



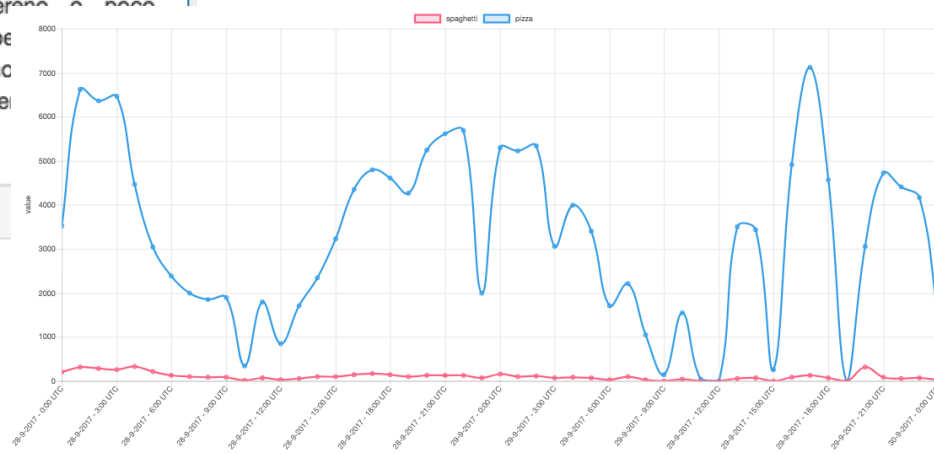
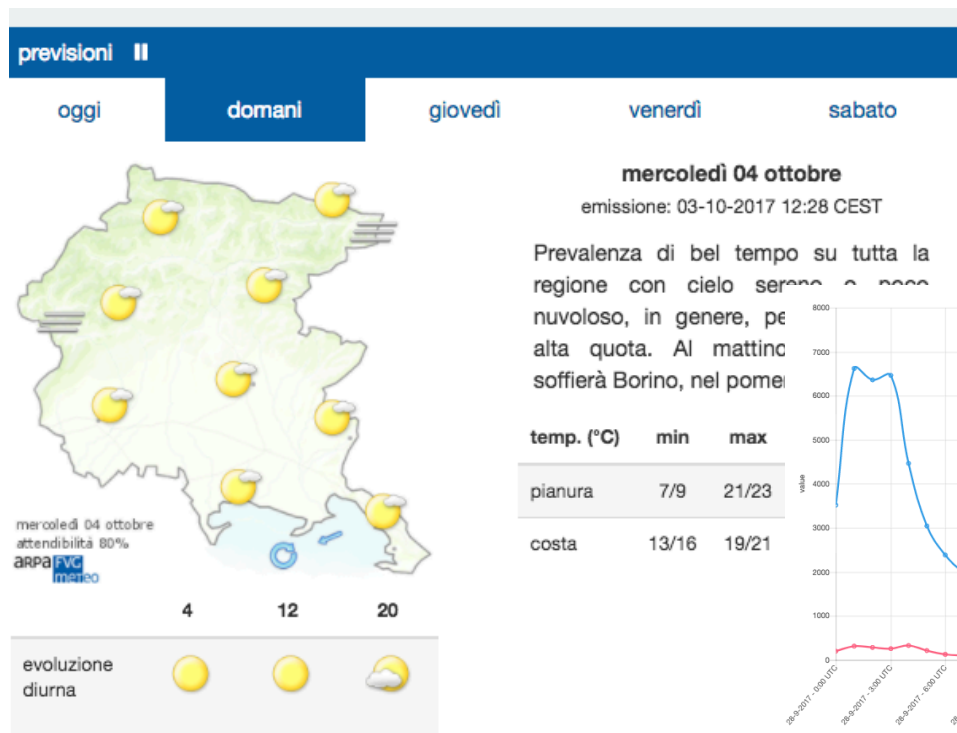
IL CALCOLO AD ALTE PRESTAZIONI: UNA TECNOLOGIA ABILITANTE PER LA SOCIETÀ SMART

Relatore: Stefano Cozzini

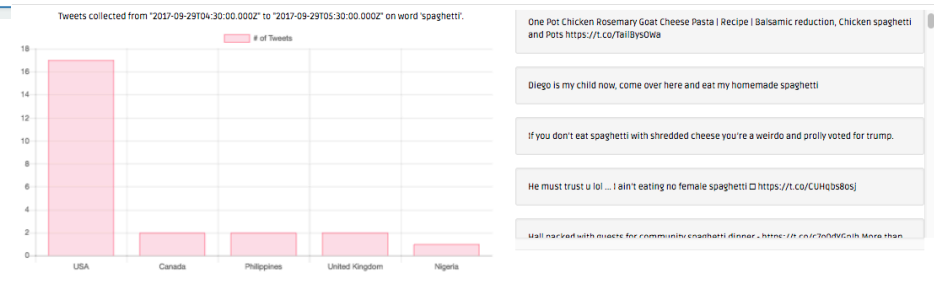
4 ottobre 2017 ore 14:30

FP1640985001
#Sharing3FVG

Cosa hanno in comune queste figure?



Entrambe sono generate grazie all'HPC...



Agenda

- Premessa: Chi sono / Chi siete
- Introduzione:
 - Perchè il calcolo ad alte prestazioni
- concetti di base dell'HPC
- Il mondo dei supercomputer
- HPC per tutti: l'approccio cloud...
- 2 casi di studio:
 - deep learning per le immagini (HPDA)
 - Simulazione CFD per le NAVI (HPC)



Chi sono io ?

- Background in fisica computazionale
- Lavoro al CNR/IOM con il compito di mettere a punto infrastrutture di calcolo e dati per la scienza
- Ho fondato una start-up(eXact-lab srl) che si occupa della stessa cosa per il mercato

HPC: Serve davvero?

Perchè l'HPC è importante ?

“The next 10 to 20 years will see *computational science* firmly embedded in the fabric of science – the most profound development in the scientific method in over three centuries” (US Department of Energy).

“A *host of technologies* are on the horizon that we cannot hope to understand, develop, or utilize without simulation” (US National Science Foundation)

Dal sito della EU

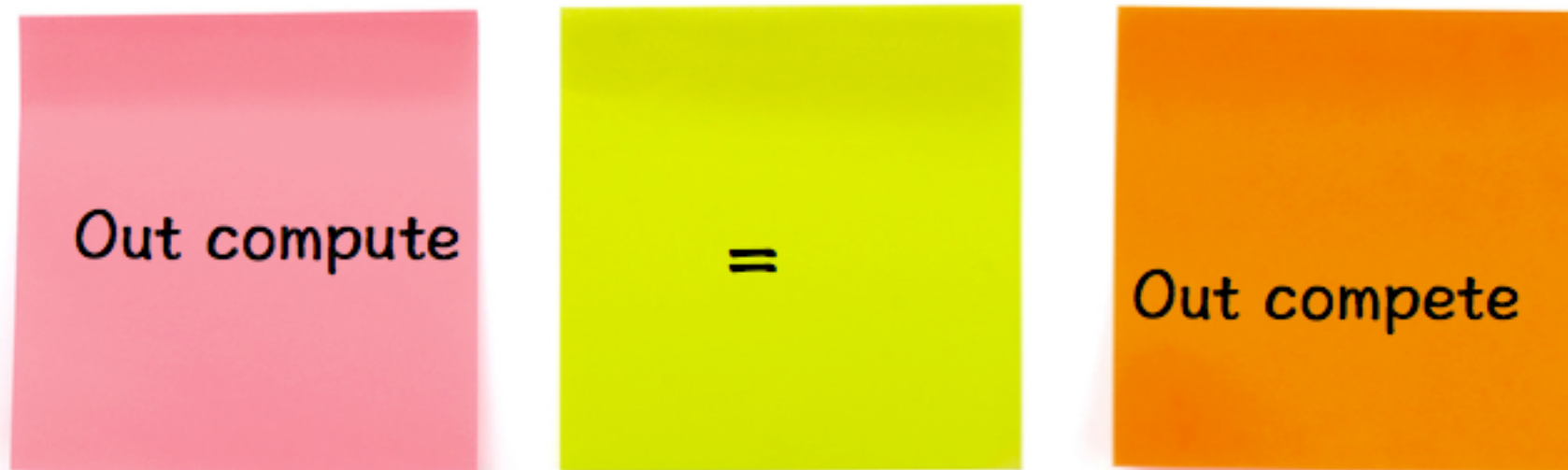
A Strategy for Research and Innovation through High Performance Computing



KEY ENABLING TECHNOLOGY

From <https://ec.europa.eu/programmes/horizon2020/en/h2020-section/high-performance-computing-hpc>

Perchè HPC ?



“Today, to Out-Compute is to Out-Compete”.

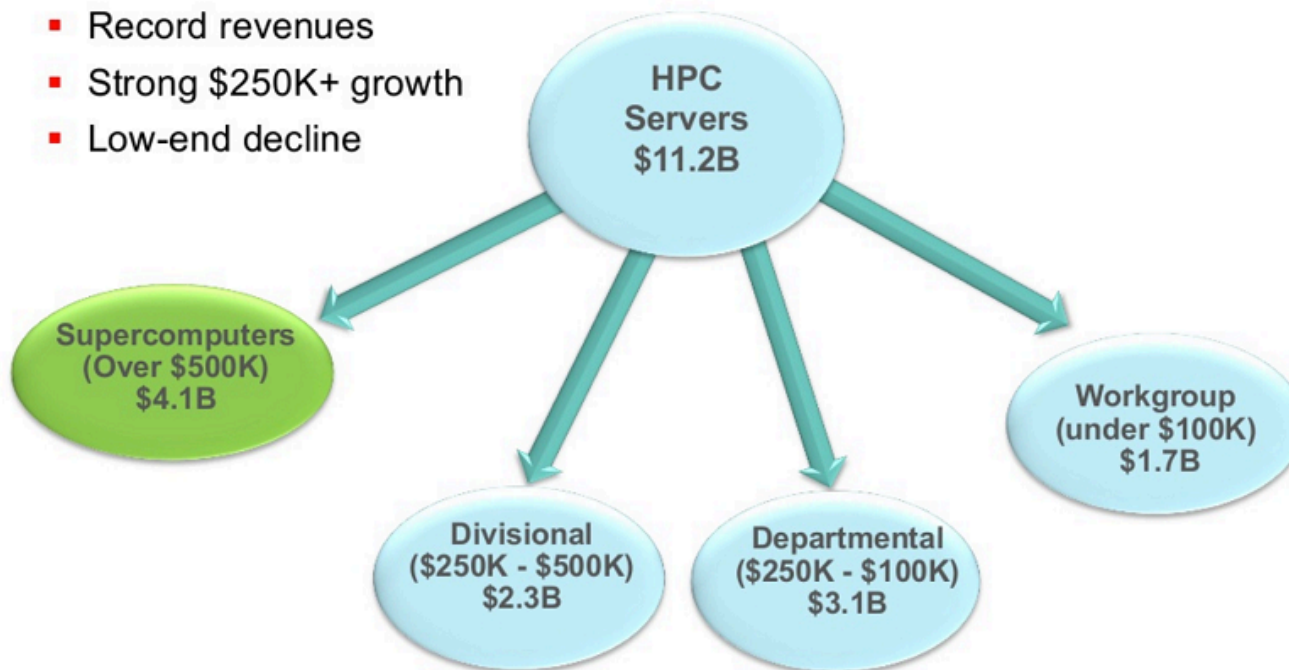
Il 97% delle aziende che ha iniziato a utilizzare i supercomputer ha riportato che senza di essi non potrebbe più competere o addirittura sopravvivere. (IDC survey 2017)

Image from UberCloud

Un mercato ampio ed in crescita..

The Worldwide HPC Server Market: \$11.2 Billion in 2016

- Record revenues
- Strong \$250K+ growth
- Low-end decline



Source: Hyperion Research

HPC: definizioni e fondamenti

HPC: una definizione

Defining HPC (from Intersect survey)

- High Performance Computing (HPC) is the **use of servers, clusters, and supercomputers** – plus associated software, tools, components, storage, and services – for **scientific, engineering, or analytical tasks** that are particularly intensive in computation, memory usage, or data management
- HPC is used by scientists and engineers both in research and in production across **industry, government** and **academia**.

Elementi fondanti dell'HPC..

- use of servers, clusters, and supercomputers → HW
- associated software, tools, components, storage, and services → SW
- scientific, engineering, or analytical tasks
→ PROBLEMS TO BE SOLVED..

ALL THE ABOVE DEFINES A
COMPUTATIONAL INFRASTRUCTURE
aka E-INFRASTRUCTURE

Elements of e-infrastructures for HPC

E-infrastructure for HPC includes:

- servers/nodes/accelerators
- High speed Networks
- High end parallel storage
- Middleware
- Scientific/Technical Software
- Research/Technical data (national and scientificv databases, individual data...)

IS ALL WHAT WE NEED ?

Last but not least: persone !

- Human capital is by far the most important aspect
- Two important roles:
 - HPC providers (plan/install/manage HPC resources)
 - HPC user

MIXING/INTERPLAYING ROLES
INCREASES COMPETENCE LEVELS

HPC Users...

High Performance Technical Computing (HPTC)

- Applications in science and engineering
- Top Market: Academia/Government Lab/ Manufacturing/ Bio Life/ Oil gas exploration

High Performance Business Computing (HPBC)

- Application included trading/pricing/ risk management/ online gaming/ analytics/fraud detection/ logistics
- Top Market: Banks/ Insurance Companies/ Online games/ Financial services/ Entertainment

A corto di esperti in HPC ? (1)

International Master in HPC

- Master di specializzazione di un anno
- Promosso da Sissa ICTP and CNR
- Quarta edizione appena partita/terza edizione in fase di arrivo
- Studenti delle prime 2 edizioni lavorano tutti !



WWW.MHPC.IT

A corto di esperti in HPC ? (2)

Master in Data science and scientific computing

Master di specializzazione di un anno

- Laurea di master di 2 anni per studenti della triennale
- Promosso da UNiTS/UNIUD Sissa ICTP
- Prima edizione partita questa settimana



DSSC.UNITS.IT

Ordini di grandezza per l'HPC

MEGA = 10^6

GIGA = 10^9

TERA = 10^{12}

PETA = 10^{15}

EXA = 10^{18}

Unità di misura per HPC

- Velocità dei processori:
Floats: floating point operation/ second
- Velocità delle rete:
bits : bit /second transmitted
– /100Mbit/1Gbit/10Gb/100Gb
- Quantità di spazio:
byte
 - Gigabite -----> RAM/hard disks
 - Terabyte -----> Disks/SAN ...
 - Petabyte -----> SAN

HPC: supercomputers e dintorni

Quali sono i computer più potenti ?

- TOP500 list www.top500.org
- published twice a year from 1993
 - ISC conference in Europe (June)
 - Supercomputing conference in USA (November)
- List the most powerful computers in the world
- Yardstick: Linpack benchmark (LU – decomposition)



And the winner is..

Sunway TaihuLight - Sunway MPP, Sunway SW26010 260C
1.45GHz, Sunway

Site:

National Supercomputing Center in Wuxi



Numero di cores 10,649,600
Flops 93 PETAFLIPS
Energia consumata 15 Mwatt

La prima macchina europea..

Site:

Swiss National Supercomputing Centre (CSCS)

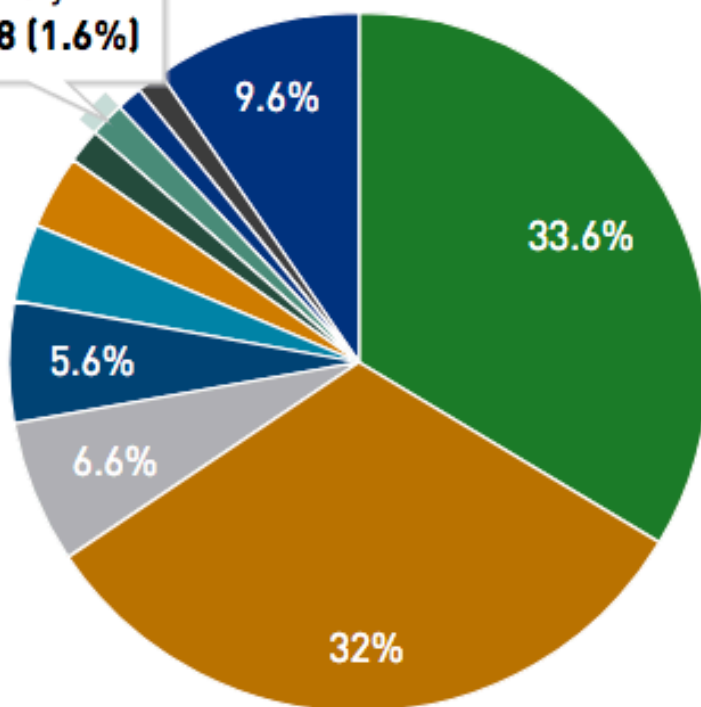


Numero di cores 10,649,600
Flops 93 PETAFLOPS
Energia consumata 15 Mwatt

E l'Italia ??

Countries System Share

Italy
8 (1.6%)



- United States
- China
- Japan
- Germany
- France
- United Kingdom
- Korea, South
- Italy
- Canada
- Poland
- Others

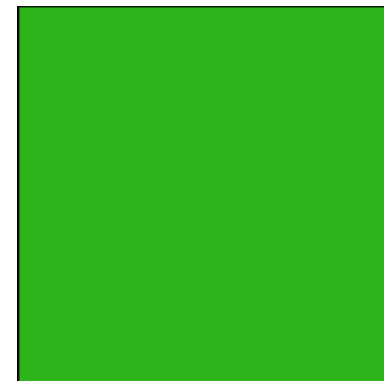
IOPS vs FLOPS

- HPC oggi è *compute-centric*
- Il Calcolo Tecnico Scientifico e non solo ha bisogno di accedere ai dati piuttosto che velocità di calcolo

computing 1 calculation
≈ 1 picojoule



moving 1 calculation
≈ 100 picojoule



Source: IDC Direction 2013

Il problema dell'energia..

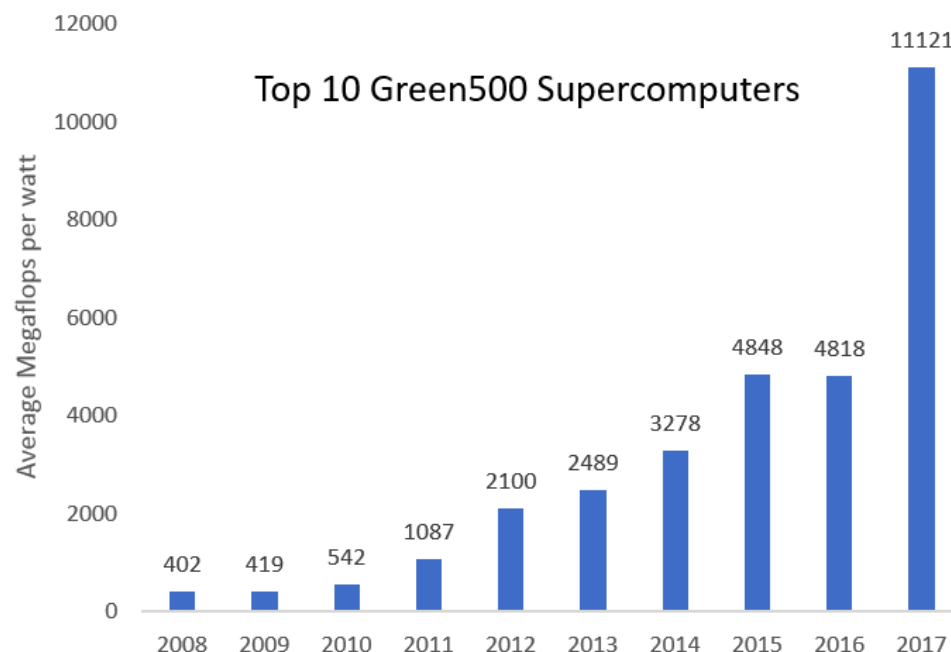
- È impossibile pensare di raggiungere l'exascale con le macchine attuali..
- I supercomputers devono diventare più green..



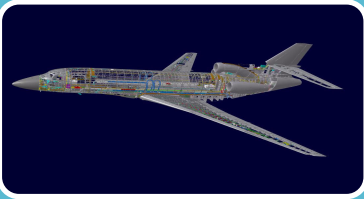
Top500&Green500



- Over the last year, the greenest supercomputers more than doubled their energy efficiency
- If such a pace can be maintained, exascale supercomputers operating at less than 20 MW will be possible in as little as two years.
- But that's a big if.

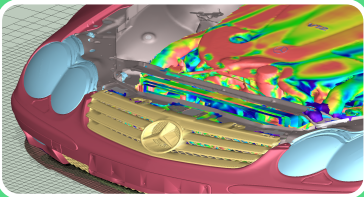


Where HPC plays a role in Industry ? Some examples..



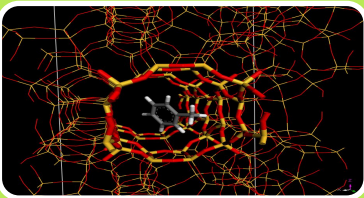
Aeronautics

where the design of airplanes more energy efficient and less noisy cannot be done without simulations involving very large models of the entire aircraft and analysis of physical phenomena at different scales,



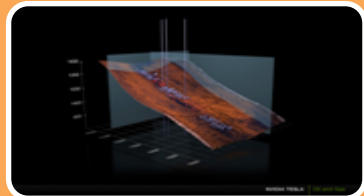
Automotive industry

which wishes to reduce consumption and CO2 emissions while increasing the level of comfort and security,



Pharmaceutical industry

where the discovery of new active molecules and new drugs is accelerated by numerical simulations,

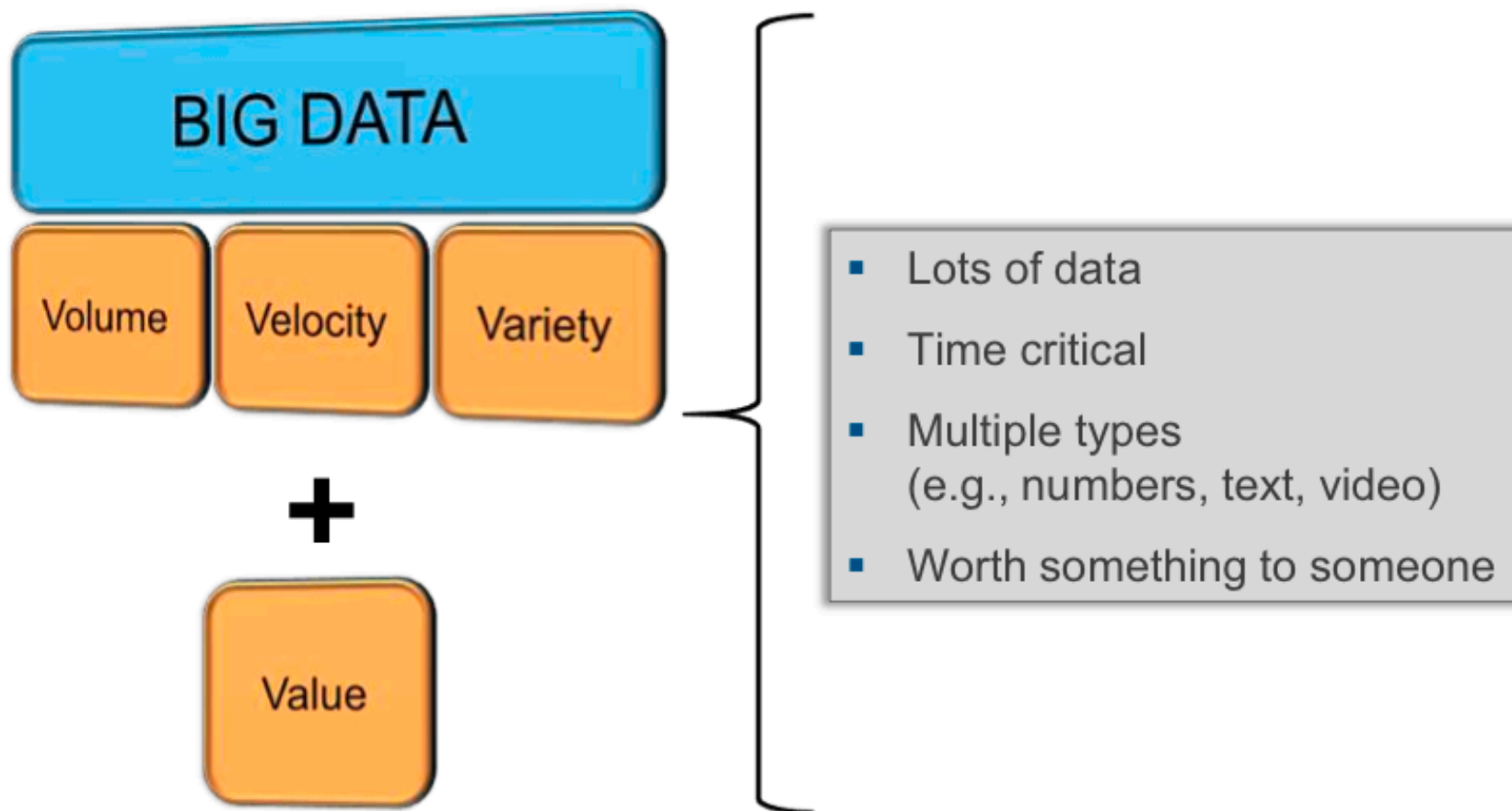


Oil industry

which needs supercomputers to discover new oil fields and to optimize production of existing reservoirs.

Il problema dei DATI...

Big Data: una definizione



The 3 V's of big data..

- **Velocity**

Data are produced at speed higher than the speed you are able to move/analyze and understand them..

- **Variety**

- Data range from simulation to remote sensing information, from instruments to market analysis etc..
- datasets come in a variety of data formats and span a variety of metadata standards

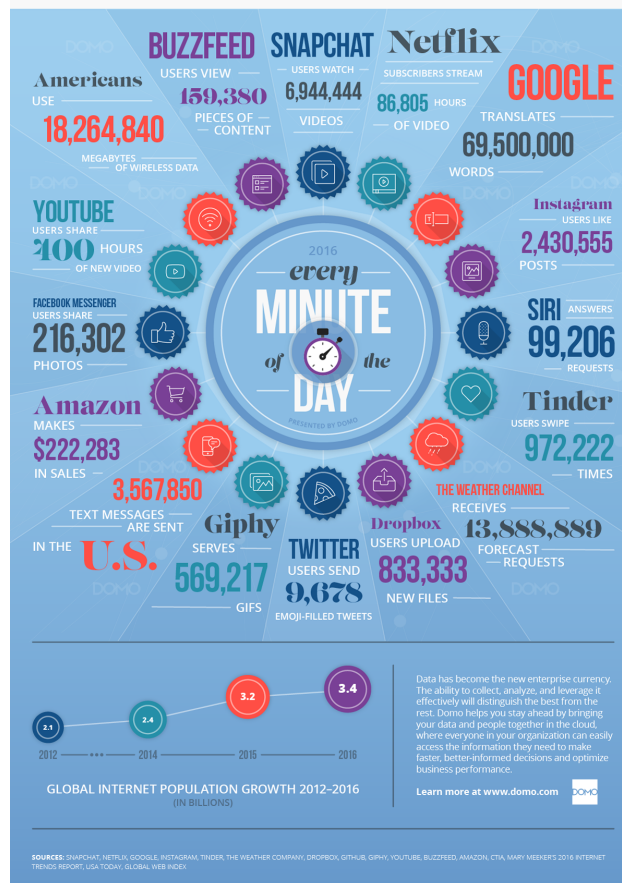
- **Volume**

Infografica

DOMO

DATA NEVER SLEEPS 4.0

How much data is generated every minute? In the fourth annual edition of Data Never Sleeps, newcomers like Giphy and Facebook Messenger illustrate the rise of our multimedia messaging obsession, while veterans like YouTube and Snapchat highlight our insatiable appetite for video. Just how many GIFs, videos, and emoji-filled Tweets flood the Internet every minute? See for yourself below.



How Much Data is Produced Every Day?



2.5 Exabytes are
are produced
every day

Which is equivalent to:

-  530,000,000 millions songs
-  150,000,000 iPhones
-  5 million laptops
-  250,000 Libraries of Congress
-  90 years of HD Video

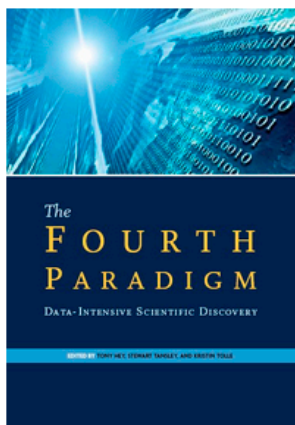
 LEVEL

Data-intensive science

- A “fourth paradigm” after experiment, theory, and computation..

The Fourth Paradigm: Data-Intensive Scientific Discovery

Presenting the first broad look at the rapidly emerging field of data-intensive science



Increasingly, scientific breakthroughs will be powered by advanced computing capabilities that help researchers manipulate and explore massive datasets.

The speed at which any given scientific discipline advances will depend on how well its researchers collaborate with one another, and with technologists, in areas of eScience such as databases, workflow management, visualization, and cloud computing technologies.

In *The Fourth Paradigm: Data-Intensive Scientific Discovery*, the collection of essays expands on the vision of pioneering computer scientist Jim Gray for a new, fourth paradigm of discovery based on data-intensive science and offers insights into how it can be fully realized.

Critical praise for *The Fourth Paradigm*

Download

- [Full text, low resolution](#) (6 MB)
- [Full text, high resolution](#) (93 MB)
- [By chapter and essay](#)

Purchase from Amazon.com

- [Paperback](#)
- [Kindle version](#)

In the news

- [Sailing on an Ocean of 0s and 1s](#) (*Science Magazine*)
- [A Deluge of Data Shapes a New Era in Computing](#) (*New York Times*)
- [A Guide to the Day of Big Data](#) (*Nature*)



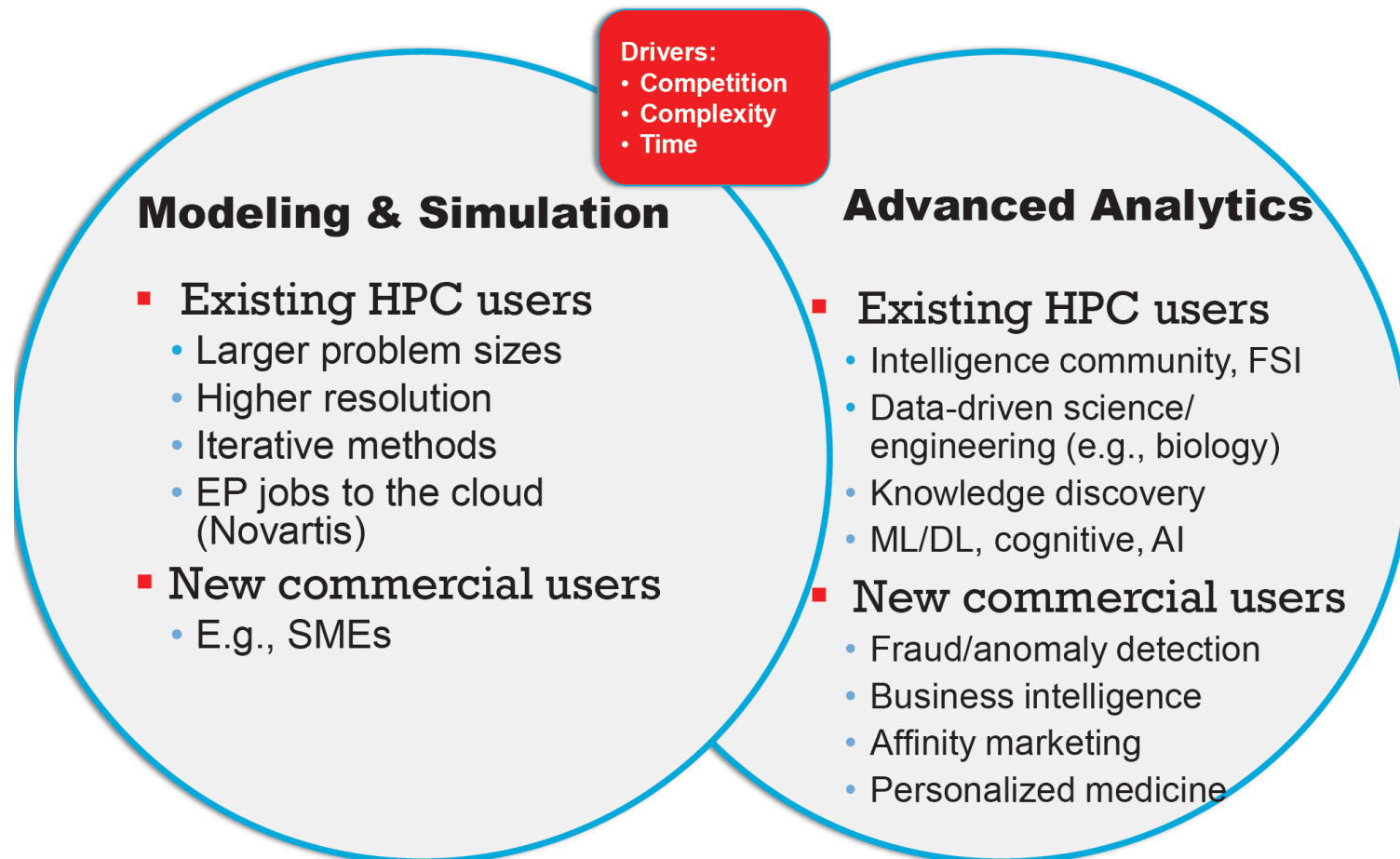
It involves collecting, exploring, visualizing, combining, subsetting, analyzing, and using huge data collections

Big data challenges

- “*BigData*” is a growing trend affecting many HPC applications touching large datacenters, and research
- Fueled by creation and availability of many data
- Organizations seek to bridge the gap between having better data and making better decisions
- Application areas include: enterprise analytics, research analytics, real-time analytics, complex event processing, data mining, visualization ...
- Growth in these application areas creates a market opportunity for providers of HPC technologies

From HPC to HPDA..

HPDA = Data-Intensive Computing Using HPC



©Hyperion Research 2017

13

HPC:
ma quanto mi
costi?

How much does it cost a computational infrastructure?

It is not just a matter of HW...

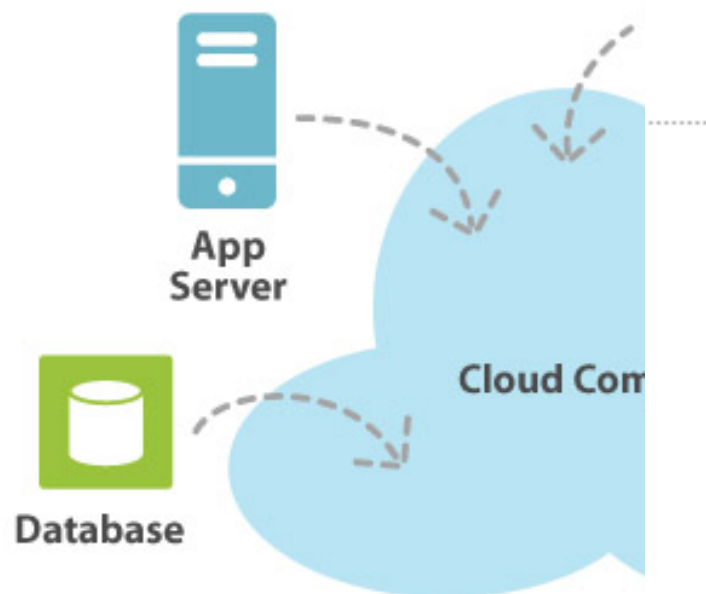
Total Cost of Ownership is the right way to calculate the budget for an HPC infrastructure..

What should be included in the TCO for HPC ?

- Investment, operation and maintenance costs:
 - Hardware: servers, storage, networking, cabling, etc.
 - Electrical equipment: power distribution units, UPS, generators, etc.
 - Cooling systems: air conditioners, water cooling, etc.
- Infrastructure for the data center, power adaptation issues, etc.
- Energy consumption of the hardware and cooling systems
- Software licenses
- Human resources
- Maintenance

HPC solo per
pochi?

Cloud Computing..



Cloud Computing
everything and the kitchen sink

TECNOLOGIA



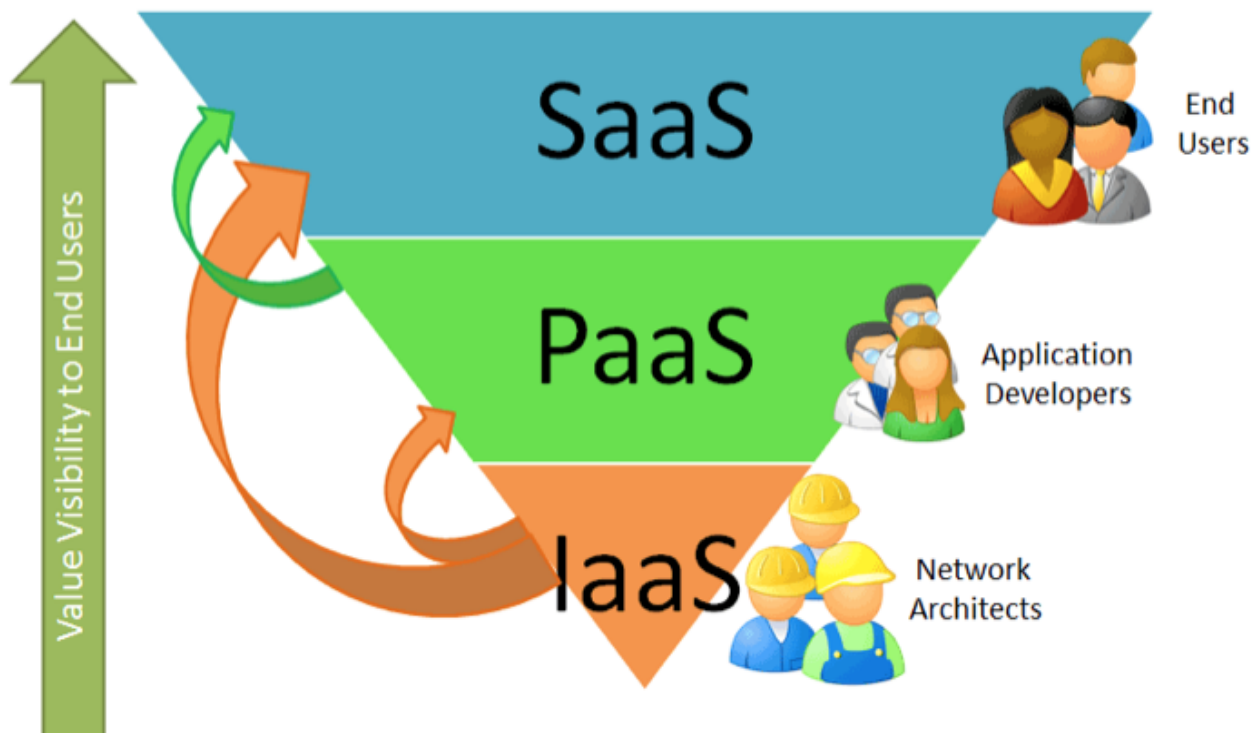
Ora l'Europa punta sul "cloud computing"

Ecco l'agenda digitale Ue:
con standard comuni, obblighi
chiari e compatibilità provider
sono previsti 2,5 milioni
di posti di lavoro entro il 2020

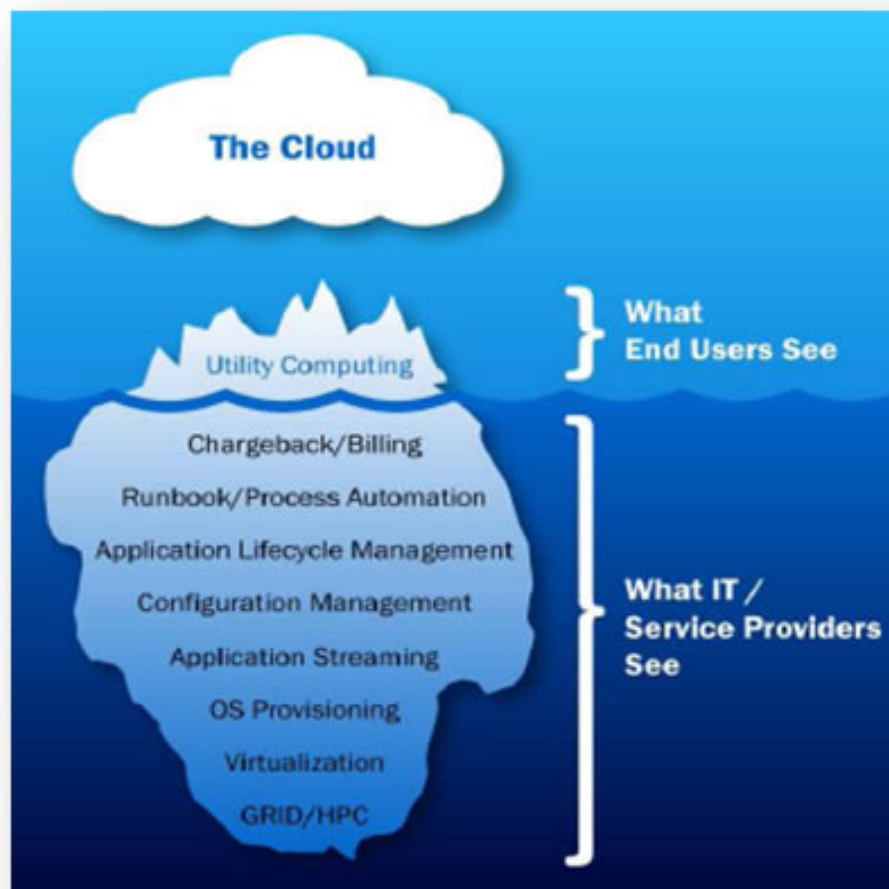
Basic Definition

- A model of computation and data storage based on “pay as you go” access to “unlimited” remote data center capabilities
- A cloud infrastructure provides a framework to manage scalable, reliable, on-demand access to applications
- Cloud services provide the “invisible” backend to many of our mobile applications
- High level of elasticity in consumption
- Historical roots in today’s Internet apps
 - Search, email, social networks
 - File storage (Live Mesh, Mobile Me, Flickr, ...)

How far are from end users?



Why the CLOUD ?



CLOUD COMPUTING PARADIGM = HPC RESOURCES for MASSES



www.c3hpc.com



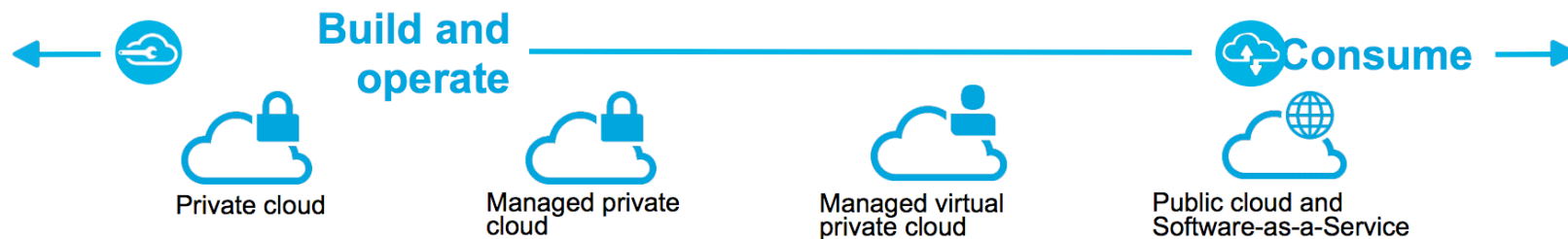
CARNIA
INDUSTRIAL
PARK

- L'infrastruttura HPC di eXact lab
- Basata su Eurotech Aurora system
- Raffreddata ad acqua: alta efficienza energetica



In linea con Industria 4.0 : servizi cloud avanzati per favorire la competitività della piccola media impresa

HPC CLOUD in eXact lab: Openstack



OpenStack as common Infrastructure Abstraction

2 CASI DI STUDIO

SIMULAZIONI HPC per
la fluidodinamica
navale



Riconoscimento di
immagini SEM di
nanoscienza



OpenViewSHIP

Sviluppo di un ecosistema computazionale per la progettazione idrodinamica del sistema elica-carena.

OpenViewSHIP

Il progetto intende sviluppare un ecosistema computazionale per l'ambito industriale dove affrontare la progettazione idrodinamica del sistema elica-carena. Basato su infrastrutture di calcolo ad alte prestazioni integrate con tecnologie innovative di visualizzazione remota, l'ecosistema permetterà la visualizzazione immediata del pre e post-processing delle grandi mole di dati prodotti dalle simulazioni CFD ad alta risoluzione. L'ecosistema sviluppato risolverà le criticità evidenziate nel progetto OpenSHIP e ne sarà promossa l'adozione presso l'industria.

DATI PROGETTO

<i>Tipologia di attività</i>	Ricerca industriale e sviluppo sperimentale
<i>Data inizio progetto</i>	1 marzo 2014
<i>Durata</i>	36 mesi
<i>Costo del progetto</i>	500.000,00 €

- Area riservata »
- Scarica brochure del progetto »

eXact

DISTRETTO TECNOLOGICO NAVALE E NAUTICO
DITENAVE
FRIULI VENEZIA GIULIA

UNIVERSITÀ
DEGLI STUDI DI TRIESTE

ma per seguir virtute e canocchia
SISSA

CETENA
CENTRO PER GLI STUDI
DI TECNICA NAVALE

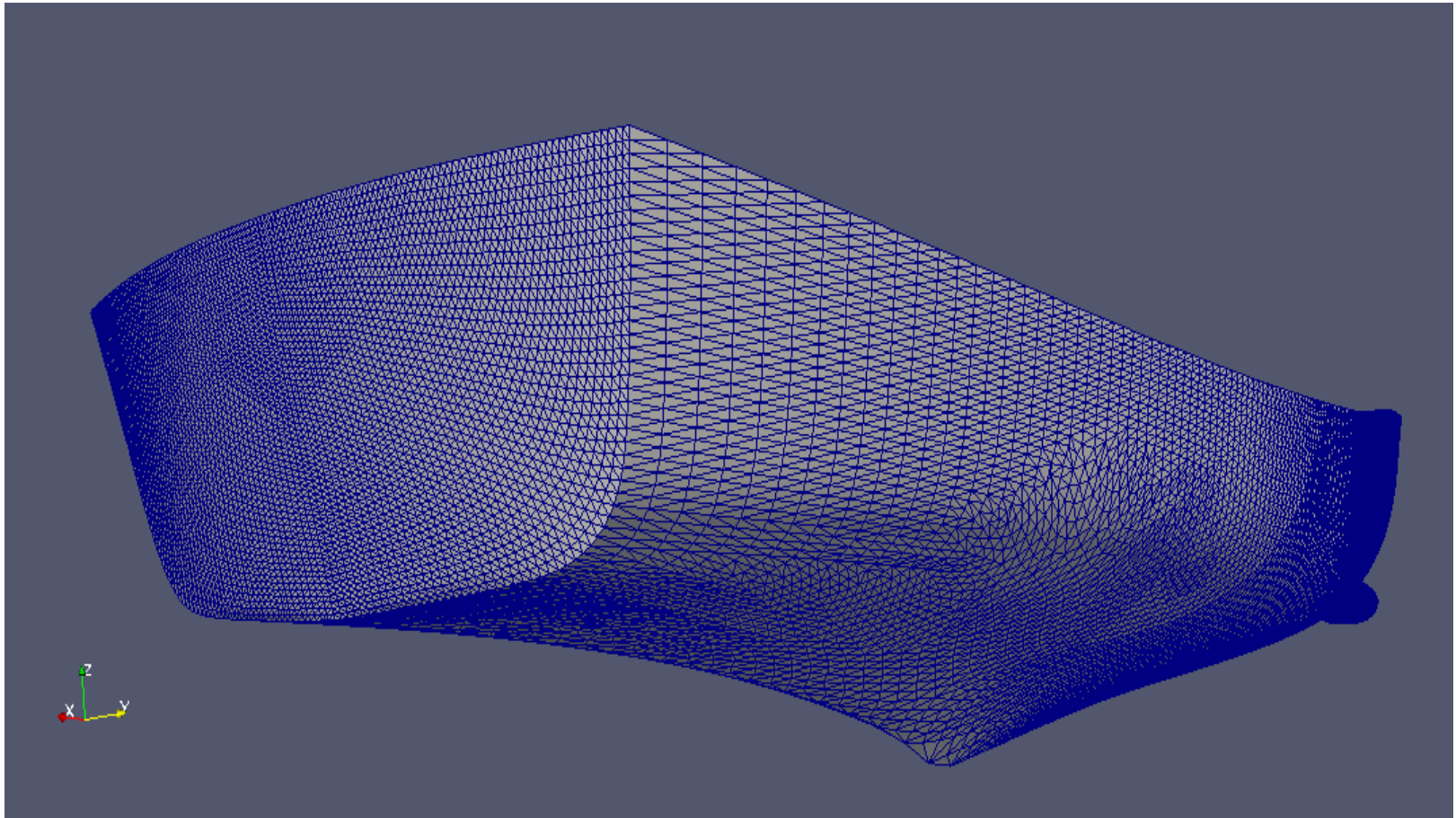
Spring
Firm S.r.l.

E quali simulazioni industriali si fanno ?



Ottimizzazione di queste carene porta ad evidenti vantaggi

La Mesh



Workflow di simulazione (originale)

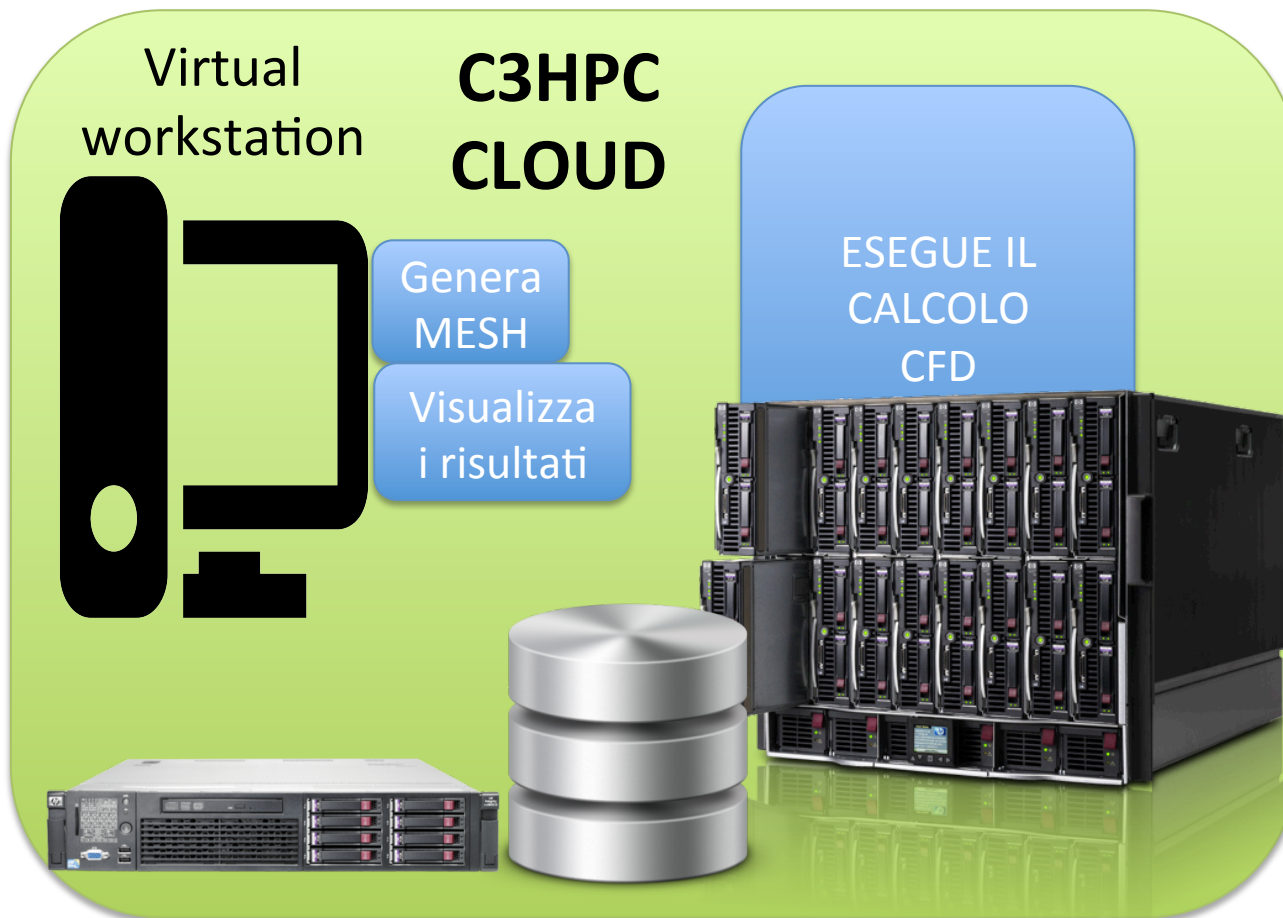
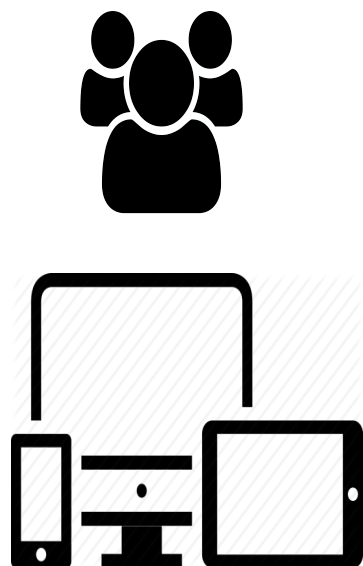
WORKSTATION



PIATTAFORMA HPC



Infrastruttura OpenViewSHIP



Infrastruttura OpenViewSHIP

Workstation virtuali

- GPU enabled!
- Elastiche in dimensione
- Accessibile da remoto con ogni device

Virtual workstation

**C3HPC
CLOUD**

IL SISTEMA HPC

- Linux based
- Nessun interfaccia grafica

NODO GRAFICO **VIZNODE**

- Cloud based
- Doppia GPU

Risorse disco
CONDIVISE

VizNode per la visualizzazione remota

- Sviluppato da eXact in collaborazione con Nvidia
- Testato dai partner OpenViewSHIP
- Disponibile in modalità Cloud su C3HPC.com
- Intera soluzione open source
- La componente a pagamento si limita al software di visualizzazione remota
 - Prodotti commerciali: Nice DCV e RGS (HPE)



Algoritmi Deep Learning per la classificazione di immagini..

- Problema da risolvere:
 - Classificare le immagini di 10 SEM del progetto NFFA
 - Solo a Trieste: 150.000 immagini !
 - Salvare le immagini in un “data repository” per permetterne l’uso e la ricerca da parte di altri utenti



SOLUZIONE: Deep Learning su piattaforma HPC

Classifying SEM images by hand

- We created and manually annotated the first sample of classified SEM images (for a total of 18,577 images).
- The classified SEM image are stored on the KITDM as a standalone

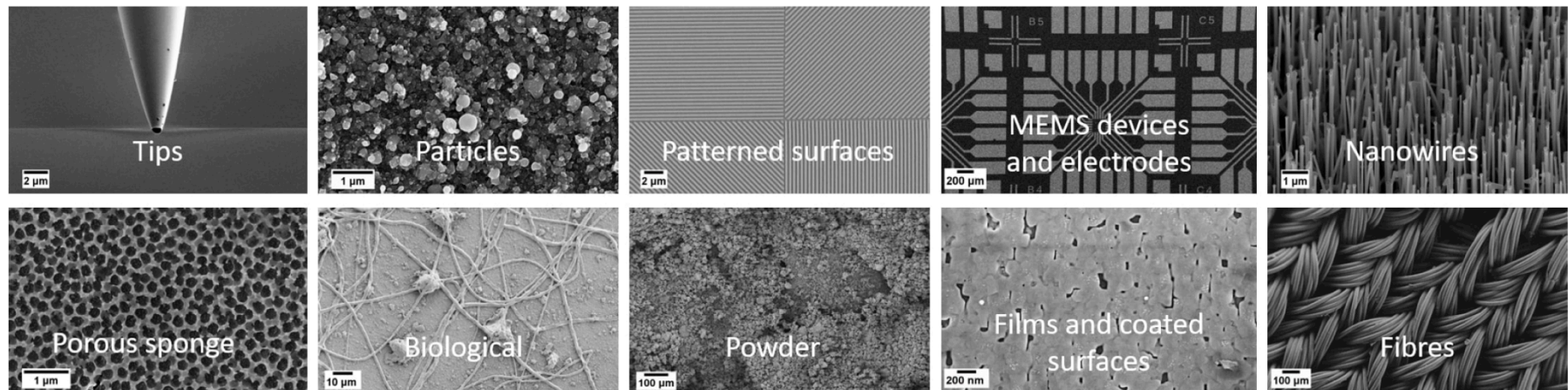
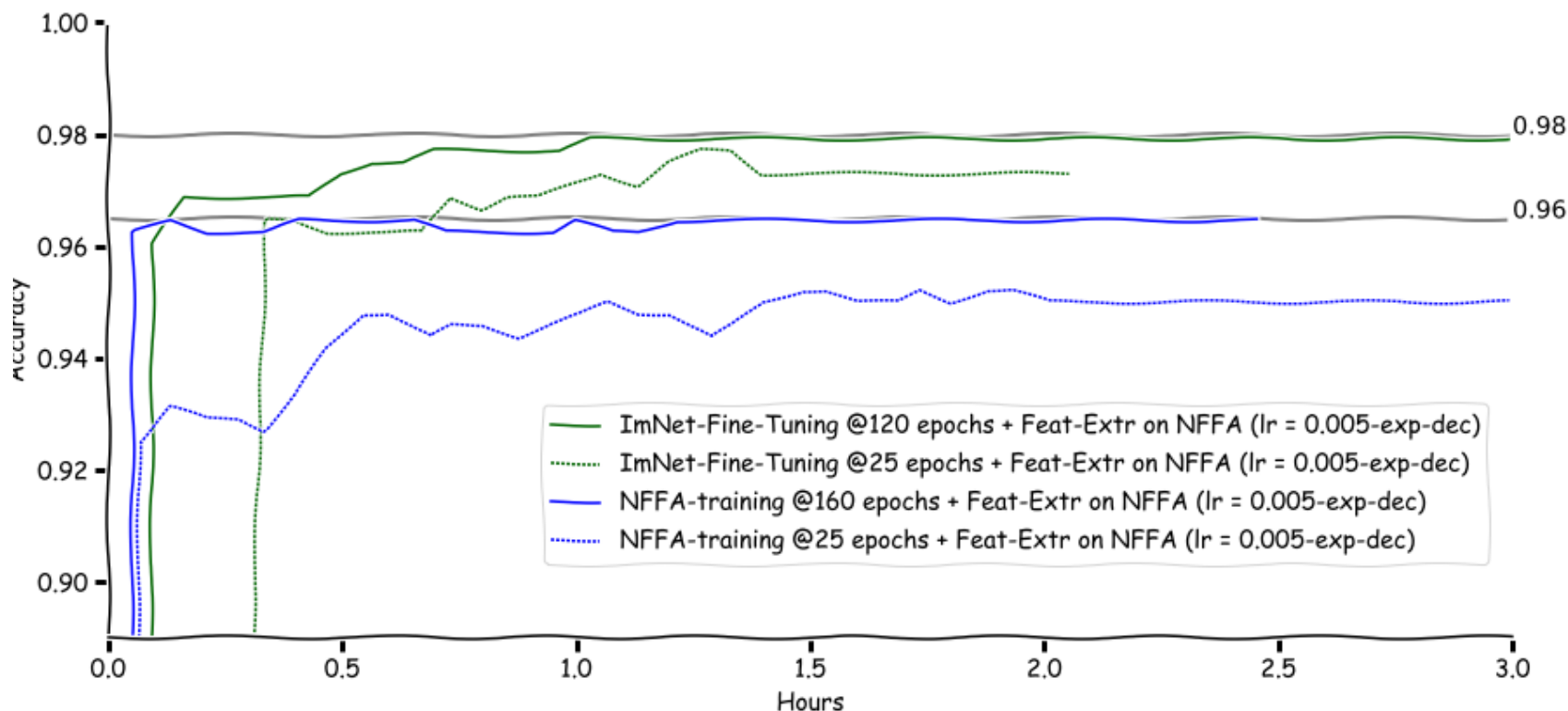
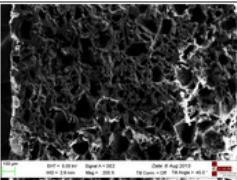
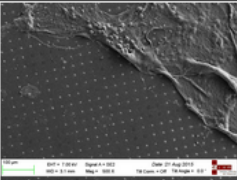


Figure 1. Categories chosen for SEM images. The dimensionality of nanoscience objects provided the basis for the choice. Other categories, such as Biological and Tips were added as these were common images found in the SEM database.

Then train the computer..



And then make the computer work for us...

Image	Nanowires	MEMS devices electrodes	Biological	Films Coated Surface	Porous Sponge	Patterned surface
	0.2% ☐	4.1% ☐	0.03% ☐	0.9% ☐	93.6% ☒ Predicted	0.31% ☐
	00.07% ☐	0.00% ☐	99.9% ☒ Predicted	0.00% ☐	0.00% ☐	0.00% ☐



Il workflow completo..



Qualche considerazione

- È indispensabile una infrastruttura di calcolo HPC per gestire le tecniche di deep learning in tempi ragionevoli.
- Tali tecniche possono essere applicate a tanti altri dataset in altri ambiti
- In corso di stesura un progetto per l'analisi di immagini in ambito medico.

Conclusioni

- HPC è strumento per gestire/indagare/capire la enorme mole di dati prodotti da una società moderna..
- HPC è strumento fondamentale per l'innovazione
- Mettere a disposizione un ambiente computazionale completo non è semplice..
- il paradigma cloud può aiutare ad abbassare la barriera iniziale per SME.
- eXact lab sta lavorando in questa direzione in collaborazione con importanti realtà regionali

Grazie per l'attenzione

Stefano Cozzini

CNR/IOM Democritos
www.democritos.it

eXact lab srl

www.exact-lab.it

stefano.cozzini@exact-lab.it