INAF

DS - ICT

Calcolo in INAF
CINECA - CHIPP - Commercial Cloud

Conclusioni… o nuovo inizio??

U. Becciani et al

ICT Workshop
23 Ottobre, 2019
Milano

Mou/Framework INAF - Cineca.

➔ Risorse annuali di calcolo per totali **50-55 Milioni di ore cpu/core Marconi** ma utilizzabili su tutti i sistemi aperti del Cineca. **150 TB spazio**

➔ Accordo valido per 3 anni: Abbiamo concluso l'ultima call per progetti classe A.

L'accordo finirà formalmente il 30 Aprile 2020 (fra 6 mesi!). Gli account saranno validi fino al 31 Ottobre 2020.

Si può ancora sottomettere a sportello progetti classe B/Test fino a esaurimento ore assegnabili. Rimaste solo 1.800.000 ore su 55.000.000 iniziali

➔ Prevista una unità di personale INAF @ Cineca. 

- Non eravamo ad assegnare un AdR presso il Cineca
- Cambio di strategia: abbiamo pagato una borsa al Cineca per un supporto offerto dal Cineca. (Fabio Pitari)

Infrastruttura HPC al Cineca

Sistemi Tier-0

MARCONI
➔ 1.40 GHz Intel KnightsLanding : 11 Pflops.

Oltre **3,600 nodi** a 68 cores per nodo 96 GB per nodo. Marconi è classificato in **Top500 list** tra i più potenti supercomputer: rank 12 in November 2016, rank 19 in the November 2018 list, **rank 21** Giugno 2019

GALILEO:
➔ Rinnovato a Marzo 2018. **Intel Xeon E5-2697 v4 (Broadwell)** , 1000 nodi con 36 cores per nodo, 118 GB per nodo. 60 nodi con GPU / K80

Infrastruttura HPC al Cineca

D.A.V.I.D.E

(Development of an Added Value Infrastructure Designed in Europe)

Fa parte del progetto europeo Prace Pre-Commercial Procurement (PCP) per sviluppare un sistema completo per HPC ad alta efficienza energetica. Basato su server OpenPOWER. **Aperto su richiesta**.

Architettura: OpenPower NViDIA NVLink

Nodi: 45 x (2 Power8+4Tesla P100) + 2 (service&login nodes)

Rete Infiniband 100 Gb/s, Peak Performance: 1 Pflops

Pico-Cloud (in dismissione)

Classi di applicazioni: "BigData", relative alla gestione e all'elaborazione di grandi quantità di dati, provenienti sia da simulazioni che da esperimenti.

Modelli supportati: ***Urgent Computing*** e (prevalentemente) ***Cloud Computing – OpenStack.***

Compute/login node	66	Intel Xeon E5 2670 v2 @2.5Ghz	20 core/node	128 GB	
Visualization node	2	Intel Xeon E5 2670 v2 @ 2.5Ghz	20 core/node	128 GB	2 GPU Nvidia K40
Big Mem node	2	Intel Xeon E5 2650 v2 @ 2.6 Ghz	16 core/node	512 GB	1 GPU Nvidia K20
BigInsight node	4	Intel Xeon E5 2650 v2 @ 2.6 Ghz	16 core/node	64 GB	32TB of local disk

Call 1 - Aprile 2017: Approvati 23 Progetti

Call 1. Progetti conclusi

N. 7 Progetti di Classe A ➔ 29 Mhours
N. 9 Progetti di Classe B ➔ 7 Mhours
N. 7 Progetti TEST ➔ 1 Mhours

Totale 37 Mhours assegnate e TUTTE consumate

Call 2. Progetti conclusi

N. 6 Progetti di Classe A ➔ 17 Mhours assegnate
N. 4 Progetti di Classe B ➔ 2.6 Mhours assegnate
N. 2 Progetti TEST ➔ 0.2 Mhours assegnate

Totale 19.8 Mhours assegnate TUTTE consumate

MA LE RICHIESTE SONO STATE MOLTO PIU' ALTE >> 50 MHours
DRASTICO TAGLIO! Over-subscription fattore 2 mantenuto!

Call 3 - Risorse Assegnate e tutte consumate

N. 4 Progetti di Classe A ➔ 18 Mhours assegnate
N. 3 Progetti di Classe B ➔ 1.8 Mhours assegnate
N. 2 Progetti TEST ➔ 0.2 Mhours assegnate

Totale 20.8 Mhours assegnate

Call 4 - Risorse Assegnate e in corso

N. 8 Progetti di Classe A ➔ 28.5 Mhours assegnate
N. 7 Progetti di Classe B ➔ 5.5 Mhours assegnate
N. 1 Progetti TEST ➔ 0.1 Mhours assegnate

Totale 34 Mhours assegnate

Call 5 - Risorse Assegnate e in corso

N. 8 Progetti di Classe A ➔ 38 Mhours assegnate
N. 7 Progetti di Classe B ➔ 5 Mhours assegnate
N. 0 Progetti TEST ➔ 0.0 Mhours assegnate

Totale 43 Mhours assegnate

CALL 5 FINALE

Progetto	Titolo	Ore KNL richieste	Macchina	TB richiesti su disco	Totale ore KNL approvato	DISCO
A36	High-energy particle acceleration in weak collisionless shocks	6.000.000	Marconi	1	4.000.000	1 TB
A37	Long term evolution of Pulsar Wind Nebulae in the light of CTA	3.000.000	Marconi	5	2.500.000	4 TB
A38	ROGER - Resimulations Of the Gigaparesc-Extended Radio web.	5.254.320	Marconi	2	4.000.000	2 TB
A39	Particle acceleration in dynamically evolving relativistic current sheets	7.500.000	Marconi	5	5.500.000	4 TB
A40	Numerical simulation of restarting radio galaxies	5.800.000	Marconi	1	4.000.000	1 TB
A41	Comparing dynamics and emission of AGN driven winds with collimated jets	6.000.000	Marconi	5	5.000.000	4 TB
	Cosmic Magnetic Fields with the Cosmic Microwave Background	2.500.000	Marconi	1	1.500.000	1 TB
A43	Investigating the origin of asymmetries revealed in X-ray observations of supernova remnant IC443	5.000.000	Marconi	5	4.000.000	4 TB
A46	Representative Sample of Highly-Resolved Massive Clusters	3.600.000	Galileo	2.5	3.600.000	2.5 TB
A48	DEMNUni Covariances II	6.000.000	Marconi	200	4.000.000	TBD
Totale Classe A		**50.654.320**		**225**	**38.100.000**	
A44	Stellar Planet Interaction	2.000.000	Marconi	0.4	1.000.000	0.5 TB
A45	Euclid - HMF-HB Calibration	1.200.000	Marconi	20	1.000.000	10 TB
A47	Hydrodynamics of gas in ultrafaint dwarf galaxies in MOND	1.500.000	Marconi	2,5	1.000.000	2.5 TB
B26	Resolving backflows near the AGN's accretion region.	876.000	Galileo	14	500.000	5 TB
B28	Dust in the Universe	384.000	Marconi	0,1	200.000	0.5
B29	Spectra and ages of the FR-I radio galaxy 3C 449	1.000.000	Marconi	1	1.000.000	1 TB
B30	Seeing the UNSeEN with CINECA-Galileo Cluster	881.280	Galileo	2,5	500.000	2.5 TB
Totale Classe B		**2.260.000**		**17,6**	**5.200.000**	

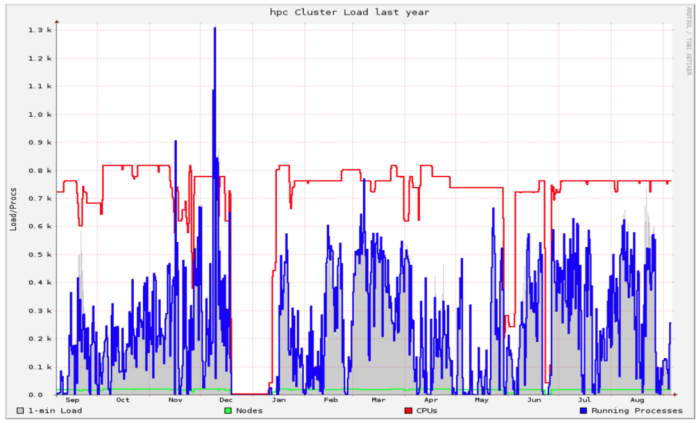
INAF - The CHIPP Project: Tier2 – Tier3 Infrastructure

Lo scopo principale del progetto CHIPP INAF è quello di fornire **risorse HTC e HPC** (per programmi di piccole / medie dimensioni) alla comunità INAF **utilizzando le infrastrutture già esistenti. Periodo 2017-2018 .. Rinnovato 2019-2020**.

Più di 50 progetti di ricerca sono stati sostenuti negli ultimi due anni.

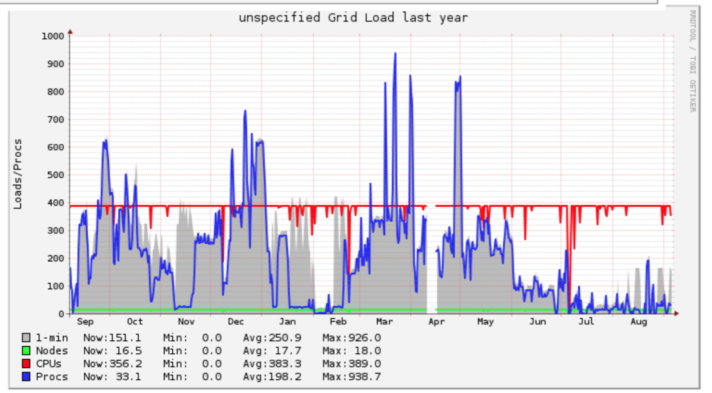
CHIPP sistema principale INAF Trieste – HOTCAT

Computing nodes: 40 Core INTEL Haswell E5-4627v3 @ 2.60GHz (4 SOCKET); 6GB RAM/Core (256GB RAM)
Numero di nodi: 20 (800 cores) RAM Totale 5.1 TB.
Storage: 250TB , 3 I/O nodi :parallel filesystem basato on BeeGFS.
Network: Infiniband ConnectX®-3 Pro Dual QSFP+ 54Gbs
Usability : 40% dedicato a CHIPP



CHIPP sistema principale INAF Catania – MUP

Computing nodes: 12 Core (24 *Hyper-Threading*) Intel® Xeon® E5-2620; 5.2GB RAM/Core (64GB RAM).
Numero di nodi: 16 (192 cores) RAM Totale: 1 TB
Storage: 60 TB parallel filesystem basato su BeeGFS pianificato.
Network: 10 Gbit network
Usability : CHIPP totalmente dedicato



CHIPP sistema sperimentale INAF Trieste - CLOUDCAT

INAF - The CHIPP Project
Tier2 – Tier3 Infrastructure

Il progetto CHIPP assegna le riosorse attraverso **bandi periodici competitivi** per l'assegnazione delle risorse, ciclicamente **ogni 6 mesi** per progetti di grandi dimensioni (~ 80.000 core/h) e un **sistema a sportello** per piccoli progetti (<10.000 core/h) di durata limitata a un mese. Tipicamente il servizio è rivolto a singoli ricercatori, piccoli gruppi o grandi progetti che necessitano di risorse informatiche al loro avvio. I progetti sostenuti da CHIPP sono scelti in base al **potenziale di innovazione, all'eccellenza scientifica e ai criteri di pertinenza**.

Dal Maggio 2017 CHIPP offre $2.5x10^6$ core/hours ogni 6 mesi per ricercatori INAF e associati

Prila call Maggio 2017:
- 25 proposal ricevute
- 16 accettate
- $2.5x10^6$ core/hours assegnate
- Project completion rate: 33%

Second call Jan 2018:
- 14 proposal ricevute
- 14 accettate
- $1.8x10^6$ core/hours assegante
- 50% completion rate

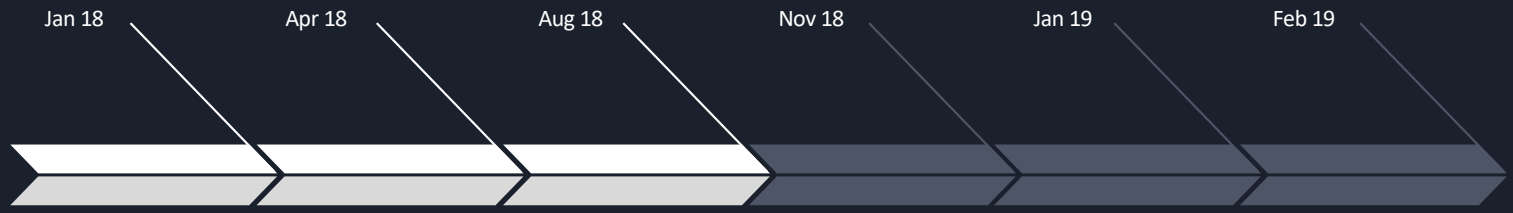
Third call Jul 2018:
- 9 proposal ricevute + 3 "rolling"
- 9 accettate
- $1.2x10^6$ cor/ hours assegnate
- 55% completion rate

Fourth call May 2019:
- 13 proposal ricevute
- 13 accettate
- $1.4x10^6$ core/hours assegnate
- On-going : 75% già partiti

Cambiamento di prospettiva nelle Call 3 e 4: Da una infrastruttura di produzione di medio-basso livello per HPC (e problemi *embarrassingly parallel*) a sistemi di sviluppo e test dedicati

Google Cloud: Proof of Concepts (PoCs)

Timeline of the Project



The idea

What about to link the investment of GSuite with Cloud Platform

Kick off

9 PoC proposed
GCP grant achieved

First 3 PoC completed!

3 out of 9 PoC have been successfully tested on the GCP.

Further 3 PoC completed!

6 out of 9 PoC completed (this time exploring also GPUs)

9 PoC implemented

The full PoCs set will be implemented on GCP

Comparison and Conclusions

Final document to summarise gained expertise and comparisons.

da Marco Landoni

INAF Pilot Project for Commercial Cloud applications

Google Cloud: Proof of Concepts (PoCs)

Valutare come una cloud commerciale risponde in diverse classi di attività computazionali

⇒ 6 use cases sono stati identificati con l'obiettivo di misurare diverse metriche su Google Cloud Platform.

Si riportano brevemente i risultati dei 6 principali POC eseguiti

1. HTC execution of embarassingly parallel code **DIAMONDS** (Corsaro). The code can be used for any application involving Bayesian parameter estimation and/or model selection problems. Platform performed correctly by executing thousand of instances of the code parallely. Instance duration 10 minutes

2. Gofio (Bignamini) (GIANO spectrograph reduction pipeline) has been deployed on the offered workflow framework, performing correctly and scaling automatically during many parallel requests of data reductions from users.

3. ***AENEAS SKA Test*** (Sciacca.) Three different Use Cases tried successfully on the platform : LOFAR prefactor calibration pipeline has deployed using real LOFAR data for 40 frequencies. We used instances with 40 vCPUs and about 256 GB of RAM. **Scalability** of pipeline has been found **very good in function of the number of cores available** on each instance. Software has been ported to the platform using ***Singularity Containers***

INAF Pilot Project for Cloud applications

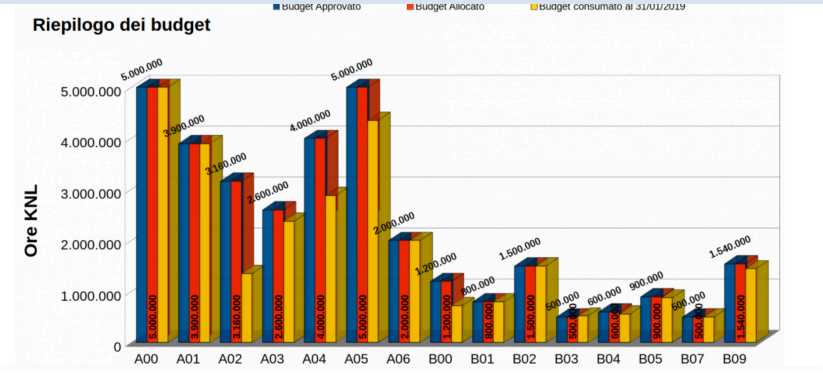
Google Cloud: Proof of Concepts (PoCs)

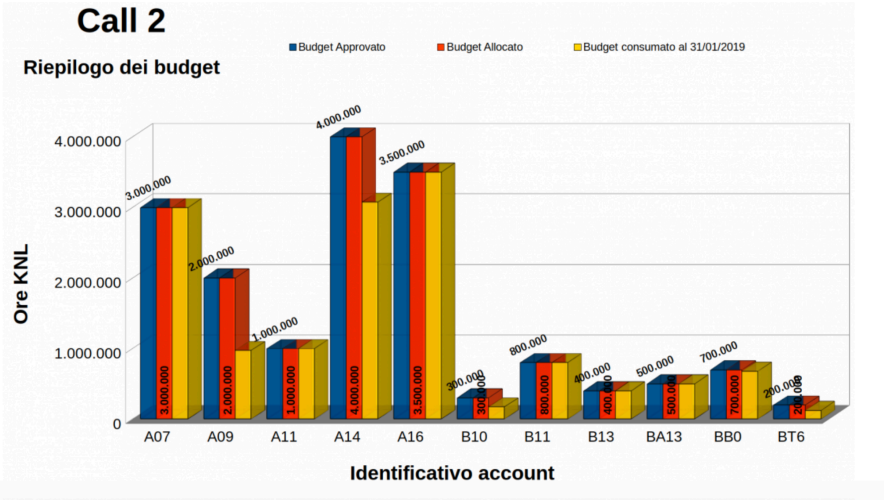
4. HPC (Taffoni): Numerical simulations of gravitationally interacting particles of both dark matter and baryonic matter. We deployed a cluster with different machine types (see figure) to run the OpenMPI based code (GADGET). **Poor results in terms of scalability**.

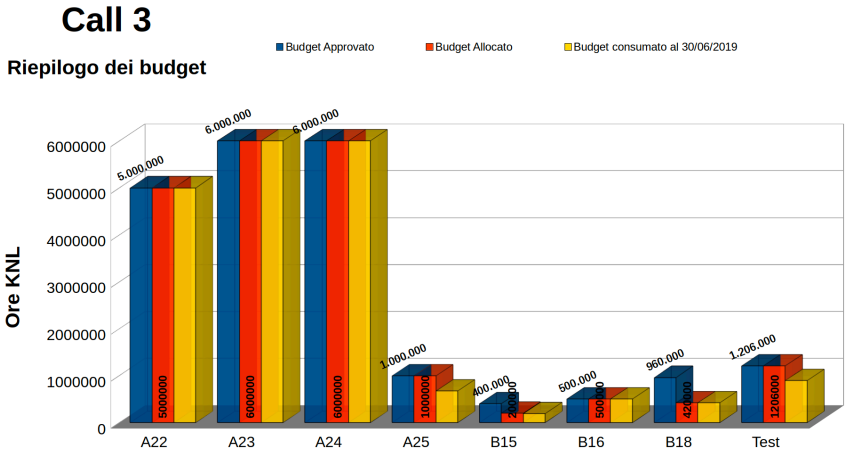
5. GPU Codes with **PASSATA** (Agapito) is a CUDA based numerical code for simulations in Adaptive Optics. Using GPU in parallel in SLI configuration to check the scalability in the context of MAORY/E-ELT project. Instances execution: 30 min - few hours. Very good scalability

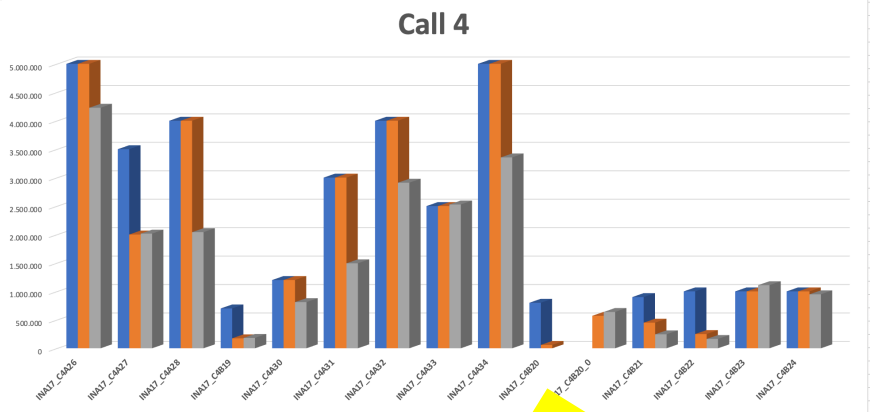
6. **ALMA** Data reduction (Massardi). CASA software for ALMA data reduction. GCP demonstrated to be successful and suitable for large dataset if compared to normal machines currently available.

MoU CINECA: Summary e Cost Analysis









- 65 Progetti A+B assegnati. 150 Unità di persone coinvolte.
- Attesi circa 180 articoli scientifici
- Spesi 290.000 Euro in 3 anni (iva inclusa)
- **Circa 0.7 K Euro / anno per unità coinvolta, o anche 1.9 KEuro per pubblicazione** [NDR: Stima dei costi, salvo miglior verifica]
- **Costo orario del calcolo: 0,0017 Euro/ora (all-inclusive)**

UN AFFARONE

CHIPP: Summary e Cost Analysis

- **52 Progetti assegnati**. Circa 150 Unità di persone coinvolte (anche in questo caso.
- **60 articoli scientifici**
- Spesi circa 360.000 Euro in 3 anni dalla DS INAF (per consumi elettrici e personale) ***Cofinanziati da OATS e OACT circa 250.000 K Euro in Hardware (presi da altri progetti di sede) per i 3 anni.***

➔ **Circa 0.8 K Euro / anno per unità coinvolta, o anche 6 KEuro per pubblicazione** [NDR: Stima dei costi, salvo miglior verifica]
➔ **Costo orario del calcolo: 0,032 Euro/h (incluso ammortamento), 0,015 Euro/h costo del solo operativo, Infrastruttura: 0,009 Hardware + consumi**

CONCLUSIONI

Non si può e non si deve paragonare il livello di richiesta e servizio offerto da CHIPP con CINECA. **CHIPP è un incubatore necessario**, CINECA è concepito per la produzione massiva.

➔ CHIPP HA DIMOSTRATO LA FORTE RICHIESTA DEI RICERCATORI DI QUESTO TIPO DI INFRASTRUTTURA. **Ha dimostrato la capacità interna di saper rispondere ai ricercatori con competenza ed altissima efficienza di gestione**

➔ **UN SERVIZIO QUINDI DA ESTENDERE E STABILIZZARE: CONVIENE INVESTIRE INQUESTA DIREZIONE**

POC GOOGLE CLOUD: Summary e Cost Analysis

- 9 Progetti sperimentati. Circa 25 Unità di persone coinvolte
- Non sono attesi articoli scientifici visto che si è trattato di un dimostratore

➔ **Costo orario del calcolo: 0,032 Euro/h (escluso personale)… quindi può anche avere un fattore moltiplicativo di 2 o 3**

CONCLUSIONI

➔ Non possono farsi paragoni tra POC su GC e CHIPP CINECA

➔ I POC hanno dimostrato che per una classe di problemi il paradigma cloud va molto bene. **Quindi una infrastruttura che sarà certamente utile con il BIG DATA**

➔ **Il costo per ora è certamente molto più alto di CHIPP e CINECA**

➔ **UN SERVIZIO QUINDI DA VALUTARE: Probabilmente sarà più conveniete investire su Cloud INAF (da realizzare): ci aspettiamo maggiore competenza ed effettivo supporto offerto ai ricercatori e costi molto più ridotti**

INAF Computing Infrastructure: Nuove Prospettive

TODAY

INAF Tier2-Tier3 Infrastructure

- Progetto CHIPP

INAF @ HPC - Cineca Partnership

- MoU
- Tier0 and Tier1 systems
- Supporto dedicato
- HPC - Cloud Computing

INAF Cloud: initial test and and future plans

- Test preliminari su Cloud Commerciali (POCs)
- EGI cloud
- EOSC

- **INAF and other dedicate infrastructures**
  - LOFAR Computing Infrastructure
  - Gaia archivi e computing
  - SRT (future)
  - … altre facilities di progetto

TOMORROW

INAF - DATA-STAR PROJECT
Coordinamento in INAF di Computing Infrastructures e Archive

FUTURO
Big Data era

EUROHPC: *Leonardo*

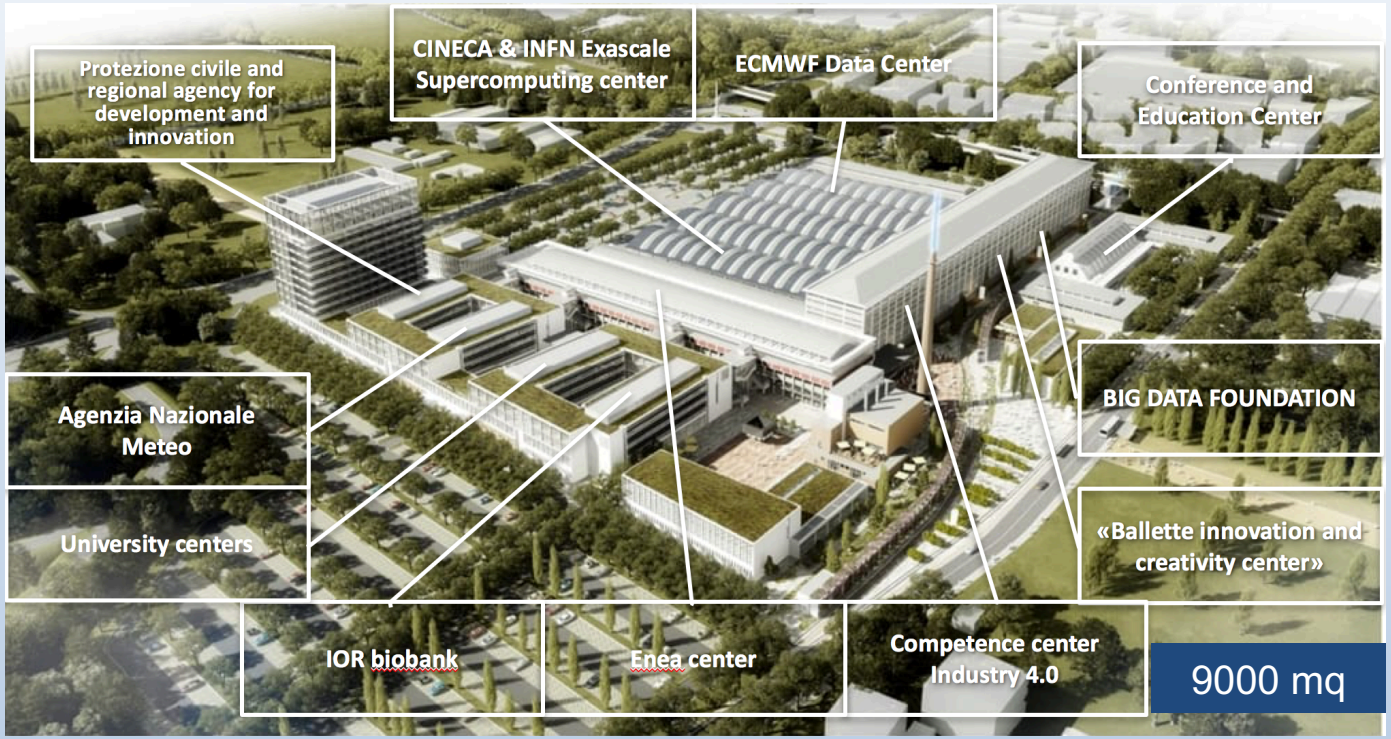
Il futuro PreExascale System in Italia

E ADESSO ? <u>Verso l'Infinito e oltre</u>

ECMWF – Bologna Science Park

La presenza di ECMWF (**centro meteo europeo)** fa diventare Bologna Science Park tra i più grandi poli di Supercalcolo e Big Data d'Europa. (ex Manifattura Tabacchi)

Cineca assume nel polo un ruolo più rilevante.



Fondamentale per INAF è essere un interlocutore sempre più attivo e principale

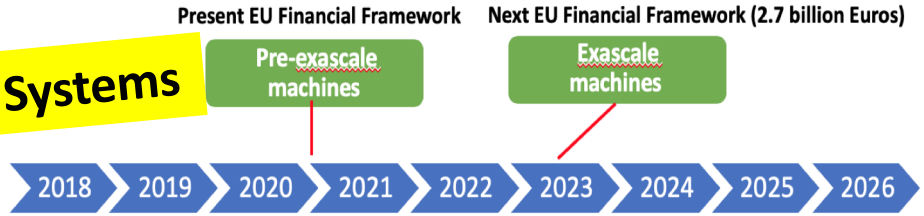
Abbiamo la consapevolezza di essere all'altezza delle sfide in gioco

Evitare di disperdere le eccellenze raggiunte delegando ad altri soggetti di rispondere alle nostre necessità attuali e future

Leonardo supercomputer: EuroHPC - JU

EuroHPC Joint Undertaking, è stata approvata dalla EC nel 2017. L'obiettivo principale è quello di consentire all'Europa di avere un ruolo leader nelle fasi HPC pre-exascale ed exascale. **La JU ha un budget di circa EUR 1.000 Milioni di Euro fino al 2022 .**

A Giugno sono stati assegnati i 3 Pre-Exascale Systems



In Italia il sistema pre-exascale è il sistema denominanto LEONARDO che sarà ospitato/gestito dal Cineca presso il Bologna Technopole ed ha un costo complessivo 240 Milioni di Euro (120 Milioni cofinanziati dal MIUR).

In Spagna la JU supporta il **MareNostrum 5**, il prossimo supercomputer del BSC.

Il supercomputer finlandese sarà ospitato dal CSC a Kajaani (Finland) e sarà gestito dal consorzio LUMI (Large Unified Modern Infrastructure)

INAF @ EUROHPC. Il supercomputer LEONARDO

Leonardo supercomputer: 270 Pflops peak performace

Notation	Description
Booster	Module of the system dedicated to capacity and capability workloads
Data-centric	Module of the system dedicated to high-memory workloads, data visualization and data management
General purpose	Module of the system dedicated to general workload, yet to be adapted to the booster module

INAF AVRA' UN RUOLO PRIMARIO NELL'IMPLEMENTAZIONE e CO-GESTIONE DEL SISTEMA

System name	Leonardo
Modules	3 (booster, general purpose, data centric)
Number of computing nodes (booster)	3500 (4 accelerators per node)
Number of computing nodes (general purpose)	1000 (> 64 physical cores per node)
Number of computing nodes (data centric)	500 (512 GB DDR and >4TB NVM per node)
Storage (scratch and work space)	Capacity: 150 PB, bandwidth: 1 TB/s
Storage (high IOPS tier and home space)	Capacity: 5 PB, bandwidth: 1 TB/s
HPL Targeted Performance (peak)	150-180 PFlops (210-250 PFlops); Top 3
HPCG Targeted Performance	2.8-3.3 PFlops; Top 3
Interconnect Bandwidth	≥ 200 Gb/s per node
Interconnect Topology	Dragonfly+ or any topology with better full bisection bandwidth
Estimated Power consumption (after PUE)	8-9 MW (8.8-9.9 MW)

L'INAF ha già pianificato le principali challenges per il nuovo sistema:

- ➔ **SKA Precursors** (ASKAP, LOFAR, MEERKAT)
- ➔ **Ground Based and Spaces Missions Observatories** (SpaceWeather, Euclid, E-ELT)
- ➔ **Simulazioni numeriche** (Black Holes e Universo Primordiale, Galassie Primordiali e Onde Gravitazionali, Struttura a Larga Scala dell'Universo)

Grazie per l'attenzione… e… adesso è il momento delle scelte, dell'organizzazione e degli investimenti