# R&D for HL-LHC from the CWP

Ian Bird
Simone Campana
Maria Girone

# Introduction

- The HSF Community White Paper defines a roadmap for HEP software and computing R&Ds for the 2020s

    - ➢ For WLCG this is crucial in preparation for HL-LHC
    - ➢ Other research communities with computing needs at the level of WLCG will coexist on the same infrastructure

- The CWP consists of 13 work packages, each one defining a set of R&Ds

- Here we focus on aspects of the "Facilities and Distributed Computing" paper, understanding that the work packages are highly dependent one from the other

# Storage Consolidation

- Many Funding agencies wish to consolidate storage in a region, to optimize operational cost, with in mind also other communities beyond HEP

- The consolidation might be physical: reduce the number centers offering a storage service

- Or logical: implement a distributed storage system across centers

- This R&D should evaluate which storage solutions are suitable to implement a distributed storage system and build prototypes based on the suitable solutions

- The R&D should also acquire experience with data processing and data management for those prototypes

# Caching Technologies

- Consolidating storage implies breaking CPU/data co-location, further than now

- This R&D should evaluate the need for a storage caching layer and for which use cases

- Assuming caching is needed, the R&D should prototype solutions and evaluate the performance benefits

- The evaluation should consider benefits in terms of performance in processing data and effort in operating the infrastructure

# Data Lakes

- Do not look at the definition in Wikipedia because this is not what we mean
  - https://en.wikipedia.org/wiki/Data_lake

- The WLCG data lake(s) is(are) an extension of the storage consolidation R&D, evaluating a distributed storage system deployed across centers connected by fat networks (10Tb in 2025) and operated as a single service

- Such storage would host a large fraction of the data and optimize the cost, eliminating inefficiencies due to fragmentation. Aspects to be studied:

  - How to leverage replication across different classes of storage (tape/disk/SSD/..) to ensure data retention and accessibility, while reducing cost

  - How to implement data "recovery" as built in storage functionality, reducing operational effort

  - How to define a security model leveraging the trust between data centers to minimize overheads while providing a secure environment

  - Which are the best protocols for internal storage management, data access and data management, data transfer (in close relation with the Data Management CWP)

# Other R&Ds in the scope of "Facilities and Computing Models"

- Build a cost Model to better understand the relationship between performance and cost

- Define common data management functionalities required by experiments

- Explore scalable and uniform means of workload scheduling, capable to incorporate dynamic and heterogeneous resources, and providing fine-grained processing capabilities

- Prototype a quasi-interactive analysis facility

# Conclusions

- The CWP roadmap document highlights lot of challenges and opportunities to prepare for HL-LHC

- WLCG started taking on board a few key aspects on the Facilities and Computing Models

- We are at the level of R&D, so we should prototype quick solutions and make this an interactive process

- Data Management R&Ds will be one of the main topics of the WLCG/HSF workshop in March and we should arrive there prepared with concrete ideas and possibly some early prototype