

Computing model and budget

Rick Snider
Fermilab

Outline

- Computing model
- Requirements and budget
- Computing resource strategy

CDF IFC Meeting
October 30, 2007

Drivers of change for the computing model

- Main issues driving evolution of the computing plan
 - Increases in raw data logging rate, total data volume
 - Expected increase in logging rate
 - Expect ultimate rate to be <30 MB/s from about 20 MB/s now
 - 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 as LHC startup nears

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

Strategic evolution of computing model

- Identified areas for continued change at past IFC meetings
 - Maximize use of “incremental” computing model
 - Most cost-effective use of CPU resources
 - Cost scales with data logging rate
 - Continue aggressively expanding use of grid-based resources
 - Secure agreements for priority access at grid sites
 - Streamline and automate operational procedures
 - Reduce the effort required to produce physics results
- All basic steps completed. Optimization needed from this point onward
Budget model uses this assumption

Incremental computing model

- “Incremental” computing
 - Processing demand that is proportional to the data logging [rate](#)
- Benefits
 - Maximizes efficiency of CPU resources
 - Fixed CPU cost per event logged
 - Can centrally manage resource utilization
 - Eliminate duplication
 - Can use production framework
 - Reduces effort and error rate
 - Leverages gains in operational efficiency
 - V7 offline release, ntupling infrastructure

Expanding use of grid-based resources

- Consolidate computing resources into grid-based infrastructure
 - Move away from supporting dCAFs
...but still need dedicated resource allocations / guarantees
 - Use a single user interface to submit jobs via OSG / LCG portals
- Benefits
 - Minimizes long-term support effort by leveraging large grid development and support teams
 - Allows seamless exploitation of opportunistic resources
 - Automated site selection across large number of installations
 - Better balances load across sites leading to improved utilization

CDF computing model

- Major hardware systems
 - Tape archive + disk cache
 - CAF
 - Full reconstruction of all data
 - Ntuple production, user analysis, MC generation
 - 1 dedicated farm + 1 CDF-purchased farm in Fermigrid (OSG) + opportunistic use of other Fermigrid computing elements
 - Dedicated off-site resources
 - MC generation at 6 remote sites — dCAFs + grid-based access
 - Opportunistic off-site resources
 - MC generation at 17 sites (as of last week — 10 more on the way soon)
 - Databases + networks + “project” disk

CDF computing model

- Major hardware systems

- Tape archive + disk cache

- **CAF**

- Full reconstruction of all data

- Ntuple production, user analysis, MC generation

- 1 dedicated farm + 1 CDF-purchased farm in Fermigrid (OSG) + opportunistic use of other Fermigrid computing elements

- Dedicated off-site resources

- MC generation at 6 remote sites — dCAFs + grid-based access

- Opportunistic off-site resources

- MC generation at 17 sites (as of last week — 10 more on the way soon)

- Databases + networks + “project” disk

Eliminated dedicated
“production farm” in Sept.
Reduces support load.



CDF computing model

- Major hardware systems

- Tape archive + disk cache

- **CAF**

- **Full reconstruction of all data**

- Ntuple production, user analysis, MC generation

- 1 dedicated farm + 1 CDF-purchased farm in Fermigrid (OSG) + opportunistic use of other Fermigrid computing elements

- Dedicated off-site resources

- MC generation at 6 remote sites — dCAFs + grid-based access

- Opportunistic off-site resources

- MC generation at 17 sites (as of last week — 10 more on the way soon)

- Databases + networks + “project” disk

Reconstruction migrated to CAF. Improves utilization efficiency.



CDF computing model

- Major hardware systems

- Tape archive + disk cache
- CAF
 - Full reconstruction of all data
 - Ntuple production, user analysis, MC generation
 - 1 dedicated farm + 1 CDF-purchased farm in Fermigrid (OSG) + opportunistic use of other Fermigrid computing elements
- Dedicated off-site resources
 - MC generation at 6 remote sites — dCAFs + grid-based access
- Opportunistic off-site resources
 - MC generation at 17 sites (as of last week — 10 more on the way soon)
- Databases + networks + “project” disk

Dedicated farm will be consolidated here within next few months.



CDF computing model

- Major hardware systems

- Tape archive + disk cache
- CAF
 - Full reconstruction of all data
 - Ntuple production, user analysis, MC generation
 - 1 dedicated farm + 1 CDF-purchased farm in Fermigrid (OSG) + opportunistic use of other Fermigrid computing elements
- Dedicated off-site resources
 - MC generation at 6 remote sites — dCAFs + grid-based access
- **Opportunistic off-site resources**
 - MC generation at 17 sites (as of last week — 10 more on the way soon)
- Databases + networks + “project” disk

Submit MC jobs to non-CDF affiliated computers via OSG/LCG.



CDF computing model

- Major hardware systems

- Tape archive + disk cache
- CAF
 - Full reconstruction of all data
 - Ntuple production, user analysis, MC generation
 - 1 dedicated farm + 1 CDF-purchased farm in Fermigrid (OSG) + opportunistic use of other Fermigrid computing elements
- Dedicated off-site resources
 - MC generation at 6 remote sites — dCAFs + grid-based access
- Opportunistic off-site resources
 - MC generation at 17 sites (as of last week — 10 more on the way soon)
- Databases + networks + “project” disk

Phasing out remaining
dCAFs, migrating
to Grid infrastructure



CDF computing model

- Major hardware systems

- Tape archive + disk cache
- CAF
 - Full reconstruction of all data
 - Ntuple production, user analysis, MC generation
 - 1 dedicated farm + 1 CDF-purchased farm in Fermigrid (OSG) + opportunistic use of other Fermigrid computing elements
- Dedicated off-site resources
 - MC generation at 6 remote sites — dCAFs + **grid-based access**
- Opportunistic off-site resources
 - MC generation at 17 sites (as of last week — 10 more on the way soon)
- Databases + networks + “project” disk

Most dCAF sites also accessible via grid submission points



CDF computing model

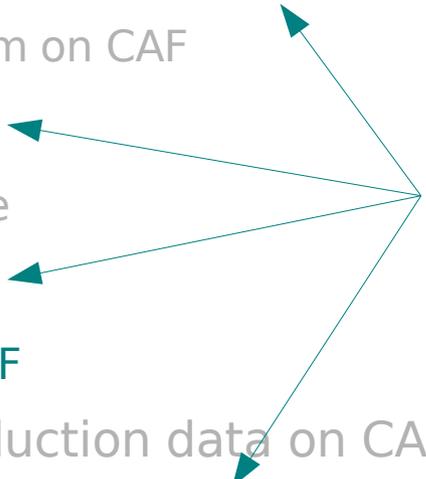
- Data flow
 - Raw data written to tape
 - Measure final calibrations (4–8 weeks of data per cycle)
 - Process about 30% of data stream on CAF
 - Full (final) reconstruction on CAF
 - Input read, output written to tape
 - Centralized ntuple production
 - All three ntuples produced on CAF
 - User analysis of ntuples and production data 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 (4–8 weeks of data per cycle)
 - Process about 30% of data stream on CAF
- Full (final) reconstruction on CAF
 - Input read, output written to tape
- Centralized ntuple production
 - All three ntuples produced on CAF
- User analysis of ntuples and production data 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 (4–8 weeks of data per cycle)
 - Process about 30% of data stream on CAF
 - Full (final) reconstruction on CAF
 - Input read, output written to tape
 - Centralized ntuple production
 - All three ntuples produced on CAF
 - User analysis of ntuples and production data 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 (4–8 weeks of data per cycle)
 - Process about 30% of data stream on CAF
 - Full (final) reconstruction on CAF
 - Input read, output written to tape
 - Centralized ntuple production
 - All three ntuples produced on CAF
 - User analysis of ntuples and production data on CAF/dCAF
 - **Monte Carlo generation/reconstruction** on remote dCAFs/Grid
 - Final analysis on user desktops / institutional machines

Moving from run-based to instantaneous luminosity weighted. Simplifies operations.



CDF computing model

- Data flow
 - Raw data written to tape
 - Measure final calibrations (4–8 weeks of data per cycle)
 - Process about 30% of data stream on CAF
 - Full (final) reconstruction on CAF
 - Input read, output written to tape
 - Centralized ntuple production
 - All three ntuples produced on CAF
 - User analysis of ntuples and production data on CAF/dCAF
 - Monte Carlo generation/reconstruction on remote dCAFs/Grid
 - Final analysis on user desktops / institutional machines

Migrating to fully Grid-based
MC production model



Computing requirements

- Computing 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, ntupling
 - 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 about 700 TB of 1.1 PB total
 - Scale analysis disk with total data volume

Computing requirements

- Changes to computing demand model
 - Reduced expected logging rates in FY2008+
 - Reduced CPU requirement for event reconstruction
 - Production farm consolidation into CAF \Rightarrow no duty cycle factors
 - Gen-7 reconstruction
 - Event reconstruction time increases by 50% (net of many speed improvements)
 - Ntupling time decreases by 10-15% (not included in current calculations)
 - Some computing moved from ntupling to reconstruction phase
 - Tevatron operations
 - Assume “Scenario IV” projections (optimistic)
 - Extended projections through 2010 operations (more optimism)
 - Updated data, CPU models based upon FY2005 – FY2007 data

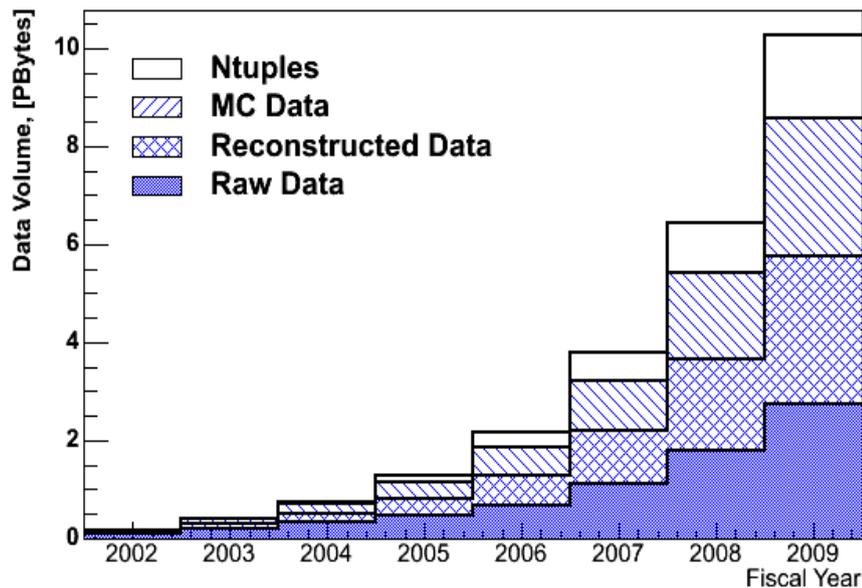
Computing model input parameters

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

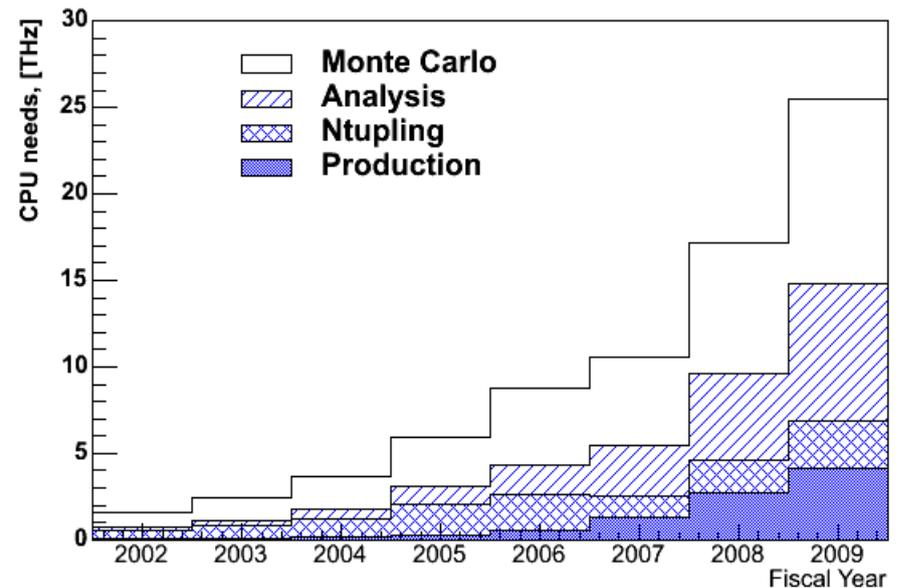
Assumed logging rates presented last year



CDF Data Volume, PBytes



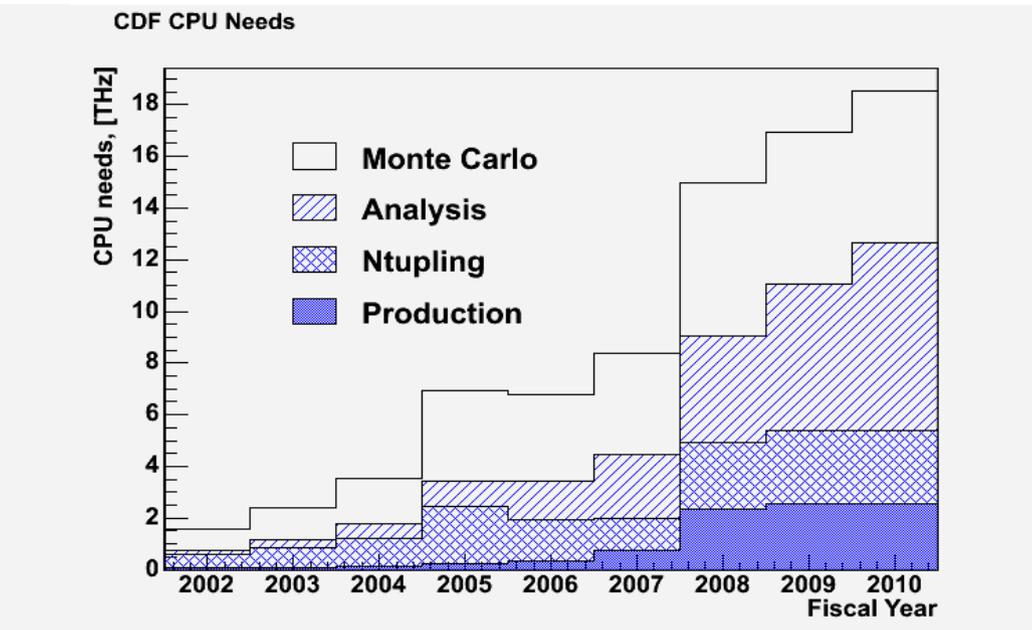
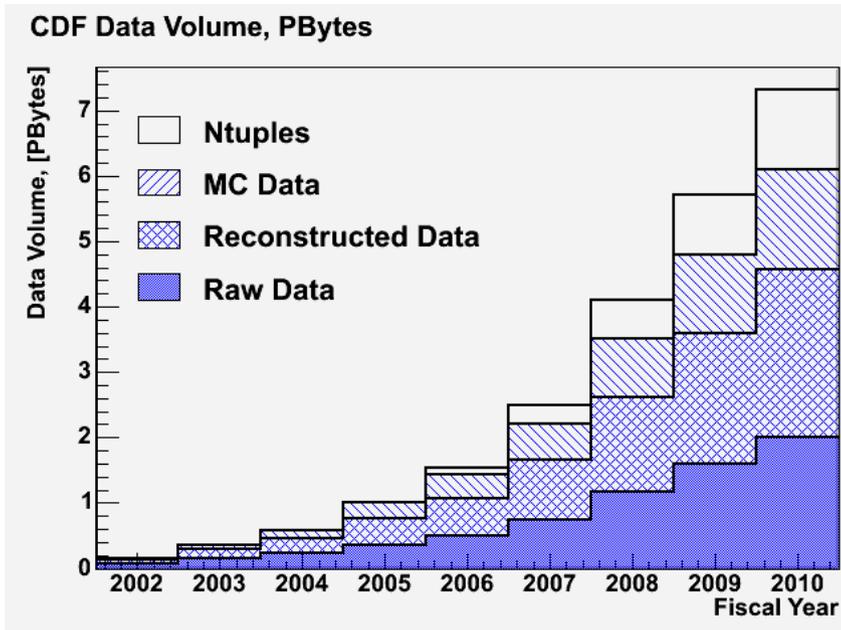
CDF CPU Needs



Computing model input parameters

Fiscal Year	2007	2008	2009	2010
Integrated luminosity (fb^{-1})	3.2	5.9	6.8	8.1
Total number of events (10^9)	5.0	8.0	11	13
Raw data logging rate (MB/s)	17	30	30	30

Expected upper limits based on curr. experience



Projected computing requirements

Projected computing requirements

	Fiscal Year	2008	2009	2010
CPU needs (THz)*		15	17	18
Disk volume (PB) [†]		1.0	1.3	1.5
Data volume on tape (PB)		4.1	5.7	7.3

* Shared between CAF, dCAF's and grid

FY2007 procurements at Fermilab

- CPU

- Shifted budget allocation from tapes into CPU
 - Tape cost dropped by 45% + lower than expected logging rate
- Added net of 1.7 THz to CPU at Fermilab (\$520k)
(includes about \$66k from Japan)
 - Will be available in November, 2007
(much earlier delivery, deployment than past years)

- Disk

- Replaced retirements in cache, expanded project disk, many new servers optimized for special uses (\$350k)

- Tape drives

- Added 7 LTO-3 drives for a total of 17 (\$126k)
- Tape library cost of about \$150k

Future procurements at Fermilab

- Equipment

- Assume slight drop in nominal dollars from FY2008 forward
 - Unofficial FY2008 budget guidance: \$1.2 M (spent \$1.3 M last year)
- Procurement strategy
 - Cost for tape storage \$263k (assumes \$66k for tape density migration in Q4 FY2008)
 - Additional \$325k for tape drives in FY2008 (assuming tape density change in Q4)
 - Target CPU required for reconstruction, ntupling and analysis
 - About \$300k / year
 - Balance for disk (\$100k to \$200k / year)
- Purchase FY200x hardware with FY(200x – 1) funds

- Operating

- FY2008 tape cost is \$154k
(Includes \$38k for tape density migration in Q4 FY2008)

Computing inventory

		Actual		Requirements		
		Fiscal Year	2007	2008	2008	2009
CPU (THz)	Estimated requirement			15	17	18
	Fermilab	7.9	9.6	10	11	12
	On-site contributions	1.7	1.7	1.7	1.7	1.7
	Remote (dedicated)	1.6	1.6	1.6	2.3	2.3
	Opportunistic	1.7	1.7	1.7	2.0	2.0
	Total available	13	15			
Disk (PB)	Estimated requirement			1.0	1.3	1.5
	Fermilab	0.7	1.0	0.98	1.2	1.4
	On-site contributions	0.1	0.06	0.06	0.06	0.06
	Remote	0.1?	0.1?			
	Total available	0.9	1.2			
Volume on tape (PB)		2.6	—	4.1	5.7	7.3

Computing inventory

		Actual		Requirements		
Fiscal Year		2007	2008	2008	2009	2010
CPU (THz)	Estimated requirement					
	Fermilab	7.9	9.6			
	On-site contributions	1.7	1.7			
	Remote (dedicated)	1.6	1.6			
	Opportunistic	1.7	1.7			
	Total available	13	15			
Disk (PB)	Estimated requirement			1.0	1.3	1.5
	Fermilab	0.7	1.0	0.98	1.2	1.4
	On-site contributions	0.1	0.06	0.06	0.06	0.06
	Remote	0.1?	0.1?			
	Total available	0.9	1.2			
Volume on tape (PB)		2.6	—	4.1	5.7	7.3

← What we have on the floor at Fermilab

Computing inventory

		Actual		Requirements		
		2007	2008	2008	2009	2010
Fiscal Year		2007	2008	2008	2009	2010
Estimated requirement						
CPU (THz)	Fermilab	7.9	9.6			
	On-site contributions	1.7	1.7			
	Remote (dedicated)	1.6	1.6			
	Opportunistic	1.7	1.7			
	Total available	13	15			
Estimated requirement				1.0	1.3	1.5
Disk (PB)	Fermilab	0.7	1.0	0.98	1.2	1.4
	On-site contributions	0.1	0.06	0.06	0.06	0.06
	Remote	0.1?	0.1?			
	Total available	0.9	1.2			
Volume on tape (PB)		2.6	—	4.1	5.7	7.3

Purchased by Fermilab

Originally purchased by institutions

Computing inventory

		Actual		Requirements		
		2007	2008	2008	2009	2010
Fiscal Year		2007	2008	2008	2009	2010
Estimated requirement						
CPU (THz)	Fermilab	7.9	9.6			
	On-site contributions	1.7	1.7			
	Remote (dedicated)	1.6	1.6			
	Opportunistic	1.7	1.7			
	Total available	13	15			
Estimated requirement				1.0	1.3	1.5
Disk (PB)	Fermilab	0.7	1.0	0.98	1.2	1.4
	On-site contributions	0.1	0.06	0.06	0.06	0.06
	Remote	0.1?	0.1?			
	Total available	0.9	1.2			
Volume on tape (PB)		2.6	—	4.1	5.7	7.3

Dedicated pools (dCAFs) + dedicated slots

Observed value in FY2007

Computing inventory

		Actual		Requirements			
		Fiscal Year	2007	2008	2008	2009	
CPU (THz)	Estimated requirement			15	17	18	← The model estimates
	Fermilab	7.9	9.6	10	11	12	← Assumes flat budget (nets all change at FNAL)
	On-site contributions	1.7	1.7	1.7	1.7	1.7	
	Remote (dedicated)	1.6	1.6	1.6	2.3	2.3	
	Opportunistic	1.7	1.7	1.7	2.0	2.0	
	Total available	13	15				
Disk (PB)	Estimated requirement			1.0	1.3	1.5	
	Fermilab	0.7	1.0	0.98	1.2	1.4	
	On-site contributions	0.1	0.06	0.06	0.06	0.06	
	Remote	0.1?	0.1?				
	Total available	0.9	1.2				
Volume on tape (PB)		2.6	—	4.1	5.7	7.3	

Computing inventory

		Actual		Requirements		
Fiscal Year		2007	2008	2008	2009	2010
CPU (THz)	Estimated requirement			15	17	18
	Fermilab	7.9	9.6	10	11	12
	On-site contributions	1.7	1.7	1.7	1.7	1.7
	Remote (dedicated)	1.6	1.6	1.6	2.3	2.3
	Opportunistic	1.7	1.7	1.7	2.0	2.0
	Total available	13	15			
Disk (PB)	Estimated requirement			1.0	1.3	1.5
	Fermilab	0.7	1.0	0.98	1.2	1.4
	On-site contributions	0.1	0.06	0.06	0.06	0.06
	Remote	0.1?	0.1?			
	Total available	0.9	1.2			
Volume on tape (PB)		2.6	—	4.1	5.7	7.3

Balance from sum of remote dedicated / priority and opportunistic.

(More on the split next...)

Computing inventory

		Actual		Requirements		
		Fiscal Year	2007	2008	2008	2009
CPU (THz)	Estimated requirement			15	17	18
	Fermilab	7.9	9.6	10	11	12
	On-site contributions	1.7	1.7	1.7	1.7	1.7
	Remote (dedicated)	1.6	1.6	1.6	2.3	2.3
	Opportunistic	1.7	1.7	1.7	2.0	2.0
	Total available	13	15			
Disk (PB)	Estimated requirement			1.0	1.3	1.5
	Fermilab	0.7	1.0	0.98	1.2	1.4
	On-site contributions	0.1	0.06	0.06	0.06	0.06
	Remote	0.1?	0.1?			
	Total available	0.9	1.2			
Volume on tape (PB)		2.6	—	4.1	5.7	7.3

← The model estimates
← Assumes no budget constraint

Cost mitigation strategies

- Opportunistic computing
 - Routinely obtained 1–2 THz during past 1.5 years
 - Do not anticipate significant problem at this level in FY2008
 - Reliance on opportunistic resources introduces risk
 - Resources will become increasingly tight as LHC startup approaches
 - Many competitors for opportunistic cycles
 - Risk mitigation
 - Secure agreements / guarantees for priority access to grid pools
 - Recent example: working to develop long-term agreement to significantly expand resources available to CDF at KISTI Tier 2 center
 - Securing additional 1 THz in priority access reduces need for opportunistic cycles to ≤ 1 THz for FY2008 and beyond

Cost mitigation strategies

- Opportunistic computing (cont'd)
 - Opportunistic demand is sensitive to data logging rate
 - Logging rate of 25 MB/s reduces projected demand by about 2 THz
 - Opportunistic demand ≈ 0 if combined with increase in guaranteed / priority queue slots equivalent to 700 GHz

Request to IFC

- Migrate existing dCAFs to LCG/OSG based pools
 - Guarantee priority access to some fraction of LCG/OSG pools
 - Particularly important after start-up of LHC
- Provide CDF access to local LCG/OSG pools if not already allowed
 - Priority access to some fraction where possible

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.
 - CDF has had access to sufficient computing to achieve physics goals as evidenced by copious output of papers and conference results
 - This computing has been provided with a combination of Fermilab, off-site dedicated and opportunistic resources
- Our plan for continued success is to continue evolving and improving the computing model while achieving the correct balance of dedicated and opportunistic resources

The end

Backup slides