CMS Status Report

Maria Girone, CERN
David Lange, LLNL

Outline

- Progress on software releases
- Resource usage
- Run II preparation and Validation of the changes
  - Data Management
  - Data Access
  - Production and Distributed Analysis tools
- Scheduling Work
  - Input to Computing from Physics
- Outlook

Scheduling Work and Releases

- The CMS software teams had a long program of work for LS1
  - Multi-core transition, speed improvements, and performance improvements for more complex events
- The current schedule is for the initial 2015 release (CMSSW_7_4) for data taking to be released for production workflows at the end of March
  - The development and validation cycle begins in January
- This constrains the time computing has to produce the 1B events needed for the beginning of the Run
  - There is not a lot of contingency for unforeseen problems

Progress on CMSSW Development Releases

- CMSSW_7_1_0 now validated for MC simulation (generator+detector simulation components)
  - Integration of latest Pythia8 and other generators complete
  - Validation of latest generator versions is on-going including Run I data comparisons

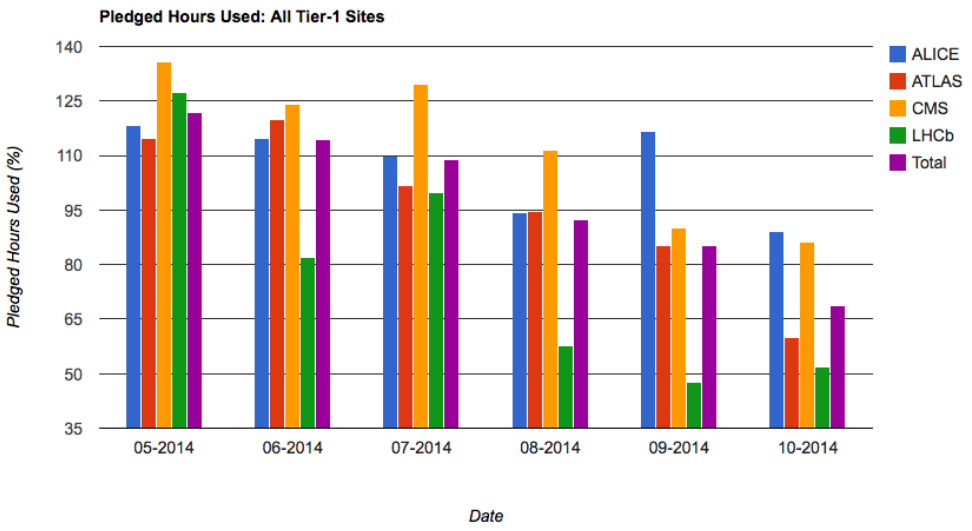
- CMSSW_7_2_0 released in October. Primary goals have been achieved:
  1. "PHYS14" exercise reconstruction:
     - Reprocessing of our CSA14 exercise simulation with the latest digitization and reconstruction (~300M events)
     - This release brings us a long ways towards the final reconstruction for 25ns bunch spacing
  2. Data taking @Point5: Both the online and Prompt Reconstruction used this release for the recent Magnet test.

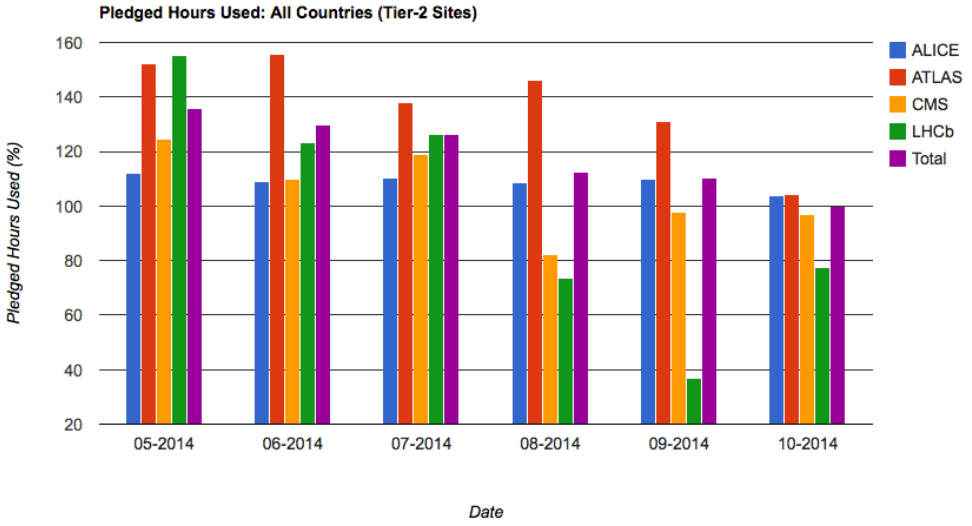
MiniAOD: New Data Format for Analysis

- Goal: Extremely small analysis data format (10% our Run 1 AOD) targeting most analysis work
- Added benefit of the miniAOD format: We have the capability to reproduce from AOD including high-level reconstruction improvements
  - Allows us to better cope with our limited ability to reprocess data in 2015 running
- Given experience in CSA14, widespread adoption is likely

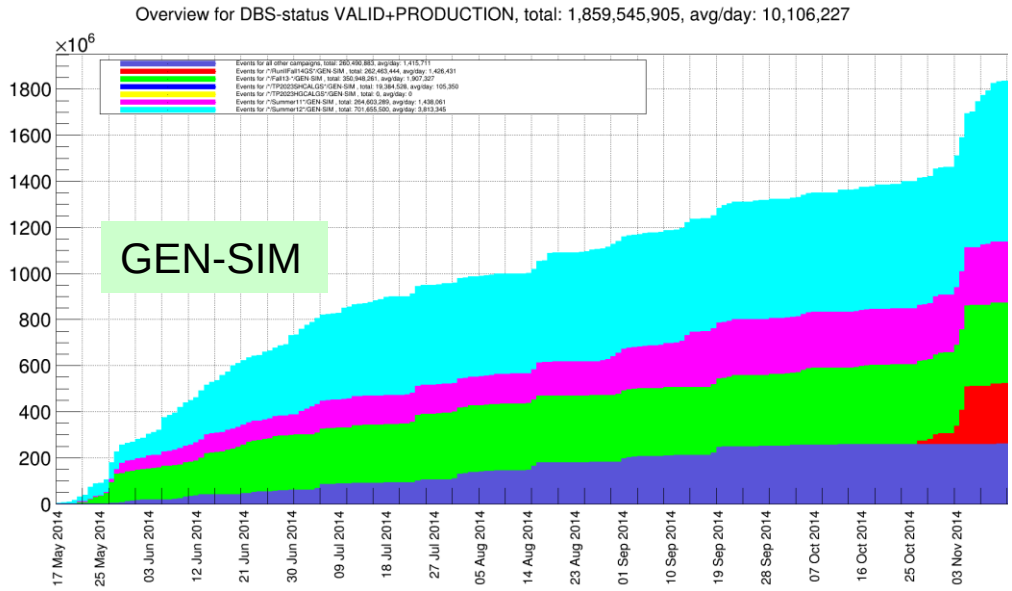
Resource Usage

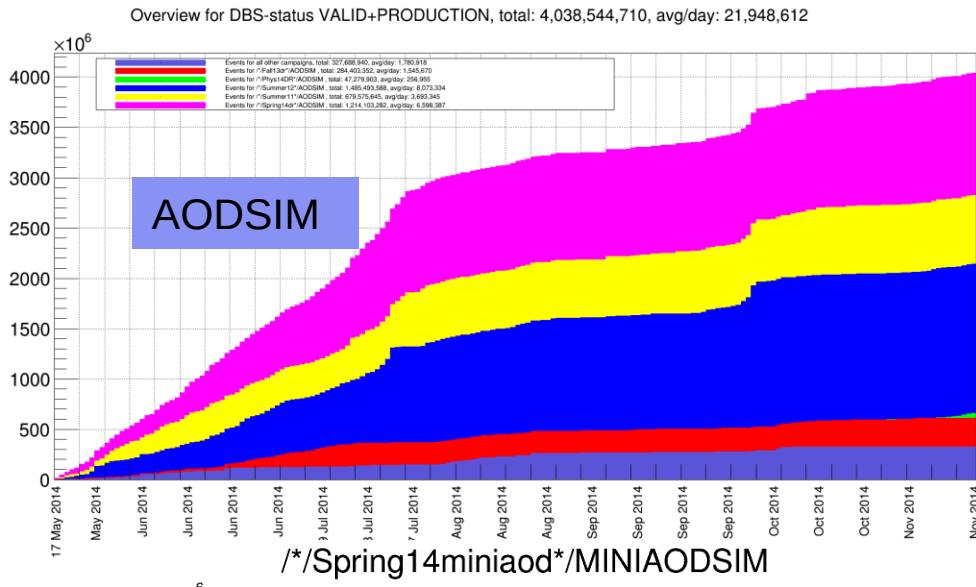
- The 13 TeV MC Reco for CSA14 ramped down at the end of August
  - Processing of MC for Run1 analysis continued together with the preparation of upgrade samples
- Tier-2s have fluctuated around 100%
  - Continued analysis
  - Generator simulation
  - Central mini-AOD production
- Production campaigns just starting
  - PHYS14: ~300M reconstructed events to target specific “early” analyses from the high priority set exploring the discovery potential with the first fb-1
  - 1Billion simulated events at 13 TeV
  - TP Upgrade samples
- Our planning for Tier-2s in 2015 is based on the ramping down of Run I analysis
  - There are many analysis ongoing, but there are not new simulation samples expected and the bulk of the computationally intensive part should be finished



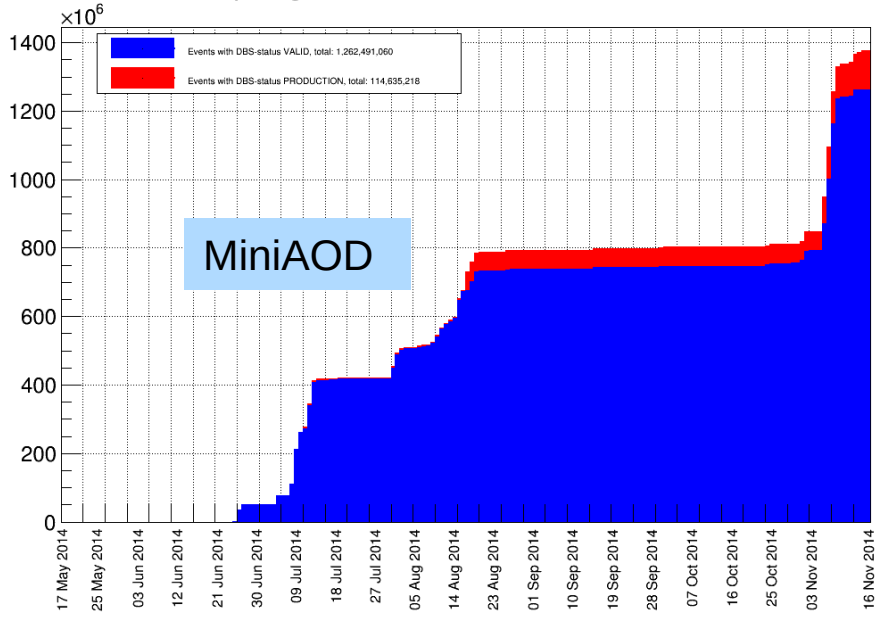


Summary of Ongoing Production Campaigns





- Simulation and Reconstruction has been steady for all of 2014
  - CSA14 and samples for Run I analysis were the bulk of the events
  - The beginning of the Run II sample preparations can be seen in the simulation plot
- CMS is validating a new format called MiniAOD
  - Fast to produce and has the potential for saving analysis computing that had been used for producing duplicate group ntuples
  - Small to store. Intended to cover the bulk of analysis use cases



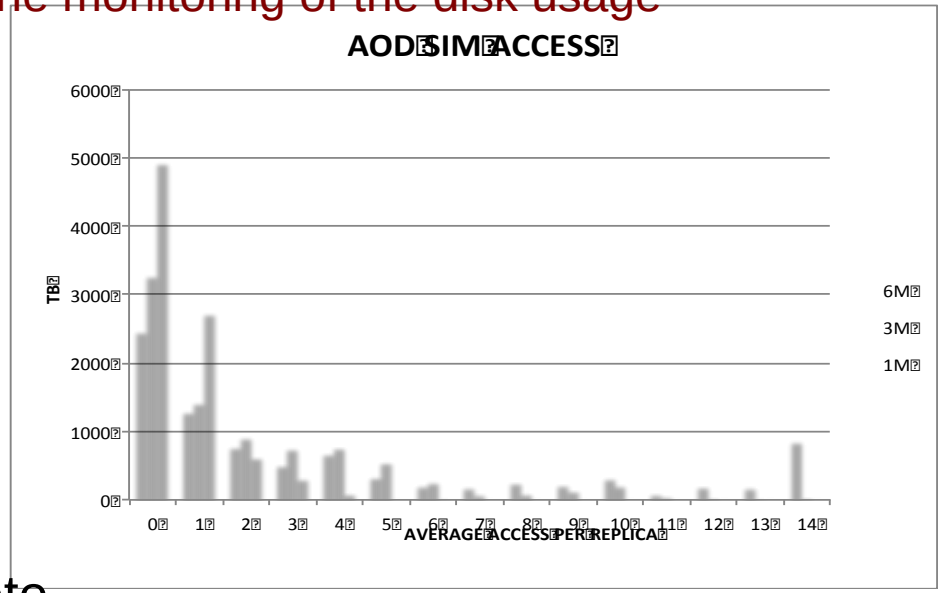
Maria Girone, CERN

Preparing for Run II

- 2014 was intended as a commissioning year in preparation for Run II
- The increase in computing capacity in Run II does not scale linearly with the increase in event rate and complexity expected in 2015.
  - Many optimizations needed to remain in the resource envelope
- Improvements were needed in data management, data access as well as production and analysis tools
  - All were intended to improve the functionality and efficiency of the system
  - Several have been significant development and commissioning efforts
    - We are still working on the validation of several areas, but there has been steady progress over the year and during CSA14
    - We are now completing the work-plan commissioning the production system for Run II

Data Management in Run II

- Dynamic Data Placement
  - Scripts to replicate data that is heavily accessed and to release the cache for under utilized samples are in place and running
    - When attempting to stress individual samples in a data federation test this fall, DDP engaged and replicated the samples
  - We are working to improve and automate the monitoring of the disk usage
    - Even at a 6M window, 27% of the disk space for AODSIM is used by un-accessed samples (10% of the total space)
    - The zero bin includes un-accessed replicas and datasets that have only **one** copy on disk

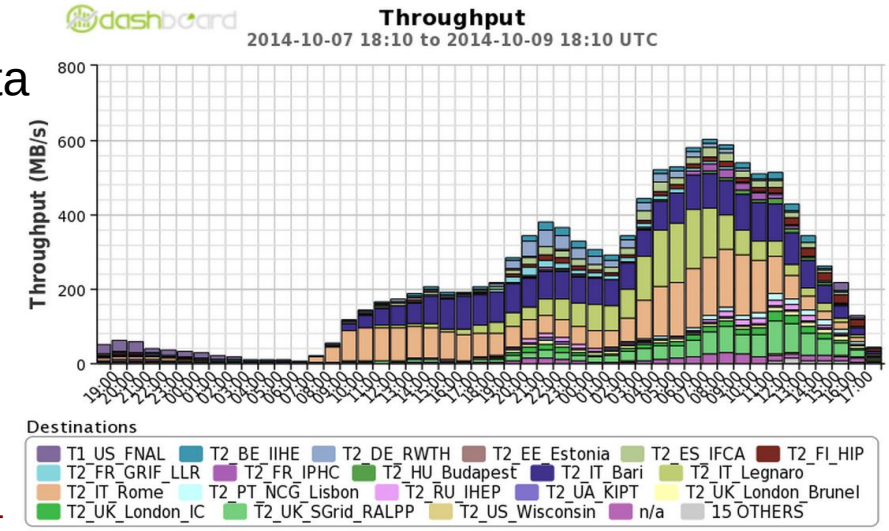


- Disk/Tape separation at Tier1s is complete
  - Allows better disk management, decoupled tape and disk functionality, analysis access to Tier1
  - All T1 sites have enabled both endpoints and we are already capable of using the new functionality

Data Federation in Run II

AAA has been a primary focus area in 2014

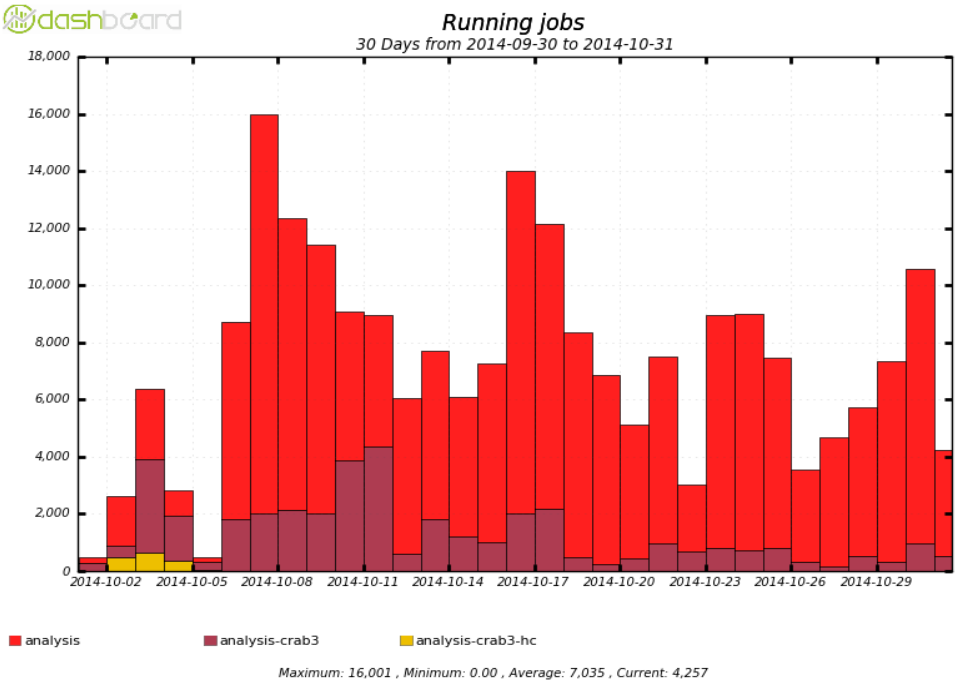
- We validated small scale use of non-local data access in CSA14
  - Fall-back when CRAB3 jobs don't find input data locally and in "ignore locality" mode
  - Very good feedback by users
- After CSA14 scale tests were performed in Europe and the US
  - 20% of jobs were able to access data over the wide area (60k files/day, O(100TB)/day)
    - Tests showed that the scale could be reached, but that the job success rates were sensitive to the health of all the sites



- Hosted xroot project leader for Offline and Computing week
  - Proposed the concept of two connected federations one production and a fall back transitional federation, which is an interesting way to involve sites that are less reliable
  - Discussing what needs to be done to transition AAA to "regular" operations. Working on how to identify site and access issues

Distributed Analysis

- CRAB3 was validated in CSA14 for specific workflows as planned
  - We have maintained a scale of about 10% since the end of the challenge with regular users
- Work plan to delivery continuous improvements with the goal of having all current functionality by the end of the year and motivating people to switch



- We need to improve the level of adoption and scale of submission
  - Reports from users are positive and system facilitated the production of miniAOD by users at similar speeds to central production
  - Overall level of adoption is smaller than we planned, and we are working to reach 50% by the beginning of 2015

Multi-Core and Tier-0

- Multi-core queues exist on all the Tier-1s and CERN
  - Small scale prompt-reco multi-core jobs work and appear to be efficient
    - Need to ramp up the scale in terms of number of sites and number of machines
- With the increased complexity and trigger rate the largest single luminosity sections can take longer than 48 hours to process using a single core
  - The improved speed of the multi-core application allows us to fit within the batch system limits
- Tier-0 workflows are running for MWGR (Midweek Global Run) including data transfer, repacking, and basic cosmic reconstruction. Scale tests using Run I saved data are scheduled for November and February
  - Lots of functionality, but a lot of validation and scaling work left to do
  - Tier-0 is not as far along in scale and hardening as we had planned at the beginning of the year

Validation of the CERN Resources

- Many changes in the CERN Site services
  - Move to a Openstack based Cloud-like virtualized resources located at CERN and Wigner
- We understand Wigner much better than we did a year ago, but we are still finding and solving issues
  - Prompt-Reco with optimized IO has a small performance hit, consistent with reading storage over the wide area (<5%). This is bulk of the Tier-0 activity
  - Digi-reco and other access to non IO optimized secondary files had a significant hit of a factor 3, but can be mitigated by insuring/increasing the replication of pile-up events in both locations
  - Somewhat unexpectedly merging has an enormous hit reading over the wide area; CMS re-activated an old functionality called “lazy download” which copies the input files to local disk in large chunks and the performance has significantly improved

Organized Production Milestone

- The final formal computing milestone is production, which was scheduled for fall
  - We have a target of improving speed of completing and announcing workflows
    - Goal is to have “tails” no longer than 25% of the processing time
  - Starting to work on improving the flexibility of resources used in workflows by enabling wide area reading of the data in production
    - Allow reconstruction of simulation at Tier-2s and sharing workflows across Tier-1s
    - The disk/tape separation was necessary to avoid duplicating tape copies

- Currently the production system performance is constrained by available computing resources, but also by available operator effort
  - We are looking to streamline and automate steps to ensure we are limited only by capacity

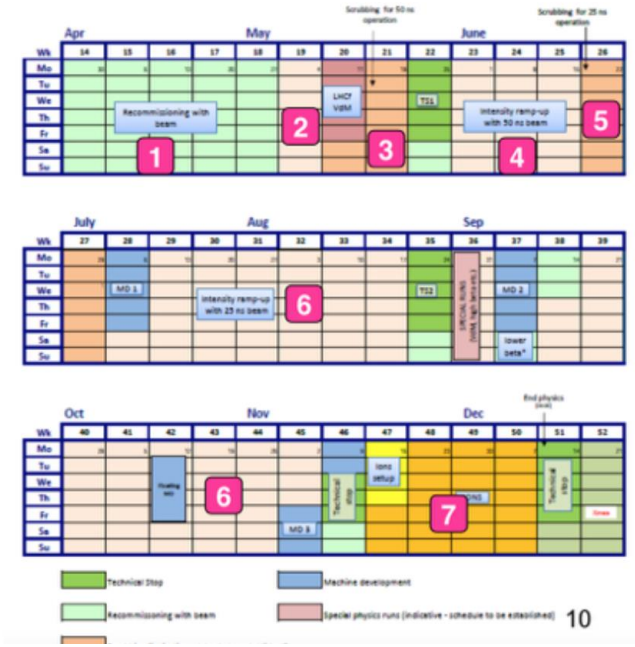
Input to Computing from Physics

- Roughly half the physics groups are expecting continued Run I analysis activity extending beyond Moriond, 2015
  - No new requests for organized processing and bulk user production for Run I should be finished. There will be some need for storage as Run I analyses close out early in Run II
- CMS is pursuing a running strategy similar to Run I
  - ~1B simulation events will be finished at the beginning of the Run with our best guess for conditions
    - Remaining sample launched when real conditions are know
  - Prompt Reconstruction will be launched with the best calibration at the time
    - We are investigating the ability to schedule reprocessing passes during the year as the running conditions change

Use of the HLT farm in 2015

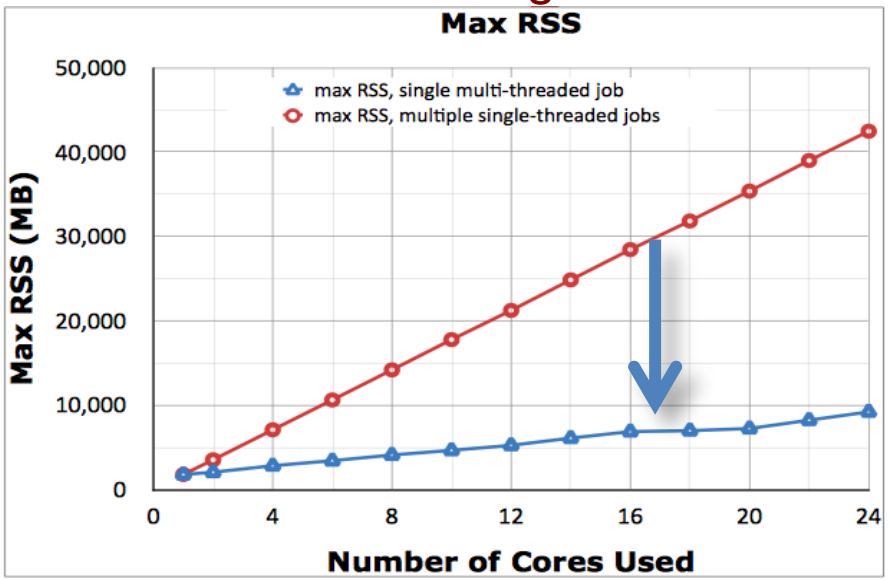
- The HLT farm has been used in production throughout 2014
  - With the upgraded network it is a large and flexible resource
- In 2015, the HLT farm is mostly busy with its primary function
  - Looking at the 2 technical stops as potential periods for reprocessing campaigns with the contribution of the HLT farm
    - Aligns well with the physics goals
  - We are also working on the capability to use the HLT farm during inter-fill periods
    - Is a good source of opportunistic computing, but will be a small absolute increase in resources
    - Requires to run workflows of a few hours

- Start June 1st
- 3 weeks @50ns ($1 fb^{-1}$)
- 3 weeks TS1 (scrubbing)
- 6 weeks @25ns high beta ($4 fb^{-1}$)
- 3 weeks TS2/special
- 7 weeks @25ns low beta ($10 fb^{-1}$)



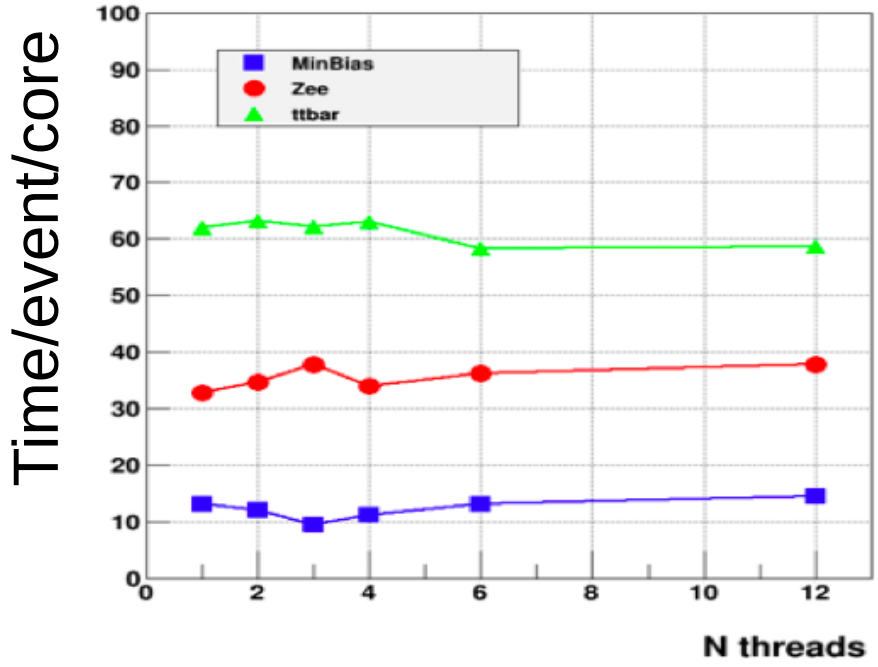
Multithreaded CMSSW applications

- Motivation for multithreaded jobs in Run II:
  - Ensure processing of luminosity sections within single job despite higher trigger rates and increased event complexity
  - Reduce number of GRID jobs to manage
  - Reduce required memory per core
  - Prepare CMSSW framework and algorithms for future technologies
- Reconstruction: Current performance meets goal for Run II
  - Continue to improve performance scaling by making more algorithms safe.
  - Big memory savings: 0.35 GB per additional thread instead of 1.8GB/job



Including multithreaded simulation

- The recent Geant4 patch release (4.10p3) addresses problems we had reported from our initial testing
  - Now we observe good CPU and memory performance in our simulation application (~800MB RSS saved per thread)



Closing out CMSSW for Run II startup

We are close to finalizing CMSSW for Run II. Important open items include

- Finalize the AOD and miniAOD content
  - Getting input from physics groups now
- Reduction of I/O for pileup simulation
- ROOT6 integration
  - Thanks to close work with the ROOT team, we nearly have a CMSSW test release with all unit tests and workflows running
  - Validation against our standard ROOT5 builds to start soon

Milestones in our release schedule

1. Freeze detector level reconstruction (~now)
2. Freeze tracking configuration in December
3. Completing the high-level reconstruction configuration and tuning in February

Outlook

- CMS just had a Offline & Computing week (Nov 2nd -6th)
  - Overall theme of this week was **"Being Ready on Day 1 of Run2"**
  - Discussions focused on development, operations for Run2 and a review of the tasks on the critical path

- Computing and software groups are facing a higher trigger rate of more complex events, a smaller amount of processing per event collected, and fewer people
  - We will need to do more with less
    - Better tools to reduce effort in production operations and improve the efficiency of analysis tools
    - Improved flexibility of how resources are used through the data federation and the efficiency of using resources through dynamic data placement
    - Improved software integration and testing processes
  - Every efficiency gain is needed contingency