Status Report and Plans for the future of CDF Italy computing

What hardware we have

How to use it

What will be next

CDF - Italy meeting
Pisa - 10 May 2002
Stefano Belforte

Summary

- What was bought
- How is used
  - Decide on usage of next chunk
- The new CAF
- Purchase plans at FNAL for 2002
- Review of overall computing plan
  - FNAL → CNAF
  - Sezioni → CNAF
- CNAF:
  - What we can have
  - Decide on who/how to use
- GRID
  - Status
  - Work to do

Introduction

- Time to do physics !
- 8 years ahead of data at knowledge frontier
  - Wise plan: invest some work now
    - ☞ Computers are an extension of the detector
    - ☞ Work hard to build
    - ☞ Relax and use it
- Keep the final goal in mind
  - Balance the work on present (poor) data with need not to endanger future higher quality years
  - Plan for 5-6 years of leisured data analysis

- Didn't we build enough already ?

The Message

- Much has been done
- Much more still has to be done

- Plentyfull computing resources
  - Always a need. Never a reality
- Competition is strong and well organised

- We need few good people : YOU !

- The group has to make clear the need and the reward
  - Support our frontline soldiers

Purchased Hardware
(2000-1-2 = 276MLit vs. 238 assigned by CSN1)

- **650GB** disk on fcdfsgi2 for MISCELLANEA 38ML
  - Symbolic links from /cdf/home/belforte/data
  - Write by Unix groups: cdfuitAd, B, C
- **1440 GB** disk on fcdfsgi2 for DATA SETS 44ML
  - In 2 weeks ?
  - Access as above
- **1440 GB** disk on fcdfsgi2 for common usage 44ML
- **8.8TB** (4 file servers) in CAF stage 1 90ML
  - By end of may ?
  - Access ? Likely by user
- **10 dual processors** (2 x 1.26GHZ) in CAF stage 1 60ML
  - Delivery to FNAL past due
  - Access ? priority queus: ilong, imed, ishort

Comments on computing budget

- All money was spent
- All money was spent on the project
- Additional money was brought on the project from Trieste
  - Tape drives financed but not bought (most)
  - Savings from other projects (little)
  - Remainings from other groups (very little)
- Excellent relationship with INFN referees and CDF offline management

- Projet financial management is working well
  - Total transparency/accountability
  - In spite of somebody's worries
- Too much enthusiasm
  - 40ML must be returned to Trieste group

Disk on fcdfsgi2

GB	Present	Managed	Used	Proposal		When	Note
100	Bottom1	Giagu	96%	SVT	Punzi ?	1.+2w	
122	Bottom2	Giagu	83%	SVT	"	1.+2w	
100	Svt_data1	Punzi	96%	SVT	"	-	
50	Svt_data2	Punzi	88%	SVT	"	-	
50	Svt_data3	Punzi	92%	SVT	"	-	
50	Svt_User	Belforte	97%	Spare/CHA/ISL	Belforte	1.+2w	
122	Top_1	Castro	91%	Spare	Belforte		
50	Spare	Belforte	1%	H→tau mu	Vataga	Now	
800	To install			Bottom	Giagu	1.	Split?
300	To install			Top	Castro	1.	
100	To install			Z→bbar + ….	Castro		
100	To install			Spare	"		Less?
144	To install			Spare	"		Split?

SVT: 200 → 422 Bottom: 222 → 800 Top: 122 → 300

Spare for: minbias, jets, …

Interactive work: a question to you

- Richard Hughes committee: Stefano Giagu for Italy
- The party line (and what most US groups will do):
  - "nothing" at FCC
  - Trailer desk= 4x1.7GHz (=1/3 old-fcdfsgi2) + 600GB
    - ☞ LCD screen 800$
    - ☞ DualAthlon, 2x160GB disk 3500$ (put 2 on each desk)
    - ☞ 8K$/desk = 10KE = 20MLit/user
- We promised (were forced) to do it in Italy (got money also)
  - each group defends his needs
    - ☞ not my problem
  - **Make it a global issue ?**
  - Big numbers require big talks, plans, reports…
    - ☞ What is the status ? What the real need ?
- More on interactive later

The new CAF

- Batch farm for analysis of:
  - 2ndary data sets (skim output)
  - Output of that (3rtiary data sets, ntuple)

- Italian contribution
  - Specification, batch configuration, batch monitor, output retrieval, betatest, money
  - Massimo Casarsa, Stefano Giagu, Igor Sfiligoi, Ombretta Pinazza, Franco Semeria, Antonio Sidoti, Paolo Mazzanti, S.B.

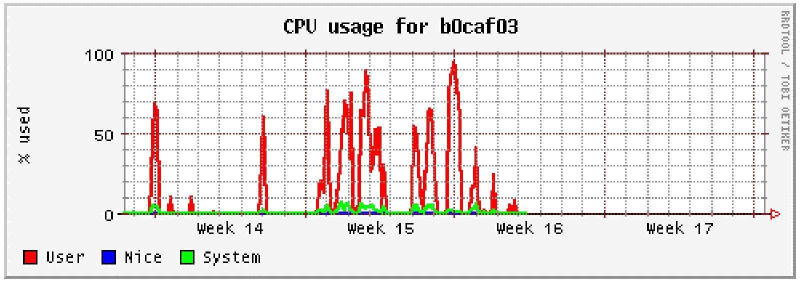
- Works ! Use it !

Interactive work on CAF

- Some/most large Root jobs can run on CAF
- CAF output on scratch/user disk accessible from trailers desktop for interactive Root
- Each user can expect O(10GB) for private use on CAF output nodes
- TB's available on CAF disk servers
- My opinion:
  - Much talk, no clear need
  - Biggest problem will be managing of large disk areas
  - Try to do it in Italy first

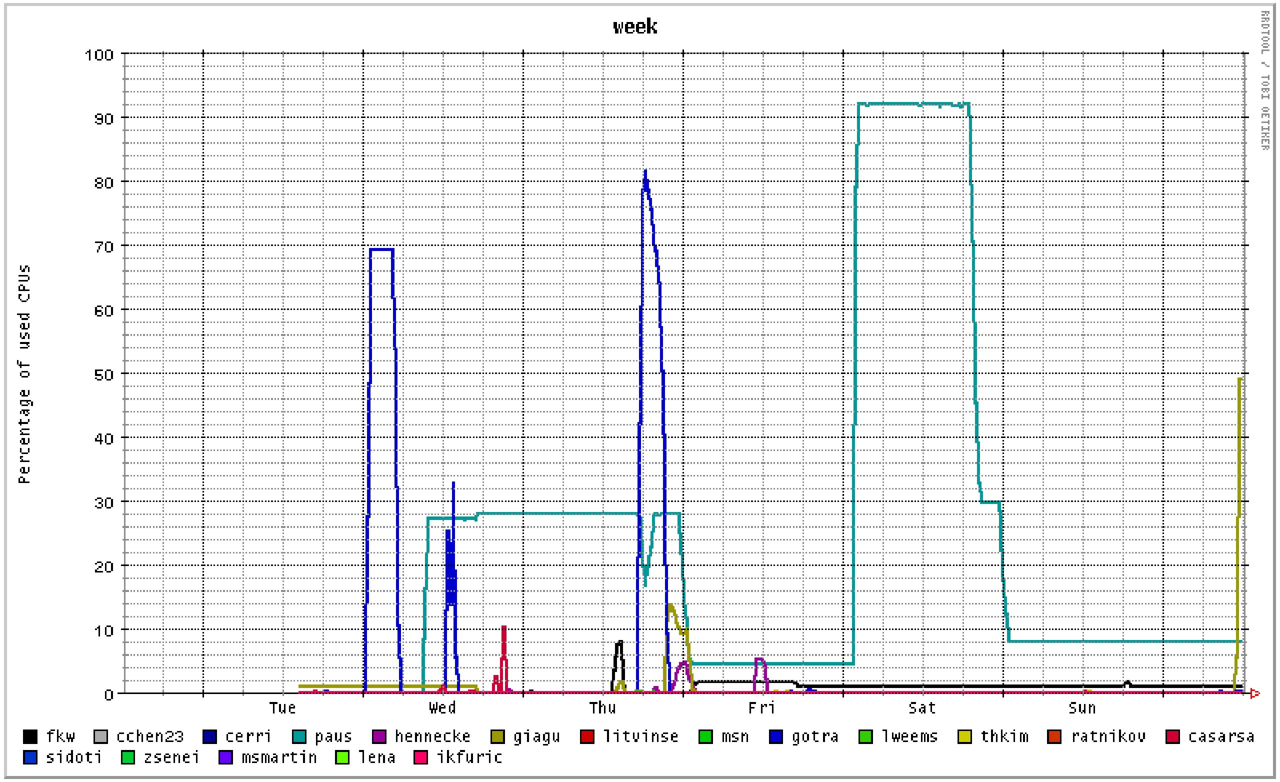
CAF prototype very little used !

- Up and running for more then a month
- 1TB disk: J/PSI and hadronic B data
- 8 dual node available there to whoever asked
- 8 with poor connection to disks used mostly for MC (C.Pauss)



- Other 7 nodes have identical profile
- Average <<50%
- Almost all work by Stefano Giagu (StreamH) and Yuri Gotra (J/Psi) in a few 3-4 few hour shots

CAF Stage 1 not better



- First week with 40 nodes
- Only Pauss is really loading it + spikes of Gotra and Giagu
- There was no hungry user waiting behind the corner

Next Purchase

- CSN1: June 24
- Letter to referees: May 30
- CAF upgrade: grow to 7 "farmlets" 240KEuro
  - +3 file servers → 7 x 2.2 = 15 TB
  - +65 dual nodes → 75 (Belforte's rule: 1 CPU / 100GB)
    - ☞ original rule was 1GHz, 50 x 1.26GHz enouh
  - File servers 11K$ each: 33 K$ = 40KEuro
  - Duals 2500K$ each : 162.5K$ = 200KEuro
- Avaibility 120KEuro s.j., needs 120 more
  - Assigned in Jan: 80KEuro
  - total request 2002 = 320KEuro
  - request September 2001: 300 KEuro
- As much disk+cpu as asked "in the plan" for all Run2a !!
  - 1999 plan was 15TB + ½ fcdfsgi2 (or 3 8-ways = 12 duals)

Following Purchase

- CSN1 September 20
- Could ask advance assignement of 2003 money
- Other O(100) Keuro possible
- What for ?

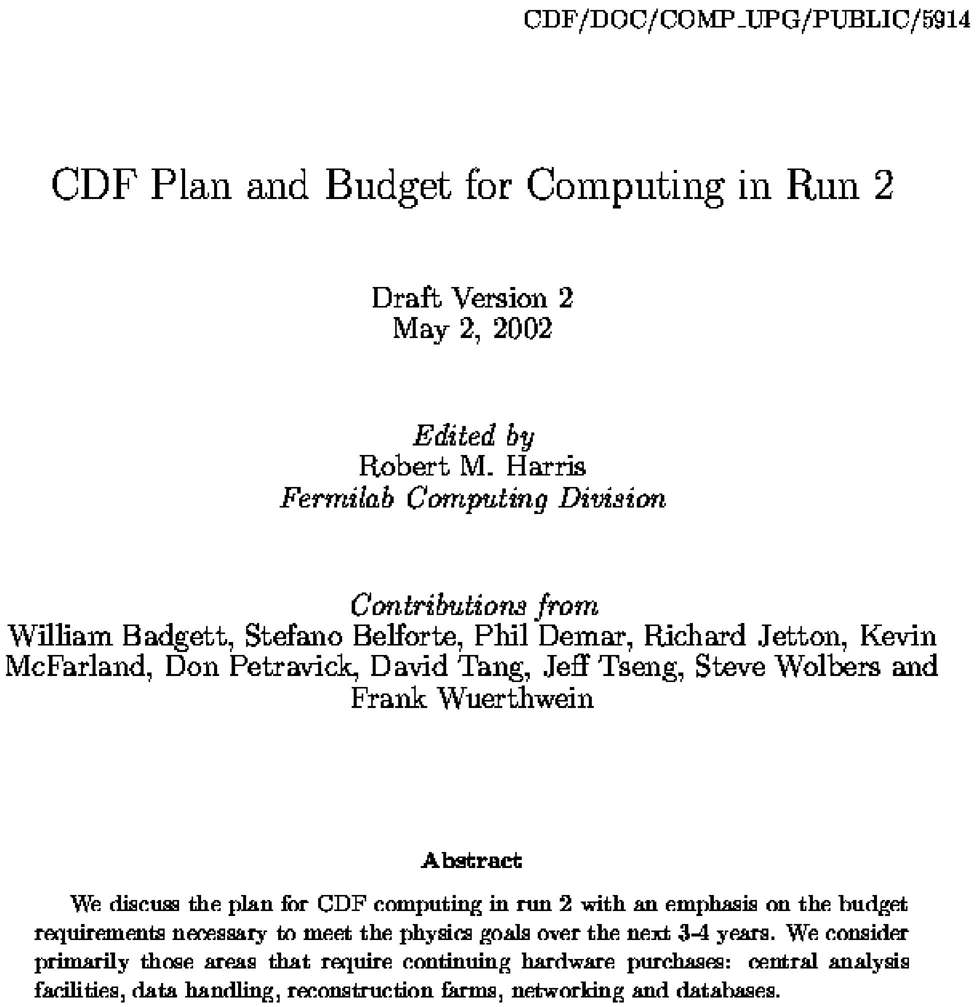
- Need
  - luminosity expectation
  - demonstration that hardware bought in 2002 is not enough to cover all of 2003 needs
  - demonstration that we can really saturate the hardware

- In 2002 INFN will already own 10% of full CAF specification for all of Run2a

Exercise for 2003

- Buy all that is needed till end of Run2a (2004+) 2.5fb-1
  - Cfr. CDF-5914 (our plan now !)
- Assume Moore's scaling: x2 every 1.5 year
  - May mean delaying purchases to end of 2003
- Add 24TB (total 38) = 7 file servers x 3.5 TB each
- Add 50 dual at 2.5GHz/CPU (total 400GHz)
- Total cost: 300KEuro
- That means buying a bit more of 10% of CDF5914 estimate for all of CDF, I.e. satisfying 20 instead of 200 users.
- This is what I want to put in 2003 requestsn
- Possibly want most of this assigned already in 2002
  - Maybe include requests to cover also 2004 needs
- To defend these numbers will need more then words

CDF 5914



CDF/DOC/COMP_UPG/PUBLIC/5914

CDF Plan and Budget for Computing in Run 2

Draft Version 2
May 2, 2002

Edited by
Robert M. Harris
Fermilab Computing Division

Contributions from
William Badgett, Stefano Belforte, Phil Demar, Richard Jetton, Kevin McFarland, Don Petravick, David Tang, Jeff Tseng, Steve Wolbers and Frank Wuerthwein

Abstract

We discuss the plan for CDF computing in run 2 with an emphasis on the budget requirements necessary to meet the physics goals over the next 3-4 years. We consider primarily those areas that require continuing hardware purchases: central analysis facilities, data handling, reconstruction farms, networking and databases.

Round Numbers

- Disco + CPU batch + 10% CPU per “interattivo”

- 2003 = 100 Keuro (1.2 fb-1)
- 2003+2004 = 300 Keuro (2.5 fb-1)
- 2003+2004+2005(1/2) = 500 Keuro (3.3 fb-1)

- Cosa metto nei moduli ?

	Ndual	GHz	tot	integ	**N-fcdfsgi2**	Nfile	TB	tot	**integ**
			GHz	cpu	**equivalent**	srver	each	disk	**TB**
2002-1	10	1.2	24	24	**1.1**	4	2.2	9	**9**
2002-2	65	1.2	156	180	**8.4**	3	2.2	7	**16**
2003/4	55	2.5	275	455	**21.3**	8	3.5	28	**44**
2005-1	45	2.5	225	680	**31.9**	7	3.5	25	**69**

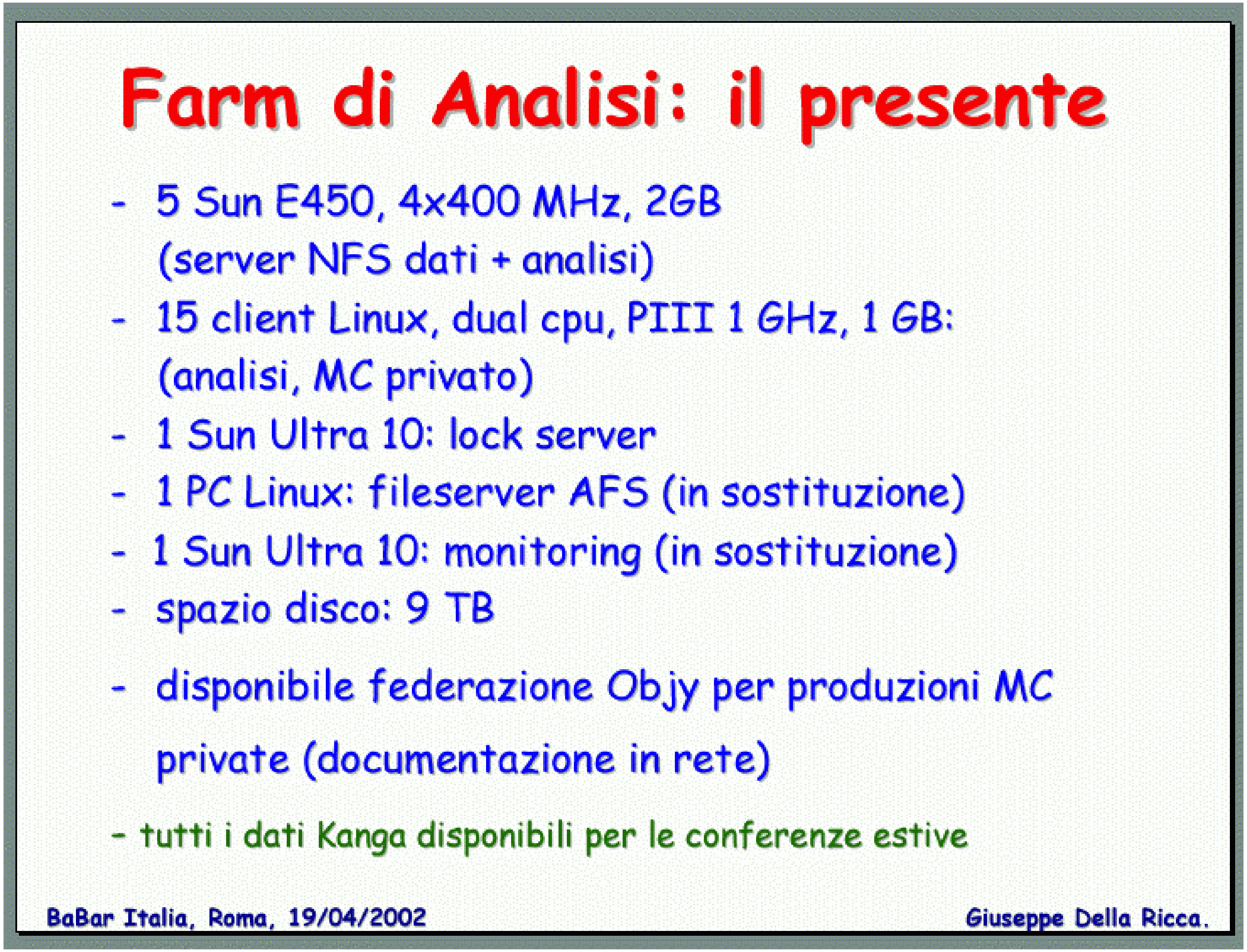
The BIG PROBLEM

- Computing for analysis is a big success
- More then enough hardware already in place
  - Luck: bad news do not come alone
    - hardware troubles : little disk so far
    - Tevatron troubles : little data so far
  - Linux saved us from C++ disaster
- Know what to buy
- System is working
- CSN1 willing to pay
- Need a case !!
  - Usage ! Usage ! ! Usage ! ! !
- Problem is not lack of computers
  is lack of people using them

One Solution

- More coordinated effort
  - Less topics, with more manpower
  - Written plans/reports: Needs, Usage, Goals, milestones
  - Demonstrations/justification of where money goes
- **IF : we build a running machine with**
  - clear direction
  - important goal
  - demonstrable progresses
- **THEN: it will be an unstoppable train**
- Learn from BaBar e.g.
- Not something I can do alone

BaBar's Analysis Farm at Caspur: 30 CPU



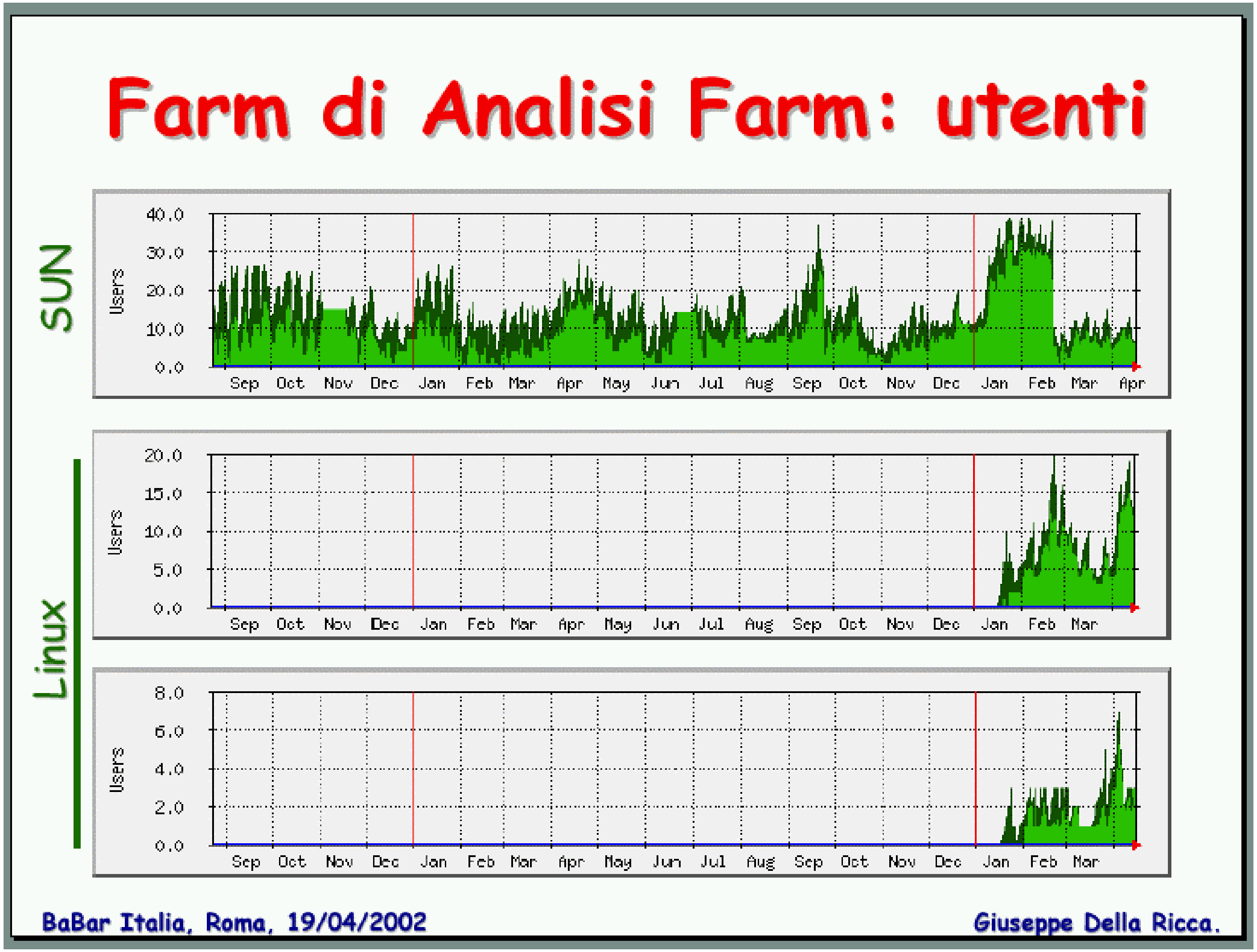
Farm di Analisi: il presente

- 5 Sun E450, 4x400 MHz, 2GB (server NFS dati + analisi)
- 15 client Linux, dual cpu, PIII 1 GHz, 1 GB: (analisi, MC privato)
- 1 Sun Ultra 10: lock server
- 1 PC Linux: fileserver AFS (in sostituzione)
- 1 Sun Ultra 10: monitoring (in sostituzione)
- spazio disco: 9 TB
- disponibile federazione Objy per produzioni MC private (documentazione in rete)
- tutti i dati Kanga disponibili per le conferenze estive

BaBar Italia, Roma, 19/04/2002

Giuseppe Della Ricca.

BaBar's Analysis Farm at Caspur: users

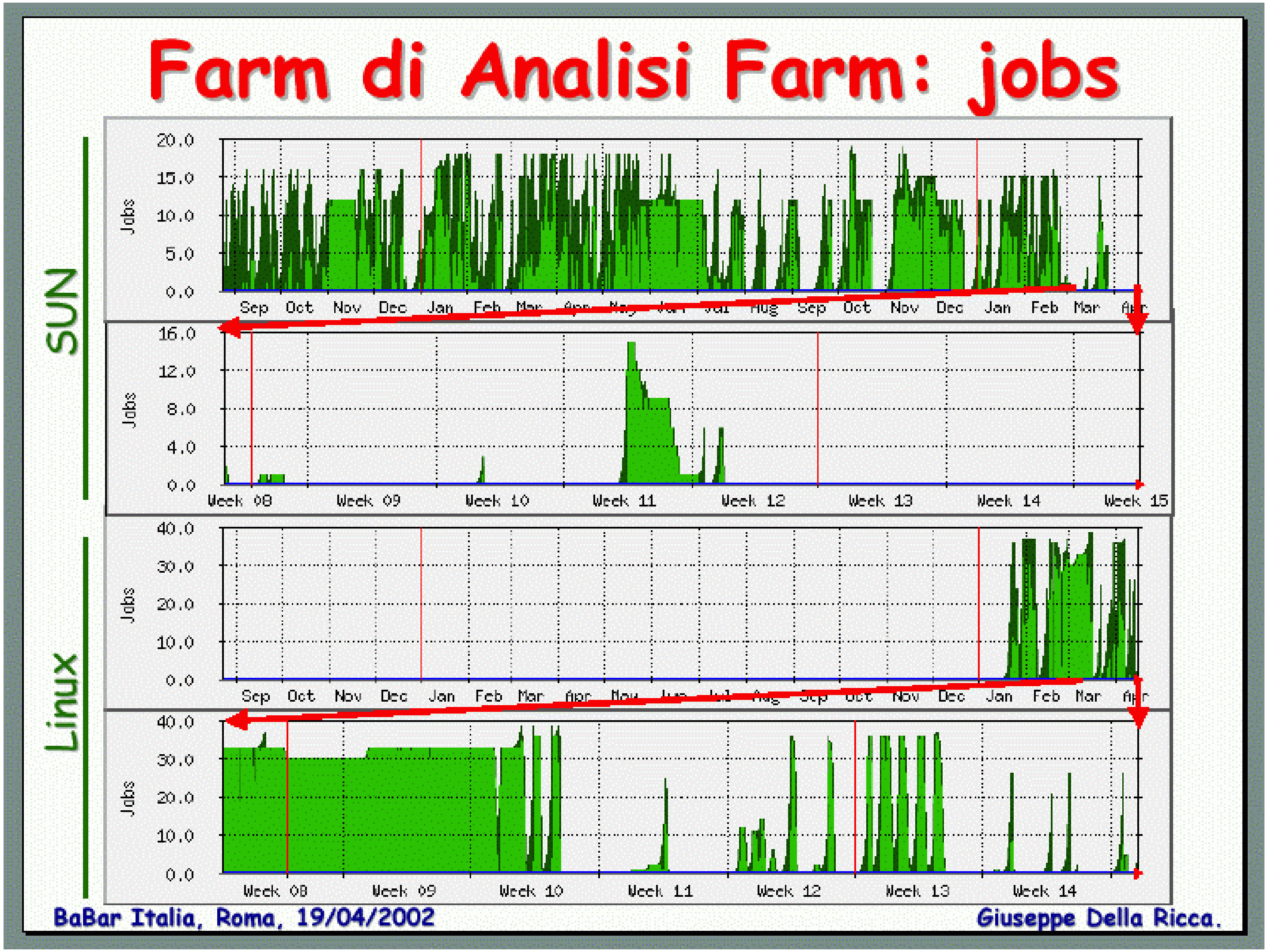


Farm di Analisi Farm: utenti

BaBar Italia, Roma, 19/04/2002

Giuseppe Della Ricca.

BaBar's Analysis Farm at Caspur: jobs



Farm di Analisi Farm: jobs

BaBar Italia, Roma, 19/04/2002

Giuseppe Della Ricca.

Another Solution

- Work on “somebody’s else hardware”:
  - CNAF Regional Center
  - GRID
- There is nothing like “idle machines waiting for us”, but in this case it “might” be easier to get resources
- Will get resources, not money
  - If we do not use them, someone else will
  - Not a bad thing
- Work on this already started and moving fast

“Move” CAF farmlets (our TierB/1) to CNAF

- Regional Center @ CNAF = INFN pet project
  - New computer room ready early 2003
  - Plans for O(1k) nodes, O(100) TB disk, big tape robot
- CSN1, CDF referees, CNAF director
  - happy to see CDF doing there most work
- Plan to build equivalent of 10~20 CAF farmlets:
  - 100~200 duals + 10~20 TB
  - Presente to CNAF group on February 5
  - Presented to Regional Center Committee on May 3
  - Received as reasonable
- Method of financing C.R. still under debate. In any case CSN1 will have to approve the requirements
- For more about CAF and CNAF:

http://www.ts.infn.it/~belforte/offline/caf/index_caf.html

Overall Hardware Needs

- Data Storage (2003 + …)
  - 10TB + 10TB/year for 2ndary/3rtiary
  - 3TB + 1TB/year for interactive
- Analysis CPU
  - 10 “1GHz” CPU / TB of data (from 2001 benchmark)
- Interactive CPU/Disk
  - 2 “up-to-date” CPU / user x 40 users
  - 300GB / user x 40 users (growing with technology)
    - ☞ size this from comparison to resources available to US students at Fermilab (typical University owned PC’s in offices: 5~7K$ per desk every 3 years)
- MonteCarlo CPU (Gen+Sim)
  - ~40 “up-to-date” CPU’s (possible underestimate)

Move interactive work to CNAF

- Reduce to mininum hardware at home
  - Last years: lot of work, little gain
  - No way to share tools/data in sight
- Common area with quotas
  - Common CPU pool
  - Easier to add new users
  - Avoid resource waste
  - Easy share of scripts/kumacs/…
- Proposed in mail to everybody one month ago
  - One enthusiastic yes

Decision time

- We will not ask money for hardware at FNAL beyond 2003
- We will ask ~20 "farmlets" at CNAF
  - Yes or No ?
- We will move interactive work at CNAF
- We will not ask for hardware in Bologna/Padova/Pisa/Roma/LNF/Pavia/Udine/Trieste/… besides simple desktops
  - Yes or No ?
- For the time being these are reversible decisions

The new plan

- 2002-2003: work at FNAL
- 2002: tests at CNAF
- 2003: try serious work at CNAF
- 2004: work symmetrically and efficiently in both places
- If no good → go back to FNAL
- Risks
  - Hardware delay
    - computing room
    - procurement
    - installation
  - Operational instabilities and/or inadequacies
    - long way from a pile of PC's to a smootly running computing centers with happy users
  - Need to define clearly what will make us say the final yes

The road to Bologna

- CDF start as "test" in June 2002
  - 5 dual nodes + 1 TB disk
  - CNAF director's gift, no review, no approval
  - Used by logging in explicitely
  - Access restricted to few users
  - PBS ?
- Share of GRID test machines at CNAF for MC possible
- Plan for resources for next year: May 30 → CSN1: June 24
- More test hardare after september (maybe)
- Need to understand:
  - Access (ssh ? Certificates ? Kerberos ?)
  - Batch (PBS ?? Not trivial problem, see CAF)
  - Performance (CPU ←→ disk, see CAF)
  - DataBase access/replica/export (MSQL?)

The first decision

- Usage of the CDF test setup at CNAF
- Proposal:
  - B-tagged multijet stream
    - top→6j H→4j
    - Bologna/Padova
- Because
  - Antonio has done/is doing a lot of work
  - The data size fits the available disk
  - The number of users (4~5) fits the requirements
  - Is all italian, no pressing need to share data with US collegues
  - Is not as pressing/fireline as Stream H

The second decision

- Want to start stealing idle cycles from GRID/LHC test beds
  - Verbal agreement with L.Perini (Atlas)
  - Some machine at CNAF allocated as "grid test bed" and not assigned to specific experiments
  - LHC work concentrated in short periods (MDCs)
- O(10) CPU for O(days) / month no local permant storage
- Proposal
  - Higgs → tau mu
- Because
  - Elena will test GRID tools
  - Can do by saving only final Ntuple (Elena dixit)
  - Size fits
  - Biggest MC production project in Italy

Long Term Future (beyond 2003-4)

- Computing at CNAF will grow with Luminosity until 2010
- CDF needs will always need to go through CSN1
- Share hardware with UK, Spain, Germany
  - Much, much better network then US
  - Can build large disk resident data sets buy avoiding overlap, x2 is already a log in some case (Stream H)
- Need tools
  - GRID
- Need agreement
  - Boring talks
  - Have to find out how "monetize" it, esp. as contribution to CDF
  - N.B. for BaBar money spent in Italian farm counts as contribution to running cost (MOF)

GRID

- GRIDs are there to stay
- It is mostly a matter of names
- CDF already has developed his own distributed job submission tool: CAF_GUI
  - Is a naïve replica of GRID tool
  - It lacks functionalities and especially design
  - It is much prettier and handier
  - It is tailored to CDF needs
  - It will not work for submission to a place other then Fnal
- Wouldn't it be nice to use the same "script" to launch a job at FNAL or CNAF ?
  - Is not simply "a script", packaging one job to run on a remote node is a full environment that has to be learnt

CDF_GRID

- CDF_GRID launched on March 13 (Italy+UK + others)
- UK active on it since > 1 year with O(10) people
- Our strength
  - Flavia
  - INFN-GRID developers have leading role in EuropeanDataGRID (EDG) and are hungry for customers
  - Good connection with management (Ghiselli, Perini …) build over the years
  - Test bed hardware already there (BO+TS)
- Our weakness
  - Very few people
  - Sluggish enthusiasm: besides Antonio
  - No full time professional
  - Test bed installation already a problem

Work on GRID so far

- CDF is officially part of DataTAG initiative
  - Contact persons: Flavia Donna , Antonio Sidoti
- CDF entity described in DataTAG databases (VO)
- 3 CDF users have got Globus Certificates (GRID passwords), Antonio, Elena Vataga, S.B.
- DataTAG test bed machines being installed at Trieste (3) and Bologna (4)
- DataTAG UI installed at FNAL on Trieste's ncdf29
- Flavia+Stefano with big help from Alex Cerri managed to
  - Run one CDF MC example on the GRID:
    - ☞ submit from FNAL, Torino
    - ☞ executes at CNAF
    - ☞ retrieve otput to Torino, CNAF
    - ☞ same result as running on fcdflnx1

CDF_GRID plan

- Data Handling (copies, replicas, staging…)
  - UK investigating SAM for remote data replica
  - CDF adopting SAM for local DH
  - Flavia + DataTAG + US (iVGDL) to integrate SAM+EDG
  - SAM station will be setup in TS (S.B.)
- Remote job submission
  - Italy will investigate EDG tools
  - Antonio/Elena on it
  - Is wat will really glue toghether European CDFers
- Authentication/Authorisation
  - Globus vs. Kerberos. Igor Sfiligoi interested
  - Distributed/Heirarchical VO: Lamberto Luminari intersted
- Details, status, docs, log of tests, hints and tricks:

http://www.ts.infn.it/~belforte/offline/grid/index_grid.html

GRID works ahead

- Push for the test bed (help?)
- Learn to use it with AFS (sysman problem)
- Learn to use it without AFS = w/o CDF offline
  - Opens the way to "running everywhere"
- Learn to use SAM "standalone"
  - May be main mode of operating CNAF for a while
  - Hardware assigned to it in Trieste
- Experiment with SAM integrated in EDG
  - DataTAG responsibility
  - CDF must provide test and feedback

Conclusion

- Remember the message:
  - This is the most important thing we are doing now
  - Need to turn analysis into a running train
  - Much much more work was done then my e-mails say, visit my web page and explore the links
- HW at FNAL under control
  - Have agreed on some resource assignement and plans
  - New CAF is also our project, make sure delivers promise
- Moving to CNAF : Requires one person to drive the work to
  - explore, comunicate, test, define, report, request
- Integrating in GRID : Require one person to drive the work
  - explore, communicate, test, define, report, request
- Important, visible, usefull, responsibilities
- Group management kindly encouraged to focus on this