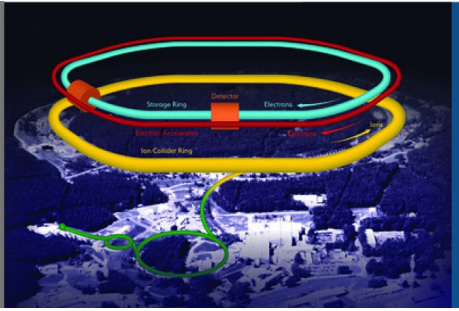


The 2022 CFNS Summer School on the Physics of the Electron Ion Collider

Computing Models

Liz Sexton-Kennedy, Fermilab
Stony Brook University, 14th July 2022

CERN

Fermilab

Ian.Bird@cern.ch, sexton@fnal.gov

Acknowledgement

- Input to this talk has been drawn from many discussions with colleagues from many HEP and closely related experiments over recent years, in particular
  - Preparation of the Community White Paper, and WLCG* strategy documents
  - WLCG and HSF-WLCG 2019 workshops

- LHC = Large Hadron Collider
- WLCG = World-wide LHC Computing Grid
- HEP = High Energy Physics

Why Talk About Computing Models?

- Computing has become as important as operating a detector in terms of delivering physics insight.
  - Why? – complexity and scale

HEP experiments?

- Why are they worth talking about to an EIC audience?
- Evolution of need
  - jLab computing is where Fermilab was late in the 2nd run of the tevatron collider
    - Local data collection with onsite storage, some MonteCarlo production on the grid or cloud
  - It is likely that EIC needs will become comparable to current LHC needs
- WLCG computing was a significant step in organization and change of computing models

Worldwide LHC Computing

WLCG combines about 1.4 million computer cores and 1.5 extabytes of storage from over 170 sites in 42 countries, producing a massive distributed computing infrastructure that provides more than 12 000 physicists around the world with near real-time access to LHC data, and the power to process it.

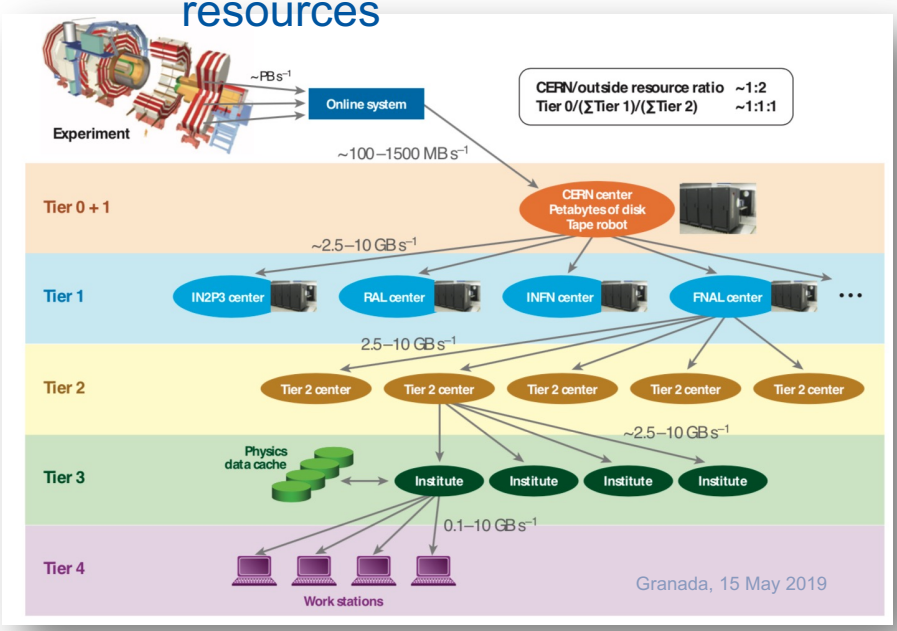
It runs over 2 million tasks per day and, at the end of the LHC's LS2, global transfer rates exceeded 260 GB/s.

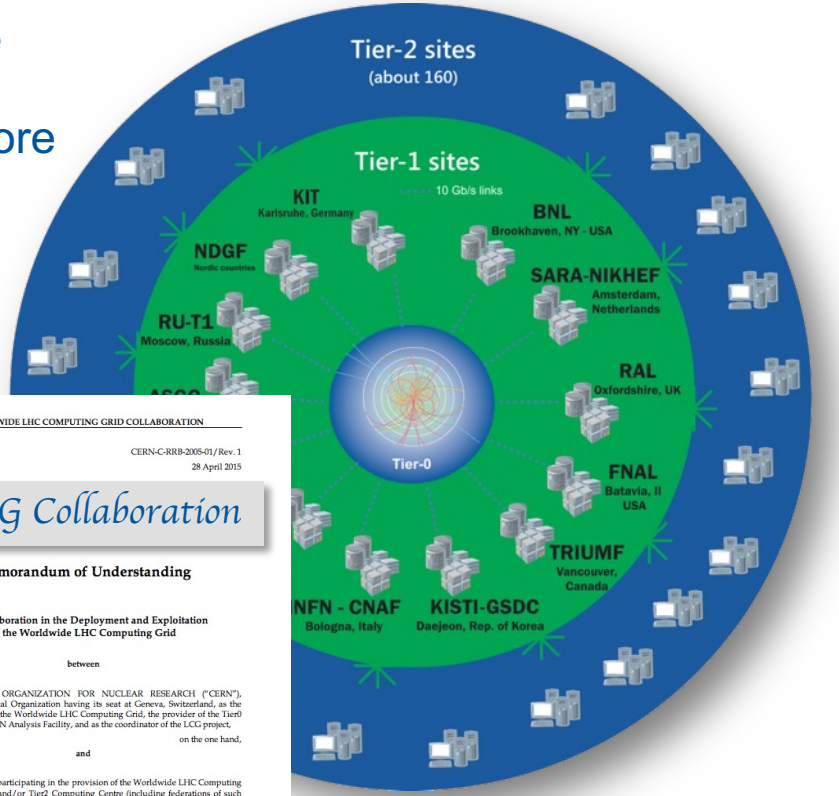
These numbers will increase as time goes on and as computing resources and new technologies become ever more available across the world.

https://videos.cern.ch/record/2640380

Today's computing models

- HEP has done distributed computing since the early 2000's
- Scale of the challenge for LHC forced a more organized and formal structure
- Built a federated system – "grid" – to integrate and make easily usable pledged resources





WORLDWIDE LHC COMPUTING GRID COLLABORATION

CERN-C-RRB-2005-01/Rev. 1
28 April 2015

WLCG Collaboration

Memorandum of Understanding

for Collaboration in the Deployment and Exploitation of the Worldwide LHC Computing Grid

between

The EUROPEAN ORGANIZATION FOR NUCLEAR RESEARCH ("CERN"), an intergovernmental Organization having its seat at Geneva, Switzerland, as the Host Laboratory of the Worldwide LHC Computing Grid, the provider of the Tier0 Centre and the CERN Analysis Facility, and as the coordinator of the LCG project,

on the one hand,

and

all the Institutions participating in the provision of the Worldwide LHC Computing Grid with a Tier1 and/or Tier2 Computing Centre (including federations of such Institutions with computer centres that together form a Tier1 or Tier2 Centre), as the case may be, represented by their Funding Agencies for the purposes of signature of this Memorandum of Understanding,

on the other hand,

(hereafter collectively referred to as "the Parties").

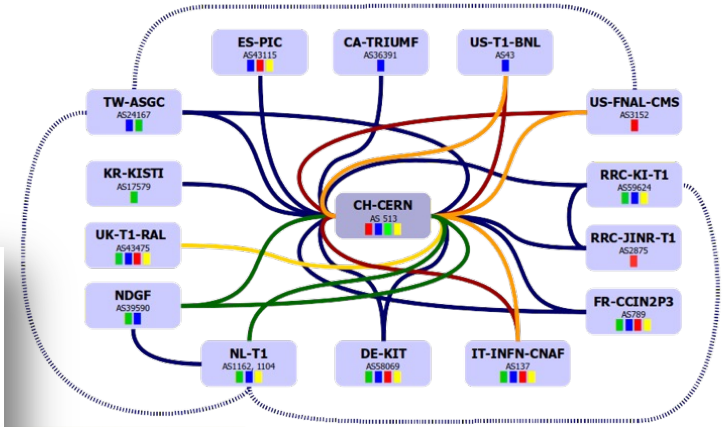
HEP Experiments Evolution

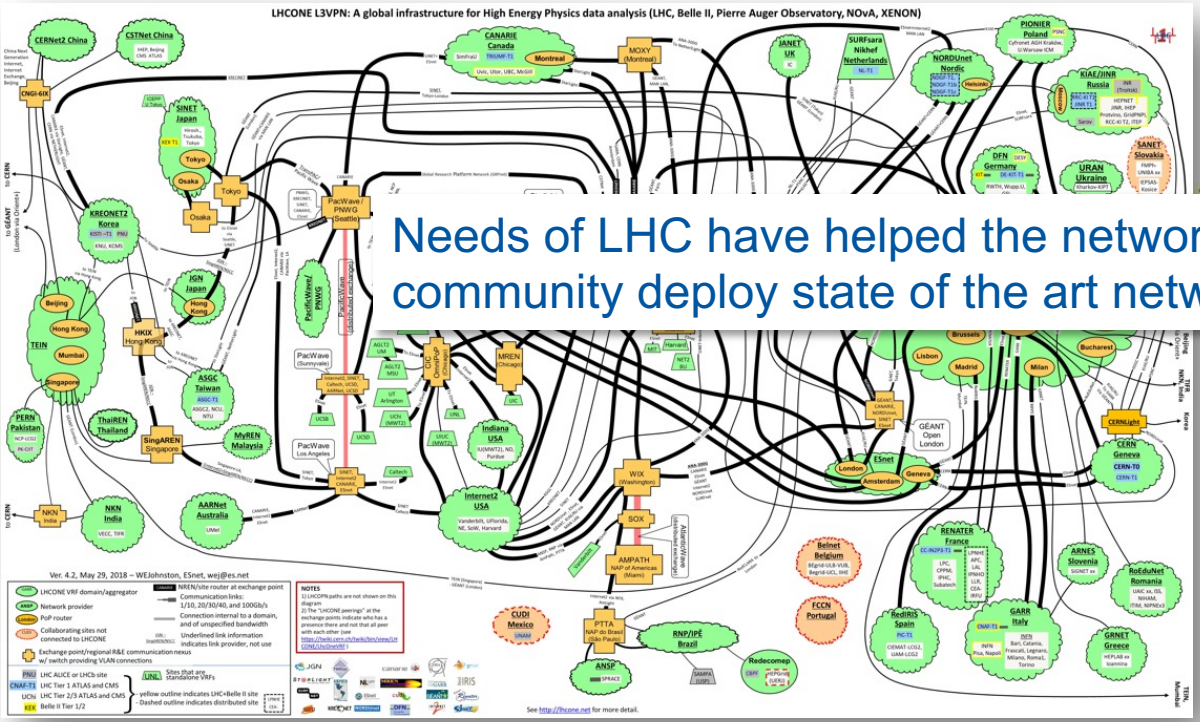
- Major features and capabilities of todays HEP computing infrastructures:
  - Networks – international and national, private and public
  - Data Organization Management and Access – key to success, data transfers, storage systems, data management tools and data organization
  - Compute – hardware and tools for provisioning of resources and workload scheduling; evolution of types of resources
  - AAI (Authentication, Authorization Infrastructure) - the mechanism of a trust federation, certificates evolving to tokens
  - Operations support – security, incident response, problem tracking, daily operations, infrastructure upgrade campaigns
  - Other common services – software delivery, databases and db replication/caching, etc.
  - Diverse experiment-specific services and tools – applications

Networks

- OPN – private connections – managed by WLCG
- LHCOne – Virtual Overlay on national academic networks – managed by NRENs
  - Also used by many other “HEP & related” experiments

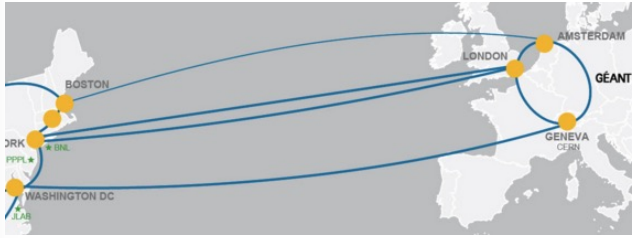
LHCOPN





Needs of LHC have helped the networking community deploy state of the art networking

Significant transatlantic capacity



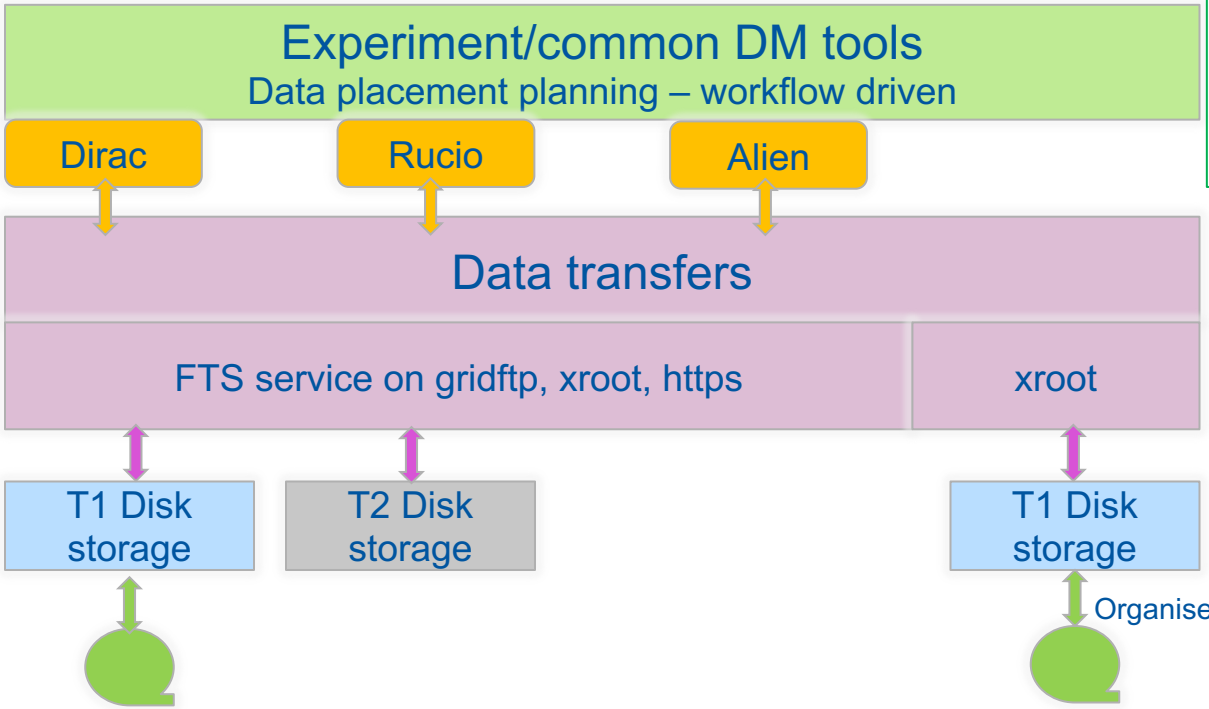
Storage & data services

- Storage tiers:
  - Tape – archive and infrequently accessed data
  - Disk – actively processed data
  - SSD – caches for certain services, databases, etc
- Storage is managed by a software components developed & proven over time
  - Tape – Castor (CERN), Enstore (FNAL), HPSS or TSM (IBM) -> CERN Tape Archive
    - Tier 0 and Tier 1s have concentrated on these solutions
  - Disk – EOS (CERN), dCache (DESY/FNAL), DPM (CERN); some others (filesystems like ceph with grid interface, StoRM)
    - All highly reliable, scalable, and robust: needs operational effort to manage
    - All compute sites today require managed storage

} Cost

- Databases
  - Conditions and other volatile data – Oracle as underlying db, with Oracle tools for replication (online→offline, Tier 0 → Tier 1s) and backup
  - Metadata served via a cache infrastructure → Frontier (dev by CMS/FNAL) used by many
    - Replaces original need to have full Oracle replication everywhere
- Software distribution
  - Was initially a clumsy problem to distribute complex software (often!) to ~100 sites
  - Solved by cvmfs – publishing, content distribution, caching – now used by many communities

Data management tools



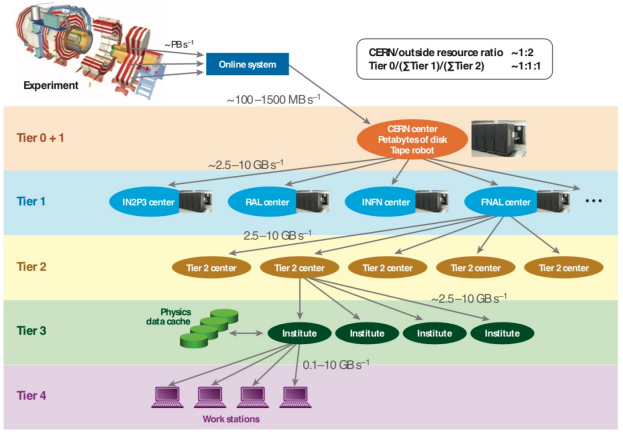
Tools have evolved over many years – experience and changing use patterns
Rucio – becoming a common tool
DIRAC – has broad user community – mainly as workflow

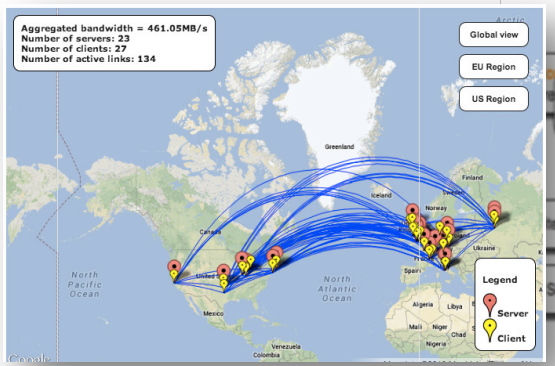
Very robust and tested services
~100 PB/month for 10 years
(~10^{11} files!)

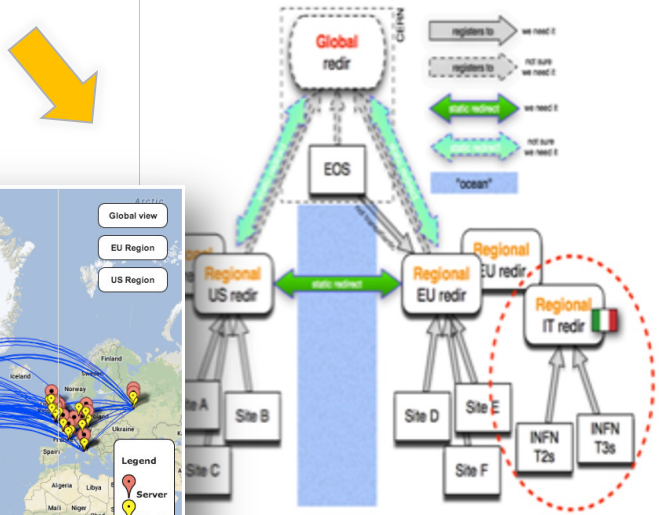
These are tools used by and interesting for many HEP and other experiments

Data organization

- Several different formats of data:
  - RAW, derived-Aanalysis, other new formats of reduced Analysis, ntuple
    - Parallel formats for simulated data
- Different formats have different policies for replication, accessibility, lifetimes, distribution
- Initially data was pre-placed according to a policy ready for processing
  - More or less strong hierarchy between Tiers
  - Today much more dynamic – peer-peer and remote access – data federations
  - Each experiment has variations on this







Workload management

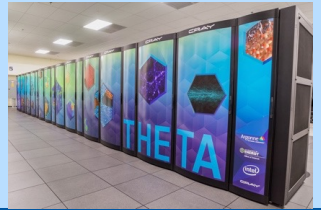
- Different experiments have different workload management systems, but have converged towards a model of “pilot” jobs
  - Late binding – a placeholder job is sent to any free resource, calls home for next (appropriate) priority work
  - This has been very effective at filling available resources, and allows dynamic prioritization within an experiment queue of work
- LHC Experiments have each developed their own workload management service
  - Panda, DIRAC, GlideInWMS/WMAgent, Alien
  - Each have broader communities in HEP, NP, astronomy, and other disciplines using these tools
  - These tools organize the workflows, dispatch chains of jobs, interact with the DM services, collect output, manage errors and resubmissions
    - They communicate with distributed compute resources
- Physicists interact via the experiment frameworks
  - Which in turn make use of the workload management and data management systems

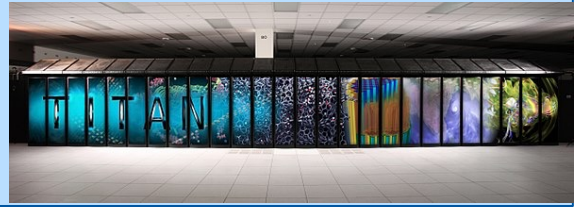
Heterogenous computing

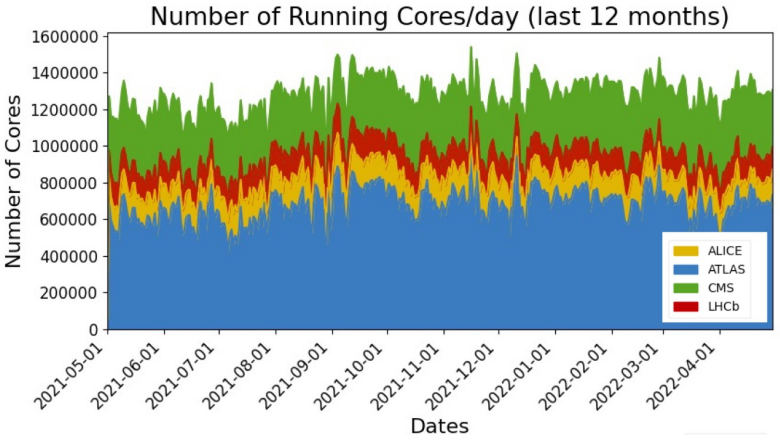
- The majority of today's HEP processing is performed on dedicated clusters of commodity processors ("x86")
- Recently: opportunistic use of many types of compute, in particular HPC systems, and HLT
- In future, this heterogeneity will expand; we must be able to make use of all types:
  - Non-x86 &GPU HPCs, clouds, HLT farms (FPGA?)

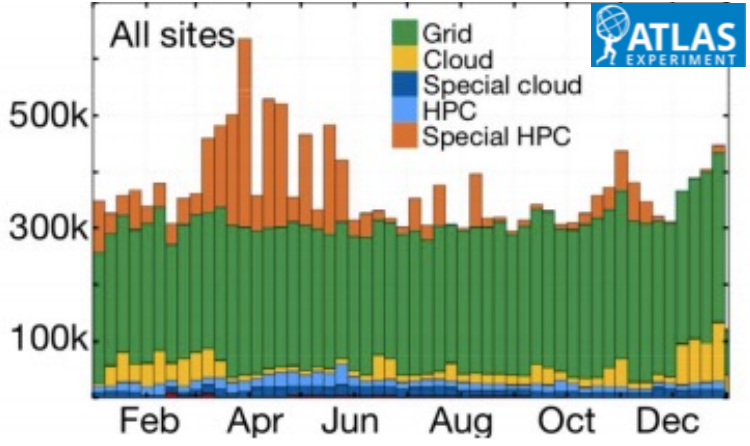
HPC Use is challenging

HEP engagement with DOE & NSF for software adaption in USA, and with PRACE and EuroHPC in Europe are providing these non-x86 processors and GPUs



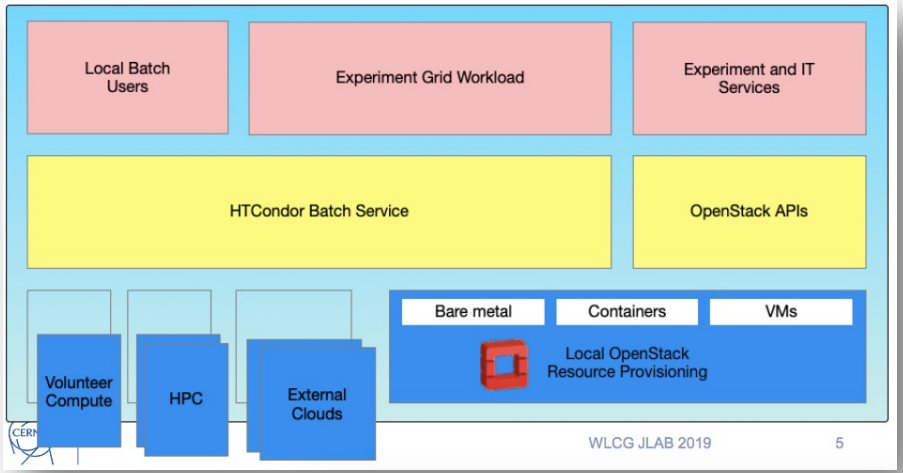






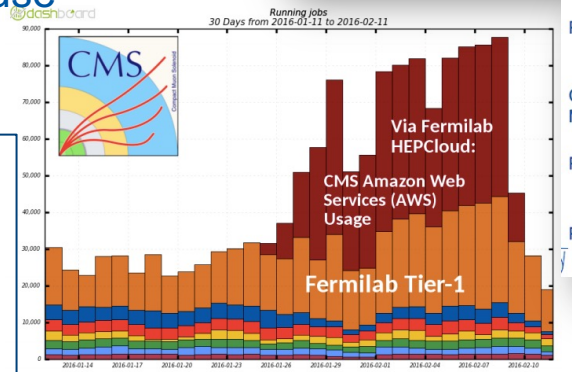
Heterogenous compute

- Requires:
  - Common provisioning mechanisms, transparent to users
  - Facilities able to control access (cost), appropriate use, etc
- HPC, Clouds, HLT will not have (affordable) local storage service (in the way we assume today)
  - Must be able to deliver data to them when they are in active use

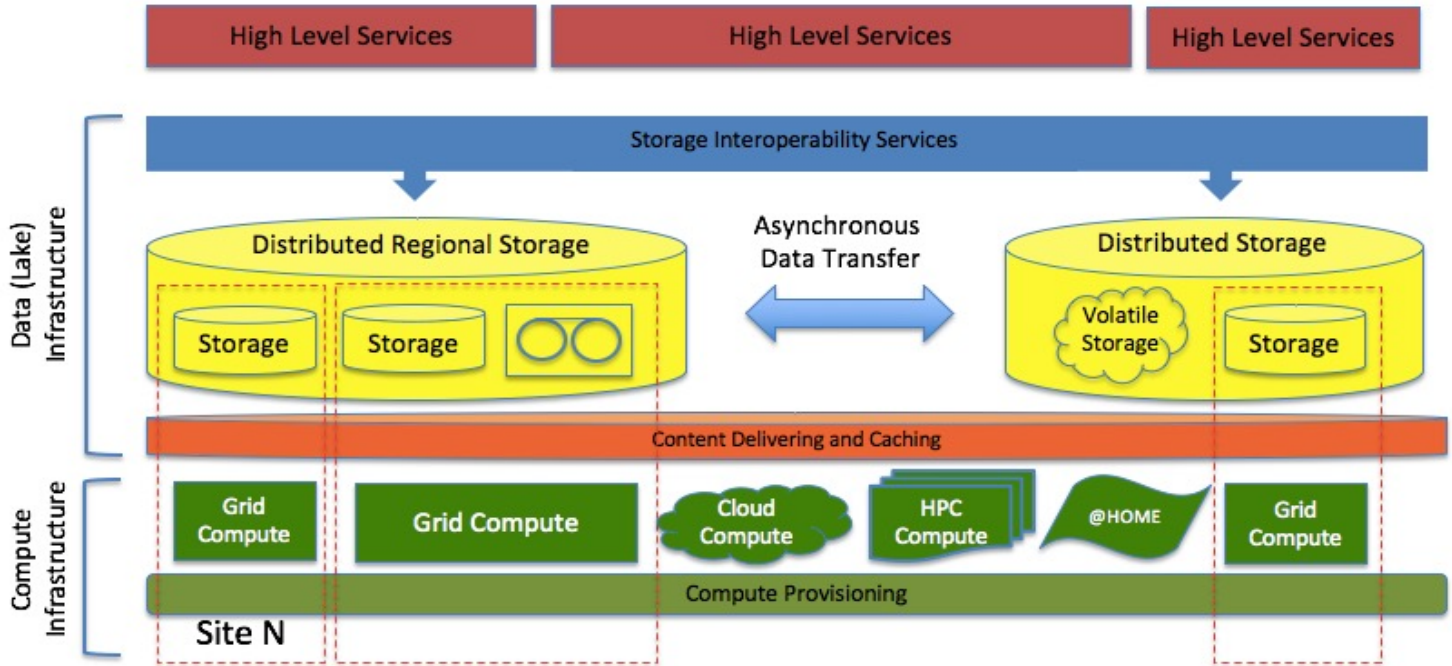


HELIX NEBULA THESCIENCECLOUD

Deployed in a hybrid cloud mode:
- Procurers' data centres
- commercial cloud service providers
- GEANT network and EduGAIN Federated Identity Management



Data delivery “data lake (cloud)”



Idea is to localize bulk data in a local cloud service (Tier 1’s ➔ data lake): minimize replication, assure availability

Serve data to remote (or local) compute – grid, cloud, HPC, ???

Simple caching is all that is needed at compute site (or none, if fast network)

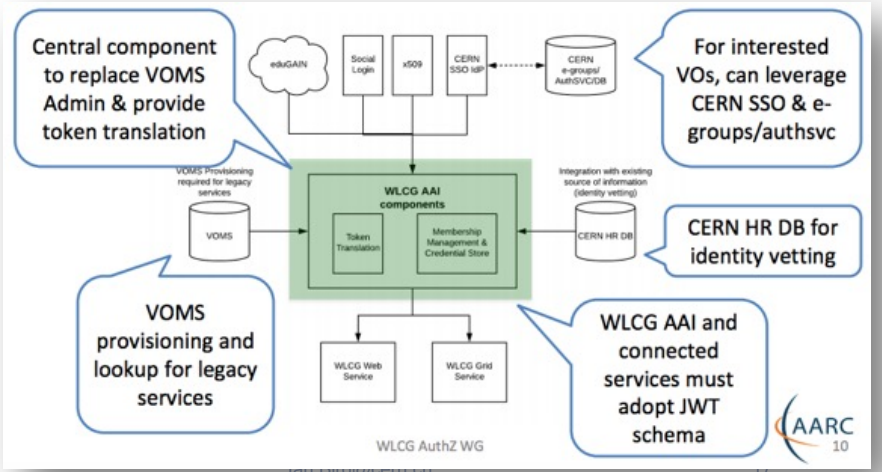
Federate data at national, regional, global scales

Data management and storage R&D

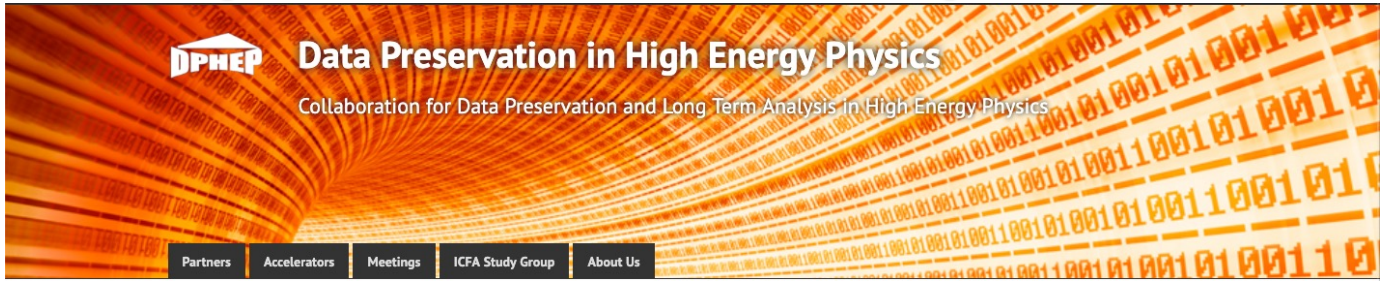
- Set of projects to prototype such a data management infrastructure – and associated tools
- Aims:
  - Reduce the global cost of storage (hardware and operations)
  - Enable a more effective use of existing storage (decrease raid level)
  - Be able to efficiently and scalably deliver data to large, remote, heterogenous, compute resources (LHC Tier centers or HPC, clouds, other opportunistic)
  - Build a common set of DM tools that can be used by a broad set of scientific experiments
    - Today LHC, DUNE, SKA, Belle-II, GW-3G, and others are all using a common set of identified high level tools and services
- Also collaboratively developing underlying data transfer and network tools (replace gridftp w/http, network protocols like IPv6, etc.)

Authentication, Authorization - trust

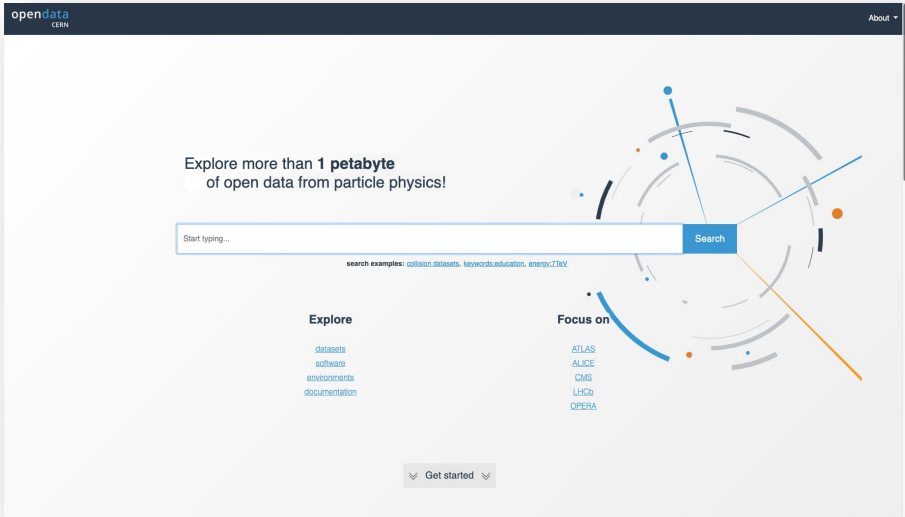
- The grid is based on the idea of identity federation
  - “single sign on” in a distributed computing environment
  - Credentials issued to a user by a trusted local authority
  - Network of trust between all participating institutions and authorities
  - Acceptance of the credentials by each resource (storage, compute, service)
  - This has been implemented using X.509 certificates as the trusted credential
- Authorization:
  - Implemented by a service – VOMS
  - Moving to IAM token server
- The success of this model and the international trust framework has been instrumental in the success of the grid model
- The “token-based” system more aligned with the rest of the world
  - Also more cyber secure



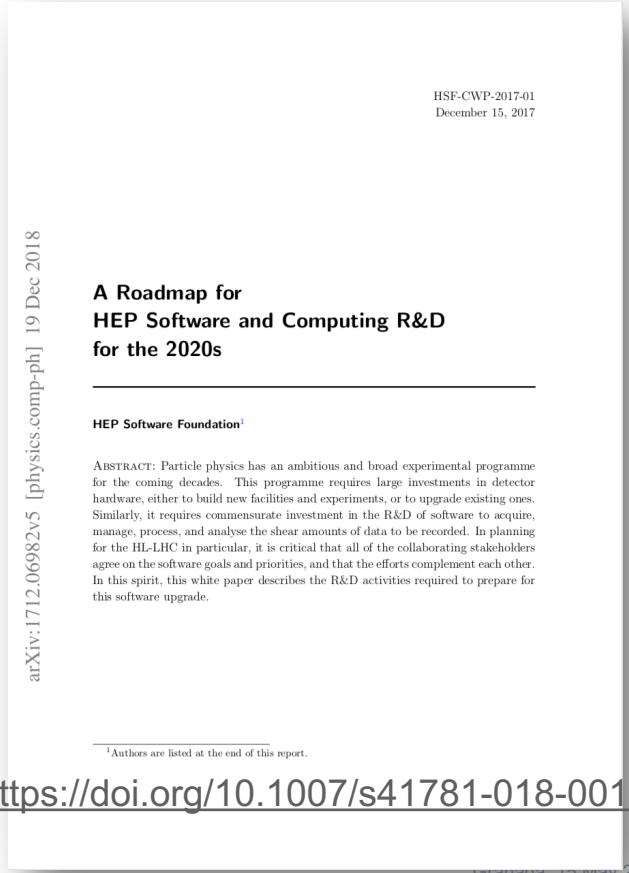
Data preservation, open access



- Bit preservation is a solved problem (modulo cost)
- Data preservation (reproducibility, accessibility) lacks consistent policies, costing, and effort
- Open access also lacks consistent policies and is not explicitly funded
  - Although required by many funding agencies
- There are many use cases of data preservation (for physics) and open access
  - Some are coincident – many are not
- Long term DP for physics requires evolving the mechanisms and tools used to do analysis
  - Cannot afford different solutions
  - Cross experiment efforts in this area have started
- Open access requires policy agreement and appropriate resources
- We can learn from other sciences who do this well



Evolution of HEP computing



HSF-CWP-2017-01
December 15, 2017

arXiv:1712.06982v5 [physics.comp-ph] 19 Dec 2018

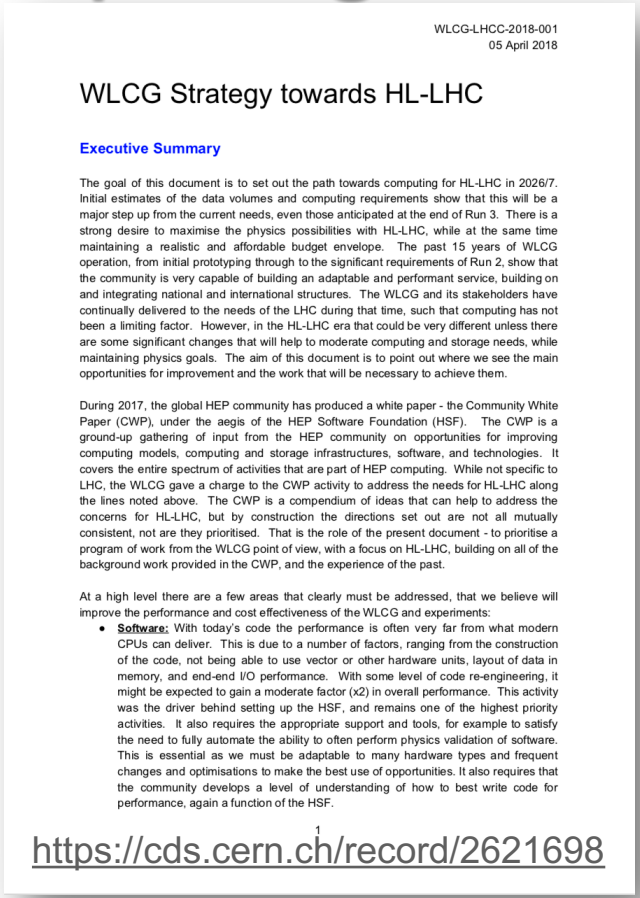
A Roadmap for HEP Software and Computing R&D for the 2020s

HEP Software Foundation[1]

Abstract: Particle physics has an ambitious and broad experimental programme for the coming decades. This programme requires large investments in detector hardware, either to build new facilities and experiments, or to upgrade existing ones. Similarly, it requires commensurate investment in the R&D of software to acquire, manage, process, and analyse the shear amounts of data to be recorded. In planning for the HL-LHC in particular, it is critical that all of the collaborating stakeholders agree on the software goals and priorities, and that the efforts complement each other. In this spirit, this white paper describes the R&D activities required to prepare for this software upgrade.

[1]Authors are listed at the end of this report.

https://doi.org/10.1007/s41781-018-0018-8



WLCG-LHCC-2018-001
05 April 2018

WLCG Strategy towards HL-LHC

Executive Summary

The goal of this document is to set out the path towards computing for HL-LHC in 2026/7. Initial estimates of the data volumes and computing requirements show that this will be a major step up from the current needs, even those anticipated at the end of Run 3. There is a strong desire to maximise the physics possibilities with HL-LHC, while at the same time maintaining a realistic and affordable budget envelope. The past 15 years of WLCG operation, from initial prototyping through to the significant requirements of Run 2, show that the community is very capable of building an adaptable and performant service, building on and integrating national and international structures. The WLCG and its stakeholders have continually delivered to the needs of the LHC during that time, such that computing has not been a limiting factor. However, in the HL-LHC era that could be very different unless there are some significant changes that will help to moderate computing and storage needs, while maintaining physics goals. The aim of this document is to point out where we see the main opportunities for improvement and the work that will be necessary to achieve them.

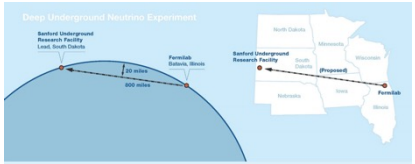
During 2017, the global HEP community has produced a white paper - the Community White Paper (CWP), under the aegis of the HEP Software Foundation (HSF). The CWP is a ground-up gathering of input from the HEP community on opportunities for improving computing models, computing and storage infrastructures, software, and technologies. It covers the entire spectrum of activities that are part of HEP computing. While not specific to LHC, the WLCG gave a charge to the CWP activity to address the needs for HL-LHC along the lines noted above. The CWP is a compendium of ideas that can help to address the concerns for HL-LHC, but by construction the directions set out are not all mutually consistent, not are they prioritised. That is the role of the present document - to prioritise a program of work from the WLCG point of view, with a focus on HL-LHC, building on all of the background work provided in the CWP, and the experience of the past.

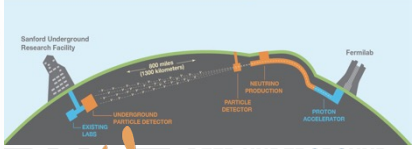
At a high level there are a few areas that clearly must be addressed, that we believe will improve the performance and cost effectiveness of the WLCG and experiments:

- **Software:** With today's code the performance is often very far from what modern CPUs can deliver. This is due to a number of factors, ranging from the construction of the code, not being able to use vector or other hardware units, layout of data in memory, and end-end I/O performance. With some level of code re-engineering, it might be expected to gain a moderate factor (x2) in overall performance. This activity was the driver behind setting up the HSF, and remains one of the highest priority activities. It also requires the appropriate support and tools, for example to satisfy the need to fully automate the ability to often perform physics validation of software. This is essential as we must be adaptable to many hardware types and frequent changes and optimisations to make the best use of opportunities. It also requires that the community develops a level of understanding of how to best write code for performance, again a function of the HSF.

https://cds.cern.ch/record/2621698

Synergies



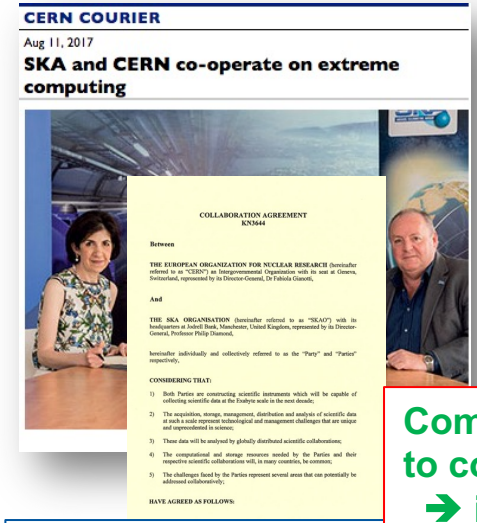


DUNE DEEP UNDERGROUND NEUTRINO EXPERIMENT

DUNE will leverage the WLCG for its computing infrastructure

First formal non-LHC

Belle II



Common theme in many contributions is the desire to collaborate with and benefit from LHC R&D work
➔ implies evolution to a global cyber infrastructure

Astroparticle Physics European Consortium (APPEC)

APPEC

APPEC Contribution to the European Particle Physics Strategy

December 17, 2018

Editorial Board:
S. Katsanevas, A. Masiero, T. Montaruli, J. de Kleuver, A. Haungs

Contact Person:
T. Montaruli (APPEC Chair from Jan. 1, 2019)
Email: teresa.montaruli@unige.ch

Website : http://www.appec.org

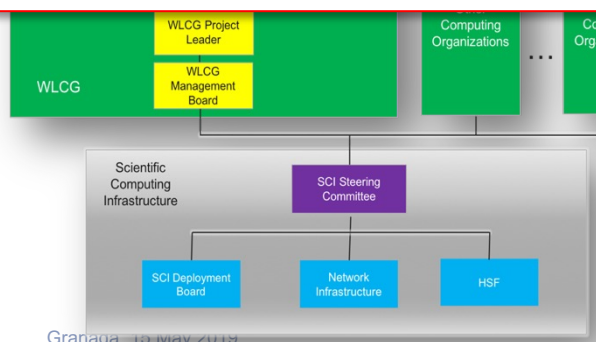
APS Division of Particles and Fields Response to E
Strategy Group Call for White Papers:
Tools for Particle Physics

DPF Executive Committee and Strategy Whitepaper Editing Group
dpfstrategy@fnal.gov

December 18, 2018

Abstract

The U.S. particle physics strategy process is summarized in a companion white paper that also describes U.S. activities related to the five P5 science drivers. Additional activities within the U.S. particle physics program that are critical to progress in our field are described here.



The International Linear Collider
A Global Project

Statement of the Pierre Auger Collaboration as input for the European Particle Physics Strategy Update 2018 – 2020

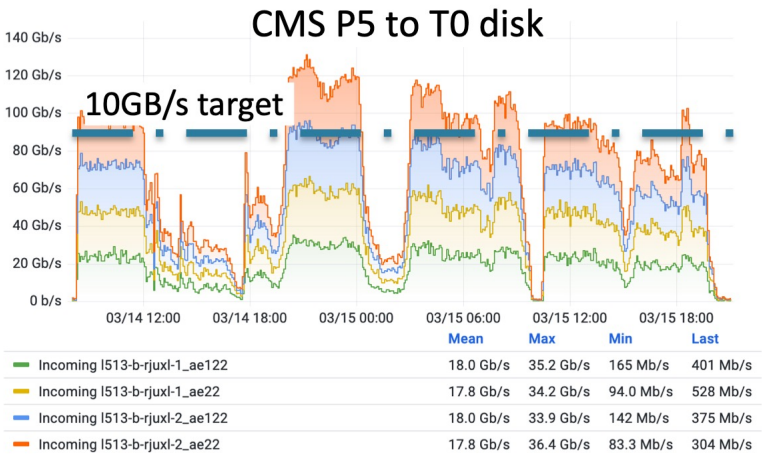
Ralph Engel (auger_spokespersons@fnal.gov),
on behalf of the Pierre Auger Collaboration

Conclusions

- The LHC computing models and WLCG have been very successful, and have evolved significantly over 15 years
- Many other experiments have collaborated and used the infrastructures and tools (WLCG, EGI, OSG, others)
- Strong desire to strengthen the synergies and collaborations for the future – infrastructure, tools, and software as the rest of the sciences also become “big data” sciences.

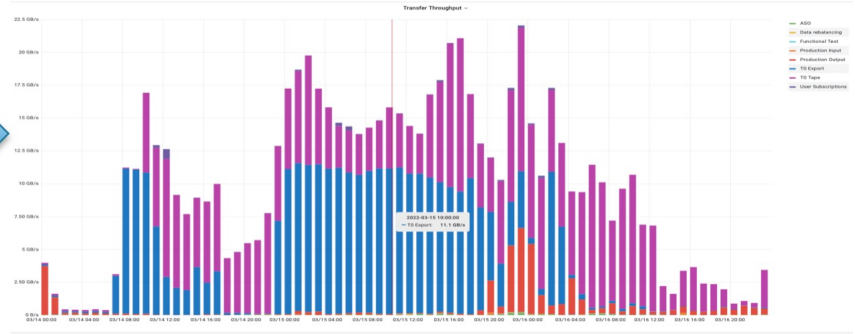
Backup

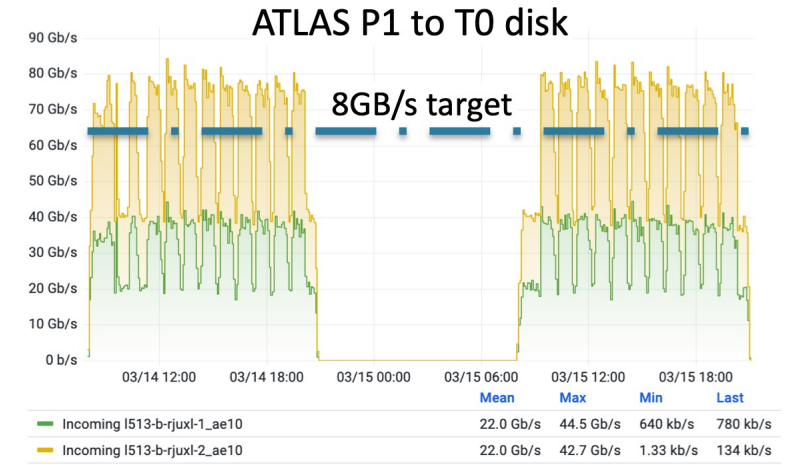
Readiness for Run-3 – ATLAS and CMS



CMS T0 disk => T1 tape @11GB/s

CMS T0 disk => T0 tape @11GB/s





ATLAS T0 disk => T1 tape @12GB/s

ATLAS T0 disk => T0 tape @16GB/s