Introduction to High Performance and Data Intensive Computing: AN EXAMPLE

Dott. Costantino Zazza, costantino.zazza@unitus.it

IR0000032 – ITINERIS, Italian Integrated Environmental Research Infrastructures System
(D.D. n. 130/2022 - CUP B53C22002150006) Funded by EU - Next Generation EU PNRR-Mission 4 "Education and Research" - Component 2: "From research to business" - Investment 3.1: "Fund for the realisation of an integrated system of research and innovation infrastructures"

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani

HPC Technologies – DIBAF cluster

Agenda

- What about ?

  HPC, Big Data, GFLOPS, PB, …

- How do they work ?

  Calcolo numerico intensivo,

  Data Processing, …

- Useful for what ?

  Chimica Computazionale, Bioinformatica,

  IoT, Image Processing, Machine Learning, …

DIBAF Cluster

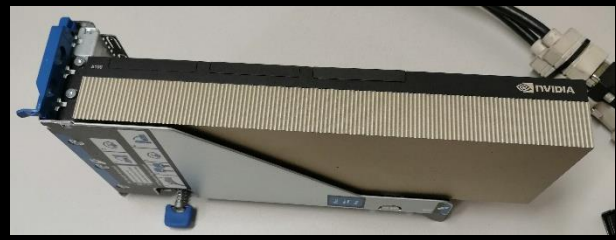
MSA1050

IB

ETH10

ETH1

The DIBAF Cluster



DIBAF Cluster

MSA1050

IB

ETH10

ETH1

DIBAF Cluster

56 / 10 / 1 Gbps

IB

ETH10

ETH1

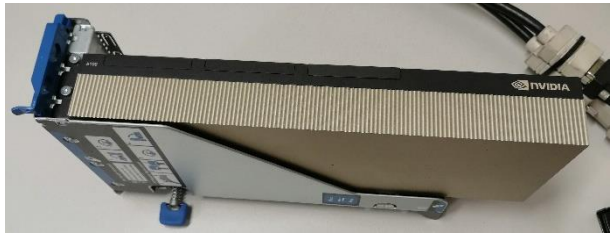
3

3

DIBAF Cluster

26

26



300 TFLOPS SP (HPC)

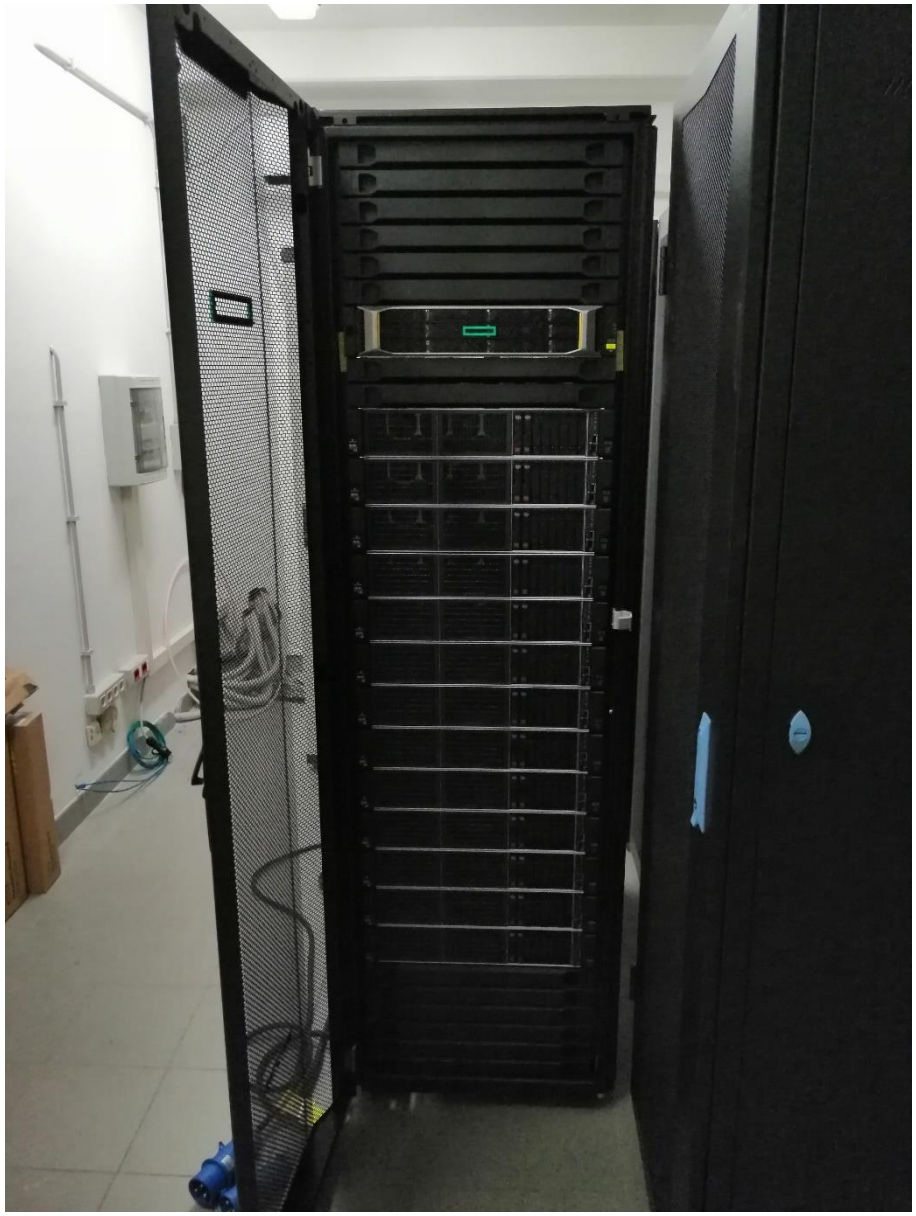
8

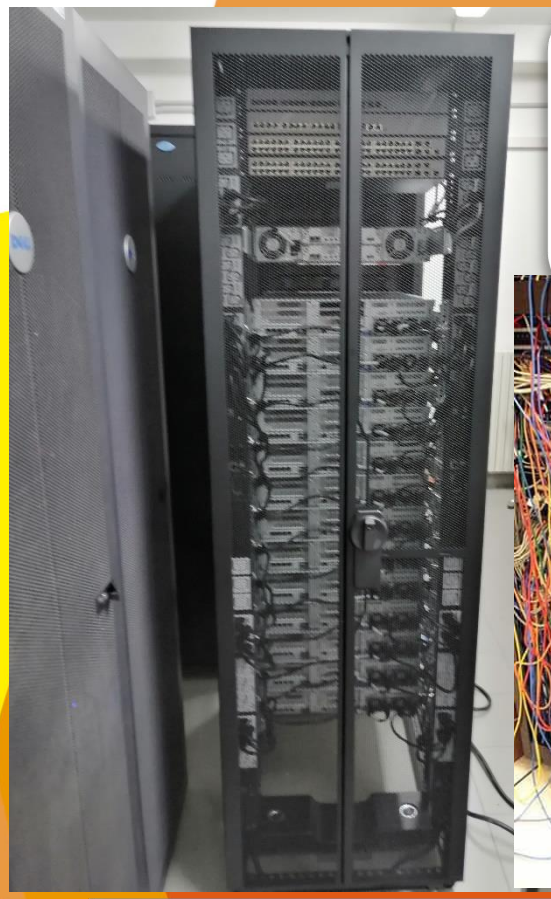
8

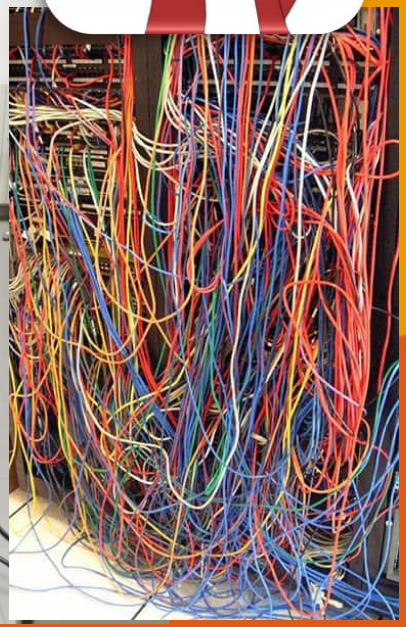
8 POPS INT4 (AI/ML)

1 PC = 0.01 TFLOPS

DIBAF Cluster

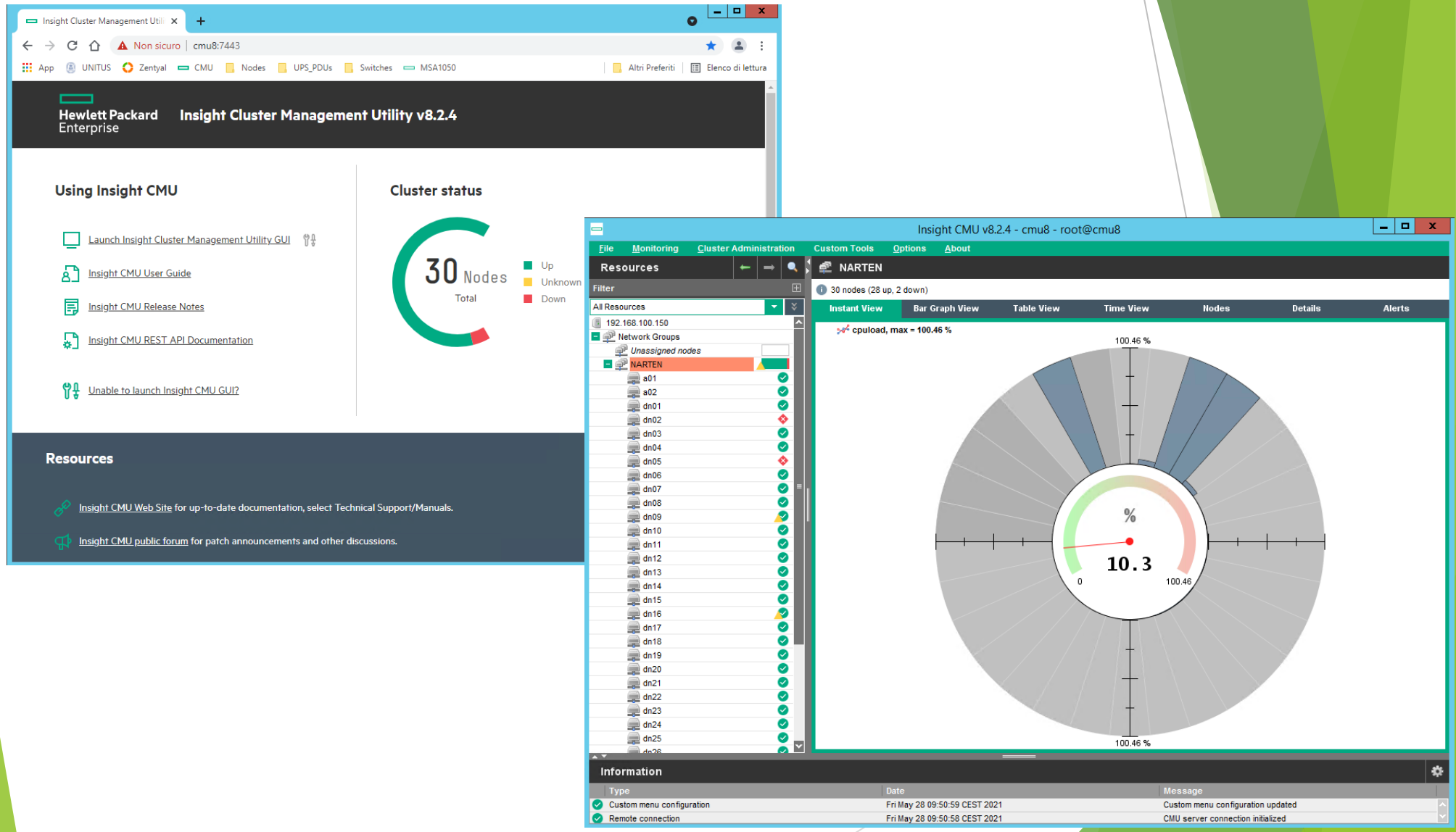




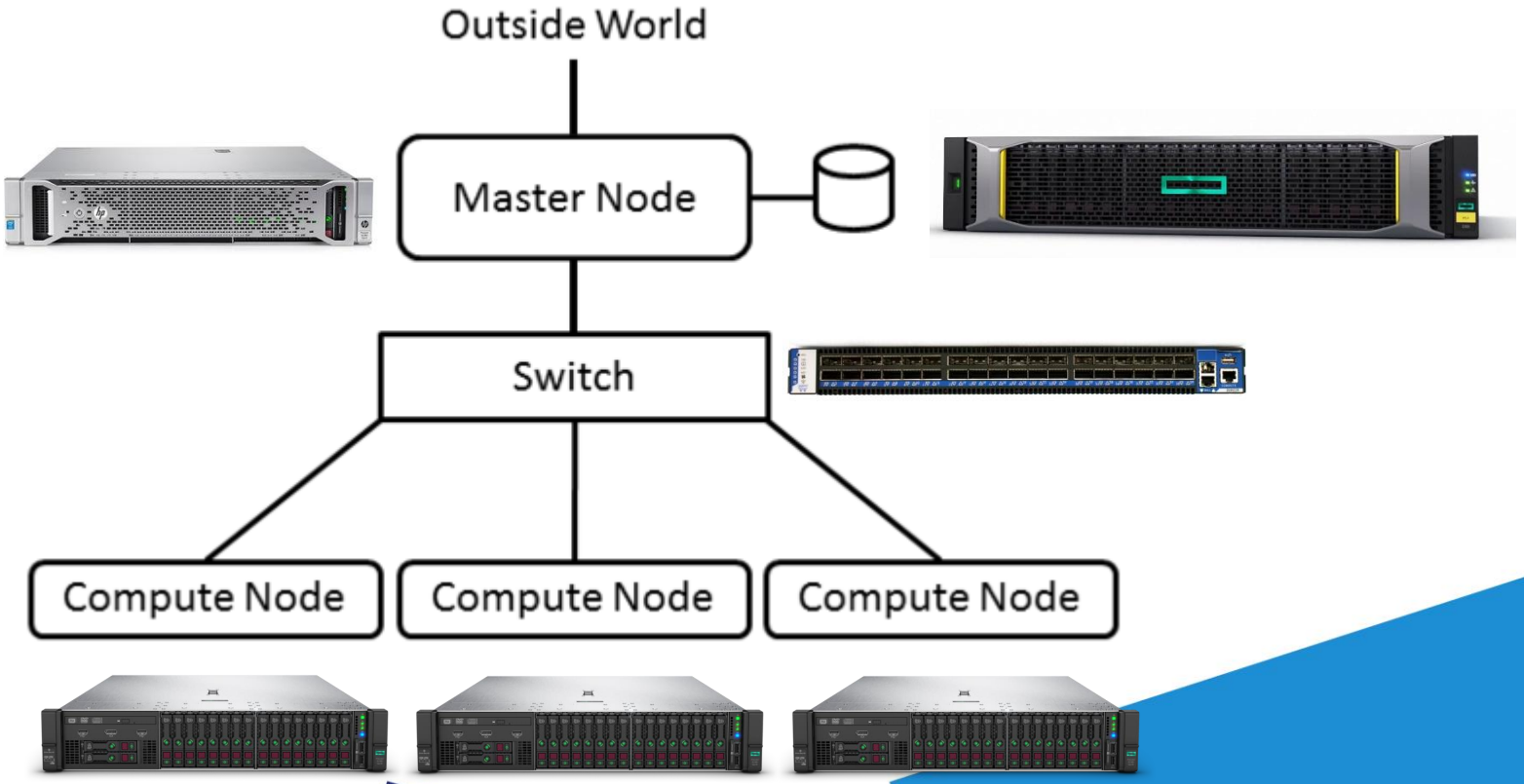


26+8 = **34** Ethernet cables
13 Infiniband cables
32 Power cables

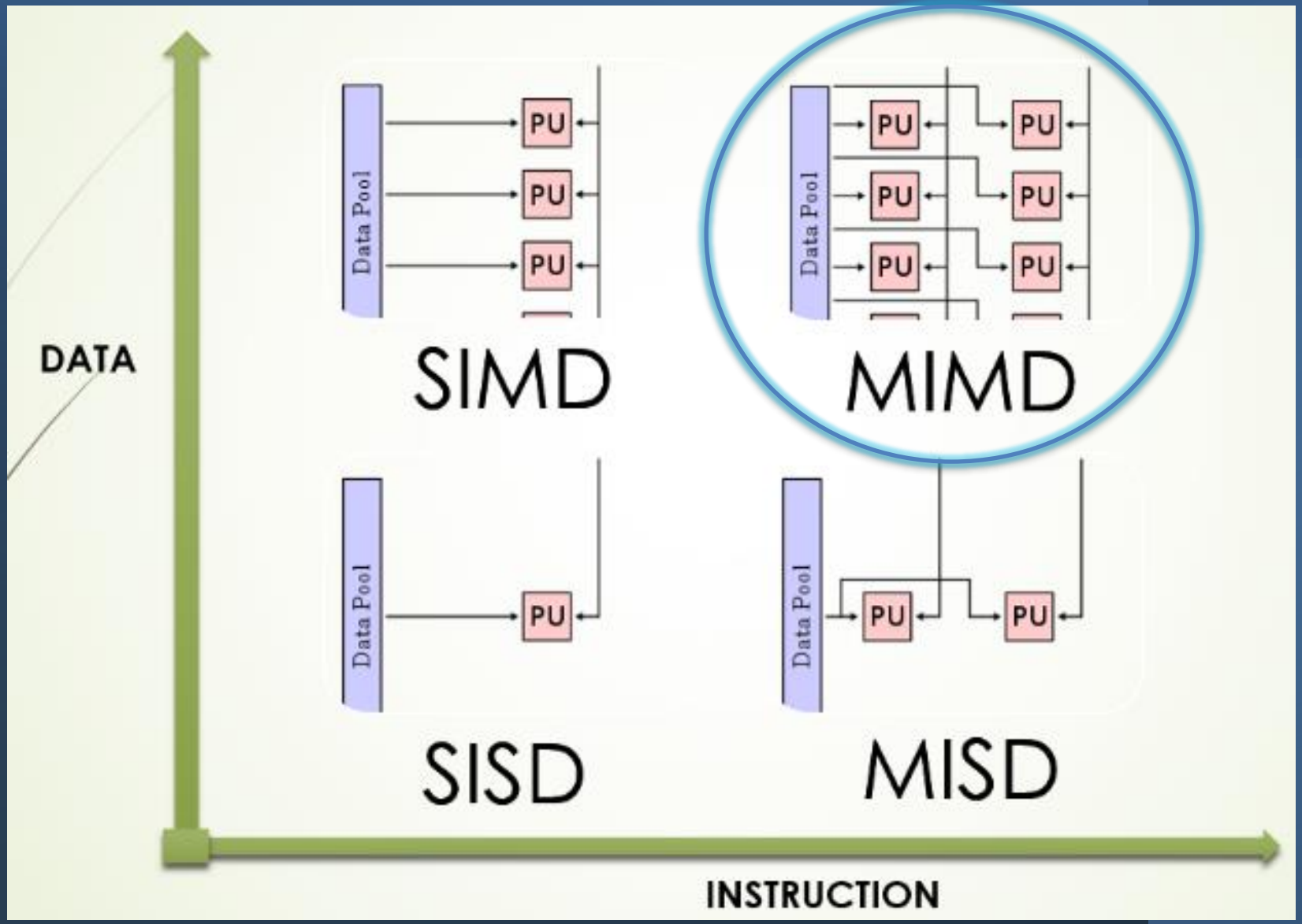
DIBAF Cluster



HPC – High Performance Computing



HPC Classification – Flynn Taxonomy

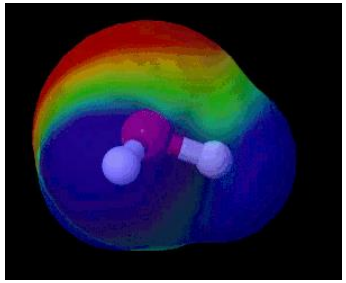


HPC – High Performance Computing

Brute force - Expected performance when scaling

Quantum Chemistry

	PC	Cluster
$O(N^3)$ / $O(N^4)$	3-5 atomi(H_2O, NH_3, CH_4)	**> 50 atomi**

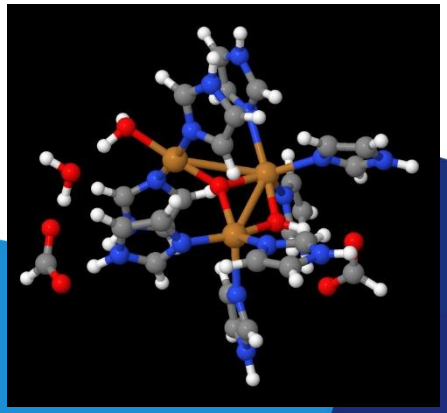


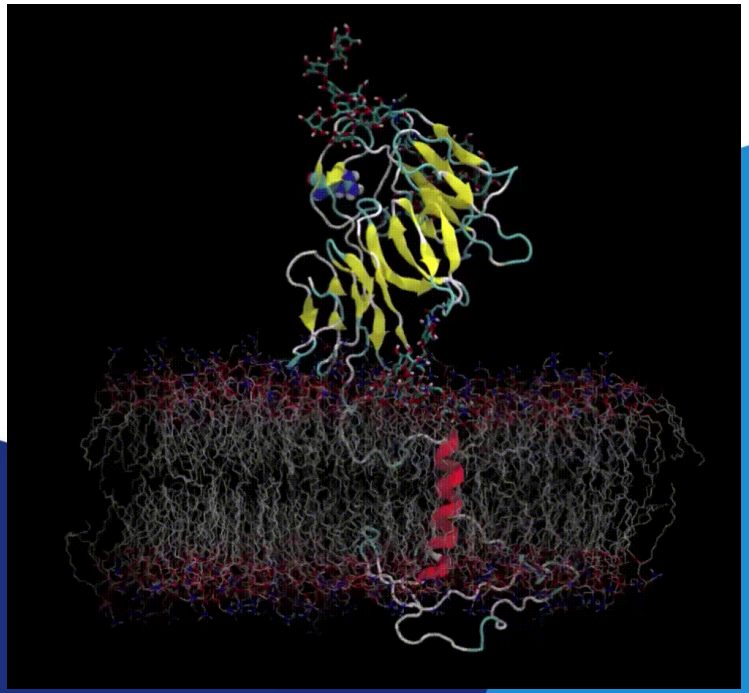
HPC – High Performance Computing

Brute force - Expected performance when scaling

Molecular Dynamics

	PC	Cluster
$O(N^2)$	20-50 atomi (peptide)	**> 1 milione di atomi**

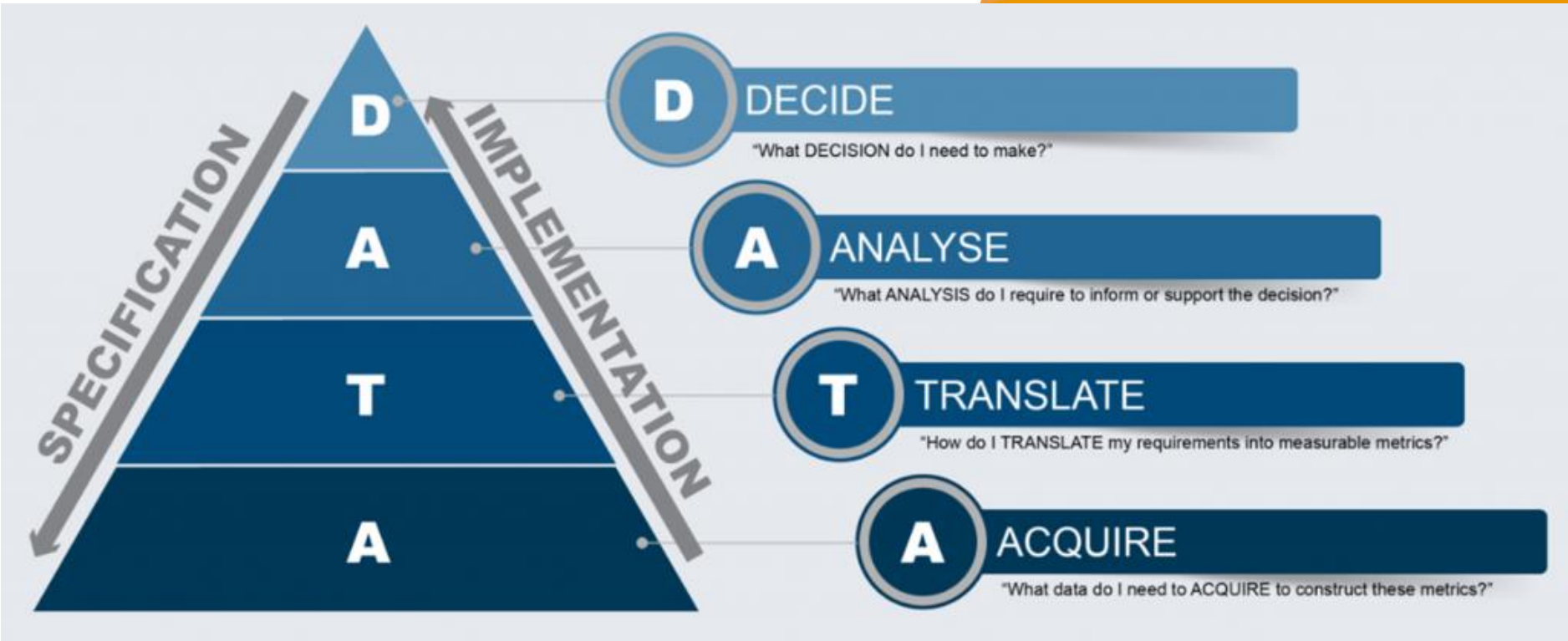




Big Data & Analytics

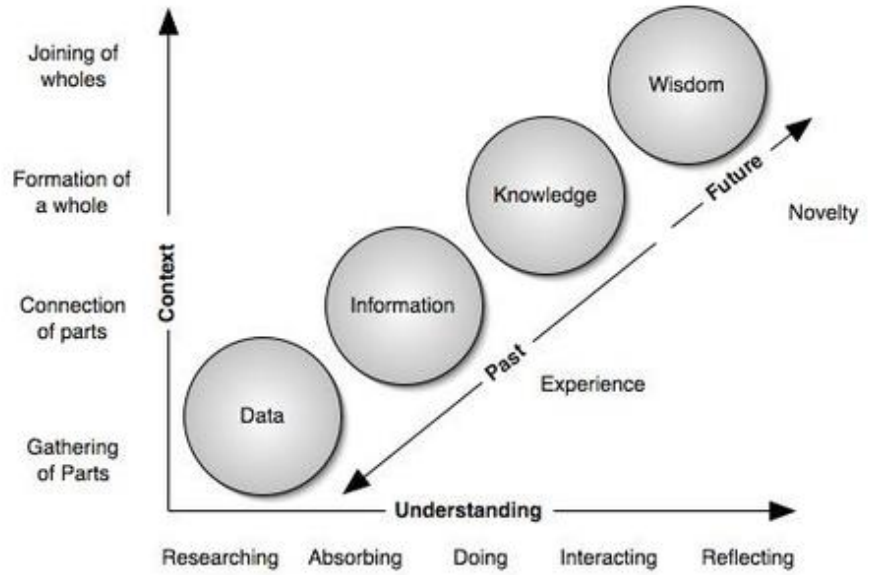
Data Driven model: make decision from data

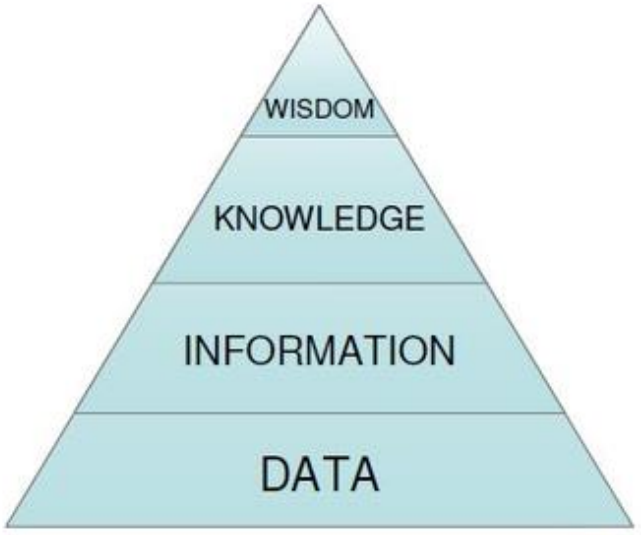
Data-driven decision-making (**DDDM**) is an approach that emphasizes using data and analysis instead of intuition to inform “business” decisions.



Big Data & Analytics

DIKW «hierarchical» model





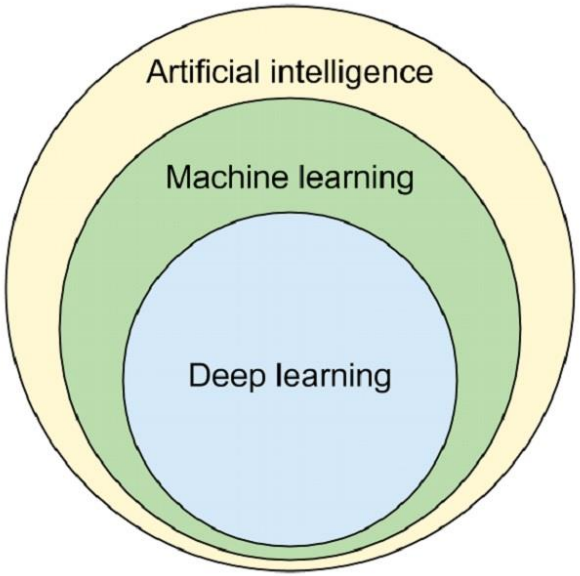
The cognitive pyramid

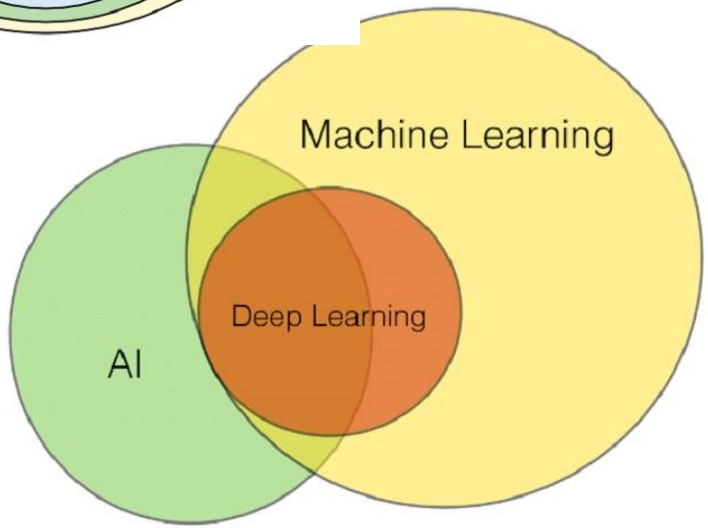
Pay attention on DATA QUALITY at the base of the pyramid !

Reference(URL):
https://en.wikipedia.org/wiki/DIKW_pyramid

Big Data & Analytics

Theoretical model: AI, ML, DL

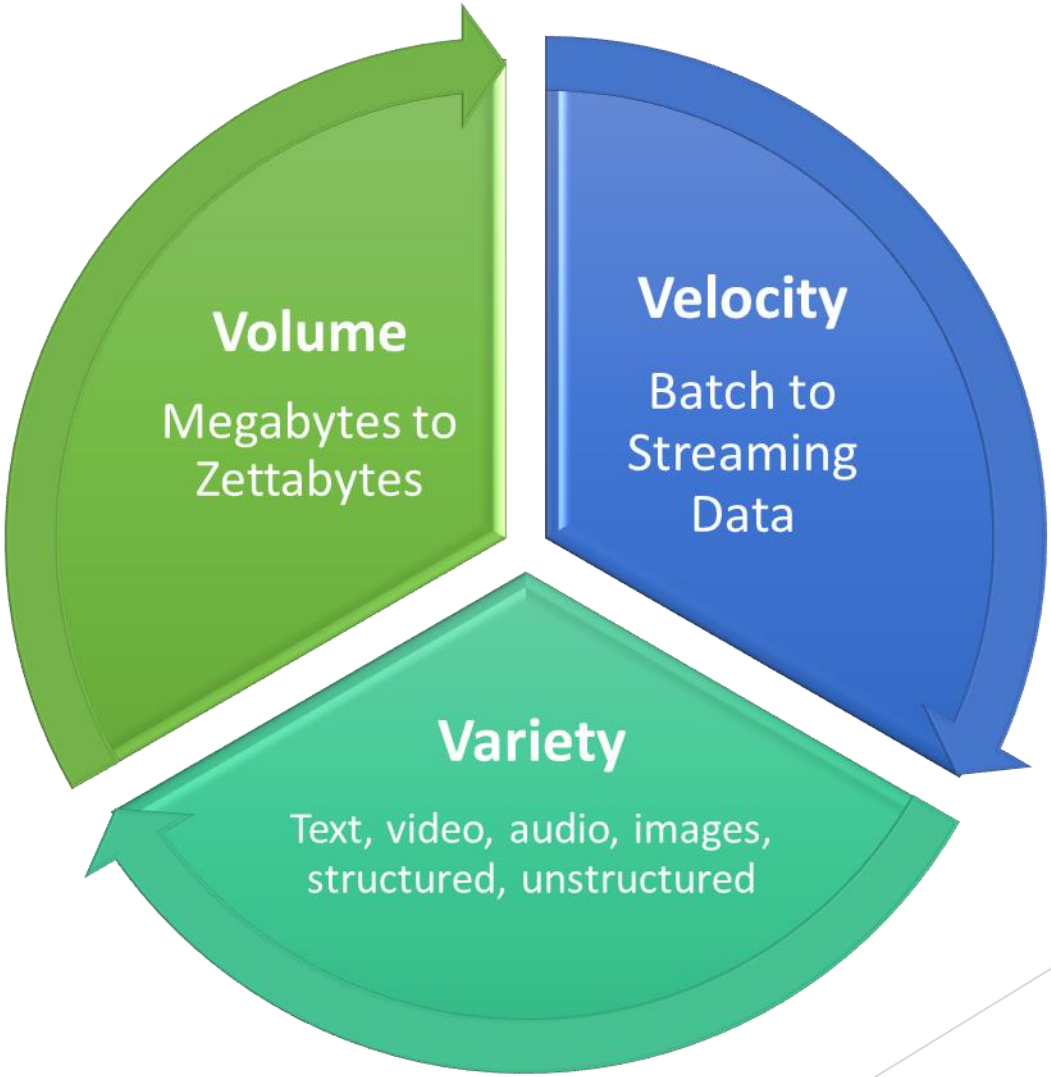




https://en.wikipedia.org/wiki/Machine_learning

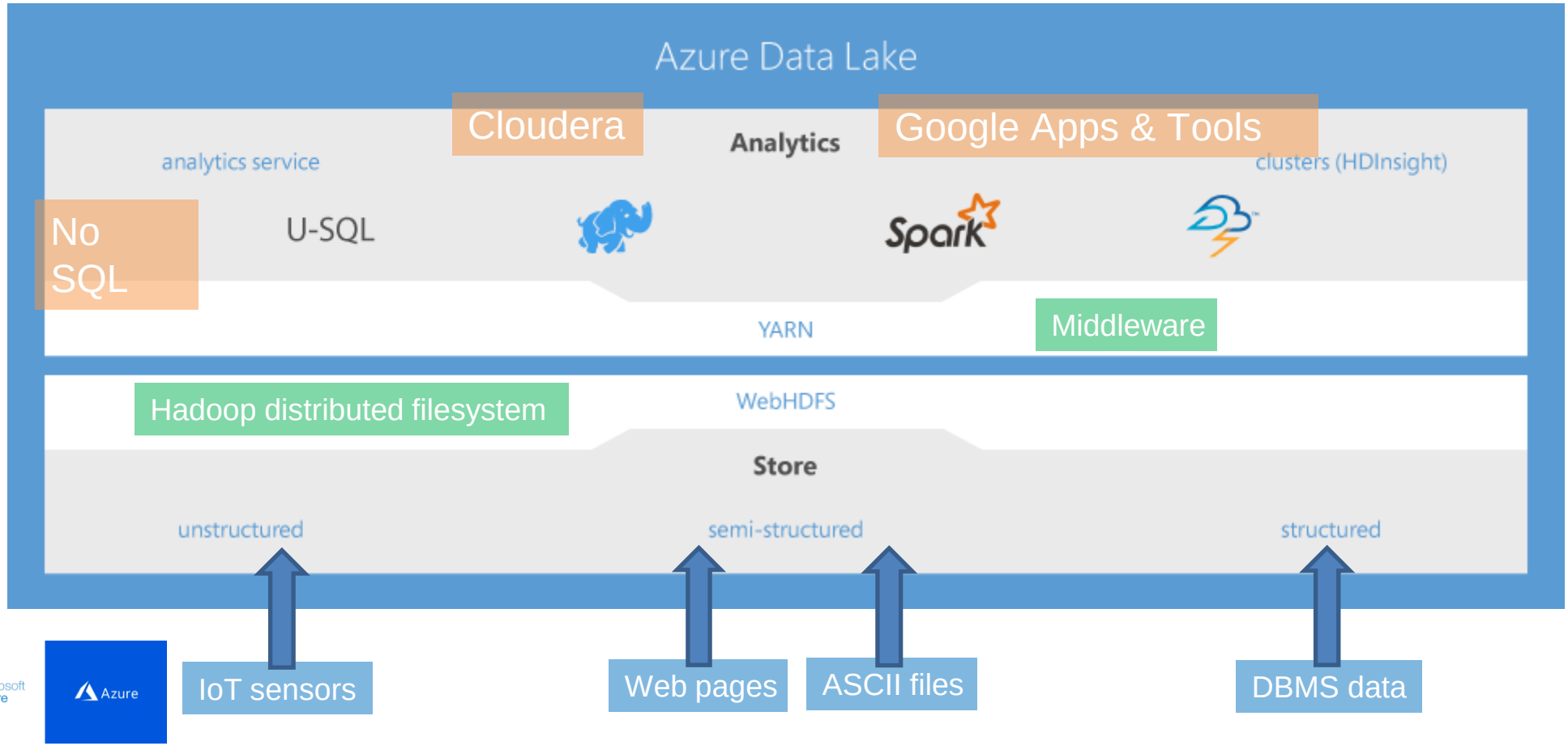
Big Data & Analytics

From Data to Big Data



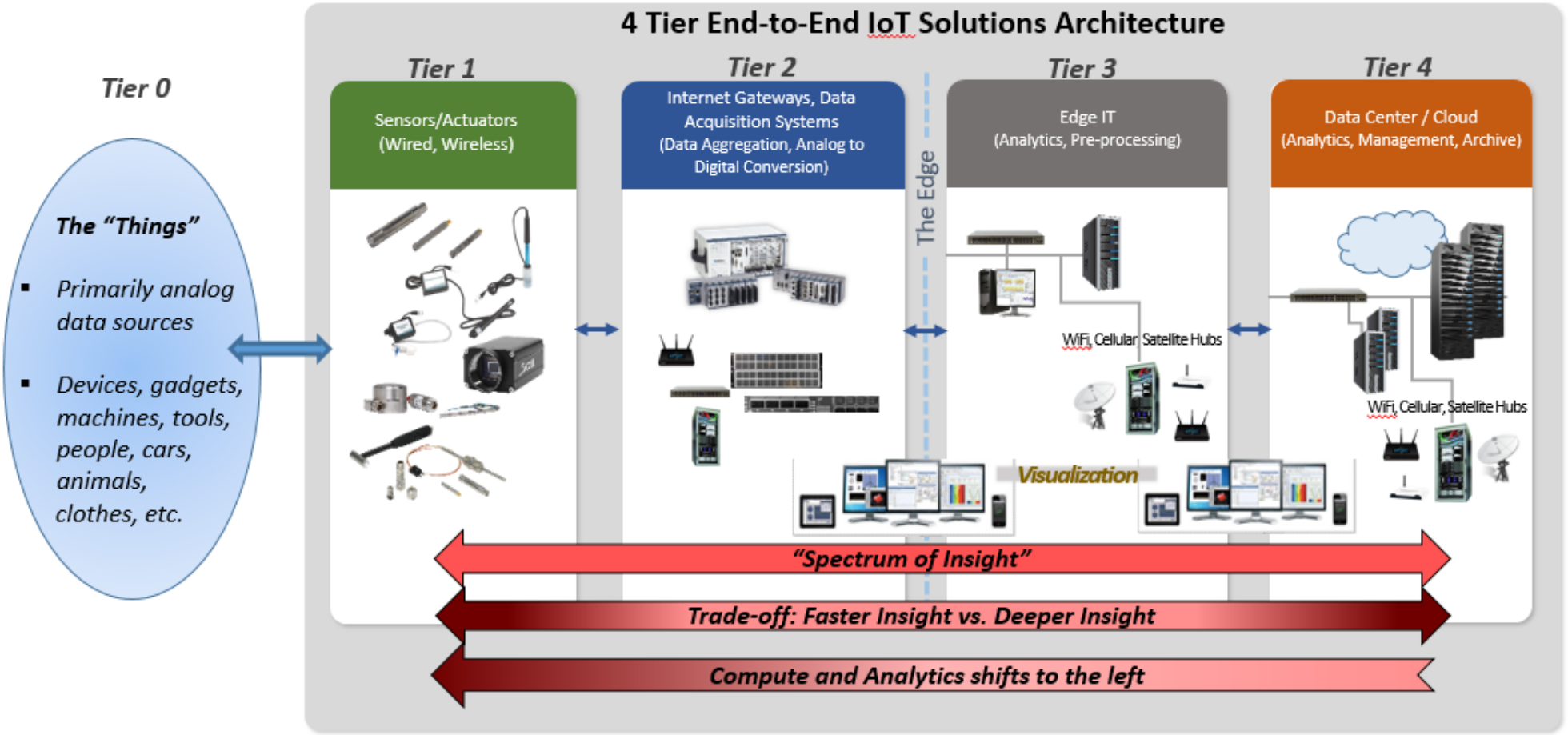
Big Data & Analytics

DIBAF BDA + AI model



Big Data & Analytics

DIBAF BDA + AI model



Hewlett Packard Enterprise

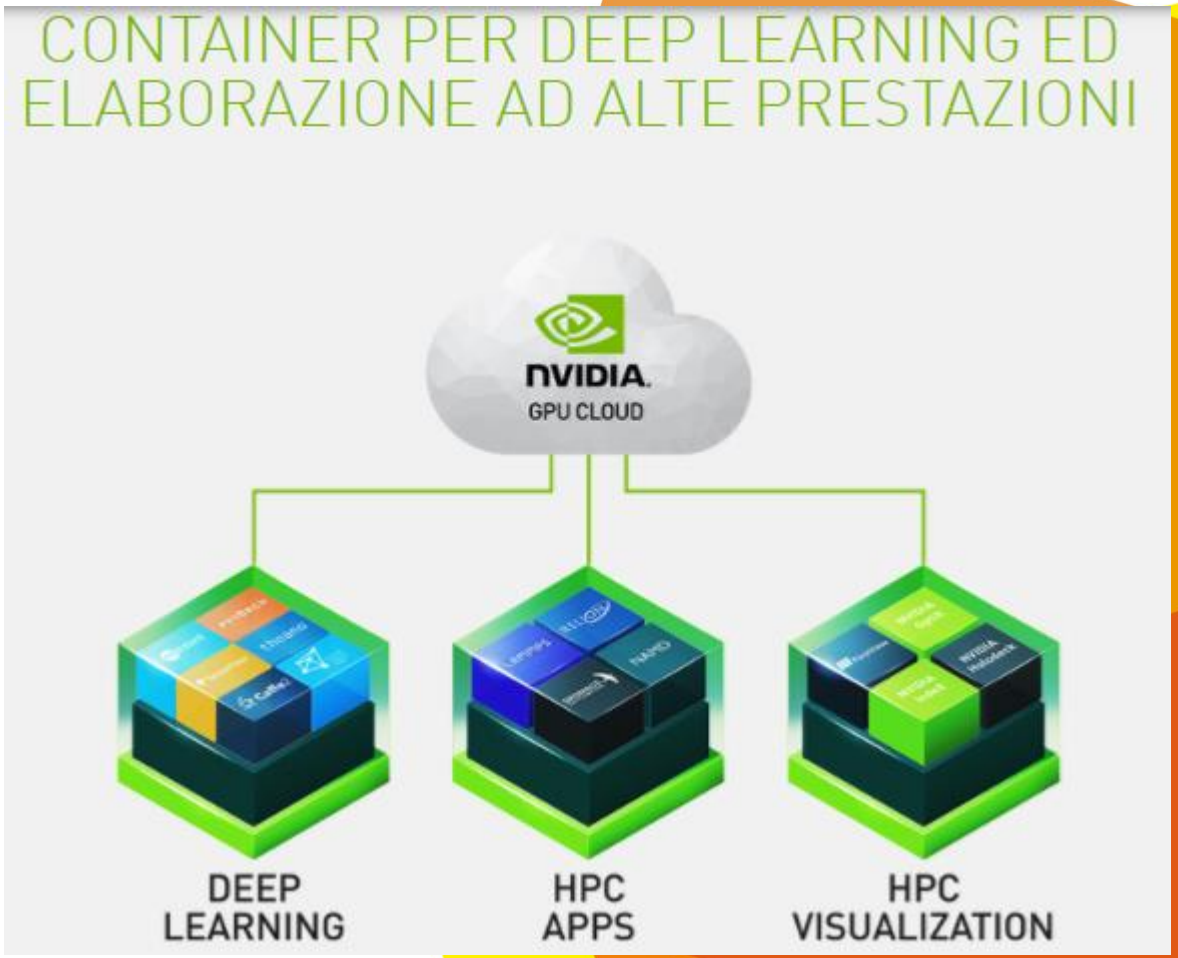
Microsoft

Big Data & Analytics

DIBAF BDA + AI model

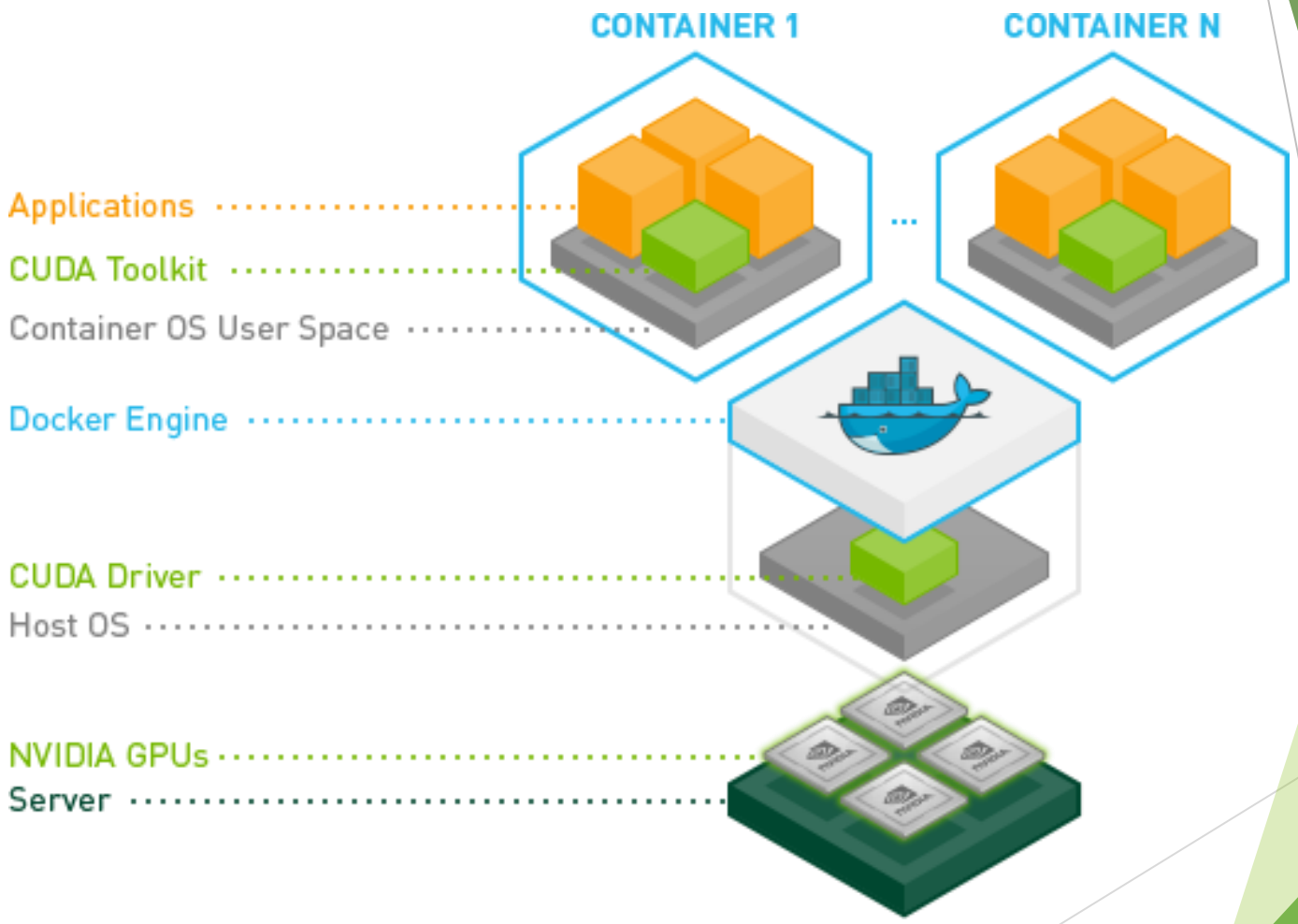
Hewlett Packard Enterprise

NVIDIA DEEP LEARNING INSTITUTE



Big Data & Analytics

DIBAF BDA + AI model

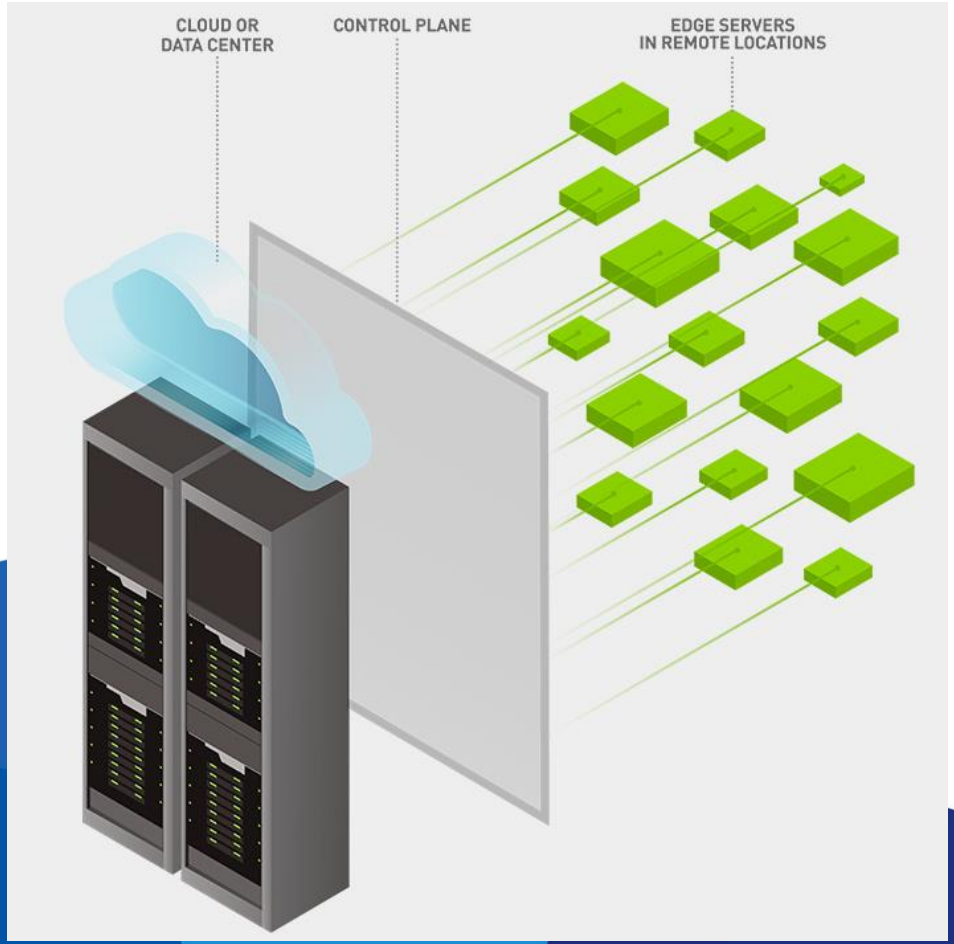


Hewlett Packard Enterprise

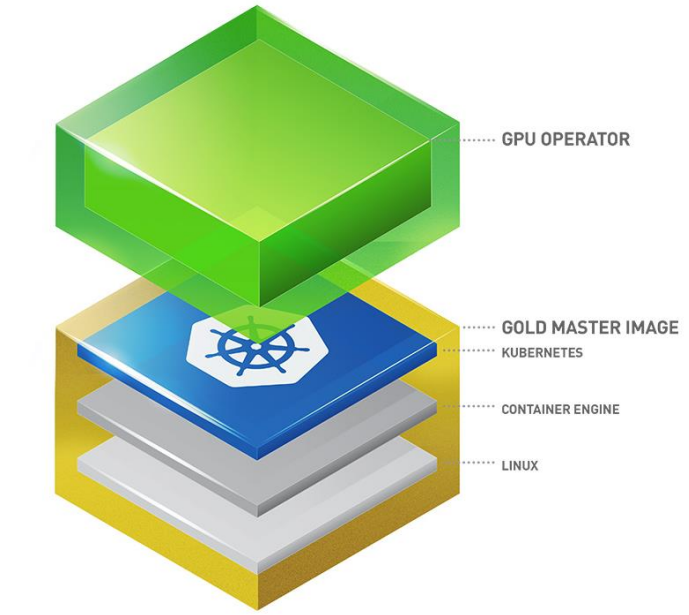
NVIDIA. DEEP LEARNING INSTITUTE

Big Data & Analytics

NVIDIA EGX platform for EDGE computir



Based on Kubernets
From Datacenter to the EDGE



Big Data & Analytics

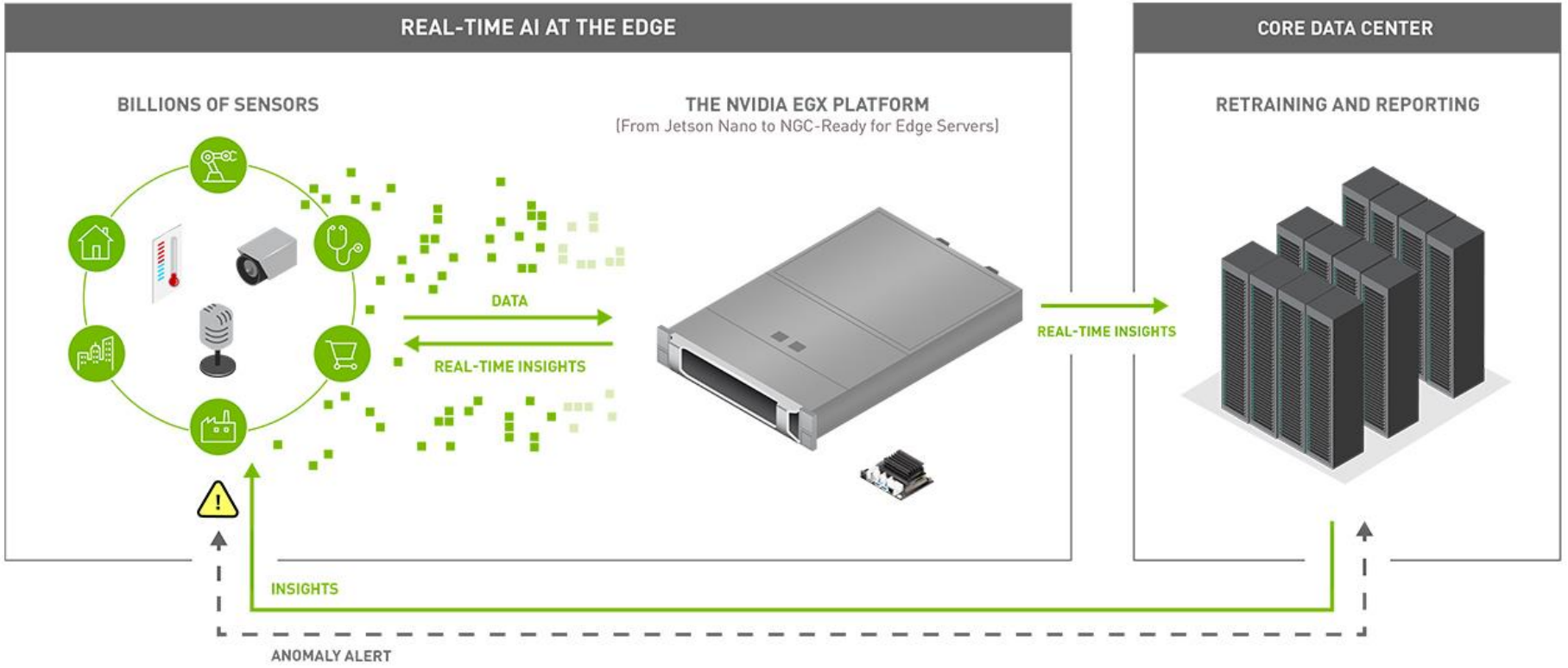
AI + IoT + 5G

Ready for field test on Jetson Nano

100$

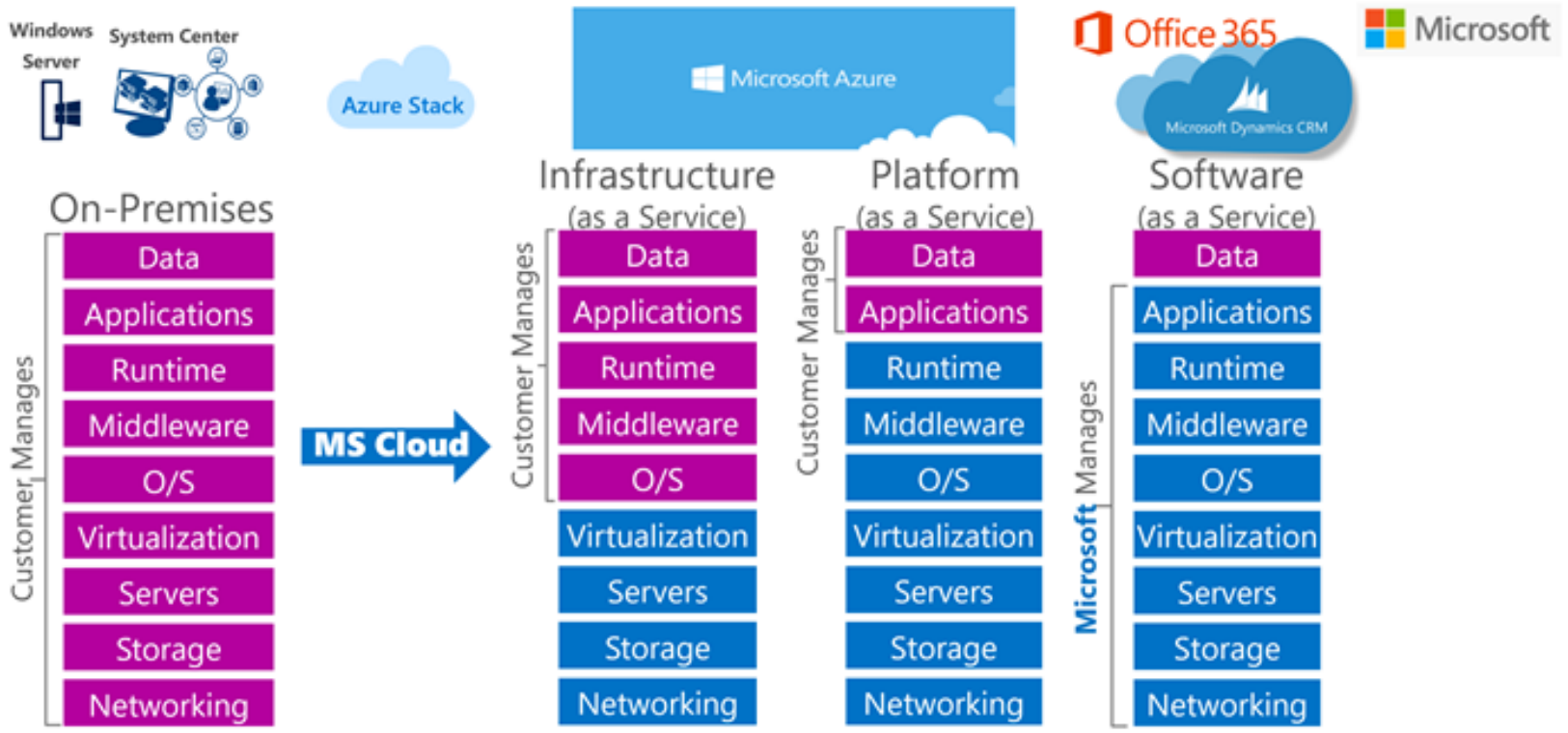
Big Data & Analytics

NVIDIA EGX platform for EDGE computing



Big Data & Analytics

DIBAF BDA + AI model



Big Data & Analytics

Supervised Machine Learning

Setting up our classifier

Field

Data in standard format					
case	Feature 1	Feature 2	...	Feature n	Class
1	xxx	x		xx	good
2	xxx	x		xx	good
3	xxx	x		xx	bad
...					...

Wheather-Climate

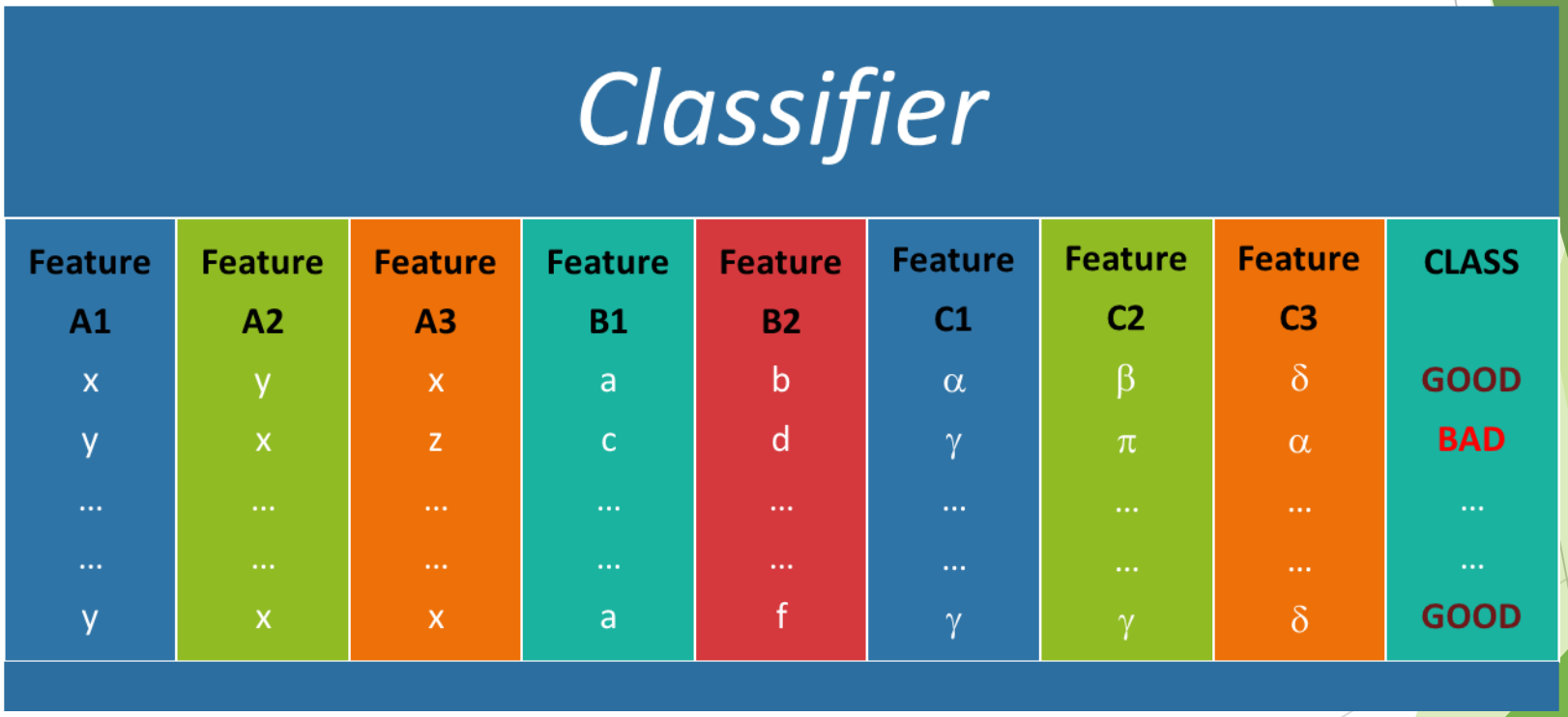
Data in standard format					
case	Feature 1	Feature 2	...	Feature n	Class
1	xxx	x		xx	good
2	xxx	x		xx	good
3	xxx	x		xx	bad
...					...

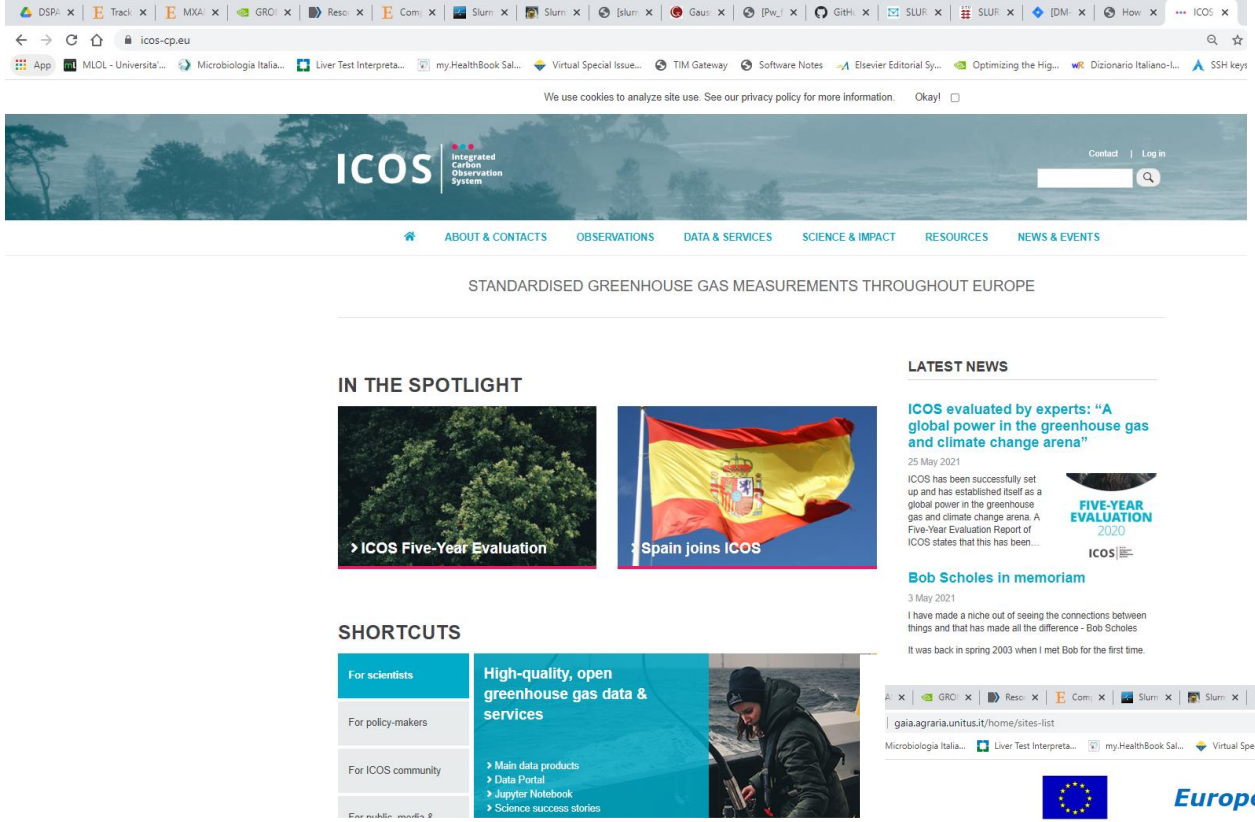
Bio-NGS/Genome

Data in standard format					
case	Feature 1	Feature 2	...	Feature n	Class
1	xxx	x		xx	good
2	xxx	x		xx	good
3	xxx	x		xx	bad
...					...

Big Data & Analytics

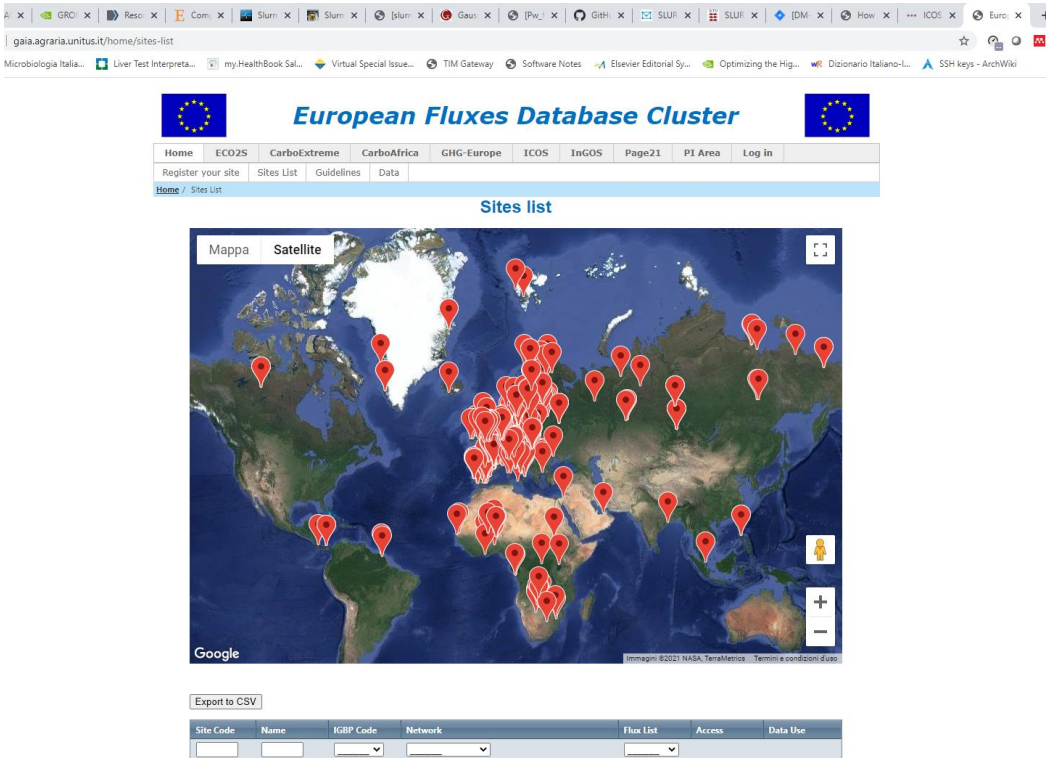
Supervised Machine Learning





ICOS Data Processing

Big Data & Analytics



Contacts Prof. Dario Papale for major details