CDF computing requirements and budget

F.D. Snider

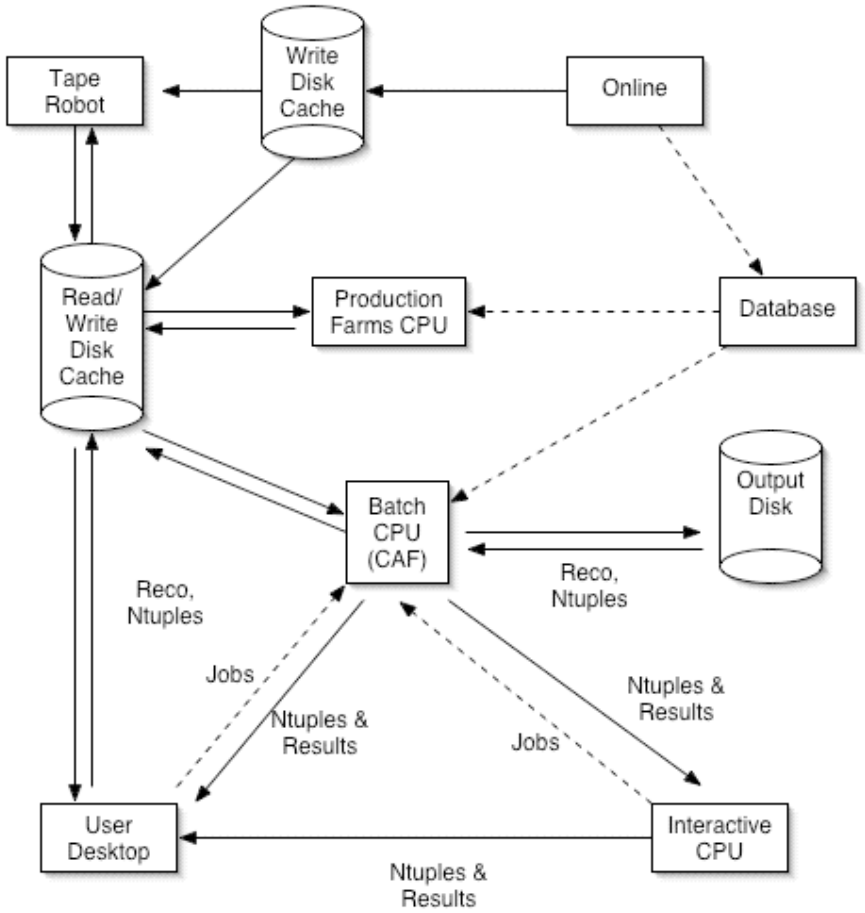
Outline

- Computing requirements
- Requirements model
- Budget estimates
- Summary

Run 2 Computing Review
September 13, 2004

Run 2 computing requirements

- Major components of CDF computing supported by FNAL
  - Central Analysis Farm (CAF)
  - Production (reconstruction) farm
  - Data archive, tapes, tape drives
  - Databases
  - Central interactive computers
  - Networks
- Use a model to project demand
  - Budget estimates based upon cost of satisfying demand
- Main issue
  - Large increase in data logging rate drives budget increases
  - Moving aggressively to expand and exploit off-site resources



Run 2 computing requirements

- Budget guidance
  - Assume approximately level funding of about $1.5 M per year
- Basic strategy
  - Estimate total computing required to meet analysis needs
  - Divide requirements between FNAL and remote contributions from collaborating institutions
    - Will show contributions in the following
  - Some institutions continue to locate equipment at FNAL
    - Used to augment existing resources
  - Official MC production has long been performed off-site
- Use same basic model as that used last year
  - Based upon a simple analysis model
  - Resource demands scale with size of dataset, event logging rate
  - Includes observed operational efficiencies, life-cycle replacements for CPU and cache disk

Computing requirements model

- Summary of requirements model
  - Data logging model
    - Upgrade logging rate to 35 MB/s in FY05, to 60 MB/s in FY06
    - Machine efficiency = 30%. Log data at 70% of peak rate
  - Analysis CPU demand scales with size of datasets
    - High-Pt datasets: allow 200 users to analyze 5 nb dataset in one day
    - Low-Pt datasets: 15 users analyze non-high Pt datasets in 25 days
  - Disk requirements scale with total number of events
    - Scale FY2004 volume.
  - Tape archive
    - I/O rate dominated by analysis
      - Scales with size of datasets assuming fixed cache hit rate
    - Volume includes raw data, production output, secondary and MC datasets, 20% contingency
  - Reconstruction farm
    - Requirements scale with data logging rate and needs of re-processing
    - Re-processing difficult to account for since it is episodic
      - Farm upgrade allows expansion into CAF, and CAF expansion into farm as needed

Computing requirements model

- Testing the model
  - Predictions of model tested against Winter 2003 resource utilization
    - Probable resource surplus during this period
  - Utilization of existing resources is high
    - Long job queues when model predicts
    - Short job queues when model predicts
    - No hoards of angry users outside the gates
  - Computing not the limitation to producing physics results
- Limitations
  - Ad hoc assumptions about usage patterns
    - Recent effort under way to understand analysis model, usage patterns
    - Promises to greatly improve underlying assumptions
  - Does not predict cache hit rate
    - Precludes optimization of disk cache and tape drives
  - Requirements for MC not explicitly included
    - MC demand scales with data volume
  - Difficult to test when resources constrained

Computing requirements model

- Current issues driving cost
  - Anticipated 50% increase in event logging rate did not materialize in FY04
  - Logging rate increases
    - 35 MB/s in FY2005
      - Further increases appear possible
    - 60 MB/s in FY2006
  - Typical 18 month Moore's law factor for CPU did not occur
    - Speed increased by only 10% across FY2004
  - Re-processed most of raw data twice
    - Three copies of production output for about 50% of the data
    - Anticipated re-processing half of data once

Total computing requirements

	Assumed conditions				Total requirements				
FY	Int L. (fb^-1)	Evts (10^9)	Peak rate (MB/s)	(Hz)	Ana (THz)	Reco (THz)	Disk (PB)	Tape I/O (GB/s)	Tape Vol (PB)
03A	0.30	0.6	20	80	1.5	0.5	0.2	0.2	0.4
04A	0.68	1.1	20	80	3.2	0.7	0.3	0.5	1.0
05E	1.2	2.4	35	220	7.2	1.4	0.7	0.9	2.0
06E	2.7	4.7	60	360	16	1.0	1.2	1.9	3.3
07E	4.4	7.1	60	360	26	2.8	1.8	3.0	4.9

A = actual E = estimated

- Analysis CPU and disk needs scale approximately with number of event
- Changes in logging rate in FY2005 and FY2005

Total equipment budget

- Estimate cost of meeting the total requirements
  - Actuals include FNAL expenditures only

FY	CAF CPU ($M)	Inter. CPU ($M)	Farm CPU ($M)	DB ($M)	Tape Drives ($M)	Disk ($M)	Network ($M)	Total ($M)
03A	0.31	0.08	0.13	0.15	0.20	0.34	0.23	1.4
04A	0.49	0.06	0.24	0.07	0.13	0.14	0.19	1.3
05E	1.2	0.10	0.13	0.05	0.43	0.50	0.31	2.7
06E	1.7	0.10	0.03	0.03	0.48	0.57	0.12	3.0
07E	1.3	0.10	0.24	0.03	0.57	0.38	0.08	2.7

- Cost dominated by analysis CPU, tape drives and disk needs

CAF procurements: Fermilab

- Cost model
  - Nodes = $2.2 k. Added $20k in FY2004 for head node replacement.
  - Nodes retired after 3 years
  - Locate 25% of capacity off-site in FY2004, 50% thereafter

FY	Total Need (THz)	Off-site (THz)	New Duals	Speed (GHz)	Total CPU	Total Cost ($M)
03A	1.5	-	159	2.2	1.3	0.31
04A	2.7	0.7	+195-31	2.8	2.3	0.49
05E	7.2	3.6	+271-200	3.9	3.6	0.42
06E	16	8.0	+386-66	6.2	8.1	0.85
07E	26	13	+332-367	9.9	12	0.73

- Logging rate upgrades drive demand beyond FNAL budget
- Estimate for FY2004 down by about 1 THz from last year's estimate

Disk procurements: Fermilab

- Cost model
  - Assume constant $15k per fileserver
  - Capacity doubles every 18 months
  - Retire servers after 3 years
  - Locate about 50% of requirements at FNAL

FY	Est. Need (TB)	New Servers	Server Size (TB)	New Size (TB)	Total Size (TB)	Total Cost ($M)
03A	180	18	5	90	204	0.34
04A	320	8	8	64	300	0.14
05E	490	+19-42	13	+240-84	480	0.29
06E	720	+18-21	20	+360-110	730	0.27
07E	1100	+11-18	32	+350-140	940	0.17

- Need more study of disk needs in distributed computing model

CAF and disk procurements: non-Fermilab

- Some CPU and disk contributed by collaboration located on-site
  - By policy, not counted against base requirements
  - Contributions expected to decline
    - Large contributors will likely create or add to remote dCAF's

	On-site contributions					Off-site contributions		
FY	New Nodes (THz)	Total CPU (THz)	New Servers	Total Disk (TB)	Cost ($M)	CPU needed (THz)	CPU (THz)	Disk (TB)
03A	63	0.65	4	90	0.19	-	-	-
04A	45	0.90	5	121	0.18	0.7	1.6	60
05E	90	1.5	5	186	0.23	3.6	2.1	80
06E	?	>1.5	?	>186	?	8.0	>2.1	?
07E	?	?	?	?	?	13	?	?

Tape drive procurements

FY	Est. Archive I/O (MB/s)	Tape Cap. (GB)	New Drives	Drive I/O Rate (MB/s)	Total Drives	Total I/O (MB/s)	Total Cost ($M)
03A	190	200	+3B	10 – 30	10A + 13B	490	0.20
04A	410	200	+5B – 10A	30	18B	540	0.13
05E	940	200	13B	30	31B	930	0.43
06E	1900	400	16X	60	31B + 16X	1900	0.48
07E	3000	400	19X	60	31B + 35X	3000	0.57

- Cost model
  - STK 9940B drives = $30k
  - Migration to new technology “X” postponed to FY2006
    - 400 GB tapes, 60 MB/s I/O rate
- Will need to find new robot space in FY2005 unless significant tape re-cycling
- Earlier migration to higher I/O probably reduces cost

Production farm procurements

- Cost model
  - Same as for CAF
  - Add $25k in FY2004 for head node replacement

FY	Est. Need (THz)	New Duals	Speed (GHz)	Total CPU (THz)	Total Cost ($M)
03A	480	+64 – 73	2.2	525	0.19
04A	1100	+80 – 64	3.0	1100	0.24
05E	1400	+80 – 64	3.9	1500	0.18
06E	1200	+0 – 64	6.2	1300	0
07E	2600	+80 – 64	9.9	2600	0.18

- Large re-processing fraction drives increase in FY2004 est.
- Drop in re-processing fraction compensates for FY2006 logging rate incr.

Network procurements

- FCC network
  - Cost driven by CAF expansion, infrastructure required for move to HDCF
  - Network topology re-assessed due to large physical separation of resources
- Trailer network
  - Planned $110k FY2004 expenditure to begin upgrade to gigE deferred to FY2005

FY	FCC Cost ($M)	Trailer Cost ($M)	Total Cost ($M)
03A	0.23	-	0.23
04A	0.19	-	0.19
05E	0.09	0.22	0.31
06E	0.06	0.06	0.12
07E	0.04	0.04	0.08

DB, interactive and miscellaneous procurements

- Databases
  - Existing replicas and new FroNtier servers adequate for life of experiment
- Interactive CPU includes login pool, code build machines, home disk, etc.
- Misc. includes some equipment needed for move to HDCF

FY	DB Cost ($M)	Int. CPU ($M)	Misc ($M)	Total Cost ($M)
03A	0.15	0.08	0.02	0.25
04A	0.07	0.06	0.07	0.20
05E	0.05	0.10	0.05	0.20
06E	0.03	0.10	0.05	0.18
07E	0.03	0.10	0.05	0.18

Fermilab equipment expenditure summary

- Total proposed expenditures for equipment at FNAL

FY	CAF CPU ($M)	Inter. CPU ($M)	Farm CPU ($M)	DB ($M)	Tape Drives ($M)	Disk ($M)	Network ($M)	Misc ($M)	Total ($M)
03A	0.31	0.08	0.19	0.15	0.20	0.34	0.23	0.02	1.5
04A	0.49	0.06	0.24	0.07	0.13	0.14	0.19	0.07	1.4
05E	0.42	0.10	0.18	0.05	0.43	0.29	0.31	0.05	1.8
06E	0.85	0.10	-	0.03	0.51	0.27	0.12	0.05	1.9
07E	0.73	0.10	0.18	0.03	0.48	0.17	0.08	0.05	1.8

Tapes and operating

FY	Archive Volume (PB)	T9940A (PB)	T9940B (PB)	X (PB)	Tape Cost ($M)	Misc Operating ($M)	Total Cost ($M)
03A	0.40	0.22	0.24	-	0.18	0.18	0.36
04A	0.98	-	-	-			0
05E	2.0	-	0.59	-	0.22	0.18	0.40
06E	3.3	-	-	1.3	0.25	0.18	0.43
07E	4.9	-	-	1.6	0.31	0.18	0.49

- Misc. operating taken from historical average
  - Covers desktops, installs, consultants, etc.
- Re-processing in FY2004 increased archive volume over 2003 estimate despite reduced logging rate
- Tape density migration would cost about $300k if started in mid-FY2005 (assuming 400 GB tapes at $75 each)

Summary

- Requirements model works well
  - Data logging rate still the main concern, origin of increased resource demand
  - Estimated cost exceeds guidance by about $300k in FY2005
- Remote resources deployed to meet significant fraction of needs
  - Effectively extends FNAL computing budget
  - Much work remains to make effective use of these resources
    - Will require GRID technologies
    - Need to focus and add effort here
- Cost mitigation strategies
  - Adopt higher density tapes ASAP
  - Aggressively re-cycle existing tapes
  - Optimize balance between disk cache and need for archive I/O
    - Needs further study
  - Improve understanding of analysis model
    - Interim report from task force already completed
  - Defer trailer networking upgrades