



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

Spoke 2 WP3 - Design and development of science-driven tools and innovative algorithms for Experimental Astroparticle Physics and Gravitational Waves

*Paolo Natoli (Uni Ferrara)
Marco Landoni (INAF)*



Agenda di oggi

- ❑ Introduzione (10')
- ❑ Risultati ricognizione use cases e Q&A (50')
- ❑ Discussione su possibili aggregazione e Q&A (20')
- ❑ Planning prossime riunioni (10')
- ❑ Piattaforma di knowledge base / ticketing (10')
- ❑ Discussione (20')



Introduzione 1/2

- ❑ La telecon di oggi ha lo scopo principale di cominciare un percorso che dovrebbe portare in ~ 2 mesi a convergere su un set di use cases per il WP3
- ❑ Gli use cases saranno formalizzati a livello di Spoke, quindi in accordo con gli Spoke leader che parteciperanno al processo di definizione
- ❑ Saranno questi use cases a ricevere “risorse” da parte del centro HPC.
- ❑ Il processo di allocazione delle risorse non è ancora definito, in attesa che Hub & Spoke validino le linee guida.
- ❑ Avverrà attraverso vari sistemi (Leonardo, Marconi, altre risorse Cineca, ...)



Introduzione 2/2

- ❑ Nel frattempo, è stata chiesta l'attivazione di una sandbox iniziale per scaldare i motori e fare team building (tempistiche TBD)
- ❑ La logica dello Spoke 2 è che la definizione delle risorse di computing necessarie avvenga anche attraverso il sostegno dei WP tecnologici 4 e 5.
- ❑ L'interazione con i WP tecnologici sarà progressiva e andrà a regime dopo la definizione degli use cases (che delineano le risorse)
- ❑ E' utile partecipare alle operazioni preliminari (e.g. la survey delle competenze)



Ricognizione degli use cases

- Grazie al form Google circolato abbiamo raccolto (ad ora) **23** use cases.
- Con le splinter sessions abbiamo effettuato fino ad ora circa una decina di teleconf individuali, necessarie a darci un quadro più chiaro della situazione e del team



Ricognizione use case - statistiche

Numero di use cases ricevuti (ad ora): 23

- Univ. Salento: 4
- UniCal: 3
- INFN: 3
- INAF: 4
- Federico II: 2
- UniMiB: 1
- Univ. Firenze: 2
- Univ. Ferrara: 2
- Sapienza: 1
- Univ. Trieste: 1



Risultati ricognizione (1 / 5)

| Titolo Progetto | Ente | Persona di Riferimento | Tematica Scientifica | Keywords tecnologiche |
|-----------------|-------------|------------------------|--|---|
| JUNO | INFN | Giuseppe Adronico | Caratterizzazione dello strumento con tecniche di calcolo avanzate | GPU + CUDA |
| Virgo/LIGO | INFN | Pia Astone | Ottimizzazione di algoritmi per detection segnali GW (continuous) | GPU + CUDA |
| LISA | UniMiB | Riccardo Buscicchio | Analisi avanzata di dati LISA (Gibbs sampling) | GPU I/O massivo |
| GRB-GW | Federico II | Fabio Garufi | Correlazione segnali GRB-GW | Neural Networks for pattern recognition |
| GAPS | Uni Firenze | Massimo Lenti | Porting codice riconoscimento eventi su GPU | GPU |



Risultati ricognizione (2 / 5)

| Titolo Progetto | Ente | Persona di Riferimento | Tematica Scientifica | Keywords tecnologiche |
|----------------------------------|-------------|-------------------------------|--|---|
| ET+Virgo | UniFi | Massimo Lenti | Searching transients GW event in Virgo+ET | GPU Machine Learning for pattern recognition |
| Virgo | UniTS | Edoardo Milotti | Transient event analysys with machine learning + code acceleration | GPU Machine Learning for pattern recognition |
| Pierre Auger for MM Astronomy | Uni Salento | Lorenzo Perrone | Optmise event reconstructions | Huge I/O Machine Learning for pattern recognition |
| EUCLID | Uni Salento | Achille Nucita | Simulations and analysis for recognition faint sources in the VIS Euclid data | Huge I/O Machine Learning for pattern recognition |



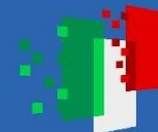
Risultati ricognizione (3 / 5)

| Titolo Progetto | Ente | Persona di Riferimento | Tematica Scientifica | Keywords tecnologiche |
|------------------------|--------------------------|-------------------------------|---|--|
| Theory of GW | Uni Salento | Claudio Corianò | Codes for the identification of signals in models of modified gravity and, within a collaborative data analysis framework, to study and optimize the existing codes | GPU (optional) |
| DAMPE+HERD | Uni Salento / INFN Lecce | Francesco de Palma | Development of new techniques for event /particle reconstructions | Machine Learning for pattern recognition |
| ASTRI + JEM-EUSO | INAF | Antonio Pagliaro | Riconoscimento di immagini e classificazione eventi negli esperimenti ASTRI e JEM-EUSO | Extension of current ML to Deep Learning GPU |



Risultati ricognizione (4 / 5)

| Titolo Progetto | Ente | Persona di Riferimento | Tematica Scientifica | Keywords tecnologiche |
|--|------------------|------------------------|---|---------------------------------------|
| JWST, LSST and VST | Uni. Federico II | Maurizio Paolillo | Simulazioni di immagini e curve di luce di Nuclei Galattici Attivi per i casi scientifici di VST, LSST e/o JWST | Code optimization (possibly with GPU) |
| Pipelines | INAF | Roberto Peron | General complex pipeline scaling | - |
| Turbulence Navier Stokes equations (MHD,...) | UniCal | Leonardo Primavera | PDE solving | GPU Machine Learning |
| BH and GW numerical simulations | UniCal | Sergio Servidio | EFE solving. Use of computational resources and optimisation of SFINGE code | GPU - CUDA |



| Titolo Progetto | Ente | Persona di Riferimento | Tematica Scientifica | Keywords tecnologiche |
|------------------------------|-----------------|--------------------------------|---|------------------------------|
| Plasma simulations | UniCal | Francesco Valentini | Code optimisation in Plasma Physics - | GPU - CUDA |
| ASTRI Stellar Interferometry | INAF | Luca Zampieri Marco Landoni | Porting of code for stellar interferometry in the context of ASTRI Mini Array | Massive I/O GPU - CUDA |
| LiteBIRD | Uni Ferrara | Luca Pagano Paolo Natoli | CMB Map Making per Lite BIRD | GPU Optimisation |
| Likelihood for CMB/LSS | Uni Ferrara | Luca Pagano Paolo Natoli | Likelihood optimisation in CMB/LSS topics | GPU - CUDA |
| LIGO/VIRGO | Roma - Sapienza | F. Pannarale | Pipeline optimization for GW signal detection | GPU Machine Learning |
| IACT telescopes | INAF | Ciro Bigongiari, F. Visconti | IACT data analysis | Machine Learning |



Criteri (da condividere) per “normalizzare” gli use case 1/2

- Massima inclusività verso le tematiche proposte, purché in linea con le finalità del CN e del WP 3.
- L'accorpamento di più UC risponde alla necessità di:
 - Valorizzare le sinergie esistenti tra gruppi con proposte affini
 - Evitare una eccessiva granularità nella gestione anche amministrativa delle attività
- Tra gli UC verranno definiti dei flagship trainanti legati alle milestones



Criteri (da condividere) per “normalizzare” gli use case 2/2

- La proposta dei coordinatori è di creare dei “contenitori” per raggruppare un numero limitato (ordine ~ 5) use case affini tra quelli da voi proposti
 - Questi “contenitori” potrebbero avere un tag semplice e relativamente ampio, da usare come interfaccia esterna, anche ai fini formali
 - Il vantaggio di questo approccio è di salvaguardare la natura, anche specifica, delle vostre proposte
 - L'evidente svantaggio è una complicazione della gestione :-)
- Il punto chiave è quale criterio adottare per questi raggruppamenti
 - Scientifico: è il più affine alla nostra mentalità, ma ci sembra che funzioni bene per alcune proposte, e peggio per altre che “fittano male”. Inoltre, questi contenitori dovrebbero rispondere a differenti esigenze tecnologiche, complicando l'interfaccia coi WP 4 e 5
 - Tecnologico: più coerente coi fini del CN, ma sparpaglierebbe gruppi affini in contenitori diversi. Inoltre, visto che la maggior parte delle richieste sono su GPU o machine learning, una divisione grossolana risulterebbe sbilanciata.
- Sarebbe utile avere un feedback



Required resources (preliminary)

| Macro tematica | Richieste CPU | Richieste GPU | Richieste storage + I/O |
|--|--------------------------------|---------------|-------------------------|
| <i>Gravitational Waves</i> | 7 milioni CPU/hrs equivalenti | TBD | 100 TB |
| <i>Particle and astroparticle physics</i> | 10 milioni CPU/hrs equivalenti | TBD | 700 TB |
| <i>Cosmology</i> | 10 milioni CPU/hrs | TBD | 20 TB |
| <i>Theory - XTalk use cases</i> | 8 milioni CPU/hrs | TBD | 1 TB |



Grand total (preliminary)

- Storage: Approx 1 PB (some use cases still missing estimates)
- CPU: 30 milioni core/hrs equivalenti (depends on the use of GPU)
- Scaling on GPU could reduce a lot the CPU requirements but this is still to be defined (also at the level of the design of each use cases, as far as we have understood).



Questions ?

- Questione dell'uso di dati proprietari. Sono protetti? Chi ha bisogno di questo servizio?
- Questioni relativi alla pubblicità del codice ? Andrà necessariamente reso pubblico ? Può rimanere all'interno del centro?

Nota: esiste una bozza di documento IP che definisce questi aspetti.

Versione finale elaborata dal Gruppo di Lavoro nominato dal CdA di ICSC – 28/10/2022

IP GUIDELINES FOR CN HPC

Recitals

a. These Guidelines apply to all Members of the CN HPC, regardless of their role within. Notwithstanding, specific rules may apply to some entities.

b. These Guidelines provide IP general principles that can be adapted within each Spoke/Research Projects and supplemented in accordance with the upcoming CN HPC Hub/Spoke regulations and/or internal procedures. In any case, any specific IP provision the Members of the Spoke may agree upon shall not be in contrast with this guidelines or other regulation and/or internal procedures at the level of the Hub. At the same way, any specific IP provision the Members may agree upon with regards to a Research Project shall not be in contrast with this guidelines or other regulation and/or internal procedures at the level of the Spoke.

c. These Guidelines apply also to the Hub considering its role in ensuring transparency and openness



Regular meeting

- Faremo circolare un Doodle per fissare degli incontri regolari su base bisettimanale (da compilare entro il 9 Gennaio).
- Sulla base dei risultati, fisseremo il giorno che riteniamo possa essere indicativamente nella settimana del 16 Gennaio.
- Criterio di massimizzazione partecipazione dei gruppi (teleconf aperta a tutti).



Finanziato
dall'Unione europea
NextGenerationEU



Ministero
dell'Università
e della Ricerca



Italiadomani
PIANO NAZIONALE
DI RIPRESA E RESILIENZA



Project management - Easyredmine (Marco Landoni)

Live demo ?