



# CMS Offline Software

Liz Sexton-Kennedy  
Jterm III  
January 16, 2008 – Fermilab



# Introduction



The goal of this talk is not to teach you how to use the software. That was what yesterday's tutorials were for. I will speak about the offline software as a project.

How is the software project broken down into subprojects and who are the current leaders.

I'll review the major accomplishments of this past year, and review the status of some subprojects.

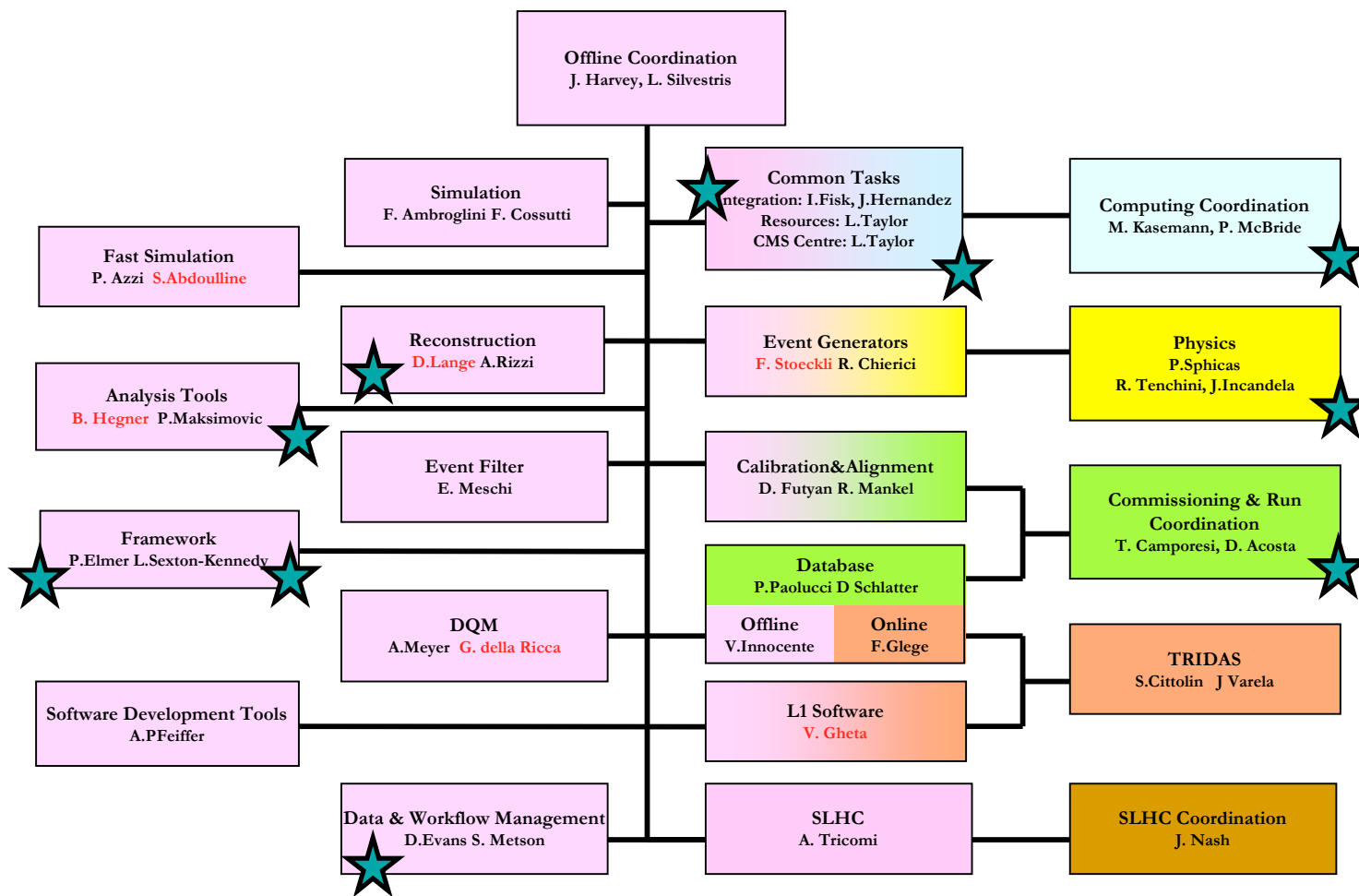
What needs to be done in 2009 including Preparation for Data Taking

Work on the offline project counts as service work, so come talk to us about getting involved.

## Summary



# Organization



## Subprojects:

- Full and Fast Simulation, Generators, SLHC
- L1 Software, EventtFilter, Calibration&Alignment, Reconstruction, DQM
- Framework, Databases, SW Dev. Tools, WMDM, Computing Integration



# This Year in Review



**CMSSW\_2\_0\_X** – CRUZET and CSA08 release. Major re-organization and migration of the DQM framework, important tracking improvements (fake rate reduction, tracking to lower pT) and much of PAT 1\_6\_12 also in 2\_0\_X.

**CMSSW\_2\_1\_X** – First data taking release! Include changes resulting from Calorimeter Simulation TF studies, new field map (3.8T), and further optimizations of memory, cpu and event size. X went to 12 this time. Caution between X=9 and X=10 LI simulation software became backward incompatible which is why there is a special 2\_1\_17 for MC production this past fall.

**CMSSW\_3\_0\_X** - Still in pre-release series; 6 is the latest. DPG & POG content frozen before CERN Dec. shutdown. The remaining pre-releases are for bug fixes from physics, and upgrades of our highly coupled external dependencies like root and geant4.

**CMSSW\_3\_1\_X** - Born from the accelerator schedule delay. Gives the physics groups more time to make improvements before data taking starts. Deadline for contributions is the end of Feb. L2s are now collecting requests and making plans.



# Accomplishments of Last Year



CSA08 has overall been a great success

Aim of CSA08: Test the full scope of offline data handling and analysis activities needed for LHC data-taking operations in 2008.

Focus is on DPG activities, with calibration and alignment forming the central component. In contrast to CSA07, perform alignment and calibration in "real-time", as needed for real data-taking mode

- ➔ Resulting constants used to re-reconstruct the data ready for physics analysis

The major objectives were achieved within the pre-defined 4 week schedule:

- ➔ Full quota of 150M events produced and reconstructed
- ➔ Calibration and alignment exercises completed on time by all sub-detectors
- ➔ Re-reconstruction at TIs using new constants completed on time
- ➔ Physics analyses successfully carried out using both prompt and re-reconstructed data

A number of problems and glitches were encountered but were quickly overcome without serious impact to the schedule. It was a great opportunity to perform a full scale dry run of offline calibration and alignment.



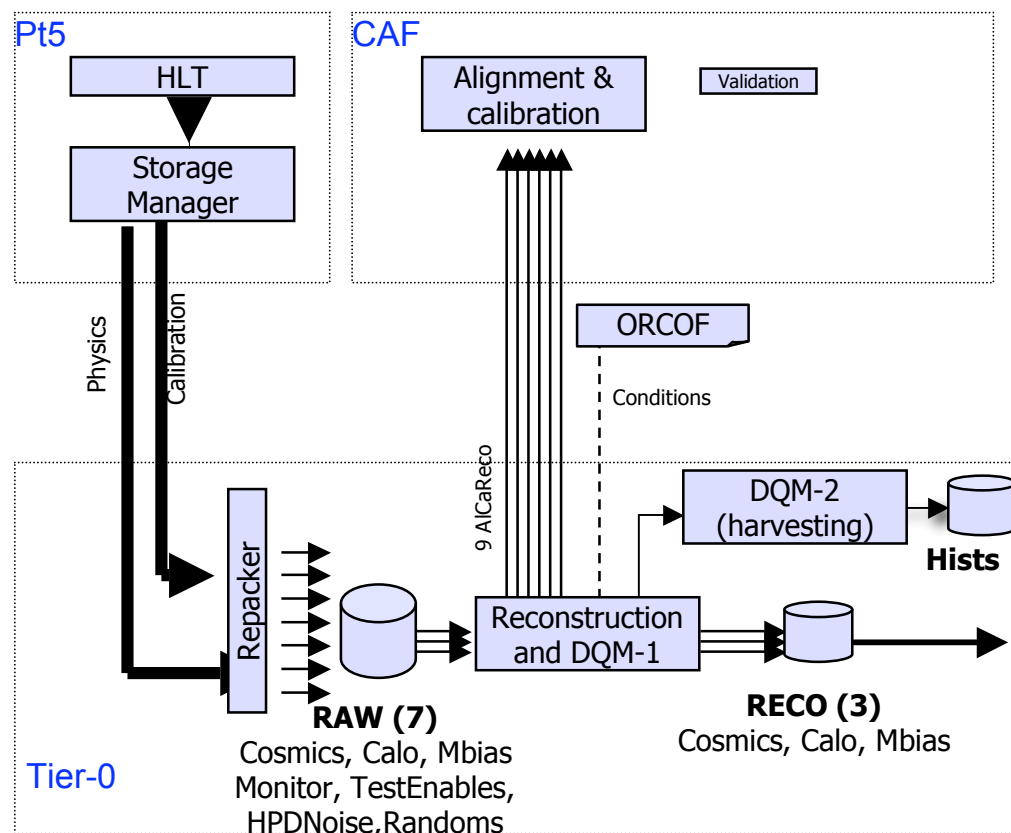
# Accomplishments of Last Year



## Successful Fall '08 data taking

The following steps were run automatically under control of the Tier-0 production system

- ➔ Repacker reassembled online streams into 3 PDS (Cosmics, Calo, Mbias)
- ➔ Prompt Reco jobs scheduled as soon as RAW files became available
- ➔ Offline DQM ran as part of reconstruction and histograms saved together with RECO data
- ➔ Latency of reconstruction was  $> \sim 6$  hours (time to process  $\sim 30$  k events)
- ➔ RECO data sent to Tier-1s for skimming
- ➔ RAW data sent to Tier-1s for future re-processing





- ➔ Cosmics/Calo datasets were skimmed to produce 9 AICaReco datasets
  - using HLT bits for the first time
- ➔ Latency between data-taking and arrival at CAF was  $1/2$ -1 day (DBS)
  - Including merging, transfer to CAF, DBS registration
- ➔ Not fully automatic - required close tracking and follow-up by Data Operations

### Alignment validation started after arrival of AICaReco at CAF

- ➔ Turnaround of algorithms in 6-16 hours
- ➔ First alignment results appeared within 2-3 days
- ➔ Constants validation sign-off procedures start to be exercised
  - Produced histograms - needs to be integrated in DQM



# Successful Data Taking



DQM-1 : hists were generated during Reco/AICaReco steps and saved together with data

DQM-2 (harvesting) : histograms are summed for each run

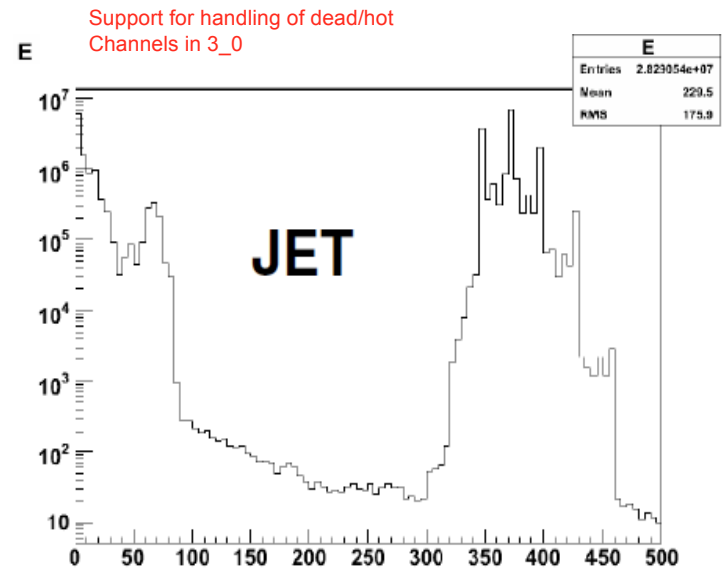
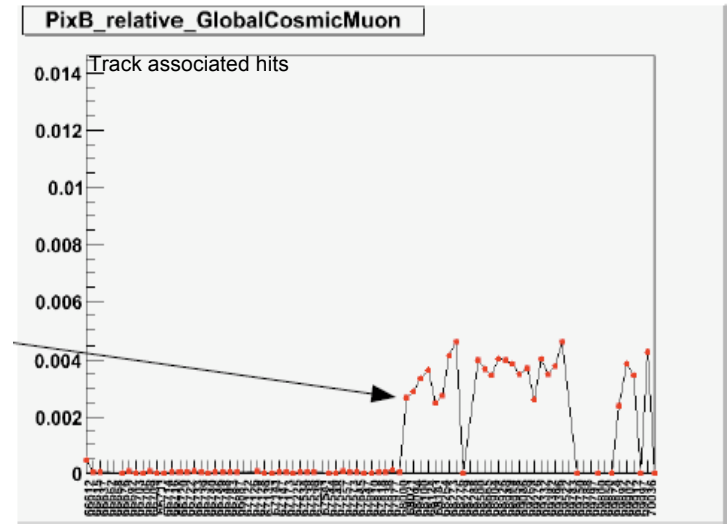
- ➔ Produce 1 ROOT file/run
- ➔ Step run every 30'
- ➔ Viewable via a web-based GUI

Helpful in revealing problems with detector, reco algorithms, conditions data ...

CRAFT used as testbench for data certification

- ➔ quality flags placed in DBS
- ➔ needs to be integrated in harvesting step and run automatically as part of Tier-0 workflow

Change in  
Algh. Pos. Error







### Generators:

- ➔ Use of matrix element based generators in production
- ➔ Formation of the generators validation group lead by Stephen Mrenna

### Simulations:

- ➔ Large FastSim samples can now be made in production
- ➔ Pileup samples can now be generated in production work flows.
- ➔ Developed software for data mixing, important for SLHC upgrade work

### Reconstruction, Calibration & Alignment:

- ➔ Deployed the use of global tag for conditions
- ➔ Made significant reduction in RECO/AOD event content and size
- ➔ Improved reconstruction using less resources (cpu, memory)
- ➔ Unification of standard tracking with iterative tracking, duplication removed
- ➔ Reduced the # of fake tracks from 10% to 1%
- ➔ Reduction of number of jet collections
- ➔ New model for ecal cluster/rechit extended information
- ➔ First implementation of e/gamma ID



## Highlights from the Sub-Groups



### Event Filter and DQM:

- ➔ See successful data taking slides

### Framework:

- ➔ Implemented reading input from two different tiers for the same set of events. This is now deployed at all levels, it was used in CSA08 and is available to CRAB users.
- ➔ Reduced merging restrictions. Reduced FW data size overhead in 2\_2\_X but still more to come in 3\_0\_X.

### Software Development Tools:

- ➔ Scram v2 was a major improvement in performance, many other developer tools added this year that make doing software work much more pleasant, and the integration of releases easier. This is also the group charged with distribution and porting to other platforms.

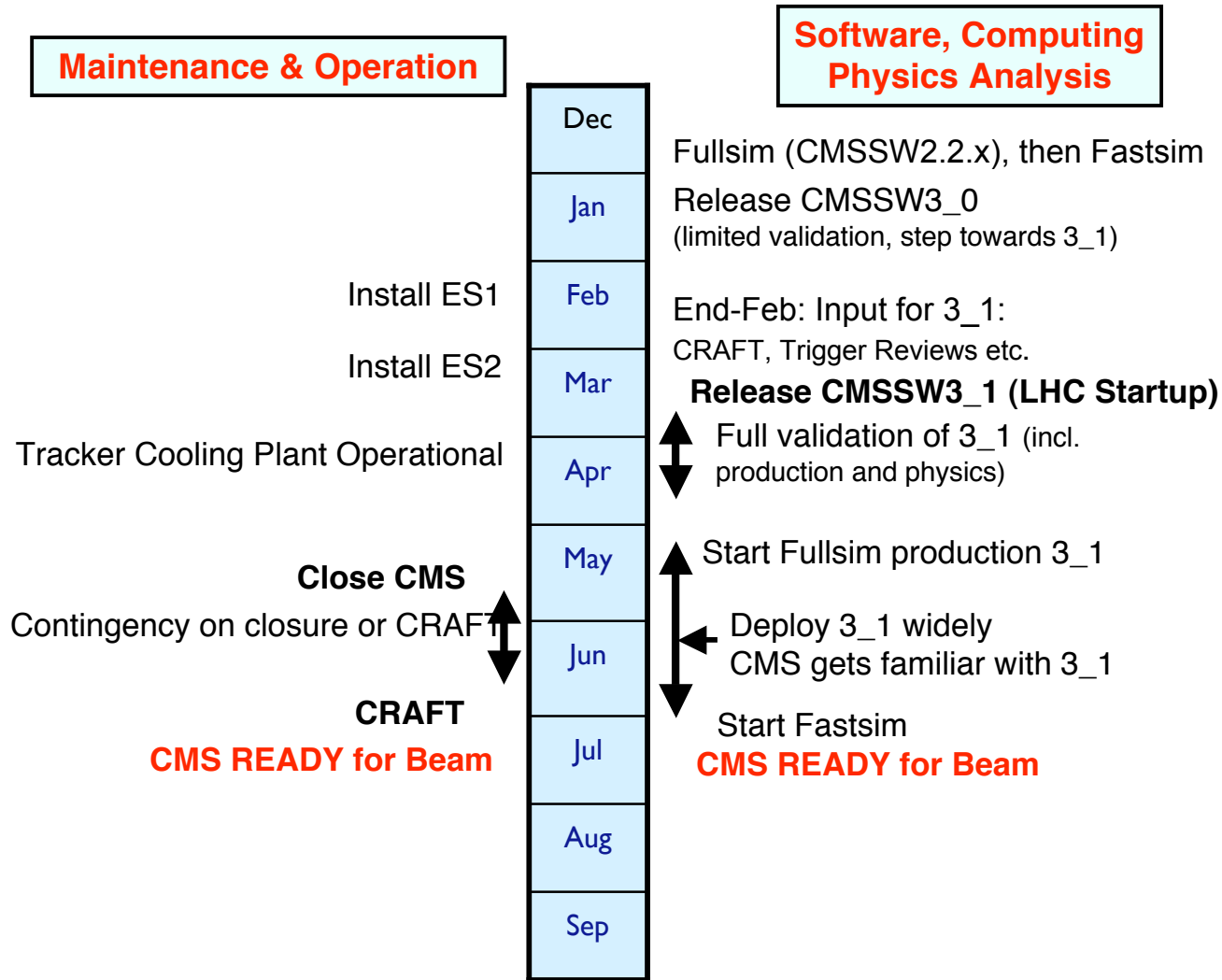
### Data and Workflow Management: See slides on data taking

- ➔ This is the group that has the largest shortfall in manpower...

### Analysis Tools: See Sal's talk



# What Needs to be Done in 2009



Main focus is now on 3\_I\_X. This is the time to fix problems observed in the fall.



## Summary & Concluding Remarks



2008 was a successful year for the offline software project.

There are still many parts of the offline which need more help. If you are a students or post-docs that has some software skills (C++ or python) and still need to fulfill your service requirement, please come see us at the LPC!

Efforts within the software project that the LPC is strong in:

- ➔ Generators
- ➔ Geant Simulation
- ➔ Tracking, Jet/Met, and Tau ID
- ➔ Data operations and WMDM development
- ➔ Framework and SWDT