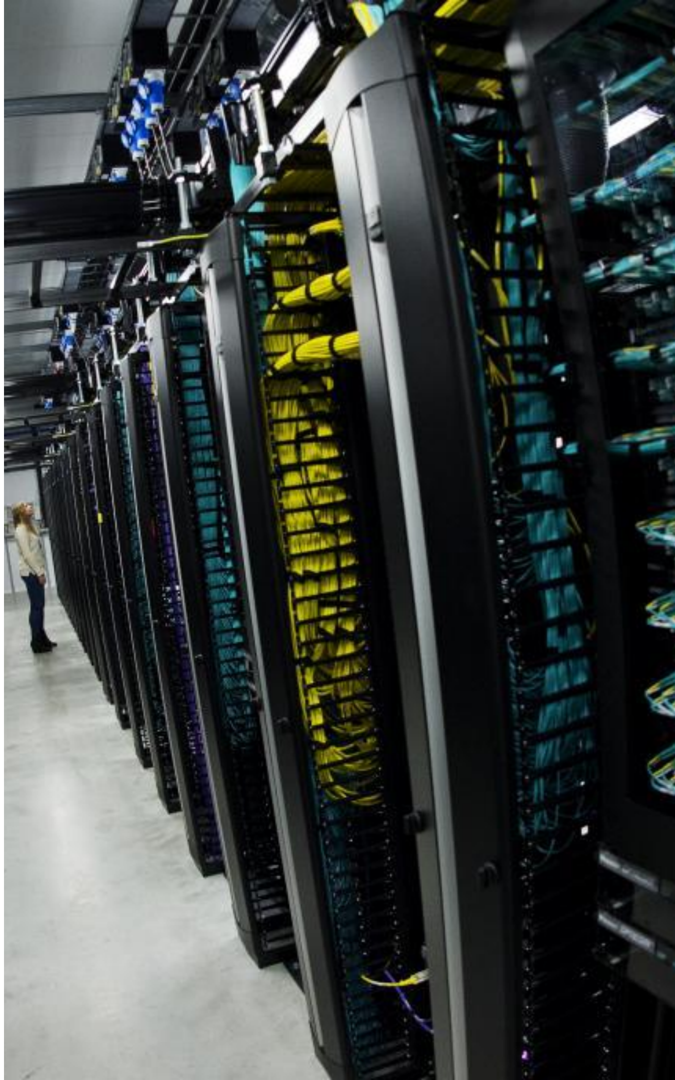
NumPEx PEPR

French contribution
to the Exascale software stack

CEA (J. Bobin), CNRS (M. Krajecki),
Inria (J-Y. Berthou)

ASNUM 2025

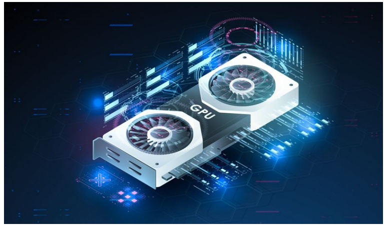
December 12th 2025 – Julien Bigot & the NumPEx team



Exascale is here

ExaFlOPs supercomputers are able to compute 2^{18} **floating point operations per second**

- double precision for HPC
- If every human on earth were to compute one operation every second
  - it would take us **2.37 centuries** to compute what this computer computes in **1 minute**
- Such computing power is a **game changer** for simulation & scientific applications

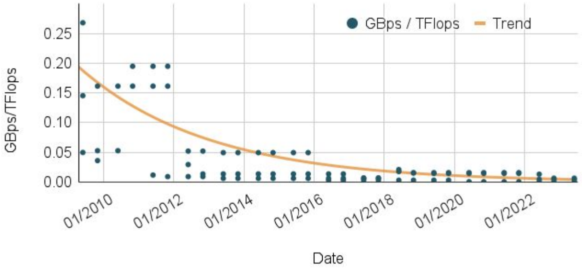


But with **great power** comes **great… complexity**

- **Computing power** is achieved by **massively parallel** nodes: **GPU**
  - One need to decompose problems with multiple levels of parallelism up to extreme fine grain
- **Memory bandwidth** is extremely **scarce** compared to computing power
  - Fetching data to compute is the new bottleneck, computing is free! (flops don't matter)
- **Disk bandwidth** & **capacity** are even more **limited**
  - You can compute huge amount of information, but don't even think to store it…

A dedicated software stack is required to leverage this

Ratio of BW/Compute for the top3 of top500



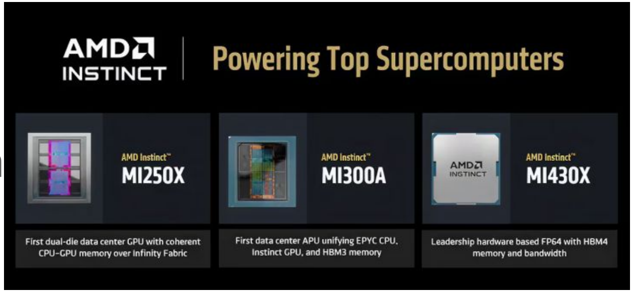
Alice Recoque, new Exascale Supercomputer

Installation in 2026, Operational in 2027

- HPL performance: **1+ Eflops HPL** (GPUs) & 30 PF CPU < 20 MW
- A system integrating **European hardware / software technologies** in terms of computing, storage, network, infrastructure, middleware, applications…
- **Addressing societal and scientific challenges** via AI, large scale numerical simulations and massive data analysis and quantum computing. A system embedded inside the digital continuum.

First NDA meeting between AMD / Eviden
and NumPEx will be organized in January 2026

Alice Recoque in a Nutshell



Eviden XH3500

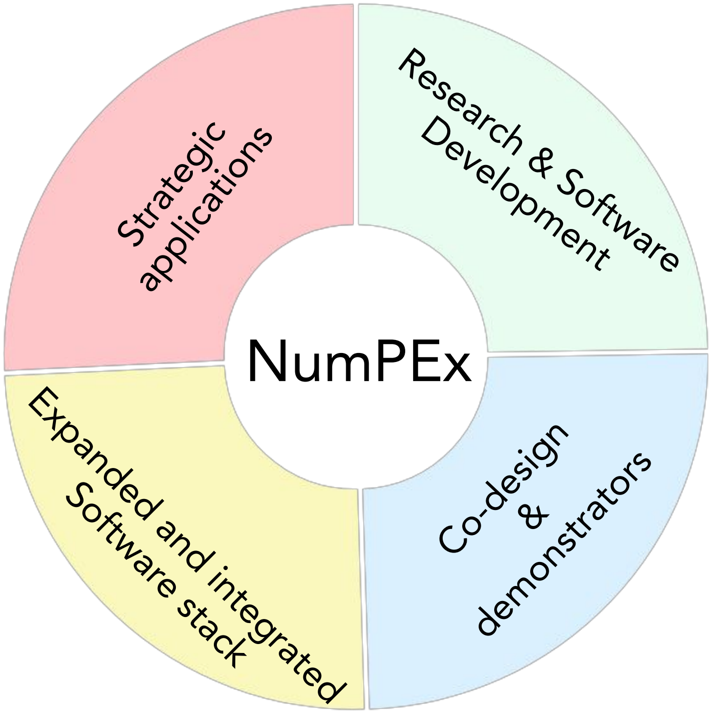
- 94 compute racks for the unified (accelerated) partition
- > 10 classic racks for scalar partition
- 100% liquid cooling (warm water cooling system)
- Power consumption range: 12-15 MW
- **Unified Compute Node**
  - Address both accelerated and scalar workloads in multi-tenant mode
  - 1 AMD Venice CPU (256c) strongly coupled with 4× AMD MI430x GPUs (4×432GB HBM4 @ 19.6To/s)
  - 1 TB of MRDIMM memory, 2 x 400 Gbps BULL BXIv3 links / GPU and 1 link per CPU
- **Scalar Compute Node**
  - Based on European ARM technology SiPEARL Rhea2 (128c)
  - > 100k cores available, specifications finalized in June

Storage (tender to follow)

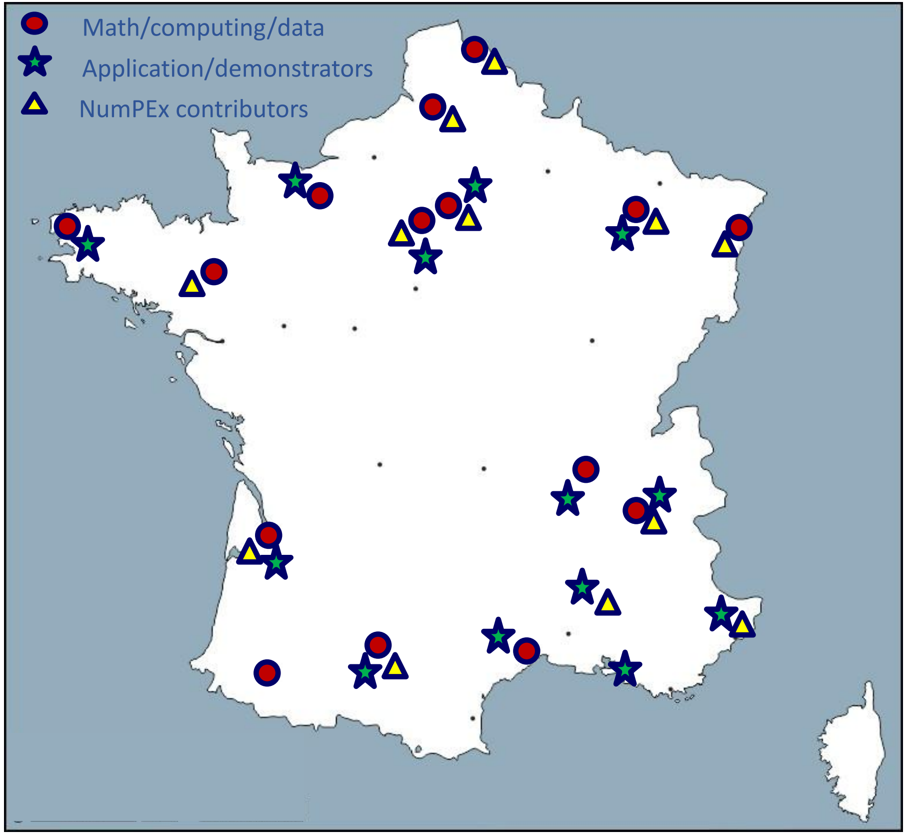
- Target: 30 PB flash, 200 PB disks

The French NumPEx Program: Objectives

- Contribute and accelerate the emergence of a **European sovereign exascale software stack** and **strategic applications exascale capability** in a **coherent framework**
- Integrate and validate **co-designed methods**, logic collection of libraries & frameworks as a **software stack** with **demonstrators of strategic applications**
- Accelerate science-driven and engineering-driven developers training and software productivity
- Foster **national and international collaborations** to prepare for the **post-Exascale era**
- Help **aggregate** the French **HPC/HPDA/IA community**



NumPEx by numbers



6 Years
41 M€*

2023-2028
* Funding 41M€=500 man.year non permanent staff
+ 170 man.year permanent staff
Total cost : 81 M€

Many Core Research Institutions

Core national Research Institutions:
CNRS, CEA, INRIA, Universities, Engineer schools, Industry

3 Focus Area

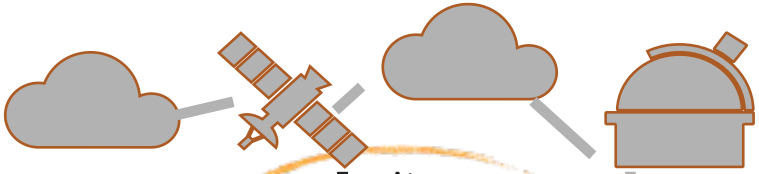
Software stack development (PC 1-3)
Wide-area workflows and architecture (PC 4)
Integration and application development (PC 5)

80 R&D teams
500 Researchers

NumPEx in a nutshell

Applications

Exa-DI
JP.Vilotte/V.Brenner
Application co-design and software integration



Exa-Atow
F.Bodin/T.Deutsch/M.Asch
Digital continuum

Exa-Dost
G.Antoniu/J.Bigot
Data

Exa-MA
C.Prudhomme/H.Barucq
Numerical methods and solvers

Exa-Soft
R.Namyst/A.Buttari
Computing

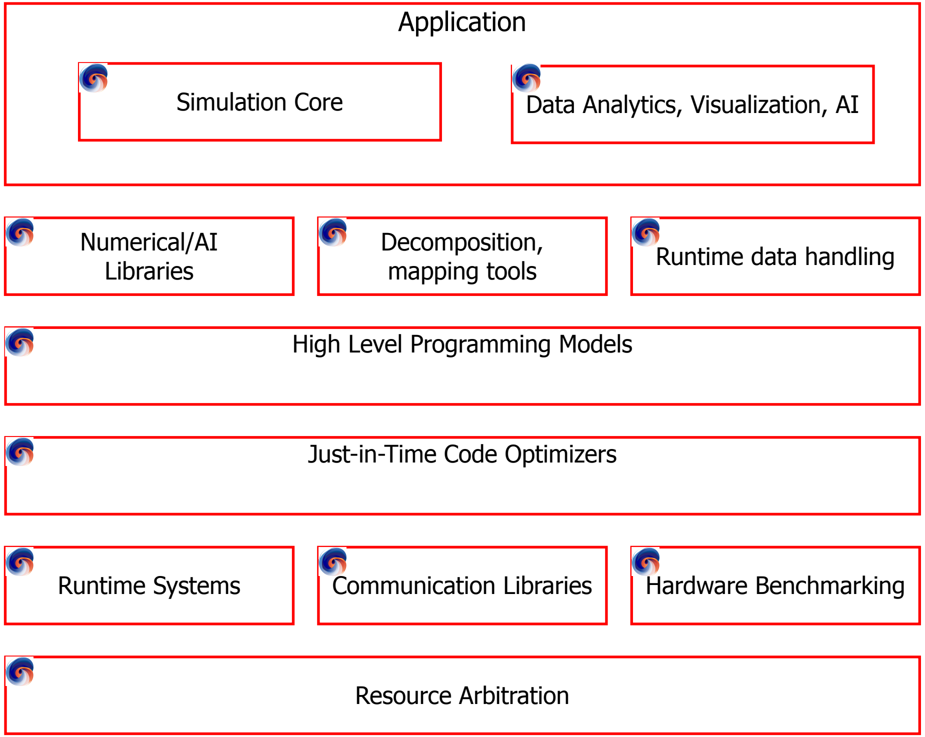
NumPEx: contributions to the stack

......

- Scalable and portable linear and multilinear algebra libraries relying on task-based parallelism
- Discretization
- Scientific Machine Learning

Runtime Systems

- Dynamic task scheduling over GPUs and CPUs
- Autotuning of task granularity
- Joblib: Lightweight pipelining in Python for embarrassingly parallel computation



Data Analytics, Visualization, AI

- scikit-learn: Machine Learning

Runtime Data Handling

- AGIOS: I/O scheduling at file level
- Damaris: asynchronous I/O
- Melissa: Online processing of data
- Deisa: Coupling MPI-Dask
- PDI: Loose coupling of simulations and libraries

Just-in-Time Optimisations

- Tiling
- Data Layout

Hardware Benchmarking

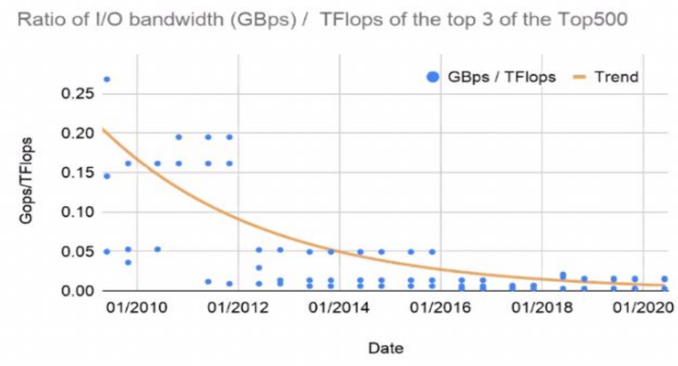
- IOPS: Automate the I/O performance evaluation process
- FIVES: Simulate high-performance storage using Lustre

Resource Arbitration

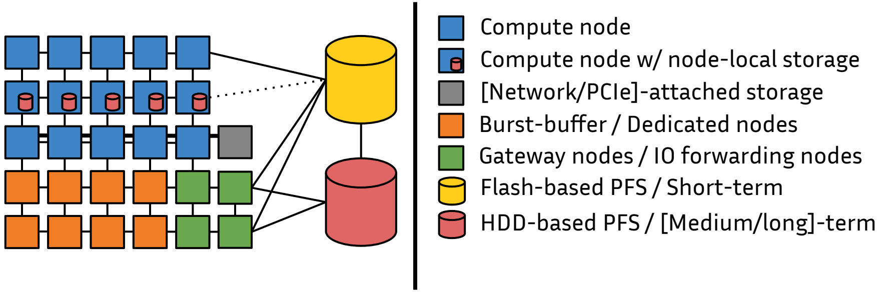
- NFS-Ganesha: High-performance and scalable NFS services
- Phobos: Long-term storage
- RobinHood: Mass action on filesystem through metadata replica

The Example of Exa-DoST: a Challenge in Hardware

- Increasing **gap between compute and I/O** performance on large-scale systems
  - Ratio of I/O to computing power divided by ~10 over the last 10 years on the top 3 supercomputers
- … and data deluge!
  - At NERSC, **data volume x41** in 10 years



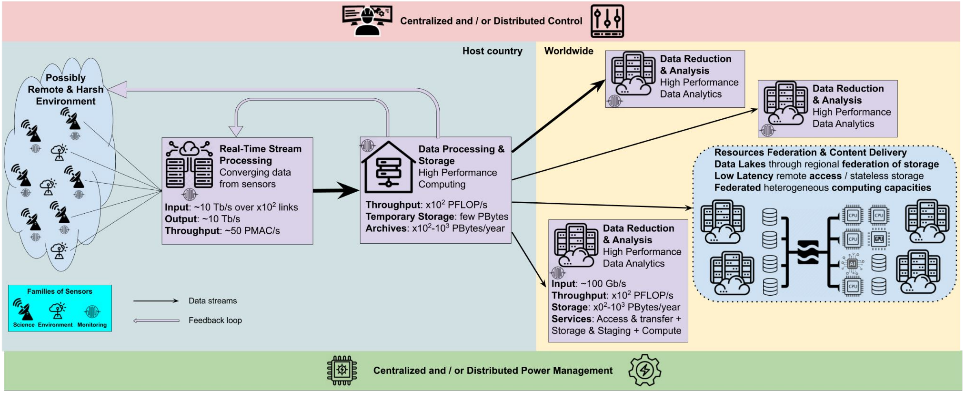
- Accelerators
  - More complex on-node memory layout
- New storage tiers and advanced architectures to try to mitigate this increasing bottleneck
  - Emerging complex applications and workflows have to adapt



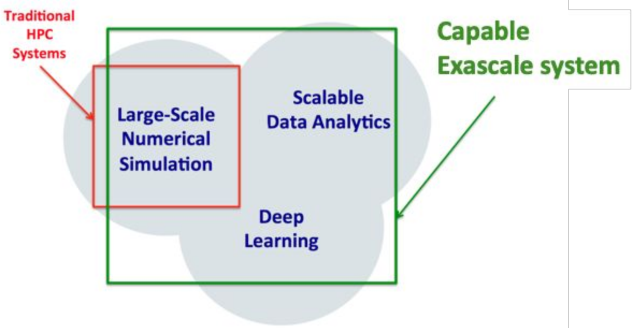
Trend in storage technologies available on extreme-scale systems

The Example of Exa-DoST: a Challenge in Usages

- HPC centers do not live in isolation anymore
  - Edge - cloud - HPC continuum
- New types of workloads
  - High-performance simulation
  - High-performance data analytics
  - Machine learning and artificial intelligence
- Interaction with data from the outside world (Cf. PC4: ExaAToW)
  - Sensors
  - Great scientific instruments
  - …



SKA data workflow from sensors to HPC centers



Exa-DoST: an ambition

Fill the gaps in the existing software stack designed by previous projects (e.g. ECP)

Take into account French & European specificities

Approach:

- **Research** on data-oriented tools for HPC
- That leads to transverse, **re-usable tools**
- Usable **in production** at exascale on Alice Recoque (BXI3, DDN, etc.) & others

Ensure French & European needs are taken into account in roadmaps

⇒ ExaDoST will produce:

- **New approaches** to handle the data challenge at exascale
- Transverse **libraries & tools** that implement these approaches

Fully application agnostic

Fully open-source

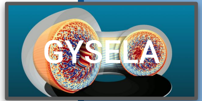
Validated in illustrators at full scale

Work Packages in ExaDoST

WP1: Exascale I/O and storage

WP2: Exascale in-situ data processing

WP3: Exascale ML-based data analytics



WP4: Shared building blocks & integrated illustrators

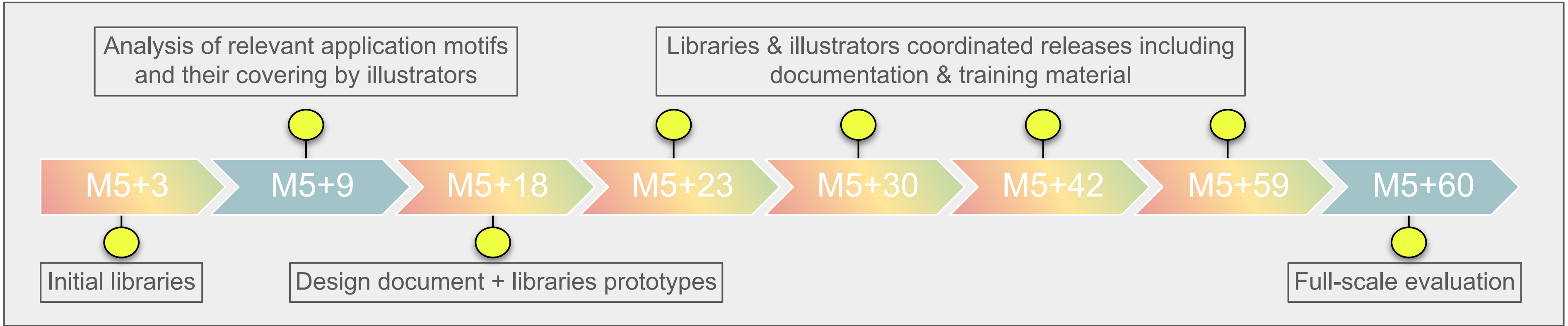
WP5: Management, dissemination and training

Exa-DoST contribution to the ecosystem

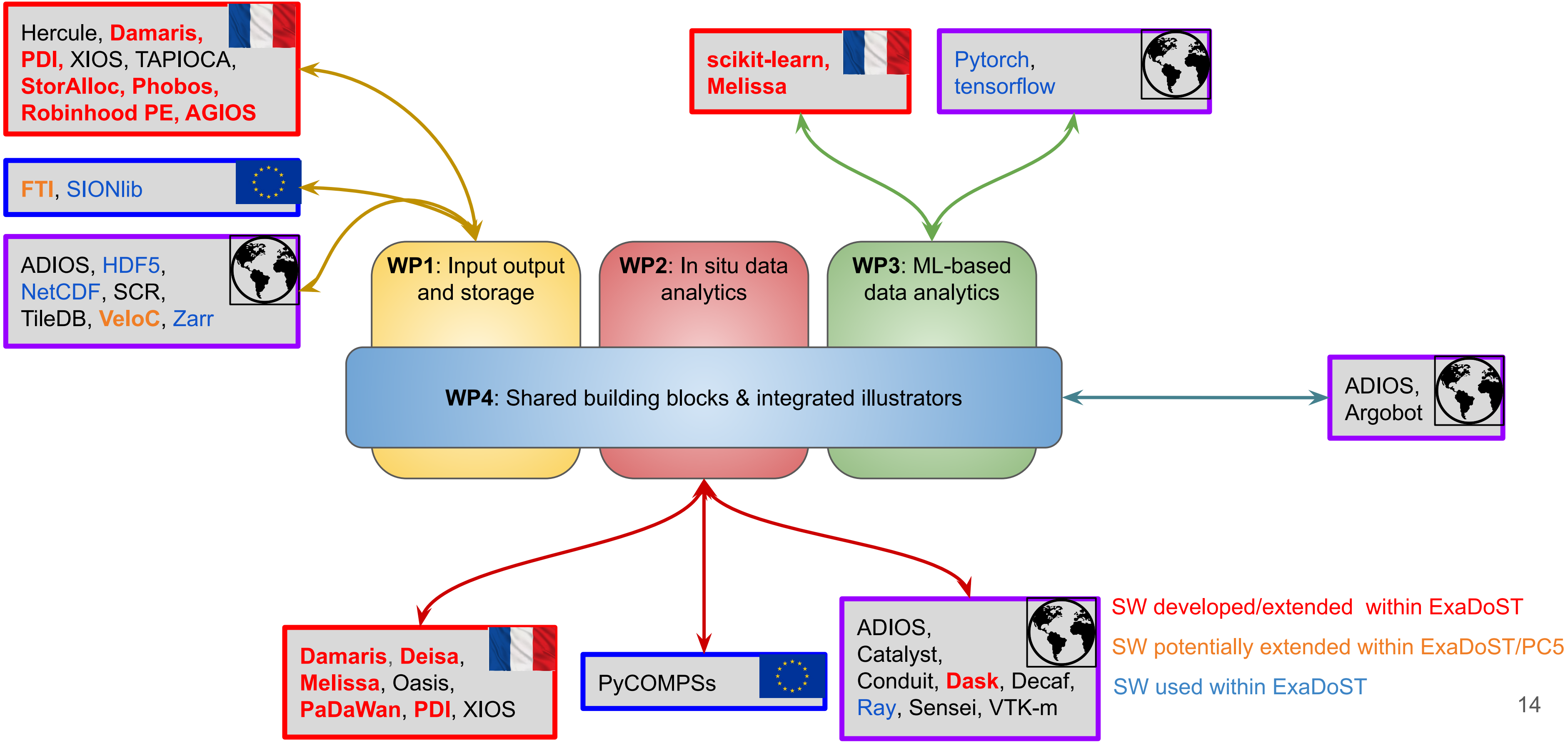
Goal: ensure French applications have the data handling software stack available to fully leverage Exascale supercomputers

- Identified libraries of interest
  - In-house and external
- Modularizing and extracting components
  - Identify and mutualize similar components
  - Add missing components
- Rebuilding libraries based on this modular approach

- Offer the community the opportunity to build taylormade data libraries & tools for their
  - Application
  - Use-case
  - Hardware
  - etc…



Exa-DoST Software Ecosystem



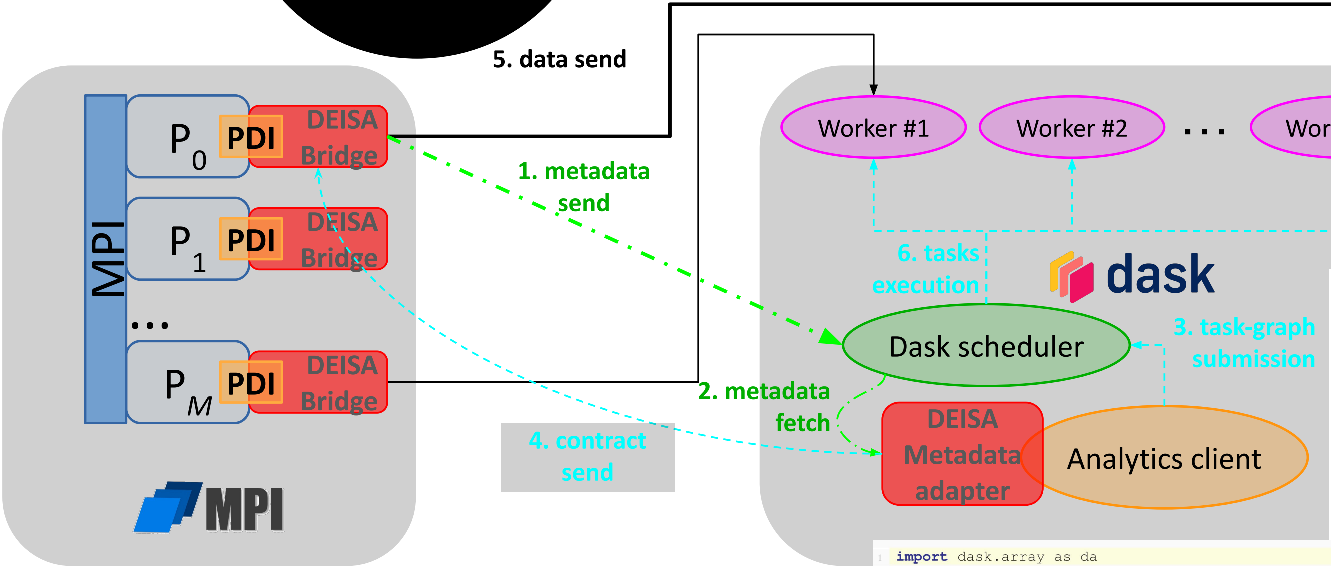
Exa-DoST: an example of work

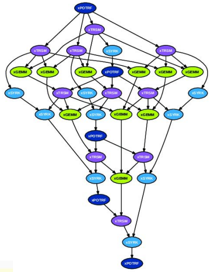
- **Post-hoc** analytics suffer from **performance issues**
  - Data transfer between simulation & analytics goes through disk
  - Network bandwidth is much better, filtering before storage reduces capacity requirements
- **In situ** analytics **solve performance issues**
  - But most frameworks use MPI-style parallelism
  - **Complex**, and badly suited to expressing analytics patterns
- Frameworks such as **Dask** have a much **nicer API for analytics parallelization**
  - Express your code in python + numpy/pandas/… with a few parallelisation hints
  - Let dask generate a task-graph automatically and schedule it over compute resources

We need to bridge the GAP between

simulation & **analytics**, **MPI** & **Dask**, **communicating processes** & **task-parallelism**

Offer users an environment for in situ analytics that is higher level than usual HPC-based ones





```python
import dask.array as da
from dask_ml.decomposition import IncrementalPCA
import yaml, json
import h5py
# Connect to Dask
sched = json.load(open('sched.json'))
client = dask.distributed.Client(sched["address"])
# load the simulation configuration
simu = yaml.load(open('simulation.yml'))
# Build a lazy array descriptor from HDF5
gtemp = h5py.File('data.hdf5',mode='r')['gtemp']
gtemp = da.from_array(gtemp, chunks=(1,4096,4096))
for step in range(0, simu['timesteps']):
  pca = IncrementalPCA(n_components=2, copy=False,
                       svd_solver='randomized')
  pca.fit(gtemp[step,:,:])
  print(pca.explained_variance_)
```

Implementation (reuse existing frameworks, “just” add the glue):

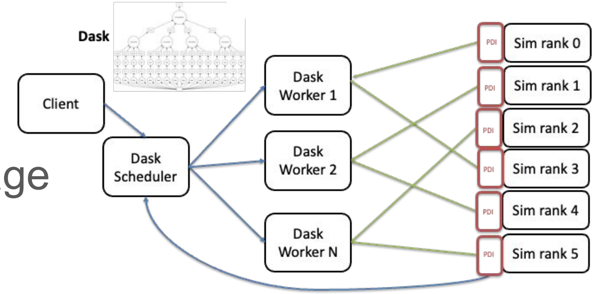
- Dask: parallel Numpy implementation based on distributed Python tasks
- PDI: app/analytics interface
- Ray: distributed Task and Actor runtime

Ongoing work on Deisa in Exa-DoST

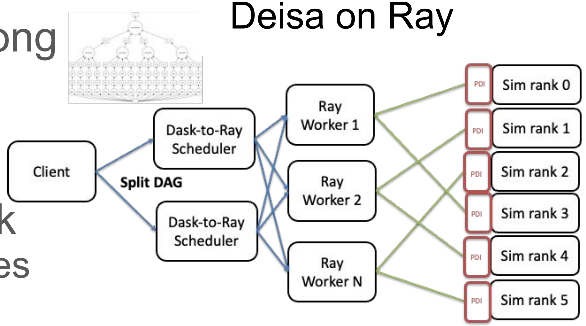
Co-design work with production simulation codes

- Gysela @ CEA & Parflow @ JSC
  - These demonstrator applications rely on us to leverage Exascale & Alice Recoque
- Exascale means scalability
  - Developing Deisa-on-Ray: leverage Dask-on-Ray (early runs at 15k cores done)
- Auto-detection of events mean uneven analytics needs along simulation
  - PhD. just started to explore load balancing & elasticity (thanks to Dask existing elasticity)
- Huge data production is even difficult to move with network
  - PhD. starting to explore advanced graph scheduling strategies

plain Deisa



Deisa on Ray



Co-funded with EU "Center of Excellence" Eocoe-III

https://github.com/deisa-project

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

FRANCE 2030 PROGRAMME DE RECHERCHE NUMÉRIQUE POUR L'EXASCALE

Transverse actions in NumPEx

- *Accelerated architectures and programming models* — *S.Thibault/M.Pérache*
- *AI* — *T.Moreau/E.Franck/J.Bobin*
- *Computing centers* — *F.Bodin/N.Lardjanne*
- *Energy management and optimization* — *A.Guermouche/G Da Costa*
- *Gender/Equity/Diversity* — *A-L Pelé/V. Grandgirard*
- *Training* — *M.Krajecki*
- *Software production and integration* — *B.Raffin/J. Bigot*
- *YoungPEx* — *PC members*
- *International collaborations* — *J-Y Berthou*

YoungPEx initiative

Leaders: T.Saigre, K. Hoogveld, M. Trochon, M. Certenais, R. Garbage

Community of about 75 people in NumPEx

- **Create a network of young people recruited within NumPEx** (interns, PhD students, postdocs, junior researchers).
- Develop a transversal community across the PCs to enlarge and enrich exchanges and foster collaborations.
- The goal is to propose original actions to be implemented within NumPEx: *Actions during NumPEx events, seminars, onboarding actions, communication, training, careers, etc.*
- Organize the actions selected by the NumPEx leaders and animate the community.
- A working group was set up with representatives from each of the targeted programs.
- Provide out-of-box ideas to build long-term vision for HPC and AI

News from the AI WG

Leaders: T.Moreau, E.Franck, J.Bobin

- **Organisation of a AISSAI/NumPEx semester dedicated to HPC/AI interplay**
- 4 events:
  - **SCOPE :** 2-days opening event, 2 focused topics : "**Foundation models for Science**" and "**AI/HPC convergence**" – Paris - march, 10-11th 2026.
  - **Ai4HPC** : 4-days workshop, dedicated to **HPC/AI hybridization and LLM for HPC** – Toulouse - may, 26-29th 2026
  - **GAP/NumPEx :** joint event with GAP, dedicated to **Inverse problems in Science** – Grenoble - June 17-19th 2026
  - July 26: Hackathon "**HPC with AI programming frameworks, focus on JAX**" – Paris - July 2026

CfP to complement NumPEx

in AI, Accelerator prog and Scientific workflows

AI4HPC – HPC4AI : SW for the efficient training of large AI models

- **DAIMOS** (900 keuro): modular, energy-efficient software stack for large-scale deep learning, integrating advanced distributed training algorithms optimized Graph Neural Network training, and reusable HPC tools.

AI4HPC – HPC4AI : Open call for AI for HPC

- **SAGE-HPC** (800 keuro): an open, scalable software platform for multi-fidelity optimization of complex physical problems on exascale HPC systems, integrating Bayesian optimization, deep reinforcement learning, and hybrid strategies—guided by AI-driven meta-learning.

Programming models for accelerated architectures

- **Koktails** (1750 keuro): an open-source software stack for Exascale GPU-based supercomputers, leveraging the Kokkos programming model and integrating AI, Python to ensure performance portability and facilitate the transition of legacy HPC applications to next-generation European architectures.

Efficient workflows for scientific data processing, the case of SKA

- **ASTRA** (550 keuro): towards interoperable distributed workflows for massive data analysis on federated infrastructures

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

FRANCE 2030 PROGRAMME DE RECHERCHE NUMÉRIQUE POUR L'EXASCALE

Koktails

Organizational Structure

- **Key Partners**: CECI, CEA, Inria, ONERA, and IFPEN.
- **Team Expertise**: Legacy code translation, GPU optimization, StarPU development, and aerodynamics simulation.

The project's DNA

- **Context**: The KOKTAILS project addresses performance portability challenges for Exascale computing on heterogeneous GPU-based systems.
- **Purpose**: Modernize legacy scientific codes to exploit diverse computing architectures while ensuring high performance, scalability, and portability.
- **European Leadership**: Reinforces Europe's position in High-Performance Computing (HPC) and digital sovereignty.

Kokotail: extending CExA beyond CEA

WP1: Tooling to support the transition to GPU with Kokkos

WP2: Leveraging AI-oriented languages and tools in Kokkos

WP3: GPU-efficient mesh management in Kokkos

WP4: Dynamic performance portability with Kokkos

WP5: Kokkos foundational support and demonstrators integration

WP0: Management, dissemination and training

CSA SPE-EU

HORIZON-CL4-2025-03-DIGITAL-EMERGING-04: Post-exascale HPC (CSA)

Expected Outcome:

- Delivery of a high-quality **roadmap** addressing the post-exascale **HPC/AI research challenges** for applications, algorithms, software, hardware and systems, including a strong emphasis on AI
- Contribution to the development of a **competitive European converged HPC/Quantum/AI ecosystem**, including AI Factories and future AI Gigafactories
- Interaction and collaboration with similar international efforts, ensuring alignment with AI-driven computing paradigms worldwide

CSA SPE-EU

Scope:

- Guide and prepare European HPC for the post-exascale era of converging supercomputing, quantum computing and artificial intelligence worlds
- Bring together the key scientific and industrial players in Europe, and should liaise with the relevant international post-exascale efforts, the EuroHPC JU private partners, relevant EuroHPC main initiatives, the hosting entities of European AI Factories and future AI Gigafactories, and other relevant European projects and initiatives
- The action should analyse the research challenges of all relevant technologies in the post-exascale/AI era and produce and maintain a high-quality research roadmap with recommendations for research actions at the European level
- Issues like hardware-supported mixed-precision, AI-driven HPC as a service, real-time HPC, next generation AI model training and inference, digital continuum, convergence of HPC/AI/Quantum/Cloud/Edge, should be part of the analysis

The International Post-Exascale (InPEx) Project

InPEx expected outcomes

- Identify future trends/disruptions, missing software components
- Contribute to the share/development of software components: deployable, maintainable, robust, sustainable => partnership factory
- Landmark documents largely exploited, worldwide, to support future post-exascale science
- Develop an international network of exascale computing experts and leaders

Actions

- Dedicated international working groups
- International Post-Exascale (InPEx) workshop series

Participants

- Researchers, engineers, industry, funding bodies

In summary

numpex.org

The world of HPC is getting **more and more complex**

Applications need to be re-**designed**

- One can not build them down to the bare metal anymore
- We need **portability**, we need **abstractions**, we need **library** and **tools**
  - **We need to extend the HPC software stack**

We have projects to contribute to the stack & MdlS takes its share of the work

- In France, a Research oriented programme: **NumPEx**
- For GPU @ CEA: **CExA** (soon to join NumPEx?)
- Worldwide to gather bits & pieces of the stack: **HPSF**

Beyond these existing projects we look for **collaborations**

- With communities that have identified **shared challenges**
- And that want to **work together** to solve them