Update on Software

Benedikt Hegner
(EP-SFT, CERN)

LHCC Referee Meeting
20.9.2016

Using input from HSF and GeantV

Outline

- Common SW Activities Update
- ROOT long-term planning
- GeantV status meeting
- HSF Community White Paper

Common SW Activities Updates

- **Gaudi Workshop at CERN (Sep 21 -23)**
  - Defining future direction of Gaudi for upgrade
  - Participation from ATLAS, LHCb, FCC, Daya Bay
  - Result will be an updated *Gaudi Design Document* and a *common plan of work*

- **LCG releases**
  - Traditionally the LCG externals have been used by LHCb and ATLAS
  - Introducing “views” the usage of those has been decoupled from CMT and other experiment-specific environment setting
  - User base now widening (FCC, CLIC, SWAN service)
  - *Effort embedded in HSF packaging working group*

Future Evolution of ROOT 1/2

- **ROOT 6 well established now**
  - Next production release end of this month
  - Closer collaboration with experiments via regular ROOT planning meetings

- **Future Directions defined beginning of this year**
  - Presented to LHC experiments and other users

- **Full exploitation of the Cling interpreter**
    - JIT compilation opens many possibilities for e.g. interactivity, serialization, etc

- **I/O Evolution**
    - Better concurrency, re-thinking file-format, support for newer C++

- **Modern C++ interfaces (“ROOT 7”)**
    - Take advantage of new C++ standards in user interface
    - Use it as opportunity to clean up the 20 year old interfaces (intentionally backwards incompatible!)

- **Re-thinking user interface in general**
    - Thin-client web-based user interfaces (JavaScript, ROOTbooks)
    - SWAN (formerly ROOT as-a-service)

Future Evolution of ROOT 2/2

- **Improving binding to other languages / eco-systems**
  - Making Python a “first class citizen” and
  - Exploit Python eco-system (lots for machine learning!)

- **Parallelization and Vectorization**
  - Exploit current and new hardware as much as possible
  - Dedicated sub-project of ROOT testing and implementing ideas

- **Engaging the community**
  - Ensuring proper prioritization via discussion with experiments
  - Vision is a ROOT eco-system allowing for contributions
    - Leaving ownership and maintenance in the hands of the contributors
    - Core stays in the full responsibility of the ROOT team though
  - Requires proper packaging and more modularization
    - Not an easy change after 20 years of code history
  - DIANA/HEP as example for cooperation with other projects

- **All these items have (prototype) milestones and deliverables**
  - One nice example is SWAN (Service for Web Based Analysis), which runs as pilot service since June 2016
    - Combining ROOT notebooks, EOS storage, and Spark

- **More details in presentations about ROOT status and plans**

CMSDimuon_py Last Checkpoint: an hour ago (autosaved)

Control Panel Logout

File Edit View Insert Cell Kernel Help Python 2

Code CellToolbar

Dimuon spectrum

This ROOTbook produces a plot of the dimuon spectrum starting from a subset of the CMS collision events of Run2010B.

Dataset Reference:
McCauley, T. (2014). Dimuon event information derived from the Run2010B public Mu dataset. CERN Open Data Portal. DOI: 10.7483/OPENDATA.CMS.CB8H.MFFA.

In [ ]:
```
import ROOT
```

A little extra: JavaScript visualisation. This command will become a magic very soon.

In [ ]:
```
%jsroot on
```

Convert to ROOT format and analyse

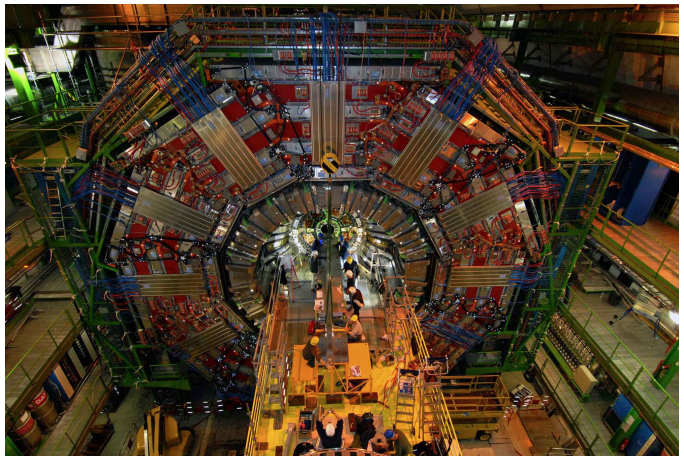
First of all we convert the csv file into ROOT format, i.e. filling up a TTree data structure. But first of all we uncompress it if it's not.

In [ ]:
```
inputFileName = '../data/MuRun2010B.csv'
import os
if not os.path.exists(inputFileName):
    import gzip
    import shutil
    with gzip.open(inputFileName+'.gz', 'rb') as f_in, open(inputFileName, 'wb') as f_out:
        shutil.copyfileobj(f_in, f_out)
```

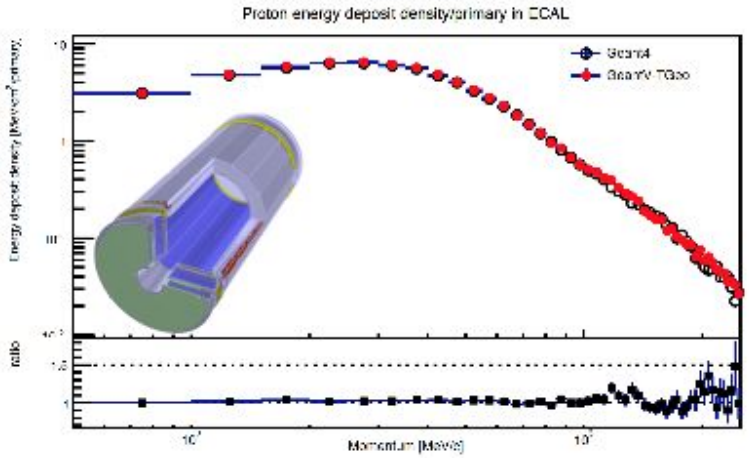
See presentation **Service for Web Based Analysis (SWAN)**

GeantV Meeting

- **After 3 years of development the GeantV team approached the HSF for expert feedback to**
  - Make sure the project is ‘on track’
  - Assess whether the goal of a ‘proof of principle’ of the design chosen was achieved
  - Look for possible show-stoppers until completion
  - Give suggestions for improvements and next steps
  - Make the project known to a wider community and potential new contributors
  - Convince clients and funding agencies



- **“HEP software community meeting on GeantV R&D”**
  - Taking place 25.-27. October
  - Presenting the GeantV status and plans to a wider audience
  - Simulation experts from LHC experiments and Belle II as “reviewers”
  - Detailed description of goals in [this document]



Community White Paper Charge

- **Discussed the need for a longer-term strategy for HEP software**

- **Prepared a CWP charge to HSF from WLCG side**
  http://hepsoftwarefoundation.org/assets/CWP-Charge-HSF.pdf

"...

More specifically the CWP should identify and
prioritize the software research and development investments required:

1) to achieve improvements in software efficiency, scalability and performance and to make use of the advances in CPU, storage and network technologies
2) to enable new approaches to computing and software that could radically extend the physics reach of the detectors
3) to ensure the long term sustainability of the software through the lifetime of the HL

...
"

Community White Paper Planning

- **Initial kick-off meeting this week (Sep 22)**
  - Involving all LHC computing coordinators
  - Draft defining working groups, priorities, and milestones

- **Discussion with wider audience as pre-CHEP meeting (October 9th)**

- **HSF Workshop in San Diego (Jan 23 - 26)**
  - Focus on CWP
  - Dates announced last week

Information and updates to be circulated via further HSF newsletters

Backup

GeantV - Basic Motivation

- **Fetching data and instructions from RAM has a cost**
  - Often much higher than expected
  - Caches misses are not something to take lightly
- **What NOT to abuse**
  - Sparse access over large data structures
  - Frequent incoherent low granularity allocations
  - Piping small data through highly branching processing logic
  - Virtual interfaces on top of fine grain data units
- **Keeping up with technology**
  - Instruction-level parallelism
  - SIMD
  - Prefetching
  - Out of order execution

- **The GeantV project tries to address all these points with a complete re-design of simulation software by grouping multiple items and operations into *baskets***