Data Preservation at the Fermilab Tevatron

Bodhitha Jayatilaka, Ken Herner, Joe Boyd, Willis Sakumoto
DPF2015 - Ann Arbor, MI
August 6, 2015
Introduction: Tevatron Run II

- 1.96 TeV proton-antiproton collider
  - Two general purpose experiments
  - CDF and DØ
- Ceased operations 30 Sep 2011
  - ~12 fb\(^{-1}\) (~10 fb\(^{-1}\) recorded)/expt
- Unique data
  - Unique initial state vs LHC
    - Asymmetry measurements, flavor physics
  - Multiple energy collisions
    - 300 GeV and 900 GeV in addition to 1960 GeV
    - “Legacy” precision measurements (e.g., \(M_W\), \(m_t\))
- Continued physics output
  - Expect long tail at both experiments
  - ~20 papers at each experiment in 2014
The Tevatron Data Preservation Project

- **Goal:** Complete analysis capability (DPHEP “level 4”) through Nov 2020 (SL6 EOL)
  - Includes ability to generate and simulate new MC
  - All necessary documentation is preserved and accessible
  - Collision data on tape remains accessible
  - Computing environment for analysis is available and accessible

- Funded project for two years (2012-14)
  - Seek common solutions between experiments whenever possible
Preserving the data

- Each experiment has ~10 PB of data on tape
  - Includes collision data and simulation
  - Migrate to current media (most recently LTO4->T10KC)
- CDF Italian institutions are migrating a subset of data (raw+ntuples) to CNAF
  - Using GARR (Italian R&E) network. Copy is now complete (~4PB)
- Non-statistical data (e.g., calibrations) for both experiments are in Oracle DBs
  - Continue to upgrade Oracle servers as needed for security needs
Preserving the computing environment

Up to September 2011

- Raw Data
  - Inclusive
  - H-Pt Leptons
  - Jets
  - Hadronic B

- Data Quality Monitoring
  - Calibration pass
  - Calib dataset

- Concurrent with data taking

- Run periodically over “run period”
  - Calibration DB
  - Calibration procedure

- Reco / split
  - Electrons
  - Muons
  - 40 GeV Jets
  - Had Bs

- N-tupling
  - Prod Data
  - Root N-tuples
  - Stn
  - Tnt
  - Bstn
Preserving the computing environment

Raw Data
- Inclusive
- H-Pt Leptons
- Jets
- Hadronic B

Concurrent with data taking
- Data Quality Monitoring
- Calibration pass
- Calib dataset

Run periodically over “run period”
- Calibration DB
- Calibration procedure

Today
- Analysis
  - Electrons
  - Muons
  - 40 GeV Jets
  - Had Bs
  - Prod Data
  - N-tupling
  - Stn
  - TNT
  - Bstn
  - Root N-tuples

Prod Data
- N-tuples
- Stn
- TNT
- Bstn
Preserving the computing environment

Data to preserve
Code to preserve

Today

Raw Data
- Inclusive
- H-Pt Leptons
- Jets
- Hadronic B

Concurrent with data taking
- Data Quality Monitoring
- Calibration pass
- Calibration DB

Calib dataset

Run periodically over “run period”
- Calibration procedure

Electrons
- Muons
- 40 GeV Jets
- Had Bs

Prod Data

N-tupling

Root N-tuples

Stn
- TNT
- Stn

Analysis

Bo Jayatilaka | Data Preservation at the Fermilab Tevatron
Adapting Intensity Frontier computing tools

• The post-Tevatron landscape of experiments at Fermilab are largely smaller experiments (e.g., neutrino detection, dark matter searches)
  - Generally described as “Intensity Frontier” (IF) experiments
  - Impractical for each experiment to develop own computing infrastructure

• FIFE project aims to develop common tools and infrastructure
  - Already utilized by a majority of current FNAL-based experiments

• FIFE toolset includes many that Tevatron experiments can adapt to replace deprecated ones (due to OS/architecture/security reasons)
  - Data access protocol (“SAMWeb”)
    • Modernized (http-based) version of the protocol originally used by CDF/D0
  - Job submission tool for grid computing (“JobSub”)
  - Code distribution (CVMFS repositories)
The software

- At the time of shutdown both CDF and D0 used 32-bit frameworks built on Scientific Linux 5 (but with compatibility libraries to older OSs)
  - CDF Plan: build legacy release that contains no pre-SL6 libraries
    - Build and test completely on SL6, drop support for all previous releases
    - Release now available for general use
  - D0 Plan: retain compatibility libraries but test running on SL6
    - Many ancillary tools dropped in favor of common ones
    - All testing completed and validated

- Both experiments move to CVMFS for code distribution
  - User setup scripts modified so that this change is invisible

- Interactive compute resources (including for building code) all moved to virtualized nodes maintained by FNAL Computing Sector
  - No new dedicated hardware for Tevatron experiments
CDF Distributed Computing

- CDF maintained a dedicated cluster “CDFGrid” at FNAL for ~10 years
  - Also an OSG site serving opportunistic hours to other VOs
  - Was the backbone of CDF computing in Run 2
- CDFGrid shut down on 1-Mar-15 and all CDF jobs now sent to “GPGrid” cluster (shared with other FNAL experiments)
  - Job submission tools for IF experiments implemented with CDF wrapper
  - End-user still runs the same commands
  - Effective rate of computing use unchanged
D0 Distributed Computing

• D0 has a PBS-based job submission system
  - Most user analysis goes to the “Central Backend” (CAB)
    • Expect full CAB retirement in late 2015
  - Simulation workload has dedicated grid entry point (SAMGrid)
    • Workflow being ported to FNAL GPGrid

• Dedicated nodes (in CAB and SAMGrid) that are retired are not replaced

• Adopt IF job submission infrastructure to use GPGrid
  - Required modifications to submission scripts on interactive nodes and to internal software environment
  - Dedicated D0 storage elements not mounted on these worker nodes (now accessible via IF data-handling tools)
  - Input file delivery, previously via dedicated caches has changed to using IF tools
Documentation

• Large internal (~10k for CDF, ~6k for D0) note catalogs
  - Migrate internal note repositories to Inspire
    • Both migrations complete
  - Early notes only exist on paper
    • Large effort to scan these (completed)
• Much remaining documentation is a patchwork of webpages
  - Keep as much of this as possible
  - Twiki/wiki pages converted to static HTML
• ~950 thesis records for both experiments
  - Electronic versions maintained at Fermilab for nearly all

Tevatron Impact Symposium, 6/11/2012

Parameters of Colliding Beam Detector

R. Diebold, A. Tollestrup, T. Collins, S. Ecklund, J.K. Walker

Constraints

1. Kissing scheme for beams.
2. Maximum energy of MR/ED beams = 150/1000 GeV for pp and 1000/1000 for p
3. Conventional magnets for the MR normal operation.
4. Low $\theta$ achieved without loss of part of the 50m long straight.

(Note: R. Diebold paper #1 of Summer Study assumes 46m is free.)
Conclusion

• Tevatron Run 2 data preservation Project completed
  - Some work still being done at the experiment level to ensure data access and usability
  - Maintain computing resources necessary to allow physicists to access and use CDF and D0 data for physics analysis through 2020
  - Practical future use case: *If a discovery is made at the LHC that in hindsight could be confirmed with Tevatron data, could we make such a confirmation?*

• Physics output from the Tevatron experiments continues
  - Multiple talks at DPF2015
  - In many cases already using the DP infrastructure