

# CDF Production Farms

Yen-Chu Chen, Roman Lysak,  
Stephen Wolbers

May 29, 2003

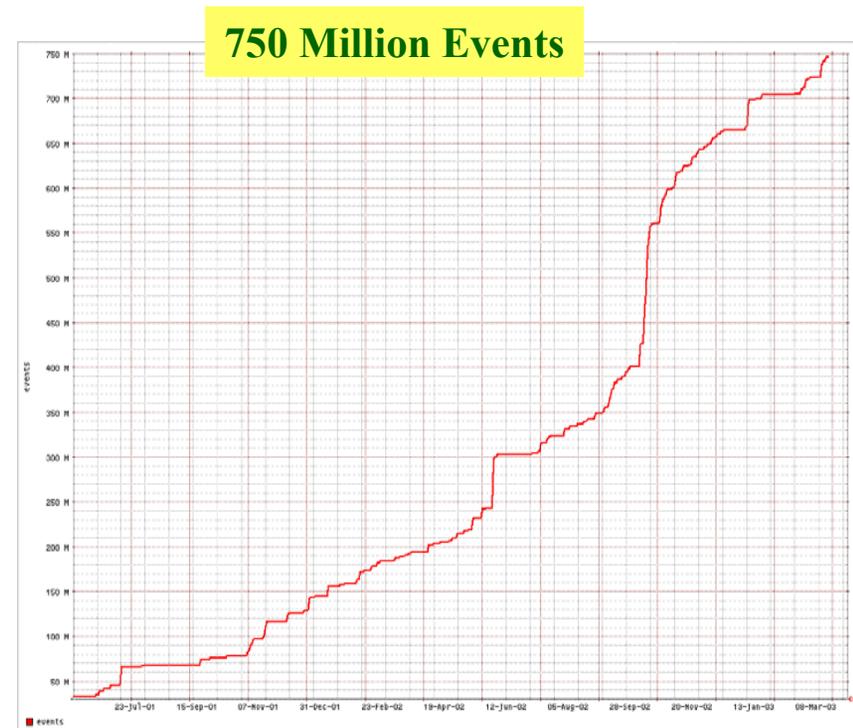
# The CDF Production Farms

- 151 PCs (317 GHz)
  - 16 of the 151 used for I/O
- Servers
- Disk Pool (dfarm - 13 TB)
- Network Switch + connections
  - Most 100 Mbit
  - 16 Gbit connections for the I/O PCs
- Software to control everything.
  - Most Important!
  - Fairly complicated system
    - 8 input streams, 35 output datasets
    - Thousands of files/day - kept in order
- People to keep it all going
  - Roman Lysak and Yen-Chu Chen
  - Lots of help from Pasha, Liz and Dmitri!



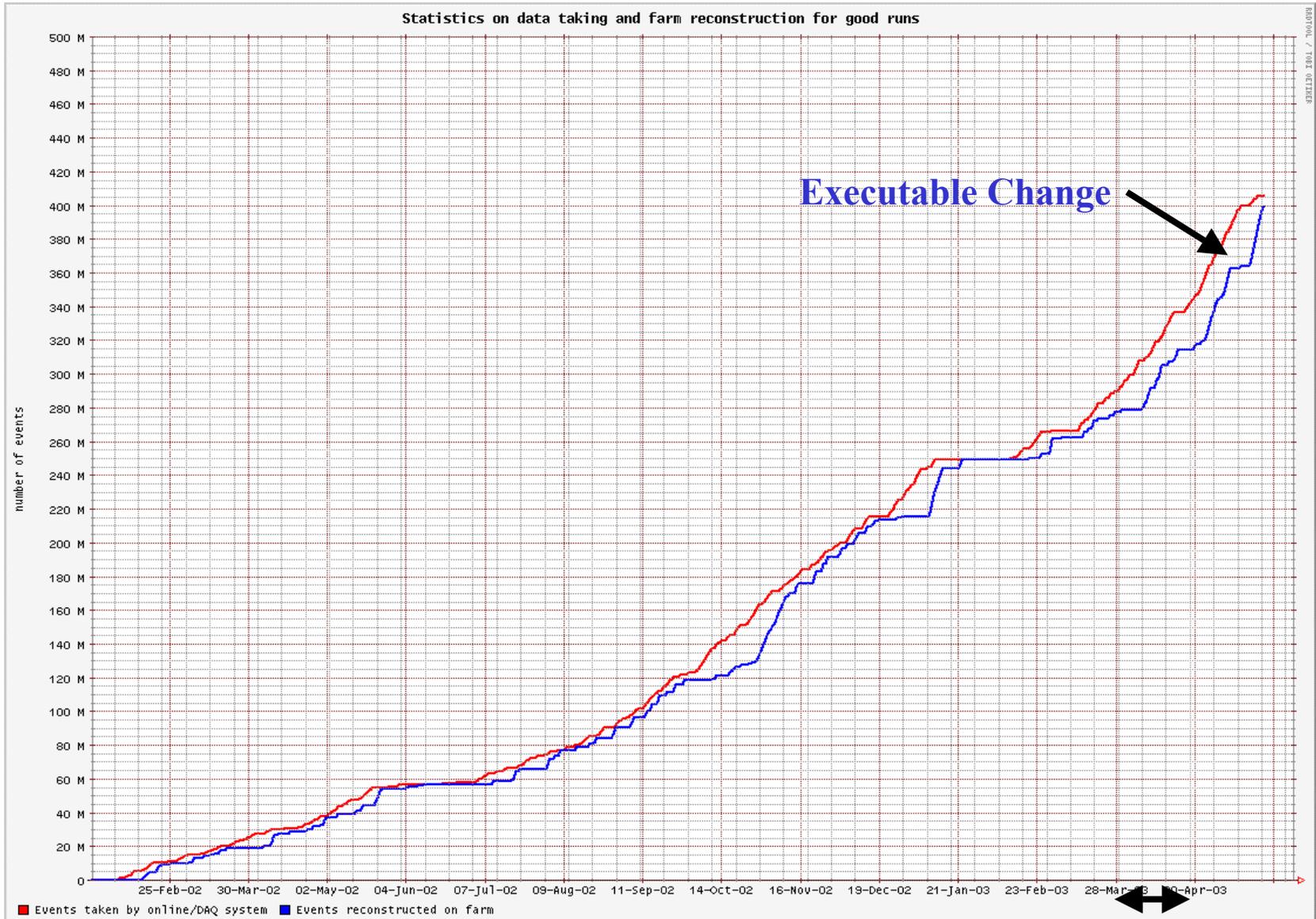
# Very Short History (up to end of March)

I (Commissioning)	3.11.0g	9,775,297
I (Commissioning)	3.12.0	9,866,564
II (1x8)	3.14.0	2,539,217
III (36x36)	3.15.0c	5,113,927
III (36x36)	3.16.0	5,113,927
IVa (June, 2001)	3.17.1	12,322,465
IVb (June, 2001)	3.18.0	21,993,586
V (August, 2001)	4.0.0i	9,882,486
V (August, 2001)	4.1.0	38,857,497
V (August, 2001)	4.2.0a	8,685,950
VI (December, 2001)	4.2.0b	6,445,760
VI (December, 2001)	4.2.0c	26,944,076
VI (December, 2001)	4.3.1	14,519,033
VII (2002 data)	4.3.1b	10,369,500
VII (2002 data)	4.3.2	10,109,874
VII (2002 data)	4.3.2a	38,690,220
VII (2002 data)	4.5.2	125,161,272
VII (2002 data)	4.8.0	16,572,810
VII (2002 data)(through Jan,2003 shutdown)	4.8.4a	257,886,088
TOTAL		630,849,549



**Includes  
Reprocessing**

# Recent Processing



Tape Problems

# Daily Reports

- Updated status of the farms as of :
  - 8:00 A.M. 5/29/03
  - Behind by approx. 6.3 Million on Raw Data
  - 7.5 Million total events processed in last 24 hours
  - 6.6 Million raw events through 4.8.4g1
  - 0.9 Million gjet08 processed through 4.10.4e
- Issues:
  - 1. 4.10.4e running. Output going to tape.
  - 2. Caught up to calibrations on most streams.
- Plan for the day:
  - 1. Process raw data with 4.8.4g1.
  - 2. Run streams to catch up to calibrations with priority on B and G.
  - 3. Run 4.10.4e reprocessing.
- -----

# Daily Reports

- | Stream   | Run    | Date run was taken | Events behind* |
|----------|--------|--------------------|----------------|
| Stream A | 162498 | 5-May-03           | 1,410,665      |
| Stream B | 163431 | 24-May-03          | 524,380        |
| Stream C | 163431 | 24-May-03          | 331,471        |
| Stream D | 162498 | 5-May-03           | 2,119,913      |
| Stream E | 163431 | 24-May-03          | 466,091        |
| Stream G | 163431 | 24-May-03          | 550,608        |
| Stream H | 163431 | 24-May-03          | 683,833        |
| Stream J | 163431 | 24-May-03          | 379,415        |

- \* This is more or less behind raw data on tape. The farms
- cannot process events if the calibrations aren't
- prepared.

- Calibrations (07) 163474 25-May-03

- Current data-taking 163527

## Recent Problems/Issues

- Tape reading problems in late March-late April/early May.
  - Bizarre but extremely disruptive error.
  - Difficult to diagnose and work around.
  - Finally got past the bad set of tapes.
    - May 8 got past the last bad tape.
    - In the end lost about 10 files.
  - The bug was diagnosed by STK and presumably is now fixed.

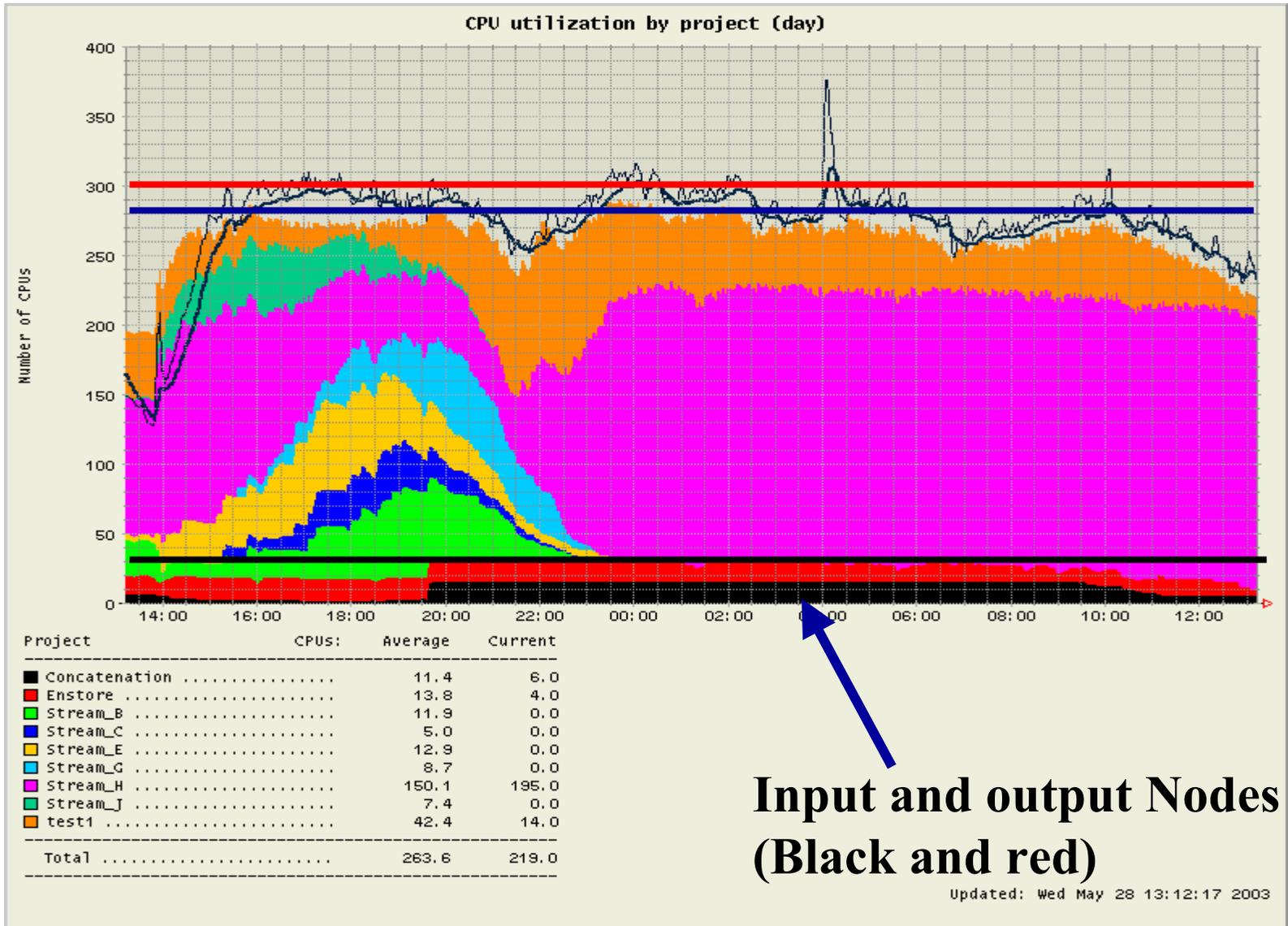
## Recent Problems/Issues

- Reprocessing with 4.10.4
  - This did not go well.
  - Lots of errors, debugging.
  - Lesson for next time: start earlier, check at each step, expect to start more than once, fix and restart.
- Change to 4.8.4g1
  - Tricky, not completely smooth.
  - 4.8.4a was used from October until last Friday.
  - New data required change - change is not trivial.
  - Took a week or so to get this correct.

# Goals/Near Future

- Keep up with the data (Up to calibrations).
- Prepare for compressed data.
  - More events/second!
- Work on changes proposed in processing:
  - Preprocessing stream  $G$ .
  - Validation after concatenation.
  - Reading/writing from/to dcache disk.
  - Gcc compiler.
- Prepare for full reprocessing for Winter '04 conferences.
- Prepare estimate for FY03 farm purchases.
  - Understand farm utilization.
  - Understand ProductionExe CPU time.
  - Model peak and steady-state requirements.
- Documentation!
- Look for more people to work on the farms.
- Good communication with physics groups, physics coordinator, online operations, etc., for setting priorities, deciding what to do on the farms are essential.

# Farms Utilization



## Goals - future

- Prepare long-term farm hardware enhancement/replacement plan.
- Plan for increased L3 output.
  - 20 Mbyte/s now.
  - ?? Mbyte/s in future.
- Other enhancements/changes as needed for CDF.