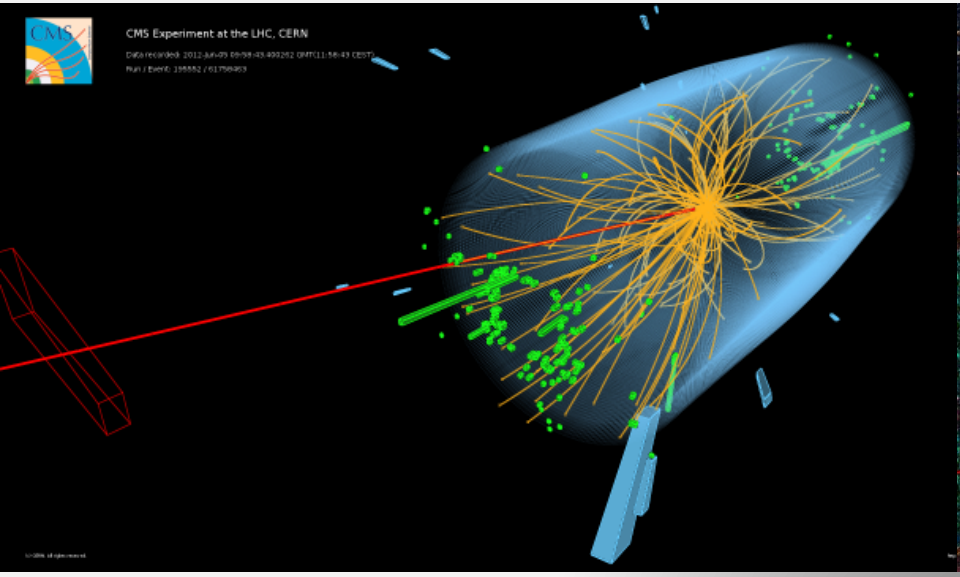
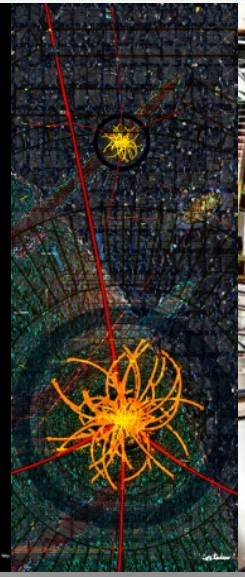
CMS Software and Computing for LHC Run 2

Ken Bloom

ICHEP 2016

5 August 2016

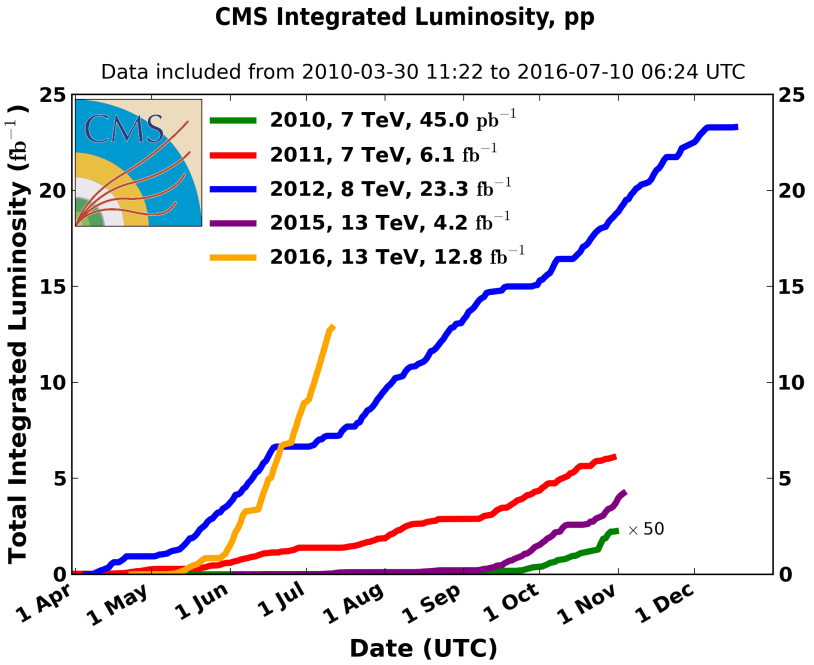


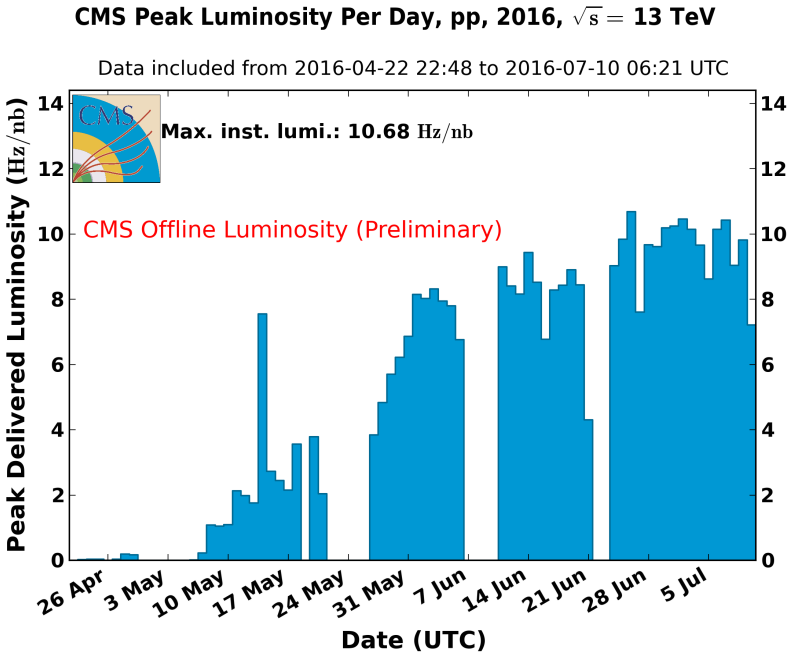


Why is this man smiling?

- Because CMS software and computing enabled the discovery of the Higgs boson!

Challenges of Run 2 (and 2016)

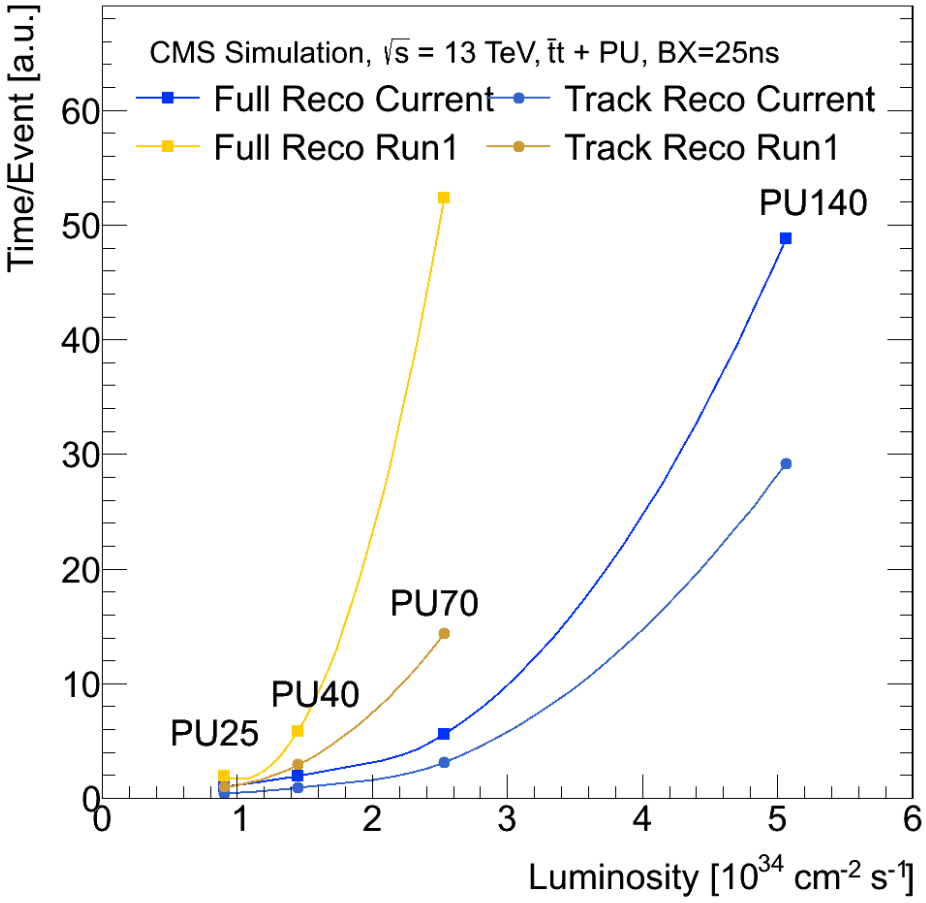




- Exploring a new energy domain with the highest luminosity ever at the LHC
  - Data arrives quickly — systems must be ready for discovery
- Computing requirements substantially larger than Run 1
  - Event rate to storage 1 kHz or more (~2.5x Run 1), pileup to reach 50
  - Without improvements, would need x6 increase in CPU for reconstruction
- Used the long shutdown to modernize CMS software and computing
  - Have delivered a system of increased agility and flexibility that enables physics discovery
  - All built off the extremely successful systems from Run 1

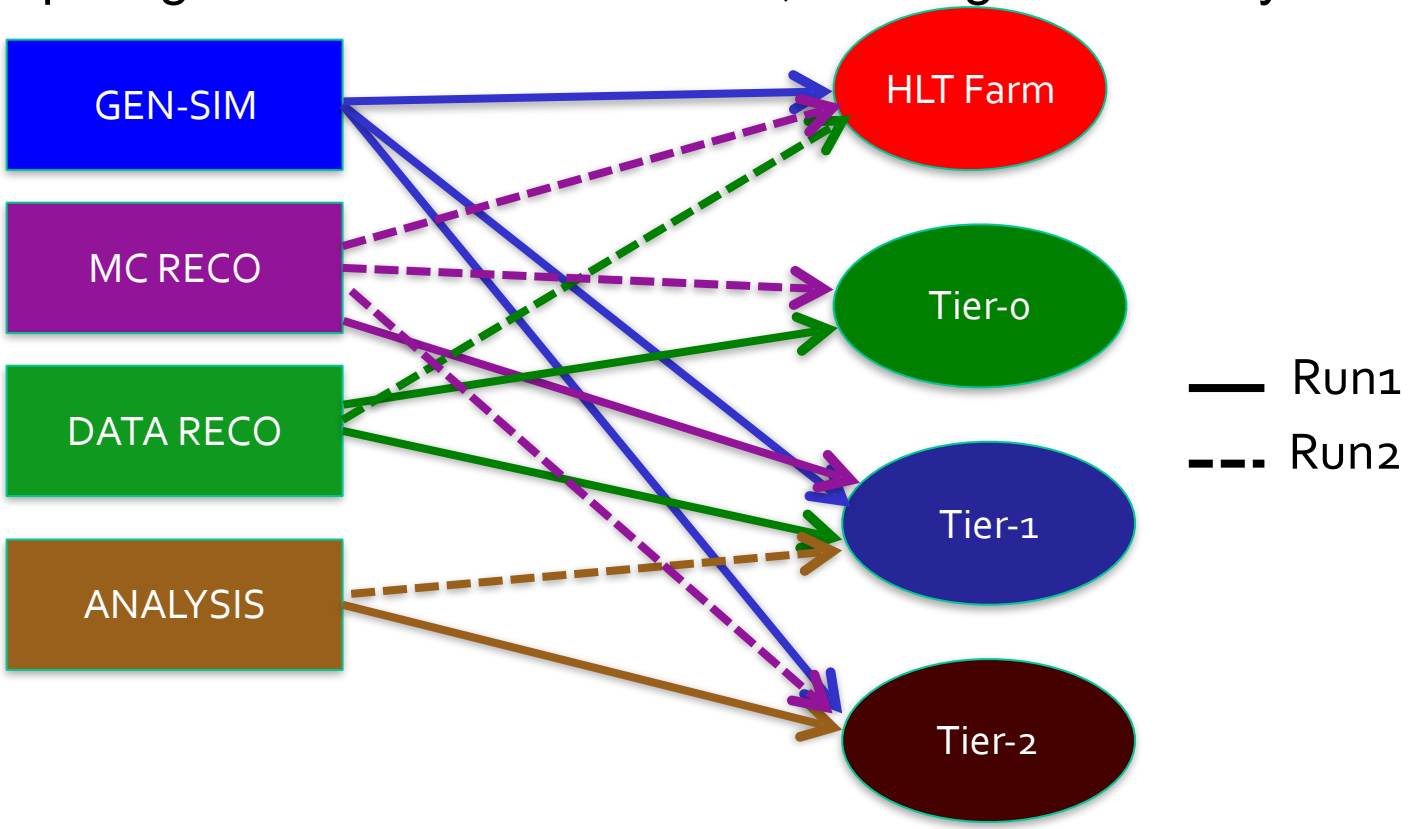
Improved reconstruction

- Event reconstruction time has been reduced while maintaining physics performance, even in more difficult event environment
  - Some improvements strictly technical/engineering
  - Others are algorithmic, e.g. changes to tracking algorithms that reduce fakes and speed execution
- Simulation time also improved by reducing time spent tracking low-energy particles in GEANT4



Flexibility for facilities and workflows

- Use computing facilities in more flexible, heterogeneous ways



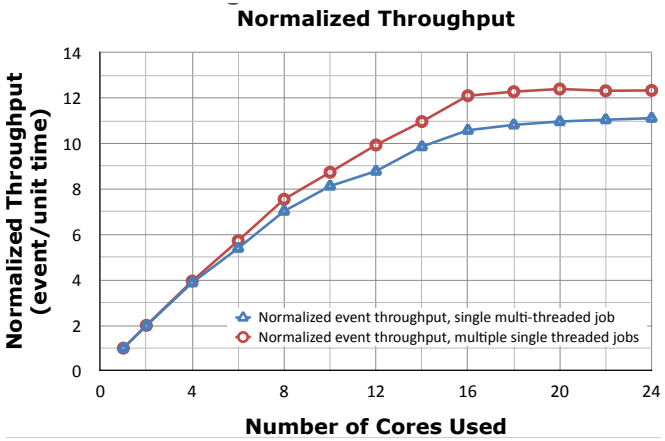
  - Use HLT (~size of T0) for organized processing during technical stops
  - Commission T2's to do reconstruction tasks previously limited to T1
  - Allow analysis jobs to run at more sites
- The more places work can run, the faster the work goes!

New services for more agile operations

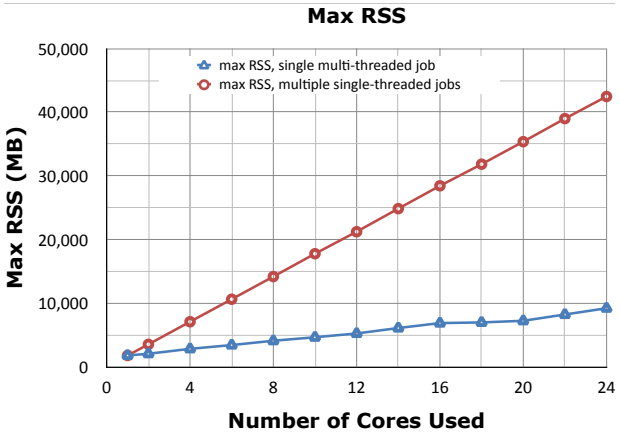
- "Any Data, Anytime, Anywhere" data federation
  - CMS applications can read data efficiently over wide-area networks
  - Relaxes constraints on locations of datasets and workflows
- Disk-tape separation at Tier-1 sites
  - Greater control over what datasets are available on disk
  - Through AAA, allows T1 data to be used in workflows anywhere
- Dynamic data management system
  - Automatic transfers of datasets on creation, deletion when not needed
  - More agile and efficient use of disk space
- Global pool for resource provisioning via glideinWMS
  - Allows central control of job priorities, simplified infrastructure
  - Scales to operate all T1/T2/opportunistic resources in single pool
- Ability to provision cloud infrastructures via glideinWMS
  - Allows use of HLT and potentially opportunistic and commercial clouds
  - Ability to burst into extra resources if necessary
- Establishment of 100 Gbps transatlantic network link via ESnet

Multi-thread, multi-core

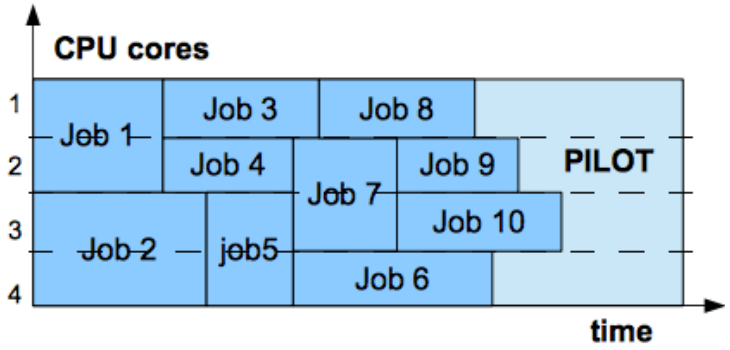
- Code for both simulation and reconstruction now multi-threaded!
- Use several CPU cores concurrently to reconstruct multiple events simultaneously
  - Less demand on computing infrastructure — fewer open files, fewer jobs….
  - Reduce time to process luminosity block of data, needed for higher trigger rates
  - Huge reduction in memory per CPU core with little efficiency loss



95% efficiency compared to single-threaded jobs at significant memory savings



- Enables the use of multi-core pilot jobs with internal dynamic partitioning of resources for greater efficiency



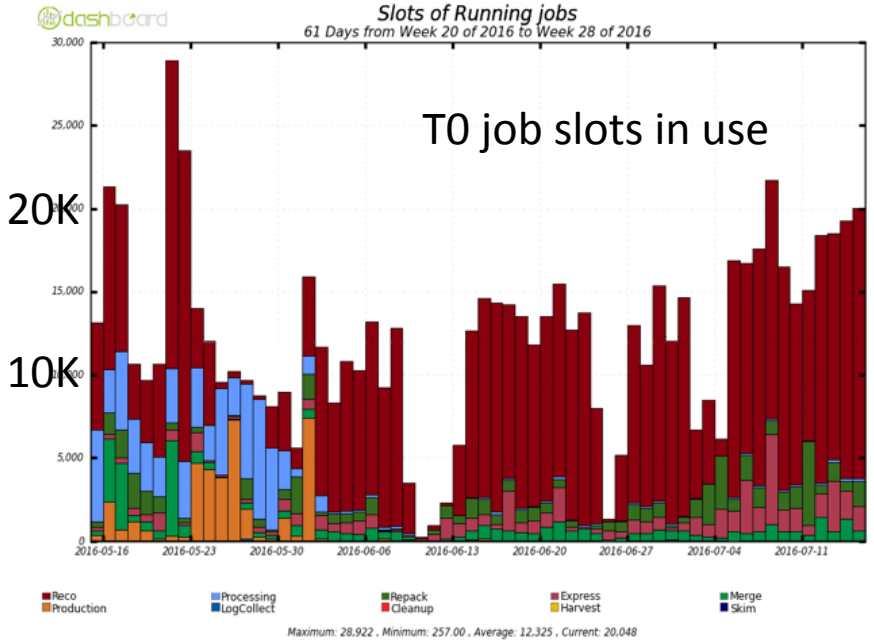
Better tools for physics users

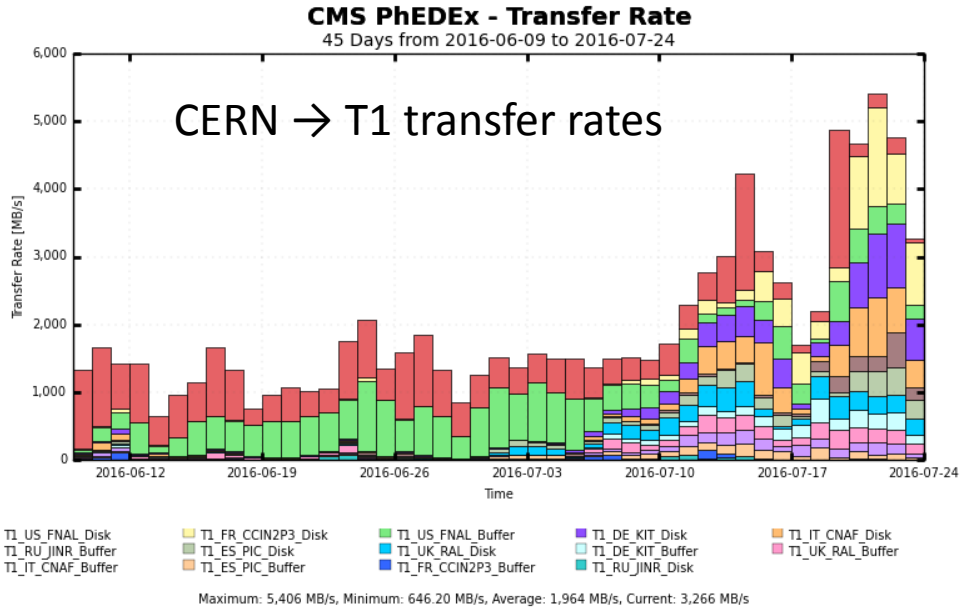
- Physics analysis is easier, more flexible, less resource-intensive in Run 2
- New analysis job submission tool (CRAB3)
  - Automatic job retries, better job tracking
  - More reliable delivery of output with centralized transfer handling
  - Thinner client, more logic on server side allows easier upgrades
  - Fully exploits HTCondor and glideinWMS systems
    - Including job overflows from busy to less-busy sites
- New miniAOD format: ~30 KB/event
  - 1/10 the size of AOD, designed to serve ~80% of analyses
  - Easier to keep more of the data at desired locations
- AAA data federation
  - Job location no longer tied to data location
  - Major enabler for university-based data analysis

Dynamic resources

- Ability to rapidly expand resources for burst needs could be a game changer for resource provisioning
  - If successful, could own CPU for average needs rather than peak
- Very successful demonstration with Amazon Web Services via Fermilab HEPCloud (see this presentation)
  - Diversity: Can do all types of CMS production workflows on all AWS resource instances in all AWS availability zones
  - Scale: Met goal of running at least 50,000 jobs simultaneously, with only 9.5% “badput" and 87% CPU efficiency
  - Knowledge: Greater understanding of how to optimize cost per unit output
  - Physics: 518M events generated in early February that went directly into results shown at major March conferences!
- Exploring possible follow-ups on all fronts
  - Other commercial providers, opportunistic cycles on Open Science Grid, “friendly” HPC sites (e.g. NERSC), XSEDE resources

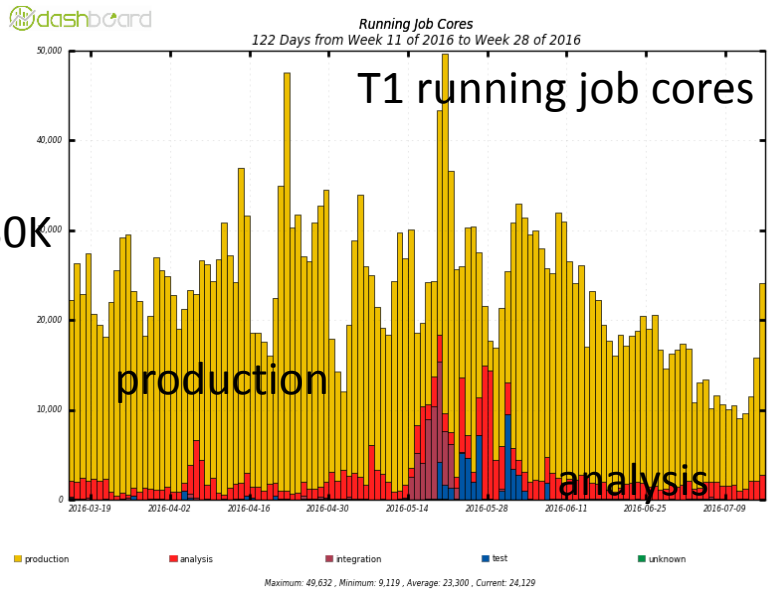
2016 performance: T0 keeps busy

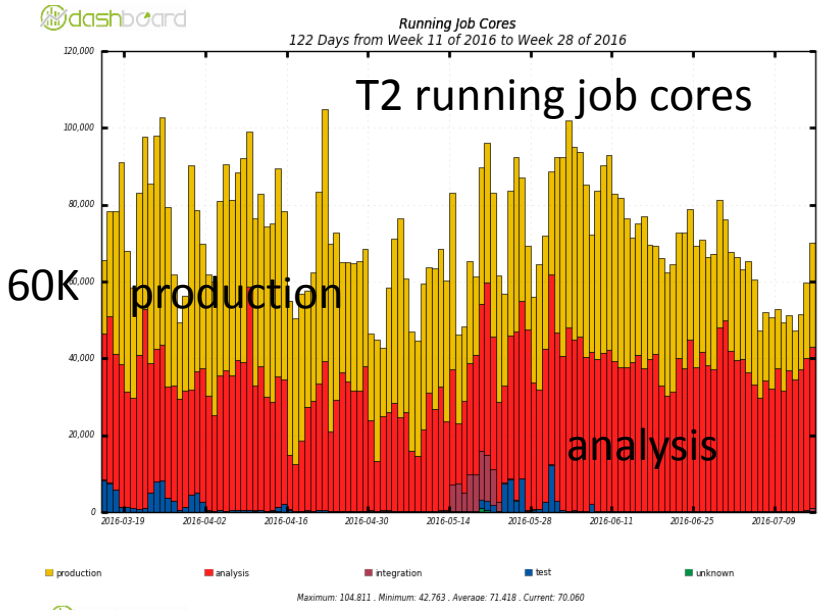


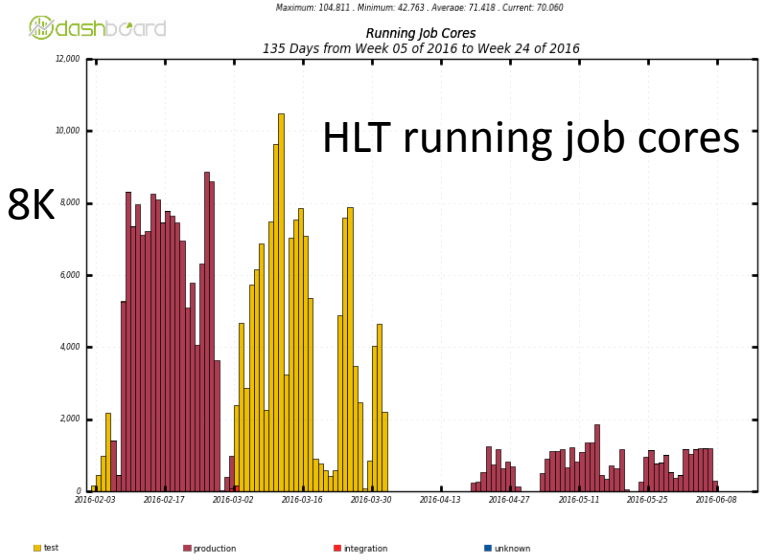


- T0 system hard at work to keep up with LHC data
- NB: plot shows number of job slots, not number of jobs
  - Fewer jobs than slots thanks to multicore processing
- Excellent LHC performance means more data must be transferred out of CERN to T1 sites
  - Much work in past weeks to improve rates

2016 performance: grid keeps busy

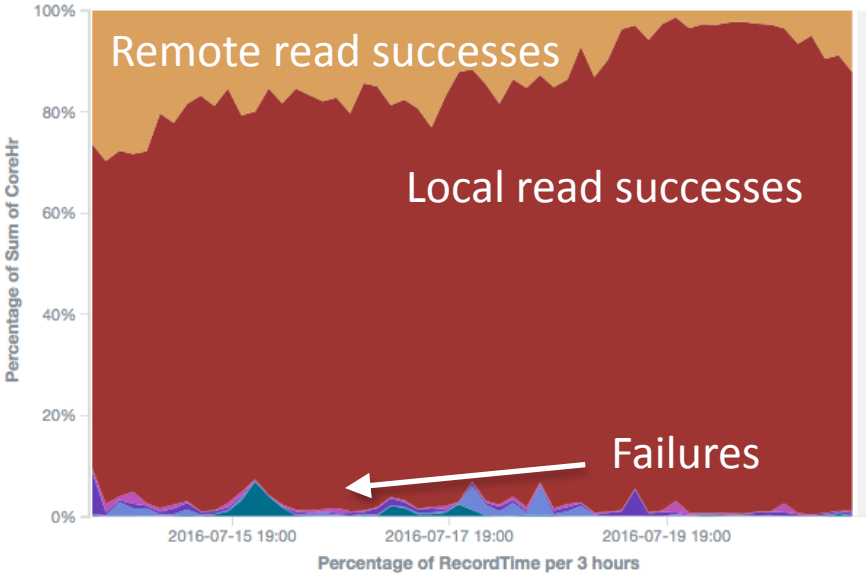




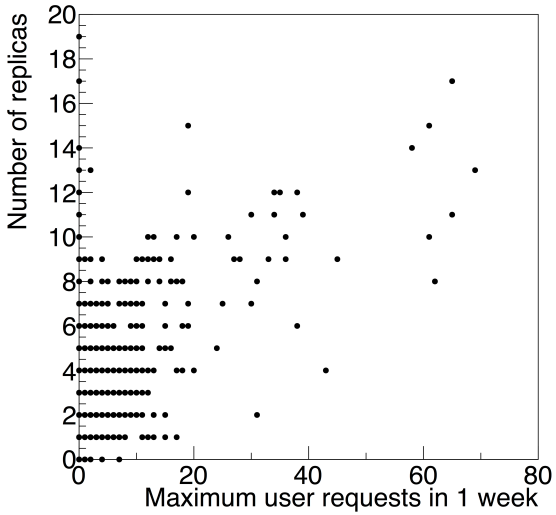


- T1 and T2 sites routinely busy with a mix of activities
  - T1s run user analysis
  - T2s run DIGI-RECO
- HLT usage can scale sufficiently; usage during technical stop periods

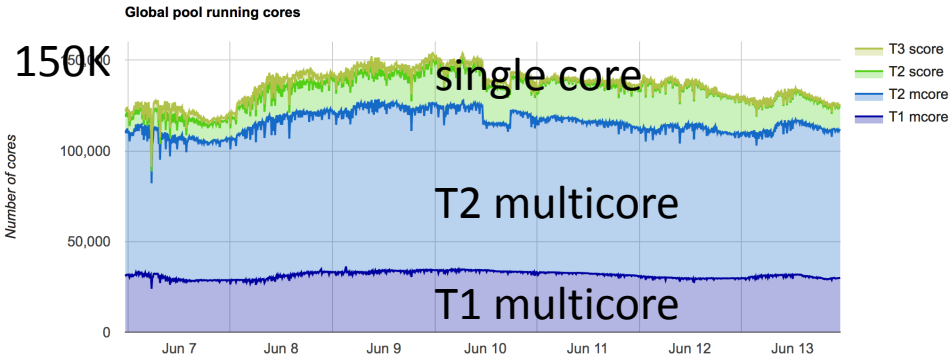
2016 performance: new services work well



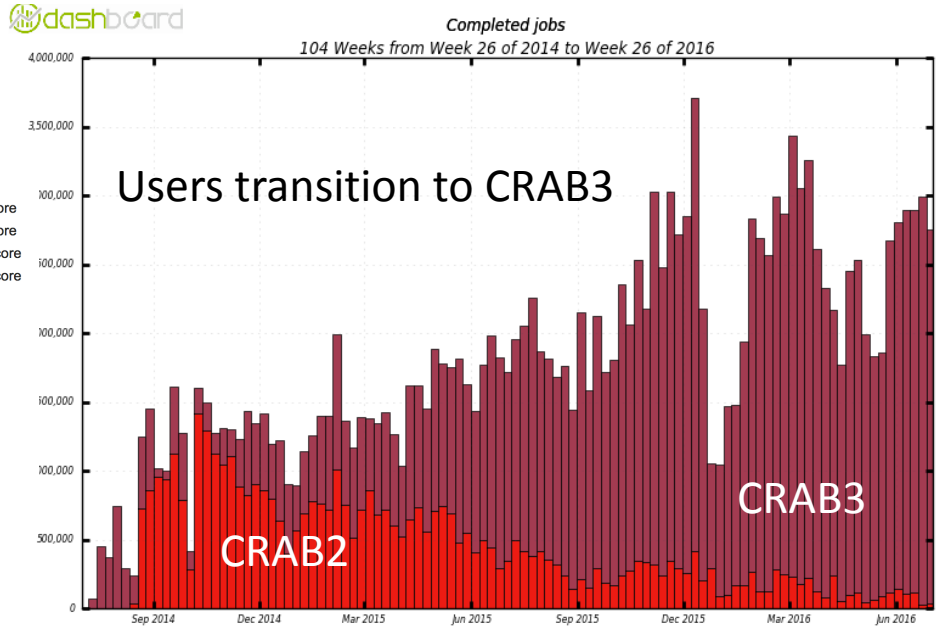
Low failure rate of AAA in MC production



Dynamic data management ensures highly-requested datasets have the most copies

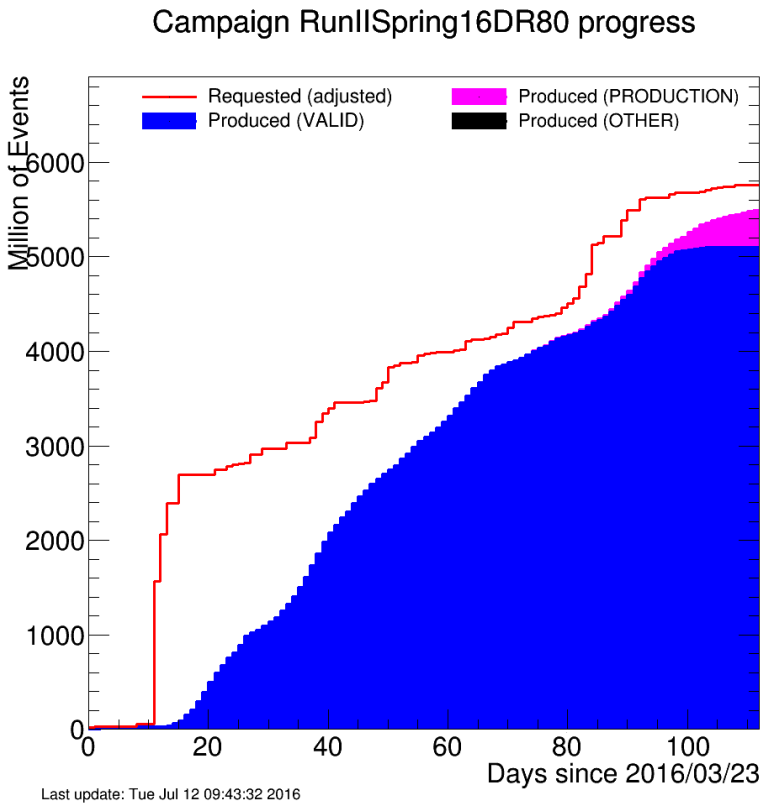


Global pool at full scale, mostly multicore



2016 performance: physics results!

- Produced billions of simulated events for ICHEP analyses
- Last data for ICHEP analyses collected on July 15; all of it was successfully ingested
  - 12.9/fb analyzed
- ~40 results on full 2016 data shown at ICHEP, demonstrating that CMS computing has sufficient throughput for quick turnaround



Billions of events simulated for this conference!
Learn more about MC production here

CMS Run 2 S&C: a great success

- CMS software and computing was very successful in Run 1, but could not — and did not — sit still during LS1
- Significant evolutionary changes to Run 1 systems
  - More flexible resource usage
  - More efficient resource usage
  - Better tools for physics users
  - Take advantage of technical developments
- These changes are now fully operational for Run 2 data taking and analysis, enabling the production of frontier physics results with fast turnaround, even with harsher experimental environment

- If Nature cooperates, CMS software and computing will have everyone smiling again!