



HEROES

Vers des places de marché pour le calcul intensif

Philippe Bricard – CEO - UCit

Atelier Teratec

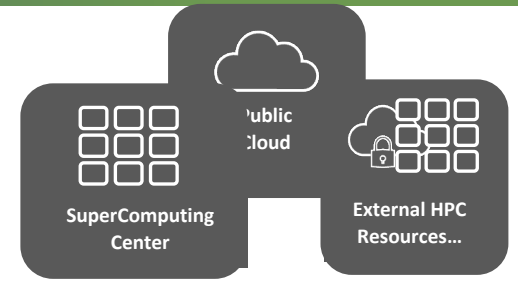
March, 24th 2022 – Paris, France

HEROES Framework at a glance

From HPC as a Service Towards HPC Marketplaces



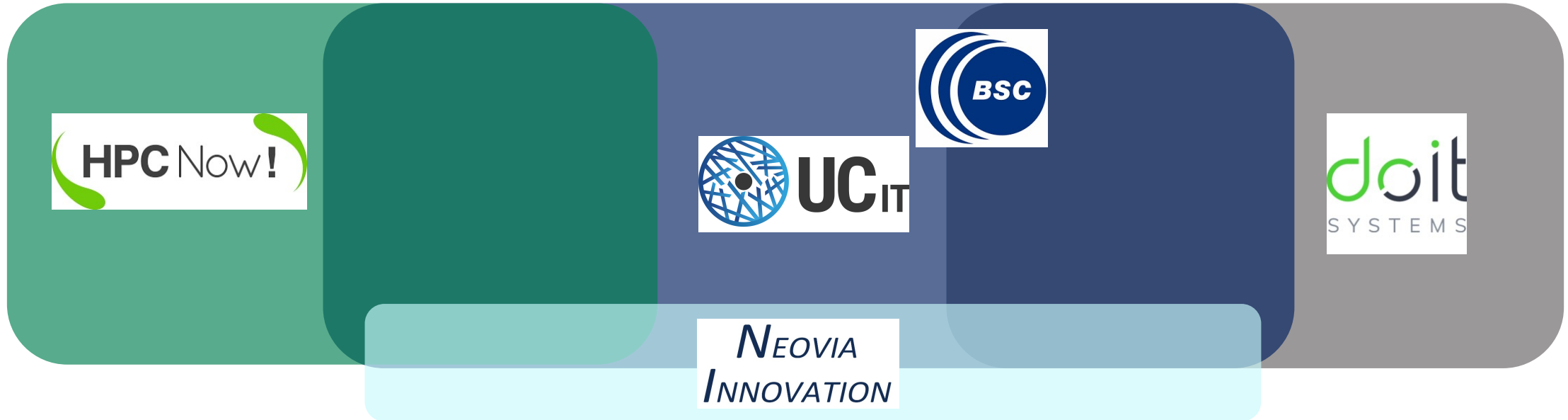
<https://heroes-project.eu>



Industry Support

Best Resource for the Job
Energy Efficient – Result in Time – Cost Optimised

Publish SuperComputing Resources



Software Framework – Ready for commercialisation
Dissemination & Communication

MARKETPLACE



A marketplace is organized : “The Mayor role”





Different
“Vendors
provide their
Products or
Resources”

Clients choose products, negotiate prices and leave the market with the products.

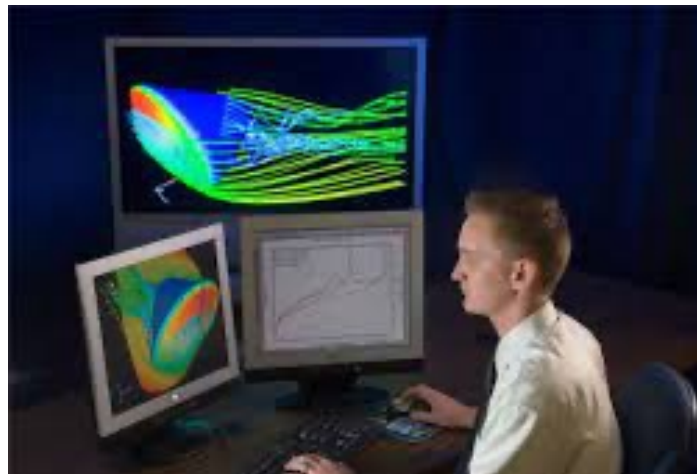
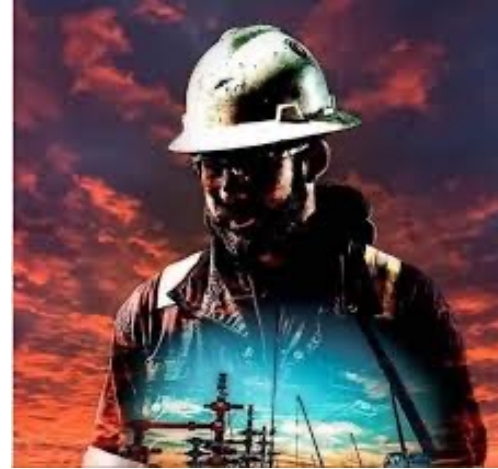


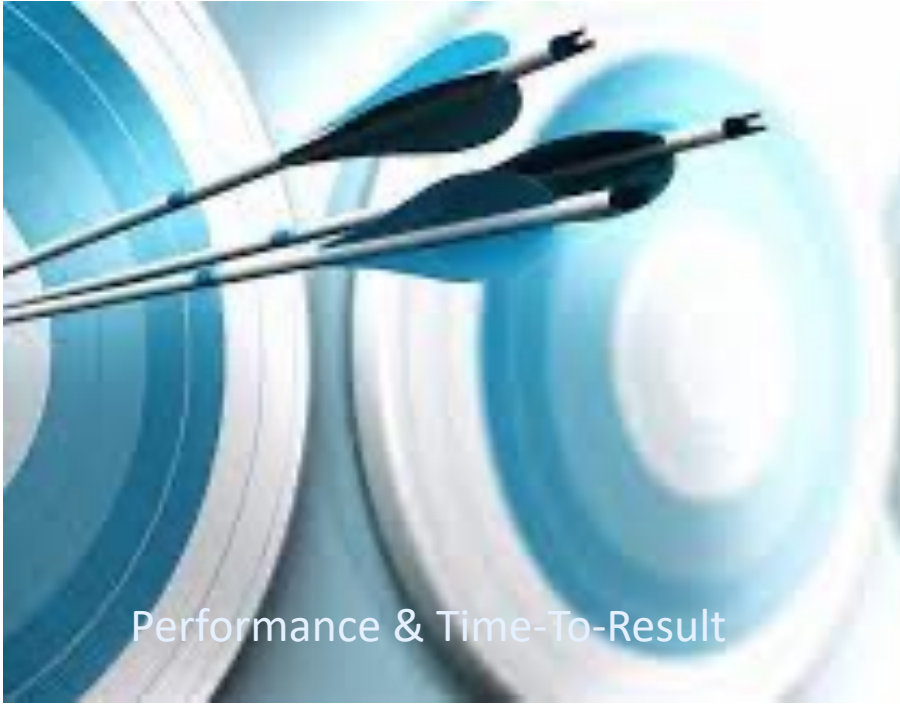
So... what would be an
HEROES Marketplace ?



The Vendor

The Client





“Mayor” : The HEROES Platform “Administrator”



“Client”: Users can bring their own workflow and if they agree on the terms, they launch the related jobs/steps

“Vendor” provides HPC Compute & Storage at their own Terms and Conditions which can vary over time

Some Vendors will provide HPC & AI Workflows

SME

Get Access to HPC Resources

Large Enterprise

Enable Hybrid HPC

Universities & Research

Get Access to HPC Resources
Publish Workflows & Software Code

HEROES Software Platform

Identification of the best Watts*€*Flops platform for the job

Improve HPC ROI

Simplify Resource Sharing

Develop New Revenue Streams

Develop New Market

On-Premises HPC Cluster

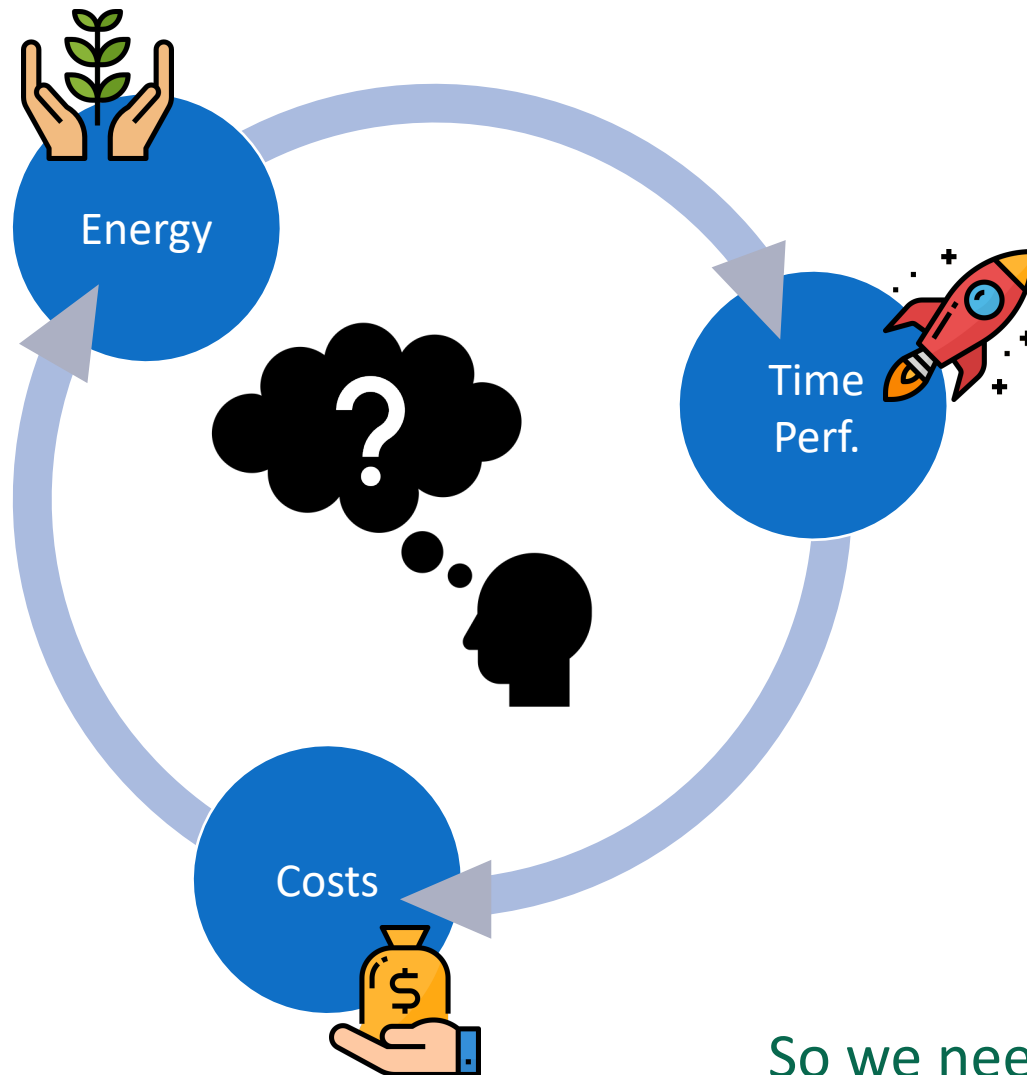
HPC Centre

Cloud Service Providers

The project is aiming at **developing in a 2 years time frame** an innovative European **software framework** allowing industrial and scientific user communities **to submit easily complex Simulation and ML workflows to HPC Data Centres and Cloud Infrastructures** as well as being able to **take informed decisions** for selecting the best platform to achieve their goals in time, within budget and with the best energy efficiency.

There will be **multiple business models** to deploy an HEROES platform (Software Model, Subscription/Support Model, Service Provider model...) and configure the Decision Module and Energy related policies

Where should I run my jobs?



First the job need to be able to run

Time-to-Result

Do I have Performance / Time constraints ?
When do I need my results ?

Costs

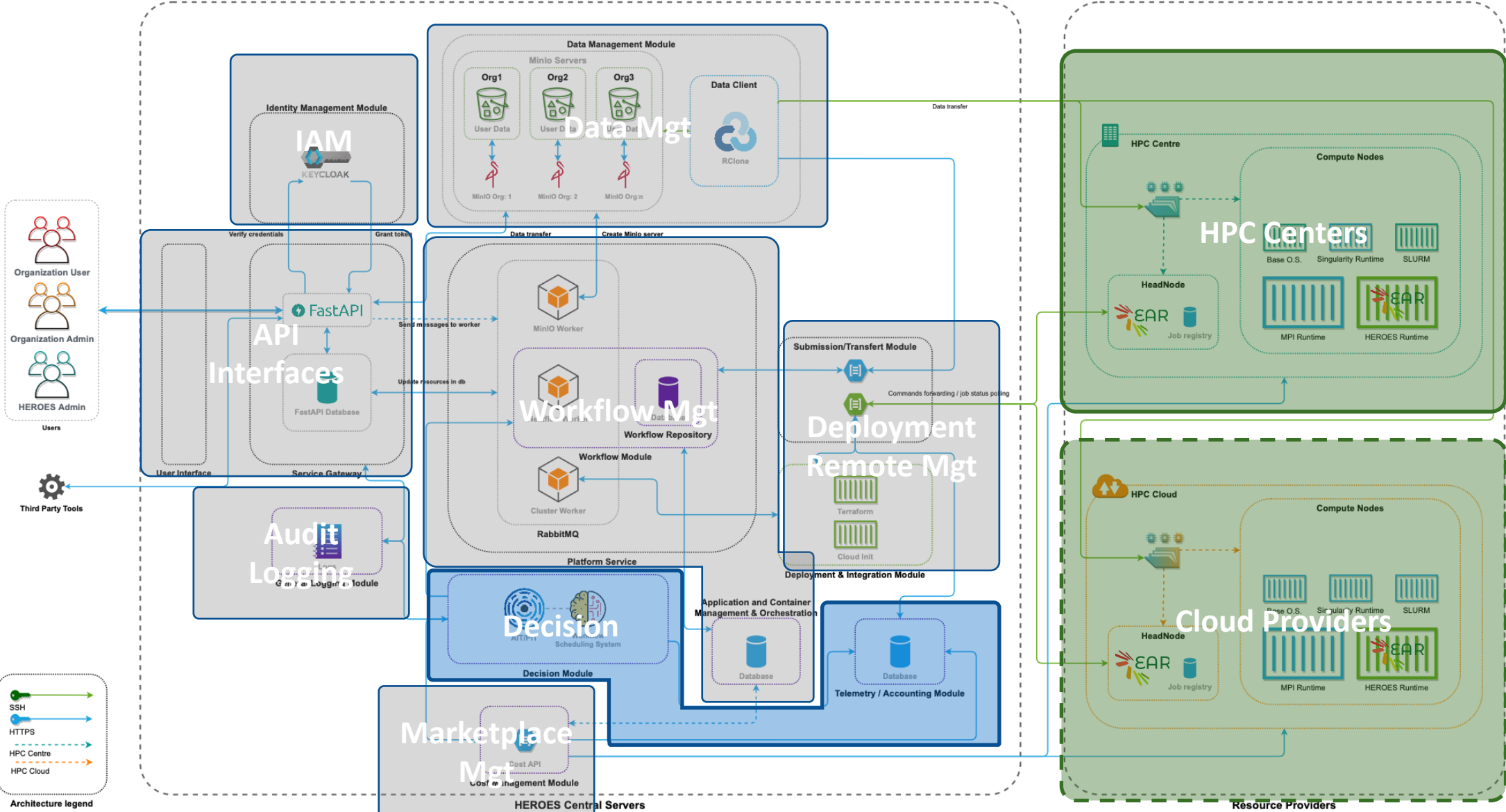
Do I have budget constraints ?
How do I express them ? €? CPU*GB*Hours ?

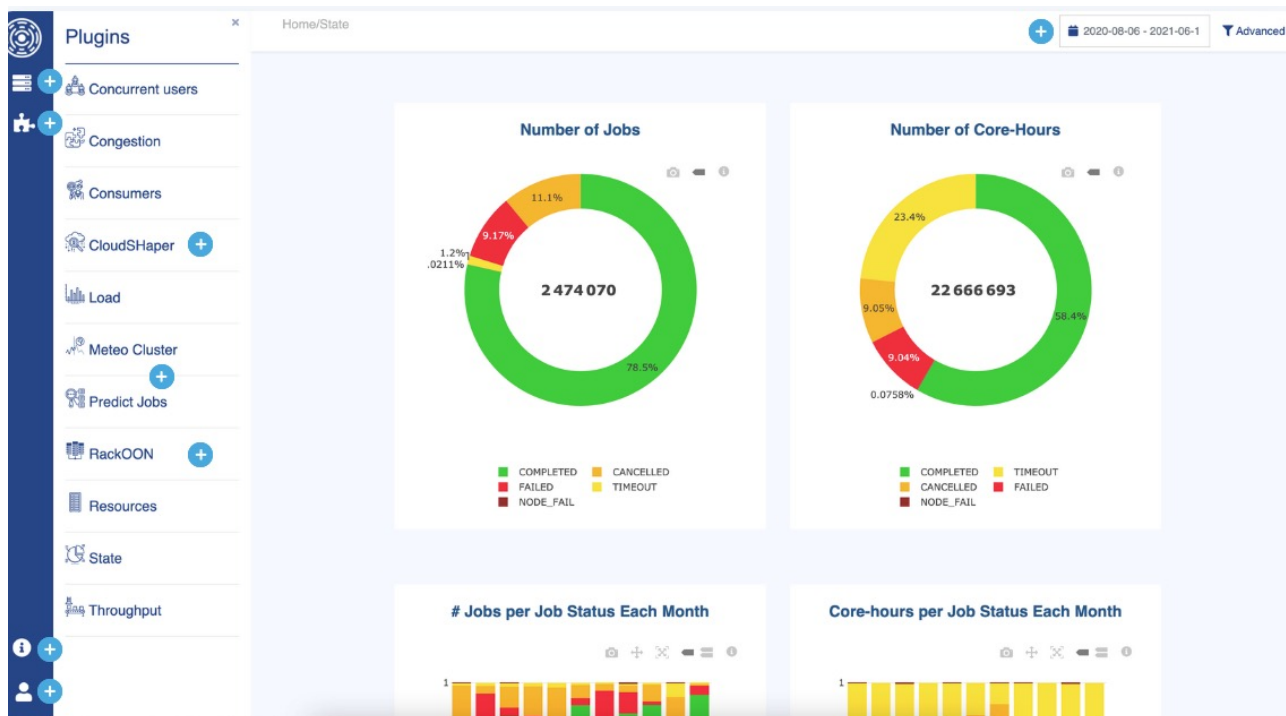
Energy/Carbon Footprint

Is it important to me ?
What is it ? How can I improve it ?

So we need to evaluate resources along these dimensions

HEROES Architecture: What about Energy ?





OKA is the Data Science platform for HPC Environments
It provides a global view of HPC clusters and their usage
OKA is extensible, adaptable and dynamic

In HEROES, OKA is the core of the Decision module:

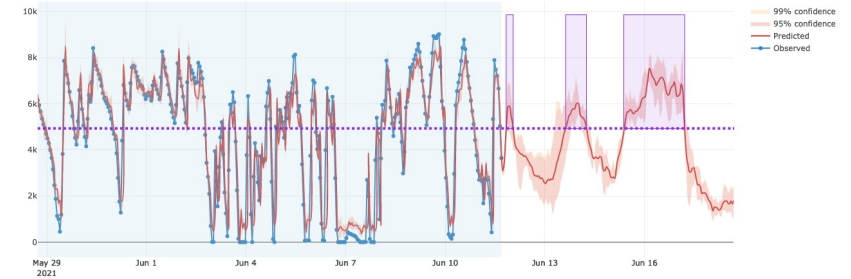
- Centralization of metrics from multiple sources
 - Jobs/Workflows accounting logs/metrics
 - Infrastructure metrics (node/core availability...)
 - Cost & Energy consumption tracking
- Provides both analytics and predictive capabilities



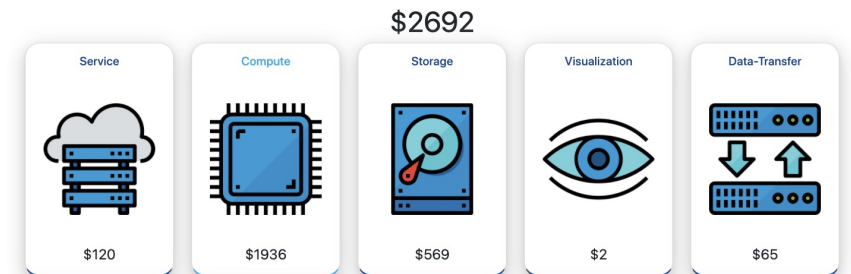
<https://oka.how>

HEROES Plugin in OKA

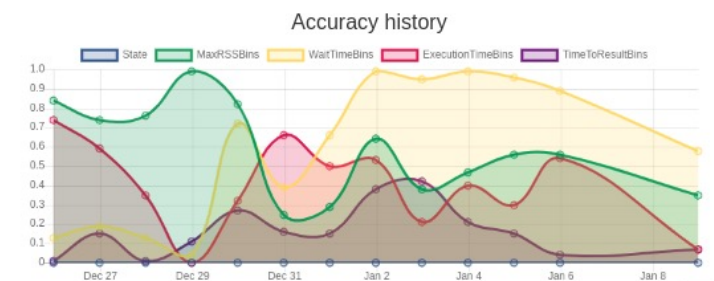
- API to get the placement (cluster selection + submission parameters) of a set of jobs based on:
 - Details about the jobs (user, job name, application, number of cores required, timelimits...)
 - An optimization choice: $f(\text{cost, performance, energy})$
- Placement algorithm will rely on
 - Constraints matching (job technical requirements, e.g., arch, min #cores/nodes...)
 - AI/ML models to predict
 - The global behaviour of HPC/Cloud platforms thanks to MeteoCluster
 - The cost associated with a workload thanks to CloudShaper
 - The needs and behaviour of individual jobs thanks to Predict-IT
- Integration options from Recommendations up to applying automated actions.



MeteoCluster



CloudShaper



Predict-IT

EAR main goal: From powerful Data Centers to Powerful and Energy-Efficient Data Centers



Be cost-effective

Consume what you need
and no more
Optimize your electricity
bill



Be eco-responsible

Resources are limited
Carbon footprint
minimization



Be energy-efficient

Understand/Optimize your
system energy
consumption
Know why!



<https://www.bsc.es/research-and-development/software-and-apps/software-list/ear-energy-management-framework-hpc>

System monitoring and Job accounting

- EAR reports power and system power consumption in heterogeneous Data Centers
- EAR reports Job energy accounting for each job submitted in the system

Powerful application performance and power monitoring

- EAR runtime library monitors performance and power dynamically without neither recompiling or application modification

Energy-Efficient system

- EAR offers runtime energy optimization
- Cluster power management
- Cluster and node powercap

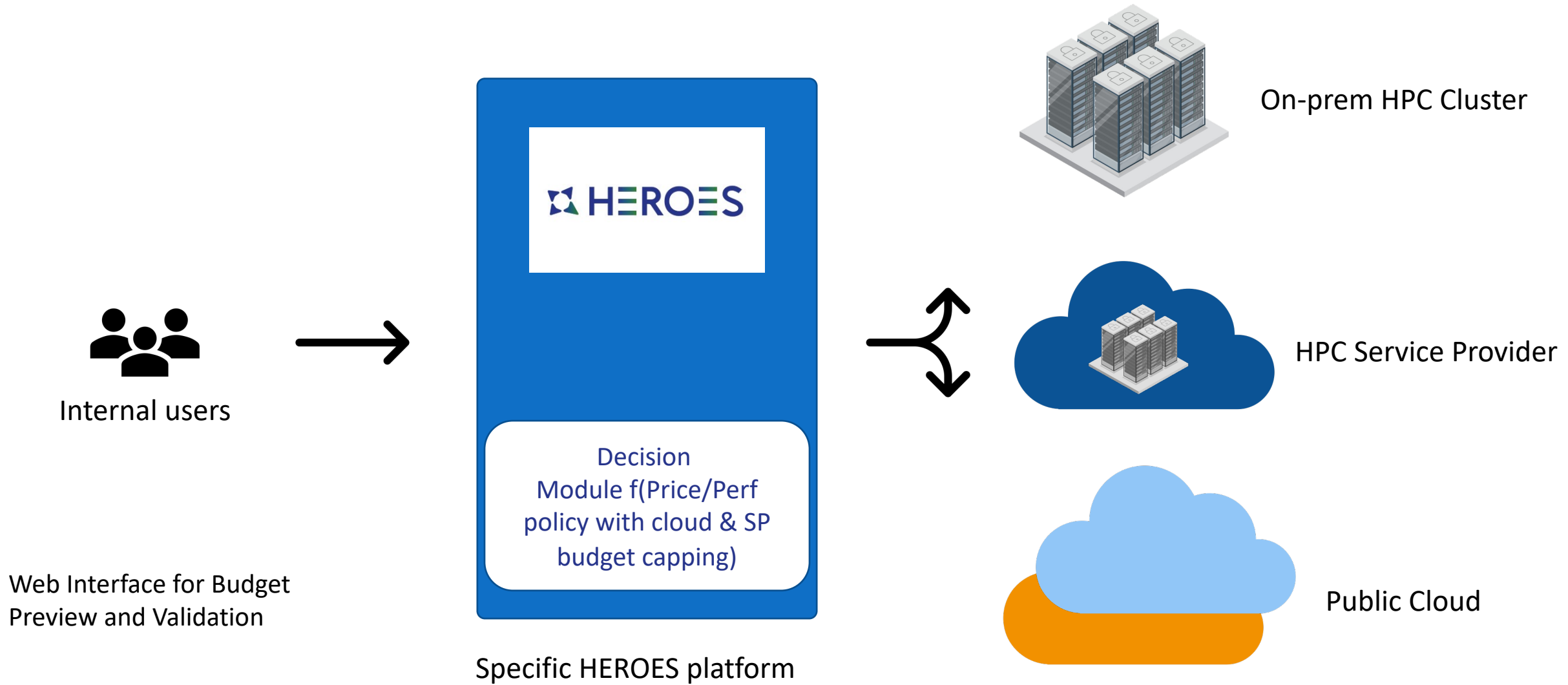
EAR architecture is being extended

- **Lightweight EAR deployment**
 - Support for Data Centers security limitations (no-root installation, no daemons...)
 - Runtime only
 - Execute seamlessly in constrained environment
 - Support for virtualized systems
 - Job monitoring only
- Provide additional **data for energy-efficient workflow scheduling**
 - Power and energy
 - Cycles, instructions, gflops, cache misses...
 - GPU usage...
- Modular and pluggable **accounting mechanism** for OKA integration

2 deployment options

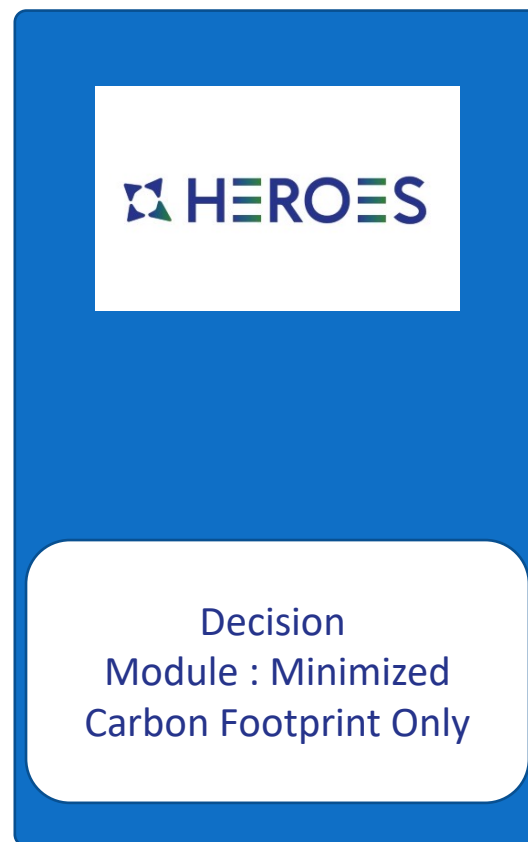
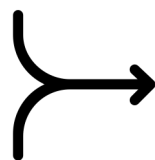
- EAR "Full", with runtime optimization of energy
- EAR "Lightweight", embedded in HEROES runtime

Example for a HEROES implementation Software infrastructure at a large client

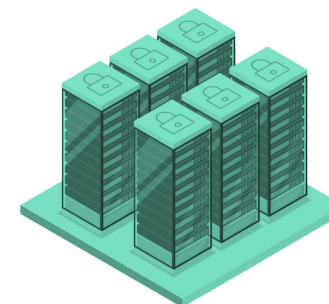
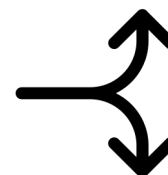


Example for a HEROES implementation

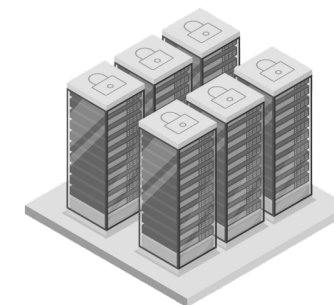
Which EuroHPC Resources is the more Eco-Responsible



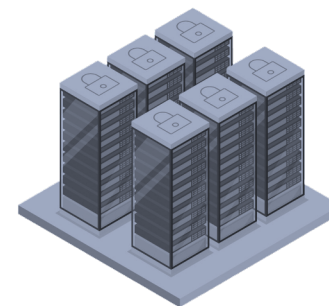
Centralized HEROES platform



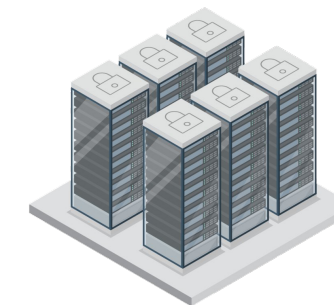
HPC Center #1



HPC Center #2



HPC Center #3



HPC Center #4

Philippe Bricard

philippe.bricard@ucit.fr



www.heroes-project.eu

heroes@ucit.fr



The HEROES project has received funding from the European High-Performance Computing Joint Undertaking (JU) under grant agreement No 956874.
The JU receives support from the European Union's Horizon 2020 research and innovation programme and France, Spain, Italy.