

Belle II Computing

F. Bianchi

INFN & University of Torino

Software/Computing Workshop

11-15 Maggio 2015

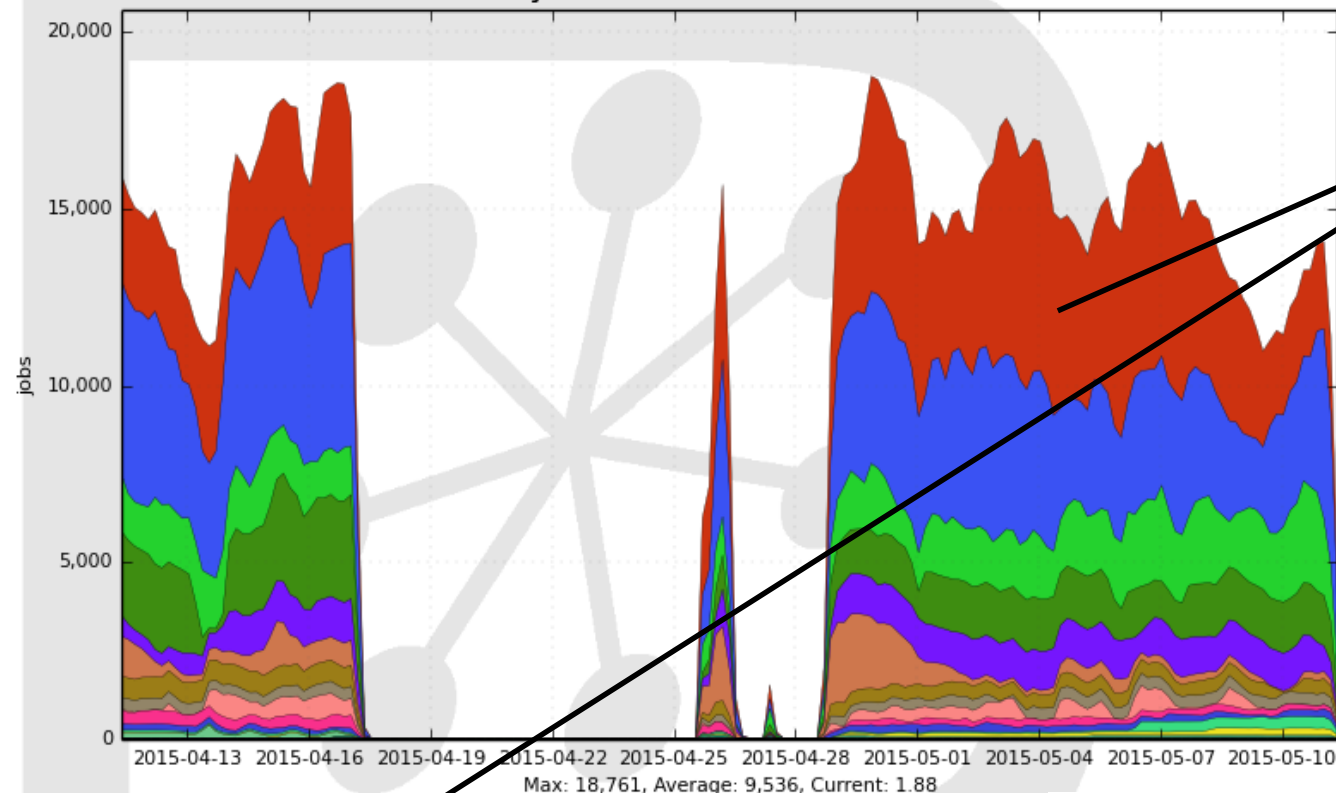
- Giornate intense piene di presentazioni e discussioni
- See:
<http://kds.kek.jp/conferenceDisplay.py?confId=17040>
- Progressi importanti, principale fattore limite e' la mancanza di manpower
 - Critica in Europa ed Italia

Attivita' in Italia

- Importanti contributi italiani allo sviluppo del software
 - Tracking, ECL, tools di analisi (Pisa, Perugia, Napoli, Frascati)
- Contributi al computing essenzialmente sotto forma di risorse di calcolo messe a disposizione e di studi sul networking (Napoli)
 - Dirac server a Napoli
 - CPU e storage pledged al CNAF ed a Napoli
 - VO Belle abilitata anche a Frascati, Pisa, Legnaro e Torino

Produzione MC in Aprile - Maggio

Running jobs by Country
30 Days from 2015-04-11 to 2015-05-11



Contributo italiano

IT	28.2%	SI	3.9%	UA	1.0%	TW	0.1%
DE	26.2%	PL	2.9%	KR	0.8%	CN	0.0%
JP	11.3%	CZ	1.9%	TR	0.5%	SG	0.0%
CA	10.8%	AT	1.8%	AU	0.5%	ANY	0.0%
US	6.4%	RU	1.4%	MX	0.2%	IN	0.0%

L'Italia è riuscita, nell'ultima campagna, a fare molto più di quanto concordato (il 28.2% invece del 10%)

KEKCC

- Macchine e personale di una ditta esterna
 - Contratto di 3 anni, dopodiche' si azzera tutto e si ricomincia
 - Importanti rigidita' nel sistema, costi significativamente piu' alti dei nostri.
- Doveva esserci un nuovo contratto nel 2015 con incremento di risorse (e di costi), rimandato per ragioni di budget di KEK, ci sara' nel 2016
 - Possibile downtime di KEKCC fino a 2 mesi
 - Necessita' di duplicare all'esterno di KEKCC i servizi essenziali al funzionamento del computing
 - Scelta strategica da fare comunque

SuperKEKB Schedule

- 2016:
 - Phase 1 commissioning of the machine (mid Jan to mid May)
 - Cosmic data taking (Sept to Nov)
- 2017:
 - Phase 2 commissioning of the machine with Beast detector (mid May to mid January 2018)
- 2018:
 - Cosmic data taking (mid Jul to mid Oct)
 - Physics data taking (mid Oct to mid Dec)

Stima delle Risorse: Disclaimer

- Sequenza:
 - Stima preparata e approvata dal CSG di Belle II
 - Review interna
 - Approvata dall'Exec Board
 - Approvata dal BPAC
 - Review esterna, G. Carlino per l'INFN
- A febbraio 2015 il BPAC ha fatto alcuni rilievi a cui non e' ancora stata data una risposta ufficiale
 - I numeri che seguono devono essere ancora approvati da Exec Board e BPAC
 - Tengono pero' conto dei rilievi del BPAC

Assumptions

- Reconstruction 45 HEPSpec s/event
- Simulation (generation + reconstruction) 89 HEPSpec s/event
- Phase 1 Background Simulation 4.5 HEPSpec s/event
- Phase 1 Background Analysis 2.25 HEPSpec s/event
- Phase 2 Background Simulation 45 HEPSpec s/event
- Phase 2 Background Analysis 22.5 HEPSpec s/event
- Phase 2 raw data simulation: 55 HEPSpec s/event

- Raw data size: 300 kB/event
- mDST size: 40 kB/event
- Skim size: 40 kB/event
- User data size: 4 kB/event

- 10 months/year of useful production
 - Site downtime, software bugs that require restart of production, operational mistakes, etc

Activities in 2016

- Requested by Physics Groups:
 - 5 ab^{-1} generic MC production (+ skimming + analysis)
 - ~ 5 times integrated luminosity of first year of data taking
 - BaBar final MC sample is 10 times integrated luminosity
 - 20 ab^{-1} -equivalent B Rare Decay MC
 - Decay with missing energy sensitive to New Physics
 - 20 ab^{-1} -equivalent Btag-skim (skim at generator level + processing)
 - For studies of modes with missing energy on the recoil of a fully reconstruct B
 - 0.3 ab^{-1} scan & non 4S
 - Quarkonia studies and search for dark sector particles
 - 0.1 ab^{-1} $ee(\gamma)$, $eeee$, $ee\mu\mu$
 - Important for calibrations and physics background studies

Activities in 2016 (2)

- Requested by Background Group (~60 s of bkg data):
 - 2.4×10^{10} Phase 1 background events (MC prod + analysis)
 - 2.4×10^{10} Phase 2 (Beast) background events (MC prod + analysis)
- Cosmic Data (processing + analysis)
- 0.5 ab^{-1} Phase 2 (Beast) raw data simulation (to be processed in early 2017 to test processing of Phase 2 detector data)

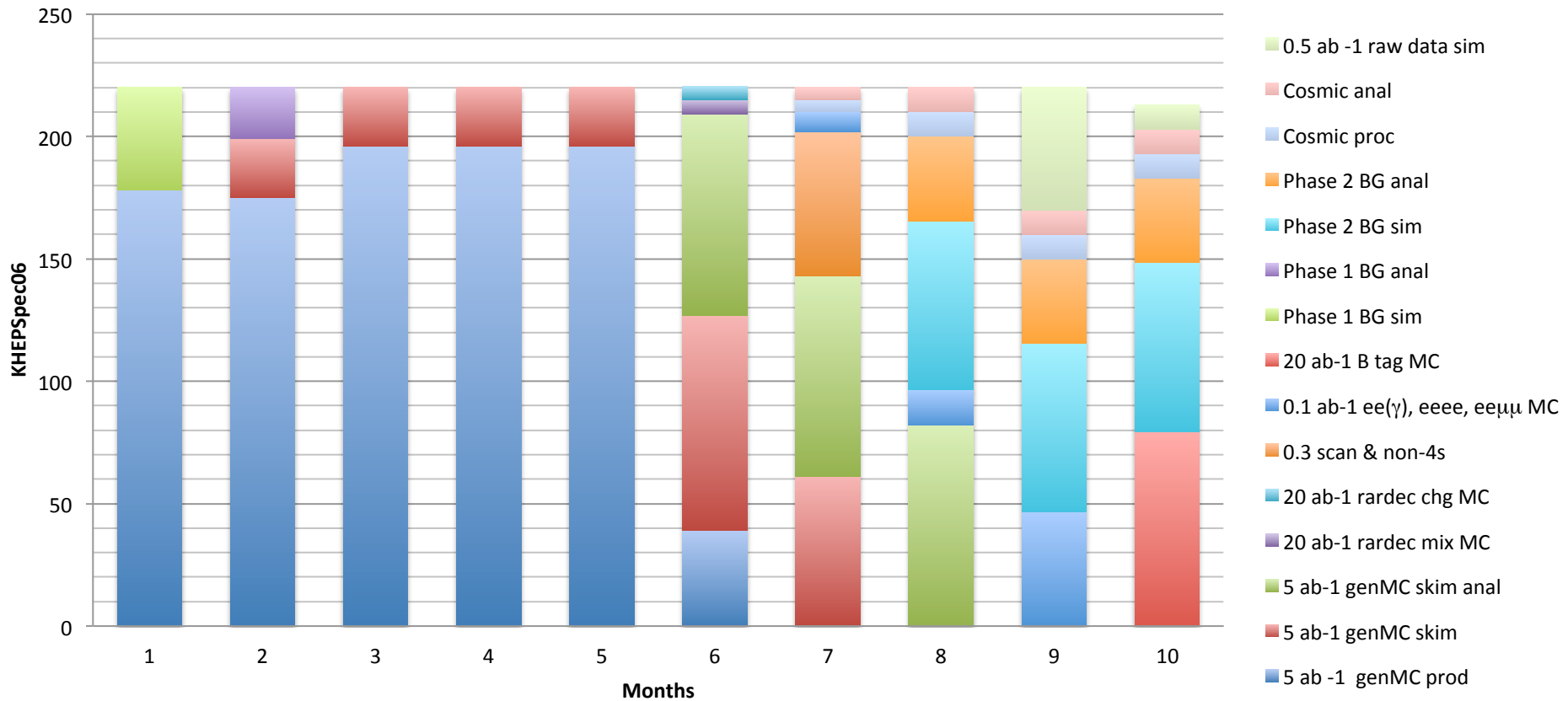
Resource Estimate 2016 (1)

Data Sample	Events ($\times 10^9$)	CPU (KHEPSpec06)	Tape (TB)	Disk (TB)
5 ab^{-1} generic MC Production	28.50	98.00		1087.19
20 ab^{-1} rare B decay	0.33	1.15		12.58
0.3 ab^{-1} scan & non 4S	1.71	5.90		65.23
0.1 ab^{-1} $ee(\gamma)$, $eeee$, $ee\mu\mu$	2.00	6.90		76.29
20 ab^{-1} B tag skim (generator level) MC	2.31	7.95		88.12
5 ab^{-1} generic skimming	28.50	24.50		1087.19
5 ab^{-1} generic analysis	28.50	24.60		108.72

Resource Estimate 2016 (2)

Data Sample	Events ($\times 10^9$)	CPU (KHEPSpec06)	Tape (TB)	Disk (TB)
Phase 1 BG simulation	24.00	4.20		915.33
Phase 1 BG analysis	24.00	2.10		91.5
Phase 2 (Beast) BG simulation	12.00	20.70		457.76
Phase 2 (Beast) BG analysis	12.00	10.35		45.78
Cosmic Processing		3.50		
Cosmic analysis		3.50		
0.5 ab ⁻¹ (Beast) Raw Data Simulation	2.85	6.00	815.39	

Time Profile of CPU Usage 2016



Resource Estimate 2016 (3)

- Grand Total:
 - CPU: 219.35 kHEPSpec to be used for 10 months
 - Tape storage: 815.39 TB
 - Disk Storage: 4035.95 TB

Activities in 2017

- Recreate the MC sample requested by Physics Groups for 2016:
 - Improved software.
 - Keep 2016 sample for comparison.
- 0.5 ab⁻¹ Phase 2 (Beast) raw data processing (simulated at the end of 2016)
- Requested by Background Group (~60 s of bkg data):
 - 2.4x10¹⁰ Phase 2 (Beast) background events (MC prod + analysis) with detector constant from Phase 2 data taking
 - Keep 2016 BG simulation data for comparison
- Phase 2 (Beast) data:
 - Process + reprocess + skim + analyze detector data
 - Produce + skim + analyze MC data (4 x Det Data)
- 1 ab⁻¹ raw data processing rehearsal for 2018 Physics Run data
 - get raw data from 5 ab⁻¹ generic MC production
 - Process and reprocess the data sample

Resource Estimate 2017 (1)

Data Sample	Events ($\times 10^9$)	CPU (KHEPSpec06)	Tape (TB)	Disk (TB)
5 ab^{-1} generic MC Production	28.50	98.00		1087.19
20 ab^{-1} rare B decay	0.33	1.15		12.58
0.3 ab^{-1} scan & non 4S	1.71	5.90		65.23
0.1 ab^{-1} $ee(\gamma)$, $eeee$, $ee\mu\mu$	2.00	6.90		76.29
20 ab^{-1} B tag skim (generator level) MC	2.31	7.95		88.12
5 ab^{-1} generic skimming	28.50	24.50		1087.19
5 ab^{-1} generic analysis	28.50	24.60		108.72

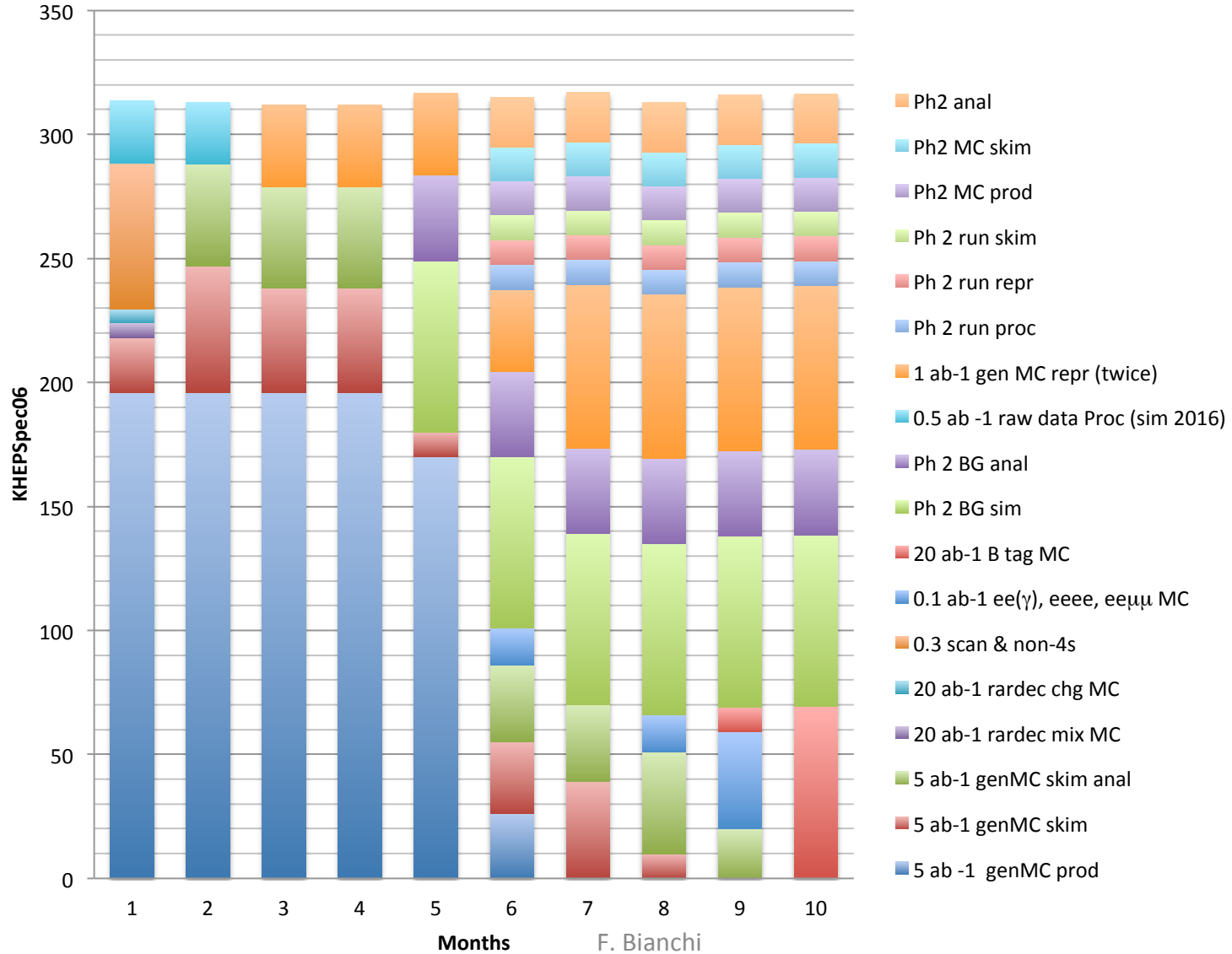
Resource Estimate 2017 (2)

Data Sample	Events ($\times 10^9$)	CPU (KHEPSpec06)	Tape (TB)	Disk (TB)
Phase 2 (Beast) BG simulation	24.00	41.40		915.33
Phase 2 BG (Beast) analysis	24.00	20.70		91.50
0.5 ab^{-1} Phase 2 MC raw data processing	2.85	5.00		108.72
1 ab^{-1} generic MC raw data storage (Phys Run rehearsal)	6.7		1916.90	
1 ab^{-1} generic MC raw data process + reproces (Phys Run rehearsal)	6.7	39.60		511.17

Resource Estimate 2017 (3)

Data Sample	Events ($\times 10^9$)	CPU (KHEPSpec06)	Tape (TB)	Disk (TB)
Phase 2 (Beast) run raw data storage	1.20		343.32	
Phase 2 (Beast) run Raw data processing	1.20	5.00		45.78
Phase 2 (Beast) run Raw data reprocessing	1.20	5.00		45.78
Phase 2 (Beast) run Detect data skim	1.20	5.00		45.78
Phase 2 (Beast) run MC production	2.00	6.85		76.28
Phase 2 (Beast) run MC skim	2.00	6.85		76.28
Phase 2 (Beast) run (Det + MC) analysis	3.20	10.00		12.21

Time Profile of CPU Usage 2017



Resource Estimate 2017 (4)

- Grand Total:
 - CPU: 314.4 kHEPSpec to be used for 10 months
 - Tape storage: 2260.22 TB (2017) + 815.39 TB (to be kept from 2016) = 3075.61
 - Disk Storage: 6488.41 TB (2017) + 4035.95 TB (to be kept from 2016) = 10524.36 TB

Activities in 2018 (1)

- Recreate the MC sample requested by Physics Groups for 2016 and 2017:
 - Improved software.
 - Keep 2017 sample for comparison. Delete 2016 sample
- Requested by Background Group (~60 s of bkg data):
 - 2.4×10^{10} Phase 2 (Beast) background events (MC prod + analysis) with detector constant from Phase 2 data taking
 - Keep 2017 BG simulation data for comparison. Delete 2016 sample
- Phase 2 (Beast) data:
 - Reprocess + skim + analyze detector data with improved software and constants
 - Re-produce + skim + analyze MC data with improved software and constants
 - Keep 2017 data for comparison
- 2 ab^{-1} raw data processing rehearsal for 2018 Physics Run data
 - get raw data from 5 ab^{-1} generic MC production
 - Process and reprocess the data sample

Activities in 2018 (2)

- Cosmic Data (processing + analysis)
- Physic run data:
 - Process + reprocess + skim + analyze detector data
 - Produce + skim + analyze MC data (4 x detector data)

Resource Estimate 2018 (1)

Data Sample	Events ($\times 10^9$)	CPU (KHEPSpec06)	Tape (TB)	Disk (TB)
5 ab^{-1} generic MC Production	28.50	98.00		1087.19
20 ab^{-1} rare B decay	0.33	1.15		12.58
0.3 ab^{-1} scan & non 4S	1.71	5.90		65.23
0.1 ab^{-1} $ee(\gamma)$, $eeee$, $ee\mu\mu$	2.00	6.90		76.29
20 ab^{-1} B tag skim (generator level) MC	2.31	7.95		88.12
5 ab^{-1} generic skimming	28.50	24.50		1087.19
5 ab^{-1} generic analysis	28.50	24.60		108.72

Resource Estimate 2018 (2)

Data Sample	Events ($\times 10^9$)	CPU (KHEPSpec06)	Tape (TB)	Disk (TB)
Phase 2 (Beast) BG MC analysis	24	21.20		91.5
1.5 ab^{-1} generic MC raw data storage (Phys Run rehearsal)	10.05		2875.30	
1.5 ab^{-1} generic MC raw data process + reprocess (Phys Run rehearsal)	10.05	29.60		766.70
3 ab^{-1} MC challenge in parallel with reprocessing rehearsal	17.10	58.80		652.30
Cosmic Run Processing		3		
Cosmic Run MC		3		
Cosmic Run Analysis		3		

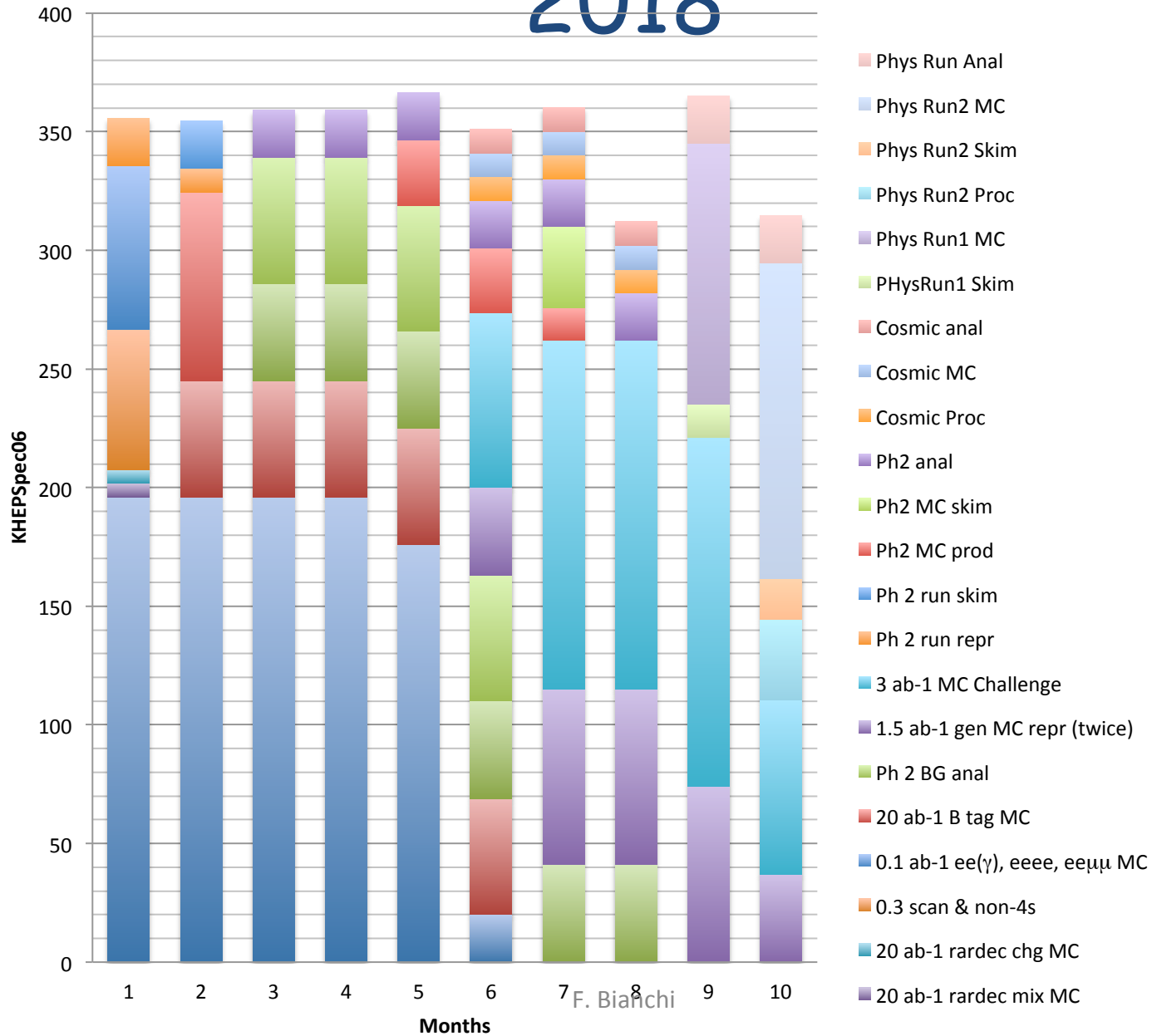
Resource Estimate 2018 (3)

Data Sample	Events ($\times 10^9$)	CPU (KHEPSpec06)	Tape (TB)	Disk (TB)
Phase 2 (Beast) run Raw data reprocessing	1.20	3.00		45.78
Phase 2 (Beast) run Detect data skim	1.20	2.00		45.78
Phase 2 (Beast) run MC production	2.00	6.85		76.28
Phase 2 (Beast) run MC skim	2.00	3.44		76.28
Phase 2 (Beast) run (Det + MC) analysis	3.20	12.00		12.21

Resource Estimate 2018 (4)

Data Sample	Events ($\times 10^9$)	CPU (KHEPSpec06)	Tape (TB)	Disk (TB)
Phys Run Month 1 processing	0.80	2.80		30.52
Phys Run Month 1 skimming	0.80	1.40		30.52
Phys Run Month 1 MC	3.20	11.00		122.07
Phys Run Month 2 processing	0.97	3.40		37.00
Phys Run Month 2 Skimming	0.97	1.70		37.00
Phys Run Month 2 MC	3.88	13.00		148.01
Phys Run Raw Data Storage	1.77		506.40	
Phys Run Analysis (Data and MC)	8.85	4.00		33.76

Time Profile of CPU Usage 2018



Resource Estimate 2018 (5)

- Grand Total:
 - CPU: 314.4 kHEPSpec to be used for 10 months
 - Tape storage: 3381.70 TB (2018) + 2260.22 TB (to be kept from 2017) = 5641.92 TB
 - Disk Storage: 4655.02 TB (2018) + 6488.41 TB (to be kept from 2017) = 11143.43TB

Total Resources for 2016-2018

Year	CPU (kHEPSpec06)	Tape (TB)	Disk (TB)
2016	220	815	4036
2017	314	3076	10534
2018	315	5642	11143

Italian Share (12%) for 2016-2018

Year	CPU (kHEPSpec06)	Tape (TB)	Disk (TB)
2016	26.4		480
2017	38		1264
2018	38		1337

- Al CNAF abbiamo 5 kHEPSpec + 150 TB
- Nei siti ReCaS:
 - 13 kHEPSpec + 800 TB pledged
 - Accesso ad altre 21 kHEPSpec non pledged
- Non abbiamo bisogno di risorse aggiuntive per il 2016
 - Segnalo pero' un problema nel throughput del networking verso Napoli (<500 Mbps verso KEK e < 900 Mbps verso PNNL)
 - Contatti ReCaS - GARR sono in corso