Distributed Scientific Computing

Xavier Espinal (CERN-IT)

CERN WLCG

Scientific computing goes distributed ~2000

- LHC brought unprecedented computing needs, too much for a single datacenter
  - Change of paradigm: scientific computing goes distributed
- Big challenge. Years of preparation to design and set-up the Worldwide LHC Computing Grid (aka **the grid**)
  - Common auth/authz, data transfers across sites with different technologies, workload management systems (grid vs. batch systems), networking, operations model, etc.
- Big success: in full production since the start of LHC enabling science worldwide

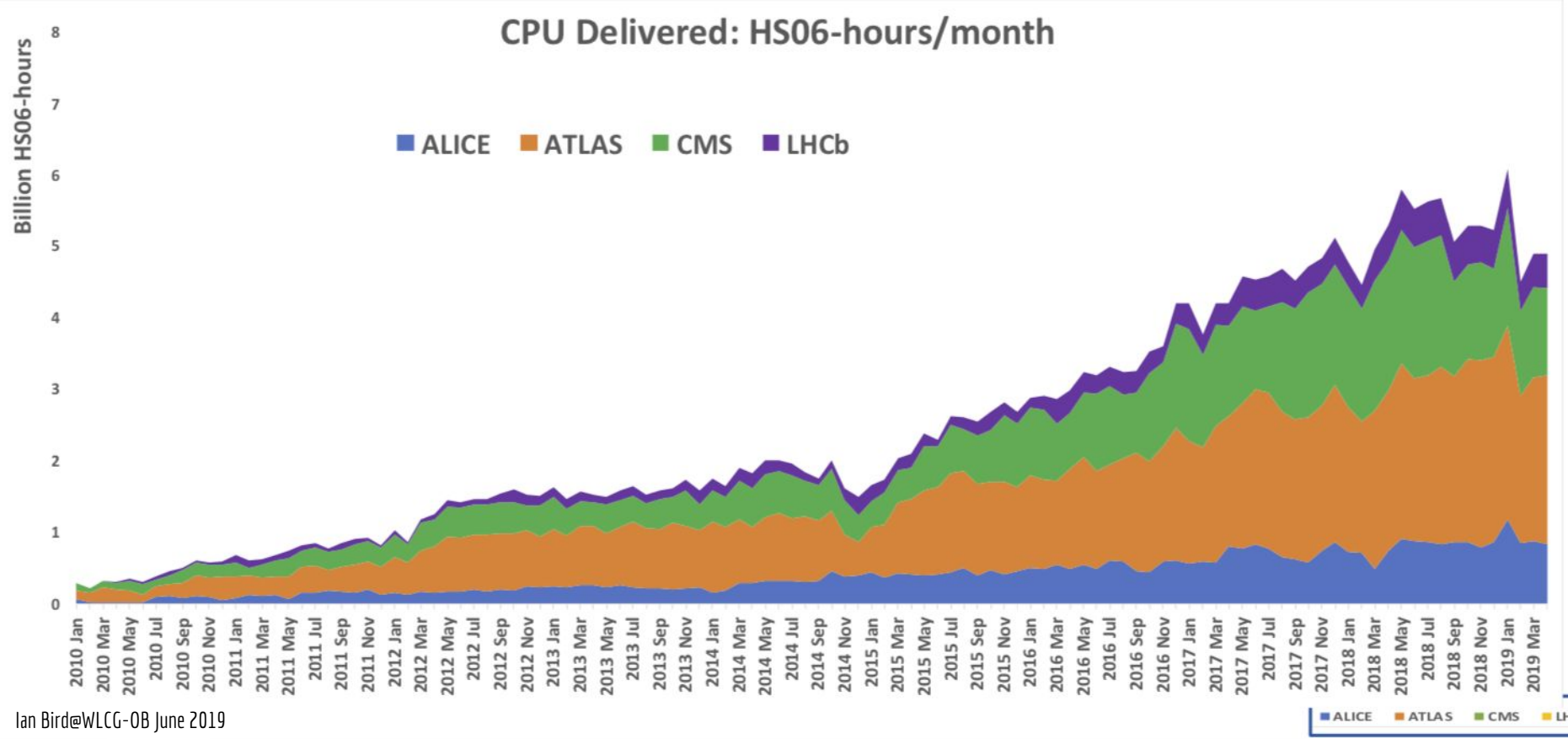
CERN WLCG

WLCG: the Worldwide LHC Computing Grid now

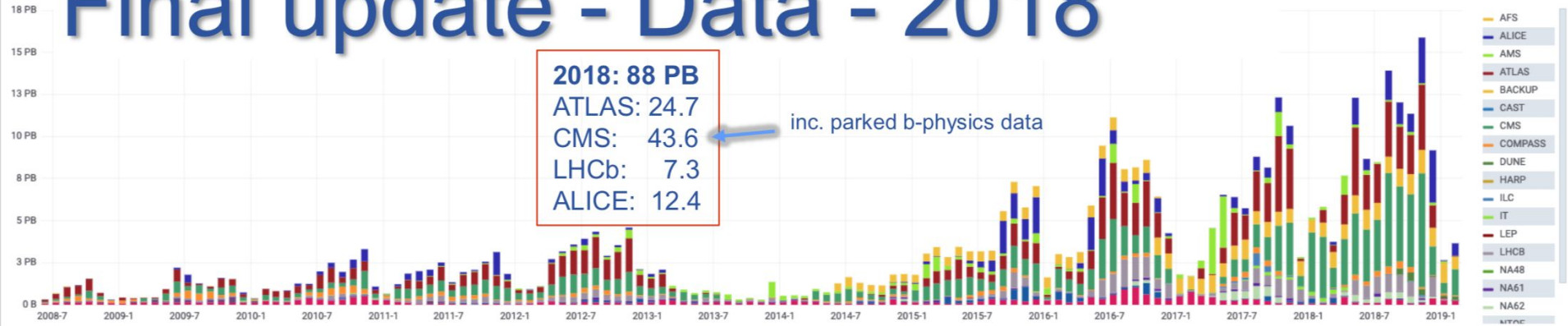
- 170+ centers in 42 countries
  - Interoperability with architectural ”freedom of choice” for the sites and common “middleware”
- 4 LHC experiments with a community of 12000+ scientists
- 900,000 cores and 1EB of storage (disk and tape)
  - CERN provides about 20% of the WLCG resources
- Global transfer rates exceeding 60 GB/s, EB scale annual global data flow

CERN WLCG

WLCG: data processing



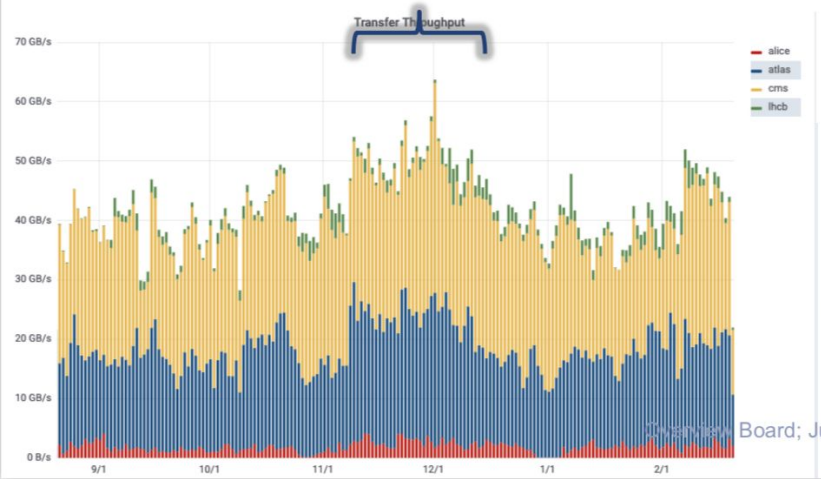
Final update - Data - 2018



Ian Bird@WLCG-OB June 2019

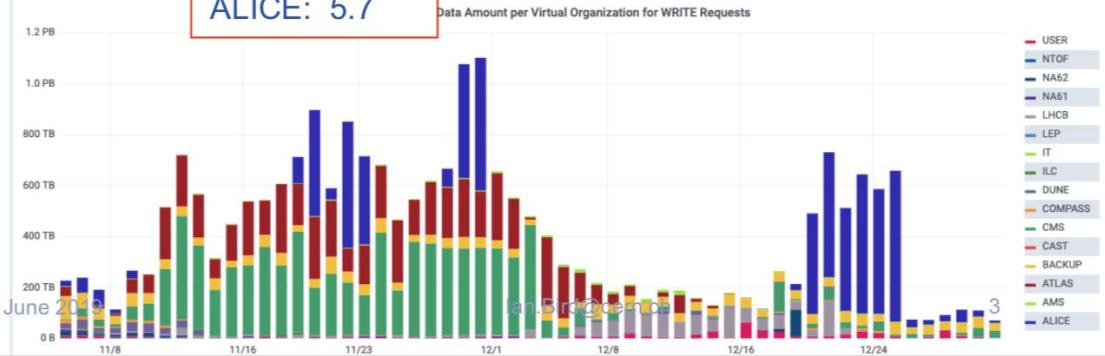
Data transfers

HI Run

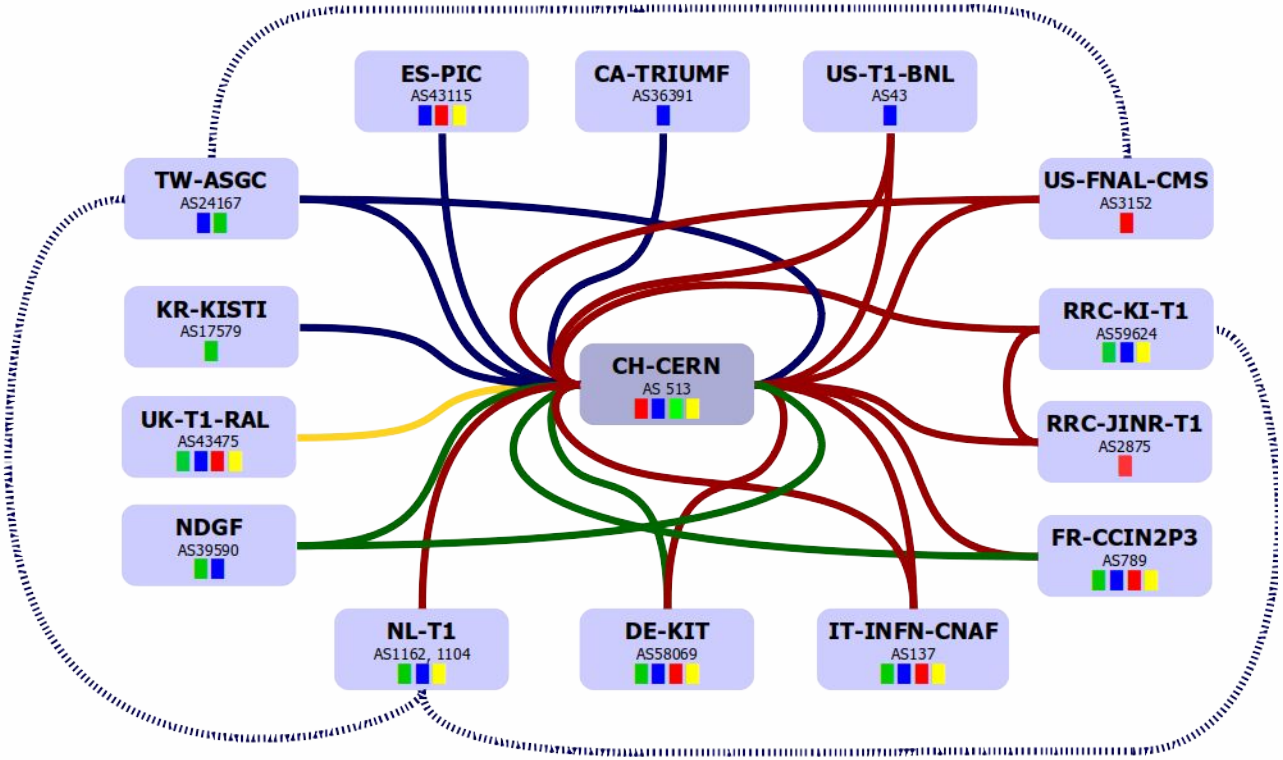


2018: 19.8 PB
ATLAS: 5.2
CMS: 7.7
LHCb: 1.2
ALICE: 5.7

HI Run



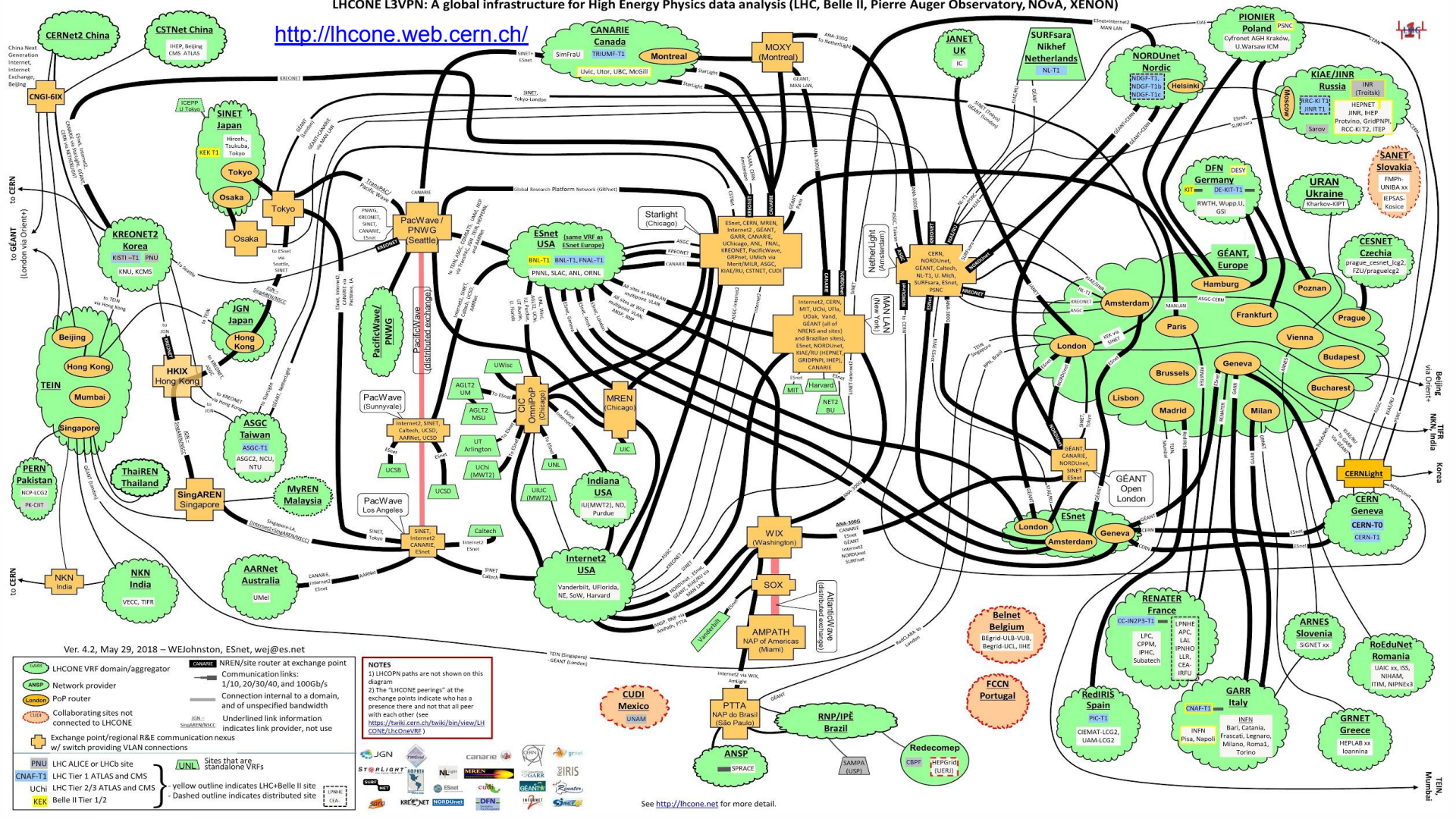
LHCOPN



http://lhcopn.web.cern.ch/lhcopn/

LHCONE L3VPN: A global infrastructure for High Energy Physics data analysis (LHC, Belle II, Pierre Auger Observatory, NOvA, XENON)

http://lhcone.web.cern.ch/



WLCG, 10 years of operations

- Several solutions arose to address same problems (DM, WM). Converging with time.
- Constant quest for optimization in computing resources and operations (manpower)
  - From sites to infrastructure to experiments
- Several tools seen as strong building blocks in future scientific computing:
  - FTS, RUCIO, token-based AAI, etc.

CERN WLCG

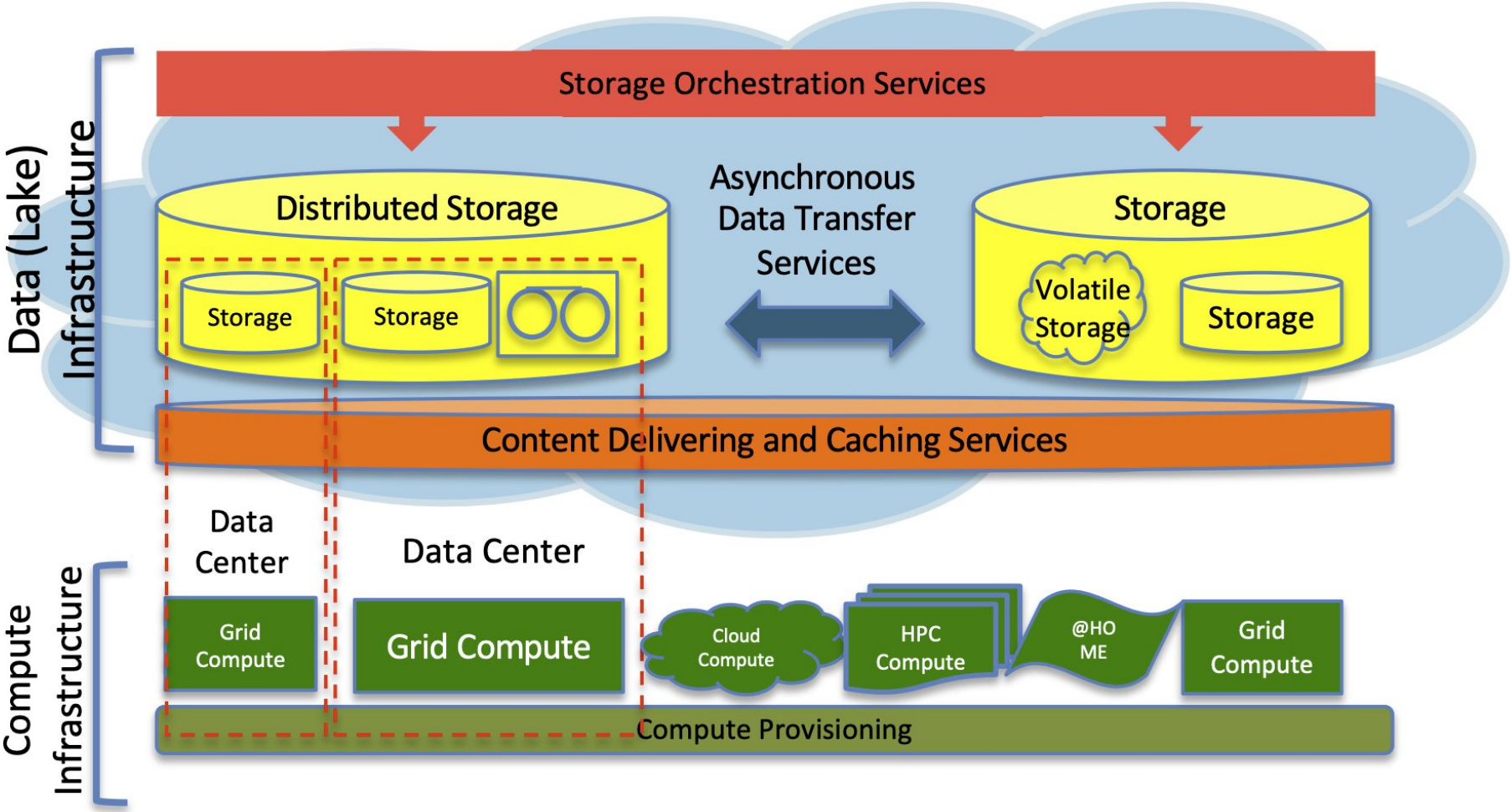
Future in Scientific Computing

- HL-LHC requirements (2025+) pose (again) a challenge in terms of data volumes and processing capacity
  - technology evolution and funding not helping
- Experiments other than LHC/HEP will be facing similar challenges
  - astrophysics, cosmology, GW and neutrino physics
- Global interest for collaboration: common concepts, tools and infrastructures for future scientific computing
  - AAI, data management, workload management, data transfers, protocols, etc.

CERN WLCG

ESCAPE

- Open Science challenge shared by 31 partners including
  - ESFRI facilities: CTA, ELT, EST, FAIR, HL-LHC, KM3NeT, SKA
  - Pan-European research infrastructures: CERN, ESO, JIV-ERIC, EGO-Virgo
- Aiming at delivering solutions to ensure integration of data, tools, services and scientific software
- Started: 1/2/2019 (end date 31/7/2022)


Data (Lake) Infrastructure
Storage Orchestration Services
Distributed Storage
Storage
Storage
Asynchronous Data Transfer Services
Storage
Volatile Storage
Storage
Content Delivering and Caching Services
Data Center
Data Center
Compute Infrastructure
Grid Compute
Grid Compute
Cloud Compute
HPC Compute
@HOME
Grid Compute
Compute Provisioning

Summary

- WLCG instrumental for LHC success
- Facing a change of scale in the computing requirements for scientific experiments
- Future in Scientific Computing towards collaboration between sciences and scientific communities
- Time for intense R+D in different areas
  - Datalakes, caching, AAI, HPC, protocols, networking, software optimization, etc.
  - Common infrastructures