S2I2-HEP Conceptualization and the Community White Paper (CWP)

Peter Elmer - Princeton University

7 December, 2016
S2I2 HEP/CS Workshop
NCSA/UIUC

S2I2-HEP

The primary goal of the S2I2-HEP conceptualization project (`http://s2i2-hep.org`) is to produce a well-defined strategy for developing the software and computing models for use in high energy physics (HEP), in particular for the experiments collecting the very large data sets anticipated in the "High-Luminosity Large Hadron Collider" (HL-LHC) era of the 2020s.

Specifically the S2I2-HEP project will identify potential areas where U.S. university personnel can lead in key areas of software development to help realize the full potential of the HL-LHC program.

However HEP and the LHC are global projects, so no long-term planning exercise can exist in isolation, thus we are also pursuing a wider HEP community roadmap for software and computing in the 2020s.

NSF 15-553 - S2I2 Conceptualization Awards

(These) are planning awards aimed at organizing an interdisciplinary community and understanding their software requirements and challenges. (...) The product of a conceptualization award will be a strategic plan for enabling science and education through a sustained software infrastructure that will be freely available to the community, and will address the following elements:

- the science community and the specific grand challenge research questions that the S2I2 will support;
- specific software elements and frameworks that are relevant to the community, the sustainability challenges that need to be addressed, and why addressing these challenges will be transformative;
- appropriate software architectures and lifecycle processes, development, testing and deployment methodologies, validation and verification processes, end usability and interface considerations, and required infrastructure and technologies;
- the required organizational, personnel and management structures and operational processes;
- the requirements and necessary mechanisms for human resource development, including integration of education and training, mentoring of students, postdoctoral fellows as well as software professionals, and proactively addressing diversity and broadening participation;
- potential approaches for long-term sustainability of the software institute as well as the software; and
- potential risks including risks associated with establishment and execution, necessary infrastructure and associated technologies, community engagement, and long-term sustainability.

NSF 15-553 - S2I2 Conceptualization Awards

(These) are planning awards aimed at organizing an interdisciplinary community and understanding their software requirements and challenges. (...) The product of a conceptualization award will be a strategic plan for enabling science and education through a sustained software infrastructure that will be freely available to the community, and will address the following elements:

- the science community and the specific grand challenge research questions that the S2I2 will support;
- specific software elements and frameworks that are relevant to the community, the sustainability challenges that need to be addressed, and why addressing these challenges will be transformative;
- appropriate software architectures and lifecycle processes, development, testing and deployment methodologies, validation and verification processes, end usability and interface considerations, and required infrastructure and technologies;
- the required organizational, personnel and management structures and operational processes;
- the requirements and necessary mechanisms for human resource development, including integration of education and training, mentoring of students, postdoctoral fellows as well as software professionals, and proactively addressing diversity and broadening participation;
- potential approaches for long-term sustainability of the software institute as well as the software; and
- potential risks including risks associated with establishment and execution, necessary infrastructure and associated technologies, community engagement, and long-term sustainability.

LHC Grand Challenge Research Questions

The goal of HEP (and the LHC) is to understand the fundamental building blocks of nature, and their interactions. The potential of the LHC has been demonstrated in its first years with the discovery of the Higgs Boson.

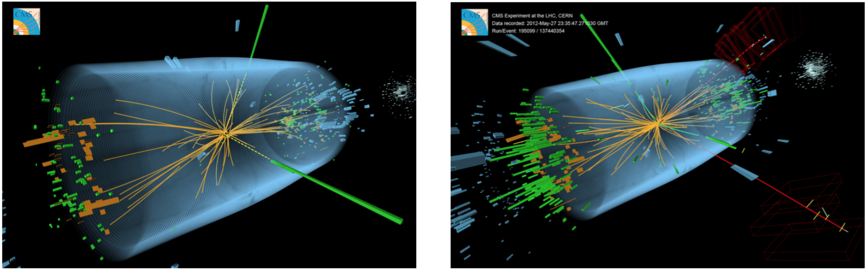
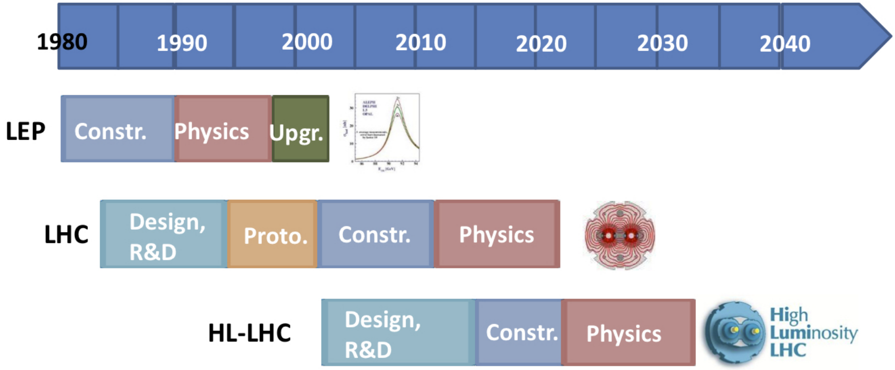


Figure 1.1: (a) Candidate for the decay Higgs $\to \gamma\gamma$, where the green lines are the two photons; and (b) Candidate for the decay Higgs $\to ZZ^*(ee\mu\mu)$, where here the green lines towards the center of the picture are the two electrons and the red lines in the center and at the upper right of the detector are the two muons.

LHC Grand Challenge Research Questions

Many fundamental questions remain, however, including: Why does nature express the symmetries embodied in the SM, and not other equally elegant symmetries? Why are there (only) three generations of basic building blocks of matter? Why are the masses of these building blocks so different from each other, both within a generation and between generations? What is the dark matter which pervades the Universe? Why is matter so dominant over antimatter in the Universe? Does space-time have additional symmetries or extend beyond the $3+1$ dimensions of which we know? What mechanism stabilizes the Higgs mass from large quantum corrections at high energy? Are neutrinos their own anti-particles? Can gravity and quantum mechanics be described in a consistent theoretical framework?

CERN Accelerator Timeline



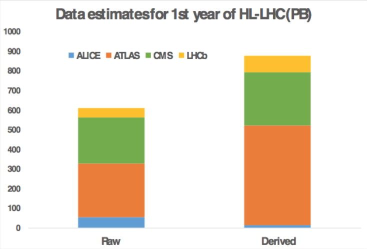
Various concepts also exist for subsequent machines.

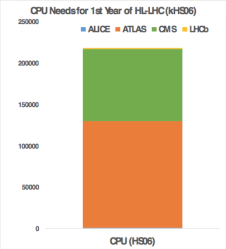
NSF 15-553 - S2I2 Conceptualization Awards

(These) are planning awards aimed at organizing an interdisciplinary community and understanding their software requirements and challenges. (...) The product of a conceptualization award will be a strategic plan for enabling science and education through a sustained software infrastructure that will be freely available to the community, and will address the following elements:

- the science community and the specific grand challenge research questions that the S2I2 will support;
- specific software elements and frameworks that are relevant to the community, the sustainability challenges that need to be addressed, and why addressing these challenges will be transformative;
- appropriate software architectures and lifecycle processes, development, testing and deployment methodologies, validation and verification processes, end usability and interface considerations, and required infrastructure and technologies;
- the required organizational, personnel and management structures and operational processes;
- the requirements and necessary mechanisms for human resource development, including integration of education and training, mentoring of students, postdoctoral fellows as well as software professionals, and proactively addressing diversity and broadening participation;
- potential approaches for long-term sustainability of the software institute as well as the software; and
- potential risks including risks associated with establishment and execution, necessary infrastructure and associated technologies, community engagement, and long-term sustainability.

Estimates of Resource Needs for HL-LHC (WLCG)





Data:
- Raw 2016: 50 PB → 2027: 600 PB
- Derived (1 copy): 2016: 80 PB → 2027: 900 PB

CPU:
- x60 from 2016

Technology at ~20%/year will bring x6-10 in 10-11 years

- Simple model based on today's computing models, but with expected HL-LHC operating parameters (pile-up, trigger rates, etc.)
- At least x10 above what is realistic to expect from technology with reasonably constant cost

(Slide from WLCG Workshop Intro, Ian Bird, 8 Oct, 2016)

HEP Software Ecosystem

indico

TMVA

ROOT
Data Analysis Framework

IgProf

FroNTier

FairRoot

Roo
Fit

XRootD

iNSPIRE HEP

FastJet

RooStats

EvtGen

The Gaudi Project

CernVM
File system

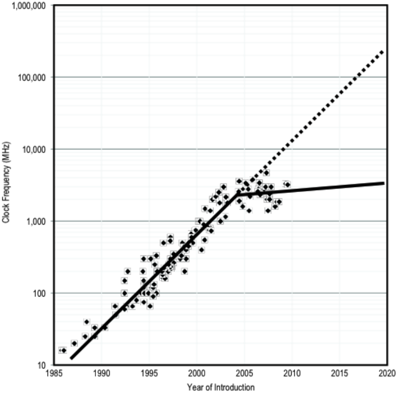
Geant 4

Examples, definitely incomplete!

dCache.org

Plus 15-20M Source Lines of Code (SLOC) of "experiment specific" codes, as well as dependencies on non-HEP scientific software.

Processor evolution and software impact



Clock Frequency vs Time

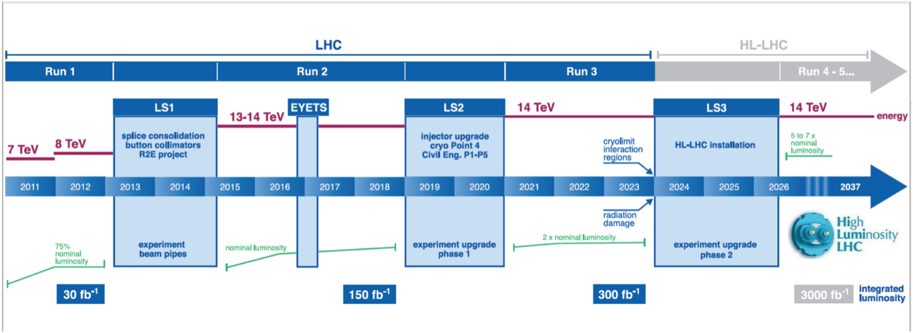
- Single core performance has stalled, leading to multi/manycore and specialization
- To even realize Moore's Law gains, we are pushed towards parallelization of algorithms and design for performance.
- The software designs and implementations themselves need to evolve, not just be recompiled

Back to heterogeneous systems?

Building the worldwide distributed LHC computing grid was largely made possible by the convergence on Linux on (commodity) Intel x86 processors around the year 2000. Building the WLCG at this scale in the heterogeneous workstation era would have been quite difficult. For better or for worse, heterogeneity is returning:

- Diversity of computing processor architectures (general purpose cores vs specialized processors)
- Owned vs commercial/cloud providers
- Some pressure to use systems traditionally designed for other types of applications (e.g. HPC/supercomputer as opposed to HTC/high-throughput systems)
- Possible further commoditizing market pressures (e.g. mobile)

Plans for upgrading the LHC and Experiment Detectors



A Software "Upgrade" for HL-LHC and 2020s HEP?

Looking forward to the next 10 years, we see a number of challenges for HEP software and computing:

- **Scale:** The HL-LHC will integrate 100 times the current data, with significantly increased data (pileup) and detector complexity.
- **Performance/cost:** Estimates of computing needs run faster than Moore's Law by factors of 3-30
- **Technology/Market evolution:** the return of heterogeneity; technology change will also make it challenging to exploit Moore's Law without software evolution.
- **Sustainability:** Most of the current software, which defines our capabilities, was designed 15-20 years ago: there are many software sustainability challenges.

Why Software? Software is *the* Cyberinfrastructure

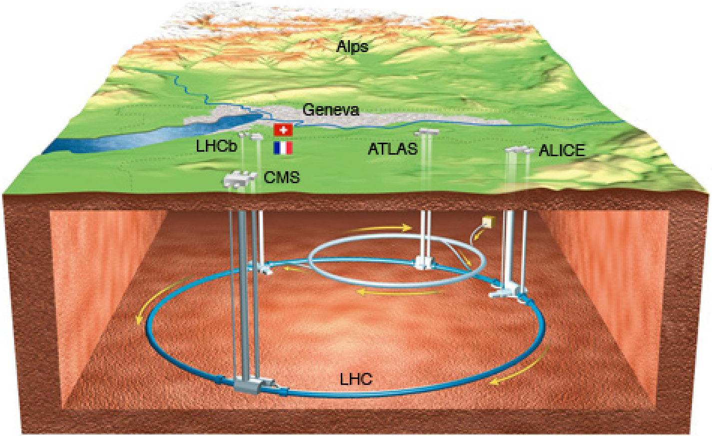
Computer hardware is a consumable.
Software is what we keep, and invest in, over time.

NSF 15-553 - S2I2 Conceptualization Awards

(These) are planning awards aimed at organizing an interdisciplinary community and understanding their software requirements and challenges. (...) The product of a conceptualization award will be a strategic plan for enabling science and education through a sustained software infrastructure that will be freely available to the community, and will address the following elements:

- the science community and the specific grand challenge research questions that the S2I2 will support;
- specific software elements and frameworks that are relevant to the community, the sustainability challenges that need to be addressed, and why addressing these challenges will be transformative;
- appropriate software architectures and lifecycle processes, development, testing and deployment methodologies, validation and verification processes, end usability and interface considerations, and required infrastructure and technologies;
- the required organizational, personnel and management structures and operational processes;
- the requirements and necessary mechanisms for human resource development, including integration of education and training, mentoring of students, postdoctoral fellows as well as software professionals, and proactively addressing diversity and broadening participation;
- potential approaches for long-term sustainability of the software institute as well as the software; and
- potential risks including risks associated with establishment and execution, necessary infrastructure and associated technologies, community engagement, and long-term sustainability.

Large Hadron Collider (LHC) and Experiments

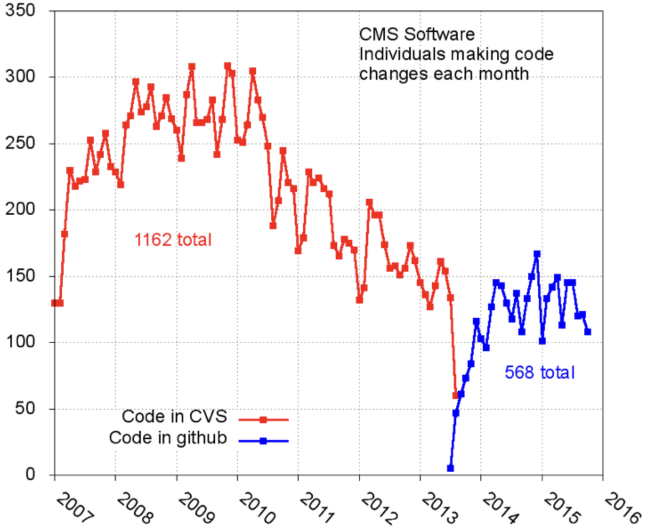


Two very large experiments (Atlas, CMS) with 3500+ people, and two large experiments (Alice, LHCb) with 500+ people

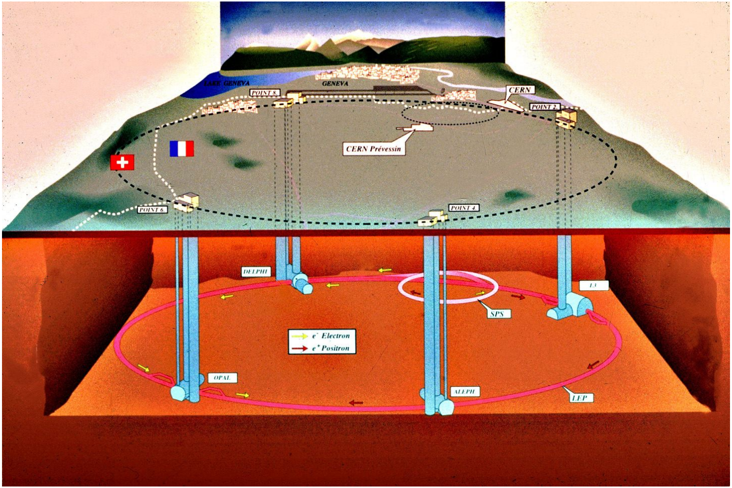
HEP Organization

- HEP is organized at a global scale ("Big Science") through experiments involving 100s to 1000s of invididuals.
- This can be daunting: how does one university group or grad student make an impact?
- Building these large experiments has pushed HEP to create organizational structures that permit individuals to contribute.
- The grad student will quickly and routinely wind up presenting to large groups of collaborators. Similarly, HEP has a footprint in 100+ computing centers around the world.
- Other advantages: realistic computing/software problems at a large scale, relatively open academic culture.

HEP Organization



CERN Experiments in the 1990s



4 Experiments, each with 400-500 people. Similar experiments were happening at Fermilab, SLAC (Stanford), BNL, DESY (Hamburg), KEK (Tsukuba)...

Careers and HEP Sociology in the LHC era

- Before the LHC-era particle physicists would typically work on multiple experiments (with different configurations of people) in their career
- Significant career incentives exist for people to stay within one of these experiments
- Many examples over the past 10 years of career paths (grad to postdoc to faculty/staff) happen within one LHC experiment
- J.Birnholz, *When Authorship Isnt Enough: Lessons from CERN on the Implications of Formal and Informal Credit Attribution Mechanisms in Collaborative Research*, `http://dx.doi.org/10.3998/3336451.0011.105`

NSF 15-553 - S2I2 Conceptualization Awards

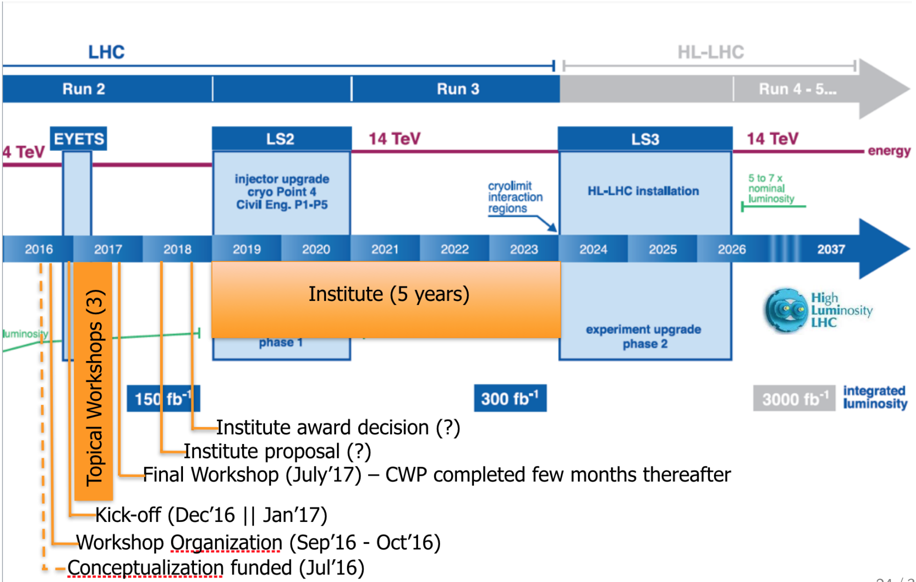
(These) are planning awards aimed at organizing an interdisciplinary community and understanding their software requirements and challenges. (...) The product of a conceptualization award will be a strategic plan for enabling science and education through a sustained software infrastructure that will be freely available to the community, and will address the following elements:

- the science community and the specific grand challenge research questions that the S2I2 will support;
- specific software elements and frameworks that are relevant to the community, the sustainability challenges that need to be addressed, and why addressing these challenges will be transformative;
- appropriate software architectures and lifecycle processes, development, testing and deployment methodologies, validation and verification processes, end usability and interface considerations, and required infrastructure and technologies;
- the required organizational, personnel and management structures and operational processes;
- the requirements and necessary mechanisms for human resource development, including integration of education and training, mentoring of students, postdoctoral fellows as well as software professionals, and proactively addressing diversity and broadening participation;
- potential approaches for long-term sustainability of the software institute as well as the software; and
- potential risks including risks associated with establishment and execution, necessary infrastructure and associated technologies, community engagement, and long-term sustainability.

HEP and CS

- What are examples of successful CS-HEP collaborations, and what properties have driven their success?
- How to align the CS research mechanisms (3 year grants, student developers, conference pubs) with the longer term needs of big science (30 year projects, production software, journal publications)?
- How to engage a broader slice of the CS community and make scientific computing more respectable within CS circles? (A commonly heard complaint in CS: scientific computing is a "'niche" research area.)
- What CS technologies, techniques, and trends could the HEP community adopt, rather than doing everything internally? (Keeping in mind the long time scales and production needs of HEP.)
- How could an HEP software institute facilitate interactions between the CS and HEP communities?
- What are the incentives for such collaboration for HEP people? For CS people? For non-CS people? E.g. recognition, funding, publications, students, new problems to solve, new places to apply technologies, new solutions to current problems, pride in working on a global-scale problem

S2I2-HEP (Success-oriented) timeline



Defining Longer-term Strategy

- HL-LHC computing requires a major ‘software upgrade’ and an eventual S2I2 institute for HEP would be a major player in that task
- Planning for such an “upgrade” cannot be done for the US (Universities) in isolation
- Thus we are intiating a larger community process to produce a Community White Paper (CWP) with an overall consensus strategy and roadmap for software and computing in HEP
  - Initiated as WLCG charge to the LHC experiments and HSF as a step towards the LHC experiment TDRs in advance of HL-LHC
  - The scope should not be restricted only to HL-LHC
  - Some early software components could be built, tested and used by experiments in LHC Run3
- Organised by the HEP Software Foundation (HSF) [next slide]
- Paper to be delivered by Summer 2017
- The S2I2-HEP Strategic Plan will be derived from this global plan

HEP Software Foundation (HSF)

The HSF (http://hepsoftwarefoundation.org) was created in early 2015 as a means for organizing our community to address the software challenges of future projects such as the HL-HLC. The HSF has the following objectives:

HEP Software Foundation

- Catalyze new common projects
- Promote commonality and collaboration in new developments to make the most of limited resources
- Provide a framework for attracting effort and support to S&C common projects (new resources!)
- Provide a structure to set priorities and goals for the work

Community White Paper (CWP)

- The CWP will identify and prioritise the software research and development investments required:
  - to achieve improvements in software efficiency, scalability and performance and to make use of the advances in CPU, storage and network technologies
  - to enable new approaches to computing and software that could radically extend the physics reach of the detectors
  - to ensure the long term sustainability of the software through the lifetime of the HL-LHC
- The HSF is engaging the HEP community to produce the CWP via a "community process"
  - Initiated as an HL-LHC planning process
  - Aiming for a broader participation (LHC, neutrino program, Belle II, linear collider so far)

Detector Simulation, Triggering, Event Reconstruction and Visualization

Challenges surrounding high pile-up simulation, including the CPU resources needed for large statistics samples needed to compare with data from high trigger rates, high memory utilization, generation and handling of the large (min-bias) samples needed to achieve accurate description of high pile-up collision events, and a flexible simulation strategy capable of a broad spectrum of precision in the detector response, from "fast" (e.g. parametric) simulation optimized for speed to full simulation in support of precision measurements and new physics searches (e.g. in subtle effects on event kinematics due to the presence of virtual particles at high scale). Software required to emulate upgraded detectors (including the trigger system) and support determination of their optimal configuration and calibration. • Software in support of triggering during the HL-LHC, including algorithms for the High-level Trigger, online tracking using GPUs and/or FPGAs, trigger steering, event building, data "parking" (for offline trigger decision), and data flow control systems. • New approaches to event reconstruction, in which the processing time depends sensitively on instantaneous luminosity, including advanced algorithms, vectorization, and execution concurrency and frameworks that exploit many-core architectures. In particular, charged particle tracking is expected to dominate the event processing time under high pile-up conditions. • Visualization tools, not only in support of upgrade detector configurations and event displays, but also as a research tool for data analysis, education, and outreach using modern tools and technologies for 3D rendering, data and geometry description and cloud environments.

Data Access and Management, Workflow and Resource Management

Data handling systems that scale to the Exabyte level during the HL-LHC era and satisfy the needs of physicists in terms of metadata and data access, distribution, and replication. Increasing availability of very high speed networks removes the need for CPU and data co-location and allows for more extensive use of data access over the wide-area network (WAN), providing failover capabilities, global data namespaces, and caching. • Event-based data streaming as complementary to the more traditional dataset-based or file-based data access, which is particularly important for utilizing opportunistic cycles on HPCs, cloud resources, and campus clusters where job eviction is frequent and stochastic. • Workflow management systems capable of handling millions of jobs running on a large number of heterogeneous, distributed computing resources, with capabilities including whole-node scheduling, checkpointing, job rebrokering, and volunteer computing. • Systems for measurement and monitoring of the networking bandwidth and latency between resource targets and the use of this information in job brokering. • Software-defined networking technologies which enable networks to be configurable and schedulable resources for use in the movement of data.

Physics generators, Data Analysis and Interpretation, Data and Software Preservation

There are many theory challenges in the HL-LHC era, among them are improving the precision of SM calculations, better estimation of systematic uncertainties, and elucidation of promising new physics signals for the experiments. Software needed to make connection between observations and theory include matrix element generators, calculation of higher-order QCD corrections, electroweak corrections, parton shower modeling, parton matching schemes, and soft gluon resummation methods. Physics generators that employ concurrency and exploit many-core architectures will play an important role in HL-LHC, as well better sharing of code and processing between LHC experimenters and phenomenologists. • Data analysis frameworks that include parallelization, optimized event I/O, data caching, and WAN-based data access. Analysis software that employs advanced algorithms and efficiently utilizes many-core architectures. • Tools and technologies for preservation and reuse of data and software, preservation and re-interpretation of physics results, analysis providence and workflow ontologies, analysis capture, and application packaging for platform abstraction. • Future software repositories and build platforms that leverage advances in these areas and improved software modularity and quality control that will allow a broader community of people to effectively contribute to software in the HL-LHC era.

Practicalities: CWP Process

- The end goal here is a single (consensus) CWP roadmap for the community.
- Finding consensus in a large community is a difficult task: broad participation and visibility/transparency are key elements
- The process being used largely mirrors that used in the "decadal survey" process in high energy physics
- Working groups self-organize, with encouragement from institutions and projects/experiments/etc.
- A series of workshops is planned over about 9 months to allow topics to be explored, sometimes overloading CWP discussions onto preexisting meetingd
- Contributions along the way can come in the form of "white papers" by individuals/groups/projects/institutions
- Based on the ideas emerging from the discusions, workshops and white papers, a consensus CWP document will be written.

Practicalities: Possible Working Groups

Detector Simulation	full and fast simulations, hi-pileup environments
Triggering	algorithms, GPUs and/or FPGAs
Event Reconstruction	new approaches to event reconstruction
Visualization	tools for data analysis, education, and outreach
Data Access and Management	scaling to the exabyte level
Workflow and Resource Management	millions of jobs in heterogenous systems
Physics generators	better models, better precision, code optimisations
Data Analysis and Interpretation	efficient use of many-core, modern techniques
Data and Software Preservation	preservation and reuse of data and software
Software Development, Deployment and Validation/Verification	improved modularity and quality, contribution
Computing Models, Facilities, Distributed Computing	range of possible models, costing
Various Aspects of Technical Evolution (Software Tools, Hardware)	
Security and Access Control	
Careers, Staffing and Training	perhaps in a separate concurrent white paper
Machine Learning	
Conditions Database	
Event Processing Frameworks	

More details in links at http://hepsoftwarefoundation.org/cwp.html

HSF Community White Paper Workshop at SDSC

- The kick-off workshop for the HSF CWP process will be on 23-26 January 2017 at SDSC/UCSD
- http://indico.cern.ch/event/570249/
- People from many HEP experiments (LHC and beyond)
- A key element in the "community process" to form a consensus
- We are looking for opportunities to introduce new ideas into the HEP discussions
- Additional topical workshops will happen in the spring and a final workshop ("near CERN" in summer 2017) is planned
- Further information about these will be posted on the s2i2-hep google group and at http://s2i2-hep.org