

Calcolo dell'esperimento ALICE

Francesco Noferini – INFN sez. Bologna

Domenico Elia – INFN sez. Bari

Responsabile Nazionale Calcolo ALICE
Francesco Noferini

Deputy
Domenico Elia

Tier-1&Tier-2 Operations:

- coordinatore Tier-2: Sara Vallero
- referente Tier-1: Francesco Noferini
- bi-monthly meetings via zoom
- 1 Workshop annuale(2012@CT, 2013@TS, 2014@LNF, 2015@CNAF, 2016@PD, 2017@BA, 2018@TO, 2019@CT, 2020@ZOOM, 2021@????)

+ 1 meeting mensile coordinamento CNAF → CdG

+ Resource Computing Board (Noferini, Elia) durante ALICE Week



- **Tier-1 al CNAF, Bologna**
 - condiviso con altri esperimenti (LHC e altri)
- **4 Tier-2 ufficiali** (→ finanziati INFN e pledged)
 - Bari, Catania, Padova-LNL, Torino
- **Altre risorse**
 - Sito di Trieste
- Progetti che hanno fornito risorse nei siti INFN negli ultimi anni:
 - **PON ReCaS (BA e CT)**, dal 2014 al 2016 → pledges
 - **PON IBiSCo (BA e CT)**, al 2019 al 2021 → pledges
→ (contributo IBiSCo disponibile anche per pledge 2022 e 2023)



Istituto Nazionale di Fisica Nucleare

PON IBiSCo BA/CT



“Calcolo” e “Storage” INFN (risorse T2):

- scadute 18/12/2020 (1.3 M€ CPU HTC + 1.1 M€ Storage)
- Gara approvata in Giunta. → **previsione in produzione entro fine 2021**

IBiSCo CT (**aggiornamento**):

- Infrastruttura: Gara approvata in Giunta (giugno), a breve inizieranno i lavori

IBiSCo BA (**aggiornamento**):

- “Calcolo” Uniba (calcolo non-HTC)
 - scaduta 30/11/2020 (3 ditte) → gara terminata, in attesa di appr. dal CDA.
- “Impianti” Uniba (cogeneratore e condizionatore)
 - scaduta 4/12/2020 (2 ditte) → gara terminata, in attesa di appr. dal CDA.
- “Impianti” INFN (rack e PDU):
 - acquisti diretti, attrezzatura arrivata e collaudata
- “Rete” INFN (Switch):
 - gara terminata, in attesa consegna materiale



DISCO IN TB	BARI	CATANIA	PD/LNL	TORINO	TOT
DISCO INSTALLATO	1984	1976	1960	1859	7779
DELTA 2021	0	0	196	193	389
TOTALE*	1984	1976	2156	2052	8168

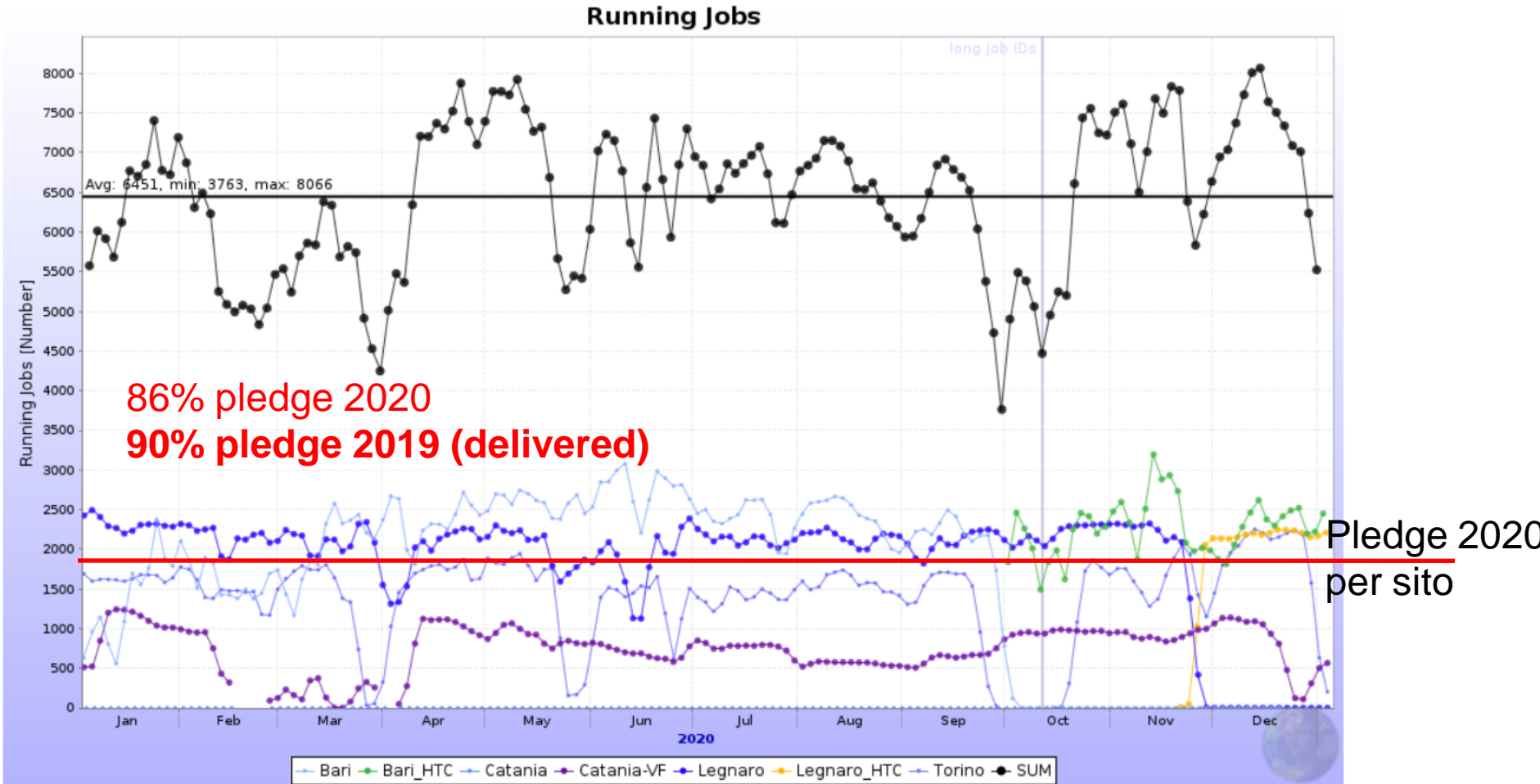
CPU HS06	BARI	CATANIA	PD/LNL	TORINO	TOT
CPU INSTALLATA	18993	18025	18993	18985	74996
DELTA 2021	0	0	2750	2700	5450
TOTALE*	18993	18025	21743	21685	80446

*Totale finanziato con anticipo 2022 (> pledged 2021)

Convenzione CONSIP

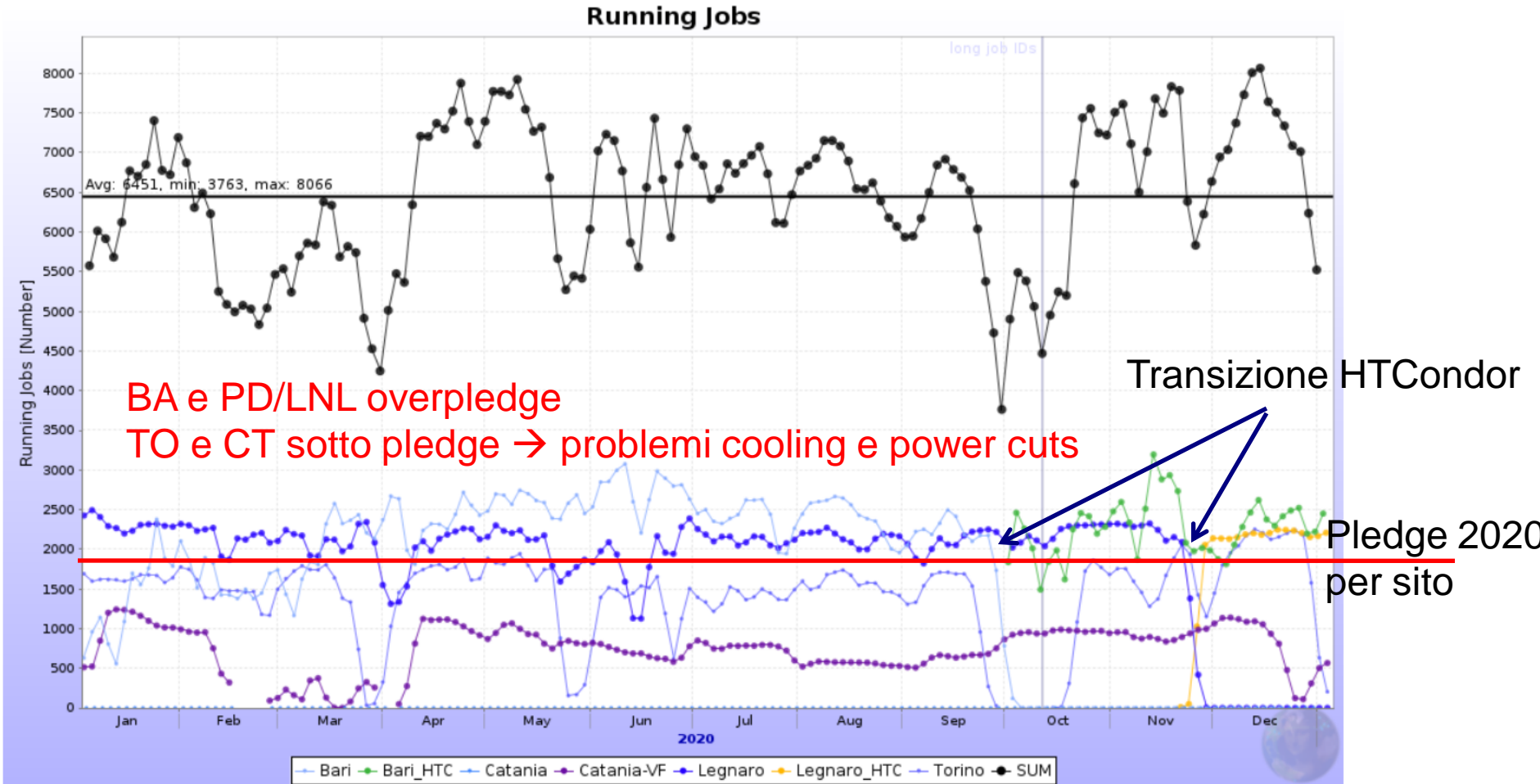
Uso risorse 2020

Incremento 2020 installato nel 2021 a causa di ritardi nelle gare (emergenza COVID)



Uso risorse 2020

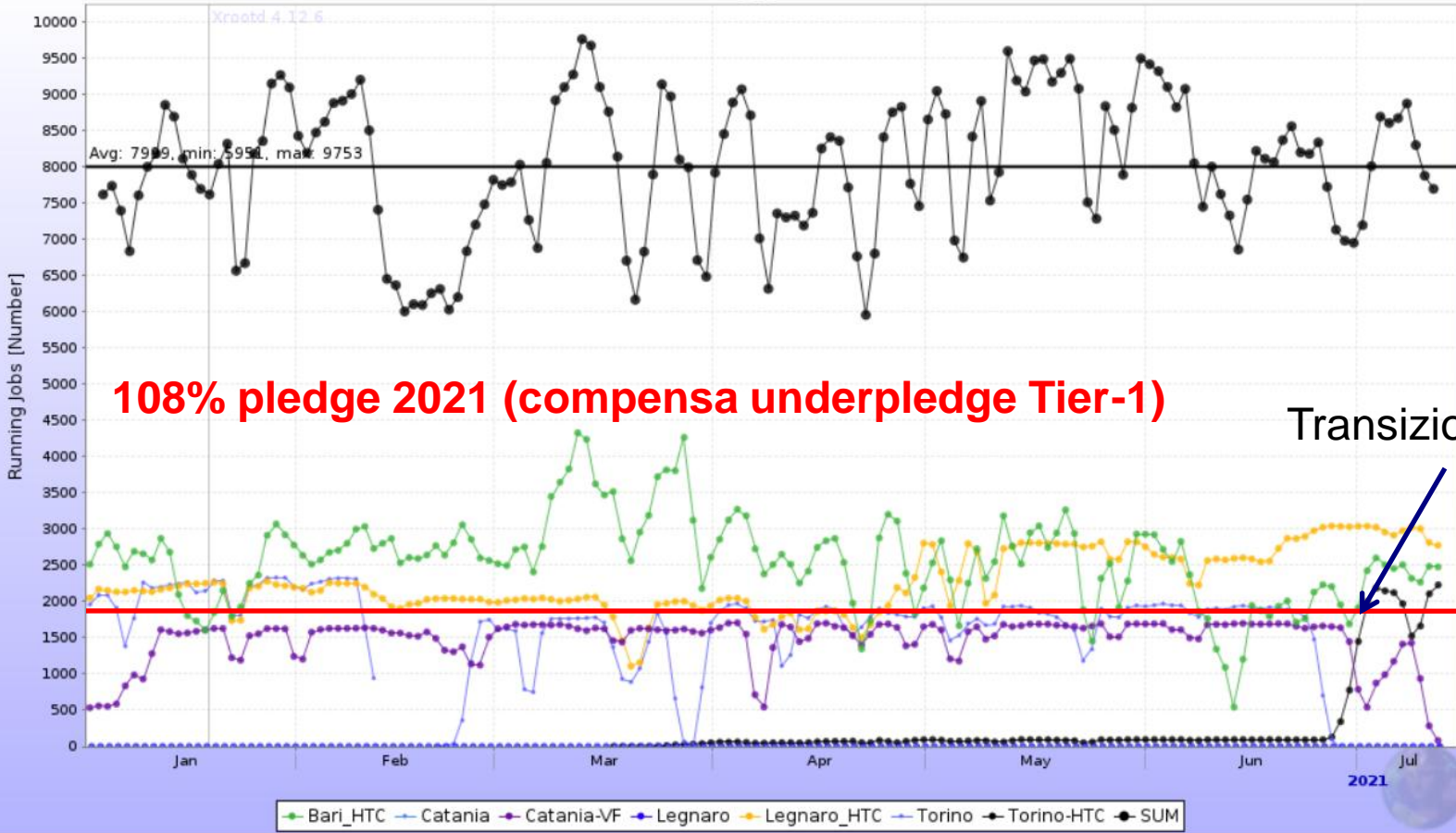
Incremento 2020 installato nel 2021 a causa di ritardi nelle gare (emergenza COVID)



Uso risorse 2021



Running Jobs



Situazione di sofferenza causa ritardi gare 2020 legati ad emergenza COVID
(anche se nel 2021 siamo a crescita zero)

DISCO (SE): 7.0 P used at 92% (limite operativo 95%) → CRITICITA'

BUFFER (TAPE): 574 T

Pledge 2021 → 8.23 PB

-1 PB rispetto ai pledge

ritardo gara 2020 → disco arrivato, installazione prevista a settembre

Setup coda
multi-cores

CPU

Pledge 2021 → 71400 HS06

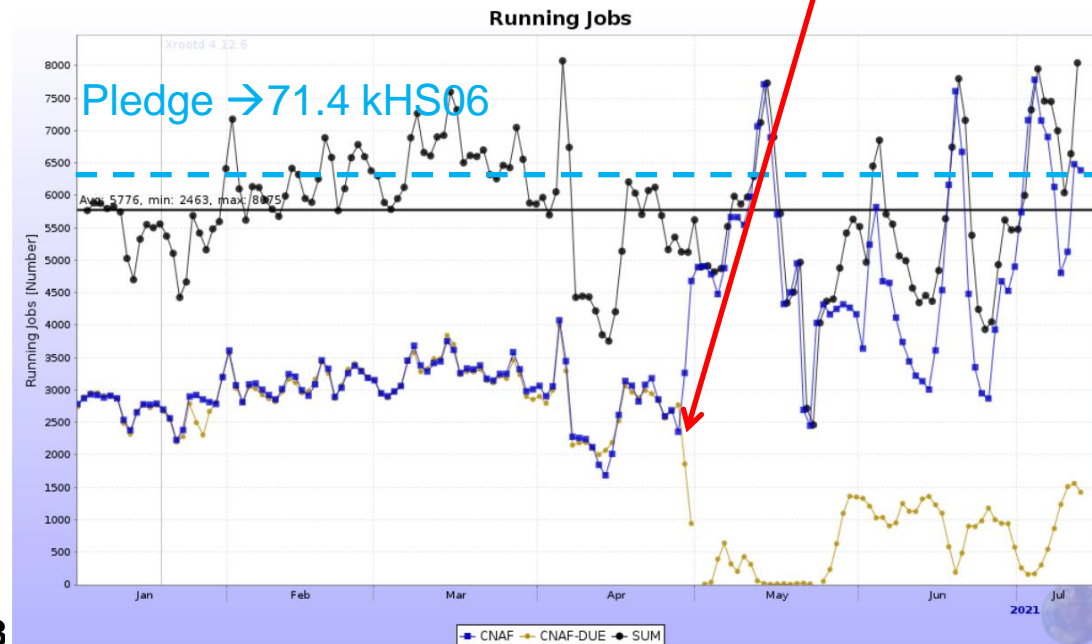
Utilizzo → 92.5%

Ritardo gare 2020 (Cineca)

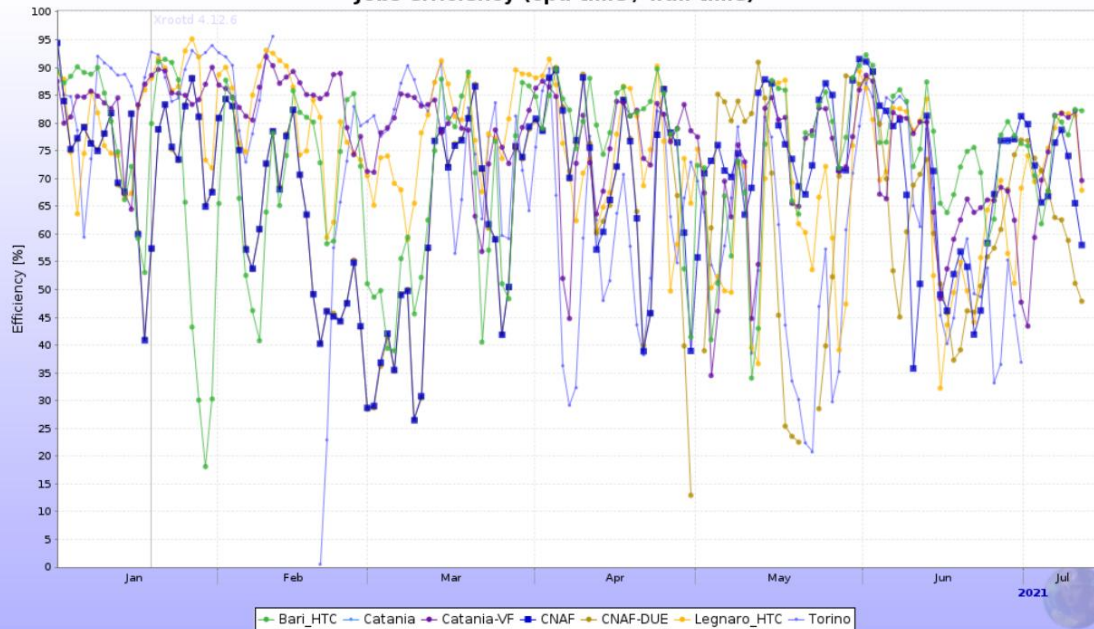
TAPE (11 PB) → no issue

Stiamo monitorando la situazione per non arrivare in ritardo nel 2022!

Migrazione al tecnopolo prevista a inizio 2023



Jobs efficiency (cpu time / wall time)



Jobs efficiency (cpu time / wall time)

	Series	Last value	Min	Avg
1.	Bari_HTC	82.21	0	72.5
2.	Catania	76.38	0.462	76.38
3.	Catania-VF	69.63	0	76.63
4.	CNAF	58.02	0	68.42
5.	CNAF-DUE	47.89	0	64.93
6.	Legnaro_HTC	67.88	0	73.92
7.	Torino	36.91	0	68.44
Total		62.7		71.6

Resource	Sim	Reco	Data Analysis
CPU	56%	10%	34%
Disk	50%	41%	9%

- attività di analisi intensa nel 2021 → efficienze più basse rispetto al 2022 (stiamo anche migrando a O2)
- in linea con la media ALICE
- primi test multi-cores al CNAF e hyperloop (analisi O2)

Update attività siti nel 2021

Nel 2021 sono state completate per tutti i siti le migrazioni a:

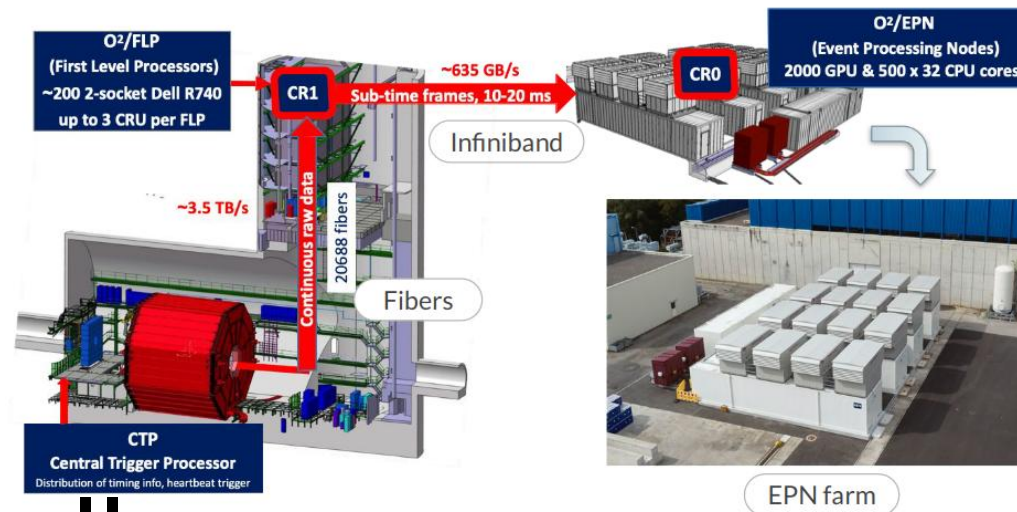
- HTCondor → batch system uniformato per tutti i siti Tier-1 e Tier-2
- IPV6 → Storage configurato in dual stack

O2 (Offline-Online) è il nuovo framework sviluppato per il Run-3 (continuous readout, triggerless), principali differenze rispetto a Run-2:

- processamento sincrono con compressione dei dati (**compressed time frame** → formato persistente su nastro, **1 replica**)
- sviluppo approccio dichiarativo per analisi
- alto livello di parallelizzazione e uso di GPU in ricostruzione

Attività correlata (fine 2020 – primo semestre 2021):

- Data sample portato nel formato O2 per validare nuovo framework. Prevista campagna di tutti i dati Run-2 a settembre 2022
- Job multi-cores: sfruttando l'alto parallelismo ALICE sta passando a job multi-cores (prima Tier-1 poi Tier-2) → passaggio trasparente con HTCondor
- Test con Marconi/Cineca su risorse HPC (simulazione O2)



Run-3 challenge

- Data reconstruction
 - Starts online: **synchronous reconstruction**, requires **online calibration**
 - Extremely demanding task → Dedicated farm Event Processing Nodes with **GPUs**
 - **Timeframes** rather than events as **fundamental unit** in reco algorithms
- Detector simulation and MC production
 - **New geometries** of the **detectors** with a **new digitization framework** to account for pileup
 - Timeframe reconstruction requires **larger memory resources**
 - **Multicore Grid operations** (until now single core jobs exclusively)
- Data analysis:
 - New analysis framework to work with increased regime of data to be processed
 - Flat data model: optimized for fast I/O and supports new data processing technologies on the market

Matteo Concas – CCR workshop 2021

https://agenda.infn.it/event/25889/contributions/135758/attachments/81871/107484/WCCR2805201_v5.pdf

Run-2 → Run-3

- Online reconstruction is possible thanks to GPU offloading, we aim for more
- Asynchronous reconstruction on GPU
 - TPC less relevant, lot of GPU resources available → offload larger set of algorithm on GPU
 - Some options already working, more to come. Different scenarios and priorities

Synchronous processing		Asynchronous processing	
Processing step	% of time	Processing step	% of time
TPC Processing	99.37 %	TPC Processing	72.01 %
EMCAL Processing	0.20 %	TRD Tracking	12.69 %
ITS Processing	0.10 %	TOF-TPC Matching	9.94 %
TPC Entropy Coder	0.10 %	MFT Tracking	1.69 %
ITS-TPC Matching	0.09 %	ITS Tracking	0.78 %
MFT Processing	0.02 %	TPC Entropy Decoder	0.73 %
TOF Processing	0.01 %	Secondary Vertexing	0.69 %
TOF Global Matching	0.01 %	ITS-TPC Matching	0.56 %
PHOS / CPV Entropy Coder	0.01 %	Primary Vertexing	0.14 %
ITS Entropy Coder	0.01 %	TOF Global Matching	0.11 %
FIT Entropy Coder	0.01 %	PHOS / CPV Entropy Decoder	0.10 %
TOF Entropy Coder	0.01 %	FIT Entropy Decoder	0.10 %
MFT Entropy Coder	0.01 %	ITS Entropy Decoder	0.06 %
TPC Calibration residual extraction	0.01 %	TOF Entropy Decoder	0.05 %
TOF Processing	0.01 %	MFT Entropy Decoder	0.05 %

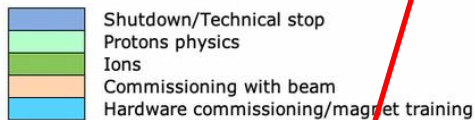
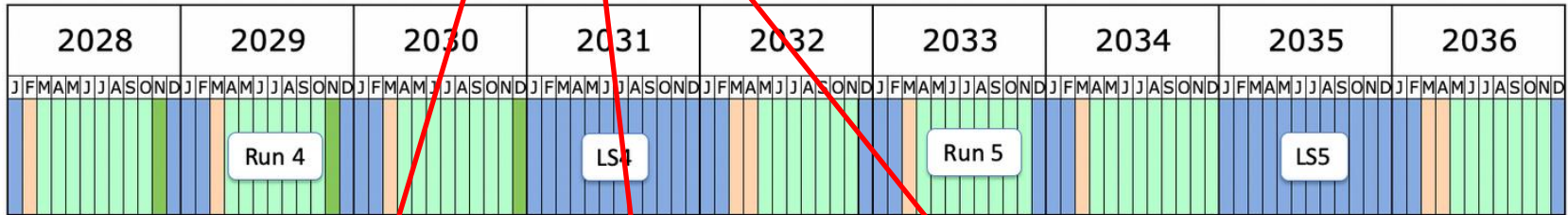
Running on GPU in baseline scenario
Running on GPU in optimistic scenario
Preliminary numbers: some algorithms not yet complete or not optimized!

D. Rohr, [vCHEP2021](#)

Miglioramento framework di analisi (speed up e ottimizzazione I/O)

- 99% del processing della ricostruzione sincrona su GPU (TPC,ITS)
- In corso di sviluppo porting su GPU anche della ricostruzione asincrona

	Run 2	Run 3	Run 2	Run 3
Benchmark	AliPhysics	O ²	AliPhysics	O ²
Histograms	3.9	38	4.5	9
Correlations	2	21	2.8	6
Spectra	4.6	14	5.4	4
	events/s		MB/s	



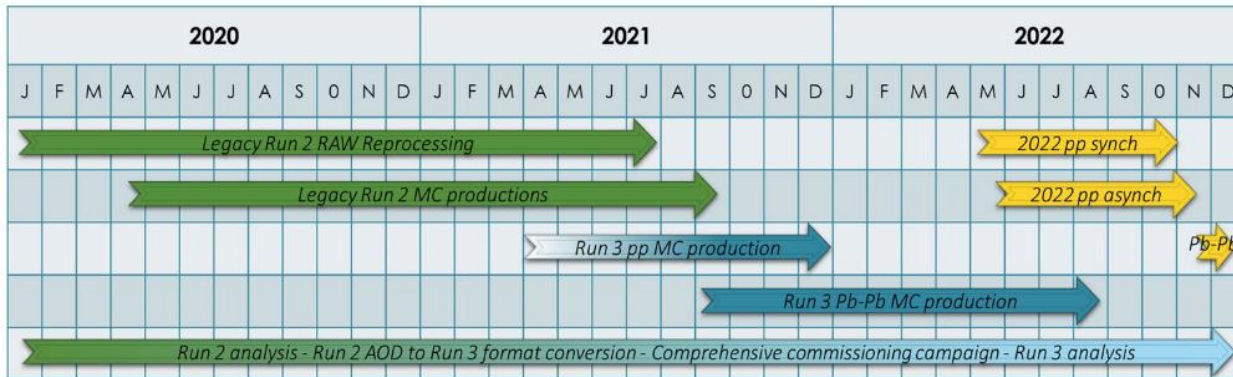
Pb-Pb@2.56/pb
ALICE

Last updated: June 2021

Pilot run
(1 week)
pp@900 GeV

pp@35/pb
ALICE

Programma ALICE (calcolo)



2022 Running scenario:

- **Commissioning:**
 - pp collisions at 20-200 kHz during ramp-up (28d)
 - pp collisions at 1 MHz for pp ref. exercise (3d)
 - pp collisions at 5 MHz IR (eq. load of Pb-Pb@50 kHz)
- **High-energy pp physics programme:**
 - 135 days pp for physics in 2022
 - 500 kHz interaction rate @ 45% efficiency
 - 35 / pb \Rightarrow $2.6 \cdot 10^{12}$ events
 - Strong event selection during asynch. pass on O2 facility (1.2%) \Rightarrow $3.1 \cdot 10^{10}$ events on disk
- **24 days of Pb-Pb in 2022:**
 - 24 days @ 50% LHC stable beams @ 95% ALICE eff
 - 2.56 / nb \Rightarrow $2.0 \cdot 10^{10}$ events

Fondamentale che le risorse di calcolo pledged siano disponibili prima della presa dati 2022

Share INFN 2020 – 2021 – 2022(?)

Lo scorso anno lo share INFN in ALICE (basato su peso relativo M&O-A)

- era sceso dal 20% al 17% (17.49 troncato a 17)
- gli share erano cambiati in modo brusco a settembre a seguito di riconteggi causando problemi nell'interazioni con i referee
- Per i Tier-2 (CSN3) avevamo deciso di mantenere le risorse approvate col precedente share come anticipo 2022 (soluzione che si è rivelata saggia)

Quest'anno gli share definitivi saranno fissati il 31/8 (prima dell'approvazione delle richieste ma dopo la discussione di luglio)

La fotografia al 30/6 prevede un aumento per lo share italiano dal 17 al 18% (18.4 troncato a 18!), per il nastro al Tier-1 dal 29% al 30%.

Abbiamo utilizzato questo nuovo share per le richieste correnti (da ridiscutere a settembre).

ALICE		2020			2021		2022			Share INFN
		C-RSG recomm.	Pledged	Used	C-RSG recomm.	Pledged	Request	2022 req/2021 C- RSG	C-RSG recomm.	
CPU	Tier-0	350	350	595	471	471	471	100%	471	
	Tier-1	365	354	405	498	412	498	100%	498	18% → 90
	Tier-2	376	435	405	515	481	515	100%	515	18% → 93
	HLT	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	
	Total	1091	1139	1405	1484	1364	1484	100%	1484	
	<i>Others</i>									
Disk	Tier-0	31.2	31.2	30.1	45.5	45.5	50	110%	50.0	
	Tier-1	44	41.8	35.9	53.3	45.4	55	103%	55.0	18% → 9,9
	Tier-2	39	43.2	36.1	44.8	50.4	49	109%	49.0	18% → 8,8
	Total	114.2	116.2	102.1	143.6	141.3	154	107%	154.0	
Tape	Tier-0	44.2	44.2	41.4	86.0	86.0	95.0	110%	95.0	
	Tier-1	37.7	44.4	35.8	57.0	59.8	63.0	111%	63.0	30% → 18,9
	Total	81.9	88.6	77.2	143.0	145.8	158.0	110%	158.0	

Table 6 ALICE resource requests for 2022 and C-RSG recommendations. For reference, the 2020 utilization and 2021 pledged resources are also given.

Share 18%/30%

	Pledge 2021	Finanziate 2021	Richieste 2022	Delta
Tier-1				
CPU (HS06)	71400	71400	89640	18240 x 10 E/HS06
Disco (TB)	8228	8228	9900	1672 x 140 E/TB
Tape (TB)	11948	11948	18900	6952 x 18 E/TB

Richieste incremento Tier-1 → CPU (182,5 kE) + Disco (234 kE) + Tape (125) = 541,5 kE

Missioni e overhead non riportate

Share 18%

Tier-2	Pledge 2021	Finanziate 2021	Richieste 2022	Delta
CPU (HS06)	73440	80446	92700	12252 x 10 E/HS06
Disco (TB)	7293	8168	8820	652 x 140 E/TB

Richieste incremento Tier-2 → CPU (122,5 kE) + Disco (91,5 kE) = **214 kE**

Dismissioni 2020	CPU 2021	CPU 1sem 2022	CPU 2sem 2022	Disco 2021	Disco 1sem 2022	Disco 2sem 2022
BARI	10512 HS	3065 HS	2466 HS	1224 TB	0	0
CATANIA	15125 HS	0	0	960 TB	0	0
PD/LNL	0	1504 HS	1504 HS	180 TB	0	0
TORINO	0	4200	1534	0	88	250

Dismissioni Tier-2 CPU (HS06) → **87,5 + 311,5 kE**

Dismissioni Tier-2 Disco (TB) → **72,5 + 306,0 kE**



160,0 kE + 617,5 kE

Capitolo	Richieste kE	Descrizione
Inventario	182.5	Crescita netta risorse CPU Tier-1: 18240 HS06 x 10 Euro/HS06
Inventario	234.0	Crescita netta risorse DISCO Tier-1: 1672 TB x 140 Euro/TB
Consumo	125.0	Crescita netta risorse TAPE Tier-1: 6952 TB x 18 Euro/TB
G. Totale	541.5	

Capitolo	Richieste kE	Descrizione
Inventario	214.0	Crescita netta risorse Tier-2 italiani: 12252 HS06 x 10 Euro/HS06 + 652 TB x 140 Euro/TB
Inventario	55.0	Dismissioni risorse Tier-2 Padova LNL: 3008 HS06 x 10 Euro/HS06 + 180 TB x 140 Euro/TB
Inventario	104.5	Dismissioni risorse Tier-2 Torino: 5734 HS06 x 10 Euro/HS06 + 338 TB x 140 Euro/TB
Inventario	96.0	Overhead risorse Tier-2 italiani 6% costo CPU + 7% costo totale
Inventario	332.0	Dismissioni risorse Tier-2 Bari: 16043 HS06 x 10 Euro/HS06 + 1224 TB x 140 Euro/TB
Inventario	285.5	Dismissioni risorse Tier-2 Catania: 15125 HS06 x 10 Euro/HS06 + 960 TB x 140 Euro/TB
G. Totale	1087.0	

Missioni

Sessione	Capitolo	Richieste (kE)	Descrizione	Person e
BARI	MISS		21 riunione comitato calcolo	x3
	MISS		12 Riunioni infrastruttura (T1+T2)	x1
	MISS	1,5	1 workshop T1/T2 ALICE (no resp calcolo)	x1
	Totale BA	4,5		
BOLOGNA	MISS		11 riunione comitato calcolo	x1
	Totale BO	1		
CATANIA	MISS	1,5	1 riunione comitato calcolo	x2
	Totale CT	1,5		
PADOVA	MISS	1,5	1 riunione comitato calcolo	x2
	Totale PD	1,5		
TORINO	MISS	1,5	1 riunione comitato calcolo	x2
	Totale TO	1,5		
TRIESTE	MISS		11 riunione comitato calcolo	x1
	Totale TS	1		
TOTALE		11	KE	

1)

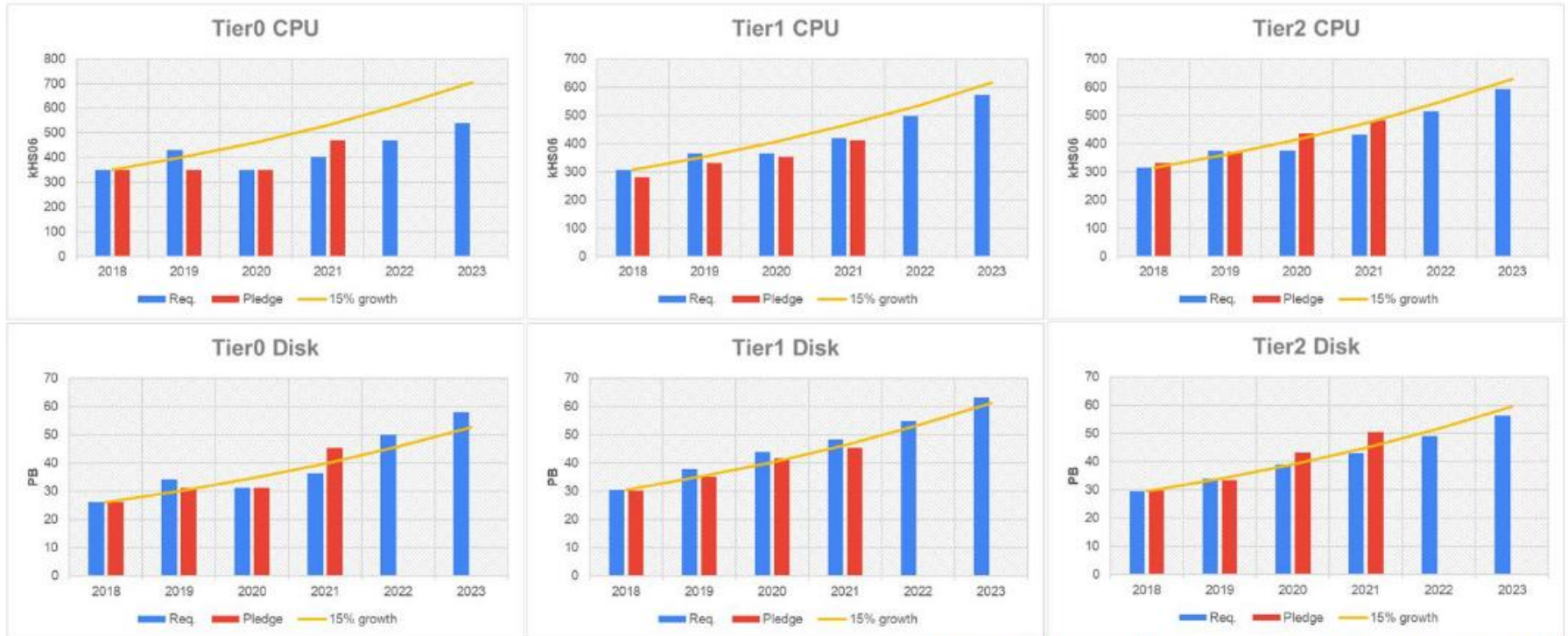
Market Trends (from WLCG report to OB)

- The most worrying trend is the worldwide shortage of semiconductors: demand is much higher than supply
- Many causes among them COVID:
 - Demand for notebooks/phones, shortage of air cargo capacity
 - Trade issue between countries
 - Environmental factors, as severe [drought in Taiwan](#)
- Effects:
 - **Large delays in procurement (from 3 to 6 months)**
 - Increase of some hardware cost:
 - +25 % for RAM memory since January
 - Hard Disks have seen a price increase of 70-100% during the last month due to the very high demand (processing of the new crypto currency Chia)
- **FA's need to consider early procurements in preparation for Run 3**



➔ Potenziale criticità
specialmente al
Tier-1

2) Il prezzo dei nastri sembra essere in diminuzioni (ultima discussion CNAF CdG).
Se confermato potrebbe diventare sostenibile il salvaggio di due repliche CTF (in discussione)



----- 15% growth: annually compounded rate from 2018 along LS2

ALICE		2022	2023	
		Req.	Est.	Est. 2023 / Req. 2022
CPU [kHS06]	Tier-0	471	541	15%
	Tier-1	498	572	15%
	Tier-2	515	592	15%
	Total	1484	1705	15%
Disk [PB]	Tier-0	50.0	58.0	16%
	Tier-1	55.0	63.0	15%
	Tier-2	49.0	56.0	14%
	Total	154.0	177.0	15%
Tape [PB]	Tier-0	95.0	126.0	33%
	Tier-1	63.0	79.0	25%
	Total	158.0	205.0	30%

Estimates for 2023:

- Moderate growth of CPU (+15% per year) and disk
 - Work to offload tasks to GPU in asynchronous reconstruction ongoing
 - This will ensure to exploit GPUs when EPN farm is not (fully) used for synchronous
 - Reassessment of the workloads computing shares among EPN and Tiers when the first estimates on GPU are available
- Strategy B helps to mitigate the needed disk space in tape:
 - For 28 Pb-Pb days: +31 PB at T0, +16 PB at T1s
 - As an exercise, p-Pb @ 1MHz will require +22 PB at T0, +11 PB at T1s
 - Needed to have running conditions and computing model parameters for p-Pb
- **A better assessment of the computing estimates for 2023 will be delineated in Autumn**

Stime presentate al Resource Computing Board di ALICE (1 luglio 2021)



Grazie!