# Sol for UK future software

# 18 February 2020

Davide Costanzo

Davide Costanzo ECHEP workshop. Sol preparation

18-February-2020

# Introduction. Setting the context

- New and upgraded detectors are coming online in the next few years
  - More channels, higher throughput rate, more data
  - More sophisticated detectors. Several 100s M£ investments
  - LHCb are upgrading their readout system in 2021
  - ATLAS and CMS will have upgraded detectors in 2026-7
  - DUNE is coming online in 2028
  - Not to mention many others. Eg HyperK, LZ, Belle2, etc
  - Not to mention other "sciences". Eg Astronomy, Biology, etc.
- Computing for HEP is a significant investment
  - About 1 EByte of data already recorded by the LHC experiments
  - ~1M CPU cores running LHC experiment jobs at any given time

#### • Why invest in Software?

- We should make efficient use of the hardware we have (workshop topic!)
- Computing hardware is evolving (parallel, concurrency). A lot of our SW stack was written in the early 2000s, eg single core power was increasing with Moore law
- Complement the investment in detectors (can we have batteries, please?)
- Not just efficiency! We want to do better physics and develop better algorithms

# Affordability of computing in the late 2020s

- The ATLAS plots demonstrate how a step-function in computing resources is needed
  - Other experiments have similar plots (I show ATLAS, I'm biased, sorry...)
  - Note: updates coming in the next 2-3 months as we improve our models
- Software R&D can fill the gap





## Timeline diagram



# Statement of Interest -- STFC bid

### Sol preparation boundary conditions. Effort requested has to:

- Do useful work for future SW for HL-LHC, DUNE, ...
- Be cross experiment, eg useful to all of us
- Match the priorities of experiments (ATLAS, CMS, DUNE, LHCb, others)
- Not request effort that already exists as part of other grants (CG, upgrade, GridPP, etc)
- Be part of an international collaboration, e.g. do not compete/replicate work done elsewhere (eg in the USA)



### Sol vs Excalibur bids

- The Excalibur-HEP proposal went to UKRI about a month ago (~70 people signed it)
  - Goal: Creation of a "design and development working group" representing HEP
  - Limited funding (~230k£) for a limited time span (15 months)
  - Material selected to match the call description, while balancing with HEP-UK priorities
- Sol is going to STFC at the end of March
  - Goal: This is a HEP proposal. Opportunity to ask for RSE effort. Fully inclusive by design
  - The need for this effort came across. We need to justify it.
  - We need to be prepared to defend our priorities through a longer review process

# Sol proposal role in the HEP community

Layer	Domain	Experiment 1	Exp. 2	Exp. 3	
6	Physicists	Analysis code.			
5	Experiment physics programmers and software engineers	Analysis frameworks, reconstruction code, calibration code			↓ Moving
4	Experiment software engineers	Production computing framework			software down the
3	Experiments/HSF WLCG/GridPP	Common software components (e.g Geant4, Generators, ROOT, Analysis tools, ACTS,)			↓ stack
2	WLCG/GridPP	Software infrastructure for distributed computing (CE, SE, CernVM-FS, data management,)			
1	WLCG/GridPP	Physical Hardware			

(Table adapted from GridPP6 proposal)

Davide Costanzo ECHEP workshop. Sol preparation

# **Proto-Work Packages**

### Organise the work in a few WP. Draft here

- WP1: Dealing with exascale data on a global scale
- WP2: Real-time exploitation of exascale data sets: classification, inference and analysis algorithms
- WP3: Confronting theory and understanding behaviours
- WP4: Simulating data and detector response
- WP5: Advanced reconstruction and pattern finding algorithms.
- WP6: Analysis systems

#### With some activity spanning across work packages such as:

- Machine learning, Al
- Heterogeneous computing, portable parallelization strategies Code maintenance over 2 decades
- Training. Lots of new tools appearing, remember most users are graduate students.

#### This is a rather ambitious programme

- We should express our ambitions in the Sol.
- But also realise that it will not fit into the budget envelope under discussion
- Prioritisation exercise is needed -- After Sol submission, when writing full proposal
- Sol submission at the end of March
- Use as a reference for future funding opportunities

#### Advisory board established to ensure priorities are agreed. Monitor evolution of the project

- Representation from experiments (ATLAS, CMS, DUNE, LHCb) and GridPP, IRIS

# A few thoughts about areas of interest

### Event generators

- Lots of interest in the UK, eg MCnet. Lots of optimisation work needed
- Cross experiment, useful to the community as a whole
- Opportunities to use other resources (eg GPUs) becoming available (eg Summit, EuroHPC, ...)

### Trigger/real time systems

- Lots of interest in the UK.
- Big challenges ahead due to the increase of data throughput
- Cost/benefit analysis is clear, experiment procure the hardware and write the software. GPU demonstrators show that we are almost there
- Lots of opportunities for cross-experiment collaboration

### Simulation

- Again, good UK involvement.
- Geant4 R&D gaining momentum. Good opportunity to participate to it
- Fast simulation synergies between ATLAS and LHCb. Tuning of full vs fast
- Crucial for the future experiments. Redesign in a parallel world

### Tracking algorithms development

- UK excels and designing and building tracking (and not only!) detectors
- With better hardware we need improved algorithms
- Better algorithms = Better physics exploitation
- What are the opportunities for cross-experiment R&D ?

# A few thoughts about areas of interest (2)

### Data Organisation Management and Access (DOMA)

- GridPP really successful in obtaining and providing computing infrastructure to the HEP experiments (and beyond)
- No long term R&D effort available in GridPP6
- DOMA is for more optimised distributed data management, caching, transport and access and "data lakes"

### Analysis systems

- Rapidly changing environment, due to the evolution of "data science" outside HEP
- We need to keep up with this tide
- How do we organise this?

#### Training and career

- How do we fit this in the Sol? Add-on package? Training needs to be integral part of the community.
- CDTs go a long way. We should ensure students involvement in this activity
- We need a long term view to ensure a career path for people working as "research software engineer"
- Importance of using industry standard tools for a large training impact.