Computing Models, Facilities, Distributed Computing

- Draft Chapter

- Scope
  - Consider changes in the computing models, and the distributed infrastructure that can help to mitigate the disparity between the anticipated requirements for storage & computing and what is probably affordable. How can the models easily adapt to changing technology, and resources, without need for continual major porting efforts? We must be able to easily use any and all resources that may be offered to us. NB. This is a cross-cutting and high level view of the overall computing models

- Charge
  - This working group focuses on how computing models might evolve and which elements might be key to developing different possibilities over the next years. Contributions originating from many other CWP WGs will of course contribute to the overall computing models.
    - How will (or should) computing models evolve in the next 5-10 years to meet the challenges for HEP in the 2020s and which elements are key to developing these possibilities?
    - Which elements of the current computing models will ***not*** need to change over the next 5-10 years?

Key challenges and opportunities (1 slide)

- Fitting requirements (>>x10-50) within constrained funding
  - Requirements significantly exceed likely hardware evolution expectations – but that is a big unknown
  - Reduce operational costs
  - Understand cost models and trade-offs
- Potential for common solutions between experiments – at all levels of the computing models
- Facilities – consolidation of resources in regions
  - Efficiency of scale and operations costs
  - Benefit from technology advances e.g. as demonstrated by large internet companies
  - Agile use of heterogenous resources, workflow-specific facilities, opportunistic resources, etc.
  - Create opportunities for very different models of processing and analysis

Addressing the CWP Charge

- Cost models
  - Try to determine metrics by which it makes sense to optimise and to be able to trade-off between aspects of the models
  - Can be used to determine best workload for a site, taking into account actual costs; guide where to put optimisation effort; guide procurement decisions
- High level computing models – likely a hybrid of many key ideas
  - “Data – lake” : a well-connected, federated data infrastructure of relatively few large sites – where bulk data is managed and stored, served to clients; requires a long-term commitment to manage the data
    - Could include non-owned facilities
  - Other facilities – providing compute and supporting many workflows, but not needing much more than cache storage
  - Other key ideas include the need to support many heterogenous types of resource and facility:
    - Centres of excellence that may have specific expertise or facilities for certain problems (e.g.ML)
    - Use of industrial expertise and trends in managing data centres
    - Use of various types of specific resource (e.g. HPC) which may support only well-adapted workloads
    - Use of opportunistic (CPU) resources as an integral part of the model: commercial provisioned – clouds/spot markets; shared backfill capacity, volunteer computing, etc.
    - Use of elastically provisioned capacity for specific tasks

Addressing the charge – 2

- The facility model must allow national/regional optimisation
  - That fits political and funding situation: e.g. local consolidation of resources and effort through a given interface
  - May distinguish a major data infrastructure from other resources
- Facility model must be inclusive and allow contributions from for smaller sites
  - Depending on the facilities available etc.

Addressing the CWP charge – 3

- Hardware evolution
  - No magic bullet coming
  - No big high-level architectural changes expected – will still have something like tape/disk/CPU
    - But consider key parameters such as space, performance/latency, etc rather than technology
  - Parameters should quantify capabilities since technologies can then map to those (e.g. “archive” rather than “tape”)
- In-depth understanding of performance is essential
- Computing models – we see potential for commonality at many levels – ongoing active discussion
  - Data management – storage and data management models – optimise once for everyone - tightly connected to facility model
  - Transfer performance, protocols, efficient I/O and data streaming
  - Common resource provisioning and basic workload management layers
  - Networks – interconnects, intelligence, etc. – How do LHCOPN, LHCONE evolve? How to efficiently use them.

Practical considerations

- Proposed R&D activities – prototypes and demonstrators
  - Data federation, policy driven replication
  - Networking and traffic management of a data-lake model, SDN, CDN, network on demand
  - Efficient end-end I/O: streaming, event services, etc
  - Ongoing prototyping of (commodity&special) hardware developments – agility across platforms, connection with commercial resources
  - Need for transfer challenges to efficiently use large network pipes?
  - Data management for finer-grained processing
- R&D
  - Cost models
  - What facilities required for data analytics platforms?
  - Object stores ?
  - Moving DM, transfers, caching into network stack?
  - Explore scales of many millions of jobs in HTCondor (or?) with integrated workflow management systems

Commonality and Leveraging S&C beyond HEP domain

Opportunities for common tools:

- As noted above – potential for much commonality at every level – try and address with early prototypes to explore key functions and edge cases

- Make use of technologies, strategies, (open source) tools from large internet companies etc.

- Importance of synergies with facilities needed by other HEP and Astronomy/astro-particle

Cross-cutting Elements

Essentially all of this cuts across the other WG’s and depends on ideas from them

CWP Chapter Status and Plans

There is a draft of most key sections – some are missing but promised
- Key ideas have been discussed and are included at some level
- This WG needs to absorb the ideas expressed in the other WG's

The document has significant content (~40 pages) & needs a global edit and clean-up (following discussions this week)
- Comments are welcome now

Finish a clean draft within 2-3 weeks (by a small editing team)

This workshop should ensure that main ideas are there and first look at consistency with other WG's

Auxiliary Material

Primary Activities

San Diego workshop – first ideas

Several ad-hoc meetings with experiments' S&C management and others

"Writing day" – end April – confirmed document structure and assigned tasks

A lot of contributions since then

<Insert Name of CWP WG>

Primary organizers of the WG are:

1. <insert name>
2. <insert name>
3. ...