Three Topics Today

- *Status report*
- *CDF-GRID*
  - a new request here
- *Metabolismo per analisi*

CNAF hardware

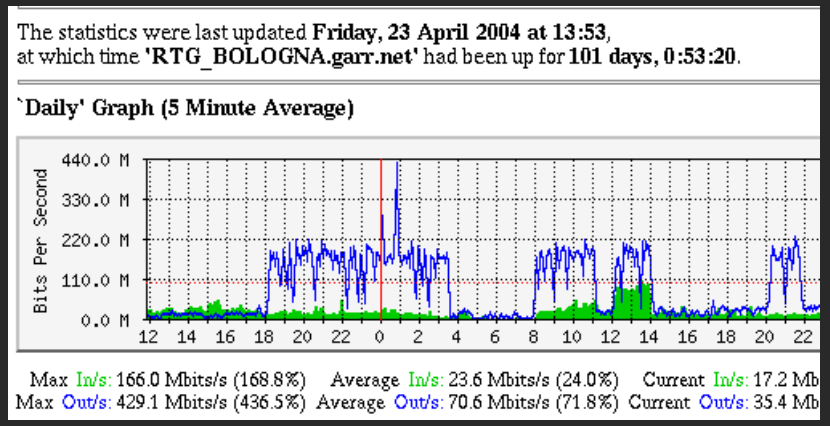
	Assign.	money	purch	delivery	Install	Status
Tier1	6 duals + 800GB	Tier1 gift 2003		Started with these in February 2003		Duals OK Disk broken
2003 money	40 duals	28KEu sep 02 + 114KEu (from s.j.) may 03	48 duals	July 03 (12) Jan 04 (36)	Jan 04	Up & Running
	4 TB		8.5TB	Nov 2003	Feb 04 (6TB) 2.5 TB ?	Up & Running. Too small
2004 money	700 GHz ~120 2x3GHz	Tier1	all	June ? Contract still to be signed	?	Wait & Hope
	30 TB		all	July ? Approved by CD one month later then cpu	?	Wait & Hope

How nice is to be at Tier1 ?

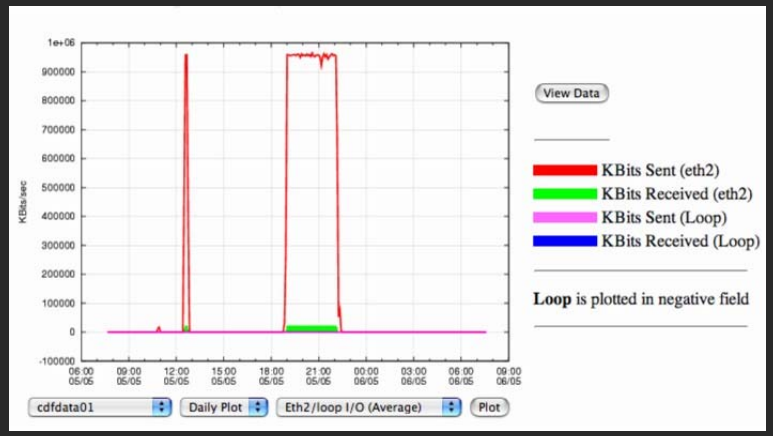
- Advantages
  - Build upon existing infrastructure: room, network, power, cooling
  - CNAF provides system-management (CDF used ~0.5FTE so far, mostly spent in setting up and debugging hw and file server performance and non-CDF specific troubles)
  - A.R. for CDF (selection in progress)
  - Informal decision process: flexibility to accommodate our reqs
- Drawbacks
  - Large acquisitions = long times
    - 2004 purchase started in Nov 2003, hope hw in by July
  - Understaffed, overloaded, overcommitted personnel
    - 3TB FC disk purchased by CDF in Nov 2003, still not operative
  - Informal decision process: never clear what will really get when
  - Constant pressure to integrate into LCG framework
    - what exactly is the deal we have ?

CNAF performance: data → CPU : OK

- Data import : 1TB/day
  - ~120Mbit/sec
  - OK
- Data export :
  - → output at FNAL
  - 200Mbits/sec achieved



- Data analysis:
  - Problem :
    - >100 processes read from same disk… performance drop to zero
  - Solution (home made):
    - Files are copied on worker node scratch disk and opened there
    - Queuing tool limits to 20 copies at the same time → file server feeds at 110MByte/sec (950Mbit/sec)
    - e.g standard at Ian Bird's lab



Technical note of the day: file fragmentation

- In september it was: data flow disk → cpu
- We are spending an awful amount of time struggling with file server performance issues
  - Well known by now that single stream data transfer is limited by link latency, not bandwidth
  - 15 parallel gridFtp used for previous slide "1TB/day"
  - Many write streams → fragmented files → slow read
  - spent one month on xfs
    back to ext3 + "hand defgragmentation"
    - Very disgusting situation
    - Help welcome

The next frontier: the common pool

- CNAF/Tier1 wants a global common pool of CPUs
  - Access via common batch system (PBS now)
  - For each experiment:
    - minimum guaranteed
    - maximum allowed
  - Start with ~50% or resources there
  - Not so secret plan to put all CPU in this pool
- CDF needs to do some work, can not rely on future grid tools
  - Present manpower on this : ~1/5 of sb
- A.R. for CDF support will take this as main task
- Still may not have this ready before new hw arrives

Bottom lines for CDF @ CNAF

- So far so good
  - Glad to have avoided a "CDF-Italy farm"
  - Do not regret "all computers at Fermilab", yet
- One question for the near future
  - We are working to change batch system from FBSNG to PBS
- If not "PBS ready" when the promised 700GHz are here, two options:
  - 1. do not use hw (CSN1 asked for this to be up by May) while working on sw
  - 2. put hw in present farm while working on sw
- Who should decide ?
  - CDF Italy ?
  - Tier1 Director ?
  - CSN1 ?

Now that we have a farm…

… let's put it on the grid

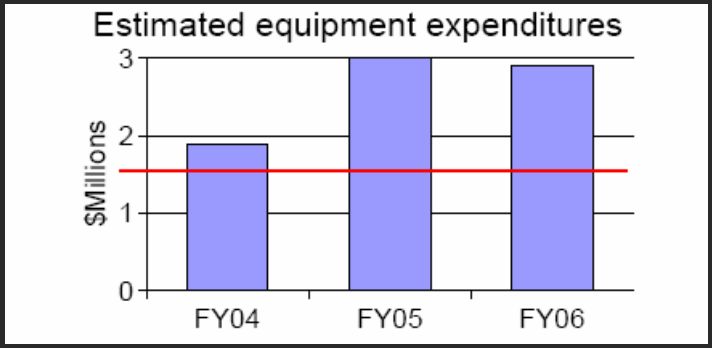
CDF-GRID

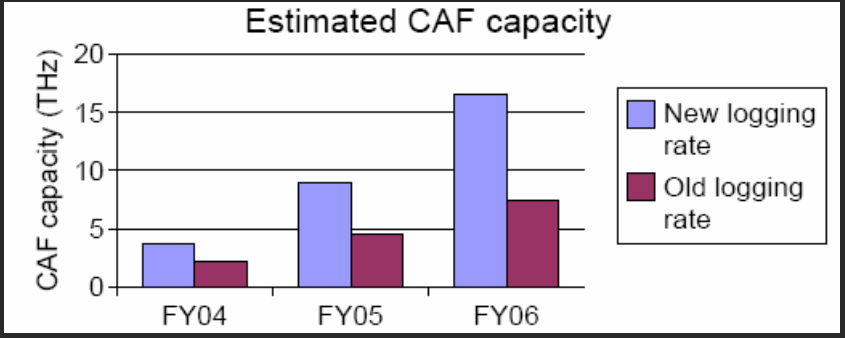
Less functionalities then LHC-Grid, but

- Works now
- Tuned on user's needs
  - ~~Object~~ Goal Oriented software
- Costs little

The landscape

- DAQ data logging upgrade
  - More data = more physiscs
  - Approved by FNAL's Physics Advisor Committee and Director
- Computing needs grow, but DOE/Fnal-CD budget flat





- CDF proposal: do offsite 50% of analysis work
  - CDF-GRID
    - We have a plan on how to do it
    - We have most tools in use already
    - We are working on missing ones (ready by end of year)
- Our proposal: do 15% of analysis work in Italy

possible !

CDF-GRID Ship is sailing

- CDF-GRID is de-facto our working environment and hypothesis
- Analysis farm built/developed to be clonable
- Large effort on tools usable both on- and off-site
  - data access (SAM, dCache)
  - remote / multi-level DB servers
  - Store from Italy to tape at FNAL
- User's MC at remote sites = reality
- Analysis on remote-copied data samples based on SAM
  - Up and working, already used for physics !
    - ~all work done like this in Germany, but access to locals only
  - INFN: limited test data so far (30TB requested in Sept 2003)
    - provides access to all CDF (680 users)
- Making analysis at CNAF as easy as at FNAL is taking all our time (possible → working → easy)

Hardware resources in CDF-GRID

site	GHz now	TB now	GHz Summer	TB Summer	Notes
INFN	250	5	950	30	Priority to INFN users
Taiwan	100	2.5	150	2.5	
Japan	-	-	150	6	
Korea	120	-	120	-	
Germany GridKa	~200	16	~240	18	Min. guaranteed CPU from x8 larger pool. Open to all by ~Dec (JIM)
Cantabria	30	1	60	2	~1 months away
UCSD	280	5	280	5	Days away. Pools resources from several US groups. Min guaranteed from x2 larger farm (CDF+CMS)
Rutgers	100	4	400	4	In-kind, will do MC production
MIT	-	-	200	-	~1 month away
UK	-	-	400	-	Open to all by ~Dec (JIM), access to larger common pool
Canada	240+	-	240+	-	In-kind, doing MC production, access to larger common pool

Evolution of farm at CNAF

- Proposal for analysis & MC farm at CNAF growth
  - **Modest increase in 2005/6 driven by increased data sample**
    - we are doing fine now : thank you !
    - future needs always uncertain
    - Tevatron OK DAQ upgrade lagging
    - Usage so far OK Large MC production still looming
    - 90% of work done at FNAL But our FNAL share will not grow
  - Count on our usage to average at ~70%
  - Donate 30% to CDF-Grid (let the other 600+ users fill our gaps)
  - **Add more CPU for CDF-GRID** (use same disk as we do)
- Plan to fill a bit less of present estimate of CDF
  - Force optimization of usage
  - **Shoot to cover 15% of needs, not of estimates**
- Be prepared to add more resources if needed
  - A large common CPU pool at CNAF will help

proposed INFN contribution to CDF-GRID

- CDF ANALYSIS HARDWARE PLAN (guideline, not Bible)

	CDF ANALYSIS NEEDS			15%		
Year	GHz	TB	K$	GHz	TB	K$
2004	3700	300	960	555	45	144
2005	9000	600	1800	1350	90	270
2006	16500	1100	1590	2475	165	239

- ROADMAP FOR CNAF FARM

	CDF FARM AT CNAF						
	for INFN physicists		for CDF grid			CNAF	Notes
Year	GHz	TB	30% of our CPU	GHz to add	GRID GHz	tot GHz for CDF	
2004	950	38.5	285	200	485	1150	"for INFN" already payed
2005	1500	90	450	600	1050	2100	discuss in Assisi
2006	2000	150	600	1500	2100	3500	discuss in 2005

- Presented to IFC meeting April 16, next slide

IFC response

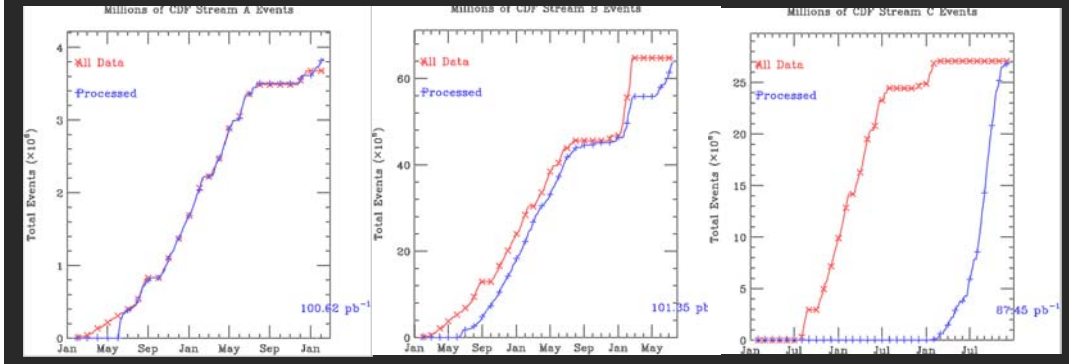
- Moving 50% of analysis offsite = Good Plan
- Contribution to CDF Grid on a voluntary base and separate from running costs
- INFN contribution to 15% of total: reasonable and welcome
- CDF needs to show real effort on curbing needs

Are computing needs under control ?

- CDF accepted criticism and will act
  - E.g. optimization of widely used vertex fitting algorithm
- Reconstruction code already OK 2sec/event (10x faster then D0)
- 3 reasons behind needs
  - Technical: OO and general unoptimized code, room for improvements, but … reconstruction time within x2 of '97 est.
  - Sociology: freedom to try, test, err, learn, explore… pays.
  - Physics: we are doing **more better physics faster**
    - >45 papers by 2004 vs ~20/year in the '90's

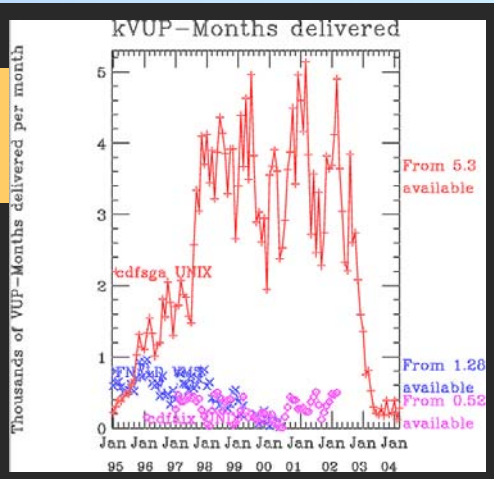
- Present resources not a constrain to physics, but 100% used
  - the way it should be.
- **It works, don't break it !**
  - Let's keep up growing with data size and keep a tight but soft rein
  - Be prepared to add (or subtract) resources if needed

Run1 (Jan'94 Feb'96) vs Run2. 2003 ~ 1996



analys cpu
~3600 Mips

x3



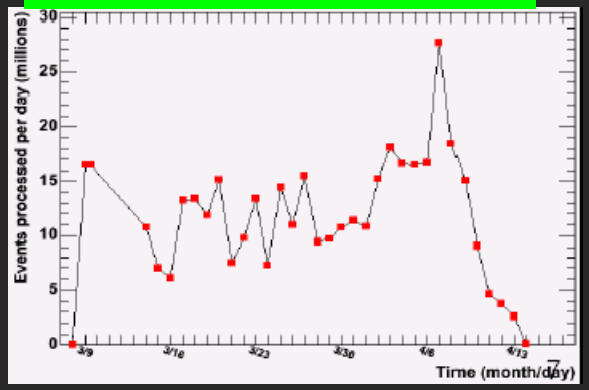
Run1b reconstruction
1.3Mev/week ~2Hz
0.6GHz-sec per event

Recofarm
~ 1200 Mips

RUN 2
- ✓ Same ana/reco ratio
- ✓ Code 4x more time
  more complex detector
- ✓ 100x more events

Run2 reconstruction
20MEv/day ~200Hz
2GHz-sec per event

Recofarm
~500GHz



x3

Analysis CPU
~1500GHz

Conclusion

- CDF is building a MC and Analysis grid
- It is a lot of work for fnal/cdf caf/sam/dh/db/jim teams
- People are working hard for this:
  - Implement and use an analysis grid 4 years before LHC
  - Working in close relation but not as part of LHC-Grid (so far)
  - LHC will benefit from feedback and user's case
  - Not obvious that code developed for CDF will be part of LHC grid nor viceversa
- Clear commitment and timelines for deployment of significant offsite resources makes this effort more appealing and adds deadlines to developer's motivation
- Integration with LHC/LCG has to be an advantage not a slowing factor

The request

- Add 200GHz in summer 2004 to dedicate to CDF-GRID
  - Keep priority for INFN physicsts on the 950 we already have
- Implemented as additional CDF quota from common Tier1 pool
- CSN1 should request this to Tier1
- On CDF the burden to become "PBS compliant"

L'interattivo: il problema

- Computers e dischi nelle sezioni per lavoro di sviluppo codice, paw/root, etc. "**l'interattivo**"
  - Lavoro **FONDAMENTALE**
  - CNAF = BATCH
- Pochi soldi, tante discussioni, tendenza al micro-management

- Ogni situazione locale e' diversa
  - PC "cicciuti", piccoli cluster locali, farm di sezione …
  - dischi USB/IDE/SCSI/FC…
- dipende da:
  - Dimensioni del gruppo
  - Storia
  - Scelte del gruppo calcolo locale
  - Collaborazione con altri gruppi (anche non in CSN1)

- Alla fine lasciare liberta' di azione paga

L'interattivo: la proposta

- Metabolismo per analisi (inventariabile):
  - Una dotazione su inventariabile piccola, ma adeguata, definita "per una persona attiva sull'analisi"
  - Assegnazione ad ogni sezione su inventariabile ottenuta moltiplicando per il numero di tali persone
  - Una quota indivisa nazionale s.j. a disposizione del coord.nazionale per risolvere emergenze e mediare fluttuazioni
  - La dotazione individuale e' stabilita dai referees
  - Il numero di persone e' indicato dal capogruppo locale e verificabile dai referees (note, presentazioni, articoli, documentazione interna, incontri…)

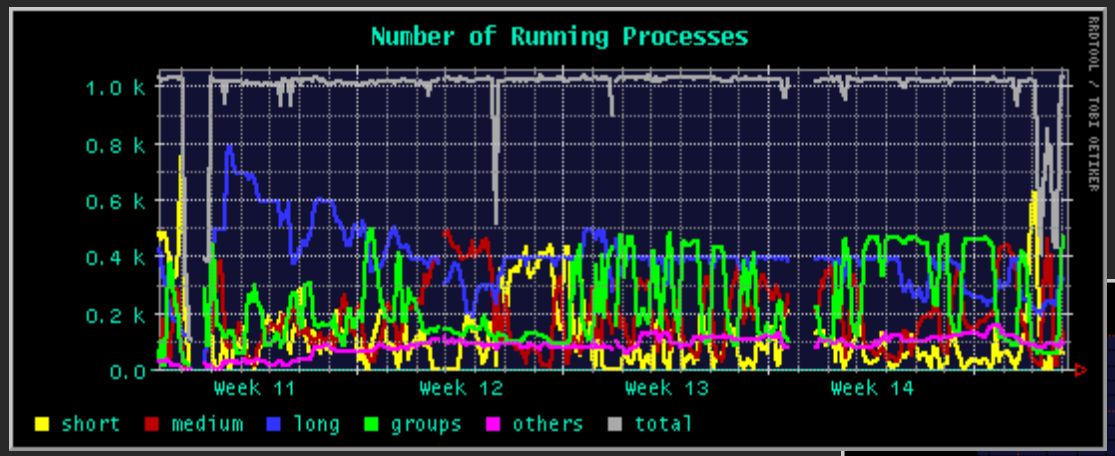
- Se la Commissione e d'accordo, prepareremo i moduli 2005 secondo queste linee e discuteremo a settembre i numeri

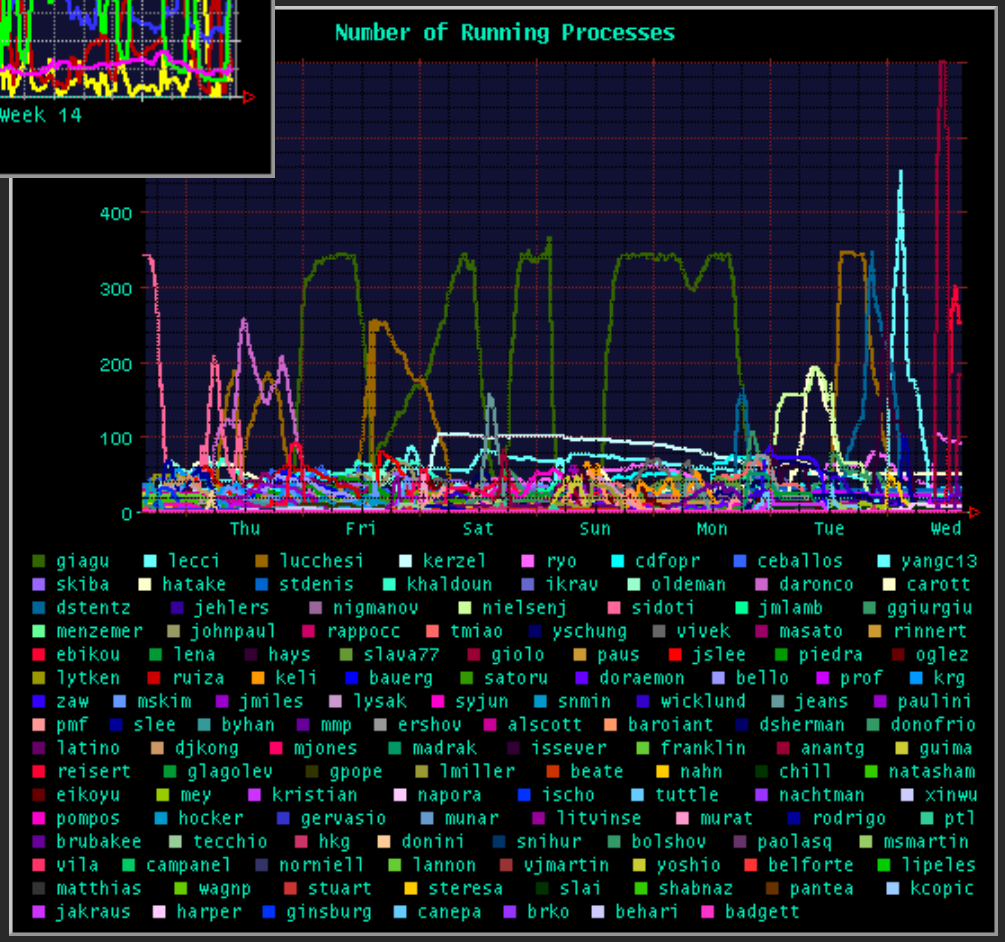
spares

The tools

- de-centralized CDF Analysis Farm
  - Develop code anywhere (laptop is supported)
  - Submit to FNAL or CNAF or Taiwan or SanDiego or…
  - Get output ~everywhere (most desktops OK)
- SAM
  - Manages metadata and data replicas
- FBSNG
  - FNAL's own batch system
  - Being replaced by Condor (US) or PBS (CNAF) or both
- JIM
  - Will move authentication from kerberos to certificates
  - Prerequisite for opening UK and German computers to all CDF'ers
  - Tying access to certificates is a major slowdown in delivering resources to users
  - CNAF (and others) who accepted kerberos are sailing fast

Monitor 1: what, who





Each remote CAF runs software that makes this kind of plots on the web

Monitor 2: to do what

- Analysis code logs data set access

```
Data access summary

Datasets: aexp08,hbot0h


INPUT data summary:

           RecRead EvtRead RO(sec) OC(sec) Size(MB) KbPerRec KbPerEvt FailOpen
Aggregate  7.8e+04 7.8e+04      26   18405  8.3e+03       --       --        0
Average    1.6e+04 1.6e+04     5.2  3681.0  1.7e+03    108.7    108.7        0


OUTPUT data summary:

           RecWrote EvtWrote OC(sec) Size(MB) KbPerRec KbPerEvt
Aggregate   2.3e+05  2.3e+05   55308  2.5e+04       --       --
Average     4.7e+04  4.7e+04 11061.6  5.1e+03    111.4    111.5
```

- CAF software collects name of data set accessed by users, amount of data read, data written, cpu time, real time

- Existing tools allow to tell
  - What resources are there
  - Who is using them …
  - … to look at which data