Hall D Offline Software Performance and Status

12 GeV Software Review III
February 10, 2015
Mark Ito

High-Level Software Packages

Function	Package	"Locally" Grown?
Raw Data Format	EVIO	yes
Offline Data Format	HDDM (compressed XML, REST a variant)	yes
Geometry Specification	HDDS (an XML)	yes
Data Acquisition Framework	CODA	yes
Simulation Engine	GEANT 3	no
Event-Based Processing Framework	JANA	yes
Event Simulation and Event Reconstruction	sim-recon (sim: GEANT 3, recon: JANA)	yes
High-Level Event Analysis	Analysis Library (part of sim-recon)	yes
Histogramming, Fitting, etc.	ROOT	no
Amplitude Analysis (PWA)	AmpTools	yes

Low-Level Software Packages

Function	Package
XML Parsing	Xerces-C
Source Code Management	Subversion
Build System	SCons (some GNU Make)
Scripting	Python (some legacy Perl)
Database	MySQL/MariaDB and SQLite
Web Authoring	MediaWiki (mostly)
Data Transfer (WAN)	SRM, Globus Online

Computing Performance Example

Computing task times: assume a 10,000 core Haswell farm.

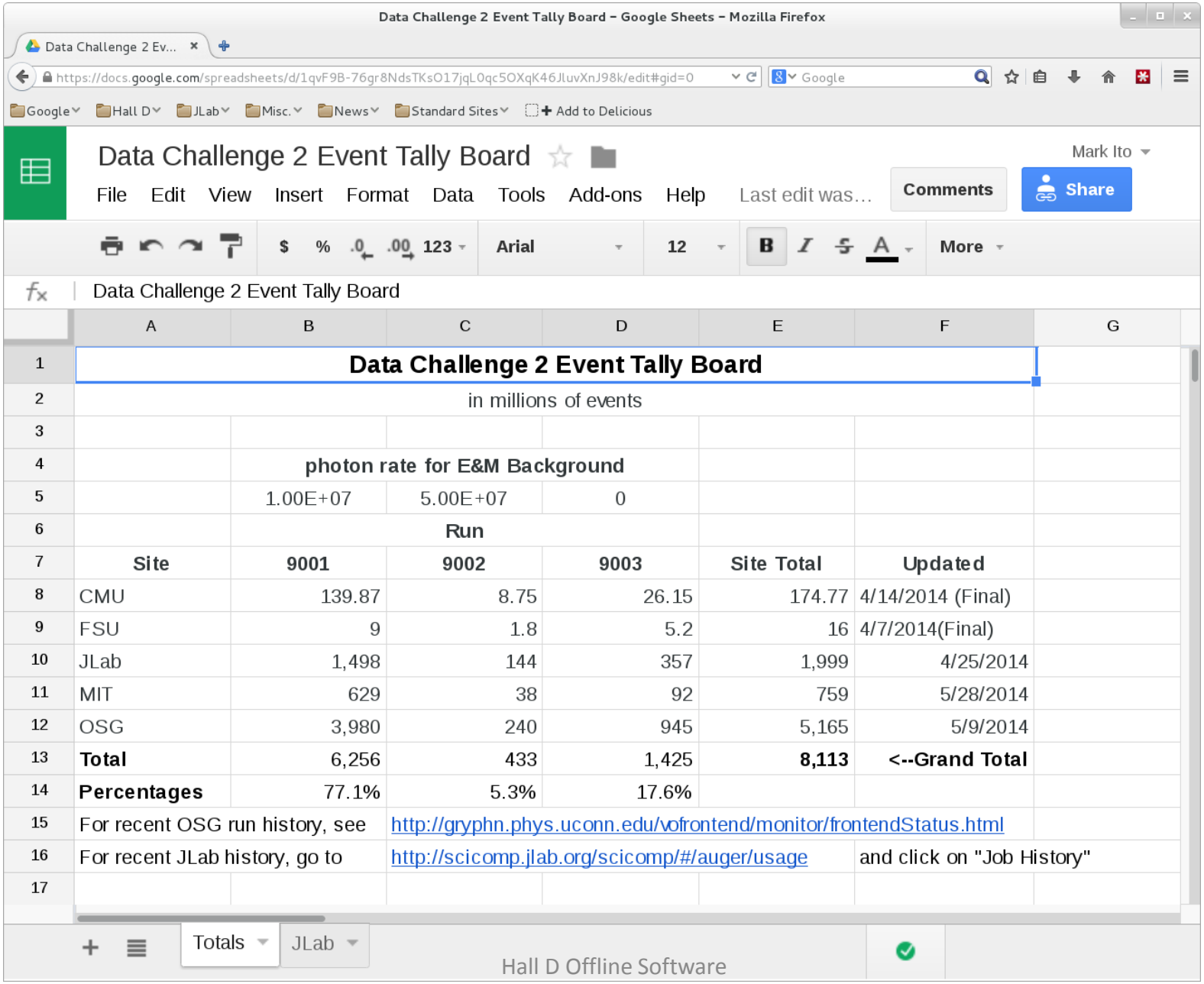
Fiscal Year	2015	2016	2017	2018	2019
Weeks of running	2	16	25	18	22
Trigger Rate (kHz)	2	20	20	20	20
Number of events (billions)	1.2	97	151	109	133
Reconstruction time (days)	0.06	5.1	8.0	5.7	7.0
Recon.+Sim. Time (days)	0.6	50.9	79.5	57.2	70.0
Total data to tape (PB)	0.05	4.0	6.3	4.5	5.6

Computing Requirements Discussion

- Reconstruction time reduced by a factor of 3 with new Haswell chip relative to assumption of last estimate.
- Running time per year has decreased
  - Peak of 25 weeks in FY17 vs. 35 weeks in last estimate.
- Simulated data creation time has gone up factor of 7 relative to reconstruction since last estimate.
  - Original estimate not well documented
  - Experimental resolution inclusion: much slower now.
  - Improvement likely possible, unimproved number used.
- Commissioning data: event reconstruction rate 50% better than that on simulated data, but…
  - “junk” events not accounted for yet (event fraction, recon. rate)
  - Beam-spectrum/trigger not realistic (different center-of-mass energy)
  - Code base slightly different (for commissioning geometry, recent improvements)

Data Challenges (DCs)

- DC1 - December 2012/ January 2013
  - 5 billion Events - OSG, JLab, CMU
  - 1200 Concurrent Jobs at Jlab
  - Data produced used to support proposal for GlueX Phase IV running
- DC2 - March/April 2014
  - 10 billion events with EM backgrounds included - OSG, JLab, MIT, CMU, FSU
  - 4500 Concurrent Jobs at JLab
  - 11,000 Concurrent Jobs on the OSG
  - Well under 0.1% failure rate
- DC3 - January/February 2015
  - Read data in raw-event format (EVIO) from tape and produce DST format (REST) files.
  - JLab only
  - Test throughput from Tape Library
  - Run Multi-threaded jobs

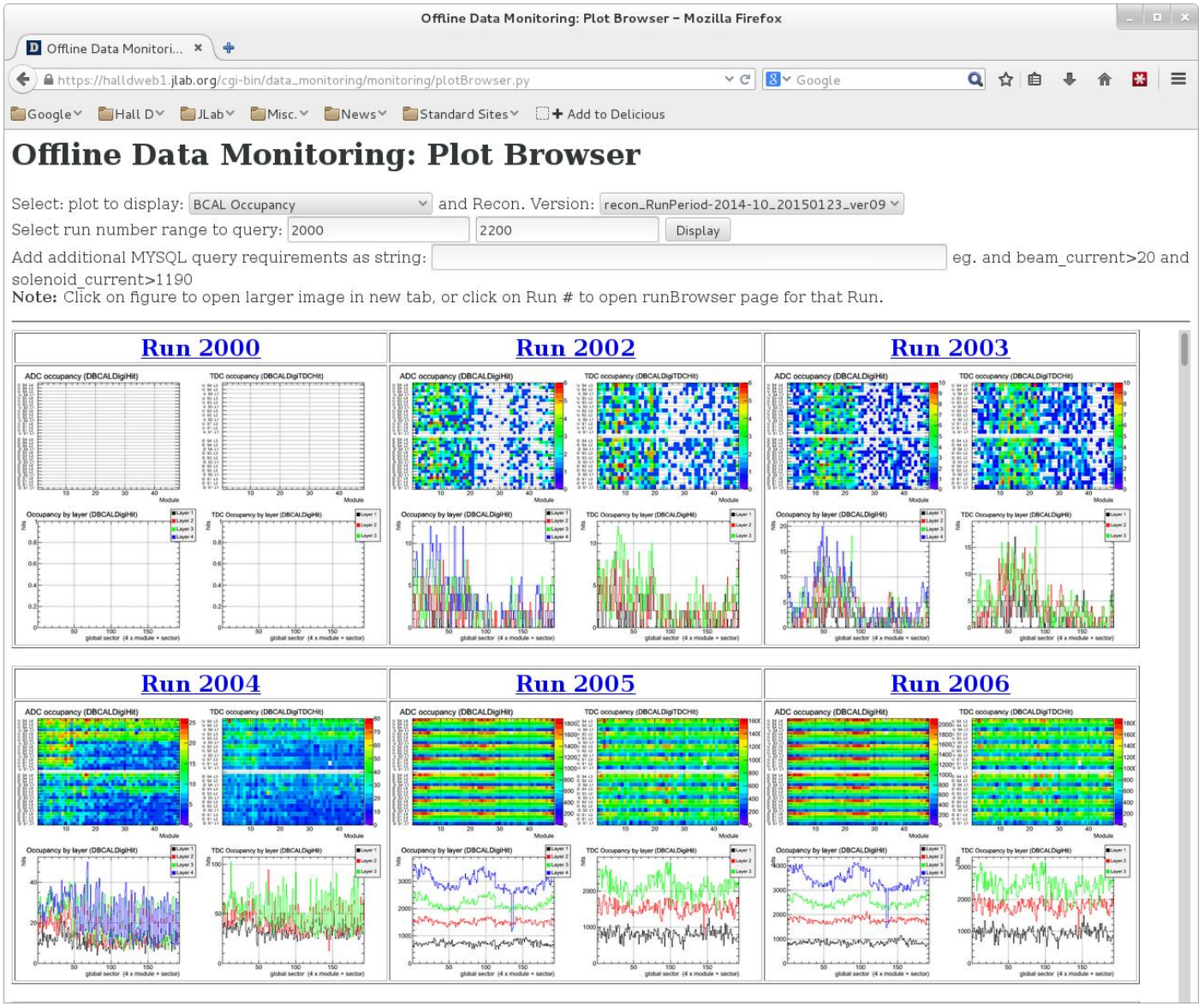


	A	B	C	D	E	F	G
1	**Data Challenge 2 Event Tally Board**						
2	in millions of events						
3							
4		**photon rate for E&M Background**					
5		1.00E+07	5.00E+07	0			
6		**Run**					
7	**Site**	**9001**	**9002**	**9003**	**Site Total**	**Updated**	
8	CMU	139.87	8.75	26.15	174.77	4/14/2014 (Final)	
9	FSU	9	1.8	5.2	16	4/7/2014(Final)	
10	JLab	1,498	144	357	1,999	4/25/2014	
11	MIT	629	38	92	759	5/28/2014	
12	OSG	3,980	240	945	5,165	5/9/2014	
13	**Total**	6,256	433	1,425	**8,113**	**<--Grand Total**	
14	**Percentages**	77.1%	5.3%	17.6%			
15	For recent OSG run history, see	http://gryphn.phys.uconn.edu/vofrontend/monitor/frontendStatus.html					
16	For recent JLab history, go to	http://scicomp.jlab.org/scicomp/#/auger/usage				and click on "Job History"	
17							

Offline Monitoring during Commissioning Run

- Weekly reconstruction pass through data
  - All data as of Friday afternoon
  - Online monitoring plots reproduced
  - Full reconstruction with updated code and constants
  - Skims of raw events done for calibration
  - DST data produced (REST format)
- Web site for browsing results
- REST data good enough to observe multi-particle final states
- Continuing on a bi-weekly schedule
- Paul Mattione (CMU), Kei Moriya (ASU), Justin Stevens (MIT), Sean Dobbs (NU)

Plots for a selected range of runs



Plots for a given run

Offline Data Monitoring: Run Browser – Mozilla Firefox

Offline Data Monitori...

https://halldweb1.jlab.org/cgi-bin/data_monitoring/monitoring/runBrowser.py?run_number=2205&ver=ver09

Google | Hall D | JLab | Misc. | News | Standard Sites | Add to Delicious

recon_RunPeriod-2014-10_20150123_ve

Display

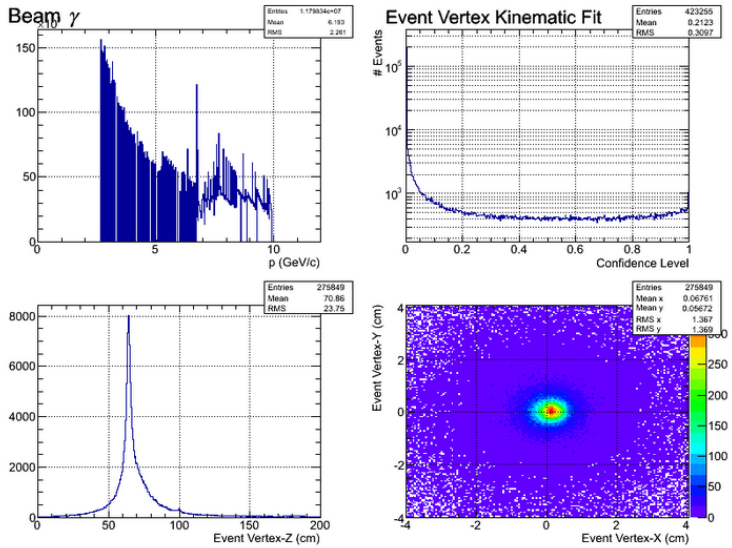
Link displays plots or ROOT opens file

- Nov 05 (Run 940-940)
- Nov 06 (Run 941-980)
- Nov 07 (Run 981-997)
- Nov 10 (Run 999-1005)
- Nov 11 (Run 1010-1036)
- Nov 12 (Run 1039-1053)
- Nov 13 (Run 1056-1081)
- Nov 14 (Run 1091-1091)
- Nov 15 (Run 1093-1093)
- Nov 16 (Run 1097-1144)
- Nov 17 (Run 1145-1229)
- Nov 18 (Run 1230-1286)
- Nov 19 (Run 1288-1301)
- Nov 20 (Run 1302-1332)
- Nov 21 (Run 1334-1393)
- Nov 22 (Run 1394-1400)
- Nov 23 (Run 1402-1436)
- Nov 24 (Run 1437-1515)
- Nov 25 (Run 1516-1569)
- Nov 26 (Run 1572-1640)
- Nov 30 (Run 1642-1642)
- Dec 01 (Run 1643-1643)
- Dec 02 (Run 1644-1656)
- Dec 04 (Run 1693-1696)
- Dec 05 (Run 1748-1796)
- Dec 06 (Run 1797-1846)
- Dec 07 (Run 1847-1894)
- Dec 08 (Run 1897-1993)
- Dec 09 (Run 1995-2009)
- Dec 10 (Run 2014-2030)
- Dec 11 (Run 2037-2078)
- Dec 12 (Run 2084-2122)
- Dec 13 (Run 2129-2153)
- Dec 14 (Run 2154-2198)
- Dec 15 **(Run 2205-2231)**
  - 2205 (10046165 events)
  - ROOT
  - 2207 (6302534 events)
  - ROOT
  - 2209 (8556301 events)

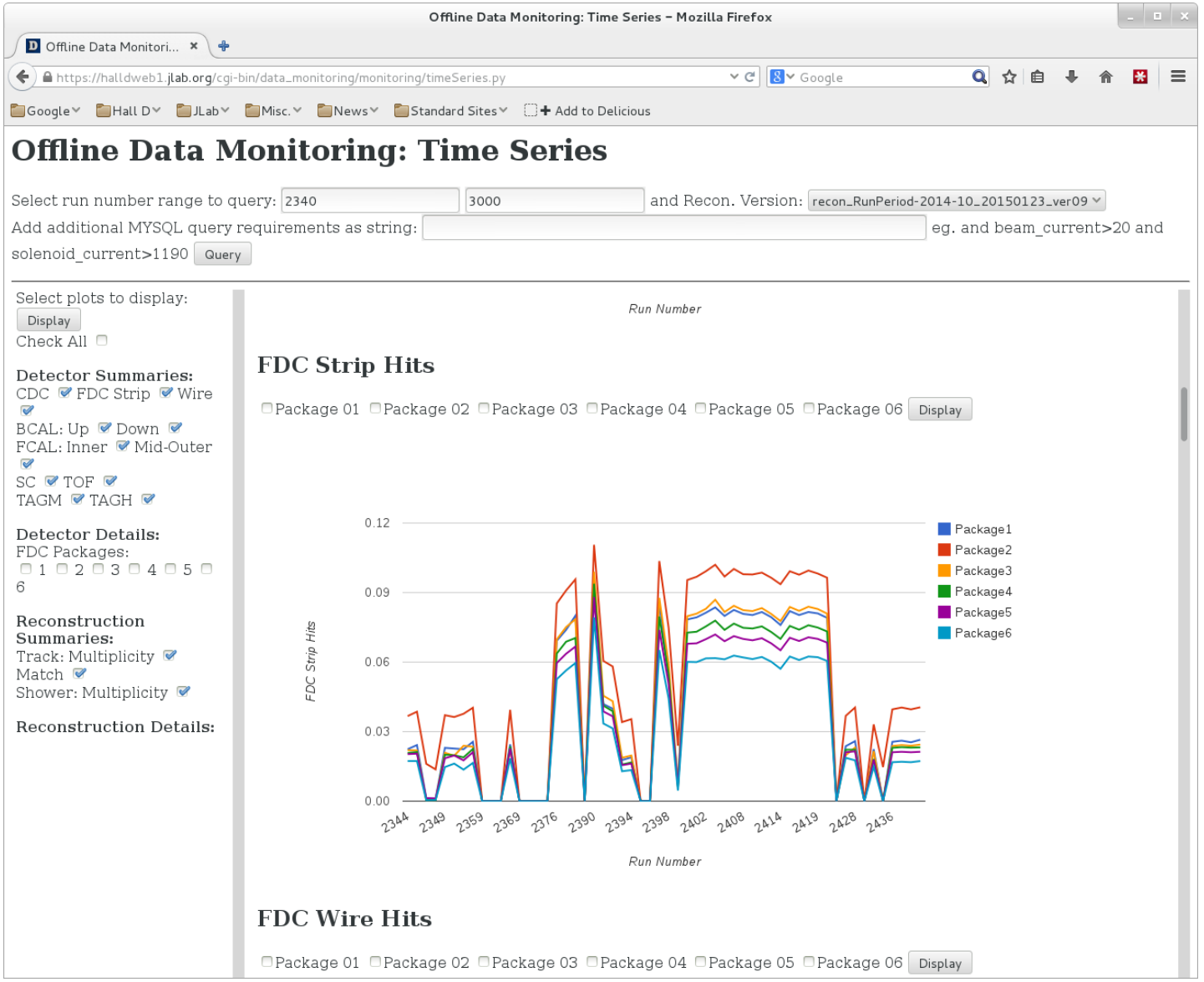
Mouse over light blue entries in table to view histograms, and **click** on an entry to freeze/unfreeze a specific historgram.

CDC:	RawInt	Time	Occupancy	Pedestal	WinDataPedMean	WinDataPedWidth	RawIntVsN	RawTimeVsN	PedVsN
FDC:	FdcStripOcc	FdcWireOcc							
BCAL:	DigiSummary	DigiTime	DigiOccupancy	Cluster	Shower				
FCAL:	DigiPulseInt	DigiOccupancy	DigiTime	HitSummary	HitTime	ClusterEnergyTime	ClusterSpace		
TOF:	Energy	Time	OccupancyPlane1	OccupancyPlane2					
SC:	DigiPulseInt	DigiTime	DigiOccupancy						
TAGM:	DigiPulseInt	DigiMultiplicity	HitOccupancy	HitTime					
TAGH:	DigiPulseInt	DigiTDCTime	DigiPedVsSlot	HitSummary					
RECO:	EventInfo	LLObjects1	LLObjects2	HLObjects	TrackMult	Tracking1	Tracking2	Tracking3	Matching1	Matching2	Kinematics1	Kinematics2
	FCAL1	FCAL2	BCAL1	BCAL2	SC1	SC2	SC3	TOF1	TOF2			

Run 2205: Beam current = 42.0696 nA, Radiator = 2x10-5 RL, Solenoid current = 1199.63 A, Trigger = fcal_bcal_m8.conf



Quantities as a function of run number



Calibration Working Group

- Bi-weekly meetings
  - chaired by Sean Dobbs of Northwestern
- Preliminary list of constants compiled in advance of run, used to guide activity
- Calibration procedures still being developed
- Substantial progress:
  - Basic timing offsets
  - Global energy scale for calorimeters determined
- All collaborating institutions involved.

Calibration Database Experience

- Fully integrated into reconstruction
- Near-complete migration of constants into database
- All detector groups making contributions (no known rogue systems)
- SQLite form of database as alternate to MySQL/MariaDB
  - Complete history and versioning support
  - Solves:
    - Distribution (remote sites, network-challenged computing)
    - Server contention from farm usage
  - Drawback: no automated back-annotation

Data Distribution Methods

- Used OSG SRM to archive DC2 OSG results to JLab Tape Library
- Raw data shipped to CMU using Globus Online

Data Management

- Raw data stays on JLab Tape Library
- Reconstructed REST-formatted data compact
  - Fall 2014 REST data only 113 GB total
  - Keep as much as possible disk-resident at JLab
  - Distribute most (all?) to collaborating institutions
- Data Catalog/Tracker needs development
  - Existing package?
  - Develop one?
  - Dependence on details of how data distributed

Online Conditions Database

- Database to store online run conditions, e. g., magnet current settings, configuration files, etc.
- Two were deployed
  - One looks great (API, modern web interface)
  - One was useful (hand work required, HTML-based web interface, CSS-free)
  - Effort underway to consolidate and expand.

Things To Do

- Real data exposed areas requiring further work:
  - Online Conditions Database
  - Data Catalog
  - Version Management
    - Made branch to deal with non-standard detector configuration
    - branch-to-trunk merge problematic
- Geant4: need to pick up the pace of development
- Still need profiling discipline!?
  - Simulation speed has decayed
- Code review system needs design and implementation
  - Size of collaboration makes this a challenge
- Never-ending data challenge never got started

Summary

- Software performed all critical tasks needed to support detector commissioning and to see physics signals.
- The collaboration feels that we reached a number of milestones that we did not expect to see until the April 2015 run.
- Many basic tasks lie ahead: calibrations, monitoring, reconstruction Q/A.
- Software infrastructure to support these activities largely in place.