Current HEP Computing Models

Ian Bird, WLCG, CERN

ESPPU Open Symposium

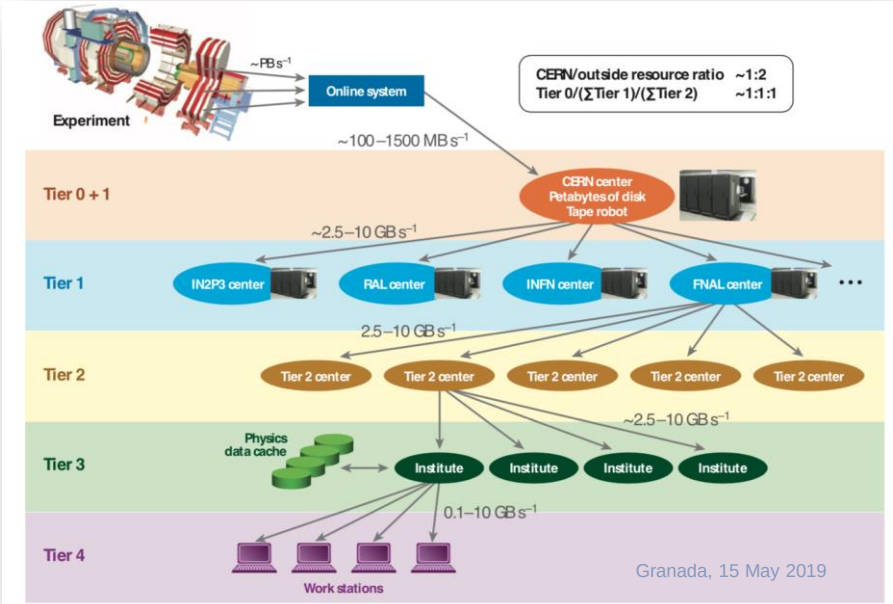
Granada, 15th May 2019

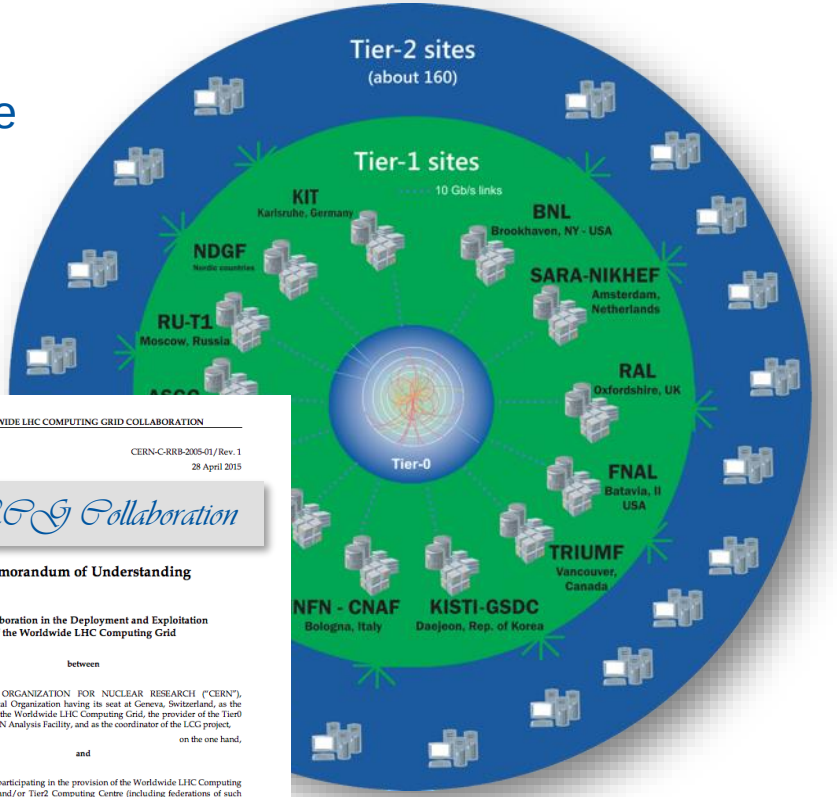
Acknowledgement

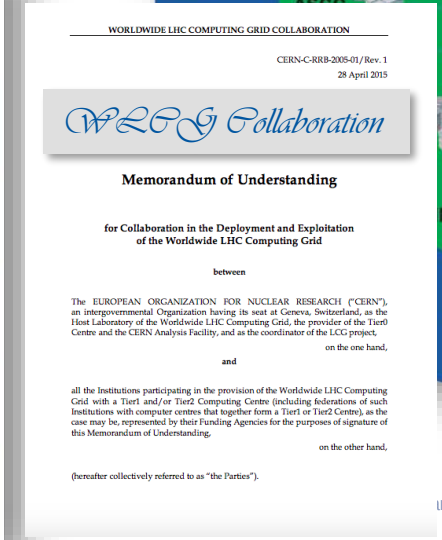
- Input to this talk has been drawn from many discussions with colleagues from many HEP and closely related experiments over recent years, in particular
  - Preparation of CWP and WLCG strategy documents
  - Recent WLCG and HOW2019 workshops
  - Contributions as input to the ESPPU

Today's computing models

- HEP has always done distributed computing
- 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").

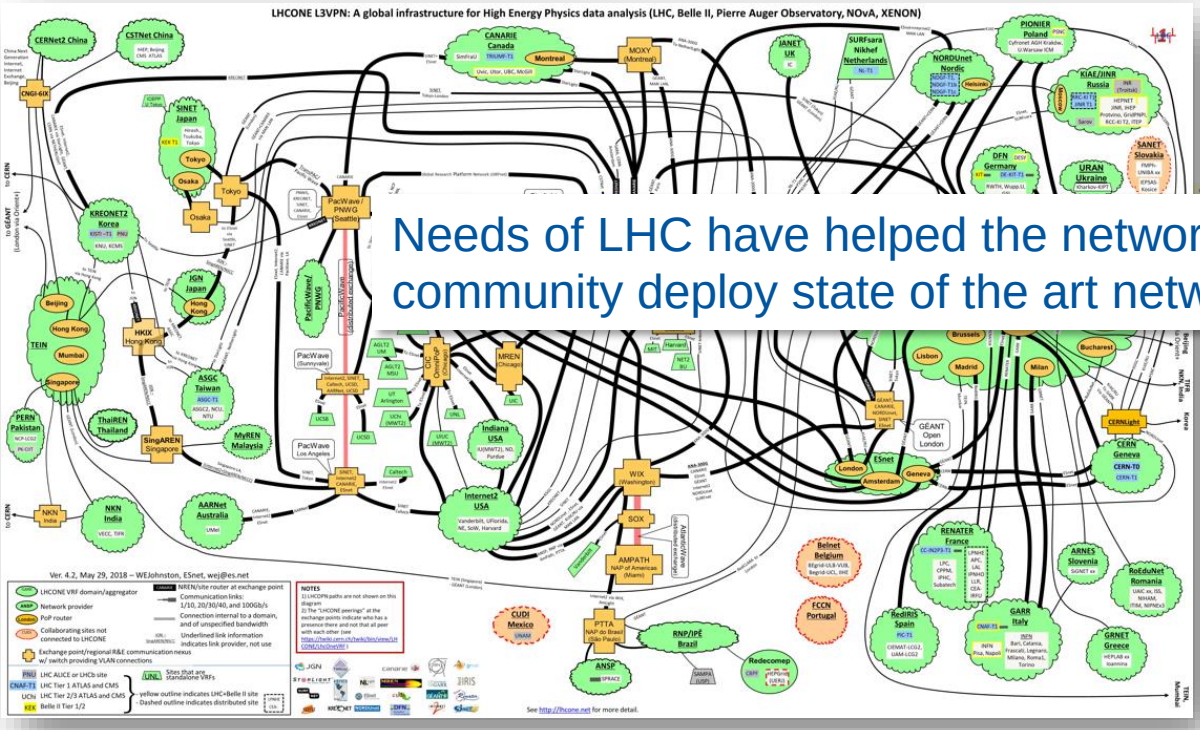
LHC and other HEP experiments

- LHC computing (WLCG) was a significant step in organization and change of computing models
- However, from the beginning of the grid projects
  - Most other HEP experiments have made use of many of the tools and benefitted from the grid infrastructures (national, regional, global)
- All experiments' computing is distributed to some degree
- Major features and capabilities of todays HEP computing infrastructures:
  - Networks – international and national, private and public
  - Data management – key to success, data transfers, storage systems, data management tools and data organization
  - Compute – provisioning of resources and workload scheduling; evolution of types of resources
  - AAI (Authentication, Authorization Infrastructure) - the mechanism of federation, single sign on, etc.
  - Operations support – security, incident response, problem tracking, daily operations, upgrade campaigns
  - Other common services – software delivery, databases and db replication/caching, etc.
  - Diverse experiment-specific services and tools – applications

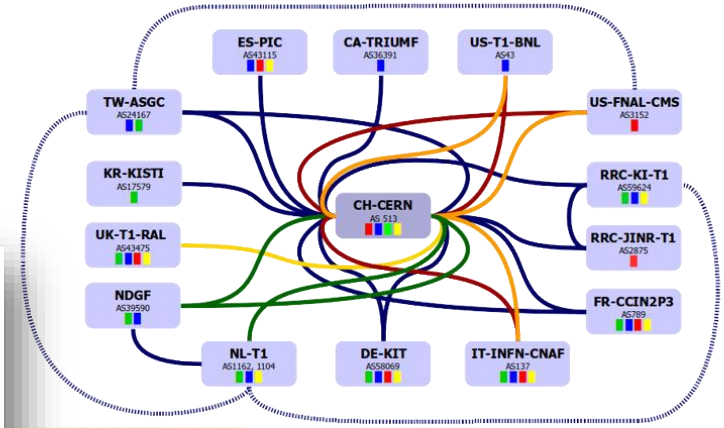
Networks

LHCOPN

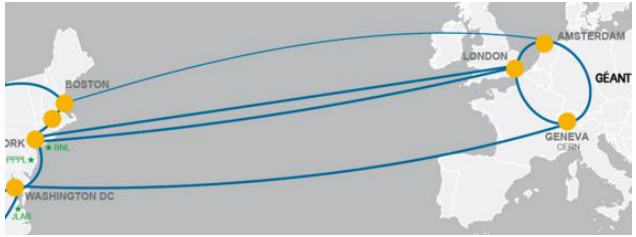
- OPN – private connections – managed by WLCG
- LHCOne – Virtual Overlay on national academic networks – managed by NRENs
  - Also used by many other “HEP & related” experiments



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



Significant transatlantic capacity



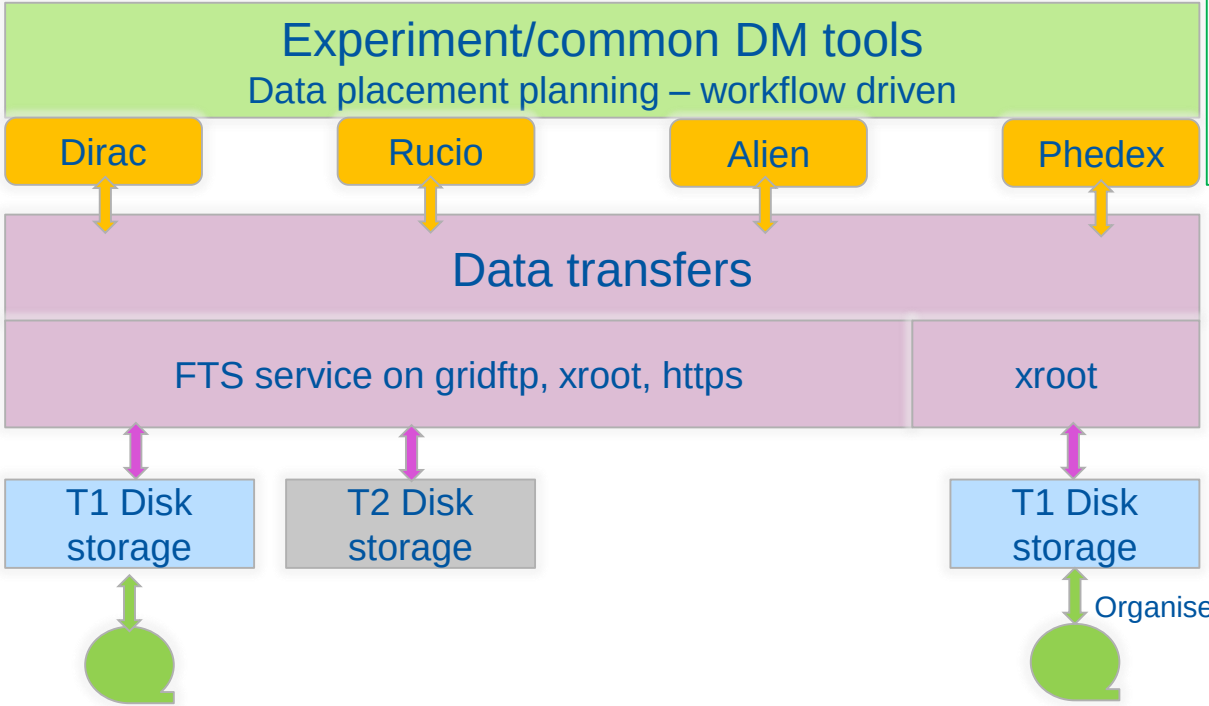
Storage & data services

- 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)
    - 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 some form of 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



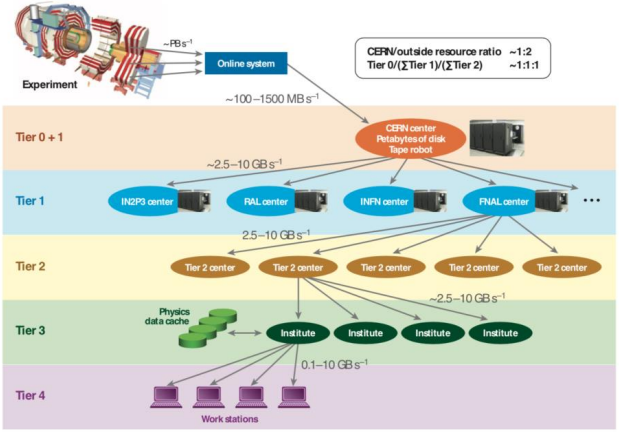
Tools have evolved over many years – experience and changing use patterns
Rucio – becoming a common tool
DIRAC – has broad user community – mainly as workflow

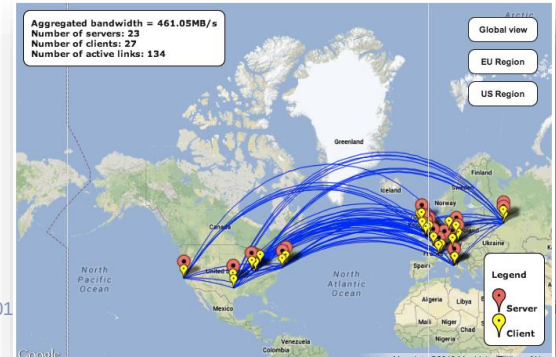
Very robust and tested services
~100 PB/month for 10 years
(~10^{11} files!)

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

Data organization

- Several different formats of data:
  - RAW, ESD, AOD, derived-AOD, other new formats of reduced AOD, 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



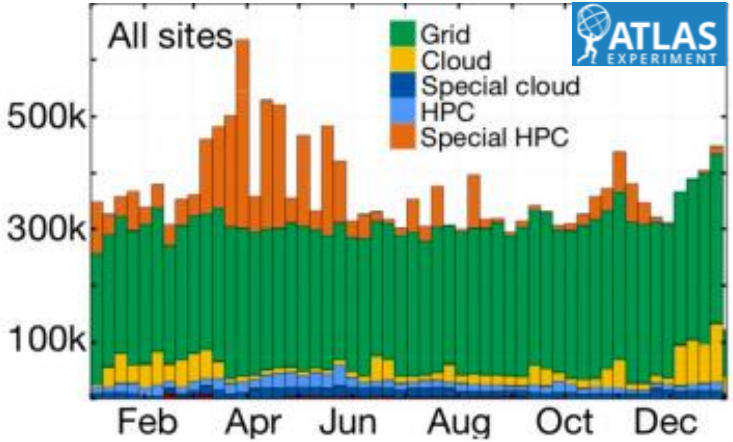


Workload management

- 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

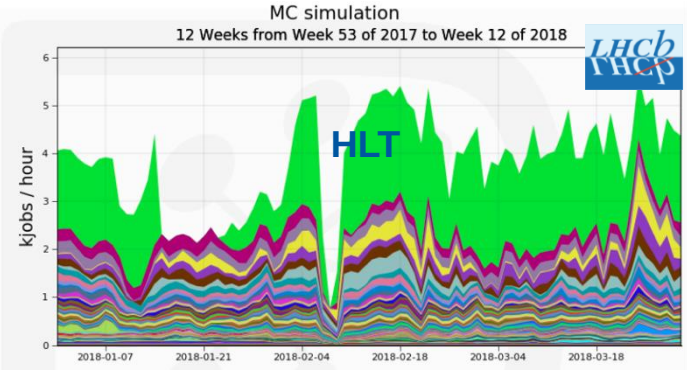
Heterogenous computing

- 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 (esp GPU), HPC, clouds, HLT farms (inc FPGA?)



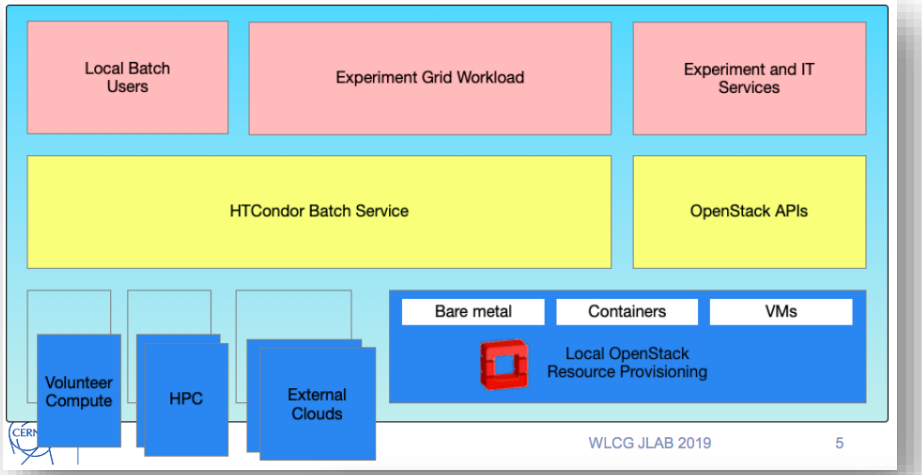
HPC Use is challenging

HEP engagement with DOE & NSF in USA and (together with SKA) with PRACE and EuroHPC in Europe and participating in BDEC2 workshops



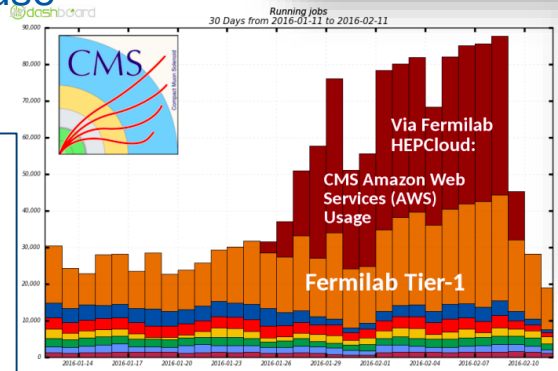
Heterogenous compute

- 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

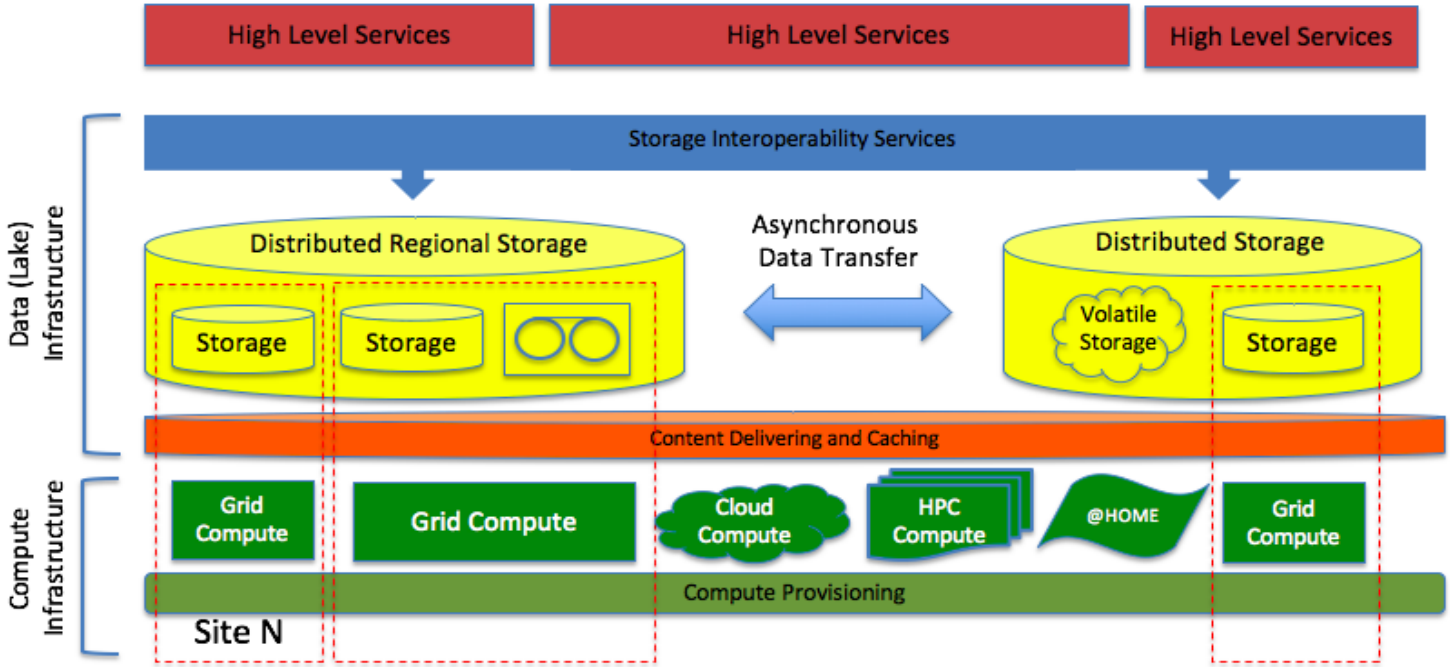


HELIX NEBULA THESCIENCECLOUD

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 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)

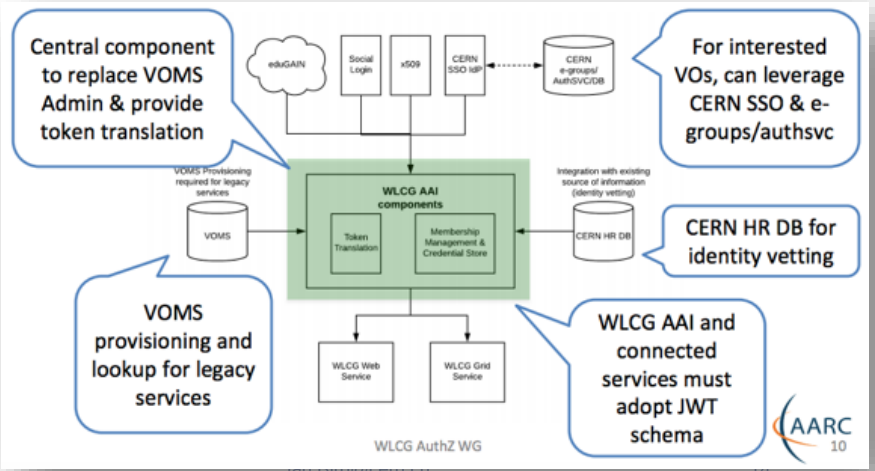
Federate data at national, regional, global scales

Data management and storage

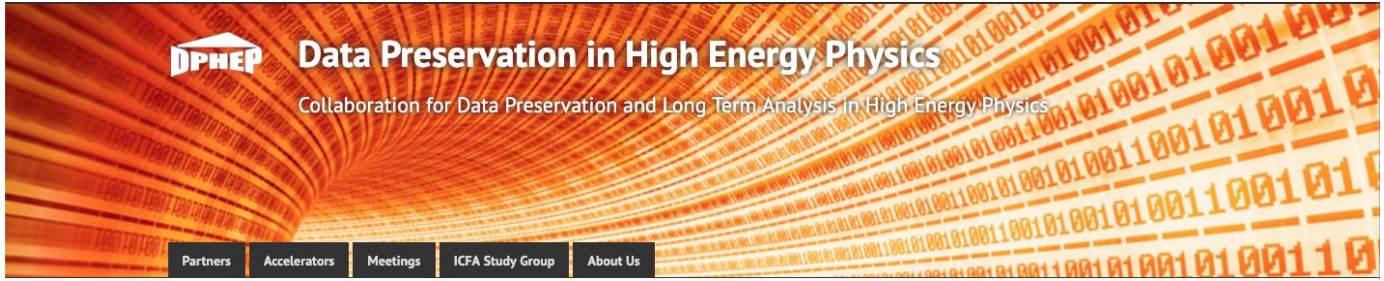
- Set of R&D projects to prototype such a data management infrastructure – and associated tools
- Aims:
  - Reduce the global cost of storage (hw and operations)
  - Enable a more effective use of existing storage
  - Be able to efficiently and scalably deliver data to large, remote, heterogenous, compute resources (LHC Tier centres 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 looking at a common set of identified tools
- Also collaboratively (LHC+SKA with GEANT) looking at underlying data transfer and network tools (replace gridftp, network protocols, etc.)

Authentication, Authorization - trust

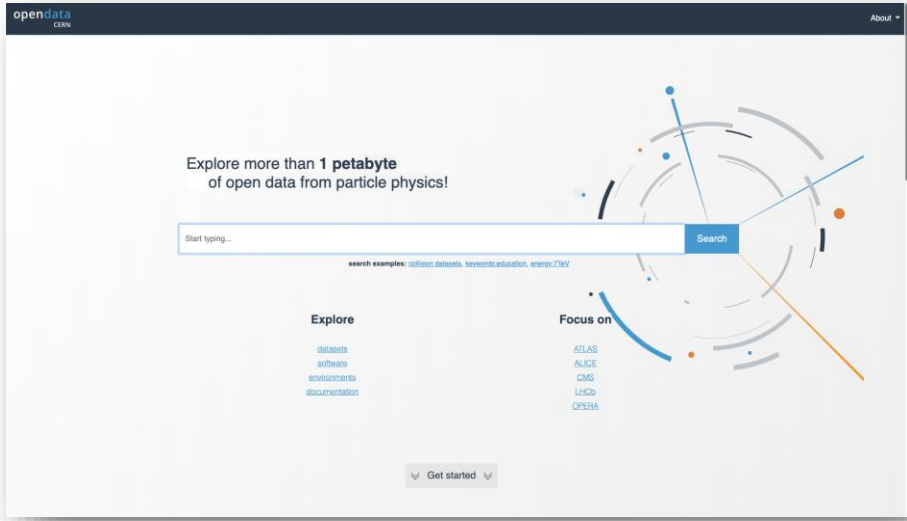
- The grid is based on the idea of identity federation
  - “single sign on” in a distributed computing environment
  - Credentials issued to a user by a trusted local authority
  - Network of trust between all participating institutions and authorities
  - Acceptance of the credentials by each resource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



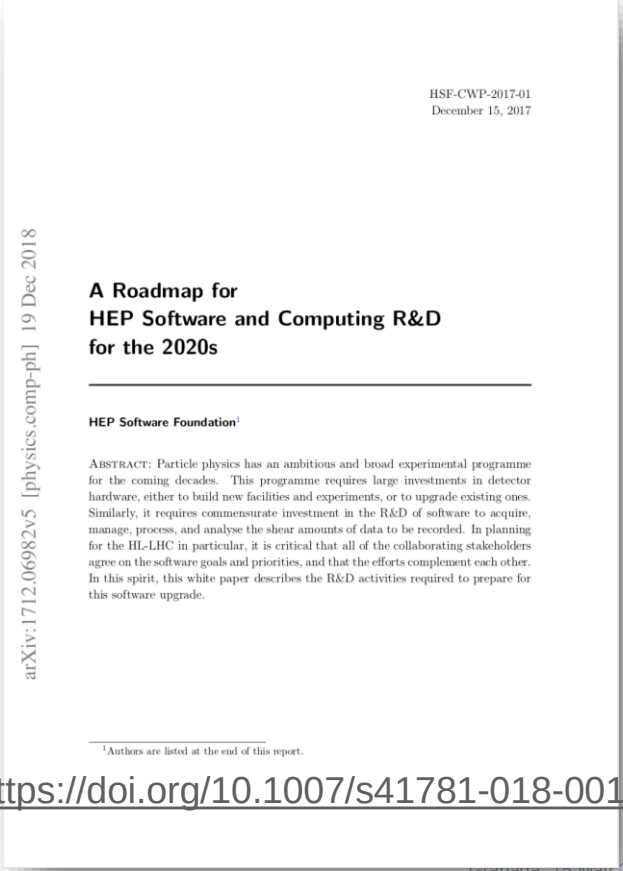
Data preservation, open access



- Bit preservation is a solved problem (modulo cost)
- Data preservation (reproducibility, accessibility) lacks consistent policies and costing
- 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
- Open access requires policy agreement and appropriate resources
- HEP 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

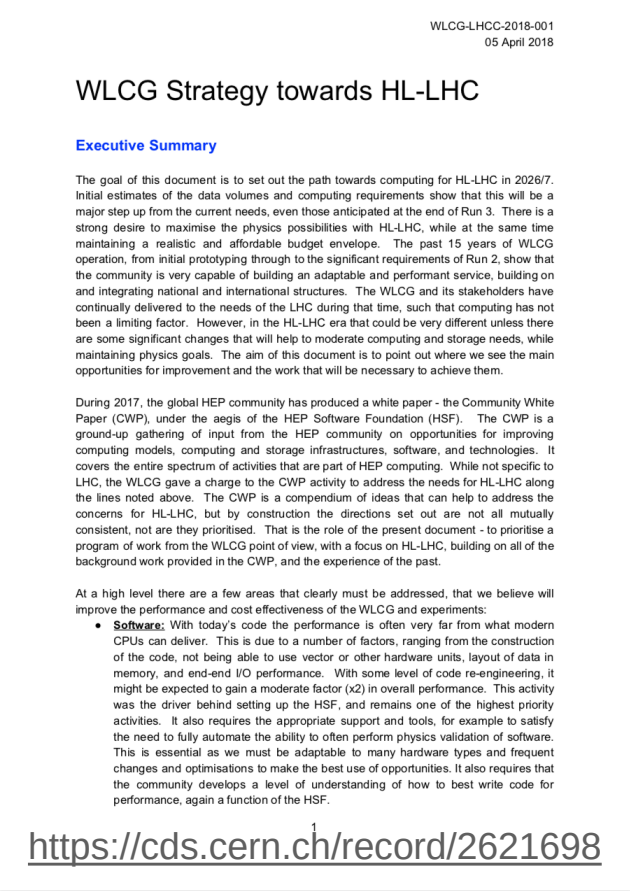
A Roadmap for HEP Software and Computing R&D for the 2020s

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.

1

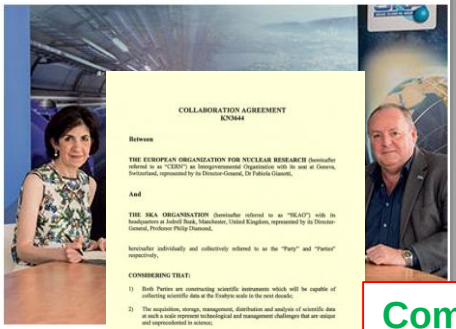
https://cds.cern.ch/record/2621698

Synergies

CERN COURIER

Aug 11, 2017

SKA and CERN co-operate on extreme computing



COLLABORATION AGREEMENT
KN3644

Between

THE EUROPEAN ORGANIZATION FOR NUCLEAR RESEARCH (hereinafter referred to as "CERN") an Intergovernmental Organization with its seat at Geneva, Switzerland, represented by its Director-General, Dr Fabiola Gianotti,

And

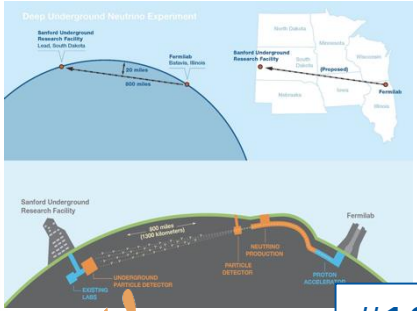
THE SKA ORGANISATION (hereinafter referred to as "SKAO") with its headquarters at Jodrell Bank, Manchester, United Kingdom, represented by its Director-General, Professor Philip Diamond,

hereinafter individually and collectively referred to as the "Party" and "Parties" respectively,

CONSIDERING THAT:

1) Both Parties are constructing scientific instruments which will be capable of collecting scientific data at the Exabyte scale in the next decade;
2) The acquisition, storage, management, distribution and analysis of scientific data at such a scale represent technological and management challenges that are unique and unprecedented in science;
3) These data will be analysed by globally distributed scientific collaborations;
4) The computational and storage resources needed by the Parties and their respective scientific collaborations will, in many countries, be common;
5) The challenges faced by the Parties represent several areas that can potentially be addressed collaboratively;

HAVE AGREED AS FOLLOWS:



DUNE DEEP UNDERGR NEUTRINO EXPERIMENT

#126

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 governance evolution

Joint Gravitational Waves and CERN Meeting

Friday 1 Sep 2017, 09:00 → 20:00 Europe/Zurich
500-1-001 - Main Auditorium (CERN)
Federico Ferrini (INFN Sezione di Pisa (INFN)), Francesco Fid

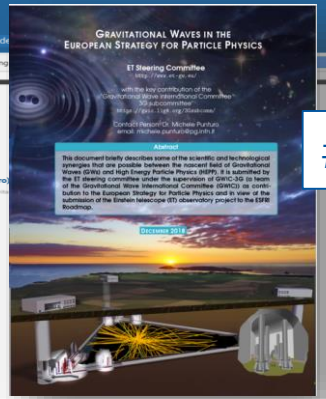
Videoconference Rooms: Joint_Gravitational_Waves_and_CERN_Meeting

09:00 → 09:20 Welcome and introduction to the meeting
Speakers: Eckhard Elsen (CERN), Federico Ferrini (EGO)

09:20 → 09:45 GW from a particle physics perspective
Speaker: Gian Giudice (CERN)
LigoVirgo.pdf

09:45 → 10:10 The New Era of Precision Gravitational-Wave (astro
Speaker: Alessandra Buonanno (Max Planck Institute for Gravit
CERN-AB.pdf

10:10 → 10:30 Discussion



#64

Astroparticle Physics European Consortium (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

APPEC

#84

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

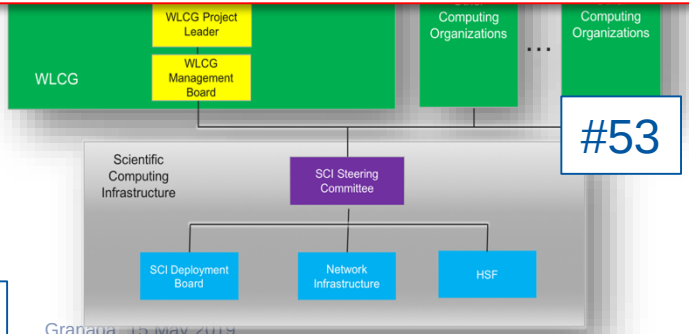
DPF Executive Committee and Strategy Whitepaper Editing Grou
dpfstrategy@fnal.gov

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.

#150



#53

#77

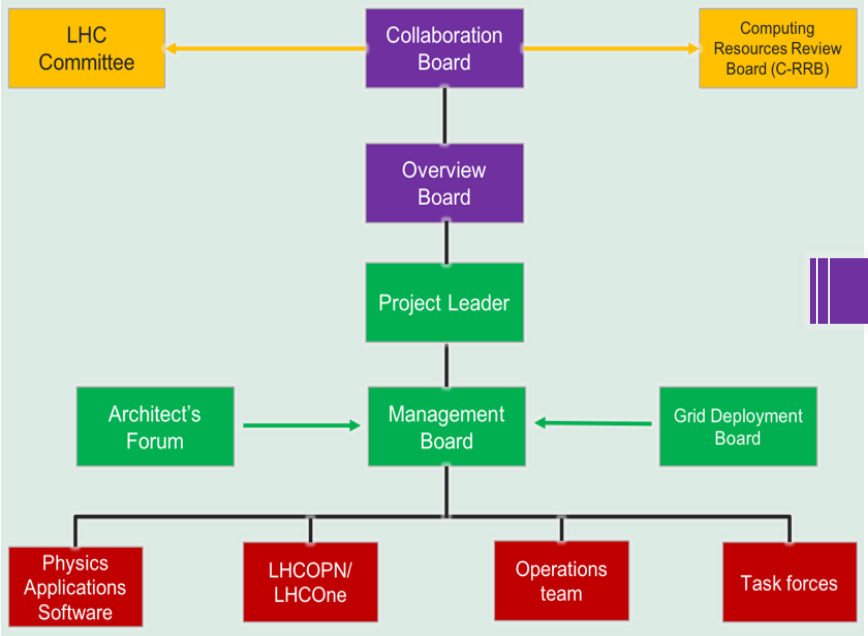
The International Linear Collider
A Global Project

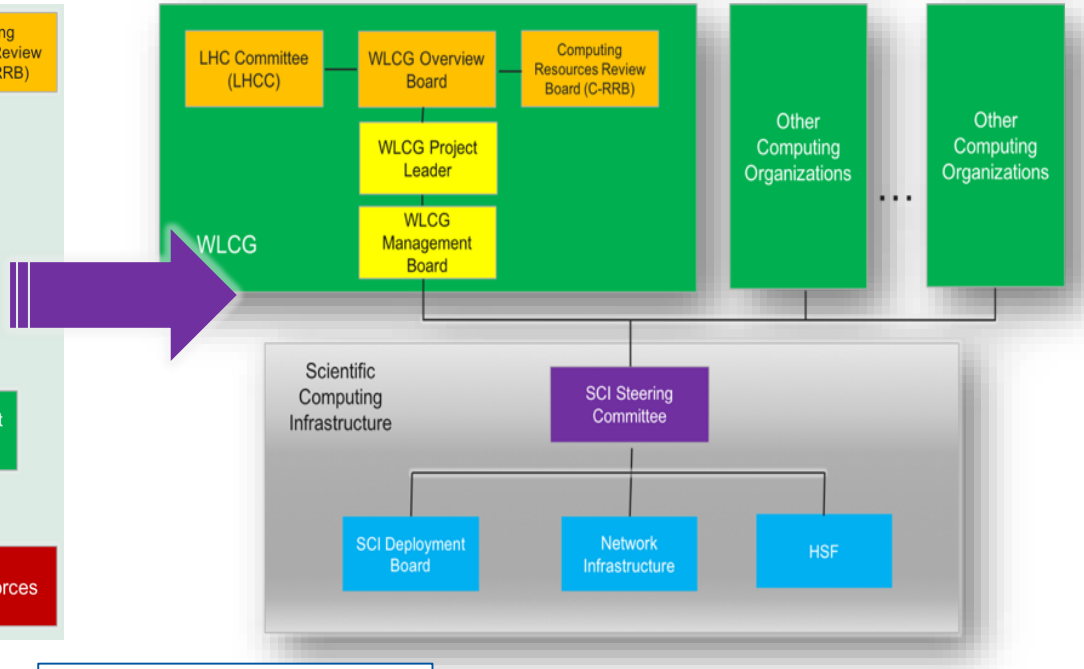
Statement of the Pierre Auger Collaboration as input for the European Particle Physics Strategy Update 2018 – 2020

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

#117

Governance / organization





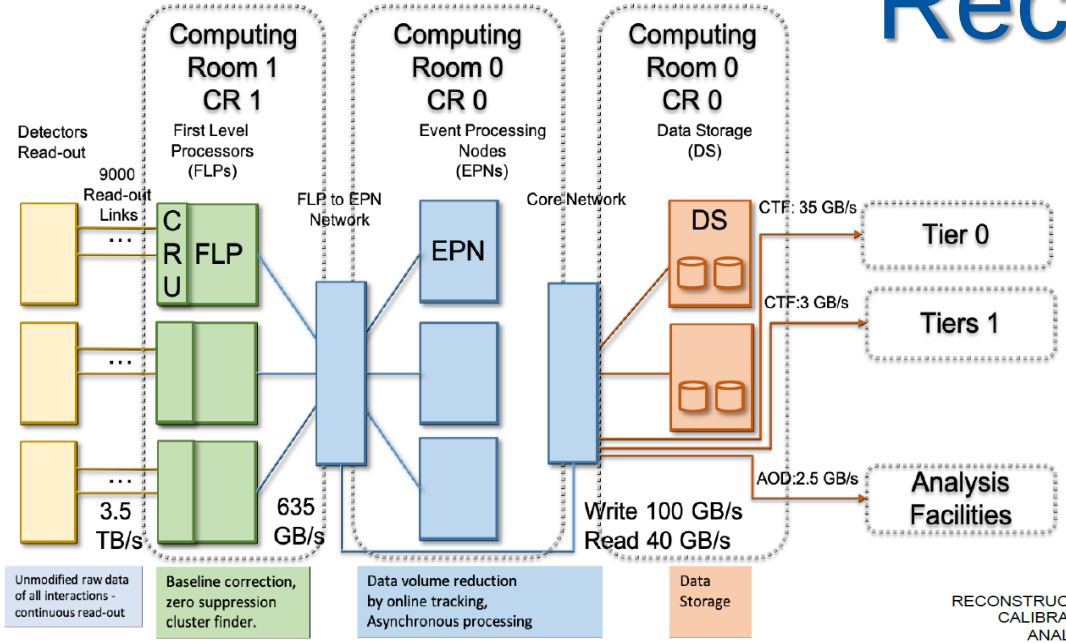
Contribution #53

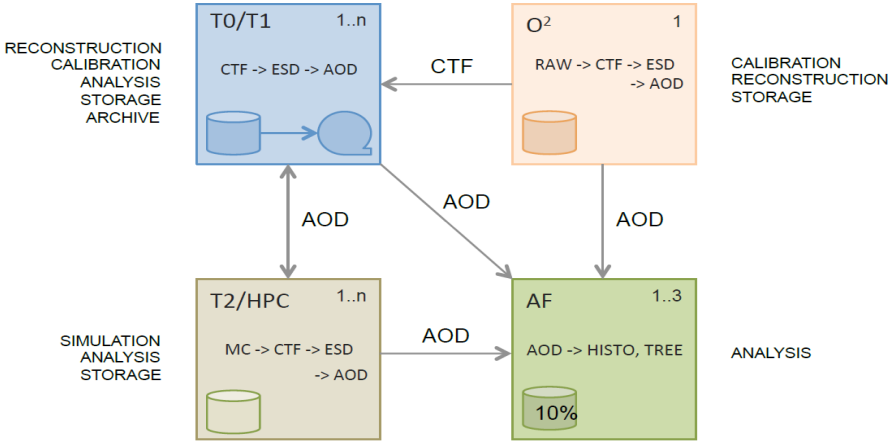
Conclusions

- The LHC computing models and WLCG have been very successful, and have evolved significantly over 15 years
- Many other HEP 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

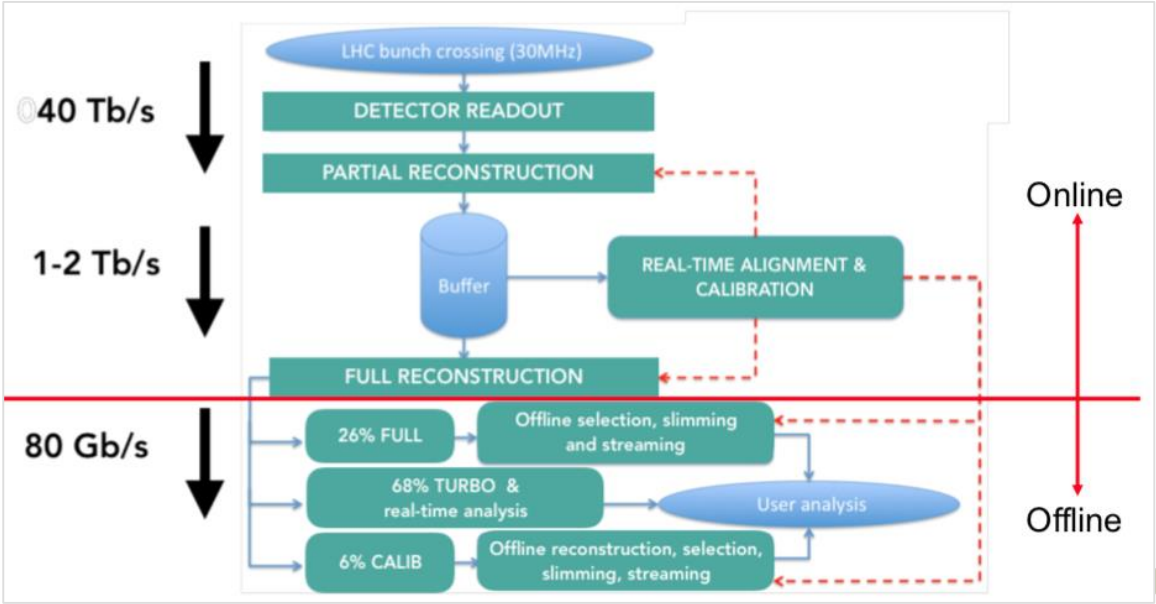
Backup

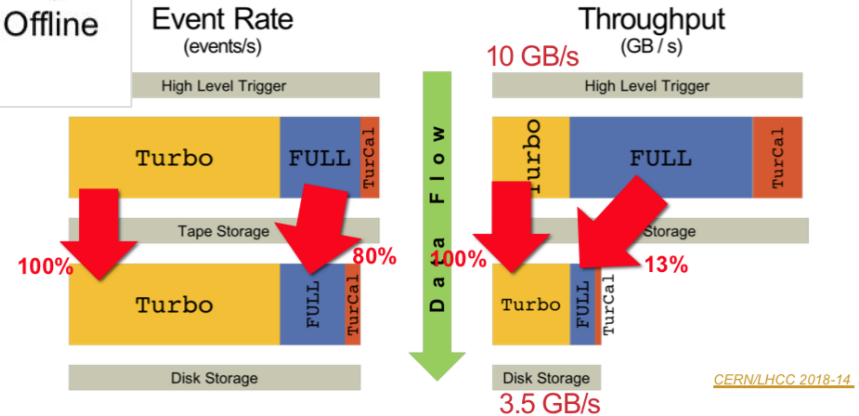
Recent Evolution - ALICE





Recent Evolution - LHCb





CERN/LHCC 2018-14