



**Vision historique:
Evolution du HPC et de l'IA depuis 10 ans**

Comité scientifique d'ORAP

Forum des 30 ans d'ORAP
BNF, 14-15 octobre 2024

ORAP 1994-2014

- **ORAP créé en 1994 : Promouvoir le calcul massivement parallèle → promouvoir le calcul haute performance**

- **10 ans, en 2004 : Retard sur les investissements calcul intensif**

Rapports de l'ORAP et de l'Académie des Technologies:

→ nécessité d'une politique nationale sur le HPC

Rapport Héon et Sartorius en 2005 et soutien du Ministère de la Recherche

Implication de la communauté scientifique dans les « *scientific cases* »



- **2004-2014 : Investissements importants au niveau national & et au niveau européen**

2006 Création de GENCI: sécurisation des investissements et soutien financier du Ministère

2008 PRACE sur la liste ESFRI et investissement national de la France dans PRACE

2012 Communication de la Commission Européenne sur l'importance du calcul haute performance

2014 Programme R&D sur le calcul à côté de PRACE (Centres d'excellence et ETP4HPC)

2014, le Forum des 20 ans d'ORAP: vision de l'avenir

- **Architectures:** un futur moins certain
- **Exaflop** à l'horizon lointain avec l'enjeu de la consommation d'énergie
- **Accélérateurs:** Importance croissante, un enjeu pour les applications
- **Le co-design** architectures/applications: vu comme la solution
- **Applications** : être plus agile, avoir des codes plus faciles à porter; des codes communautaires
- **Méthodes et algorithmes:** besoin d'innovations pour traiter le volume des données, besoin de techniques de mise en œuvre (parallélisation, workflow, interface d'utilisation).
- **Le big data:** un enjeu important pour de nombreux domaines applicatifs
- **La formation** & le développement des compétences

2014-2024

Réalisé

- **Architectures:** un futur moins certain
S'est concrétisé et de plus en plus incertain
- **Exaflop** à l'horizon lointain avec l'enjeu de la consommation d'énergie
Exaflop est là et l'énergie mieux maîtrisée (à proportion)
- **Accélérateurs:** Importance croissante, un enjeu pour les applications
De plus en plus critique
- **Le co-design** architectures/applications: vu comme la solution
Reste limité dans les applications HPC
- **Applications** : être plus agile, avoir des codes plus faciles à porter; des codes communautaires
Succès de PRACE; développement des Centres d'Excellence
Mais accélérateurs toujours un enjeu : rapport « SP3 » de 2022
- **Méthodes et algorithmes:** Besoin d'innovations et de techniques de mise en œuvre
Des progrès: minimisation des communications, algorithmes
Des workflows correspondant aux usages et à la circulation des données
- **Le big data:** un enjeu important pour de nombreux domaines
Plein développement des centres de données
- **La formation & le développement des compétences**
Toujours clé; s'est développé en Europe avec les PATC
Puis à partir de 2020, premiers efforts EuroHPC (EUMaster4HPC, Centres de compétences)

2014-2024 au-delà des anticipations

- **L'investissement massif de l'Europe** dans le cadre d'EuroHPC
 - > Machines pre-exascales, exascales ainsi que multi-petascales
- **L'explosion de l'IA :**
 - Très large succès des méthodes d'apprentissage
 - 'La malédiction de la grande dimension': données et calculs massifs
 - > IA nouvel usager du calcul intensif & des équipements ciblés (Jean Zay)
 - IA dans de multiples usages: IA pour IA mais aussi IA pour la science (modèles numériques)
 - Succès spectaculaire du co-design chez les hyperscalers
 - > IA un moteur du développement du HPC et des architectures
- **Applications :** restent un enjeu majeur, en particulier pour les « *legacy codes* »
 - Vers de nouvelles approches : type « Separation of concerns » ou hybridées avec l'IA
- **Technologies d'usage:**
 - Containerisation, jupyterlab , workflows...
- **Compétences:** restent un enjeu majeur
 - Formations initiale et continue, de la théorie à la technologie et aux usages
- **De nouveaux thèmes :**
 - Données FAIR, reproductibilité, sobriété/frugalité

Les 10 prochaines années ?

- **Modèles:**

- Vers de nouvelles approches (stochastiques, grande dimension) suite au succès de l'IA
 - Vers une variété de représentation de l'information (multi-précision, quantique, compression)
 - Nouveaux besoins: vérification, validation, explicabilité, modèles en dimension réduite

- **Applications/algorithmes :**

- Vers des codes capables d'utiliser une hiérarchie d'architectures ?
 - Vers une meilleure intégration multidisciplinaire (ex NUMPEX) ?
 - Vers Un meilleur soutien en compétences HPC ? La pérennité pour les outils développés ?

- **Quel rôle de l'IA générative ?**

- Dans le domaine de la simulation (génération de codes, pré-post traitements) ?

- **Vers de nouveaux usagers du HPC :**

- Simulation / IA / traitements massifs de données d'expériences ou d'observations
 - Vers des workflows plus complexes et multidisciplinaires & Jumeaux numériques

- **Vers des infrastructures plus intégrées :**

- Données / calcul intensif / quantique / cloud / edge / liaison avec les infrastructures expérimentales

- **Quelles évolutions des architectures ?**

- Rôle des hyperscalers, de l'IA; émergence du quantique ?

- **Enjeu environnemental:**

- Réduction, réflexion sur les usages, post-exascale ...

- **Ressources humaines, un enjeu majeur:**

- Former pour renforcer nos capacités d'innovation et notre potentiel de compétence ! Attractivité !