

Report CNAF reloaded

22/04/2022

Outline

- News da accordi CINECA
- News variante di gara
- News dal progetto
 - Stato milestones etc
- AoB

News su accordi CINECA

- Accordo per utilizzo delle risorse in attesa di firma
- Per quanto riguarda l'uso di Leonardo per le pledge 2023-25 (news: copre anche 2026 adesso!) nessuna sorpresa
- Nodi Booster ad uso oggi dei teorici ma domani anche di WLCG: allocazione molto soddisfacente
- Inoltre: accordo raggiunto anche su CPU a CINECA/Casalecchio (prox slide)

Tabella (tab. 1) per la durata dell'accordo.

Partizione	Risorse 2023	Risorse 2024	Risorse 2025	Risorse 2026
General Purpose (CPU)	Fino a 165 nodi	Fino a 300 nodi	Fino a 300 nodi	Fino a 300 nodi
Booster (GPU)	Fino a 3 milioni di NODE-hours/anno	Fino a 3 milioni di NODE-hours/anno	Fino a 3 milioni di NODE-hours/anno	Fino a 3 milioni di NODE-hours/anno

- f) Cineca conferma il proprio impegno a concedere in uso dedicato la macchina Intel ~~Broadwell~~ citata nelle premesse del presente accordo al punto 1.2 ~~fino al decommissioning del sistema~~ secondo il profilo di utilizzo indicato nella tabella seguente:

Anno 2022	Anno 2023	Anno 2024	Anno 2025
216 nodi	216 nodi	144 nodi	72 nodi

g)

- g) Cineca renderà inoltre disponibili ad INFN nella medesima infrastruttura di cui al punto f) ulteriori 288 nodi ~~Broadwell~~ ~~fino alla fine del 2026~~; ~~fino al decommissioning del sistema~~;

CPU a CINECA - Casalecchio

- Dal 2018 una (grande) parte delle risorse pledge CNAF è fornita da 3 rack a CINECA/Casalecchio (~60 kHS06 per rack, $3 \times 72 = 216$ nodi) → ~180 kHS06
 - Proposto inserimento nell'accordo il profilo di dismissione (verbalmente approvato dal CINECA) di queste macchine (costo elettricità non indifferente)

Anno 2022	Anno 2023	Anno 2024	Anno 2025
216 nodi	216 nodi	144 nodi	72 nodi

- Prevista per settembre 2021, poi fine Q2 2022, l'aggiunta di altri 4 rack ($4 \times 72 = 288$ nodi, 240 kHS06), per un totale di 7 rack → ~420 kHS06
 - In produzione almeno fino al 2026

Nuove stime CPU

- Per il 2022 i referee del calcolo hanno finanziato un quantitativo INFERIORE di quello previsto dal piano di investimenti di CNAF reloaded (anche se tecnicamente non siamo ancora nella fase CNAF reloaded)
 - Pledge di WLCG minori del previsto
 - NON finanziato il +10% Cloud
 - Ritardate le dismissioni di CPU del 2022
- Alla luce di questo, e del fatto che per il 2023 sappiamo già che le richieste WLCG saranno inferiori al +20%/y modellizzato, abbiamo rifatto il piano

Year	Y2021	Y2022	Y2023	Y2024	Y2025	Y2026
Standard Pledge	557000	600000	690000	900000	1080000	1296000
Pledge +10%	557000	600000	759000	990000	1188000	1425600
CNAF						
installed	198238	149333	86333	86333	0	0
To be replaced	48905	27000	0	86333	0	0
Available	149333	122333	86333	0	0	0
Bari						
installed	10000	0	0	0	0	0
To be replaced	10000	0	0	0	0	0
Available	0	0	0	0	0	0
CINECA						
installed	417667	417667	417667	417667	357667	297667
To be replaced	0	0	0	60000	60000	60000
Available	417667	417667	417667	357667	297667	237667
Tecnopolis						
installed	0	0	96000	96000	96000	59600
To be replaced	0	0	0	0	36400	0
Available	0	0	96000	96000	59600	59600
From Leonardo	0	0	330000	600000	600000	600000
Total to be replaced	58905	27000	0	146333	96400	60000
Total available	567000	540000	930000	1053667	957267	897267
To be acquired	0	60000	0	0	230733	528333
Operational cost (E/HS06) -20%	8	6.4	5.12	4.096	3.2768	2.62144
security hosts			60000			
Total cost	0	384000	60000	0	756065.8944	1384993.26

Outcome:

- Per il momento il modello in cui 2023 e 2024 NON servono acquisti diretti è **confermato**
- Nel 2023 c'è spazio per il +10% cloud, ma non sappiamo se avverrà l'allocazione da parte dei referees

Un criticità generale in vista?

- Da modellizzazione WLCG avevamo assunto un calo dei prezzi unitari delle CPU del 20%/y
- 2022: aspettati 6.4 €/HS06, ma non pare realistico. Gara bandita a 10 €/HS06, come l'anno scorso.
 - Incertezza dovuta a situazione internazionale
 - Questo non implica che poi i prezzi di aggiudicazione saranno più vicini a 10 che a 6.4 €, ma la situazione è preoccupante, soprattutto se estrapolata agli anni successivi
- La stessa considerazione vale per tutto. Anticipiamo (vedi dopo) aumenti di prezzo generalizzati ...

	2021	2022			2025	
unitary cost (E/HS06) -20%	8	6.4	5.12	4.096	3.2768	2.62144
security hosts			60000			

Variante di gara

- Documento di variante consegnato a CINECA
 - Effettuata richiesta ufficiale a CINECA per la gestione della variante.
 - Incontro con ditta esecutrice (CEFLA) e CINECA il 27/4 per nuova stima dei costi. Ci aspettiamo rincaro non banale (vedi dopo)
- Accordo attuativo per formalizzare assegnazione lavori a CINECA per la gestione della variante pronto
 - Goal: approvazione da CD INFN e CdA CINECA entro Maggio
- Una piccola descrizione dei cambiamenti e delle criticità ...

Variante ...

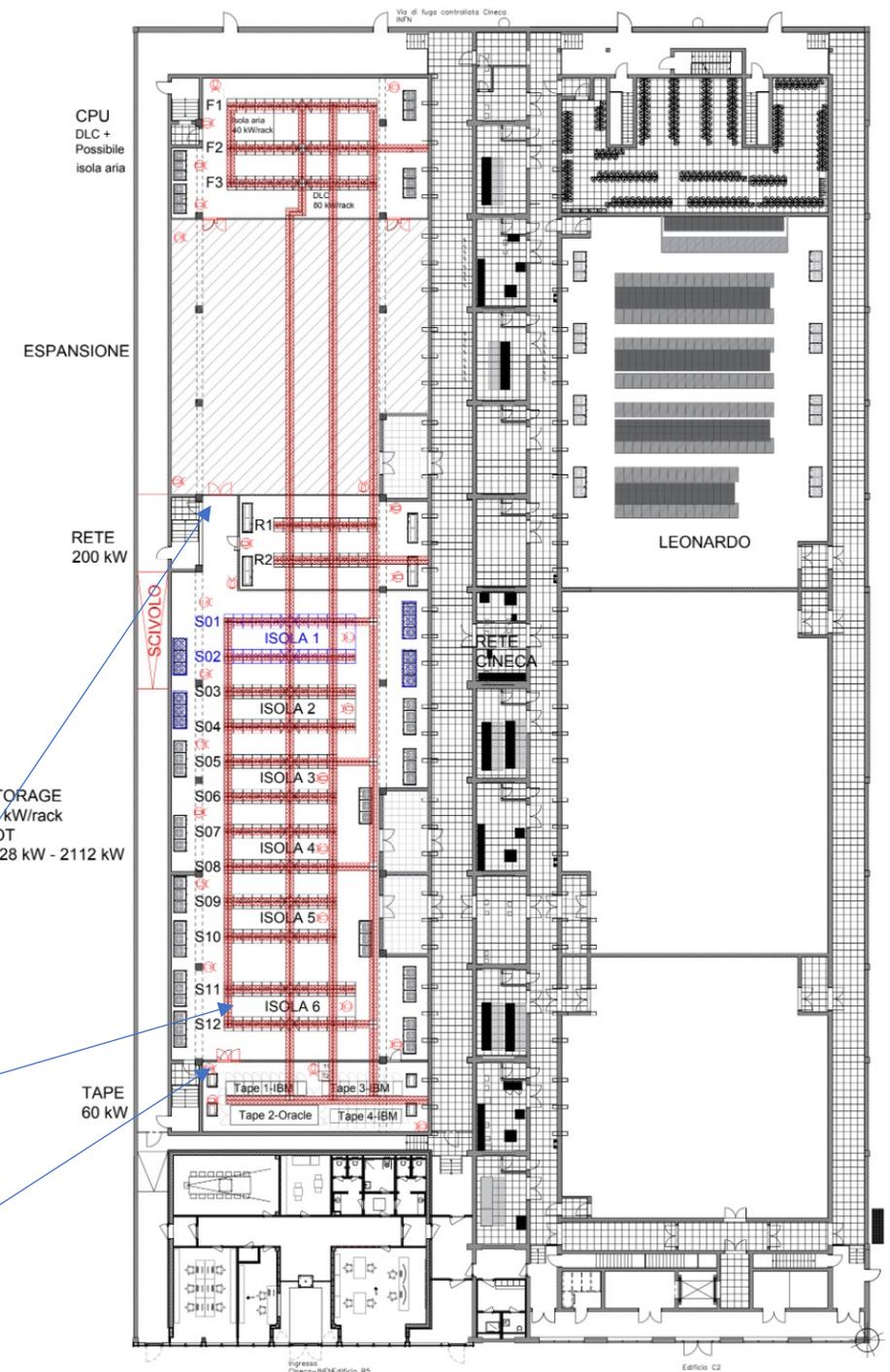
- Documenti [qui](#)
- Parti colorate modifiche rispetto a Progetto esecutivo
 - Rosso = modifiche richieste
 - Blu = modifiche in opzione (quotazione separata)

La questione dell'accesso ...

Porta aggiunta

Canaline come da richiesta rete

Porta allargata



Modifiche salienti rispetto al TDR (1/2)

- Zona alta densità NON equipaggiata
 - 2023 – 2024 non serve acquisire CPU
 - Da contatti con vari vendor abbiamo capito che non è conveniente installare ora la parte di distribuzione idraulica (tubi per acqua temperata) perché da qui a 2 anni gli standard per rack DLC potrebbero cambiare
- 6^a isola zona low-density in opzione, insieme a 4 CRAH
 - Predisposizione per la 6^a, installabile poi a sistema attivo
 - Motivazione: capire eventuale risparmio se necessario, inutile acquisire ora CRAH e tenerli inattivi, ...
- Porte e canaline ridisegnate
- Sistema di videosorveglianza integrato con quello generale del Tecnopolo (serve anche per ISO 27001)

Modifiche salienti rispetto al TDR (2/2)

- Integrazione della gara rack nella variante per i lavori
 - Avere una soluzione certificata e integrata rack e CRAH è preferibile
 - Rack: **capitolato pronto**, in attesa di capire se verrà inserito in variante o andrà bandita gara
 - Se NON possibile mettere in variante stiamo valutando la possibilità di fare affidamento diretto per avere lo stesso fornitore
- CINECA ha deciso di disporre dovunque un pavimento flottante con maggiore portata, anche per problemi di tenuta sismica (costo non in variante ma da considerare...)
- Nostra previsione su costo finale
 - Precedente stima costo variante (quella con raffreddamento ad aria) era 2,550 k€
 - Zona ad alta densità non popolata, e solo metà dei CRAH → - 200 k€
 - Rack alta densità non presenti → - 100 k€
 - Diminuzione della portata delle blindosbarre → (1000 A → 400 A) ~ -100 k€
 - **A valle di questo, PREZZI GENERICAMENTE AUMENTATI DEL 50%! → ci aspettiamo ~3 M€ + ~1 M€ se riusciamo a includere i racks**

L'accesso è un problema (NON E' IN VARIANTE, strettamente) dovuto al fatto che la struttura è protetta dalle belle arti. Piano originale prevedeva un unico accesso per CINECA e CNAF mediante montacarichi (neanche molto grande!)(*)

Situazione non soddisfacente

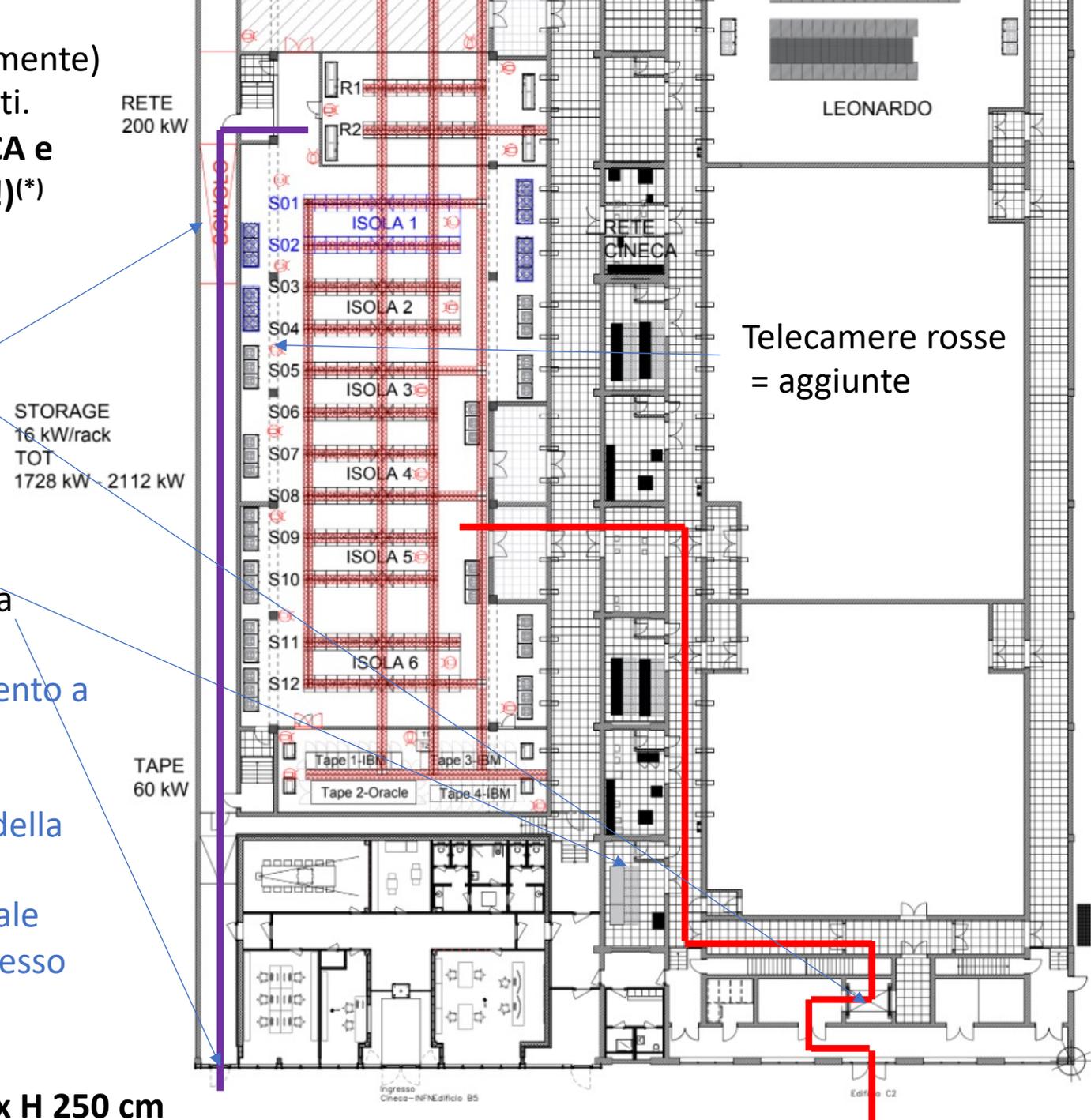
- Con montacarichi rotto, nessun accesso
- Curve curve e curve

Richiesto accesso diretto con **scivolo** (anche il CINECA ha poi deciso di fare scivolo a dx)

Ok, ma c'è il problema per ora insoluto della porta sulla facciata

- 120 cm di luce, richiesta alle belle arti per ampliamento a ~180 cm ma "difficile"
- Tecnicamente il pezzo + grosso che abbiamo (base della libreria ORACLE) è 236x170x110 cm³ e passa sia dai montacarichi sia dalla porta di 120 cm, ma in generale avere un centro nuovo con queste limitazioni di accesso non pare saggio

(*) Dimensione montacarichi: Cabina W 150 x L 270 x H 250 cm



Tempistica

- Alla scorsa riunione, tempistica per la consegna hall B5 “istrumentata con secondari”: Marzo 2023, con possibile spostamento a Luglio
- Ultima stima: a Marzo 2023 «hall pronta per installazione secondari», quindi almeno +3 mesi di ritardo si stanno consolidando
- Problema delle consegne materiali può essere complesso:
 - E' assolutamente necessario che tutta la procedura burocratica per la variante si consolidi e ordine parta prima dell'estate
 - Per arrivare a questo mancano i seguenti step
 - Preventivo CEFLA (forse già il 27 Aprile)
 - Passaggio in CD e CdA CINECA Maggio o Giugno al limite
 - Lato ns, chiesto a DG di mettere in calendario per GE e CD di Maggio, in deroga ai tempi standard
 - Ordine CEFLA → fornitore in Estate

Stato milestone e Change Request

- Gare
 - Storage
 - Gara 2021-22 approvata GE fine Marzo.
 - Preparazione bando in stand-by (problemi in AC): pubblicazione metà maggio
 - AQ storage 2023-24 – in attesa consegna hw per test: capitolato fine luglio
 - CPU – gara 2022 approvata GE metà aprile
 - Rete - da TDR:
 - *DCI per periodo intermedio (Apr-22)*: verrà acquisito dal GARR, gara non necessaria
 - *Cabling + management network (Apr-22)*: capitolato pronto
 - *Core switch + transceivers (May-22)*: capitolato in fase avanzata di scrittura
- Altre milestone
- Change requests 143 e 144
 - 143: Eliminazione CE dedicato a Leonardo
 - 144: Modifica della configurazione per includere risorse pledged da Leonardo

Change requests

- **143:** Già discussa l'altra volta. L'uso opportunistico di Leonardo sarebbe dovuto avvenire tramite un Computing Element dedicato; siamo riusciti a uniformarlo con i Computing Element standard (meglio!)
 - Nessun effetto previsto su nessun'altra attività, ma usato per testare la procedura di CR
- **144:** "Configurazione V1" tiene conto del fatto che le CPU di Leonardo sono disponibili a pledge. Questo porta al 2025 ad una netta diminuzione dei consumi lato INFN
- Problemi →

Configurazione Data Center INFN al Tecnopolo		Version	XXX
Sistemi di livello 1		Sistemi di livello 2	Sistemi di livello 3
Sistemi di livello 3		Sistemi di livello 4	Link
Power consumption		Rack	
Infrastrutture tecniche			
Power Available for IT (kW)		3000	
Cooling with chilled water (kW)		6000	
Spazio racks bassa densita (inclusa rete)			
Spazio racks alta densita (includa rete)			
Racks totali			
IT Infrastructures			
Data Center			
Farming			
HTC		192	
HPC		128	
External hosting		208	
Tier-1 services		96	
Disk		430	
Tape		60	
Networking		220	
Cloud		160	
EPIC Cloud		160	
Information System + national services		80	
Consumo residuo CRAH		260	
TOTAL		1994	

1994 kW vs 2500 kW da TDR

Problemi CR

- **Il flusso è davvero complesso:** come da descrizione TDR, 20 persone (Task Leader) devono dare acknowledgement, sia che la CR abbia implicazioni per il loro task o no
- Sarà anche il periodo pasquale, ma anche con ripetuti reminder il processo è lungo e non ancora concluso. E' un processo davvero troppo lento per essere usabile
- **Proposta di revisione: dopo un timeout di N giorni (7?) basta l'ok del WP leader (non servono i Task Leader)**

Previous configuration link

<https://istnazfisnucl.sharepoint.com/>

Inserire il link SharePoint che punta al documento contenente la configurazione di partenza

New Configuration Link

<https://istnazfisnucl.sharepoint.com/>

Inserire il link SharePoint che punta al documento contenente la nuova configurazione

Aknowledge Status

- Boccali
- Cesini
- Chierici
- Ciaschini
- Dal Pra
- dell'Agnello
- Duma
- Falone
- Fattibene
- Giacomini
- Guizzunti
- Longo
- Martelli
- Salomoni
- Sapunenko
- Scarponi
- Spiga
- Vistoli
- Zani

What the submitter thinks are the possible consequences (editable)

Assignee Luca Dell'Agnello

Affected Tasks

- WP1.Project-Office
- T2.1.1.Farming
- T2.1.2.DataManagement-Storage
- T2.1.3.Network
- T.2.1.4.Cloud
- T.2.1.5.EPIC-Cloud
- T.2.1.6.Operations
- TT.2.1.7-InformationSystem
- T.2.1.8.SSNN
- T.2.1.9.UserSupport
- T.2.2.AAI
- T2.3.ITSecurity
- T2.4.HTC-HPC
- T2.5.ExternalProjects
- T2.6.Migration
- WP3-Technical-Infrastructure
- WP2.IT-Infrastructure

People who need to say at least "I saw it", and possibly comment further

All [Comments](#) [Work Log](#) [History](#) [Activity](#)

Stefano Zani added a comment - 1 week ago

Ho visionato la nuova configurazione. Indipendentemente dall'utilizzo di Leonardo, il dimensionamento della rete in termini di numero di porte non richiede modifiche in quanto è stata dimensionata per consentire comunque il collegamento nodi di calcolo attraverso switch di TOR. Rimane l'incertezza sulla possibilità di sfruttare a piene performance gli skyway ma assumendo che gli skyway funzionassero fino ai loro valori di targa, passare da circa 300 a 600KHEPspec potrà rendere necessario un raddoppio degli skyway stessi oppure un aggiornamento delle porte degli skyway prima di raggiungere il totale delle risorse di calcolo se vogliamo garantire i 5 MBps per core se assumiamo che 600KHEPspec siano forniti da 60.000 core, la banda passante teorica sarebbe di 2.4 Tbps (60000x40)

PROJECT INDICATORS - EARNED VALUE MANAGEMENT

PROJECT INDICATORS - MILESTONES ADVANCEMENT

CPU		RACKS	
PV		PV	
AC		AC	
% Avanzamento		% Avanzamento	
EV		EV	
SPI		SPI	
CPI		CPI	
General Status		General Status	

DISKS		NETWORK	
PV		PV	
AC		AC	
% Avanzamento		% Avanzamento	
EV		EV	
SPI		SPI	
CPI		CPI	
General Status		General Status	

TAPES	
PV	
AC	
% Avanzamento	
EV	
SPI	
CPI	
General Status	

NOT APPLICABLE AT THIS STAGE

MILESTONES ADVANCEMENT

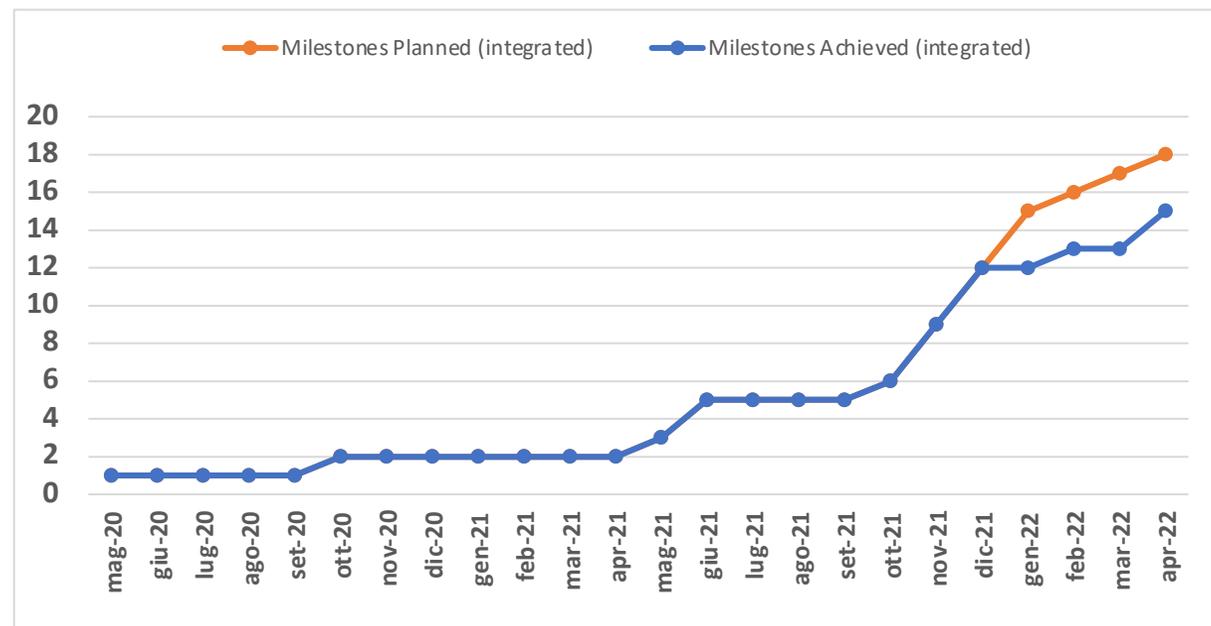
#Milestones Planned = 18
#Milestones Achieved = 15

Milestone accomplishment:

MPI=83%

Milestones delayed:

- WP1 - M1. Tender for Storage - *Diviso in due slot. 2022 ok. 2023/2024 uscirà a Settembre 2022.*
- WP1 - M7. Tender CPU - *Tender per le risorse 2022 approvato. Si copriranno le esigenze ulteriori con l'uso di Leonardo.*
- WP1 - M5. Tender Racks - *Capitolato pronto: in fase di verifica se possibile inserirlo in variante lavori. 80%*
- WP1 - M3. Tender for networking - *Cablaggio pronto, 50%*
- WP3- M1. Tender for secondary plants - *Capitolato definito (da firmare accordo con CINECA). 80%*



AOB

- **Rischio rete:** siamo oggettivamente preoccupati dal fatto di NON aver ancora potuto testare i device che connetteranno Leonardo allo storage/rete CNAF (i famosi Skyway)
- Sulla carta gli oggetti fanno quello che serve come macro descrizione, ma sono una tecnologia nuova e di nicchia, e ci sono dei dettagli che preoccupano un po' (capacita' di usare tutta la banda, MTU massimo, ...)
 - Discusso con il Presidente: unica soluzione CERTA sarebbe dire "non usiamo Leonardo" e compriamo tutte le risorse
 - Direzione presa: andiamo avanti cosi', e cerchiamo di testare prima possibile questi oggetti → contatti Zani → CINECA

