

WLCG → HEP Computing Infrastructure

Ian Bird, Simone Campana; CERN

WLCG Overview Board

CERN, 30th November 2018

Background

- Ideas presented at OB 2 years ago
 - <https://indico.cern.ch/event/468477/>
- Then discussed at SCF in Feb 2017
 - <https://indico.cern.ch/event/581096/>
- Then
 - CWP and HL-LHC strategy for computing
 - ESCAPE proposal
- More recently:
 - Discussion with DUNE on how to use “WLCG” infrastructure and be able to benefit from tools and developments
 - Mentioned in SCF Sept. 2018
 - Also potential interest from SKA, 3G-GW community, and others (Belle II)



Introduction

- Leverage
 - WLCG experience
 - Capabilities in the internet sector (large distributed DC's, clouds, etc.)
 - New ideas of how to manage Exabyte scale data
- Towards a HEP-wide (at least) scientific data and computing environment for the future
 - Similar needs from related fields – astro, gw, ...

S.W.O.T

- Strengths
 - WLCG established, trust network among stakeholders
 - Lightweight decision making processes
 - Consolidated infrastructure – common building blocks, policies, federation, resource sharing, etc.
- Weaknesses:
 - Lack of investment in software for the future (general concern); initial investment in WLCG applications has levelled off; future need more weight on software
 - WLCG does not have development effort – reliance on external providers – some developments and efforts were of marginal use
- Threats:
 - Flat funding but expectation of more and more computing and data; Significant investment in data management and software performance is essential to manage this
 - Other sciences now compete with WLCG for resources at similar scales; WLCG may lose influence
 - Need to use facilities not particularly suited to our problems – HPC
- Opportunities
 - Leverage experience and be central to evolution of scientific computing and become the reference; benefit other sciences, resource providers, funding agencies
 - Retain significant leadership role in setting direction

What?

- The idea of broadening the WLCG infrastructure to HEP++ (e.g. parts of Astro community, GW, etc.)
- See 3 key areas:
 - “HEP” computing infrastructure
 - Facilities and services – Data Cloud/Lake and tools
 - Software investment
- Would imply a change in organizational structure
 - Factorise WLCG into:
 - LHC-specific part (MoU, pledges, etc.)
 - HEP-wide Infrastructure – services, tools, operations, etc.
 - The latter foreseen as a collaboration with a steering group made up from stakeholders

WLCG:

- Collaboration
- Resource management
- LHCOPN
- (operations?)

DUNE:

- Collaboration
- Resource management



Belle 2?
Etc.
GW 3G?

Astro(-particle)?

HEP Distributed Computing Infrastructure

- AAI, security
- Common DM tools
- Common provisioning (HPC, cloud, grid)
- Accounting, monitoring, etc. (operations?)

Tool box of compatible services

International Steering Group

- ### Networking
- LHCOne
 - Etc.

HEP Software Foundation

- Application software
- Software skills
- Software sharing
- Etc.

Tool box of applications

Data management (“data lake”)

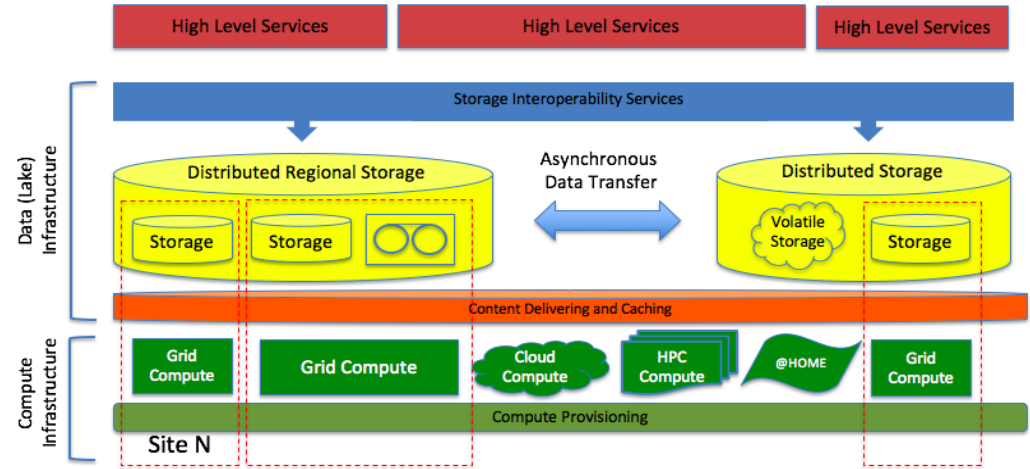
Data Organisation, Management, Access (DOMA)

□ Several activities and working groups

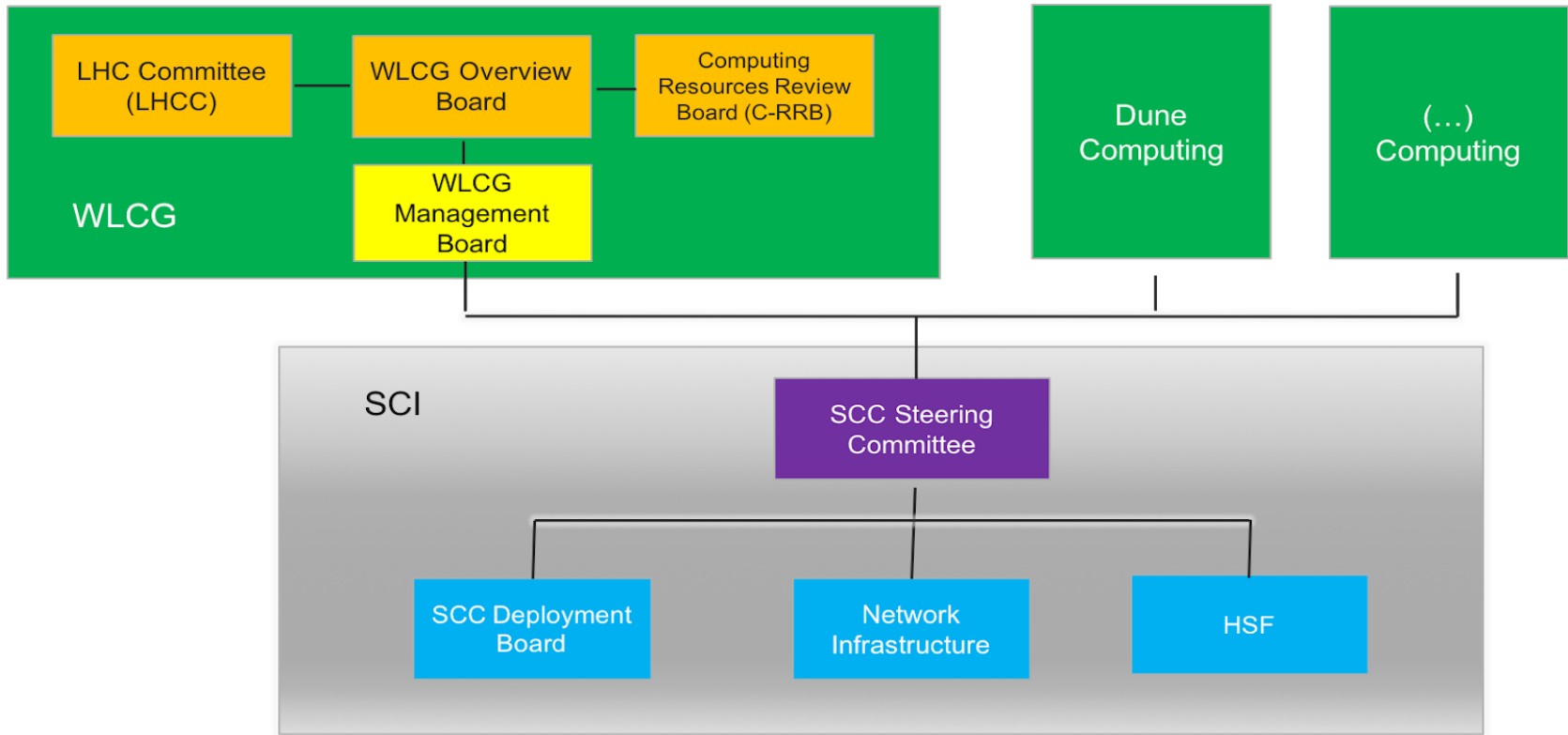
- Storage consolidation
- Caching and data access
- Data transfer and access protocols:
 - 3rd party copy
 - Replacement of gridftp
- Quality of Service
 - Performance/reliability vs capacity
 - Use of high-performance storage?
- Use of networks and Investigation of low level protocols and optimization of data movement (with SKA, Geant, others)
 - Between parts of the data lake
 - Serving data

□ A prototype “data lake” has been set up and can be used to explore technology and R&D questions

- Several Tier 1s participating in the prototype



- Idea is to localize bulk data in a cloud service (→ data lake): minimize replication, assure availability
- Serve data to remote (or local) compute – grid, cloud, HPC, etc.
- Simple (unmanaged) caching is all that is needed at compute site
- Works at national, regional, global scales



Advantages

- We take the initiative and lead the development
- Model we imagine is also interesting to other communities:
 - DUNE, GW, SKA, etc
 - Model includes potential use of commercial facilities, HPC, etc.
 - Is inherently scale-out
 - Is the model emerging from the CWP and Strategy process
 - Maintain the formal part of the collaboration where it is required (resource pledging, etc.)
- ESCAPE project picks up on the infrastructure ideas as input to EOSC for Exascale data infrastructure
- The governance is science community-led and steered
 - This is key
 - Different from EGI (and EGEE, PRACE, EUDAT, etc.)

Next steps

- Paper attached to agenda
- Would like to submit this to ESPP open call as ideas for future HEP computing infrastructure
 - Deadline Dec 18.
- OB (+CB) should discuss how we move towards such a model in a non-disruptive way
 - Preferably without the need to re-sign the MoU