

CDF Offline Production Farms

Stephen Wolbers
for the CDF Production Farms Group
May 30, 2001

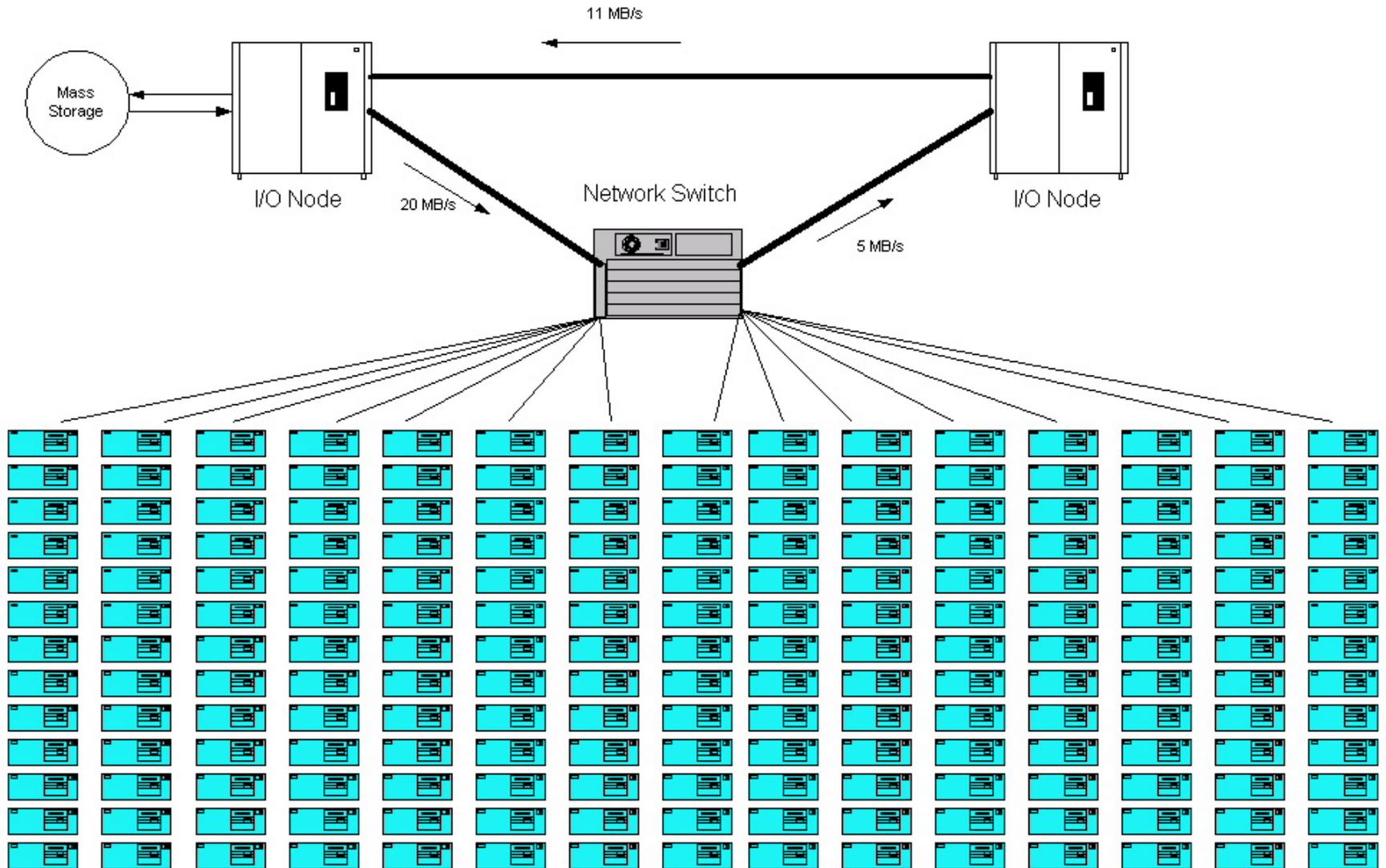
Outline

- Farms Status, Size, Capacity
- Processing History
 - Commissioning
 - 1x8, 36x36
- Ramp up to full data-taking rates
- Monte Carlo Production Possibilities

CDF Run 2 Farms

- 48(50) PIII /500 duals running Linux, 512 MB memory, 42 GB disk, 100 Mbit ethernet
- 40 PIII /800 duals running Linux, 512 MB memory, 50 GB disk, 100 Mbit ethernet
- SGI O2200 I/O node + SGI O2000 I/O node
 - Disk on both and tapedrives on the O2000
- Cisco 6509 switch to connect it all together
- 64 PIII /1 GHz duals have been purchased and they will be in production in July/August.
- Total: 152 duals, equivalent to 240 duals (500 MHz)

Run II CDF PC Farm



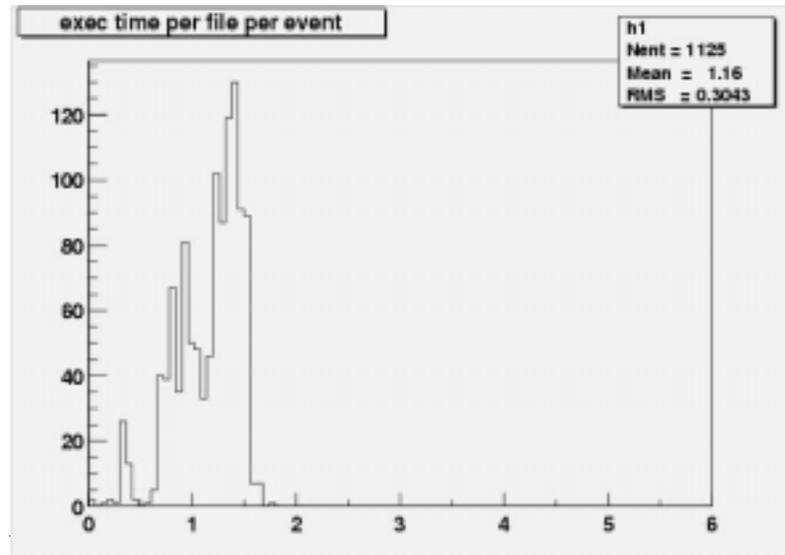


Total Capacity

- Assume 5 s/event on a PIII/500 machine.
 - Full code, fully optimized, R2 luminosity
- At 100% utilization:
 - Events/s = $240 \times 2 / 5$
 - = 96 Hz
- Even if one assumes a more realistic utilization it will be easy to keep up with new data at peak rates.
- In addition there will be reprocessing and Monte Carlo.
- More machines can and probably will be purchased in FY02.

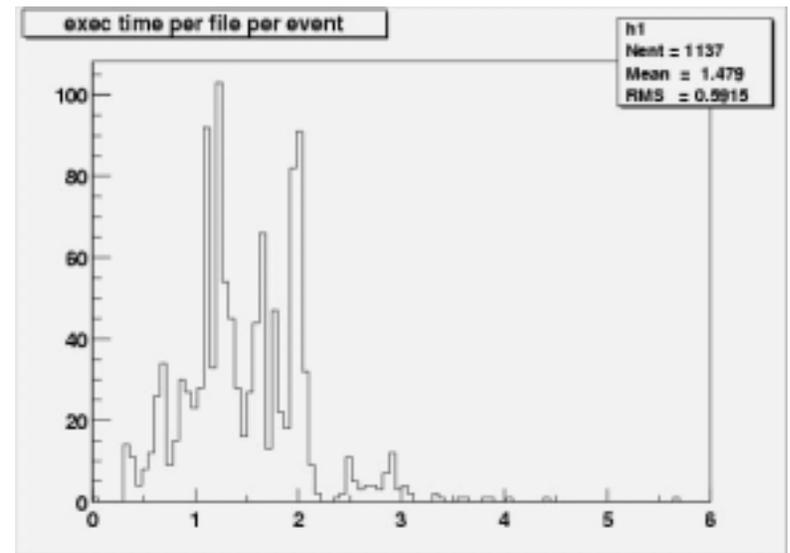
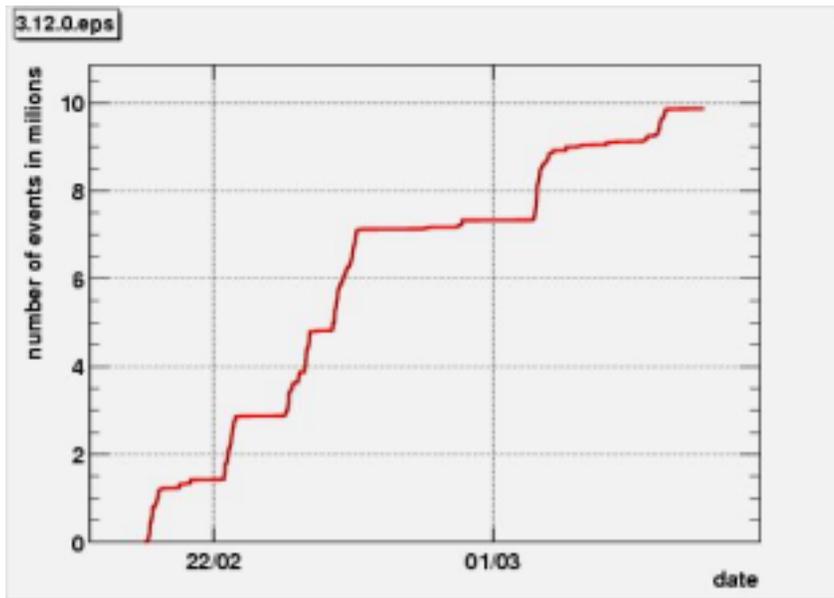
Processing History

- Commissioning Run, first pass: 3.11.0g
 - Was run 4 weeks after the data were collected
 - 9.8 Million Events, 730 GB input, 1080 GB output
 - 179 code crashes
 - CPU/event = 1.2 seconds (PIII /500)



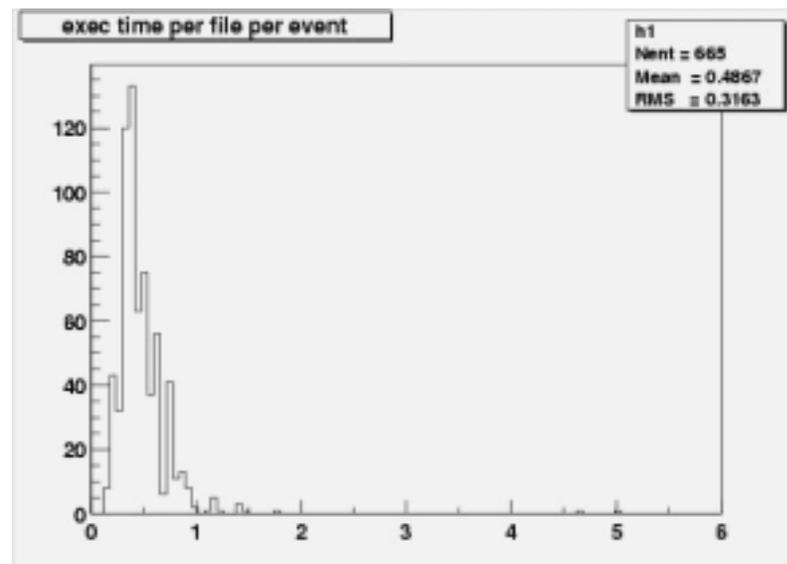
Processing History

- Commissioning Data, second pass: 3.12.0
 - Was run 12 weeks after data-taking
 - 9.9 Million Events, 914 GB input, 1.5 TB output
 - 32 code crashes
 - CPU/event = 1.5 seconds (PIII /500)



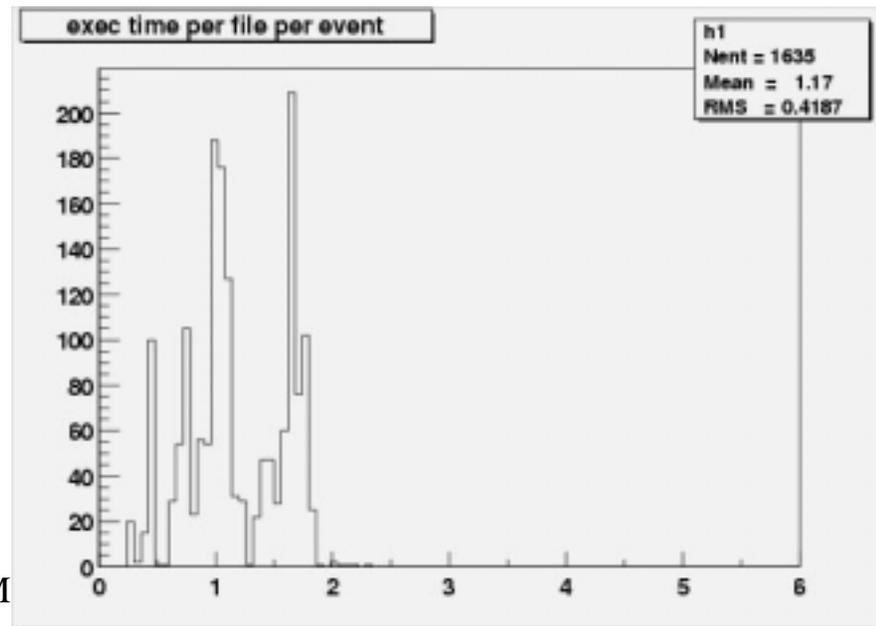
Processing History

- 1x8 Data: 3.14.0_int1_patch
 - Run in quasi-real time, as the data was taken
 - 2.5 Million Events, 400 GB input, 498 GB output
 - 0 crashes (exception handling was off)
 - 4 streams were processed (a,b,c,d)
 - CPU/event = 0.5 seconds (PIII/500)

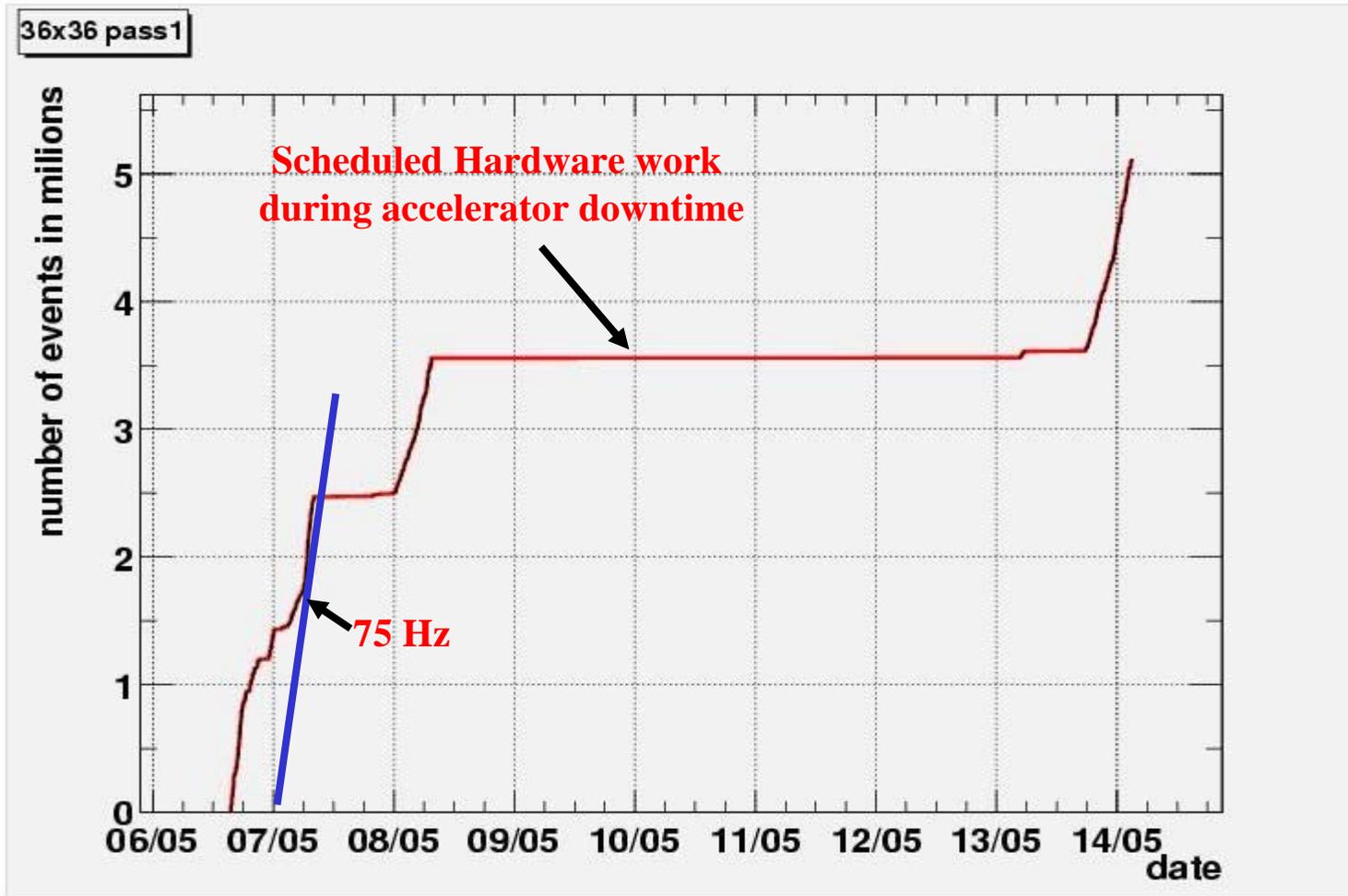


Processing History

- 36x36 Pass 1: 3.15.0c ProductionExe
 - Ran about 1 week after data was taken
 - 5.1 Million events, 1.2 TB input
 - 1.6 TB output
 - 15 code crashes
 - CPU/event = 1.0 seconds (PIII /500)

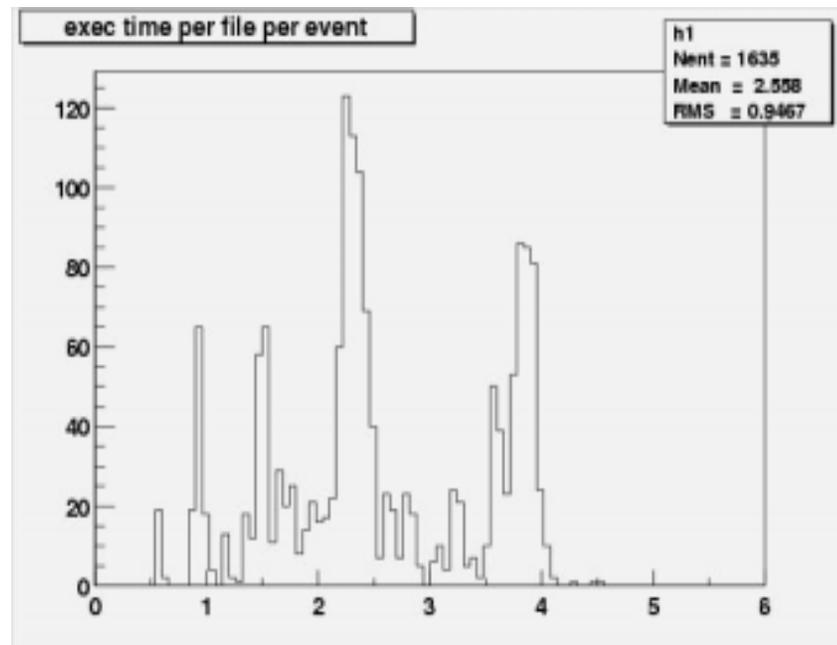
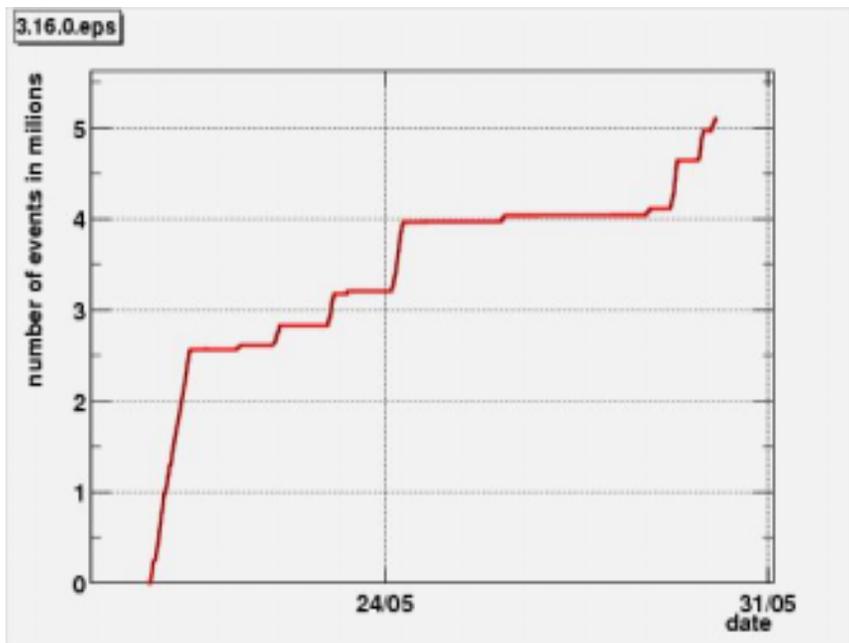


First Pass of 36x36 Production



Processing History

- 36x36 pass 2
 - ProductionExe 3.16.0
 - Started May 20
 - CPU/event = 2.3 seconds (PI II /500)



Processing Summary

- Commissioning 3.11.0g 9.8 Million 1.2 sec/event
- Commissioning 3.12.0 9.9 Million 1.5 sec/event
- 1x8 3.14.0 2.5 Million 0.5 sec/event
- 36x36 3.15.0c 5.1 Million 1.0 sec/event
- 36x36 3.16.0 5.1 Million 2.3 sec/event

- TOTAL 32.5 Million events
- TOTAL CPU 44.7 x 10E6 CPU seconds
- = 2.3 days on current farm

- CPU time/event is affected by luminosity, trigger mix, and code

Issues for Ramp up to full data taking rates

- Hardware
- Software
 - Expressline
- External to farms
- People

Hardware

- Plenty of CPU
- Disk, networking, OK
- Tapedrives and tape I/O - part of DH system
- Overall -- OK

Software

- Farm Software – mostly ready.
- Need to run with multi-streams, input and output.
- Essentially ready, no problems expected.
- Expressline: Will dedicate a portion of the farm to expressline at all times.

External to Farms

- Farms rely on many things:
 - ProductionExe
 - Calibrations
 - Data Handling System
 - Networks
 - Dataset definitions
- The farms group will continue to work with CDF offline, data handling group, CDF L3, etc. to ensure that timely production is assured.

People

- **Production Coordinators (shiffters)**
 - Working well (thanks!)
 - Debugging, checking output, checking logs, checking status, running jobs, etc.
 - Work, responsibilities, will ramp up with the data-taking rate increasing.
 - Look at fnppc.fnal.gov to see the status

How much Monte Carlo?

- The original farm design document (CDF 4810) envisioned a part of the farm for Monte Carlo generation and reconstruction.
- Assume that this is about $\frac{1}{4}$ of the farm, on the average.
- This would allow about 8 Hz of Monte Carlo, assuming that simulation+reconstruction is about 15 s/event on a PIII/500 (700,000 events/day).
 - This compares to 28 Hz average data-taking rate for CDF
 - Degrade by some factor for efficiency
 - Will be reduced if time/event increases
=> 500,000 events/day
- Needs CDF physicist support! Effort to coordinate and verify that the simulation is correct for CDF.

Summary

- Farms CPU Capacity is in place.
- Farms Infrastructure is in place.
- Effort needs to go into smoother operations, better communications, quicker startup, better quality control, better documentation as we head toward full-rate data-taking.
- All of CDF offline has had a big role in making the farms work.