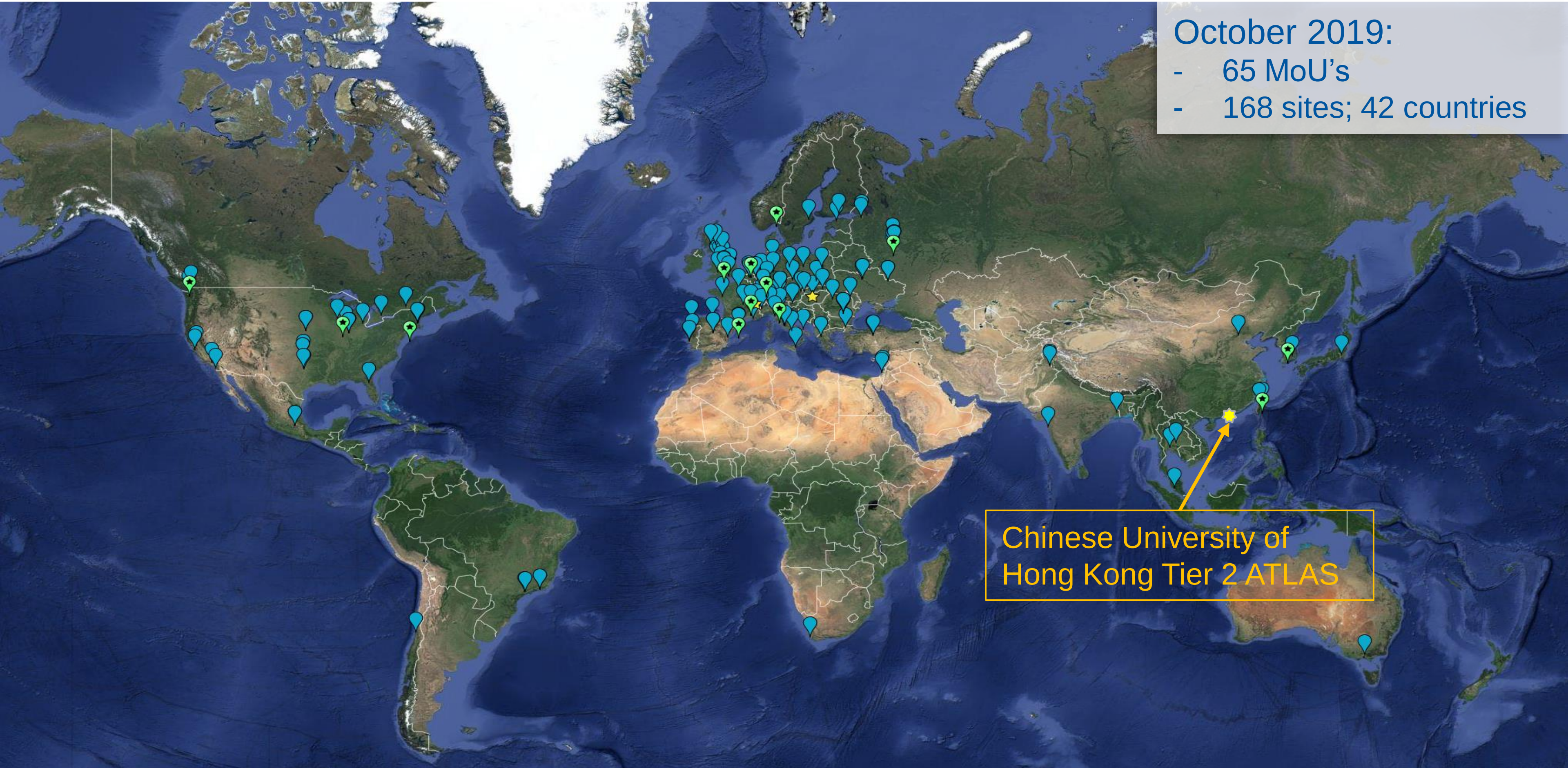
CERN-RRB-2019-123

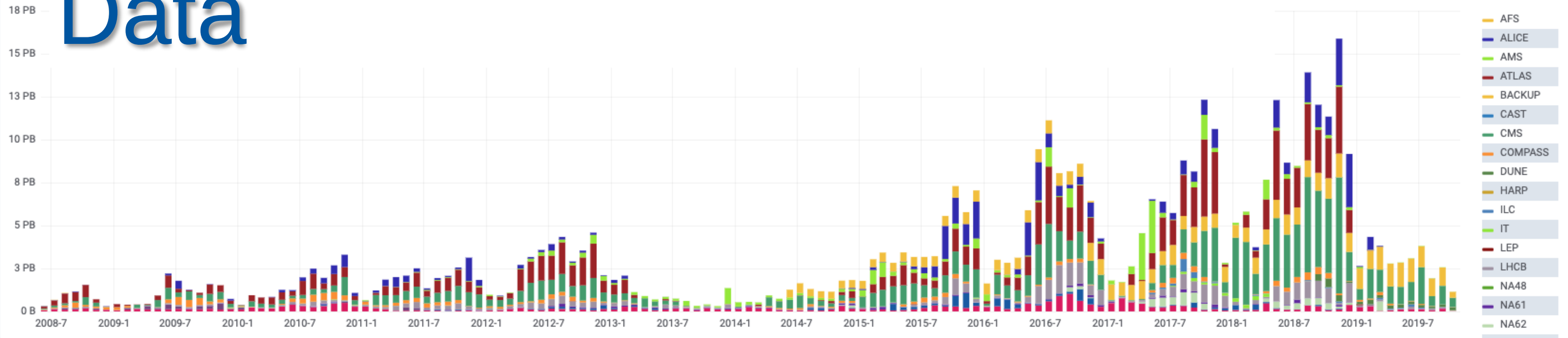
WLCG Status Report

Ian Bird
Computing RRB
CERN, 20th October 2019

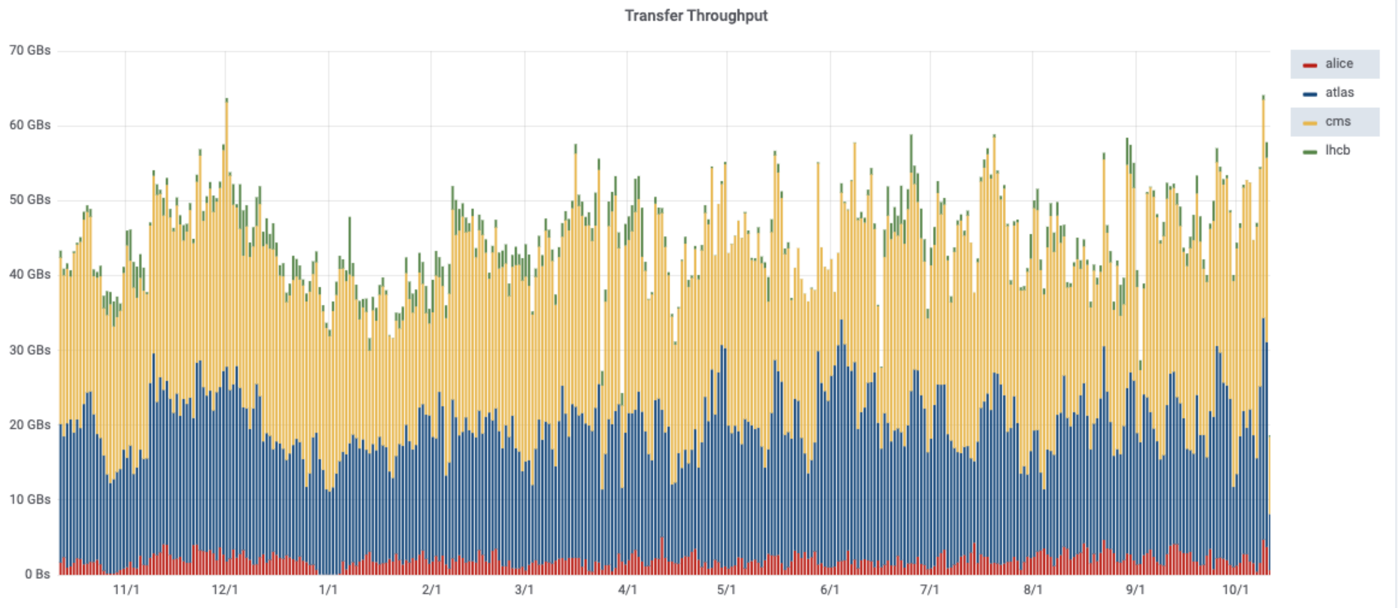
WLCG Collaboration



Data

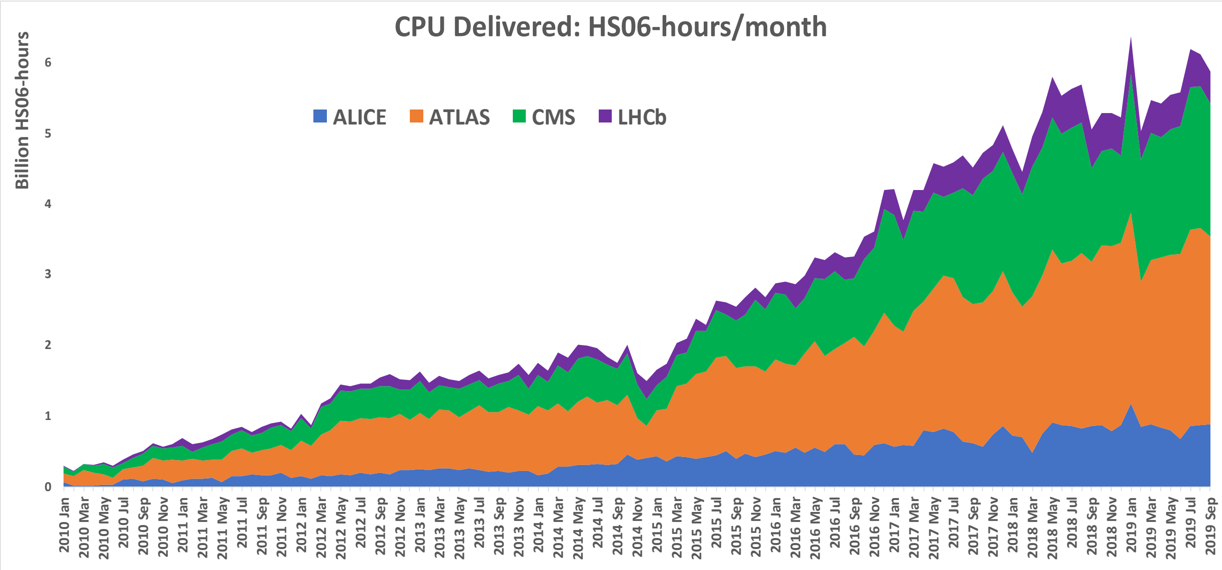


Data transfers



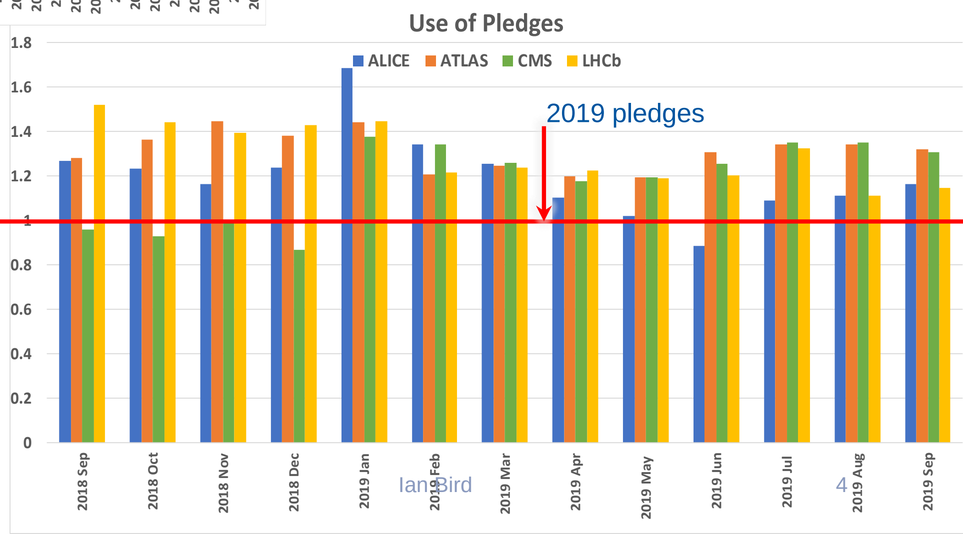
CERN | WLCG

CPU Delivered

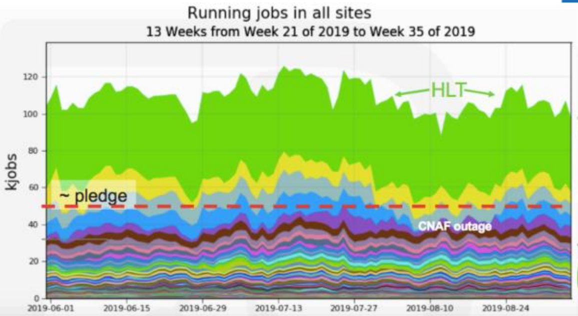


New peak: ~270 M HS06-days/month

~ 860 k cores continuous



LHCb



Production update: Summer 2019

ATLAS EXPERIMENT

Smooth operation over the summer

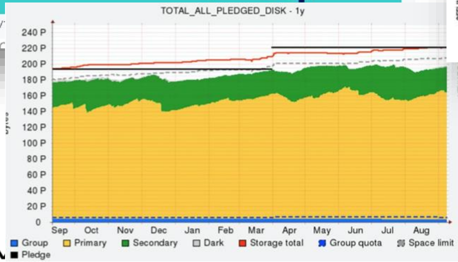
Successful production on Grid, Tier0, HLT (Mostly MC production to support Run-2 analysis)

Expect to operate in a very similar way for the next 18 months (Great thanks to our operations team!)

Storage:

Applied lifetime model deletion on disk before the summer (and one planned soon)

Applied lifetime model deletion on tape last summer (so tapes can be repacked before Run-3)



Release 22: deployment of AthenaMT

Currently running in multithreaded mode (offline and online) with one thread

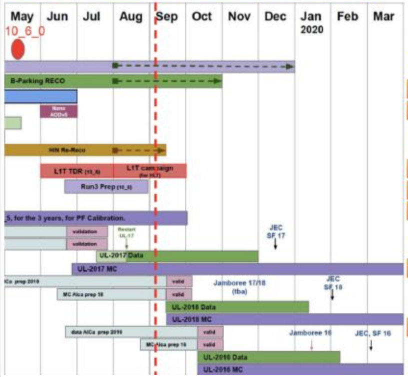
Multiple threads expected towards the end of the year

Performance and optimisation in 2020

Run-2 reprocessing planned with release 22 in 2020-21

Same release for Run-2 and Run-3 analysis

CMS Production and Release Plan



Activities

Concluded

- Samples to prepare Run III (this round)
- L1 TDR samples

Finalizing

- Run II MC pileup library
- B parking datasets
- Heavy Ion datasets

Ongoing

- 2017 ultra legacy samples (data and MC)

About to start

- 2018 ultra legacy samples (data and MC)

Legacy processing on-going. To be finalised in the next 1.5 years

12/09/2019

ALICE computing resources

Physics analysis and Pass2/3 of 2018 RAW data are ongoing

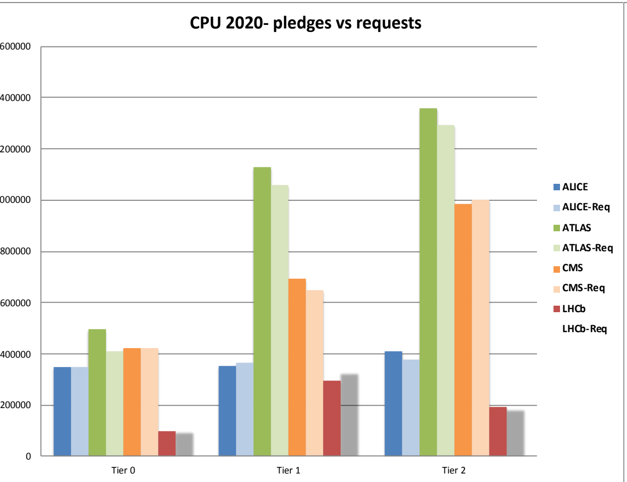
Continuous full utilization of the computing resources

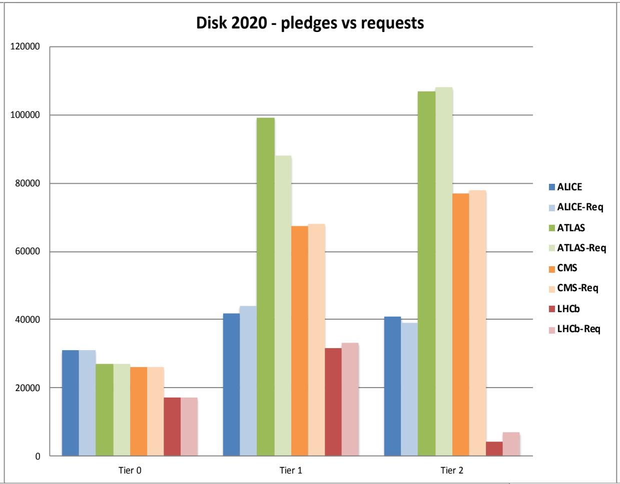
ALICE O2

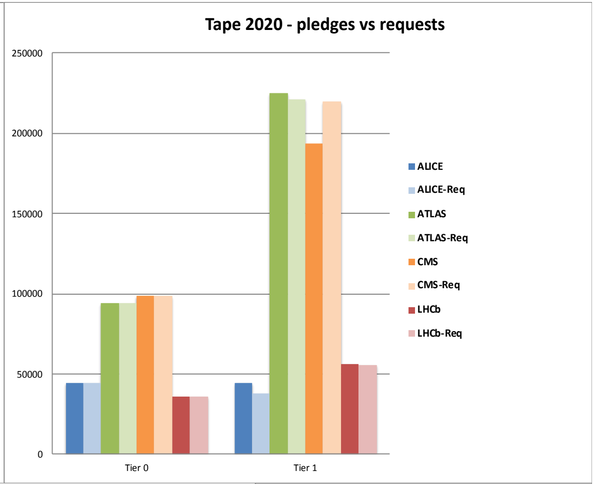
Development almost completed for all large and majority of small detectors

Code to be tested during the vertical test exercise (Infiniband purchase delayed)

2020 Pledge situation – no change





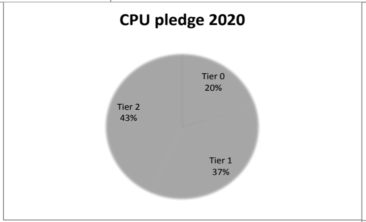


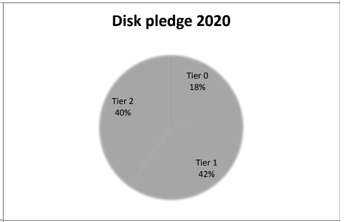
2020 pledges wrt requests:

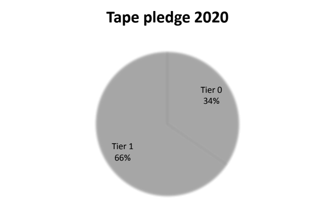
As given in REBUS

NB Very delayed pledges; still missing:

- Russia NRC-KI (T1)
- Greece: U. Ioannina, Kavala Inst.
- Mexico: UNAM
- Pakistan: ALICE Federation
- Spain: LHCb Federation
- Taiwan: CMS Federation
- Thailand: National e-Science consortium
- Turkey: Tier-2 Federation







Run 3 expectations

- Bunch intensities ramp up from 0 to 1.4e11ppb over the year
  - with limited availability of the injectors/LHC resulting in **only 20% machine efficiency**.
- For contingency planning, the machine efficiency assumed to reach normal value of 50%. This results in the following luminosity envelope:

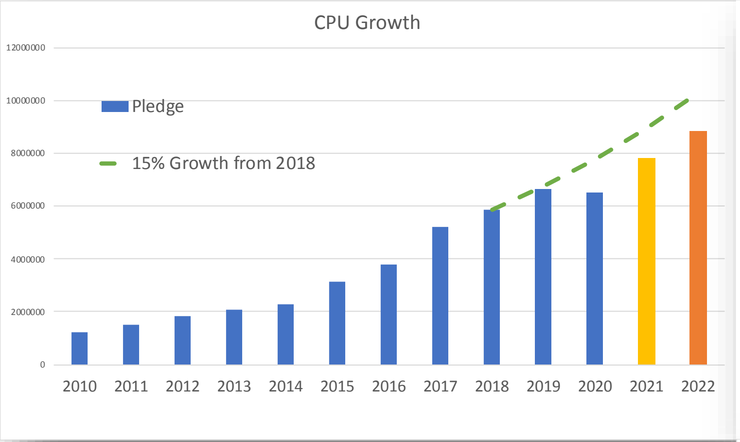
	Baseline	Upper limit
ATLAS / CMS	17 fb^{-1}	42 fb^{-1}
LHCb	3 fb^{-1}	7 fb^{-1}
ALICE	36 pb^{-1}	90 pb^{-1}

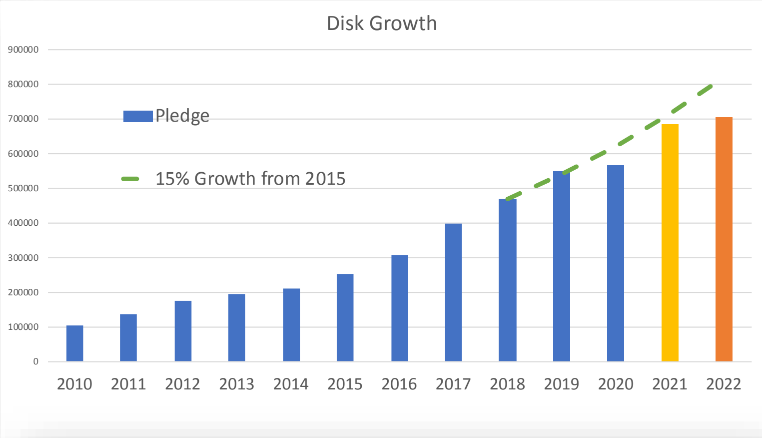
- NB. The upper limit is **contingency planning only** (i.e. raw data tape storage), **not physics**.
- Pb-Pb assumed to be a full production year: >2 nb^{-1} for ATLAS, ALICE and CMS.
- 2022: We assume a full production year with 1.5 x 2018 resource levels

- To be updated once running conditions better specified (End Nov)
  - In particular different assumptions on pileup (55 vs 45) will make noticeable difference

Resource evolution

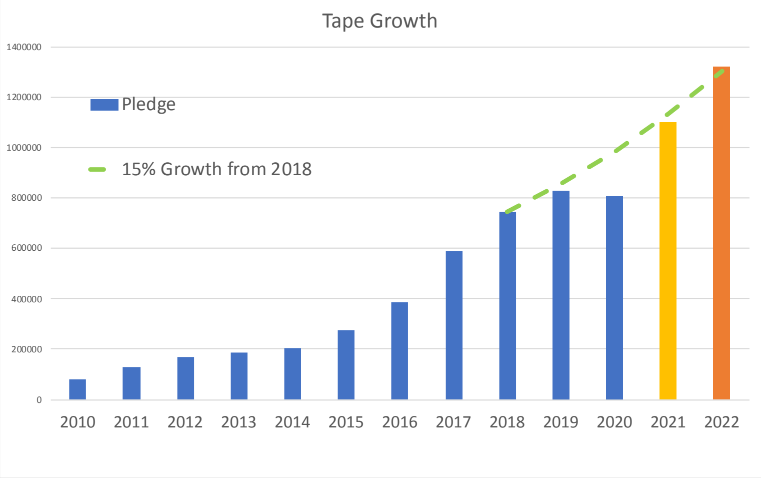
- Actual installed
- 2021 request to C-RSG
- 2022 = 1.5*2018





NB: Run 3 probably manageable overall, ***but*** constant budget growth until Run 4 is essential for HL-LHC

NB2: 15% maybe an optimistic assumption – many indications that flat budget is much worse (~ 10%)



WLCG Funding & Expenditure

LHC Computing Funding and Expenditure

Result 2018, estimates 2019...2022

All figures in MCHF; data extracted on 04 October 2019

	2018	2019	2020	2021	2022
Funding					
From CERN budget[1]					
Personnel	17.4	18.6	17.9	17.3	17.2
Material[2]	11.4	14.5	20.5	25.9	16.6
Total funding	**28.8**	**33.1**	**38.4**	**43.2**	**33.8**
Expenditure					
Personnel[3]	17.4	18.0	17.6	17.0	16.7
Material	11.4	15.5	15.3	29.2	15.6
Total expenditure	**28.8**	**33.5**	**32.9**	**46.2**	**32.3**
Balance personnel	**0.0**	**0.6**	**0.3**	**0.3**	**0.5**
Balance material	**0.0**	**-1.0**	**5.2**	**-3.3**	**1.0**

1) Internal budget 2019
2) Includes carry-forward/carry-back, EUR/CHF exchange rate penalty and negative CVI
3) Excluding data centre operations

Personnel: balanced situation

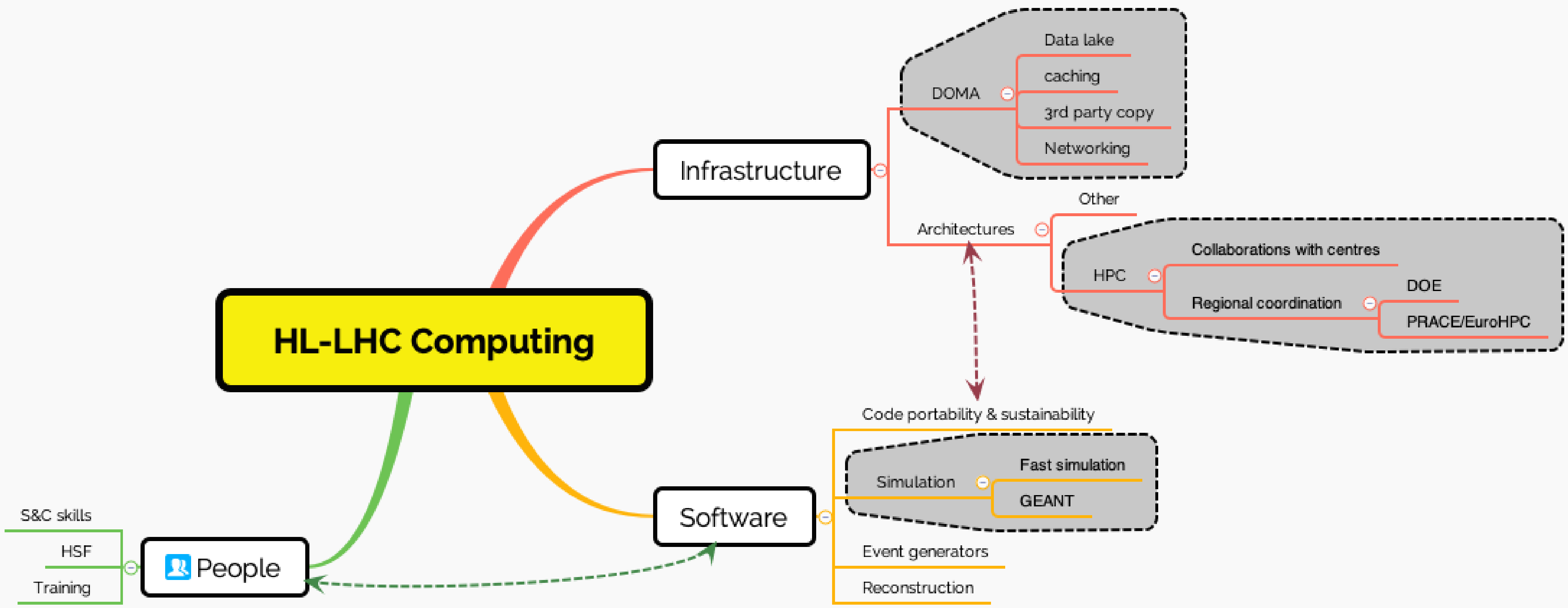
Materials planning based on currently understood parameters:

- CERN plan for 2019,20 is minimal purchases –
- 2021 – as described – to re-assess
- 2022 assumes 1.5 x 2018
- Cost extrapolations updated based on recent experience;
- Large uncertainties and variations
- Savings in LS2 have been re-profiled – overall balance for ramp up to Run 3 now OK

Assumptions on Run 3 running conditions will be updated at the end of the year

Longer term planning

Towards HL-LHC



DOMA in a nutshell

DOMA project
(Data Organization, Management, Access)

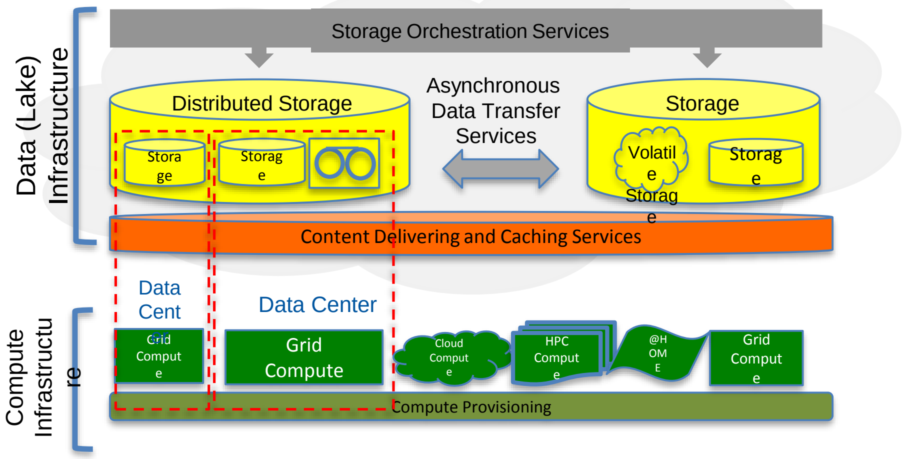
A set of R&D activities evaluating components and techniques to build a common HEP data cloud

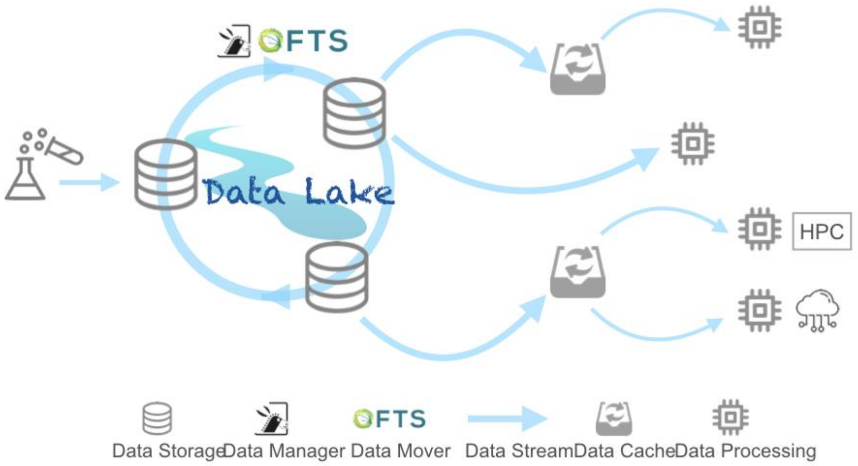
Three Working Groups

- ACCESS for Content Delivery and Caching
- TPC for Third Party Copy
- QoS for storage Quality of Service

And many activities, reporting regularly

https://twiki.cern.ch/twiki/bin/view/LCG/DomaActivities

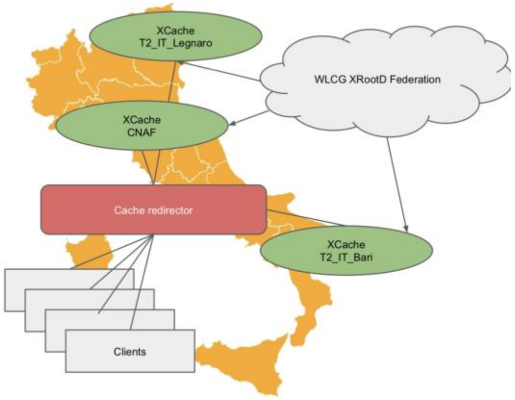


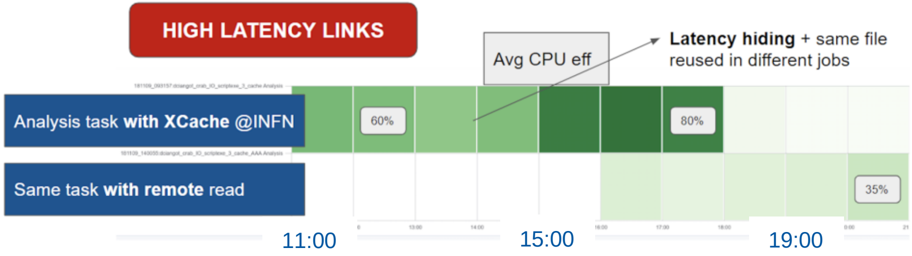


From Simone Campana @ LHCC 10/09/19

ACCESS: caching layer prototype

A distributed caching system in INFN



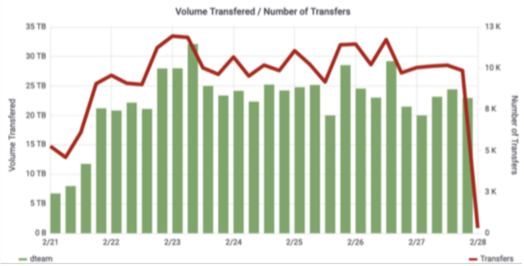


From Simone Campana @ LHCC 10/09/19

TPC

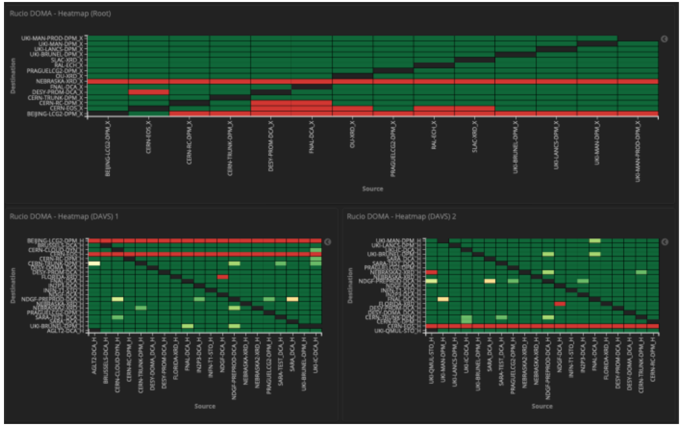
Goal: commission non-gridFTP protocols for asynchronous data transfer (Third Party Copy)

- Phase-2 (deadline June 2019): all sites providing > 3PB of storage to WLCG should provide a non-gridFTP endpoint in production



Functional and Stress testing

Capable to fill available bandwidth



Point-point functional testing

- Phase-3 (Dec 2019): all sites to have a non-gridFTP endpoint

NB: some features needed for TPC are available only in recent versions of storage

From Simone Campana @ LHCC 10/09/19

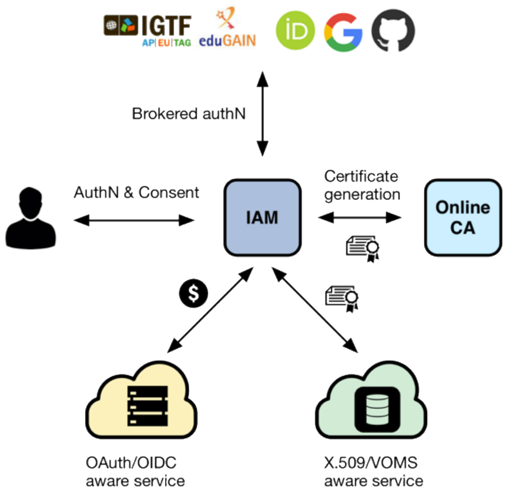
TPC and AAI

WLCG is planning to evolve AAI toward token based Auth/AuthZ and Federated Identities

The WLCG task force is finalizing the token profile as last item

While this is has a much broader scope than DOMA, TPC offers a well confined use case to start with

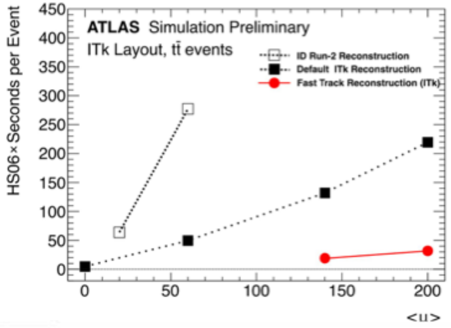
Rucio is integrating tokens. Storage is preparing to manage them.

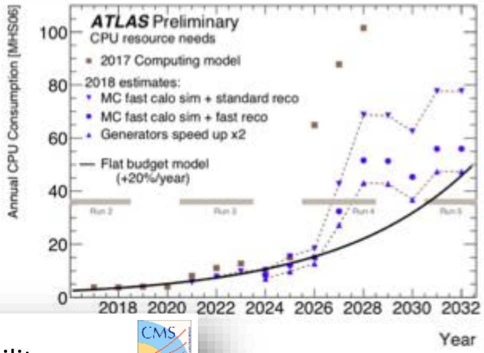


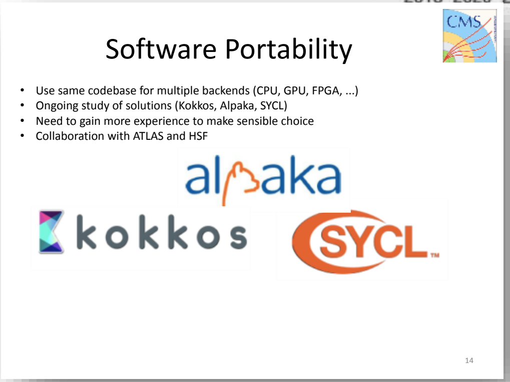
From Simone Campana @ LHCC 10/09/19

Software topics

- Several active HSF working groups
  - Event generators
    - Several workshops and meetings
  - Reconstruction and software triggers
    - Common topics: GPUs, real time analysis, links to other communities
  - Data Analysis working group
    - From DOMA to final analysis
    - Future analysis models, role of ML, etc.
  - Software frameworks
    - Just set up, conveners nominated
- Lots of work in experiments on software portability and performance
  - Use of HPC
  - Lots of work on tuning simulation; fast simulation (and where it is appropriate)
  - Performance and portability:
    - Adaptation of frameworks to accommodate heterogenous code (CPU+accelerators)
    - Portability libraries: Kokkos, Alpaka, SYCL, etc
      - Can there be one codebase for all architectures?







Simulation

- Is a major cost driver (~50% total computing cost)
- Long term supportability/portability/performance is essential
  - Must ensure code modernization & long term supportability, adaptability to changing computing landscape,
    - In a sustainable way - Not as one-off to e.g. GPU-version-x
  - Lot of effort in the world on portability to new architectures
  - Need a major effort on simulation for the future to tie in all of these R&D efforts
  - This is going to be a many-year effort
- This is where we really need to invest effort in the future
  - And is a significant opportunity

Modernisation and Tuning of Software

- **Modernise and improve the performance of the CMS sw stack**

Why?

- Accommodate within computing resources an **ambitious Run3 Physics program**
- **Be ready for Run4**
  - Use Run3 also to test solutions targeting Run4

How?

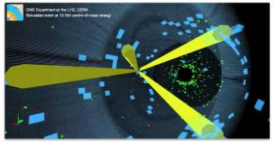
- **Optimisations**: technical (e.g. compiler flags) and algorithmic in CPU code
- **Size reduction** of AOD(Sim) and RAW on storage media
  - E.g. compression settings/algorithm, precision, content of tiers, row Vs. columnar storage
- **Accommodate in CMSSW heterogeneous code**, i.e. CPU + accelerator (e.g. GPU)
  - Evolve respecting present CMSSW architecture
  - Identify the right tools for **performance portability: one codebase for all architectures**
  - Start from framework and high level trigger code

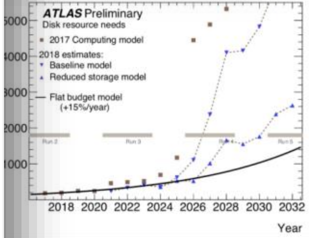
Running on GPUs and other accelerators for HL-LHC

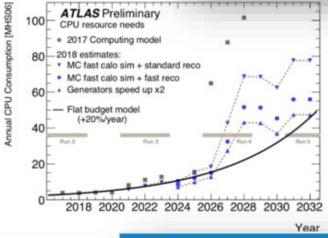
- Resources available to us on acceleration hardware (opportunistic)
  - Much debated topic in the past year or so
  - Complementary to non-x86 resouces that could be used by recompiling the SW stack
- Prototype of Fast Simulation in CUDA
  - Self-contained kernel, collaboration with computing scientists
- ACTS module for GPUs in initial design phase (IRIS-HEP)
- Cross experiment initiatives:
  - Prospects of running event generators on GPUs.
  - Geant4 GPU kernels (?)
- Focus on frameworks for running on heterogenous resources
  - Two ATLAS senior developers charged with accelerators R&D
  - Current prototypes in CUDA running on NVIDIA GPGPUs.
    - Issue with sustainability, code duplication and validation
    - How practical is this outside Online or other contained environments?
    - How do we keep both CPU and GPU busy?
  - We need to focus on **portability**, kokkos, SYCL
  - Not all HEP code suitable for GPUs
  - The technology is evolving (We will soon evaluate Intel's OneAPI beta)

Showcase: Improvements in Simulation

- Continuing efforts to improve performance of FullSim
  - **Preliminary result: 20-30% runtime reduction possible for Run3**
- Several elements to achieve this success:
  - Switch from Geant4 10.4 to 10.6
  - Tune energy-dependent propagation through EM fields (*smart tracking*)
  - Optimize usage of the VecGeom library
- Investigating technical solution to run simulation efficiently on HPCs with accelerators

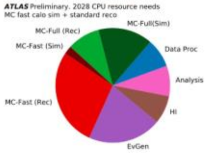


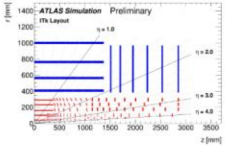




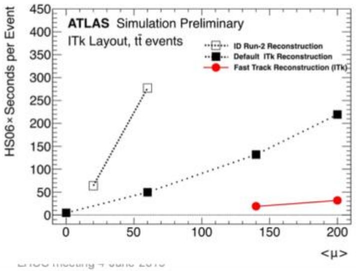
Tracking reconstruction improvements

Need to speed up reconstruction at high <mu>
Optimised tracker (ITk) with x10 more channels





- Optimised track selection
- Improved seeding algorithm (for ITk)
- Omission of ambiguity resolving (to be partly recovered by the new fitter)



Upgrading the O&C Software Toolset

- Sustainability of software tools on the Run 4 timescale is a concern
- Strategy: turn to common solutions, put in production the products already for Run3

CRIC Computing Resource Information Catalogue (used by Atlas et al.)

- Access physical and CMS logical computing resources
- Replace Information System
- Already there

DD4HEP (used by ILC/CLIC, evaluated by LHCb)

- Detector description tool, EU financed (AIDA 2020)
- Review and optimize current detector description tool!
- Steady progress, replacement planned next year

Rucio (originated in Atlas, rapidly growing adoption!)

- Data management solution replacing Phedex / Dynamo
- Steady progress, looking for power users this fall
- One big step forward: transfer ownership of NanoAOD to Rucio

Potential mitigation of costs and improved sustainability: common solutions with industry and other experiments

HPC Challenges

- Draft discussion document on challenges related to being able access and use large HPC
  - Policy & technical
- Working group on how to value HPC cycles for pledges and accounting
  - Very complex
  - Hand-in-hand with next round of benchmarking using suite of experiment codes
  - “HPC” here means GPU and non-x86
- Heading for a future where not all workloads will be efficient on some architectures → complexity and inefficiency
- In addition there is the software portability and sustainability challenge

Common challenges for HPC integration into LHC computing [1]

Motivation

With the detector upgrades for ALICE and LHCb in Run3 (2021-2023) and the HL-LHC accelerator upgrades in Run4 (2026-2028), which will increase the data rates in ATLAS and CMS, the LHC experiments are facing unprecedented computing challenges in the near future. With no changes to the computing models of the experiments the resource gap would be at least a factor of 10 over the gains expected from technology alone and a constant budget: the experiments are working on advances to the software and operating models that will improve the situation, still significant additional processing and storage resources will be needed on the time scale of Run3 and Run4.

High Performance Computing Centers (HPC) are some of the largest processing resources accessible to science applications. They are centers of expertise for computing at large scale with low latency local networking. The efficient usage of HPC facilities may provide a substantial contribution to the success of future LHC data processing by providing much needed computing capacity. R&D investigations are being performed in order to harness the power provided by these facilities and evolve the experiments’ computing models to include their usage. HPC facilities represent a unique challenge and opportunity as they are early adopters of technology including heterogenous accelerated computing architectures.

The experiments have compiled HPC-related documents, including the summary of a joint meeting on this subject [1][2][3]. This document intends to extract the commonalities between experiments with the aim of developing a joint roadmap and strategy for enabling the exploitation of HPC resources. To develop common approaches between experiments and HPC sites, a foundation and understanding of the problems is needed. This is built on a summary of technical challenges, described in section 2. They are broken into two main categories: computing resource challenges and software and architecture challenges. Computing resource challenges describe issues related to operations, facility access, provisioning, and monitoring; while software and architecture challenges are related to adapting HEP applications to make effective use of alternative architectures often found on HPC. In order to explore potential solutions a number of pilot demonstrators are proposed in section 3 below.

1. Status

All LHC experiments report using some HPC resources with varying degrees of both success and technical difficulty. Accessing HPC sites with both workflows and data needs a level of customization. Development of applications for HPC centers has been more successful when the site architectures are the most similar to the generic x86 systems used

[1] **DRAFT WLGB/MB/2019-3.** Editor: Maria Girone (maria.girone@cern.ch). Contributions: Gavin McCance, Xavier Espinal, Domenico Giordano, Hannah Short

1

Summary

- LS2 is busy for the experiments & facilities
  - Ongoing processing, analysis, etc.
  - Preparations for Run 3 – simulations, software preparation, etc.
- Run 3 looks like an evolution of Run 2 for ATLAS and CMS
  - LHCb & ALICE major changes – but sw & computing preparations in hand
  - Resource outlook seems realistic
- Data preservation , open access workshop to be held 26 Nov
  - Initial discussion to align and agree policies, strategies, goals, and resource needs
- Further outlook to HL-LHC
  - Many R&D topics progressing well
  - Significant work in experiments closing the gap between requirements and likely resources
    - Although the cost evolution of hardware is a major concern
  - Software challenges are potentially significant – but are opportunities for the longer term sustainability
- LHCC will hold a review of HL-LHC computing preparations in ~Spring 2020