CDF Production Status and Plans

Ashutosh Kotwal

Director's Review of Run 2 Computing

13 September 2004

Outline

- Production tasks:
  - Reconstruction of collider data
  - Generation, detector simulation and reconstruction of coordinated Monte Carlo samples
- Current schemes and upgrades: two motivations
  - Migrate to common, global mode
  - Improve operational efficiency
- ...so that CDF physics needs are met when
  - We have order of magnitude more data
  - Fewer people to operate the experiment (LHC involvement)

Upgrade Plans

- Single-pass data reconstruction scheme
- Data-processing farm upgrade
- Monte Carlo Production upgrade

Current Data Reconstruction Scheme

- Most detector calibrations (drift chamber, beam-line parameters etc.) needed for reconstruction are produced in quasi-real-time
- Reconstruction farm processes and splits data into physics data-sets, with latency ~3 days from data-taking
- Time-dependent and tower-dependent calorimeter constants are extracted from offline analysis of processed data
- On a time scale of few-many months, data are reprocessed on farms with final calorimeter constants

Current Data Reconstruction Scheme

- Some B-physics analyses already start using output of first-pass processing, as they are not sensitive to calorimetry
- High-p_T physics groups typically use output of second pass for analysis
- Validation procedures for beam-lines etc. are stable and automated
- Analysis procedures for extraction of calorimeter calibrations are becoming stable
- Natural next step: move analysis and extraction of calorimeter calibrations on similar, rapid time scale as other calibrations
  - Process and produce final "physics" quality data for all physics groups in single pass

Advantages of Single-Pass Reconstruction Scheme

- Conserve human resource
- Conserve CPU: immediate reduction by factor of 2 of CPU usage for reconstruction
- Faster availability of physics-quality data for all analyses

Single-Pass Reconstruction Scheme

- CDF Workshop on August 19 to work out details
- Build on beam-line generation semi-automated machinery
  - Tower gains from minimum bias data
    - already reconstructed for beam-lines
    - Produce calorimeter analysis ntuple from beam-lines executable
  - Calorimeter data for electron sample
    - half of min-bias sample
- Automated validation plots for calorimeter calibrations
- Total estimated latency: one month
- Estimated disk for calibration samples: 10 TB (for 35MB/s peak logging rate)

Software Infrastructure Upgrades for Data Processing and Monte Carlo Production

- Long-term (~2 year) goal: create a uniform, grid-enabled computing platform for all computing
  - Data processing
  - Monte Carlo production
  - User analysis
- A strategy for deployment of tools to get there:
  - Identify projects where operational efficiency improves “right away”
  - Immediate impact on physics output for winter conferences
  - Introduces new “culture/style” of computing/data-access to physics users
  - Increased user involvement => more help in deployment

SAM on CAF

- The CAF has become a very successful computing platform for run 2:
  - 6 papers published
  - 15 paper drafts circulating/submitted
  - 27 advanced analyses
  - FNAL CAFs saturated (partly due to its reliability), but users generally happy
- User Analysis on CAF is dominant CPU usage, compared to data-processing and "official" Monte Carlo production
- CAF is the natural choice in the short term for a uniform computing platform
  - Simplifies migration to GRID for all CDF computing simultanously
- Deploying SAM on CAF is the next short-term goal

Farm Upgrade Plan

- Proof of principle exists: DCAFs (offsite CAF's) already SAM-enabled
  - B physics datasets reprocessed at UCSD
- Migrate Data Production Farm software infrastructure:
  - Run data-processing on CAF
  - Use SAM for data-flow control
  - When fully tested, recycle current farm hardware to add to CAF
  - Maximizes CPU utilization: allow sharing of common CAF resources between data-processing and user analysis
  - Allows optimization of resource usage: eg. CPU *vs* Disk *vs* Tape
- Share experience & knowledge gained with MC Production Upgrade

MC Production Upgrade Plan

- Use SAM to automate book-keeping of output before and after concatenation
  - Reduce human intervention => maximise operational efficiency
- Improved "productivity" of MC Production Group implies
  - More efficient usage of shared, validated samples by physics groups
  - Increase efficiency of physics users: less time spent generating and validating private samples, more time for physics analysis
  - Reduction in CAF usage for private MC generation (currently about equal to official MC production)
  - Net effect: better organized, more disciplined user analysis culture
- Time-scale: implement in time for large scale MC production for Winter Conferences

Consolidation of Production Executable

- As experience is gained with Run 2 detector and analysis, physics algorithms mature and stabilize
- Plan to incorporate stable algorithms into production executable, to
  - Reduce coding effort for individuals
  - Reduce user analysis CPU and duplication
  - Standardize calibrations, efficiencies, backgrounds
- Examples for Winter 2004 release:
  - COT-based cosmic ray finder
  - Improvements in Silicon Layer00 reconstruction
  - Improvements in Time-of-Flight reconstruction

Summary

- CDF computing has been able to satisfy the needs of the CDF physics program
- Preparation for future demands means that we must continue to improve
- Upgrade plans based on
  - Staged deployment of SAM and GRID tools
  - Take advantage of considerable synergies to improve operational efficiency