Draft v0.1, 22 August, 2017

Optimal HL-LHC ATLAS Software Focus Areas for NSF Support

Overview

During 2016-17, the Hep Software Foundation (HSF), a community driven volunteer effort which includes a large cross section of software experts from the ATLAS experiment at the LHC, worked on developing a Community White Paper (CWP) for the HL-LHC. While the CWP process is ongoing and should not be considered as completed, a number of areas have already emerged as being important for success of the HL-LHC physics program. These focus areas are broadly classified as (alphabetically):

Careers, Staffing and Training
Computing Models, Facilities and Distributed Computing
Conditions Database
Data Access, Organization and Management
Data Acquisition Software
Data Analysis and Interpretation
Data and Software Preservation
Detector Simulation
Event Processing Frameworks
Machine Learning
Physics Generators
Software Development, Deployment and Validation/Verification
Software Trigger and Event Reconstruction
Various Aspects of Technical Evolution (Software Tools, Hardware, Networking)
Visualization
Workflow and Resource Management

While the CWP process continues, the team working on the conceptual plan for a proposed NSF funded S2I2 Software Institute (S2I2-SI) for the HL-LHC is charged to identify areas that could be initial high priority working areas for the proposed institute. We present here the areas, based on the CWP work, that we deem to be the highest priority for US participation in the ATLAS experiment in the HL-LHC era. This list is preliminary, and may be updated in the future.

The focus areas proposed by US ATLAS for the S2I2-SI are not exhaustive. In fact, many high priority areas are skipped since they are expected to be covered by other collaborating countries participating in the WLCG, or other funding agencies in the US (i.e. DOE, or other NSF initiatives). Also, the list below is focussed to ensure success in those areas that match interests, expertise, and leadership by US universities.

Some software areas are considered cross-cutting, and therefore not included in the US ATLAS list of priorities. For example, Machine Learning and Deep Learning. ML and DL techniques are expected to provide significant improvements to the algorithms used by ATLAS in the future, and provide a fertile area for innovation and optimization. Therefore, ML is not explicitly listed as a focus area, because it will be the core of almost all the focus areas picked below.

ATLAS is a unique detector (i.e. just as unique as CMS and LHCb). Some software are therefore unique to ATLAS. We address why some areas are unique in the brief descriptions below. A more detailed exposition will be included in the S2I2-SI conceptualization report.

In order to optimize the selection of the software priorities for the S2I2-SI, all LHC experiments used a common set of guidelines. They are:

1. Impact - Physics: Will efforts in this area enable new approaches to computing and software that maximize, and could potentially radically extend, the physics reach of the detectors?
2. Impact - Resources: Will efforts in this area achieve required improvements in software efficiency, scalability and performance and make use of the advances in CPU, storage and network technologies?
3. Impact - Sustainability: Will efforts in this area guarantee the long term sustainability of the software through the lifetime of the HL-LHC?
4. Interest/Expertise: Does the U.S. university community have a strong interest and expertise in the area?
5. Leadership: Are the proposed focus areas complementary to efforts funded by the US-LHC Operations Program, DOE or international entities?
6. Value: Is there potential to provide value to more than one LHC experiment and to the wider HEP community?
7. Research/Innovation: Are there opportunities for combining research and innovation as part of partnerships between the HEP and Computer Science communities?

Based on the above criteria, we would like to suggest the following areas of focus for the proposed NSF S2I2-SI.

High Priority 1: Data Analysis/Data Preservation/Scalable Platforms

This area has high impact factor for the physics reach of the HL-LHC, enabling data analysis by individual physicists, and to become a center of innovation through ML. Many US ATLAS universities have expressed interest in this area.

High Priority 2: High Level Trigger and Reconstruction Algorithms

Success of HLT and reco algorithms will directly fuel the physics success of the HL-LHC and is a critical factor in resource management. Application of ML will enhance partnerships between HEP and CS. Sustainability is important to the algorithms developed. US universities are leading the effort in these areas in ATLAS.

High Priority 3: Workflow, Workload and Resource Management (including networking)

This area has a large potential for increasing the physics reach of the HL-LHC, enabling access by a larger number of physicists, and developing solutions useful to other data science communities. This is a natural area of collaboration for HEP and CS, to develop sustainable solutions that are applicable to multiple LHC experiments. The US universities have long held leadership in this area in ATLAS.

High Priority 4: Detector Simulation (including Machine Learning)

Detector simulation is very important for exploiting the physics at the HL-LHC. For ATLAS, given the more complex calorimeter geometry, improvements in simulation performance can provide huge gains in performance and resource optimization. ML techniques can provide huge efficiencies, and involve CS communities. Many US universities are interested in contributing to this area.

High Priority 5: Event Visualization

Event visualization is an important tool for physicists at the HL-LHC. The complexity of the events will provide a challenging environment for visualization. Net architectures may be useful. US universities have traditionally played an important role in ATLAS.

Medium Priority 1: Data Access and Data Management

While this area is very important for HL-LHC physics, sustainability, and resource management, it has been led by international and DOE groups in the past in ATLAS. Additional effort from NSF university groups is necessary to solve this challenging software and computing problem for the HL-LHC.

Paolo Calafiura & Kaushik De
On behalf of US ATLAS