Architecture and prototype of a WLCG data lake for HL-LHC

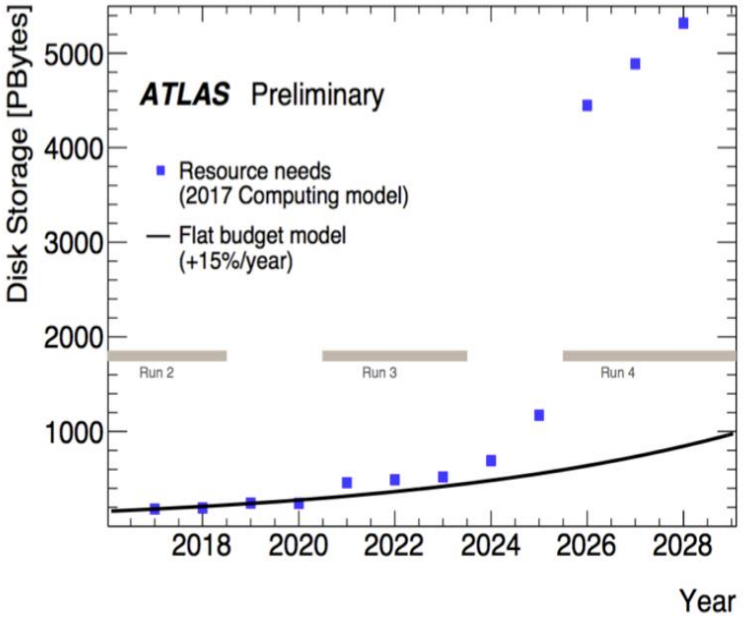
Gavin McCance, Ian Bird, Jaroslava Schovancova, Maria Girone, Simone Campana, Xavier Espinal Currul

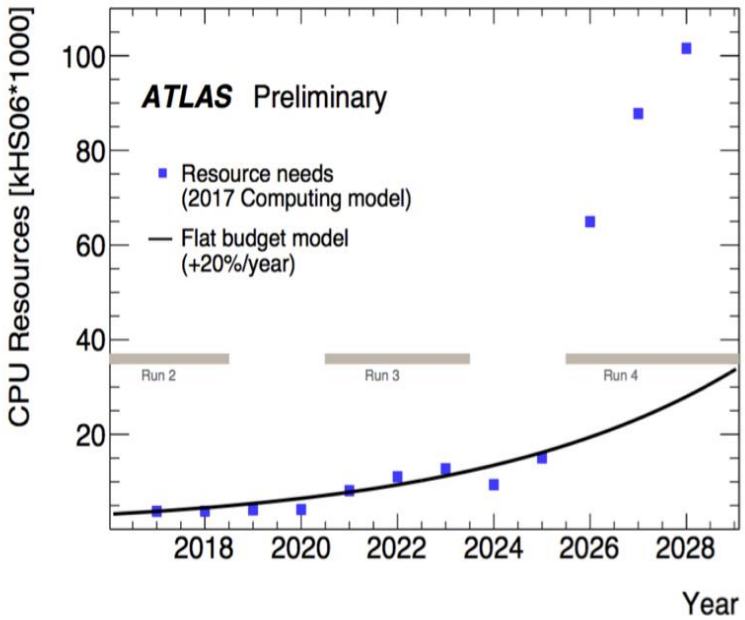
(CERN)

Motivation

- HL-LHC needs are above the expected technology evolution (15%/yr) and funding (flat)

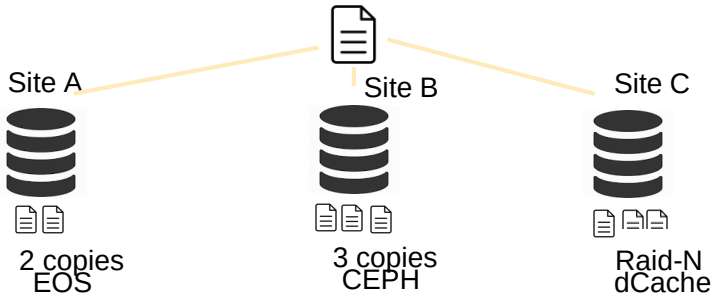
- We need to optimize hardware usage and operational costs





Some ideas for reducing cost

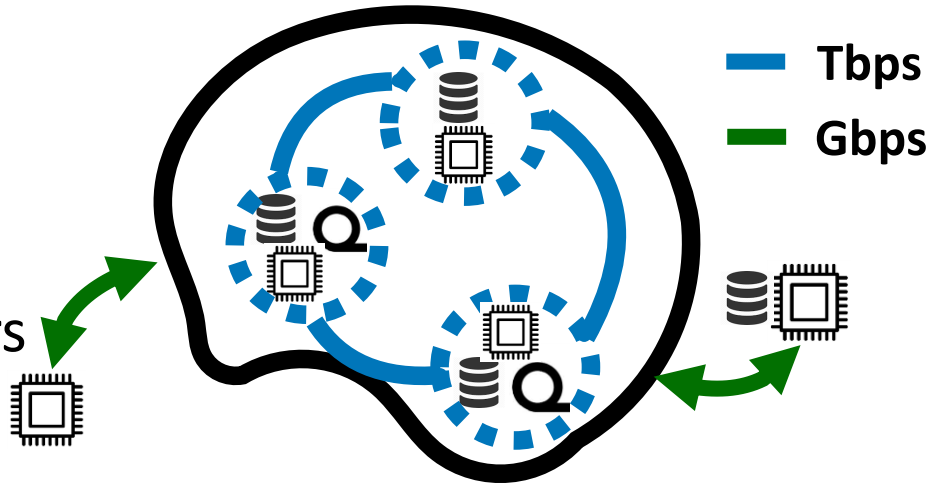
Reduce hardware cost:
introduce the concept of QoS
(Quality of Service)



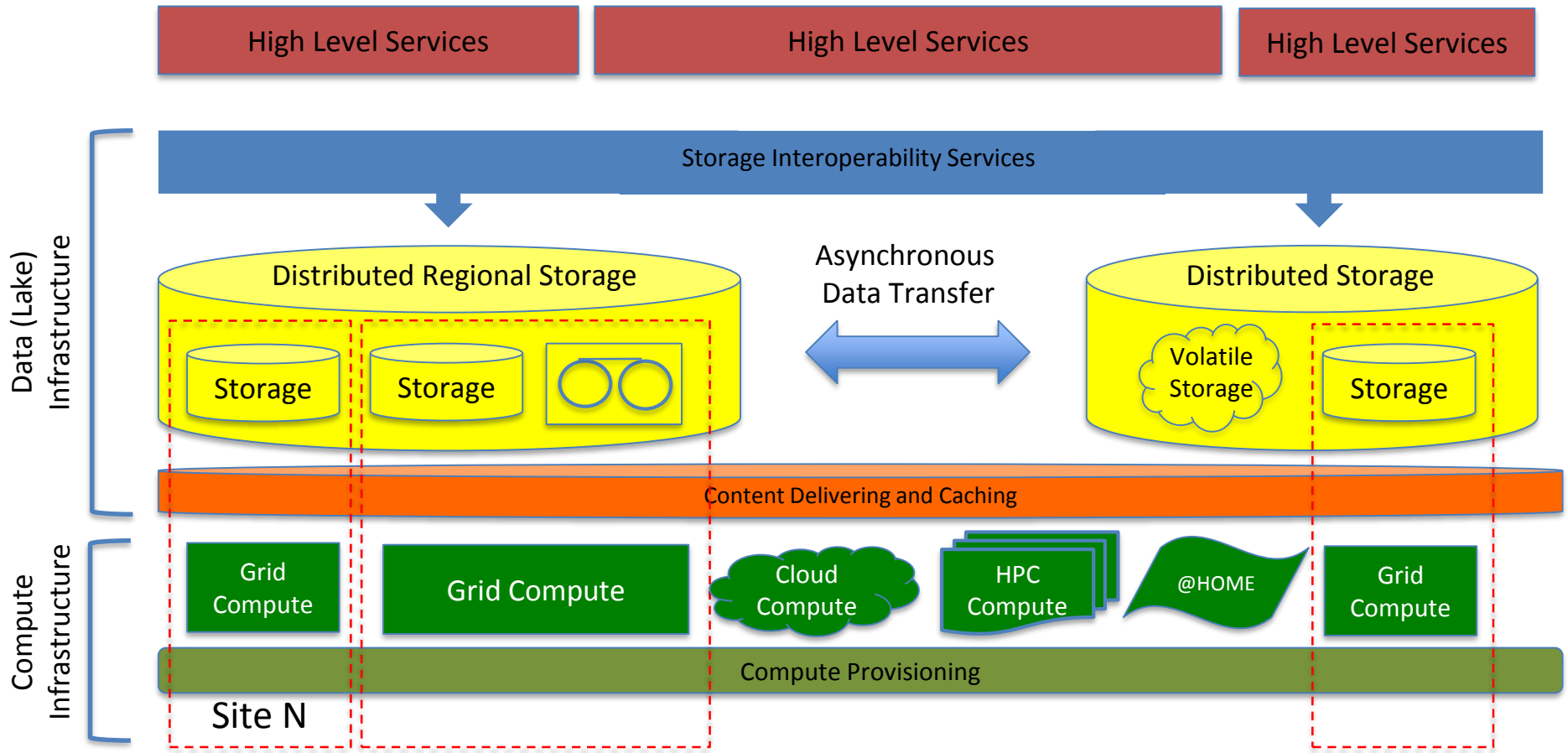
Today we store more than we think

Reduce Operational Cost:
deploy fewer (larger) storage
services.

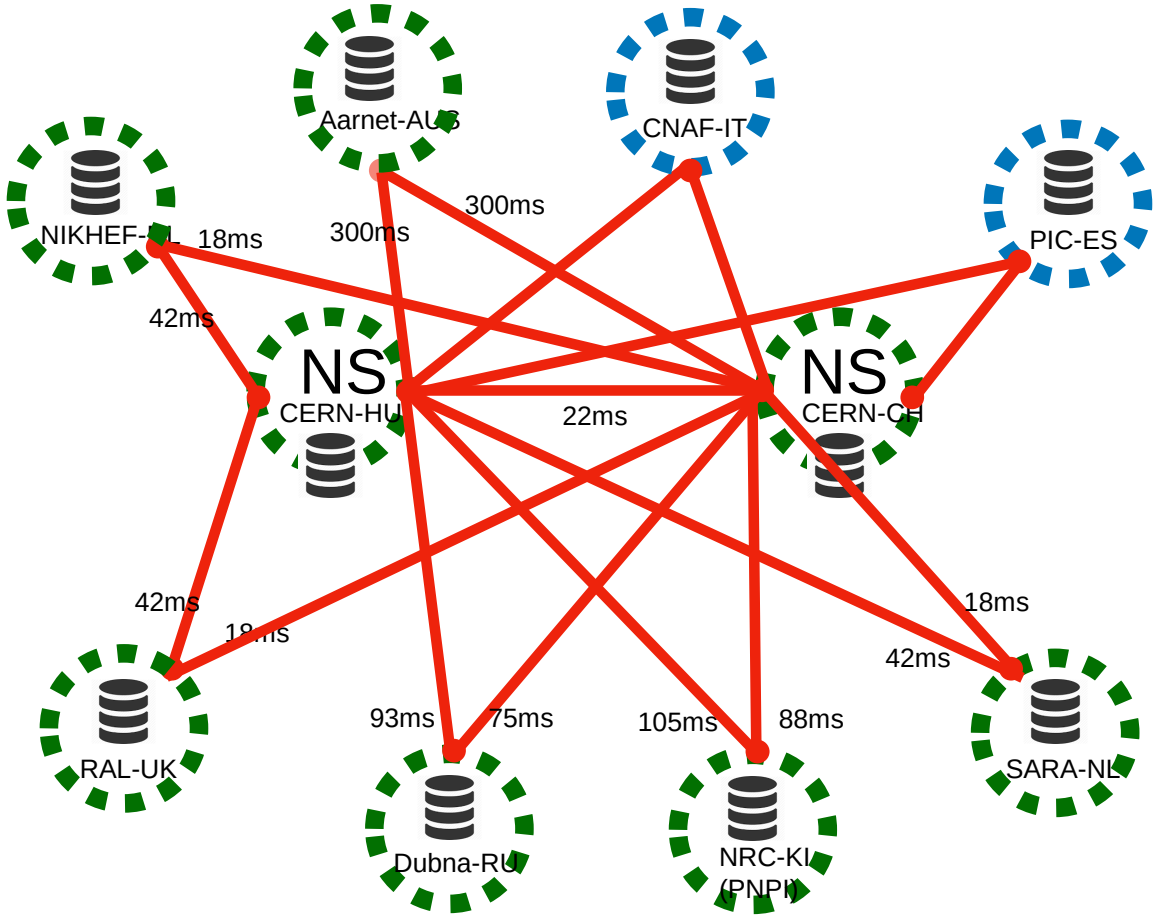
Co-location of data and processors
is not guaranteed



Data and Compute Infrastructures



The eulake prototype



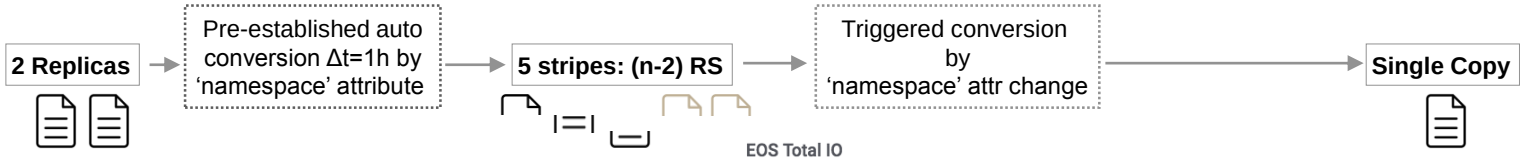
Goal: test and demonstrate some of those ideas

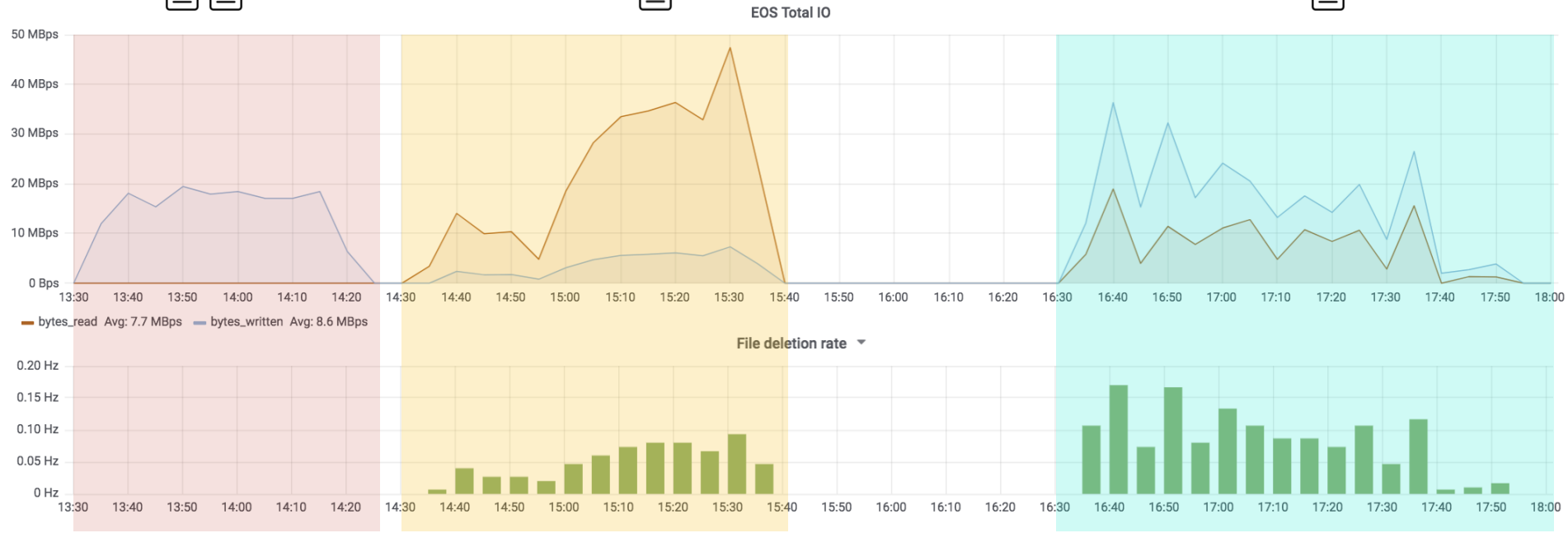
We deployed a Distributed Storage prototype

Based on the EOS technology

Quality of Service and Transitions

Three different QoS: 2 copies, 3+2 chunks, 1 copy. Automatic and triggered transition





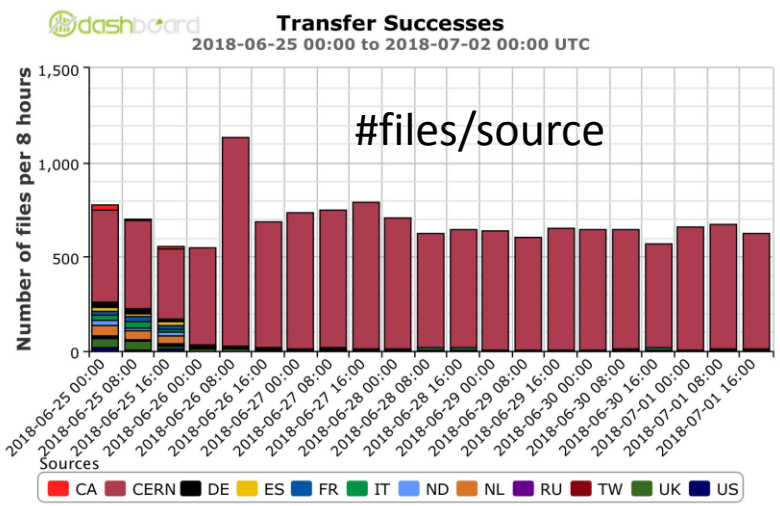
Dataset:100 files of 1GB - Single client writing (VM)

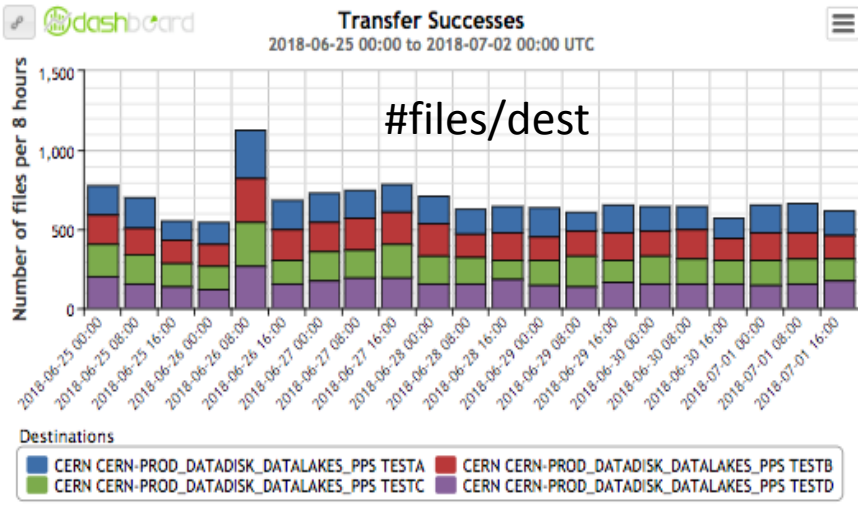
Integration of eulake with ATLAS Data Management

We expose eulake to the ATLAS data management system as storage endpoint

Data can be transferred from any ATLAS site into eulake

We imported 4 input samples in different eulake areas for the next tests

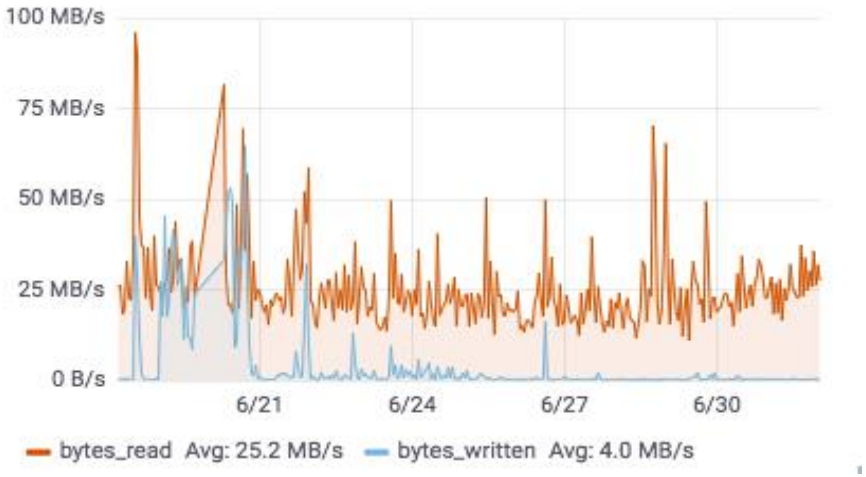




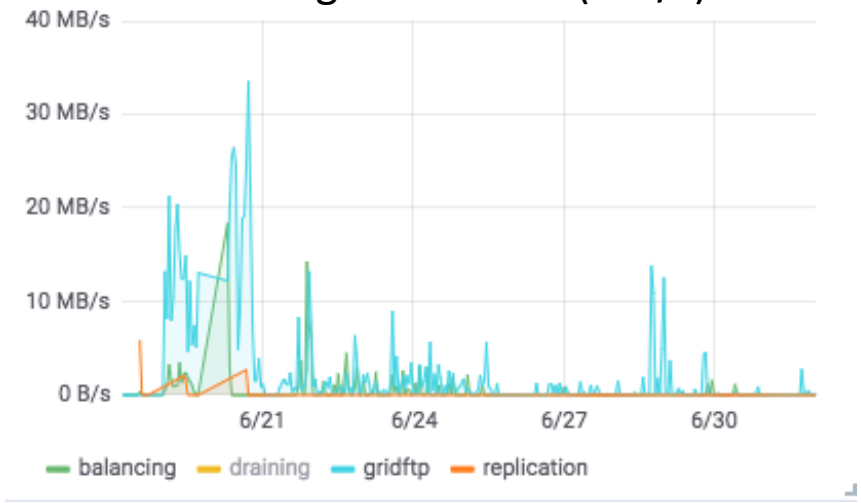
Activities

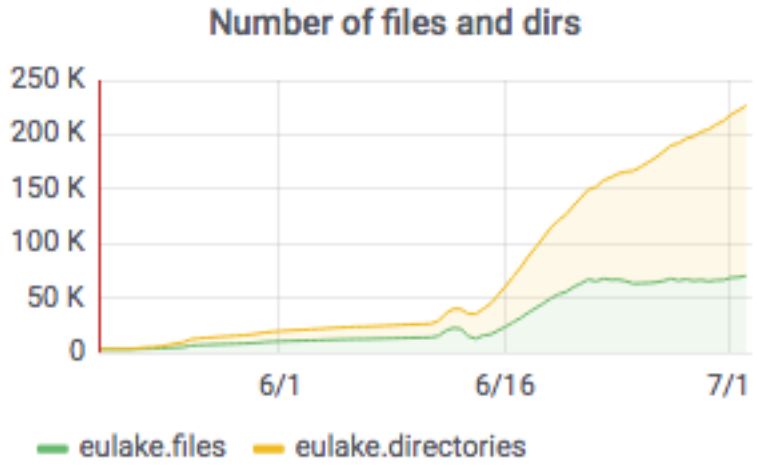
1. deployment and commissioning
2. transfer tests and input data replication
3. data access

eulake I/O (MB/s)



writing into eulake (MB/s)





HammerCloud

Running Tests backed by the WLCG Data Lake

State	Id	Host	Template	Start (Europe/Zurich)	End (Europe/Zurich)	Sites	subm jobs	run jobs	comp jobs	fail jobs	fail %	tot jobs
running	20116971	hammercloud-ai-12	1006: benchmark derivation AthDerivation/21.2.8.0 1k events - WLCG Data Lakes - local data clone.977 EULAKE folder CERN	01/Jul, 10:24	02/Jul, 11:52	CERN-PROD_DATALAKES, CERN-PROD_DATALAKES_MCORE, CERN-PROD_DATALAKES_TESTA, 3 more...	1	6	121	8	6	137
running	20116974	hammercloud-ai-12	1007: benchmark digi+reco derivation Athena/21.0.53 5 events - WLCG Data Lakes - local data clone.987 EULAKE folder CERN	01/Jul, 12:58	02/Jul, 13:50	CERN-PROD_DATALAKES, CERN-PROD_DATALAKES_MCORE, CERN-PROD_DATALAKES_TESTA, 3 more...	0	6	71	42	35	119
running	20116984	hammercloud-ai-12	1012: A.F.T. AtlasDerivation 20.7.6.4 clone.808 clone.845 EULAKE folder CERN	01/Jul, 19:02	02/Jul, 16:39	ANALY_CERN-PROD_DATALAKES, ANALY_CERN-PROD_DATALAKES_TESTA, ANALY_CERN-PROD_DATALAKES_TESTB, 2 more...	5	0	0	0	0	5
running	20116988	hammercloud-ai-12	1005: P.F.T. mc16 Sim_tf 21.0.16 - WLCG Data Lakes - local data clone.989 EULAKE folder CERN	02/Jul, 2:58	03/Jul, 5:01	CERN-PROD_DATALAKES, CERN-PROD_DATALAKES_MCORE, CERN-PROD_DATALAKES_TESTA, 3 more...	0	6	63	6	8	77

Running Tests backed by the standard storages, copy-to-scratch

State	Id	Host	Template	Start (Europe/Zurich)	End (Europe/Zurich)	Sites	subm jobs	run jobs	comp jobs	fail jobs	fail %	tot jobs
running	20116979	hammercloud-ai-12	977: benchmark derivation AthDerivation/21.2.8.0 1k events - WLCG Data Lakes - local data	01/Jul, 13:42	02/Jul, 14:21	CERN-PROD-preprod, NIKHEF-ELPROD, SARA-MATRIX, 6 more...	0	4	96	0	0	100
running	20116980	hammercloud-ai-12	987: benchmark digi+reco derivation Athena/21.0.53 5 events - WLCG Data Lakes - local data	01/Jul, 13:54	02/Jul, 12:20	CERN-PROD-preprod, NIKHEF-ELPROD, SARA-MATRIX, 6 more...	0	3	96	0	0	100
running	20116986	hammercloud-ai-73	845: AFT AtlasDerivation 20.7.6.4 clone.808	01/Jul, 21:42	02/Jul, 19:38	ANALY_AGLT2_SL6, ANALY_AGLT2_TEST_SL6-condor, ANALY_ARNES, 142 more...	180	237	6603	267	4	7339
running	20116991	hammercloud-ai-12	989: P.F.T. mc16 Sim_tf 21.0.16 - WLCG Data Lakes - local data	02/Jul, 8:04	03/Jul, 6:16	CERN-PROD-preprod, NIKHEF-ELPROD, SARA-MATRIX, 6 more...	0	2	8	0	0	10

We integrated eulake with the HammerCloud framework

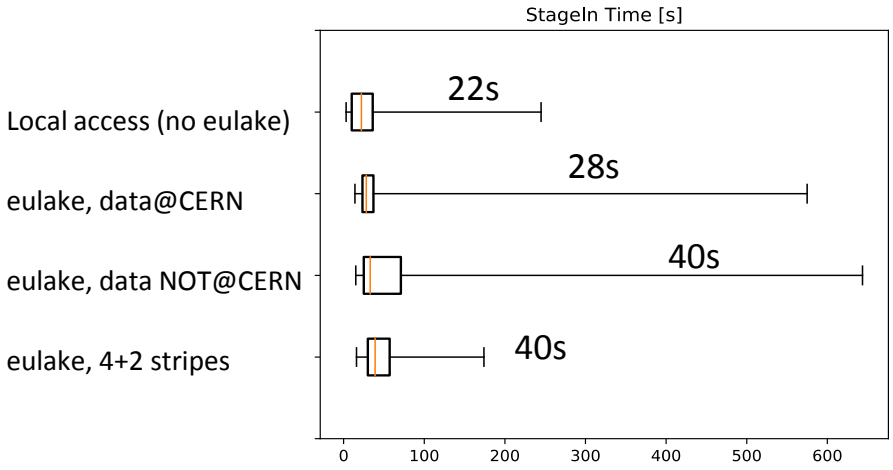
Allows test real workflows and data access patterns. Initial focus on ATLAS

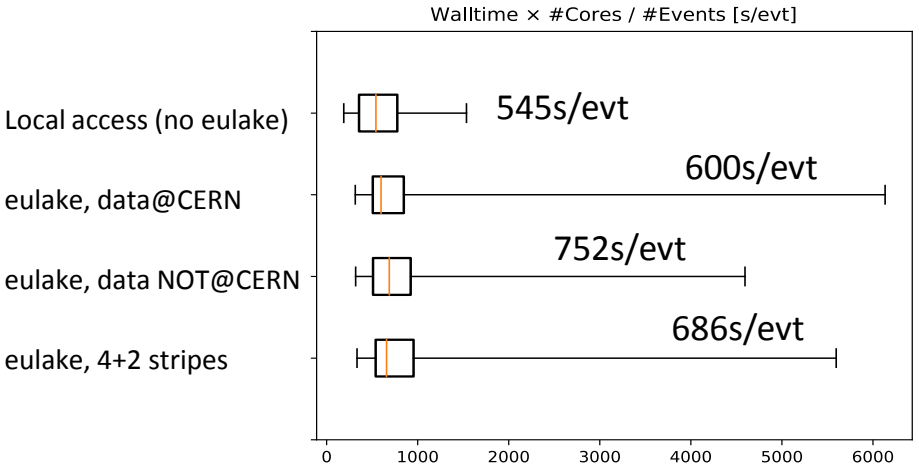
Four test scenarios. Read from:

1. Local access (no eulake)
2. eulake, data@CERN, WN@CERN
3. eulake, data NOT@CERN, WN@CERN
4. eulake, 4+2 stripes, WN@CERN

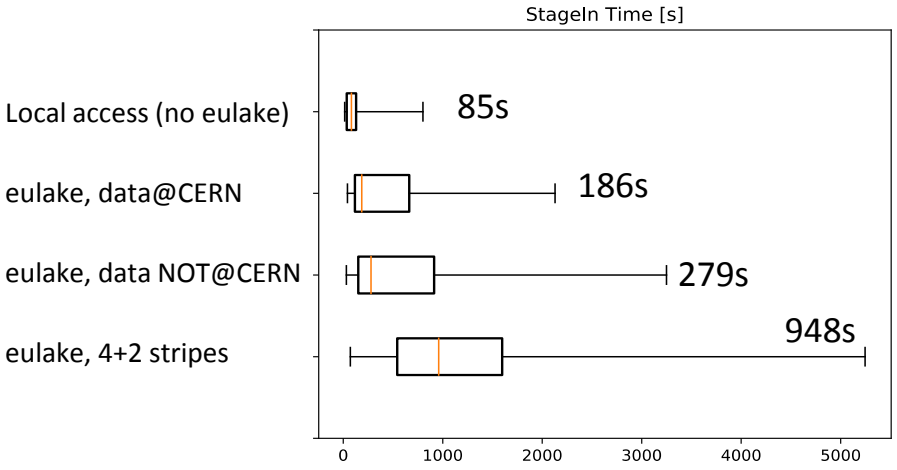
Data is copied from storage to WN

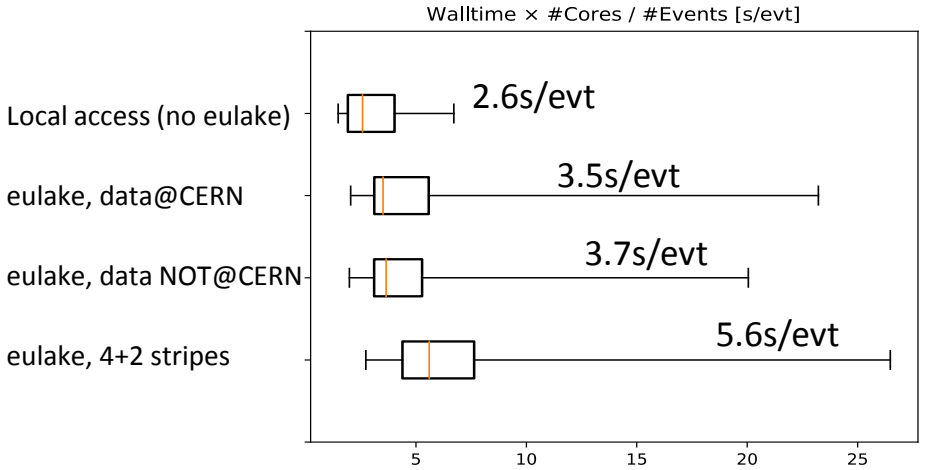
Low I/O intensity workflow: ~40MB input (1 file), 2 events, ~5 mins/event





High I/O intensity workflow: ~6GB input (1 file), 1000 events, ~2 seconds/event





Conclusions

- We set up a distributed storage instance based on EOS technology
  - Small in terms of storage volume, large in the geographical sense
  - Deployment varieties: Native EOS, EOS on Docker containers, volume export (CEPH)
- We integrated such instance with the ATLAS distributed computing services and HammerCloud
  - Next step is integration with CMS
- We have a prototype in place and understand what we want to measure
  - We don't yet have enough stats to draw any firm conclusions
- The prototype we set up serves to test many ideas in preparation for HL-LHC

Acknowledgements

- This work has been done in collaboration with many participant WLCG sites. Credit goes to their work
- We would like to thank the ATLAS Distributed Computing and particularly the Rucio team for the help in the integration
- We thank the EOS experts at CERN
- Special thanks to Ivan Kadochnikov for the help in the final part of this work