Belle II Computing Resource Estimate 2016-2018

F. Bianchi

INFN & University of Torino

Software/Computing Workshop

- 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

Important outcomes from the WS

**Schedule and plan of the Belle II distributed computing
(including data distribution, distributed analysis)**

→ see the chart in page 2

→ see the short-tem schedule in page 3

Keep using LFC+AMGA for e.g. 5 years from now on at least.

→ KISTI AMGA team will support it as far as Belle II uses (@ KISTI in this February)

→ need to exchange an MoU with CERN IT for the LFC support ASAP (@CERN in June)

→ We had a meeting with LHCb DIRAC developers during CHEP2015

→ ensure the LFC can coexist with DIRAC without any problem

→ need to confirm @ Ferrara 5/27-5/29)

Synergy with LHCb

→ many listed items, to be discussed at DIRAC workshop with DIRAC developers

@ Ferrara 5/27-5/29

Resources

→ Fabrizio will update the documents and plots in 10 days

→ Martin will prepare a pedagogical explanation on the requirement form physics
by the middle of the week from May 18.

Others

→ VOMS server at PNNL (at least before the replacement of the current KEKCC in summer 2016)

→ DIRAC server at Napoli by Silvio

Machine 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)

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
- Cosmic Processing ?? HEPSpec s/event
- Cosmic Analysis ?? 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(γ), eeee, ee$\mu\mu$
    - Important for calibrations and physics background studies

Activities in 2016 (2)

- Requested by Background Group (~60 s of bkg data):
  - 2.4x10^{10} Phase 1 background events (MC prod + analysis)
  - 2.4x10^{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(γ), 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 (x10^9)	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

Legend:
- 0.5 ab -1 raw data sim
- Cosmic anal
- Cosmic proc
- Phase 2 BG anal
- Phase 2 BG sim
- Phase 1 BG anal
- Phase 1 BG sim
- 20 ab-1 B tag MC
- 0.1 ab-1 ee(γ), eeee, ee$\mu\mu$ MC
- 0.3 scan & non-4s
- 20 ab-1 rardec chg MC
- 20 ab-1 rardec mix MC
- 5 ab-1 genMC skim anal
- 5 ab-1 genMC skim
- 5 ab -1 genMC prod

Axis labels: KHEPSpec06 (y-axis), Months (x-axis)
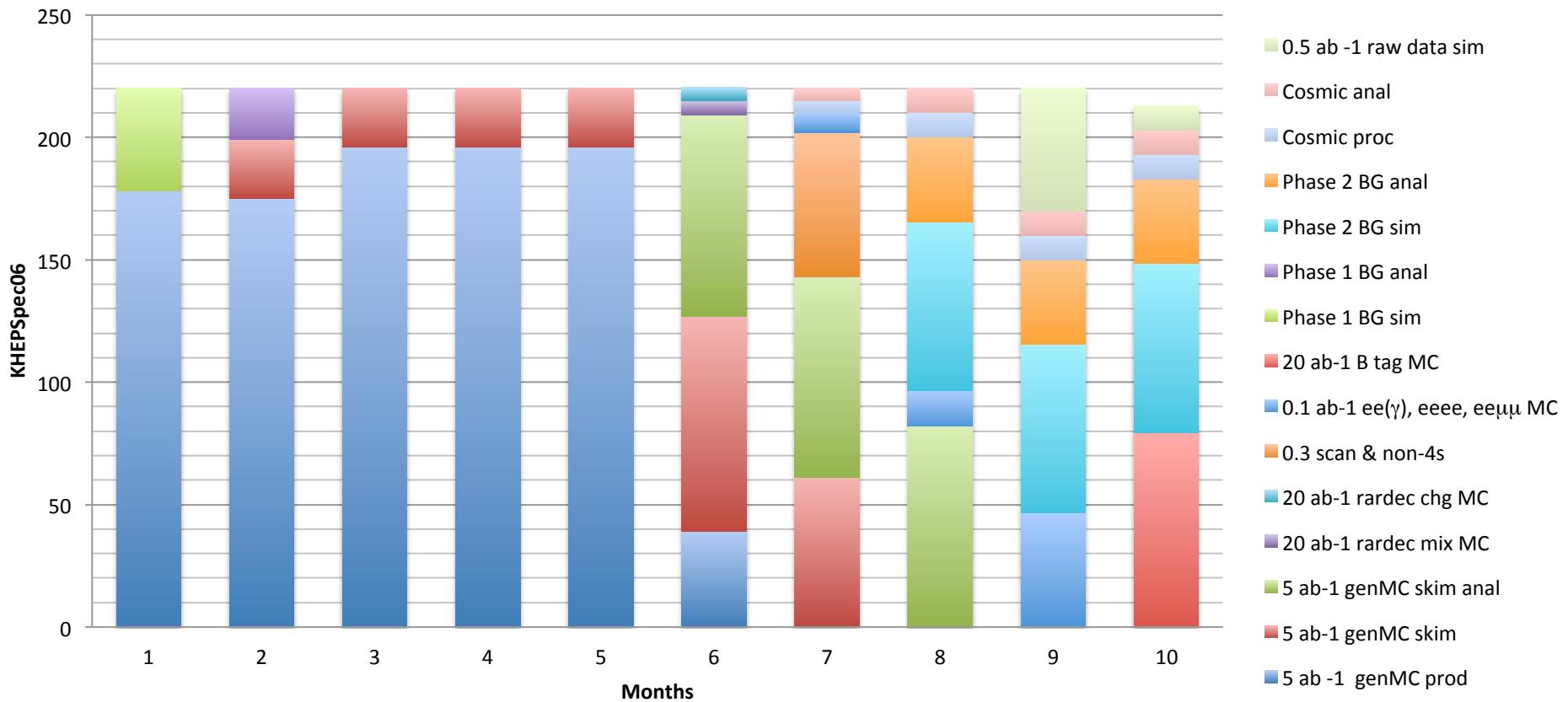
Resource Estimate 2016 (3)

- Gran 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):
  - 2.4×10^{10} 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 (x10^9)	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(γ), 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 (×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 (x10^9)	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

F. Bianchi
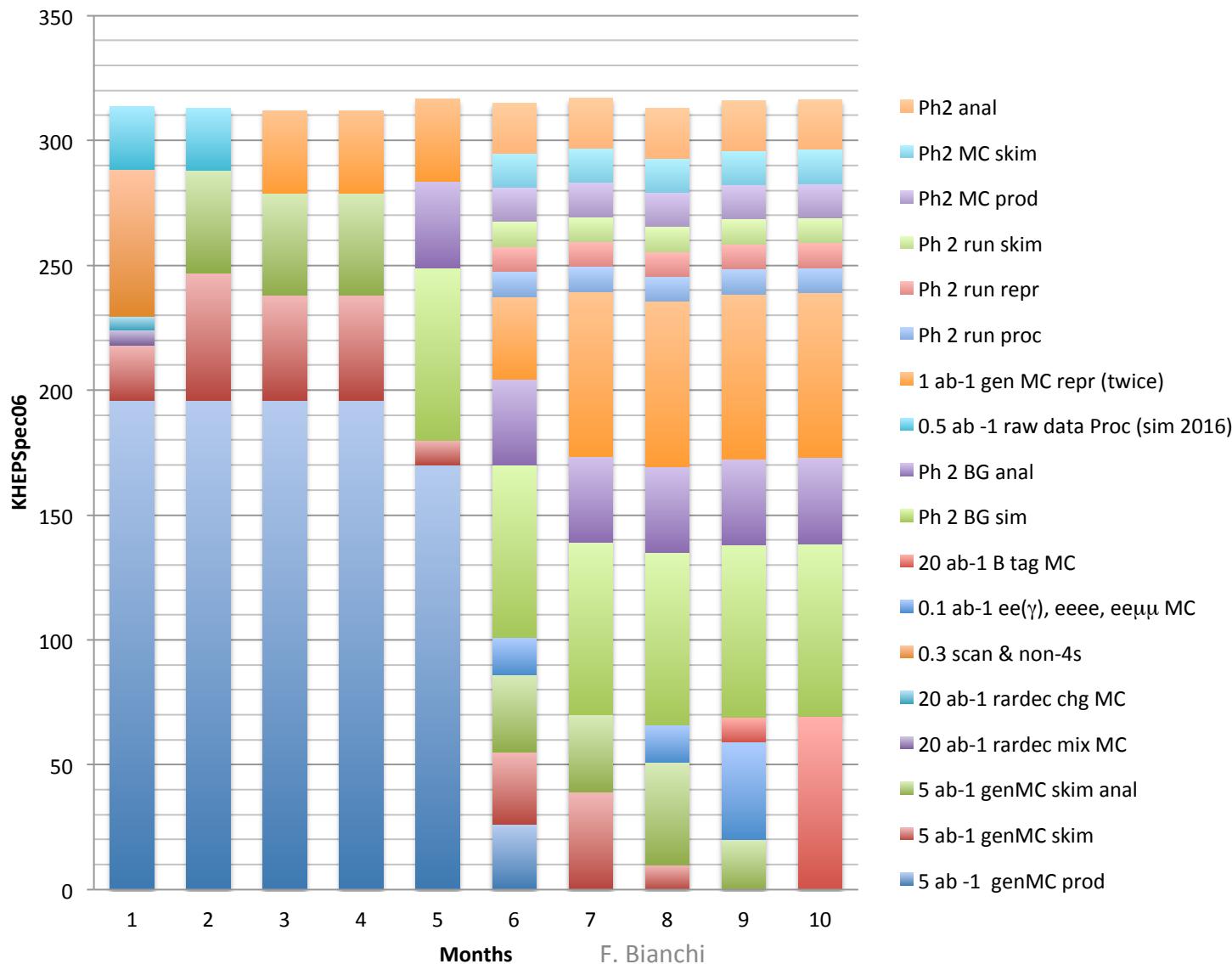
Resource Estimate 2017 (4)

- Gran 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(γ), 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 + reproces (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 (x10^9)	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 (x10^9)	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

F. Bianchi
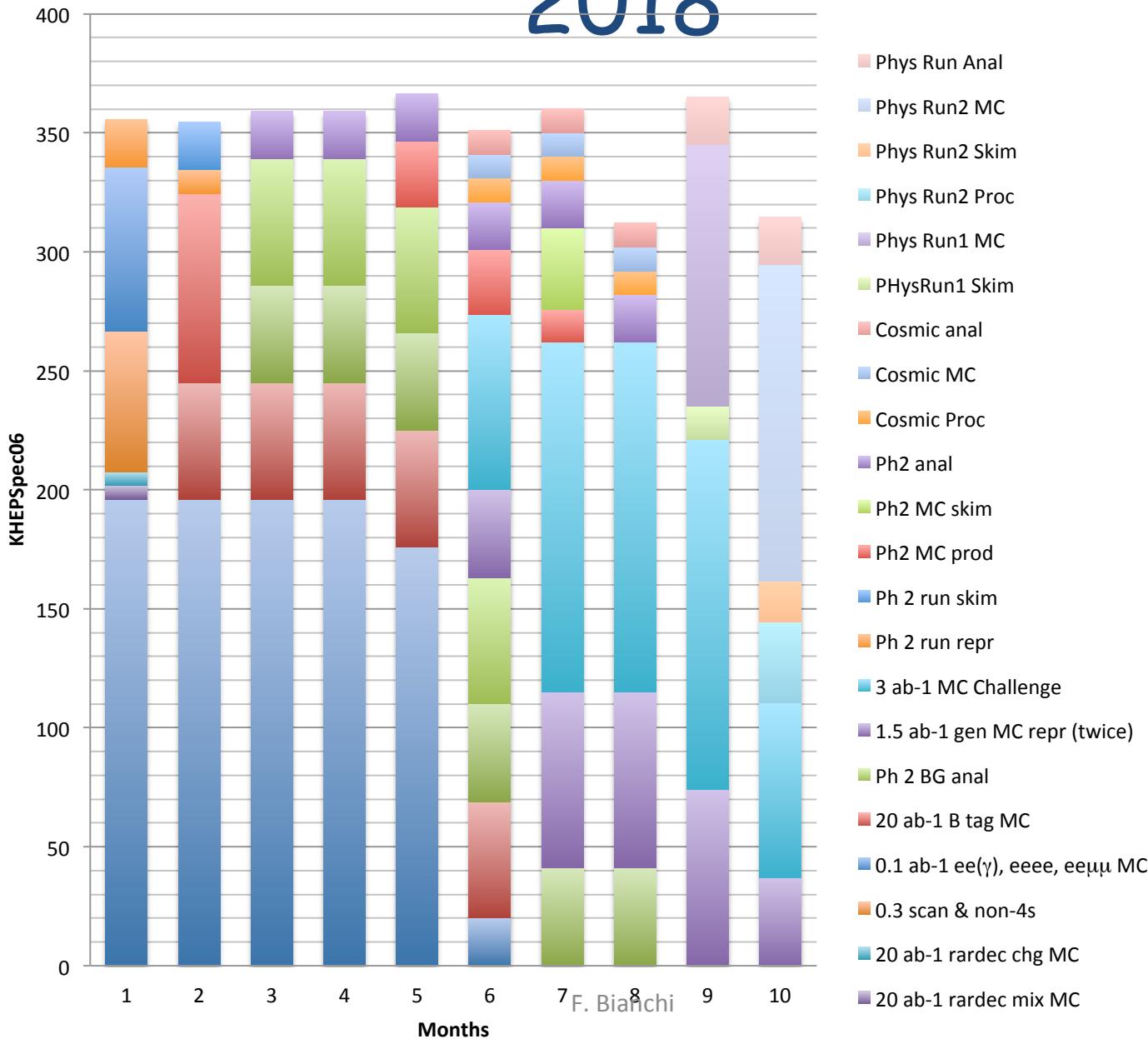
Resource Estimate 2018 (5)

- **Gran 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
- Il resto potremmo averlo nei siti ReCaS