

# DOMA ACCESS WG update

Xavier Espinal on behalf of the WG coordinators

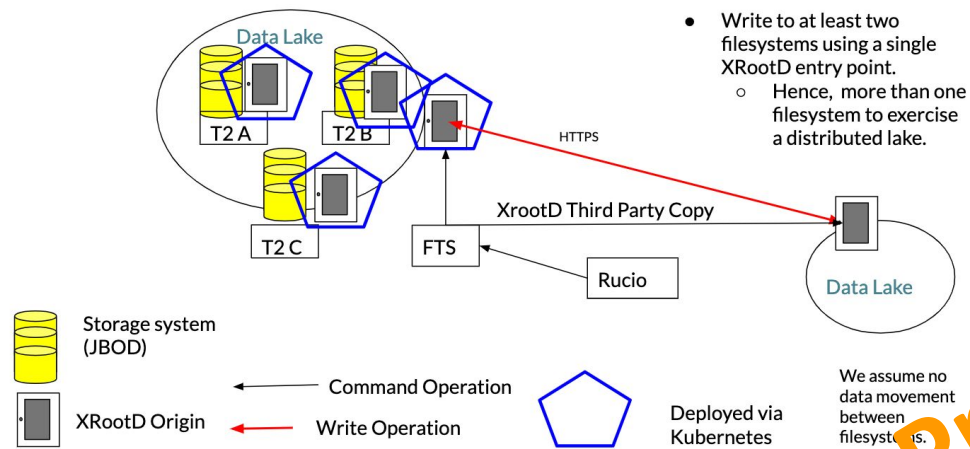
# Some background

- Mar 2020: [Document](#) wrapping up the first mandate of the WG.
  - Input for the HL-LHC review
- Mar 2020: DOMA ACCESS evolution proposed
  - <https://indico.cern.ch/event/901410/>
- May 2020: Proposal to focus on Datalake Prototypes and Data Challenges
  - <https://indico.cern.ch/event/923189/>
- Jun 2020: Proposal Data Challenge for Production Processing (10PB/day) -
  - <https://indico.cern.ch/event/932079/>
- Jul 2020: Brainstorming discussion on the datalake challenge and CMS and ATLAS use cases
  - <https://indico.cern.ch/event/939368/>
- **Now:** Datalake prototyping proposals by CMS (and ATLAS soon)
  - <https://indico.cern.ch/event/953032/>

# Planning for September/Q4

- September 15th: Data Lake prototyping (<https://indico.cern.ch/event/953032/>)
- September 29th: Archival bandwidth at T1s (<https://indico.cern.ch/event/954845/>)
  - The 10PB/day challenge: ATLAS and CMS will process ½ exabyte of RAW data each during 100 days on overlapping infrastructure
- October 13th & 20th: Impact of Data Lake Model on Total Cost of Ownership
  - We want to listen opinions from several Tier-2s, regional centers.
- Preparations for the Storage Workshop in November (*tentative dates: 19/20 & 23/24*) in parallel to WLCG-HSF meeting

# Adding data into Data Lake

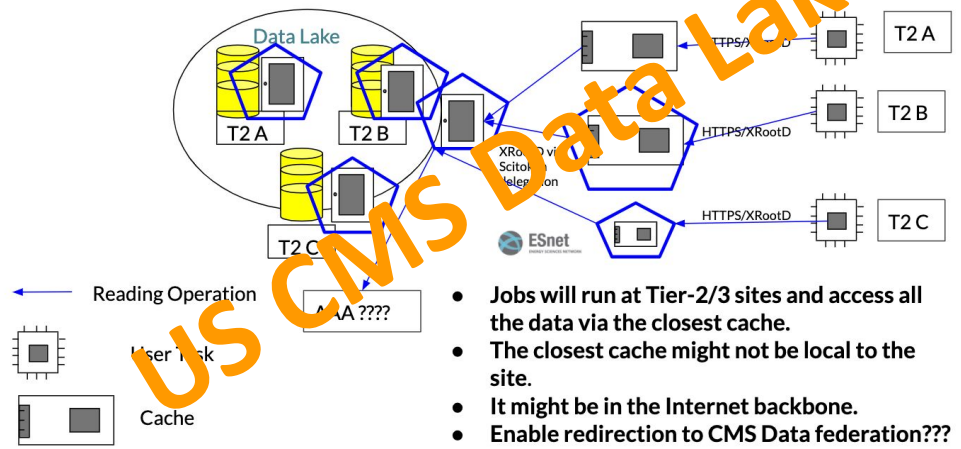


- Write to at least two filesystems using a single XRootD entry point.
  - Hence, more than one filesystem to exercise a distributed lake.

# Proposed Timeline for prototype Deployment

Task	By
All hardware for prototype in Kubernetes cluster	September 2020
Setup the XRootD origins and configure them with a the data lake single entry point.	October 2020
Configure caches to read from the data lake and use Scitokens for authentication.	November 2020
Setup a ... (R...E) in RUCIO (UST2DataLake) and have ... register all NANOAOB to it.	December 2020
Setup submission infrastructure to be cache aware.	January 2021
Data lake testing, benchmarking and DevOps	January 2021 - September 2021

# Accessing data in lake for processing



- **Jobs will run at Tier-2/3 sites and access all the data via the closest cache.**
- **The closest cache might not be local to the site.**
- **It might be in the Internet backbone.**
- **Enable redirection to CMS Data federation???**

# Benchmarking Goals

1. Exercise deletions and measure missed deletions as a function of:
  - a. Scale
  - b. Disconnecting an XRootD origin
2. Exercise data input and data removal via FTS
  - a. Scale
  - b. Success rates
3. Exercise NanoAOD application access.
  - a. Recruit students and postdocs with realistic applications
  - b. Cpu efficiency as a function of RTT to the closest cache.
  - c. Data access pattern (?to be thought about more carefully?)

- Many datalake components already included in ATLAS Grid (Rucio, FTS,...)
  - Adiabatic changes from existing infrastructure
  - Evolution to be endorsed within ATLAS
- Autumn 2020 : Steps towards datalake (US, EU) carefully monitored
  - Datalake include few sites
  - Existing monitoring to be consolidated (Network)
  - Measure performances :
    - benchmark jobs (local analysis, HammerCloud) : Existing at local level
    - VP within US
- 2021 :
  - Embark more Grid/local sites
  - Larger diversity of jobs
- Presentation in DOMA ACCESS to be scheduled

# Planning for September/Q4

- September 15th: Data Lake prototyping (<https://indico.cern.ch/event/953032/>)
- September 29th: Archival bandwidth at T1s (<https://indico.cern.ch/event/954845/>)
  - The 10PB /day challenge: ATLAS and CMS will process  $\frac{1}{2}$  exabyte of RAW data each during 100 days on overlapping infrastructure
  - Evaluation of all potential read/write concurrent activities using Tape drives (RAW reprocessing, RAW data expert during data taking, archived AOD → DAOD)
- October 13th & 20th: Impact of Data Lake Model on Total Cost of Ownership
  - We want to listen opinions from several Tier-2s, regional centers.
- Preparations for the Storage Workshop in November (*tentative dates: 19/20 & 23/24*) in parallel to WLCG-HSF meeting

# Technical Challenges



- Process 10PB of data in a single day
  - Tape recall
    - How much bandwidth can we achieve from tape?
    - What's reasonable for buffer sizes and tape bandwidth?
  - Manage the limited disk buffer at archival T1
    - Tape recalls will be carousel style, i.e. buffer much smaller than the exabyte dataset.
  - Manage 1Tbit/sec network to an HPC center
    - Network bandwidth needs to be managed with tools like SENSE and AutoGOLE
  - Manage the disk buffer at the HPC center
- Co-schedule processing and all of the above.

# FABRIC Core



# Proposal



- We get organized and apply to use this tested for a variety of tests that build up over time to the 10PB/day data processing challenge. E.g:
  - Learn how to tag traffic.
  - Learn how to use SF (SF etc. to schedule networks.
  - Benchmark out entire data transfer chain at Tbit/sec (Rucio, FTS, TPC, SENSE ...)
  - Learn how to co-schedule tape, disk, network and processing
- Do all of the above as a program of work over the next 4 years, with the 10PB/day processing as crowning achievement.
- Do it jointly between ATLAS, CMS, IRISHEP, ...

	Run2 repro 2020	All data (2020)	Repro year 2028
Number events (Billions)			70 (1300k * 700k)
Event size (MB)			4.4 (estimate)
Total RAW size (PB)	18		300
Reprocessing time (days)	75 (not permanently full speed)		100
Reprocessing PB per day			3
File size (GB) (estimation)	2.2		10
Staging (GB/s)	15 (sum of max of each site)		40 (steady) → 60 (peak)
Fraction reprocessed outside T1	50% ?		50% ?
WAN traffic of RAW (GB/s)	10	10-20	150
Nb transferred files / hour		100k	12 k
Processing time tbar (HS06.sec)			300 (estimate)
Nb core (1 core=10 HS06)	~ 150k	400k	250 k

# Planning for September/Q4

- September 15th: Data Lake prototyping (<https://indico.cern.ch/event/953032/>)
- September 29th: Archival bandwidth at T1s (<https://indico.cern.ch/event/954845/>)
  - The 10PB/day challenge: ATLAS and CMS will process ½ exabyte of RAW data each during 100 days on overlapping infrastructure
- October 13th & 20th: Impact of Data Lake Model on Total Cost of Ownership
  - We want to listen opinions from several Tier-2s, regional centers
- Preparations for the Storage Workshop in November (*tentative dates: 19/20 & 23/24*) in parallel to WLCG-HSF meeting



# Summary and scope

- Datalake prototypes and data challenges activities planned. Implementations starting.
- Latency hiding and caching infrastructures in place and running - ready to eventually join the prototypes. Strategic component.
- Network and data challenges in parallel. Ensure mechanisms are in place to leverage prototypes and eventual data/computing challenges.
- ESCAPE project prototype progressing on some key aspects:
  - Storage QoS, RUCIO-multiVO, XCache-multiVO, end-to-end access tokens integration

These activities start addressing some of the recommendations received from the HL-LHC computing review: i.e. flexibility for adapting to new political scenarios, R&D on new concepts potentially helping TCO, effort on software components fundamental for the progress of DOMA activities.