



Data Preservation at the Tevatron

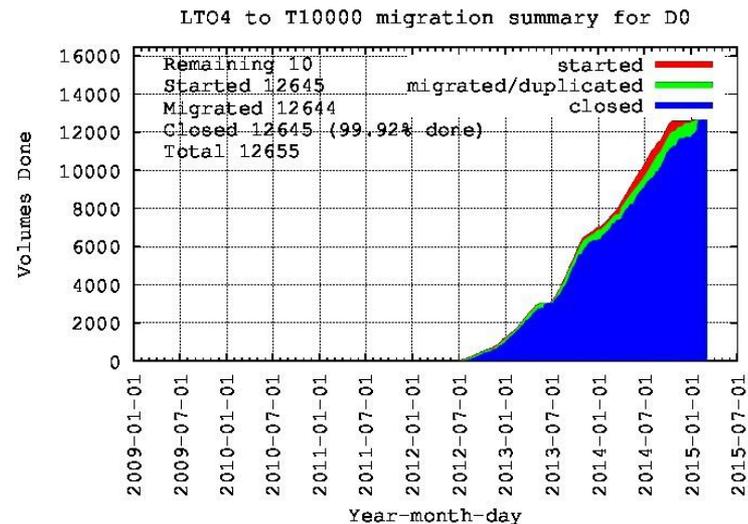
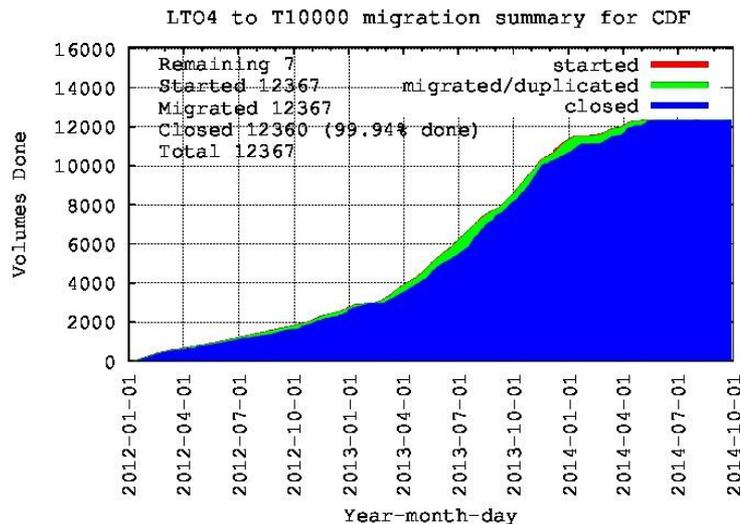
Ken Herner, Mike Hildreth, Bo Jayatilaka for the R2DP Project
DPHEP Meeting
14 Mar 2017

Run II Data Preservation Overview

- Overall project goal was DPHEP Level 4 preservation of both CDF and D0 through at least 2020 (end of SL6)
 - Efforts began in 2011 within each experiment
 - Fermilab SCD established R2DP project in 2012 to accomplish specific pieces of the experiments' programs; ran 2013-2015
 - Dedicated experts from CDF and D0, Fermilab SCD technical lead, and project manager
 - Several efforts from the experiments outside of project scope as well
 - **Budget pressures may move the end date forward**
- Technical work to change experiment code bases, move to CVMFS, running on IF computing infrastructure complete
- Some work pushed back, but now has to be done due to EOL for SL5

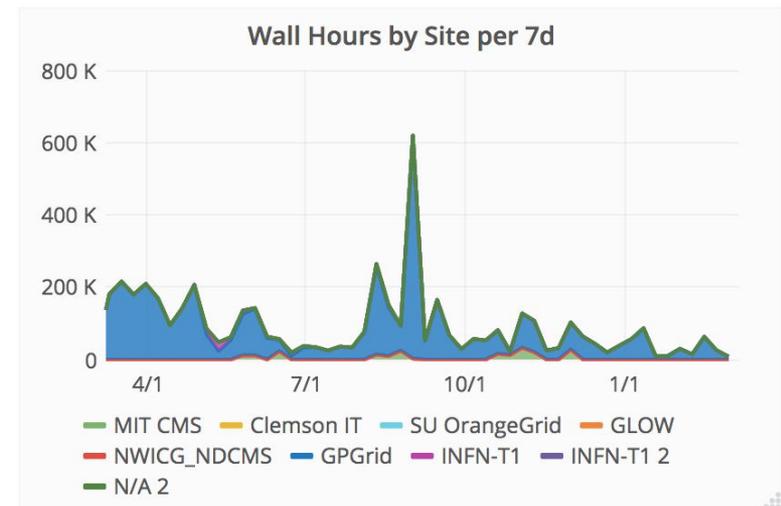
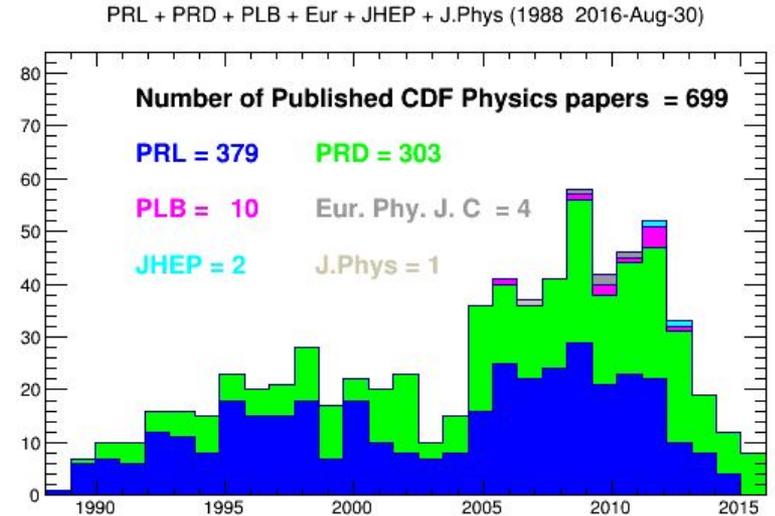
Tape Migration

- Migrated all raw data, reconstructed data, simulation to from LTO4 to T10K tapes-- expected to be readable through 2020
 - roughly 10 PB/experiment
 - Copy of CDF raw data to CNAF complete
- With T10K going away; FNAL is looking at other technologies such as LTO8
 - Current planning **does** include an additional migration to the new technology



Current CDF Activities

- Ongoing activities
 - W mass measurement with full dataset (expected to be world's best)
 - Final top mass measurement
 - Other SM and flavor physics measurements
- 8 publications in 2016
 - Several others submitted
- Jobs running on FNAL GPGrid
 - Completely transitioned in 2016
 - 4.8M hours in the past 12 months



Current D0 Activities

- Presently perhaps 30 collaborators active
 - Mostly top, QCD, and electroweak analysis now
 - 6 Ph.D. and 2 undergraduate students still active
 - 10 publications in 2016; 2 so far in 2017
- Ability to run jobs on FNAL GPGrid exists, but D0 has chosen to continue using dedicated farm (no longer doing hardware maintenance/replacement)
 - Still several million CPU hours in 2016
- Expect some activity to continue to Sep. 2018
 - Final Tevatron top combinations
 - Lingering QCD analyses
 - Final W mass measurement + CDF combo

New Technical Work

- Want to move CDF and D0 SAM (file delivery system and metadata catalog) backends off of Oracle and to Postgres
 - **Initially studied in 2013, but cost-benefit analysis said staying put was right then. Costs to maintain system have significantly risen since then, however.**
 - All other FNAL expts. using Postgres backend now
 - Work to move to web interface already done.
 - Oracle -> Postgres SAM migration tested for other expts. with copy of D0's DB, so procedure is well understood
- SL5 End of Life: 31 March 2017
 - D0 MC production at Lyon will cease working: last non-FNAL MC generation site
 - D0 provided effort to migrate MC production chain to FNAL SL6 machines

About that paper we promised...

- We finally got it ~~finished~~ published!
- Describes the overall goals of the project, the work we ended up doing, how the system is performing now
- NIM A Review process went rather well
- S. Amerio et al., "Data Preservation at the Fermilab Tevatron", Nucl. Instrum. Methods Phys. Res. Sect. A, 851, 1 (2017).
 - DOI: <http://dx.doi.org/10.1016/j.nima.2017.01.043>
 - [arXiv:1701.07773](https://arxiv.org/abs/1701.07773)
- Hopefully it helps pave the way for other such papers

Summary

- 5.5 years after shutdown, the Tevatron experiments are still publishing
 - Several more publications expected in 2017; few more PhDs to award
 - Expect some physics effort from both experiments to continue through FY 2018
- There is a little new technical work to migrate SAM DB servers and update final SL5-based pieces to SL6
- NIM Paper describing Tevatron DP effort published
- We hope that other Fermilab experiments can learn from this effort as they mature

Aside: other Fermilab experiments

Newer experiments are of course following DOE guidelines (must have a “Data Management Plan”, etc.) but no other major efforts of this type ongoing right now

Lessons learned are certainly of value to all IF experiments

Important to stay flexible and consider that even things you thought were “done” can change later (e.g. Tevatron SAM backend)

Discussions about having some generic data preservation lessons learned talks for the other experiments at some point