



# **Proposed New Scope of DOMA Access**

**Frank Würthwein  
UCSD/SDSC**

**June 24th 2020**



# Dual Scope: DL & DC

- Data Lake Prototyping (DL)
- Towards a Data Challenge for Production Processing (DC)

# Data Challenge (DC) for Production Processing

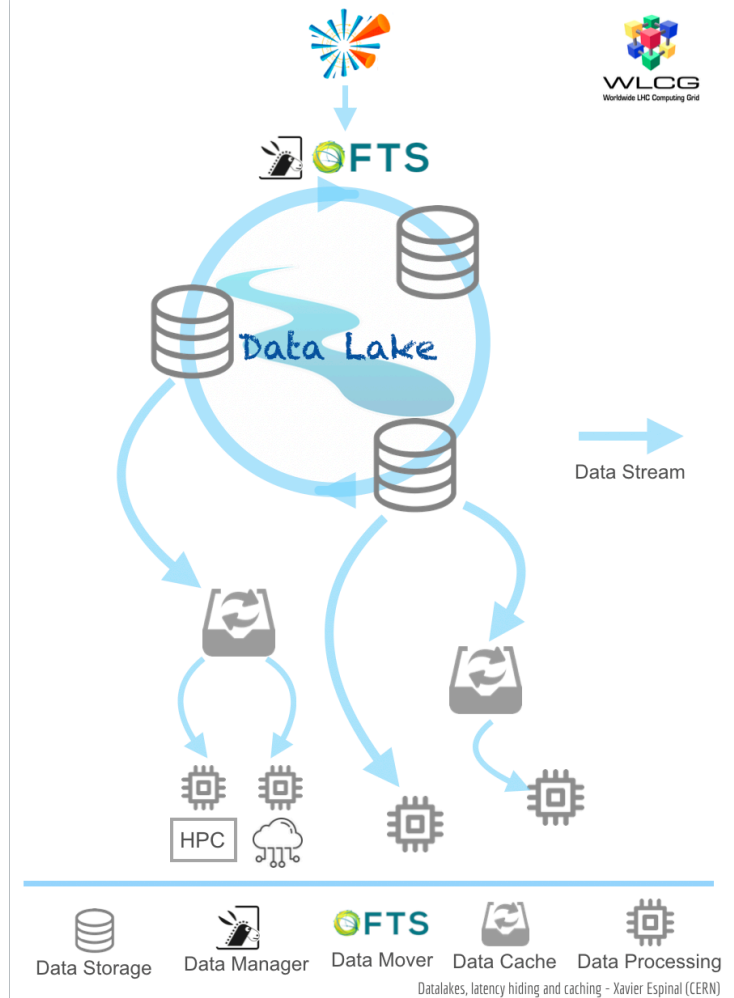
- During nominal year of HL-LHC, ATLAS and CMS will take  $\frac{1}{2}$  an exabyte of RAW data each.
  - Want to process an exabyte “as quickly as possible” after the end of the annual data taking (well ... early in the year after)
- Exabyte/100 days  $\sim$  10PB/day  $\sim$  1Tbit/sec
- See DOMA Access meetings of May 5<sup>th</sup> and June 2<sup>nd</sup>



# DC Goals for 2020

- Write a document that describes a straw proposal for how ATLAS and CMS will process  $\frac{1}{2}$  exabyte of RAW data each during 100 days on overlapping infrastructure.
  - Document relevant numbers from the experiments planning, as presented to the LHCC.
  - Make up & document some simple assumptions
  - Document the implications for the infrastructure from archives at the T1s and CERN to disk staging areas to networks to disk staging at processing centers, incl. large HPC systems, ... **“conversation starter”**
- Have reports on
  - scalability of existing tools like Rucio, FTS.
  - Tape retrieval bandwidth measurements at T1s
  - past RAW data processing campaigns with regard to filesizes etc.
  - “on the fly” reformatting from large files to event service ideas and status.
  - .... (your favorite topic here)

- Example topics:
  - Regional prototypes of
    - Caching, remote access, diskless sites, ...
    - New caching algorithms
  - Scale testing and hardening
  - Integration of HPC
  - Cross lake replication
  - DOMA aspects of Analysis Facilities prototypes
  - ...





# DL Goals for 2020

- Continue documenting progress with caching infrastructures
  - Site level
  - Regional level
  - Status of software versions
    - Functionality, scale testing, production experience.
- Reports on ideas & prototyping for DOMA aspects of Analysis Facilities
  - “on-the-fly” reformatting from archived files to columnar data stores
  - Storage implications for columnar data stores

# Comment & Questions

As usual, everybody is invited to contribute to these meetings with their ideas, prototypes, ...