Status of Computing

Stefano Spataro

UNIVERSITÀ DEGLI STUDI DI TORINO
ALMA UNIVERSITAS TAURINENSIS

ISTITUTO NAZIONALE DI FISICA NUCLEARE
Sezione di Torino

Tuesday, 2nd December 2015

Computing extensively discussed in the last 12 months

December 2014 – Status of Computing (myself) - 1h20'
March 2015 – FAIR Computing activities (Thorsten Kollegger)
September 2015 – Extensive computing review – almost 3h

- Stefano Spataro – General Status of Computing
- Radek Karabowic – Status of Core Developments
- Gianluigi Boca – Review of Tracking Software
- Paul Buehler – Review of Data Production Systems
- Klaus Goetzen – Review of Analysis Tools
- Mark Palizaeusz – Overview of Analysis activities

Tracking

TARGET SPECTROMETER

- ✓ Pattern Recognition in Central Tracker Working since 2011
- ✓ Basic assumption: external t0
- ✓ STT Region well covered

Recent Activities

- ✓ Definition of standard QA for tracking
- ✓ Automatic QA in dashboard
- ✓ Secondary tracking algorithm, covering GEM region
- ✓ Genfit2 improves low momentum (fixed memory problems)
- ✓ Needed Track Cleaning
- ✓ CA Tracking developments ongoing (last months -> standby)

FORWARD

- ✓ Pattern Recognition in Forward Tracker only Ideal since 2012
- ✓ No progresses this year

Particle Identification

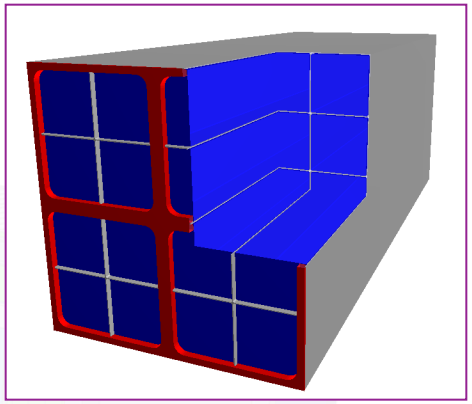
	Det.	PID Variables	Track Correlation	Algorithm
TARGET SPECTROMETER	MVD	dE/dx	YES	Bayesian
	STT	dE/dx truncated mean	YES	Bayesian
	GEM	-	-	-
	EMC	E/p, Shower Shape	YES	Bayesian, TMVA
	DIRC	Cherenkov angle (fast digi)	YES	Bayesian
	DISC	Cherenkov angle (fast digi)	YES	Bayesian
FORWARD	SciTil	Time of Flight	YES	NO
	MDT	# layers, # hits, track χ^2	YES	Hard Cut
	FTS	-	-	-
	FTOF	Time of Flight	YES FTOF NO DTOF	NO
	FSC	E/p, Shower Shape	YES	NO
	RICH	-	NO	NO

Good for TS
Bad for FS

New (advanced) efforts for RICH, not yet in svn

Available
Mosty available
Not available

What about EMC and neutrals?



Geometry

- Mostly just crystals – almost no passive structures

Digitization

- Time based simulation recenty updated

Reconstruction

- Clustering, bump splitting, energy corrections, etc…
- Correlation to charged tracks
- Uncertainty in calibration, covariance matrices, stability, random maintenance
- Improve algorithms, split-off handling, covariance matrices
- Need to improve neutral reconstruction, noise suppression

Recent Activities

- MC Matching fixed
- Studies on Preshower in the DISC (by SciTil group)

Proposed EMC software coordination in Bonn, but BMBF stopped

What can I do with PandaRoot?

Charged channels analysis

- In the Target Spectrometer: everything!
  - Good tracking
  - Realistic PID (almost)
  - Secondaries: improved, but tests are needed
- In the Forward Spectrometer: well…
  - Only Ideal tracking
  - No PID algorithms, only PID variables

Channels with neutrals

- No passive elements, but full reco chain
- Uncertainty in the response (cov matrix? Calibration?)

Outcome from the SG

- We need more computing manpower from the institutes
- Low threshold for physics analysis (using virtualization)
- More efforts on the time-based simulations
- A Computing TDR
- A realistic funding plan and risk analysis

- Strong reliable code
- Centralized data production
- Analysts free from installing software

Time based simulation

	Det.	Time Stamp	Pile-Up simulated	Reco using time information
TARGET SPECTROMETER	MVD	Yes	Old digi kept – energy modified	MVD Riemann track finder
	STT	Yes	Throw away new data	Two different track finders, not fully in PandaRoot
	GEM	Yes	Old digi kept – energy modified	GEM track finder
	SciTil	Yes	Yes	Ongoing
	DIRC	Yes	Old digi kept	Yes
	DISC	(Yes?)	??	(Yes?)
	EMC	Yes	Modified waveforms	Bump builder
FORWARD	Muon	Yes	Modified waveforms ongoing	Ongoing
	FTS	Yes	Throw away new data	No
	FTOF	No		
	FSC	Yes	?	?

In December 2014 a list of software tasks was provided to the CB

Neutral calorimetry
Characterization/improving of bump splitting algorithms, pi° reconstruction
Hadronic and em split-off recognition
PreShower detection and correction
Calibration of photon energy response
Parametrization of photon covariance matrix
Detection of charged signals in EMC using tracking information
Detection of charged signals in EMC using SciTil/GEM/MDT information
Criteria for good neutral candidate selections
Validation of hadronic signal in calorimeter with G3 and G4
Validation of electromagnetic signals in forward calorimeter with G3 and G4
Alignment algorithms
Automatic Quality Assurance

Tracking
Validation of Pattern Recognition codes
Secondary Vertex track reconstruction and fitting (Lia)
Low Pt tracking investigation for barrel
Low momentum tracking for Forward Tracker (inside the dipole)
Implementation/interface of new propagation tools different from Geane
Tracking in the GEM region (Lia)
Correlation between central and forward tracker
Algorithms to find t0
Alignment algorithms
Check and correct the covariance matrixes of fitting tools
Automatic Quality Assurance (Lia, Tobias)
Tracking with muon chambers
Tracking with different particle type
Alternative Kalman Filter algorithms
Track cleaning
Association of MC particles to reconstructed tracks (Lia, Tobias)

Analysis
Check of correct combinatory, overlap logic
Validation of fitter results (on idealized 'toy' candidates and on “real” data)
Validation of MC truth mapping, and MC tree match
Estimation of PID quality (loose, tight...) efficiencies/purity
Diagnostics of PID raw values distributions
Validation of complete analysis chain with complicated channels
Automatic Quality Assurance

Particle Identification
Improve detector correlation with tracking
Implement MVD de/dx with different methods (see TDR)
PID for Shashlik
PID using time-of-flight detectors
Automatic Quality Assurance

Event Generators
Comparison between different background generators (DPM, Pythia, Fluka, FTF)

Code Management
Release manager
Memory profiling and leakages finding
Cleaning of data structure

General tasks
Quality Assurance system for all the systems
Tag algorithms for event selection
Development of a framework inside Panda(Fair)Root to handle event times
Include time as an individual parameter into the reco algorithms of the detectors
Development of algorithms to extract the t0 time from the detector data
Propagation of the time in GEANE
Event building
MC Truth matching with FairLinks (Tobias)
Usage of different hardware types (FPGA / GPU) in combination with PandaRoot
Magnetic field handling at different beam momenta
Database interface of all the detectors and table design

Geometry
ROOT to STEP converter
Update of yoke geometry
Passive volumes description in all the detectors

Distributed computing
GRID Software management
GRID Data management
GRID Central administration
Exploration of new distributed systems

- No additional manpower was sent for code developments
- The few tasks in blue attacked by the existing manpower
- People leaving, maybe ending PhD, and scarce new forces

A Common Task Committee has been nominated, but activities were somehow blocked due to our uncertain future

Destructive interference coming from the Heuer review outcome

- Difficulty/impossibility to have contracts for PANDA
- People discouraged thinking about something else
- Difficulty to do long term plans
- Computing TDR activities in stand-by
- INFN froze Italian activities

Italian software involvement in standby

Torino

- Computing coordination – ongoing
- Code and release management – reduced
- MDT developments – stopped (contract almost expired)
- Secondary and GEM tracking – stopped (contract expired)

Pavia

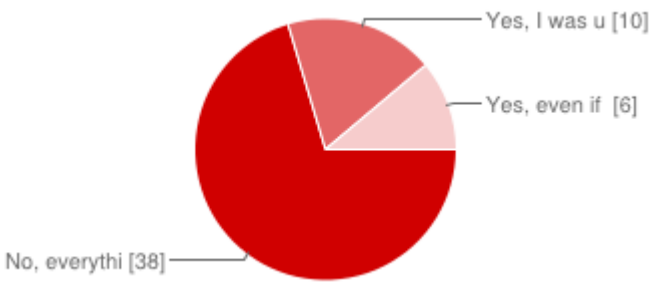
- Tracking coordination - ongoing
- STT developments – stopped
- Central tracking – finished – no further activities planned

Ferrara

- FTS developments – stopped
- Forward tracking studied - stopped

Questionnaire: Installation

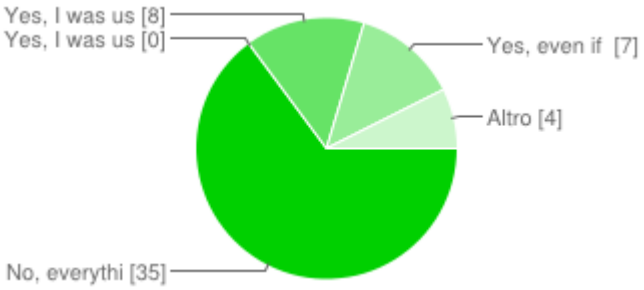
Did you have problems with the installation of external packages?



No, everything smooth	**38**	**70%**
Yes, using a non supported system	10	19%
Yes, even if the system was supported	6	11%

Several people like to try new OS (our beta tester)
Sometimes problems of the cern repository

Did you have problems with the installation of PandaRoot?



No, everything was smooth	**35**	**65%**
Yes, using a non-supported release	0	0%
Yes, a version of the trunk crashing	8	15%
Yes, the version was the suggested one	7	13%
Altro	4	7%

- Mostly trunk versions crashing due to recent updates, fixed quickly in the next hours
- Confusions about the "suggested version" (we have improved it > wiki)
- In general problems are fixed in a short time, thanks to forum or collegues
- Installation is not an issue for software users! ☺

Nevertheless Virtual Machine were provided for releases oct14 and mar15:

- ✓ 1 request for password for oct14
- ✓ 0 requests for password for mar15

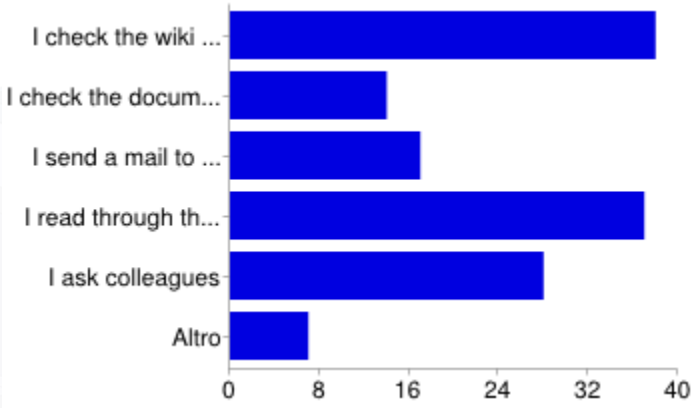
People prefer to install the software by themselves

The idea to organize a new Computing Week, but due to the review…

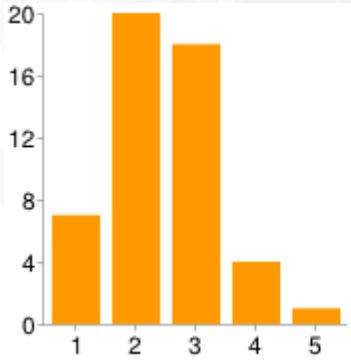
Documentation

How do you find usually information about the software?

I check the wiki documentation	38	70%
I check the the Computing Week INDICO	**14**	**26%**
I send a mail to the Computing Coordinator	17	31%
I read the forum and, if needed, write a message	37	69%
I ask colleagues	28	52%
Altro	7	13%

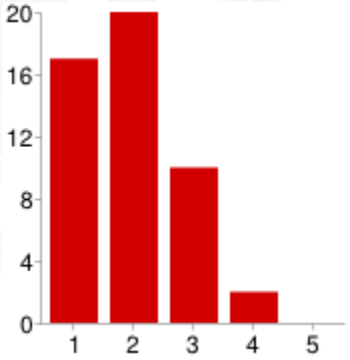


How do you rate the wiki pages?



1 very good
2 good
3 satisfactory
4 bad
5 very bad

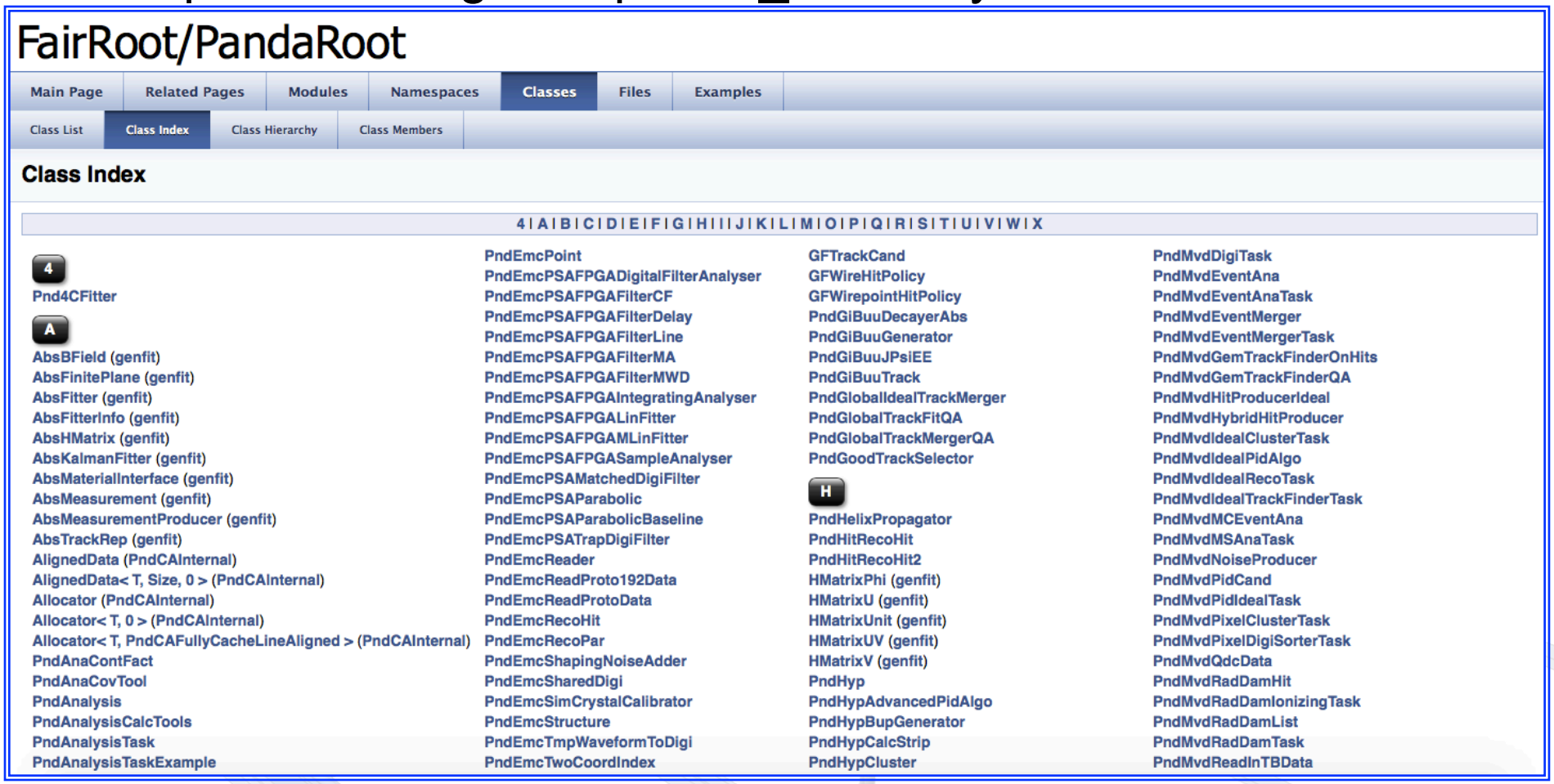
How do you rate the forum?



- People are happy of the forum, answers are received in a short time ☺
- People not so happy with documentation, and wiki covers few aspects
- People noticed (and are happy of) an improvement of wiki documentation ☺

Documentation

http://cbmroot.gsi.de/panda_doc/daily/html/classes.html



FairRoot/PandaRoot

Main Page | Related Pages | Modules | Namespaces | Classes | Files | Examples

Class List | Class Index | Class Hierarchy | Class Members

Class Index

4 | A | B | C | D | E | F | G | H | I | J | K | L | M | O | P | Q | R | S | T | U | V | W | X

4

Pnd4CFitter

A

AbsBField (genfit)
AbsFinitePlane (genfit)
AbsFitter (genfit)
AbsFitterInfo (genfit)
AbsHMatrix (genfit)
AbsKalmanFitter (genfit)
AbsMaterialInterface (genfit)
AbsMeasurement (genfit)
AbsMeasurementProducer (genfit)
AbsTrackRep (genfit)
AlignedData (PndCAInternal)
AlignedData< T, Size, 0 > (PndCAInternal)
Allocator (PndCAInternal)
Allocator< T, 0 > (PndCAInternal)
Allocator< T, PndCAFullyCacheLineAligned > (PndCAInternal)
PndAnaContFact
PndAnaCovTool
PndAnalysis
PndAnalysisCalcTools
PndAnalysisTask
PndAnalysisTaskExample

PndEmcPoint
PndEmcPSAFPGADigitalFilterAnalyser
PndEmcPSAFPGAFilterCF
PndEmcPSAFPGAFilterDelay
PndEmcPSAFPGAFilterLine
PndEmcPSAFPGAFilterMA
PndEmcPSAFPGAFilterMWD
PndEmcPSAFPGAIntegratingAnalyser
PndEmcPSAFPGALinFitter
PndEmcPSAFPGAMLinFitter
PndEmcPSAFPGASampleAnalyser
PndEmcPSAMatchedDigiFilter
PndEmcPSAParabolic
PndEmcPSAParabolicBaseline
PndEmcPSATrapDigiFilter
PndEmcReader
PndEmcReadProto192Data
PndEmcReadProtoData
PndEmcRecoHit
PndEmcRecoPar
PndEmcShapingNoiseAdder
PndEmcSharedDigi
PndEmcSimCrystalCalibrator
PndEmcStructure
PndEmcTmpWaveformToDigi
PndEmcTwoCoordIndex

GFTrackCand
GFWireHitPolicy
GFWirepointHitPolicy
PndGiBuuDecayerAbs
PndGiBuuGenerator
PndGiBuuJPsiEE
PndGiBuuTrack
PndGlobalIdealTrackMerger
PndGlobalTrackFitQA
PndGlobalTrackMergerQA
PndGoodTrackSelector

H

PndHelixPropagator
PndHitRecoHit
PndHitRecoHit2
HMatrixPhi (genfit)
HMatrixU (genfit)
HMatrixUnit (genfit)
HMatrixUV (genfit)
HMatrixV (genfit)
PndHyp
PndHypAdvancedPidAlgo
PndHypBupGenerator
PndHypCalcStrip
PndHypCluster

PndMvdDigiTask
PndMvdEventAna
PndMvdEventAnaTask
PndMvdEventMerger
PndMvdEventMergerTask
PndMvdGemTrackFinderOnHits
PndMvdGemTrackFinderQA
PndMvdHitProducerIdeal
PndMvdHybridHitProducer
PndMvdIdealClusterTask
PndMvdIdealPidAlgo
PndMvdIdealRecoTask
PndMvdIdealTrackFinderTask
PndMvdMCEventAna
PndMvdMSAnaTask
PndMvdNoiseProducer
PndMvdPidCand
PndMvdPidIdealTask
PndMvdPixelClusterTask
PndMvdPixelDigiSorterTask
PndMvdQdcData
PndMvdRadDamHit
PndMvdRadDamIonizingTask
PndMvdRadDamList
PndMvdRadDamTask
PndMvdReadInTBData

Doxygen

It exists since the beginning of PandaRoot
But many people were not aware

Code Coverage

Automatic code coverage checks each night in Dashboard

lxg0156.gsi.de	Debian GNU_Linux 6.0.10 (squeeze)-linux-x86_64-gcc4.4.5-fairsoft_mar15-fairroot_v-15.03	9	0	0	0	500 -1	0	3 +1	66 +1	Sep 04, 2015 - 03:01 CEST
briareos	Ubuntu14.10-linux-x86_64-gcc4.9.1-fairsoft_mar15-fairroot_v-15.03	9	0	0	0	500 +389 -365	0	3 -1	66 +3	Sep 04, 2015 - 02:01 CEST
host236-pool27	Darwin_14.4.0-darwin-x86_64-gcc4.2.1-fairsoft_mar15p1-fairroot_v-15.03	9	0	0	0	12	0	3	66	Sep 04, 2015 - 10:48 CEST

Continuous

Site	Build Name	Update	Configure		Build		Test			
		Files	Error	Warn	Error	Warn	Not Run	Fail	Pass	Build Time
lxbuild04.gsi.de	Wheezy64-GNU_Linux-x86_64-_usr_bin_g++4.7-fairsoft_mar15-fairroot_v-15.03	14	0	0	0	0				Sep 04, 2015 - 15:15 CEST
lxbuild04.gsi.de	Wheezy64-GNU_Linux-x86_64-_usr_bin_g++4.7-fairsoft_mar15-fairroot_v-15.03	7	0	0	0	0				Sep 04, 2015 - 14:15 CEST

Coverage

Site	Build Name	Percentage	LOC Tested	LOC Untested	Date
host236-pool27	Darwin_14.4.0-darwin-x86_64-gcc4.2.1-fairsoft_mar15p1-fairroot_v-15.03	35.54%	36910	66942	Sep 04, 2015 - 10:48 CEST
lxsub13.gsi.de	Debian-linux-x86_64-gcc4.4.5-fairsoft_mar15-fairroot_v-15.03	33.67%	33430	65846	Sep 04, 2015 - 02:22 CEST
briareos	Ubuntu14.10-linux-x86_64-gcc4.9.1-fairsoft_mar15-fairroot_v-15.03	36.63%	37521	64898	Sep 04, 2015 - 02:01 CEST
lxbuild04.gsi.de	Wheezy64-linux-x86_64-_usr_bin_g++4.7-fairsoft_mar15-fairroot_v-15.03	31.53%	32192	69894	Sep 04, 2015 - 00:20 CEST

Kitware

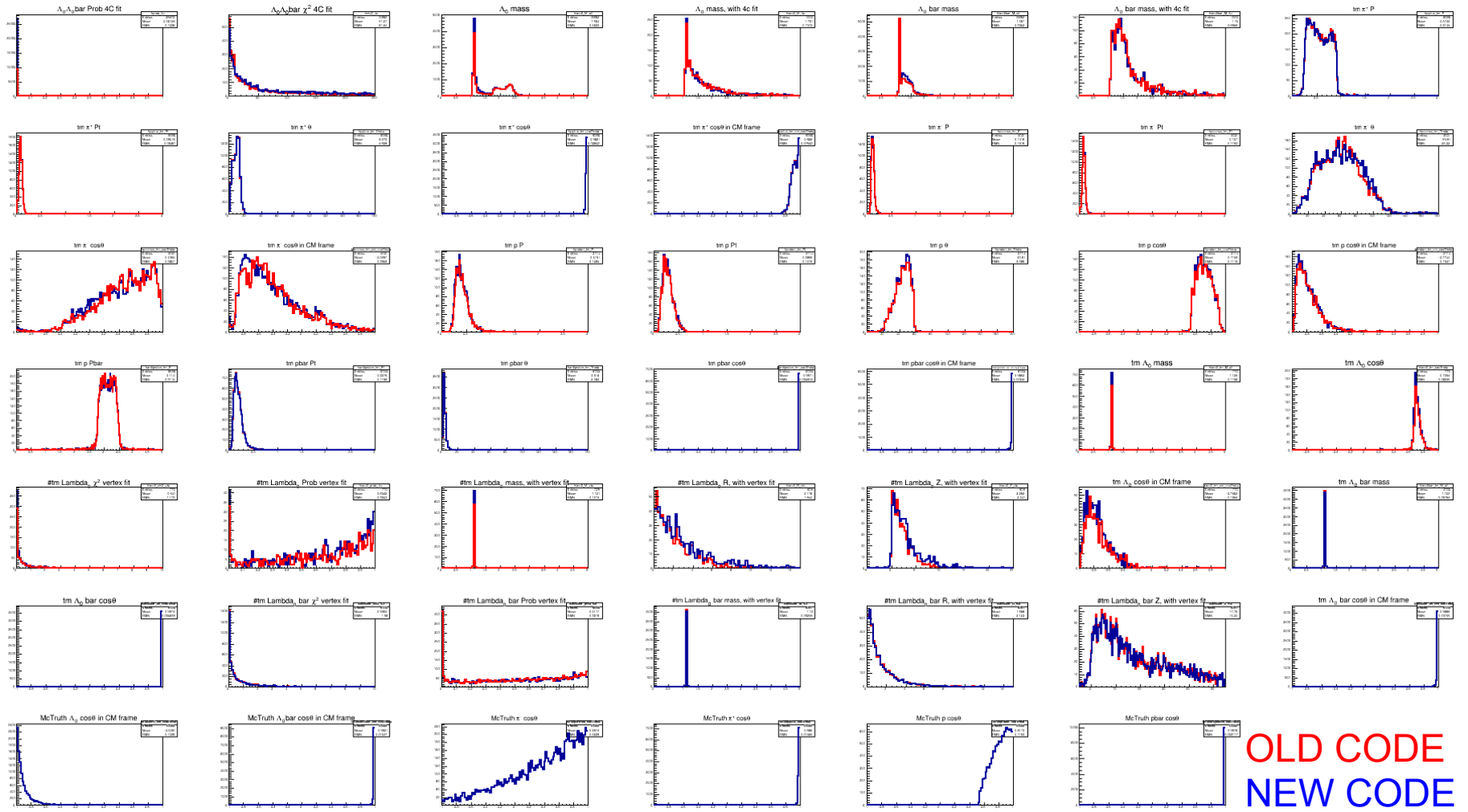
We started to add more QA tests on the dashboard

Now 36% coverage (we started with < 20%)

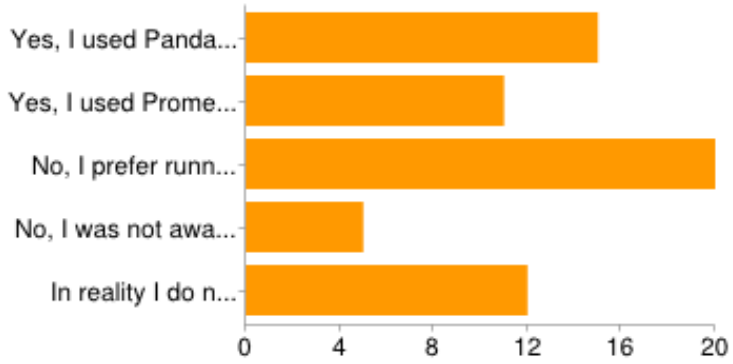
Quality Assurance - Analysis

Set of physics variables, compared to a sample file

Invariant mass, momentum, theta, fitter χ^2, $\cos\theta_{CM}$, MC truth, different pid selection



Have you ever used the production systems for massive simulations?



Yes, PandaGrid	15	28%
Yes, Prometheus at GSI	11	20%
No, I run in my local farm	**20**	**37%**
No, I was not aware	5	9%
I do not need so many events	12	22%

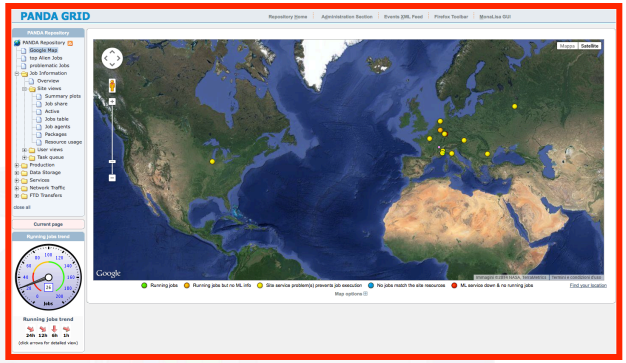
Coordinated data production

- ✓ contact
  - Production Manager Paul Bühler - paul.buehler@oeaw.ac.at
  - Computing Coordinator Stefano Spataro - spataro@to.infn.it.
- ✓ together define procedure and perform production of data set
- ✓ list of available data sets is updated on

https://panda-wiki.gsi.de/foswiki/bin/view/Computing/Productions

Data Production Systems

PandaGrid



- ✓ Around 1200 cores
- ✓ >100 TB disk space
- ✓ Empty most of the times

Prometheus farm @ GSI (Kronos becoming available)

- ✓ 10k cores for all the GSI/FAIR experiments (max queue 2000)
- ✓ 160 TB disk space
- ✓ Always busy and disk full

New FAIR-Russia Research Center @ ITEP (Moscow)

- ✓ 10000 cores
- ✓ 1PB disk space
- ✓ Almost empty (only theory runs there)

Probable
Tier1 Center

Paul tested the performances, we can start to use it

Speaking about Computing TDR

The Original plan

- End 2015 – FAIR Computing TDR (with Panda requirements)
- End 2016 – Panda Computing TDR

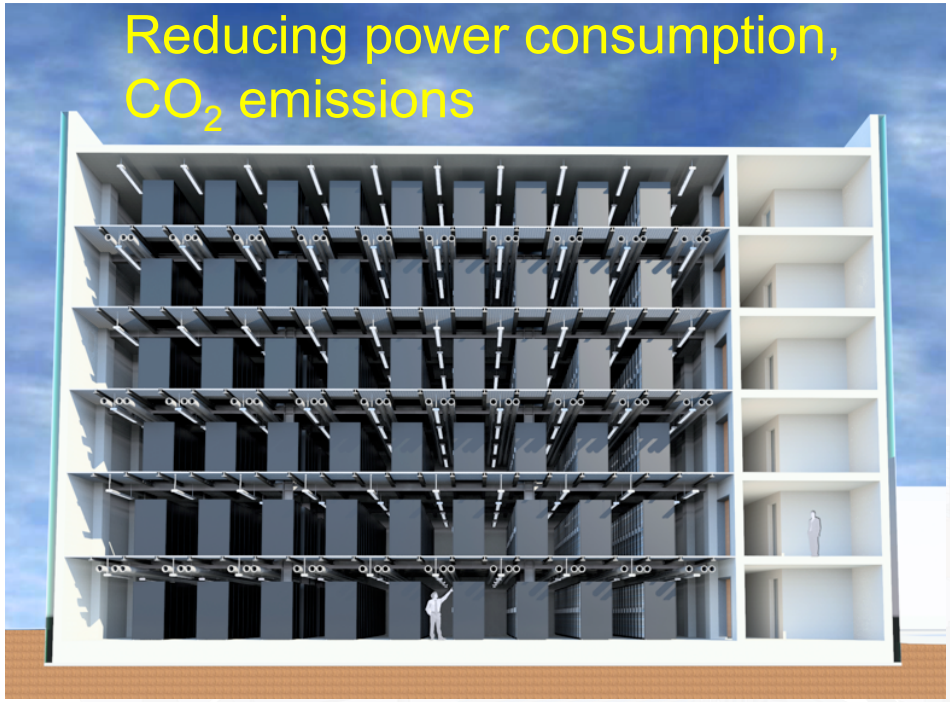
- ✓ Started requirements calculations (see my talk Dec 2014)
- ✓ Started design of distributed models (see Thorsten talk mar 2015)
- ✓ Started search for candidates Tier-1 (negotiation with INFN for CNAF)
- ✓ PANDA strongly involved in the FAIR Computing TDR activities

Heuer review froze all the computing activities

Uncertainty on day-1 date (202?)
Lost CNAF as possible Tier-1
INFN sites in PandaGrid?

- Monalisa server @ Torino
- Replica of DB @ Torino
- 200 cores, 40 TB @ Torino
- 100 cores,10 TB @ CNAF lost

The FAIR Tier-0 Green Cube



- 6 floors (starting with 2)
- 128 racks each floor (8 rows with 16 racks)
- Each rack can provide:
  - Data: 0.6 PB
  - Cores: 1800

- Construction work started in fall 2014
- Building finished summer 2015
- Tests and migration
- Starting normal operations end of 2015

Successfull prototype: “Prometheus” (mini-cube)

Computing Requirements

In dec 2014 first estimation was present, with ideas about a distribution model

FULL LUMINOSITY

1MHS06, 58 PB disks, 69 PB/year tape

(let's forget about continuous data stream)

(if 1 core = 10 HS06 → 100k core)

Afterwards, compression of MC truth data:
45 PB disks, 34PB/year tape

What about running at low luminosity for several years?

LOW LUMINOSITY

Ideally, let's scale resources of a factor 10

@ 2 MHz performances close to "event based" simulation
(maybe we don't need time-based simulation?)

100kHS06, 5 PB disks, 4 PB/year tape

GreenCube (Tier0) + ITEP (Tier1) + PandaGrid could be enough

Summary

Target Spectrometer: advanced status (different level of details)

Forward Spectrometer: a lot of things to do (but RICH!)

Several SG criticalities were attacked (even before SG)

- ✓ Documentation (wiki + doxygen) - improved
- ✓ Code quality (automatic code coverage and QA)
- ✓ Virtualization, but installation is no problem for users
- ✓ Improved time-based-simulation -> TS
- ✓ Centralized data production – stressed again

Computing TDR

- ➢ Make no sense too much early – technology dependent
- ➢ With low luminosity, computing not so demanding

Resources and manpower are the real problem!!