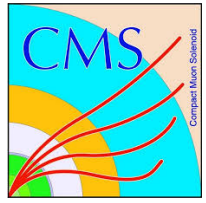
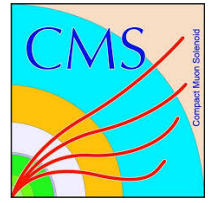


CMS Experiment Training and Career Program

Sudhir Malik (University of Puerto Rico Mayaguez)

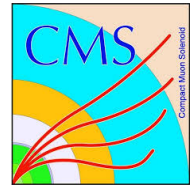
Joint WLCG & HSF Workshop 2018, 26-29 March, Napoli, Italy

CMS Training Paradigm



- **Huge collaboration, e**normous resources, manpower, long life span of the experiment, enormous data rate
- **Most users not resident at CERN**, possible financial and logistic constraints to be at CERN
- Highly **distributed environment** for - Computing (Grid), Physics analysis
- Support and training reach every user wherever they may be, be taken up in **organized and central** way
- CMS User Support is organized under two **Level-2 coordinators under CMS Computing**
- Training Model
  - **of the Users, by the Users, for the Users**, almost no dedicated personnel, clear manifestation of a big collaborative spirit, expert in different physics analysis tools help other users, users become expert and provide feedback, get recruited for further help and the cycle goes
  - **Engages collaboration** in meeting the pre−requisites to perform physics analyses, Distribute the expertise besides CERN to other centers/institutes
- Training program geared towards **ultimate goal of Physics Analysis**
- **CMS Training Program is robust** despite challenges and there is a lot of support for training activities, including **service points for documentation work**, rarely an issue to find people to train (LHC Physics Center at Fermilab, Terascale (DESY), CERN etc.), **Facilitator in training activities is considered an additional bonus for your CV** when looking for faculty jobs in terms of having gained teaching experience.

CMS Training Program (Evolution)



- **Training is of two types** - (1) **self-training** any time (WorkBook, twikis and indico talks) and (2) **hands-on training**

- **Started in 2006 (long before data taking)**
  - 2-day program - talks on CMS Physics, sub-detectors, physics objects, hands on with simulated data, computing tools, grid jobs, **CMS StarterKit** etc
  - evolved later to cosmic data to collision data (before 2010)

- In **2009** started 5-day long hands-on analysis training program, called **Physics Analysis Toolkit (PAT) tutorials** based on common analysis data format used by CMS and learn configuring analysis in python, 10 PAT tutorial so far (once or twice a year), **trained 600 students**.

- Since **2010** (after data taking commenced) started 5-day long hands-on **CMS Data Analysis School (CMSDAS)** which follow the model of PAT tutorials where student perform full fledge physics analysis, 18 CMSDAS so far (once or twice a year), **trained >1000 students**.

- In **2017** also started 5-day long hands-on **CMS Physics Object School (CMSPOS)** to best use of Physics Objects, performance to reconstruct and identify them, performance of sub-detectors with a goal towards detector upgrade and future physics

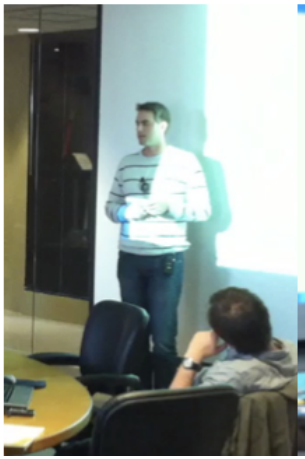
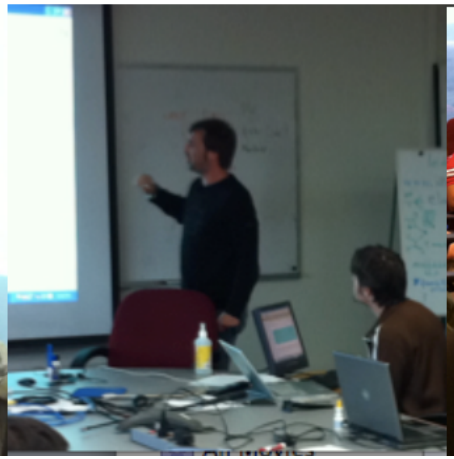
- All schools are supervised by **CMS Schools Committee (senior people across the CMS world)**

- 1-day **Hands-On Advance Tutorial Sessions (HATS)** on advance topics - several of them held each year

- **Advanced Physics Workshops** bring together experts and theorists beyond CMS to assess the state-of-the-art of current LHC physics, future physics topics and shape the future of the field

CMS Physics Analysis Toolkit

PROGRAM	GOAL	TIMETABLE	TOPICS	MATERIAL
Pre-School Exercises	Beginners to Experienced, jump start Physics Analysis	The preparatory Exercises start a month before the school, Exercises prepared and checked before that by a team	**CMS Basics - software, access to code data, run Grid jobs, Github, ROOT, Python, PyRoot, Fitting**	twiki, espace
5-Day Hands on sessions, Students work in class-like settings with student/teacher ration of ~10:1 **Held mostly at CERN and Fermilab** **Survey and Feedback from users**	Learn Python and Physics Analysis Toolkit (PAT), common analysis data formats and common analysis algorithms, approved algorithms and sensible defaults to start analysis 600 students trained so far	Each day devoted to one or two topic	**Document Navigation** **PATTuple Creation** **PATConfiguration Exercise** **Access PAT candidates** **Cross Object Collection** **PAT to Particle 2PAT** **Monte Carlo Matching** **Using Trigger** **Full Analysis Example**	hands-on, twikis Indico (talks)


CMS Data Analysis Schools

PROGRAM	GOAL	TIMETABLE	TOPICS	MATERIAL
Pre-School Exercises	Beginners to Experienced who want to jump start Physics Analysis	The preparatory Exercises start a month before the school, Exercises prepared and checked before that by a team of facilitators	**CMS Basics - software, access to code data, run Grid jobs, Github, ROOT, Python, PyRoot, Fitting**	twikis, espace to answer questions
5-Day Hands on sessions, Students work in class-like settings with student/teacher ration of ~5:1 **Held at several places typically 2-3 times per year - Fermilab, Taiwan, CERN, Italy, DESY, India, Korea** **Survey and Feedback from users**	Common interface to the algorithms developed by physics objects groups, single entry point to information associated to physics objects, Approved algorithms and sensible default, configure ones analysis in python language **1000 students trained so far**	Lectures - 1/2 day	**LHC Machine, CMS Physics, CMS Detector, CMS Software Tools, Physics Analysis Design**	Indico (talks)
		Short Exercises - 2 days Each Exercise - 2hrs, Can take up to 6 exercises	**Roostats, Generators, Tracking, Vertexing, Electrons, Muons, Jets, b-tagging, PFlow, Pileup, Event**	hands-on, twikis Indico (talks)
		Long Exercises - 2.5 days Physics Analysis: 6-8 students per analysis	**Examples - Dark matter (with Higgs boson to four-leptons), Mono-Photons, B2G Boosted Z'->ttbar semileptonic, SUSY hadronic, Z to tau-tau, Top mass measurement etc**	hands-on, twikis Indico (talks),
		Mini-symposium	The student groups present their work and compete for the “first prize” judged by panel of senior CMS physicists	Indico talks

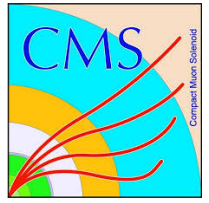



CMS Object Schools

PROGRAM	GOAL	TIMETABLE	TOPICS	MATERIAL
5-Day Hands on sessions, Students work in class-like settings with student/teacher ration of ~5:1 **Held in Italy, next at RWTH Aachen** **Survey and Feedback from users**	Contribute to performance CMS subdetectors and the reconstruction and identification of Physics Objects **50 students trained so far**	The preparatory Exercises start a month before the school, Exercises prepared and checked before that by a team of facilitators	**CMS Basics - software, access to code data, run Grid jobs, Github, ROOT, Python, PyRoot, Fitting**	twikis, espace to answer questions
		Lectures - 1/2 day	**LHC Machine, CMS Physics, CMS Detector, CMS Software Tools, CMS Upgrade, Datasets and Performance**	Indico (talks)
		Short Exercises - 2 days Each Exercise - 2hrs, Can take up to 6 exercises	**RPC/DT, GEM, Tracker, ECAL, HCAL, Physics Performance and Datasets**	hands-on, twins Indico (talks)
		Long Exercises - 2.5 days Physics Analysis: 6-8 students per analysis	**Muon, Electron/Photon, Tau, Tracking, b-tagging, Demonstration of Complete Physics Analysis, signal and control region and statistical analysis**	hands-on, twins Indico (talks)
		Mini-symposium	The student groups present their work and compete for the “first prize” judged by panel of senior CMS physicists	Indico talks




CMS - Advance Tutorials (1-day) (complements CMSDAS)



- CMS Connect HATS
- Framework 101
- Big Data with Spark
- Statistics for Particle Physicists Course
- PyRoot
- B-tagging
- Tracking and Vertexing
- Multivariate Analysis Techniques
- Visualization
- Generators
- CRAB3
- Trigger
- Electron and Photon
- Jets Energy Corrections and Pile-Up Mitigation
- Tau HATS@LPC (Jun 16, 2017)
- MET HATS@LPC (Jun 15, 2017)
- Jets Algorithms and Substructure
- Muon
- Git/GitHub
- HCAL HATS
- Machine Learning
- HEP analysis in the Python ecosystem
- Histogrammar
- PyRoot and rootpy
- Delphes simulation

- **1/2-day or 1-day typically**
- **hands-on**
- **several topics**
- **complements CMSDAS**
- **Vidyo for remote participation**
- **indico talks, twikis**

CMS Career Guidance (non-academia)

- CMS Collaboration Board endorsed the composition of the **CMS Career Committee** (members from all geographic regions) in 2012 - four working groups:
  - **Networking** with CMS alumni and to the non-academic job market in general,
  - **Collecting and providing information** on academic jobs
  - Reflecting on the **recognition of individual achievements**
  - Organizing **information sessions** on career related topics

- Our skills, especially computing, are much sought after in industry
- Committee organizes career events **bringing CMS alumni from companies** in a diverse range of fields (industry, finance, IT) at CERN, Fermilab etc.
- Lately this idea was recognized liked by other LHC experiments and since then we have been **organizing career events with ALICE, ATLAS, CMS and LHCb**

  - https://indico.cern.ch/event/561880
  - https://indico.cern.ch/event/440616/

- We also maintain a **CMS job twiki** - academic and industry jobs and guidance, highly popular, jobs advertised free of cost, only for HEP community - https://twiki.cern.ch/twiki/bin/view/CMSPublic/JobOpportunities



ATLAS Software Training Program

Davide Costanzo (The University of Sheffield)
Louise Anne Heelan (Carleton University)

ATLAS Software Training Program

NAME	AUDIENCE	PROGRAM	TOPICS	MATERIAL
Software Tutorial	Beginner (+Refresher)	5 days @CERN 4x/year (since 2012 900 people trained)	ATLAS-specific software (formats, frameworks, analysis model), the Grid, git, cmake, statistical tools	Indico (talks) twiki git-style docs recorded videos
Migration Tutorials	Experienced	1-2 days (as needed)	Examples: svn→git, cmt → cmake, Run 1→Run 2 analysis model	Indico (talks) twiki git-style docs recorded videos
Developer Tutorial	Intermediate/ Advanced	5-days @CERN ~1/year	code quality, writing code in ATLAS, multi-threading	Indico (talks)
Workbooks	Beginner	updated by dedicated responsible person as needed	Computing Physics Analysis Software Developers	twiki

Training challenges

- manpower: difficult to get already overworked software experts to create, maintain and deliver training materials
  - software tutorials: 1 coordinator + ~5 consistent contributors/helpers + 20 other expert speakers based on availability
  - developers tutorials: given by experts, usually organized by software coordinators
  - workbooks: typically single responsible person to maintain
  - has been interest in developing other training sources (ex. a StarterKit like LHCb or frequent one-hour topical tutorials) but these have not happened yet
- sustainability of material
  - software tutorials: frequent, so material continuously updated
  - can be difficult to find the latest and best reference for documentation

Training challenges

- material format
  - most often material contained in talks and twikis
  - twiki: allows anyone to easily update (pro and con), poor search engine, aesthetically unpleasant, too many twiki pages
  - exploring git-style documentation: slight learning curve, slower to update, google searchable, nice looking, developers can review before posting live

- lacking intermediate-level training
  - ATLAS sub-detector systems describing reconstruction/calibration chain, monitoring tools, databases, etc.
  - ROOT and other statistical packages