Liz Sexton-Kennedy

Resource Utilization & Optimization

Toronto

INFN Bologna

Glasgow

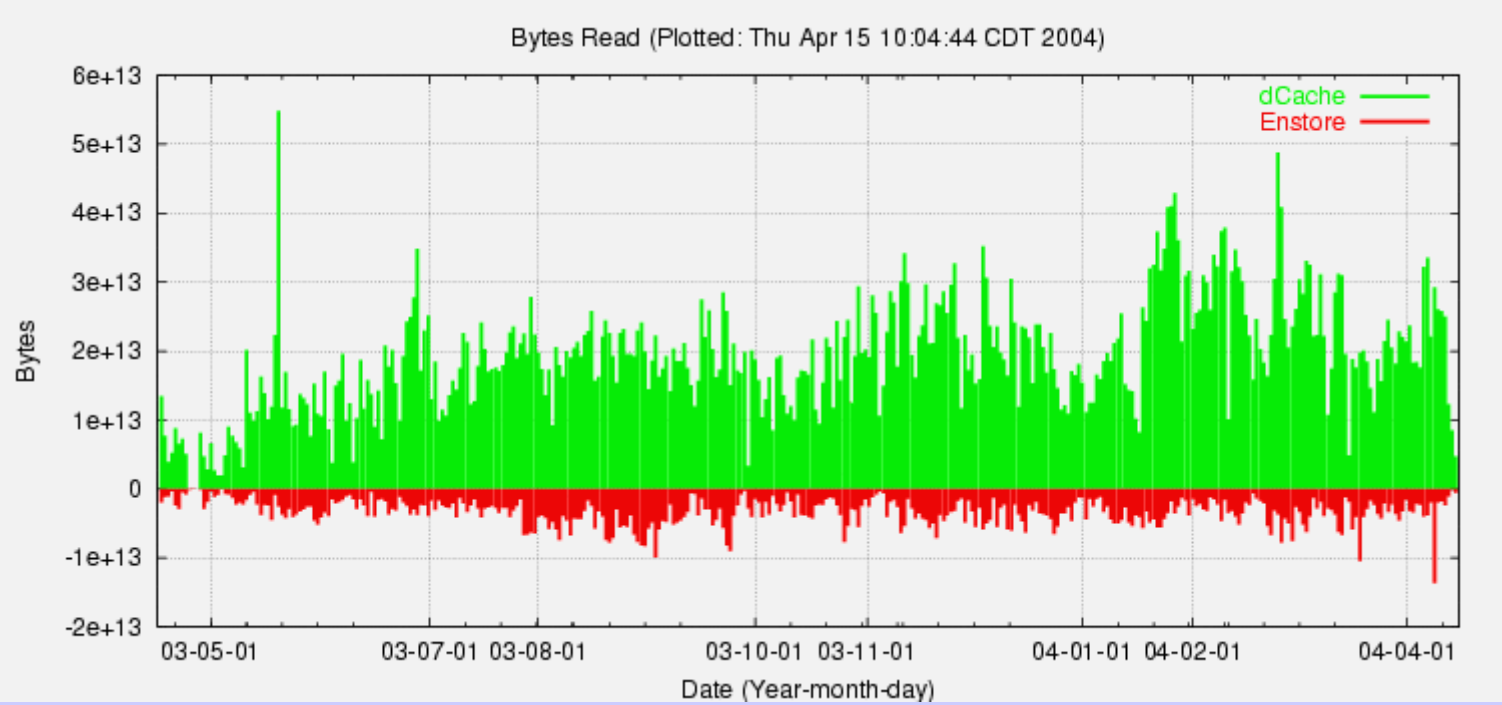
UC San Diego

- What we've done
- What we want to do

All MC done at remote sites, and 2 remote CAFs

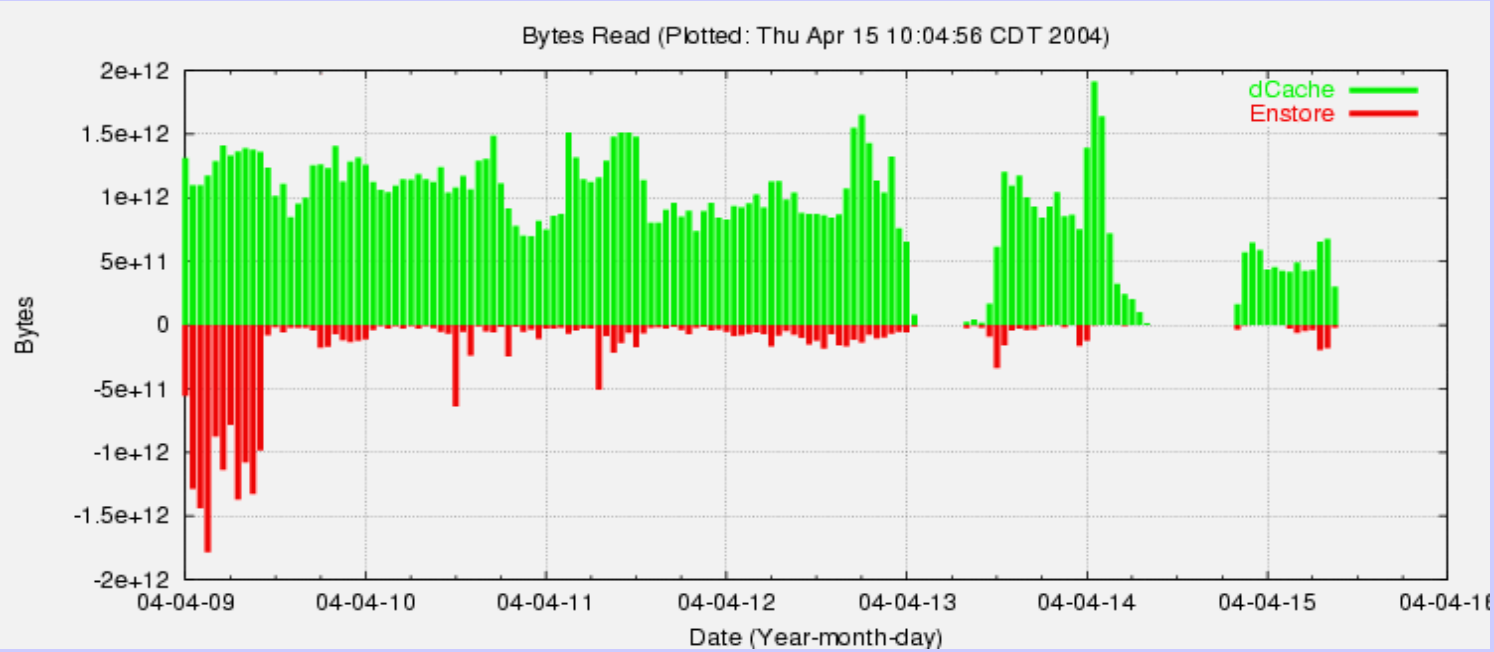
Resource Utilization - DH

- 685M unique events have been processed
- 747 TB on tape, 4 copies x 250kB/event + MC
- 48TB/day moved, at 900MB/sec sustained

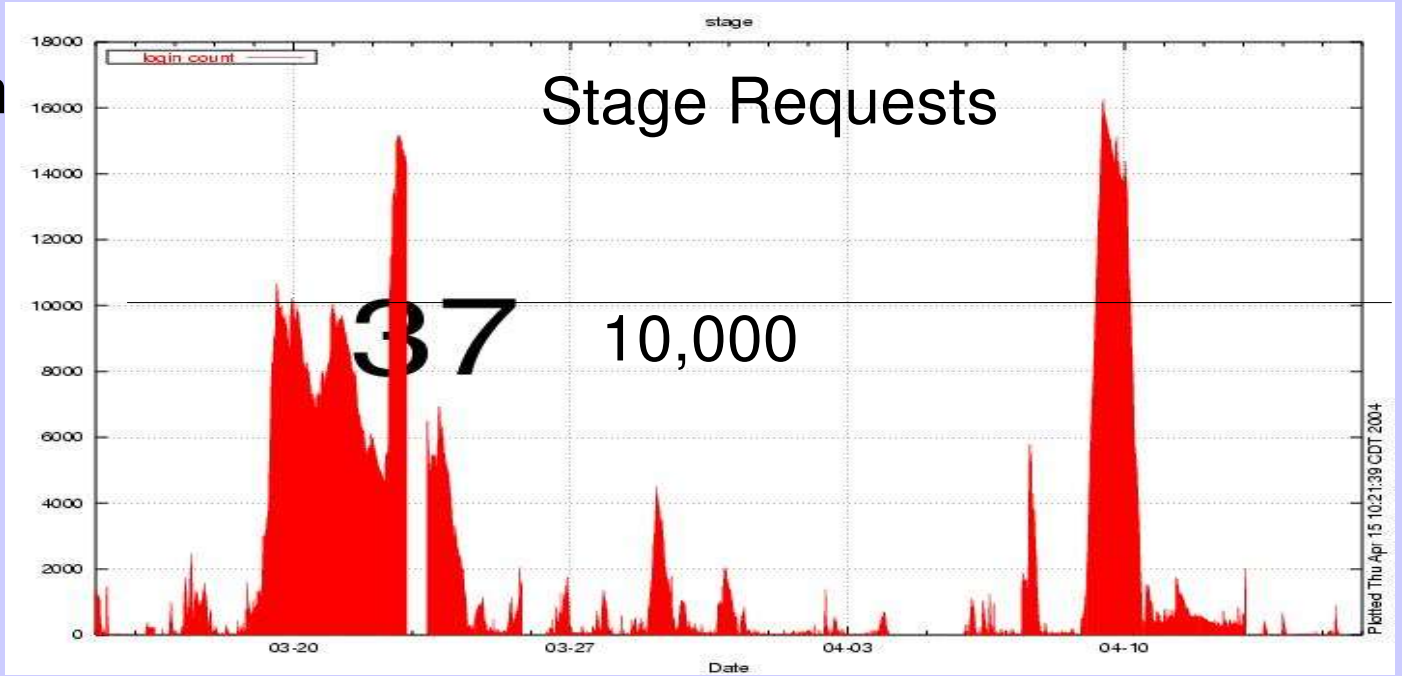


Resource Utilization - DH

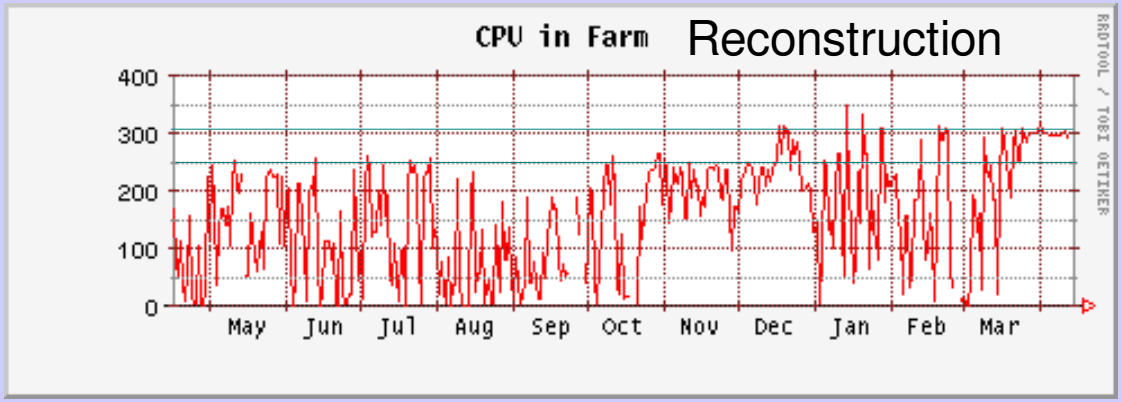
Current system can clearly benefit from resource management



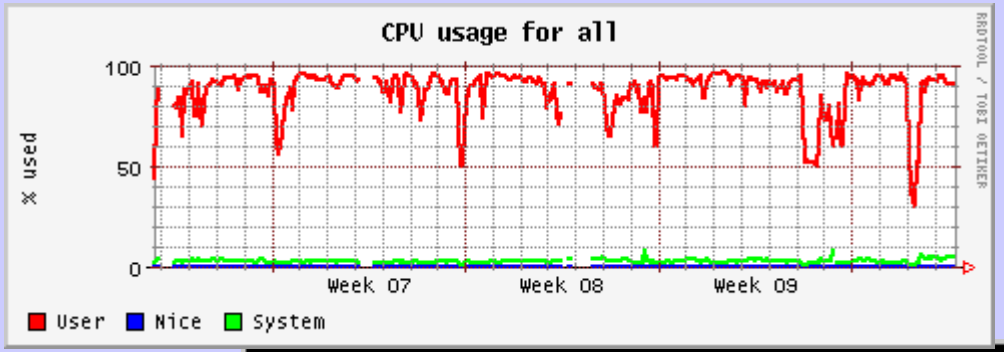
When this system fails it's a service outage for all.

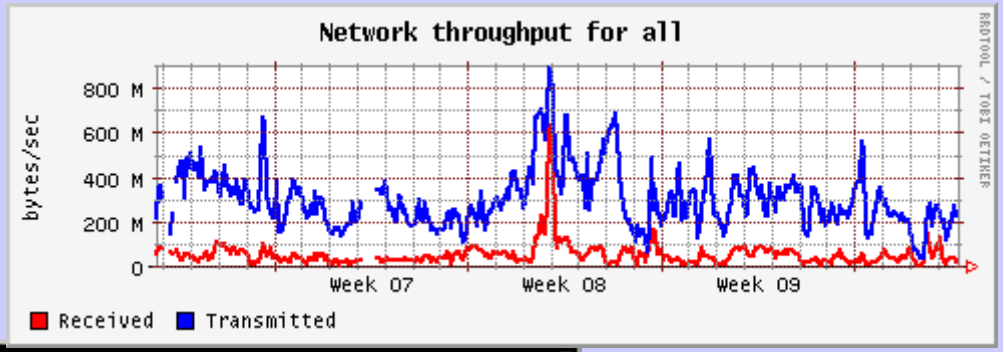


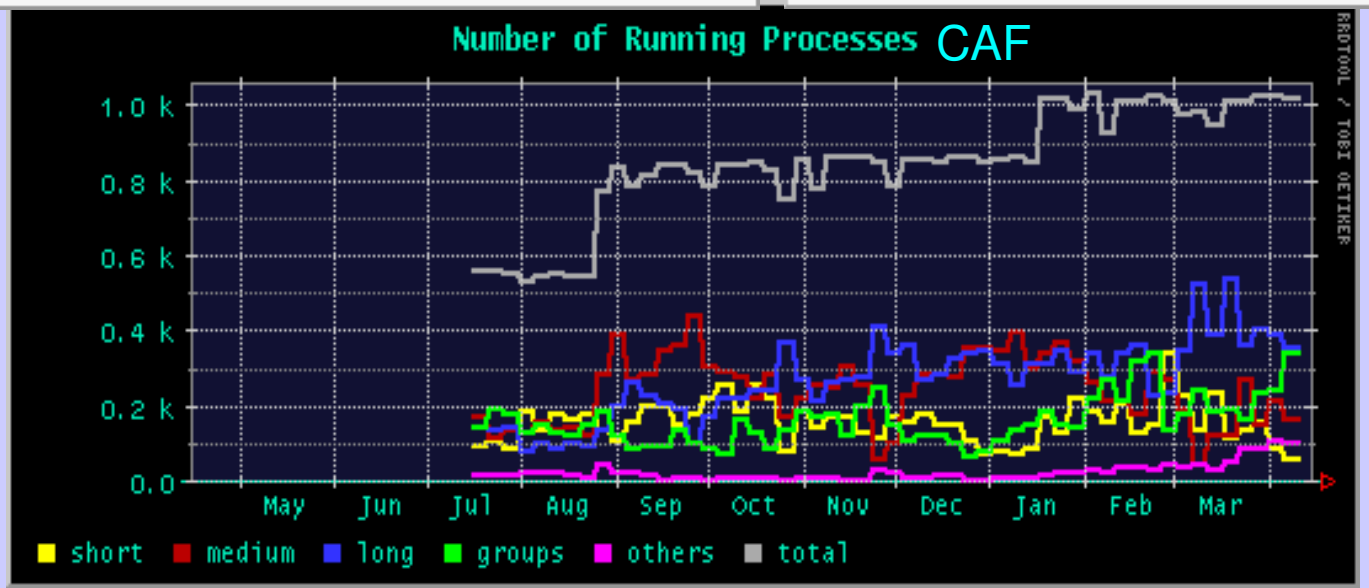
Resource Utilization - Farms



~250 cpus last fall
~300 cpus in winter



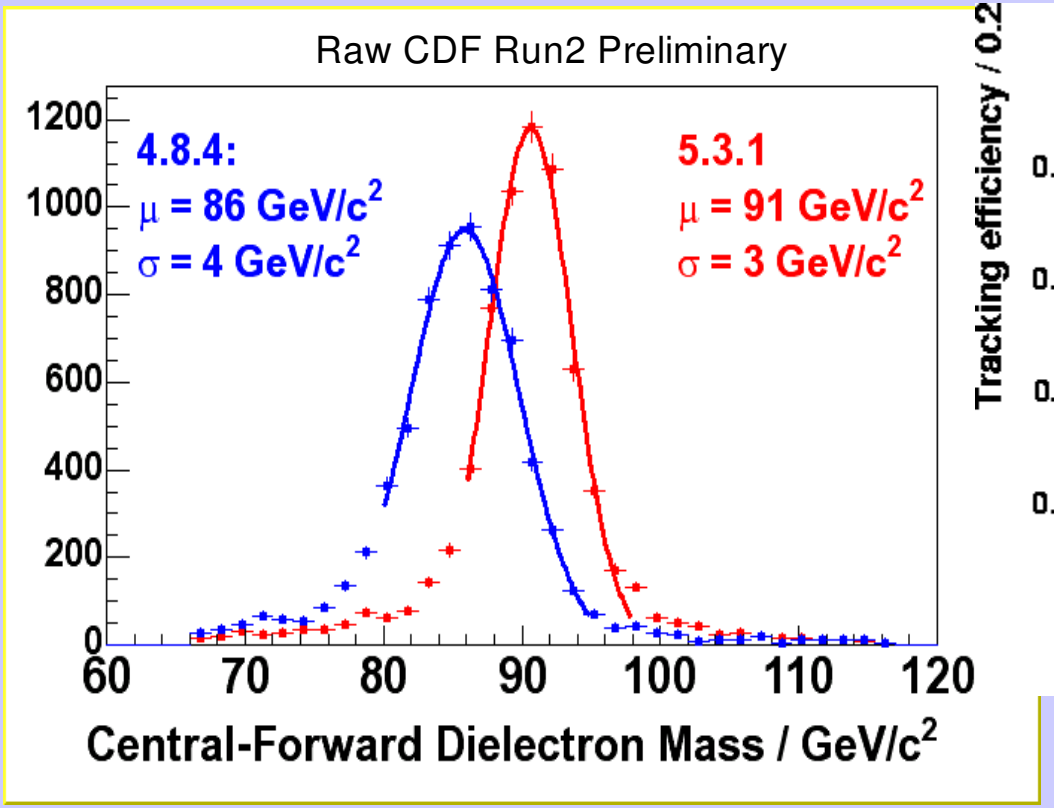


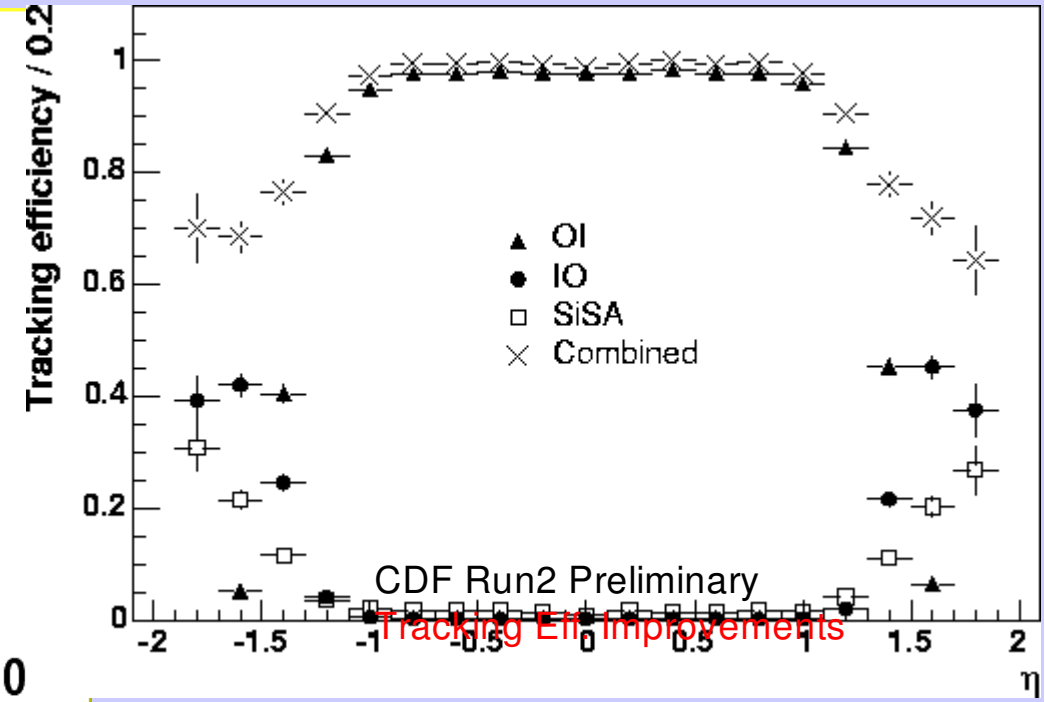


3*384 nodes
cpus=2*384
(Condor CAF
not shown)

What We've Done -Reprocess

- Reconstruction passes(all data, 3 times):
  - 4.8.4 ran non-optimized, no forward tracking or IMU
  - 5.1.0 optimized, alignment improved, beamline used
  - 5.3.0 uses final CAL calib., high forward tracking eff.
  - Output of production is immediately useful



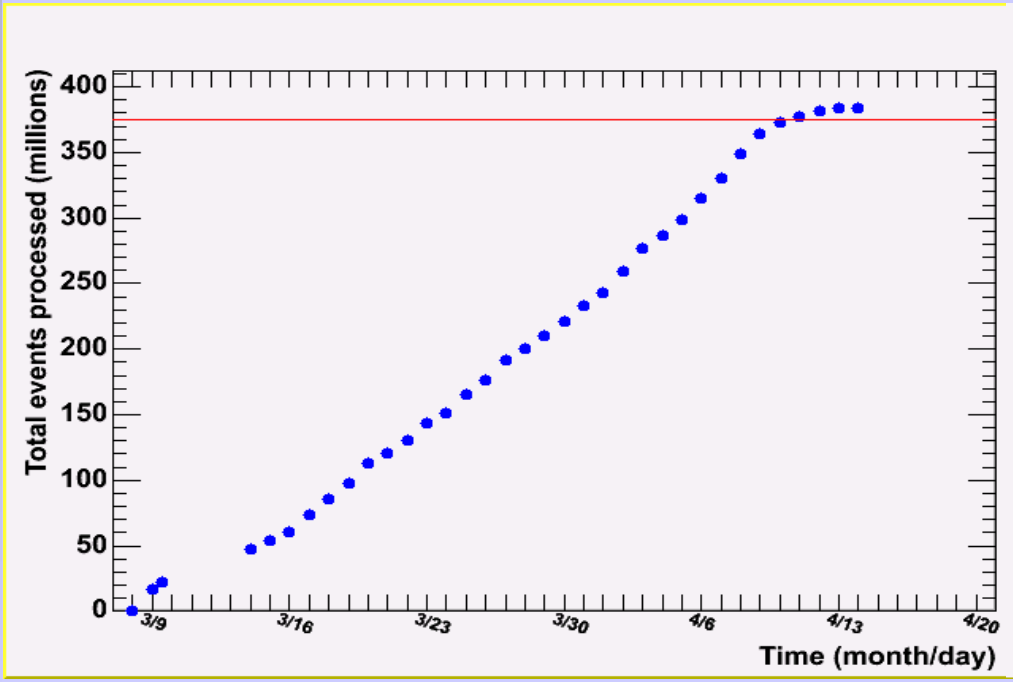


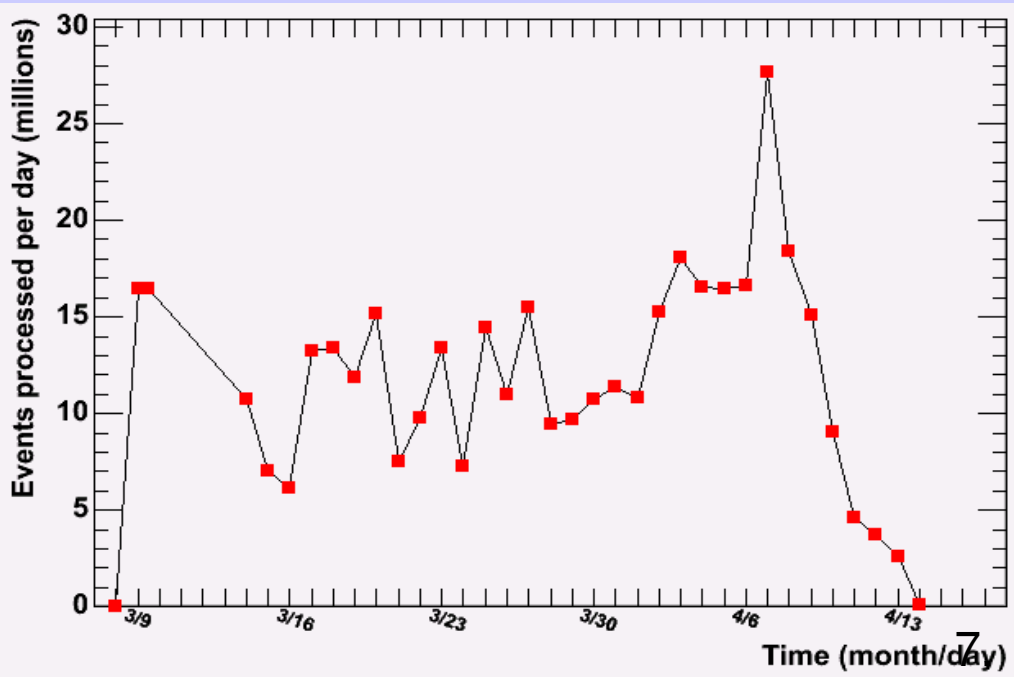
Major upgrade of Simulation

- Detector geometry description improvements
  - Added detectors
  - Tuned passive material with photon conversions
- Parameterized Si charge deposition models
- COT drift model improved
- Calorimeter response tune – the results jet energy scale task force
- Much wider use of “realistic” simulation
  - Inactive regions or channels are ignored
  - Misalignments are applied
  - Fully digital trigger is simulated
  - Big improvement over run1 but cost resources

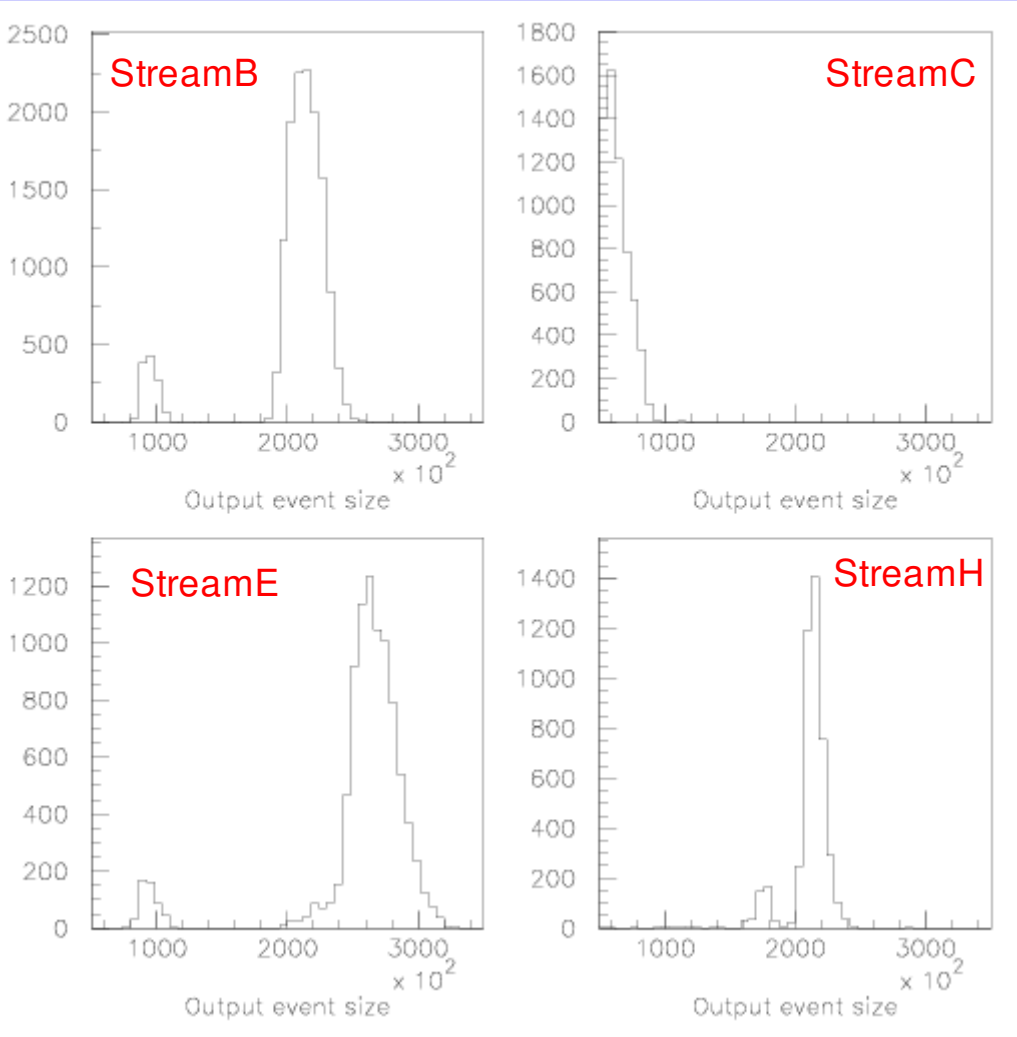
Success of 5.3.1 Reprocessing

- Split and conquer
  - Run 2 datasets are split into 42 different physics streams
  - An example: high pt muons = 3M events in 200pb-1
  - The first time the data is split it is very costly in terms of farms operations
  - Subsequent reprocessing can be done very quickly



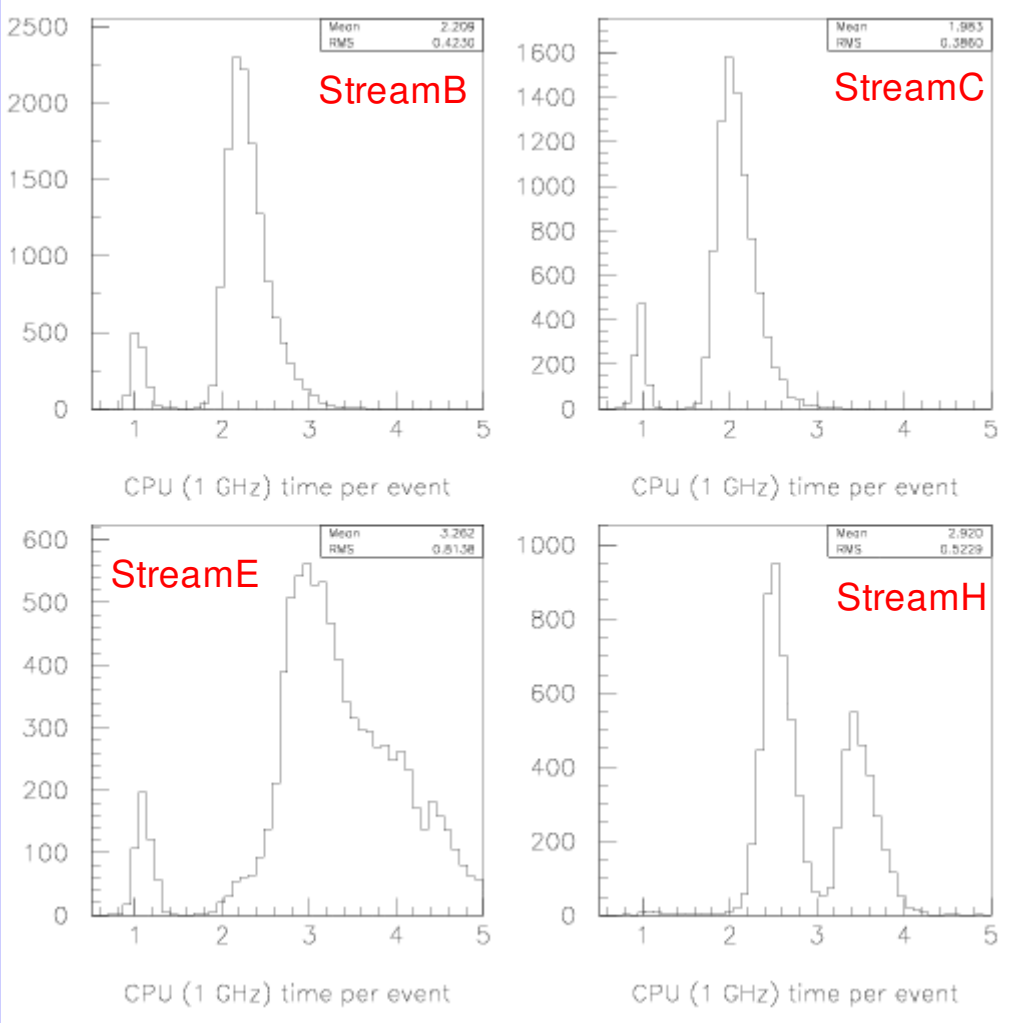


Output Event Size



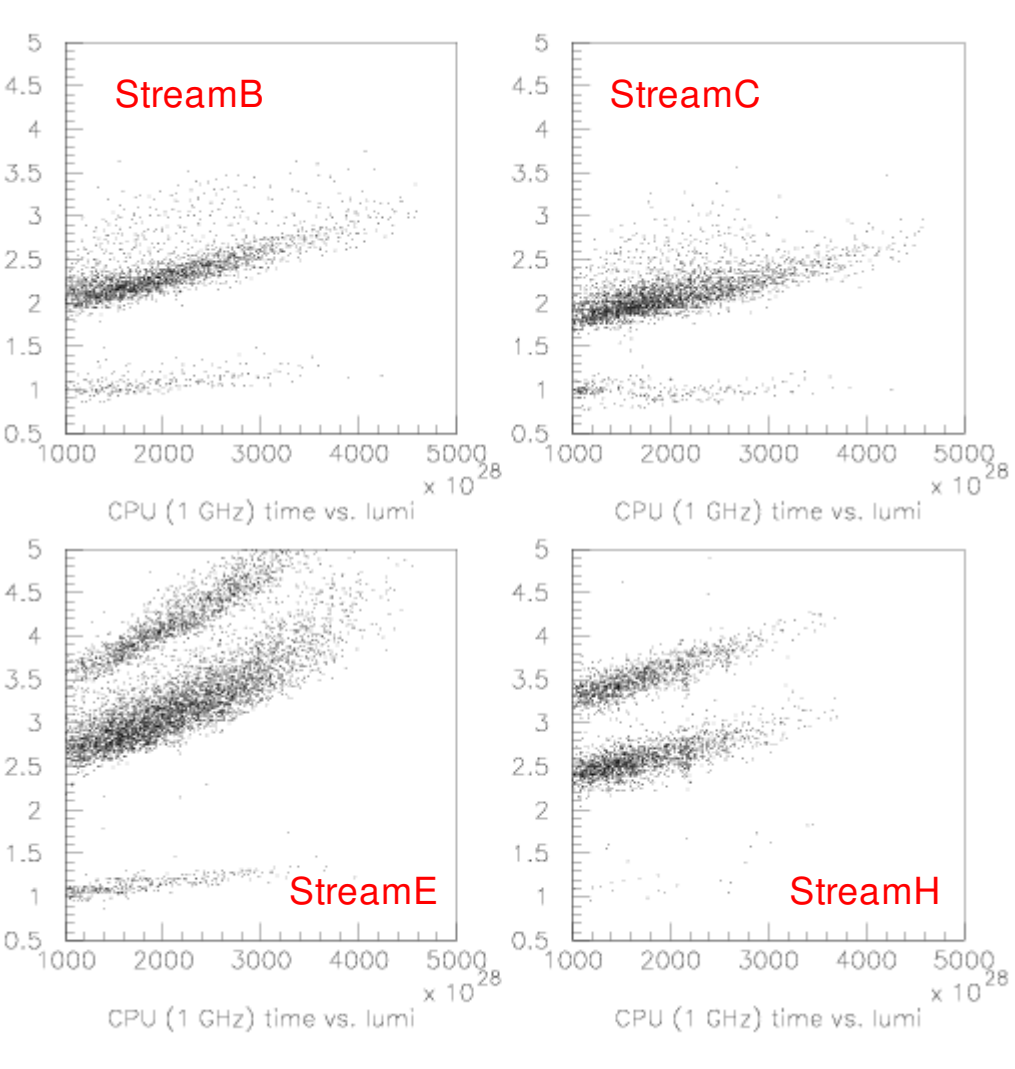
- B=high PT leptons
- C=photons
- E=exotics
- H=all hadronic Bphys

CPU time/event



- B=high PT leptons
- C=photons
- E=exotics
- H=all hadronic Bphys

CPU vs. Luminosity



- B=high PT leptons
- C=photons
- E=exotics
- H=all hadronic Bphys

What we plan to do

- Improve data format further – trigger info.
- Reduce the number of copies of data
- Continue to improve tracking – may prompt another reprocessing
- Prepare code base for required GRID infrastructure without disrupting physics
- Continue to support core software packages
- Improve support for analysis packages eg. Stntuple, Btag*. They should be validated for frozen releases as they mature.

Performance of Analysis Code?

- As shown in the CAF plots, physics will fill the available CPU resources.
- Limited only by the imagination of the physicist
  - Encourage them to “do it right” otherwise pain
  - Encourage optimization
  - Without harming physics program of CDF
- You can't tell them how to write their code but you can concentrate on optimizing shared tools, examples:
  - SecVtx, ctvmft (constrained track vertex mass fitter)
- Not much else you can control

Summary

- CDF has put a lot of effort into reconstruction improvements, robustness and optimizations
- Further improvements in CPU optimization can come from operational stability and job success rate gains
- Over the next year more attention should be paid to analysis tools. This is the one remaining place where we could still make big gains.