



The CDF Run 2 Computer Farms

Stephen Wolbers

Fermilab

**For the CDF Farms Group and the CD
Farms Group**

September 3, 2001



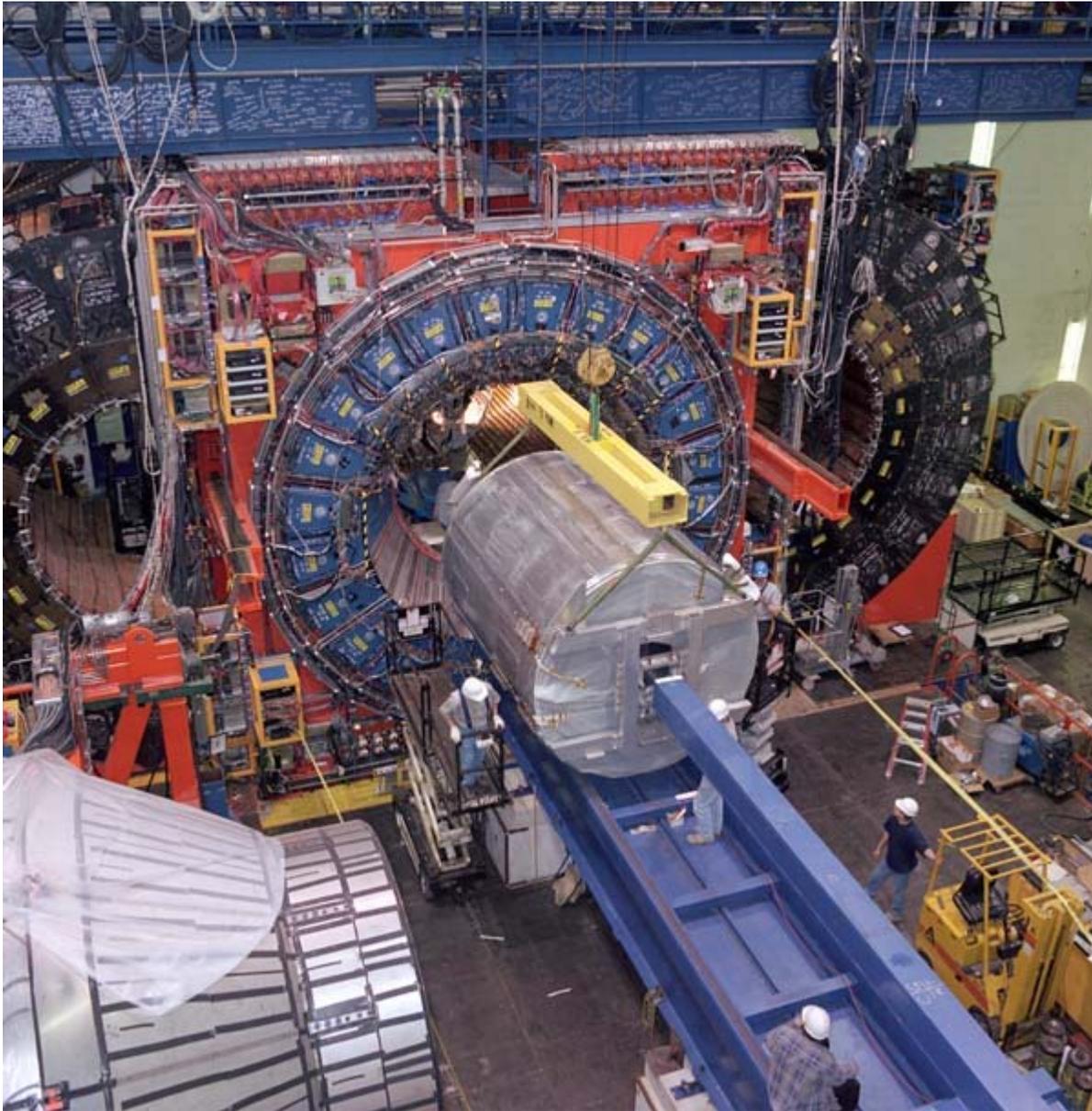
Outline

- Introduction to Run 2 Data Rates/Processing Needs
- Architecture of the CDF Run 2 Farms
- Experience with the farms in Run2
- Future
 - Run 2a
 - Run 2b



Introduction

- CDF Run 2 Data Rates are substantially larger than Run 1 (factor 20 higher).
 - 20 Mbyte/sec to tape peak
 - Well over 100 Tbyte/year to tape
- This data must be processed as quickly as it is collected (with a short time delay for preparing constants or code).
- The data has to be organized into well-defined physics data sets.
- In addition, reprocessing and simulation are also required.



September 3, 2001

Stephen Wolbers, CHEP2001,
Beijing, China



CDF Run 2a Farm Computing

- CPU for event reconstruction of about 5 sec/event on a PIII/500 MHz PC (Each event is 250 Kbyte).
- Assuming 20 Mbyte/sec peak (approx. 75 Hz)
 - Requires 375 PIII/500 processors to keep up
 - Faster machines -> Fewer processors required
 - So 180 PIII/500 duals will suffice.
 - Or 90 PIII/1 GHz duals.
- Requirement is reduced by accelerator/detector efficiency and increased by farms inefficiency.



Run I : Data Volume

Category	Parameter	D0	CDF
DAQ rates	Peak rate	53 Hz	75 Hz
	Avg. evt. Size	250 KB	250 KB
	Level 2 output	1000 Hz	300 Hz
	maximum log rate	Scalable	80 BM/s
Data storage	# of events	600M/year	900 M/year
	RAW data	150 TB/year	250 TB/year
	Reconstructed data tier	75 TB/year	135 TB/year
	Physics analysis summary tier	50 TB/year	79 TB/year
	Micro summary	3 TB/year	-
CPU	Reconstr/event	25 - 65 SI95xsec	30 SI95xsec
	Total Reconstruction	2000-4000 SI95	2000-4000 SI95
	Analysis	2000-4000 SI95	2000-4000 SI95
Access for analysis	# of scientists	400 - 500	400 - 500

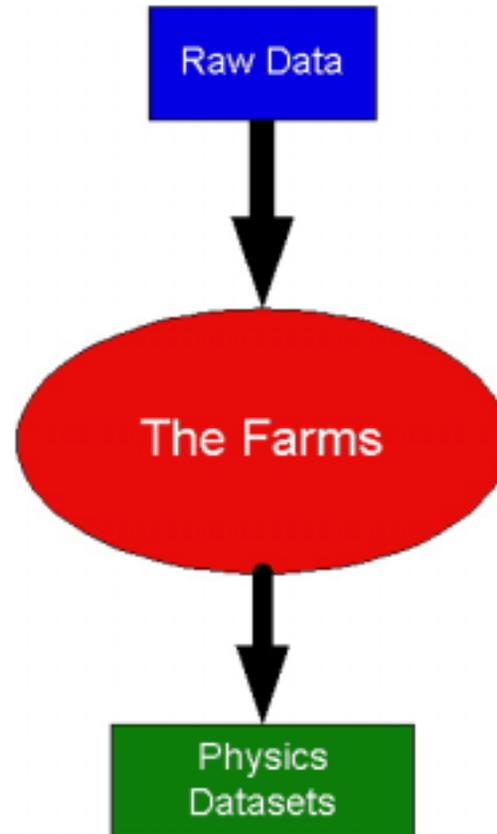


CDF Offline Production Farms for event reconstruction

- The CDF farms must have sufficient capacity for Run 2 Raw Data Reconstruction.
- The farms also must provide capacity for any reprocessing needs.
- Farms must be easy to configure and run.
- The bookkeeping must be clear and easy to use
- Error handling must be excellent.

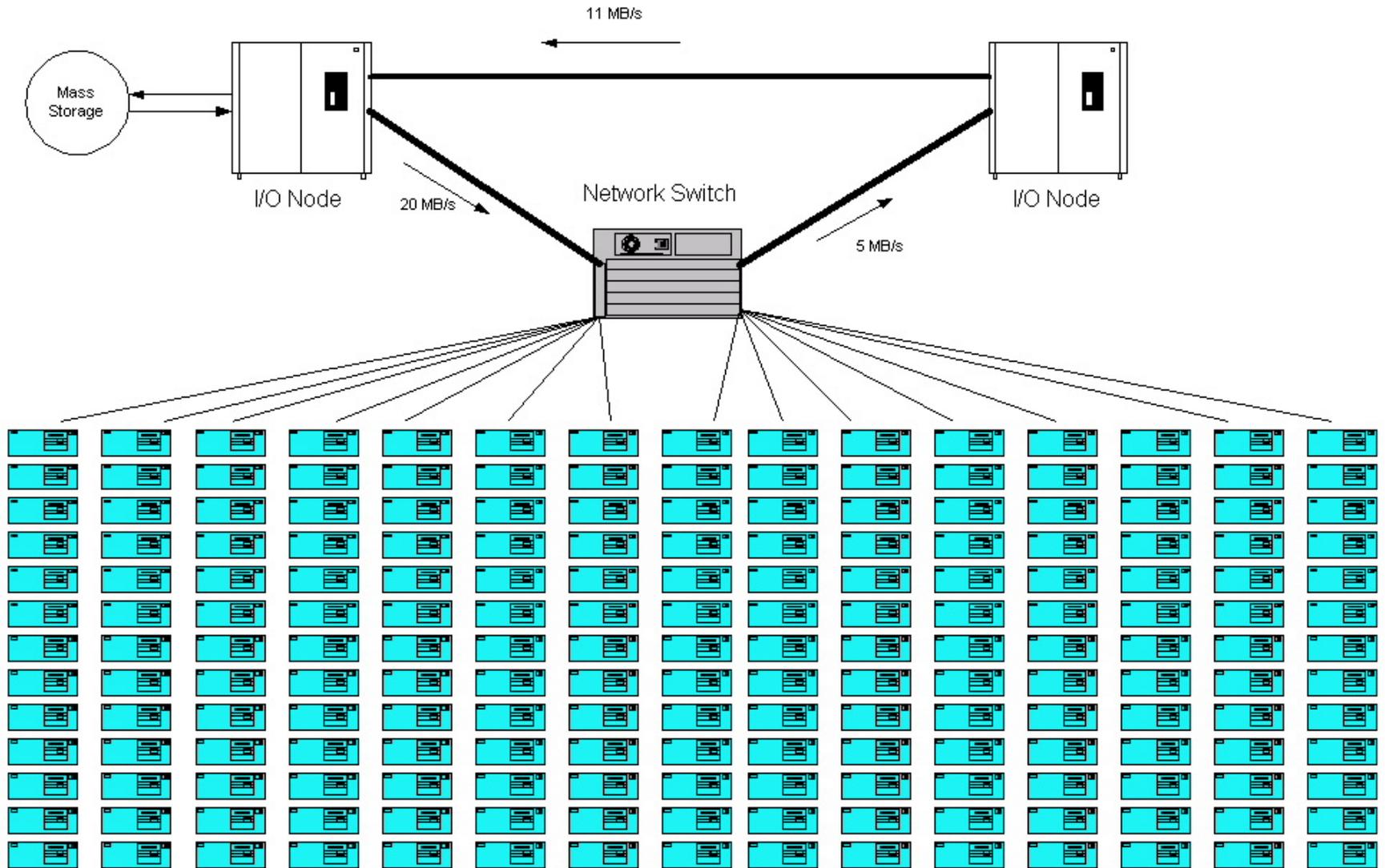


Simple Model





Run II CDF PC Farm



Beijing, China



Design/Model

• Hardware

- Choose the most cost-effective CPU's for the compute-intensive computing.
- This is currently the dual-Pentium architecture
- Network is fast and gigabit ethernet, with all machines being connected to a single or at most two large switches.
- A large I/O system to handle the buffering of data to/from mass storage and to provide a place to split the data into physics datasets.



September 3, 2001

Stephen Wolbers, CHEP2001,
Beijing, China

11



September 3, 2001

Stephen Wolbers, CHEP2001,
Beijing, China

12

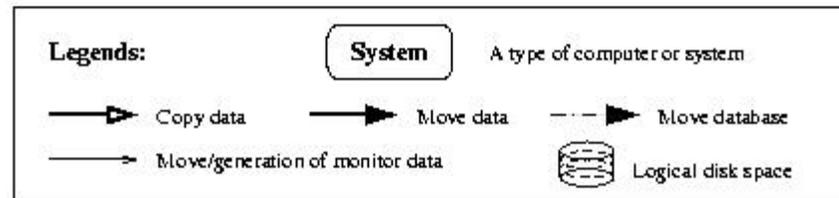
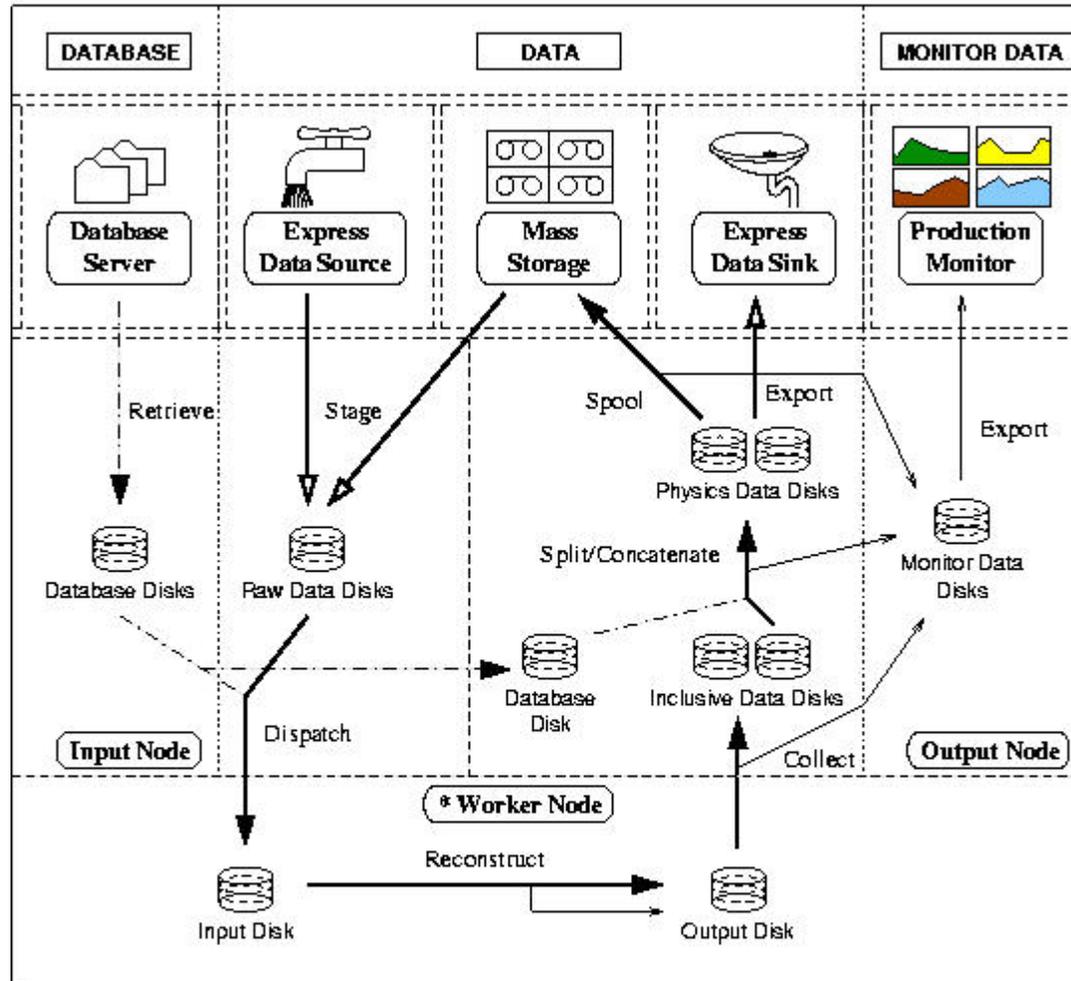


Software Model

- **Software consists of independent modules**
 - Well defined interfaces
 - Common bookkeeping
 - Standardized error handling
- **Choices**
 - Python
 - MySQL database (internal database)
 - FBSNG (Farms Batch System)
 - FIPC (Farms Interprocessor Communication)
 - CDF Data Handling Software



Conceptual Model of Run 2 Production System



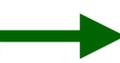
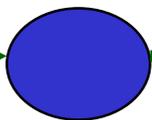


Physics Analysis Requirements and Impact

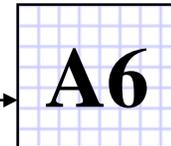
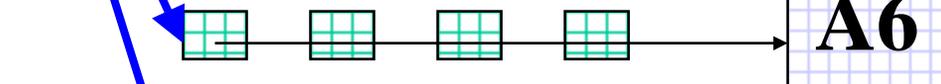
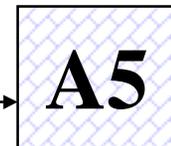
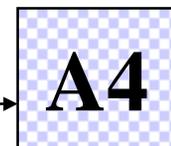
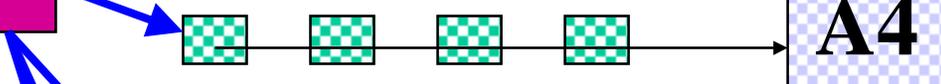
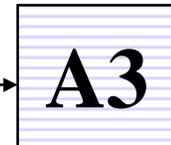
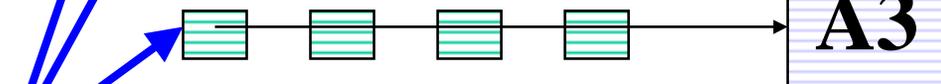
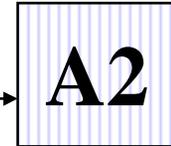
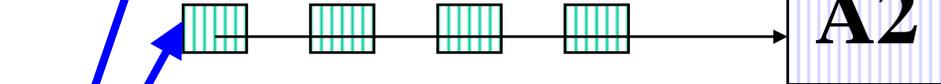
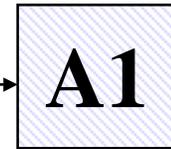
- Raw Data Files come in ~8 flavors, or streams
 - 1 Gbyte input files
- Reconstruction produces inclusive summary files
 - 250 Mbyte output files
- Output Files must be split into ~8 physics datasets per input stream
 - Target 1 Gbyte files
 - About 20% overlap
- Leads to a complicated splitting/concatenation problem, as input and output streams range from tiny (<few percent) to quite large (10's of percent)



Input Stream (x8)



Farms



September 3, 2001

Stephen Wolbers, CHEP2001,
Beijing, China



Status of CDF Farms

- 154 PC's are in place.
 - 50 PIII /500 duals
 - 40 PIII /800 duals
 - 64 PIII /1 GHz duals
- I/O nodes are ready (more disk has been added recently for buffering).
- The CDF Data Handling System has sufficient capacity to handle the I/O to/from tape.

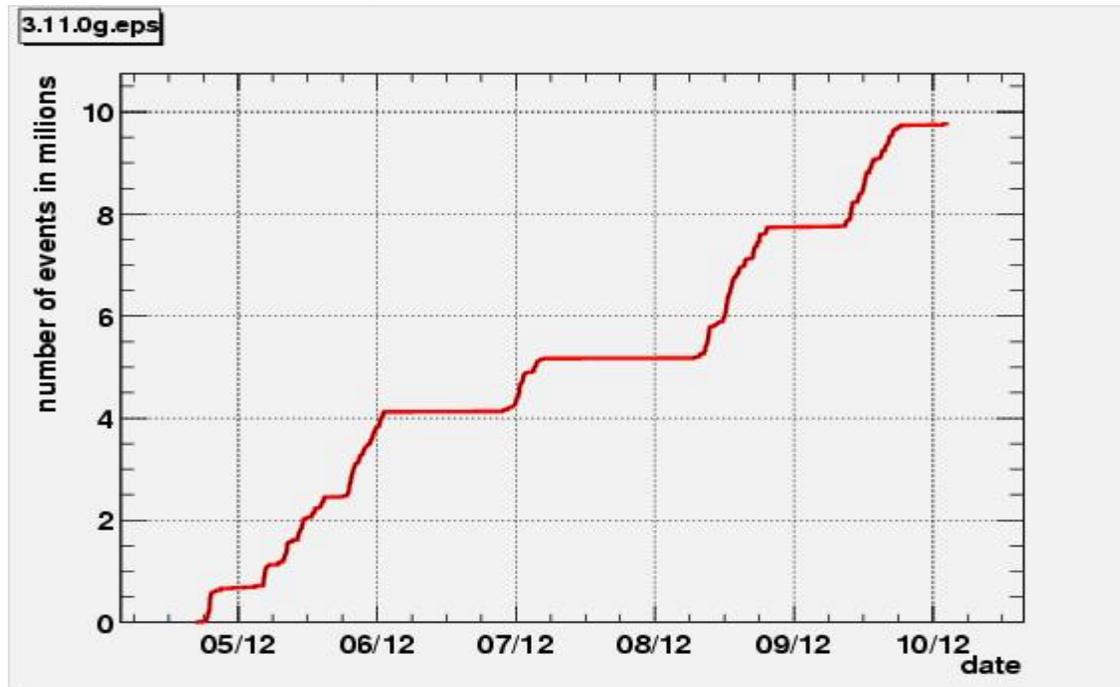


Experiences so far in Run 2



Early Processing Experience

- Commissioning Run (October, 2000)
 - Tevatron Collider luminosity was small
 - CDF detector was not yet complete
 - Best data was processed on the farms.





Lessons from Commissioning Run

- Data size was not an issue. Farms could easily keep up.
- I/O was problematic. It was easy to flood the system, filling buffers, etc.
- Reconstruction code was a big issue. Many modifications were requested, delaying the reconstruction.



Early Processing Experience

- April 2001 Data
 - First 36x36 bunch collisions
 - Most of CDF detector was complete
 - Small amount of data, additional experience





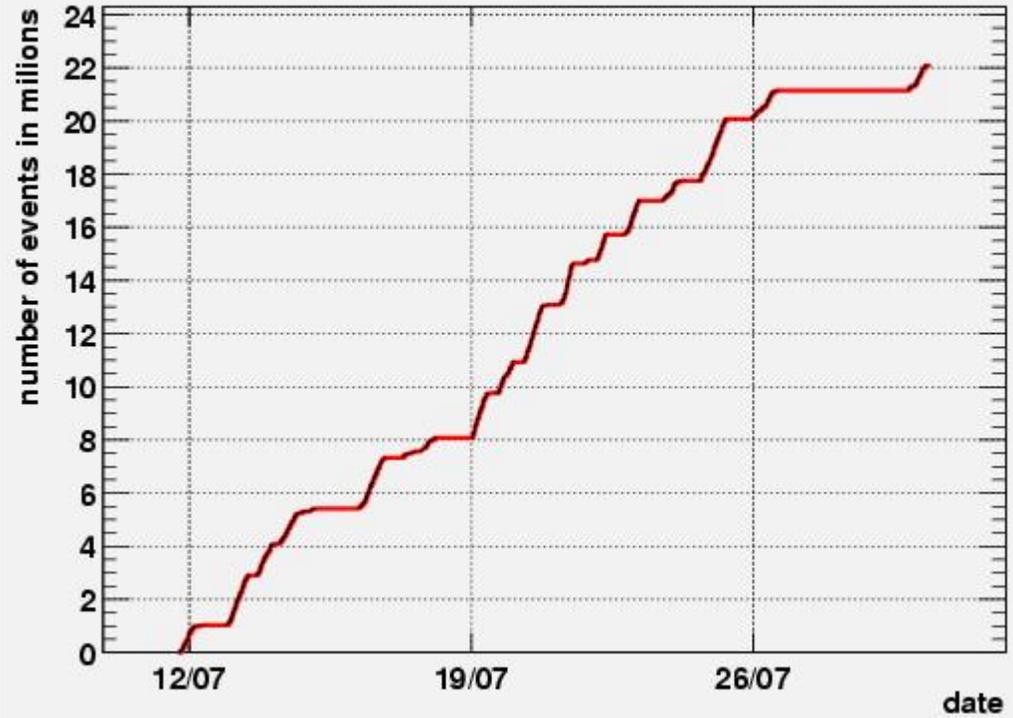
June-July 2001 Data

- **First substantial data taken in Run 2**
 - Approximately 34 million events (“good runs”)
 - Processed with two versions of the code
 - Long accelerator downtime (unplanned) allowed the farms to catch up with the backlog of data
 - I/O system was still not complete at this time, leading to a reduction of the overall rate of the farms
 - Code modifications (mainly due to detector changes) were common.
 - Calibrations were an issue
 - Full splitting into many output datasets was first tested
 - This put more of a load on the I/O system



June/July Processing

3.18.0.eps



3.17.1.eps





August-September Data

- More data was taken in August-September, 2001.
- The CDF detector was still changing, making calibrations and code changes more important.
- A “super-expressline” was invented to get data to physicists as quickly as possible.



Unexpected Issues in Run 2

- I/O problems.
 - The output was very large (in bytes/event – 400 KB/event vs 60-100 KB/event planned for) during the early Run 2 running. This was for debugging of the detector, algorithms, etc.
 - The rapid collection of data made it hard to keep up with this very large output.



Run 2b at Fermilab

- Run 2b will start in 2004 and will increase the integrated luminosity to CDF and D0 by a factor of approximately 8 (or more if possible).
- It is likely that the computing required will increase by the same factor, in order to pursue the physics topics of interest:
 - B physics
 - Electroweak
 - Top
 - Higgs
 - Supersymmetry
 - QCD
 - Etc.



Run 2b Computing

- **Current estimates for Run 2b computing:**
 - 8x CPU, disk, tape storage.
 - Expected cost is same as Run 2a because of increased price/performance of CPU, disk, tape.
 - Plans for R&D testing, upgrades/acquisitions will start next year.
- **Data-taking rate:**
 - May be as large as 80 Mbyte/s.
 - About 1 Petabyte/year to storage.



Summary

- CDF Production Farms are commissioned, tested and have processed tens of millions of events.
- Run2a will be a major task for the farms.
- Run2b is potentially substantially larger than Run2a, and some changes to the farms will likely be needed to address this.