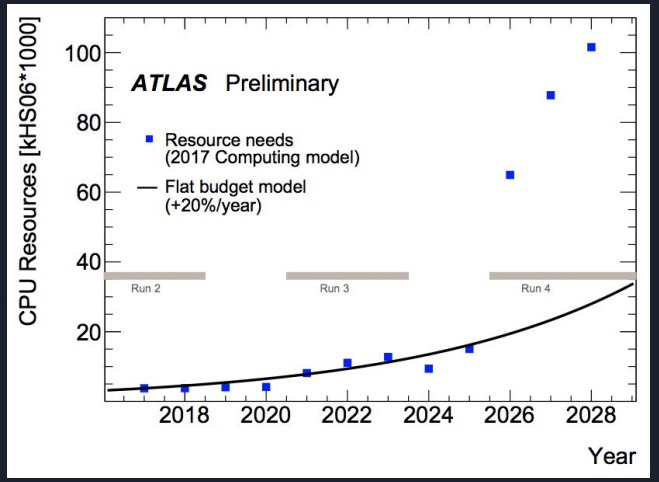
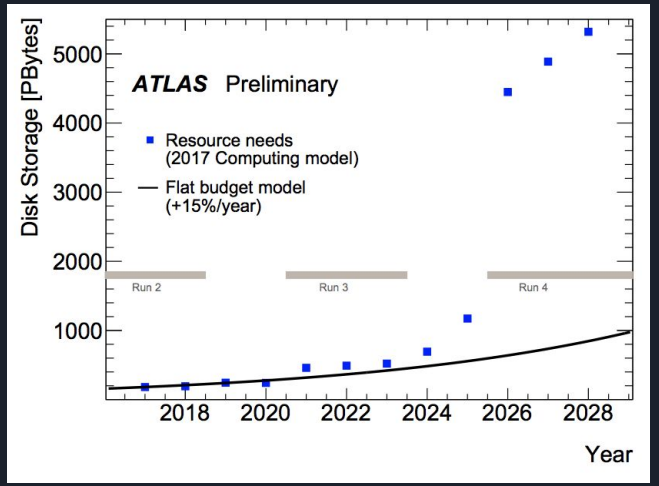
US Software and Computing R&D projects

Taylor Childers - ANL
Peter Elmer - Princeton University
(With contributions from Oliver Gutsche, Rob Gardner, Brian Bockelman, Ken/Paolo/Kaushik/Oli (JOG))

Challenges faced in 5-10 years

- Hi-Lumi LHC will drive computing needs to 10x or more compared to current resources. It will take us to 100 times the data of Run1 (and beyond) and will run through the 2030s. As for many countries, a flat budget scenario is expected at best in the US over this time period. Drivers for R&D activities:
- Resources/Cost: Will efforts in this area lead to improvements in software efficiency, scalability and performance and make use of the advances in CPU, storage and network technologies, that allow the experiments to maximize their physics reach within their computing budgets?
- Physics: Will efforts in this area enable new approaches to computing and software that maximize, and potentially radically extend, the physics reach of the detectors?
- Sustainability: Will efforts in this area significantly improve the long term sustainability of the software through the lifetime of the HL-LHC?





US Software/Computing R&D

Long running funded efforts which include some elements of development/technology-investigations:

- US-ATLAS and US-CMS Operations programs (joint NSF/DOE funding)
- Open Science Grid
- Institutional (University and Labs) physics research programs

Additional R&D efforts:

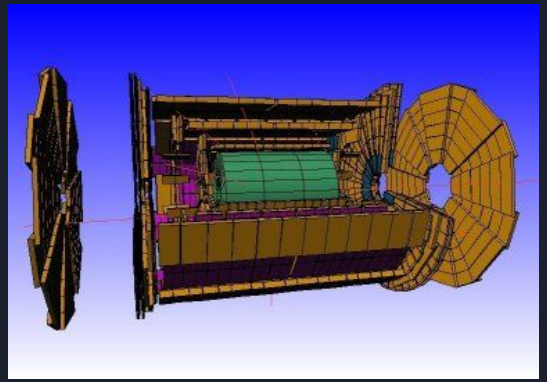
- Standalone NSF-funded projects, DOE SciDAC projects, DOE Computational HEP projects, Lab Directed R&D projects, industry collaborations, etc.
- Sometimes, but not always, exactly aligned or connected to specific experiments, and they include both LHC and IF-directed efforts.
- Possibility for an eventual Software Institute (NSF)?

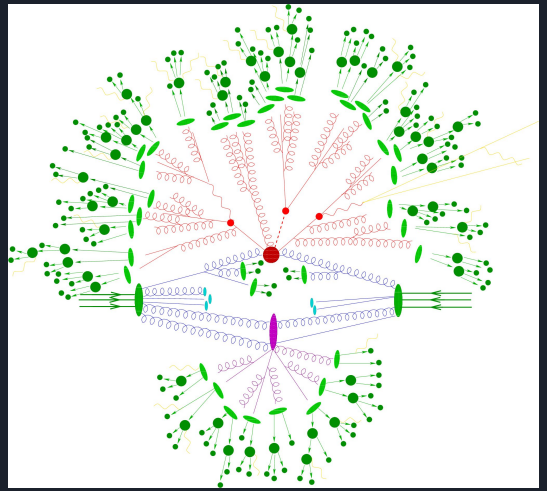
In these slides we focus on the CWP-relevant R&D projects for the medium/long term and have left out some more evolutionary development efforts in the interest of time


Software

Simulation

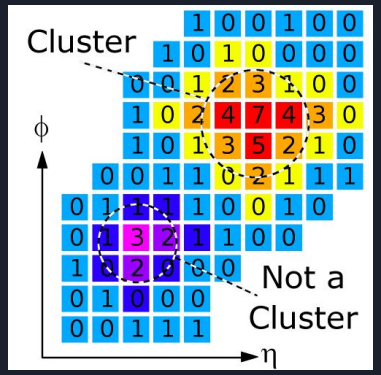
- Geant
  - Geant4:
    - R&D for incremental improvements continue to be supported through DOE Office of HEP Computing and the Center for Computational Excellence (Recent R&D)
    - Parallelization using MPI+multithreading recently implemented and tested
  - GeantV:
    - R&D to study reengineering Geant4 libraries to support vectorization and parallelization
    - CMS deployed VecGeom, the GeantV geometry library, in Geant4 in scalar mode and yielded speedups of 4-13%
- Generators
  - 3yr SciDAC-4 project: Develop novel scalable integration toolkit as a plugin for Madgraph/Sherpa/etc.
  - Part of larger 5yr SciDAC-4 Project: Exploit HPC facilities (compute, memory, and storage) to tune Pythia, updating workflow and data management especially for HPC machines

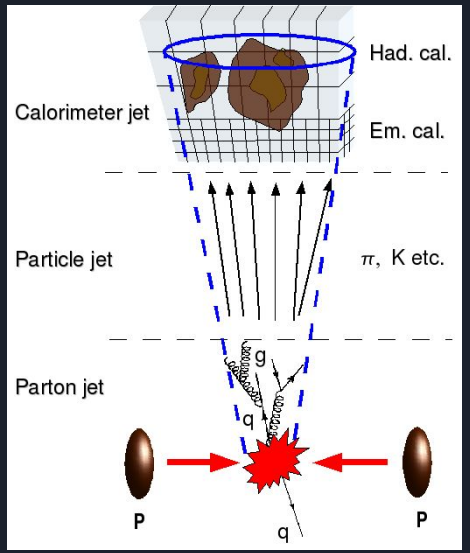




Reconstruction/Trigger

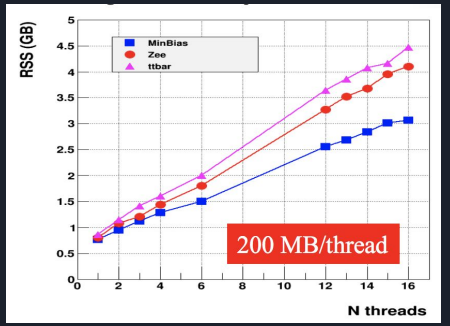
- HEP Event Reco (SciDAC)
  - Study experimental software stacks using profilers
  - Identify sections suitable for vectorization or parallelization
  - Targets Xeon Phi, GPU, other non-Xeon architectures
- Parallel Kalman Filter Tracking
  - R&D to develop parallelized and vectorized algorithms to implement the analog of the iterative Kalman Filter tracking used today. Focused on multicore Xeon, Xeon Phi, KNL and GPUs.
  - Began 2015 with 3 year NSF funded project (Princeton/UCSD/Cornell) plus USCMS, FNAL SCD & SciDAC collaboration.
  - The NSF part nominally ends this year, but with no-cost extensions will probably run into 2019.
  - Embedded in this is an effort to look at segment linking and optimization of the tracker layout for HL-LHC
- HEP.TrkX (HEP-CCE)
  - Pilot project funded to explore the use of non-traditional algorithms for tracking, e.g. Machine Learning, Deep Learning, … (FNAL/LBNL/Caltech)
- ML for LHCb Trigger
  - 3 year NSF project (MIT/Cincinnati) exploring use of ML for LHCb trigger



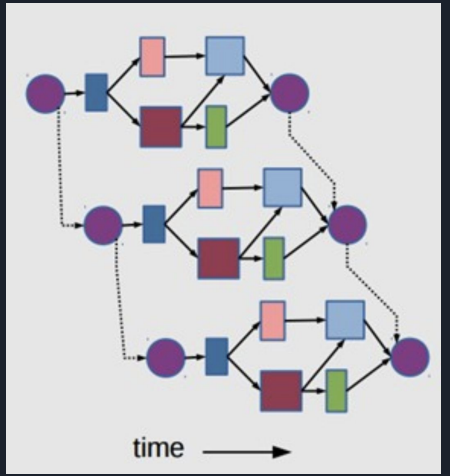


Multi-threading Software Frameworks

- CMSSW
  - Core software team continues to increase parallel thread counts
  - Current limitations in RECO to go to multi-hundred parallel threads
  - Beginning to look into event processors



- AthenaMT
  - Underlying framework changes largely complete
  - Geant4 simulation is complete, validation ongoing
  - Experiment-wide effort underway to migrate reconstruction algorithms
  - Event Generation not yet investigated



Analysis

DIANA/HEP

- Project focused on the development of analysis tools: interoperability, performance, collaborative analysis.
- 4 year NSF funded project (Princeton/U.Nebraska/NYU/Cincinnati). Nominally ends mid-2019, but may continue in part with no-cost extensions into 2020.
- Recent activities: ROOT I/O improvements, interoperability of HEP tools (ROOT) with Spark, ROOT I/O in Python w/o ROOT (uproot), Scikit-HEP toolkit, ROOT modularity, etc.
- DIANA/HEP R&D adds energy to the ROOT I/O development as well as contributing to a number of ideas in the analysis facility part of the CWP
- Note in particular the frequent DIANA/HEP “topical” meetings on R&D topics: https://indico.cern.ch/category/7192/
- http://diana-hep.org

“Big Data” Projects

- FNAL/Princeton/CERN(IT,Openlab)/Intel, USCMS Ops partnership, recently adding INFN-Padova and UChicago
- Investigate industry “Big Data” tools for HEP analysis
  - https://cms-big-data.github.io
- Analysis Thrust
  - Full analysis in Apache Spark
- Intel Big Data Reduction Facility
  - “1 PB to 1 TB in 5 hours”
- Fermilab LDRD (lab directed R&D)
  - Striped data representation with multiple layers of caching
- Using developments from DIANA and CERN-IT
  - spark-root
  - spark-xrootd-connector
  - Uproot
- UChicago exploring auto-provisioning of Spark Clusters

HEP Data Analytics

- Integration of new analysis tools in CMSSW ecosystem
  - Infrastructure facilitating R&D
  - Python analysis tools from the Data Science community, promote and enable python as analysis language for HEP (eg, Tensorflow now available from CMSSW)
  - Site tuned installations of toolkits provided to CMS collaborators by OSG & CMS connect (facilitating R&D on GPUs, etc)
  - New interest: FPGA for ML inference and beyond
- UChicago ATLAS and OSG efforts
  - ELK service with GPU analytics servers
  - Indexing of PanDA, Rucio, FTS, perfSONAR & LHCONE networking metrics
  - Analysis of facility and services performance
  - ML for infrastructure - intelligent alerting & alarming

Facilities

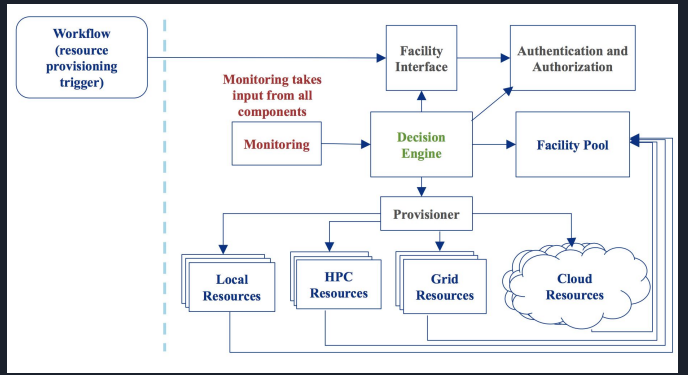
HPC


- Significant push by US funding agencies (especially DOE) to utilize HPC
  - HPC center architectures range from x86-compatible resources to PowerPC + GPU
  - ATLAS/CMS both focused on running full software workflows on these systems
  - ATLAS derived 10% of their total computing time from these systems in 2017 running simulation jobs
- Job Management & Software Distribution
  - ATLAS/CMS working toward common container production and distribution system, working with CVMFS-team for container population
  - ATLAS employing an edge service (Harvester) to interface between outside world and secure internal HPC network; discussions ongoing about expanding it for CMS
- Effort needed to exploit the capabilities of the new architectures
  - Exascale machines arrive in 4 years with 100x current computing power
  - Bring large computing resources but free isn't always free
  - Variation in architectures likely to increase as industry drives the hunger for Machine Learning friendly architectures

Job Management

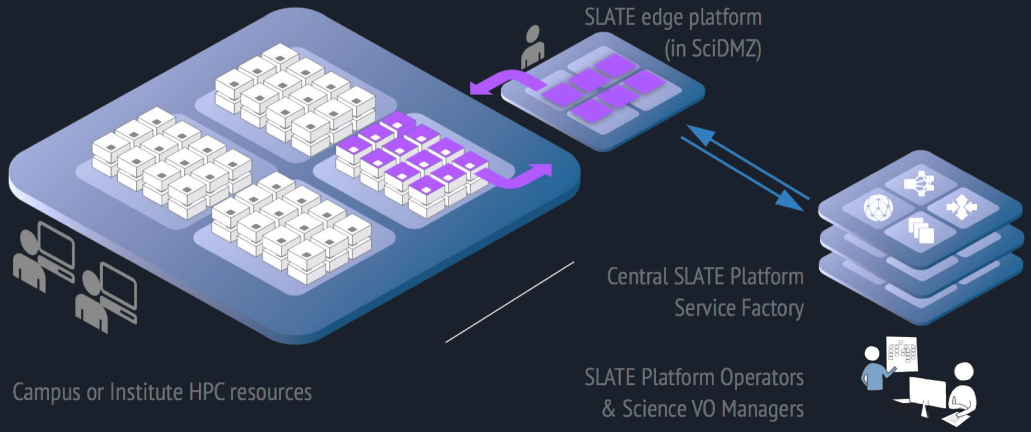
- BigPanda (SciDAC)
  - Pilot 2.0 being developed for heterogeneous resources, storage, and science domains
  - Harvester service development for fine grain job control
  - Incorporating and Tracking Containers
  - Local caching capabilities to improve data management

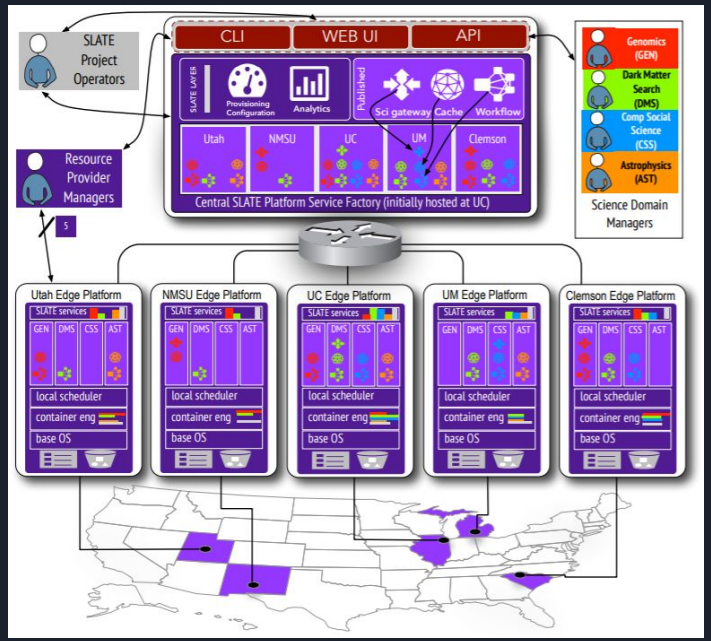
- HepCloud
  - Portal to an ecosystem of diverse computing resources commercial or academic
  - USCMS strategy for elastic scale out of compute resources
  - T1_US_FNAL endpoint scales into
    - Allocations at US supercomputer centers (HPC)
    - Commercial cloud providers
    - Opportunistic resources
  - Multiple demonstrations on AWS and Google Cloud
  - Intelligence to make decision where to send work



SLATE - Service Layer at the Edge

- 3-year NSF-funded project (UChicago/UMichigan/UUtah)
- Deployment of domain-specific services a burden on local sysadmins, which limits use of local resources as part of the national cyberinfrastructure
- SLATE vision: a programmable “underlayment” edge platform for deploying higher level services






Infrastructure Software

Data Management Evolution - Rucio

- Rucio manages data replication and enforces data placement policy for ATLAS.
  - ATLAS demonstrates functionality and scale, but data replication / management is a common problem across all HEP experiments.
- Over the past year, the Rucio team has worked to make Rucio more generic and fit multiple experiment's needs.
  - Already working in production for Xenon1T and AMS.
- OSG has been working with several US organizations to provide a platform for Rucio evaluation.
  - Provides expertise, surrounding infrastructure (FTS), and hosted instances.
- CMS has been testing this at scale as one option for post-Run2 data management.
  - Already, nearly 1PB moved with Rucio - including nearly 600TB to NERSC as a simple scale test.
  - Other option is Dynamo.
- Rucio would provide a significant, visible platform for cross-LHC collaboration.
  - And beyond: evaluation ongoing at FNAL for neutrino experiments.

Data organization, management and access R&D

- Caching Proxies and Delivery Services;
  - Work on expanding current prototypes at UCSD and Caltech to the point where a processing site could *only* have caches.
    - Current activities focus on understanding IO patterns and working set size.
  - Similar activities in ATLAS (particularly SLAC & Chicago), including using cache as an edge service at an HPC site
  - Delivery service from data lakes (c.f. http://bit.ly/atlas-lakes)
- ROOT-aware software-defined storage:
  - Idea: using a programmable storage layer (such as Ceph) that allows the storage layer to optimize data locality or compression approaches for the most popular objects in a file.
  - Initial exploration being done at UCSC (core Ceph expertise), Nebraska, UCSD.
- Google Data Ocean:
  - R&D partnership between Google & ATLAS developed over ~6 months
  - Goals include creating a cloud analysis service
  - Reduce dependency on data locality
  - Google + BNL/UTArlington/CERN


Training

Training

- To support the R&D activities above, developer training is needed
- USCMS has participated in CMSDAS, of course, which is one eventual vector for introducing new tools for analysis to users.
- The U.S. has however lacked a HEP-focused school for advanced development training for software. Schools like ESC (INFN, Bertinoro) have trained many people involved in aspects of advanced development in CMS over the years. In the U.S. this has fallen a bit between the cracks of various programs and missions.
- NSF funded CoDaS-HEP (http://codas-hep.org) for 2017 and 2018 (applications soon!), which aims to provide a broad introduction: parallel programming, Big Data Tools, Machine Learning technology and methods, performance, development tools.
- We are exploring how to sustain this going forward.


Community

S2I2-HEP Conceptualization

- “Planning” project whose deliverables included enabling the process to produce the CWP document as well as a “Strategic Plan” for a possible NSF-funded software institute.
- 2 year NSF funded project (Princeton/UIUC/Cincinnati), nominally completed with these deliverables:
  - A Roadmap for HEP Software and Computing R&D for the 2020s (Community White Paper) - arXiv 1712.06982
  - Strategic Plan for a Scientific Software Innovation Institute (S2I2) for High Energy Physics - arXiv 1712.06592
- Next step is now up to NSF
- However some participant cost funds remain and these are being spent to support travel to CWP follow-on activities, e.g. HSF/WLCG and other small planning workshops.
- http://s2i2-hep.org