Au service de
la recherche scientifique,
l’innovation,
et la compétitivité des entreprises

GENCI update
3 national centers (CINES, IDRIS, TGCC)

- Occigen 3.5 pflops since 2015
  Montpellier
- JOLIOT CURIE (50% for PRACE)
- BULL Sequana 9.5 Pflops
  - 6 cells Skylake 24 c : 79 488 c
  - 3 cells KNL 68 c : 56 304 c
  - 400 To de mémoire,
  - Débit disque global 500 Go/s.
  - Upgrade >20 Pflops end 2019

14,5 Pflop/s

*1.9 every year

Multi year investment plan
- 2018 TGCC renewal / 2019-2020 extension
- 2019 IDRIS renewal / 2021 extension
- 2020 CINES renewal / 2023 extension (?)
- ...
- 2022 Exascale system

Evolution every 10 months
Complementary architectures
CONVERGENCE HPC - AI

In France a fast move toward AI

☐ Context

- End March 2018: Publication of Villani’s report and conference “AI For Humanity” in Paris
  - Creation of several AI multi-disciplinary institutes (3IA)
  - Talent generation and retain
  - Foster AI use in health, transport, defence and environment
  - Address issues about AI explicability and ethics
  - **Dedicated HPC facilities for French AI researchers**
- Ministry asked to GENCI to integrate dedicated AI facilities into ongoing tender
  - French AI researchers as first circle
  - AI users then as a second step: climate, fusion, combustion, cosmology, materials, …

☐ Approach

- Working group GENCI, IDRIS, AI experts -> needs, access mode, benchmarks…
- Joint definition of converged node
  - Able to run HPC and/or AI workloads
  - Interest of newest GPU tech. (HBM, nvLINK/nvSWITCH, tensorcores), 32 GB mem, …
  - Mono GPU → multi GPU → multi nodes
  - **Target: scale up of xNN learning**
- Containerized AI stacks
EVOLUTION OF IDRIS HPC FACILITIES

Introducing Jean Zay

- **HPE SGI 8600** solution
  - **Scalar partition**: 5.08 Pflops peak
    - 1528 bi sockets compute nodes
    - Intel Cascade Lake 6248 20-core 2.5 GHz
    - 192 GB of DDR4-2667 memory (4.8GB/core), 1 OPA100 link
  - **Converged partition**: 9.05 Pflops peak (FP64), 130 Pflops peak (FP16)
    - 261 dual socket scalar nodes with each 4 nVIDIA V100 32 GB, 4 OPA 100 links
    - A total of 1044 GPU
  - **Interconnect**: Intel OPA100
    - Topology: *Enhanced Hypercube*
    - 100 Gb/s
  - **Fat nodes for pre/post processing**
    - 4 Superdome nodes – 4 SKL 12c 3.2 GHz, 3 TB, 1 NVMe 1.6 Bo, 1 GPU nVIDIA V100 with possibility to aggregate into one supernode
  - **1st level storage**: 6 couplets DDN GS18k – IBM Spectrum Scale
    - Full flash (SSD), > 300 GB/s, >1 PB
    - Completed end 2019 by 2nd level storage on discs >15PB
EVOLUTION OF IDRIS HPC FACILITIES

Jean Zay

- **Software environment**
  - Based on RHEL 7.2
  - HPE Performance Cluster Manager
  - Compilers Intel, PGI, outils ARM Allinea
  - SLURM as scheduler
  - AI support: integration of containers (Caffe, TensorFlow, Scikit-learn, PyTorch and Keras)
  - Ongoing evaluation of portals (ATOS Codex AI, Activeon ProActive, …) and Grid’5000 tools

- **Physical Integration**
  - 32 racks (18 compute racks + 9 C-Racks + 5 CDU) + 6 racks for storage and service
  - 76 sq, 27 tons
  - Max electrical consumption: 1371 kW, 1MW typical
  - Cooling DLC – hot water 32° for 90% of the configuration
New access modes

- Process of attribution of resources at IDRIS based on 2 tracks: DARI and Dynamic Access
  - Outside usual DARI process, permanent open call
  - Application process lighter and faster
  - Validation by the center
  - **No additional expertise** if request < 10kh GPU and <10% of the AI partition otherwise confirmation by an expert of CT10
  - **No constraint of regular consumption**
  - Possible to make advanced reservations

- Smart optimisation of resources and adjustment between HPC and AI partitions on the GPU nodes
EVOLUTION OF IDRIS HPC FACILITIES

Jean Zay

Next steps
- April 2019 : delivery & et installation
- March-May : pilot study toward portals within 3 centers
- June 2019 : start of acceptance and Grands Challenges
- October 2019 : machine in full production

Contract of progress
- Objective : port during 1yr a set of representative applications on GPU
  - 6 apps = DYNAMICO, YALES2, AVBP, RAMSES, BMW QCD and ThinkerHP
  - OpenMP and OpenACC preferred
  - Ongoing work with chemistry community (ABINIT, GAMESS-US, GROMACS, NAMD, GAUSSIAN, NWChem, TeraCHEM, BigDFT, etc) already ported
  - Performance expected : 4x node to node
- Work on the integration of AI tools (containers, portals, workflows, Kubeflow, …)
- Involvement of IDRIS and GENCI’ funding of HPE’ user support HPE (4 ETP)
- Decision end 2019 : To grow scalar partition or converged one ?

End 2019 : new GPU fat nodes for ML oriented needs
Later : evolution of the configuration planned end 2020
ONE MORE THING!
Some PRACE news

Best Practice Guide - Deep Learning
Damian Podarcanu, SURFsara, Netherlands
Valeriu Codoreanu, SURFsara, Netherlands
Sandra Aigner, TUM, Germany
Caspar van Leeuwen (Editor), SURFsara, Netherlands
Volker Weinberg (Editor), LRZ, Germany
Version 1.0 by 12-02-2019

PRACE (Partnership for Advanced Computing in Europe)
818 abonnés
1 sem.

Best Practice Guide on modern interconnects: This guide will give an overview of the most common types of interconnects in the current generation of HPC systems. It will introduce the key features of each interconnect. 

Voir la traduction

PRACE CALL 19 is open → 30 April 2019
2 billion core hours for Open R&D projects in academia and research
6 Tiers0 available
10% resources pre-reserved for industry, pilot for interactive data services with FENIXRI

19/03/2019