



## Calcolo per CDF

- 1) Rivediamo le richieste 2005
  - ❖ Le 4 slides che ho circolato a Luglio con i numeri nei moduli
- 2) Stato attuale CNAF
- 3) Meditazioni sul futuro farm CNAF alla luce dell'esperienza summer 2004 = 2nda iterazione sulle previsioni 2005 che erano state date a maggio 2004 alla CSN1
- 4) Stato e prospettive del SW e di unificazione con GRID
- 5) CDF Compting Usage Task Force



## Richieste 2005

- Dati I tempi delle gare etc. sono in realta' richieste per computers da usare nel 2006
- + 900 GHz
- + 50 TB
- Motivazioni etc. sono state circolate a Luglio
- Richieste incluse nella programmazione Tier1
- Accettabile considerare 900GHz = 100 duals, anche se la CPU non sara' 4.5 GHz



## Stato CNAF

- 250 GHz + 8.5TB comprati da CDF (fondi 2003)
  - 7.5GTB dati 1TB user space
- Tier1 doveva fornire 700GHz + 30TB nel 2004, avuti 170GHz
- Uso in ascesa via via che disco/cpu aumenta, ma non abbiamo saturato
  - Le farm di FNAL sono piene, ma i tempi di attesa ancora ragionevoli (~giorni), gli users sono poco motivati a fare il piccolo sforzo necessario per lavorare in Italia
  - Inoltre non abbiamo potuto copiare i dati in Italia
- Transizione a Condor inizia ora
  - CondorCAF stabile da pochi mesi + attesa manpower
  - Subir Sarkar (AR CNAF 100% CDF) + Igor Sfiligoi (tecn. LNF, 50% CDF)
- I/f con INFN Grid = next frontier (vedi dopo)



## Quale farm al CNAF nel 2006

- Le richieste dei moduli sono ancora valide alla luce della esperienza delle conferenze estive 2004 e dei primi risultati della CUTF
- Ci sono incertezze nel futuro, ci sono momenti di picco e di minor uso
- La strategia di confluire in un pool comune Tier1 e' la difesa contro gli uni e gli altri, nessuna CPU vada sprecata
- L'hardware che abbiamo al FNAL e' considerevole ed e' stato il nostro principale sostegno fino ad oggi:
  - 12TB disco
  - 500 GHz
  - Sono macchine acquistate nel 2002/3. Nel 2006 saranno fuori manutenzione ed obsolescenti (un RAID array e' gia' stato sostituito e.g.)
- La richiesta attuale non include il rimpiazzo di questo hw



## Sviluppo SW

- CAF on LCG = next frontier
  - Joint project INFN - UCSD - FNAL
  - SB, Igor Sfiligoi, Subir Sarkar, AR ad-hoc from INFNGrid (bando pubblicato)
  - Salvare semplicità ed user-friendliness della CAF dando agli users accesso alla GRID, simmetria di accesso a EGEE e OSG
  - Only address job management, keep using SAM for data handling
- Sviluppo sw utilities per rendere manggevole ed efficiente lavorare su campioni >>1TB, usando SAM
  - Skim inclusive hadronic B stream (bookeeping, concatenation, load into data catalog) (D.Lucchesi PD, M.Casarsa TS)
  - MC production offsite (temporary storage for validation, concatenation, load into data catalog)
  - Improved bookeeping and retry of failed job sections
  - SB, Armando Fella (bors. Pisa 50%) + Francesco Delli Paoli (bors. Padova 100%) + I.S. + S.S.



## Piu' Efficienza ?

- Computing Usage Task Force creata a Maggio da CDF, SB chair
- Molto lavoro nel monitorare e capire (farms a FNAL ~sature)
- Bottom line: scoperta dell'acqua calda
  - Gli users provano, cercano, imparano, sbagliano
  - Tutti i lavori fatti portano conoscenza e sono utili
  - Quanto accelerano la produzione di articoli, non si sa
  - Molto, molto piu' MC di quanto ufficialmente dichiarato
  - Gia' ora molto molto lavoro fatto su ntuple su desktops
- Lo scenario: tanti users che fanno tante cose diverse, molto difficile da ottimizzare, nessun "caso eclatante"
- La CUTF ha individuato alcuni esempi di razionalizzazione ed ottimizzazione possibili (centralizzare alcune attivita'):
  - Guadagno in termini di CPU non quantificabile
  - Implementazione richiede grande lavoro di validazione di fisica, non e' un problema di computer professionals
- Molto aiuto dal Bs skimming e.g. (INFN work)



Dettagli...

*Dettagli...*



## Slide luglio 1: Richieste 2005

- Parto da quanto presentato alla CSN1 a Maggio 2004
- Ci sono due parti della richiesta
  - Coprire l'analisi degli italiani con risorse a noi dedicate
  - Contribuire ai crescenti bisogni di tutto CDF con una parte in cui noi non abbiamo accesso prioritario rispetto agli altri di CDF, mirante a coprire tra 2 anni il 15% del bisogno di calcolo per analisi e MC di tutto CDF.
  - Come si gestiscono le 2 "quote" e' un dettaglio tecnico, si fara'.
- Rispetto a quello indicato a Maggio prevedo che:
  - I bisogni per la analisi degli italiani rimangano quelli indicati, il ns. Lavoro va avanti bene, la luminosita' cresce, le code batch dedicate che abbiamo a fermilab hanno smesso di crescere e diventano a mano a mano irrilevanti. QUindi aumentare la quota "per italiani" del 50%
  - I bisogni di CDF overall saranno rivisti un po' al ribasso in autunno perche' l'upgrade del DAQ tarda e la TaskForce sulla ottimizzazione qualche cosa riuscirà ad ottimizzare
- Seguono 2 slides estratte dalla mia presentazione di maggio alla CSN1
- Segue una slides con le richieste che metterò nei moduli





## Slide luglio 2: 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



# Slide luglio 3: 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
	GHz	TB	30% of our CPU	GHz to add	GRID GHz	tot GHz for CDF	
Year	GHz	TB					
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



## Slide luglio 4: Richieste 2005

- Le richieste addizionali per il 2005 sono le seguenti.
  - + 900 GHz = 100 duals = 200KEuro
  - + 50TB = 150KEuro
- Con umilta' ed ottimismo ho ridotto i 2100GHz di maggio a 1850. Sul disco non vedo molti margini perche' l'aumento di efficienza dell'analisi non ridurra' le necessita' di disco e se vogliamo spostare in Italia il lavoro che ora viene fatto a Fnal dobbiamo replicare molti data sets. Per la conversione in KEuro:
  - 1 nodo 2x4.5GHz = 2KEuro (mia stima)
  - 1TB = 3KEuro (dall'ultima gara CNAF, comunic. da Federico)
- All'atto pratico nello spirito del pool comune di CPU, questo corrisponderebbe a che CDF abbia in totale nel 2005
  - Quota minima = 1400 GHz
  - Quota massima = 1900 GHz
  - Si intende che lo sviluppo del sw della farm di CDF gia' da quest'estate e' teso a fare si' che altri possano usare i nostri cicli cpu idle, if any, e noi possiamo usare nei momenti di picco eventuali cpu libere oltre tale quota



## Stato Estate 2004

- Stato CNAF:
  - A settembre 2003:
    - ☞ + 700GHz
    - ☞ + 30TB
  - A settembre 2004:
    - ☞ abbiamo avuto da ~1mese  $28 \times 2 \times 3.06 = 170$  GHz
    - ☞ il resto ancora sotto esame per problemi calore ?
    - ☞ niente disco (installazione IBM da completare)
- Non possiamo aggiornare richieste sulla base della esperienza dell'utilizzo di quanto "dato" un anno fa
  - Ma la carenza di CPU non ci ha fatto soffrire
- Ma possiamo prevedere che la discussione attuale riguarda non quale hardware avremo nel 2005, ma quale avremo nel 2006
- Ad oggi le analisi si fanno su ~200pb-1
- Da ~Novembre su ~350
- Estate 05: ~500pb-1
- Estate 06: ~1fb-1 = oggi x 5



## Revisione annuale previsioni

- Come ogni anno CDF ha una nuova edizione del computing plan con aggiornamento previsioni bisogni, richieste, strategia, tools etc. ed una nuova Director's Review da un comitato esterno il 13/15 Settembre (review di tutto il computing di Fnal)
  - <http://cdinternal.fnal.gov/RUNIIRRev2004/runIIMP.asp>
- Documento CDF non ancora non disponibile (nemmeno a me) fino al 13 Settembre. Non ve lo presento. Le cose importanti sono:
  - Upgrade DAQ rallentato
    - ☞ 2004: 20 MB/sec (come 2003/2/1)
    - ☞ 2005: 35 MB/sec (dimostrato in test run agosto 2004)
  - Rimane costante la linea di fare tutto MC centralizzato e 50% della analisi offsite.
  - Il computing model e' ancora quello 2004:
    - ☞ high-pt scala con luminosita', 200 users
    - ☞ low-pt scala con tempo reale, 20 users
  - La realta' e' diversa, ma le risorse previste in base al model sono 100% utilizzate



## Calcolo CDF 2005

- Come analizzare nel 2006 5 volte I dati attuali ?
  - Attuale: 400GHz, 6TB
  - I 6TB sono del tutto insufficienti, avevamo chiesto 30, facciamo conto di avere 20 (il minimo per lavorare: 8TB Hadronic B, 4TB high Pt, 4TB skimed Had.B, 1TB J/psi, 3TB MC)
  - 400GHz OK (vedi plots uso CPU)
- Aumento hw, aumento efficienza.
- La richiesta attuale prevede aumento hw fino a:
  - 90 TB (scaling:  $30 \times 5 = 150$ , min:  $20 \times 5 = 100$ )
  - 1500 GHz (scaling:  $400 \times 5 = 2000$ )
- D'altra parte:
  - La parte low-pt non scala con L (x5) ma con T (x3)
  - La parte low-pt consuma moltissime risorse ed e' il punto di forza di >50% del gruppo italiano
  - CUTF → nessun magic bullet per ridurre x2, x3 ...



## Stato SW

- Cdf Analysis Farm at CNAF (CNAFCAF)
  - Da Luglio abbiamo AR dedicato al CNAF (Subir Sarkar)
  - SAM ancora responsabilita' di SB
  - Passaggio a Condor scheduled ~fine settembre
    - ☞ Igor Sfiligoi (LNF) main CondorCAF developer + SS
  - Confluenza nella farm PBS/LCG del Tier1 attraverso Condor Glide-In = lavoro primario per Subir
    - ☞ non fa felice dirigenza Tier1, ma unica soluzione possibile ora, discussione con nuova dirigenza Tier1 ancora da fare, trattative con INFN-Grid (A.Ghiselli, F. Prelz, F.Donno) in corso
- Merging con GRID beyond Glide-In
  - A.R. per integrazione CDF-GRID, bando pubblicato da <1 settimana, su progetto speciale INFN-Grid (SB in EB)
- SW per users: rendere facile lavorare su TB di dati, in modo da migliorare efficienza, ridurre errori, risparmiare risorse
  - Armando Fella (borsista Pisa) + F.Delli Paoli (borsista Padova)



## SoftWare per users

- Armando Fella (borsista a Pisa da 1 anno)
  - 50% manutenzione cluster locale
  - 50% CNAFCAF (ora passato a Subir Sarkar) e sviluppo SamUpload = Tool per permettere ad uno user di salvare e catalogare in SAM un set di files da una locazione esterna
- Benefici:
  - Dati users salvabili su nastro a FNAL
  - Dati users condivisibili
  - MC prodotto/storato in UK usabile in Italia senza rigenerarlo
- F.Delli Paoli (borsista a PD da oggi)
  - 100% supporto applicazioni CDF = automatizzazione di jobs "labor/bookeeping intensive" (skimming HadronicB stream for Bs candidates e.g.)
- Benefici:
  - Users del B-adronico possono lavorare su campioni ridotti "certificati"
  - Concatenazione per salvataggio su nastro per MC





## CUTF 0 : Membership

- Stefano Belforte (chair)
- Patrizia Azzi (top)
- Song Min Wang (exotics)
- Ray Culbertson (exotics)
- Ashutosh Kotwal (EWK and offline)
- Matt Herndon (bottom and tracking)
- Konstantin Anikeev (bottom)
- Elliot Lipeles (CAF)
- Igor Sfiligoi (CAF)



## CUTF 1

- May 2004: CDF Spokespeople form Task Force on Computing Usage to report on degree of optimization and suggest fixes if any, chaired by SB
- June = review of status talking with phys.groups
- July = setup monitor of CAF to verify/quantify
- August = collect data and ideas
- September = write report
- Conclusion so far:
  - Thanks to delayed DAQ upgrade, CPU was not missed
  - Much more user MC done/needed than planned
  - B analysis are enormously CPU consuming due to combinatorial load in secondary/tertiary vertex finding
  - Overall scenario : many users doing many different things. No way to simulate a few typical cases and attack them. Very difficult to optimize.
  - Users are using CPU to try, test, explore, learn, do physics
    - ☞ monitoring with no-privacy-regard demonstrate no abuse



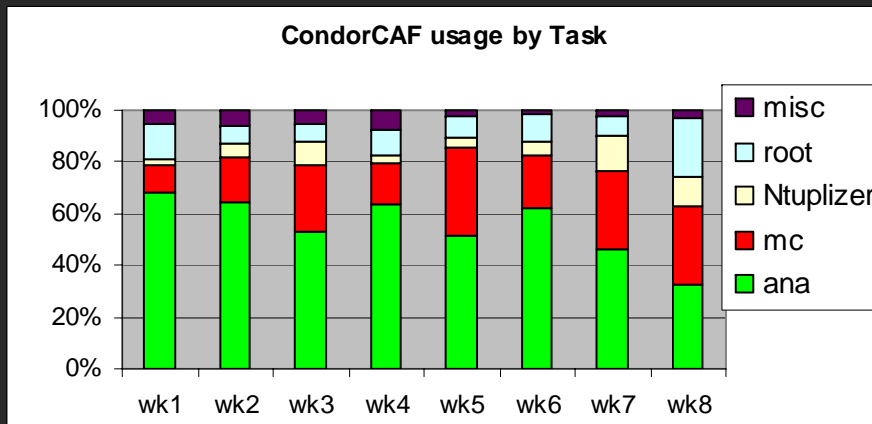
## CUTF 2

- Path to efficiency improvement outlined
  - Move more work from scattered users effort to planned effort
  - Since planning requires effort over and beyond writing a thesis, not clear where the needed manpower may come from
  - Move "tasks" from user's executable to production (cosmic ray, dedx, beam constraint): CPU saving unclear, need validation, see above comment



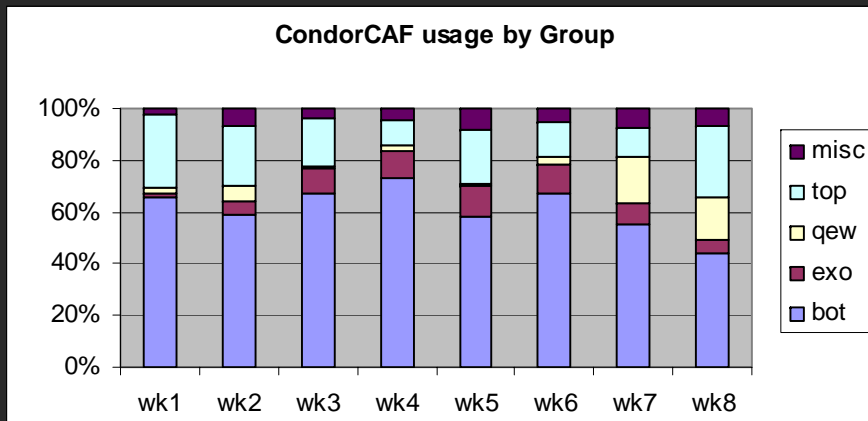
# Last 2 months on CondorCAF

## Lots of MC, Lots of B



- Up to 50% is MC (root=MC)
  - this is a lot of MC !! All organized MC production uses less

Now → → → 2 months ago



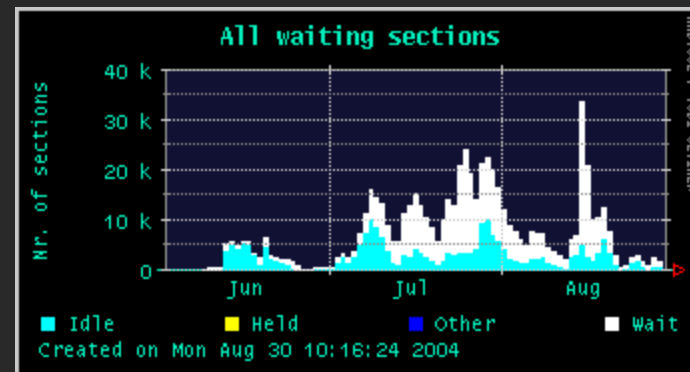
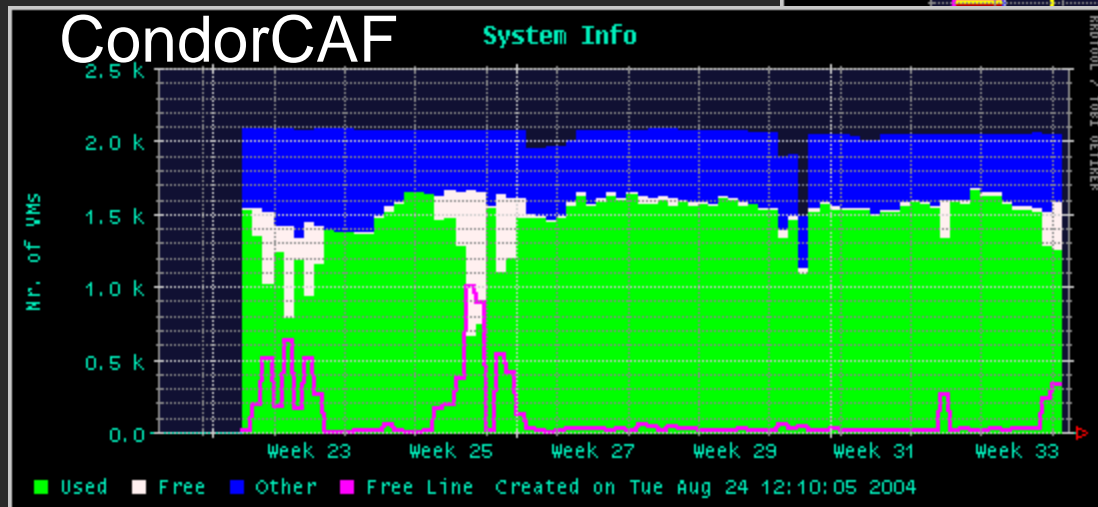
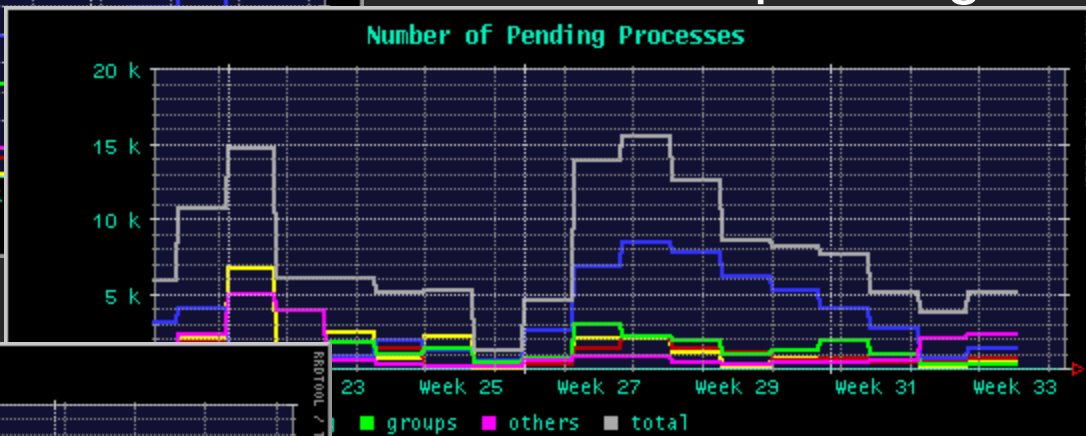
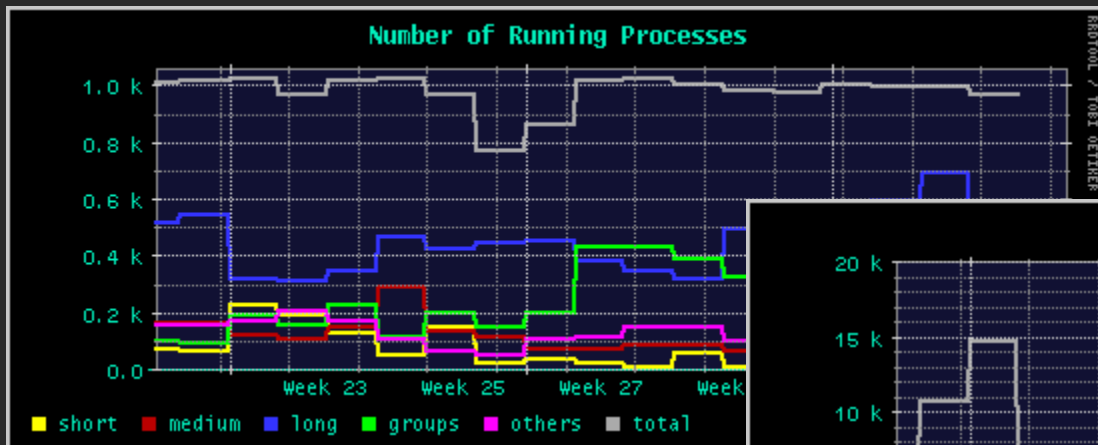
- B rules
- what is all that activity by top/Exo ?
- it is not making common ntuples



# Are our users CPU-limited ? (Last 3 months)

- Looking at FNAL, seems we lack CPU

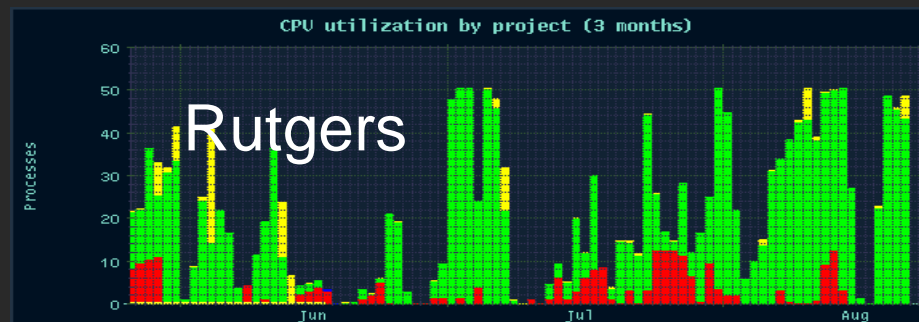
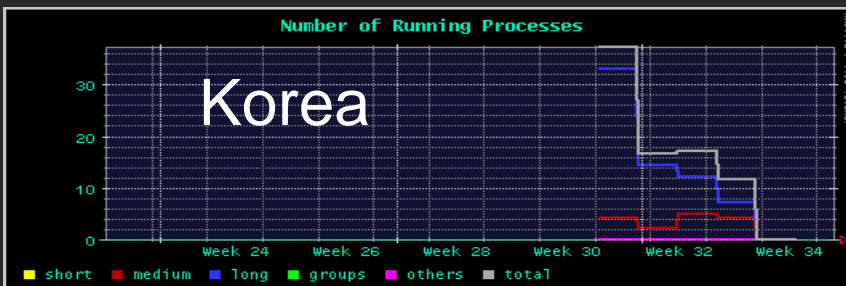
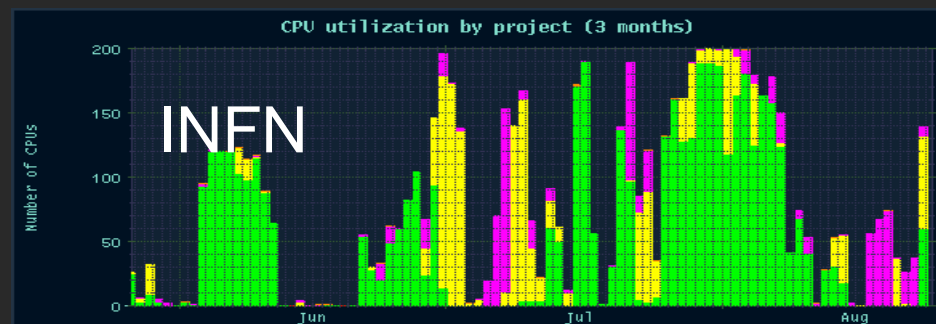
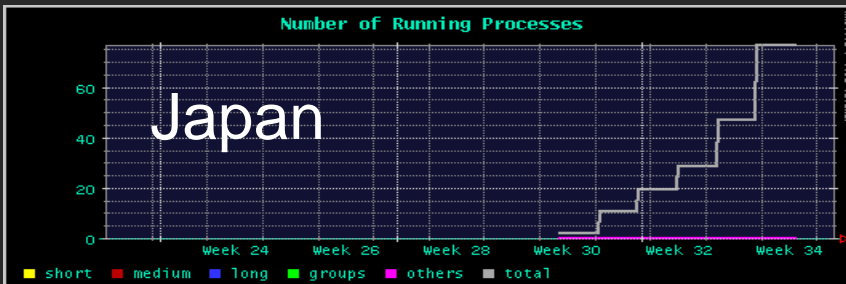
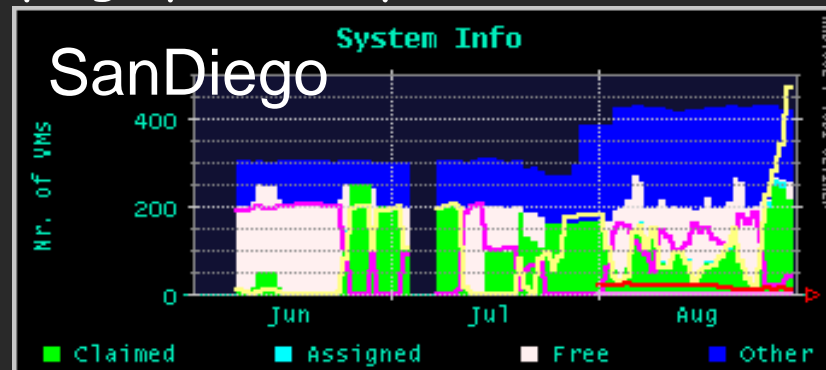
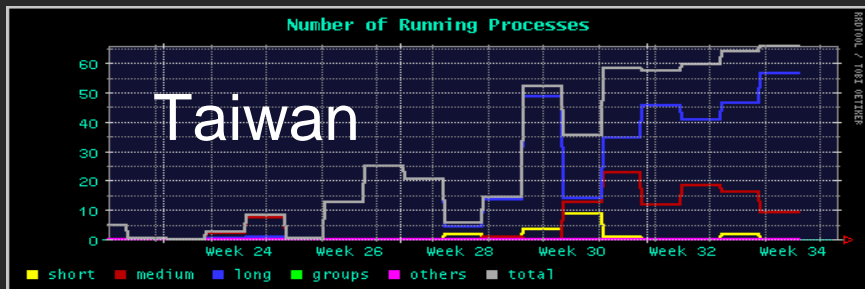
CAF:  
Always running  
Thousands pending





# Are our users CPU-limited ? (Last 3 months)

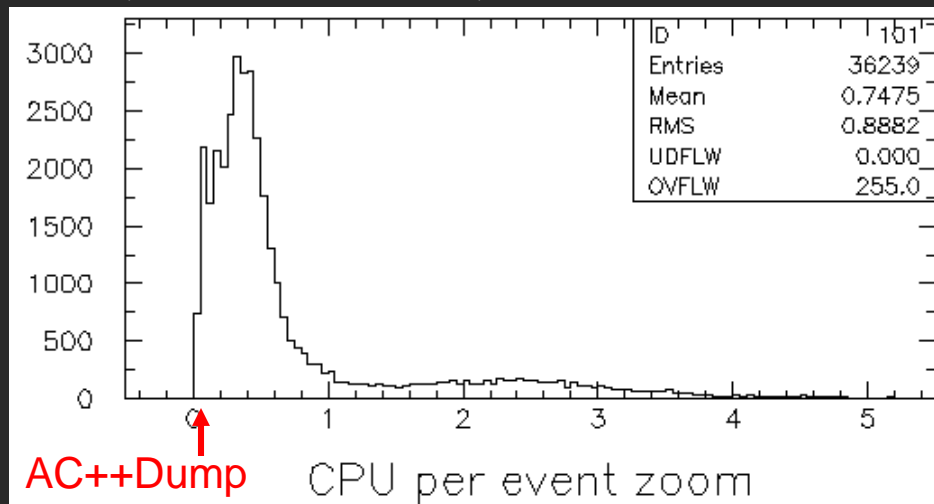
- Looking offsite, usage has been ramping up steadily





## Last 8 month of CAF (non Condor)

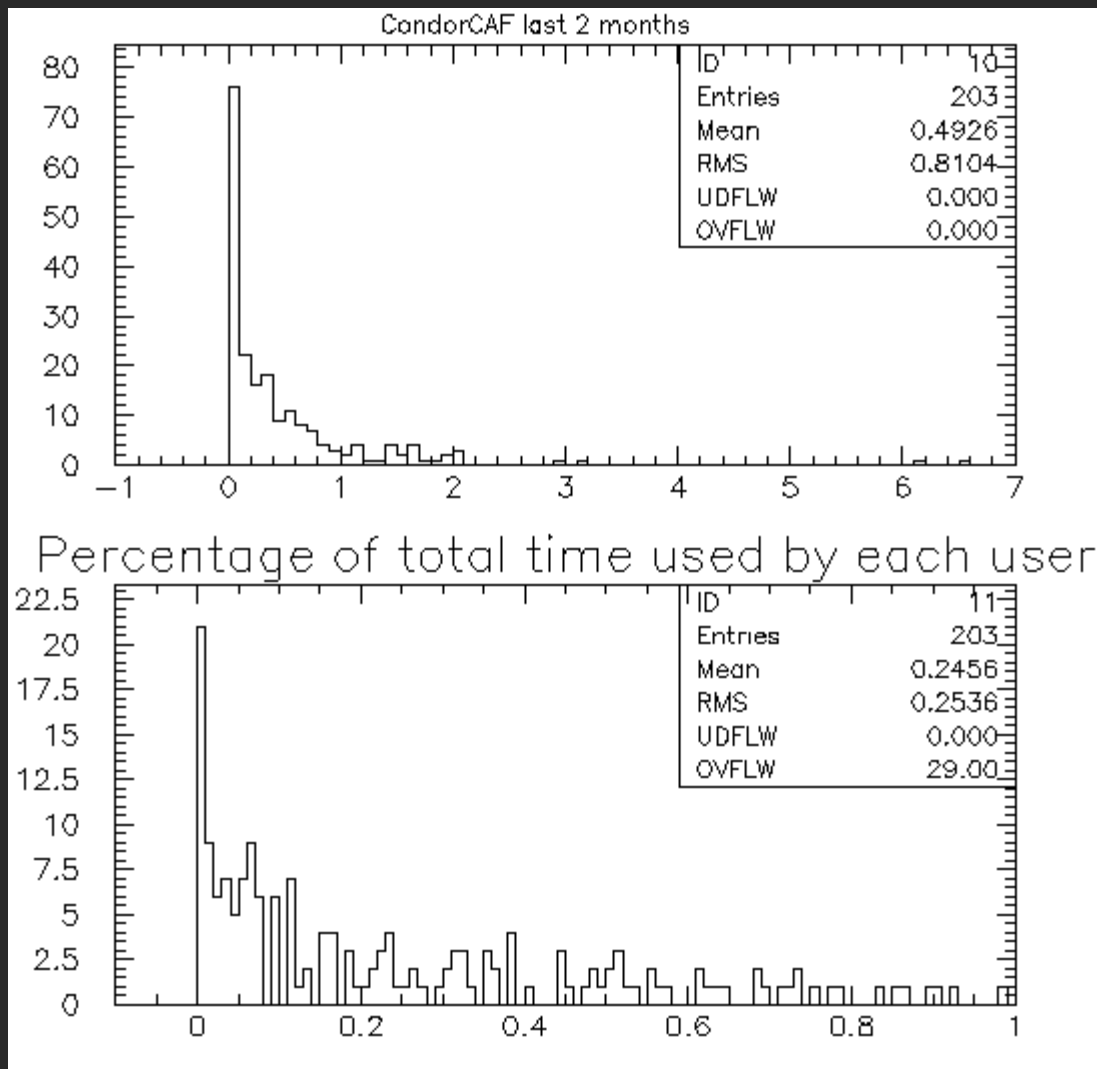
- 240K cpu-days available, 80K jobs "logged"
- Have input-dataset/cpu-time info for 36K jobs: 56K cpu-days, ~25%



AC++Dump sits  
at ~0.06  
sec/event

- 7K jobs use >1 sec/event
  - 400 of them account for 40% of total cpu: 23K cpu-days. 18K hadronic B + J/Psi (CVTMFT? Have ideas on how to help here but <10% total)
- 30K jobs use <1sec/event, each job minuscule fraction of total
  - Account for ~60% of "analysis CPU"
  - Not I/O limited, user code is the "culprit"
  - SB could not find a way to summarize/break-down them
  - No idea how to tackle them "C++-wise"

# CPU By user

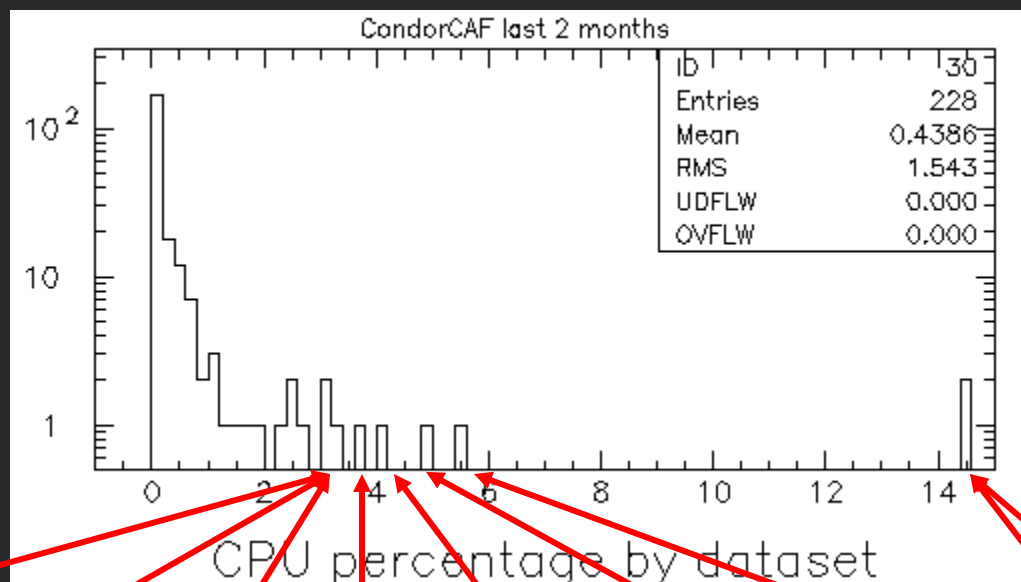


- 203 users on CondorCAF in last 2 months
- Not as simple as a few users using all CPU (we could learn a lot about patterns and needs talking to them)





## CPU by dataset



xpmm0c

xbhd0e

jpmm0dgjt10d

xbmu0d

xbel0d

xpmm0d

xbhd0d

xbhd0c

- 228 cdf dataset names "hit" in last 2 months on CondorCAF
- Usual suspects show up, **red names** makes up ~60% of total
- 30% of usage for the big inclusive B-Charm xbhd0d + xbhd0c
- 40% is scattered in >200 dataset names