



Finanziato  
dall'Unione europea  
NextGenerationEU



Ministero  
dell'Università  
e della Ricerca



Italiadomani  
PIANO NAZIONALE  
DI RIPRESA E RESILIENZA



Centro Nazionale di Ricerca in HPC,  
Big Data and Quantum Computing



# ICSC

Centro Nazionale di Ricerca in HPC,  
Big Data and Quantum Computing

## INAF e il Centro Nazionale HPC, BigData e Quantum Computing

*Ugo Becciani*

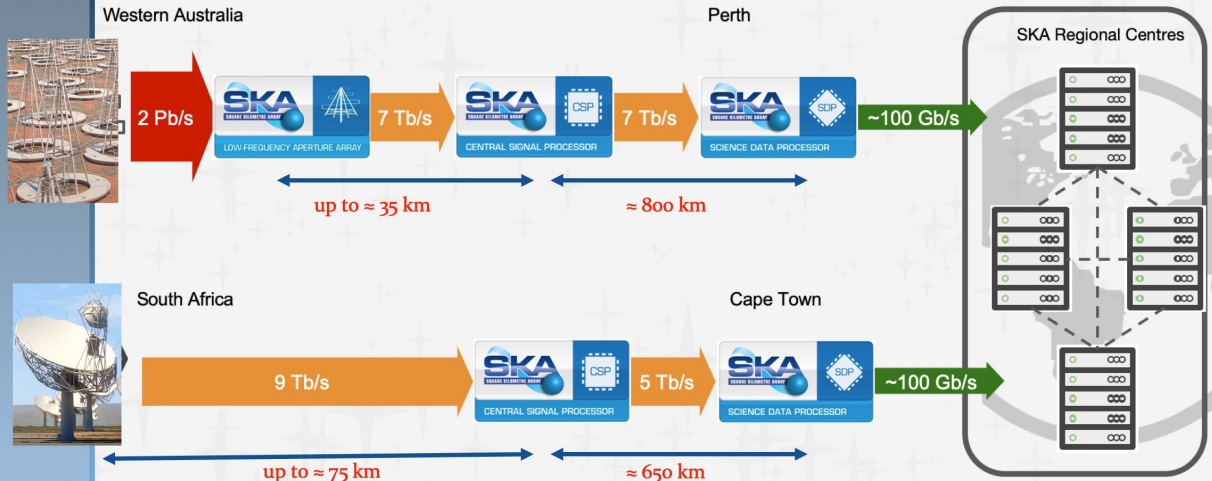
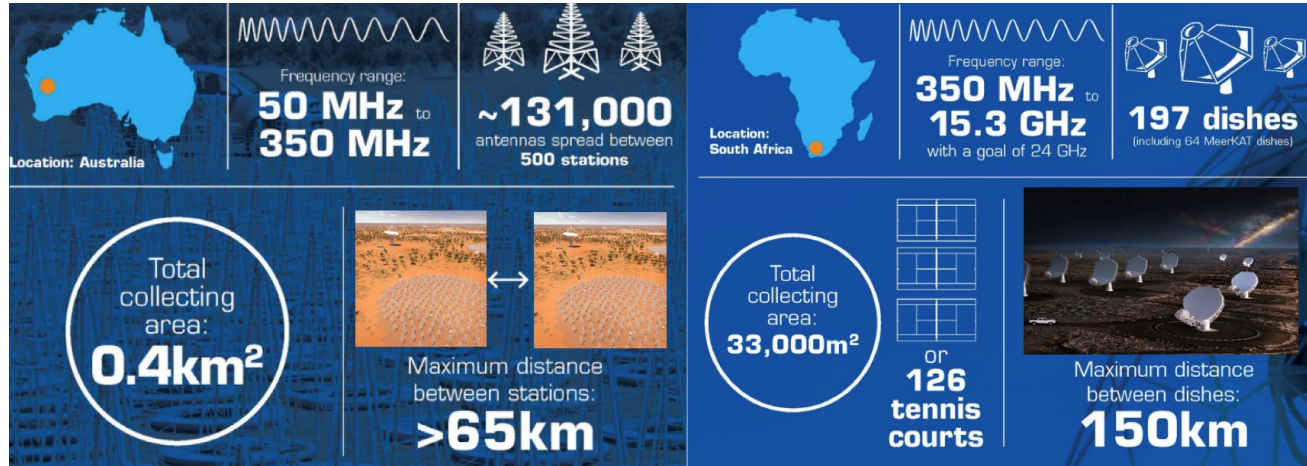
**INAF USC VIII - Calcolo Critico**

15-16 giugno 2023,

Dipartimento di Fisica e Astronomia

Catania

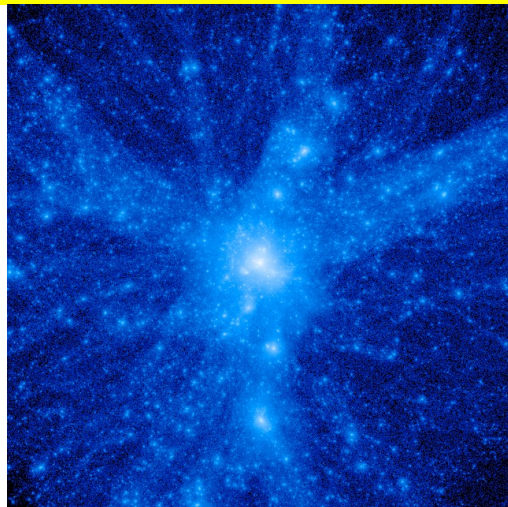
# INAF: Le sfide odierne verso l'HPC, Big Data ed Exascale



**SKA**  
*La sfida dell'immediato futuro*  
**700 PB/anno a regime.**  
*Primi dati science previsti nel 2025*

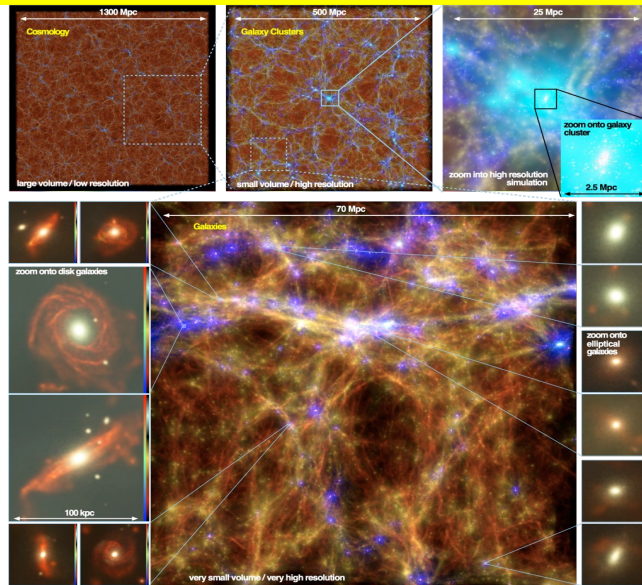
**A&C Italy**  
**SKA-RC al Tecnopolo e al CN**

# Le sfide odierne verso l'HPC, Big Data ed Exascale



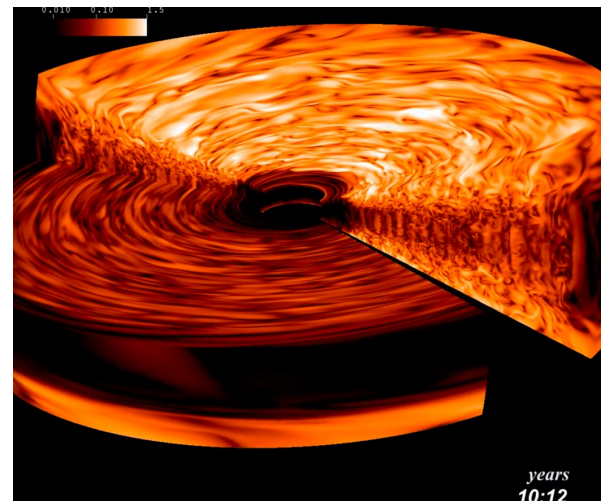
Distribuzione della Dark Matter in una simulazione high-res di un ammasso di galassie

*Eseguita con il codice OpenGadget*



Mappa di gas density da Magnetium

*Eseguita con il codice OpenGadget*



MRI-induced turbulent velocity field in a protoplanetary disk simulations.

[Credits: M. Flock (2011)]

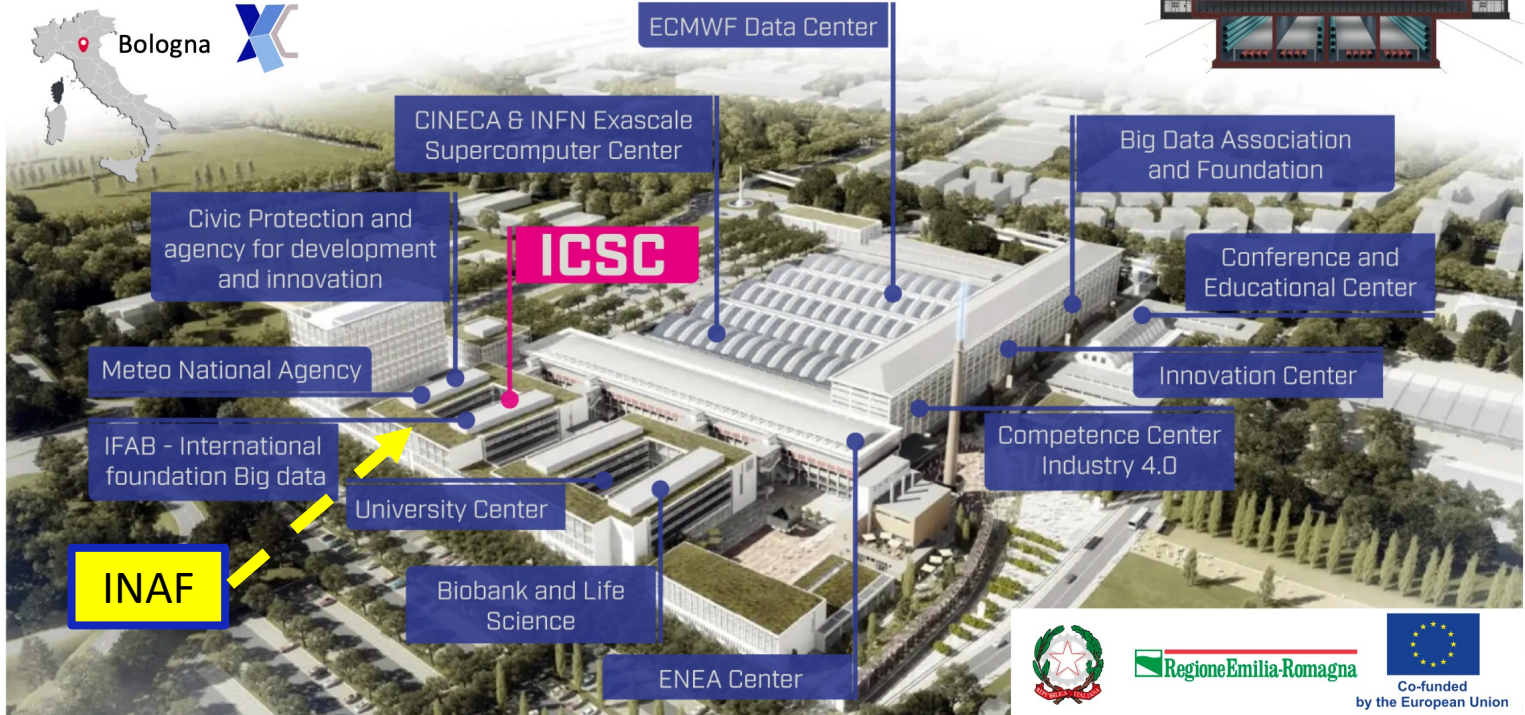
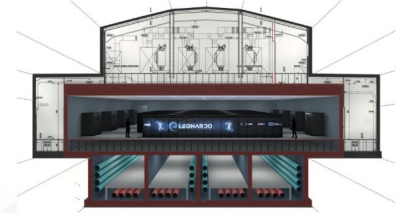
*Eseguita con il codice PLUTO*

L'interpretazione dei dati dei nuovi strumenti in preparazione PASSA anche dai modelli e simulazioni cosmologiche che hanno sempre più necessità di alta Risoluzione

**Challenge Simulation: 1.000-5.000 nodes e 0.5 PB- 5 PB**

**FONDAMENTALE l'uso di Sistemi (pre) Exascale ed anche post-Exascale → A&C Italy @ CN**

# The Big Data Technopole, Bologna



Regione Emilia-Romagna



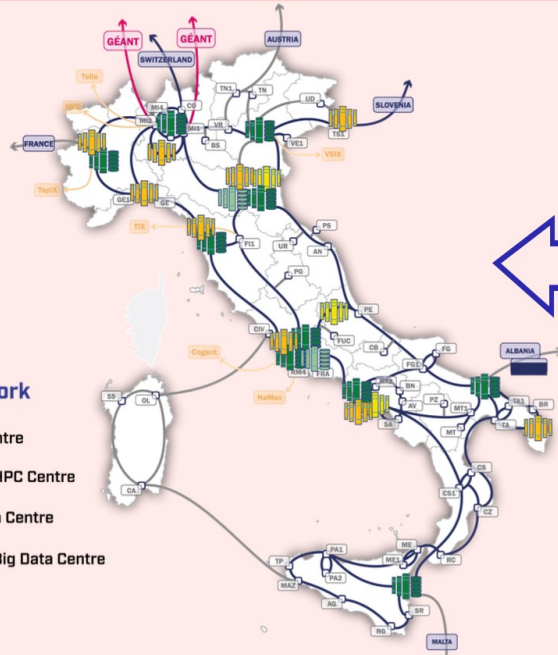
Co-funded by the European Union

# ICSC - Centro Nazionale HPC

## 10 thematic Spokes and 1 Infrastructure spoke

INAF Spoke co-Leader

### 0 SUPERCOMPUTING CLOUD INFRASTRUCTURE



#### Garr Network

- HPC Centre
- Future HPC Centre
- Big Data Centre
- Future Big Data Centre

equipped with high-level teams of experts integrating the Spokes working groups (mixed cross-sectional teams)

INAF Affiliato

INAF Spoke Leader

ER, POLICY, OUT

EDUCATION & TRAINING, ENTREPRENEURSHIP, KNOWLEDGE T

1

FUTURE HPC & BIG DATA

2

FUNDAMENTAL RESEARCH & SPACE ECONOMY

3

ASTROPHYSICS & COSMOS OBSERVATIONS

4

EARTH & CLIMATE

5

ENVIRONMENT & NATURAL DISASTERS

6

MULTISCALE MODELING & ENGINEERING APPLICATIONS

7

MATERIALS & MOLECULAR SCIENCES

8

IN-SILICO MEDICINE & OMICS DATA

9

DIGITAL SOCIETY & SMART CITIES

10

QUANTUM COMPUTING

INAF Affiliato

# Soci della Fondazione ICSC

36 Entità

## ENTI E UNIVERSITÀ



## AZIENDE E ISTITUTI PRIVATI

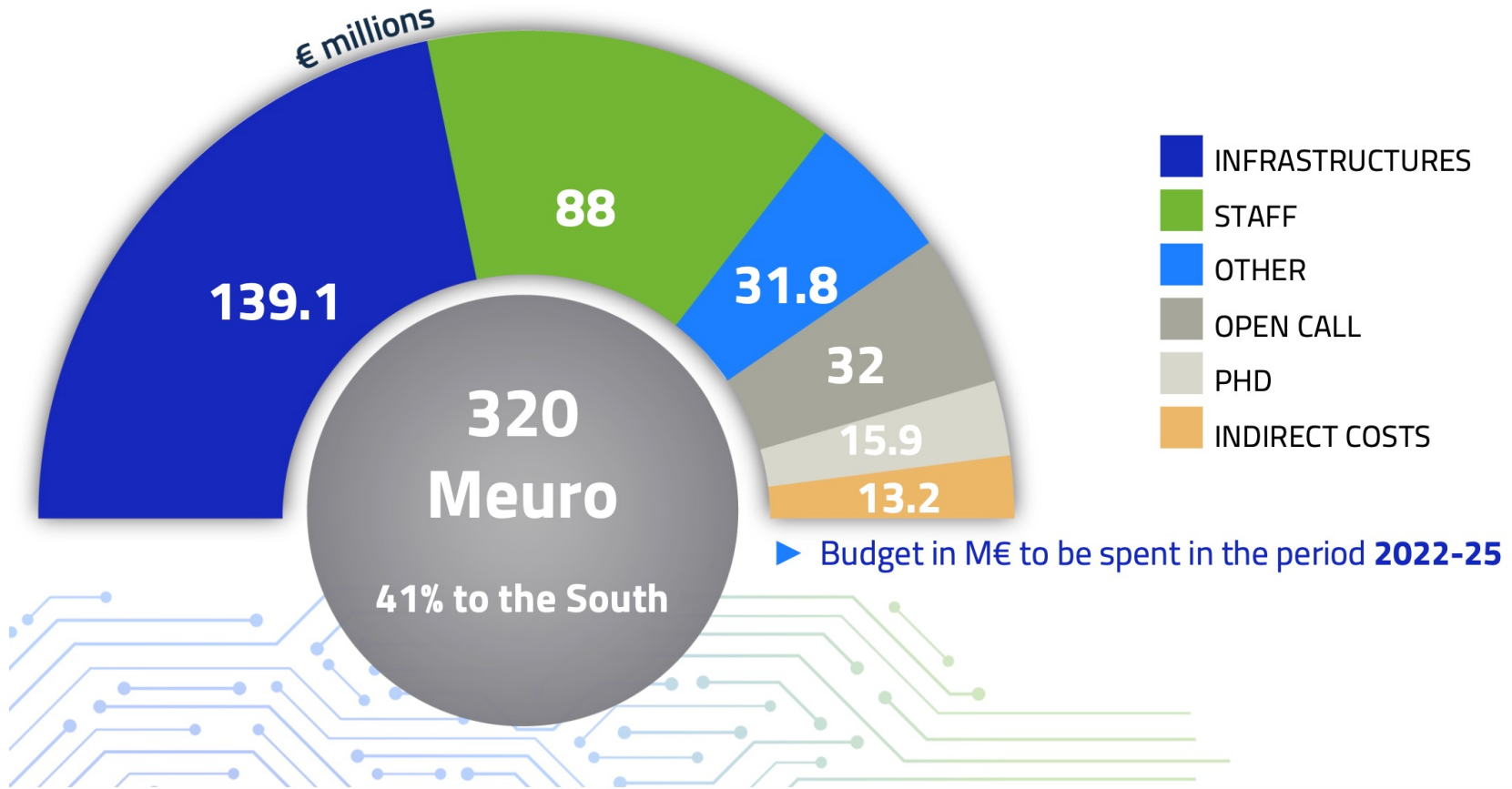


15 Entità

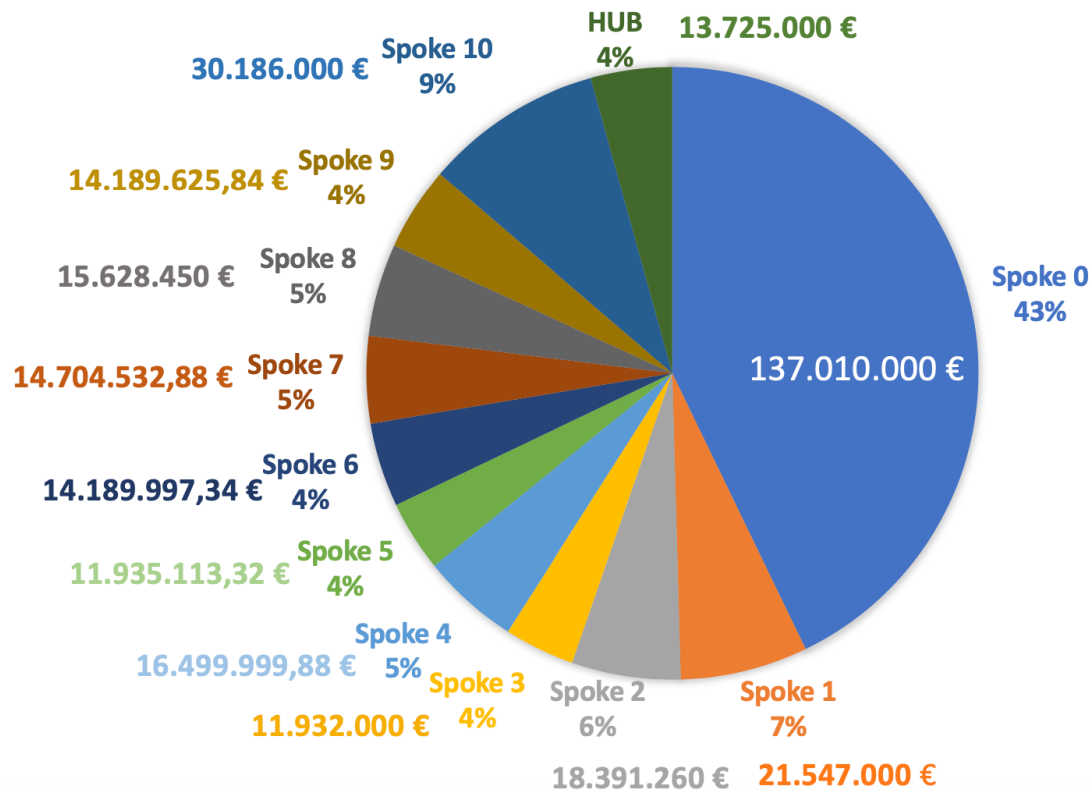
Soggetti pubblici nell'Hub/Spoke e/o Affiliati di cui 11 Spoke leader (e 11 co-leader\*)  
3 soggetti pubblici solo nell'Hub, soggetti privati (attualmente) solo nell'Hub

\* Formalmente i soggetti «co-leader» sono affiliati.

# ICSC Budget



# Distribuzione Agevolazione per Hub-Spoke



Costo totale:  
325.938.979,26 €

Agevolazione totale:  
**319.938.979,26 €**  
(di cui ~ 40.6% al Sud)



**ICSC**  
**Final budget for the  
computing infrastructure**

		TOT (M€)	% South
SPOKE 0	Leonardo Upgrade	32.50	0.00
SPOKE 0	Quantum computing	10.00	0.00
SPOKE 0	Tier1 CNAF & Tier2 INFN	25.00	10.00
SPOKE 0	Garr T	15.00	14.10
SPOKE 0	Tier1 CNR/INAF al Tecnopolo	7.00	0.00
SPOKE 0	Tier1 Napoli Cineca/CNR	13.00	13.00
SPOKE 0	Tier1 UnivAq a L'Aquila	5.00	5.00
SPOKE 0	Tier1 ESA a Frascati	4.00	0.00
SPOKE 0	storage/cpu ISO 27001	2.50	2.50
SPOKE 0	storage/cpu Cloud backbone	2.50	2.50
SPOKE 4	Tier1 CMCC a Lecce	6.40	6.40
SPOKE 1	Living Labs	3.15	0.00
SPOKE 9	Living Labs	1.35	1.35
SPOKE 10	Instrumentation	11.70	4.68
	<b>TOTAL Investment</b>	<b>139.10</b>	<b>59.53</b>
			<b>43%</b>

# INAF e l'Infrastruttura di calcolo attuale e in prospettiva a breve termine: ICSC e USC VIII

## PLEIADI e-infra

Sistema di calcolo **distribuito su tre sedi** (Bologna, Catania, Trieste). Management centralizzato e gestione a livello nazionale. **Board tecnico**: F. Vitello (coordinatore), F. Bedosti, G. Maggio.

Sono disponibili per la comunità scientifica risorse di calcolo HPC/HTC per un totale di quasi **60 Milioni di core/hour** su oltre 7000 core, con nodi da 128 o 256 GB di ram e 12 nodi GPU. **Circa 72 nodi per sito**

→ **Sistema** che soddisfa la stragrande maggioranza dell'utenza INAF:

- Il gruppo di gestione del sistema è sempre a disposizione dell'utenza (3 persone full time + 3-4 staff a disposizione)
- Progetti che usufruiscono dell'infrastruttura : 50 progetti supportati
- Paper che sono stati pubblicati attraverso l'uso dell'infrastruttura: circa 100
- Utenti sul sistema oltre 150. La richiesta massima 1.200 core (circa 36 nodi). L'utenza media INAF al Cineca è stata di 20 nodi, **50-60 per fascia alta**.

**Call #1 giugno 2022 e # 2 aprile 2023**, messe a bando **25 M core/h**. In call 1: **Richieste ~8 M core/h**, suddivise in **23 progetti approvati**. Ulteriori 5 richieste a sportello per ~ 150 k core/h

**Progetti speciali** supportati (**35 M core/h**): SKA Data Challenge 3, LOFAR, VST, EUCLID, AlmaGal..**Altri si erano proposti in una prima fase.. E poi non hanno dato seguito alle intenzioni!**

**MoU INAF-Cineca 2022-2023. Disponibilità 9.4 M core/hours su Marconi 100**

*In fase di proroga per altri 6 mesi con l'operazione Tier1@Cineca*

## INAF e l'Infrastruttura di calcolo attuale e in prospettiva a breve termine: ICSC e USC VIII

Nuova infrastruttura per INAF all'Osservatorio Astronomico di Cagliari

### Infrastruttura dedicata al progetto PON-SRT.

#### Cluster OACa Infrastruttura dedicata al progetto PON-SRT

- 14 nodi di calcolo 32 core, 512 GB ram (6 nodi con 2 GPU NVIDIA A40 da 48 GB ram) rete Infiniband
- 2 nodi di storage scratch: 1 PB spazio disco (di cui 48 TB in SSD)
- 2 nodi di storage long term: 7 PB spazio disco HDD



## INAF e l'Infrastuttura di calcolo attuale e in prospettiva a breve termine

Accordi strategici di medio/lungo termine con il Cineca per la creazione dello SKA-RC

**1. Acquisto diretto di infrastruttura INAF al Tecnopolo (accordo Dicembre 2022) di circa 2.5+1 Mln/Euro (sistema Tier -3) dal DM450.**

**Il Sistema viene acquisito su fondi Ska e deve soddisfare le necessità di storage in primis dello startup degli SRC**

### Sistema di storage+calcolo taglia TIER-3

In fase di definizione del bando di gara: Azione congiunta Cineca – INAF.

Sistema finalizzato alla messa in operatività di **storage e hardware ad esso associato da usare in ambito SKA-precursori, Path-finders e relativi ambienti di sviluppo.**

- ➔ **STORAGE LONG TERM:** Tape Library cassette LTO, ≈ **6 PB** espandibile a 100 PB
- ➔ **STORAGE di bassa Velocità** su disco: ≈ **4-5 PB**
- ➔ **STORAGE scratch ad alta velocità** su disco: ≈ **1-2 PB**
- ➔ **Nodi di calcolo.** 40 nodi FAT: 2 CPU x86 (>= 48 core) e 1TB RAM; 14 nodi Accelerati: 2 CPU x86 (>=48 core) e 512 GB RAM, 2 CPU x nodo con 80 GB memory, CUDA compatibile



## INAF e l'Infrastuttura di calcolo attuale e in prospettiva a breve termine

**2. Nuovo sistema Tier-1 del CN e porzione INAF dedicata a SKA-RC.  
Ipotizzabili 4-5 Meuro in 5 Anni (DM450). === DA DEFINIRE ===**

**Sistema di calcolo taglia TIER-2 integrato in sistema di taglia TIER-1 del CN**

**A fronte di un Investimento del CN-PNRR «per le esigenze di INAF e CNR»: 7 Mln di Euro complessivi di acquisto. Utilizzo garantito (sebbene non esclusivo) per INAF.**

**Roadmap che potrebbe impegnare INAF per 5 anni per complessi 4-5 MEuro oltre IVA. (potrebbe essere variato)**

**Vorremmo raggiungere circa 4 Pflops (Nodi Data Centric e Nodi Booster) e circa 2 PB storage alta velocità**

**INAF is developing the activities within four spokes (1, 2, 3, 10) of the CN.  
Capital Human resources, described below**

**Spoke 1 (affiliate)  
HPC & BigData**

**Main Target**

Programming models for modern HPC applications (shared-memory, message-passing, with-accelerators (e.g., GPU and FPGA); technologies and tools HPC-enabled digital twins; workflow management systems; high-performance I/O; parallel algorithms and libraries for scientific computing. Co-Design

Proponent/Supervisor	Topic	Involved Institutes	New Position	FTEs
Tornatore	Codesign of algorithms for cosmological simulations. Green Computing	OA Trieste	TD	2
Granato	Development of algorithms toward exascale	OA Trieste	1 PhD	3
Zanni	Developing the PLUTO code supporting hardware codesign and Green Computing	OA Torino, OA Brera, OA Padova, OA Catania	TD	2
Sciacca	Advanced high performance I/O and complex workflow for SKA	OA Catania, IAPS Roma, OA Trieste	1 PhD	3
Vecchiato	Highly scalable linear algebra for Gaia data	OA Torino, OA Catania	TD	2
Antonuccio	ottimizzazione di codici CFD e N-body su architetture HPC. Ottimizzazione di codici CFD e N-body su architetture HPC	OA Catania	TD	1
<b>TOTAL</b>		<b>4-5 FTE (Staff)</b>	<b>4TD, 2 PhD</b>	<b>12 FTE</b>

## Spoke 2 (co-Leader) Fundamental Physics and Space Economy

### Main Target

The activities of the spoke 2 will be developed in the context of the state-of-the-art research in basic science, and, in particular, of the domains of theoretical and experimental physics with accelerators and with space- and ground-based detectors for astroparticle physics and gravitational wave investigations

Proponent Supervisor	Topic	Involved Institutes	New Positions	FTEs
Pagliaro	Image and pattern recognition with AI. Classificazione di grandi dataset.	IASF Palermo	<b>TD</b>	<b>2</b>
Visconti	Porting e ottimizzazione pipeline analisi dati da IACT. Piattaforma di servizi di benchmarking, training e prediction per modelli di machine learning.	OA Roma	<b>TD /PhD</b>	<b>2+2</b>
Landoni	Interferometria di intensità ottica ASTRI, CTA. Ottimizzazione analisi serie temporali su BigData e simulazioni.	OA Brera, OA Padova, OA Arcetri	<b>TD, PhD</b>	<b>1+2</b>
Peron	Ottimizzazione e scalabilità di codici paralleli per la determinazione orbitale di precisione di satelliti geodetici e di navigazione, elaborazione dati accelerometrici.	IAPS Roma	<b>TD</b>	<b>1</b>
Zampieri		OAPD	<b>ADR</b>	<b>1.5</b>
Fioretti	Calcolo multi-threading e distribuito e profiling e ottimizzazione pipeline di simulazione su calcolo parallelo applicato a modelli geometrici CAD (Geant4) di strumenti su missioni X e gamma. I/O innovativo.	OAS Bologna	<b>PhD</b>	<b>3</b>
Natalucci	Integration of different reduction and analysis pipelines on heterogeneous (MWL, multimessenger) dataset in a single workflow, allowing flexible user inputs. Graphical dashboard.	IAPS Roma	<b>TD</b>	<b>1</b>
<b>TOTAL</b>		<b>6-7 FTE (Staff)</b>	<b>4-5 TD 4 PhD 1 AdR</b>	<b>15.5 FTE</b>

## Spoke 3 (Leader) AstroPhysics and Comos Observations

### Main Target

High and Extreme performance computing. Big data processing and visualization. Analysis of large and complex data volumes.

High Performance storage, Big Data management, and archiving

Proponent Supervisor	Topic	Involved Institutes	New Positions	FTEs
Calura	N-body cosmological simulations, gravitational lensing for EUCLID	OA Bologna	<b>TD</b>	<b>2</b>
Mulas	Simulations of molecular systems	OA Cagliari, OA Palermo	<b>TD</b>	<b>2</b>
Brunetti	Code development for efficient processing of LOFAR data	IRA Bologna, OA Trieste, OA Catania	<b>TD, PhD</b>	<b>2+3</b>
Bachetti	HPC enabling of astrophysical time series algorithms (from radio to gamma-ray)	OA Cagliari	<b>TD</b>	<b>2</b>
Taffoni	Development of innovative algorithms for cosmological numerical simulations	OA Trieste	<b>ADR</b>	<b>2</b>
Possenti	Development of innovative data processing/analysis algorithms for SKA-Regional Center data	OA Cagliari, OA Trieste, OA Catania, IRA Bologna	<b>TD,PhD</b>	<b>2+2.5</b>



# Human Capital: ICSC the bigger INAF investment on new generation and formation for HPC , BigData and QC

## SPOKE 3 – Cont.

Proponent Supervisor	Topic	Involved Institutes	New Positions	FTEs
Riggi	Development of source finding and classification software tools for SKA	OA Catania	PhD	3
Vitello	High performance remote visualization of big data (LOFAR, MeerKAT, SKA...)	IRA Bologna, OA Catania, OA Trieste, OA Torino	TD	1
Carbone	Development of Machine Learning data processing tools for EUCLID and SKA	IASF Milano, OA Brera	TD	2
Bonomo	Development of Machine Learning tools for planetary science (Gaia, Tess, K2, Cheops, Plato)	OA Torino, OA Brera, OA Arcetri, OA Catania	TD	2
Knapic	Data archiving solutions, data standardization and adoption of VO and FAIR standards	OA Trieste, IRA Bologna	CTER	2
Busonero	Development of innovative data archiving solutions for Gaia	OA Torino, OA Padova, OA Capodimonte, OAS Bologna, OA Abruzzo	CTER	2
Politi	Development of innovative tools for real time access to big data	IAPS Roma, OA Padova, OA Trieste, OA Torino	2 TD	2
	Technical Project Supervisor	TBD	TD	2.5
<b>TOTAL</b>		<b>11-12 FTE (Staff)</b>	<b>11 TD 2 CTER 1 AdR 3 PhD</b>	<b>34</b>

## Spoke 10 (Affiliate) Quantum Computing

### Main Target

Quantum algorithms characterized by a speedup, even limited, if compared with the corresponding classical algorithm. Quantum approach to perform calculations on state-of-the-art classical computers (quantum-inspired algorithms, emulators, tensor network computations) in order to gain significant benefits in terms of algorithms efficiency without using a quantum hardware.

Proponent Supervisor,	Topics to be <i>possibly</i> studied with Quantum Computing	Involved Institutes	New Positions	FTEs
Murante	Cosmological simulations (Poisson equation)	OAS Bologna, OA Catania, OA Roma, OA Trieste, IRA Bologna	<b>PhD</b>	<b>3</b>
Bulgarelli	Machine Learning		<b>TD</b>	<b>2</b>
Meneghetti	Lensing reconstruction			
Burigana & Trombetti	Cosmological background models and separation			
Cardone	Cosmological models (Montecarlo)			
Schillirò	Radio Astronomy data reduction (QFTE)		<b>TD</b>	<b>2</b>
Scaramella	Cosmology, radio and data analysis		<b>TD</b>	<b>2</b>
Testa	Image analysis			
<b>TOTAL</b>		<b>4-5 FTE (Staff)</b>	<b>3 TD 1PhD</b>	<b>9</b>

## Innovation Projects

INAF	Spoke/Project	New Positions	FTEs
OAC IASF-PA	Spoke 3 – <b>Intesa Sanpaolo</b> Data Quality & Fraud detection	ADR	4
OACT+IRA	Spoke 1, 2, 3 – <b>UnipolSai – Sogei – IFAB</b> Hazard Mapping and vulnerability Monitoring (HaMMon)	TD/ADR	8
OATS	Spoke 2,3 – <b>Leonardo, Thales Alenia Space.</b> Interoperable Data Lake (IDL)	TD/ADR	4
OACT	Spoke 2 – <b>ENI</b> Predictive Maintainance	TD/ADR	2
<b>TOTAL</b>	<b>2 FTE (Staff)</b>	<b>8 TD/ADR</b>	<b>18</b>

### ICSC: Global INAF Effort on HPC, BigData and Quantum Computing

→ 88 FTEs => 45-50 TD/ADR/PhD new recruitment

→ 29 FTEs Staff



## Conclusioni

- **Grosso Investimento di INAF nel coinvolgimento della nuova classe di Ricercatori/Tecnologi esperti in HPC.**
- **Occorre investire nella formazione verso le nuove infrastrutture exascale a di tipo innovativo oltre che accelerate con GPUs**
- **Molti ricercatori sono TD/PHD/ADR, come non disperdere questo patrimonio?**
- **Stiamo facendo un investimento infrastrutturale. Come questo investimento può/deve diventare un patrimonio per tutta INAF?**
- **Come dobbiamo affrontare il post “Centro Nazionale”? (Altri progetti nazionali ed europei, INAF resources, Dottorato Nazionale in HPC ecc)**
- **Abbiamo come interlocutore il Cineca. Come devono cambiare I nostri rapporti con il Cineca?**
- **Il valore aggiunto che creiamo con i sistemi “in house” di INAF sono un volano per la formazione/gestione ed creazione di competenze. E’ importante continuare questa esperienza. Quale strada intraprendere?**