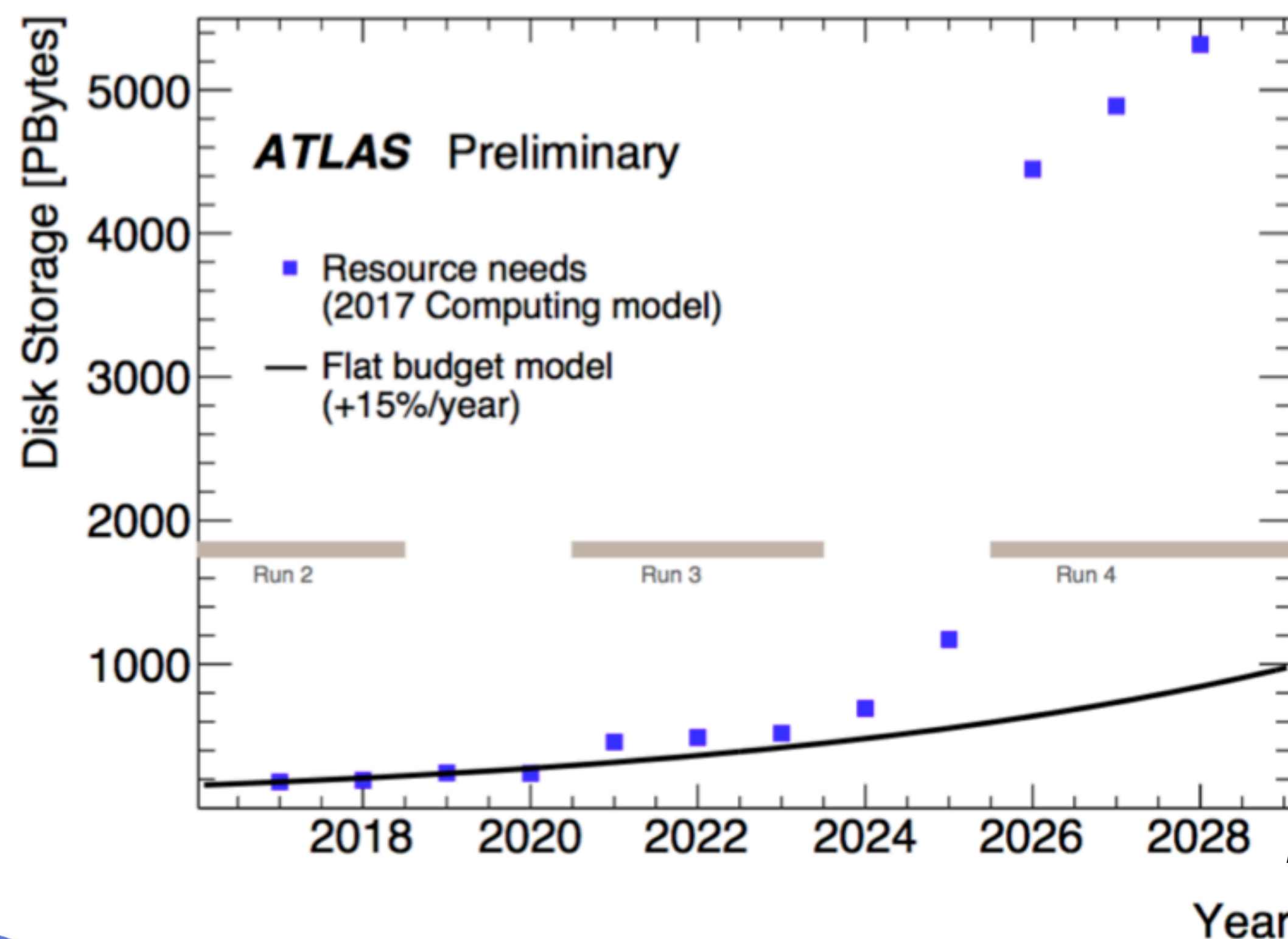


# Data Lake R&D: high level goals

Xavier Espinal  
on behalf of the WLCG Data Lake R&D group

# A Data Lake - why?

- HL-LHC storage needs are above the expected technology evolution (15%/yr) and funding (flat).
- We need to optimize storage hardware usage and operational costs.

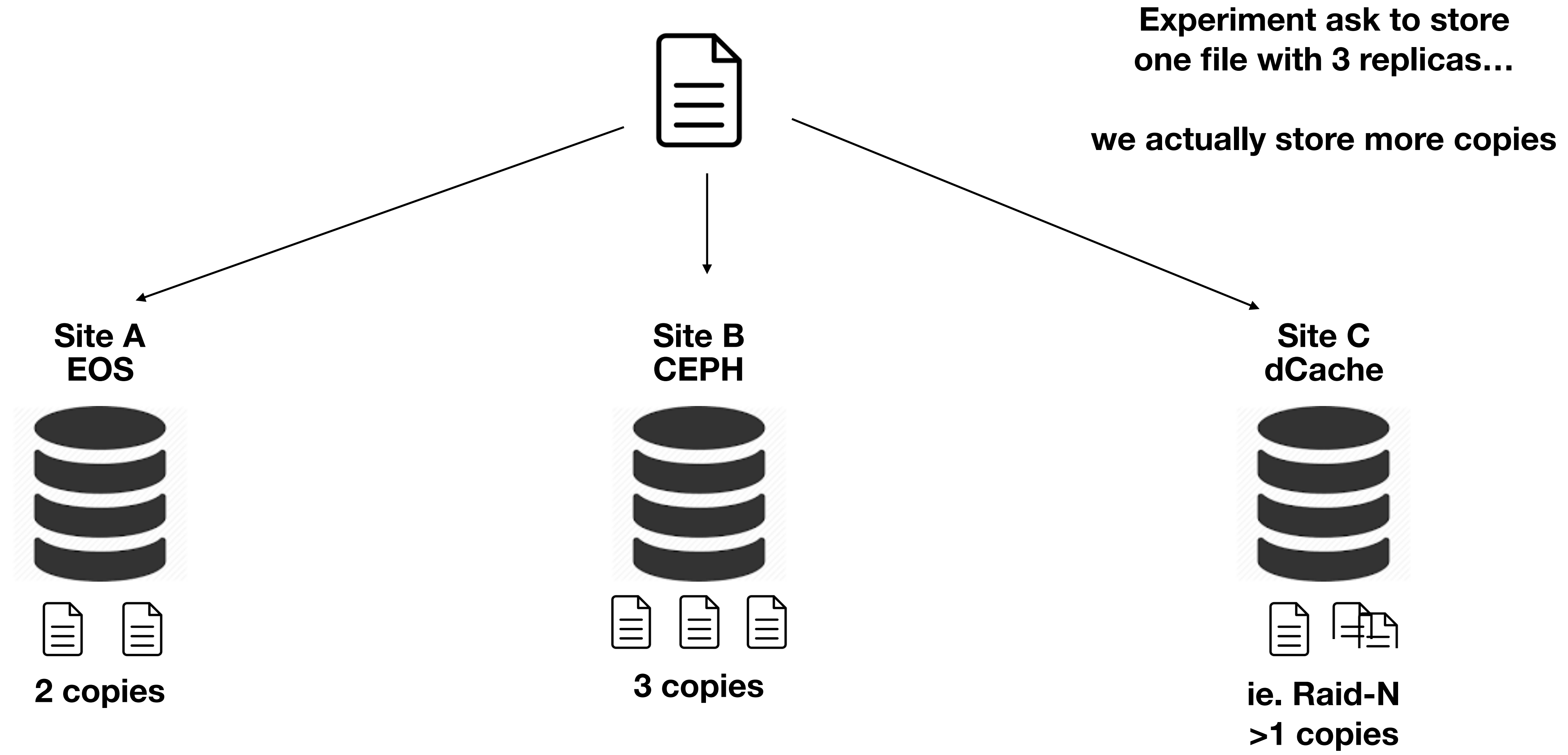


# Data Lake - how ?

- Explore distributed storage **evolution** to improve overall costs (storage and ops):
  - Common namespace and interoperability
  - Co-existence of different QoS (storage media)
  - Geo-awareness
  - File transitioning based on namespace rules
  - File layout flexibility
  - *Distributed* redundancy
  
- This can be achieved by a variety of storage technologies.

# Data Lake - how ? (2)

## Distributed redundancy challenge



# Data Lake - how ? (3)

  Disk Storage System with arbitrary QoS






 Tape Storage

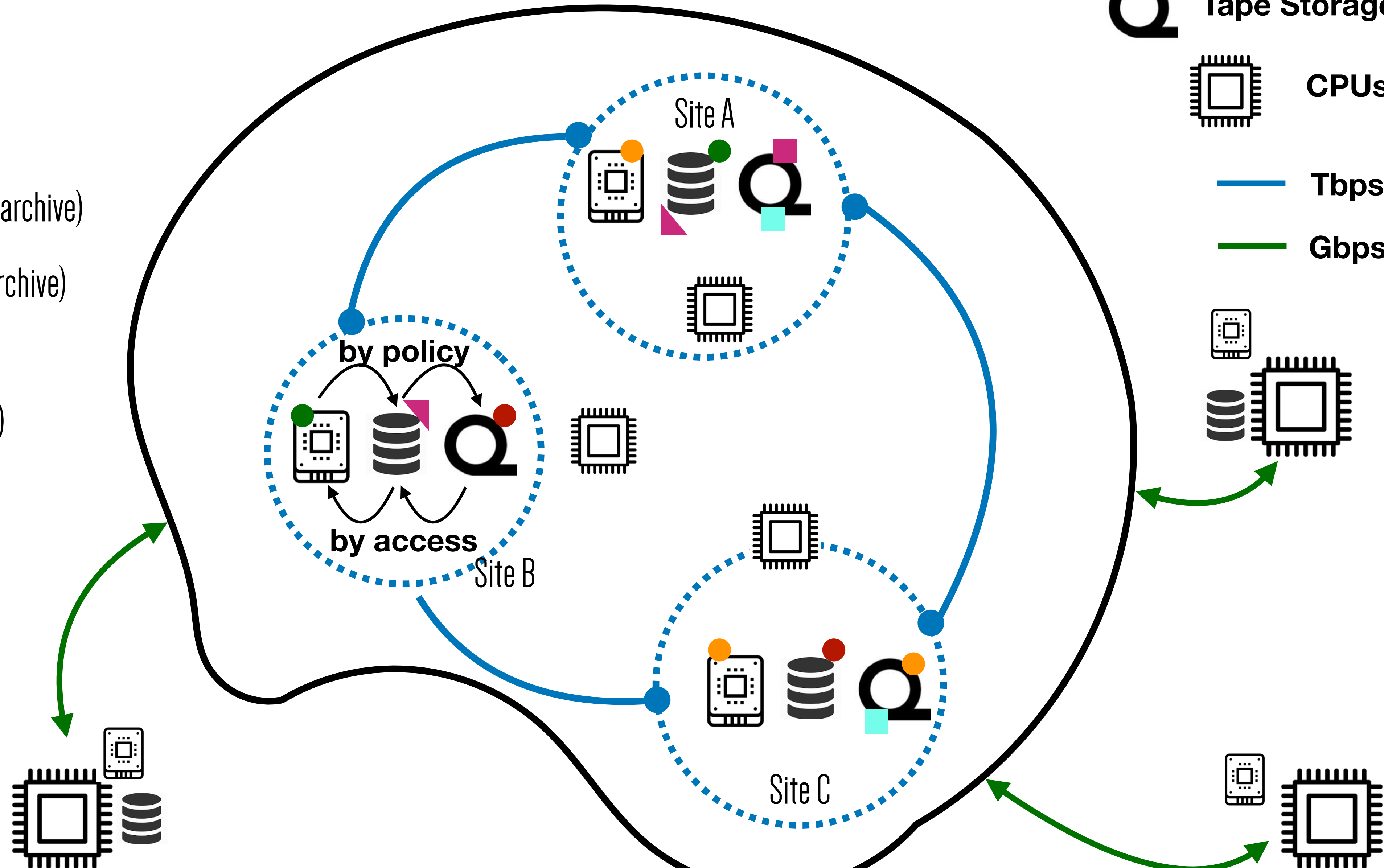
 CPUs

 Tbps

 Gbps

## File placement by QoS

-  Hot custodial file (2 fast copies+archive)
-  Warm custodial file (disk copy+archive)
-  Cold custodial file (archive)
-  Hot ephemeral file (2 fast copies)
-  Warm ephemeral file ("Rain")



# Data Lake - new challenges:

- **Dynamic** storage: impact on performance, reliability, resilience and redundancy.
- **Networks** <sup>(even more)</sup> instrumental: distributed metadata, cpu/data detach, file workflows, latencies.
- New **security** model: reduce overheads, remain highly secure.
- Storage systems **inter-operability**.
- Computing models **revision** (within a lake and across lakes): file transitioning, caches usage, data retention.

# eulake R&D goals

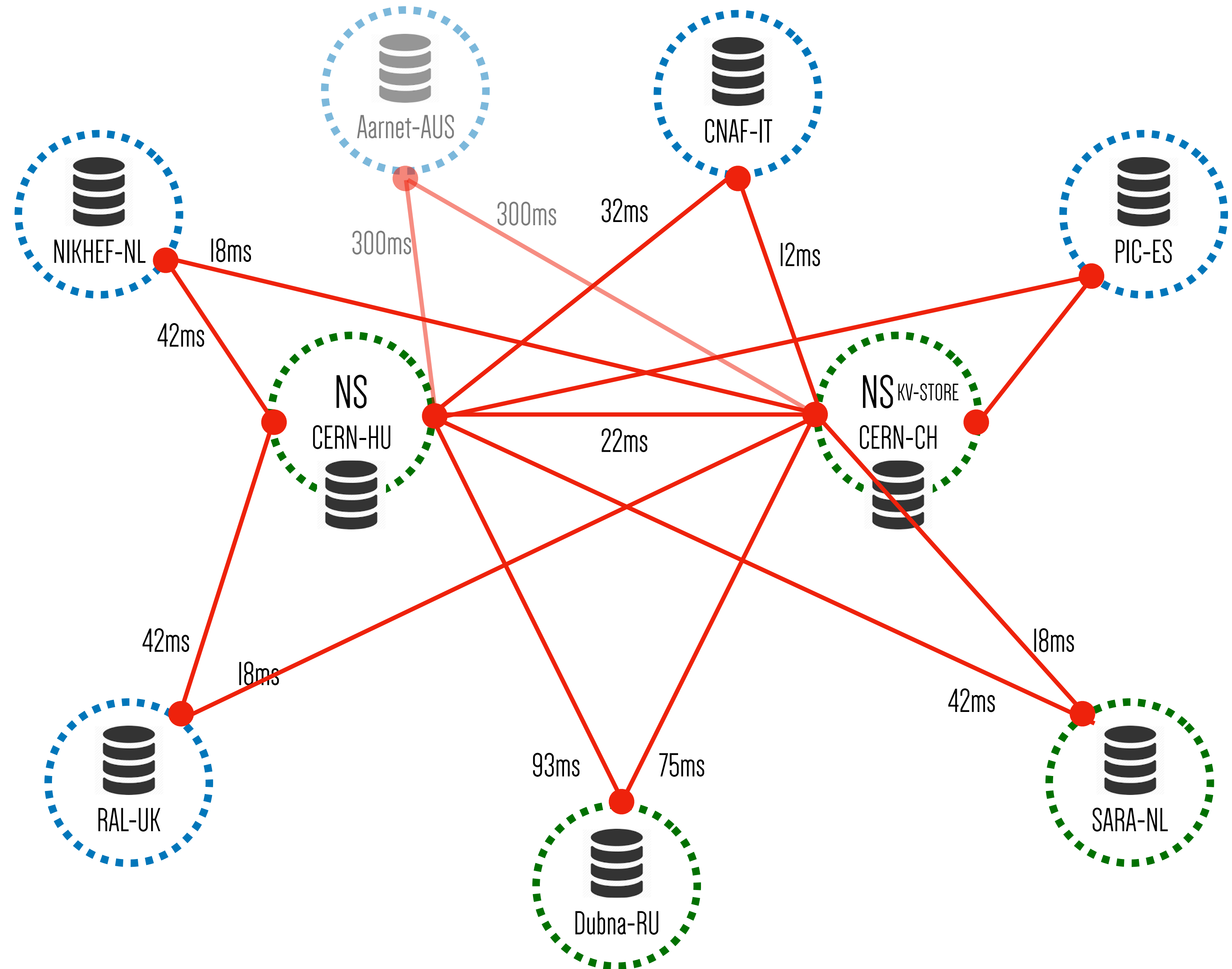
- Our R+D project aims to demonstrate that a dynamically distributed storage system with a common namespace:
  - Has the potential to **lower the cost** of stored data.
  - Has the potential to **ease** (local)**administration** and world-wide **operations**.
- The R+D should also demonstrate:
  - That the **efficiencies** in performance, reliability and resilience **are acceptable**.
  - The **compatibility** with HL-LHC **computing models**.

# eulake prototype

- To achieve the goals of this R&D we built a prototype.
- Our prototype is based on EOS (given our expertise). Parallel session to discuss various storage solutions and their interoperability.
- Distributed storage endpoints at SARA, NIKHEF, RAL, DUBNA, Kurchatov, PIC, CNAF and Aarnet.
- eulake storage endpoints running a simple EOS-fileserver daemon.
- Monitoring in place (grafana) and performance tests ready to be run in continuous mode (see Jarka's presentation)

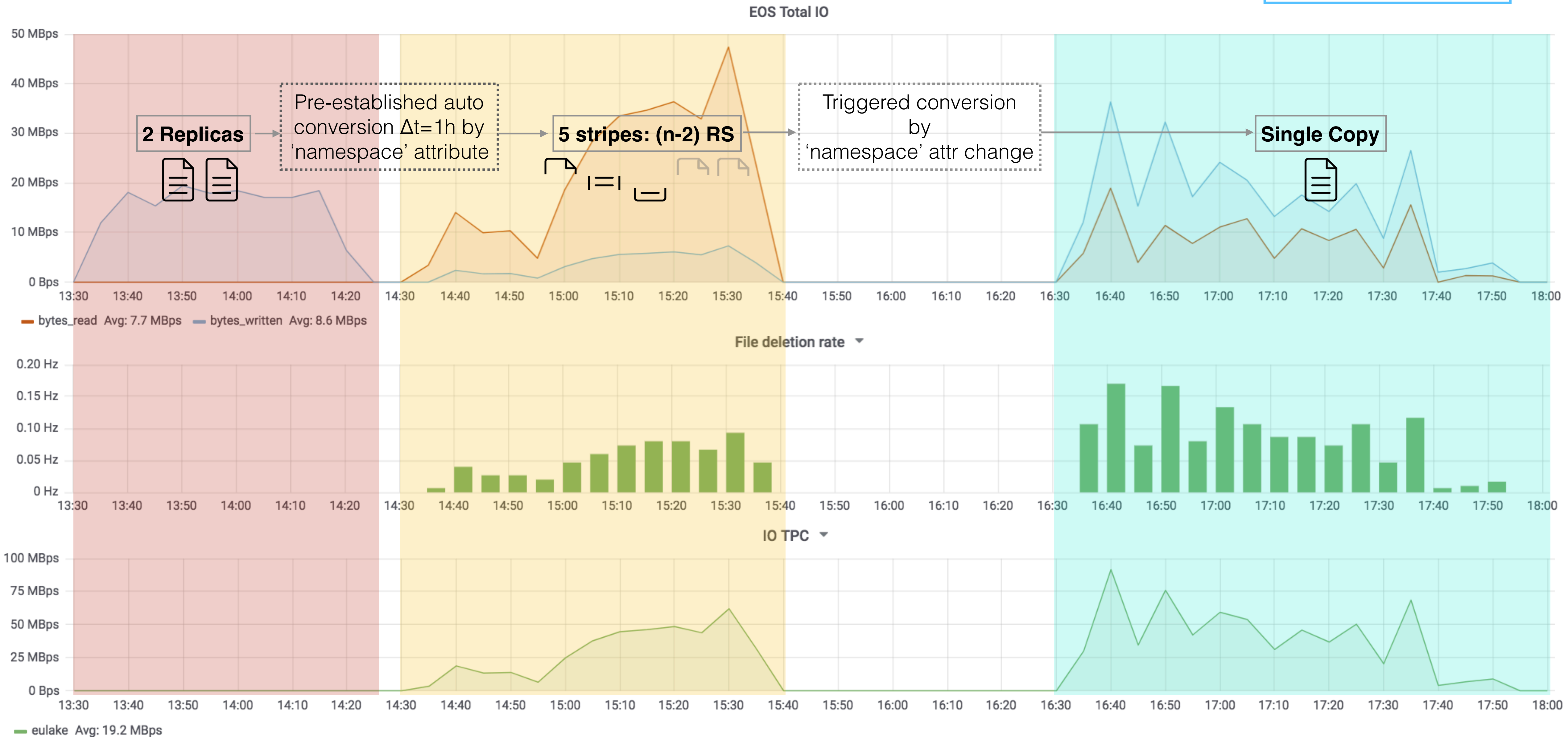


# eulake prototype (2)



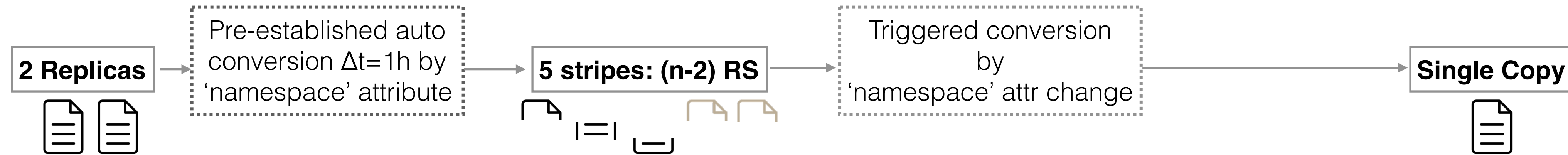
# eulake prototype: proof of concept :: file transition

Dataset: 100 files of 1GB  
Single client writing (VM)  
Conversion threads=2



# eulake prototype: proof of concept :: file transition

Dataset: 100 files of 1GB  
Single client writing (VM)  
Conversion threads=2



180315 14:04:36 func=open path=/eulake/lcg/test/conversion/2replicas-to-rain32/file-workflow-2r-rain32.175.file  
**op=write** target[0]=(p05799459m56401.cern.ch,33) target[1]=(p05798818t49625.cern.ch,80)

180315 15:04:58 time=1521123718.328306 func=open path=/eulake/lcg/test/conversion/2replicas-to-rain32/file-workflow-2r-rain32.175.file  
**op=read** target[0]=(p05799459m56401.cern.ch,33) target[1]=(p05798818t49625.cern.ch,80)

180315 15:04:58 func=open path=/eos/eulake/proc/conversion/0000000000001819:default#20640442  
**op=write** eos.layout.nstripes=5&eos.layout.type=raid6  
target[0]=(fst2.grid.surfsara.nl,130) target[1]=(p05496644k62259.cern.ch,1) target[2]=(dvl-mb01.jinr.ru,122) target[3]=(p05798818t49625.cern.ch,97)  
target[4]=(fst1.grid.surfsara.nl,124)

180315 17:22:17 func=open path=/eulake/lcg/test/conversion/2replicas-to-rain32/file-workflow-2r-rain32.175.file  
**op=read** target[0]=(fst2.grid.surfsara.nl,130) target[1]=(p05496644k62259.cern.ch,1) target[2]=(dvl-mb01.jinr.ru,122)  
target[3]=(p05798818t49625.cern.ch,97)

180315 17:22:17 func=open path=/eos/eulake/proc/conversion/00000000000018e2:default#00100001  
**op=write** eos.layout.nstripes=1&eos.layout.type=plain tpc.stage=copy redirection=p05799459m56401.cern.ch?

# Summary and outlook

- To be able to store the estimated HL-LHC data volume we need to improve storage costs at all levels: usage, administration and operations.
- To achieve this **we**\* need to: be creative, be flexible and be open.
  - \* all of us: wlcg, sites, storage providers, experiments and users.
- WLCG eulake R&D project to evaluate the potential of a new distributed storage solution.
- Initial phase started with few volunteering sites involved in a small prototype to proof the concept and evaluate viability of the data lake approach.