LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Offline Computing and Software Summary

on behalf of all the excellent presentations in both sessions

Martin Ritter
LMU München

The 2020 International Workshop on the High Energy Circular Electron Positron Collider

CEPC Software Status

Prof. Xingtao Huang (Shandong University)

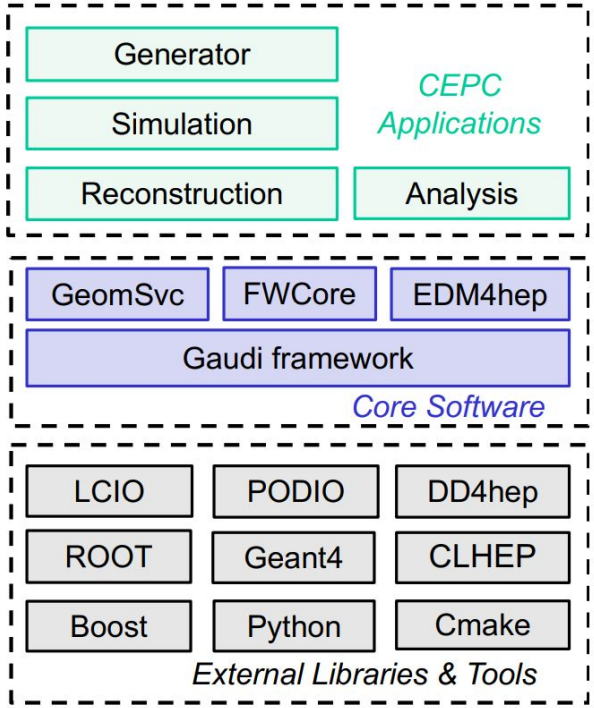
CEPC Software now using

Common Software Stack (Key4HEP)

- shared effort between CEPC, FCC, ILC, CLIC, …
- common core software and external dependencies
- experiment dependent applications on top.

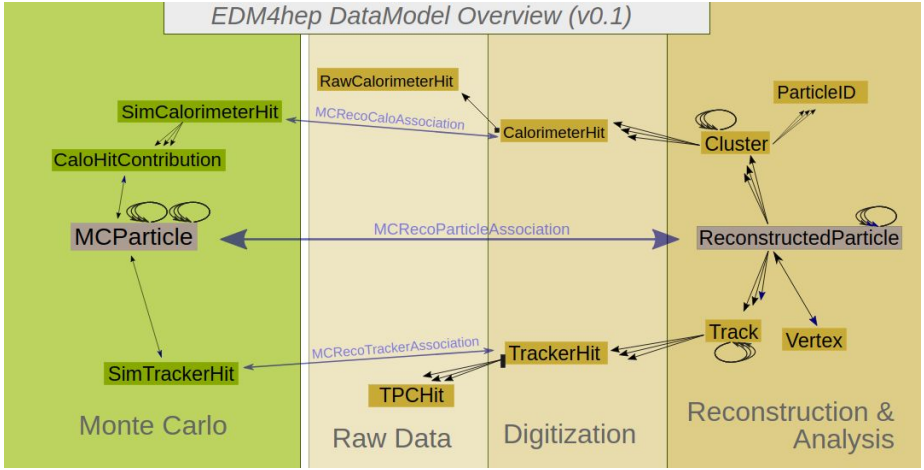
⇒ Impressive amount of progress

- fully validated with respect to previous iLCSoft effort
- active contribution to Key4HEP by CEPC
- public and managed on Github
- available via CVMFS on all CEPC sites.

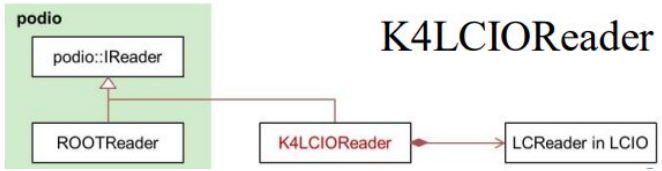


New Event Data Model

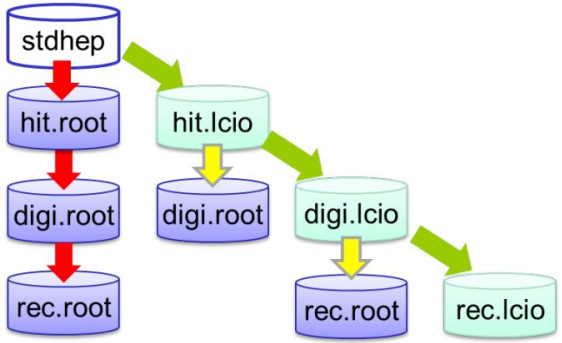
Now fully adopted EDM4hep as event data model



- Developed additional software convert previous LCIO files.
- tools becoming part of Key4hep



- Very useful when porting algorithms from Marlin to CEPCSW and using the LCIO data to compare the results
  - **The Green arrows:**
    - LCIO+Mokka+Marlin
  - **The Yellow arrows:**
    - Input: LCIO
    - Output: EDM4hep
  - **The Red arrows:**
    - EDM4hep+DD4hep

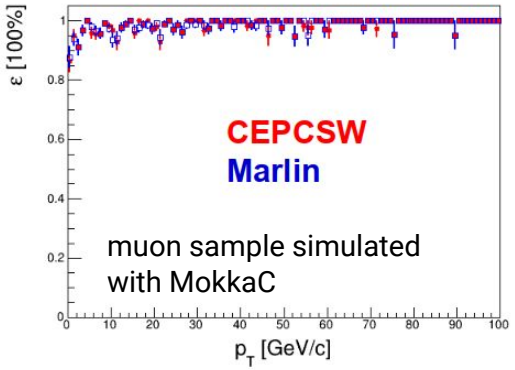


Tracking in TPC and silicon

Yebo Chen (IHEP)

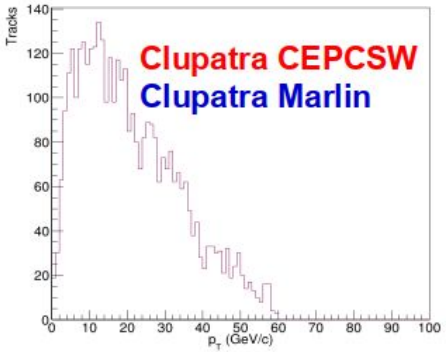
Moving from iLCSoft to CEPCSW

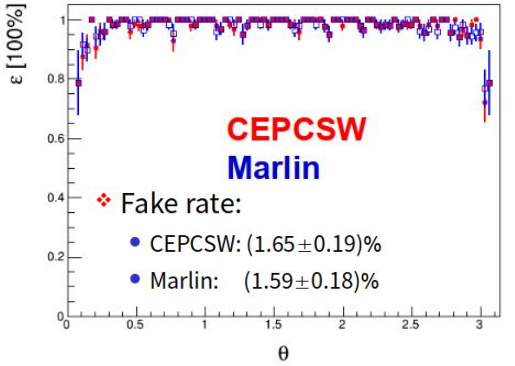
- benchmark for the new framework
- **all** steps of the reconstruction chain converted successfully
- consistent results in both frameworks
- fully identical if same digitization is used



Future plans:

- improve realism of digitization
- clustering of hits
- non-uniform magnetic field
- background mixing





Status of Drift Chamber Software

Geometry has been implemented

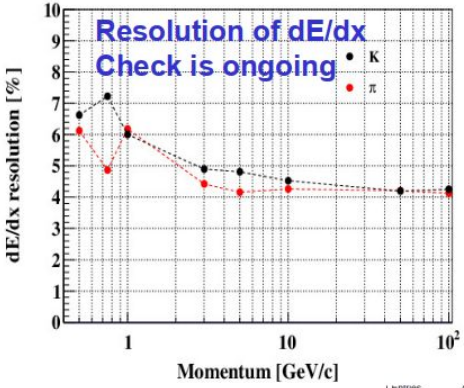
- inner + outer chambers
- baseline: 1.8 m, 130 layers, He:iC_4H_{10} = 90:10
- cell partitioning with segmentation method

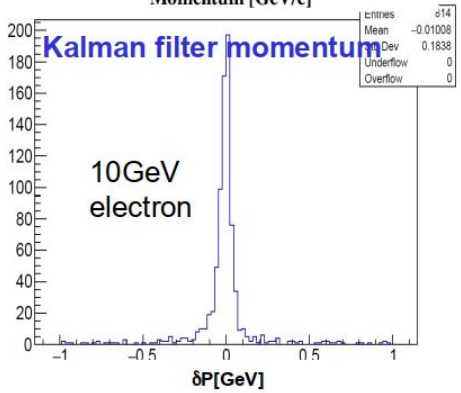
Particle Identification

- Implemented dE/dx simulation method
- preliminary studies show possible improvement using cluster counting instead

Reconstruction

- Track fitting with standalone Genfit2 works as expected





Plans for Drift Chamber Software

First Release of drift chamber software

- baseline with axial wires
- simple dE/dx

Simulation and Geometry and Waveforms

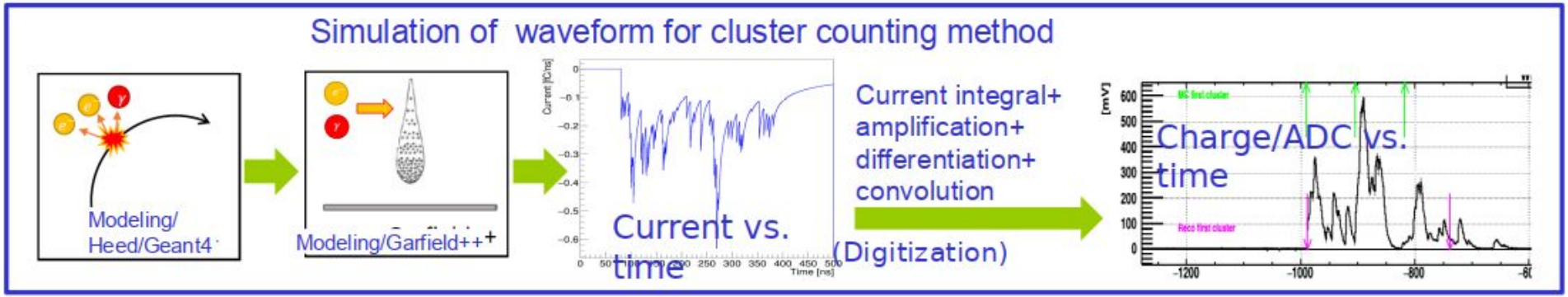
- implement stereo wires
- implement fast Garfield++ waveform

Standalone drift chamber studies

- cluster counting with Garfield++
- Trackerr for momentum resolution

Development of drift chamber tracking

- combine silicon and drift chamber
- track finding (seeding, self-tracking, ML)



ACTS

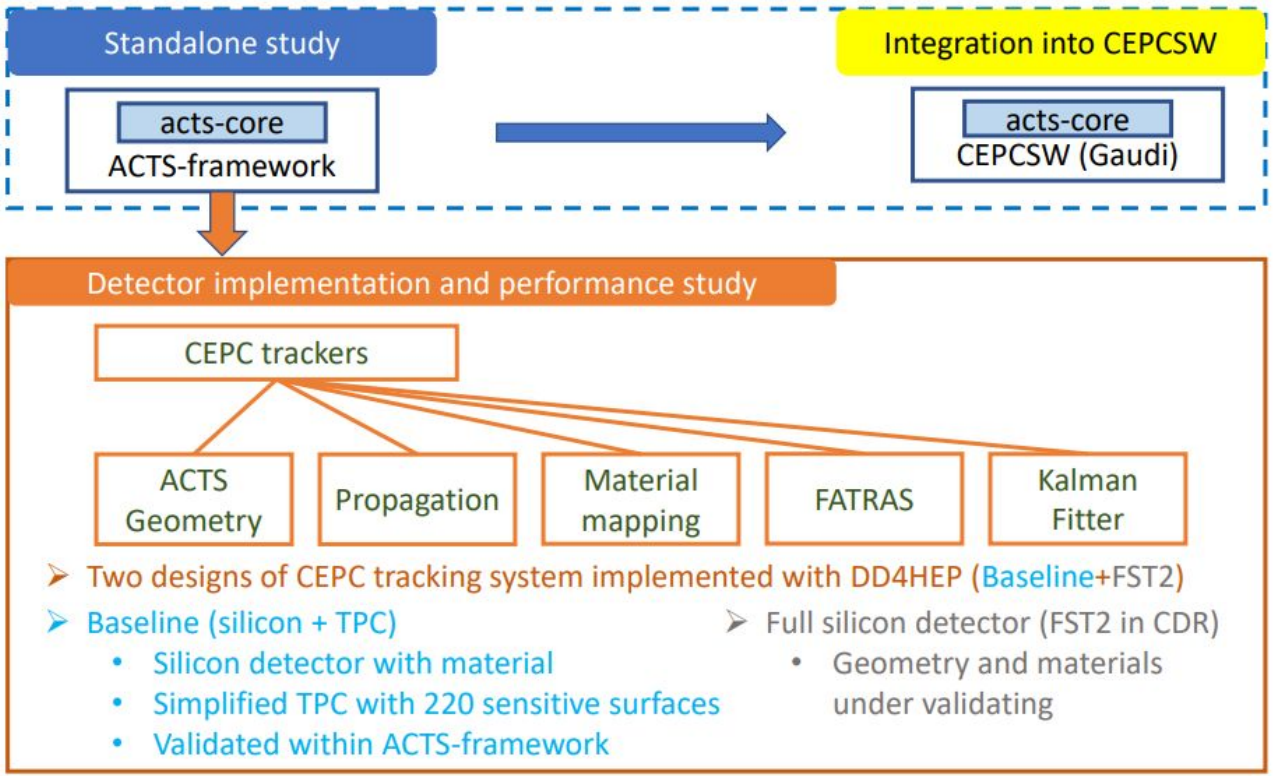
Yebo Chen (IHEP)

ACTS: A Common Tracking Software project initiated by ATLAS

- under evaluation by multiple experiments
- modern technologies
- active development

First step: standalone study

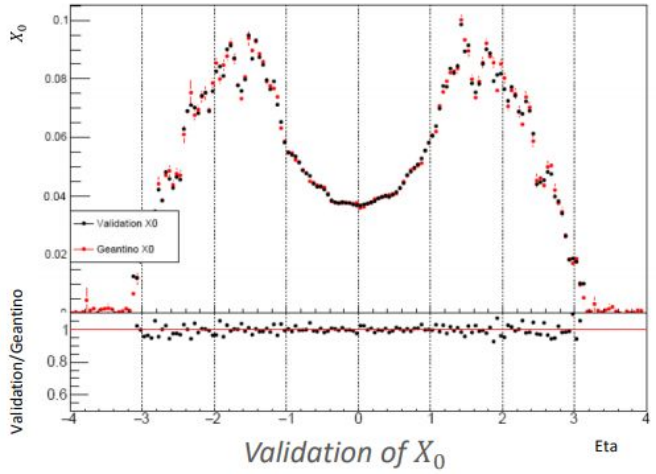
Second step: integration to CEPCSW



ACTS Studies

First result with ACTS fast simulation (FATRAS)

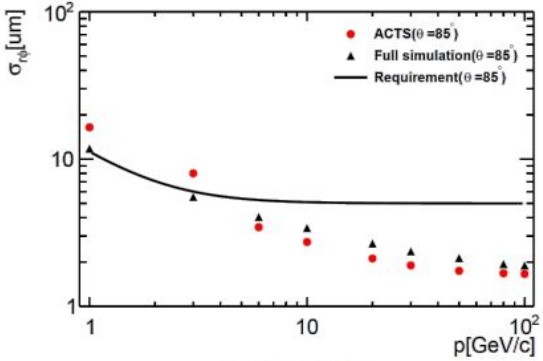
- truth track finding
- general agreement with full simulation
- more validation needed.



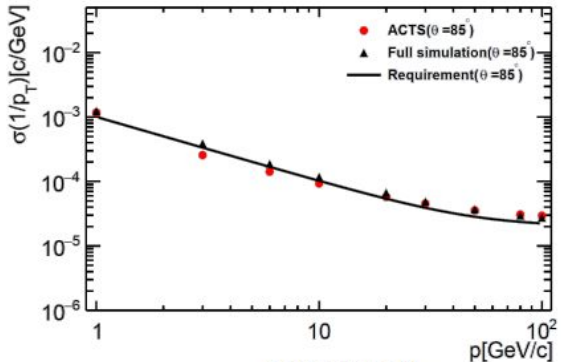
Validation of X_0

Next steps

- Optimize and improve geometry description
- Layout and physics studies
- Integration into CEPCSW



Resolution of D_0



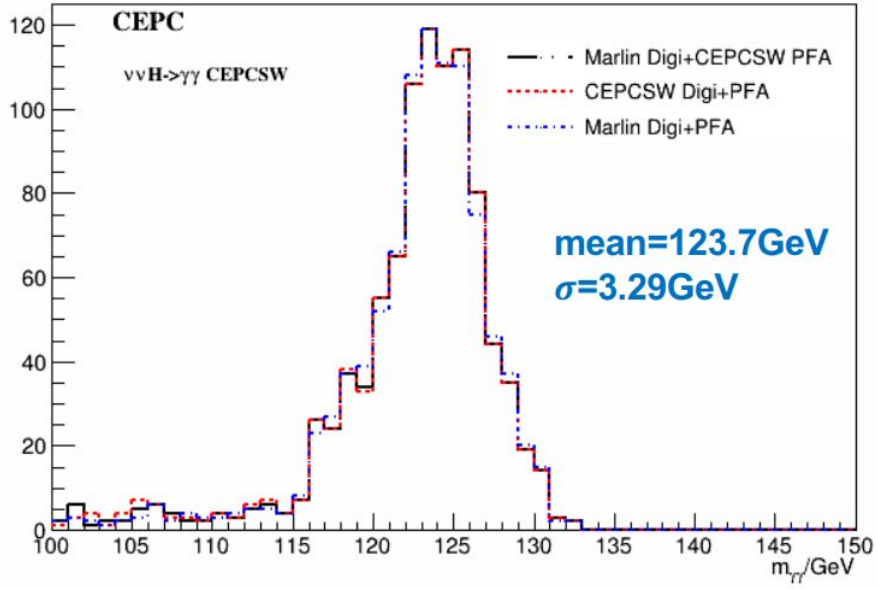
Resolution of $1/p_T$

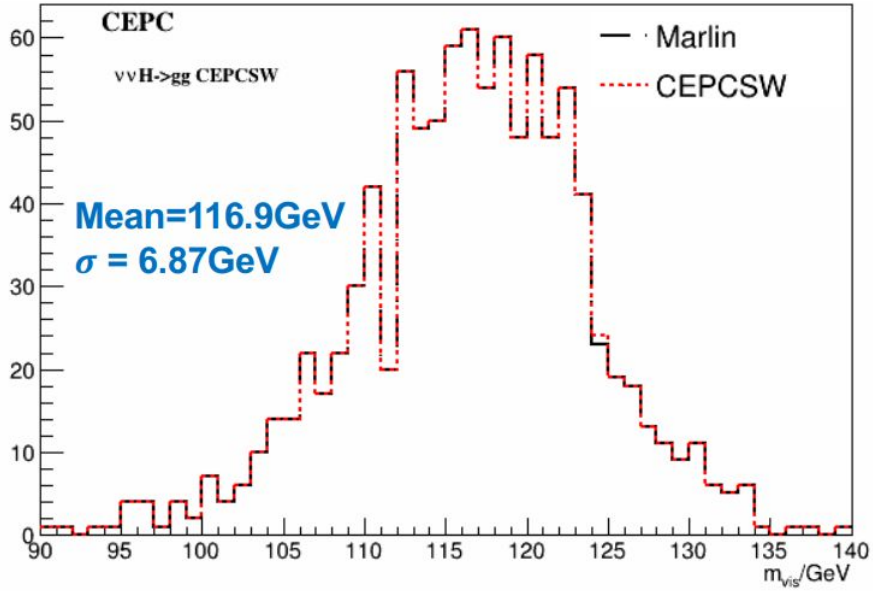
Simulation and Reconstruction in ECAL

Dr. Sheng-Sen Sun (IHEP)

Full simulation and digitization of ECAL works well in CEPCSW

- compatible responses with previous software
- slight differences persist in simulation, being worked on.





Further ECAL developments

Pandora also ported to Key4HEP

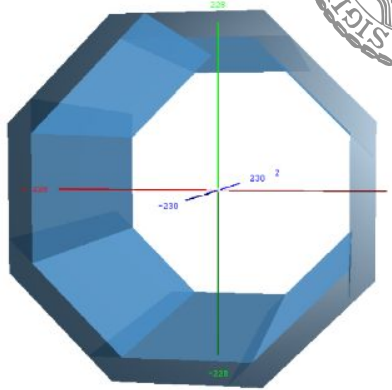
- validation with gamma events so far.
- extend to hadrons and jet energy resolutions

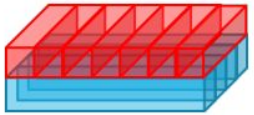
Fast simulation for ECAL showers (Frozen showers)

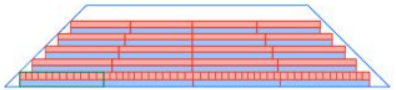
- for low energies use pre-generated low energy showers.
- target speedup of factor 2

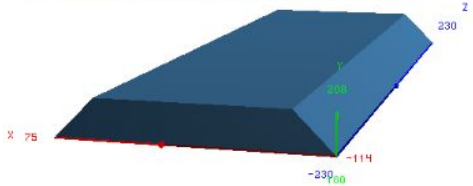
Crystal Bar Solution for the ECAL

- A crystal bar ECAL design for optimal intrinsic energy resolution and reduction of number of channels.
- preliminary: can extract 3D information for single gamma events
- major focus: multiple incident particles hitting one super cell









Analysis Tools and Infrastructure

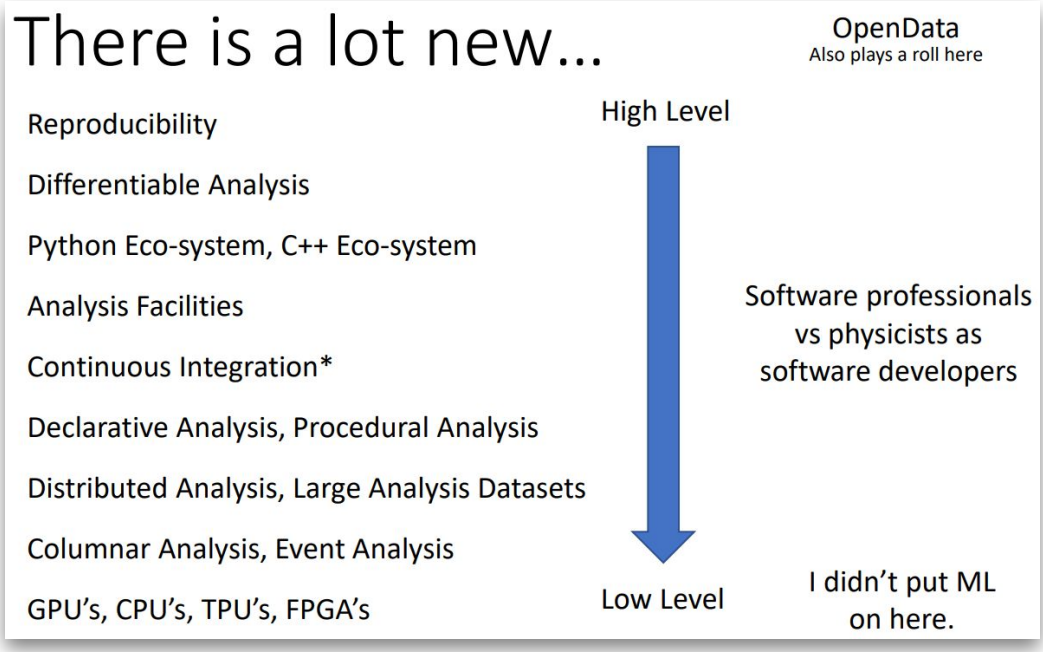
Gordon Watts (University of Washington)

Nice overview of current and upcoming challenges in analysis

- very hard to summarize
- how to scale to the upcoming data?
- declarative analysis to the rescue?

Analysis is one of the fastest moving parts of HEP

➔ very active community
- The HEP Software Foundation
- In the USA, IRIS-HEP
- ROOT workshops
- And many more



ROOT: Status and Plans for ROOT

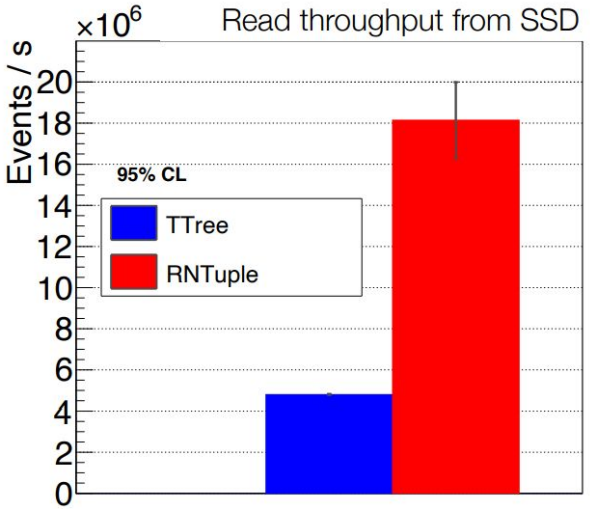
Axel Naumann (CERN)

HEP specific framework

- allows to predict changes, adapt and benefit
- tailored to our very own problems
- focusing to make ROOT easier to use.

ROOT 7: massive multi year modernization project

- focused on I/O, analysis, graphics and histograms
- most relevant for physicists
- early versions available:
  WebGUI, RNTuple, RHist,



New Features in ROOT

```
ROOT::EnableImplicitMT();
ROOT::RDataFrame df(dataset);
auto df2 = df.Filter("x > 0")
             .Define("r2", "x*x + y*y");
auto rHist = df2.Histo1D("r2");
df2.Snapshot("newtree", "out.root");
```

PyROOT

- completely new backend
- simultaneous python 2 and 3 support

RooFit

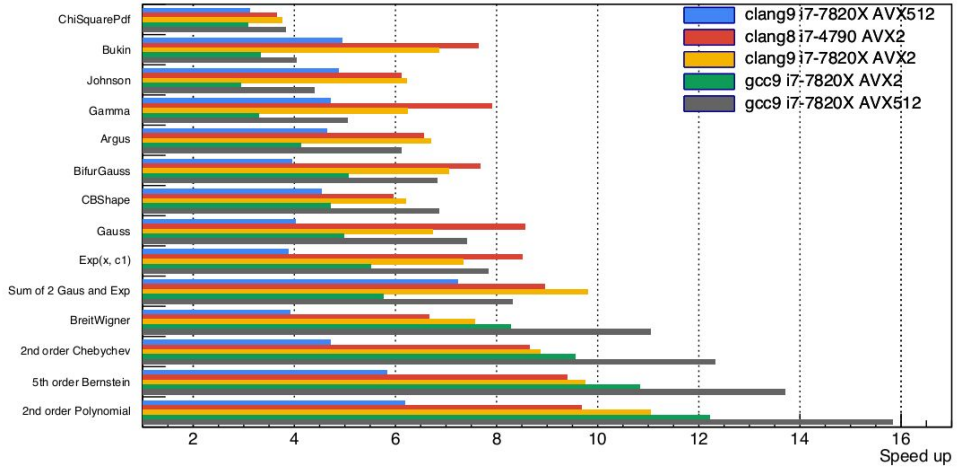
- speedup by large factors
- improving PyROOT integration

RDataFrame

- efficient declarative analysis

TMVA

- Adapters to external backends: TensorFlow, Keras, scikitlearn

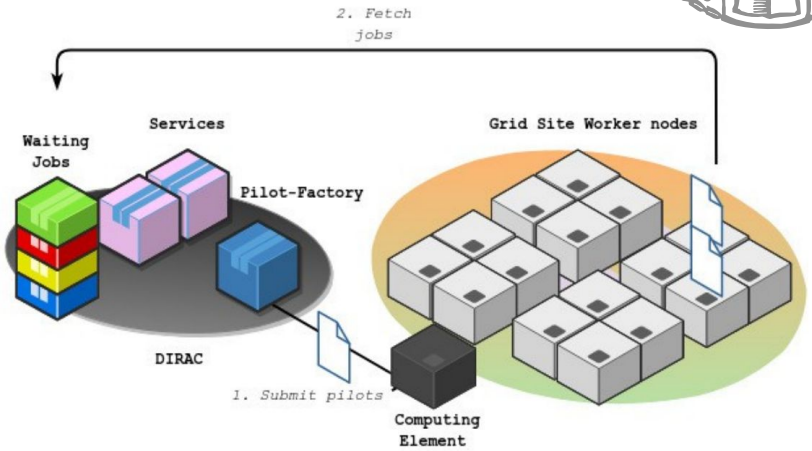


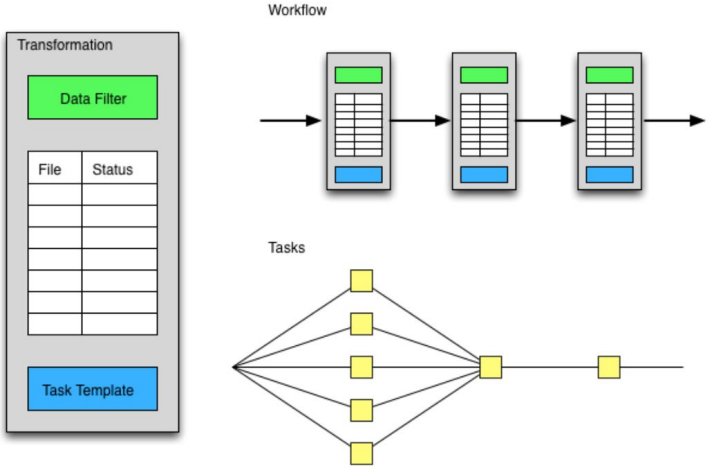
DIRAC: The Interware

Frederico Stagni (CERN)

A software framework for distributed computing

- layer between users and computing resources
- used by multiple experiments/projects including CEPC
- experiment agnostic and extensible





Handles both Workflow and Data mangement

- Generic system for requests and transformations
- Allows to chain productions, scaling to millions of files

DIRAC: Upcoming Features

- Replace proprietary protocols with https
- Support Python 3

New underlying DIRACOS2

- CentOS 8 + Python 3
- easy to extend by experiments

External Authentication

- Introduce OAuth/OIDC
- both web based and CLI

DIRAC ↔ Rucio Bridge

DIRAC ↔ Rucio

Cedric Serfon (BNL)

Rucio: Scientific Data Management

- initially developed for the ATLAS experiment
- in production since 2014
- used by a large and growing community

Belle II uses DIRAC and a custom Data management

- consolidate efforts by moving to Rucio
- adapt/extend Rucio to fit Belle II needs
- integrate Rucio into DIRAC
- validation successful, move to production soon

Extended Rucio for Belle II

Chained data subscriptions

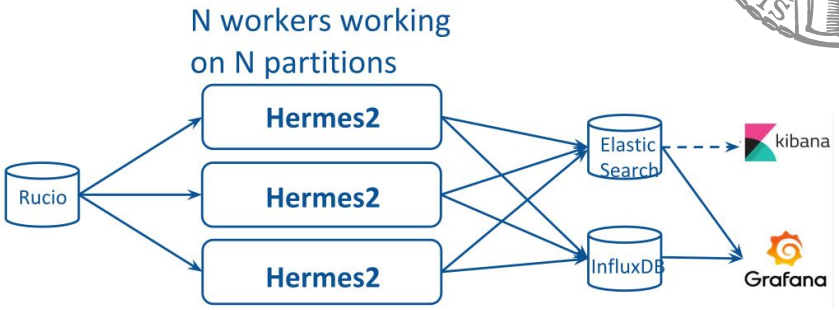
- move data from A→B→C
- for example first to disk, then to tape

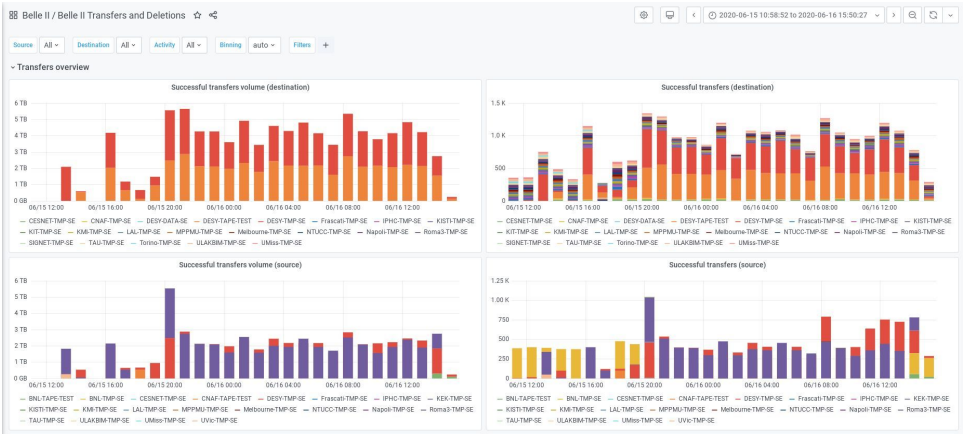
Simplified Monitoring System

- Rucio daemon to directly aggregate monitoring information
- Easy to deploy for any collaboration

Rucio File Catalog plugin

- allows DIRAC to directly talk to Rucio
- will be integrated into DIRAC





Software Testing & Validation at LHCb

Chris Burr (CERN)

The amount of progress on CEPCSW is very impressive

- very active in Key4HEP community
- modern development flow with code review and CI builds
- successful validation of performance with respect to iLCSoft

But automated testing and validation still in early stages

- very comprehensive overview over the testing and validation at LHCb
- total of 700 CPUs involved in building and testing

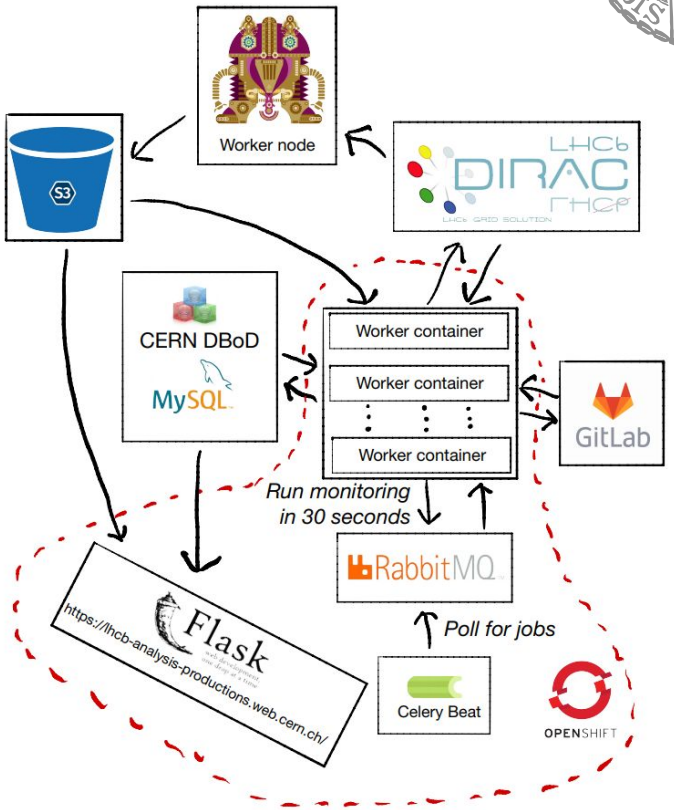
Software Testing & Validation at LHCb (cont.)

Cover a large variety of use cases

- basic compile and unit tests
- performance benchmarks
- nightly validation
- physics validation on demand
- large release validation campaigns
- special builds for GPU HLT (Allen) validation
- analysis job validation using gitlab + DIRAC

Complex pipelines with advanced monitoring.

- Utilize standard tools where possible
- custom extensions where necessary



Software Testing & Validation at LHCb (cont.)

Example “analysis validation”

Analysis production job for D02HH_Starterkit

2016_MagDown_PromptMC_D02KK

Status	Commit	Requested	Processed	Runtime	Kept
Success	53210696	-1 events	2548 events	0:00:34	True

Input

Name	Size	Total
/lhcb/MC/2016/ALLSTREAMS.DST/00070793/0000/00070793_00000206_7.AllStreams.dst	544.8 MB	0.4 TB

Output

Name	Size	Total (estimated)
00012345_00006789_1.d02kk.root	0.2 MB	0.1 GB

Browse output file Show

Reproduce on lxplus:

```
lb-ap reproduce 1958460 "D02HH_Starterkit" "2016_MagDown_PromptMC_D02KK" 0
```

Explain job log Show

Job log Show

DIRAC log Show

If the logs show "Error 404 while fetching file: NOT FOUND" it is sign the test job has not started

For support see the DPA WP2 Analysis Productions channel on mattermost.

LbAnalysisProductions 0.0.2.dev2+g2875ced · LbAPCommon 0.0.4

Browse output file Hide

Read a ROOT file

JSROOT version 5.8.1 7/05/2020

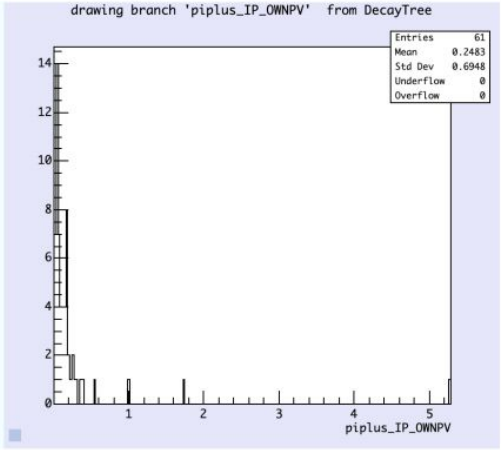
open all | close all | clear

00012345_00006789_1.d02kk.root

TupleDstToD0pi_D0ToKK

DecayTree;1

- Dstar_ENDVERTEX_X
- Dstar_ENDVERTEX_Y
- Dstar_ENDVERTEX_Z
- Dstar_ENDVERTEX_XERR
- Dstar_ENDVERTEX_YERR
- Dstar_ENDVERTEX_ZERR
- Dstar_ENDVERTEX_CHI2
- Dstar_ENDVERTEX_NDOF
- Dstar_ENDVERTEX_COV_
- Dstar_OWNPV_X
- Dstar_OWNPV_Y
- Dstar_OWNPV_Z
- Dstar_OWNPV_XERR
- Dstar_OWNPV_YERR
- Dstar_OWNPV_ZERR



Explain job log

Fatal errors:

Suggestions:

Other explanations:

- **Line 356:** Histograms are not being saved as no filename has been specified for storing them. This message is harmless and normally ignored.

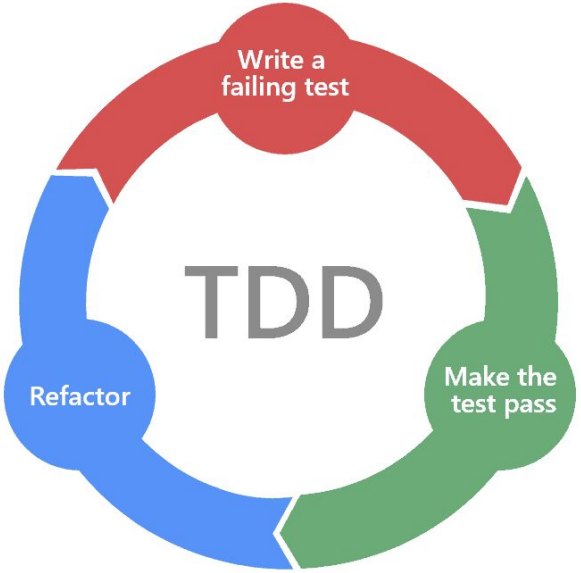
Job log Hide

Software Testing & Validation at CEPC

Important to invest early into testing and validation infrastructure

A comprehensive testing system like LHCb takes years of work

- CEPC should invest some effort to start right away
- Otherwise the effort will be huge
- ... and resources will be scarce



Now would be a very good time to focus some effort

This should not be limited to offline

- all software parts can benefit greatly from test infrastruct
- see Allen in LHCb.

Conclusions

Huge progress on CEPCSW

- new framework, new edm, new everything
- huge effort to convert and improve the algorithms
- validated with respect to previous software

Many interesting contributions from “outside”

- ROOT plans for the coming years
- DIRAC overview and plans.
- DIRAC Rucio plugin coming in the future
- Data analysis …
- Software testing and validation
  ⇒ good time to invest effort!