Computing model and budget

Rick Snider
Fermilab

Outline

- Computing model
- Requirements and budget
- Computing resource strategy

CDF IFC Meeting
October 28, 2006

Drivers for the computing model

- Main issues driving the computing plan
  - Increases in raw data logging rate, total data volume
    - Final planned increase in logging rate – 60 MB/sec
      - Capability in FY2007 (originally FY2006)
    - Drives increasing complexity of computing problem with time
  - Limited Fermilab computing budget
    - Anticipated CPU demand exceeds local supply
    - Continuous need to reduce costs, increase operational efficiency
  - Evolving grid infrastructures, access policies
  - Anticipated decline in effort available starting around FY2007

Must evolve the computing model to meet challenge of resource limitations and exploit new opportunities

Evolution of computing model

- Identified several key areas for continued change at last IFC
  - Increased use of “incremental” computing model
  - Make incremental component an explicit part of budget estimates
    (more on this later...)
  - Continue aggressively expanding use of grid-based resources
  - Streamline and automate operational procedures
    - Reduce the effort required to produce physics results
  - Assess demand for 10 “core” analyses as defined by R2TF
    - Have not yet done this...

Evolution of computing model

- Identified several key areas for continued change at last IFC
  - Increased use of “incremental” computing
  - Make incremental component an explicit part of budgeting model (more on this later...)
  - Continue aggressively expanding use of grid-based resources
  - Streamline and automate operational procedures
    - Reduce the effort required to produce physics results
  - Assess demand for 10 “core” analyses as defined by R2TF
    - Have not yet done this...

Evolution of computing model

- Increase incremental computing
  - Incremental computing
    - Processing demand that is proportional to the data logging rate
  - Why is this good?
    - Maximizes efficiency of CPU resources
      - Effectively processing events only once: fixed CPU cost per event logged
    - Can centrally manage resource utilization
      - Eliminate duplication
      - Can use production framework
        - *Reduces effort and error rate*
      - Search for improvements in efficiency
        - *V7 offline release* *(more on this later...)*

Evolution of computing model

- Continue aggressively expanding use of grid-based resources
  - Meet increasing demand by increasing use of opportunistic resources off-site
    - Move away from model of increasing size of dedicated remote pools
  - Use grid tools to automate site selection for both opportunistic and dedicated resources
    - Better balances load across sites, leading to improved utilization

CDF computing model

- Major (dedicated) hardware systems
  - Tape archive + disk cache
  - Production farm (uses CAF infrastructure)
    - Performs full reconstruction of all data
  - CAF
    - Ntuple production, user analysis, MC generation
    - 1 dedicated farm + 1 CDF-purchased farm in Fermigrid (OSG) + opportunistic use of other Fermigrid computing elements
  - dCAFs
    - 11 remote sites — 4 in shared pools accessed via glide-ins
  - Databases + networks + static “project” disk

CDF computing model

- Major (dedicated) hardware systems
  - Tape archive + disk cache
  - Production farm (uses CAF infrastructure)
    - Performs full reconstruction of all data
  - CAF
    - Ntuple production, user analysis, MC generation
    - 1 dedicated farm + 1 CDF-purchased farm in Fermigrid (OSG) + opportunistic use of other Fermigrid computing elements
  - dCAFs
    - 11 remote sites — 4 in shared pools accessed via glide-ins
  - Databases + networks + static “project” disk

Eliminated 2nd farm in Sept.

CDF computing model

- Major (dedicated) hardware systems
  - Tape archive + disk cache
  - Production farm (uses CAF infrastructure)
    - Performs full reconstruction of all data
  - CAF
    - Ntuple production, user analysis, MC generation
    - 1 dedicated farm + 1 CDF-purchased farm in Fermigrid (OSG) + opportunistic use of other Fermigrid computing elements
  - dCAFs
    - 11 remote sites — 4 in shared pools accessed via glide-ins
  - Databases + networks + static “project” disk

Merging “production farm” into CAF. Improves overall utilization.

CDF computing model

- Major (dedicated) hardware systems
  - Tape archive + disk cache
  - Production farm (uses CAF infrastructure)
    - Performs full reconstruction of all data
  - CAF
    - Ntuple production, user analysis, MC generation
    - 1 dedicated farm + 1 CDF-purchased farm in Fermigrid (OSG) + opportunistic use of other Fermigrid computing elements
  - dCAFs
    - 11 remote sites — 4 in shared pools accessed via glide-ins
  - Databases + networks + static “project” disk

New purchases destined for Fermigrid. CAF will migrate there if possible.

CDF computing model

- Data flow
  - Raw data written to tape
  - Measure final calibrations (4 weeks of data per cycle)
    - Process about 30% of data stream on production farm and CAF
  - Full (final) reconstruction on production farm
    - Input read, output written to tape
  - Centralized ntuple production
    - All three ntuples produced on production farm
  - User analysis of ntuples and production output on CAF/dCAF
  - Monte Carlo generation/reconstruction on remote dCAFs/Grid
  - Final analysis on user desktops / institutional machines

CDF computing model

- Data flow
  - Raw data written to tape
  - Measure final calibrations (4 weeks of data per cycle)
    - Process about 30% of data stream on production farm and CAF
  - Full (final) reconstruction on production farm
    - Input read, output written to tape
  - Centralized ntuple production
    - All three ntuples produced on production farm
  - User analysis of ntuples and production output on CAF/dCAF
  - Monte Carlo generation/reconstruction on remote dCAFs/Grid
  - Final analysis on user desktops / institutional machines

Since last IFC (was 2 of 3)

CDF computing model

- Data flow
  - Raw data written to tape
  - Measure final calibrations (4 weeks of data per cycle)
    - Process about 30% of data stream on production farm and CAF
  - Full (final) reconstruction on production farm
    - Input read, output written to tape
  - Centralized ntuple production
    - All three ntuples produced on production farm
  - User analysis of ntuples and production output on CAF/dCAF
  - Monte Carlo generation/reconstruction on remote dCAFs/Grid
  - Final analysis on user desktops / institutional machines

Incremental, centrally managed components

CDF computing model

- Data flow
  - Raw data written to tape
  - Measure final calibrations (4 weeks of data per cycle)
    - Process about 30% of data stream on production farm and CAF
  - Full (final) reconstruction on production farm
    - Input read, output written to tape
  - Centralized ntuple production
    - All three ntuples produced on production farm
  - User analysis of ntuples and production output on CAF/dCAF
  - Monte Carlo generation/reconstruction on remote dCAFs/Grid
  - Final analysis on user desktops / institutional machines

Almost completely automated process

CDF computing model

- Data flow
  - Raw data written to tape
  - Measure final calibrations (4 weeks of data per cycle)
    - Process about 30% of data stream on production farm and CAF
  - Full (final) reconstruction on production farm
    - Input read, output written to tape
  - Centralized ntuple production
    - All three ntuples produced on production farm
  - User analysis of ntuples and production output on CAF/dCAF
  - Monte Carlo generation/reconstruction on remote dCAFs/Grid
  - Final analysis on user desktops / institutional machines

New and increasingly important resource

Computing requirements

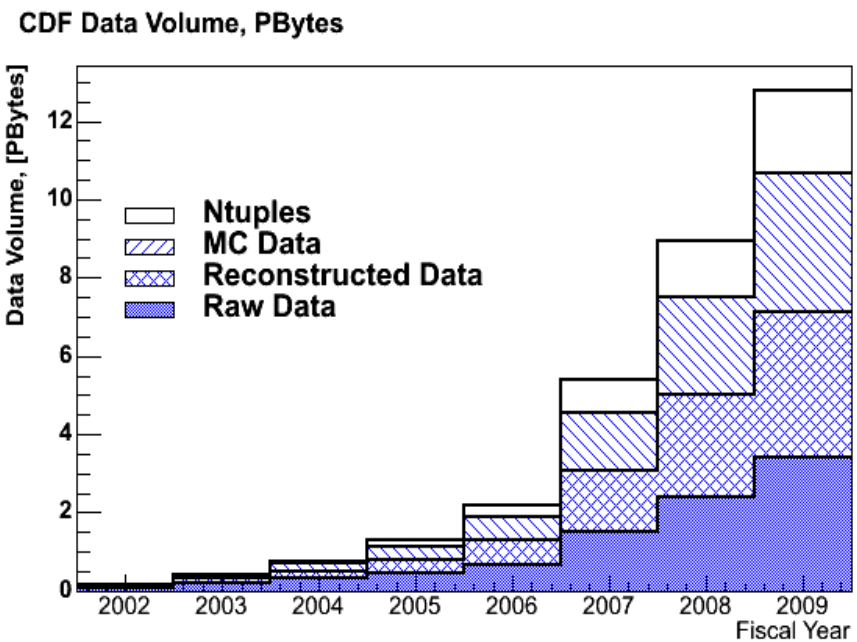
- Several improvements/changes to the demand model
  - Explicitly calculate CPU needs for all incremental computing
    - Takes into account
      - Measured CPU required for each type of process, type of data processed
      - Event processing overlaps, where appropriate
      - Measured luminosity dependence of reconstruction
  - Analysis CPU
    - Equal to total minus incremental (marked to FY2005 inventory)
    - Scale with data volume (...a pessimistic assumption)
  - Disk requirements model
    - Scale cache disk by available CPU
      - Currently 450 TB of 700 TB total
    - Scale analysis disk proportional to data volume

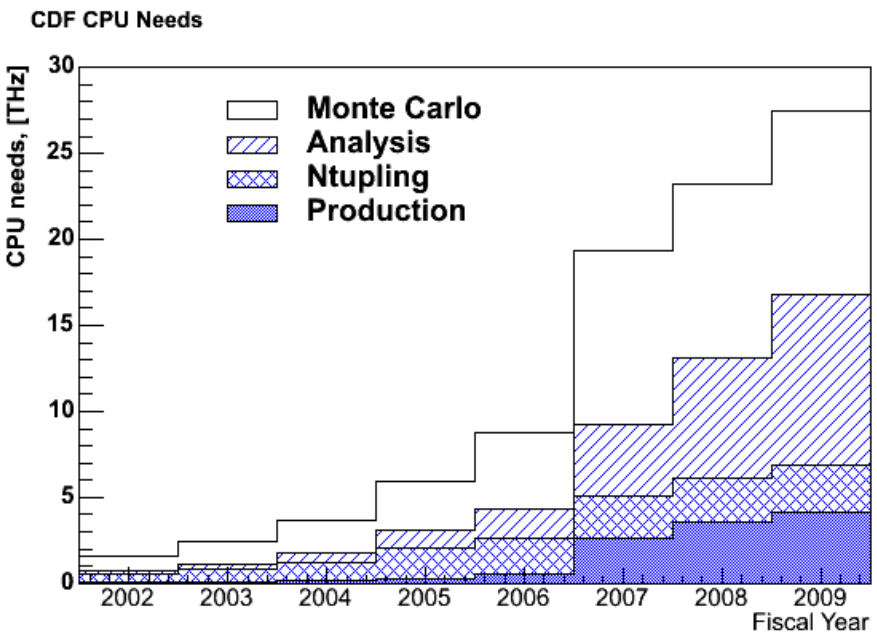
Computing model input parameters (spec)

Fiscal Year	2007	2008	2009
Integrated luminosity (fb^{-1})	3.9	5.9	8.1
Total number of events (10^9)	5.7	9.9	14
Raw data logging rate (MB/s)*	60*	60*	60*

These rates reflect the system specifications

* Revised upward since last IFC

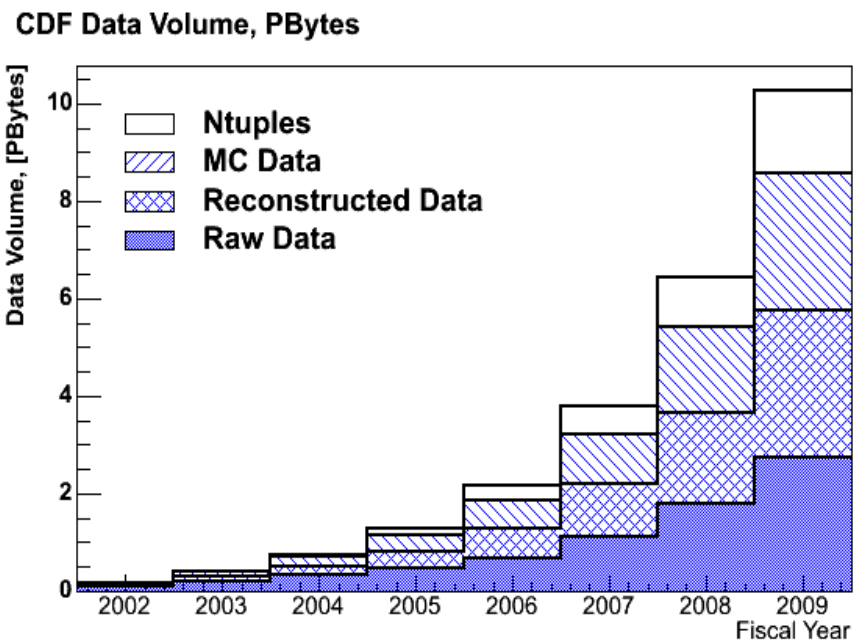


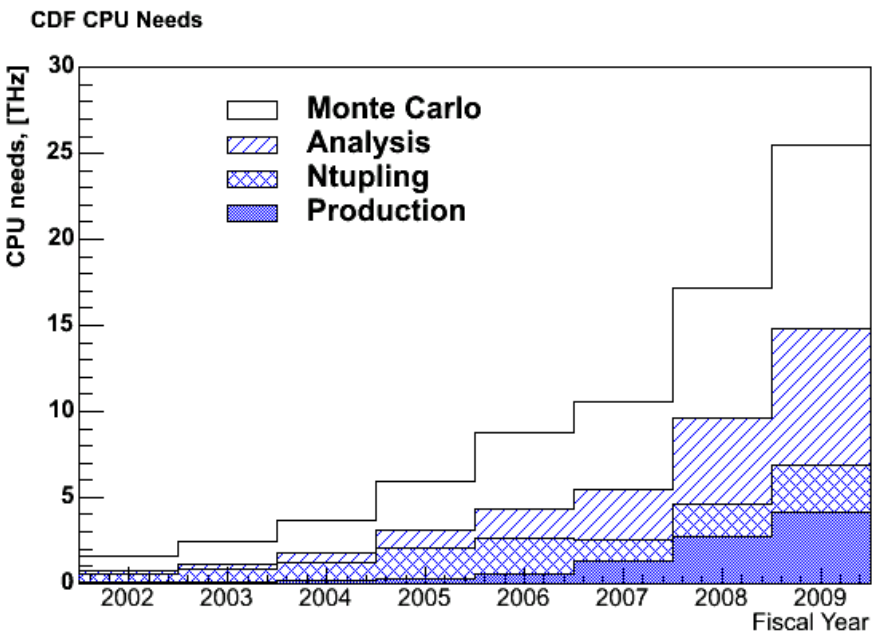


Computing model input parameters

Fiscal Year	2007	2008	2009
Integrated luminosity (fb^{-1})	3.9	5.9	8.1
Total number of events (10^9)	5.7	9.9	14
Raw data logging rate (MB/s)	30	45	60

More realistic assumptions about rates that will be achieved





Projected computing requirements

Fiscal Year	2007	2008	2009
CPU needs (THz)*	11	17	25
Disk volume (PB)†	1.0	1.6	2.5
Data volume on tape (PB)	3.8	6.4	10.3

* Shared between CAF, dCAF's and grid
† Hosted at Fermilab

Fermilab budget

- Equipment
  - Procurement profile
    - Assume constant nominal budget
      - FY2007 budget guidance: $1.25M
    - Defined cost for tape storage ($400k/year)
    - Target CPU required for reconstruction, ntupling and analysis
      - Spend $300k/year
    - Balance for disk ($300k/year)
  - Purchase FY200x hardware with FY(200x – 1) funds
- Operating
  - Tape cost is $225k logging at 30 MB/s

Request to IFC

- No further expansion of dedicated resources requested
  - Maintain existing dedicated resources where feasible
  - Put additional resources into expanding generic pools accessible via grid
- Provide CDF access to generic grid pools (if not already allowed)
  - Priority access to some fraction of generic pools where possible

Computing inventory

		Actual		Requirements		
	Fiscal Year	2006	2007	2007	2008	2009
CPU (THz)	Fermilab	4.8	7.6	7.6	7.8	7.9
	On-site contributions	1.7	1.7	1.7	1.7	1.7
	Remote (dedicated)	2.3	2.3	2.3	2.3	2.3
	Opportunistic		1.0	1.0	5.0	13
	Total	8.8	13	13	17	25
Disk (PB)	Fermilab	0.6	0.6	0.6	0.9	1.6
	On-site contributions	0.2*	0.1	0.1	0.1	0.1
	Remote	0.1?	0.1?			
	Total	0.9?	0.9?	0.7	1.0	1.7
Volume on tape, Oct. 1, 2006 (PB)		1.5	1.6	3.8	6.4	10.3

* 50% retires at end of FY2006

Cost mitigation strategies

- Opportunistic computing
  - Is it plausible to obtain 1.0 THz in FY2007, 5.0 THz in FY2008 from opportunistic sources?
  - Routinely getting up to 3 THz today on just 5 NamCAF sites
    - OK in FY2007
  - Estimated CPU for LHC
    - Opportunistic FY2008 (FY2009) demand is about 2% (3%) of LHC total
      - This is uncomfortably large, but not obviously implausible
      - Becomes more problematic LHC begins operating
      - Target 1%

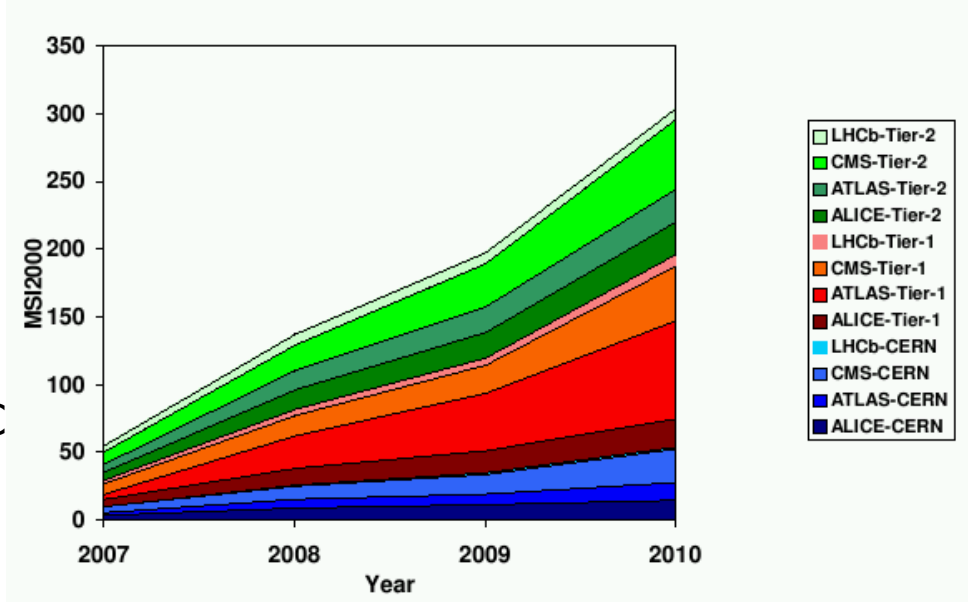


Figure 1 CPU requirements at the Tier-0, Tier-1 and Tier-2 centres time period is from 2007 to 2010.

Cost mitigation strategies

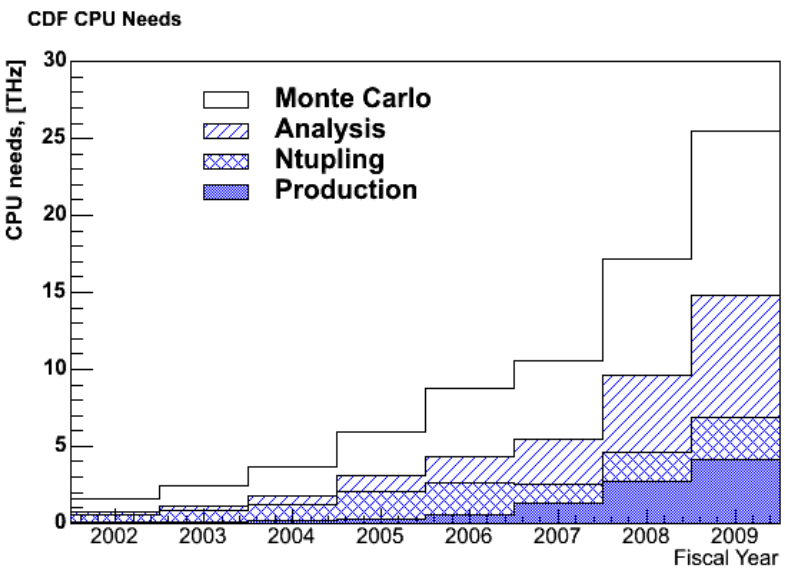
- Improve utilization efficiency
  - V7 offline release
    - Improved tracking
      - Can eliminate most CPU-intensive part of ntuple creation, some analyses
- Prioritize computing tasks
  - Perform fewer analyses
    - May happen anyway with anticipated reductions in available effort
  - Defer processing some data
  - Reduce MC statistics

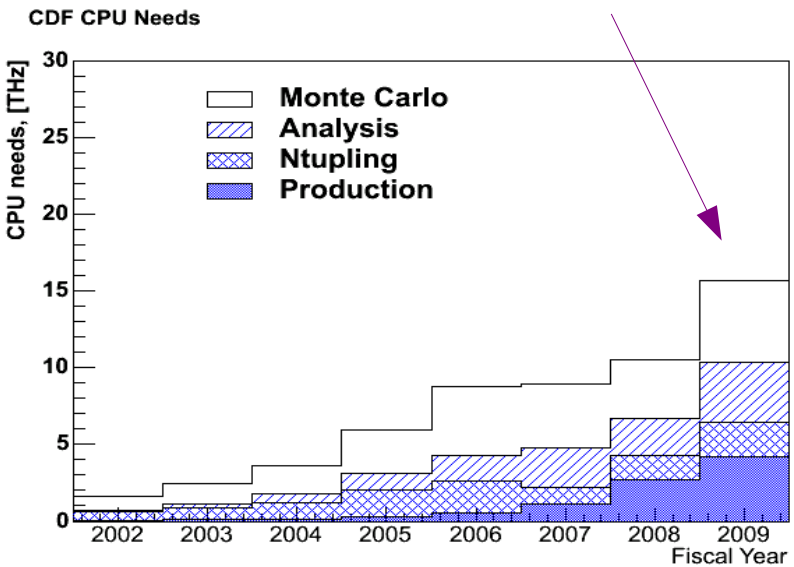
Could imagine reduction of a factor of two in combined CPU for MC, ntupling and analysis, but at some cost to the physics

Cost mitigation strategies

- Prioritize computing tasks (cont'd)

Opportunistic demand in this scenario= 1.7% of LHC computing





MC, analysis CPU demand reduced by factor of 2

Summary

- The success of CDF physics program owes much to the strong support and contribution of the Computing Division and IFC to the CDF computing project.

- To make best use of computing resources, CDF continues to evolve the computing model to reduce the cost and effort required to perform computing tasks

The end

Backup slides

FY2006 procurements at Fermilab

- CPU
  - Shifted budget allocation for disk, tape drives into CPU
    - Realities of off-site growth rate
    - Better scaling model for disk
    - Logging rate increases did not materialize as expected
  - Added net of 2.8 THz to CPU at Fermilab ($920k)
    - Will be available in January, 2007
- Disk
  - Just replaced retirements, optimized for specific use cases ($240k)
- Tape drives
  - None added pending availability of higher density LTO-3

Equipment budget details

Projected total computing cost ($M)

Fiscal Year	2007	2008	2009
CPU*	0.3	0.3	0.3
Disk	0.3	0.3	0.3
Tape drives + library†	0.4	0.3	0.4
Interactive computing	0.02	0.02	0.02
Databases	0.03	0.03	0.03
Miscellaneous	0.05	0.1	0.1
Networking	0.1	0.1	0.1
Total cost	1.2	1.2	1.2

* Cost shared between Fermilab and remote institutions
† Change in tape density in FY2008