

Le groupe « Cellule de Veille Technologique », coordonné par GENCI, fédère des experts de la communauté HPC française dont : le CEA, le CNRS, Inria, la Maison de la Simulation, le Groupe Calcul ainsi que les centres de calcul nationaux (CINES, IDRIS, TGCC). L'objectif de ce groupe de travail est d'évaluer la facilité d'utilisation de standards (langages, outils) de programmation afin de porter et d'optimiser des applications sur les nouvelles architectures de calcul.

Par un accès à des systèmes prototypes en avance de phase, ces travaux permettent aux communautés utilisatrices de se préparer à utiliser ces architectures et à GENCI et aux centres nationaux de pouvoir mieux appréhender l'adéquation de ces nouvelles architectures aux besoins des communautés.

## 1. Quels sont les projets concernés ?

L'ensemble des porteurs de projet de la communauté scientifique nationale peut demander un accès aux plateformes prototypes via un accès préparatoire. Pour cela, il vous suffit de déposer une demande d'accès préparatoire sur le site <https://www.edari.fr> **pour chacune** des machines suivantes :

- Frioul
- Ouessant

Les ressources de calcul seront allouées à toute équipe souhaitant porter son code sur ces deux architectures et s'engageant à faire un retour d'expérience à l'équipe GENCI en charge du projet.

## 2. Quel est le montant et la durée des allocations sur chaque machine ?

L'allocation d'heures est une enveloppe fixée par le groupe de travail « Veille Technologique » et correspond à un nombre d'heures suffisant pour effectuer des tests sur la machine complète pendant plusieurs jours.

Afin de tenir compte de l'architecture particulière de ces machines, l'unité d'allocation est l'« heure.nœud » (et non pas l'habituel « heure.cœur »). Un accès préparatoire permet d'obtenir une allocation forfaitaire d'un montant de :

- **5 000 heures.nœuds** sur Frioul
- **1 250 heures.nœuds** sur Ouessant

La durée d'allocation est fixée à **6 mois**.

### 3. Quelles sont les architectures disponibles ?

Deux premières machines prototypes sont actuellement disponibles dans le cadre du projet de veille technologique. Les prototypes Ouessant et Frioul sont respectivement installés à l'IDRIS et au CINES. Ces deux machines aux architectures complètement différentes présentent tout de même le point commun d'exprimer un très fort parallélisme, induit par le nombre très important de cœurs de calcul.

Leurs caractéristiques techniques sont synthétisées dans le tableau suivant :

	Ouessant	Frioul
Vendeur	IBM	Atos
Type de nœud	OpenPOWER « Minsky »	Intel Knight Landing (KNL 7250)
Cœurs par nœud	20 cœurs Power8 + 4 GPU nVIDIA P100 couplés via Nvlink	68 cœurs x86
Réseau d'interconnexion entre les nœuds	Infiniband EDR	Infiniband EDR
Nombre de nœuds disponibles	12	54
Puissance crête	260 Tflops	164 Tflops

Le passage à l'échelle des applications sur ces architectures est au centre du travail demandé lors de cet accès préparatoire.

### 4. Objectif : les directives

L'objectif de ce projet est de permettre aux utilisateurs de prendre la main sur de nouvelles architectures de calcul mais aussi d'appréhender de nouveaux paradigmes de programmation.

La solution commune retenue pour un ensemble d'architectures hétérogènes est l'utilisation du langage par directive. Le portage en OpenMP 4.x ou OpenACC est un travail qui devra être effectué sur les machines cibles.

Pour la plateforme OpenPOWER, des résultats de portage brut en CUDA seront intéressants (pour fournir des performances théoriques atteignables) mais l'objectif reste tout de même de permettre à la communauté d'utiliser des machines hétérogènes avec une unité de l'implémentation.

### 5. Retour d'expérience

Les retours d'expérience demandés par GENCI ont pour objectif de permettre à l'ensemble de la communauté scientifique de prendre en main ces architectures de nouvelle génération en s'appuyant sur ces enseignements.

Les questions que tout projet doit se poser sont les suivantes :

- ***Quel est le bilan de l'utilisation des plateformes pour mon application ?***
- ***Quelles sont les bonnes pratiques à suivre pour permettre à mon application de tirer le meilleur parti des plateformes ?***

Afin de répondre à cette question, un certain nombre de points peuvent être évoqués :

- Combien de temps prend un portage sur cette architecture ?
- Par rapport à une architecture actuelle, ai-je un gain de performance ?
- Quel est l'environnement logiciel disponible qui permettrait de facilement obtenir de la performance ?
- Comment dois-je agencer mes données pour tirer parti de ces architectures ?
- ...

Ces retours d'expérience seront effectués en fin d'accès préparatoire. En plus des équipes support de chacun des centres hébergeurs, les équipes d'Atos, Intel, IBM et Nvidia seront en mesure d'apporter du support spécifique aux utilisateurs des plateformes.