CHERNE 2019 -15th
Workshop on European Collaboration in Higher Education on Radiological and Nuclear Engineering and Radiation Protection

NEW FRONTIERS IN COMPUTING AND DATA ANALYSIS. THE EUROPEAN PERSPECTIVES

HPC-HTC and Cloud in the era of BigData

U. Becciani, C. Petta

INAF – OACT, Phys. Dep. Unict

Portopalo di Capo Passero
2-5 June 2019

Main Topics

General Considerations

This is not an epoch of changes but the start of a change of epoch.

- Infrastructures are **rapidly evolving** in a new generation of more and more powerful computing elements
- Software and analysis tools **must be re-engineerized** to be still a good instrument for the new generation of Computing infrastructure

➔ Big Data

- **The new challenges of large experiments**
- **New tools for data analysis: Data Mining and Artificial Intelligence**
- **Numerical Simulations**

➔ Convergence of different scientific fields

- **an example: SKA and HL-LHC**

➔ New European Programs

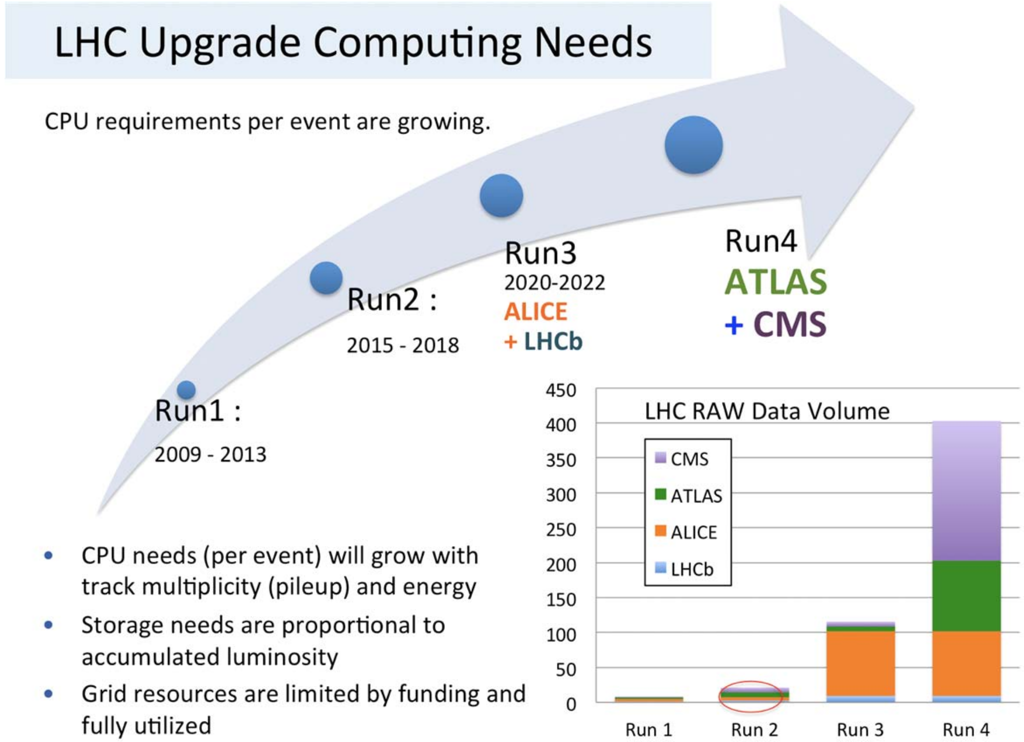
- **HPC and new Generation system towards the Exascale**
- **Large Cloud Computing for Research: EOSC platform**

Big Data era: HEP & LHC

BigData and Computing Challenges in Nuclear Physics

The LHC Run1 (2009-2013) and the first years of LHC Run2 (2015-2018) have convinced physicists that ***their codes need fundamental re-engineering to address the realities of future commodity***, highly parallel processors, and that, if they can achieve this re-engineering, they could emerge with codes ready to exploit ***High Performance and Supercomputing facilities*** quite well.

ALICE and ATLAS will collect a factor of 10-100 more data during the next 3-5 years



ATLAS – from 2022

Simulation, and data analysis:

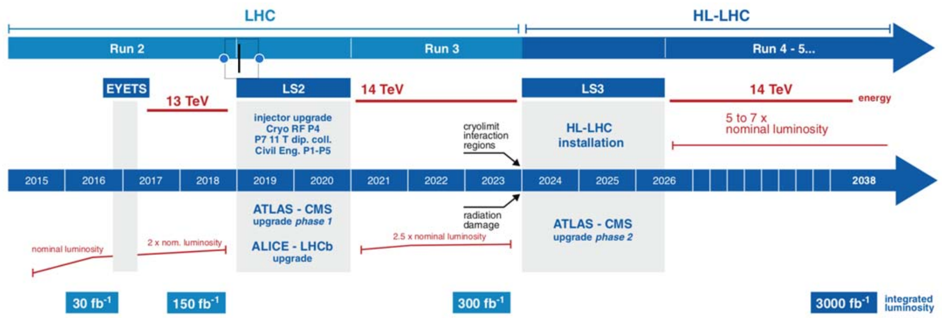
➔ computing: geographically distributed resources of approximately 150,000 cores continuously and over 1 billion core/hours per year

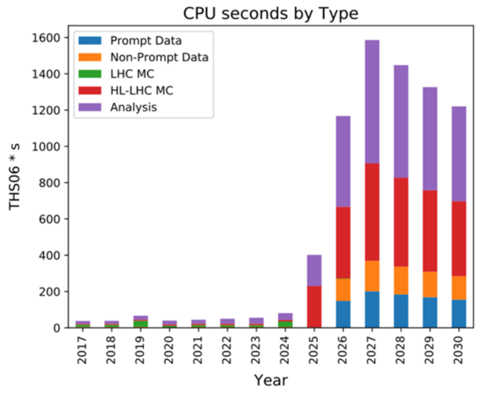
➔ data volume: more than 260 PB.

Big Data: WLCG & HL-LHC

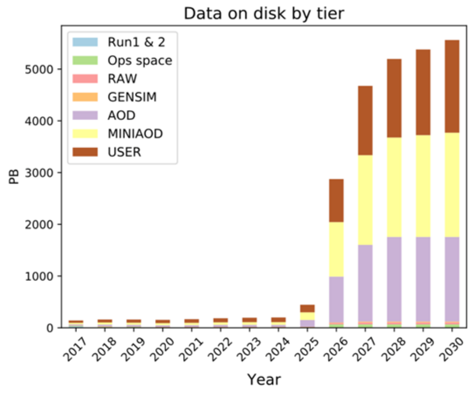
Worldwide LHC Computing Grid (WLCG) HEP WHITE PAPER: Roadmap for HEP Software and Computing R&D for the 2020s

An important message of this report is that ***a "software upgrade" is needed to run in parallel with the hardware upgrades planned for the HL-LHC*** in order to take full advantage of these hardware upgrades and to complete the HL-LHC physics programme.





(a)

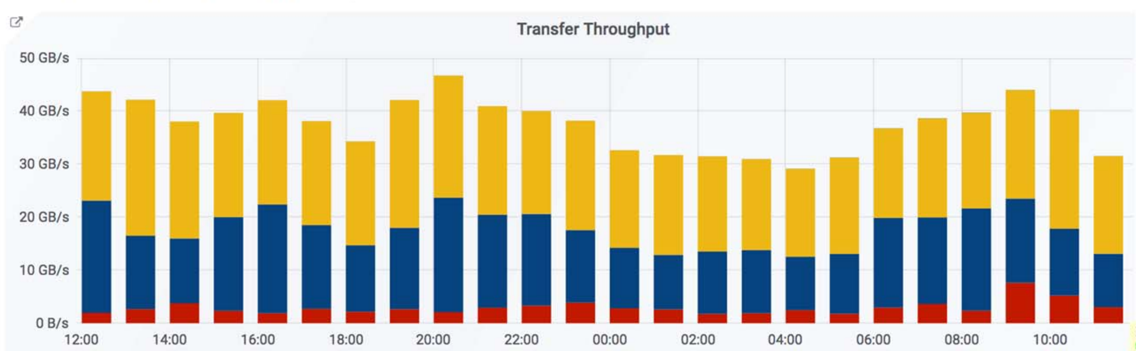


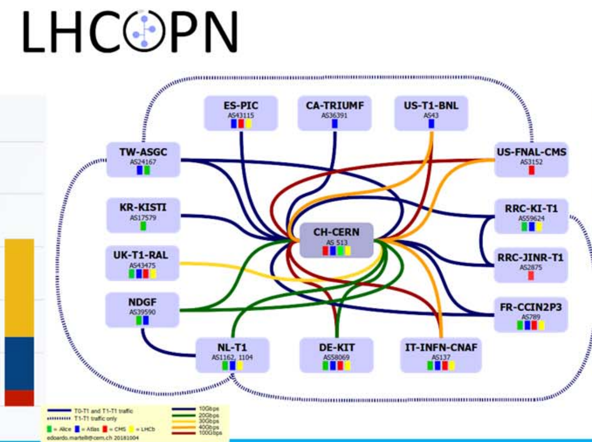
(b)

The four main component layers of the Worldwide LHC Computing Grid (WLCG) are **networking, hardware, middleware and physics analysis software**.

Networking
Distribution of data to the hundreds of collaborating institutes worldwide. In 2018, we are seeing an average of 35 Gigabit per second (Gb/s) global transfer rates

Last 24 hours WLCG network throughput





Big Data: Astronomy and SKA

SKA

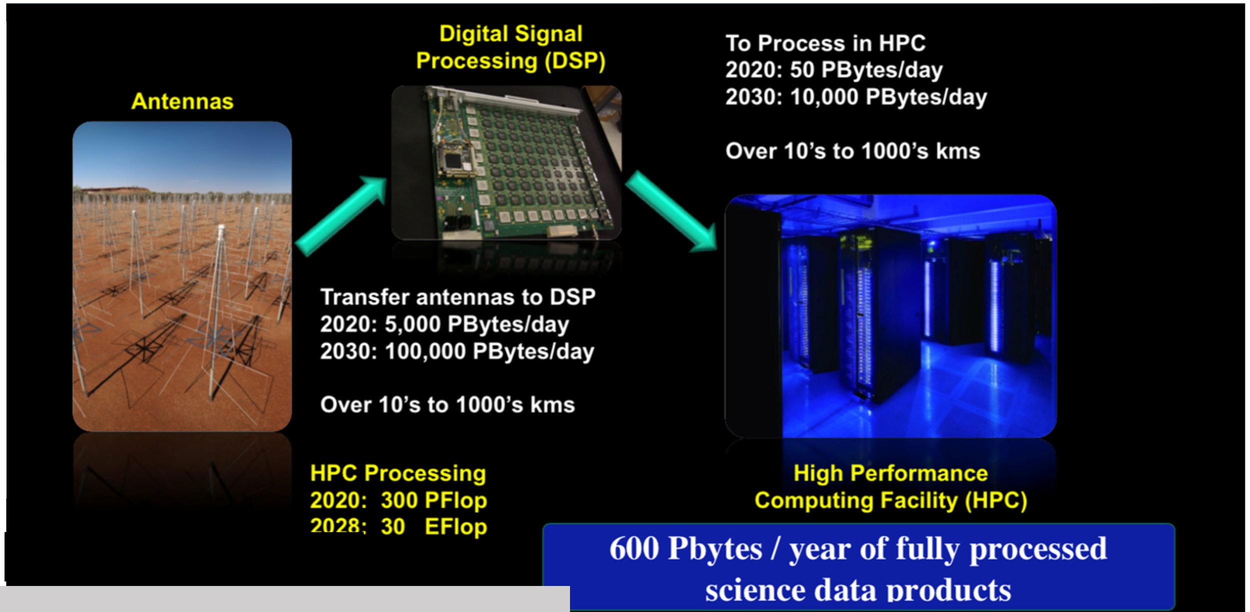
➔ the world's largest radio telescope capable of transformational science and discoveries: one of the priorities for the scientific communities and biggest challenge in data management, computing, networking
impossible with current facilities.

➔ built over two sites in Australia and Africa, over a million square metres of collecting area through ***<u>many thousands of connected radio antennas</u>***.

➔ constructed in two phase with existing technology and then upgraded

Table 1: List of SKA Science Working Groups.

SKA Science Working Groups			
1	Extragalactic Spectral Line	7	Our Galaxy
2	Solar, Heliospheric & Ionospheric Physics	8	Epoch of Reionization (EOR)
3	Cosmology	9	Extragalactic Continuum
4	Cradle of Life	10	HI Galaxy Science
5	Magnetism	11	Pulsars
6	Transients		



The SKA project has <u>eleven science working groups</u>, with multiple scientific objectives. These science objectives were ranked by the SKAO and a <u>list of thirteen High Priority Science Objectives (HPSOs) was extracted.</u>

Worldwide SKA Regional Center (SRC)

Coordinated by a specific committee

What does an SRC need to do to maximise science delivery from the SKA?

How can we exploit the new capabilities of HPC-HPDA to afford the new science challenges for SKA?

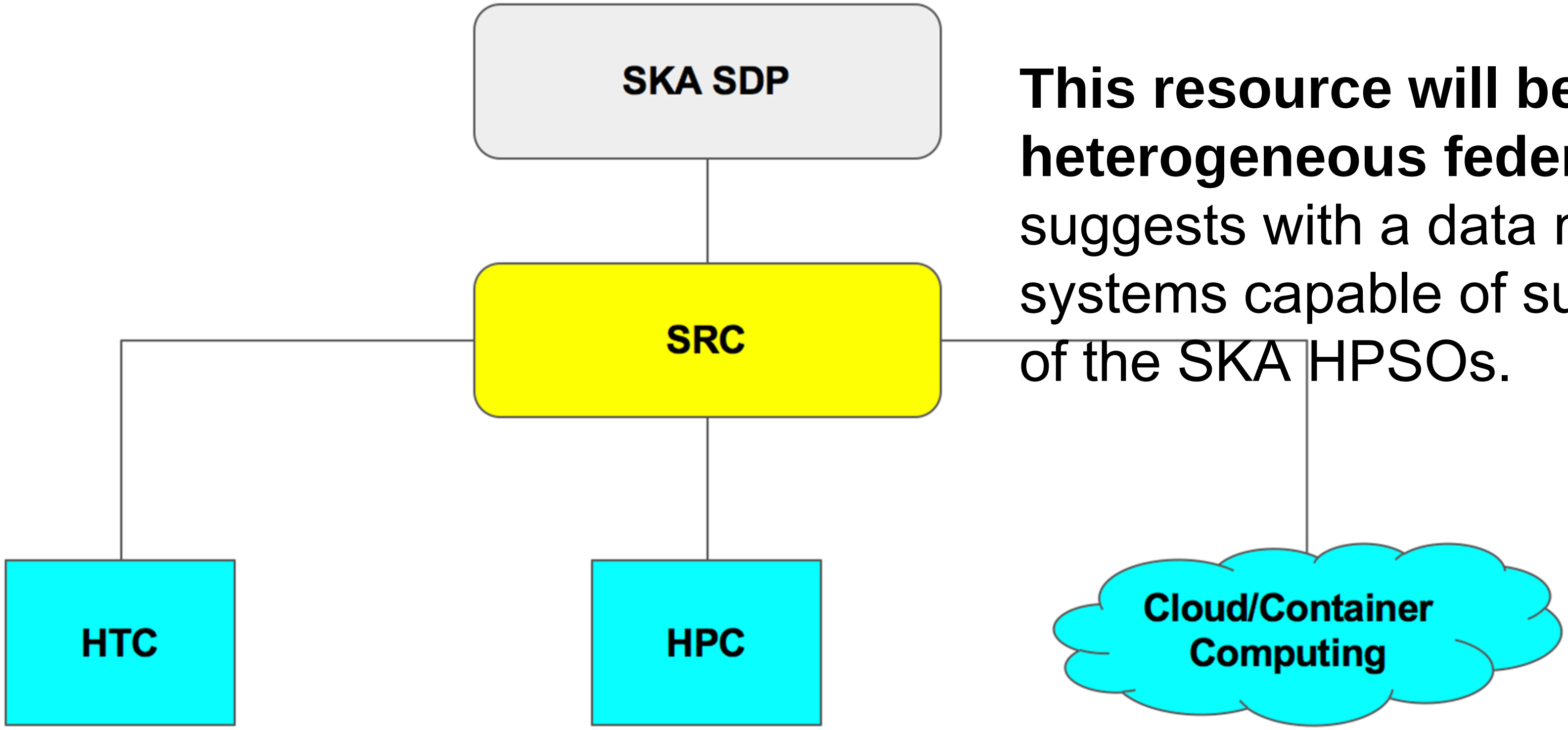
SKA SRC Computing Challenges

- Storage growing **at 750PB per year for the first 10 years for operation**, and **8.5 ExaBytes for a 15 years.** Assuming that a European SKA data centre will hold a full mirror of the SKA HPSO SDP primary data products and their associated derived secondary products.
- Minimum specified sustained processing capacity of **29** PFlops (***8 PFlops of re-processing and 21 PFlops of post-processing***) Assuming that the expected number of post-processing runs for each dataset and that processing is completed at the same rate that new data is ingested.
- ***Minimum ingest data rate of 60 Gbit/s to support SKA key science*** and PI science data products.

SRC Computing Main Schema connections

The SRC will need to provide **a minimum of 29 PFlops of sustained processing** to achieve the key science goals of the SKA.

This resource will be distributed over a heterogeneous federated system and suggests with a data management systems capable of supporting the needs of the SKA HPSOs.



There is also an ***highly probablity that containerisation (using Singularity) will be supported*** at a site level within a federated SRC.

HPC: resources/limits for BIG Data Science

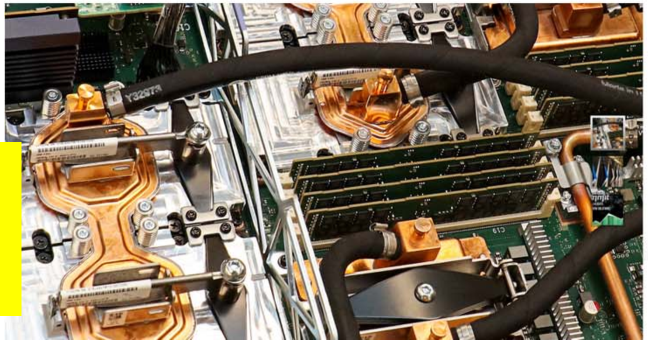
HPC infrastructure today are based on nodes with CPU **1.5-3 Ghz** (Intel Xeon, Power9, CRAY… other specific CPUs) with high number of cores on the same nodes and **accellerators** GPUs. Memory per node **256-512 GB**

The main systems have peak performance **20-190 Pflops.**

Network Infiniband/Mellanox

High Power consuming: **2-18 MWatt**

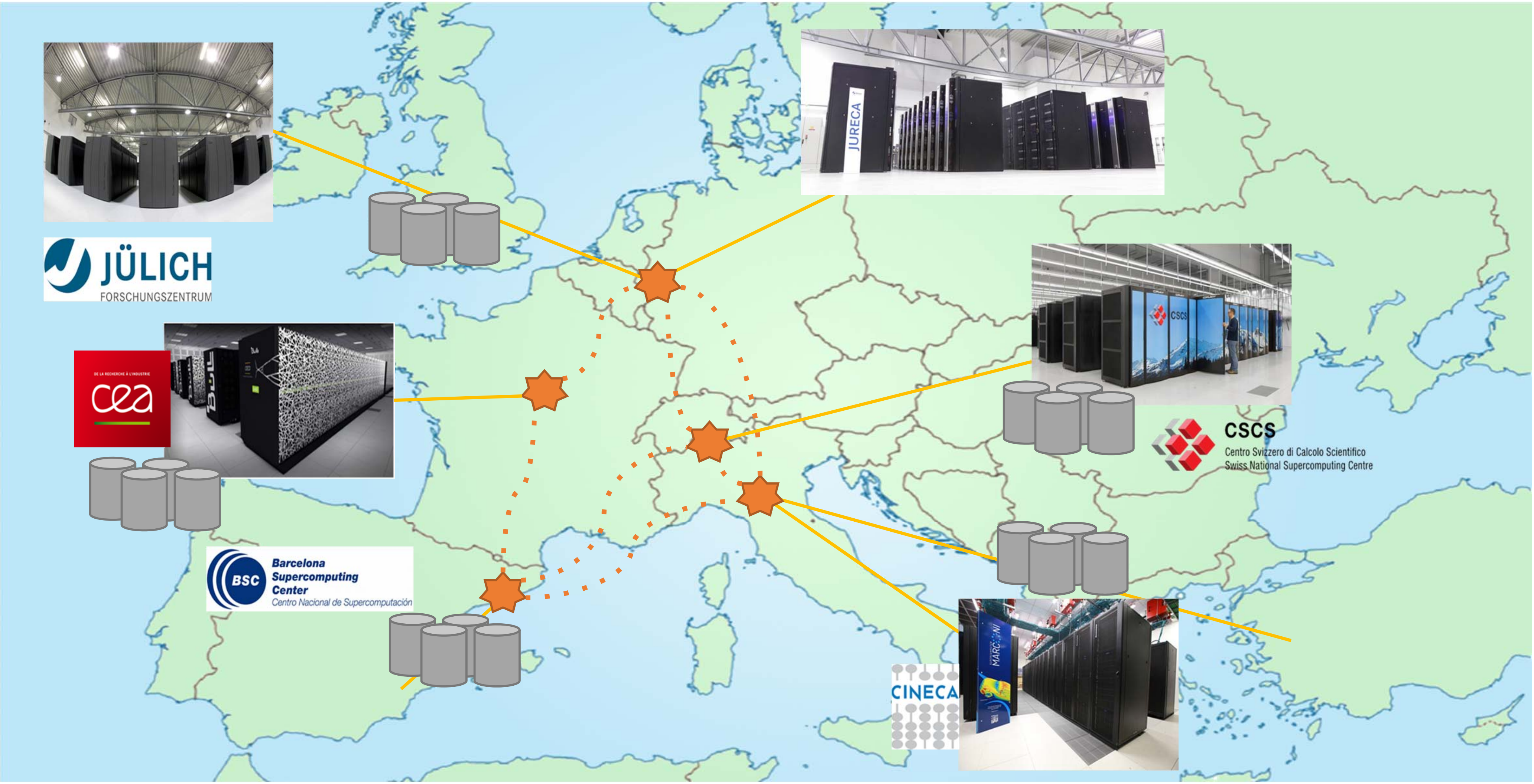
Liquid **cooling system**



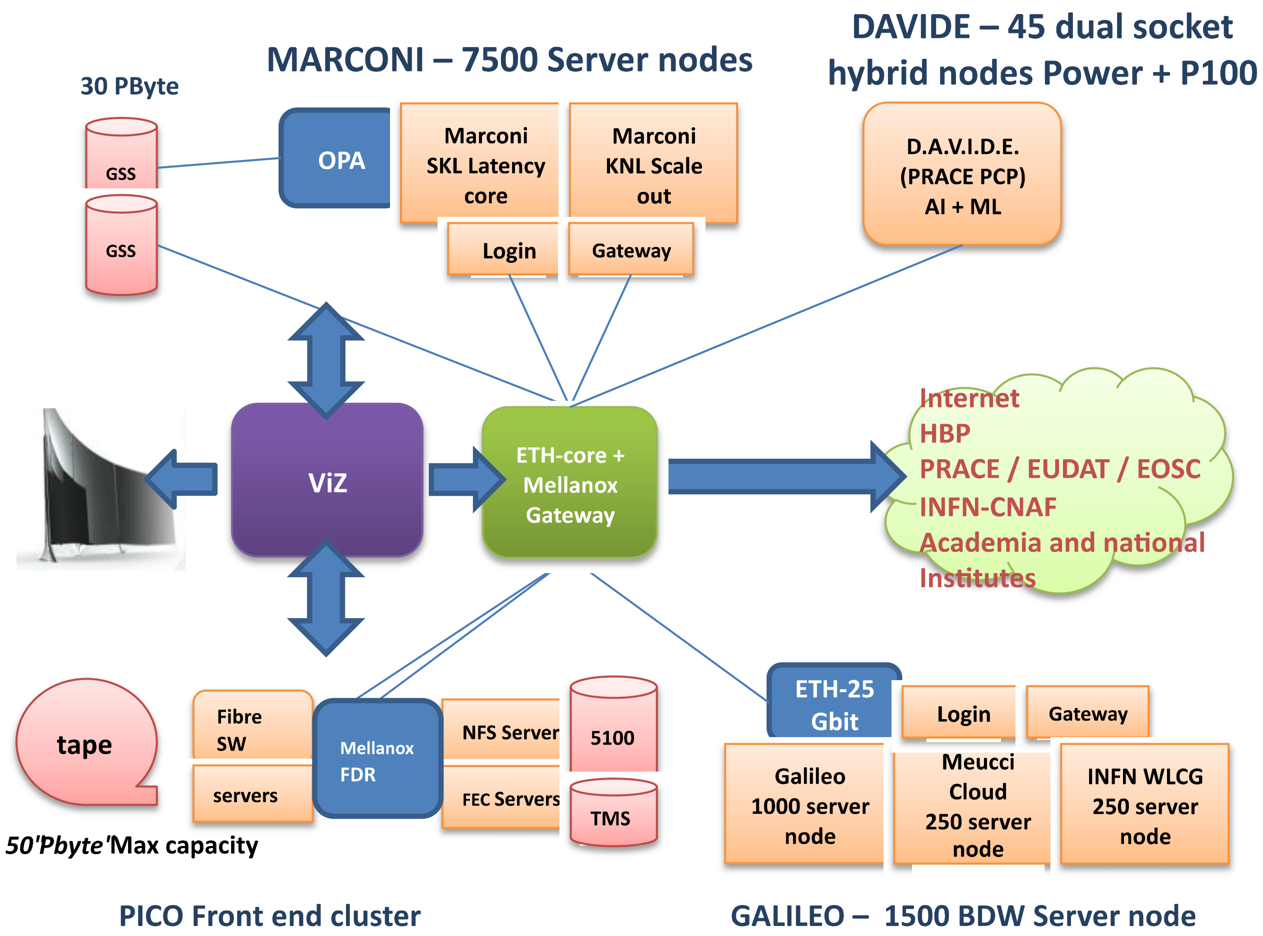
These features are also the ***limits*** of the current technology.

Rank	Site	System	Cores	Rmax (TFlop/s)	Rpeak (TFlop/s)	Power (kW)
1	DOE/SC/Oak Ridge National Laboratory United States	**Summit** - IBM Power System AC922, IBM POWER9 22C 3.07GHz, NVIDIA Volta GV100, Dual-rail Mellanox EDR Infiniband IBM	2,282,544	122,300.0	187,659.3	8,806
2	National Supercomputing Center in Wuxi China	**Sunway TaihuLight** - Sunway MPP, Sunway SW26010 260C 1.45GHz, Sunway NRCPC	10,649,600	93,014.6	125,435.9	15,371
3	DOE/NNSA/LLNL United States	**Sierra** - IBM Power System S922LC, IBM POWER9 22C 3.1GHz, NVIDIA Volta GV100, Dual-rail Mellanox EDR Infiniband IBM	1,572,480	71,610.0	119,193.6	
4	National Super Computer Center in Guangzhou China	**Tianhe-2A** - TH-IVB-FEP Cluster, Intel Xeon E5-2692v2 12C 2.2GHz, TH Express-2, Matrix-2000 NUDT	4,981,760	61,444.5	100,678.7	18,482
5	National Institute of Advanced Industrial Science and Technology (AIST) Japan	**AI Bridging Cloud Infrastructure (ABCI)** - PRIMERGY CX2550 M4, Xeon Gold 6148 20C 2.4GHz, NVIDIA Tesla V100 SXM2, Infiniband EDR Fujitsu	391,680	19,880.0	32,576.6	1,649

PRACE: HPC in Europe



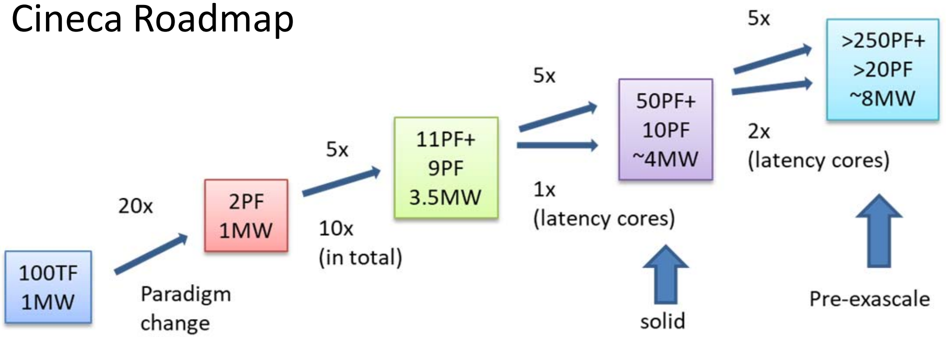
Current%HPC%architecture%infrastructure%



- Global%peak%performance%in% excess%of%25%Pflops%
- Global%produc, on%capacity% >%3000%servers%nodes%
- Modularity%compu, ng% combining%different% microprocessors%technology% based%on%the%same%ISA%and% different%soT ware%stack% enabling%effec, ve% produc, on%work%flows%
- Architecture%development% driven%by%the%scien, fic% challenges%and%applica, ons% domains%

HPC : Sustainable Roadmap

Cineca Roadmap

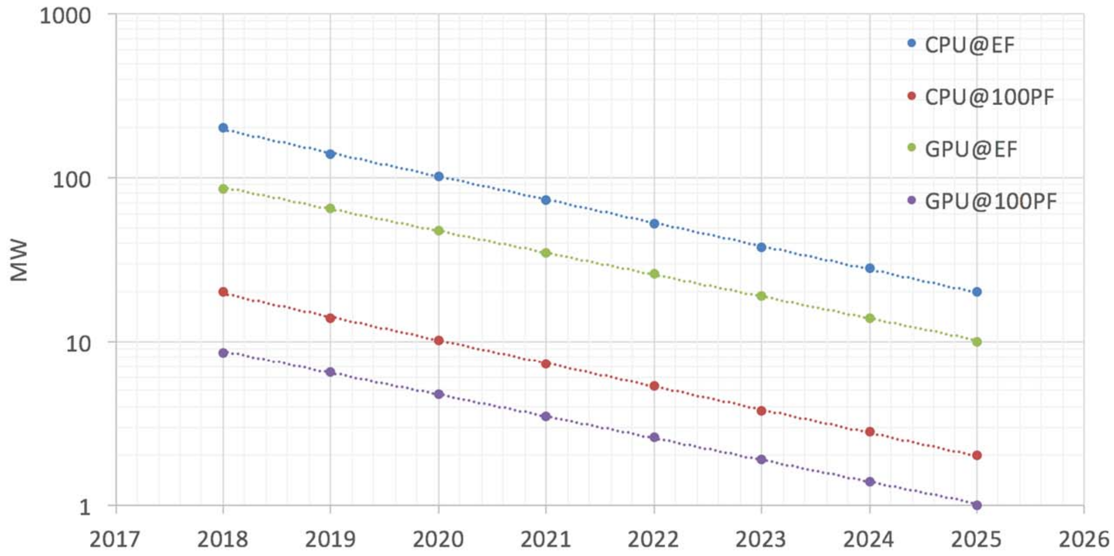


2009	2012/2013	2016/2017	2019/2020	2021/2022
IBM SP6 Power6	Fermi IBM BGQ PowerA2	Marconi Lenovo Xeon+KNL	Marconi + Public Procurement of Innovative Solutions for HPC	EuroHPC

Main current limits

- Disk storage capacity
- Network Latency,
- I/O latency and limited disk performances, etc.
- CPU core/freq./performances
- Power consuming

Power projection



Peek Perf (DP) @ 10MW

	2018	2019	2020	2021	2022	2023	2024	2025	2026
CPU	50PF	70PF	100PF	140PF	200PF	250PF	330PF	500PF	750PF
GPU	125PF	166PF	200PF	300PF	385PF	525PF	715PF	1EF	1.3EF

The EuroHPC Declaration

EuroHPC Declaration – Participating States

EuroHPC Joint Undertaking · CINECA

Declaration signed in Rome 23/03/2017 by:

For Italy signed the Ministries of Education University and Research and of Economic Development

France · Germany · Italy · Luxembourg · Netherlands · Portugal · Spain

21 more countries signed the Declaration:

Austria · Belgium · Bulgaria · Croatia · Czech Republic · Denmark · Estonia · Finland · Greece · Hungary · Ireland · Latvia · Lithuania · Norway · Poland · Romania · Slovakia · Slovenia · Sweden · Switzerland · Turkey

		European Commission	Partecipating States	Total	Private Mem.
Infrastructure Acquisition Operating machines	Pillar 1	270	290	560	
Research & Innovation Applications & Skills	Pillar 2	206	186	392	422
JU Admin/Running costs		10	10	20	

Present EU Financial Framework — 2 Pre-exascale machines (2020–2021)

Next EU Financial Framework (2.7 billion Euros) — 2 Exascale machines (2023)

2018 · 2019 · 2020 · 2021 · 2022 · 2023 · 2024 · 2025 · 2026

Exascale consideration

- Exascale is not (only) about scalability and Flops performance!
  - In an exascale machine there will be **10^9 FPUs**, bring data in and out will be the main challenge.
  - **10^4 nodes**, but **10^5 FPUs inside the nodes**!
- Exascale future nodes:
  - heterogeneity of hw in the nodes
  - deeper memory hierarchies

- From GPU to *specialized core* (tensor core **)
- Specialized memory module HBM
- Specialized ***non volatile memory NVRAM***

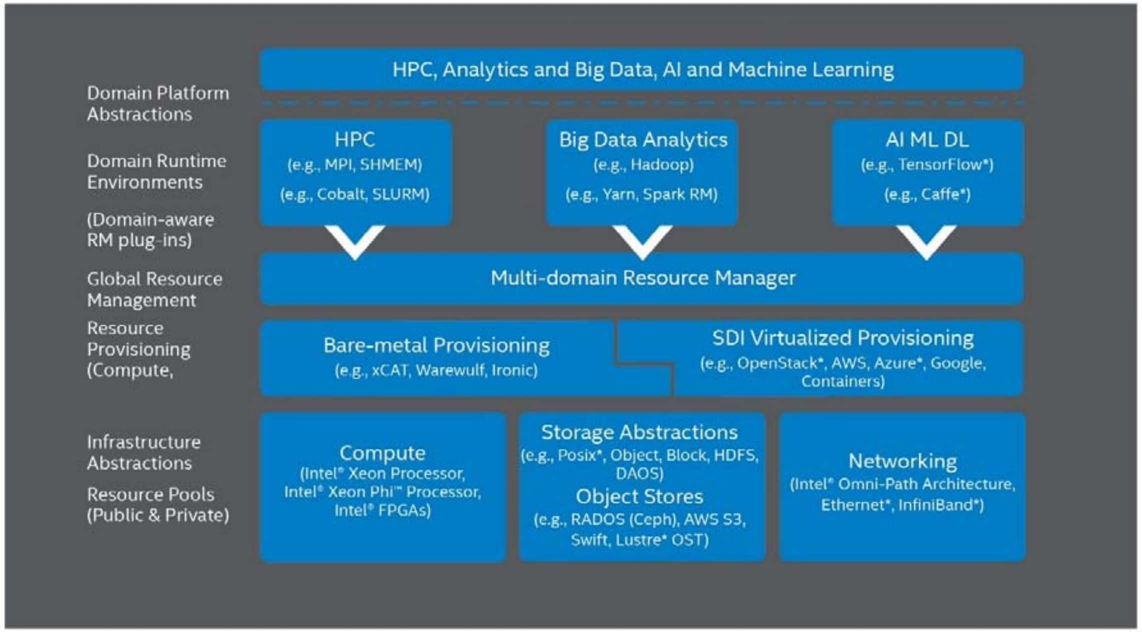
*** A revolutionary technology that delivers groundbreaking AI performance. Tensor Cores can accelerate large matrix operations, which are at the heart of AI, and perform mixed-precision matrix multiply and accumulate calculations in a single operation*

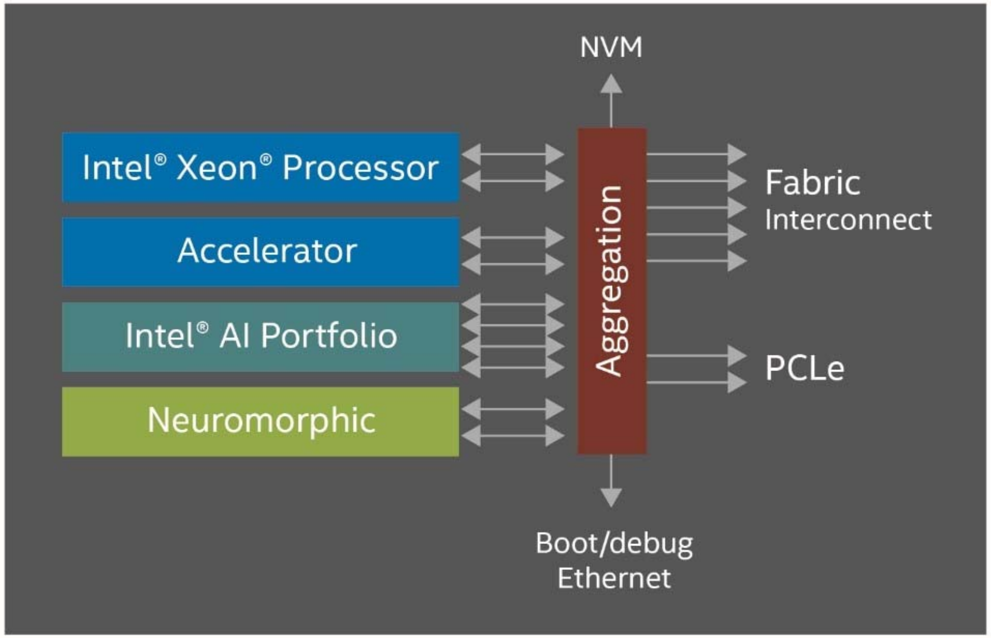
HPC and computing models in the next generation

"3 Pillars" of Exascale:

- Artificial Intelligence (AI) and Machine Learning (ML);
- Data Analytics and Big Data;
- High Performance Computing (HPC).

What this means is that ***users of the future will program using mixing models*** that interact through memory.





More concretely, **Intel** is working towards exascale systems that are highly configurable that can support upgrades to fundamentally new technologies including scalable processors, accelerators, neural network processors, neuromorphic chips, FPGAs, persistent memory, 3D NAND, and custom hardware.

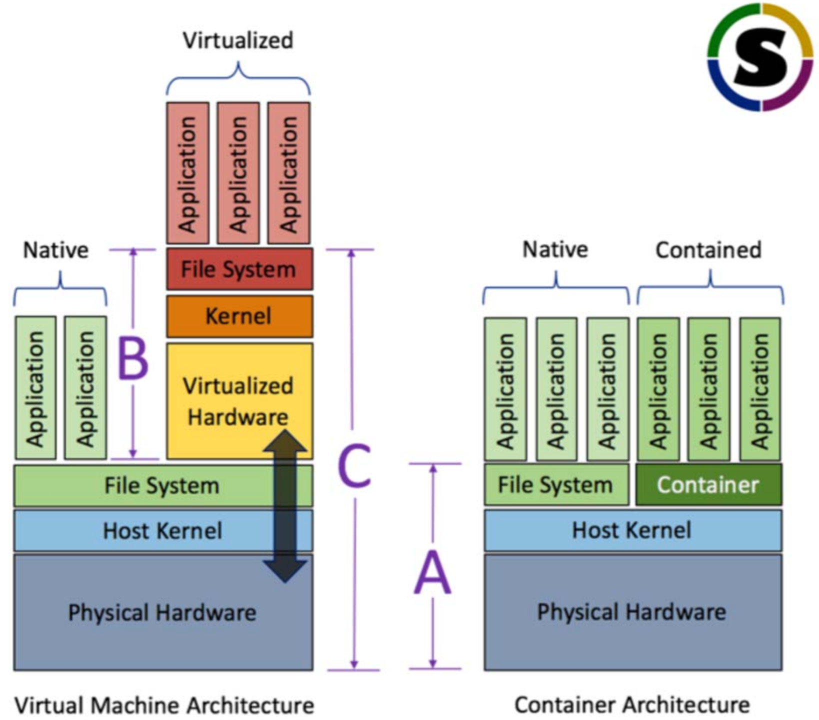
Big Data and Cloud Computing

A container is simply a ***runtime environment*** - a particular OS distribution, and user's application together with all its dependencies bundled into a single package. Containers sit directly on top of the host kernel and are much more lightweight than virtual machines.

Users would port their applications within containers for reproducibility, distribution and ease of deploying their application across varied sites and middlewares seamlessly.

Docker and Singularity are two popular containerisation technologies with Docker having more appeal for provisioning resources on demand, and ***Singularity for packaging scientific compute workflows***.

When working in a distributed environment, having an application containerised means ***more sites will be eligible to run the software analysis*** since all the dependencies are already packaged into the container.



The vision

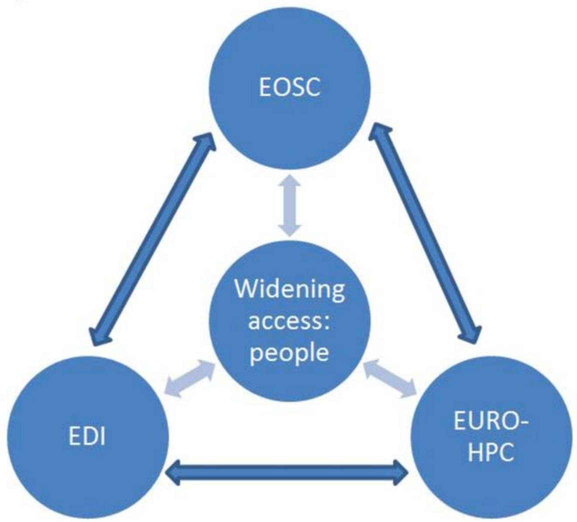
The European Open Science Cloud (EOSC) is currently a process to build a digital platform that is inspired by the ***FAIR*** principle.

FAIR stands for: Findable, Accessible, Interoperable and Reusable.

It is the idea between the recent actions taken by the European Commission and various funding agencies in favour of open data and open science. The EOSC platform, also called the **EOSC-Hub**, intends *in fine* to make possible to the whole European scientific Community and beyond: **exchange of data, easy access to knowledge and access to all useful infrastructures for all scientific disciplines**.

The ***EOSC*** will be ***linked*** to the future pan-European ***HPC***, as well as with the future European Data Infrastructure ***(EDI)*** which are funded alongsie the EOSC by the European Commission.

The European Commission has put ***375 million Euros*** for "implementing the European Open Science Cloud" phase 1 (H2020)

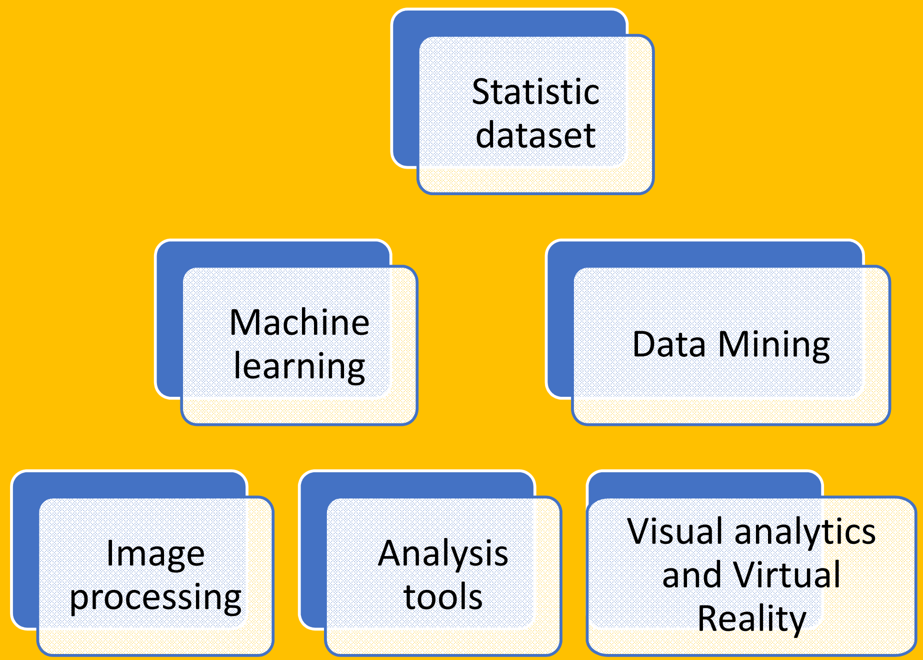


Big Data: Machine Learning & Deep Learning

Collection of multi-dimensional data → complete set of examples in the filed to investigate → *good* statistical dataset → good model for that domain.

→ Large number of examples
→ Good statistical dataset
→ Good models

Big Data Science will requires the integration of HPC-HTC-HPDA and Cloud resources. But also integrated method for a complete data analysis:



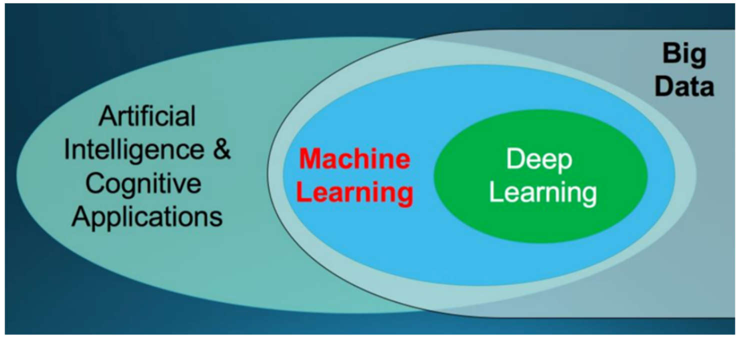
ML typical Steps

1. **Characterize the known**
   - Feature selection, Parameter space analysis & exploration
2. **Assign the new from the known**
   - Regression, classification, supervised learning
3. **Explore the unknown**
   - Clustering, unsupervised learning
4. **Discover the unknown**
   - Outlier detection and analytics

- Gain from large datasets: Best statistics of "typical" events, cross-correlation

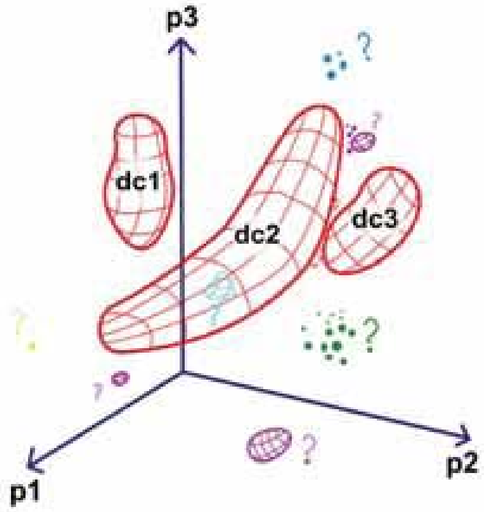
Big Data: Machine Learning & Deep Learning Challenges

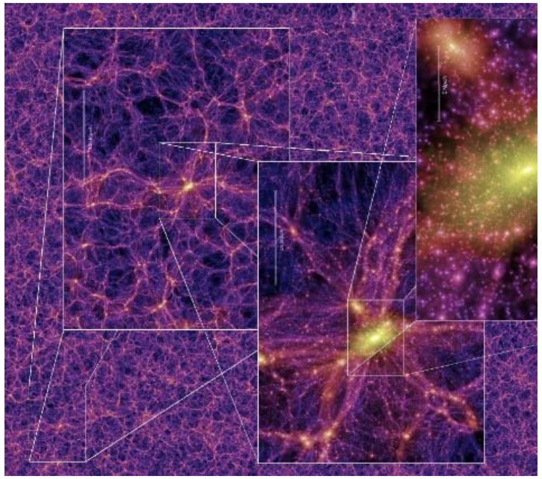
Recommendation 1: Big Data ➔ ***code must run close to data***

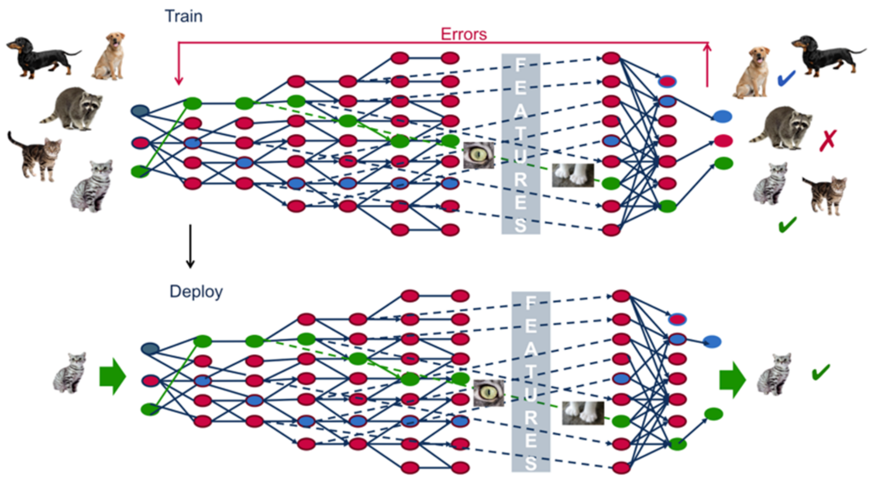


***Recommendation 2**: automatic self-adaptive methods ➔ explore and cross-correlate data*

***Recommendation 3**: improve simulations and exascale infrastructure to make intensive tests*







Conclusions

- HPC is rapidly evolving in the world: new Exascale infrastructure and Big data will change our traditional methods and tools
- The current (and traditional) infrastructure start to become **obsolete in a very few years**
- Are we ready to ***re-design*** with new infrastructure capabilities and ***re-engeneer*** our codes?
- ***International collaboration*** with mixed team is a big opportunity we will have: center of competence with joint expertize ***Researchers –ICT experts*** must to be created to win the new challenges .
- **Stay open to new cross-disciplinary collaborations**.