PhysX CoE: LHC Data-intensive workflows and data-management

Wahid Bhimji, Pete Clarke, Andrew Washbrook – Edinburgh

And other CoE WP4 people…

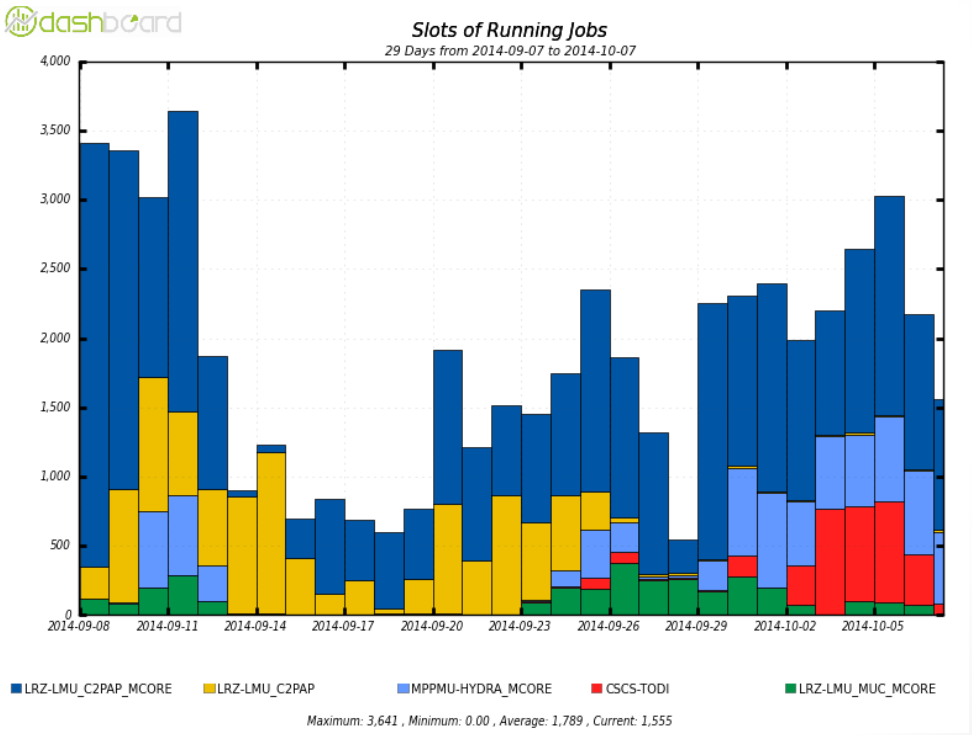
THE UNIVERSITY *of* EDINBURGH

Outline

- Current ‘State-of-the-art’ technology and brief summary of goals of LHC ‘task 4.3’ of proposed CoE:
- “Exploitation of HPC for [LHC] data-intensive workflows”
  - HPC Workflows for LHC
  - Data Processing (I/O)
  - Data Management (transfer, meta-data, storage …)
- Focus on the data management aspects
  - Evolution of WLCG storage / data tools
  - (Possible) relevance for LQCD..
    - but here I’m guessing until the discussion…

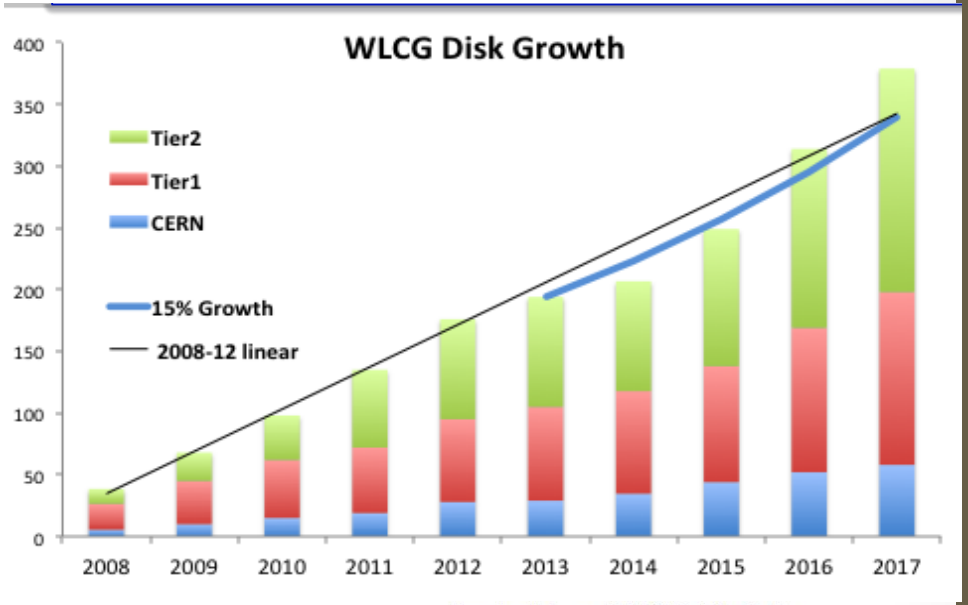
LHC Workflows on HPC

- See https://agenda.infn.it/conferenceDisplay.py?confId=8521
- Existing work – some recent successes
  - Mostly focus on simulation (event generation and detector simulation)
- HPC centre constraints (e.g. network, compute node storage ..):
  - Vary between centres and evolving . – ad-hoc solutions.
- CoE aims:
  - Production services
  - Extend to data-intensive workflows.



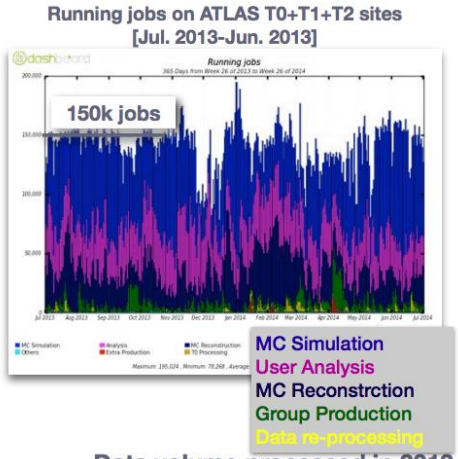
Data in LHC - Context

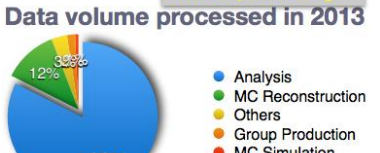
- WLCG 200 PB data stored on disk
- Doubling during Run 2
- 1.2 EB data read by ATLAS jobs in 2013
- WLCG and experiment tools evolving:
  - Scaling; Simplification; Flexibility



ATLAS Grid activity

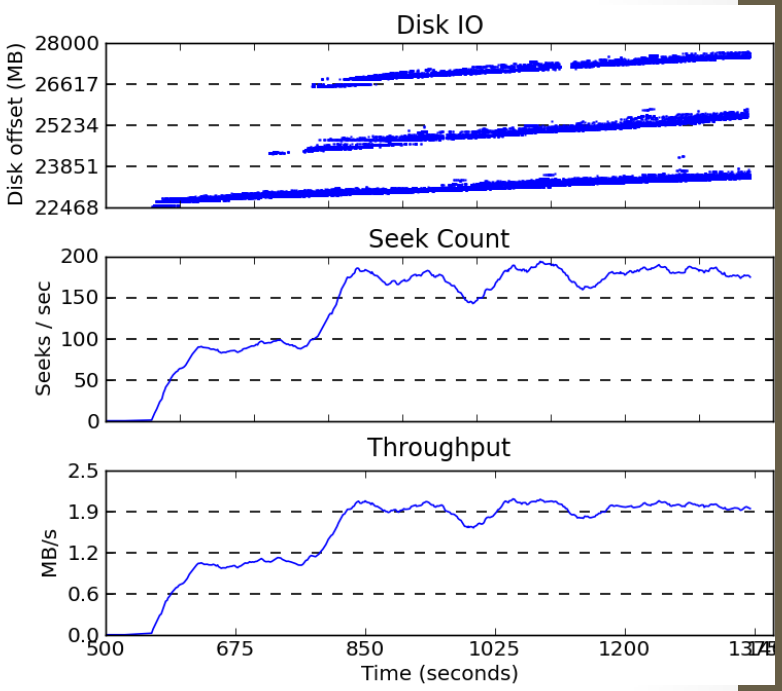
- ~150K concurrent jobs running
- 350M jobs completed in 2013
  - Analysis: >50% of the jobs
- **1.2 EB** of data read-in by ATLAS grid jobs in 2013
  - 82% by analysis jobs
- **Analysis** is the main driver of storage & network I/O capacity





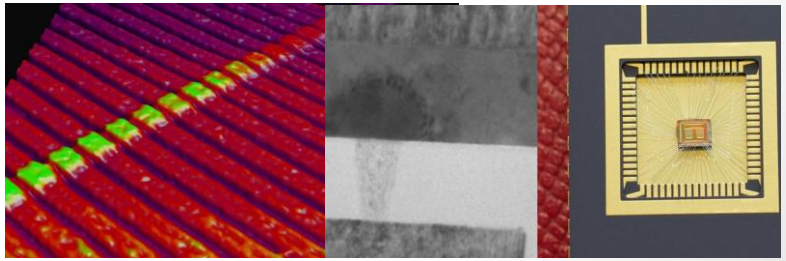
Data Processing (I/O)

- Nearly all LHC I/O is ROOT I/O
- A lot of work by ROOT and LHC experiments on reading only parts of what's needed
- CoE goals:
  - ROOT IO inc. parallelism
  - Use of caching (see details of http / xrootd later)
  - Data delivery solutions extending existing data structures
    - c.f Atlas 'event server' but also beyond…
  - And new hw technologies e.g NVRAM



Memristor

Phase change memory



Data Management – LHC tools

- Placed data - still optimum for known data requirements.
  - new tools FTS3, WebFTS
- Dynamic data ‘federations’ - instead of transferring ahead of time (or in case of broken local replica), exploit improved WAN performance by reading from application over WAN. Present a common namespace across remote sites - without need for a central catalogue.
  - New tools:
    - xrootd-based – now in production, high-performance protocol,
    - http-based: ‘dynamic-federations’, ATLAS Rucio redirector
- Improved (client) tools:
  - Lcg-utils -> Gfal2 – hides multi-protocol complexity
  - Davix – client for performant http(s) I/O
  - Xrootd4
- Decreasing LHC use of SRM , LFC (though these will be available for a while)

Data management – CoE goals and LQCD overlaps

- CoE Goals:
  - 'Transparent' interoperation of diverse data stores
    - Based on federation and caching
- LQCD needs (according to draft CoE text..)
  - Output: 4.4.1: Updated LDG services and tools enabling the users to upload, search and download gauge configurations in a unified, transparent and safe way.
  - Output 4.4.2: Enabling integration of data sharing capabilities into HPC infrastructure
- Possible use of WebFTS to transfer data from HPC centres
- With http-based federation and gfal2 client tools for access?
- Catalogue options (if needed): Rucio (Atlas catalogue) Dirac (LHCb etc.)

Summary

- The proposed Centre of Excellence would take LHC to routine running on HPC
  - Including data-intensive workflows not currently considered
- Requires improvements in I/O, caching and data management

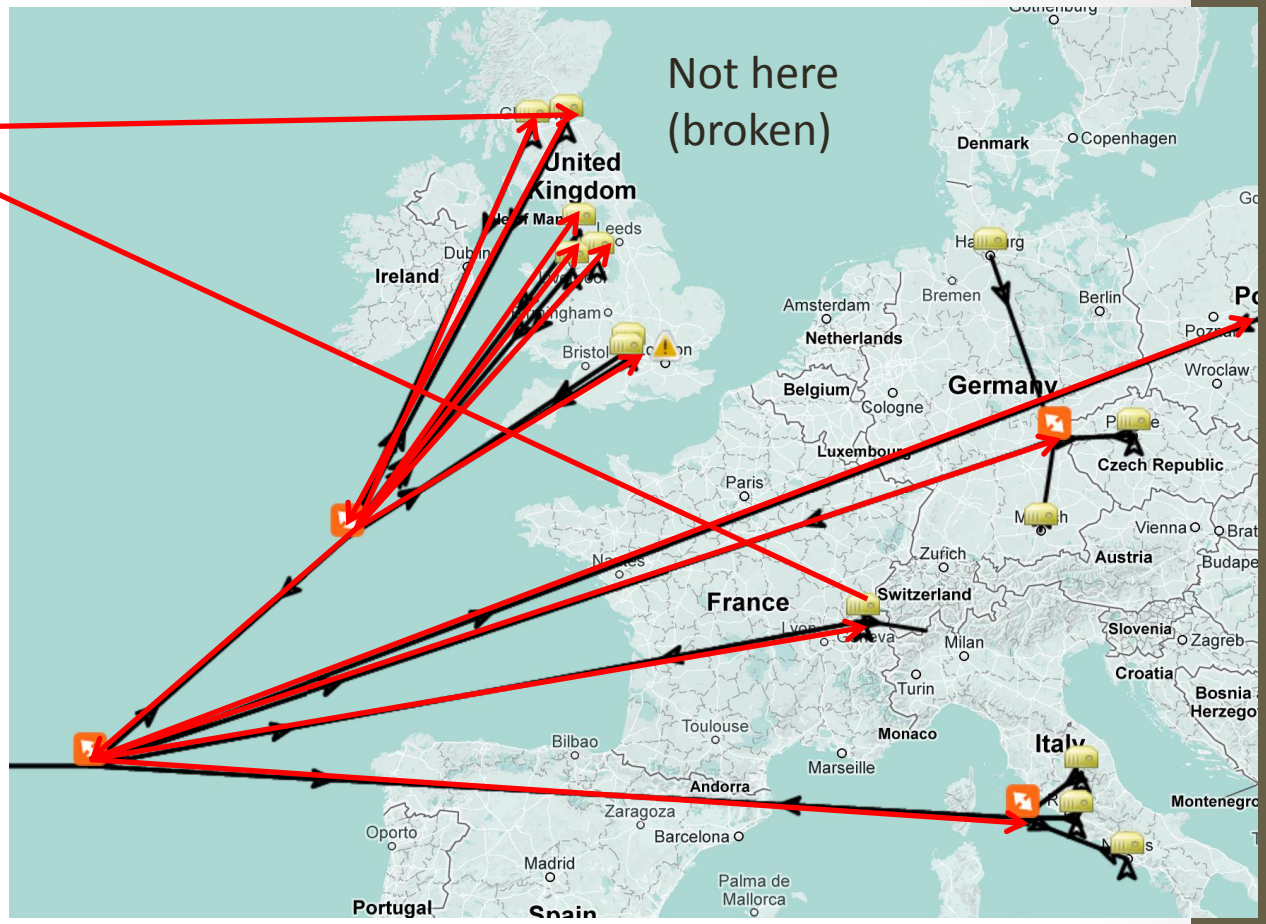
- Probably the evolved WLCG /LHC tools can work for LQCD aims:
  - Overlap between actual CoE activities in exportation of data from HPC centres as well as expertise in these tools.

EXTRAS…

Edinburgh User

I want
/atlas/blah/higgs.root

Open
root://diskX.cern.ch/
where/blah/higgs.root



Read over WAN
No need for central catalogue
In production use for all LHC experiments.

Comparing CERN -> CERN and CERN -> Edinburgh (ECDF) reading (for most extreme IO application) (with xrootd and http)