Data Movement for the Large Hadron Collider

- The WLCG is a global collaboration of more than 170 computing centres in 42 countries arranged with four tiers:
  - Tier 0 at CERN (and Wigner in Hungary)
  - 13 Tier 1s (mainly national physics laboratories like RAL in UK)
  - 149 Tier 2s (generally university physics laboratories)
  - Tier 3s

- Data transfer within the WLCG occurs mainly through two mechanisms:
  - First: bulk data transfers using movement orchestrators such as Phedex for CMS, Rucio for ATLAS and Dirac for LHCb
    - Data pre-placed at appropriate sites; popular data is replicated dynamically across sites to improve access
    - Reprocessing campaigns can involve large data transfers – required in short period of time (~2-3 days)
  - Second: remote reading from a global data federation using the xrootd protocol (CMS version called ‘AAA’, ATLAS ‘FAX’)

- Remote reading is slightly less efficient;still generally try to send jobs to data where possible

Changing Computing Models

- Increasing data volume and rate drives changes in the computing models for the experiments
- Consolidation of data storage to fewer, larger Tier 2 sites
  - More sites offering processing plus smaller caches
  - Lightweight management for caching sites
- Implied changes for the data movement
  - More use of streamed and cached data at the lightweight sites
  - Increased traffic to and from the Tier 2 storage sites and Tier 1s
- Other drivers:
  - More intense colliding beams, increased energies and increases trigger rates (rates of events accepted for storage and processing)

The ATLAS experiment as an example

- Over the next 4 years we expect a growth of processing and analysis capacity of a factor 2 (roughly 20% growth/year).
- These processes will either fetch data from storage, produce new datasets for storage or both.
- The overall computing model in terms of formats, versions and selections recently revised.
- Expect stability for another 4-5 years; this model requires a factor 2 more network bandwidth compared to the current baseline.

Implications of the Model Changes

- Storage Tier 2s:
  - Example from UK; the large sites saturate the bandwidth in periods of reprocessing at the 10-20Gbps level.
  - Doubling from rate and other changes over the next 2-4 years
  - Doubling because of consolidation of data to these sites
  - Typical sites will need 40-80Gbps in 2-4 years
  - With overheads and other university traffic, top end is advisable
  - 100Gbps network for large T2s is well motivated within 5 years, while in 10 years 1Tbps should be considered
  - Expect storage to be dual stack IP by the end of 2019

- Lightweight sites do not see these pressures.
  - Naive models show 1-4Gbps sufficient now
  - Grow to ~20Gbps over half a decade
  - Lightweight sites need at most 100Gbps in 10 years

Longer term – HL-LHC and Run 4

- Optimistic scenarios & very reduced online storage project another fourfold increase in data volume
- Tier 0 to Tier 1 links must also scale up
  - Anticipate 10 Tbps links by 2027
  - Data movement will use all affordable network uplift
  - New working models & computing models required