Data Processing Frameworks, Scope

Scope:

1. The framework holds the **protocols, tools and concepts for defining,** developing, and deploying **physics algorithms**, along with all the ancillary tools for providing services to the algorithms. This includes algorithm scheduling components, the event data model, handling of input and output from physics applications that utilize a framework, interfaces for non-event data, and configuration of framework applications.
2. **Processing model**: mechanism to execute/apportion work. Tasks, threads, processes, and communication between them
3. **Programming model** elements that permit efficient and maintainable algorithms to be developed. This includes interactions with multiple languages.
4. **Interfaces with the computing model**, including data, job, and workflow management.
5. Consideration of the needs of experiments beyond HL-LHC.
6. The framework should provide **hooks for doing monitoring and logging services** for performance and other purposes.

Data Processing Frameworks, Charge Questions

1. How do **compute facility changes**, both in resource type and availability, affect how and what a framework does?
2. How ought the framework affect or **inform decisions concerning the compute model** and all the services that it is hooked up to?
3. What ought it be like to **develop an algorithm** within the context of a framework ?
4. What will it be like to **configure and run** a framework executable ?
5. To what extent can the **HEP community come together on framework** concepts, tools, libraries, and services? Will we be tied to our history for the next 10 years? Will the agencies tolerate this?
6. How much influence ought the framework have on the **programming model**?
7. Should the framework be better integrated with a set of overarching, common workflow management tools?
8. Does the framework dictate the programming language?

Challenges and Opportunities

Changes needed in the **programming model** that are necessary to handle the **massive parallelism** that will be present throughout all layers in the **available compute facilities**.

Providing common **framework services and libraries** that will meet with the expanding compute and data needs

Tighter integration with the **computing model** required by fundamental changes in compute facilities

Support structures that permit frameworks to be **collaboratively developed** and maintained across a large number of experiments.

The challenge present in all of these elements is to **increase productivity given the increased complexity and scale** of the upcoming experiments and computing facilities. This means decreasing program development and debugging time, and increasing efficiencies in the use of diverse facilities.

Addressing the CWP Charge

The Data Processing Framework plays a role in all aspects below:

1) to achieve improvements in software efficiency, scalability and performance and to make use of the advances in CPU, storage and network technologies … the Framework must provide facilities and services that allow user algorithms to utilize new technologies with efficient abstractions that hide details, while still getting most of the compute power delivered by the HW.

2) to enable new approaches to computing and software that could radically extend the physics reach of the detectors… Frameworks that enable machine learning techniques in event reconstruction have already demonstrated this, and it is likely there will be future examples as well.

3) to ensure the long term sustainability of the software through the lifetime of the HL-LHC… sustainability in the face of continuous funding cuts are a major driver of consolidation and sharing in the Framework domain.

Big Ideas?

At the level of concepts there is already a lot of commonality among experiments.

In fact we wrote an (incomplete) table of concepts to call out this commonality:

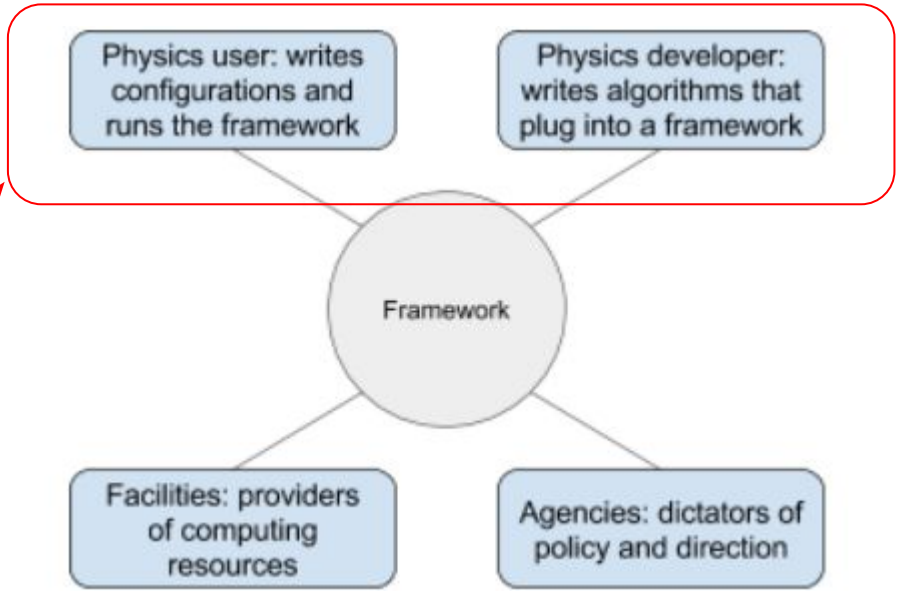
ALICE/FAIR	ATLAS/LHCb	ILC/CLIC	Belle II	CMS / art	This paper
	Configurable	Parameter	Parameter	Parameter	Parameter
	Algorithm	Processor	Module	Module	Processor
	Trigger Line		Path	Path	Path
	Event Store	Event	DataStore	Event	Event Store
	Conditions Store		DataStore	EventSetup	Conditions Store

Commonality

There are opportunities given this commonality, yet it is also listed as a charge question and a challenge.

The barriers are history and commitment from the stakeholders identified in the document: developers, facility providers, and most importantly funding agencies.

Collaboration Stakeholders which have different time scales



More Ideas

Increase framework modularity, both to enable shared components between frameworks, as an intermediate step towards more commonality; and to allow building mini-frameworks that can be used in contexts where there are less resources, for example to access data using the EDM on devices like tablets and smartphones…

Blurring the lines between workflow/workload management and the framework

Heterogeneous resources that will be in close proximity, and management of the memory space that is going to be available

The movement completely away from file-based I/O, and the introductions of hierarchical “layers” in the event model, and offering operators that help with automatic concurrency are a few ideas that we need to explore.

Leveraging S&C beyond HEP domain

In what ways can this WG area leverage software and computing techniques and technologies outside of the traditional HEP domain (e.g. from industry and CS)?

- In many ways the current frameworks already do, for instance both GaudiHive and CMS' Framework leverage the work Intel has put into TBB.
- This success must be continued, but it is not easy to pick the right technology. It requires R&D to find suitable industry tools for use in HEP.

Cross-cutting Elements

The document explains the connections of the Framework to the following cross-cutting areas:

Computing Model

Data Analysis

Data and Workload Management

Conditions Database

Visualization

Simulation and Reconstruction

CWP Chapter Status and Plans

What is the status of the CWP Chapter for this working group? Are the key ideas and R&D in place?

- There is 26 pages in this paper with a few unfilled sections
- The headings do cover the key ideas
- We have not defined a roadmap

What additional work is required to get the prose in good shape for a viable CWP chapter and for others outside of your WG to read and comment?

- It is in good enough shape to be read by others

CWP Chapter Status and Plans

How do you plan to complete your chapter?

- Obviously we need to brainstorm about a possible R&D roadmap
- We have proceeded so far by assigning parts of the outline to different volunteer contributors and iterating from there.

What do you expect to accomplish by the end of this workshop?

- Since most of the challenge is social (requiring commitment from stakeholders) it would be great to get input from the wider community.

Auxiliary Material

Primary Activities

Useful session at the Jan. kickoff workshop, which produced the following document: https://docs.google.com/document/d/1pMWia3IXX6w5075PGs_IeEPnWZhkwQgoDQnHGqbPG-c/edit

As planned we had one 2 day workshop in June and a meeting in May where we iterated on the CWP document:

https://docs.google.com/document/d/14NMDpzlIKaR_L3dhdRKh3mswA008ahuNxwTs7hiNlBc/edit#

We heavily relied on asynchronous communications given the geographically distributed group membership.

Event Processing Frameworks

Primary *organizers* / members of the WG are:

1. *Jim Kowalkowski*
2. *Liz Sexton-Kennedy*
3. Benedikt Hegner
4. Paolo Calafiura
5. Simon Patton
6. Charles Leggett
7. Marco Clemencic
8. Chris Jones
9. Marc Paterno
10. Thomas Kuhr

- Giulio Eulisse
- Mohammad Al-Turany
- Hadrien Grasland
- Sebastien Binet
- Chris Green
- Kyle Knoepfel
- David Malon
- Vakho Tsulaia
- And more...

Practical Consideration for Progress in the WG Area

How will the proposed activities empower HEP physicists to get the most physics out of the experiments during the HL-LHC era? What new physics capabilities might these bring?

Already covered..

What are the proposed R&D activities over the next 5 years toward these applications?

We have hints of the answer for this, but we need more time to work on it.

How will the software be deployed by the experiments and sustained for the duration of the HL-LHC?

Sustainability is the major driver for attempting a common framework

What are the primary future applications you see in this WG area?
What is the likely impact that the techniques and applications will have on overcoming the challenges of the HL-LHC era?

Are there risks associated with going in the direction of each of the proposed ideas/R&D? What are the associated costs and is the development and implementation realistic in this regard?

We It is too early to say...