Finanziato dall'Unione europea
NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

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

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

Spoke 3 - Astrophysics and Cosmos Observations

INAF – Spoke Leader Ugo Becciani

Bologna, Kick-off Meeting, 25/26 Novembre 2022

Spoke Leader INAF
Co-Leader INFN

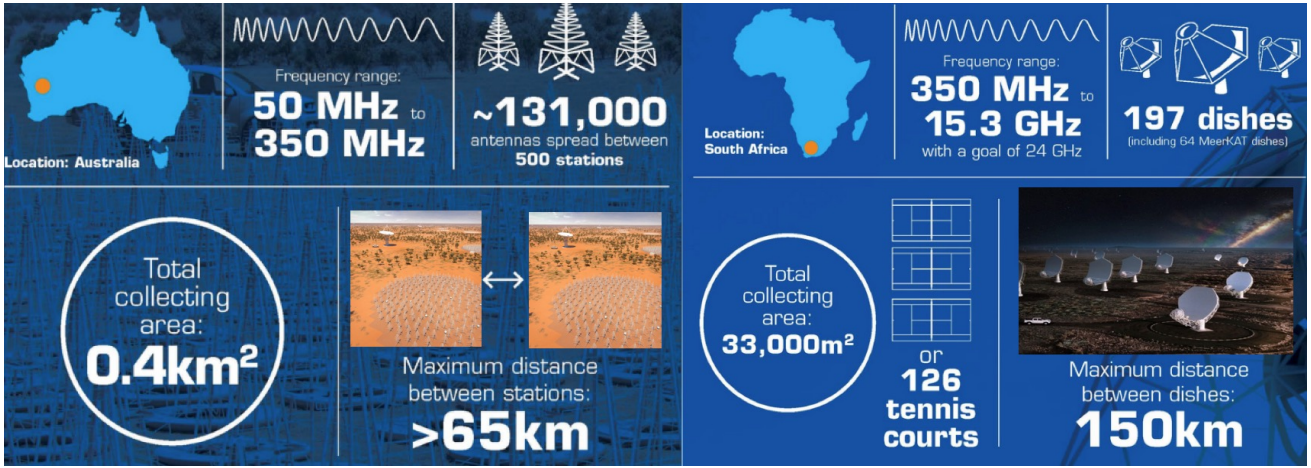
Soggetti Pubblici Affiliati	
Università Roma Tor Vergata	
Università di Trieste	
Università di Torino	
Università di Catania	
Scuola Normale Superiore - Pisa	
Sissa - Trieste	

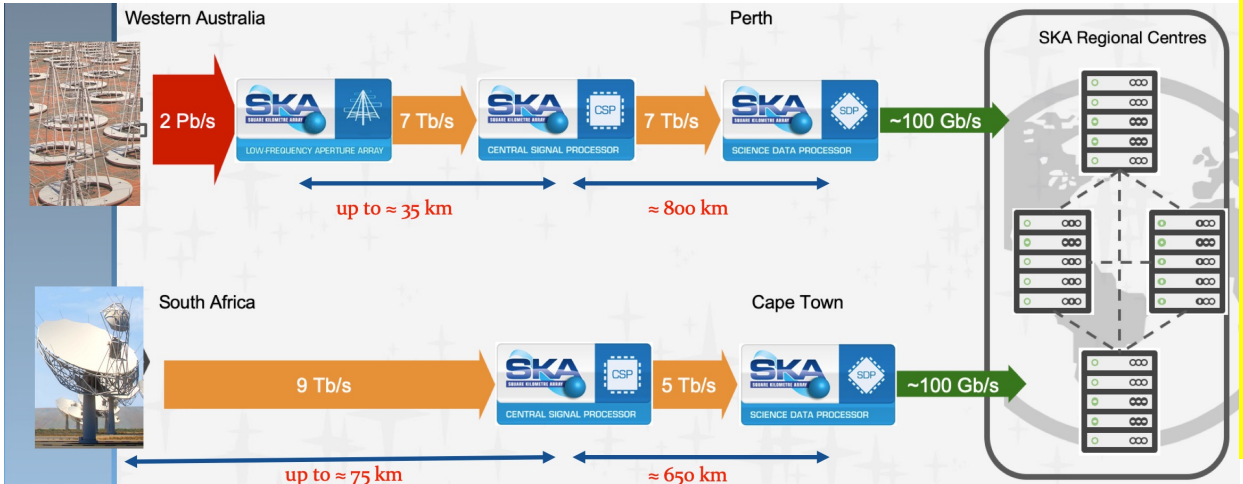
Imprese Aderenti allo Spoke	
Intesa SanPaolo	
UnipolSai	
Sogei	
IFAB	

Dalla Ricerca all'Impresa *(Persistenza)*

LEONARDO

Spoke 3 – A&C le sfide odierme verso l'HPC, Big Data ed Exascale





SKA

La sfida dell'immediato futuro

700 PB/anno

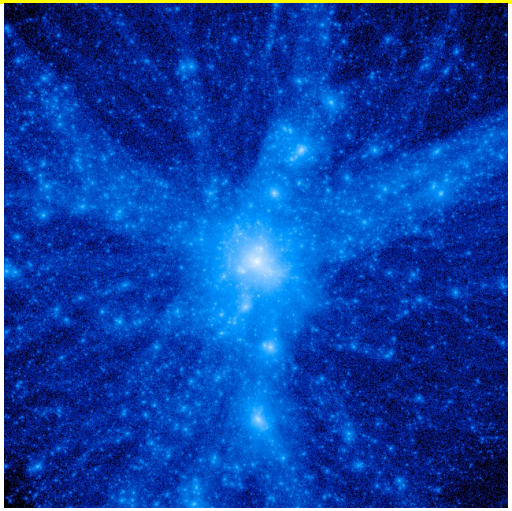
nei primi anni di operatività.

Primi dati science previsti nel 2025

A&C Italy

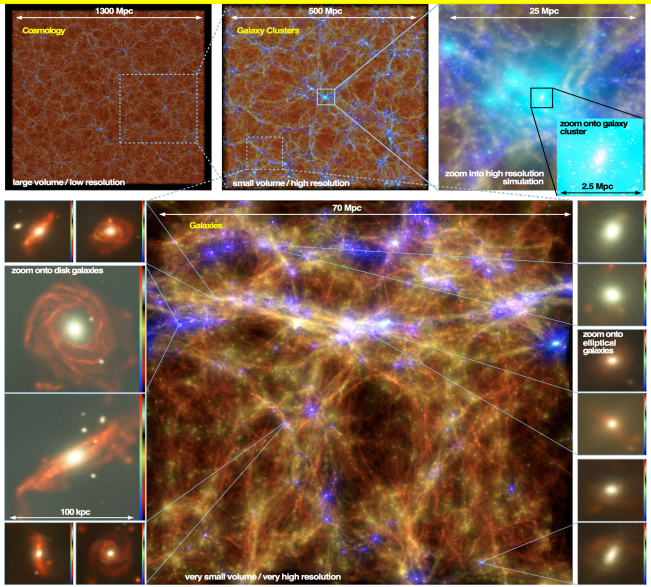
SKA-RC al Tecnopolo e al CN

Spoke 3 – A&C le sfide odierne verso l'HPC, Big Data ed Exascale



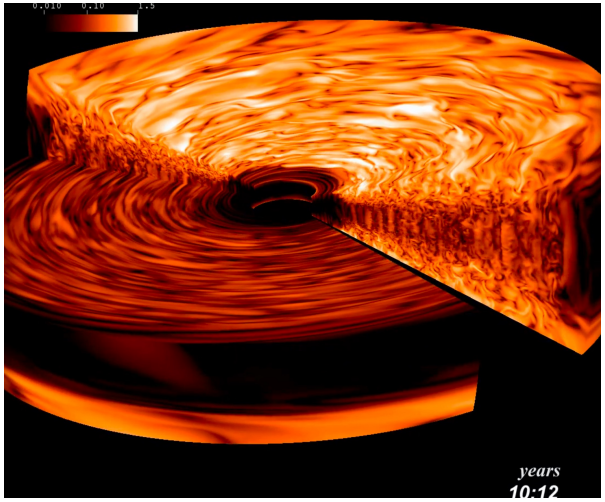
Distribuzione della Dark Matter in una simulazione high-res di un ammasso di galassie

Eseguita con il codice OpenGadget



Mappa di gas density da Magneticum

Eseguita con il codice OpenGadget



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

[Credits: M. Flock (2011)]

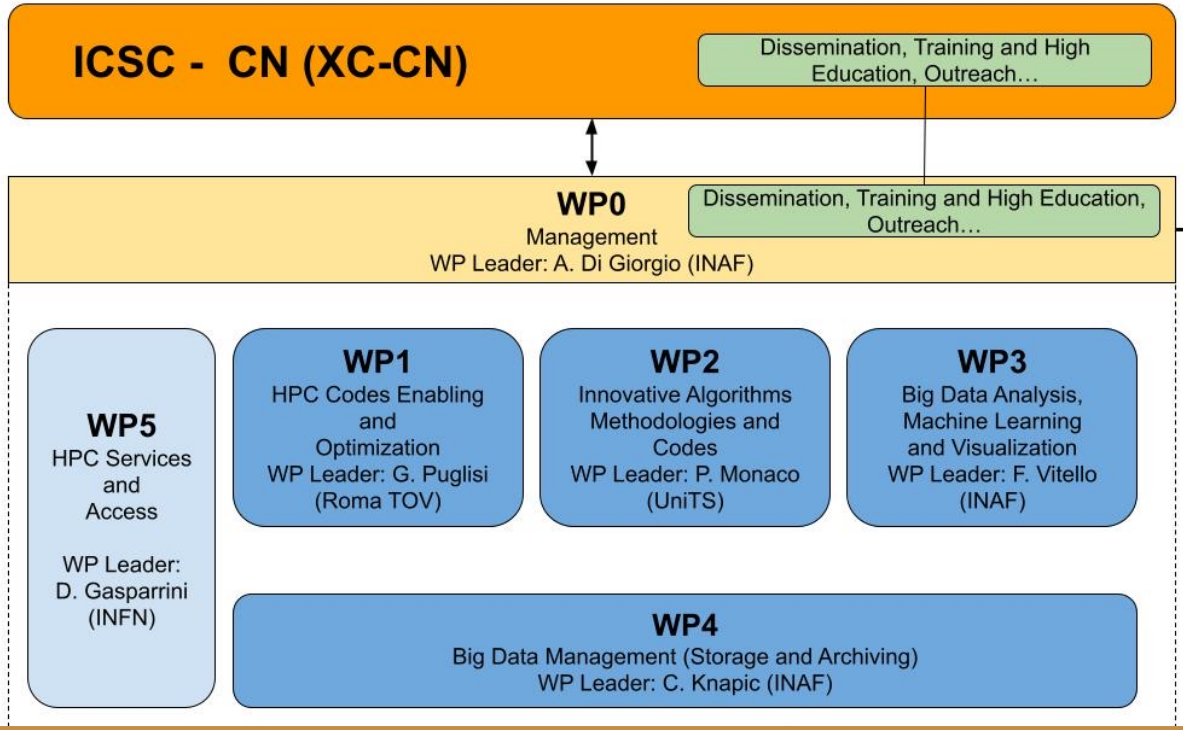
Eseguita con il codice PLUTO

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

Spoke 3 – Struttura dei WP



Project Coordination Board

Spoke Leader: **U. Becciani**
co-Leader: **P. Lubrano**
System Engineer: **G. Taffoni**
Technical Secretariat Res.: **C. Gheller**

Spoke 3: Principali obiettivi del programma scientifico

WP1 - HPC Codes Enabling and Optimization

Porting e ottimizzazione di codici già in produzione (HPC, GPUs, ARM ecc).

- **Selezione, Analisi e Test di codici, algoritmi e programming models**
  - Esempi: PLUTO, OpenGADGET, RAMSES, AREPO, XMAP, TOAST, Pinocchio, COMPLETE and 21cmFAST

- **Sviluppo Software, Refactoring e Ottimizzazione**
  - Analisi delle Sezioni critiche e "computational intensive" che ***saranno riscritte***. Si creerà un team di sviluppo composto da Scientist del dominio ed Esperti in HPC per raggiungere l'obiettivo.

- **Integration, Verification and Validation.**
  - ***Integrazione*** delle nuove parti nei codici in uso della comunità e ***re-ingegnerizzazione*** dei codi per la fase di ***produzione***

Spoke 3: Principali obiettivi del programma scientifico

WP2 - Design of Innovative Algorithms, Methodologies, Codes toward Exascale and beyond

Innovazione verso l'Exascale

- **Progettazione di soluzioni per nuove architetture**
  - Sfruttamento delle infrastrutture exascale e post-exascale (GPUs, FPGAs, acceleratori vettoriali, NVM, HBM, ARM)

- **Sviluppo o trasferimento di sezioni computazionalmente intensive di codici del dominio astrofisico**
  - Uso di ***piattaforme*** con acceleratori di ***nuova generazione*** (nuove GPUs, FPGAs e accelleratori vettoriali) – azione in possibile relazione con Spoke 1

- **Utilizzo di metodologie basate sul co-design**
  - Applicazioni a codici di dominio astrofisico, azione congiunta tra esperti HPC in Spoke 3 e aziende che sviluppano hardware innovativo - azione in possibile relazione con Spoke1

- **Efficienza energetica e sostenibilita' del calcolo ad alte prestazioni nei regimi exascale e post-exascale (green computing)**

Spoke 3: Principali obiettivi del programma scientifico

WP3 - Big Data Analysis, Machine Learning and Visualization

Soluzioni avanzate di Intelligenza Artificiale e Visualizzazione per il Big Data

- **Requirements per la comunità A&C e metodi innovativi di ML**

- Valutazione dei requisiti principali delle simulazioni cosmologiche e dei dati osservativi, considerando in particolare i casi d'uso definiti dallo ***SKA-RC (Regional Center)*** e dei suoi precursori e path-finder ➔ ***Big Data*** con complessità rilevante, che richiedono lo studio e l'implementazione di prototipi con soluzioni innovative.

- Processi per le trasformazioni, validazione e integrazione al fine di ***migliorare i data science*** e studio di adeguati ***modelli di ML*** con implementazioni efficienti e scalabili

- **HPC/Cloud Visualization Services**

- ***Visualizzazione multi-dimensionale e interattiva***, utilizzando servizi e fase di rendering nei sistemi HPC e/o server collegati ***close to the data***, ma con fruizione in tempo reale sul client utente

Spoke 3: Principali obiettivi del programma scientifico

WP4 - Big Data Management, Storage and Archiving

Framework innovativi per il Big Data Challenge (FAIR e Open Science)

- **Big Data Management, Storage and Archiving**

- ***Sviluppo di interfacce efficienti*** per la ***semplificazione della ricerca di dati da processare*** e l'immediato salvataggio dei prodotti dell'elaborazione secondo i principi del ***VO*** (Virtual Observatory) e del ***FAIR*** (Findable Accessible Interoperable Reusable) in condivisione tra gruppi di utenti
- Analisi delle prestazioni e studio di ***soluzioni innovative per l'I/O*** nella prospettiva del Big Data (ad es. SKA)

Spoke 3: Principali obiettivi del programma scientifico

WP5 - HPC Services and Access

Gestione e implementazione di un ambiente integrato che fornisca i tools di support per lo sviluppo efficiente del lavoro descritto nei WP.

- **Sviluppo di una piattaforma per *il collaborative software* ed il *continuous integration***
- Implementazione di soulzioni per il ***Continuous Integration/Continuous Delivery*** ➔ speeds up dei processi del software delivery e repetition.

- **Progettazione, implementazione e validazione di una architettura per l'interoperabilità**
- ***Sviluppo di data services*** (tools, web portals ecc) ***che permettano l'interoperabilità dei dataset*** secondo quanto definito dal FAIR (Findable Accessible Interoperable Reusable) e dagli standard del VO (Virtual Observatory)

- **ML e Visualizzazione Scientifica: deployment di servizi e HPC/Cloud integration**
- Deployment dei servizi sviluppati nel WP3 orientate all'utente finale e sviluppo di ***soluzioni container-based.***

Spoke 3: WP0 - Management

Coordinazione e management dello Spoke inclusi aspetti tecnici generali, amministrativi, legali e fianziari

Task 0.1: Overall coordination, governance, monitoring, communication flow and methods.

Riceve e controlla la documentazione contabile di tutti gli enti dello Spoke e dialoga con l'Hub; ***Organizza le attività day-by-day*** e supervisiona l'avanzamento delle milestone e la preparazione dei risultati finali; Organizza gli incontri di coordinamento dello Spoke.

Task 0.2 Data Management Plan production

Redige il piano di gestione dei dati (DMP) seguendo le linee guida fornite dall'HUB e tenendo in considerazione quanto segue: "Guidelines on Data Management in H2020", "Horizon Europe Data Management Plan Template", e "EU General Data Protection Regulation (GDPR)""

Task 0.3: Training, dissemination plan and outreach activities

Training, dissemination and outreach activities sono coordinati a livello centrale dall'HUB
Spoke 3 prepara e collabora nello sviluppo di attività di interesse per lo spoke, ad es:

- **HPC schools**: - Entry level e advanced school (verso l'Exascale)
- **HPC Code Enabling Hackathon**: Favorire l'attività di porting verso sistemi avanzati di HPC (parallelizzazione, GPU enabling etc.). Workshop (full immersive) con ricercatori e sviluppatori di codici ed esperti HW/SW per la progettazione e implementazione dei codici paralleli e verso l'Exascale.
- **Master di secondo livello** in Space Science and Technology

Task 0.4: Open Calls, Innovative Grants

Coordina tutte le azioni relative all'impiego dei fondi delle Open Call e Innovation Funds

Spoke 3: A&C Long Term Sustainability

Spoke 3: Ricadute fondamentali verso progetti di frontiera, per gli sviluppi scientifici attesi dalla comunità A&C

Tema Scientifico connesso allo Spoke 3	Progetti Principali
Radioastronomia	Square Kilometer Array (**SKA**), Low Frequency Array (LOFAR2.0) e **Meerkat+**
Observational Astrophysics e time-domain	Missioni Spaziali: **Euclid** e **Gaia** Legacy Survey of Space and Time (Osservatorio **Vera Rubin**) Extremely Large Telescope (**ELT**)
High-Energy	CTA, ASTRI, FERMI, SWGO, e altre infrastrutture
Large Scale Simulation	**HPC Theory** (P-GADGET3 -> OpenGADGET, PLUTO, plasma physics simulations, ecc..)

A&C al CN: Long Term Sustainability

1. **Accordi MoU A&C-Cineca** Utilizzo di sistemi Tier 0 (Marconi 100 e Galileo 100)

2. **E-Infra PLEIADI – INAF** Il sistema di calcolo INAF denominato PLEIADI offre alla comunità scientifica la possibilità di utilizzare risorse di calcolo HPC/HTC per un totale di quasi ***50 Milioni di core/hour su oltre 7000 core***

3. **Tier 3 – INAF** - ***In fase di redazione capitolato (2.5 Mln/Euro) .*** Costruzione prototipale del nucleo Italiano dello SKA-RC. Basato su ***nodi di calcolo con GPU***, dischi per l'operatività e lo ***storage*** ad alte prestazioni e ***Tape Library*** da 5-10 PB espandibile e in prospettiva anche a 500 PB.

4. **Tier 1 SKA-RC** (CN – Spoke 0). Ci aspettiamo la ***connessione*** e ***l'espansione*** del sistema INAF Tier3. ***Un sistema che ponga la comunità Italiana in posizione leader nell'European SKA-RC***

SOSTENIBILITA' FUTURA PER I PROSSIMI ANNI

INAF → DM-450 fondo pluriennale per SKA e CTA

Casi d'uso aziendali
Banca Intesa SanPaolo

Corporate credit risk

Problem definition:

Predict the *status* of a «**performing**» client after 6 months from the reference time of observation (i.e. predict if the client will be able to repay his loans/debts).

Challanges:

- **Default rate** of «performing» clients is low ranging from **1 to 3%**.
- **Feature distribution can vary significantly** in test set if compared to training set due to unexpected major events (e.g. covid crisis, bank merge and acquisitions, …)
- The nature of clients in **perimeter is heterogeneous** (natural persons, sole proprietorships, shared accounts, ..)
- The output of the model must be **explainable**

Current approach: classification model (XGBoost), binary target, tabular data

Possible alternative approaches:

- Use a generative model to assess **covariate shift**
- **Clustering** on clients (graph on clients or in a latent space) and subsample on cluster
- **Explainable** Time series

FDoT – sampling strategy

Fraud Detection on Transactions there is a huge amount of transactions that are no fraud. It is required an undersample of this set to exploit the eterogeneity of the non-fraud transactions. The data are tabular, the target is binary

Solution now: plain subsampling

Possible solutions: Clustering on clients (graph on clients or in a latent space) and subsample on cluster

Churn analysis

Problem definition: The **Private Bankers** (PB) are the professionals that **manage** the **portfoglio** of several **clients**. Rarely happen that the **PB quits** to take an opportinuty in another bank. The **aim** is to **indetify** in the operations of clients (time series) the pattern to anticipate **this behaviour**.

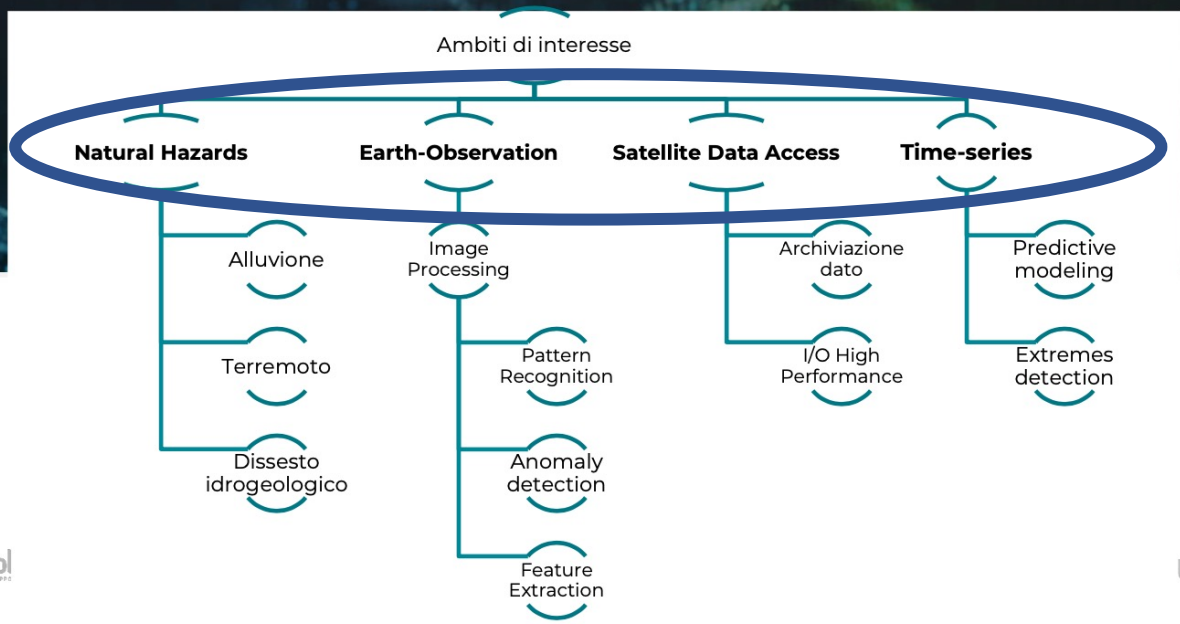
Solution now: classifier on tabular data based on cumulated montly operation per each client with binary target (1 if the PB of that client quit)

Possible solution: Exploit the data directly in time series. Use method for rare events: Anomaly detection? E.g. CVAE

Casi d'uso aziendali UnipolSai

Spoke 3: Astrophysics and Cosmos Observations

Unipolsai - Leithà



Unipol

Leithà

Casi d'uso aziendali SOGEI

USE CASE 1 ➔ Sversamenti in mare

Azione primaria : porting di algoritmi esistenti (seriali e molto lenti) su piattaforme HPC
Scopo: individuazione delle truffe
➔ Obiettivo 1 riduzione dei tempi per il ***monitoraggio e allarme intervento***
➔ Obiettivo 2 maggiore risoluzione nella ***individuazione dell'inquinamento e conseguenti azioni***

USE CASE 2 ➔ URGENT DECISION MAKING

Azione primaria : controllo e riduzione degli accessi illegali ai sistemi (cybernetic attacs)
Scopo: raffinare il sistema in ***tempi rapidi per bloccare gli attacchi***
➔ Obiettivo Raffinazione dei modelli in uso per i controlli: avvenuto un attacco con successo, riallocare con urgenza risorse di calcolo per riaddestrare il software di controllo anti-attacco per prevenire e bloccare altri attacchi simili

USE CASE 3 ➔ PATTERN RECOGNITION

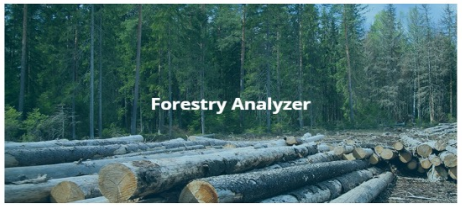
Azione primaria : controllo del territorio e zone censite
Scopo: determinazione con precisione delle ***zone censite catastali*** e ***monitoraggio dei cambiamenti***
➔ Obiettivo: Maggiore aderenza del catasto al territorio reale e delle truffe catastali

Casi d'uso aziendali IFAB

E3CIndex: Indice sintetico che misura gli eventi climatici estremi

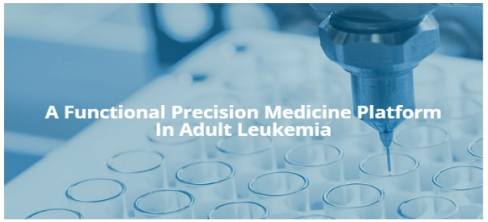
Casi d'uso reali per applicazioni Quantum Computing

Applicazioni di machine learning per ottimizzazione codici e algoritmi



Strumento per misurare I trend della deforestazione

Piattaforma che utilizza AI e dati del traffico in zone industriali, per minimizzare inquinamento e migliorare la sicurezza



Applicazione di Big data per testare ex ante l'efficacia di specifiche terapie contro la Leucemia

iFAB

Spoke 3 – Relazioni con Altri Spoke

Spoke 0 ➔ Infrastruttura disponibile e servizi

Spoke 1 ➔ co-Design di infrastrutture, green computing e piattaforme innovative, POC imprese

Spoke 2 ➔ U. Becciani e D. Spiga nei Technical Board corrispettivi. Moltissime aeree comuni sulla ricerca di base ed applicazioni innovative

Spoke 5 ➔ Use Case con aziende della Fondazione

Spoke 10 ➔ Quantum advantage per simulazioni astrofisiche

Budget Complessivo: 11.932 Meuro

- Personale 6.5 MEuro. ➔ 55%
- Open Call 3.2 MEuro ➔ 27%
- Innovators 1,0 MEuro ➔ 8%
- Formazione 1.2 Meuro ➔ 10%

Investimento su Personale e Formazione

- Oltre **100 Annualità** di Nuovo Reclutamento:
  - *Circa 60 FTE tecnologo*
  - *Circa 40 annualità di PhD*
- Circa **40 Annualità** di personale Staff

Spoke 3 – Budget Overview

I Passi già fatti … verso il Kickoff di Spoke 3

➔ **Pre-Kickoff 17 Giugno 2022** presso Roma TOV sede cittadina

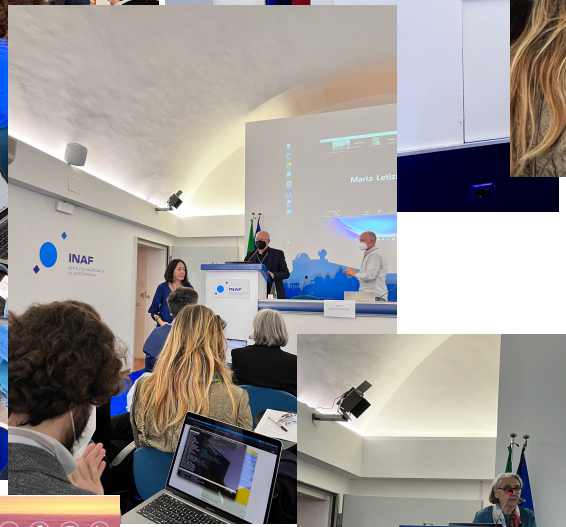
➔ **Visite alle singole aziende** e pre-discussione dei casi di interesse comune Luglio 2022

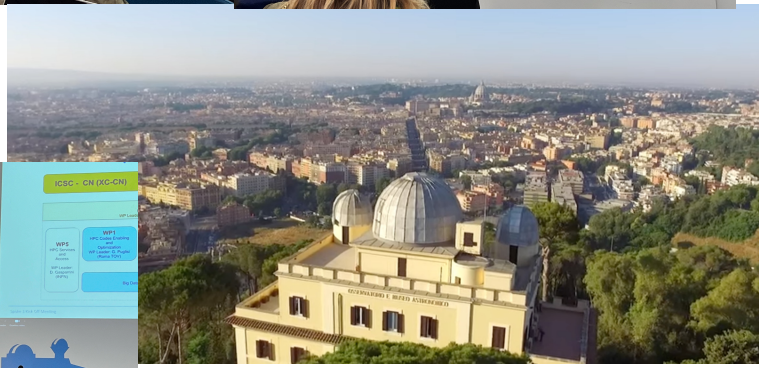
➔ 15 Settembre a Roma INAF Sede Centrale **Riunione dei WP Leaders** con ***inizio delle attività*** per il detailed draft plan dei WPS

➔ 29 Settembre a Bologna (sede UnipolSai) **meeting dedicato alle Aziende**

➔Dal 30 Settembre ***regular meeting in telco e in presenza, con i WP Leaders*** e i partecipanti ai WPs

Spoke 3 – 27-28 Ottobre Kickoff – INAF Headquarters





https://www.youtube.com/watch?app=desktop&v=_HwWgK479RI