



MISTRAL

Meteo Italian Supercomputing Portal

MISTRAL, la piattaforma italiana degli open data meteorologici e le sue funzionalità
Evento Online - 23 Novembre 2020

Mistral e il ruolo del supercalcolo in ambito meteo e clima

Sanzio Bassini

Direttore del dipartimento High Performance Computing
CINECA

UN CONSORZIO PRIVATO DI DIRITTO PUBBLICO

CINECA is a **Consortium of Universities**, founded in 1969 by (MIUR) to support scientific research
CINECA is a not-for-profit organization

Members:

- MUR & MPI
- 67 Italian Public Universities
- 9 Institutions,...

In-house provider:

- ❖ Supervised by MIUR
- ❖ No private capital
- ❖ 80% of activities towards Consortium members
- ❖ 20% Technology Transfer
- ❖ Annual Budget: ~ 100M€
- ❖ Employees: ~ 800



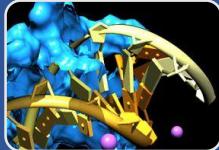
CINECA IN SHORT



Services both for Universities and Ministry

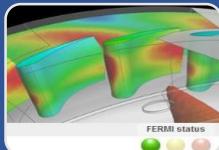
- Solutions & Services for the University Administration
- Information system for the Ministry for the management, assessment, evaluation, funding of the research

Research and Innovation



- Promote the use of the most advanced High Performance Computing systems to support public and private scientific and technological research

Innovation and technology transfer



- Numerical experiment, virtual prototyping
- Big data, ML, AI
- Scientific visualizations, computer graphics

HIGH PERFORMANCE COMPUTING



Italian inter-university consortium CINECA — one of the European and world's most important supercomputing center

MARCONI100

Open in production in February this year.
Number 9 in June 2020 Top500



LEONARDO pre-exascale

Procured by EuroHPC JU
In production: 1Q22; 250 PF HPL; 10 EF AI



BOLOGNA BIG DATA TECNOPOLO

A European, National and Regional investment for HPC Big Data and AAI



A machine to compute the weather



John von Neumann



The ENIAC Computer, Princeton, NJ

Il meteo al Cineca: una lunga storia

Dall'archivio storico CINECA...

Il primo contratto operativo è stato firmato **dall'ERSA** (Ente Regionale Sviluppo Agricolo – RER) **nel 1987**, a seguito della disponibilità del primo Cray, installato al CINECA nel 1985.

«ERSA riuscì ad avvisare in tempo gli agricoltori di Vignola di una gelata primaverile tardiva con i ciliegi già in fiore.

Non si poteva fare nulla, salvo innaffiare le piante per consentire che si alzasse di qualche metro da terra il livello dello zero termico atmosferico. »

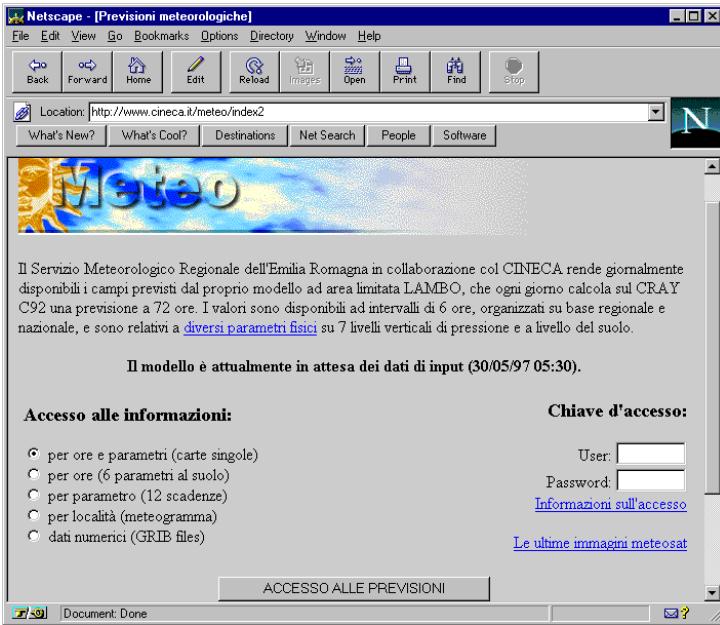
Questa dimostrazione di efficacia con i “cerasari” di Vignola convinse la Regione (ER) a compiere un investimento sul servizio.

Da una fase pre-operativa si passò quindi a una fase operativa e di servizio, fino a che la Regione divenne Centro di competenza nazionale per le previsioni numeriche.



10 anni dopo!

Servizio di elaborazione e diffusione delle previsioni meteorologiche



Utenti del servizio

Regioni: Friuli, Toscana,
Sardegna, Marche, Abruzzo

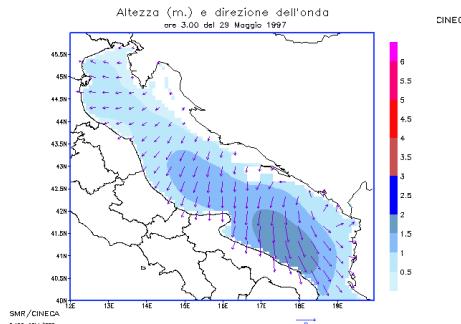
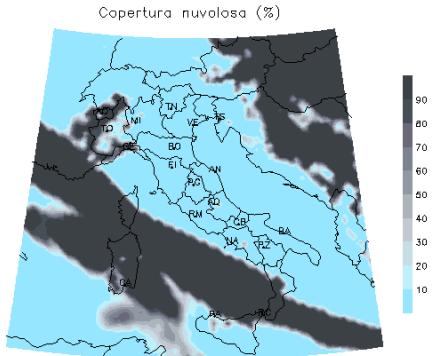
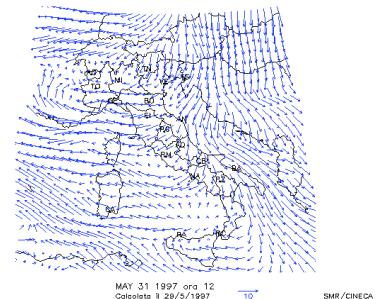
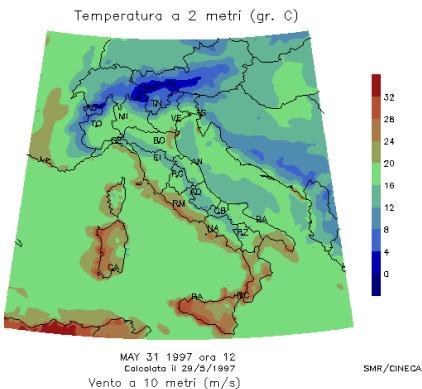
Ministero dei Trasporti -
Direzione Marittima

Sviluppo di applicazioni relative ai problemi di qualità dell'aria in ambito urbano

ENEA / Comune di Roma Sistema
“ATMOSFERA”

Provincia di Firenze

Servizio di elaborazione e diffusione delle previsioni meteorologiche



Clima, Italia nella Top 10 dei Paesi con più danni legati a catastrofi naturali

Negli ultimi 20 anni aumentate del 151% le perdite economiche mondiali dovute ai disastri ambientali e climatici

Climate-related disasters dominate the picture over the past 20 years, accounting for 91% of all 7,255 recorded events between 1998 and 2017.

Within this total, floods were the most frequent type of disaster, 43% of all recorded events (Figure 3).

Studio dell'ONU per la riduzione del rischio dei disastri, (l'Unisdr) e del Centro di ricerca sulla epidemiologia dei disastri (Cred).

Figure 2

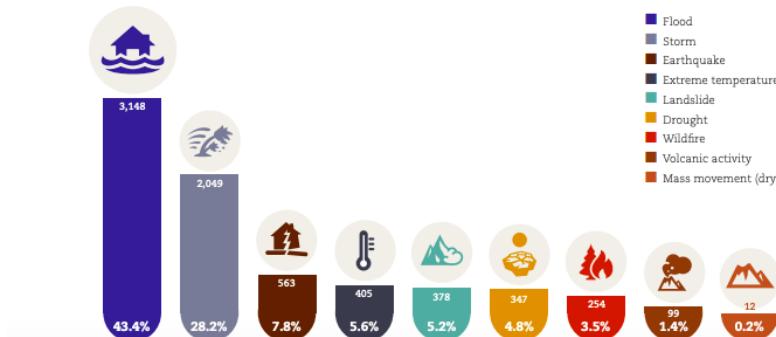
Number of disasters by major category per year 1998-2017



Previsioni meteo di eventi meteo avversi sono fondamentali per offrire un allarme tempestivo alle autorità e al pubblico, dando il tempo di attuare piani di emergenza.

Figure 3

Numbers of disasters per type 1998-2017



Catene Operative che il Cineca gestisce per Arpae e Protezione Civile

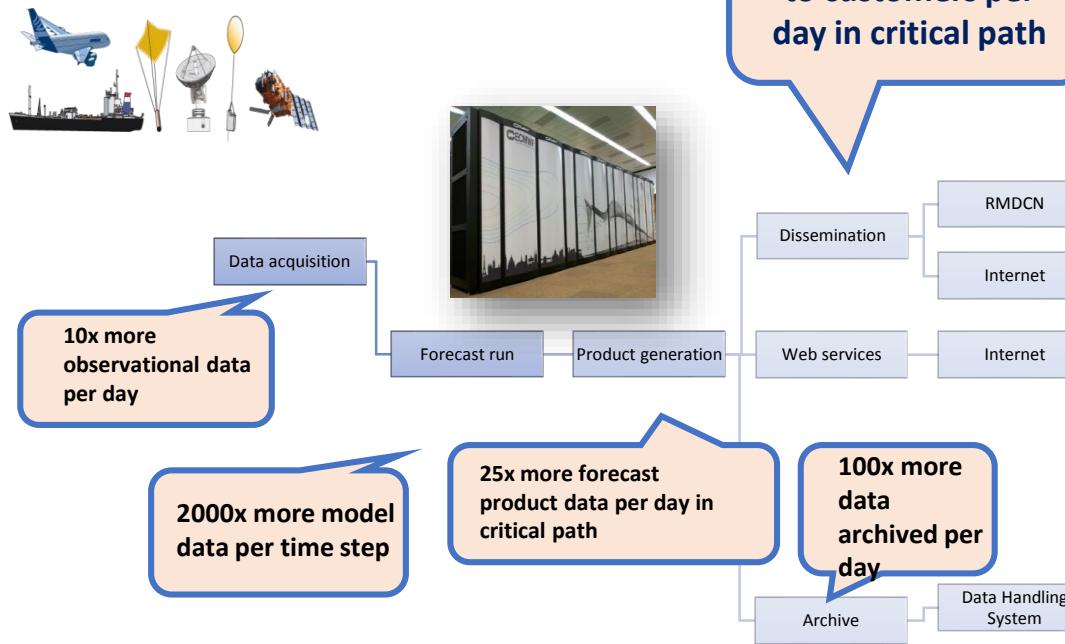
Catena Operativa	Risoluzione	Area	Corse giornaliere	Previsioni a
Cosmo-5M	5 Km	Mediterraneo	2 corse: 00 e 12 UTC	-18 ore (assimilazione) e +72 ore (previsione).
Cosmo-2I	2,2 Km	Italia	2 corse: 00 e 12 UTC	+48 ore (previsione)
SWAN-MED-ITA-RE		Stato del Mare Mediterraneo, Italia e su 5 aree geografiche a ridosso delle coste italiane	1 corsa: 00 UTC	-24 ore (assimilazione) e +72 ore (previsione)
LAMI-ENDA	2,2 Km	situazione meteorologica attuale e previsioni a brevissimo termine	corse di 3-18 ore di durata, distribuite lungo tutto l'arco della giornata.	
LAMI-ENS	2,2 Km	previsione probabilistica mediante tecniche di ensemble forecasting	20 (40) corse	

Risorse di Calcolo per le catene operative meteo

Catena Operativa	Risoluzione	Risorse di calcolo	Storage (dati creati al giorno)	Archivio dati (arkimet)
Cosmo-5M	5 Km			
Cosmo-2I	2,2 Km			
SWAN-MED-ITA-RE		<ul style="list-style-type: none"> • Meucci : 39 nodi (1404 cores) 12.3 M core hours/year • Galileo : 30 nodi (1080 cores) 9.5 M core hours/year 	<ul style="list-style-type: none"> • Meucci : 0.5 TB day • Galileo : 0.5 TB day 	<ul style="list-style-type: none"> • Meucci: 109 GB day (~40 TB year) • Galileo: 109 GB day (~40 TB year)
LAMI-ENDA	2,2 Km			
LAMI-ENS	2,2 Km	<ul style="list-style-type: none"> • Galileo : fino a 95 nodi (3420 cores) 30.0 M core hours/year • Meucci: fino a 61 nodi (2196 corse) 19.2 M core hours/year 	<ul style="list-style-type: none"> • Meucci : 160 GB day • Galileo : 20 GB day 	
		Tot = 71 M core hours/year		

Weather & Climate “is” Big Data

Growth factors of data volumes along forecast chain in next 10 years.



Today: ~2 Pbyte per week!
Total archive: ~ 250 Pbyte

Courtesy: Peter Bauer - ECMWF

Numbers

Glossary:

s : denotes simulated days per day, a measure of throughput

f : denotes floating point operations per second. (Yes, not transferrable across applications, or across machines of different classes, but roughly scales performance of the same application across similar classes (memory bandwidth, interconnect speeds etc) of machines.

- Full scale benchmarking at 5 km gives 50 s/Pf on traditional (CPU), and 25 s/Pf on hybrid (GPU) architectures.
- In reach, and transformative, would be $s=100$ on a 1.25 km (Δx grid). This demands an additional 2^7 in computational capacity (with no vertical refinement), 2^3 of which must come from strong scaling.
- This is feasible on existing machines, practical with exa-scale (already on Fugaku?).
- Getting the clouds (not just V) right may require global large-eddy simulation ($\Delta x=0.15\text{km}$) and two fold vertical refinement.
- This requires a performance boost of 2^{11} . Because 2^4 must come from strong scaling this seems unfeasible. But... $s=0.2$ might be enough for this problem. This gives back a factor of 2^9 , and eliminates need for improvement in strong scaling. So we might just get the clouds as well.

Some transformative calculations are within reach, but the end is also in sight.
Courtesy: Bjorn Stevens - Max-Planck-Institute for Meteorology

Connecting Europe Facility (CEF)

Public Open Data Programme of the EU



MISTRAL

Meteo Italian Supercomputing portal

Durata: 01/10/ 2018-31/01/2021

Coordinatore: Cineca

Partners: DPC, ARPAE, ARPAP,, Dedagroup, ECMWF

Budget: **2.342.243 €**

Obiettivi: Creare il portale nazionale di dati metereologici e di servizi



Co-financed by the Connecting
European Facility of the European Union

Grazie

- Sanzio Bassini
- CINECA, Direttore dipartimento HPC