Laboratoire d'Annecy de Physique des Particules

ATLAS Data Lake : DOMA ACCESS exercise

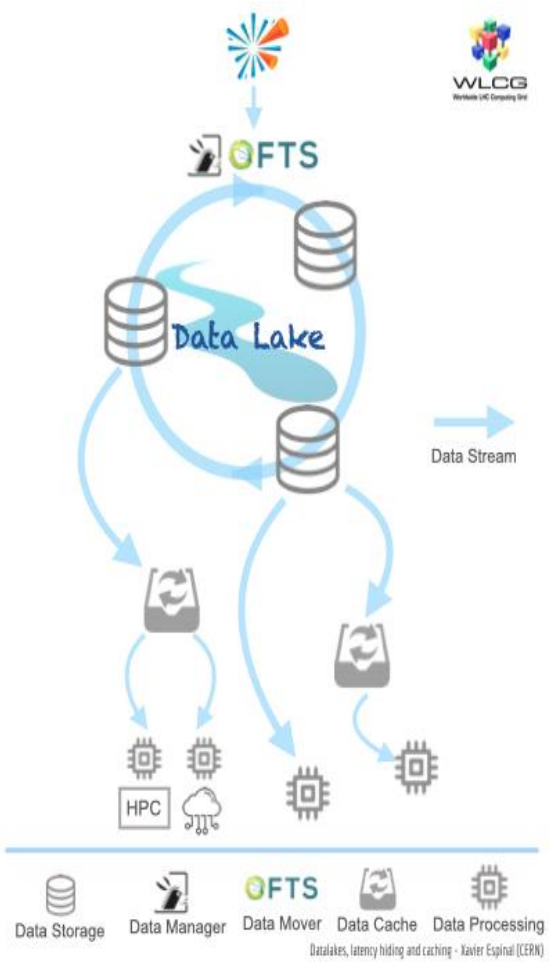
S. Jezequel

Wednesday 28 October 2020

cnrs

UNIVERSITÉ SAVOIE MONT BLANC

Inputs for ATLAS Data Lake model implementation



- IO per core for mixed Prod + Analy: Mean 0.5 MB/s/core
  - Production : Up to 0.4 MB/s/core for derivation
  - Analysis : Observed mean value : 2 MB/s/core
- File replica management within/between Data Lake
  - WLCG Datalake : Rucio RSEs with Grid storage
  - Replica location registered in Rucio catalog
  - Similar as production today : Rucio + FTS
    - Expect new things from QoS part
- File replica in caches :
  - VP - Maximise Hit rate by restricting allowed cached copy to few Xcache
  - Preplacement à la Arc-Cache
- NETWORK :
  - Becomes hard to address only at application level

- No specific data format stored in each Data Lake (eg. DAOD_PHYS(LITE))
- Goal : Measure access performances to Data Lake storages
  - Focus on scalability issues
- US : Organised within VP running analysis jobs
  - Analysis jobs reading any requested file
  - Measure Hit rate and IO rates
- RU/FR : Based on HC measurements reading few files
  - Local vs Remote
  - Academic measurement of performances from some sites
  - RU : Gain with/without Xcache accessing central storage
  - FR : Evaluation data access within distributed federation and further away
- DE :
  - Stress tests with Xcache layer
  - Access performance for short/long distance

CPU_only : PQ + remote Grid storage

- Standard way for jobs submitted through Panda
  - If necessary, input files transfered to associated RSE (read/write)
- Scalability test for production only or prod+ analysis
  - Prod only (copy2scratch(c2s)) :
    - Simul only : UKI-LT2-IC (4k cores), Romania, Austria, Russia
  - Prod(c2s) +analy (direct access- No TtreeCache) : 2.5k cores : Israel
- Still missing :
  - No report of access in rucio monitoring
  - Network conditions critical (bandwidth occupancy, latency)
    - Resilience to degraded network conditions
- Proposition for Data Lake exercise (2020-2021)
  - Run large sites with large IO (prod + analy) : 5k → 10k slots
  - Measure performances with proto Data Lakes

- Read synchronously input files
- Explored through
  - Arc : NDGF-T1
  - VP in production for analysis in few US sites
  - BHAM : Production only (only 300 slots)
  - Russian Data Lake : HC test
- Missing
  - No monitoring if Xcache integrated in Grafana DDM
  - VP : Enforce data accessible only in same Data Lake (Cloud)
- Proposition for datalake exercise 2020-2021:
  - Consolidate VP results
  - Consolidate performance and Hit rate measurements with Xcache
  - Include HPC/Cloud computing

- Local analysis facility
  - Measure degradation of performances compared to local copy
  - Different access methods
    - Xcache
    - Data format (uproot)
- Proposition for Data Lake exercise 2020-2021
  - Run in site at large scale (5k events)
    - Identify large local analysis facility not yet done
    - Analysis code under indentification