Report sull'infrastruttura di calcolo HPC ai LNGS

Sandra Parlati

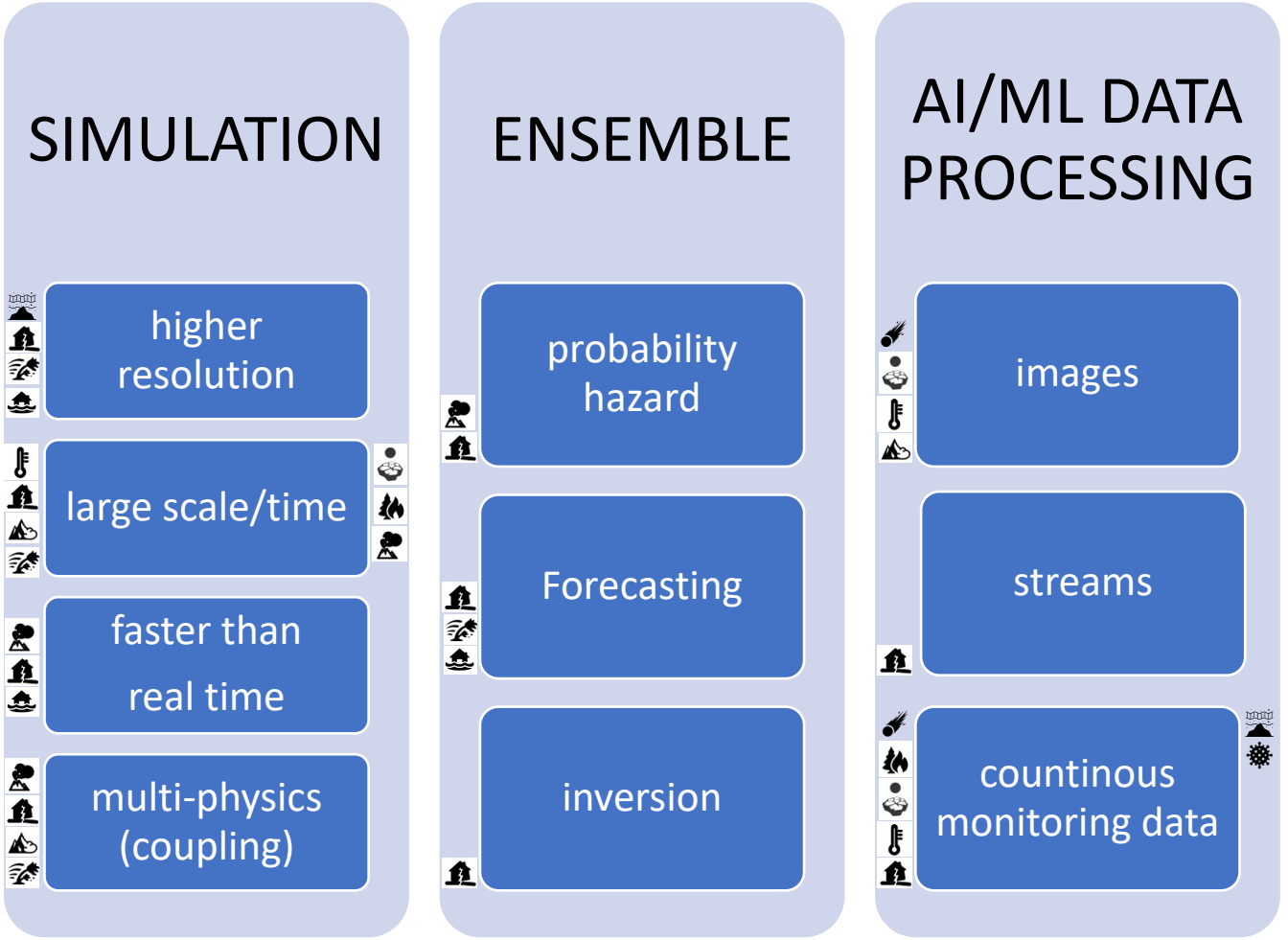
Workshop sul calcolo nell’INFN – Loano 22-26 Maggio 2023

Riassunto delle puntate precendenti…

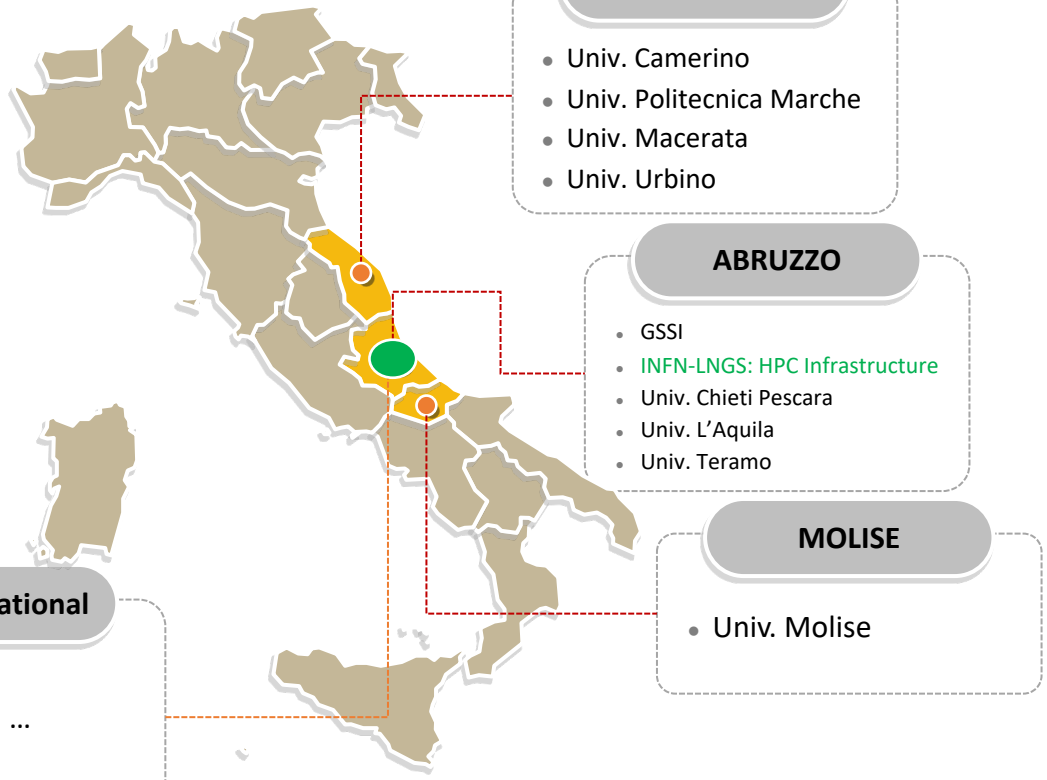
- La proposta di avere un centro HPC ai LNGS nasce nel 2020 nell'ambito del progetto HPC4DR
- Il progetto HPC4DR nasce nel 2020 da Università e centri di ricerca delle regioni Abruzzo, Marche e Molise, duramente colpite dai terremoti e da altri eventi catastrofici nel 2016/2017
- L'idea è quella di "realizzare un **centro di competenze** per la riduzione dei rischi connessi ai disastri dovuti a fenomeni naturali e di origine umana, dotato di un'infrastruttura tecnologica di calcolo ad alte prestazioni"

HPC and AI per capire il rischio

- **Floods**
- **Storms and Cyclones**
- **Landslide and hydrogeological events**
- **Earthquake and tsunami**
- **Extreme weather**
- **Drought**
- **Wildfires**
- **Volcanoes**
- **Near-earth asteroids and space debris**
- **Space weather**
- **Epidemics**



HPC4DR PARTNERSHIP



MARCHE

- Univ. Camerino
- Univ. Politecnica Marche
- Univ. Macerata
- Univ. Urbino

ABRUZZO

- GSSI
- INFN-LNGS: HPC Infrastructure
- Univ. Chieti Pescara
- Univ. L'Aquila
- Univ. Teramo

MOLISE

- Univ. Molise

Local vs. National

- INGV
- INFN – LNGS, …
- INAF
- CNR – ITC, …

Dal CINECA ai LNGS

- Opportunita' di avere nodi HPC per l'INFN dalla dismissione di GALILEO
- Il consorzio HPC4DR ha scelto i LNGS come sede per l'infrastruttura HPC
- A gennaio 2022 e' stato firmato l'accordo tra **INFN** e **CINECA** per la cessione gratuita di
  - n. 5 rack GALILEO ciascuno dei quali contenente 72 nodi
  - N. 1 rack GALILEO contenente 36 nodi di calcolo
- Ad aprile 2022 i nodi sono stati consegnati ai LNGS
- Al workshop di Paestum: i nodi erano stati collegati all'impianto elettrico del CED da pochi giorni!

Dal CINECA ai LNGS

L’infrastruttura di calcolo e’ Lenovo NextScale

Ogni rack contiene fino a 6 Lenovo NeXtScale n1200 Enclosure (6RU)

Ogni enclosure contiene 12 NeXtScale nx360 M5 Compute Node

- 2*Intel Xeon E5-2697 v4 @ 2.30GHz 18-core each (Broadwell)
- 128 GB di RAM/nodo, 3.5GB RAM/core

- 5 rack * 72 nodi, 1 rack * 36 nodi -> 396nodi, 14256 core
- Ogni server ha una potenza di calcolo di picco di 1.3Tflop/s.
- La potenza di calcolo totale e’ di circa 0.5PFlops
- I nodi di calcolo sono interconnessi da una rete Intel Omnipath a 100Gb/s e una rete ethernet 1Gb/s

Piano per l'utilizzo delle risorse

- Ogni rack assorbe circa 30KW a pieno regime
- Compatibilita' con gli impianti del CED: solo 1 rack con 72 nodi e 1 rack con 36 nodi sono stati portati al CED e installati: siamo in una fase iniziale di test!
- Il rack con 36 nodi sara' utilizzato per i servizi di accesso cloud
- Storage e server di gestione dei nodi non erano compresi nel materiale dismesso da Galileo
- Acquistati con fondi LNGS e installati
  - 2 server 'Master' per la gestione centralizzata del cluster HPC
  - 3 server CEPH (qualche decina di TB)
  - 2 storage server
  - 1 storage ~200TB netti (20x16TB + 4x2TB SSD)
- I rimanenti 4 rack di calcolo saranno installati dopo l'upgrade degli impianti elettrici e di condizionamento del CED.

Dal CINECA ai LNGS

- Difficolta’ che abbiamo affrontato:
  - Tutti gli apparati (server e rete) sono stati resettati dopo la dismissione e sono arrivati privi di sistema operativo
  - Scarsa documentazione: es. schemi fisici/logici della rete ethernet e Omnipath, configurazione XCAT
  - Rete Omnipath e modalita’ di gestione completamente nuove per noi
  - Rete Omnipath 100Gb/s non piu’ supportata da Intel; venduta a Cornelis che sta sviluppando Omnipath 400Gb/s
  - La presenza della rete Omnipath ha impattato sulla scelta del SO dei nodi HPC (rocky Linux 8.5) e sui server di storage
  - Frequenti e fruttuosi contatti con CINECA su configurazione rete, server, installazione centralizzata e gestione ambiente di calcolo


progress

Stato attuale

- Gli switch di rete ethernet sono configurati e la rete (IPMI, 1Gb/s, 10Gb/s e connessione alla dorsale LNGS) e' funzionante
- Rete Omnipath funziona
- E' stato realizzato di un ramo di rete, separato dalla LAN e dalle reti degli esperimenti, che collega il cluster HPC al router di frontiera
- Sono operativi i servizi di rete DNS, DHCP, VPN
- E' stato creato un dominio DNS hpc.lngs.infn.it per i nodi del cluster
- I nodi di calcolo sono stati installati centralmente con XCAT
- Il sistema operativo dei nodi e' rocky8.5 compatibile con Omnipath (software Cornelis disponibile per Rocky8.5)
- Puppet per la configurazione centralizzata di nodi e servizi

Stato attuale

- E’ stato installato un Filesystem Lustre (visto dai nodi di calcolo su Omnipath)
- Sono state create su Lustre le aree $HOME, $DATA e altre aree condivise per gli utenti
- Configurazione iniziale nodi di calcolo sullo stile CINECA
  - Gestione del software con ‘spack’
  - Gestione dell’ambiente utente con ’module’
  - Sono stati installati i compilatori OpenMPI e IntelMPI
  - Sono stati installati software di comune utilita’
- Sono stati creati due nodi di public login per l’accesso ssh
- Creato il tool per la registrazione degli utenti locali e inserimento in AAI
- Siamo sostanzialmente pronti a ospitare i primi utenti di test
- La modalita’ di accesso alle risorse è, ad oggi, simile a quella CINECA

Stato attuale

- Stiamo utilizzando dei tool per la documentazione interna: cloud storage gsbox.lngs.infn.it, confluence
- Git per il software
- Utilizziamo Servicedesk per il ticketing
- Sistema di monitoring (check_mk) e di asset management (OpenDCIM? Insigh?) sono in fase di valutazione
- Ancora da valutare e implementare: sistemi di gestione delle priorita’ e accounting
- Grazie al Servizio Calcolo e reti dei LNGS per aver svolto tutte queste attività!

Stato attuale

- Primi contatti con la comunità scientifica di HPC4DR
- Queste comunità usano di solito risorse CINECA e/o piccoli cluster locali
- Primi use case:
  - propagazione di onde sismiche e simulazione numerica di terremoti
  - monitoraggio strutturale
  - uso di AI per allerta precoce
  - AI e social media per gestione rischi e conseguenze di calamita' naturali
  - monitoraggio space weather
  - modelli per le previsioni meteo a breve e medio termine e simulazione di eventi severi
- Alcuni use case con MPI, implementabili da subito; altri su GPU.
- Possibilita' a breve di sperimentare sistemi con GPU grazie ad un server non INFN ospitato nell'infrastruttura HPC.

ICSC – Italian center for Super Computing

The ICSC will include ten **thematic Spokes** and one ***Infrastructure* spoke**

EDUCATION & TRAINING, ENTREPRENEURSHIP, KNOWLEDGE TRANSFER, POLICY, OUTREACH

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

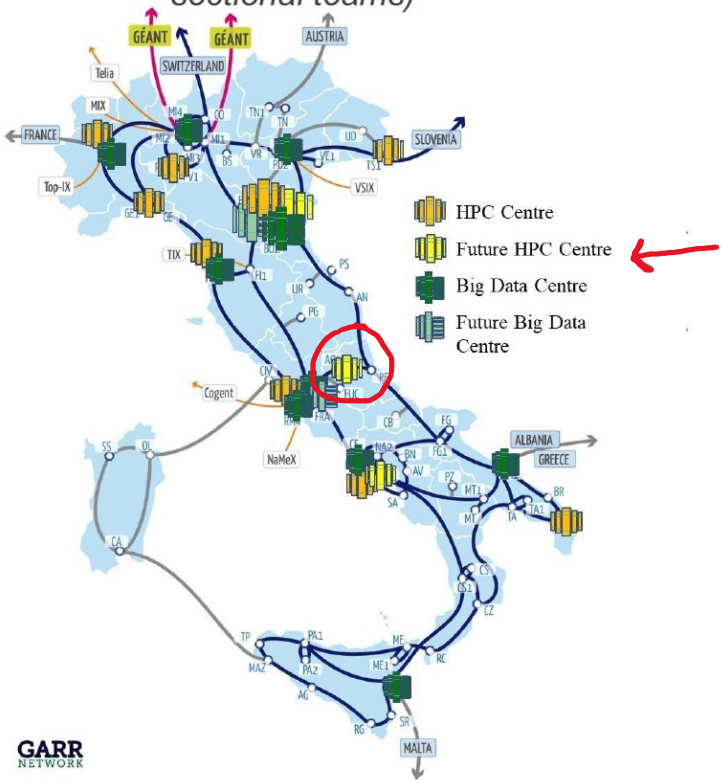
8 IN-SILICO MEDICINE & OMICS DATA

9 DIGITAL SOCIETY & SMART CITIES

10 QUANTUM COMPUTING

0 SUPERCOMPUTING CLOUD INFRASTRUCTURE

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



Attività INFN

Spoke 0 Infrastructure (INFN co-leader)

Hardware Tier-1 e Tier-2	30.0 M€
Hardware INFN cloud	10.0 M€
Hardware Tier-1 HPC4DR ai LNGS (per spoke 5 -Env.)	5.0 M€
Hardware Tier-1 ESA ai LNF (per spoke 2)	5.0 M€
Personale strutturato (incl. costi indiretti)	1.0 M€
Personale a tempo determinato (incl. costi indiretti)	5.2 M€

Spoke 2 Fundamental Research & Space Economy (INFN leader)

Personale strutturato (incl. costi indiretti)	1.4 M€
Personale a tempo determinato (incl. costi indiretti)	1.1 M€

Finanziamenti ICSC per..

- Ricondizionamento rack (36 nodi) per installazione cloud e servizi di accesso alle risorse HPC – gara in corso
- Acquisto nuovo storage da inserire nel cluster Lustre – gara in corso
- Upgrade del CED:
  - Creazione di un'isola a corridoio freddo capace di ospitare 12 rack e adeguamento impianto di condizionamento
  - L'isola dovra' avere un pavimento in grado di sostenere il peso dei rack
  - Installazione nuovi UPS in grado di sostenere un carico previsto di circa 400KW, duplicazione linee di alimentazione degli UPS
- Upgrade della rete del cluster HPC – gara in corso
- Acquisto router di frontiera e upgrade collegamento al GARR 100Gb/s
- Acquisto di altro storage (sia veloce sia per conservazione a lungo termine dei dati) e server con GPU (AQ Terabit)
- Acquisto libreria di nastri per backup

Finanziamenti ICSC per…

- Saranno assunti con fondi PNRR ICSC 2 tecnologi e 2 tecnici che si occuperanno di queste attività
  - Gestione dell'infrastruttura (rete calcolo storage, monitoring, accounting,..)
  - Installazione e configurazione di SLURM
  - Gestione del middleware cloud per accesso alle risorse HPC e integrazione in DATACLOUD
  - Aumentare esperienza sulla gestione del calcolo HPC (MPI e GPU)
  - Supporto alle comunita' scientifiche di HPC4DR per l'accesso e l'uso efficiente delle risorse
  - **Supporto agli esperimenti dei LNGS per l'uso delle risorse HPC!!!**
  - Informazione agli utenti e formazione
  - Partecipazione alla attivita' di DATACLOUD
  - Supporto all'installazione delle ulteriori risorse HPC Cineca e delle nuove risorse acquistate con fondi ICSC

..alla prossima puntata…

…spero con prime applicazioni funzionanti in ambito Disaster Resilience