

Strategie INFN sul calcolo e su partecipazione ai progetti

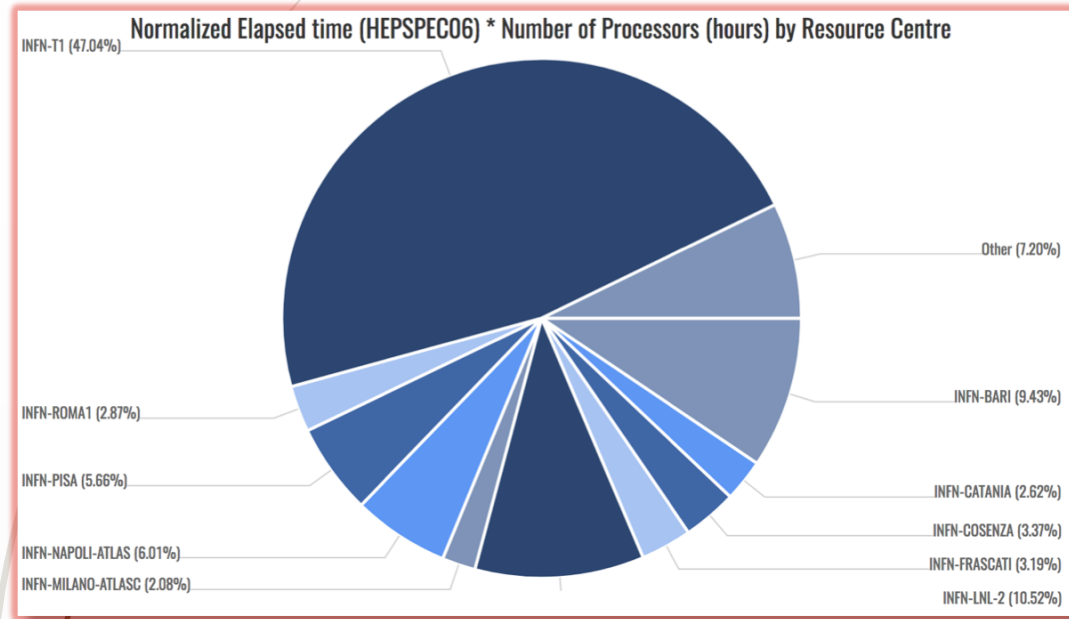
Donatella Lucchesi
University and INFN Padova

ReCaS - Riunione n. 2 Comitato di gestione Accordo ReCaS
Bari, February 19^{en} , 2018

In questa presentazione:

- quadro della situazione attuale sia per le risorse HTC e HPC che per i progetti in corso
- Evoluzione delle richieste da parte degli esperimenti
- Strategie INFN (alcune solo mie)

Le Risorse INFN e la “Cloud”



Tier1 corrisponde attualmente al 47%

Donatella Lucchesi - INFN & Universita' Padova

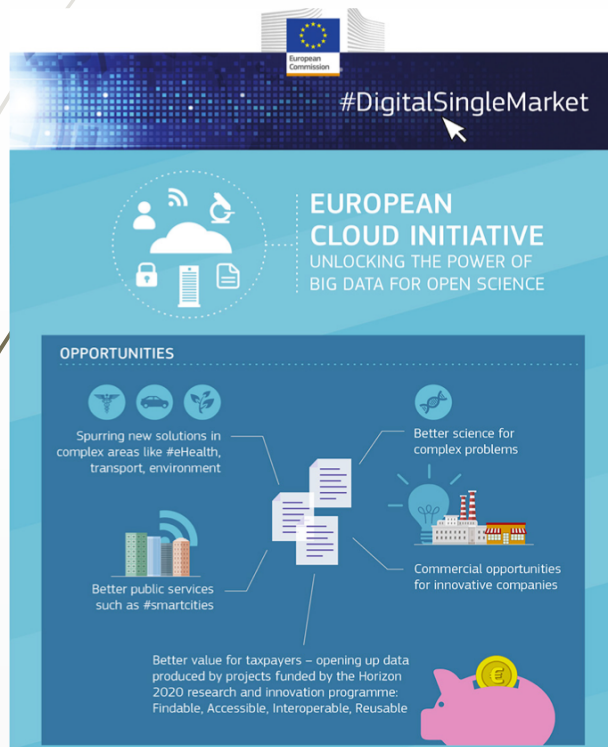
Passaggio da GRID a Cloud favorito dall'uso di fondi Europei, progetto INDIGO-DataCloud



L'Europa e la cloud, cioè' EOSC

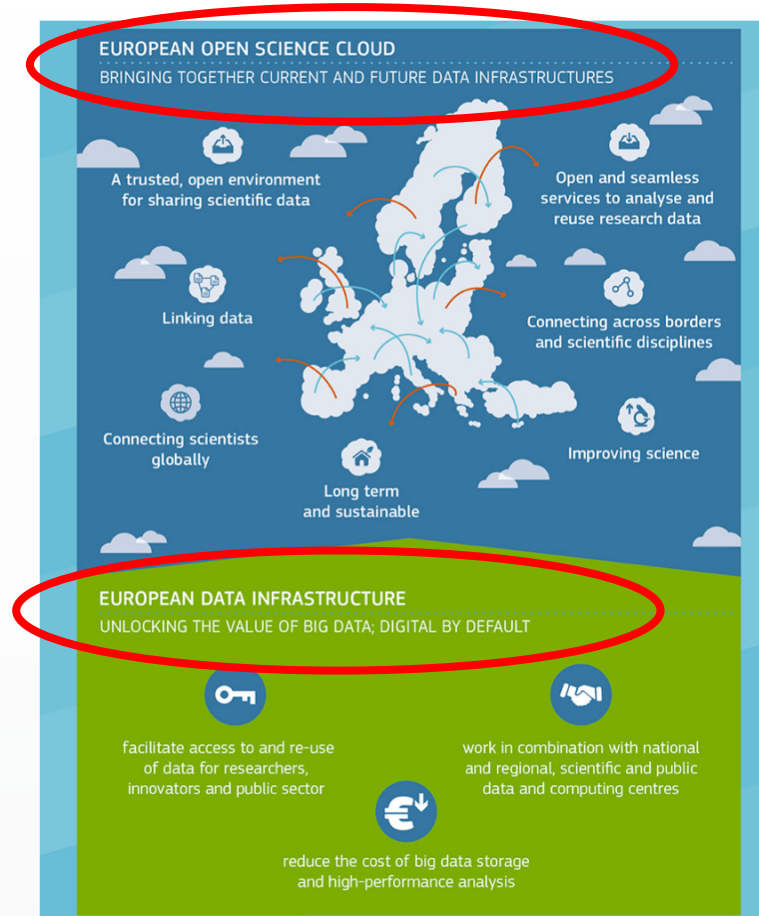
European Open Science Cloud, EOSC, nasce nel 2015 per volontà' della Commissione Europea per costituire un'infrastruttura per sostenere e sviluppare Open Science e favorire l'innovazione.

Nel 2016 viene pubblicato Blueprint
European Cloud Initiative



Donatella Lucchesi - INFN & Università' Padova

EOSC



EDI

Come si prosegue verso la “Cloud”



Vorrebbe definire la governance di EOSC
Realizzare dei dimostratori di EOSC
PI: UK INFN: C. Duma



Contribuisce alla EOSC rendendo possibile l'accesso a dati e servizi
PI: EGI INFN: L. Gaido



Sviluppa tecnologie per federare risorse di storage e per data management in sistemi distribuiti
PI: INFN D. Cesini

DEEP HybridDataCloud
Designing and Enabling
E-Infrastructures for intensive
Processing in a Hybrid DataCloud

Promuove uso di intensive computing (es. GPU) per diverse discipline sviluppando servizi e fornendo supporto
PI: CSIC (Spain) INFN: G. Donvito

Progetti in preparazione

6

INFRAEOSC-01: Access to commercial services through the EOSC-hub

- Call Budget: 12M€ submission deadline 22nd March 2018
- Consortium: GEANT (coordinator), CERN & RHEA
- Proposal: OCRE – Open Clouds for Research Environments
- Combine the work of GEANT IaaS tender and HNSciCloud PCP
- Establish a common process to procure commercial services integrated with public e-infrastructures

INFRAEOSC-04: connecting ESFRI RI through clusters the EOSC

ASTERICS-2 Proposal, Astronomy ESFRI & Research Infrastructure Cluster

WP per “ESRFI”

1. “Data lake” (HL-LHC, SKA, CTA)
2. “Software and service repository” (KM3NeT, CTA/LAPP, SKA)
3. “Data (interoperable archives and preservation)” (CDS, ESO/E-ELT).
4. “Open-science Platform” (SKA, CTA).

INFN (T. Boccali e A. De Salvo) partecipa al 1.

Progetti in preparazione -2

7

ICT-34-2018 Pre-Commercial Procurement

- Focus: Data Archiving in the cloud
- DARE, Data Archiving for Research Environments proposto dal CERN
- Complement in-house archiving/preservation with remote commercial cloud services providing long-term data storage
- Interesse da parte dei gruppi che si occupano di data preservation non-LHC

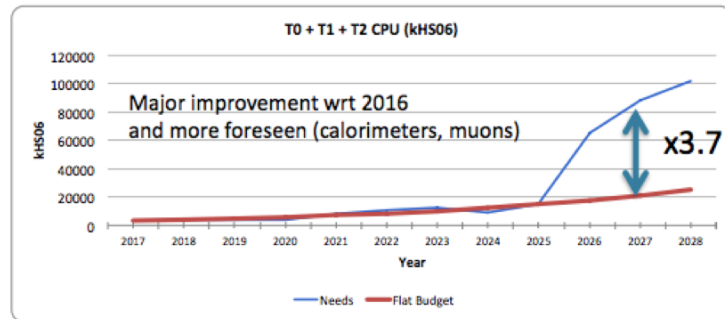
INFRAEOSC-05: Setup of an EOSC coordination structure

- Deadline 18-giugno 2018
- Key Features: "Support EOSC Executive Board and advise on EOSC implementation"
- Project budget: 10 M€
- Support decision making process; innovative business models; help set rules of participation; help structure EOSC user community convening EOSC Stakeholder Forum annually

Risorse necessarie per LHC High Lumi, Astrofisica e Astronomia

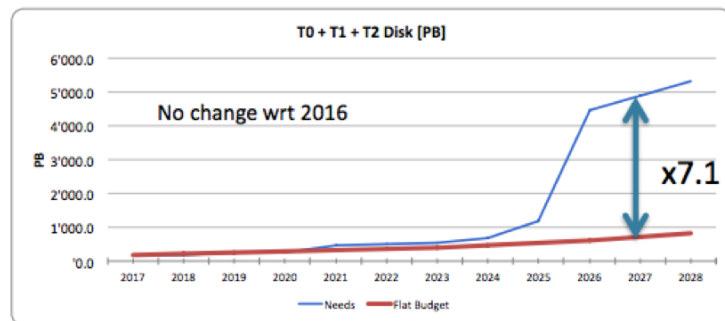
8

- Miglioramento tecnologico: limite ottimistico +20%/y → 6x in 10 anni
- Miglioramenti fino a 10x in CPU con architetture parallele e/o acceleratori (GPU/FPGA) → riscrittura sostanziale del codice e mancanza di competenze adeguate



CPU:

6x (CMS)
4x (ATLAS)



Spazio disco:

4x (CMS)
7x(ATLAS)

Astrofisica e Astronomia

INFN ha firmato una convenzione quadro con INAF per usare e sviluppare assieme una unica infrastruttura. Si proce poi, esperimento per esperimento con accordi attuativi:

- CTA gia' firmato
- EUCLID in discussion
- SKA?

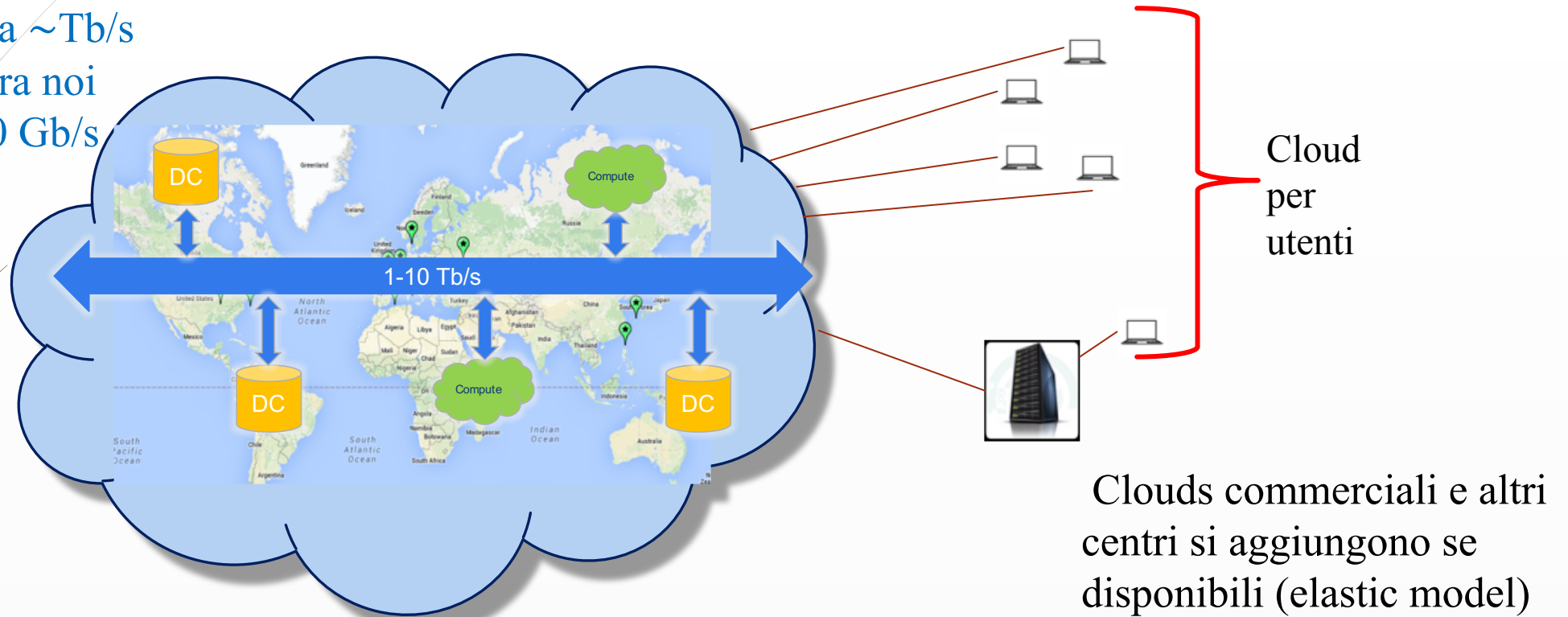
Queste risorse contribuiranno ~ quanto 1/2 LHC

Modello di infrastruttura proposto WLCG

9

Rete veloce
collegamenti a ~Tb/s
In Italia per ora noi
andiamo a 100 Gb/s

Compute Centers, ospitano CPU con
spazio disco sufficiente ad essere operativi

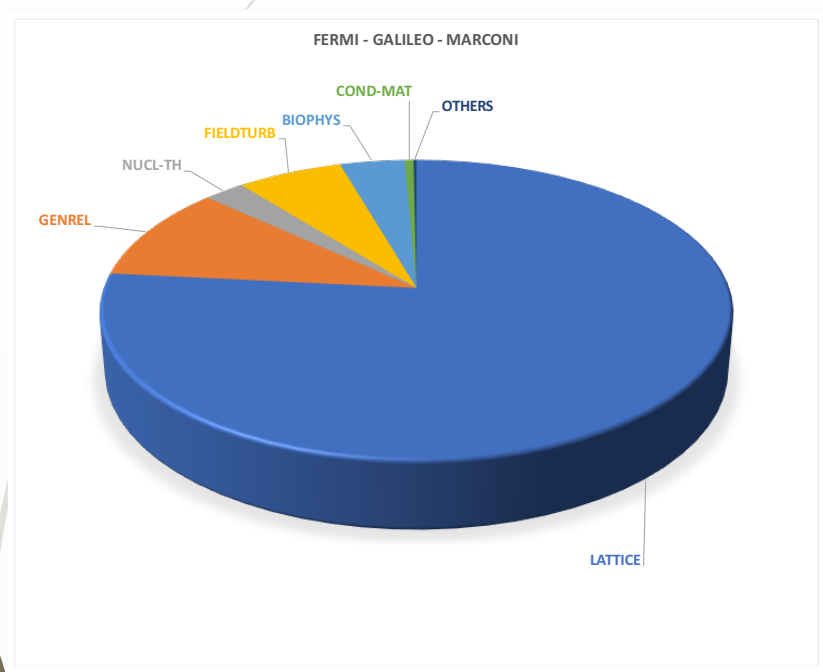


Data Centers (DC): ospitano la maggior parte dei tape e dello spazio disco con CPU necessaria, cuore della infrastruttura

Risorse HPC

10

Calcolo per Fisica Teorica Uso delle risorse HPC INFN al CINECA Settembre 2012-oggi



Usate anche altre risorse via
PRACE e ISCRA

Donatella Lucchesi - INFN & Universita' Padova

Risorse richieste per il 2018-2020

	2018	2019	2020
LGT: hadron physics	54	108	180
LGT: QGP and BSM	207	432	648
LGT: flavor physics	117	234	387
Colliders Phenomenology	1	2	3
General Relativity	142	182	227
Cosmology and Astroparticle Physics	3	4	6
Nuclear Theory	18	27	36
Fluid Dynamics	50	80	110
Quantitative Biology	9	18	27
Disordered systems	4	6	8
Condensed matter	2	4	6
Grand Total (Mcore-h)	607	1097	1638
Grand Total (Eq. Pflops)	4.6	8.4	12.5

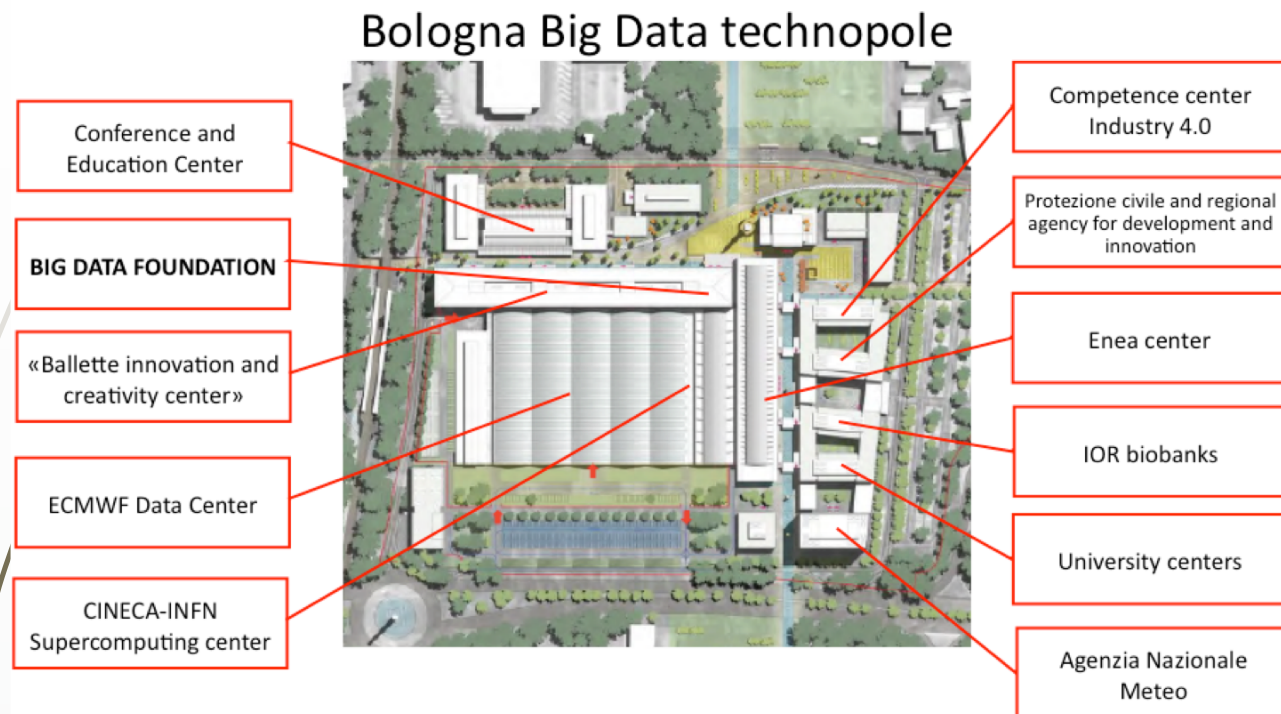
EUCLID: 20 Mcore-h

Convenzione INFN - CINECA per 3 anni:
6% MARCONI + 15 Mcore-h GALILEO per
un totale 164 Mcore-h (~1.4 Pflops)

Evoluzione dell'infrastruttura INFN: data center?

11

INFN puo' puntare ad avere un data center grazie al Tecnopolo di Bologna che ospitera' nuovo centro Centro Europeo per le Previsioni Meteorologiche di Medio Termine (ECMWF)



Area del tecnopolo e' piu' grande delle necessita' ECMWF, quindi puo' ospitare:

- CNAF e Tier-1 ampliato
- Cineca

Allocati 40 M€ dal MIUR per la ristrutturazione edifici

Evoluzione della infrastruttura INFN: altri centri

Survey sull'Infrastruttura di calcolo scientifico INFN

* Required

Analizzate 30 infrastrutture molte delle quali multifarms

30 Infrastrutture	
Grandi (T1+T2+CS+TS)	Piccoli
10	20
95% risorse	5% risorse

Analisi completa in corso ma:

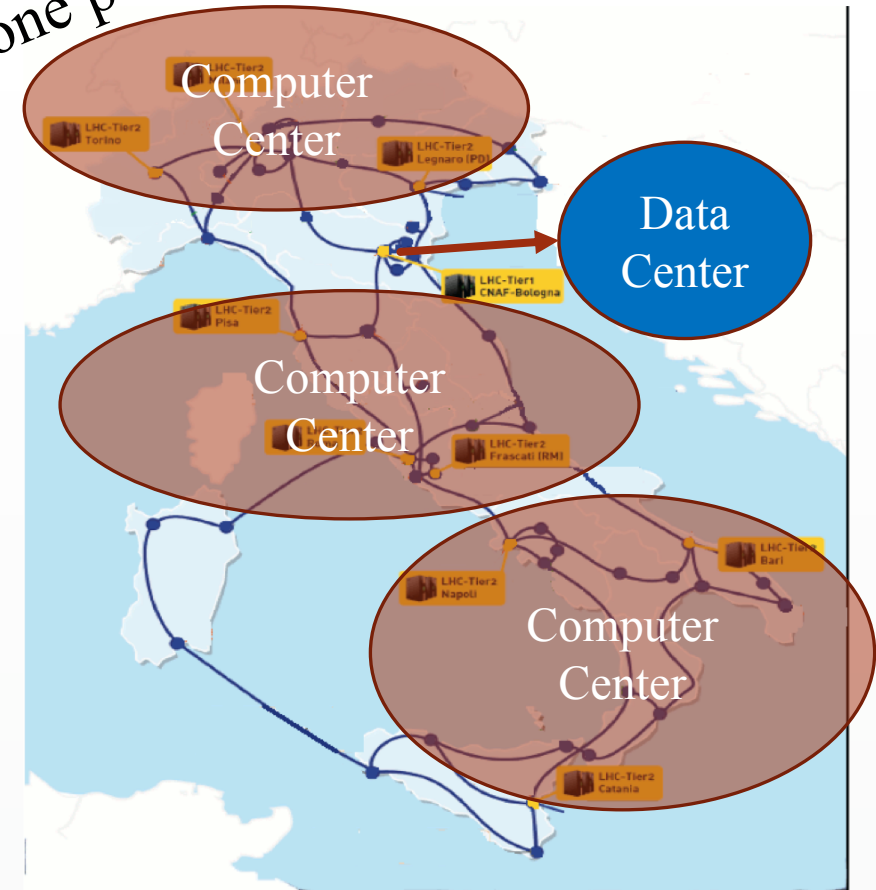
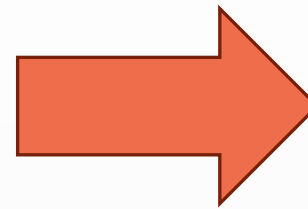
- Frazione consistente di personale e' dedicato ai centri piccoli (ben piu' del 5%)
- Sinergia molto importante con le universita' che contribuiscono a ospitare, operare e mantenere i siti. In valutazione il costo che avrebbe INFN se supporto venisse a mancare

Evoluzione della infrastruttura INFN: altri centri



Donatella Lucchesi - INFN & Universita' Padova

Opinione personale



PON infrastruttura DHTCS puo' essere occasione per il computer center al sud

In US “Big data meets HPC” e’ una realta’ da tempo!

US **CORAL**, Collaboration of Oak Ridge Argonne and Livermore

US to Build Two Flagship Supercomputers

OAK RIDGE National Laboratory Lawrence Livermore National Laboratory

SUMMIT **SIERRA**

150-300 PFLOPS Peak Performance
 IBM POWER9 CPU + NVIDIA Volta GPU
 NVLink High Speed Interconnect
 40 TFLOPS per Node, >3,400 Nodes

2017

Major Step Forward on the Path to Exascale

Progetto da ~500 M\$ finanziato da DOE:

- 3 sistemi da 100-200 PFlops nel 2018-2019 (pre-ExaScale)
- sistema ExaScale 2023

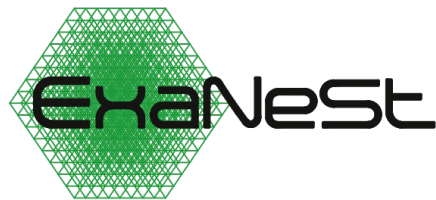
Donatella Lucchesi - INFN & Universita' Padova

In Europa EuroHPC a oggi 9 stati membri hanno firmato la dichiarazione.

EuroHPC **Joint Undertaking (JU)**, *legal and funding entity for the Union's and national resources on HPC with the aim of:*

- *acquiring and providing a world-class pre-exascale supercomputing infrastructure to Europe's scientific and industrial users, matching their demanding application requirements by 2020,*
- *developing exascale supercomputers based on competitive EU technology that the Joint Undertaking could acquire around 2022/2023, and that would be ranking among the top three places in the world.*

Verso ExaScale HPC nell'INFN



ExaNeSt

European Exascale System Interconnection Network & Storage
Sviluppa soluzioni testandole con prototipi per problemi cruciali verso ExaScale

PI: Foundation for Research & Technology, GR

INFN: P. Vicini



EuroEXA

Project ID: 754337

Co-designed Innovation and System for Resilient Exascale Computing in Europe: From Applications to Silicon

Propone co-design di una architettura bilanciata per applicazioni compute- and data- intensive

PI: Institute of communication and computer systems, GR

INFN: P. Vicini

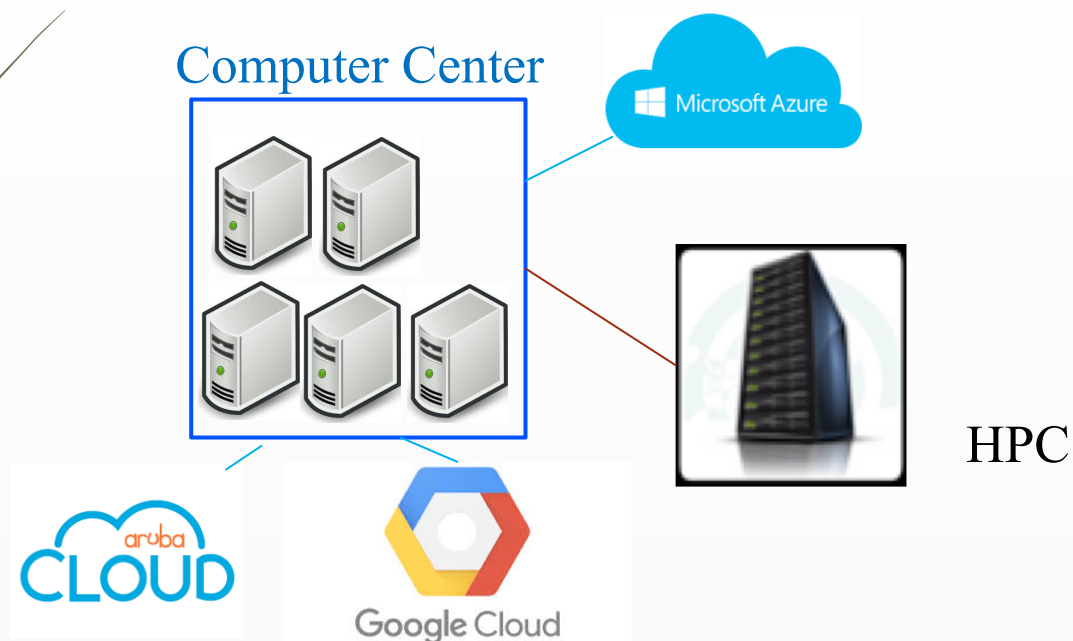
Progetto CIPE

Acquisizione di sistemi many core "next gen" e loro utilizzo per il computing HPC e HTC in ambito INFN

P. Vicini + altri

HPC per “noi” (HEP)

- In US “Big data meets HPC” HPC e’ *partial solution* per soddisfare le richieste di LHC
- HPC non puo’ esser lasciato “solo” al CINECA
- LHC, Virgo, esperimenti astroparticle fanno sempre piu’ uso di strumenti di calcolo parallelo
- I nostri Computing Centers avranno infrastruttura ibrida per poter rispondere a tutte le esigenze delle ricerca e dovranno espandersi in modo “elastico”



Collaborazione nel forum EU_T0 per sviluppare insieme strumenti necessari per accedere in modo trasparente dai nostri Tier HPC

INFN risponde

- 1) Borse Calcolo: conferite 12 borse di studio post-lauream per un anno rinnovabili
 - Novel networking techniques for data acquisition, trigger and data processing
 - Machine and deep learning for HEP in reconstruction, selection and analysis algorithms
 - New technologies in software: architecture dependent optimizations, task based frameworks, data acquisition systems
 - Innovative Workflow and Data Management solutions for trigger-less data acquisition
 - High performance data analysis and algorithms: trigger and reconstruction algorithms, use of hardware accelerators
 - Innovative Workflow and Data Management solutions for Large Scale science: large datasets, large workloads, heterogeneous resources
 - Evolution of simulation frameworks: handling of generation tools, development of fast parametrized detector simulation

- 2) Creazione di un gruppo di lavoro
 - Presentazione dei risultati ottenuti alla commissione scientifica nazionale 1 22 febbraio p.v.
 - Workshop 13 marzo a Torino (da confermare):
 - Mattino: discussione delle attività di calcolo scientifico dei borsisti per farli e farsi conoscere a tutta la nostra comunità
 - Pomeriggio: attività di calcolo scientifico HPC

ICDI: Italian Computing and Data Infrastructure

18

Iniziativa che vede al momento:

- Le infrastrutture: CINECA, GARR, INFN
- Data archives in many scientific domains: Health, BioMedicine, Agriculture, Physics, Earth Observation, Cultural Heritage, Archaeology, Geophysics, Astronomy and Astrophysics
- Data & Computing platform & services (INDIGO-DataCloud, OpenAIRE, D4Science, PRACE, EUDAT)

Osservazioni

- In Italia, come in molte altri paesi europei le risorse di calcolo e di storage sono frammentate fra diverse agenzie di finanziamento e tra le infrastrutture di ricerca
- Ciascuna infrastruttura di ricerca parla per se stessa sia in Italia che in Europa ed e' meno efficace di un'azione coordinata

Azioni

- Federare le infrastrutture esistenti e le e-infrastrutture
 - Ottimizzera' il processo di finanziamento accedendo e distribuendo in modo coerente e armonico fondi regionali, nazionali e europei
 - Rinforzera' il ruolo italiano all'interno di EOSC, di EDI, di EuroHPC e in generale con la commissione europea.

Donatella Lucchesi - INFN & Universita' Padova

Obiettivi iniziali di ICDI

Formulare congiuntamente la visione strategica del coinvolgimento della rete di ricerca italiana nell'European Open Science Cloud (EOSC) e nell'European Data Infrastructure (EDI), definire le azioni che valorizzino ed espandano il ruolo dell'Italia nell'EOSC e EDI.

Finalizzare la partecipazione italiana alle call H2020 e a future call presenti nel programma quadro, FP9, per la realizzazione di un piano coerente e pienamente rappresentativo per l'evoluzione e lo sviluppo delle infrastrutture di calcolo italiane

Favorire il massimo ritorno di sistema della partecipazione da parte di alcuni membri della ICDI ai progetti EOSCPilot, EOSCHub in corso di svolgimento.

Condividere con i rappresentanti del MIUR gli elementi riguardanti la partecipazione nazionale alle attività europee in modo che i progetti già in corso di svolgimento o la partecipazione italiana a progetti o accordi intergovernativi possano consentire il massimo del ritorno in termini di capacità di attrarre finanziamento comunitario, guidare progetti strategici, favorire l'attrattività e la competitività del sistema scientifico e dell'innovazione.

ICDI stato attuale

- ❑ Memorandum di Intesa in fase di scrittura il cui *scopo* è un accordo tra le Parti per collaborare in aree di mutuo interesse e per portare avanti le attività che sono necessarie per raggiungere gli obiettivi dell'associazione che sono descritti nell'allegato 1.
- ❑ Memorandum contiene
 - Obiettivi iniziali
 - Governo
 - Risorse disponibili per le single RI
 - Attivita' iniziali
- ❑ Formato un executive board ad-interim
- ❑ Sito web in costruzione
- ❑ Mailing list icdi@garr.it raccoglie le persone interessate
- ❑ In preparazione Position Paper su EOSC, altri position paper sono possibili

Per concludere

- Calcolo scientifico in rapida evoluzione, cambiamenti importanti:
 - nel software per analisi e gestione dati
 - nel middleware
 - nell'infrastruttura

- Distinzione tra HTC e HPC sempre piu' flebile, si sta passando dall'era HTC e HPC all'era del "scientific computing"

- Importanti iniziative per la definizione di una infrastruttura "federata" di calcolo per la ricerca in Europa e in Italia