



Tera

La saga des
supercalculateurs



Tera

La saga des supercalculateurs

À la mémoire de Patrick LASCAUX, Directeur scientifique de la DAM
de 2001 à 2003 et ardent promoteur du calcul scientifique
et des mathématiques appliquées en France.

Sommaire

ÉDITO	P.01
-------------	------

CHAPITRE 1

LE PROGRAMME SIMULATION ET SES ENJEUX

1.1 Comprendre et prévoir par la simulation	P.03
1.2 Le programme Simulation et les supercalculateurs Tera de 2001 à 2019	P.10
1.3 De nouveaux enjeux pour la prochaine décennie	P.13

CHAPITRE 2

LES SUPERCALCULATEURS

2.1 Le supercalculateur dans son centre de calcul	P.16
2.2 Surveillance du centre de calcul Tera	P.20
2.3 Attribuer les ressources du centre de calcul	P.23
2.4 Protéger les centres de calcul des cyberattaques	P.24

CHAPITRE 3

LA SAGA DES CALCULATEURS TERA

3.1 Les supercalculateurs, éléments clés de la dissuasion	P.28
3.2 De Tera1 à Tera1000 - Plus de 15 ans de défis technologiques et de retombées industrielles	P.30
3.2.1 Co-conception avec l'industriel: les grandes thématiques et l'organisation de la collaboration	P.35
3.2.2 Le réseau d'interconnexion - élément clé d'un supercalculateur parallèle	P.36
3.2.3 Au cœur de Tera1000-2: le processeur Intel KNL	P.37
3.2.4 Tera1000-2, un concentré d'innovations	P.40
3.3 Le challenge énergétique	P.41

CHAPITRE 4

LES LOGICIELS

4.1 La « pile logicielle » du calcul scientifique	P.46
4.2 L'évolution des modèles de programmation parallèle: le projet MPC	P.51
4.3 Préparer un calcul et développer les logiciels de simulation	P.56
4.3.1 Développer les logiciels de simulation: la plateforme Arcane	P.56
4.3.2 Préparer une simulation numérique: maillage du modèle	P.59

4.4 Traiter les données et visualiser les résultats	P.63
4.4.1 Modélisation et gestion des données pour la simulation numérique: la plateforme logicielle Hercule	P.63
4.4.2 Traiter et visualiser les résultats d'une simulation numérique	P.66
4.4.3 Analyse et visualisation des données <i>in situ</i> : PaDaWAn, une solution au traitement des volumes de données à l'échelle de l'exascale	P.69

CHAPITRE 5

AVANÇÉES EN SIMULATION ET SUPERCALCULATEURS

5.1 L'apport du HPC à la simulation des matériaux à l'échelle microscopique	P.74
5.2 Simulation de phénomènes électromagnétiques sur des supercalculateurs parallèles	P.79

CHAPITRE 6

LA CONSTRUCTION D'UN ÉCOSYSTÈME DU CALCUL INTENSIF

Le CEA/DAM, moteur d'un pôle de compétences en calcul et traitement de données à haute performance	P.84
--	------

CHAPITRE 7

PRÉPARER L'AVENIR EN COOPÉRATION : EN ROUTE VERS L'EXASCALE

Préparer l'avenir en coopération : en route vers l'exascale	P.94
---	------

CHAPITRE 8

LE CALCUL HAUTE PERFORMANCE : UN TRAVAIL D'ÉQUIPES

Les métiers du HPC au CEA/DAM	P.100
-------------------------------------	-------

GLOSSAIRE	P.105
-----------------	-------

LIENS UTILES	P.114
--------------------	-------

Le programme simulation au cœur de la renaissance d'une filière industrielle de calcul à haute performance



V. SALVETTI
Directeur des applications militaires du CEA

Si la France dispose aujourd'hui d'une filière de calcul à haute performance, c'est bien grâce à la dissuasion. Après l'arrêt définitif des essais nucléaires, le lancement du programme Simulation, en 1996, a profondément transformé la manière de concevoir et garantir la sûreté, la fiabilité et les performances des armes nucléaires qui équipent les deux composantes de la dissuasion française.

Ce programme repose, d'une part, sur des équipes scientifiques de haut niveau et, d'autre part, sur de grands équipements indispensables pour résoudre et valider les équations modélisant le fonctionnement des armes nucléaires : installation de radiographie Epure, laser Mégajoule et supercalculateurs. En particulier, il a conduit la Direction des applications militaires (DAM) à se doter de capacités de calcul exceptionnelles. En effet, entre la fin des années 1990 et aujourd'hui, soit en une vingtaine d'années, la puissance de calcul mise à disposition des concepteurs et physiciens a été multipliée par 500 000 !

Cet accroissement a été obtenu en quatre étapes, Tera1, Tera10, Tera100 et Tera1000. Chacune correspond à l'installation d'un nouveau supercalculateur et est intimement liée aux exigences d'un programme d'armes. De plus, la nécessaire maîtrise de la capacité à disposer, de manière permanente et sans entraves, de ces équipements a engagé les équipes du CEA/DAM dans la co-conception de ces supercalculateurs en relation avec un industriel français. Le programme Simulation est ainsi à l'origine d'une renaissance industrielle qui aujourd'hui mûrit en une filière compétitive de supercalculateurs, vendus à travers le monde. C'est un exemple supplémentaire du rôle structurant de la dissuasion sur le tissu industriel français.

Conjointement avec la montée en puissance des capacités de calcul, un énorme effort de conception et de développement de logiciels nouveaux a été entrepris. Cet effort n'a pu être mené que grâce à un management des compétences et des ressources humaines, démontrant les capacités d'adaptation du CEA/DAM.

Dans le domaine de la simulation et du calcul, comme dans d'autres, l'innovation se nourrit des échanges avec les pairs, académiques ou industriels. Le programme Simulation a aussi été à l'initiative de la réalisation d'un technopôle européen installé à Bruyères-le-Châtel, à côté du centre DAM Île-de-France. Son Très Grand Centre de calcul est au cœur de la communauté du calcul à haute performance française et européenne et fait de la France un leader européen du supercalcul.

Cette publication présente la « saga » des calculateurs Tera et montre qu'il est possible de revenir au meilleur niveau mondial à condition de s'inscrire dans une mission qui remonte au plus haut niveau de l'État avec des objectifs ambitieux.

La simulation, le traitement de volumes gigantesques de données et le calcul à haute performance évoluent très rapidement. Pour répondre aux enjeux de ses programmes, le CEA/DAM disposera dans les années 2020 d'un calculateur de classe exaflopique. La force du CEA/DAM repose sur l'engagement de femmes et d'hommes au service d'une mission, la dissuasion française. C'est un environnement de travail unique pour relever de nouveaux défis scientifiques et techniques, en partenariat avec l'industrie française. ■

CHAPITRE 1

LE PROGRAMME SIMULATION ET SES ENJEUX

La décision de 1996 d'arrêt complet des essais nucléaires a profondément transformé la manière de garantir la sûreté et le fonctionnement des armes nucléaires. Le programme Simulation de la Direction des applications militaires du CEA a été élaboré pour prendre en compte ce nouveau contexte. Dans ce cadre, de plus grandes exigences en matière de logiciels de simulation numérique ont placé sur le devant de la scène un nouveau grand instrument : le supercalculateur.

1.1 Comprendre et prévoir par la simulation

A

près une brève introduction à la simulation numérique, le contexte particulier de la simulation du fonctionnement des armes nucléaires est introduit : la complexité du défi que représente la garantie de fonctionnement et de sûreté sans essais nucléaires

nouveaux est abordée et permet de comprendre l'évolution des machines Tera, support matériel du programme Simulation développé depuis plus de vingt ans au CEA/DAM, ainsi que les perspectives sur les machines futures. L'apport de ces moyens exceptionnels à d'autres missions de la DAM est illustré, que ce soit dans le contexte de la défense conventionnelle, où la question de l'efficacité des armements est primordiale pour leur mise en œuvre, ou encore de la détection des essais nucléaires dans le cadre des traités en vigueur.

INTRODUCTION À LA SIMULATION NUMÉRIQUE

La simulation numérique, associée à la puissance de calcul des superordinateurs et à l'analyse mathématique, est partout : elle permet notamment de garantir la sûreté et la fiabilité des armes nucléaires. La simulation numérique, ou simulation sur ordinateur, a pris une importance considérable dans le développement des sciences de l'ingénieur. Elle assiste ce dernier pour concevoir et optimiser un objet ou un dispositif complexe. La simulation numérique permet des expériences virtuelles qui remplacent ou complètent les expérimentations lorsque celles-ci sont coûteuses ou dangereuses (p. ex. accidents, crash tests), trop longues (p. ex. climatologie) ou trop courtes (p. ex. physique atomique), trop petites (p. ex. protéines) ou trop grandes (p. ex. formation des étoiles et des galaxies), interdites (p. ex. essais nucléaires). Elle permet de réaliser à coût raisonnable des études paramétriques. Grâce aux progrès des calculateurs, la simulation numérique s'est généralisée à toutes

les disciplines au point de devenir le « troisième pilier » de la méthode scientifique, aux côtés de la théorie et de l'expérimentation. Les retombées économiques de l'utilisation intensive de la simulation numérique sont évidentes sur les temps et les coûts de développement. Peu à peu, elle s'est donc imposée dans toutes les branches de l'industrie, en diffusant des grands groupes vers les PME (petites et moyennes entreprises).

La simulation est aussi essentielle en sciences fondamentales et appliquées, pour prédire le comportement d'objets inaccessibles, parce que trop petits, trop grands, ou trop lointains dans l'espace ou le temps. Comment comprendre le déploiement du prion, les instabilités de certaines supernovae, la formation de l'Univers primordial, l'évolution du climat sans recours à la simulation ? La simulation numérique est devenue ainsi omniprésente dans la recherche en astrophysique, climatologie, biologie, pour ne citer que ces quelques exemples.

Les besoins de simulation dans les secteurs de la recherche ou de l'industrie augmentent donc. De surcroît, afin de se rapprocher de plus en plus de la réalité, les simulations numériques doivent être de plus en plus précises, elles sont donc de plus en plus longues et nécessitent de plus en plus de puissance de calcul, d'où l'importance cruciale de moyens de calcul haute performance.

La simulation numérique est une représentation approchée de phénomènes du domaine réel obtenue par une série de calculs sur ordinateur. Elle aboutit à la description des différentes phases ou étapes d'un phénomène physique, comme s'il s'était réellement déroulé. Les ingénieurs et les chercheurs utilisent les données et images (ou vidéos) produites par la simulation pour améliorer leurs connaissances ou pour prévoir les conséquences des phénomènes étudiés (voir chapitre « Avancées en simulation et supercalculateurs »).

La simulation se construit par étapes avec différentes équipes aux métiers complémentaires :

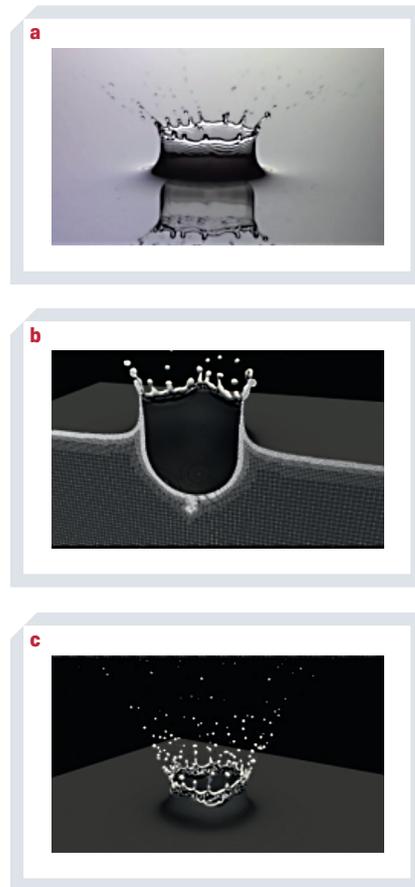


FIGURE 1
 (a) Photo de la réalité physique; (b) à un instant donné, aperçu du maillage en coupe (discrétisation de l'espace); (c) Image obtenue par simulation.

représentation du phénomène (**figure 1**) selon les lois de la physique (physicien) – transcription du phénomène physique en équations mathématiques (mathématicien) – résolution numérique et traduction en langage informatique (numéricien et informaticien) – calcul numérique et comparaison avec l'expérience ou l'observation (ingénieur d'étude et expérimentateur). Ces étapes sont illustrées sur un exemple emblématique, celui de la simulation des écoulements fluides.

LES ÉTAPES DE LA SIMULATION NUMÉRIQUE

1. Réalité

Pour représenter un phénomène, les physiciens et les ingénieurs prennent en compte de nombreuses lois physiques, par exemple celles qui régissent les relations entre les quantités physiques décrivant un écoulement dans un fluide : densité ρ , vitesse u , pression p , tension de surface σ .

2. Modélisation

Ces lois physiques sont traduites en équations mathématiques, faisant intervenir l'ensemble des grandeurs nécessaires à la modélisation : on

obtient alors le modèle mathématique d'un phénomène physique. Il fait intervenir les grandeurs physiques que l'on cherche à déterminer et qui sont les **inconnues** du problème. Pour notre exemple, des équations régissant le mouvement d'un fluide, elles s'écrivent, l'inconnue étant la vitesse \vec{u} :

$$\begin{aligned} \frac{\partial(\rho\vec{u})}{\partial t} + \nabla - (\rho\vec{u} \otimes \vec{u}) &= \nabla - \vec{\gamma} \quad (\text{force visqueuse}) \\ &= \nabla p \quad (\text{force de pression}) \\ &= \sigma\kappa\delta_s\vec{n} \quad (\text{force de tension de surface}) \end{aligned}$$

3. Résolution numérique et programmation

Mais ces équations mathématiques sont souvent trop complexes pour être résolues « à la main » et doivent être traitées à l'aide d'un ordinateur. Les mathématiciens les discrétisent dans l'espace et le temps :

- le domaine de calcul est découpé en petits éléments, les **mailles** ; on dit qu'il est maillé ; l'ensemble des mailles représente le **maillage** ;
- le temps est découpé en petits intervalles appelés pas de temps.

$$\frac{(\rho\vec{u})_i^{n+1} - (\rho\vec{u})_i^n}{t^{n+1} - t^n} + \dots$$

Ces équations sont ensuite résolues pour chacun des points dans l'espace et des instants dans le temps ainsi découpés. L'enchaînement des calculs à réaliser s'appelle un algorithme, qui est défini par les formules mathématiques spécifiques d'un **schéma numérique**.

L'algorithme est décrit dans un langage informatique, éventuellement à l'aide de logiciels spécifiques appelés plateforme de développement.

4. Exécution de simulations

Avant de soumettre une simulation sur un supercalculateur, l'ingénieur d'étude ou le physicien doit préparer les données spécifiques à ces calculs (voir article « Préparer une simulation numérique : maillage du modèle »). Une fois la simulation terminée, celle-ci pouvant prendre plusieurs jours pour les simulations les plus complexes, il peut visualiser les résultats sur son poste de travail de façon interactive (voir article « Traiter et visualiser les résultats d'une simulation numérique »).

5. Expérimentations

Une phase d'expérimentation est nécessaire afin de valider les modèles, de mesurer les paramètres d'entrée et leur influence. Si les expérimentations diffèrent de la réalité, de nouvelles simulations sont lancées, en faisant varier les paramètres, en

intégrant ou affinant d'autres phénomènes dans la modélisation. L'enjeu est de réduire le plus possible l'écart entre la réalité physique et le modèle numérique qui s'en approche.

Afin d'approcher au mieux la réalité, les simulations numériques se doivent, souvent simultanément :

- ⊙ d'intégrer un nombre croissant de modèles physiques ;
- ⊙ d'utiliser des résolutions (p. ex. temporelles, spatiales, angulaires, fréquentielles) de plus en plus fines, en vue de réduire et quantifier les erreurs de **discrétisation** propres aux méthodes utilisées ;
- ⊙ de multiplier les calculs pour quadriller l'espace des solutions et réduire les incertitudes.

Face à de tels besoins, les puissances de calcul et capacité de stockage des centres de calcul dédiés à la simulation numérique sont en constante augmentation. Celle-ci est favorisée par les incessantes évolutions technologiques.

LA SIMULATION APPLIQUÉE AUX TÊTES NUCLÉAIRES

Le fonctionnement d'une tête nucléaire repose sur des processus physiques complexes, couplant une large gamme d'échelles d'espace et de temps et génère des conditions extrêmes de pression et de température. Avec l'arrêt définitif des essais nucléaires français, il ne peut plus être validé en vraie grandeur. Le CEA/DAM s'appuie donc, pour comprendre, prévoir et garantir ce fonctionnement, sur le triptyque modélisation physique-validation expérimentale par partie-simulation numérique.

Le fonctionnement d'une arme comporte deux phases : la première sans réactions nucléaires, la seconde avec réactions nucléaires, que nous allons brièvement décrire. Nous verrons dans la suite que la résolution des équations associées demande une puissance de calcul considérable. La DAM s'est donc dotée d'importants moyens de calcul, en impulsant et en accompagnant leur évolution en coopération étroite avec l'industrie informatique nationale. Enfin, nous expliquerons comment construire des méthodes numériques approchant de manière fidèle les équations de l'hydrodynamique.

Une première phase décrite par les équations de l'hydrodynamique

La première phase de fonctionnement, sans réactions nucléaires, est décrite par les équations de la mécanique des milieux continus, ou équations de l'hydrodynamique, présentées plus haut. Ces équations

relient les dérivées partielles, en temps et en espace, de la densité, de la vitesse et de la densité d'énergie, et permettent de calculer l'évolution au cours du temps de ces quantités en chaque point du milieu simulé. Pour être exploitables par un ordinateur, ces équations continues doivent être ramenées à un ensemble fini d'équations discrètes. Pour cela, comme expliqué plus haut, l'espace est divisé en petits volumes, appelés mailles, le temps est fractionné en pas de temps. Les quantités à un pas de temps sont calculées explicitement en fonction de celles au pas de temps précédent. Pour ces températures assez basses, le comportement élastoplastique des matériaux, c'est-à-dire comment ils se déforment sous contraintes, doit être modélisé et c'est un domaine de recherche en soi ! La conception de schémas numériques simulant les équations de l'hydrodynamique est un problème difficile de mathématiques appliquées.

Une puissance de calcul phénoménale est nécessaire

Le nombre d'opérations à réaliser à chaque pas de temps dépend d'une part du nombre de mailles, déterminé par la finesse du maillage h rapporté à la taille L du domaine simulé et de la dimension d'espace du problème résolu, et d'autre part de la précision des modèles physiques. Une simulation monodimensionnelle, à une seule dimension d'espace, demandera L/h mailles. Pour passer en bidimensionnel (deux dimensions), il en faudra $(L/h)^2$, et en tridimensionnel, ou 3D, $(L/h)^3$: la puissance de calcul nécessaire est multipliée par L/h quand on change de dimension. Si par exemple $L/h = 1000$, un changement de dimension de la simulation conduira à un besoin en puissance de calcul 1000 fois supérieur (soit une quinzaine d'années de développement des calculateurs, suivant la **loi de Moore**, qui prédit un doublement de la puissance de calcul en une année et demie). Il a été ainsi possible de réaliser très tôt des simulations d'hydrodynamique monodimensionnelles. Aujourd'hui, l'essentiel des simulations est effectué en 2 ou 3 dimensions. Cela requiert non seulement une énorme puissance de calcul et de stockage de données, mais aussi des capacités de visualisation spécifiques. Passons maintenant à la phase avec réactions nucléaires, qui demande de résoudre une autre catégorie d'équations.

Phase nucléaire et équations de transport

Les réactions nucléaires produisent une population de neutrons et de particules chargées (électrons, protons, particules alpha), qui, en ralentissant par interaction avec les matériaux, les portent à des

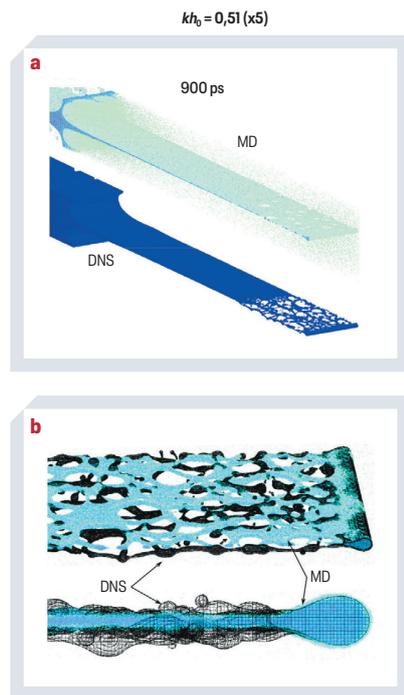


FIGURE 2

Éjection de matière suite au débouché d'une onde de choc plane à l'échelle nanométrique sur une interface perturbée (étain) - Deux codes de simulation sont comparés, le code ExaStamp de *dynamique moléculaire* (résultat dynamique moléculaire (DM)) et le code HERA *hydrodynamique* (résultat simulation numérique directe (ou DNS en anglais)) après recalage de l'équation d'état et des coefficients de viscosité et de tension de surface au potentiel interatomique. (a) vue d'ensemble de la nappe éjectée, (b) zoom sur la tête de la nappe et accord sur les positions et les rayons de courbure.

températures très élevées. Les matériaux passent en conséquence à l'état de plasma, de plus en plus ionisés au fur et à mesure que la température augmente. Des phénomènes nouveaux interviennent alors. La matière chaude rayonne : elle émet (et absorbe) des photons. Il faut décrire l'évolution au cours du temps de cette population de photons, pilotée par les équations dites de transport. Les populations de neutrons sont également décrites par des équations de transport. Les équations de transport sont posées dans un espace avec plus de dimensions que l'espace usuel : il faut en effet connaître, en chaque point de l'espace, dans quelles directions se propagent les photons et les neutrons (ainsi que les particules chargées), et comment ils se répartissent en énergie. En somme, deux dimensions de direction (par exemple les deux angles d'Euler) et une dimension d'énergie s'ajoutent aux dimensions de l'espace physique. Cela se traduit en pratique par une augmentation considérable du nombre de quantités à calculer dans chaque maille. Par exemple, si la sphère des directions est divisée en une centaine de secteurs, et la répartition en énergie décrite à l'aide d'une centaine de groupes, c'est-à-dire de mailles dans la dimension énergie, nous devons calculer

une dizaine de milliers de quantités dans chaque maille. La résolution des équations de transport est donc coûteuse et s'appuie sur des moyens de calcul haute performance.

Données physiques pour la simulation

Il faut aussi mentionner que la simulation doit être alimentée par des données physiques : lois de comportement déjà évoquées, équations d'état donnant la pression en fonction de la température et de la densité, opacités mesurant l'interaction du rayonnement avec la matière, sections efficaces nucléaires, mais aussi conductivités et viscosités des milieux. Le calcul de ces quantités s'appuie sur une importante activité de recherche en physique du solide, des plasmas, physique atomique et physique nucléaire, qui permet d'établir des modélisations précises et cohérentes. Mais les modèles élaborés ne peuvent être simulés qu'avec de puissants moyens de calcul. Si, par exemple, on remplace une loi de comportement semi-empirique par un modèle multi-échelle, le besoin en puissance de calcul augmente considérablement. Parmi les travaux les plus exigeants en ressources de calcul, on peut citer par exemple l'équation d'état du plutonium avec traitement des électrons corrélés, ou encore l'évaluation des sections efficaces des actinides avec covariances et contraintes intégrales et la simulation de la fission dynamique [1].

Encadré

L'éjection de matière consécutive au débouché d'une onde de choc sur une interface perturbée correspond à une physique complexe où l'apport de la simulation numérique est multiple, selon le type de modélisation retenue (approche multi-échelle).

En configuration *nanométrique*, l'approche *dynamique moléculaire* consiste à traiter un grand nombre d'atomes et leurs interactions, « toute » la physique étant alors contenue dans le potentiel d'interaction utilisé (plus ou moins précis, selon le coût de calcul visé). L'approche *mécanique des milieux continus* repose quant à elle sur les équations de Navier-Stokes avec effets capillaires, moins coûteuses et utilisables à l'échelle *micro-métrique*. Après ajustement des paramètres des relations de fermeture du calcul Navier-Stokes au code de dynamique moléculaire (équation d'état, coefficients de viscosité et de tension de surface), un excellent accord est observé entre les deux approches jusqu'à la phase de fragmentation de la

nappe éjectée (figure 2), ce qui permet de progresser dans la compréhension de tels phénomènes (voir article « L'apport du HPC à la simulation des matériaux à l'échelle microscopique »).

Une machine multi-pétaflopique comme Tera1000 s'avère actuellement indispensable au code de dynamique moléculaire qui nécessite, selon le potentiel d'interaction moléculaire utilisé, de l'ordre d'un mois calendaire sur 28 000 cœurs de calcul pour 1 milliard d'atomes.

LA SIMULATION NUMÉRIQUE APPLIQUÉE À LA DÉFENSE CONVENTIONNELLE

Dans ce cadre, et à des fins opérationnelles, il est primordial d'être capable de prévoir avec une grande précision les effets des explosions de munitions très diverses sur des structures elles aussi de nature et de forme variées et parfois mal connues. Il est alors difficile de balayer la variété ou diversité de situations possibles au travers d'essais coûteux et qui *in fine* ne représenteraient qu'une petite partie des conditions pouvant être rencontrées. C'est

pourquoi le recours à la simulation numérique est fortement croissant et requiert une connaissance fine des propriétés des matériaux, des mécanismes d'explosion et des phénomènes d'interaction entre l'explosion et les structures qui motivent de nombreuses études de physique à différentes échelles. Si l'explosion d'une charge conventionnelle à proximité d'un bâtiment correspond à une physique *a priori* bien connue (ondes de souffle, équations d'Euler de la dynamique des gaz), elle est très exigeante en ressources de calcul en raison d'une configuration intrinsèquement tridimensionnelle. Quatre simulations ont ainsi été menées avec des résolutions grandissantes de 8 cm, 4 cm, 2 cm et 1 cm sur l'ensemble du bâtiment, une « sonde virtuelle » étant placée en extrémité du couloir du rez-de-chaussée (figure 3). Le niveau du premier choc est bien capturé dès 2 cm de résolution spatiale, mais le deuxième choc est encore éloigné de la convergence numérique en résolution 1 cm, simulation pourtant 16 fois plus chère (facteur 8 sur le nombre de mailles, facteur 2 sur le nombre de pas de temps).

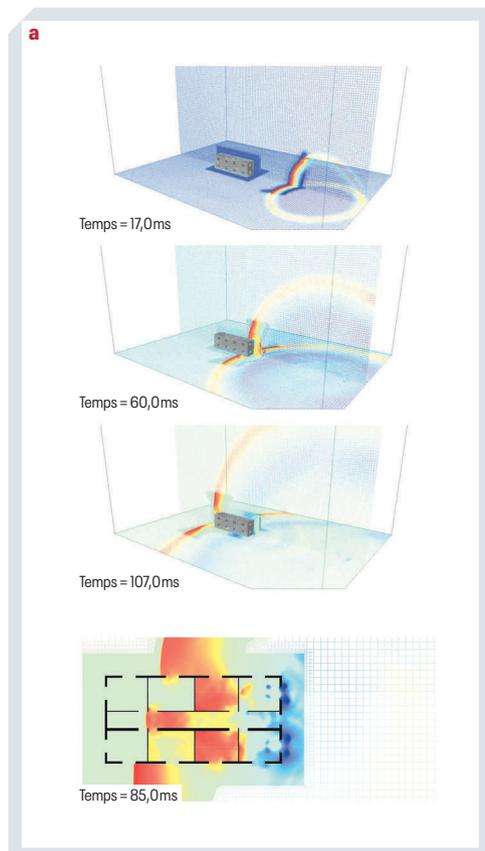
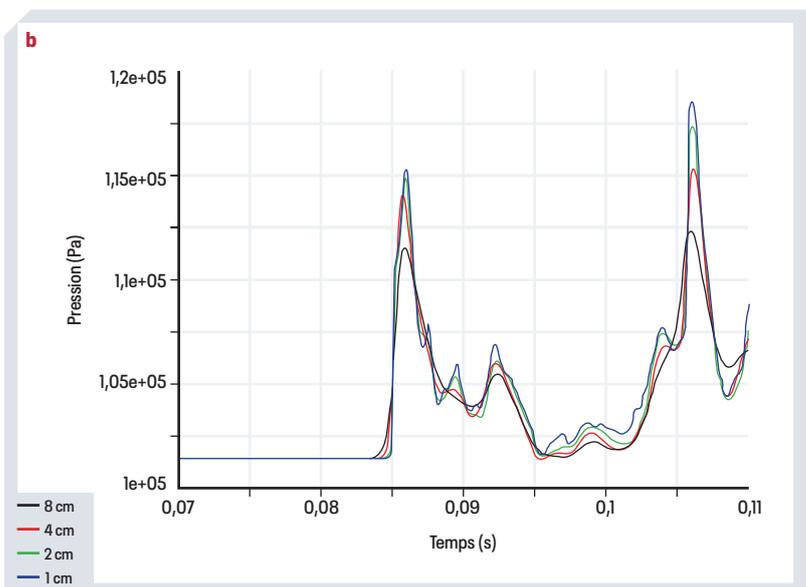


FIGURE 3

Explosion d'une charge conventionnelle à proximité d'un bâtiment - La simulation des ondes de souffle exige en 3D des ressources de calcul importantes. Le logiciel de simulation utilisé ici met en œuvre du raffinement adaptatif de maillage qui adapte la taille de la maille de calcul en fonction de l'évolution du phénomène et du besoin de précision des calculs. (a) maillage dans 2 plans de coupe orthogonaux à différents instants et champ de pression en rez-de-chaussée, (b) convergence de la simulation numérique selon la résolution spatiale, relevés de pression sur une « sonde » située en couloir rez-de-chaussée pour les simulations 8 cm, 4 cm, 2 cm et 1 cm (7,23 milliards de mailles).



Pratiquement inaccessible jusqu'alors, la simulation à l'échelle centimétrique a été rendue possible avec l'installation de Tera1000. Ainsi, à cette résolution et pour un total de 7,23 milliards de mailles, la simulation a été achevée en 7 jours sur 1/30 de la machine.

LA SIMULATION NUMÉRIQUE APPLIQUÉE À LA DÉTECTION ET À LA SURVEILLANCE DE L'ENVIRONNEMENT

Les missions opérationnelles de détection et de surveillance de l'environnement reposent sur un grand volume de données et sur des modèles physiques complexes. Afin de caractériser rapidement des événements d'intérêt, pour réaliser des études environnementales de plus en plus précises, les moyens numériques jouent un rôle croissant sur la quantification des incertitudes et la montée en résolution des modèles.

Par exemple, la simulation, grâce aux calculateurs hautes performances, permet de compenser la méconnaissance relative des milieux d'intérêt, généralement inaccessibles directement (croûte terrestre, haute atmosphère). L'inversion des données, en s'appuyant sur les nombreux modèles numériques disponibles, permet d'estimer statistiquement les incertitudes ou probabilités sur les grandeurs physiques recherchées (localisation, magnitude, signal acoustique...), et, en retour, informe aussi sur la variabilité des milieux traversés (hétérogénéités crustales, vitesses des vents). La prise en compte des hétérogénéités (figure 4) avec une puissance informatique accrue, autorise ainsi la résolution de problèmes inverses complexes, notamment lorsqu'il s'agit de déterminer les propriétés des sources sismiques. Ces approches ouvrent aussi la voie au traitement de phénomènes physiques non linéaires et multi-échelles, comme pour rendre compte des phénomènes de conversion aux interfaces, d'interaction d'ondes acoustiques avec des topographies complexes (figure 5). Les progrès récents permettront de fournir des estimations de plus en plus précises pour les études, voire de contribuer à des prévisions en temps réel, par exemple hauteurs des tsunamis attendues dans les ports dans le cadre de la mission du Centre national d'alerte aux tsunamis (Cenalt).

Le calcul haute performance, support de la simulation

La complexité des équations en présence et la précision demandée pour leur résolution nécessitent donc non seulement des méthodes numériques et informatiques spécifiques, mais

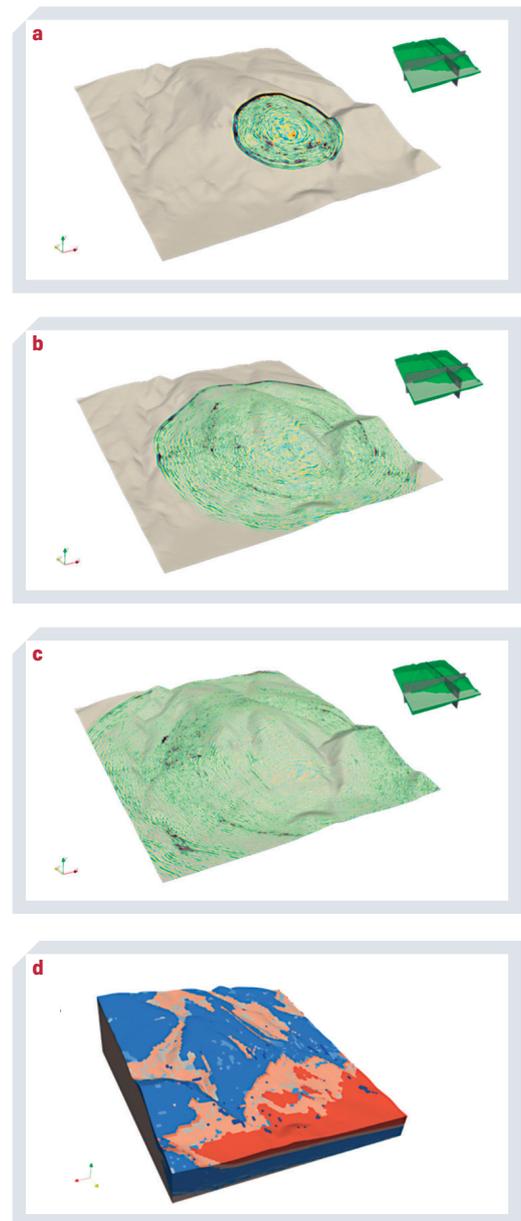


FIGURE 4

Simulation de la propagation des ondes sismiques avec le code d'élastodynamique linéaire SEM-3D (éléments finis spectraux) pour trois instants 0,2 s (a), 0,4 s (b) et 0,6 s (c) consécutifs à une explosion. Les ondes sont émises par une source isotrope fictive située dans le massif du Rustrel (Vaucluse). Les matériaux constitutifs sont regroupés au sein de quinze classes, chacune étant représentée par une couleur distincte (d). Le domaine de calcul couvre un volume de $52 \times 2 \text{ km}^3$, ce qui représente $64 \cdot 10^9$ degrés de liberté. Simulation réalisée sur le calculateur Tera1000-2 (Yoccoz) lors d'un grand challenge CEA/DAM ayant mobilisé 16384 cœurs.

également les meilleurs et les plus puissants ordinateurs du moment. La simulation numérique est indissociable du calcul haute performance : chaque progrès incrémental dans le domaine du calcul permet à la simulation d'être

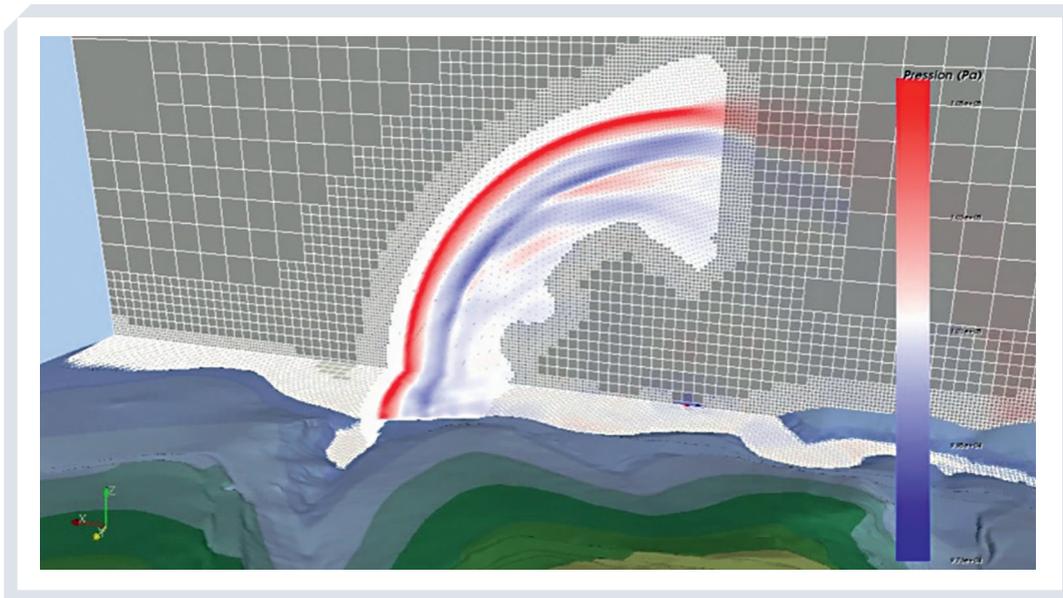


FIGURE 5

Exemple de la simulation 3D sur maillage adaptatif de l'interaction d'une onde de souffle avec une topographie variable. Les plans de coupe mettent en évidence le champ de pression et le maillage dynamique. Ce calcul reproduit le comportement de l'onde générée par le soudage de plaque par procédé explosif dans un site industriel du sud de la France. L'utilisation de Tera100 a permis de passer les premiers cas en 3D (1 milliard de mailles sur mille cœurs), et les études sont devenues plus systématiques grâce à Tera1000. Des simulations de l'ordre de 8 milliards de mailles **AMR** (raffinement adaptatif de maillage) sur 8000 cœurs sont aujourd'hui accessibles.

plus précise. Toute avancée décisive des moyens de calcul permet à la simulation, comme nous l'avons vérifié à chaque changement de calculateur, d'aborder de nouveaux problèmes ou d'utiliser des modèles plus prédictifs pour un problème donné.

CONCLUSION

La simulation est un domaine multidisciplinaire, à la croisée des compétences en modélisation physique, analyse numérique et conception de logiciels adaptés aux machines massivement parallèles. Cette pluridisciplinarité, les évolutions rapides du domaine sont les gages d'un travail stimulant, attractif pour les ingénieurs et les chercheurs impliqués. La progression volontariste des capacités de calcul et l'utilisation de schémas numériques plus sophistiqués ont contribué à améliorer significativement la simulation numérique, en permettant la mise en œuvre de modèles de plus en plus phénoménologiques capables de restituer une physique et des géométries toujours plus proches des situations réelles. C'est pourquoi la simulation numérique, qui s'appuie ou qui repose sur le calcul haute performance, est une composante essentielle de la garantie de fonctionnement sans essai et de la sûreté des armes nucléaires. ■

.....

→ Référence

1 / Ouvrage *60 ans de sciences à la Direction des applications militaires : avancées récentes et défis de demain*, <http://www-dam.cea.fr/> rubrique « Un peu de science ».

.....

1.2 Le programme Simulation et les supercalculateurs Tera de 2001 à 2019

— La formidable montée en puissance des calculateurs impulsée par le programme Simulation a permis de garantir sans essais nouveaux la sûreté et la fiabilité des armes nucléaires.

La Direction des applications militaires (DAM) du CEA a pour mission de concevoir, fabriquer, maintenir en condition opérationnelle, puis démanteler les têtes nucléaires qui équipent les forces nucléaires aéroportée et océanique françaises. Les têtes nucléaires, appelées à remplacer les armes en service arrivant en fin de vie, sont désormais garanties sans essais nucléaires nouveaux. Le programme Simulation, lancé en 1996, répond à cet objectif de garantie de leur fiabilité et de leur sûreté. Ce programme est décliné en un ensemble de projets fondés sur une démarche scientifique, technologique et industrielle appliquée à la conception et la réalisation des armes. Il se compose de trois volets complémentaires :

- ⊙ la physique des armes, c'est-à-dire la mise au point des modèles physiques prédictifs ;
- ⊙ la simulation numérique, qui nécessite la réalisation d'un ensemble de logiciels de simulation numérique (appelés « codes de calcul » ou « applications ») et d'exploitation des données. Ceux-ci prennent en compte les modèles physiques et doivent être adaptés aux supercalculateurs ;
- ⊙ la validation expérimentale, par parties ou plus globalement, des logiciels de simulation. Celle-ci repose tout particulièrement sur de grands équipements indispensables pour valider les équations modélisant le fonctionnement des armes nucléaires : l'installation de radiographie éclair (**Epure**) et le laser Mégajoule (**LMJ**).

Deux spécificités de ce programme, importantes par leurs retombées, doivent être mentionnées.

La première concerne son ambition et son ampleur. En effet, la simulation numérique dépend de la disponibilité et de la puissance des calculateurs, et simuler les différentes étapes du fonctionnement d'une arme avec une précision importante requiert des capacités de calcul considérables. Dans la première phase du programme, celles-ci ont été évaluées comme devant être multipliées par 10 000 en dix ans, soit bien plus que ne le permettraient les seuls progrès de la microélectronique.

Il a donc fallu anticiper un changement radical de l'architecture des supercalculateurs, mais également engager un important effort dans le domaine du logiciel. Plus généralement, c'est l'ensemble du système « calculateur – centre de calcul – logiciels – applications » qui a dû évoluer dynamiquement et de concert.

UNE DÉMARCHE « QUALITÉ » : LES STANDARDS DE CALCUL

En effet, les programmes de conception et de garantie des armes s'appuient sur un outil de simulation qui doit avoir un haut niveau de précision. Cet outil, appelé « standard de garantie », capitalise l'ensemble des connaissances du programme Simulation à un instant donné. Il est constitué d'une chaîne de logiciels (ou « codes ») reproduisant par le calcul les différentes phases

de fonctionnement d'une arme nucléaire. Sa mise en œuvre nécessite l'enchaînement de modèles physiques (équations) développés dans le cadre d'études scientifiques spécifiques ainsi que l'utilisation de puissants moyens de calcul pour résoudre ces équations. À chaque étape du programme, les supercalculateurs sont dimensionnés pour répondre aux besoins de conception et de garantie des armes nucléaires, et *in fine* à la précision requise pour la simulation de leur fonctionnement. La garantie de l'énergie et de la sûreté d'une arme nucléaire nécessite d'avoir auparavant validé expérimentalement ce standard de garantie. Cette validation est obtenue en confrontant les résultats des calculs aux mesures recueillies lors des essais nucléaires passés et aux expériences de validation réalisées sur les installations Epure et LMJ.

À la fin des années 1990, le CEA/DAM s'est donc engagé dans l'ère de la simulation avec la conception et le développement d'une nouvelle génération de codes de calcul particulièrement complexes. Mais pour élaborer un nouveau standard de garantie, qui nécessite des milliers de simulations, il faut compter de l'ordre de deux ans sur l'une des machines les plus puissantes au monde. Une fois cet outil de simulation validé, il faut compter à nouveau des milliers de simulations pour concevoir et garantir une arme nucléaire.

Le CEA/DAM a donc mis en œuvre un programme d'équipement en supercalculateurs depuis plus de vingt ans, conçu pour satisfaire les besoins croissants des standards de simulation.

LES SUPERCALCULATEURS, COMPAGNONS INDISPENSABLES DES STANDARDS DE CALCUL

La construction d'un nouveau standard de garantie avec des descriptions de la physique plus complètes nécessite une représentation spatiale plus précise de l'objet simulé, c'est-à-dire un **mailage** plus fin. Ces simulations, notamment parce qu'elles contiennent plus de **mailles** à calculer, sont plus consommatrices de temps de calcul. Pour conserver la même enveloppe temporelle de réalisation des programmes, il faut alors utiliser des « supercalculateurs » plus puissants que ceux utilisés pour mettre au point le standard de calcul précédent.

C'est l'objectif du projet Tera, qui a conduit à l'installation des supercalculateurs Tera1 en 2001, puis Tera10 en 2006 et Tera100 en 2010. La deuxième phase du programme a été lancée en 2011 et comprend le calculateur Tera1000 mis en service de 2016 à 2017 et le futur calculateur EXA1, dont la première partition sera opérationnelle en 2021.

Toutefois, pendant qu'il mène à bien une étape de son programme Simulation avec le supercalculateur dont les spécifications ont été décidées au moment où celle-ci a été lancée, le CEA/DAM doit anticiper la conception de la nouvelle génération de supercalculateurs qui lui permettra de réaliser l'étape suivante.

En conséquence, le CEA/DAM a établi, depuis le début des années 2000, une série d'accords de partenariat avec la société Bull, devenue depuis Atos/Bull, dans le domaine de la R&D en vue de la conception et la réalisation de supercalculateurs. Ce partenariat est fondé sur une méthodologie de co-conception qui prend en compte les besoins de performances du CEA/DAM, les contraintes technologiques, mais aussi les aspects économiques en vue de la réalisation de lignes de produits commercialement compétitives.

Outre qu'elle donne au CEA/DAM l'assurance de disposer, en temps voulu, des supercalculateurs dont il aura besoin pour la réalisation de son programme, elle lui permet d'optimiser le couple « codes de calcul »-supercalculateur, tout en maîtrisant son coût total d'exploitation et de maintenance. Cette optimisation est réalisée en relation avec l'industriel par une démarche itérative afin de sécuriser les objectifs de performances et de coût du nouveau supercalculateur.

Pour développer les couples codes-calculateur les plus performants, les compétences des équipes du CEA/DAM sont mises à contribution à plusieurs niveaux : la structure du code, l'interaction code-calculateur et l'architecture même de la machine. Cette démarche de partage de la conception se décline selon les cas sous la forme de développements menés conjointement avec des industriels, ou bien dans des projets collaboratifs impliquant académiques et industriels, comme le laboratoire Exascale Computing Research (**ECR**) avec la société Intel et l'université de Versailles-Saint-Quentin-en-Yvelines ou le Centre d'expertise en simulation des matériaux (**Cesimat**) pour la simulation des matériaux qui regroupe des experts du CEA/DAM, Michelin, L'Oréal, Safran et Faurecia.

Cette démarche s'effectue aussi en contribuant activement à la communauté **open source** sur les technologies logicielles pertinentes. Par ailleurs, les durées de vie moyennes des codes de calcul étant bien supérieures aux durées de vie des calculateurs, les grands codes de calcul sont développés en séparant au maximum les aspects liés aux technologies matérielles et les algorithmes liés aux modèles physiques. Ainsi, la plateforme ARCANE, initiée au début des années 2000 par les seules équipes du CEA/DAM, est depuis près de 15 ans co-développée avec l'**IFPEN** dans le

cadre d'un partenariat visant à enrichir les fonctionnalités informatiques et numériques. Il vise à réduire le coût de développement des logiciels de simulation utilisant les supercalculateurs dans les domaines respectifs des deux organismes.

Cette stratégie permet aux équipes du CEA/DAM d'orienter et d'anticiper le travail important de développement des nouvelles versions des codes de calcul qui seront déployés et utilisés sur la future machine dans l'objectif de tirer pleinement parti de la puissance de celui-ci.

Les supercalculateurs livrés au CEA/DAM par Atos/Bull (Tera10 et Tera100) ont pleinement atteint les objectifs de performances fixés par le programme Simulation et ont été classés, à la date de leur mise en service, premiers en Europe et respectivement 5^e et 6^e au classement mondial du **Top500**. Ces supercalculateurs ont de plus donné naissance à une ligne de produits commercialement compétitive, pérenne et souveraine. Plus généralement, le programme Simulation du CEA/DAM est un facteur essentiel du développement du **HPC** en France, et ce dans chacun des trois piliers qui le soutiennent :

- ① 1 – Le développement des technologies matérielles et logicielles, qui incluent la R&D et la conception des calculateurs. Ce pilier est assuré d'une part par le partenariat entre le CEA/DAM et Atos-Bull, porté depuis 2013 par le plan supercalculateur français, et d'autre part par le technopôle européen de compétence en simulation numérique haute performance **Teratec**, chargé de promouvoir l'ensemble de la chaîne de valeur du HPC, afin de positionner la France comme l'un des principaux acteurs mondiaux dans ce domaine.
- ② 2 – La mise en œuvre et l'exploitation de grandes infrastructures de calcul et de stockage de données, avec le complexe de calcul du CEA dont le CEA/DAM porte la responsabilité. Celui-ci comprend deux volets, l'un tourné vers l'industrie, l'autre vers la recherche académique. Le volet industrie s'appuie sur le **CCRT**, dont les moyens sont utilisés par un grand nombre d'industriels français partenaires. Le volet recherche académique s'appuie sur des moyens de calcul financés par la société civile **Genci**. Son utilisation est réservée, sur appel à projets, aux organismes de recherche européens, notamment dans le cadre du programme **Prace**.
- ③ 3 – Le développement des applications et des usages, qui englobent les développements applicatifs et les formations nécessaires à l'utilisation efficace des supercalculateurs, ainsi que les projets de simulation eux-mêmes. Un grand nombre d'entre eux sont impulsés et fédérés à Teratec et au CCRT.

La vision globale puis la mise en œuvre du programme Simulation ont structuré, nourri et fait grandir l'écosystème du calcul haute performance en France. Ce programme a positionné et continue de positionner le CEA/DAM comme un acteur majeur de la recherche et du développement sur toute la chaîne de valeur du calcul haute performance. Il est de plus à l'origine de la renaissance et le socle d'une filière industrielle compétitive de supercalculateurs.

Dans la même période, la création de Teratec et les différents instruments et dispositifs de soutien à l'innovation mis en place par les pouvoirs publics français – pôles de compétitivité, Programme d'investissements d'avenir... – ont permis à la France de devenir un leader et un territoire attractif pour le calcul intensif et le développement de ses technologies.

Le programme Simulation se poursuit en une nouvelle phase d'extension et d'exploitation. Des outils de simulation plus précis nécessiteront la mise en œuvre, dans la prochaine décennie, de calculateurs 1000 fois plus puissants que ceux de la génération TERA100 qui a permis d'atteindre les objectifs de la première phase du programme Simulation en contribuant à la garantie des performances de la tête nucléaire océanique (**TNO**). Ces supercalculateurs prendront en compte la contrainte de plus en plus forte liée à la consommation d'énergie. Ainsi, les futures architectures seront hétérogènes, optimisées énergétiquement, et modulaires afin de répondre au mieux aux différents besoins tout en gardant une souplesse d'évolutivité et de réactivité dans le temps.

Depuis 2018, l'objectif du CEA/DAM est rejoint par les objectifs de la politique numérique européenne en matière de supercalculateurs. Le CEA/DAM et ses partenaires sont particulièrement fiers d'être au cœur de ce dispositif. ■

1.3 De nouveaux enjeux pour la prochaine décennie

— En route vers la simulation à l'échelle du milliard de milliards d'opérations par seconde : de Tera1000 à Exa1...

La démarche de simulation appliquée initialement à la garantie des performances de la charge nucléaire seule s'étend progressivement à l'ensemble des fonctions de la tête nucléaire, impliquant de nouveaux enjeux scientifiques pour le calcul numérique (aérodynamique, électromagnétisme...). Par ailleurs, l'explosion des volumes de données générés par les grands instruments du programme Simulation, supercalculateurs et installations expérimentales **Epure** et **LMJ**, que ce soit en matière de résultats ou de données d'exploitation, va nécessiter la montée en puissance des méthodes d'intelligence artificielle ou du **HPDA** (High Performance Data Analytics), en s'appuyant en particulier sur la chaire « Industrial Data Analytics & Machine Learning » créée avec Atos au sein de l'ENS Paris-Saclay.

Ainsi, pour répondre aux enjeux de ses programmes et satisfaire les besoins exprimés par la défense, le CEA/DAM devra disposer à l'horizon de la décennie en cours (**figure 1**) d'un ordinateur de classe **exaflopique**, dénommé Exa1. Dès 2013, l'analyse faite par ses experts et ceux d'Atos/Bull a permis d'esquisser les architectures des calculateurs à l'horizon de la prochaine décennie. Celles-ci se révèlent différentes de celle de Tera100, d'une part compte tenu de l'évolution de la feuille de route de la microélectronique (ralentissement de la **loi de Moore**) et d'autre part du fait de l'adaptation aux besoins des algorithmes d'intelligence artificielle et au traitement de volumes de données croissants. Pour réussir ce saut technologique, il faut relever plusieurs défis majeurs.

Pour le supercalculateur, les priorités sont de maîtriser la performance énergétique, la fia-

bilité globale et l'interconnexion d'un système modulaire et hétérogène composé d'un très grand nombre de processeurs. Les méthodes et outils de développement des logiciels de simulation devront faciliter l'adaptation aux différents niveaux de « **parallélisation** ». Ces optimisations nécessitent de maîtriser le placement des données et l'utilisation des différents niveaux de mémoire, ce qui conduira sans doute à concevoir et développer une nouvelle génération de codes de calcul.

Enfin, les simulations de plus en plus détaillées, en particulier tridimensionnelles, produisent de gigantesques volumes de données dont l'exploitation et la gestion bouleversent les approches traditionnelles. Les pratiques usuelles de stockage sur disque de l'ensemble des données produites pour les analyser ne seront plus possibles car beaucoup trop onéreuses. L'analyse devra se faire davantage *in situ* en mémoire vive du calculateur pour n'en extraire que les données pertinentes. Cette problématique conduit naturellement le CEA/DAM à introduire des technologies de type « big data » et d'intelligence artificielle pour l'analyse des données et l'accélération des processus de simulation.

Le projet Tera a positionné une étape intermédiaire en 2017 avec l'acquisition du calculateur Tera1000, démonstrateur pré-exaflopique des différentes partitions du calculateur Exa1.

Tera1000 comporte deux partitions :

- Tera1000-1, calculateur doté d'une architecture analogue à celle de Tera100, offrant une puissance de calcul double avec une empreinte énergétique réduite d'un facteur 5. Il assure la continuité des travaux de conception et de garantie des armes nucléaires ;

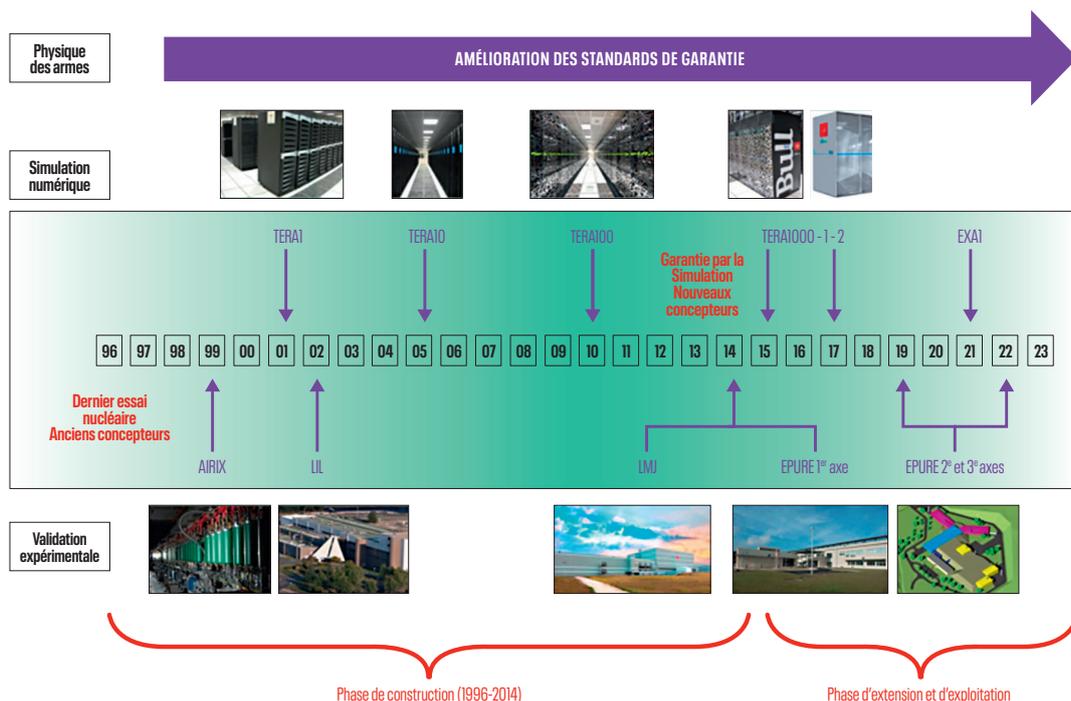


FIGURE 1

Frise chronologique du programme Simulation. AIRIX : prototype de l'installation EPURE. LIL : prototype de l'installation LMJ. 2021 marque la mise en service de la première partition du supercalculateur Exa1.

© Tera1000-2, le système prototype du calculateur Exa1, met en œuvre une technologie de processeurs « **manycore** », interconnectés par le réseau à très hautes performances **BXI** (voir l'article « Le réseau d'interconnexion – élément clé d'un supercalculateur parallèle »), co-développé par Atos/Bull et le CEA/DAM, intégrés dans la nouvelle plateforme Bull Sequana. Cette machine atteint 25 **pétaflops** pour une consommation électrique de 4 MW, soit un gain en efficacité énergétique de 25 par rapport à Tera100.

En juin 2018, Tera1000 a été classé en 14^e position dans le **Top500** mondial et est devenu le supercalculateur généraliste le plus puissant d'Europe. Tera1000 permet de réaliser des simulations numériques d'une précision inégalée et de préparer l'avenir en préfigurant, notamment avec son architecture hétérogène, l'architecture d'Exa1.

Au-delà de la sphère nationale, le CEA est membre fondateur de la plateforme européenne ETP4HPC, dont l'objectif est d'accélérer le développement des technologies européennes à tous les niveaux de la chaîne du **HPC**, technologies indispensables dans la plupart des secteurs industriels. En particulier, le CEA participe activement au projet de **processeur** européen **EPI**, réalisé dans le cadre de l'initiative **EuroHPC**.

À plus long terme, une évaluation des possibilités des futurs calculateurs « quantiques » est engagée et, dans ce cadre, le **CCRT** accueille le simulateur **QLM30**, construit par Atos, qui, avec des technologies classiques, émule un processeur

quantique et permet d'en évaluer l'intérêt sur différentes applications.

Plus généralement, le CEA/DAM continue à mettre à profit l'expertise de ses équipes et la qualité de ses partenariats pour relever ces nouveaux défis et répondre aux besoins de la défense pour une dissuasion crédible et pérenne, tout en continuant à jouer son rôle de catalyseur au profit de la recherche académique et du développement industriel. ■

CHAPITRE 2

LES SUPER-CALCULATEURS

Un supercalculateur est constitué par l'interconnexion de milliers d'ordinateurs. Il produit une somme gigantesque d'informations sous la forme de données qui doivent être stockées et analysées en vue de répondre aux questions posées par les ingénieurs et les chercheurs. Comme tout grand instrument, il nécessite une infrastructure d'accueil spécifique, un centre de calcul, qui doit être surveillé en permanence afin d'en superviser le fonctionnement et d'en assurer la sécurité.

2.1 Le supercalculateur dans son centre de calcul

— Un supercalculateur, assemblage de très nombreux processeurs reliés par un réseau ultrarapide, assorti de moyens de stockage massifs de données est implanté dans une structure d'accueil, le centre de calcul, qu'il faut surveiller en permanence pour s'assurer de son bon fonctionnement et le protéger contre tout type d'attaque informatique.

Pour réaliser le plus grand nombre d'opérations simultanées, les supercalculateurs effectuent les calculs « en parallèle », c'est-à-dire en les répartissant sur différents **processeurs**. Ceux-ci sont organisés en « grappe » de « **nœuds de calcul** » (voir infographie), connectés

par un réseau ultrarapide. Les nœuds de calcul mettent en commun leur mémoire pour former une mémoire « distribuée » de très grande taille ; ils sont reliés à des espaces de stockage de plus grande taille encore. L'architecture détaillée des nœuds de calcul est un élément déterminant pour optimiser leur fonctionnement. Les performances d'un supercalculateur décrivent ses capacités à exécuter des calculs, mais aussi à traiter de très grands volumes de données. On parle de « calcul haute performance » (ou « HPC », pour High Performance Computing en anglais) et la vitesse de traitement des opérations s'exprime principalement en **flops** (Floating Point Operations Per Second, opération de base pour du calcul numérique, soit addition ou multiplication décimale).

Physiquement, les supercalculateurs sont constitués de nombreuses armoires, regroupant plusieurs nœuds de calcul, reliées entre elles par des kilomètres de câble réseau (pour l'interconnexion des nœuds de calcul) et installées dans des centres de calcul. Un centre de calcul comprend aussi de gigantesques capacités de stockage local de données auxquelles les supercalculateurs doivent pouvoir accéder rapidement (dizaines de « **pétaoctets** », contrepartie des dizaines de « **pétaflops** » de puissance de calcul).

LES CENTRES DE CALCUL

Comme ces machines sont de plus en plus puissantes et denses, leur consommation électrique devient très importante et dégage énormément de chaleur – tant dans les processeurs que dans les mémoires et réseaux de communication. Il faut donc mettre en place un système de refroidissement efficace, lui-même le moins énergivore possible – par exemple par circulation d'eau dans les portes des armoires ou au plus près des nœuds de calcul – ainsi qu'une climatisation dans la salle machine. Bien optimisés, ces systèmes de refroidissement consomment une fraction minoritaire de la consommation électrique globale, l'essentiel de l'énergie apportée au centre de calcul peut alors servir directement aux calculs et traitements de données, et les coûts de fonctionnement sont mieux maîtrisés.

Un centre de calcul et de traitement des données est conçu pour offrir différents services aux utilisateurs, en premier lieu de la puissance de calcul et des ressources de stockage de données. Il doit aussi proposer des moyens de communication et de transfert de données entre les éléments de calcul, ainsi qu'avec les ressources de stockage et les postes des utilisateurs situés à l'extérieur (soit sur le site local, soit sur des sites distants).

Pour résumer, les différents éléments d'un centre de calcul et de traitement de données sont les suivants :

- pour la partie informatique : des supercalculateurs, des systèmes de stockage, des réseaux, des serveurs spécialisés ;

Centre de calcul et de traitement de données

Machine de calcul fonctionnant selon une architecture massivement parallèle : il décompose les tâches à effectuer en milliers de sous-tâches. Chacune d'elles est ensuite réalisée, de manière simultanée, par un processeur (**unité de calcul**). En agrégeant un grand nombre de processeurs, ce procédé permet d'atteindre de grandes puissances de calcul. De même, en agrégeant un grand nombre d'unités de stockage, l'ensemble offre un volume de stockage sécurisé et parallèle de taille très importante.

Séquençement de la simulation

- 1 Demande de calcul par l'utilisateur.
- 2 Attribution d'une capacité de calcul, donc d'un nombre de processeurs avec de la mémoire, pendant un certain temps, selon les besoins et les disponibilités du supercalculateur.
- 3 Réalisation des calculs par le système de calcul.
- 4 Transition des résultats de calcul par le réseau de connexion vers le système de stockage.
- 5 Récupération et visualisation de la simulation numérique par l'utilisateur.

Consommation et alimentation électriques

- 6 MW en régime de production.
- 25 km de câbles.

Système de calcul Tera000-2

- **Architecture** : plusieurs centaines de milliers de **cœurs de calcul** regroupés dans une dizaine de milliers de processeurs multicœurs qui équipent des serveurs mono ou bi-processeurs interconnectés par un réseau ultrarapide. Le tout est contenu dans une centaine d'armoires informatiques.
- **Puissance de calcul** : quelques dizaines de pétaflops de puissance crête, soit une capacité de calcul de plusieurs dizaines de millions de milliards d'opérations par seconde.
- **Mémoire vive** : quelques milliers de **téraoctets**, soit la capacité de dizaines de milliers d'ordinateurs classiques.
- **Surface occupée au sol** : plusieurs centaines de m².

Refroidissement

Système mixte eau tiède/air ; le système à eau refroidit les serveurs de calcul, soit 90 % de toute la consommation de la machine. L'eau à 40 °C, par contact direct, refroidit les éléments chauds de la machine (processeurs, mémoires, alimentations électriques) :

- 1 kilomètre de tuyauteries installé dans les faux planchers ;
- système de climatisation de la salle par air pour le complément.

Réseau d'interconnexion

Structure en « arbre » reliant de proche en proche les éléments de calcul et utilisant deux niveaux de commutateurs. Fonctions : assurer la communication ultrarapide entre les processeurs ; envoyer les informations traitées vers le système de stockage.

Débit de données

Plus d'un téraoctet par seconde pour l'entrée et la sortie des données, qui ont lieu simultanément. Équivalent : capacité de transfert pour plusieurs millions de personnes regardant en même temps un film en haute définition.

Système de stockage

- Capacité : plusieurs dizaines de pétaoctets à l'aide de dizaines de milliers de disques. Équivalent : quelques dizaines de milliards de livres.
- Durée du stockage : indéfinie par migration au cours du temps (voir encadré).

→ Définitions

Puissance crête

Puissance théorique correspondant à l'addition de la puissance maximale de tous les processeurs de la machine. La puissance réellement mesurée sur une application étant la puissance dite « soutenue ».

Commutateur

Équipement qui relie plusieurs segments dans un réseau informatique.

Un exemple de supercalculateur fonctionnant selon une architecture massivement parallèle. © F. Mathé/CEA

© pour l'infrastructure d'accueil : un réseau d'alimentation électrique et des dispositifs de climatisation.

Cela doit être complété par un dispositif assurant la supervision et la sécurité de l'installation, qui doit notamment comprendre un dispositif d'alerte, *via* des sondes thermiques, et un mécanisme d'extinction en cas d'incendie.

LES SUPERCALCULATEURS DU CEA/DAM

Les supercalculateurs Tera sont conçus sur un même principe : agréger et relier entre eux des milliers de processeurs afin d'atteindre de très grandes puissances de calcul. Chaque unité de traitement est associée à une mémoire en propre et est hébergée dans un serveur. Des milliers de serveurs sont répartis dans des armoires et sont refroidis, pour la génération actuelle, par un réfrigérant liquide *via* un circuit parcourant les cartes électroniques (avec des dissipateurs positionnés sur les composants les plus chauds). Des **logiciels**

« **open source** », développés en partie par les experts du CEA/DAM, permettent de fédérer ces ressources et d'en faciliter leur administration et leur utilisation.

LES SYSTÈMES DE STOCKAGE

Utilisant aussi le principe de l'agrégation et du **parallélisme**, les systèmes de stockage sont constitués de dizaines de milliers d'unités de stockage selon le principe de l'architecture **RAID**. Ces unités sont regroupées dans des armoires mettant en œuvre des mécanismes de tolérance aux pannes basés sur des principes de redondance. Les données sont associées à des **sommes de contrôles** qui, en cas de perte d'une ou de plusieurs unités de stockage, permettent de reconstruire les données inaccessibles.

Les données réparties sur les unités de stockage sont fédérées grâce à un système de fichiers parallèle (au CEA/DAM le logiciel **Lustre** – voir chapitre « Les logiciels ») qui présente une

Dominique Martinet optimise le stockage des données

Pour utiliser au mieux les données, il faut les stocker intelligemment. C'est la tâche de Dominique Martinet, qui développe des logiciels de stockage spécifiques et s'attache à aider au mieux les utilisateurs au quotidien.

— Dès qu'on parle de capacités de stockage des grands calculateurs, les chiffres donnent le vertige. Le Très Grand Centre de calcul (TGCC) du CEA/DAM, par exemple, stocke 50 pétaoctets de données, l'équivalent de ce que contiennent 1 000 000 téléphones portables modernes. Cela nécessite le recours à 20 000 cartouches de bandes magnétiques et près de 6 000 disques durs. Les transferts de données sont tout aussi impressionnants : tous ces disques sont connectés par 88 câbles de 100 gigabits, qui pourraient permettre à 5 000 000 personnes de visionner des vidéos YouTube en même temps.

Robin des bois

De telles quantités de données requièrent des méthodes de gestion spécifiques. C'est le travail de Dominique Martinet, ingénieur dans l'équipe de stockage du TGCC. Parmi les experts en stockage, certains développent des logiciels pour répondre aux besoins des utilisateurs, d'autres administrent le stockage (ils suivent les incidents, résolvent les problèmes...). Dominique Martinet, lui, assure les deux. « J'aime toucher à tout, voir toute la chaîne jusqu'à l'utilisateur, explique-t-il. C'est intéressant de discuter pour améliorer les choses, et avoir des retours de la part des utilisateurs. J'ai travaillé dans d'autres centres de calculs, au Japon notamment, où les tâches sont plus séparées, et les développeurs ne voient pas si, en pratique, les gens sont contents. »

Parmi les logiciels permettant d'organiser le stockage, l'un a été joliment nommé Robinhood (Robin des bois), le logiciel qui prend de l'espace de stockage aux « riches » pour le donner aux « pauvres ». Il s'appuie sur l'analyse statistique de la manière dont les utilisateurs accèdent

aux fichiers et permet à ceux-ci de gérer leurs stockages selon leurs besoins d'accès.

Simplicité d'utilisation

En plus de gérer la répartition des données sur les différents types de stockage, Dominique Martinet régit la manière dont cela est présenté aux utilisateurs. Mais il l'avoue : « Nous pouvons beaucoup progresser sur la simplicité d'utilisation ! Pour les machines Tera, ce n'est pas très grave car nous avons affaire aux mêmes utilisateurs, nous pouvons leur apprendre. En revanche, sur TGCC, il y a plus de turnover, de nouveaux doctorants et chercheurs, et l'on voit davantage d'erreurs dues à la complexité d'utilisation. » Après sept ans passés au CEA/DAM, Dominique Martinet continue d'adorer la variété de ce travail. « Je ne vois pas ça chez mes condisciples qui sont aujourd'hui dans le privé », expose-t-il. Il aime aussi se mettre au service des utilisateurs, apprendre à leur contact, avoir leur retour pour améliorer le stockage. « On fait bien-sûr des tests pour éviter les problèmes, mais il en reste forcément. Et les utilisateurs font des choses auxquelles on ne s'attend pas », s'amuse-t-il. Son entrée au CEA s'est faite un peu par hasard, grâce à un maître de stage très convaincant. Dominique Martinet ne l'a jamais regretté, et souhaite à son tour promouvoir le CEA/DAM auprès des étudiants. « Je leur montre que l'on peut faire des choses qu'on ne fait pas ailleurs. Il y a beaucoup de superbes matériels à tester : les premiers prototypes fonctionnels de supercalculateurs sont chez nous. On recrute par contacts, je propose des stages et des thèses. On convainc grâce à l'intérêt du travail. »

arborescence de fichiers aux utilisateurs. Les supports physiques d'information sont variés : bandes magnétiques (**tape**), disques rotatifs (**HDD**) ou plus récemment des disques flash (**SSD**).

LES RÉSEAUX

Toutes les ressources de calcul ou de stockage sont reliées *via* des réseaux ayant différentes

caractéristiques selon l'usage pour lequel ils ont été conçus. Par exemple, les réseaux reliant les nœuds de calcul sont à haut débit et à faible **latence**. Ils sont optimisés pour les **échanges de messages** et peuvent être spécifiques à un type de supercalculateur. En revanche, les réseaux qui relient les supercalculateurs et les systèmes de stockage sont définis selon des standards internationaux afin de garantir une interopérabilité entre systèmes provenant de différents fournisseurs. ■

Organiser le centre de calcul autour des données

Grâce aux logiciels de simulation numérique, les supercalculateurs sont des instruments qui produisent des volumes de données très importants. Le volume de ces données a été multiplié par un facteur 100 depuis 2001 (Tera1) et c'est actuellement plus de 100 To (téraoctet) par jour qu'il faut savoir traiter, stocker temporairement puis extraire du cluster pour enfin les analyser.

Ce flux important de données nécessite une architecture et des serveurs spécialisés : ainsi, au sein du supercalculateur, une centaine de nœuds, appelés les nœuds routeurs, sont ainsi dédiés pour transférer à 1 To/s les données vers un espace centralisé de stockage d'une dizaine de Po. Cet espace de stockage est situé au cœur du centre de calcul, dont l'architecture générale est organisée autour des données (**figure E1**).

Ce système de stockage doit être dimensionné pour être capable d'absorber ce déluge d'informations. Différents types de données sont cependant à distinguer :

- ⊙ les données en cours d'utilisation, dont le contenu évolue de façon instantanée (ou du moins à l'échelle de la durée d'une simulation) ;
- ⊙ les données constantes, qui sont fréquemment lues mais jamais modifiées ;
- ⊙ les données dormantes, qui ne sont pas ou peu utilisées.

Ces différents types de données ne doivent pas être gérés de la même manière afin d'optimiser leur stockage. C'est ainsi que le système de stockage a été conçu : un système de stockage à plusieurs niveaux. Le premier niveau est un système de fichiers rapide, une sorte de zone tampon permettant d'absorber le flux de données et toutes les opérations extrêmement rapides engendrés par le calculateur. Le second niveau est

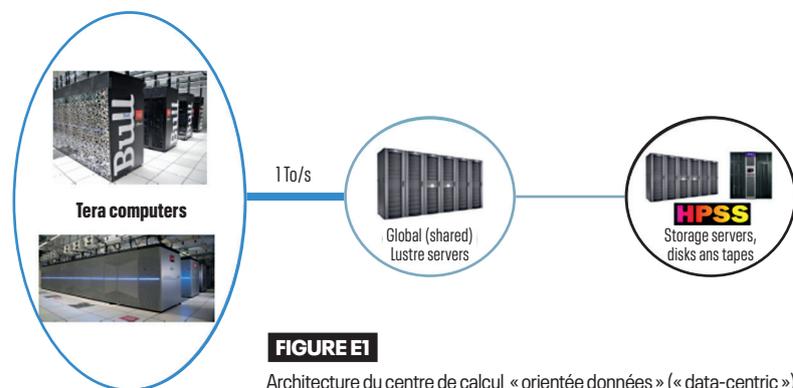


FIGURE E1

Architecture du centre de calcul « orientée données » (« data-centric »).

composé de disques plus lents mais plus capacitifs permettant d'accueillir les données moins fréquemment utilisées. Un troisième niveau de stockage, basé sur l'utilisation de bandes magnétiques (de capacité 13 To par bande actuellement), permet de conserver à long terme un grand volume de données sans consommer d'énergie. Les données décanteront alors progressivement du système de fichiers rapide vers le système de bandes en passant par le second niveau de disque, au fur et à mesure qu'elles seront de moins en moins utilisées.

Gérer la vie de la donnée

Face à la variété des types de support, les algorithmes au cœur des systèmes de stockage prennent la décision d'enregistrer ou de déplacer les données sur le média offrant le meilleur rapport performance/service. Ces choix sont effectués en fonction des profils d'utilisation des données au cours du temps. Ainsi, un fichier peu utilisé est hébergé sur une bande, alors qu'un autre utilisé tous les jours est placé sur des disques rapides. Cette hiérarchie de stockage est adossée à des politiques de migration qui permettent de choisir le meilleur endroit pour conserver la donnée.

Ce fonctionnement est mieux décrit en associant une « température » à chaque fichier. Ainsi, un fichier utilisé fréquemment par l'utilisateur est considéré comme « chaud » car on y accède très régulièrement. Au fil de son existence, étant de moins en moins utilisé, le fichier se « refroidit » petit à petit.

Un fichier nouvellement créé et qui accueille sans cesse de nouvelles informations est considéré comme « brûlant ». Il commence sa vie sur disques SSD (Solid State Drive). Un peu moins utilisé par la suite, il peut être déplacé de manière automatique sur disques rotatifs HDD (Hard Disk Drive). À terme (quelques jours ou semaines), peu utilisé, il devient plus « froid » et est conservé sur bandes magnétiques. Ce processus de « sédimentation » des informations, qui voit les données les plus « froides » descendre le long des différentes couches de la hiérarchie de stockage, est au cœur des systèmes de stockage de données du HPC. Lors d'une utilisation ultérieure de la donnée, celle-ci est automatiquement remontée dans le premier niveau de la hiérarchie. Afin d'optimiser des transferts et d'éviter les oscillations entre les niveaux, des copies multiples et cohérentes sont conservées entre les différents niveaux de la hiérarchie.

2.2 Surveillance du centre de calcul Tera

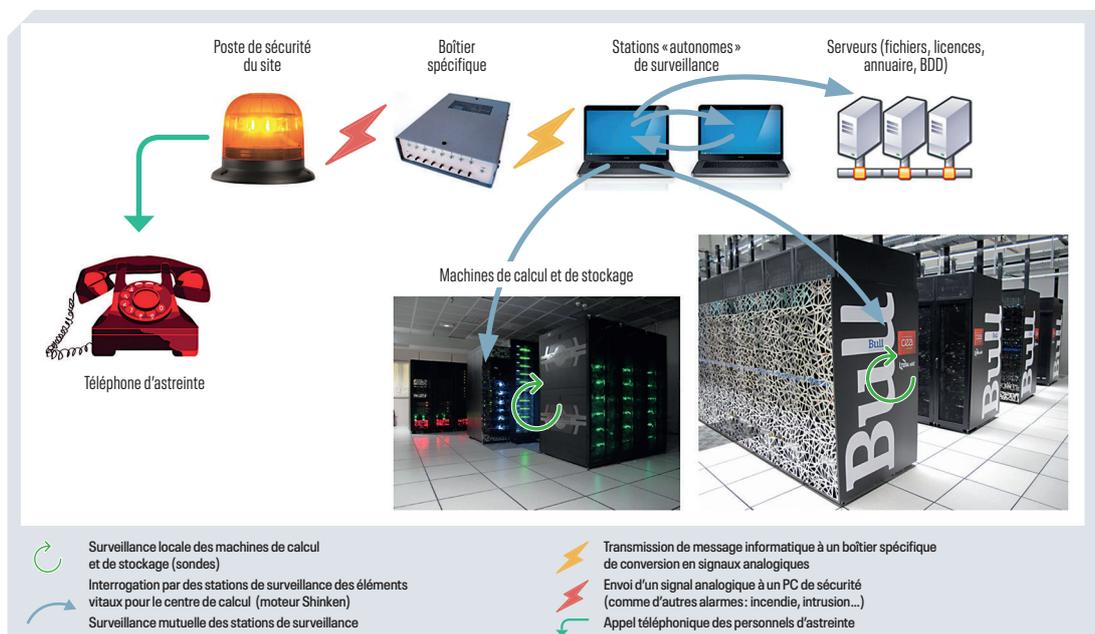
— Pour tirer le meilleur parti du système extraordinairement complexe qu'est un centre de calcul, il faut en permanence en vérifier le fonctionnement. Ceci est obtenu par le recueil et l'analyse automatique des informations sur le fonctionnement des éléments les plus critiques.

La surveillance est un élément fondamental de l'exploitation d'un centre de calcul. Son but est d'assurer le meilleur fonctionnement possible des éléments nécessaires à la bonne exécution des calculs. Elle regroupe deux types d'activités différentes : la supervision et la métrologie.

LA SUPERVISION

La supervision (figure 1) est un ensemble de mécanismes automatiques qui fonctionnent en permanence et vérifient que le centre de calcul rend les

services attendus par les utilisateurs : se connecter, soumettre des calculs, les effectuer et, enfin, lire et écrire d'importants volumes de données. Chaque fonctionnalité est vérifiée par l'intermédiaire de sondes spécifiques (voir encadré « Sonde de surveillance ») dont les résultats sont agrégés suivant des règles simples (pourcentage de nœuds opérationnels, nombre de serveurs disponibles...) qui permettent de déclencher 24 h/24, soit des comportements préprogrammés (redémarrage de service par exemple), soit des alarmes informatiques qui sont transmises à l'équipe de surveillance, sur place ou à distance. La supervision permet donc de



réagir au plus vite en cas de comportement anormal du centre de calcul.

Trois parties principales forment le système de supervision :

- ① **1. Le moteur d'ordonnancement des tâches** définit la fréquence du sondage, agrège les résultats et déclenche des actions. Il utilise généralement des **logiciels libres**, définis et développés en mode communautaire (**open source**), comme Shinken sur Tera100.
- ② **2. Les sondes** exécutées sur les éléments à surveiller. Une partie d'entre elles arrivent avec le moteur d'ordonnancement, mais le fournisseur du supercalculateur ou les équipes du CEA/DAM peuvent en écrire de nouvelles afin de s'adapter aux spécificités du centre de calcul. Ces sondes sont ajoutées dans le moteur sous forme de grefons (« **plugin** »).
- ③ **3. L'interface web** permettant d'interagir au quotidien avec le moteur pour changer des paramètres, désactiver un contrôle ou simplement consulter l'état du système de surveillance. Des logiciels libres remplissent ce rôle et donnent la possibilité de créer des tableaux de bord synthétiques.

LA MÉTROLOGIE

La métrologie permet de mesurer et d'enregistrer en continu l'activité du centre de calcul afin d'étudier le comportement des différents composants et permettre ainsi d'anticiper les problèmes, d'ana-



MATTHIEU HAUTREUX

« Je dois garantir le fonctionnement d'un grand nombre de serveurs.

C'est le même métier que ceux qui font fonctionner les data centers, mais en plus nous sommes très focalisés sur l'optimisation des performances des utilisateurs, souligne-t-il. C'est du data center de pointe. »

lyser une situation *a posteriori* ou d'optimiser les paramètres des systèmes (figure 2).

Les premières données collectées il y a quelques années sur les premières machines Tera étaient envoyées dans un outil qui gérait une base de données de taille constante. Pour cela, les données les plus anciennes étaient moyennées sur des périodes de plus en plus longues (jour, semaine, mois, année). Cette approche ne permettait pas une analyse fine d'un phénomène

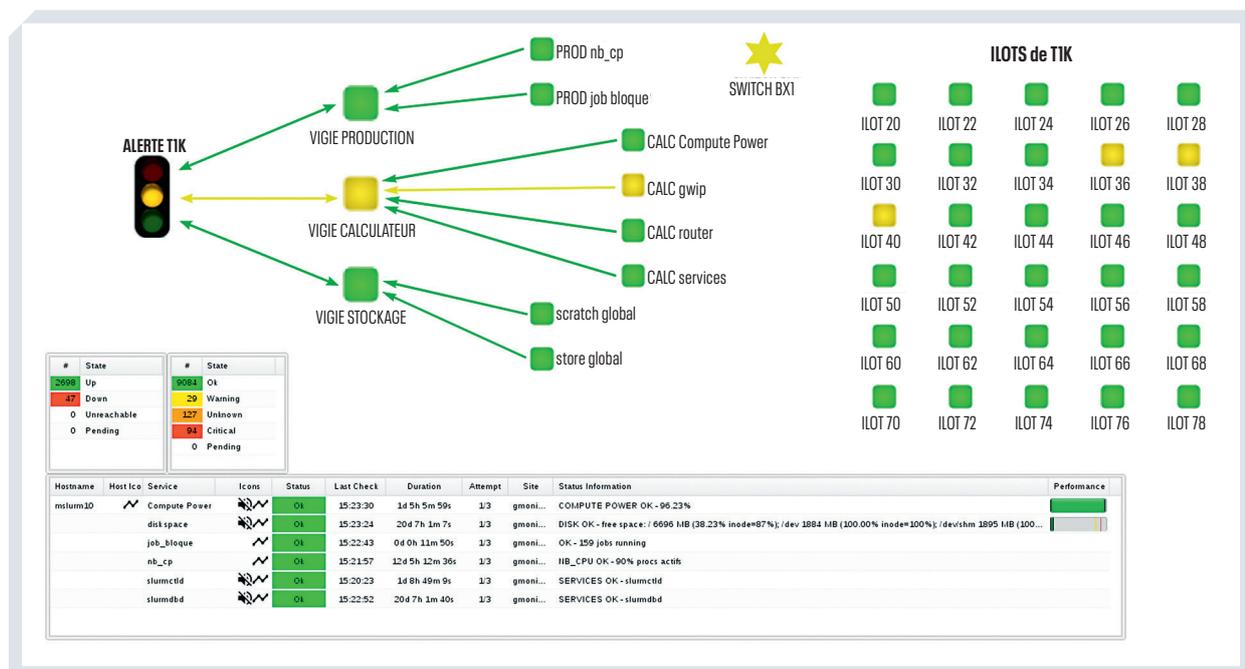


FIGURE 2

Synoptique, sous Panorama, de l'état de Tera1000 intégré à l'outil Shinken. Le code couleur est le suivant: vert = tout va bien, jaune orangé = warning = petit problème ou absence d'information, et rouge = équipement inopérant.

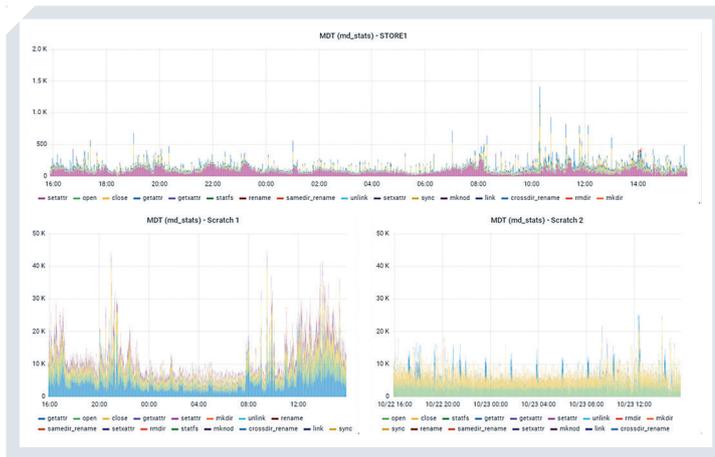


FIGURE 3

Exemple de tableau de bord représentant, via Grafana, des opérations (ouverture de fichiers, fermeture, lecture, écriture, modification des droits d'accès) effectuées sur les systèmes de stockage des calculateurs. Ce type de données permet la détection rapide de dégradations de performances liées soit à un problème technique, soit à une mauvaise utilisation.

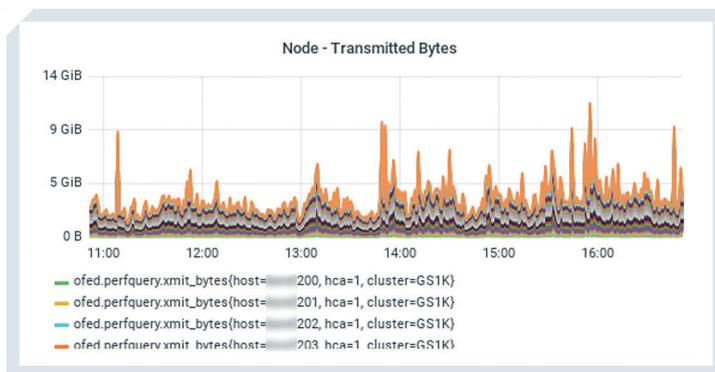


FIGURE 4

Exemple de graphe représentant, via Grafana, les opérations de transfert de données effectuées en parallèle sur les systèmes de stockage des calculateurs. Ce type de données permet de visualiser simplement les pics de transferts et ainsi de pouvoir les corréliser à d'autres événements du centre de calcul (baisse de performance, temps d'attente augmentés...).

ancien ; aussi, au fur et à mesure des nouvelles générations de supercalculateurs, un ensemble de serveurs dédié à la collecte et à l'enregistrement sans perte de données a été mis en service en utilisant une nouvelle base de données issue des technologies de traitement des grands volumes de données. Ce sont désormais plusieurs dizaines de milliers d'informations qui sont collectées chaque seconde. Ces informations sont ensuite accessibles interactivement aux équipes d'exploitation par l'intermédiaire d'interfaces web sous forme graphique, numérique ou textuelle grâce à des logiciels libres comme Grafana (figures 3 et 4).

CONCLUSION

Ces mécanismes de surveillance sont installés sur le centre de calcul depuis les premières machines Tera et reposent sur un ensemble de briques logicielles libres permettant de construire des solutions opérationnelles complètes. Ces logiciels ne sont pas spécifiques aux supercalculateurs ou aux centres de calcul. Ils sont largement utilisés pour surveiller également des parcs de machines individuelles ou des centres de données. Cela permet d'une part d'interagir avec une large communauté d'utilisateurs chevronnés en cas de problème ou de besoin nouveau et d'autre part d'être indépendant des constructeurs informatiques. Ce dernier point apporte de la portabilité et de la pérennité aux systèmes mis en place et a l'avantage de dissocier l'installation des calculateurs des systèmes de surveillance. ■

Sonde de surveillance

Une sonde de surveillance est un petit programme informatique dont le but est de donner rapidement une information simple. Elles sont soit développées par le fournisseur de la machine (Atos/Bull) quand elles sont liées au matériel, soit par le CEA/DAM quand elles sont liées à un service propre au centre de calcul. Ces sondes peuvent être exécutées de quelques fois par jour à plusieurs fois par minute, voire par seconde. Il est donc primordial que le temps de traitement soit le plus court possible et surtout adapté à la période d'exécution. De même, ces sondes s'exécutent sur différents types de machines (nœud de calcul, nœud d'administration, serveurs...), générale-

ment dédiées à des activités autres que la surveillance, elles doivent être les plus légères possibles en ce qui concerne les ressources informatiques (mémoire, CPU, réseau...). Elles sont classiquement écrites dans des langages interprétés, tels que Bash, Python ou Ruby, plus rapides à modifier et à porter sur des machines et des systèmes d'exploitation différents que des langages compilés. L'information donnée par une sonde doit être simple pour être facilement affichable par des outils génériques ou être agrégée avec d'autres informations. Le plus souvent, cette information est un booléen (on/off, service démarré ou pas, réponse à une requête réseau, chaîne

de caractères trouvée dans un fichier...) ou une valeur numérique (débit réseau, nombre d'écritures par seconde...). Lorsque ces informations sont nombreuses (période d'exécution faible et/ou nombre de machines élevé), il est possible de les agréger par des règles permettant de les rendre humainement interprétables (pourcentage, moyenne par minute...) ou de les comparer à un critère de seuil afin de déclencher une alarme, par exemple si le nombre de nœuds opérationnel est inférieur à 95% du maximum. Ces informations nombreuses peuvent également servir à alimenter les bases de données du système de métrologie qui serviront à leur tour à générer des tableaux de bord.

2.3 Attribuer les ressources du centre de calcul

— S'il peut être entièrement mobilisé au profit d'une seule simulation, un supercalculateur est la plupart du temps partagé par un grand nombre d'utilisateurs qui cherchent à s'attribuer une partie des ressources. Un logiciel d'ordonnancement est donc nécessaire pour en optimiser l'utilisation et satisfaire au mieux l'ensemble des utilisateurs.

Pour faire exécuter un travail de simulation numérique sur les ressources du centre de calcul (supercalculateur, stockage de données, réseaux...), l'utilisateur doit réaliser plusieurs étapes : tout d'abord, il lui faut indiquer l'endroit où sont stockées les données d'entrée de la simulation, puis réserver les ressources de calcul (nombre de processeurs par exemple), enfin commander l'exécution de sa simulation. À l'issue de celle-ci, il pourra récupérer, sur son espace de travail, les résultats de calcul sur lesquels il effectuera, s'il le souhaite, de nouveaux traitements *via* des ressources informatiques spécialisées.

Sur un calculateur à architecture parallèle, la commande d'exécution d'un code de calcul n'est pas triviale. Contrairement à un ordinateur domestique, celui-ci est composé de nombreux et différents éléments (**noeuds de calcul**, réseau d'interconnexion, stockage, services...). L'utilisateur doit donc réserver un sous-ensemble de ces différentes ressources afin de pouvoir commander l'exécution de son calcul.

La demande d'allocation des ressources à une simulation doit notamment préciser : la durée de réservation, le nombre de noeuds de calcul, la taille de la mémoire et l'espace de stockage requis.

Cette demande est mise dans une file d'attente et traitée en fonction de la priorité de la demande, des quotas spécifiques étant attribués à tous les utilisateurs.

Le mécanisme d'allocation des ressources est automatique et doit prendre en compte les paramètres ou contraintes suivants :

- ⊙ l'ensemble des demandes d'allocation faites par les utilisateurs ;

- ⊙ les ressources disponibles et leurs spécificités ;
- ⊙ la politique de définition des priorités en fonction des études et des quotas alloués à une équipe ;
- ⊙ la durée optimale d'allocation des ressources en fonction de la charge définie par les exploitants du centre de calcul ;
- ⊙ et éventuellement d'autres contraintes, comme l'optimisation énergétique ou la topologie du calculateur.

L'allocation précise des ressources est déterminée par des algorithmes d'ordonnancement sous contraintes.

Cependant, la durée de simulation demandée par les utilisateurs est souvent supérieure à la durée maximale de réservation des ressources définie par les exploitants. En effet, certaines simulations peuvent nécessiter plus d'une semaine de calcul, alors même qu'il n'est pas toujours possible de bloquer les ressources au profit d'un seul utilisateur.

Il est donc indispensable que le code de calcul dispose d'un mécanisme de sauvegarde et de récupération de l'état de ses données en cours d'exécution ; les équipes du CEA/DAM fournissent aux utilisateurs les fonctions à appeler à cette fin si nécessaire. Cette fonctionnalité permet aussi à l'utilisateur de faire avancer sa simulation par étapes, les résultats d'une étape donnée étant les entrées de l'étape suivante. L'utilisation des ressources de calcul, le suivi et la relance automatique de l'exécution des travaux sont facilités par un logiciel développé par les experts du CEA/DAM.

Une interface graphique est proposée à l'utilisateur indiquant le nombre d'exécutions du code et l'avancement du calcul. L'utilisateur peut alors, à distance *via* son navigateur, suivre l'état d'avancement de son calcul. ■

2.4 Protéger les centres de calcul des cyberattaques

— Un centre de calcul, nécessairement en interaction avec l'extérieur, ne serait-ce que par ses utilisateurs, est vulnérable. Il doit en conséquence être protégé contre les malveillances et les erreurs humaines.

La sécurité informatique a toujours été un élément à intégrer dans la conception et l'exploitation des centres de calcul et des supercalculateurs. La protection des éléments informatiques composant un calculateur et, de manière plus générale, un centre de calcul a été très vite considérée comme un point stratégique. Si la première protection des calculateurs est la protection physique et, pour ce qui concerne les calculateurs de la dissuasion, la non-connexion au réseau Internet public, il n'en reste pas moins que l'évolution et la sophistication des techniques d'attaque nécessitent des mesures de protection complémentaires. Avec l'arrivée d'Internet, de plus en plus d'informations sur les techniques et les outils informatiques (malveillants ou non), ainsi que la médiatisation de plus en plus grande des intrusions informatiques, des vols d'information ou des dégâts provoqués par les logiciels malveillants (**virus**, **bombe logicielle**, **vers**...) ont renforcé cette volonté de sécuriser ces systèmes. Pourtant, le domaine du calcul haute performance qui tend à privilégier la vitesse de calcul et les échanges d'informations ne semblait pas en adéquation avec le domaine de la sécurité informatique, plus propice à limiter et à contrôler ces mêmes échanges ainsi qu'à interdire tout ce qui n'est pas explicitement autorisé. Mais la synergie a opéré et la sécurité informatique a réussi à conquérir tous les composants des supercalculateurs et des centres de calcul en s'y implantant pour couvrir les principes de base que sont la confidentialité (une information ne doit être accessible que si l'on en a besoin), l'intégrité (une information ne doit pas être modifiée

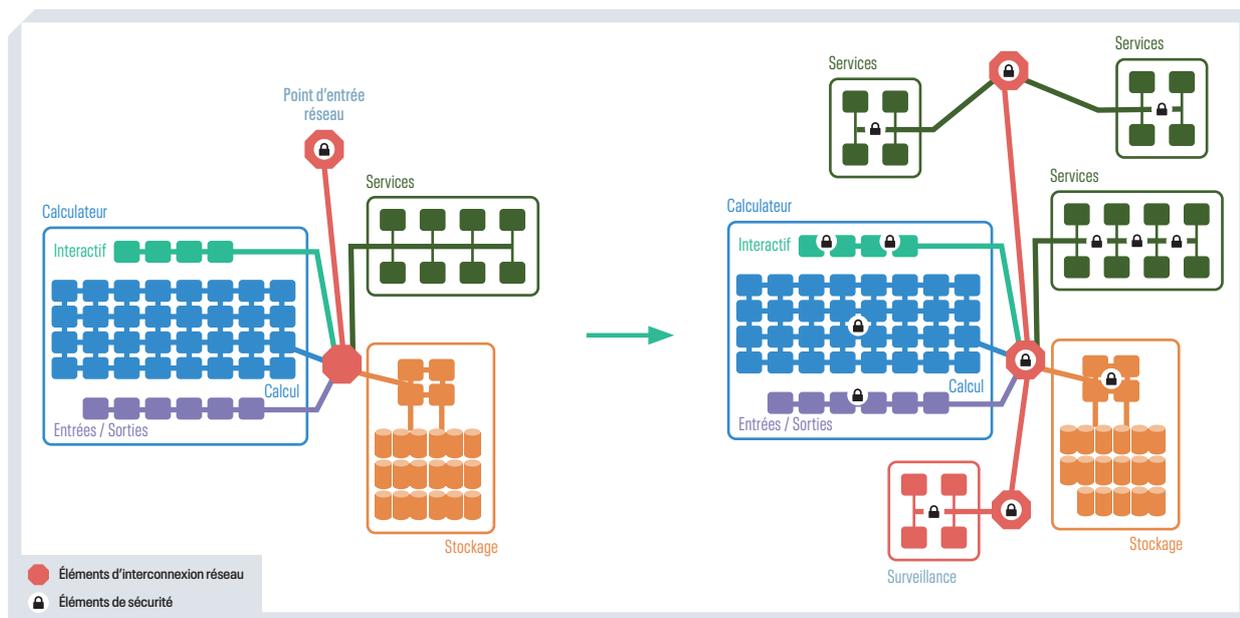
sans raison valable), la disponibilité (un système doit toujours être capable de rendre le service pour lequel il est conçu) et la traçabilité (savoir qui a fait quoi).

AU-DELÀ DE LA PROTECTION PÉRIMÉTRIQUE

Les premiers supercalculateurs intégraient des mécanismes de contrôles et de limitations directement dans leur système de base, ce qui permettait d'assurer une certaine sécurité interne en limitant les impacts d'une **action malveillante** à l'intérieur du calculateur. Cependant, la **sécurité périmétrique** permettant de limiter le périmètre d'attaque des interfaces avec l'extérieur était assez peu développée.

Avec l'arrivée des supercalculateurs architecturés autour de **grappes de serveurs** génériques interconnectées à l'aide de réseaux de plus en plus performants et basés sur des **logiciels libres**, la tendance s'est, dans un premier temps, inversée en se concentrant sur les aspects de la protection périmétrique, laissant ainsi la puissance de calcul intacte, les éléments de sécurité ne venant pas dégrader les performances. Puis, les logiciels libres ont pu montrer tout leur potentiel en permettant, plus simplement, d'ajouter, de modifier, d'améliorer les différents composants et d'y intégrer des éléments de sécurité.

À cette première avancée de la cybersécurité, il faut ajouter la sécurisation de l'ensemble du centre de calcul (composé d'un ou plusieurs supercalculateurs, de dispositifs de stockage autorisant la gestion de plus, et de services informatiques de base comme une référence tempo-

**FIGURE 1**

Passage de la sécurisation périmétrique à une sécurisation en profondeur du centre de calcul.

relle commune ou une base de comptes utilisateurs), car les infrastructures informatiques permettant le fonctionnement du calculateur pouvaient, également, apporter des éléments de sécurisation importants dans ce processus.

Le premier élément fut le réseau global du centre de calcul et sa segmentation pour produire une **architecture cloisonnée** aidant à limiter les impacts d'une **compromission**. Le deuxième élément important fut la mise en place de méthodes d'authentification (comment prouver son identité) de plus en plus robustes (difficiles à falsifier, à rejouer, à voler, à deviner) et, petit à petit, généralisées à tous les services et les ressources. Enfin, l'arrivée des outils, notamment **SSL** ou **TLS**, implémentant les concepts de chiffrement (rendre une information incompréhensible pour qui ne détient pas la clé de déchiffrement) et de scellement de données (signature pour s'assurer de la non-modification et de l'émetteur de cette donnée) a permis de renforcer la confidentialité et l'intégrité des échanges informatiques.

SURVEILLER ET DÉTECTER

Grâce à toutes ces briques de base pour la sécurité informatique, il a été possible de développer de nouveaux modes de contrôle et de commencer à réaliser de la surveillance orientée sécurité. En effet, il apparaît assez vite que toutes les vulnérabilités (failles logicielles permettant une attaque informatique) et leurs exploitations dans un système informatique ne peuvent être toutes évitées, corrigées ou mitigées (rendre plus difficile la réalisation d'une

attaque) techniquement. Dans ce cas, il est nécessaire de pouvoir au moins détecter cette exploitation pour pouvoir prendre des mesures *a posteriori* (comme les systèmes de détection d'intrusions permettant de détecter qu'une attaque informatique est en cours). Cette surveillance oblige à récupérer les informations de fonctionnement du centre de calcul à tous les niveaux (réseaux, services, authentification...) et à les traiter pour en extraire des informations permettant la détection d'actions ou de comportements suspects déclenchant des alertes, conduisant à une réponse appropriée plus ou moins automatique.

LA DÉFENSE EN PROFONDEUR

Aujourd'hui, quasiment toutes les couches du système d'information composant un supercalculateur et un centre de calcul possèdent des mécanismes de sécurité adaptés à son niveau de service attendu (performance, robustesse, traçabilité...). Cependant, ces mécanismes peuvent être faillibles et comporter eux aussi des vulnérabilités, tout en n'étant pas immunisés contre les erreurs humaines. Pour limiter une nouvelle fois ces impacts, il est nécessaire de disposer de plusieurs mécanismes permettant le même contrôle ou la même surveillance, mais à des niveaux différents ou fournis par des composants différents. De cette manière, un défaut de fonctionnement d'un mécanisme peut être rattrapé par un autre et, ainsi, conserver un niveau de sécurité équivalent. C'est le principe de **défense en profondeur** (figure 1). Ce prin-

Optimiser le calcul haute performance et ses moyens grâce à l'intelligence artificielle

Les supercalculateurs s'avèrent parfois indispensables pour calculer les phases d'apprentissage et d'inférence des méthodes d'apprentissage automatique. À l'inverse, les algorithmes d'apprentissage peuvent être mis à profit dans la gestion des supercalculateurs. En voici quatre illustrations sur des études menées au centre de Bruyères-le-Châtel.

Afin d'optimiser la consommation d'un supercalculateur, il est souhaitable de prévoir le comportement d'une tâche avant que celle-ci ne lui soit soumise. Pour effectuer cette prévision, le système de gestion des soumissions de tâches ne dispose que de peu de données sur la tâche à accomplir : en général, le nom de l'utilisateur, celui du programme de lancement, un temps maximum pour le déroulement de la tâche et des caractéristiques de parallélismes, comme le nombre de processus MPI et le nombre de threads par processus MPI. Si ces informations ne suffisent pas à un expert pour anticiper la consommation d'une tâche, un algorithme d'apprentissage statistique – qui se fonde sur les données précédemment observées – est capable de conjecturer finement la puissance moyenne consommée de la tâche de calcul. Cette prédiction permet au système de soumission des tâches d'optimiser la puissance consommée totale du calculateur. Néanmoins, le résul-

tat de l'algorithme est dépendant des données en entrée et ce qui sera appris sur un supercalculateur n'est pas généralisable pour l'exploitation d'autres supercalculateurs : il faut refaire une rapide phase d'apprentissage pour adapter les prédictions à l'usage et aux caractéristiques propres à l'utilisation de chaque calculateur.

Dans le même ordre d'idées, on peut utiliser un réseau profond de neurones à convolutions pour prédire précisément la durée de vie d'un fichier (le temps entre sa création et sa destruction), en connaissant uniquement son nom complet. L'objectif : optimiser les mécanismes de migration de fichiers, ceux dont on prédit une durée de vie courte n'ayant pas vocation à être transférés vers les systèmes de sauvegarde.

Par ailleurs, les algorithmes d'apprentissage peuvent également s'avérer utiles pour la maintenance prédictive des supercalculateurs. En effet, les pannes sont systématiquement précédées par des dysfonctionnements mineurs qui laissent des traces infimes dans les journaux système. Ces journaux – qui se présentent la plupart du temps sous la forme d'un fichier texte classique – sont augmentés de 15 000 à 900 000 lignes par jour et par nœud de calcul. Or, un supercalculateur est composé de plusieurs milliers de nœuds : il y a donc de

plusieurs millions à quelques milliards de lignes de texte à surveiller chaque jour ! La détection des signaux faibles, un des cas d'application typique de l'apprentissage automatique, appliquée à ces journaux sert à repérer les dysfonctionnements et ainsi remplacer préventivement les composants matériels ou logiciels avant des pannes plus importantes.

Enfin, l'équipe qui gère les infrastructures du centre de calcul va utiliser l'apprentissage automatique pour optimiser le fonctionnement des servitudes des supercalculateurs hébergés par le centre. Des systèmes de mesures vont récolter des données qui serviront à entraîner des algorithmes d'apprentissage. Le résultat permettra d'affiner les différentes consignes et leurs durées de validité qui s'appliquent aux équipements constituant ces infrastructures. En ligne de mire : réaliser des économies et améliorer les indicateurs d'efficacité énergétique du centre de calcul.

Ces différents exemples illustrent bien l'apport de l'apprentissage machine et, plus largement, de l'intelligence artificielle dans l'optimisation des supercalculateurs. Ils s'inscrivent naturellement dans les travaux menés dans le cadre de la chaire « Industrial Data Analytics & Machine Learning » hébergée au sein de l'ENS Paris-Saclay.

cipe permet, également, la superposition d'une multitude de mécanismes de sécurité obligeant un attaquant ou un code malveillant à franchir plusieurs barrières pour arriver à son but pendant une attaque informatique. Cette multiplicité de barrières est à la fois dissuasive pour l'attaquant, diminue la probabilité de pouvoir exploiter une vulnérabilité sur chaque barrière et augmente la probabilité de détection de ces actes.

Fort de tous ces éléments, de tous ces mécanismes de sécurité, de leur intégration dans les

systèmes d'information et de la surveillance de ceux-ci, le CEA/DAM n'oublie pas que la sécurité informatique parfaite n'existe pas, que de nouvelles vulnérabilités et de nouveaux risques apparaissent chaque jour, et que la R&D dans le domaine de la cybersécurité, la veille technologique et l'intégration de plus en plus de mécanismes de sécurité à tous les niveaux sont obligatoires pour continuer d'assurer le meilleur niveau de sécurité possible pour les supercalculateurs et les centres de calcul qu'il exploite. ■

CHAPITRE 3



LA SAGA DES CALCULATEURS TERA

Le programme Simulation, en synergie avec les besoins de l'industrie et de la recherche, a impulsé la renaissance d'une lignée de supercalculateurs français, compétitive sur le plan mondial.

3.1 Les supercalculateurs, éléments clés de la dissuasion

— La décision d'arrêt complet des essais nucléaires et le maintien sur le long terme d'une capacité de dissuasion fiable et sûre ont fait entrer le CEA/DAM dans l'ère de la simulation. Son objectif est de reproduire par le calcul les différentes phases du fonctionnement d'une arme nucléaire. L'atteinte de cet objectif dépend donc de la disponibilité et de la puissance des calculateurs, car représenter de manière précise les différentes étapes du fonctionnement d'une arme nécessite des capacités de calcul que seuls les supercalculateurs les plus puissants peuvent produire.

Le supercalculateur, à l'instar du laser de haute énergie ou de la radiographie éclair, est ainsi devenu l'un des trois grands instruments du programme Simulation. Dans le calcul à haute performance (ou **HPC** en anglais), la plupart des développements technologiques nécessaires ne sont pas spécifiques de la simulation du fonctionnement d'une arme. Cependant le CEA/DAM a choisi d'être un acteur du domaine afin de pouvoir disposer librement, dans les coûts et les délais, des moyens de calcul adaptés à ses programmes. Cette démarche lui permet de plus de mieux anticiper les ruptures technologiques et les évolutions des architectures informatiques, et une meilleure maîtrise de la sécurité informatique. Grâce à la collaboration très tôt mise en place avec ATOS/BULL, basée sur le principe de la co-conception, le besoin dissuasion a conduit à la renaissance d'une industrie nationale de supercalculateurs. Cependant le CEA/DAM ne pouvait assurer à lui seul les débouchés suffisants à la soutenabilité économique de cette industrie. Celle-ci a été confortée par la prise en compte d'une quatrième contrainte, endogène : la compétitivité coût/performance des supercalculateurs du CEA/DAM, conçus comme devant être des pilotes d'une ligne de produits commercialisables dans de larges secteurs de l'industrie ou de la recherche. Cette

collaboration s'est donc, dès l'origine, inscrite dans un écosystème de partenaires industriels et académiques. Les travaux de co-conception ont toujours veillé à prendre en compte d'une part les standards ouverts et d'autre part les besoins de l'industrie, de la recherche et d'autres activités de la défense.

Cette démarche est encore valable aujourd'hui pour préparer le saut vers l'exascale. Mais elle doit être aménagée afin de prendre en compte de nouveaux enjeux techniques ou stratégiques. Sur le plan technique, il s'agit d'un changement d'échelle qui doit notamment prendre en compte, d'une part, la problématique des données massives et de leur exploitation, et, d'autre part, les évolutions des architectures des mémoires et des processeurs. Sur le plan stratégique, la technologie du HPC est de plus en plus considérée par les grands pays (États-Unis, Chine, Japon) comme une technologie souveraine dont il peut être tentant d'en limiter l'accès.

L'entreprise commune **EuroHPC**, mise en place par la Commission européenne et 32 États participants, vise elle aussi à développer les technologies et les composants d'une filière HPC indépendante. Les experts du CEA/DAM et de ses partenaires se sont préparés à inscrire leurs travaux dans ce nouveau paysage. ■



Jacques-Charles Lafoucrière, pionnier des supercalculateurs

Relever des défis technologiques, travailler avec des instruments de pointe, collaborer avec les meilleurs scientifiques du moment : tel est ce qui motive Jacques-Charles Lafoucrière, en charge du développement des nouvelles machines Tera.

— Jacques-Charles Lafoucrière a connu tous les supercalculateurs Tera. Arrivé au CEA/DAM en 1990 en tant que scientifique du contingent durant son service militaire, il a observé l'incroyable évolution de ces machines, d'abord en s'occupant des aspects de stockage, jusqu'à gérer les projets complets. Aujourd'hui, les équipes qu'il anime conduisent l'ensemble du projet, depuis la R&D jusqu'au démantèlement, en passant par la définition de l'architecture, les achats, l'installation, et l'exploitation.

Propriétaire ou open source

« La première machine Tera avait été achetée chez un constructeur américain, se souvient-il. La machine comme le logiciel étaient propriétaires et non **open source** : impossible de savoir ce qui était dedans. Lorsque nous avons eu des soucis avec cette machine, nous ne pouvions donc pas les résoudre nous-mêmes. C'est pourquoi, pour la machine suivante, nous avons décidé de prendre part à sa conception, en partenariat avec un industriel, Atos. »

Depuis lors, toutes les machines Tera ont été co-conçues avec Atos. Le CEA/DAM s'est donc doté d'un axe fort de développement de logiciel open source, en collaboration avec d'autres centres de calcul.

« Nous disposons ainsi de machines adaptées à nos besoins et utilisables par d'autres, souligne Jacques-Charles Lafoucrière. Mais seuls les logiciels du système sont ouverts, les codes de calcul pour nos missions scientifiques restent, eux, secrets. » Ces calculateurs développés conjointement sont adaptés à un marché très compétitif, afin qu'Atos puisse les commercialiser à l'international.

Têtes de série

Le CEA/DAM met en œuvre des têtes de série, avec tout ce que cela implique de difficultés, imprévus et réglages. « Les premiers instants sont toujours difficiles, il faut quelques mois pour que le calculateur soit utilisable de façon stable et simple. Il faut également gérer les retards des fournisseurs de technologies sur lesquelles s'appuie Atos (**processeurs**, mémoires). Ce n'est pas un long fleuve tranquille, car le dernier calculateur

utilise toujours les technologies les plus avancées, à la limite de ce qui est faisable à ce moment. C'est ça qui est intéressant ! » s'enthousiasme Jacques-Charles Lafoucrière.

Pas le temps de s'ennuyer, puisqu'il faut tout remettre à plat pour concevoir les nouveaux calculateurs tous les 4 à 5 ans. En effet, les besoins des utilisateurs sont croissants, et au bout de ce temps, il est plus intéressant de disposer d'une machine plus récente que d'améliorer l'ancienne, qui consomme plus d'électricité et coûte cher en maintenance.

Environnement stimulant

Jacques-Charles Lafoucrière ne regrette pas un instant son choix du CEA/DAM pour toute sa carrière, à l'heure où les jeunes sont plus enclins à changer de poste tous les 5 à 10 ans. « Ce qui est intéressant, c'est que nous travaillons sur des projets de longue durée, puisque aujourd'hui on réfléchit déjà à ce qu'on fera en 2030, voire 2040, explique-t-il. C'est autre chose que d'utiliser des technologies "sur étagère" ! On n'est pas sur du travail rapide et bâclé, on a le temps de construire de belles choses. Et on travaille sur des instruments exceptionnels : il existe moins de cinq centres comme le nôtre en Europe, et pas plus d'une dizaine dans le monde. Enfin, repartir sur une nouvelle machine tous les 5 ans crée une dynamique : les technologies informatiques bougent très vite, c'est très stimulant, motivant intellectuellement. Et collaborer avec les utilisateurs est particulièrement riche : nous travaillons avec le haut de gamme de la recherche, la science mise en œuvre est très pointue. »

Mais il ne faut pas imaginer qu'il s'agit d'un travail de solitaire, voire de « geek » asocial. C'est avant tout un travail en équipe, aussi bien au sein du CEA/DAM (notamment avec les utilisateurs des machines) qu'avec les industriels et les autres scientifiques, au sein de projets européens et internationaux. « Nos centres de calcul sont d'une telle complexité que personne n'est expert de tout, souligne-t-il. Résoudre un problème implique forcément que plusieurs personnes s'y attellent. C'est un mode de fonctionnement humainement passionnant. » Le recrutement est en phase : seules les personnes aimant travailler en équipe sont embauchées.

3.2 De Tera1 à Tera1000 - Plus de 15 ans de défis technologiques et de retombées industrielles

— Plus qu'un programme d'équipement, le programme Tera a été une véritable aventure industrielle menée conjointement par les équipes du CEA/DAM et celles de l'industriel Bull, aujourd'hui Atos/Bull.

C

onduit en quatre étapes, Tera1, Tera10, Tera100 et Tera1000, le projet Tera a été au cœur de la renaissance d'une industrie française et européenne des supercalculateurs. Et ce initialement dans un contexte technologique propice, toute phase de transition technologique favorisant en effet

l'émergence de nouveaux acteurs. Des challenges programmatiques et technologiques sans précédent ont conduit à la mise en place d'une approche innovante : la co-conception entre industriels et utilisateurs finaux des technologies aptes à répondre à leurs besoins. Cette méthodologie suppose une gestion cohérente et simultanée des projets de réalisation des technologies et des enjeux d'adaptation des codes ou processus de traitements qui devront en tirer parti. L'adéquation des technologies aux besoins est assurée par des interactions bidirectionnelles et itératives entre les projets de développements technologiques et ceux d'usages finaux. Les domaines en évolution rapide, comme les technologies de l'information (**HPC, IA...**), sont particulièrement bien adaptés à ce mode de développement. Il garantit d'une certaine manière à l'industriel de cibler durablement les besoins d'un marché en expansion et aux utilisateurs d'anticiper les évolutions technologiques. C'est ainsi que les technologies développées par Bull pour Tera10 ont rapidement trouvé un débouché pour des usages industriels et académiques, comme l'illustre le succès du **CCRT**. Il en va de même pour les innovations technologiques de Tera100 et Tera1000 qui

ont été mises en œuvre non seulement au **TGCC** pour des usages en recherche et industrie, mais aussi pour un large éventail de clients à l'international. C'est ainsi qu'Atos/Bull détient aujourd'hui, dans un contexte très concurrentiel, 8 % du marché mondial du HPC alors qu'aucun constructeur européen n'était présent sur ce marché au début des années 2000.

TERA1 (2001), LE PASSAGE AU PARALLÉLISME EN MÉMOIRE DISTRIBUÉE

Dans les années 1990, les supercalculateurs vectoriels sont arrivés en limite de technologie, car ils ne permettaient pas de paralléliser les calculs sur plusieurs machines dans des conditions économiques réalistes. De plus, les progrès de la technologie silicium n'étaient pas assez rapides pour répondre aux besoins des simulations de plus en plus exigeantes. Dans le même temps, les microprocesseurs développés pour des usages plus généraux (serveurs, stations de travail et même ordinateurs personnels) devenaient de plus en plus puissants en calcul scientifique. L'avènement de systèmes d'interconnexion performants entre ces **processeurs** a donc permis l'émergence d'une nouvelle architecture de supercalculateurs : les clusters de machines multiprocesseurs, associés à un nouveau modèle de programmation – le **parallélisme** en mémoire distribuée par échange de messages (voir chapitre « L'évolution des modèles de programmation parallèle »).

Après les premières expérimentations réalisées à partir de 1993 sur machines Cray T3D puis T3E, des clusters prototypes furent installés à la fin des années 1990 pour finir de valider le concept à échelle réduite pour une nouvelle génération de codes de calcul implémentant ce modèle de programmation.

Ces phases préparatoires ont permis d'aboutir en 2001 à l'installation de Tera1 (**figure 1**), calculateur composé de 640 nœuds quadri-processeurs Compaq Alpha-EV6 à 1 GHz disposant de 4 Go de mémoire. L'interconnexion était assurée par un réseau rapide développé par la société anglaise Quadrics. La puissance de calcul totale était de 5 **téraflops** et la capacité de stockage disque de 50 **téraoctets**. Elle fut classée 4^e machine mondiale (1^{re} européenne) en juin 2002 au classement mondial du **Top500**.

Cette machine a permis de poursuivre le développement et de valider une nouvelle génération de codes de simulation du CEA/DAM. Son exploitation au sein du centre de calcul et la gestion des flux de données issus des grandes simulations numériques parallèles ont tout particulièrement mis en évidence la nécessité de maîtriser le passage à l'échelle des logiciels régissant le fonctionnement du supercalculateur.

TERA10 (2005), LE CALCULATEUR DE LA TNA, OU L'APPROPRIATION DES TECHNOLOGIES

Au milieu des années 2000, la DAM devait relever un défi majeur : garantir par la Simulation, pour la première fois au monde, la sûreté, la fiabilité et la performance en énergie d'une arme nucléaire – la tête nucléaire aéroportée (**TNA**) destinée à équiper les missiles **ASMP-A** emportés par les avions de chasse Rafale et Mirage 2000N. Du point de vue de la puissance de calcul, il fallait gagner un ordre de grandeur par rapport à Tera1.

Dès le début des années 2000, le CEA/DAM s'était rapproché de l'industriel Bull pour l'inciter à reprendre une activité de conception d'ordinateurs scientifiques afin d'avoir au moins un compétiteur crédible face aux constructeurs américains, dont les machines avaient fait plusieurs fois l'objet d'embargos dans le passé. Un projet collaboratif (TeraNova) avait été mené en 2003-2004 avec pour objectif la réalisation d'une machine téraflopique. Le succès de cette opération permit à Bull de développer un produit commercial extrêmement général pouvant être utilisé à la fois pour le marché de la gestion et celui du calcul scientifique. Mais le plus important était probablement les compétences acquises qui mettaient Bull au niveau des plus grands industriels, lui permettant ainsi de répondre et de remporter l'appel d'offres Tera10.



FIGURE 1

Tera1, installé en 2001.



FIGURE 2

Tera10, installé en 2005.

Cette deuxième étape du programme Simulation a donc conduit en 2005 à l'installation d'un supercalculateur de 60 téraflops (**figure 2**), conçu par le constructeur Bull et fabriqué à Angers. L'architecture était basée sur 624 **nœuds de calcul** Novascale, chacun disposant de 8 processeurs Intel Itanium Montecito bi-cœur. L'interconnexion était réalisée par une nouvelle génération du réseau Quadrics. La majorité des nœuds de calcul disposait de 48 Go de mémoire, soit 3 Go par cœur. L'espace de stockage disque atteignait quant à lui le **pétaoctet** (le million de milliards d'octets).

D'importantes caractéristiques du système Tera10 reposent sur l'expérience acquise avec Tera1. Les problématiques de passage à l'échelle des logiciels propriétaires sur une machine exploitée en environnement classifié de défense ont conduit à favoriser le choix de solutions **open source** pour la pile logicielle et à une forte implication des équipes de la DAM dans ces communautés. Les limitations constatées sur les entrées/sorties avec Tera1 ont également conduit à renforcer cette exigence sur Tera10. Ainsi, à la date de son installation, Tera10 était le plus puissant supercalculateur utilisant le système d'exploitation ouvert **Linux** et disposait du débit d'entrées/sorties vers les systèmes de stockage de données le plus élevé au monde (100 Go/s).

Forts de cette réalisation, le CEA et la société Bull décidèrent de proposer la thématique « supercalculateur » comme l'un des axes de développement du pôle de compétitivité **SYSTEM@TIC**, alors en gestation. Le projet FAME2 (Flexible Architecture for Multiple Environment), élaboré pour développer une nouvelle unité de calcul en vue d'atteindre le **pétaflops**, devint ainsi l'un des projets fondateurs de SYSTEM@TIC.

TERA100 (2010), LE CALCULATEUR DE LA TNO, OU LA CONFIRMATION DU SUCCÈS

Au même titre que Tera10, Tera100 (figure 3) était dimensionné pour offrir la garantie sans essai d'une nouvelle charge nucléaire, en l'occurrence la tête nucléaire océanique (**TNO**), destinée à être embarquée sur le missile M51 dans les **SNLE** de type Le Triomphant. L'évaluation de

la puissance de calcul nécessaire conduisit à la mise en place dès 2010 d'une machine au-delà du pétaflops, performance en limite de technologie compte tenu du calendrier du projet. La gestion des risques conduisit donc à mettre en œuvre et à piloter une phase de R&D avec un industriel afin de sécuriser la disponibilité des technologies matérielles et logicielles nécessaires à l'horizon 2010. Grâce à l'expérience acquise, Bull proposa la meilleure offre de R&D partenariale avec le CEA/DAM qui permettait de développer en deux ans les technologies nécessaires à une machine pétaflopique de production. À l'issue de cette phase essentielle, le CEA/DAM fut en mesure de lever l'option que constituait l'acquisition du calculateur Tera100.

Conformément au planning, Bull acheva le développement en 2010 et installa au CEA/DAM le troisième calculateur du programme Simulation. Celui-ci reposait sur 4 300 serveurs quadri-processeurs Intel Xeon Nehalem octo-cœurs à 2,27 GHz embarquant 64 Go de mémoire (soit 2 Go/cœur). L'interconnexion était assurée par un réseau InfiniBand développé par la société américaine Mellanox. Sa puissance de calcul était de 1250 téraflops et sa capacité de stockage disque de 23 pétaoctets. En novembre 2010, elle fut classée 6^e machine mondiale (avec la meilleure efficacité du Top10 sur le test du **Linpack**) et 1^{er} supercalculateur européen à dépasser le pétaflops, confirmant ainsi le savoir-faire en HPC de Bull et du CEA/DAM sur le plan mondial.

Tera100 a donné naissance à une véritable ligne de produits compétitive, positionnant son constructeur comme une alternative européenne crédible à ses concurrents américains, chinois et japonais. Ce calculateur a clos une première phase du programme Simulation, dont un objectif majeur était de garantir, sans essai nucléaire nouveau, le renouvellement des deux composantes de la dissuasion nucléaire française (voir chapitre « Le programme Simulation et ses enjeux »).

TERA1000 (2015-2017), UN TREMPLIN VERS L'EXAFLOPS

À l'horizon des années 2020, la deuxième phase du programme Simulation exige des outils de simulation encore plus précis et donc la mise en œuvre de calculateurs des centaines de fois plus puissants que ceux de la génération Tera100, plus communément appelés « de classe **exaflopique** ». Pour rester dans des conditions économiques acceptables, en particulier en ce qui concerne la consommation électrique, de nouvelles ruptures étaient indispensables vis-à-vis de l'architecture du calculateur et des méthodes numériques à mettre en œuvre dans

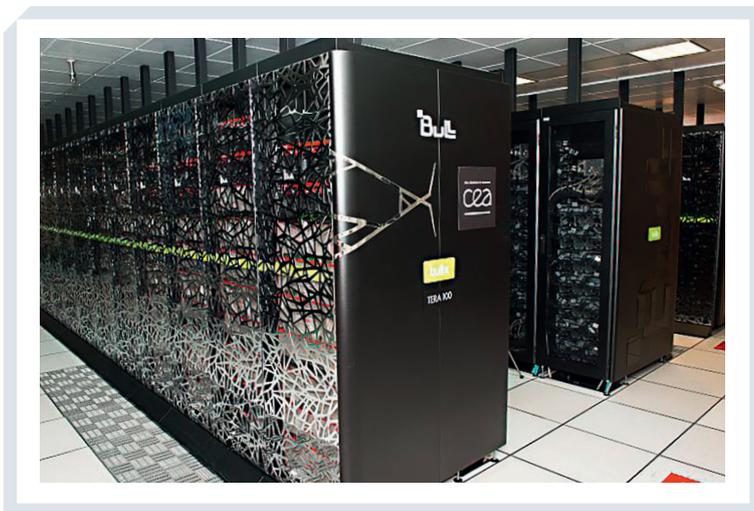


FIGURE 3

Tera100, installé en 2010.

les codes. Ainsi, l'explosion du nombre de **cœurs de calcul** par nœud (et donc du nombre total de cœurs de la machine) conduit à compléter le parallélisme en mémoire distribuée par du parallélisme en mémoire partagée au niveau du nœud de calcul. De même, l'apparition d'instructions massivement **SIMD** ne peut être exploitée qu'avec des structures de données adaptées (un peu comme sur les anciennes machines vectorielles des années 80-90). Enfin, tout ce qui concerne l'accès et le déplacement des données doit être optimisé avec, en particulier, la prise en compte de mémoires rapides connectées directement au processeur.

Il n'était donc pas envisageable d'aller directement à l'**exascale** sans aménager une phase intermédiaire permettant d'anticiper et de valider ces ruptures sur un prototype représentatif. C'est précisément une des fonctions de Tera1000, et tout particulièrement de la partition Tera1000-2 mise en service fin 2017. La première partition TERA1000-1 (**figure 4**, photo du haut), mise en service en 2015, assure quant à elle la continuité de la production des codes de calcul existants, en offrant un accroissement significatif des capacités de calcul accessibles aux concepteurs du CEA/DAM.

Tera1000-1 est en quelque sorte un « rafraîchissement technologique » de Tera100 qui permet une performance double avec un budget énergétique divisé par cinq. Son mode d'utilisation reste entièrement compatible avec les machines précédentes, permettant ainsi une utilisation efficace des codes sans adaptation substantielle. Tera1000-1 dispose en outre d'une sous-partition comportant 400 **GPU**. Ceux-ci sont utilisés d'une part pour visualiser les résultats des simulations (voir chapitre « Traiter les données et visualiser les résultats »), et d'autre part pour certains calculs intensifs, notamment des calculs matriciels (voir chapitre « Simulation de phénomènes électromagnétiques sur des supercalculateurs parallèles »).

Tera1000-2 (**figure 4**, photo du bas) est beaucoup plus novatrice puisqu'elle est constituée de 8 004 nœuds de calcul comprenant chacun un processeur Intel/**KNL** doté de 68 cœurs à 1,4 GHz et 192 Go de mémoire (soit près de 3 Go/cœur). Les processeurs embarquent également 16 Go de mémoire rapide pour alimenter plus efficacement leurs unités fonctionnelles. Enfin, le réseau d'interconnexion marque le retour à une technologie européenne puisqu'il s'agit du **BXI** conçu et développé par Atos/Bull dans le cadre de son programme de R&D avec le CEA/DAM. L'ensemble de la machine développe une performance crête de 25 pétaflops, soit 20 fois la puissance de Tera100 avec une consommation électrique en baisse de 20 %. La capacité de stockage disque totale atteint les 40 pétaoctets.



FIGURE 4

Tera1000-1 installé en 2015 (photo du haut). Tera1000-2 installé en 2017 (photo du bas).

Conçu en partenariat entre la DAM et Atos/Bull, Tera1000 initie chez l'industriel une nouvelle gamme de machines « pré-exascale » mondialement plébiscitée par de nombreux clients. Les travaux de développement et d'optimisation de codes, démarrés dès 2015 sur des « mini-applications », se sont poursuivis sur les applications utilisées en production, démontrant une meilleure efficacité énergétique sur Tera1000-2 que sur Tera1000-1.

	Tera1 (2001)	Tera10 (2005)	Tera100 (2010)	Tera1000-1 (2016)	Tera1000-2 (2017)
Architecture	640 serveurs quadri-processeurs	624 serveurs octo-processeurs	4 300 serveurs quadri-processeurs	2 196 serveurs biprocesseurs	8 000 serveurs monoprocesseur
Processeurs	1 GHz mono-cœur	1,6 GHz bi-cœurs	2,27 GHz octo-cœurs	2,3 GHz 16 cœurs	1,4 GHz 68 cœurs
Cœurs	2 500	10 000	140 000	70 000	500 000
Calcul	5 Tflops	60 Tflops	1 250 Tflops	2 500 Tflops	25 000 Tflops
Capacité et bande passante mémoire	2,5 To 5,12 To/s	30 To 16 To/s	300 To 430 To/s	300 To 300 To/s	DRAM : 1 540 To 920 To/s MCDRAM : 128 To 3 800 To/s
Capacité et bande passante disques	0,05 Po 7,5 Go/s	1 Po 100 Go/s	23 Po 500 Go/s	30 Po 500 Go/s	40 Po 800 Go/s
Consommation	0,65 MW	1,3 MW	5 MW	1 MW	4 MW

FIGURE 5

Principales caractéristiques des calculateurs Tera.

Ces résultats confortent les orientations architecturales vers des supercalculateurs hétérogènes et modulaires qui sont anticipées pour les machines exascale qui seront exploitées par le CEA/DAM dans la prochaine décennie. Le premier système de cette nouvelle génération est le calculateur Exa1, dont la première partition est mise en service en 2021.

PASSER À L'ÉCHELLE ET MAÎTRISER LA CONSOMMATION

Pour atteindre les performances visées, la complexité des systèmes est en constante augmentation (nombre de serveurs, nombre de cœurs de calcul...), ce qui a imposé aux couches logicielles et aux applications de prendre en compte ce passage à l'échelle massif.

Même si la croissance de la capacité mémoire a sensiblement suivi celle de la puissance de calcul (figure 5), la bande passante mémoire

(c'est-à-dire la capacité d'alimenter les processeurs avec des données externes) n'a pas suivi la progression puisqu'on est passé de plus de 1 **Byte/Flop** sur Tera1 à 0,037 **Byte/Flop** sur la mémoire **DRAM** de Tera1000-2. Pour tempérer cette évolution, le recours à de la mémoire rapide devient indispensable : sur Tera1000-2, cela permet de « remonter » à 0,15 **Byte/Flop** avec la **MCDRAM**.

La maîtrise de la consommation électrique (et donc des coûts de possession de ces supercalculateurs) a été obtenue grâce à une augmentation phénoménale de l'efficacité énergétique des systèmes : amélioration d'un facteur 800 entre Tera1 et Tera1000-2 (voir chapitre « Le challenge énergétique »).

L'amélioration de cette efficacité reste un axe majeur de R&D pour les futurs supercalculateurs nécessaires aux programmes du CEA/DAM ainsi qu'à de nombreux usages scientifiques et industriels.

3.2.1 Co-conception avec l'industriel : les grandes thématiques et l'organisation de la collaboration

Au-delà de l'acquisition de supercalculateurs nécessaires au programme Simulation, le projet Tera a pris en compte et assuré une partie importante du financement de la R&D nécessaire au développement de la base technologique et industrielle permettant la maîtrise des briques technologiques des supercalculateurs et leur conception. Au début des années 2000, le CEA/DAM et la société Bull (aujourd'hui Atos/Bull) ont initié la collaboration en donnant naissance à une plateforme térafloppique prototype, marquant le retour de l'industriel dans le domaine du calcul scientifique. C'est sur ces fondations que les équipes de Bull et du CEA/DAM bâtirent le succès de Tera10.

En 2006, le CEA/DAM et la société Bull furent à l'initiative de la prise en compte de la thématique « supercalculateur » comme axe de développement du pôle de compétitivité SYSTEM@TIC, officialisant et renforçant ainsi un partenariat en vue de consolider une filière commerciale pérenne de supercalculateurs.

Par la suite, la collaboration s'est encore accrue et a été structurée sous la forme d'accords-cadres pluriannuels définissant les grands axes de R&D. Un premier protocole d'accord signé le 30 juillet 2007 couvrait la période 2007 à 2010. Le 6 juillet 2011, un second accord-cadre est établi pour couvrir une nouvelle période de 4 ans.

Le 19 septembre 2014, l'État confirme son soutien à cette démarche ; le cadre de la collaboration entre les équipes du CEA et celles d'Atos/Bull est fixé par la convention entre l'État et le Commissariat à l'énergie atomique et aux énergies alternatives relative au programme d'investissements d'avenir (action Calcul intensif). Dans ce cadre, l'État demande au CEA « de mobiliser son expertise

pour mettre en place un programme de recherche et développement en vue de disposer à l'horizon 2020 de la capacité de concevoir et réaliser des ordinateurs de grande puissance de manière durablement compétitive¹ ».

Ce programme de recherche conduit par les équipes du CEA/DAM et d'Atos/Bull combine, d'une part, le savoir-faire des équipes de la DAM dans les domaines applicatifs, le **middleware** et la conception et l'exploitation de grands centres de calcul et, d'autre part, l'expertise des équipes d'Atos/Bull sur les technologies de base, la conception de grands systèmes et l'industrialisation de solutions pour le calcul haute performance.

La phase en cours définit les recherches à conduire pour concevoir et développer les systèmes futurs. Elle s'articule en cinq axes techniques :

- ⊙ maîtrise de l'architecture du processeur et du nœud de calcul,
- ⊙ réseau d'interconnexion et intégration,
- ⊙ gestion des données,
- ⊙ logiciel pour le contrôle et le suivi des ressources et de l'énergie,
- ⊙ modèles de développement et d'exécution.

Ces cinq axes recouvrent les enjeux technologiques essentiels pour maîtriser les architectures exascale :

- ⊙ **l'efficacité énergétique** : avec l'objectif d'atteindre un ratio de Gflops/W conduisant à des supercalculateurs de classe exafloppique consommant moins de 20 MW ;
- ⊙ **la gestion des données** : pour accélérer l'accès aux données pour le calcul et l'exploitation des

1. Cf. Journal officiel du 21 septembre 2014 <http://www.legifrance.gouv.fr/eli/convention/2014/9/19/PRM1419351X/fo>



INTERVIEW

MATTHIEU HAUTREUX

« Je travaille aujourd'hui sur la définition des prochains calculateurs.

Il existe de plus en plus de calculateurs haute performance, note-t-il. Nous devons avoir une contribution intéressante et pertinente. Nous avons une vraie différence : nous comprenons les utilisateurs et leurs problématiques... »

résultats tout en minimisant les mouvements de données coûteux en énergie ;

⊙ **la maîtrise du parallélisme** : les architectures massivement parallèles posent des défis aussi bien pour l'exploitation de ce parallélisme matériel par les applications que pour la gestion de l'ensemble de ces ressources par les logiciels de contrôle du calculateur ;

⊙ **les nouveaux modèles d'usage** : avec la **convergence** entre le HPC et le Big Data (ou l'IA).

Les experts du CEA/DAM et les équipes de R&D d'Atos/Bull collaborent en mode agile sur les sujets suivants :

- ⊙ nœuds de calcul pour les grands systèmes,
- ⊙ stockage,
- ⊙ réseaux,
- ⊙ logiciel cluster,
- ⊙ intégration en salle machine,
- ⊙ outils de développement,
- ⊙ sécurité,
- ⊙ intégration des composants et processus de validation transverse.

Ce partenariat a démontré son efficacité à travers le développement des gammes « pétascale » et « pré-exascale » d'Atos/Bull. Le programme en cours doit conduire à une gamme « exascale » disponible pour les utilisateurs à l'horizon 2022.

3.2.2

Le réseau d'interconnexion - élément clé d'un supercalculateur parallèle

Un supercalculateur est obtenu par la fédération d'un très grand nombre de serveurs (ou nœuds), comportant un ou plusieurs processeurs, qui coopèrent *via* un réseau d'interconnexion. Tous les nœuds du supercalculateur, nœuds de calcul, de stockage, d'administration, etc., sont connectés à ce réseau qui constitue les artères par lesquelles

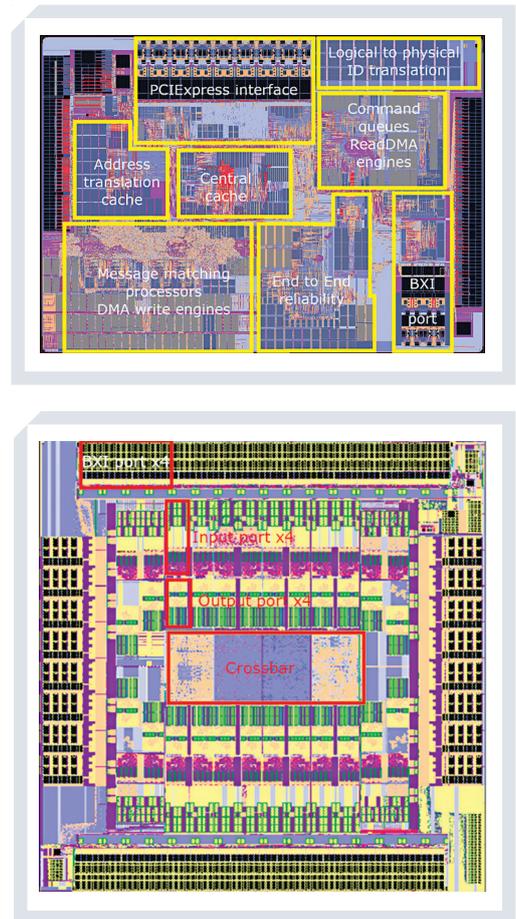


FIGURE 1

Description des deux circuits « switch » et « NIC » du réseau d'interconnexion BXI.

vont passer toutes les données d'une simulation. L'augmentation de la taille des calculs et du nombre de processeurs utilisés induit en général une augmentation des phases d'échanges et du nombre de messages qui y transitent. Ce réseau est donc un élément primordial de l'architecture et de la performance du supercalculateur, dont il peut représenter jusqu'à 30 % du coût. Il doit être tolérant aux pannes de ses constituants, en assurant prioritairement la conservation de la connectivité, quitte à diminuer son débit par une gestion plus importante.

La topologie, c'est-à-dire la forme du graphe des connexions, décrit les liaisons matérielles entre les nœuds. Elle détermine la manière dont ceux-ci sont fédérés et a une grande influence sur les performances. En théorie, la configuration optimale consisterait à relier directement tous les nœuds un à un. Elle est de fait irréaliste, car beaucoup trop coûteuse pour des supercalculateurs comportant des milliers de nœuds. La topologie du réseau est donc le résultat d'un compromis coût/performance. Elle détermine alors la dis-

tance entre les couples de nœuds (c.-à-d. la longueur du chemin que l'information doit parcourir entre un nœud émetteur et un nœud récepteur), le diamètre du réseau, et induit une notion de localité dans l'architecture du supercalculateur. Paramètre important, le routage décrit la route à emprunter d'un émetteur vers un récepteur. La gestion du réseau doit donc définir des routes pour aller d'un nœud à un autre, identifier les liens et nœuds en panne et, dans ce cas, trouver des routes alternatives.

D'autres paramètres sont aussi à prendre en compte dans la caractérisation d'un réseau d'interconnexion : le débit de communication (ou bande passante du réseau) et la **latence** (ou durée minimale pour établir une communication). Vu du nœud de calcul, la capacité à émettre ou recevoir plusieurs millions de messages par seconde est un critère primordial. En effet, ce taux d'émission/réception doit être en relation avec la vitesse de calcul des processeurs.

La conception et la réalisation d'un réseau d'interconnexion pour supercalculateurs étaient jusqu'à il y a quelques années maîtrisées par très peu d'acteurs, localisés aux États-Unis, au Japon et en Chine. Le fait de disposer d'une capacité équivalente en France est donc une avancée considérable dans la maîtrise des technologies critiques pour le développement des supercalculateurs, et constitue un élément majeur de la stratégie de souveraineté inhérente à la mission de dissuasion, en permettant à la France de disposer d'une solution indépendante sur un élément clé de l'architecture des supercalculateurs. Le programme de recherche-développement conjoint Atos-CEA/DAM a conduit à la réalisation d'une nouvelle génération de réseau dénommé Bull eXascale Interconnect (BXI), dont la première version est opérationnelle depuis 2017 dans le supercalculateur Tera1000.

Le réseau BXI s'appuie sur deux circuits intégrés spécifiques (**ASIC**) qui ont été spécialement conçus et réalisés à cette fin (**figure 1**) :

- ⊙ un contrôleur d'interface réseau (ou **NIC** – Network Interface Controller) qui gère l'émission et la réception des messages du côté du nœud de calcul ;
- ⊙ un commutateur (ou switch) qui a la charge d'aiguiller les messages au cours de leur progression dans le réseau.

BXI dispose d'un ensemble complet de logiciels de gestion et de communication. Il a été spécifiquement conçu pour les plus grands supercalculateurs en permettant de connecter jusqu'à 64 000 nœuds.

Cette brique technologique sera primordiale pour le développement des futurs systèmes exaflopiques.

3.2.3

Au cœur de Tera1000-2 : le processeur Intel KNL

Le processeur Knights Landing, ou KNL, commercialisé par la société Intel en 2016, est la deuxième génération de processeurs « **manycore** » (« pluricœurs » en français) de type Xeon Phi. Avec un très grand nombre de **cœurs** (jusqu'à 72) et une nouvelle architecture de la mémoire par rapport au processeur Xeon classique, ce processeur a pour objectif de fournir le meilleur ratio performance/watt pour le calcul scientifique. Pour réaliser cet objectif ambitieux, les ingénieurs d'Intel ont choisi :

- ⊙ D'utiliser un grand nombre de cœurs. La version de KNL utilisée dans Tera1000-2 possède 68 cœurs utilisables. Chaque cœur, dont la fréquence de fonctionnement nominale est de 1,4 GHz, peut faire fonctionner deux unités vectorielles¹ (appelées **VPU** pour « Vector Processing Unit » sur la **figure 1**). L'essentiel de la puissance de calcul est concentré dans ces deux unités.
- ⊙ D'insérer à proximité des cœurs de calcul des mémoires à très grande bande passante (MCDRAM sur la photo), capables de fournir des débits plus de 4 fois supérieurs (~400 Go/s) à ceux de la mémoire classique de type **DDR** (~90 Go/s).

Réalisé avec une finesse de gravure de 14 nm, ce processeur utilise plus de 8 milliards de transistors.

1_ Une unité vectorielle du processeur Intel KNL effectue ses calculs sur des groupes (vecteurs) de 8 nombres en virgule flottante double précision (ou 16 nombres flottants en simple précision) en une seule instruction.

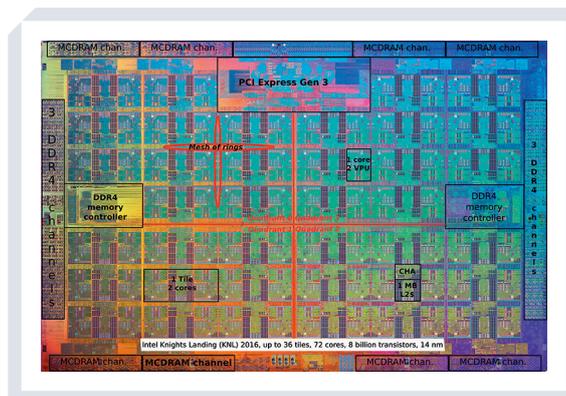


FIGURE 1

Photo du processeur Intel KNL (crédit Intel) – le processeur de Tera1000-2.

Pierre Barnabé, directeur général adjoint d'Atos



— Quelles sont vos fonctions au sein d'Atos et quel est le périmètre de la division que vous dirigez ?

Je suis en charge de la division Big Data et Security du groupe Atos. Cette division BDS en forte croissance, qui réalise plus de 1 milliard d'euros de chiffre d'affaires, regroupe les produits et technologies répondant aux défis critiques de la confiance et de la souveraineté numérique : le contrôle et la valorisation des données, l'intelligence artificielle, les besoins toujours plus importants en puissance de calcul pour la modélisation, et la cybersécurité.

— Quelle est la place du développement des supercalculateurs au sein de BDS et comment voyez-vous leur évolution ?

Le développement de la division BDS se concentre sur la puissance de calcul et la maîtrise des algorithmes, notamment en cryptographie. Les supercalculateurs et le calcul intensif ont été au centre de notre histoire et de notre activité. C'est aussi un élément crucial de l'indépendance numérique française. Un savoir-faire sur lequel nous avons pu bâtir notre avenir, l'accès et l'usage de la puissance de calcul s'étant « démocratisés ». Sur cette base, nous avons construit un portefeuille de produits très performants incluant les serveurs d'entreprise haut de gamme de calcul « in memory » qui équipent les plateformes des principaux fournisseurs mondiaux de « cloud computing » avec, depuis l'année 2019, une déclinaison dans le « Edge Computing », c'est-à-dire le calcul en périphérie de réseaux.

— La compétition internationale est très vive entre les États-Unis, la Chine et d'autres pays comme le Japon, comment voyez-vous la place de l'Europe dans cette compétition ?

Au-delà d'être des objets à très haute valeur technologique, les supercalculateurs sont pour nos clients aussi emblématiques et stratégiques que de grands projets d'infrastructure, de maîtrise des

énergies ou des développements dans le domaine de la défense. Dans ce cadre, paradoxalement, la surpuissance américaine et chinoise nous a servi et continue de nous servir.

Cette course à la puissance « pure » engagée par les États-Unis et la Chine demande que les industriels soient sursubventionnés pour créer des machines incroyablement puissantes et conçues pour des applications très spécifiques... mais qui sont peu utiles en dehors de ce pour quoi elles ont été conçues.

De notre côté, nous sommes une société industrielle et nous concevons des produits qui sont multi-usages, modulaires, qui prennent en compte les exigences de développement durable. Tout en restant dans la nécessaire course à l'excellence technologique et la puissance, nos machines sont plus « universelles ». Je suis très attaché à cette modularité et cette approche par le besoin client. Nous avons ainsi développé un actif technologique dont la valeur s'est considérablement accrue depuis dix ans. Un très grand nombre de puissances économiques et industrielles dans le monde souhaitent pouvoir utiliser des technologies neutres par rapport aux superpuissances américaine et chinoise. À ce titre, un constructeur européen apparaît comme un fournisseur « non aligné ». Nous offrons la seule alternative... si ce n'est le Japon qui présente une situation insulaire très singulière.

— Cette compétition et l'évolution permanente des technologies requièrent une recherche-développement importante. Pouvez-vous nous donner quelques

informations sur votre activité en recherche-développement, les principales thématiques et la localisation des équipes ?

La R&D d'Atos est constituée d'une équipe pluridisciplinaire principalement composée d'ingénieurs et de docteurs. Elle a par exemple développé des savoir-faire très complets :

- ⊙ en développement logiciel pour : la cybersécurité, les systèmes critiques, l'Internet des objets, le calcul parallèle et l'intelligence artificielle... ;
- ⊙ en conception de systèmes matériels complexes et de circuits intégrés spécifiques (ASIC) : supercalculateurs, cartes électroniques, stockages hiérarchiques de données, circuits pour l'interconnexion ou la cohérence de cache ;
- ⊙ ainsi qu'en informatique quantique (algorithmique, simulation) où nous sommes devenus un acteur de référence mondiale...

L'équipe R&D est impliquée dans de grands programmes d'innovations mondiaux ou européens. En outre, elle collabore avec les grands clients du groupe Atos dans le cadre de projets de co-développement, évidemment avec le CEA/DAM pour les supercalculateurs, ou avec Siemens sur l'intelligence artificielle, l'analyse de données et la conception de serveurs compacts pour l'« Edge ». Pour accélérer notre avancée sur ces nouveaux usages et ces programmes, Atos a entamé une transformation de sa R&D autour de trois axes :

- ⊙ l'automatisation : à fin 2019, 95 % des tests de **non-régression** sont automatisés et permettent aux équipes de test et validation de se focaliser sur des tâches à plus forte valeur. Cela permet de réduire le cycle de développement global des solutions et de le rendre plus prédictible ;
- ⊙ la conception du logiciel, en redéveloppant en profondeur plusieurs blocs logiciels soit pour basculer dans une conception modulaire (microservices et **API**), soit pour permettre le support du multi-tenant (« as a service ») dans ses solutions d'identification et d'authentification et dans ses produits de supercalcul ;
- ⊙ l'intelligence artificielle, avec plus de 50 ingénieurs et docteurs spécialisés, organisés

autour d'un laboratoire de R&D dédié implanté à Échirolles (France). Nos travaux en IA sont plus particulièrement appliqués aux développements de produits en cybersécurité, dans le domaine de l'analyse d'images, ainsi qu'à l'optimisation du fonctionnement des infrastructures de calcul. La DAM apporte des compétences techniques et humaines extrêmement précieuses au développement d'Atos, et particulièrement pour la conception du supercalculateur BullSequana X, qui est devenu un best-seller mondial. C'est un partenaire à la fois bienveillant et exigeant, qui, depuis près de vingt ans maintenant, nous a toujours tirés vers le haut, chacun restant dans son domaine de compétence. Je vous prends mon billet que ce partenariat existera toujours dans vingt ans !

— Pour les dix prochaines années, quels semblent être pour vous les défis à relever et comment la Direction des applications militaires pourrait-elle vous accompagner pour les relever ?

J'identifie trois grands défis à relever dans les années à venir, et notre relation avec la DAM est indispensable pour tenir la dragée haute à nos concurrents américains et asiatiques. Tout d'abord, pour faire face aux besoins toujours plus gargantuesques de calcul et de traitement de données, nous devons travailler sur des machines toujours plus puissantes, qui vont devoir s'hybrider avec des accélérateurs quantiques, intégrant des processeurs européens et des jeux d'instructions toujours plus complexes. Ensuite, nous devons arriver à maintenir, sinon à réduire notre empreinte CO₂ et la consommation en énergie face à cette explosion des données. Dans ce domaine, Atos est très clairement en pointe, largement grâce à la DAM. Enfin, nous devons transformer notre puissance en un mode « as a service ». Nous devons être particulièrement attentifs à cette nouvelle manière de vendre de l'énergie de calcul, dérivée du « cloud computing », car elle sera à l'origine de beaucoup des innovations de ces prochaines années.

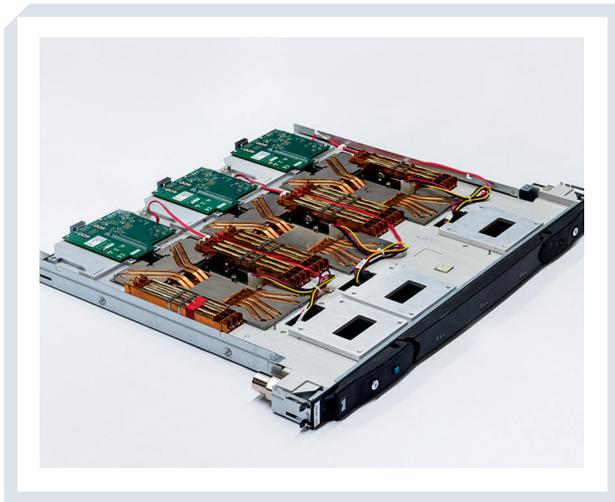


FIGURE 1

Lame de calcul regroupant 3 nœuds de calcul, chacun comportant un processeur Intel KNL et un ASIC d'interface avec le réseau d'interconnexion (NIC).

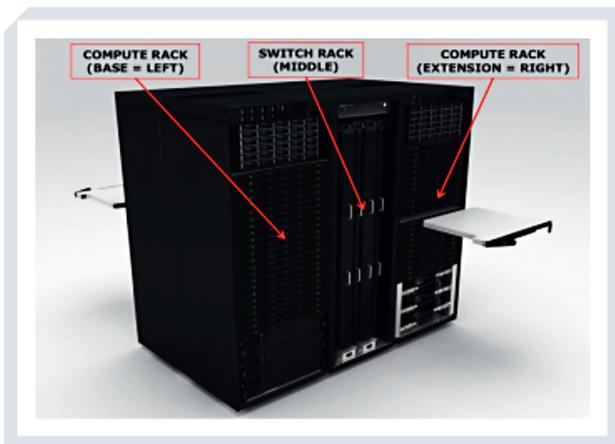


FIGURE 2

Plateforme Sequana.

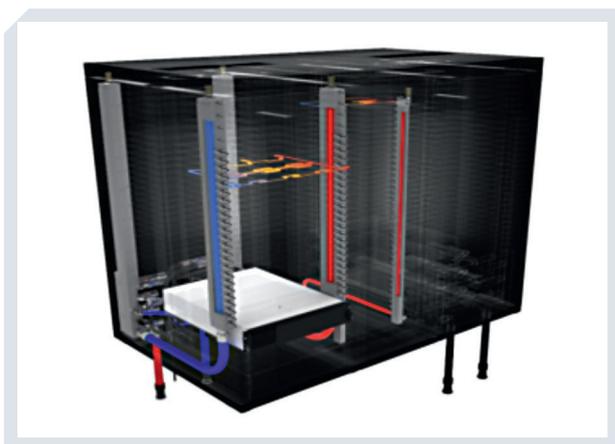


FIGURE 3

Système de refroidissement.

3.2.4 Tera1000-2, un concentré d'innovations

Tera1000-2 rassemble plus de 8 000 processeurs Intel KNL (voir article « Au cœur de Tera1000-2: le processeur Intel KNL »), pour environ un demi-million de cœurs de calcul, reliés par un réseau d'interconnexion BXI (voir article « Le réseau d'interconnexion – élément clé d'un supercalculateur parallèle ») avec une topologie basée sur un concept d'îlots pouvant fonctionner de manière indépendante.

Trois nœuds de calcul sont regroupés sur une même carte (figure 1), avec son refroidissement intégré. La plateforme Sequana, conçue et réalisée dans le cadre du projet Tera1000, permet d'accueillir 96 cartes, soit 288 nœuds de calcul, leur réseau BXI et leur refroidissement (figures 2 et 3).

Des logiciels optimisent la consommation associée aux différentes activités réalisées par le supercalculateur, soit par adaptation de la répartition sur les nœuds de calcul, soit par réglage de la fréquence des processeurs.

Les briques logicielles indispensables au fonctionnement et à l'exploitation du supercalculateur ont été adaptées ou optimisées. Tout particulièrement :

- ⊙ le gestionnaire et l'allocateur de ressources ;
- ⊙ la couche de communication **MPI**, adaptée pour le réseau BXI ;
- ⊙ le système de fichiers parallèles pour le partage des données ;
- ⊙ des mécanismes de tolérance aux pannes.

Ainsi, le CEA/DAM et Atos/Bull ont développé une « pile logicielle » optimisée, offrant un démarrage complet du supercalculateur en moins d'une heure, une gestion efficace des différentes versions du système d'exploitation, la possibilité de lancer plus de 1000 travaux par seconde, des systèmes de fichiers cohérents (**Lustre**) avec un débit global de plus de 600 Go/s...

Cette stratégie a conduit à une évolution majeure du métier des administrateurs de la machine : le système d'exploitation n'étant plus une « boîte noire » (c.-à-d. un système non modifiable), les équipes peuvent maintenant directement analyser les problèmes, créer des correctifs et adapter le système au profil de charge du supercalculateur. ■

3.3 Le challenge énergétique

— La consommation électrique représente un poste important du budget de fonctionnement d'un centre de calcul. Au cours du temps la densité énergétique (kW/armoire) et la consommation électrique des supercalculateurs ont fortement augmenté, nécessitant la mise en place d'une démarche d'optimisation énergétique au sein du centre de calcul.

A

vec plus de 5 MW de puissance électrique, l'arrivée du supercalculateur Tera100 a été le déclencheur de la mise en place d'une démarche d'optimisation énergétique des centres de calcul sous responsabilité des équipes du CEA/DAM. En particulier, le challenge était de multiplier par 20 la puissance de calcul de Tera10 tout en restant dans une enveloppe énergétique raisonnable et dans le contexte du ralentissement de la **loi de Moore**. Cette démarche a été reconnue à l'échelon européen par les promoteurs du « Code of Conduct on Data Centre Energy Efficiency ».

LA DÉMARCHÉ D'OPTIMISATION ÉNERGÉTIQUE

Depuis 2009, le CEA/DAM a ainsi mis en place une démarche d'optimisation énergétique au sein du complexe de calcul situé à Bruyères-le-Châtel afin de réduire les consommations électriques. Elle s'est traduite par l'installation de l'outil MAPE (Maîtrise et Analyse de la Performance Énergétique) développé par EDF et dédié à l'acquisition et au traitement des données énergétiques (**figure 1**). La finalité de cette démarche est l'identification de gisements d'économie d'énergie et le ciblage des actions d'amélioration de la performance énergétique. Les différentes composantes analysées sur l'installation sont listées ci-dessous :

- ⊙ les consommations électriques des équipements informatiques ;
- ⊙ les consommations électriques des servitudes (refroidissement, distribution de l'énergie) ;

- ⊙ les paramètres de fonctionnement du système de refroidissement (température et débit d'eau, cascade des groupes frigorifiques, vitesse de rotation des ventilateurs des tours aéroréfrigérantes...);
- ⊙ les mesures d'ambiance (température et hygrométrie).

Le suivi et l'analyse de ces données permettent d'identifier les actions à mener, de tracer l'impact des améliorations mises en place et de quantifier les gains obtenus. Parmi les actions réalisées qui ont amélioré la performance énergétique des centres de calcul du CEA/DAM, on peut citer :

- ⊙ la diminution des débits d'eau dans les circuits de refroidissement ;
- ⊙ l'augmentation de la température dans les salles informatiques ;
- ⊙ l'amélioration du comportement du système de production de l'eau de refroidissement en fonction des variations de charge des calculateurs ;
- ⊙ la maîtrise de l'intégration de nouvelles solutions techniques plus efficaces (les ultra-capacités pour le stockage d'énergie électrique en cas de microcoupure et les portes froides puis la technologie Direct Liquid Cooling pour le refroidissement des composants électroniques).

Les optimisations énergétiques ainsi réalisées ont également permis au CEA/DAM d'être labellisé par la Commission européenne dans le cadre du « **European Code of Conduct for Data Centres** ». Ce code de conduite a pour objectif de proposer des bonnes pratiques afin de maîtriser

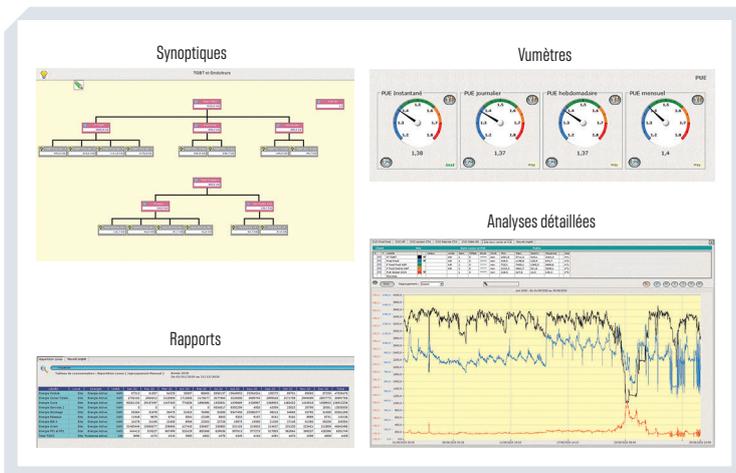


FIGURE 1
Outil de supervision énergétique.

gain réalisé sur Tera100 est dû à l'utilisation de portes froides (**figure 2**) pour le refroidissement des armoires de calcul et à l'utilisation d'ultra-capacité (**figure 3**) pour remplacer les onduleurs. Le gain réalisé sur Tera1000 est dû à la mise en œuvre de la technique DLC (Direct Liquid Cooling) (**figure 4**) qui consiste à amener le refroidissement par eau « tiède » au plus proche des composants électroniques du serveur en utilisant des plaques d'échange thermique (dispositif « Cold Plate »). En effet, les processeurs pouvant fonctionner à température élevée, l'utilisation d'une eau de refroidissement à une température de 35°C permet de supprimer les groupes mécaniques de production d'eau glacée qui sont très énergivores.

et d'optimiser la consommation énergétique des centres de calcul. Le centre de calcul du CEA/DAM a été labellisé en 2012 et a obtenu un prix en 2013 récompensant la meilleure démarche énergétique des nouveaux centres européens référencés.

LE PUE, UN INDICATEUR D'EFFICACITÉ ÉNERGÉTIQUE DES CENTRES DE CALCUL

L'efficacité énergétique d'un centre de calcul se mesure avec l'indicateur **PUE** (*Power Usage Effectiveness*) défini comme suit :

$$PUE = \frac{\text{Puissance électrique totale de l'installation}}{\text{Puissance électrique des équipements informatiques}}$$

Un PUE de 1,3 signifie que pour faire fonctionner une configuration informatique de 1 MW, il est nécessaire d'apporter un complément de puissance de 300 kW pour alimenter les infrastructures techniques associées au calculateur (production et distribution d'eau de refroidissement, transformateurs, onduleurs, systèmes de sécurité...). L'amélioration de l'efficacité énergétique d'une génération de supercalculateurs à l'autre est un des thèmes prioritaires des travaux de R&D menés en coopération avec Atos/Bull.

Ainsi, les bons résultats obtenus sur le centre de calcul du CEA/DAM qui accueille les calculateurs Tera sont caractérisés par l'amélioration de son PUE depuis 2010 : 1,6 pour Tera10, 1,3 pour Tera100 et 1,15 pour Tera1000. L'évolution de ce PUE indique que la consommation électrique des servitudes de l'installation est divisée par deux entre chaque génération de supercalculateurs. Le

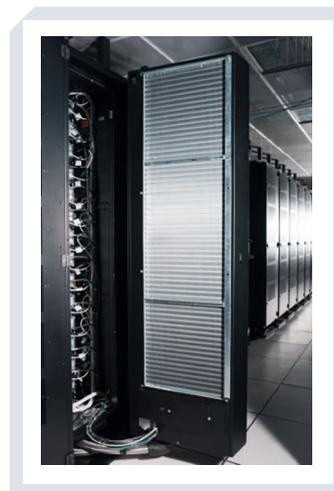
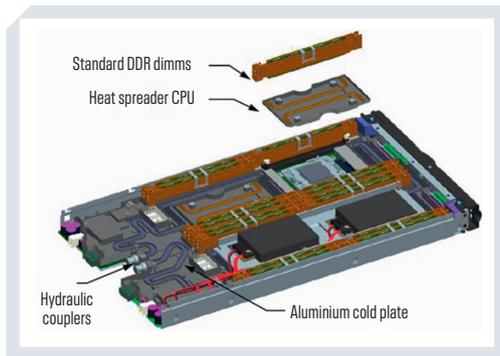


FIGURE 2
Porte froide.



FIGURE 3
Ultra Capacitor Module.

**FIGURE 4**

Cold Plate (DLC).

LA SIMULATION AU SERVICE DU CENTRE DE CALCUL

Mais la maîtrise de la consommation énergétique ne garantit pas le bon refroidissement d'une salle machine. Il est nécessaire d'éviter ou de réduire la présence de points chauds qui pourraient apparaître lors du fonctionnement du supercalculateur. C'est ainsi que la simulation numérique du dégagement de chaleur a été utilisée pour décider de l'agencement des armoires du supercalculateur Tera1000 en salle machine (figure 5).

RÉUTILISER LES CALORIES DES SUPERCALCULATEURS POUR LE CHAUFFAGE

Le régime de température de la boucle de refroidissement de l'installation Tera1000 (entrée d'eau à 35 °C et retour à 45 °C) a été une opportunité pour récupérer les calories produites et

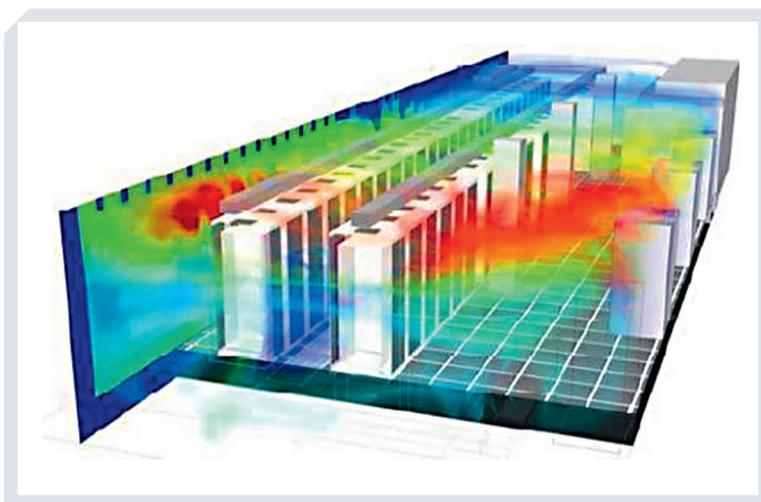
perdues afin de les réutiliser pour le chauffage de bâtiments. Ce projet TERATHERMIE, mis en œuvre dans le centre DAM Île-de-France par les équipes de la DAM, a permis depuis l'hiver 2017 de diminuer considérablement la consommation de gaz usuellement utilisé pour le chauffage des bâtiments du centre et contribue ainsi directement à l'amélioration de son bilan carbone. Cette stratégie perdurera avec les prochaines générations de supercalculateurs.

UNE DÉMARCHÉ PARTAGÉE PAR LES CENTRES DE CALCUL À TRAVERS LE MONDE

L'amélioration continue de l'efficacité énergétique des centres de calcul exploités par le CEA/DAM nécessite un certain investissement et une collaboration étroite avec nos partenaires et nos homologues européens. Cela se traduit par :

- ⊙ une participation active à des groupes de travail internationaux où le CEA met en valeur sa démarche énergétique et partage les retours d'expérience d'experts du domaine (Workshop annuel des partenaires européens dans le cadre du projet **Prace**, groupe de travail **EE-HPC** dédié à l'amélioration de l'efficacité énergétique des centres de calcul haute performance) ;
- ⊙ une contribution à la définition de solutions innovantes et efficaces en matière de refroidissement et d'alimentation électrique au travers de projets collaboratifs de R&D avec des industriels du monde du HPC.

Ainsi, le projet Cool-IT (mis en œuvre au sein du pôle **SYSTEM@TIC**) fut dédié à l'optimisation

**FIGURE 5**

Exemple de simulation de la répartition spatiale de la température de la salle machine produite par un supercalculateur en fonctionnement (Datacenter ressources).

de l'énergie totale nécessaire au fonctionnement d'une infrastructure informatique. Neuf contributeurs dont Atos/Bull, l'**Inria**, le CEA et 6 PME ont travaillé selon les trois axes suivants :

- ⊙ l'étude et développement de nouvelles méthodes de refroidissement ;
- ⊙ l'élaboration de nouvelles méthodes d'asservissement, d'intégration et d'optimisation de la chaîne d'alimentation électrique ;
- ⊙ le développement de la collecte et de la fusion d'informations énergétiques pour une prise de décision au niveau du centre de calcul.

Le projet Efficient DC (aussi mis en œuvre au sein du pôle SYSTEM@TIC) fut consacré à la stratégie de réduction des coûts d'infrastructure d'un cluster HPC en introduisant le refroidissement à « eau tiède » dans les blocs d'alimentation électrique des serveurs de calcul. Cette collaboration entre Atos/Bull, la société française

BrightLoop et le CEA a permis d'aboutir à une nouvelle génération d'alimentation intégrant un refroidissement à eau, et présentant un haut niveau de rendement ainsi qu'une densité volumique de puissance équivalente aux alimentations industrielles refroidies par air.

PERSPECTIVES POUR LA FUTURE MACHINE EXA1

Pour la future machine Exa1, il devient de plus en plus difficile d'améliorer le PUE de l'installation au vu de la faible valeur obtenue avec Tera1000. Ce challenge sera tout de même relevé avec notamment la mise en place d'un refroidissement naturel direct qui permettra de supprimer totalement la production d'eau glacée pour la climatisation des nouvelles structures d'accueil modulaires. ■

CHAPITRE 4

LES LOGICIELS

L'ampleur et l'ambition du programme Simulation ont élevé le niveau des exigences du CEA/DAM en matière de calcul scientifique. Une nouvelle génération de codes de calcul modélisant le fonctionnement des armes et de logiciels de traitement de données a été mise en œuvre. D'une part, ils prennent en compte des modélisations plus précises, et d'autre part ils tirent parti du saut de performances apporté par les supercalculateurs parallèles du programme Simulation.

À partir de 2003, avec le retour d'expérience de la mise en service de Tera1, cet effort a aussi porté sur l'ensemble de la pile logicielle du calcul scientifique. Ceci a notamment conduit à une plus grande maîtrise par le CEA/DAM des logiciels critiques régissant les nouvelles générations de supercalculateurs permettant ainsi d'optimiser les performances du couple supercalculateur/codes de calcul. Elle s'appuie sur l'utilisation de logiciels open source, mais aussi sur la participation aux efforts de la communauté des développeurs, voire à la réalisation de nouveaux logiciels. Elle s'effectue notamment dans le cadre d'accords de collaboration avec de grands organismes ou laboratoires publics et notamment celle établie dans le domaine du «computer science» avec les équipes du département de l'Énergie aux États-Unis (DOE/NNSA).

4.1 La « pile logicielle » du calcul scientifique

— L'association du supercalculateur, monstre d'électronique, et des logiciels de simulation ne peut être réalisée que grâce à un ensemble de logiciels intermédiaires qui est brièvement décrit ici.

La pile logicielle représente l'ensemble de tous les logiciels nécessaires pour assurer l'exploitation, le maintien en condition opérationnelle et la performance applicative de l'ensemble des codes de simulation (applications) ainsi que la pérennisation de leurs données et résultats. C'est la pierre angulaire du bon fonctionnement du calcul scientifique sur les supercalculateurs. Les différentes couches de la pile logicielle sont présentées sur la **figure 1**.

De manière simplifiée, une pile logicielle est constituée de plusieurs couches interagissant plus ou moins (fond de couleur gris sur la **figure 1**) dont la partie système :

- ⊙ une couche logicielle de base : le système d'exploitation pour chaque équipement informatique ;
- ⊙ une couche logicielle système qui fédère l'ensemble de l'accès aux équipements informatiques du supercalculateur, et des composants logiciels qui intègrent les outils d'installation, de configuration et d'administration. Cette couche logicielle couvre l'ensemble du centre de calcul, en intégrant les différents services : le calcul, le stockage, les entrées/sorties, les infrastructures réseaux gestion et calcul, les serveurs dédiés... ;
- ⊙ une couche logicielle « **intergiciels** » (« **middle-ware** » en anglais) fournissant aux logiciels applicatifs une abstraction de la partie infé-

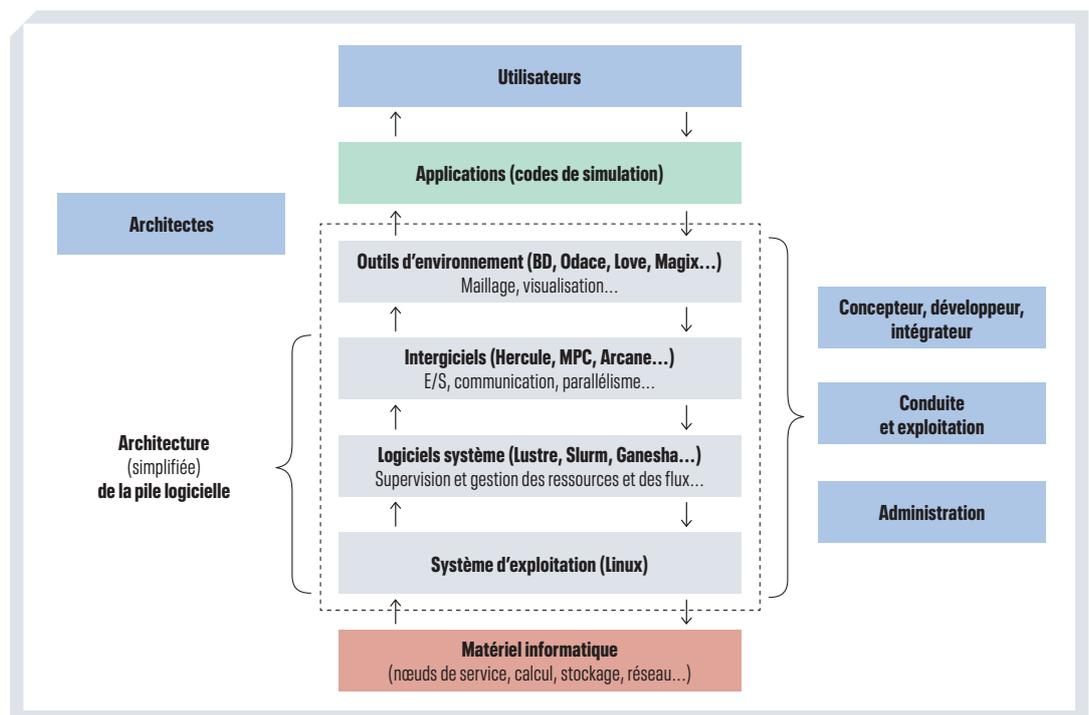


FIGURE 1
Couches de la pile logicielle.

Les bénéfices de l'open source

Qu'est-ce que l'**open source** ? C'est la définition d'une licence d'utilisation d'un logiciel qui garantit plusieurs libertés à l'utilisateur : celle de pouvoir utiliser un logiciel, de pouvoir le modifier, de pouvoir le distribuer tel quel ou modifié. De ce type de licences découlent de nombreux avantages. Tout d'abord, le logiciel n'est plus développé par une seule personne ou une seule équipe mais par plusieurs. Il peut ainsi être amélioré par la communauté, personnalisé ou tout simplement testé plus largement et corrigé si nécessaire. En tant qu'utilisateur, cela garantit une certaine autonomie et une certaine pérennité de la solution

retenue. Si le développement du logiciel est arrêté, n'importe qui peut le poursuivre et le maintenir soi-même. On peut aussi ajouter l'absence de coût d'utilisation.

En tant que développeur, cela permet de partager et valoriser des travaux, mais aussi de bénéficier des retours d'utilisateurs, voire à terme de leurs contributions. De manière globale, cela consolide des logiciels communs qui permettent de favoriser l'innovation et la recherche. Et cela est primordial pour la recherche ; du fait de son ouverture, chaque logiciel open source fait progresser le niveau général des connaissances, permet

d'échanger et par conséquent d'avancer plus vite collectivement.

Dans le domaine du **HPC**, le logiciel open source est au cœur de la stratégie du CEA/DAM depuis plus de quinze ans. Les avantages de cette stratégie d'ouverture à la communauté sont tangibles et de deux ordres : d'une part, elle permet de bénéficier des avancées de la communauté scientifique en simulation numérique, et d'autre part, des contributions nombreuses et de qualité « industrielle » permettent de s'intégrer dans une communauté de pairs, renforçant ainsi l'expertise des équipes du CEA/DAM.

rieure (matériel et logiciels système) facilitant la conception et le développement des codes. Les « intergiciels » permettent d'améliorer la productivité des développeurs de logiciels de simulation, tout en prenant en compte les éléments clés des architectures matérielles dans un contexte complexe de calcul haute performance ;

- ⊙ une couche logicielle « environnement » qui offre à l'utilisateur nombre de solutions pour exploiter, mettre en forme les données et interpréter les résultats des simulations numériques sur de gros volumes (modélisation géométrique et **maillage**, **dépouillements**, visualisation, bibliothèque de données de constantes physiques...). Cette couche comporte une partie « métier » qui la rend spécifique à la communauté des centres de calcul.

SIMULATION NUMÉRIQUE ET LOGICIELS

Une simulation numérique se construit un peu comme une expérience de physique. Dans ce cas, le dispositif expérimental est l'ensemble des programmes informatiques. Les logiciels (ou codes) de calcul sont la traduction des formules mathématiques des modèles physiques étudiés (voir article « Comprendre et prévoir par la simulation »). Ces codes de calcul s'appuient sur des logiciels chargés de dialoguer avec les couches « système » qui agissent directement avec le matériel (**processeur**, réseau...). La première brique logicielle est le système d'exploitation (système **Linux** depuis la machine Tera10). Le système

d'exploitation est le dispositif central de la couche système. En effet, le système d'exploitation permet en particulier d'accéder au réseau rapide d'interconnexion comme le **BXI** (voir **encadré** « Le réseau d'interconnexion – élément clé d'un supercalculateur parallèle »). *Via* les pilotes (drivers), le système d'exploitation est en charge de l'initialisation des connexions. Il va permettre aux couches logicielles intermédiaires comme **MPI** de réaliser les **échanges de messages** nécessaires à l'exécution d'un code de calcul parallèle.

L'implémentation MPI regroupe l'ensemble des fonctions de communication que peut utiliser un code de calcul pour communiquer entre les différents processeurs d'un supercalculateur... Le système d'exploitation fournit aussi les mécanismes et les interfaces pour gérer les flux de données en entrée et en sortie des systèmes de stockage. Il s'agit en particulier des interfaces vers le logiciel **Lustre** (voir **encadré** « Le CEA/DAM, un contributeur majeur au développement du logiciel Lustre »), dans le cas des calculateurs du CEA. Le CEA/DAM a de plus développé le logiciel Hercule (voir article « Modélisation et gestion des données pour la simulation numérique : la plateforme logicielle Hercule »), qui fournit une couche d'abstraction supplémentaire pour la gestion des données. Il simplifie le travail du développeur de codes de calcul en lui permettant de programmer les flux de données à partir d'entités spécifiques à la simulation numérique, comme des **mailles** ou des particules... Le logiciel Hercule est fondé sur un **modèle de données** partagé par tous les codes de simulation.

Le CEA/DAM, un contributeur majeur au développement du logiciel Lustre

Le logiciel « open source » Lustre, élément fondamental des calculateurs Tera depuis 2005, est un système de fichiers parallèle conçu et développé à la fin de la décennie 1990. C'est un logiciel qui permet d'agréger des dizaines de milliers d'unités de stockage physiques (disques rotatifs aussi appelés **HDD** ou disques flash appelés **SSD**) afin de créer un espace de stockage unique hautement performant. Cet espace est visible depuis toutes les machines d'un centre de calcul sous la forme d'une arborescence de répertoires et de fichiers avec la même ergonomie que sur une machine de bureau.

Ce système de fichiers est dit parallèle car il permet aux données d'être accessibles depuis des centaines de milliers de programmes simultanément. Ce **parallélisme**, en agrégeant les performances des éléments de stockage et des liens réseaux, permet d'atteindre les très hautes performances en débit d'**E/S** (entrées/sorties) requises par les supercalculateurs. Lustre a été initialement conçu dans cet objectif par quelques experts de la communauté du calcul haute performance. Le retour d'expérience de Tera I, qui utilisait un logiciel propriétaire, a conduit le CEA/DAM à mettre, dès 2005, le logiciel

Lustre au cœur de la gestion de données du centre de calcul.

Cependant, les experts du CEA/DAM ont très vite compris que ce logiciel devait être enrichi de nouvelles fonctionnalités de gestion des données et se sont engagés dans leurs développements.

La première fonctionnalité, les « OST Pool », est un mécanisme interne à Lustre qui permet de regrouper des fichiers dans une sous-partie des ressources matérielles de stockage afin de les localiser. Ainsi, si Lustre s'appuie sur des ressources matérielles hétérogènes comme un mélange de SSD et de HDD, cela permet de garantir que certains fichiers seront enregistrés sur un ensemble homogène de ressources. La deuxième fonctionnalité qui a été développée est « Lustre HSM ». Ce développement, qui a duré plusieurs années, a permis de mettre en œuvre au sein de Lustre des fonctionnalités de gestion de la vie de la donnée, permettant en pratique d'offrir un espace de stockage quasi infini. Ce mécanisme permet de transférer de façon transparente à l'utilisateur et à l'administrateur le contenu des fichiers gérés par Lustre vers un espace de stockage externe, comme des bandes magnétiques. Ainsi, les données les plus utilisées sont conservées

sur le stockage rapide et les données plus anciennes et peu utilisées sont conservées sur du stockage plus lent mais aussi plus capacitif, plus fiable et ne consommant aucune énergie (voir **encadré** « Organiser le centre de calcul autour des données »).

Ce développement a aussi conduit le CEA/DAM à concevoir et développer un nouveau logiciel compagnon de Lustre, Robinhood, une intelligence artificielle qui, à base d'événements et de règles, va déplacer les données au sein du centre de calcul.

Enfin, le dernier exemple de développement autour de Lustre est un outil de configuration hautement parallèle, Lustre shine, qui permet de gérer tous les systèmes de fichiers du centre de calcul et leur déploiement sur toutes les machines.

L'ensemble de ces développements a été mis en « open source » et est aujourd'hui utilisé dans de nombreux centres de calcul HPC.

Aujourd'hui, les administrateurs système et les développeurs responsables du complexe de calcul du CEA continuent d'être très actifs dans la communauté du calcul haute performance et du traitement des données scientifiques.

L'écriture des parties « calculatoires » d'un code de simulation sur ordinateur parallèle peut bien entendu reposer sur les outils « standard » de la programmation parallèle. Mais l'aspect générique de ceux-ci ne permet pas la prise en compte de notions spécifiques à la résolution numérique d'équations aux dérivées partielles sur un domaine de calcul discrétisé par un maillage. C'est la raison du développement d'un troisième élément logiciel : la plateforme de développement Arcane (voir article « Développer les logiciels de simulation : la plateforme Arcane »). Cette plateforme facilite l'écriture de logiciels de simulation

en prenant en compte directement des notions plus abstraites, comme les maillages et les opérations sur les mailles. Elle facilite aussi l'écriture informatique des formules mathématiques pour la résolution des équations aux dérivées partielles qui modélisent un ensemble de phénomènes physiques.

Les logiciels mentionnés ci-dessus forment la couche d'intergiciels (middleware) d'un code de calcul, car ils sont directement utilisables par le programmeur. Ces intergiciels permettent au code de calcul de s'exécuter *via* l'appel aux fonctions du système d'exploitation du ordinateur.



M. HAUTREUX

Cette structuration en couches logicielles permet une montée en abstraction qui assure la pérennité des développements réalisés dans les codes de simulation. Elle permet l'évolution du matériel (processeur, réseau d'interconnexion...) sans conséquences sur les logiciels de simulation.

En amont et en aval de cette phase d'exécution, des logiciels de pré- et post-traitement sont chargés de conditionner les données d'entrée d'une simulation (voir article « Préparer une simulation numérique : maillage du modèle ») et d'analyser les résultats (voir article « Traiter et visualiser les résultats d'une simulation numérique »).

LE DÉVELOPPEMENT DES LOGICIELS « OPEN SOURCE » AU CEA/DAM

Pour réaliser les objectifs du programme Simulation, les ingénieurs et chercheurs du CEA/DAM ont choisi de s'intégrer aux différentes communautés scientifiques qui composent le paysage de la simulation numérique et du calcul à haute performance. L'utilisation d'un logiciel par une vaste communauté d'utilisateurs conduit à une amélioration continue de la fiabilité de celui-ci. De plus, l'accès au code source du logiciel permet son analyse et la détection d'éventuelles failles de sécurité.

Les travaux dans ce domaine touchent tous les niveaux de la pile logicielle HPC, depuis les outils d'analyse et de visualisation scientifiques, utilisés quotidiennement par les physiciens pour analyser les résultats des simulations numériques, jusqu'aux logiciels système spécifiques aux supercalculateurs.

L'ENVIRONNEMENT LOGICIEL POUR LA SIMULATION NUMÉRIQUE

Ainsi, pour la préparation des données de simulation, la bibliothèque open source de **CAO Open Cascade** a été retenue puis enrichie avec des structures de données (GMDS) et des algorithmes (FRAME) rendus accessibles à la communauté scientifique, facilitant la collaboration avec d'autres organismes de recherche et universités. Pour la visualisation 3D et parallèle des résultats de simulation, les experts du CEA/DAM sont actifs au sein de la communauté du logiciel VTK. Un logiciel de visualisation de données à la volée, simultanément à leur production par le code de simulation, a été développé (voir article « Analyse et visualisation des données *in situ* : **PaDaWAn**, une solution au traitement des volumes de données à l'échelle de l'**exascale** »).

« Tera100 a vu l'arrivée d'un nouveau système de batch.

Nous avons basculé d'un système propriétaire à un système open source, afin que le CEA contribue au bon fonctionnement du logiciel, décrit Matthieu Hautreux. Cela permet également de mieux adapter l'outil aux spécificités du CEA, notamment la taille de nos machines. Nous avons ainsi été les premiers à passer certaines barrières. Sans notre logiciel de traitement par lots baptisé Slurm, nous aurions eu plus de difficultés à utiliser efficacement nos systèmes. Slurm fonctionne aujourd'hui sur toutes nos machines. »

Pour faciliter la conception et l'écriture des codes de simulation eux-mêmes, plusieurs briques logicielles sont mises à disposition de la communauté scientifique :

- ⊙ la bibliothèque **Alien** fournit des interfaces unifiées pour l'utilisation de solveurs linéaires open source comme **Hypre** ou ceux de **Trilinos** (tous deux issus des laboratoires américains du département de l'Énergie) ;
- ⊙ le langage **Nabla**, un **DSL**, est dédié à l'écriture de **schémas numériques** et l'outil **NabLab** fournit un environnement dédié à l'écriture de programme en langage Nabla ;
- ⊙ l'outil de **partitionnement** de maillages Coupe vient compléter la liste des outils de partitionnement open source disponibles, comme **Scotch** ou **Metis**.

L'ENVIRONNEMENT LOGICIEL DES SUPERCALCULATEURS

Depuis la fin des années 1990, l'architecture des supercalculateurs utilisés dans les centres de calcul est fondée sur l'interconnexion d'un grand nombre de serveurs. À la différence d'une machine monolithique, ces supercalculateurs requièrent un ensemble de logiciels distribués sur les serveurs (appelés nœuds) qui assurent la synchronisation des composants matériels et la fédération de l'ensemble des services afin d'of-

frir aux utilisateurs la vision homogène d'une machine unique. Parmi ceux-ci, mentionnons :

- ⊙ le système de fichiers : il permet de gérer les données rémanentes du centre de calcul ;
- ⊙ le gestionnaire de ressources : il permet d'allouer les ressources de calcul en suivant des règles d'exploitation et de partage ;
- ⊙ les logiciels de surveillance et de mesure, pour la détection des pannes et anomalies, et la collecte des métriques et des statistiques sur les différents composants du centre de calcul (voir article « Surveillance du centre de calcul Tera »).

Tirant l'expérience de Tera1 et avec Tera10, une stratégie active d'utilisation ou de développement de logiciels open source a été mise en place, avec l'adoption du système d'exploitation Linux, du système de fichiers Lustre (voir encadré « Le CEA/DAM, un contributeur majeur au développement du logiciel Lustre ») et du gestionnaire de ressources **Slurm**, logiciel dont le Lawrence Livermore National Laboratory, un laboratoire du département de l'Énergie (**DOE**) des États-Unis, est l'initiateur.

À cette fin, le profil des ingénieurs et chercheurs du centre de calcul a notablement évolué (voir article « Les métiers du HPC au CEA/DAM »). Outre un investissement important dans le développement du logiciel Lustre et de logiciels « compagnons », cette stratégie a aussi conduit à la conception et au développement de logiciels à l'initiative du CEA/DAM. En voici deux exemples récents :

- ⊙ Le logiciel **PCOCC** (pour Private Cloud On a Compute Cluster) permet aux utilisateurs d'un calculateur à architecture distribuée d'instancier simplement des **machines virtuelles** pour y exécuter des applications parallèles. Les utilisateurs disposent ainsi d'un contrôle complet de leur environnement d'exécution et peuvent le personnaliser selon leurs besoins. L'outil s'intègre au gestionnaire de ressources Slurm de façon à pouvoir héberger les machines virtuelles sur les nœuds du cluster en concurrence avec les autres tâches de calcul. Les machines virtuelles d'un utilisateur sont interconnectées par un réseau privé et isolé du réseau du calculateur afin de ne pas perturber son fonctionnement.
- ⊙ Le logiciel Selfie (pour Self and Lightweight profiling Engine) est un outil de **profilage** léger de codes de calcul parallèles se présentant sous la forme d'une bibliothèque dynamique à précharger avant exécution. Cet outil modulaire permet de récupérer des informations de profilage

simple (temps d'exécution, mémoire maximum utilisée...) et peut intégrer des extensions pour remonter d'autres informations de profilage sur les entrées/sorties, les échanges de messages ou d'autres informations récupérées des compteurs matériels.

CONCLUSION

L'ensemble des composants constituant une pile logicielle doit répondre aux besoins des différents types d'acteurs : administrateurs, développeurs et utilisateurs. Une partie des composants logiciels de la pile logicielle implémentée sur les calculateurs Tera est partagée par toute une communauté scientifique et technique au-delà du CEA/DAM. Les experts du CEA/DAM y contribuent en vue de :

- ⊙ corriger les dysfonctionnements (correctifs d'erreur de programmation) ;
- ⊙ développer des fonctionnalités nouvelles ;
- ⊙ adapter la pile logicielle aux nouvelles architectures matérielles (processeur, mémoires, réseaux d'interconnexion, entrées/sorties, technologies de stockage...);
- ⊙ innover sur des nouveaux concepts en cohérence avec les besoins du programme Simulation (performance des outils de simulation numérique, gestion des données et des flux, gestion et allocation des ressources...). ■

4.2 L'évolution des modèles de programmation parallèle : le projet MPC

— La puissance de calcul d'un supercalculateur est obtenue en parallélisant à la fois sur les nœuds de calcul et sur les cœurs dont ils sont composés. L'outil MPC permet de gérer efficacement ces deux niveaux de parallélisme.

Un logiciel de simulation numérique a pour objectif de simuler un phénomène physique dans un espace ou sur un domaine déterminé. Une première étape consiste par exemple à construire une **discrétisation** du domaine physique sous la forme d'un **maillage** que l'on peut représenter schématiquement comme une grille (voir article « Préparer une simulation numérique : maillage du modèle »). Cette grille ou domaine de calcul est le fondement sur lequel va s'appuyer la résolution, dans l'ordinateur, des équations modélisant le phénomène physique. Une des spécificités du calcul à haute performance est la nécessité de savoir tirer parti des caractéristiques architecturales des supercalculateurs. À partir des années 1980, la **vectorisation**, ou programmation vectorielle, s'est révélée indispensable. Au début des années 1990, le **parallélisme** de tâches a fait son apparition avec les premiers calculateurs multiprocesseurs à mémoire partagée. Une telle organisation des calculs est indispensable pour répartir ceux-ci sur les différents processeurs sans avoir à se soucier de l'emplacement des données, puisque tous les processeurs y ont accès.

Avec l'arrivée des calculateurs parallèles à mémoire distribuée, cette vision globale des données d'une simulation est perdue au profit d'un formidable accroissement des capacités de calcul. La programmation parallèle doit tout d'abord prendre en compte la répartition des données sur les **nœuds de calcul** du calculateur

puis les nécessaires communications, appelées **échanges de messages**, entre ceux-ci. À cette fin, le maillage va préalablement être découpé en plusieurs **sous-domaines**, chaque sous-domaine étant affecté à un nœud de calcul. Le calcul aux frontières d'un sous-domaine nécessite la communication des informations portées par les frontières des sous-domaines voisins, ceux-ci étant répartis sur des nœuds de calcul différents. Ces communications entre nœuds de calcul doivent être explicitement insérées au sein du code de simulation. Pour ce faire, la communauté du calcul scientifique a défini et fait régulièrement évoluer un standard international : Message Passing Interface (**MPI**), qui définit une bibliothèque de fonctions utilisables avec les principaux langages de programmation.

L'évolution des architectures des différents supercalculateurs Tera a eu un impact sur les méthodes de programmation des codes de simulation. En effet, dans les premières années, la programmation parallèle se faisait principalement grâce à un modèle basé sur les échanges de messages. Cependant, l'augmentation du nombre de **cœurs de calcul** par nœud de calcul et la diminution de la quantité de mémoire disponible par cœur ont poussé à l'utilisation d'un mode « hybride », mêlant deux paradigmes de programmation parallèle : un paradigme de programmation parallèle à mémoire distribuée et un paradigme de programmation parallèle à mémoire partagée (pour une gestion plus efficace des ressources partagées).

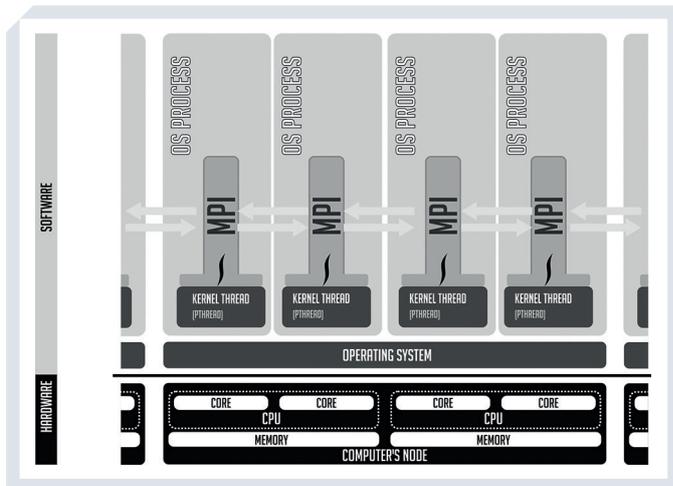


FIGURE 1

Modèle d'exécution MPI.

LA PROGRAMMATION PARALLÈLE À MÉMOIRE DISTRIBUÉE : L'INTERFACE MPI

À l'arrivée du supercalculateur Tera1, le modèle de programmation prédominant pour le parallélisme est MPI (*Message Passing Interface*). S'appuyant sur un paradigme à mémoire distribuée, MPI propose une interface définissant un ensemble de fonctions à appeler dans le code pour effectuer des échanges de messages. Ainsi, le programme est exécuté en parallèle par un ensemble de ressources appelées « processus MPI » (figure 1). Afin de ne pas effectuer de calcul redondant, le programmeur doit spécifier sur quelles données chaque processus MPI doit travailler grâce à des fonctions permettant d'avoir accès à son **rang**. Dans la grande majorité des codes de simulations (proposant des méthodes itératives), il est alors possible de découper le système concerné en sous-parties, ou sous-domaines, réparties sur les différents processus MPI.

Si cette approche permet de résoudre le système en parallèle, cela nécessite également des échanges de données entre les sous-domaines. En effet, en faisant évoluer le système global, il est commun que des éléments transitent à travers le système et puissent ainsi passer d'un sous-domaine à un autre. Pour effectuer ces transitions, la base des échanges de messages dans MPI est le modèle point à point : les deux parties (l'émetteur et le récepteur) doivent participer en appelant une fonction du standard. L'émetteur appelle une fonction de type envoi (« Send ») en spécifiant le destinataire ainsi que le contenu du message (adresse mémoire de départ, nombre et type d'éléments à envoyer). De son côté, le récepteur invoque une

fonction de type réception (« Receive ») signifiant qu'il est prêt à recevoir un message, en indiquant le rang de l'émetteur, ainsi que la zone où stocker le message une fois la réception effectuée. Cette approche est illustrée figure 1 à travers les flèches entre les processus MPI.

Au-delà des communications point à point disponibles depuis le standard MPI-1 de novembre 1993, MPI propose des opérations collectives qui fournissent aux utilisateurs une implémentation efficace de schémas de communication plus complexes, nécessitant la participation de tous les processus MPI. On peut citer par exemple la fonction « broadcast », pour diffuser les données d'un processus MPI, la fonction « scatter », pour répartir les données, ou encore les fonctions de type « reduce », permettant la réduction de données réparties sur tous les rangs.

MPI est le standard *de facto* pour les communications en mémoire distribuée dans le domaine du **HPC**. Il est géré par un comité composé d'une trentaine d'organisations internationales, dont le CEA, appelé MPI Forum. Ce comité se réunit quatre fois par an et organise des conférences téléphoniques hebdomadaires sur des sujets spécifiques pour continuer à faire évoluer le standard MPI.

LA PROGRAMMATION PARALLÈLE À MÉMOIRE PARTAGÉE ET HYBRIDE : OPENMP

Les différentes générations de supercalculateurs Tera ont mis en avant deux évolutions majeures : l'augmentation du nombre de cœurs par nœud et l'avènement des processeurs multicœurs. Cela a un impact direct sur les modèles de programmation parallèle utilisés par les applications. En effet, les différents cœurs de calcul vont partager certaines ressources telles que la mémoire du nœud. Lorsqu'on utilise un modèle à mémoire distribuée (comme MPI) sur ce type de nœud, chaque entité du modèle distribué va allouer ses structures internes, entraînant un surcoût mémoire. Pour un nombre important de cœurs de calcul sur le même nœud, cela peut empêcher le bon fonctionnement de l'application parallèle. Pour ces raisons, des modèles de programmation parallèles à mémoire partagée ont été développés, tels que le standard **OpenMP**, dont la première version sort en 1997. Ainsi, il n'est plus nécessaire de spécifier explicitement les échanges de messages (comme dans MPI) : les entités de calcul peuvent partager les accès aux données.

OpenMP est un modèle fonctionnant par l'insertion de directives de compilation (ou

« pragmas »). Au lancement de l'application, le code est exécuté séquentiellement jusqu'à ce qu'il arrive sur une directive lançant l'exécution parallèle. Cela ouvre une région parallèle en créant un certain nombre d'entités de calcul parallèle (« threads »). Le modèle d'exécution parallèle est illustré **figure 2**. À la fin de cette région, les threads générés sont alors détruits ou endormis, et l'exécution reprend de façon séquentielle jusqu'à la prochaine directive d'exécution parallèle. Ce modèle repose alors sur une vision « fork-join » du parallélisme : le programme alterne entre des parties séquentielles et parallèles. Pour répartir le travail sur les threads, il existe deux possibilités : (i) spécialiser le travail de chaque thread en fonction de son rang ou (ii) répartir le travail sur tous les threads à l'aide de directives dédiées.

Cette vision partagée de la mémoire permet ainsi une répartition plus directe du travail, mais elle apporte également des contraintes comme la concurrence des accès. En effet, il est alors possible qu'un thread essaie de lire une donnée qu'un autre thread est en train d'écrire. En fonction de l'ordre de ces accès, le résultat obtenu pourra être différent. Pour imposer un ordre entre différents threads, il est nécessaire de les synchroniser. Ainsi, on peut s'assurer qu'une lecture de donnée aura bien lieu après que celle-ci a été mise à jour par un autre thread. Cette synchronisation peut se faire soit avec l'insertion d'une **barrière** explicite, soit à l'aide d'un **verrou** permettant de limiter le nombre de threads exécutant de façon simultanée un ensemble d'instructions.

OpenMP est un des langages les plus utilisés dans le HPC pour la programmation en mémoire partagée. Il est géré par un comité composé lui aussi d'une trentaine d'organisations internationales, appelé Architecture Review Board (ou OpenMP ARB). Ce comité se réunit trois fois par an et organise des conférences téléphoniques régulières sur des sujets spécifiques pour continuer à faire évoluer le standard OpenMP.

PROGRAMMATION PARALLÈLE HYBRIDE MPI/OPENMP : LE PROJET MPC

Si la programmation parallèle en mémoire partagée offre certains avantages pour la programmation intra-nœud dans les supercalculateurs Tera, elle a le grand inconvénient de ne pas permettre une programmation parallèle sur l'ensemble du supercalculateur. C'est pourquoi il est nécessaire de la combiner avec une programmation paral-

lèle à mémoire distribuée tel que le MPI. Cette approche s'appelle programmation hybride.

Leur utilisation conjointe est cependant bien plus complexe. Chaque modèle de programmation est implémenté dans un support exécutif, s'exécutant en arrière-plan pour permettre une progression parallèle. Le problème majeur de la combinaison de tels modèles est le fait que ces supports exécutifs sont développés par des personnes/équipes/organismes différents, et ils n'ont pas forcément connaissance de l'existence d'autres modèles. Ils mettent alors en place des politiques agressives d'utilisation des ressources (par exemple attente active) pour améliorer les performances. Ces ressources sont donc occupées par un support exécutif pour du temps d'attente, alors qu'elles pourraient être utilisées par le deuxième support exécutif pour faire du travail « utile », et recouvrer ainsi ce temps d'attente. Ce manque de collaboration entre les supports exécutifs est un sujet d'évolution des modèles depuis quelques années. Plusieurs axes sont développés. Le premier se situe au niveau des standards de programmation parallèle. Par exemple, le standard MPI intègre depuis plusieurs années des options permettant d'appeler des fonctions du standard à l'intérieur de threads générés par un modèle à mémoire partagée. Un autre axe se situe directement au niveau des supports exécutifs, proposant des options pour limiter la consommation de ressources. Certains

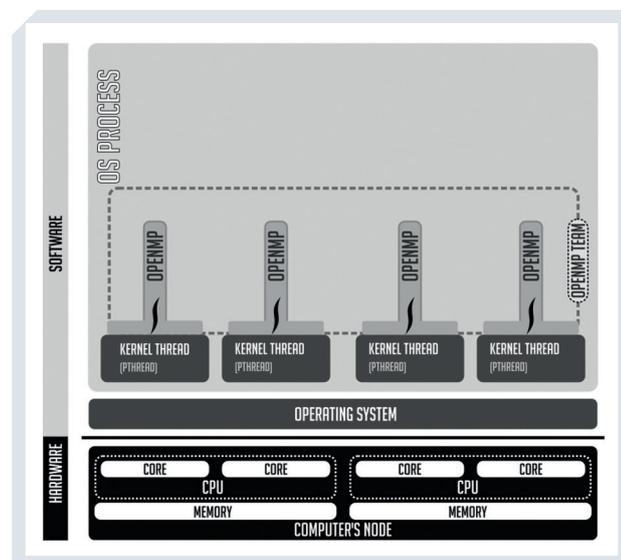


FIGURE 2

Modèle d'exécution OpenMP.

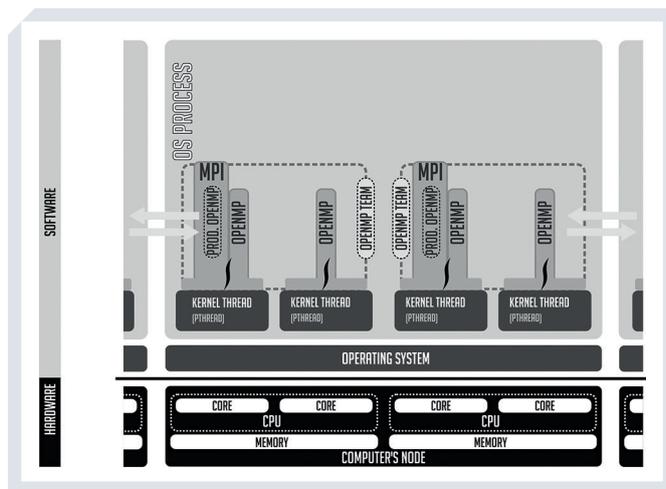


FIGURE 3
Modèle d'exécution MPI/OpenMP sur MPC.

supports vont plus loin en combinant l'implémentation de plusieurs modèles. Cela permet d'avoir une vision globale du parallélisme présent dans l'application, et de prendre les bonnes décisions sur les politiques d'attente et de recouvrement lors de synchronisation.

Dans ce cadre, le CEA/DAM développe le projet open source MPC (Multi-Processor Computing) disponible à l'adresse suivante : <http://mpc.hpcframework.com>. Cette plateforme logicielle propose une implémentation unifiée des modèles de programmation MPI et OpenMP afin d'améliorer les performances du parallélisme hybride des applications. En effet, grâce à la vision globale de ces modèles de programmation, il est alors possible d'améliorer le placement des flots d'exécution sur les cœurs de calcul et de réduire l'empreinte mémoire des structures internes. Le modèle d'exécution hybride MPI/OpenMP via MPC est présenté **figure 3**. Le projet MPC est porteur de nombreux axes de recherche avec l'encadrement de stagiaires et de doctorants, et la publication de nombreux articles dans des revues et conférences internationales. Cette dimension recherche du projet MPC lui permet de participer à de nombreuses collaborations avec des partenaires français, tels que l'**Inria**, mais aussi des partenaires étrangers, tels que les laboratoires nationaux des États-Unis (**DOE**), ou encore le centre de recherche du RIKEN au Japon.

À travers les membres du projet MPC, le CEA/DAM participe activement au MPI Forum et à l'évolution du standard MPI, primordial pour la programmation parallèle des supercalculateurs. Cela lui permet de partager ses expériences avec d'autres organisations et de mettre en avant ses besoins. Fort de son expérience dans le projet MPC, le CEA/DAM est un membre très actif dans les problématiques

de topologie matérielle, de communications asynchrones et de compatibilité avec les modèles de programmation pour mémoire partagée.

Pour exploiter les futures générations de supercalculateurs, il sera nécessaire d'exploiter en profondeur le modèle hybride en s'appuyant notamment sur les capacités des supports exécutifs à orchestrer l'exécution des différentes parties du code. En effet, exprimer l'application sous forme de multiples parties de calcul avec une gestion fine des communications nécessaires permettrait d'optimiser l'exécution en recouvrant les communications par du calcul et de pallier le surcoût induit par le réseau de communication. ■



Marc Pérache à l'interface de plusieurs informatiques

Les informaticiens spécialistes du « middleware » s'assurent que les simulations utilisent le plus efficacement possible le matériel et les logiciels. Une tâche complexe au service des autres.

— « Simplifier la vie des développeurs de code » : telle est la mission des informaticiens du « **middleware** » (**intergiciel** en français), à l'interface entre ceux qui développent les logiciels de simulation et ceux en charge des logiciels système. C'est le travail de Marc Pérache au CEA/DAM. Concrètement, il s'agit de gérer la manière dont les simulations utilisent les processeurs, le système de fichiers et le réseau d'interconnexion entre les nœuds de calcul.

Sept ans en avance

Par exemple, il est possible d'utiliser pour les calculs non seulement les processeurs, mais aussi les cartes graphiques : c'est le système **GPU**, développé par le constructeur Nvidia, et présent sur Tera100 et Tera1000. Mais cela engendre des modèles de programmation différents : certains calculs sont adaptés, d'autres non. C'est typiquement le rôle de Marc Pérache que de veiller à ce que les calculs le prennent en compte.

C'est une tâche de longue haleine, qui commence bien avant l'arrivée d'une nouvelle machine. « Environ sept ans avant son arrivée, on regarde quels sont les besoins des utilisateurs, quelles technologies seront probablement disponibles sept ans après, et ce qu'il est possible de faire dans ces intergiciels, explique Marc Pérache. Par exemple, si l'on souhaite utiliser des processeurs que l'on ne sait pas gérer avec nos logiciels, soit on change d'avis, soit on développe les logiciels adéquats. » Deux ans avant la date prévue d'arrivée de la machine, on fige les choix scientifiques, mais le développement informatique continue, aussi bien sur les processeurs que sur les fichiers et l'interconnexion.

Faire évoluer les codes

« C'est un travail sur la durée, il faut se projeter sept ans plus tard, apprécie Marc Pérache. Nous disposons de Tera1000 depuis 2017, mais nous nous posons déjà des questions sur les machines suivantes. Cela

n'empêche pas de continuer à travailler sur Tera1000, il y a toujours des améliorations à faire. » Un code de calcul évolue très lentement, sa durée de vie est d'une quarantaine d'années : vingt ans de conception, vingt ans d'utilisation. En effet, les problèmes à résoudre sont si complexes qu'il est impossible de tout réécrire, et il faut donc faire évoluer ces codes.

Vive le doctorat

La motivation de Marc Pérache, c'est la difficulté. « Les challenges sont très ambitieux, on utilise une technologie qui n'existe presque pas ailleurs, se réjouit-il. J'aime aussi la durée des projets : on n'est pas sur du "one shot", on a le temps de construire de belles choses, depuis les balbutiements jusqu'au produit fini. J'apprécie aussi la veille technologique, et l'encadrement des doctorants : des jeunes arrivent avec de nouvelles idées, on s'attelle pendant trois ans sur des points durs... Bref, je m'éclate ! »

Pourtant, l'informatique n'était pas sa passion première puisque, plus jeune, il voulait être... aiguilleur du ciel ! Mais l'influence de ses professeurs d'université a été déterminante. « Va à l'ENS, tu t'y plairas », lui a dit l'un. « Va au CEA, tu t'y plairas », a assuré un autre. Nous sommes alors dans les années 2000, à la fin de la « bulle informatique », lorsque tout le monde croyait qu'il n'y avait plus de débouchés dans ce secteur. Aujourd'hui, c'est l'inverse : les étudiants en informatique sont trop peu nombreux par rapport aux besoins. Marc Pérache tente donc de motiver les jeunes à aller vers l'informatique, par exemple lors de la Fête de la science.

Il conseille aux jeunes de ne pas craindre la durée des études. « La thèse est un vrai travail, rémunéré, et elle est indispensable pour avoir une visibilité à l'international, souligne-t-il. Elle n'est pas pénalisante si l'on souhaite aller dans l'industrie. » Dans son laboratoire, une large proportion des personnes recrutées avait fait auparavant une thèse.

4.3 Préparer un calcul et développer les logiciels de simulation

— Les logiciels de simulation, qui capitalisent un ensemble considérable de connaissances, doivent être fonctionnels durant des décennies, donc s'adapter à l'évolution des machines. Ils doivent pouvoir être conçus et réalisés par des physiciens, et être en même temps efficaces sur plusieurs générations de machines de puissance. La plateforme Arcane fournit un ensemble de services répondant à ces exigences *a priori* contradictoires.

4.3.1 Développer les logiciels de simulation : la plateforme Arcane

Le programme Simulation comprend un volet « logiciel » qui vise à concevoir une chaîne de simulation numérique fiable tirant parti des performances des générations de supercalculateurs. Pour répondre à ce défi, le CEA/DAM a démarré en 2000 l'élaboration d'une plateforme logicielle pour faciliter le développement de codes de simulations numériques 2D et 3D, fondés sur la résolution de systèmes d'équations aux dérivées partielles (voir article « Comprendre et prévoir par la simulation »). Cette plateforme logicielle a été dénommée Arcane. Depuis 2007, elle est développée et enrichie en partenariat avec l'organisme de recherche IFP Énergies Nouvelles (**IFPEN**) qui s'est engagé dans une démarche similaire pour ses besoins propres. La réalisation d'un code de simulation

numérique requiert l'association de spécialistes en physique, mathématiques et informatique. La structure en couches logicielles d'Arcane permet à chacun de se concentrer sur son savoir-faire (**figure 1**).

Les objectifs d'Arcane sont les suivants :

- ⊙ prendre en charge les parties techniques des codes de simulation pour simplifier le développement des modules physiques et mathématiques. Cela comprend notamment la gestion du **maillage**, du **parallélisme**, de la mémoire, des entrées/sorties... ;
- ⊙ masquer la programmation parallèle au développeur du code de simulation tout en assurant de bonnes performances sur les supercalculateurs massivement parallèles ;
- ⊙ accélérer le cycle de développement en offrant un ensemble de logiciels compagnons spécialisés facilitant la conception, l'implémentation, le débogage et la vérification des codes de simulation numérique.

Arcane est écrit en **langage C++** et comprend plusieurs centaines de milliers de lignes de code. La plateforme comporte deux parties : les classes de base et un ensemble de services standard comme la lecture du maillage ou les sorties de post-traitement. Les outils d'Arcane, comme la gestion des calculs sur les supercalculateurs ou l'environnement de test, sont écrits en **langage C#**.

Les principaux concepts manipulés par Arcane sont le maillage, les modules, les variables et les services numériques, ces derniers ayant pour fonction de faciliter l'écriture des algorithmes de schémas numériques.

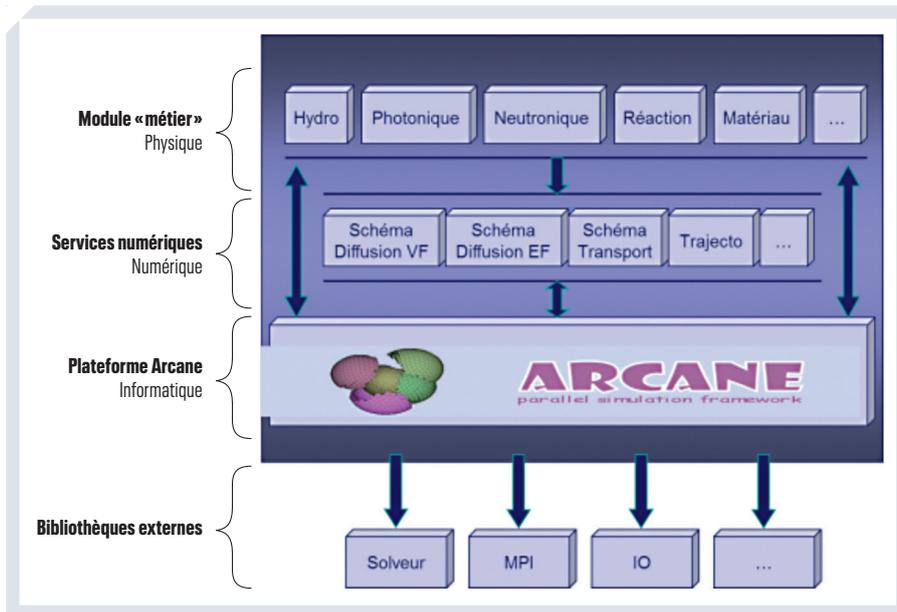


FIGURE 1
 Architecture typique d'un code de simulation exploitant la plateforme informatique Arcane.

Ainsi, Arcane gère les maillages de dimension 1, 2 ou 3 de type non structuré (irrégulier) et différents types de **maillages** : triangle, quadrilatère, tétraèdre, hexaèdre. Ces maillages sont utilisés dans les simulations de type Adaptative Mesh Refinement (**AMR**) où les mailles sont raffinées (subdivisées) et dé-raffinées (fusionnées) au cours du calcul en fonction de différents critères. Le maillage est de plus dynamique : il est possible d'ajouter ou de supprimer des entités durant la simulation. Dans ce cas, Arcane assure automatiquement la cohérence des objets de la simulation. Par exemple, en cas d'ajout de mailles, les variables associées aux mailles sont automatiquement redimensionnées. Les structures de maillage sont conçues pour que le codage soit indépendant de la dimension lorsque cela est possible.

Un code numérique basé sur Arcane est composé de plusieurs modules « métiers » traitant différentes équations de la physique : hydrodynamique, neutronique, thermique... Un module contient les modèles et algorithmes mathématiques. Les modules sont indépendants et ne partagent pas de code entre eux. Chaque module est caractérisé par :

- ⊙ des variables qui représentent ses paramètres physiques, par exemple la température ou la pression ;
- ⊙ les traitements et opérations que fournit le module, par exemple le calcul des forces ou l'écriture des informations de post-traitement ;
- ⊙ des options de configuration qui définissent les données que l'utilisateur devra fournir en entrée.

Les modules Arcane forment le cœur du code de simulation numérique.

Arcane implémente aussi le concept de variable globale, dans un contexte où les données sont réparties sur les mémoires de **nœuds de calcul** différents. Une variable est identifiée par son nom et caractérisée par :

- ⊙ un type de données, par exemple un entier, un réel, un vecteur ;
- ⊙ une dimension, par exemple un scalaire (0), un tableau 1D (1) ou une matrice (2) ;
- ⊙ éventuellement un genre d'entité sur laquelle repose la variable, par exemple les mailles, les nœuds ou les particules.

À titre d'exemple, la température est une grandeur réelle scalaire portée par chaque **maille** du maillage. Pour toutes les variables, Arcane prend en charge :

- ⊙ la gestion de la mémoire : Arcane alloue les variables définies sur les éléments du maillage (mailles, nœuds...) et fournit des mécanismes permettant d'itérer globalement sur ces variables sans se soucier de leur distribution au sein du calculateur. Ainsi, le code utilisateur reste indépendant de la manière dont sont stockées les variables. Si une répartition dynamique de la charge entraîne un déplacement des éléments de maillage au sein des nœuds de calcul, les variables sont réallouées automatiquement sur chacun des sous-domaines d'exécution ;
- ⊙ le parallélisme : Arcane fournit des méthodes pour synchroniser ou mettre à jour les variables entre les différents sous-domaines ;

⊙ la persistance des données : Arcane prend en charge la lecture/écriture lors des arrêts/redémarrages du code de simulation et permet d'écrire les variables dans différents formats de sortie en vue du post-traitement. Cette stratégie de persistance se paramètre au moment de la définition de la variable, dans le descripteur de module.

Par conception, les modules ne peuvent pas s'appeler entre eux. Cela évite de coupler trop de parties de code et permet de simplifier la maintenance et l'évolutivité des modules. Dans certains cas, il est cependant nécessaire de pouvoir partager du code entre des modules. Par exemple, un **schéma numérique** peut être utilisé par plusieurs modules. Pour répondre à ce besoin, Arcane fournit la notion de service. Un service peut être technique, comme le service de gestion du parallélisme, ou réaliser un algorithme de calcul de schémas numériques. Dans les deux cas, il définit un contrat représenté par un ensemble d'opérations rassemblées sous la forme d'une interface. Les modules peuvent alors utiliser une instance du service comme option du jeu de données en spécifiant l'interface souhaitée.

Les simulations 3D peuvent contenir plusieurs dizaines de millions de mailles non structurées. Pour que les temps de calcul restent compatibles avec les contraintes de conception, il est nécessaire d'utiliser un grand nombre de **cœurs de calcul**. Arcane a été conçu pour pouvoir utiliser efficacement au moins 10 000 cœurs de calcul.

La première approche utilisée est celle du **partitionnement** de maillage : le maillage initial est partitionné en plusieurs sous-domaines associés chacun à un cœur de calcul. Chaque entité du maillage appartient à un et un seul sous-domaine appelé propriétaire de l'entité.

Les communications entre sous-domaines utilisent un paradigme d'échange de messages. Arcane fournit plusieurs modes de parallélisme :

- ⊙ **MPI** : le mode le plus usuel, utilisé dans les codes de production ;
- ⊙ Thread ou **OpenMP** : mode permettant l'utilisation de la mémoire partagée mais limité à un nœud de calcul ;
- ⊙ hybride thread/MPI : les threads sont utilisés à l'intérieur d'un nœud de calcul et MPI pour les communications entre les nœuds (voir article « L'évolution des modèles de programmation parallèle : le projet **MPC** »).

Les calculs des valeurs des variables physiques portées par les mailles nécessitent souvent de connaître les valeurs sur les mailles voisines. Chaque sous-domaine effectue les calculs

sur ses propres entités et, lorsqu'il a terminé, les variables modifiées sont synchronisées : chaque sous-domaine envoie les valeurs nécessaires aux autres sous-domaines et reçoit celles dont il a besoin. Arcane facilite ce travail en permettant de synchroniser directement les variables en une seule instruction, par exemple, pour une variable appelée 'temperature' : `temperature.synchronize()`. Ainsi, le développeur du logiciel de simulation n'a plus à se soucier de la programmation des **échanges de messages** entre sous-domaines.

Certains codes basés sur Arcane utilisent des méthodes de type **Monte-Carlo** dans lesquelles des particules sont créées à chaque itération du calcul et se déplacent dans le maillage. Dans le cas des **maillages 3D non structurés**, le calcul des trajectoires et la détection des intersections entre les trajectoires et le maillage exigent beaucoup de temps de calcul. De plus, la distribution des particules au sein du maillage n'est pas uniforme et évolue au cours du calcul. Ainsi, certains sous-domaines réalisent une quantité importante de calculs tandis que d'autres attendent.

Arcane possède un mécanisme pour évaluer ce déséquilibre et le diminuer en échangeant automatiquement les entités du maillage entre les sous-domaines. Pour évaluer le déséquilibre, il faut mesurer le temps passé dans le calcul et celui passé dans les communications, et le distribuer par maille : la distribution peut être uniforme ou définie par une fonction. Arcane construit ainsi un graphe valué.

Un algorithme de repartitionnement va alors essayer de découper ce graphe en faisant en sorte que la somme des poids de chaque partie soit identique tout en minimisant le nombre d'arêtes coupées. Ensuite, Arcane recalcule à quel sous-domaine chaque entité va appartenir et échange entre les sous-domaines les entités nécessaires. Cette gestion est transparente pour le développeur de codes.

Au cours du temps, les équipes du CEA/DAM ont su adapter Arcane aux évolutions des architectures des supercalculateurs. Alors que sur Tera1, Tera10 et Tera100, le modèle de programmation prédominant reposait sur la mémoire distribuée, le calculateur Tera1000 met en œuvre des nœuds de calcul comprenant un très grand nombre de cœurs, rendant souhaitable l'utilisation de la mémoire partagée. Cela a conduit à réviser certains algorithmes pour qu'ils puissent utiliser efficacement à la fois la mémoire partagée et la mémoire distribuée. Par ailleurs, les nœuds de calcul peuvent se voir

adjoindre des accélérateurs, comme les cartes graphiques (**GPU**) ou les coprocesseurs dédiés. Ainsi, les architectures matérielles deviennent hétérogènes, rendant plus complexe l'obtention de la performance.

En vue de l'exascale, il est nécessaire de pouvoir rapidement s'adapter à une diversité et une complexité plus grandes des architectures de supercalculateurs. Cependant, pour des raisons de coûts et de délais, il n'est pas toujours possible de réécrire les codes de simulation pour chaque nouvelle architecture. Par ailleurs, ces codes deviennent plus riches en incorporant des modèles physiques et numériques de plus en plus précis nécessitant des ressources de calcul plus importantes. Il faut donc trouver un moyen d'adapter le plus simplement et le plus rapidement possible les codes à ces nouvelles technologies.

Pour cela, l'équipe Arcane travaille avec les physiciens et les mathématiciens dès la phase de spécification dans l'objectif de capturer les besoins des utilisateurs sous forme de modèles **UML** (Unified Modeling Language) ou **DSL** (Domain Specific Language). Ces modèles sont ensuite transformés par étapes successives pour finalement produire quasi automatiquement le code de simulation final. À terme, des mécanismes devraient être introduits pour générer des codes optimisés, dédiés à certaines architectures matérielles.

En conclusion, la plateforme Arcane a permis de structurer les codes de simulation en introduisant des types de base tels que les variables, les modules ou les points d'entrée (fonctions). Elle facilite également la maintenance des applications en arbitrant les interactions entre les différents modules de la simulation et en interdisant les couplages directs entre les modules. Ainsi, un module peut aisément être extrait d'un code de simulation pour être utilisé dans un autre. Il est également possible d'intégrer deux modules équivalents au sein du même simulateur pour effectuer des comparaisons entre différents algorithmes ou simplement pour comparer deux versions d'un même module lors de tests de **non-régression**. Grâce à ces notions et à la prise en charge des parties techniques du codage (parallélisme, gestion de la mémoire...), Arcane réduit la durée du cycle de développement des codes de simulation et permet d'aborder de nouvelles problématiques métier en structurant la complexité.

Arcane a permis de réaliser des simulations performantes sur un grand nombre de cœurs de calcul tout au long du programme Tera. Les codes de simulation n'ont jamais été affectés par

le passage d'un ordinateur Tera à un autre : les portages (compilation et exécution du même code sur une nouvelle machine) se sont faits instantanément et aucun module métier n'a dû être modifié. Les seules adaptations ayant dû être réalisées l'ont été au cœur de la plateforme. Au CEA/DAM comme à IFP Énergies Nouvelles, Arcane a permis de mettre en œuvre une nouvelle génération de simulateurs numériques fiables et performants.

4.3.2 Préparer une simulation numérique : maillage du modèle

Le maillage est déterminant pour la précision de la simulation, mais générer un bon maillage est délicat et chronophage, particulièrement en trois dimensions. L'utilisateur peut heureusement s'appuyer sur des outils performants d'aide à la conception et à la réalisation de maillage.

La préparation des données consiste pour partie à décrire de manière discrète une simplification de l'objet ou du domaine physique étudié. Selon les besoins des codes de simulation et les phénomènes étudiés, cette **discrétisation**, appelée maillage, et cette simplification sont plus ou moins difficiles à obtenir.

QUELQUES NOTIONS DE MAILLAGE

Pour s'exécuter, la plupart des codes de simulation nécessitent de discrétiser le domaine de calcul en éléments simples que l'on appelle usuellement des « mailles ». Leur union forme alors un maillage du domaine de calcul. On considère différents types de maillage selon leur dimension (2 ou 3 le plus souvent), leur type de mailles (triangulaires,

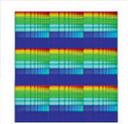
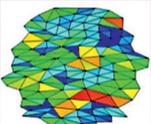
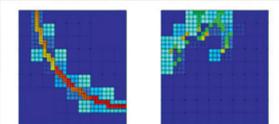
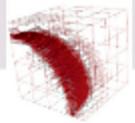
	Structuré	Non-Structuré	AMR
2D			
3D			

FIGURE 1

Différents types de maillages utilisés par les codes de simulation du CEA/DAM.

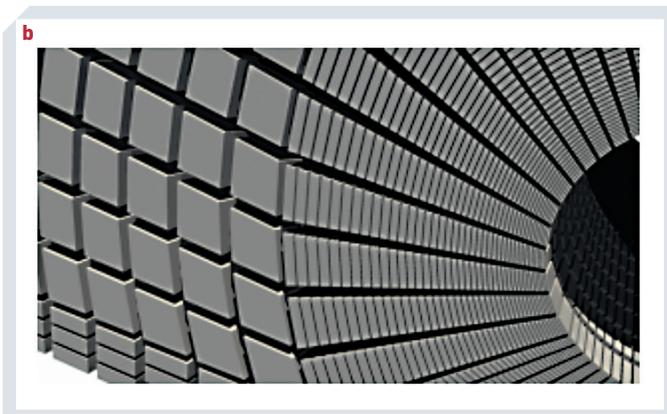
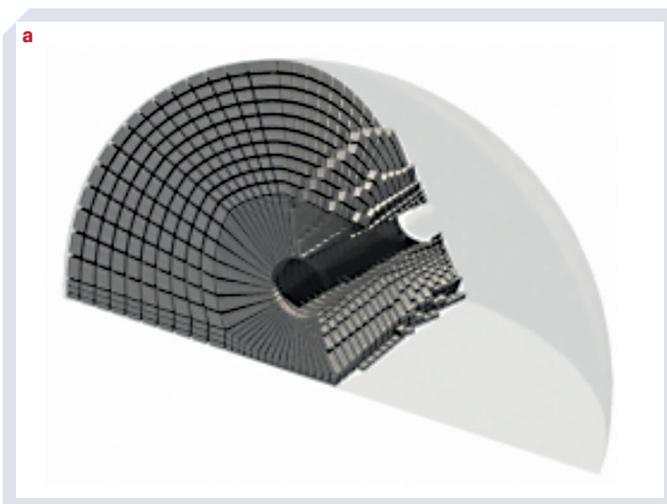


FIGURE 2

Exemple de **maillage structuré par blocs** avec, en (a), une vue partielle interne du maillage et, en (b), un zoom où l'épaisseur des mailles est contrôlée le long d'une surface cylindrique.

tétraédriques, quadrangulaires, hexaédriques, polyédriques, etc.) et leur structure. En général, on distinguera les **maillages structurés**, qui correspondent à des grilles régulières, les **maillages AMR**, qui sont des cas particuliers de maillages structurés où chaque cellule de la grille peut être localement subdivisée, les **maillages non structurés** qui sont des assemblages quelconques de cellules et enfin les **maillages structurés par blocs** qui sont des maillages non structurés grossiers où chaque maille est subdivisée en grille régulière. Différents exemples de ces maillages sont présentés sur les figures 1 et 2.

Ces différents types de maillages sont utiles à différents types de codes de simulation. Deux catégories de codes de simulation et leurs besoins relatifs au maillage peuvent ainsi être distingués en mécanique des milieux continus : les codes dits « eulériens », dans lesquels la position des nœuds du maillage n'est pas modifiée au cours de la simulation, et ceux dits « lagrangiens », où le maillage va évoluer au cours du temps, pour, par exemple, suivre un déplacement de matière et représenter un phénomène physique. Puisque le maillage reste inchangé, les premiers vont généralement utiliser des maillages structurés ou AMR où chaque maille pourra contenir plusieurs matériaux. Ces maillages sont faciles à construire. Lors de la simulation, les matériaux se déplaceront au cours du temps alors que les mailles resteront fixes. Les seconds utiliseront des maillages plus flexibles, structurés par blocs ou non structurés, et chaque maille ne contiendra cette fois qu'un unique matériau ; ce sont donc les mailles qui vont devoir se déformer pour suivre le déplacement des matériaux.

GÉNÉRATION DE MAILLAGES HEXAÉDRIQUES

La génération automatique de maillages structurés par blocs ou non structurés hexaédriques est actuellement un problème académique non encore totalement résolu. Or, de nombreux codes nécessitent de tels maillages. Pour assurer un bon comportement du code de simulation, les maillages hexaédriques produits doivent respecter un certain nombre de critères (figure 2) :

- ⊙ ils doivent être structurés par blocs avec des couches de mailles alignées le long des bords géométriques ;
- ⊙ il doit être possible de contrôler l'épaisseur des mailles et les mailles doivent le plus possible correspondre à un parallélepède rectangle ;
- ⊙ en aucun cas, les mailles ne doivent être non convexes sous peine d'erreurs numériques pour les codes.

L'exemple de la figure 3 illustre l'utilisation de la simulation numérique pour préparer une expérimentation avec le laser Mégajoule du CEA/DAM. La déformation du maillage lagrangien démontre le bénéfice de contrôler l'alignement et la taille des mailles aux interfaces et au bord du domaine pour éviter qu'elles ne deviennent de trop mauvaise qualité. On peut constater la large déformation des

mailles vertes correspondant à un matériau de la cavité au centre de laquelle se trouve la microbille (représentée en bleu sur la figure) contenant le matériau fusible.

PRÉPARATION DU MODÈLE D'ÉTUDE (CAO)

Pour être utilisable par un code de simulation, un maillage doit donc satisfaire certains critères de qualité. Selon la complexité du domaine de calcul, il peut être nécessaire de recourir à une étape de simplification géométrique du domaine afin de pouvoir le mailler correctement. C'est par exemple le cas pour un modèle de CAO, représentant un objet, issu du bureau d'études, où l'on réalise des plans de pièces mécaniques réelles. Un code de simulation peut difficilement s'accommoder de la complexité du modèle, ou alors en tenir compte fait diminuer le pas de temps de manière drastique, et augmente ainsi le temps de simulation au-delà des impératifs de l'étude.

Préparer un modèle pour une étude numérique est alors un processus itératif, dont le résultat est un compromis entre le modèle complet, les détails négligeables ou non pertinents pour le phénomène que l'on souhaite étudier, les capacités du code de calcul, les fonctionnalités

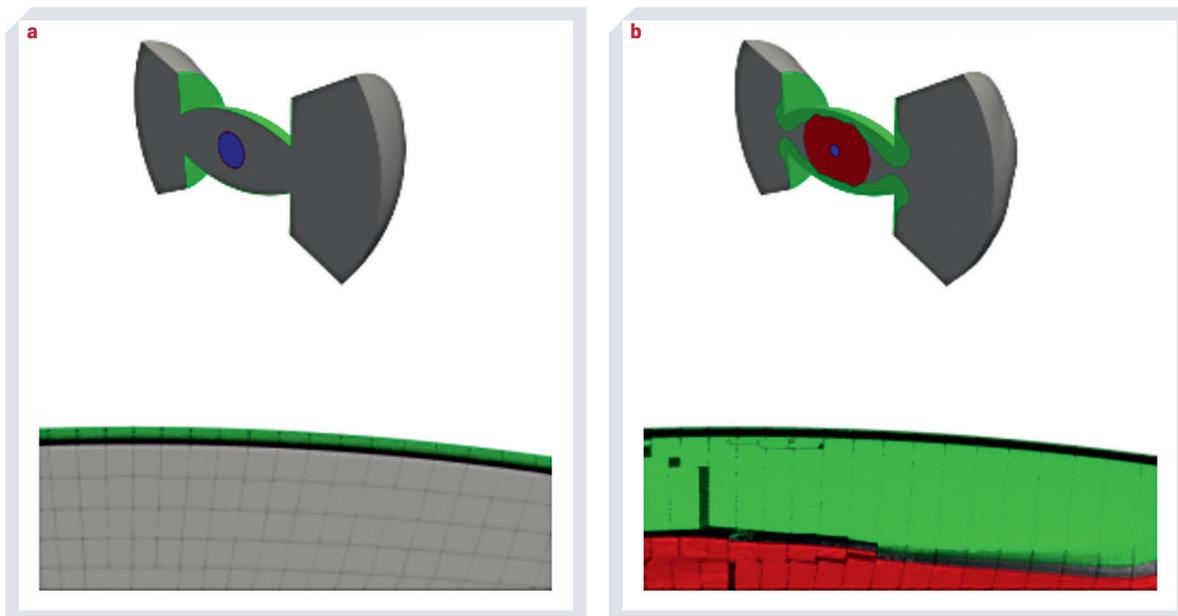


FIGURE 3

Illustration de la déformation d'un maillage lagrangien lors d'une simulation d'expérimentation laser. En (a), le maillage à l'instant initial; en (b) le maillage s'est largement déformé après plusieurs pas de temps.

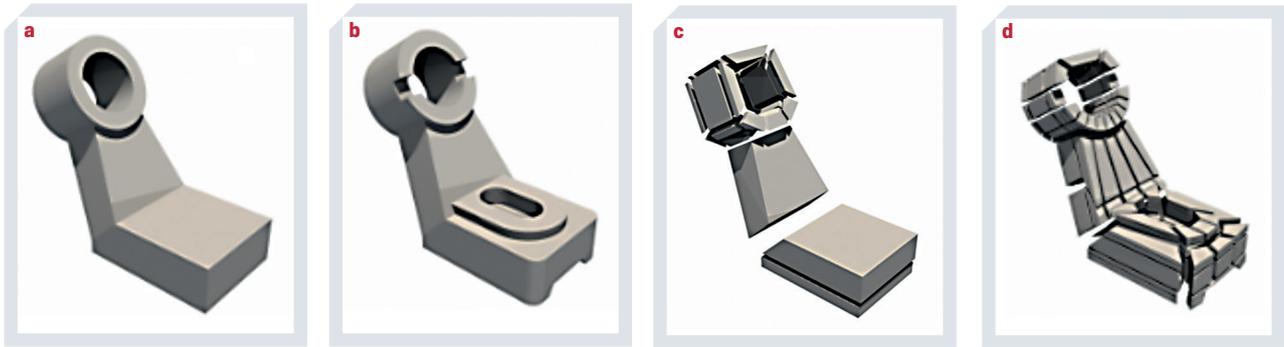


FIGURE 4

En (a) et (b) respectivement un modèle géométrique simple et une décomposition par blocs associés ; en (c), le modèle est enrichi de détails géométriques et la structure de blocs associée est grandement complexifiée en (d).

des algorithmes et outils de maillage à disposition, et bien sûr le temps alloué à l'étude. Sur l'exemple de la **figure 4** est représenté un modèle sur lequel un maillage hexaédrique structuré par blocs est généré : en (a) et (b) un modèle géométrique simple et une décomposition associée en 29 blocs ; en (c) le modèle est plus détaillé et la décomposition présentée en (d) comporte 174 blocs. Le temps ingénieur nécessaire pour créer le modèle géométrique en (c) et effectuer la découpe en blocs proposée en (d) augmente grandement par rapport au modèle simple.

C'est pourquoi une activité de R&D constante est maintenue. L'objectif est de réduire le temps ingénieur consacré à la simplification du domaine d'étude et à la génération de structures de blocs hexaédriques. Cela passe par la définition de nouveaux algorithmes et méthodes mêlant de nombreuses compétences scientifiques en informatique et en mathématiques appliquées : parallélisme, théorie des graphes, recherche opérationnelle et méthodes d'apprentissage, géométrie algorithmique, analyse numérique et algèbre linéaire, etc. ■

VERS L'OBTENTION DE SOLUTIONS DÉDIÉES LES PLUS AUTOMATIQUES POSSIBLES

Les maillages hexaédriques structurés par blocs sont les plus complexes à produire actuellement. Si on ajoute à cela le fait qu'il faille générer et manipuler ces maillages dans un contexte de parallélisme hybride, la tâche à réaliser est colossale. Et c'est à celle-ci que s'attellent les experts en maillage du CEA/DAM en proposant d'une part des outils interactifs dédiés aux besoins des codes afin de faciliter la génération des maillages, et en menant d'autre part une activité de recherche pour aller vers des solutions les plus automatiques possibles.

Les outils interactifs développés sont principalement dédiés à l'obtention de maillages hexaédriques structurés par blocs. Ils offrent des mécanismes de **scripting** pour automatiser les études, des fonctionnalités sur mesure, certaines indispensables et sans lesquelles les études ne pourraient s'effectuer, d'autres facilitant la génération de blocs à partir du modèle d'étude. Cette étape de génération est complexe et peut prendre plusieurs mois de temps ingénieur pour des domaines d'études complexes.

4.4 Traiter les données et visualiser les résultats

— L'interprétation des simulations numériques repose en grande partie sur la visualisation des résultats. Mais les volumes de données à traiter sont considérables. Il faut donc disposer d'outils performants de gestion de ces données (4.4.1). Il faut aussi que le physicien puisse visualiser les résultats sans se préoccuper de la complexité informatique du système sous-jacent (4.4.2). Enfin, les mouvements de données, coûteux en temps et en énergie, doivent être autant que possible limités (4.4.3) en réalisant l'essentiel des traitements *in situ*.

4.4.1 Modélisation et gestion des données pour la simulation numérique : la plateforme logicielle Hercule

Les volumes de données générées par les simulations, en particulier tridimensionnelles (3D), sont considérables et de natures diverses. Le logiciel Hercule propose une solution unifiée à ces difficultés.

Les codes de simulation numérique produisent des résultats qui représentent les approximations des valeurs des champs des variables principales des équations des phénomènes simulés. Ils permettent aux physiciens de « visualiser » et analyser ces phénomènes. Plus la simulation est complexe et précise, plus les résultats sont volumineux.

Dans de nombreuses situations d'intérêt pour le CEA/DAM, les phénomènes à simuler sont

trop complexes pour être appréhendés par un seul code de simulation en une seule exécution. Il est donc nécessaire de chaîner les exécutions en relançant plusieurs fois le même code ou en enchaînant des codes (l'exécution courante utilisant des données issues de l'exécution précédente) différents, chacun représentant une étape particulière d'un ensemble de phénomènes à simuler. À la fin de chaque maillon de la chaîne, des données sont produites et consommées par le maillon suivant (figure 1).

UN MODÈLE DE DONNÉES UNIQUE POUR L'ENSEMBLE DES CODES DE SIMULATION

Dans toutes ces situations, le **modèle de données** est au cœur d'une bonne communication entre un code « producteur » de données et un ou plusieurs codes « consommateurs » des données. Un modèle de données unique fédérant l'ensemble des représentations des variables utilisées par les codes de simulation facilite les dialogues entre codes et permet d'optimiser les chaînes de traitement. Ce modèle de données est un élément central du simulateur numérique.

Ces principes s'appliquent aussi bien au flux d'informations pour le post-traitement (visualisation et analyse), pour l'« **interCode** » (couplage entre deux codes). Seule la situation d'un flux de **protection/reprise** ne nécessite pas un grand formalisme car, dans ce cas-là, le code « consommateur » est aussi le code « producteur » de données.

Depuis près de vingt ans, la plateforme logicielle Hercule fournit les services pour modéliser, écrire et lire, manipuler et extraire les données de simulation des codes de calcul sur les supercalculateurs Tera.

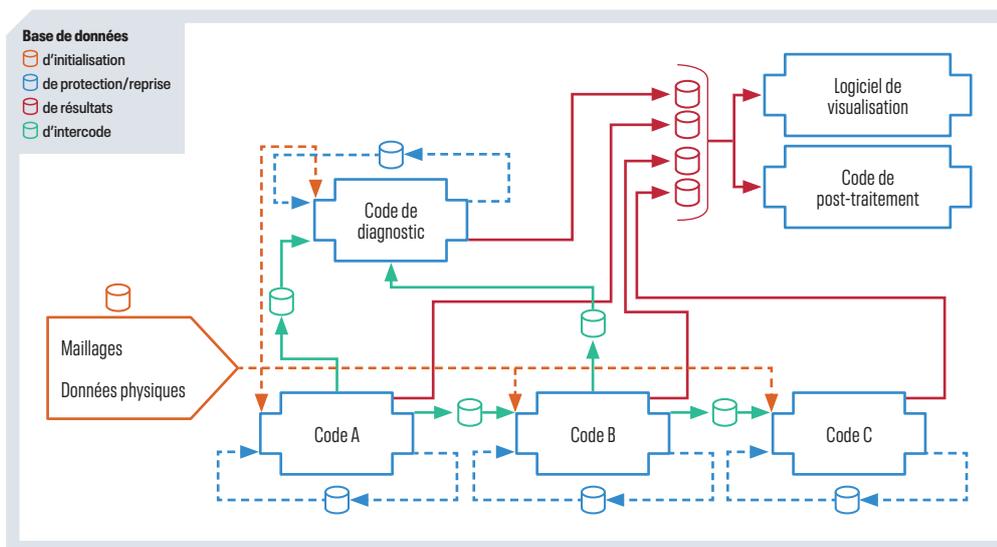


FIGURE 1
Schéma des flux de données d'une chaîne de simulation « type ».

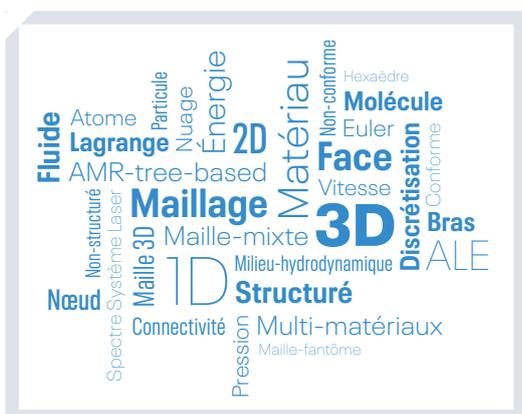


FIGURE 2
Diversité de concepts utilisés dans la modélisation.

Cette plateforme s'est enrichie au cours du temps et les différents domaines de modélisation des données sont à ce jour très variés. En effet, les caractéristiques des codes de simulation utilisant Hercules sont nombreuses et diversifiées (figure 2) : modélisation de maillages multiphysiques, multi-fluides ayant une **discrétisation** non structurée, structurée ou adaptative (**AMR**). Les domaines de modélisation se sont étendus récemment avec la prise en compte de rayons laser (pour la simulation des expériences sur le laser Mégajoule) ou d'atomes et molécules (pour des simulations de dynamique moléculaire).

LES ALGORITHMES DU BIG DATA APPLIQUÉS À LA SIMULATION NUMÉRIQUE

Dans le monde du *big data*, un des principes est d'effectuer des traitements simples sur des blocs de données indépendants puis d'agréger les résultats de ces traitements (p. ex. traitements de type **MapReduce**).

Selon les mêmes principes, appliqués aux simulations numériques, Hercules structure le code de simulation « producteur » pour que chaque processus décrive les données de son sous-ensemble propre du calcul. Les données sont augmentées d'informations plus globales pour appréhender la simulation dans sa globalité. Cette technique permet de situer les données locales dans le contexte général (figure 3) et reconstituer, si nécessaire, un sur-ensemble qui peut aller jusqu'à la simulation complète.

Hercules permet aussi d'appliquer des traitements sur les données des codes. En effet, ayant conservé le découpage du domaine de simulation du code « producteur » en sous-domaines, il est possible d'effectuer un traitement sur les données de simulation en appliquant un traitement aux données de chaque sous-domaine individuellement (*Map*) et d'effectuer une fusion (*Reduce*) des résultats intermédiaires pour produire les résultats du traitement demandé. Hercules propose ainsi de nombreux traitements pour ventiler, transformer, calculer, compléter et/ou extraire les informations demandées à partir des données du code (en lecture de fichiers ou en *in transit*).

Des optimisations internes à Hercules à base de caractérisation (**métadonnées**) des données produites (min/max des valeurs, min/max des coordonnées...) permettent de n'appliquer les traitements que sur les sous-domaines utiles au traitement demandé par l'utilisateur (figure 4).

IMPLÉMENTATION SUR SUPERCALCULATEURS PARALLÈLES

Les communications entre les codes de simulation ont été initialement développées *via* la lecture

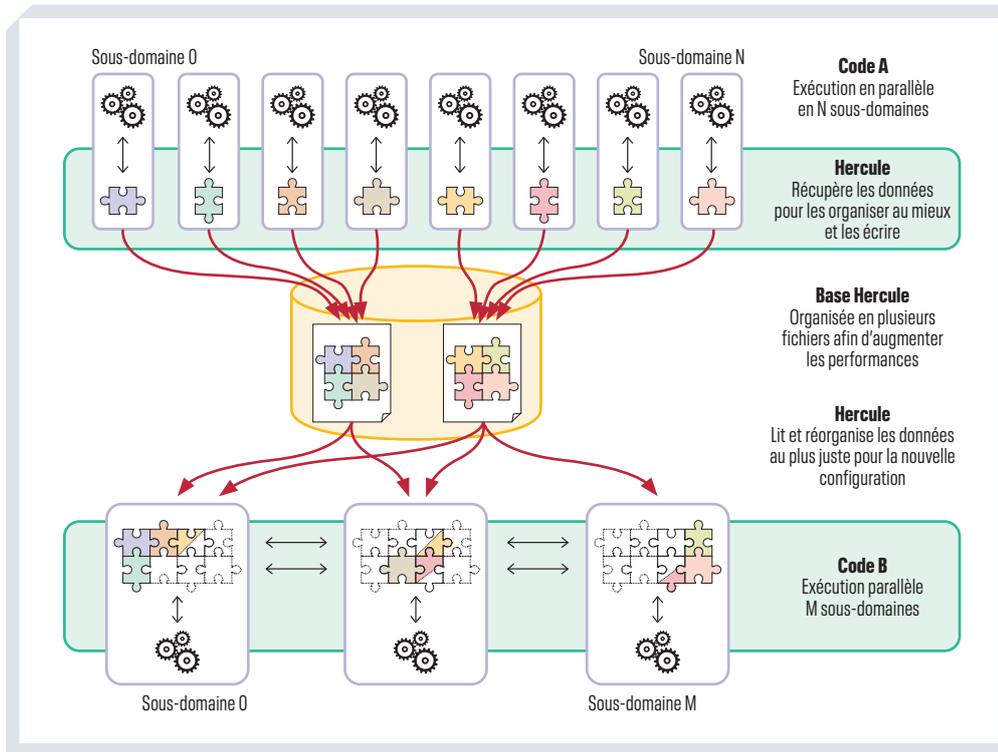


FIGURE 3
 Schéma de communication de données entre deux codes avec réorganisation des données adaptée à l'enchaînement de deux codes différents.

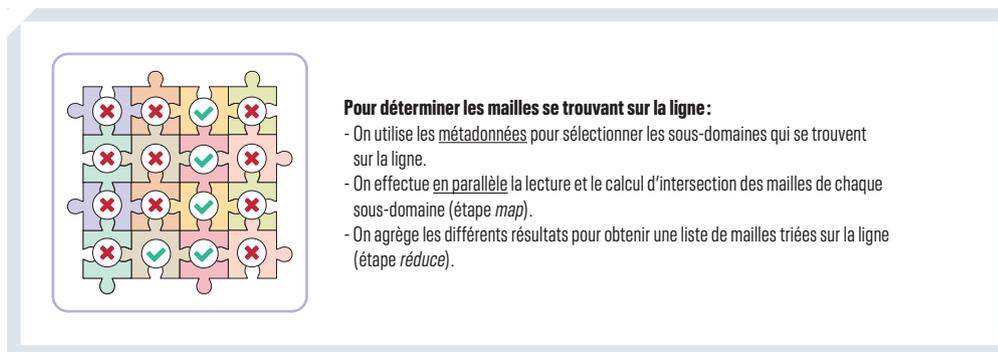


FIGURE 4
 Exemple de traitement optimal qui utilise les métadonnées pour réduire les lectures et le parallélisme pour accélérer les lectures et les traitements.

et l'écriture des bases de données sur fichiers. Mais le traitement des données peut aussi utiliser l'approche *in transit* ou *in situ* au sein même du supercalculateur, comme il est décrit dans le paragraphe consacré au logiciel **PaDaWAN**.

L'atteinte des performances en entrées/sorties passe par la prise en compte des contraintes matérielles et logicielles, comme les caractéristiques des **nœuds de calcul** et des systèmes de fichiers parallèles. Dans Hercule, le nombre de processus écrivains est généralement paramétré égal à 16 ou 32. Ainsi, 10 000 processus produisent dans cette configuration 312 fichiers simultanément (cumulant ainsi les performances de chaque fichier). Pour ne pas augmenter le nombre de fichiers au cours de la simulation, les processus écrivent toujours dans les mêmes fichiers jusqu'à ce que leur taille ne

soit plus compatible avec le système d'archivage hiérarchique **HPSS**.

La conception d'Hercule permet la **parallélisation** des traitements sur les mêmes principes que ceux du big data. Il propose des services d'extraction et recherche d'informations pour l'analyse des données, en utilisant le **parallélisme OpenMP** (intraprocessus) et/ou le parallélisme multiprocessus distribué.

Hercule continue de s'adapter aux nouvelles générations de supercalculateurs avec sa variante « en mémoire » (voir article consacré à PaDaWAN) qui anticipe les évolutions des supercalculateurs et celles des codes de calcul avec la prise en compte d'un accroissement d'un facteur 10 à 1000 de la taille des maillages et de 10 à 100 du nombre de processus informatiques utilisés par les simulations numériques.

4.4.2 Traiter et visualiser les résultats d'une simulation numérique

Une des finalités des simulations numériques est d'améliorer la compréhension des interactions entre des phénomènes physiques complexes. Pour cela, il est nécessaire de disposer de moyens performants et adaptés pour analyser, traiter et visualiser les résultats des simulations numériques. Dans le contexte des études menées au CEA/DAM, des contraintes structurantes sont à prendre en compte :

- ⊙ d'une part, les données à traiter sont de taille très importante et réparties sur les mémoires distribuées des supercalculateurs ;
- ⊙ d'autre part, elles peuvent être de natures diverses liées à la variété des codes de simulation du CEA/DAM et des différents types de **maillages** utilisés.

Ces contraintes orientent le développement des outils d'analyse interactifs, qui permettent d'analyser et de visualiser des résultats de simulation. Notons que pour les **études paramétrées**, ces outils interactifs disposent tous d'un mode « Script » permettant d'automatiser des enchaînements de commandes.

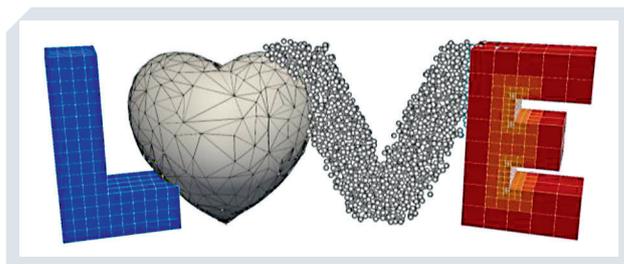


FIGURE 1

Des types de données manipulables avec Love: maillages hexaédriques structurés, tétraédriques non structurés, nuages de points et maillages AMR.

ANALYSER ET VISUALISER DES DONNÉES SPATIALES ET TEMPORELLES

Les données de grandeurs physiques à analyser et visualiser sont généralement portées par un **maillage**. Par exemple, une pression ou une température pourra être représentée par une valeur scalaire en chaque **maille**, alors qu'une vitesse ou un gradient de température pourra être représenté par un vecteur en chaque sommet du maillage.

Pour la plupart des simulations menées au CEA/DAM, ce maillage est « découpé » en **sous-domaines**, un sous-domaine correspondant alors à une partie du maillage et aux données associées. Chaque sous-domaine peut alors être traité par un processus différent, ce qui complique les traitements à effectuer. Un autre facteur de complexité est dû au fait que ce découpage peut varier au cours de la simulation : d'un pas de temps au suivant, une maille peut être déplacée d'un sous-domaine à un autre pour respecter des contraintes d'équilibrage de charge dictées par le code de simulation. De ce fait, la première caractéristique des données traitées est leur représentation spatiale en sous-domaines.

La seconde caractéristique est de nature temporelle. Il faut analyser le résultat de simulations instationnaires où les mailles vont se déformer et les données scalaires et vectorielles portées par le maillage vont évoluer au cours du temps, etc. Il faut donc être capable d'analyser non pas un maillage et ses données associées, mais une série de maillages et leurs données respectives.

Les données traitées sont donc ordonnées à la fois dans la dimension spatiale et dans la dimension temporelle, ce qui induit des analyses très différentes à gérer :

- ⊙ une analyse spatiale ou qualitative où l'on analyse et visualise à un instant donné de la simulation l'ensemble du domaine de calcul et donc du maillage. Cela permet par exemple de repérer des phénomènes physiques localisés en un point particulier ;
- ⊙ une analyse temporelle ou quantitative où l'on considère une partie du domaine de calcul et où l'on analyse l'évolution de données de cette partie au cours du temps. On regardera par exemple l'évolution au cours du temps de la température en une maille que l'on aura repérée comme intéressante.

Il est également possible de visualiser l'évolution du maillage complet au cours du temps. Ces deux types d'analyse prennent corps par

l'intermédiaire de deux logiciels développés au CEA/DAM : *Odace* pour l'analyse temporelle et *Love* pour l'analyse spatiale.

ANALYSE QUALITATIVE AVEC LE LOGICIEL ODACE

L'outil *Odace* repose sur l'utilisation et la proposition de scripts en **Python**. De nombreux modules Python **open source** comme le module numérique **numpy** et le module graphique **matplotlib** pour le tracé de courbes sont utilisés dans *Odace*. À l'aide de ses propres modules, *Odace* permet un accès simplifié et optimisé aux données temporelles et spatiales des codes du CEA/DAM en proposant de nombreux traitements prédéfinis et de nombreuses représentations graphiques pour l'analyse : 1D, 2D, 3D, nappes, histogrammes, statistiques, etc. En une unique commande, il est par exemple possible d'obtenir la courbe représentant l'évolution de la pression en une maille sur la durée complète de la simulation.

Puisque les données sont volumineuses et accessibles en mémoire distribuée sur les supercalculateurs du CEA/DAM, *Odace* fonctionne en mode **client/serveur**. Autrement dit, il s'exécute depuis une station de travail en étant connecté à des serveurs sur le calculateur où se trouvent les données calculées durant la simulation. Seules les données résultantes des calculs d'analyse sont rapatriées sur la station de travail.

ANALYSE QUANTITATIVE AVEC LE LOGICIEL LOVE

L'outil *Love* repose sur les logiciels open source VTK et Paraview et fournit un outil métier adapté aux besoins des physiciens du CEA/DAM. Totalement scripté et fonctionnant aussi bien sur

station qu'en mode client/serveur dans de nombreuses configurations, *Love* permet d'analyser et visualiser les données complètes d'une simulation pour un pas de temps donné. Il gère une grande variété de typologie de maillages (**figure 1**). Parmi les traitements spécifiques figure l'extraction d'**iso-contours** sur une donnée sur l'ensemble du domaine de calcul (**figures 2 et 3**) et la comparaison de cas, c'est-à-dire la possibilité de comparer les résultats de deux simulations à la fois spatialement et tem-

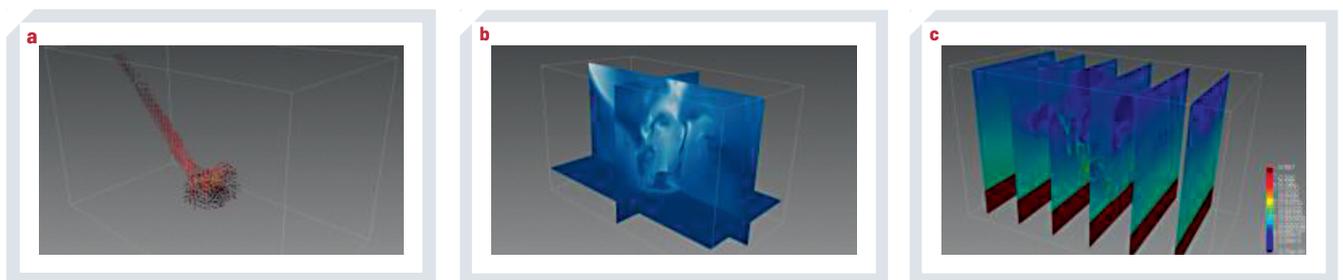
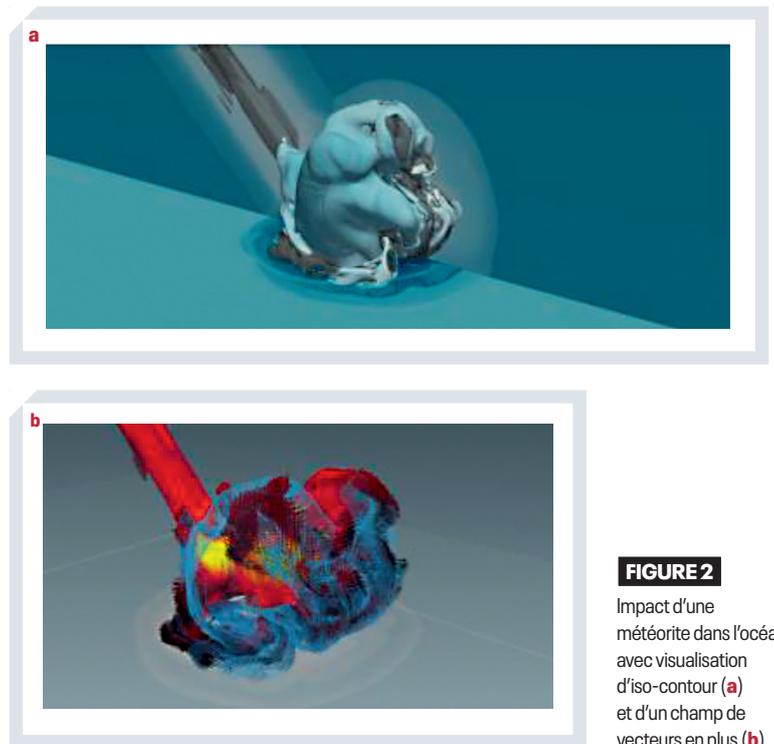


FIGURE 3

Impact d'une météorite dans l'océan avec visualisations permettant de comprendre les phénomènes : champ de vecteurs (**a**) pour appréhender les mouvements ; coupes croisées (**b**) et parallèles (**c**).

Le logiciel VTK et la contribution du CEA/DAM

Depuis plusieurs années, la bibliothèque open source VTK est utilisée par les équipes du CEA/DAM. Cela permet à la fois de disposer de nombreux traitements mis à disposition par la communauté scientifique et d'y contribuer activement. Une contribution majeure du CEA/DAM consiste en une solution efficace pour l'analyse et la visualisation de **maillage AMR** Tree based (**figure E1**) qui sont utilisés par de nombreux codes de simulation. Cette solution, appelée *Hyper-Tree Grid*, permet de visualiser interactivement des maillages pouvant comporter plusieurs dizaines de milliards de mailles sur peu de ressources de calcul en comparaison des autres approches disponibles.

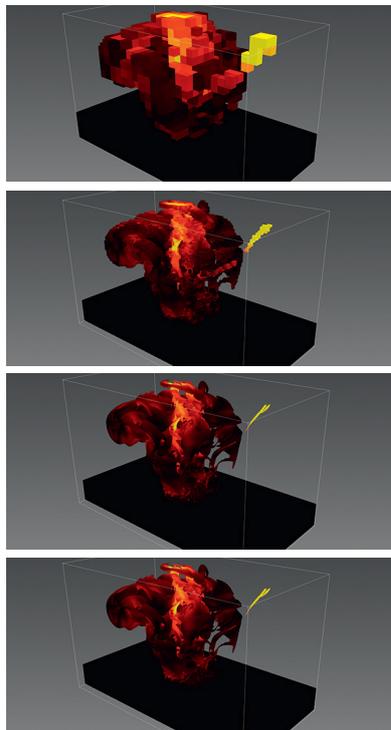


FIGURE E1

Visualisation de grilles AMR en ne retenant que les niveaux les plus grossiers en haut puis des niveaux intermédiaires pour arriver aux niveaux les plus fins en bas. C'est une facilité exploitable sur ce type de maillages afin d'accélérer la restitution en acceptant de perdre en résolution.

porellement. Love offre la possibilité de sélectionner un ensemble de mailles par application d'un plan de coupe ou bien encore de visualiser les mailles d'un sous-domaine de calcul.

DES OUTILS MÉTIERS POUR L'ANALYSE ET LA VISUALISATION

Outre la mise à disposition de fonctionnalités dédiées aux besoins des physiciens, les logiciels Odace et Love permettent de tirer parti des capacités des supercalculateurs Tera en masquant aux utilisateurs la complexité informatique sous-jacente : distribution des données, architecture client/serveur, accès aux bases de données, lecture de formats divers, déport de l'affichage graphique 3D, affichage sur un mur d'images (voir **encadré** «Le mur d'images Mirage») ou réalisation de films de résultats de simulation, fonctionnement sur tous types de machines, etc. Ce choix de simplification a pour but de permettre au physicien de s'abstraire des spécificités du matériel sur lequel ces logiciels fonctionnent. Ces mécanismes sont sans cesse enrichis et évoluent avec les évolutions matérielles du centre de calcul.

VERS DES PHÉNOMÈNES MULTI-PHYSIQUES DE PLUS EN PLUS COMPLEXES À ANALYSER

Les simulations 3D réalisées au CEA/DAM traitent de phénomènes multiphysiques de plus en plus complexes et de plus en plus gourmands en mémoire. Cela a un impact direct sur les outils d'analyse et de visualisation qui devront traiter interactivement des volumes de données de plus en plus importants et de plus en plus distribués en mémoire. Cette problématique est renforcée par le caractère 3D des données analysées et la perception humaine que l'on peut en avoir. Nous visualisons sur un écran 2D plan des données 3D, ce qui nécessite des solutions innovantes pour bien comprendre les phénomènes physiques. Cela passe par une R&D active en visualisation scientifique. Dans ce contexte, fournir des outils interactifs pour l'analyse et la visualisation sera un défi important à relever dans les années à venir.

Le mur d'images Mirage

Troisième configuration depuis 2002, le *mur d'images*, Mirage (**figure E2**), est une installation permettant des visualisations haute performance sur un écran de grande taille (5,5×3 m) ayant une résolution élevée (5088×2800 pixels, résolution proche des caractéristiques des projecteurs 8K des salles de cinéma les plus modernes, mais ici permettant d'être regardé de près). Un calculateur de visualisation est spécifiquement dédié à cette installation afin de permettre la visualisation interactive en 2D mais aussi en 3D (stéréo passive), ainsi que l'animation à grande résolution permettant d'embrasser du regard toute la géométrie, globale et détaillée, et suivre son comportement au cours du temps. Cette installation est particulièrement utilisée pour le travail collaboratif, la formation et les communications.

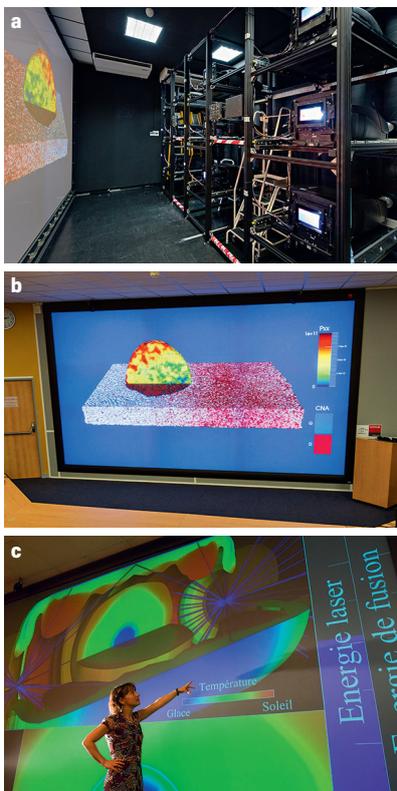


FIGURE E2

Côté coulisses (a), neuf projecteurs. À l'écran (b), le résultat d'un calcul simulant la transition graphite/diamant. Travail collaboratif (c) autour d'une simulation d'une expérience sur le laser Mégajoule.



INTERVIEW

J. DUBOIS

« Les données issues des simulations numériques.

Notre métier premier est de gérer la production, mais sans une part de recherche et développement, nous n'aurions pas pu relever les défis afin de maintenir une production de qualité et proposer des approches plus performantes à nos utilisateurs. Certaines de ces innovations ont fait l'objet de publications ou de mise à disposition dans la communauté. »

4.4.3

Analyse et visualisation des données *in situ* : PaDaWAn, une solution au traitement des volumes de données à l'échelle de l'exascale

L'approche traditionnelle pour le traitement des données produites par une simulation numérique consiste à les écrire sur disques au cours de la simulation. Une fois ces données écrites, elles sont ensuite lues pour être analysées (**figure 1**). L'évolution des architectures des supercalculateurs en vue d'atteindre des capacités de calcul **exaflopiques** rend cette approche de *post-traitement* peu efficace pour les simulations produisant de gros volumes de don-

nées. En effet, les évolutions technologiques apportent une augmentation de la puissance de calcul beaucoup plus rapide que celle des capacités des systèmes de stockage sur disques. Par conséquent, il devient possible de produire des données beaucoup plus rapidement qu'il n'est possible de les stocker. De plus, les codes de simulation devenant de plus en plus sophistiqués, ils produisent généralement plus de données que la génération de codes précédente. Cette augmentation massive du volume de données produites remet en cause les approches traditionnelles de traitement des données issues des simulations.

Une alternative consiste à s'affranchir de l'utilisation de ce stockage intermédiaire sur disques en effectuant le traitement des données « à la volée ». Dans cette approche, que l'on qualifiera d'*in situ* au sens large, le traitement est exécuté de façon concurrente avec la simulation et sur des données résidant encore en mémoire, c'est-à-dire avant leur écriture éventuelle sur disque. Les avantages du traitement de données *in situ* sont multiples. En intervenant avant la phase d'écriture, cela permet d'envisager une haute résolution des analyses en s'accommodant des flux de données importants produits par la simulation. Cette approche permet également de réduire le volume de données en appliquant des traitements de compression, de filtrage ou d'extraction de caractéristiques par-

ticulières en vue de leur utilisation future. Elle offre aussi la possibilité de visualiser l'état courant de la simulation sans attendre les phases d'écriture. L'utilisateur peut ainsi vérifier quasiment en temps réel si la simulation converge vers un état désiré. Dans le cas contraire, avec cette information, l'utilisateur peut décider d'arrêter la simulation, économisant ainsi du temps de calcul.

Les traitements *in situ* peuvent être effectués selon deux mécanismes. Pour le premier, dit synchrone, les traitements sont exécutés sur les mêmes unités de calcul que la simulation et celle-ci est interrompue le temps de la réalisation des traitements. Le second mécanisme est asynchrone, et consiste à envoyer les données à traiter sur des unités distinctes. Les traitements sont alors exécutés en parallèle de la simulation. Ce deuxième mécanisme est également appelé *in transit*, car des traitements peuvent être appliqués sur les données entre leur émission par le code de simulation et leur stockage. C'est ce deuxième mécanisme, qui offre plus de flexibilité et qui minimise l'intervention dans les codes de simulation, que propose l'infrastructure de PaDaWAn (figure 1).

Élaboré par le centre DAM Île-de-France du CEA/DAM, PaDaWAn (Parallel Data Workflow for ANalysis) est une infrastructure logicielle accélérant les échanges de données entre applications d'un **workflow**. Ces applications

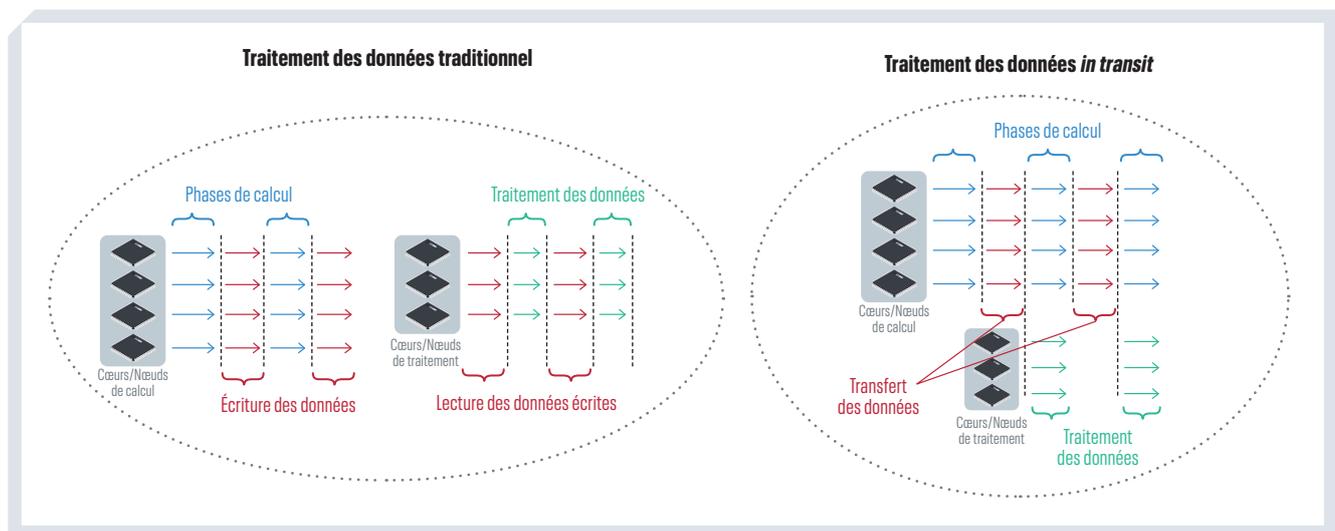
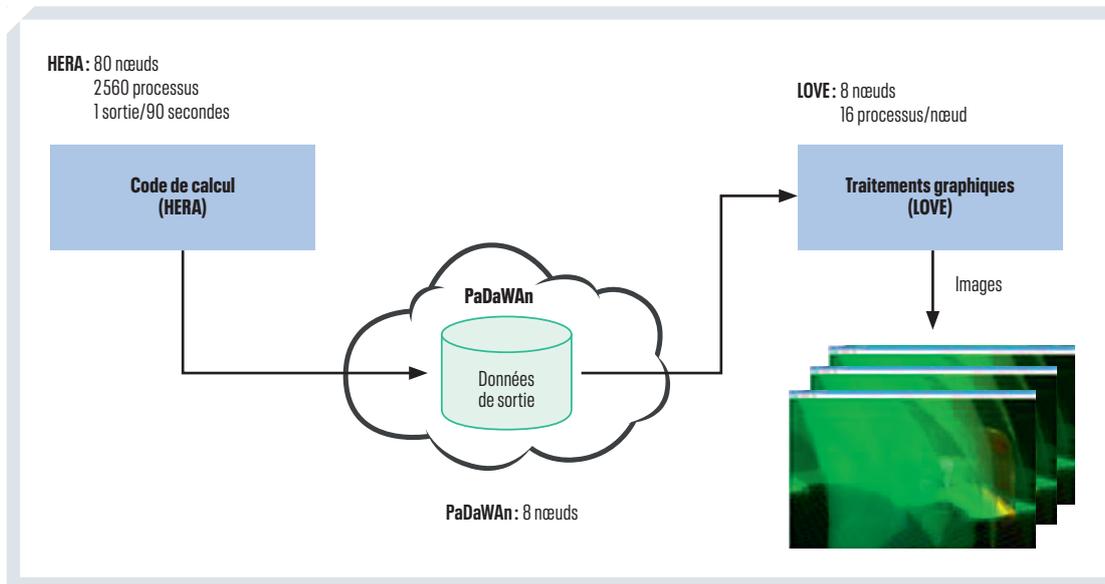


FIGURE 1

Principe de fonctionnement du logiciel PaDaWAn.

**FIGURE 2**

Exemple d'utilisation de PaDaWAn.

étant exécutées en parallèle, on parle alors de *workflow in situ*. Cette infrastructure repose sur l'utilisation du réseau d'interconnexion rapide des supercalculateurs pour transférer les données intermédiaires, et sur l'utilisation d'unités de calcul dédiées pour stocker temporairement ces données en mémoire, appelées aussi **unités de staging**. Ces unités font office de tampon entre les applications *productrices* de données (p. ex. une simulation numérique), et les applications *consommatrices* (un traitement). Cette infrastructure assure ainsi un découplage entre les applications (chacune utilisant ses propres ressources et chacune avançant à son rythme) tout en accélérant les transferts de données en s'affranchissant des aléas et des limitations des systèmes de stockage sur disques.

Par le biais de l'infrastructure PaDaWAn et de son interface de programmation (**API**), il est possible d'exécuter tous types d'opérations sur les données en cours de production par la simulation : une analyse quantitative, un filtrage, un traitement graphique, une écriture asynchrone sur disque, etc. Il est également possible d'enchaîner plusieurs simulations. L'interface de programmation s'appuie sur l'interface d'*entrées/sorties* (E/S ou I/O en anglais) existante dans les codes de simulation pour « intercepter » les données, limitant ainsi les modifications nécessaires à apporter et facilitant l'utilisation de l'infrastructure.

L'utilisation d'une infrastructure telle que PaDaWAn permet de bénéficier des avantages de l'approche *in situ* présentés dans le paragraphe précédent. En levant les contraintes liées au stockage des données intermédiaires et en accélérant les transferts de données, il est ainsi possible d'augmenter la fréquence de production des données par le code de simulation et par conséquent d'affiner la résolution des analyses. L'empreinte finale des données stockées sur disques est sensiblement réduite puisque les données intermédiaires ne sont plus systématiquement sauvegardées du fait de l'application de traitements à la volée. Cette infrastructure permet également d'accélérer les études scientifiques en exécutant, en parallèle des simulations, les traitements sur les données produites.

Un exemple d'utilisation de PaDaWAn a été la réalisation *in situ* d'un film de simulation de la propagation d'ondes de choc obtenu à l'aide du code de calcul **AMR** hydrodynamique HERA. L'infrastructure de PaDaWAn a été utilisée pour faire transiter les données entre le code de calcul et l'outil de visualisation Love. La simulation a nécessité 80 nœuds de calcul, et l'outil de visualisation, 8 nœuds de calcul pour produire les images. Initialement, la production des images était réalisée après l'écriture de toutes les données sur disque, ce qui participait à la saturation du système de stockage et entraînait la manipulation de fichiers volumineux.

L'utilisation de PaDaWAN (**figure 2**) a permis de limiter l'empreinte permanente sur disque à 1,5 GB, c'est-à-dire aux images produites, contre les 50 TB de données écrites initialement par le code. C'est plus de 30 000 fois moins ! Pour obtenir ce gain, seulement 10 % de ressources de calcul supplémentaires ont été nécessaires (8 unités de *staging* pour les données, soit un total de 1 TB de mémoire). Enfin, les traitements graphiques ayant été réalisés en parallèle du code de simulation, la réalisation du film a été accélérée par rapport à l'approche initiale.

PaDaWAN permet ainsi d'accélérer les études scientifiques et de réduire significativement la quantité de données écrites sur disques par une simulation dans le contexte du couplage de codes. L'évolution de la complexité des simulations conduisant à la construction de chaînages de codes de plus en plus complexes, l'un des prochains objectifs de PaDaWAN est la gestion de ces workflows complexes. Enfin, PaDaWAN va évoluer pour tirer parti des architectures hétérogènes de supercalculateurs en permettant par exemple la réalisation de couplages mélangeant des unités de calcul, des unités à large mémoire et des **GPU**. ■

CHAPITRE 5

AVANCÉES EN SIMULATION ET SUPER- CALCULATEURS

Les avancées en simulation numérique vont de pair avec les évolutions de la puissance, de l'architecture et des modèles de programmation des supercalculateurs. Les deux exemples suivants illustrent la démarche des ingénieurs et chercheurs du CEA/DAM.

5.1 L'apport du HPC à la simulation des matériaux à l'échelle microscopique

— Les fonctions d'ondes des électrons gouvernent la structure et les propriétés de la matière. Leur détermination demande des puissances de calcul considérables.

Si l'augmentation de la puissance des ordinateurs a eu un impact considérable sur les capacités de simulation numérique à l'échelle macroscopique, elle a aussi permis de franchir une barrière dans le domaine de la modélisation de la physique fondamentale sous-jacente. Une bonne connaissance des phénomènes physiques qui ont lieu à l'échelle microscopique est en effet nécessaire pour mieux comprendre, et donc mieux prévoir, le comportement macroscopique des systèmes complexes étudiés au CEA/DAM. Avec les supercalculateurs parallèles, la matière n'a plus été seulement représentée comme un milieu continu, mais l'analyse de son comportement a pu être menée de manière plus précise par une approche microscopique *via* sa représentation comme un assemblage d'atomes.

Dans les codes de calcul du programme Simulation, la connaissance des phénomènes fondamentaux se traduit le plus souvent sous la forme de « données microscopiques ». Celles-ci rendent compte des propriétés de structure des éléments constitutifs de la matière (noyaux, ions, molécules...) et de leurs propriétés d'interaction entre eux et avec le rayonnement électromagnétique (comme les données de réactions nucléaires et les opacités). Plusieurs branches de la physique sont ainsi concernées : la physique nucléaire (modélisation à l'échelle du nucléon de la structure du noyau), la physique atomique et la physique de la matière, de l'état condensé au plasma (modélisation à l'échelle de l'électron

de la structure de l'atome ou de la molécule). Si ces différents domaines ont largement bénéficié et bénéficient encore de l'augmentation de la puissance de calcul, seuls les progrès réalisés en simulation de physique microscopique et en couplages d'échelle avec les machines Tera sur quelques exemples de modélisation des matériaux sont illustrés ici.

Les simulations des matériaux à l'échelle microscopique ont en effet pu bénéficier dès Tera1 de l'augmentation de la puissance de calcul, grâce au développement par les équipes du CEA/DAM de la version **HPC** du code de structure électronique **ABINIT**. Et ce dans l'objectif de modéliser un solide par un **calcul *ab initio***, c'est-à-dire sans paramètre ajustable d'origine expérimentale.

En physique de la matière condensée, la modélisation se fait à l'échelle des électrons, qui sont régis par la mécanique quantique. Décrire la matière à partir des premiers principes n'a été possible que grâce à une écriture permettant de transformer le problème quantique à N corps, impossible à résoudre exactement au-delà de l'hydrogène, en un problème plus simple ne faisant intervenir que la densité électronique totale du système. Cette réécriture, connue sous le nom de théorie de la fonctionnelle de la densité (**DFT**), a valu le prix Nobel à son concepteur Walter Kohn. Si cette théorie est exacte, la résolution pratique des équations de la DFT nécessite des approximations, qui sont prises en défaut lorsque le système présente de fortes variations de densité électronique. Cette approche a ainsi

engendré de nombreuses méthodes conduisant à une description de plus en plus précise de la matière, mais de plus en plus exigeante en temps calcul.

La DFT et les méthodes dérivées s'appuient sur la résolution approchée d'un système d'équations couplées. Les avancées de la modélisation et l'accès à Tera1 ont permis, grâce à une réécriture astucieuse des algorithmes les plus coûteux et l'utilisation du **parallélisme**, de simuler un ensemble d'une dizaine de particules. Dans le domaine des matériaux, de premières simulations microscopiques ont ainsi été réalisées sur Tera1 pour étudier les vibrations collectives (appelées phonons) d'un ensemble d'atomes en interaction, qui constituent un ingrédient fondamental pour la modélisation d'un solide. Sur une telle **boîte de calcul**, les ordinateurs précédents ne permettaient pas d'atteindre la **convergence** nécessaire sur les fréquences propres et les vecteurs propres associés (spectre de phonons du solide). Le spectre de phonons d'un solide permet en particulier de déterminer certains des paramètres de l'équation d'état du matériau, qui fournit l'état thermodynamique du matériau à la simulation à l'échelle macroscopique. Plus précisément, l'obtention sur Tera1 d'un spectre de phonons du tantale pour une pression donnée nécessitait environ 7500 heures de temps **CPU** sur 128 **processeurs**. Les spectres de phonons calculés à différentes pressions (jusqu'à 10 mégabars) avaient ainsi permis d'obtenir des données thermodynamiques du matériau dans le domaine haute pression où il n'y a pas de référence expérimentale. À pression ambiante, les résultats obtenus étaient très voisins des mesures disponibles.

Cette procédure permettait de simuler de nombreux solides, mais ne pouvait toutefois pas être appliquée à tous les matériaux, en particulier aux éléments comme l'uranium ou le plutonium, dont la structure électronique très complexe nécessitait des développements théoriques supplémentaires, pour obtenir une modélisation à la fois suffisamment précise de la densité électronique tout en restant compatible avec la puissance de calcul disponible. De nouvelles approches théoriques permettant de prendre en compte l'interaction entre électrons ont été développées et implémentées dans le code ABINIT à cette période.

Les méthodes *ab initio* étaient par ailleurs toujours limitées pour des raisons de coût calcul à des simulations de « boîtes » de quelques dizaines d'atomes. En matière de développe-

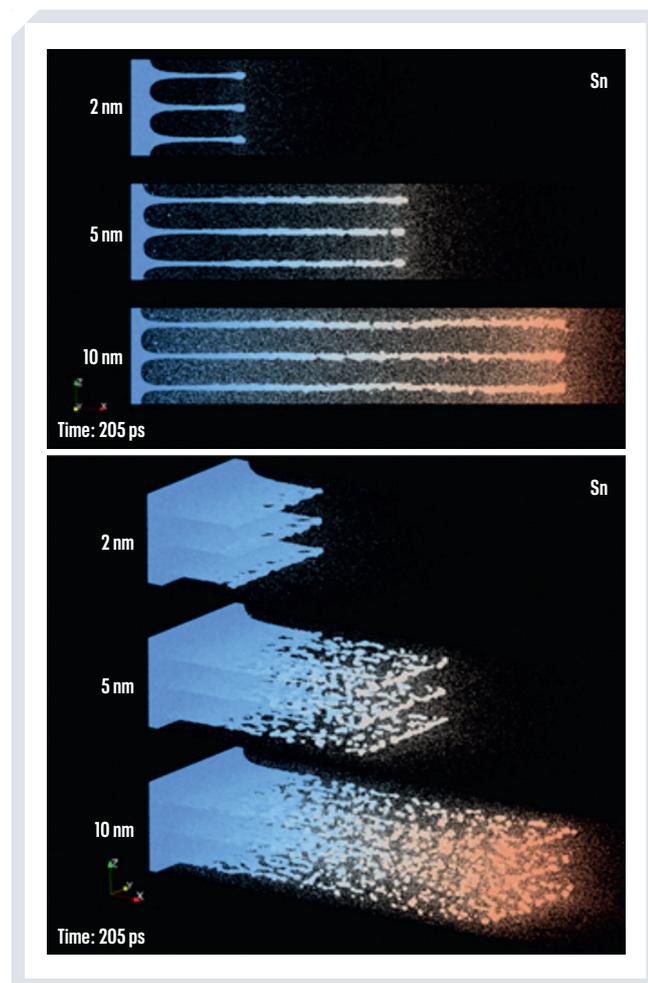


FIGURE 1

Étude de l'éjection de matière (étain) sous choc sur Tera100 : simulation de dynamique moléculaire Stamp (l'une ou l'autre).

ment numérique, l'implémentation dans le code ABINIT de la méthode Projector Augmented Wave a constitué une étape décisive pour les calculs de structure électronique. Cette méthode de calcul permet en effet de traiter des systèmes comprenant beaucoup d'atomes, et en particulier des métaux lourds, pour un coût calcul limité. En même temps, un effort important a porté sur la **parallélisation** du code, en particulier grâce à un nouvel algorithme de résolution d'équations aux valeurs propres. Cela aboutira à un facteur d'accélération quasi linéaire jusqu'à 1000 processeurs. Ces progrès sur le code ABINIT et l'accès au calculateur Tera10 vont permettre d'utiliser ces techniques *ab initio* pour une simulation systématique non seulement des propriétés de nombreux corps purs, mais aussi de leurs composés, comme les oxydes et certains alliages.

Par la suite, le développement dans ABINIT de la méthode du champ moyen dynamique **DMFT**, qui se rapproche de la résolution exacte du problème à N corps initial, permettra de réaliser sur Tera100 des simulations de référence pour des matériaux aussi complexes que le plutonium. Cette méthode est par contre trop coûteuse numériquement pour être exploitée de manière intensive sur Tera100.

Parallèlement au développement de ces techniques, les années 2000 à 2010 ont vu le développement de la simulation multi-échelle des matériaux. Couplant pratiquement toutes les techniques de simulation de physique et mécanique du solide disponibles au CEA/DAM, ces méthodes permettent d'appréhender des phénomènes aussi complexes que le comportement mécanique d'un métal ou la détonation d'un matériau énergétique. La dynamique moléculaire constitue en effet une alternative de qualité pour simuler l'évolution dynamique de systèmes complexes inaccessibles aux calculs *ab initio*, dont les boîtes de calcul restent de l'ordre du nanomètre. Cette technique de simulation construit les trajectoires des atomes et des molécules du système en appliquant le principe fondamental de la dynamique, à partir des forces qui s'exercent sur chaque particule et qui sont calculées *via* un **potentiel d'interaction atomique**. La physique du système étudié (et éventuellement sa chimie) est ainsi renvoyée dans ce potentiel d'interaction dont la mise au point est un facteur clé pour la qualité de la

simulation. Les potentiels sont basés sur une fonction d'énergie potentielle plus ou moins empirique, dont les paramètres sont ajustés à l'aide de données expérimentales microscopiques et de calculs *ab initio*. Notons que, compte tenu du domaine thermodynamique d'intérêt au CEA/DAM, les paramètres disponibles dans la littérature sont le plus souvent inadaptés. La montée en puissance des calculs ABINIT permet ainsi d'enrichir les bases de données matériaux nécessaires à la construction de ces potentiels d'interaction.

Par ailleurs, avec le code Stamp développé par les équipes du CEA/DAM, des simulations de dynamique moléculaire de plusieurs dizaines de millions d'atomes ont été réalisées sur la machine Tera10. Par exemple, de premières simulations ont été effectuées pour étudier la création de microjets à la surface d'un échantillon métallique soumis à une onde de choc. Cette étude d'un phénomène physique fondamental pour le CEA/DAM a été poursuivie sur la machine Tera100, sur laquelle l'effet conjugué de l'accroissement de la puissance de calcul et de la parallélisation à plus grande échelle permet d'augmenter la boîte de calcul et le temps de simulation du phénomène physique de manière à réaliser des simulations de plus en plus proches des conditions expérimentales (figure 1). Le passage de la machine Tera100 à la machine Tera1000, et plus particulièrement à la partition Tera1000-2, a été un défi important pour les codes matériaux. Cette nouvelle architecture de machine a conduit à une réécriture importante du code ABINIT et une refonte totale du code Stamp, qui a donné naissance au code ExaStamp. Ce nouveau code est actuellement enrichi par de nouveaux outils de visualisation et d'analyse des masses de données produites, cette activité étant en effet très consommatrice en temps ingénieur et en temps calcul. Ces différentes avancées permettent de réaliser et d'exploiter des simulations de dynamique molé-

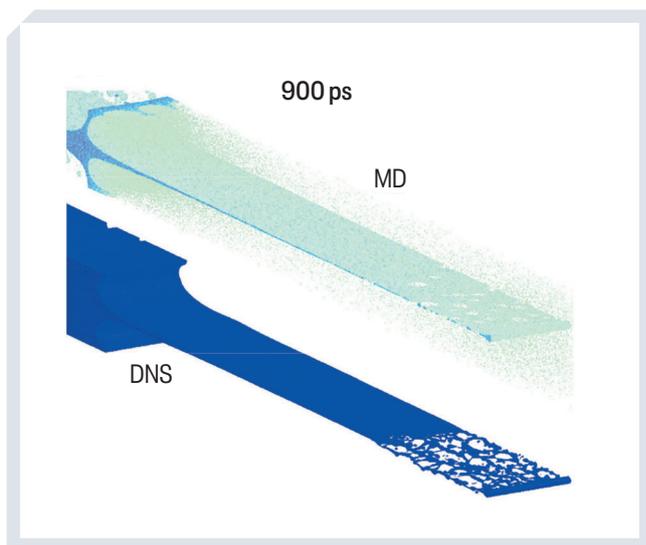


FIGURE 2

Simulation de l'éjection de matière sous choc sur Tera1000-2 : comparaison des approches atomistique (en haut, code de dynamique moléculaire ExaStamp) et continue (en bas, code hydrodynamique du CEA/DAM) un milliardième de seconde (nanoseconde) après le début de l'éjection.



Laurent Colombet, le maillon entre les physiciens et les calculateurs

Adapter les codes de calcul pour que les scientifiques mènent efficacement leurs simulations : tel est le métier de Laurent Colombet. Une tâche passionnante par sa diversité et l'interaction permanente avec les meilleurs chercheurs.

— C'est bien beau d'avoir de magnifiques supercalculateurs, puissants et au top de la technique, mais si les logiciels ne savent pas les exploiter au mieux, ce n'est pas très utile. C'est là qu'intervient Laurent Colombet, ingénieur-chercheur en calcul haute performance. « J'étudie les codes pour qu'ils s'adaptent le mieux à ces nouvelles machines, indique-t-il. Je fais le lien entre les souhaits des physiciens en matière de simulation, et ce que peut rendre un code de simulation, tout en diminuant les temps de calcul. »

Reproduire numériquement les expériences

Par exemple, comprendre les propriétés mécaniques d'un matériau (sa viscoélasticité, c'est-à-dire la manière dont il se tord et casse) met en jeu des phénomènes physiques multiples, à différentes échelles. À partir des atomes qui composent le matériau, il faut arriver à remonter à leur comportement. Il faut donc faire des hypothèses, des calculs (en un temps raisonnable), et regarder si les premiers résultats confirment ou infirment les hypothèses. « L'idée est de reproduire au mieux les expériences, explique Laurent Colombet. Nous devons donc comprendre les hypothèses des physiciens et des mathématiciens. De l'autre côté, nous devons aussi discuter avec les informaticiens qui parallélisent les codes, ou les spécialistes du traitement des données. »

Laurent Colombet a connu les tout débuts du calcul massivement parallèle. Après sa thèse, l'embauche a été rapide car le CEA avait un grand besoin d'experts dans ce domaine qui émergeait. « Il s'agissait de concevoir les premiers codes massivement parallèles, faire en sorte que le programme se découpe en tâches, que les processeurs travaillent en parallèle et collaborent, se souvient-il. Les machines n'avaient que quelques centaines de processeurs, mais pour l'époque, c'était beaucoup. »

Attirer les jeunes talents

Laurent Colombet aimerait transmettre sa passion aux jeunes. Or, les étudiants en informatique se dirigent surtout vers les domaines à la mode : les mathématiques financières il y a quelques années, la sécurité informatique ensuite, l'intelligence artificielle aujourd'hui. Et certaines entreprises se lancent dans le calcul massivement parallèle, attirant les étudiants avec des espèces sonnantes et trébuchantes. « Je donne des cours, je vais à la rencontre des étudiants pour leur montrer qu'il reste plein de belles choses à découvrir, que l'informatique n'est pas refermée sur elle-même, s'enthousiasme le chercheur. Nous devons expliquer notre travail et son intérêt concret pour tous les jours : des technologies que nous mettons au point se retrouvent dans nos téléphones. C'est un superbe métier pour les gens curieux et ouverts. Et si varié : j'ai travaillé sur l'imagerie, la science des matériaux, qui sait sur quel sujet je travaillerai demain ? »

Vers un jumeau numérique

Parmi les défis qui le passionnent : ajouter de plus en plus de **machine learning** pour améliorer les programmes. L'idée est de traiter in situ le code pour qu'il donne le plus vite possible des données exploitables par les physiciens. Autre défi, faire le lien entre les différentes échelles de simulation, pour aller jusqu'au « jumeau numérique » : l'équivalent informatique d'un objet. Pour un moteur, par exemple, il s'agit de construire numériquement un moteur purement virtuel sur lequel on simule tout : que se passe-t-il si on dérègle, le changement de matériaux des soupapes joue-t-il sur l'usure, etc. ? « Pour atteindre cela, il faudra encore augmenter la puissance de calcul, ainsi que les modèles physiques, estime Laurent Colombet. Mais cela commence à poindre, et devrait arriver dans les cinq à dix ans pour les systèmes simples comme le mécanisme d'un robot envoyé sur Mars. » Encore de la belle science en perspective !

culaire avec ExaStamp de plusieurs dizaines de milliards d'atomes tout en augmentant le temps de restitution du phénomène physique. Reprenons l'exemple de l'étude par dynamique moléculaire du comportement sous choc d'une plaque d'étain effectuée précédemment avec le code Stamp sur Tera100 et illustrée **figure 1** ; les simulations ExaStamp réalisées sur Tera1000 peuvent maintenant être comparées à des simulations réalisées à l'échelle macroscopique avec un code hydrodynamique (**figure 2** et voir aussi l'article « Comprendre et prévoir par la simulation »), fournissant ainsi un premier recouvrement d'échelles, entre la description à l'échelle moléculaire et celle à l'échelle des milieux continus. L'accès à la puissance de calcul de Tera1000 a de plus autorisé la mise en œuvre de potentiels d'interaction atomique restituant mieux la physique mais dont le coût calcul était prohibitif sur Tera100.

Toutefois, la simulation de certains phénomènes physiques reste difficile, comme le comportement sous choc en dynamique moléculaire de certains métaux nécessitant une description à l'échelle électronique. Pour pouvoir traiter ces phénomènes à la bonne échelle (temps caractéristique, taille de l'échantillon de matière), il est en revanche indispensable de s'affranchir de la description quantique et de travailler au moins à l'échelle atomique, soit en dynamique moléculaire classique. Aussi, des travaux théoriques sont-ils actuellement menés pour améliorer la qualité du potentiel d'interaction atomique, ingrédient fondamental de la simulation de dynamique moléculaire. Des techniques d'apprentissage sont en particulier utilisées pour obtenir un potentiel numérique « intelligent » au sein d'ExaStamp à partir d'une base de données microscopiques issues de calculs ABINIT utilisant la méthode DMFT, jusqu'alors inexploitable en production. Ce couplage d'échelles est permis par l'exploitation intensive et judicieuse des moyens de calcul.

L'adaptation constante du code ABINIT aux caractéristiques des supercalculateurs a permis d'apporter une offre logicielle améliorée fonction de la montée en puissance des machines, de Tera1 à Tera1000, et de produire des données de base de plus en plus précises utilisables dans les codes aux échelles supérieures, que ce soit des codes de dynamique moléculaire ou des codes de simulation du fonctionnement des armes. Le code de calcul ABINIT est ainsi devenu un outil quotidien de nombreux physiciens du CEA/DAM.

En près de vingt ans, la simulation haute performance de la matière condensée a permis des avancées notables dans la compréhension du comportement des matériaux sur des problématiques d'intérêt pour le CEA/DAM, mais aussi dans des domaines plus généraux intéressant l'industrie française. C'est en partie l'objet du projet Simulation Matériaux Industrie Calcul Exaflop (SMICE) financé par le Programme d'investissements d'avenir « Calcul intensif et simulation numérique », qui, outre le CEA/DAM, réunit des industriels partenaires, Michelin, Faurecia, Safran et L'Oréal. ■

5.2 Simulation de phénomènes électromagnétiques sur des supercalculateurs parallèles

— Comment associer modèles de programmation, algèbre linéaire et utilisation efficace des accélérateurs pour booster l'efficacité des calculs de diffraction d'ondes radar...

Les dispositifs permettant de réduire la détectabilité d'un objet soumis aux ondes émises par un radar constituent un domaine de grand intérêt pour le CEA/DAM. Dans ce contexte, une mesure de la susceptibilité d'un objet d'être détecté par un radar est quantifiée par la **surface Équivalente radar** (ou **SER**, voir **encadré** « Surface équivalente radar (SER) », p. 81). Elle correspond au rapport entre l'énergie incidente de l'onde émise par le radar et l'énergie rétrodiffusée par l'objet. Depuis les années 1990, le CEA/DAM développe des codes massivement parallèles afin de calculer la SER d'objets complexes. S'il s'agissait principalement de simulations 2D axisymétriques à l'origine, les simulations de cas 3D sont désormais rendues accessibles grâce à l'augmentation considérable de la puissance de calcul. Un de ces codes dédiés aux études de **furtivité** résout les équations de Maxwell en régime fréquentiel en 3D de manière très précise. Les équations qui modélisent ce phénomène sont des **équations intégrales** portant sur toutes les surfaces de l'objet et dont la résolution numérique nécessite de très grandes capacités de calcul.

Plusieurs facteurs expliquent le coût calcul important de ces méthodes. Après le découpage de la surface de l'objet en triangles dont les côtés sont plusieurs fois inférieurs à la longueur d'onde du radar, la **discrétisation** numérique de l'équation intégrale conduit à un **système d'équations linéaires** de très grande dimension comportant éventuellement plusieurs dizaines de millions

d'inconnues (voir article « Comprendre et prévoir par la simulation »). La matrice associée à ce système est dite dense car tous les termes qui la composent sont *a priori* non nuls. Le nombre d'inconnues est proportionnel au carré du rapport entre la longueur d'onde du radar et la taille de l'objet. Ainsi, un doublement de la fréquence conduit à devoir résoudre un système linéaire ayant 4 fois plus d'inconnues. La **figure 1** illustre par exemple les calculs nécessaires à l'évaluation de la SER d'un drone illuminé à une fréquence de 2,5 GHz. Il est par ailleurs nécessaire de résoudre ce système d'équations pour plusieurs seconds membres, chacun représentatif de l'angle d'illumination du radar, et de la polarisation de l'onde. La résolution du système linéaire utilise un solveur direct, qui consiste à factoriser la matrice en un produit de deux matrices triangulaires (algorithme de Cholesky (voir **figure 2**)), dont le nombre d'opérations élémentaires est proportionnel au cube du nombre d'inconnues N . On parle alors d'algorithme avec une **complexité** cubique ou $O(N^3)$.

LES TÂCHES DE CALCUL : UN MODÈLE DE PROGRAMMATION POUR LES MACHINES MASSIVEMENT PARALLÈLES ET LES ACCÉLÉRATEURS

La mise en œuvre d'un solveur direct capable de résoudre de tels systèmes sur une architecture massivement parallèle est un défi majeur qui a été abordé avec une grande efficacité dès le début des années 1990 et jusqu'au milieu des années 2000,

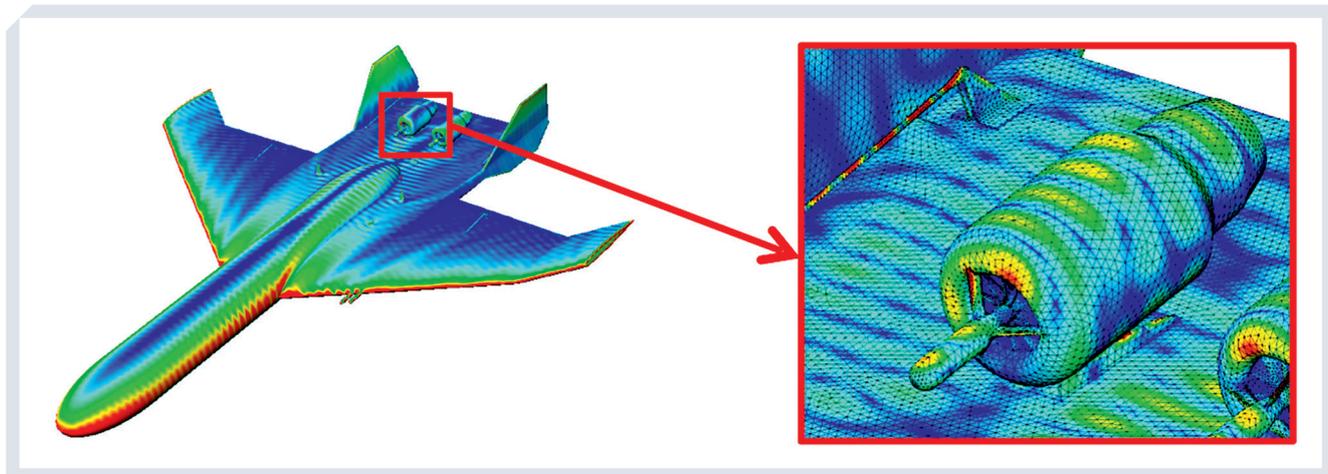


FIGURE 1

Pour calculer la surface équivalente radar (SER), on calcule d'abord les courants électriques et magnétiques induits à la surface pour en déduire le niveau de SER, en calculant les champs diffractés par ces courants à l'infini. Sur le zoom à droite, on distingue les triangles qui constituent le maillage de la surface de ce drone. Le système à résoudre comporte deux inconnues par arête.

de Tera1 jusqu'à Tera100. L'augmentation significative de la taille et de la complexité des architectures de calcul a nécessité une modernisation profonde de ce solveur à partir de la machine Tera100. Ces machines sont en effet devenues trop complexes pour occuper des centaines de milliers de cœurs avec des calculs interdépendants.

Le principal changement a consisté à adopter un paradigme de programmation basé sur des tâches de calcul. Une tâche constitue une portion de code, appelée noyau de calcul, dont on spécifie explicitement les données d'entrée et de sortie. Un algorithme parallèle peut alors s'écrire comme un ensemble de tâches interdépendantes, qu'un moteur d'exécution est en charge de mettre en œuvre tout en respectant les éventuelles dépendances. Cela permet d'exprimer autant de parallélisme que possible et de décrire cet algorithme parallèle sous la forme d'un graphe de tâches,

indépendant de la machine sous-jacente. La figure 2 illustre comment cette approche permet, à partir d'un algorithme séquentiel, d'extraire automatiquement un graphe de tâches exécutable en parallèle.

Cette structuration permet d'isoler les différentes problématiques que sont la conception d'applications scientifiques d'une part, l'optimisation de noyaux de calcul en fonction de l'architecture des processeurs d'autre part, et enfin la répartition de ces tâches sur des machines parallèles.

Ainsi, sur 49 152 cœurs de la machine Tera1000-1, les performances du code original et de sa version modernisée avec des tâches sont respectivement 0,58 Pflop/s et 1,2 Pflop/s sans aucune modification des noyaux de calcul. Cela correspond à 85% des performances maximales atteignables en pratique.

a pour k de 1 à NBLOCKS
 $A_{kk} = L_{kk} L_{kk}^T$ (noyau P)
 pour j de (k+1) à NBLOCKS
 $A_{jk} = A_{jk} L_{kk}^T$ (noyau T)
 pour j de (k+1) à NBLOCKS
 $A_{jj} = A_{jj} - A_{jk} A_{kk}^{-1} A_{jk}^T$ (noyau S)
 pour i de j à NBLOCKS
 $A_{ji} = A_{ji} - A_{jk} A_{kk}^{-1} A_{ki}^T$ (noyau G)

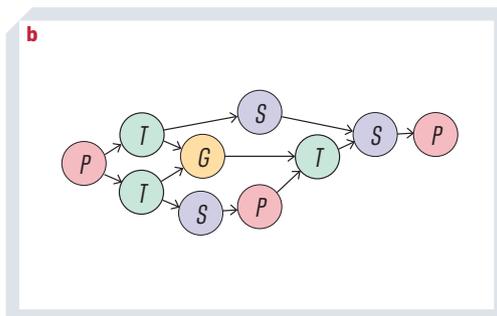


FIGURE 2

Algorithme de factorisation LL^T (Cholesky) à base de tâches (a), et le graphe correspondant généré automatiquement pour NBLOCKS=3 (b).

Sur 3500 **KNL** (238 000 cœurs) de la machine Tera1000-2, dont l'architecture est complètement différente de celle de Tera1000-1, cette approche a également permis d'obtenir une vitesse de 4,93 Pflop/s sur un problème à près de 9 millions d'inconnues ! La résolution de tels problèmes était totalement inaccessible avant de disposer de la puissance de calcul de Tera1000-2 et de codes adaptés à un parallélisme aussi massif.

L'adoption d'un paradigme de tâches est également adaptée aux architectures hétérogènes comprenant des accélérateurs tels que les cartes graphiques (**GPU**) qui disposent d'une mémoire locale. En effet, une tâche est non seulement caractérisée par son noyau de calcul, mais également par les données sur lesquelles elle opère. Le moteur d'exécution des tâches peut alors déplacer les données entre la mémoire principale et celle des accélérateurs de manière automatique. Cela permet d'exhiber une quantité de parallélisme suffisamment importante pour alimenter les nombreux cœurs de ces architectures dites **manycore**, qui sont bien plus efficaces que les architectures de processeurs classiques. En 2012, 360 GPU de type Nvidia K20 (Fermi) de la machine Tera100 ont ainsi permis d'atteindre 83 Tflop/s sur un problème de 247 000 inconnues. En 2017, 160 GPU de type Nvidia K80 (Kepler) obtenaient 207 Tflop/s pour un problème de 450 000 inconnues sur la machine Tera1000-1, c'est-à-dire une puissance de calcul comparable à celle de 10 000 cœurs classiques de cette même machine.

MÉTHODES DE COMPRESSION : UNE APPROCHE NUMÉRIQUE POUR RÉDUIRE LA QUANTITÉ DE CALCUL

La modernisation du modèle de programmation a également rendu possible la mise en œuvre efficace de techniques numériques bien plus évoluées que dans le code initial, où une partie importante des efforts de développement et de maintenance était dédiée à la gestion du parallélisme.

De par la nature des équations intégrales utilisées pour résoudre les équations de Maxwell, la quantité d'informations contenue dans les blocs de matrices représentant l'interaction de deux parties éloignées sur la surface de l'objet simulé est généralement plus faible que pour deux par-

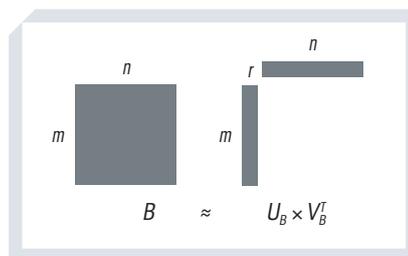


FIGURE 3

Approximation d'un bloc B de taille $(m \times n)$ sous la forme $B \approx U_B V_B^T$ avec U_B et V_B de tailles $(m \times r)$ et $(n \times r) \ll \frac{mn}{m+n}$.

Surface équivalente radar (SER)

La surface équivalente radar (ou SER) permet de quantifier la visibilité d'un objet sous l'effet d'une illumination par une onde radar à une fréquence donnée. Plus cette grandeur est faible, moins l'objet est susceptible d'être *visible* par un radar. Comme l'illustre la **figure E1**, elle se mesure en $\text{dB} \cdot \text{m}^2$ et correspond au rapport entre la quantité d'énergie incidente (E_{inc}) et celle rétrodiffusée par l'objet (E_d) à une distance r infinie, et est obtenue avec la formule suivante :

$$SER = 10 \log \left(\lim_{r \rightarrow \infty} \left(4\pi r^2 \left| \frac{E_d(r)}{E_{inc}(r)} \right| \right) \right)^2$$

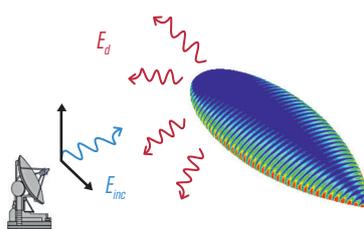
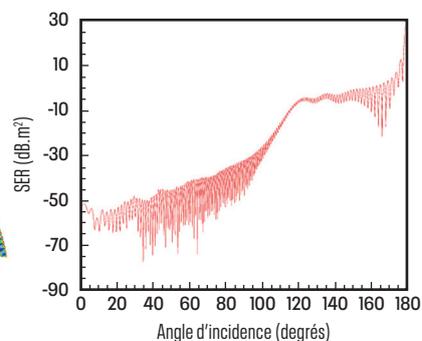


FIGURE E1

Le calcul de la surface équivalente radar peut s'effectuer en évaluant les courants induits à la surface d'un objet illuminé par une onde radar selon un certain angle d'incidence, afin de déduire les champs diffractés par ces courants à l'infini.



ties proches. Bien qu'il s'agisse toujours de blocs denses, cela signifie en pratique que leur spectre contient peu de valeurs singulières importantes. La **figure 3** montre un exemple de bloc B compressé sous la forme $B \approx U_B V_B^T$, où le nombre de colonnes de U_B et de V_B correspond au **rang** numérique r pour une précision ε donnée.

On parle alors de méthode « Block Low-Rank » lorsque certains blocs sont compressés, ou de H-matrice lorsque ces blocs sont compressés de manière hiérarchique, comme illustré par la **figure 4**. Le caractère hiérarchique est obtenu en subdivisant les blocs non compressibles jusqu'à ce qu'ils deviennent compressibles, ou jusqu'à atteindre une taille donnée. Le véritable défi consiste alors à résoudre de manière directe un très grand système linéaire stocké sous une forme compressée. La complexité théorique en nombre d'opérations d'un tel solveur étant de $O(N \log N)$, au lieu de $O(N^2)$, l'introduction de ces méthodes de compression est une avancée majeure pour résoudre des problèmes de taille absolument inaccessible, même avec un supercalculateur. L'algorithme de factorisation LL^T de la **figure 2** dont la structure reste inchangée a été repris pour opérer directement sur des données stockées sous une forme compressée en adaptant uniquement les noyaux de calcul.

La puissance de calcul disponible sur Tera1000-1 a ainsi permis de traiter un cas de 11,8 millions d'inconnues sur 3360 cœurs en seulement 4 h 27 min, sachant qu'un minimum de 168 000 cœurs aurait été requis sur cette même machine pendant une semaine entière avec la matrice non compressée. Il a également été possible de mettre en œuvre des techniques évoluées dans le moteur d'exécution des tâches, telles que de l'équilibrage de charge distribué sans que cette complexité n'affecte la conception du solveur hiérarchique. Dans l'exemple précédent, cela a permis de réduire le temps de factorisation de 4 h 27 min à 3 h 57 min. L'apport des méthodes de compression est considérable puisqu'elles permettent ici de faire un calcul avec 1000 fois moins de ressources, ou bien de calculer 1000 fois plus de problèmes avec les mêmes ressources.

CONCLUSION ET PERSPECTIVES

La modernisation du modèle de programmation n'a donc pas seulement permis d'adapter ce code à des machines massivement parallèles. Elle a rendu possible l'utilisation d'architectures complexes, telles que les accélérateurs de calcul (GPU, KNL). Cela a par ailleurs permis d'introduire des tech-

niques de compression dont la mise en œuvre à grande échelle aurait été particulièrement difficile à concevoir et à maintenir sans un paradigme de programmation aussi flexible et performant que les tâches de calcul. La simulation de cas 3D de très grande taille en électromagnétisme est donc ici rendue accessible par l'utilisation conjointe de ces méthodes de compression, d'un modèle de programmation adapté et de supercalculateurs massivement parallèles. Si un cas 3D de plusieurs centaines de milliers d'inconnues était un défi à l'arrivée de Tera100, il est désormais possible de résoudre des problèmes de dizaines de millions d'inconnues sur Tera1000-1 et Tera1000-2. Une telle capacité permet de résoudre des problèmes de manière plus précise. ■

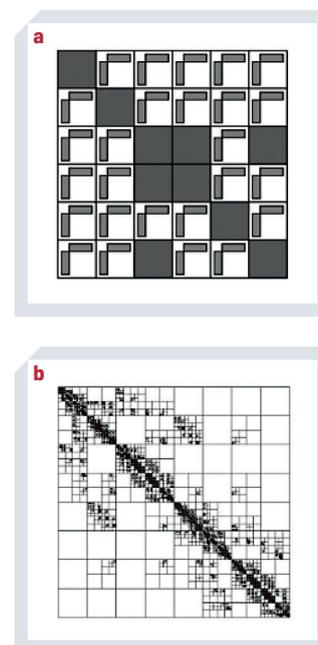


FIGURE 4
Matrice constituée de blocs compressibles (Block Low-Rank (a)), ou de blocs compressés hiérarchiquement (H-matrices (b)).

CHAPITRE 6

LA CONSTRUCTION D'UN ÉCOSYSTÈME DU CALCUL INTENSIF

Dès la mise en place du programme Simulation, la Direction des applications militaires du CEA s'est attelée à faire profiter le secteur industriel des avancées réalisées dans les domaines des technologies du calcul et du traitement intensif des données. En complément du renforcement de la base industrielle, technologique, le pôle de compétences créé à proximité du centre DAM Île-de-France a dynamisé tout le secteur du calcul et de la simulation français.

Le CEA/DAM, moteur d'un pôle de compétences en calcul et traitement de données à haute performance

— Le calcul au service de l'industrie et de la recherche française et européenne: CCRT, Genci, Prace, Teratec...

MOTIVATION

Le CEA/DAM mène, depuis le début des années 2000, une série d'initiatives cohérentes et coordonnées dans le domaine du **HPC**. Elles visent bien entendu à répondre à son besoin en termes de puissance de calcul sur le long terme, mais également, au-delà de ses programmes, à contribuer au développement du HPC dans l'industrie et la recherche ainsi qu'à la construction d'une offre industrielle compétitive, pérenne et souveraine.

À travers cette politique, le CEA/DAM cherche notamment à démontrer la crédibilité de ses compétences en matière de conception et de pilotage de grands systèmes HPC, à répondre aux besoins de l'ensemble du CEA, à soutenir l'industrie et la recherche, et à contribuer au développement d'une offre et d'un marché du HPC en Europe.

Tout d'abord, en 2003, la création du Centre de calcul recherche et technologie (**CCRT**) au sein du centre DAM Île-de-France a permis de regrouper en un seul endroit l'ensemble des grands moyens de simulation du CEA et des équipes associées pour répondre aux besoins de calcul non classifiés de défense, notamment dans le cadre de collaborations académiques ou industrielles ainsi que dans le cadre du développement de **logiciels open source**. Dès l'origine, ce centre de calcul et de compétences a été conçu comme un instrument de partena-

riat avec l'industrie. Outre trois directions du CEA, y participent aujourd'hui dix-neuf partenaires extérieurs.

Puis, en 2005, avec un groupe de grands industriels, le CEA/DAM a été à l'origine de la création de l'association **Teratec**. Cette association comprend aujourd'hui plus de quatre-vingts membres académiques et industriels impliqués dans le développement du HPC et de l'intelligence artificielle (**IA**). Elle a pour objectif d'animer l'écosystème français sur l'ensemble de la chaîne de valeur du HPC et du **HPDA**, et de promouvoir ses résultats et initiatives à l'échelon européen.

Enfin, 2011 a vu la mise en service, au sein du Très Grand Centre de calcul du CEA (**TGCC**), infrastructure construite à cet effet, d'un supercalculateur, opéré par les équipes du CEA/DAM, destiné aux chercheurs français et européens, respectivement *via* la société civile **Genci** (Grand Équipement national pour le calcul intensif) et *via* l'infrastructure **Prace** (Partnership for Advanced Computing in Europe).

Le TGCC fait partie d'un ensemble immobilier situé à proximité immédiate du centre CEA/DAM Île-de-France. Cet ensemble comprend aussi un campus accueillant des entreprises et des laboratoires mixtes industrie-recherche. Il rassemble ainsi des fournisseurs de technologies HPC (matérielles et logicielles) et

des industriels utilisateurs, depuis les petites et moyennes entreprises jusqu'aux grands groupes, afin de créer une boucle vertueuse entre les besoins et les solutions.

Fort de ses résultats et des coopérations qu'il a mises en place, le CEA est devenu un acteur majeur dans l'élaboration de la stratégie nationale et européenne du calcul haute performance.

En cohérence avec les travaux de coopération menés par les équipes du CEA/DAM et la société Atos, le CEA s'est vu confier une nouvelle mission dans le cadre de la convention État-CEA de novembre 2014 sur le calcul intensif. En partenariat avec des industriels leaders mondiaux du secteur, il s'agit de poursuivre le développement des technologies pour les calculateurs de haute performance indispensables à la souveraineté technologique et à la compétitivité des industriels nationaux et européens. L'objectif est de contribuer ainsi à la maîtrise des technologies dites **exaflopiques** (10^{18} opérations/seconde) et au développement de la filière industrielle HPC- simulation-traitement

de données massives (big data) en inscrivant son action dans les initiatives mises en place sur le plan européen.

Les perspectives de développement sont importantes, avec notamment un renforcement du rôle de l'Europe grâce à la création d'une entité juridique (**EuroHPC**) destinée à centraliser les financements européens en matière de HPC, tant en matière d'infrastructures que de R&D ou d'applications.

L'OUVERTURE AU MONDE INDUSTRIEL : LE CCRT, UNE PLATEFORME POUR LE CALCUL ET LE TRAITEMENT À HAUTE PERFORMANCE

Le Centre de calcul recherche et technologie (CCRT) a été créé en 2003 suite à la décision de regrouper dans un même site (DAM Île-de-France) les compétences et les moyens HPC du CEA. Dès l'origine, il a été pensé pour être ouvert à des partenariats industriels. L'objectif du CCRT est triple : accompagner les industriels français dans le développement de leurs projets de simulation numérique et de traitement de la donnée, démontrer le savoir-faire du CEA/DAM dans le domaine du HPC et contribuer ainsi à la crédibilité de son programme Simulation, s'inspirer des besoins et démarches de simulation des partenaires pour nourrir/renforcer/conforter les travaux de chacun (**figure 1**).

Le CCRT a mis en service en 2016 le calculateur Cobalt fourni par Atos/Bull. Ce calculateur, opéré par les équipes du CEA/DAM, dispose, en 2019, d'une puissance de calcul crête de 2,4 **Pflop/s**. Cobalt sera remplacé mi-2021 par un nouveau calculateur de plus de 8,8 Pflop/s, appelé Topaze. Le CCRT propose à ses partenaires qui souhaitent accélérer le développement de leurs applications de simulation numérique des ressources et des services HPC de très haut niveau.

Bien plus que l'accès à des moyens de calcul, de stockage et de visualisation au meilleur niveau mondial, le CCRT a aussi pour mission d'aider ses partenaires à s'approprier de nouvelles technologies et à préparer les outils de simulation de l'avenir. Des séminaires thématiques et des sessions de formation sont organisés à cet effet. Depuis 2003, le CCRT a ainsi connu quatre générations de supercalculateurs et, ce faisant, une augmentation de sa puissance de calcul, qui a été multipliée par plus de 600. De nouveaux services ont été déve-

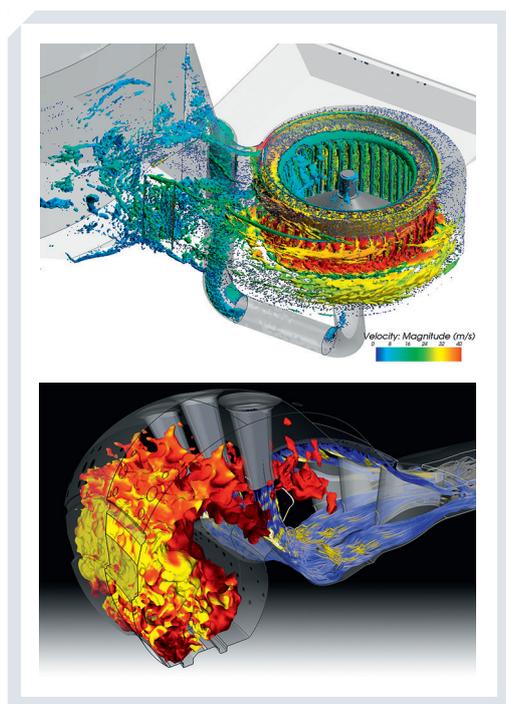


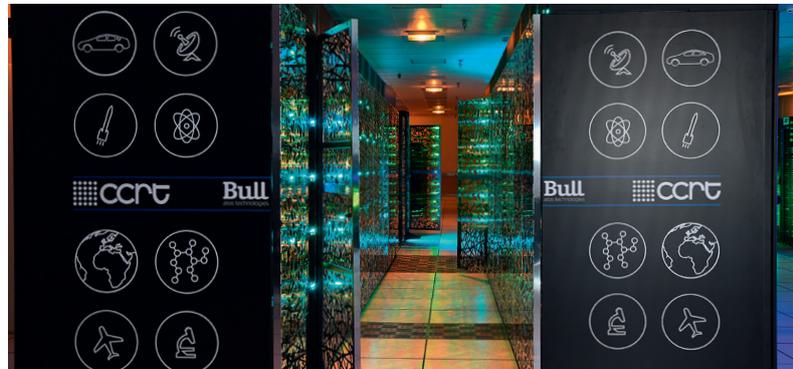
FIGURE 1

Illustration de travaux de simulation réalisés au CCRT (Courtesy from Valeo et Safran).

Le CCRT et ses partenaires en 2021

En 2021, le CCRT exploite un calculateur Atos/Bull de 2,4 Pflops, dénommé Cobalt, et met en œuvre un ensemble de services au sein d'un environnement fiable et sécurisé.

À cette date, les 19 partenaires du CCRT, outre le CEA, sont les suivants : ArianeGroup, EDF, **IPEN**, Ineris, Ingélicance, IRSN, Onera, L'Oréal, Michelin, Safran Aero Boosters, Safran Aircraft Engines, Safran Helicopter Engines, Safran Tech, GIE Synchrotron Soleil, Total, Thales, Thales Alenia Space, Technicatome et Valeo.



loppés. Au calcul se sont ajoutés la gestion de données, la visualisation distante puis, plus récemment avec l'infrastructure mise en place pour le projet **France Génomique**, le traitement de masses de données. Cette formidable évolution s'est accompagnée d'une extension des domaines applicatifs traités par les partenaires industriels qui sont aujourd'hui d'une très grande diversité (aéronautique, spatial, énergie, transport terrestre, sécurité environnement, santé, cosmétique, industrie pharmaceutique...).

En accord avec ses partenaires, le CCRT s'ouvre à de nouveaux usages via un accès virtualisé aux ressources. Il étend l'utilisation de ses ressources HPC à l'apprentissage automatique (« réseaux de neurones » et autres algorithmes d'**intelligence artificielle**) et propose un accès à un simulateur quantique, afin d'en explorer les potentialités (voir **encadré** « Se préparer à l'informatique quantique au CCRT »).

L'OUVERTURE AU MONDE ACADÉMIQUE: L'INSERTION DANS L'ESPACE FRANÇAIS (GENCI) ET EUROPÉEN (PRACE) ET L'INFRASTRUCTURE TGCC

Partenaire et associé du Genci, société civile créée en 2007, le CEA héberge et opère les moyens de calcul que le Genci finance, et ce à destination des chercheurs publics français. Le Genci est une société civile créée en 2007 dont les associés sont le **MESRI** (49 %), le CEA

(20 %), le **CNRS** (20 %), les universités (10 %) et l'**Inria** (1 %). Le Genci apporte aussi la contribution française à l'infrastructure européenne Prace, créée en 2010, dont l'objectif est de permettre à l'Europe du calcul intensif de rester dans la compétition internationale.

Afin de disposer en France d'une infrastructure d'accueil de capacité exceptionnelle, le CEA a construit le Très Grand Centre de calcul

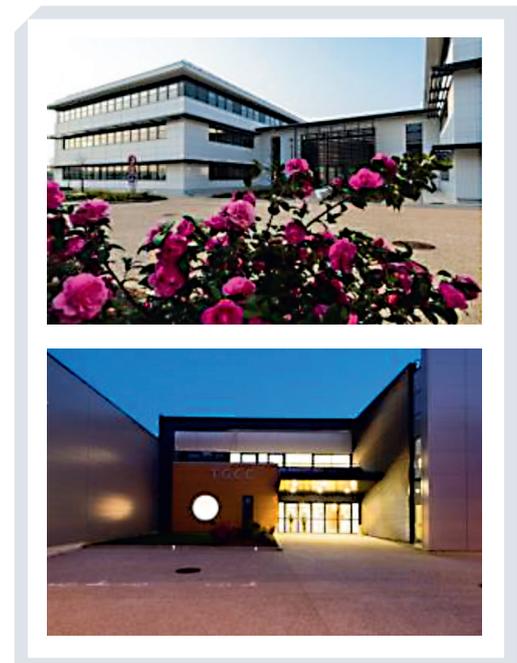


FIGURE 2

Campus Teratec (en haut) et entrée du TGCC (en bas).

du CEA (TGCC, **figure 2**). Ses principales caractéristiques sont les suivantes :

- ⊗ salles informatiques de 2 600 m², pouvant être refroidies par air et/ou par eau, associées à une surface de près de 4 000 m² pour les servitudes ;
- ⊗ servitudes dimensionnées pour des systèmes informatiques jusqu'à 7,5 MW facilement extensibles au-delà. Desserte électrique par ligne très haute tension de 225 kV d'une capacité de 60 MW... ;
- ⊗ accès réseau très haut débit grâce à un nœud du réseau de la recherche français (**RENA-TER**) ;
- ⊗ amphithéâtre de 200 places, espace polyvalent et salles de réunion pour les échanges scientifiques.

Le TGCC réunit des caractéristiques techniques d'exception, ce qui le place à la pointe des grands data centers HPC. Le TGCC a également vocation à être un lieu d'échanges scientifiques et de promotion du calcul intensif. Il offre un espace permettant l'organisation de

conférences, de séminaires et de formations dans ce domaine.

Le CEA s'implique fortement dans l'optimisation énergétique de telles infrastructures. Il participe, sur le plan européen, à l'échange de bonnes pratiques en organisant, avec d'autres grands centres de calcul européens, parmi lesquels LRZ (Munich, Allemagne) et CSCS (Lugano, Suisse), la conférence annuelle de référence dans le domaine en Europe : « Annual workshop on HPC centre infrastructure ».

Actuellement, le TGCC héberge, outre le supercalculateur Cobalt et son successeur Topaze du CCRT, le supercalculateur Joliot-Curie d'Atos/Bull, d'une puissance de 22 Pflop/s, qui matérialise la contribution de la France à l'infrastructure européenne de supercalculateurs Prace (**figure 3**). Cette machine est partagée, sur la base d'appels à projets et d'une sélection basée sur l'excellence scientifique, à 50 % pour les organismes de recherche européens *via* le programme Prace et 50 % pour les laboratoires de recherche français, *via* la société civile Genci.

Safran, un partenaire du CCRT

— TÉMOIGNAGE DE M. FRÉDÉRIC FEYEL (SAFRAN TECH)

DIRECTEUR MODÉLISATION & SIMULATION ET REPRÉSENTANT DU GROUPE SAFRAN AU COMITÉ DE PILOTAGE DU CCRT

Dans un environnement hautement compétitif entre les motoristes aéronautiques, la simulation est un élément différenciant essentiel. Elle permet de réduire le temps de conception, tout en testant toujours plus d'architectures et configurations produit différentes, en limitant le recours aux essais. Elle permet par ailleurs, pendant la vie du produit, une réactivité maximale pour résoudre les crises éventuelles.

Les moyens de calcul haute performance (HPC) sont indispensables pour traiter le plus efficacement possible ces simulations, toujours plus complexes : modèles multiphysiques, prise en compte d'effets technologiques très fins (c.-à-d. détails géométriques), etc. Maîtriser le fonctionnement de ces moyens de calcul et les utiliser au mieux contribuent, pour

Safran, à faire jeu égal avec les meilleurs acteurs mondiaux.

Sur les dix dernières années, la croissance de l'utilisation du HPC dans l'entreprise a été exponentielle, et les projections indiquent que ce rythme va rester soutenu.

Quatre sociétés du groupe Safran sont partenaires du CCRT : SAE, SHE, SAB, Safran Tech.

Afin d'amplifier les efforts de recherche associée, Safran a créé en 2015 « Safran Tech », un centre de recherche transverse groupe, qui a notamment pour objectif de mettre au point les modèles, algorithmes et codes de calcul qui seront ensuite utilisés dans les bureaux d'étude. À ce titre, Safran Tech utilise massivement les infrastructures du CCRT pour valider et améliorer ces outils de simula-

tion. Outre la mise à disposition de puissance de calcul, cet environnement facilite le transfert vers les bureaux d'études du groupe, qui peuvent directement utiliser au CCRT les technologies mises au point à Safran Tech.

Le CCRT est par ailleurs plus qu'un puissant moyen de calcul : c'est aussi – et c'est essentiel – une communauté d'experts riche d'échanges et de partage sur toutes les thématiques du calcul et des innovations associées. La récente et très rapide montée en puissance des besoins autour de l'intelligence artificielle est par exemple prise en compte par cette communauté et accompagnée par le CCRT, qui a su très rapidement décliner et mettre à disposition dans son infrastructure les briques matérielles nécessaires (p. ex. **GPU**).



Christine Ménaché, à l'interface de la technique et de l'humain

La technique n'est rien si l'on ne sait pas échanger avec les utilisateurs. L'échange n'est rien si l'on ne connaît pas la technique sous-jacente. La responsable du TGCC/CCRT jongle avec ces deux aspects avec enthousiasme.

— Pour être responsable d'un grand centre de calcul comme le TGCC/CCRT, il faut à la fois être compétent sur les questions techniques et aimer les relations humaines. Christine Ménaché est donc la personne idéale. Le calcul haute performance, elle connaît : elle a été responsable technique du CCRT à sa création en 2003, et jusqu'en 2009.

Partenariats

Entre 2009 et 2010, elle évolue vers le développement de nouveaux partenariats avec des industriels. Être force de proposition, vérifier que le CCRT répond à leurs besoins... « C'est super intéressant, s'exclame-t-elle. On s'ouvre par exemple au domaine de la cosmétique, pour répondre aux besoins de partenaires comme L'Oréal. » Le CCRT n'est pas un fournisseur de services, ne fait pas de bénéfices. Le partenaire type est un organisme faisant déjà du calcul numérique, que le CCRT aide à changer d'échelle. Les enjeux sont énormes. Par exemple les constructeurs de moteurs d'avions, qui ont besoin de calculs poussés pour réduire le bruit et la consommation de carburant. Les connaissances développées par le CEA/DAM pour son programme de simulation numérique ont des retombées sur le CCRT. « Des industriels comme EDF ou Total possèdent de gros calculateurs, mais ils ont besoin d'échanger sur ces sujets de plus en plus complexes, indique Christine Ménaché. Par exemple, les processeurs sont de plus en plus spécialisés, il faut faire une veille technologique. Les industriels viennent notamment nous voir car ils n'ont pas d'experts sur la partie amont du calcul, la R&D. »

Connaissance technique

Au quotidien, elle se partage entre les questions les plus scientifiques – participer aux réunions d'attribution des heures de calcul pour le volet « recherche académique », donc analyser la qualité d'ordre scientifique ou technique des projets – et les aspects plus relationnels. Ce grand écart lui plaît. « La connaissance technique est indispensable

pour expliquer ce que l'on fait aux partenaires, souligne-t-elle. Je me suis détachée de la technique, mais je garde beaucoup de liens avec les équipes opérationnelles. Il faut avoir vu de l'intérieur comment fonctionne un centre de calcul. » Aujourd'hui, l'informaticienne s'enthousiasme déjà pour le futur du TGCC qui devrait être le site d'accueil, si la candidature française est retenue, d'une machine « **exascale** » européenne à l'horizon 2023.

À la recherche des informaticiennes

Son unité est également chargée de développer les logiciels qui optimisent l'usage du calculateur, comme l'ordonnanceur : « C'est comme le jeu Tetris, s'amuse-t-elle. Le contenant, c'est l'ordinateur, et les calculs sont les briques de différentes formes. Il faut un chef d'orchestre pour organiser les calculs, afin que l'ordinateur soit rempli le mieux possible. » Plus la machine est grosse, plus les tâches sont nombreuses. Mais le risque que des processeurs tombent en panne ou la difficulté de gérer les grands volumes de données augmentent avec la puissance du calculateur. Lorsqu'elle décrit les métiers du calcul de haute performance, son enthousiasme est communicatif. « C'est passionnant car on touche à des technologies très innovantes, on a une marge de manœuvre en créativité pour développer de nouveaux services, de nouvelles applications et pourtant très peu de femmes candidatent dans ce domaine, et c'est bien dommage... Il existe des disciplines très différentes, depuis le plus fondamental jusqu'au plus applicatif pour ceux qui aiment mettre la main à la pâte. » Mais elle se désole de la trop faible présence de femmes en informatique. « L'image du geek solitaire agit comme repoussoir, alors que les métiers visés vont du développement autour de nouvelles technologies aux interactions avec les utilisateurs, et à la gestion de projets dans des domaines nouveaux comme la santé..., métiers où les femmes ont toute leur place et sont les bienvenues. »

Depuis fin 2020, les équipes du CEA ont mis en service au TGCC de nouveaux équipements dans le cadre du projet ICEI¹. Ces derniers sont complétés par des services innovants du type « cloud computing » souverain, « stockage objet » et « calcul interactif ».

L'ASSOCIATION TERATEC : PROMOUVOIR L'ENSEMBLE DE LA CHAÎNE DE VALEUR DU HPC

C'est dans l'objectif de promouvoir la simulation haute performance, une priorité à la fois du monde industriel et du monde académique mais aussi une opportunité pour les entreprises informatiques, que se sont réunis au sein de l'association Teratec les acteurs majeurs de ces différents domaines auxquels se sont associées les collectivités locales concernées, soucieuses du développement de leur territoire. Cette association, selon la loi de 1901 et dont le CEA/DAM est un membre fondateur, est dirigée par un président élu qui s'appuie sur un conseil d'administration.

L'association Teratec a pour objet de mobiliser ses partenaires et tous les acteurs intéressés en vue d'accélérer le développement de l'ensemble de la chaîne de valeur du HPC et de l'IA. Celle-ci repose sur trois piliers :

- ⊙ 1. Les technologies matérielles et logicielles, qui incluent la R&D et la conception des calculateurs.
- ⊙ 2. Les grandes infrastructures, qui comprennent les calculateurs et les centres de calcul.
- ⊙ 3. Les applications, qui englobent les projets de simulation et d'apprentissage par les données, les développements applicatifs et les formations nécessaires pour utiliser ces calculateurs.

Ainsi, Teratec a pour objectifs :

- ⊙ de participer à l'émergence et la mise en place de projets de R&D collaboratifs entre les différents partenaires ;
- ⊙ de favoriser l'accès à des moyens de traitement parmi les plus puissants du monde ;
- ⊙ d'assurer le développement des compétences par des actions de formation, des séminaires, des conférences ;

1_ICEI : l'infrastructure informatique de calcul interactif (Interactive Computing e-infrastructure, ICEI) est mise en œuvre pour le projet Human Brain (HBP) par les principaux centres de calcul européens, à savoir BSC (Espagne), le CEA (France), Cineca (Italie), CSCS (Suisse), et ISC (Allemagne).

La dynamique d'échanges entre partenaires du CCRT

Si le fait de partager des ressources de calcul, de stockage, de traitement et une qualité de service exceptionnelle constituent le fondement de l'adhésion des partenaires du CCRT, la possibilité d'échanger entre partenaires et notamment avec les experts du CEA/DAM dans le domaine du calcul haute performance (HPC) est un atout inestimable.

Cet environnement d'échanges s'entretient, et les experts du CEA/DAM sont force de proposition sur les thématiques débattues lors des différents « workshops » qui alimentent ces échanges. De sujets techniques, comme « l'usage du C++ en simulation haute performance » ou « les outils pour optimiser la qualité numérique des codes », à d'autres plus fonctionnels, comme « quels besoins en termes de nouveaux services aux utilisateurs sur le CCRT ? », en passant par ceux plus orientés technologies, comme « l'état de l'art de l'informatique quantique » ou le « machine learning », les thèmes de ces réunions de travail sont très variés.

Chaque partenaire est invité à y apporter son expérience ou ses interrogations, et à s'enrichir des retours de chacun des participants dans un mode ouvert et sans arrière-pensée commerciale...

Dans un monde où l'utilisation de la simulation numérique est un avantage compétitif pour les entreprises, ces lieux d'échanges permettent aux participants d'anticiper les évolutions technologiques contribuant à l'élaboration de leur feuille de route, d'évaluer des méthodes de développement ou d'utilisation de logiciels, et d'être informés de solutions logicielles « open source », notamment développées par les partenaires.

De plus, la journée scientifique annuelle permet aux partenaires de présenter à un large auditoire des résultats marquants obtenus grâce aux moyens du CCRT. Chaque année, un thème différent structure les présentations.

- ⊙ de contribuer au développement et à l'animation du campus Teratec ;
- ⊙ d'être un acteur reconnu et écouté en France et en Europe pour l'élaboration des programmes de R&D et d'infrastructures dans les domaines du HPC et de l'IA.

Le technopôle Teratec, qui se compose du campus Teratec et du TGCC, est dédié à la simulation, au calcul haute performance et à l'intelligence artificielle. Il vise, dans un contexte de proximité géographique, à favoriser le développement de collaborations entre acteurs du HPC et de l'IA par :

- ⊙ la mise en place de laboratoires de recherche communs capables de mener des recherches au plus haut niveau dans le domaine du



FIGURE 3
 Une vue de l'infrastructure européenne de supercalculateurs en 2018.

Se préparer à l'informatique quantique au CCRT



Photographie de l'Atos Quantum Learning Machine au CCRT.

Le Centre de calcul recherche et technologie (CCRT), situé au centre CEA/DAM de Bruyères-le-Châtel, et Atos collaborent pour mettre à la disposition des utilisateurs industriels du CCRT

un des simulateurs de calculateurs quantiques les plus performants au monde. Baptisé Atos Quantum Learning Machine (QLM), ce simulateur anticipe la technologie de rupture que pourrait

représenter l'informatique quantique en offrant la possibilité de tester et valider des algorithmes qui utiliseraient des **bits** quantiques (qubits) physiques ou logiques.

Conçu et commercialisé par la société Atos, il permet aux partenaires du CCRT comme EDF, Safran, l'IFPEN, et au CEA d'évaluer les potentialités des technologies quantiques à partir de quelques cas d'usage représentatifs de leurs besoins dans les domaines de la simulation numérique, de l'optimisation ou du « machine learning ».

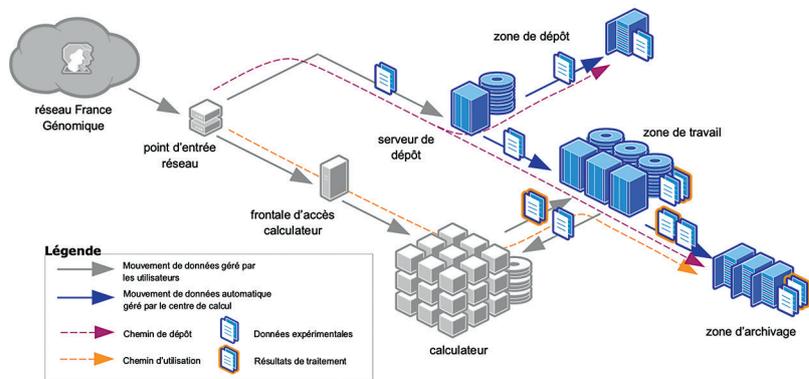
Cette initiative complète celle annoncée par Teratec de la création, sur le même site de Bruyères-le-Châtel, d'un Centre de compétences en calcul quantique qui développera l'expertise dans le domaine de l'utilisation de l'informatique quantique.

L'infrastructure d'analyse des données génomiques du projet France Génomique installée au CCRT

Le séquençage massif du génome, désormais possible, ouvre la voie à de nombreux champs de recherche dans toutes les disciplines (biologie, médecine, agronomie, biodiversité, etc.). Des traitements thérapeutiques adaptés à chaque patient ou des médicaments ciblés peuvent dès lors être envisagés. Face à cette évolution qui transforme radicalement l'approche des sciences du vivant, France Génomique, infrastructure nationale pour les projets en génomique, a mis en place une infrastructure « big data » capable de répondre aux besoins de ce projet national. Cette infrastructure, installée au Centre de calcul recherche et technologie (CCRT) au sein du Très Grand Centre de calcul du CEA (TGCC), est décrite ci-dessous. Le TGCC héberge tout autant les résultats de traitements numériques du projet France Génomique que ses données expérimentales. Ces dernières sont issues des expériences menées par les chercheurs sur les séquenceurs ADN et ne sont pas simplement reproductibles, au contraire des sorties des codes de simulation exécutés sur les supercalculateurs. Par conséquent, elles doivent bénéficier d'un traitement spécial pour garantir leur préservation. Ces données arrivent via Internet sur des machines de dépôt dédiées. Des tests

permettent de certifier la conformité et l'intégrité des données reçues. Elles sont alors placées automatiquement dans un système d'archivage informatique afin de les « sanctuariser ». Au sein de celui-ci, elles sont référencées dans une base de données et conservées en double copie sur plusieurs bandes magnétiques. En parallèle, les données sont copiées sur l'un des systèmes de fichiers parallèles du TGCC afin d'y être accessibles depuis les supercalculateurs. Les fichiers du projet France Génomique sont là encore identifiés comme critiques et des copies multiples (disques

et bandes) sont conservées dans ce second système. La conservation des résultats expérimentaux et la préservation de leur intégrité sont ici les maîtres-mots de l'architecture informatique dédiée mise en place. Des pannes informatiques (corruption silencieuse du contenu d'un support informatique, panne matérielle sur un équipement informatique) ou des erreurs humaines (fausse manipulation) peuvent survenir. Il est ainsi toujours possible d'accéder à une copie de la version déposée initialement dans le système par les chercheurs.



calcul intensif et de la simulation numérique haute performance ;

- Ⓞ l'implantation d'acteurs majeurs sur un campus fortement créateur d'emplois et de compétences autour de ces thématiques ;
- Ⓞ l'aide à la création et au développement de startup et de jeunes entreprises.

Avec ce technopôle, implanté dans une zone de R&D du pôle de compétitivité mondial « **SYSTEM@TIC Paris-Région** », Teratec constitue un acteur important du calcul et du traitement intensifs en Europe.

PERSPECTIVES

Avec la création de l'entité juridique EuroHPC, le financement par la Commission européenne du HPC en Europe va gagner en volume et en efficacité. Pour le CEA/DAM, cela signifie de nouvelles opportunités pour accélérer sa stratégie en la matière. Tout d'abord au niveau des infrastructures, avec le projet d'accueillir au TGCC en 2023/2024 un des systèmes « **exascale** », basé sur des technologies européennes, co-financés par EuroHPC. Ensuite au niveau des technologies, avec notamment le projet **EPI**

FENIX, une fédération de services pour le calcul et le stockage à l'échelle européenne

Le CEA/TGCC est des cinq membres fondateurs de la fédération de services Fenix. Cette fédération offre de nouvelles ressources de calcul et de stockage pour répondre aux besoins des communautés de recherche européennes. Dans le cadre du projet européen ICEI, une extension du centre de calcul TGCC a été déployée pour répondre aux besoins spécifiques de la recherche sur le cerveau et d'autres domaines de recherches nécessitant des services flexibles de calcul interactif et de gestion de données. Elle aide notamment les chercheurs à cartographier et simuler tout ou partie d'un cerveau humain, mais aussi à tester des intelligences artificielles de robots ou véhicules autonomes dans des environnements virtuels. Cette infrastructure de type « Cloud » permet

le développement de services par les chercheurs eux-mêmes pour des communautés de chercheurs, tels que des services web, des bases de données partagées ou des plateformes de données ouvertes. Son déploiement au sein du TGCC permet d'offrir la puissance des capacités de calcul et de traitement de données tout en gardant une interface ergonomique de type « cloud ».

À sa création Fenix fédérait cinq systèmes similaires en Europe, en France, en Allemagne, Espagne, Italie, Suisse et depuis 2021 un sixième situé en Finlande. Cette fédération gère l'allocation des ressources de calcul et de stockage entre les différents centres de calcul. Un mécanisme de contrôle d'accès unifié et fédéré est également mis en œuvre pour faciliter l'accès des utilisateurs aux dif-



férentes ressources quelle que soit leur localisation. Fenix propose également des services avancés de gestion de données, tels que la migration de données entre sites et la localisation de données. Fenix reçoit un financement du programme européen de recherche et d'innovation **Horizon 2020**, à travers le projet ICEI, sous la convention n°800858.

qui ambitionne la conception et la réalisation d'un **processeur** européen. Enfin, au niveau des applications, avec des efforts particuliers sur les applications ayant vocation à tourner sur des systèmes exascale.

D'autre part, le CEA se positionne comme un acteur majeur en matière de technologie du numérique pour la santé. Dans ce domaine, on peut citer l'initiative française du **PFMG25** (Plan France Médecine Génomique 2025). Le PFMG25, piloté par **AVIESAN** et soutenu par l'État, doit d'ici dix ans positionner la France dans le peloton de tête des grands pays engagés dans la médecine génomique. S'il répond à un enjeu de santé publique, il ambitionne aussi de faire émerger une filière médicale et industrielle nationale en médecine génomique et d'exporter ce savoir-faire.

Dans le cadre de ce projet, le TGCC pourrait héberger une partie du Collecteur analyseur de données (**CAD**). Ce dernier devra être capable de traiter et d'exploiter le volume considérable de données générées et d'offrir les premiers ser-

vices à destination des professionnels de santé dans le cadre du parcours de soins (analyses *in silico* et outils d'aide à la décision en matières diagnostic, pronostic et d'élaboration de stratégies thérapeutiques) et à destination de la recherche.

Une autre ambition du CEA est de créer, au TGCC, une plateforme de R&D en santé pour développer de nouveaux services pour la recherche autour des données d'imagerie du centre de neuro-imagerie cérébrale NeuroSpin, des données de génomique et d'autres données issues de laboratoires de recherche. ■

CHAPITRE 7

PRÉPARER L'AVENIR EN COOPÉRATION : EN ROUTE VERS L'EXASCALE

Dans les années 2020, pour répondre aux enjeux de ses programmes, la Direction des applications militaires du CEA doit mettre en œuvre un calculateur de classe exaflopique. Celui-ci, d'une puissance de plus d'un ordre de grandeur supérieure à celle de la génération des calculateurs TERA, présente des défis technologiques majeurs qui mobilisent les équipes. Afin de relever ces défis, les travaux sont menés en coopération avec de nombreux partenaires, en premier lieu en Europe, mais aussi aux États-Unis et au Japon.

Préparer l'avenir en coopération : en route vers l'exascale

— Ce chapitre présente quelques-uns des défis à relever pour les calculateurs exaflopiques. Pour y répondre, les équipes du CEA/DAM s'insèrent dans un réseau international de coopérations académiques et industrielles dont les principaux projets et accords sont mis en exergue.

Entre Tera1, cluster de 5 **téraflops** installé en 2001, et Tera100, installé en 2010, la puissance de calcul a été multipliée par 250, bien que l'augmentation de la fréquence des **processeurs** n'y ait contribué que pour un facteur inférieur à 2,5. La barre du **pétaflops** a été franchie majoritairement par l'augmentation du nombre de composants électroniques, en tout premier lieu les processeurs, conjointement à une forte augmentation du nombre de **cœurs de calcul** au sein de ceux-ci. Cette mise en œuvre de plusieurs cœurs de calcul au sein d'une seule puce électronique (appelée processeur multicœurs) est apparue en 2004 pour pallier la stagnation de la fréquence de fonctionnement et maîtriser l'énergie électrique dissipée par effet Joule. Cette évolution architecturale a marqué, à niveau de **parallélisme** inchangé, la fin de la période d'accroissement sans effort de la performance des logiciels. Avec Tera1000, le CEA/DAM est entré dans une nouvelle ère qui verra sa concrétisation dans la machine Exa1, le supercalculateur de classe **exaflopique** du CEA/DAM des années 2020.

LA RANÇON DU SUCCÈS

Quels sont les enseignements à tirer pour l'avenir des supercalculateurs et de la simulation au CEA/DAM ? D'une part, la performance des supercalculateurs Tera a transformé le processus de simulation. Des simulations plus précises, de phénomènes multiphysiques tridimensionnels et

institutionnaires produisent des volumes de résultats considérables (voir chapitre « Avancées en simulation et supercalculateurs »). La problématique du traitement des masses de données et de la régulation des flux est entrée définitivement au cœur de la réflexion des équipes du CEA/DAM et plus généralement des acteurs du **HPC** (voir **encadré** « Organiser le centre de calcul autour des données », p. 19). L'agrégation de plus en plus importante de composants élémentaires, permise par l'architecture parallèle distribuée, a conduit à une augmentation proportionnelle de la consommation électrique. Les coûts de possession et d'infrastructures ne peuvent plus être ignorés, si bien que la contrainte énergétique conditionne désormais les développements du HPC (voir article « Le challenge énergétique »).

LA MICROÉLECTRONIQUE N'EST PAS TOUTE-PUISSANTE

La fameuse **loi**, empirique et économique, de **Gordon Moore** qui prédisait le doublement du nombre de transistors par puce tous les deux ans est aujourd'hui remise en question. La finesse de gravure des circuits intégrés les plus avancés pourrait encore diminuer pour atteindre quelques milliardièmes de mètre, mais le rythme de décroissance ralentit avec l'augmentation des difficultés technologiques et des budgets d'investissement nécessaires. Depuis plusieurs années, les limites de la technologie de la microélectronique et ses

Quelques projets collaboratifs auxquels participent les équipes du CEA/DAM

Le programme européen de financement de la recherche et de l'innovation **Horizon 2020** a saisi l'importance du HPC comme technologie clé pour la recherche et l'industrie. Il finance de nombreux projets collaboratifs impliquant des

partenaires de plusieurs pays européens, réunis en consortium, sur la base d'un programme de travail de plusieurs années. Le CEA/DAM est engagé dans plusieurs d'entre eux qui sont mentionnés dans le tableau ci-dessous.

Projet	Objectif	Enjeux du CEA/DAM
SAGE2 (2017-2020)	Développer une technologie de stockage de données « intelligent » qui, contrairement à un stockage classique, sera capable de mettre en œuvre des traitements sur un sous-ensemble de données au plus près des unités de stockage.	Disposer de technologies de stockage compatibles avec la production des supercalculateurs du CEA/DAM de la prochaine décennie.
MAESTRO (2017-2020)	Développer des intergiciels adaptés à l'utilisation de différents types de mémoires et aux différents types de données.	Fluidifier les mouvements des données produites par les grands codes de calcul dans un cadre exaflopique en mettant en œuvre de nouveaux paradigmes.
EXDCI2 (2018-2020)	Animation et développement de la communauté du HPC en Europe et élaboration d'une vision commune.	Orienter les développements de la communauté et partager une vision sur l'évolution du domaine.
MB2020 (2016-2019)	Développer des blocs de base indispensables à un processeur européen.	Disposer de composants compatibles avec une utilisation HPC.
EPI - SGA1 (2018-2021)	Créer un processeur européen utilisable pour le HPC et l'automobile.	Disposer d'une source européenne souveraine de processeurs garantissant au CEA/DAM la capacité de poursuivre sa mission, quelles que soient les tensions internationales et leurs conséquences sur les approvisionnements en supercalculateurs.
HBP - ICEI (2015-2023)	Simuler le fonctionnement du cerveau humain.	Concevoir une architecture de stockage des données compatible avec l'exascale.
DEEP-SEA (2021-2024)	Développer des environnements de programmation pour les architectures modulaires.	Disposer de ces environnements pour les grands systèmes exascale européens.
IO-SEA (2021-2024)	Développer une nouvelle architecture hiérarchisée de stockage « objet » des données.	Disposer d'une architecture de stockage adaptée aux grands systèmes exascale européens.
RED-SEA (2021-2024)	Assurer l'interopérabilité et le passage à l'échelle d'un réseau d'interconnexion pour calculateur hétérogène.	Préparer la prochaine génération du réseau d'interconnexion BXI développé par Atos.

contraintes (dissipation thermique et réduction de la tension d'alimentation) ne permettent plus d'espérer une réduction significative du cycle d'horloge des processeurs. Dès lors, l'évolution de l'architecture des processeurs et la combinaison de paramètres tels que le nombre de **cœurs** ou le nombre d'instructions par cycle d'horloge restent la source principale de l'accroissement de la performance des microprocesseurs.

La réduction de la consommation d'énergie provient aussi de la maîtrise des mouvements de données nécessaires à l'alimentation des unités de calcul. Rapprocher la mémoire du processeur et en augmenter le débit grâce aux technologies d'empilement 3D sont devenus indispensables. En matière de « moteurs de calcul », les principales architectures en lice sont les suivantes :

- ⊙ d'une part les architectures multicœurs (p. ex. 10 à 20) et **manycore** (ou pluricœurs) (p. ex. 60 à 100) au sein desquelles des cœurs homogènes tirent leur performance d'une augmentation du nombre d'instructions par cycle (le retour du « vectoriel » au niveau du cœur) ;
- ⊙ d'autre part les architectures hybrides, aujourd'hui de type accélérateur (généralement une association **CPU/GPU**), mais qui pourraient préfigurer l'arrivée d'architectures hétérogènes mélangeant des cœurs de même type, mais de performances très différentes ou de parties plus spécifiquement adaptées à certains types de traitements.

Le HPC est entré dans une ère de diversité, voire de spécialisation, architecturale dont les invariants sont la recherche de la localité, la spécialisation, et la mise en œuvre de toutes les sources et tous les niveaux possibles de traitements parallèles. Aucun modèle de programmation ou langage universel ne pouvant prendre en compte la diversité et la complexité croissante des architectures matérielles, c'est de plus en plus sur le **middleware** ou **intergiciels** en fran-

çais, les couches logicielles intermédiaires entre le logiciel système et le code de calcul que reposeront la productivité et l'efficacité des simulations et des traitements de données.

Mais il en va de même dans le domaine des mathématiques appliquées et des algorithmes : ceux-ci évoluent conjointement et se nourrissent de l'évolution des calculateurs et des puissances de calcul. Il y a souvent autant à gagner dans l'amélioration des algorithmes et des méthodes que dans l'augmentation de la puissance de calcul par une nouvelle génération de calculateurs. Une démultiplication des effets est donc possible. L'histoire des mathématiques appliquées et des algorithmes ne s'est pas arrêtée il y a vingt ans et est loin d'être terminée. L'investissement dans ce domaine reste plus que jamais indispensable.

TROUVER UN NOUVEL ÉQUILIBRE AU SEIN DU SUPERCALCULATEUR

L'intégration de nouveaux, voire de différents moteurs de calcul au sein du supercalculateur n'est pas sans conséquence sur l'architecture

Accords internationaux de coopération dans le domaine du HPC

Le CEA/DAM a établi un réseau de partenaires au meilleur niveau mondial dans le domaine du HPC. Il élabore avec ceux-ci un programme de travail pluriannuel qui conduit au développement ou au partage de logiciels et à une vision commune sur les enjeux du domaine en vue de calculateurs exaflopiques.

Ainsi ont été mis en place :

- ⊙ tout d'abord et depuis près de vingt ans, une collaboration sur la « computer science » avec les homologues américains du CEA/DAM, c.-à-d. les laboratoires américains de la **NNSA** du département de l'Énergie (**DOE**). Les thèmes de travail traitent de l'architecture des supercalculateurs, du middleware, des modèles de programmation, des algorithmes de **maillage** et de visualisation, et des logiciels et algorithmes de traitement des données ;
- ⊙ depuis plus de dix ans, une collaboration avec

le Jülich Supercomputing Centre (JSC) en Allemagne. Ce centre, qui héberge aujourd'hui un supercalculateur Atos/Bull parmi les plus puissants, partage une communauté d'intérêt avec le CEA/DAM pour le renforcement d'un acteur européen du HPC. Les travaux portent sur les architectures modulaires de supercalculateurs ainsi que sur les spécifications d'un microprocesseur européen ;

- ⊙ depuis 2017, les experts du CEA/DAM participent aux travaux du CEA dans le cadre de la collaboration avec l'institut japonais Riken. Le périmètre de cette collaboration couvre le développement de composants **logiciels open source** sur des architectures x86 et ARM. Les sujets prioritaires concernent les environnements et les langages de programmation, les supports d'exécution et les ordonnanceurs de travaux optimisés en fonction de l'énergie.

Relations avec les universités, les laboratoires de recherche et participation aux cursus de formation

La préparation de l'avenir de la simulation numérique au CEA/DAM s'appuie sur le système de formation et de recherche français. Celui-ci est mondialement reconnu dans les domaines des mathématiques appliquées, de l'informatique et des sciences du numérique. Voici quelques-unes des collaborations les plus exemplaires avec le système de formation et de recherche :

- ⊙ avec l'École normale supérieure Paris-Saclay sur l'intelligence artificielle et les méthodes de résolution des équations aux dérivées partielles (ou EDP) ;
- ⊙ avec l'université de Versailles-Saint-Quentin sur les outils logiciels pour l'analyse et l'optimisation des performances des codes de calcul ;
- ⊙ avec l'université de Reims Champagne-Ardenne pour le développement d'algorithmes pour GPU et les supercalculateurs hétérogènes ;
- ⊙ avec l'université Pierre-et-Marie-Curie (Laboratoire Jacques-Louis-Lions) sur les méthodes de résolution des EDP et les schémas numériques ;
- ⊙ avec l'Institut polytechnique Bordeaux INP et l'Inria Bordeaux-Sud-Ouest sur la modélisation et le calcul numérique. De plus, les experts du CEA/DAM ont participé à l'élaboration de cursus de formation dans le domaine du calcul haute performance, des mathématiques appliquées et de la sécurité informatique. Notamment :
 - ⊙ À l'université de Versailles-Saint-Quentin, le master SeCRets (Sécurité des Contenus, des Réseaux, des Télécommunications et des Systèmes) propose une solide formation en cryptographie, en sécurité des réseaux, des télécommunications ou des systèmes d'exploitation (Unix, Windows...), tandis que le master CHPS (Calcul Haute Performance et Simulation) vise à former des cadres scientifiques de haut niveau à même de maîtriser deux évolutions technologiques majeures: l'utilisation des ordinateurs à haute performance et la simulation numérique.
 - ⊙ À l'ENSIE (École nationale supérieure d'informatique pour l'industrie et l'entreprise), le parcours « Calcul intensif et données massives » prépare les futurs ingénieurs aux technologies des supercalculateurs et du traitement des données.
 - ⊙ À l'École polytechnique, le cours d'informatique « algorithmique parallèle et distribuée ».

Les étudiants qui le souhaitent sont invités à faire un stage de fin d'études ou poursuivre par une thèse au sein des laboratoires du CEA/DAM sur des sujets d'exception dans un environnement informatique de classe internationale.

générale de celui-ci. Pour le réseau d'interconnexion, les caractéristiques habituelles débit/**latence**, certes importantes, doivent être complétées par la capacité à gérer et agréger un très grand nombre de messages par seconde, en relation avec l'explosion du nombre de cœurs. Le découplage des fonctions calcul et entrées/sorties au sein du supercalculateur paraît aussi indispensable. Les entrées/sorties pourraient être assurées par des nœuds spécialisés facilitant et régulant les échanges avec les nœuds de calcul sur la base de protocoles de communication adéquats ; la mise en œuvre de ce concept est expérimentée dans le projet européen Sage2 (voir **encadré** « Quelques projets collaboratifs auxquels participent les équipes du CEA/DAM »). Plus généralement, une plus profonde hiérarchisation du système de stockage associée à l'utilisation de nouveaux médias et contrôleurs programmables doit permettre la distribution et l'exploitation des résultats parallèlement à leur production.

LA RECHERCHE DES VOLUMES

Cependant, la technologie ne peut faire abstraction du facteur économique. Malgré l'élargissement de son marché, notamment dans l'industrie, le HPC ne peut soutenir à lui seul les investissements nécessaires à son évolution. C'est la raison pour laquelle ce secteur est largement subventionné par les États, que ce soit aux États-Unis, en Chine ou au Japon. Le HPC devra aussi s'adapter aux besoins des marchés en forte croissance du traitement et de l'analyse des masses de données portés par les systèmes autonomes, le déploiement de capteurs au sein des systèmes de production ou le suivi médical des patients. Ces nouveaux marchés ouverts grâce aux avancées de l'**intelligence artificielle** conditionneront l'évolution des architectures.

Par ailleurs, aux deux bouts de la chaîne de conception/production de microprocesseurs, de nouveaux acteurs spécialisés se renforcent, d'un côté dans la conception d'architectures et

la commercialisation de la propriété intellectuelle associée, de l'autre dans la seule fabrication de semi-conducteurs. Ce mouvement ouvre des perspectives qui pourraient conduire à la fin du monopole d'un seul pays, aujourd'hui les États-Unis, sur la conception de microprocesseurs à haute performance. Pour mieux appréhender l'évolution du contexte économique et technologique qui sous-tend l'évolution du HPC, et en saisir les opportunités, le CEA/DAM s'appuie sur un réseau international de partenaires et participe à un grand nombre de projets collaboratifs dans le cadre du programme européen Horizon 2020 (voir encadrés « Quelques projets collaboratifs auxquels participent les équipes du CEA/DAM » et « Accords internationaux de coopération dans le domaine du HPC »).

Le supercalculateur de classe exaflopique qui succédera à Tera1000, Exa1, a été défini par les équipes d'Atos/Bull et du CEA/DAM. Il capitalise l'expérience acquise avec Tera1000 et intègre certains des résultats issus des nombreuses collaborations déjà mentionnées. Pour le CEA/DAM, concevoir avec Atos un nouveau supercalculateur de plusieurs ordres de grandeur plus puissant s'inscrit dans un schéma plus global d'évolution du volet « numérique » du programme Simulation qui le conduit à coordonner un ensemble de travaux dans les domaines du logiciel, des algorithmes et des méthodes numériques. Le CEA/DAM s'inscrit depuis de très nombreuses années au sein d'un réseau dense de coopérations qu'il continue à renforcer et élargir, et qui l'assure du renouvellement de son potentiel d'idées, d'innovations et de son capital humain. ■

Accélérateur de Data Science – la chaire « Industrial Data Analytics & Machine Learning »

L'ENS Paris-Saclay, Atos et le CEA ont créé fin 2016 la chaire « Industrial Data Analytics & Machine Learning ». La formation, la recherche et le soutien au développement de startup sont les piliers d'un projet unique de création d'une filière Data Science en France, au cœur du plateau de Saclay. Ce partenariat s'est depuis lors élargi à l'École nationale supérieure d'informatique pour l'industrie et l'entreprise (ENSIIE) et aux sociétés Bertin IT, Michelin, SNCF et à la Banque de France, qui ont rejoint les fondateurs au sein de la chaire. Il s'inscrit dans une stratégie du CEA, et en particulier de sa Direction des applications militaires (la DAM), pour soutenir et dynamiser les compétences scientifiques et industrielles françaises nécessaires à la réalisation de ses missions. Ainsi, depuis le début des

années 2000, la DAM collabore avec Bull, maintenant Atos, pour développer des calculateurs de puissance performants, qui répondent aux besoins de la défense, mais qui profitent également à la recherche et à l'industrie. Ces supercalculateurs génèrent des données massives qui posent la question cruciale de leur traitement et de leur analyse, c'est-à-dire la question du big data. La chaire « Industrial Data Analytics & Machine Learning » mobilise sur le sujet du big data et de l'intelligence artificielle les étudiants, enseignants, chercheurs et industriels au meilleur niveau. Elle s'articule autour de trois axes :
 © le premier concerne la formation (initiale et continue) avec pour pilier le master « Mathématiques, Vision et Apprentissage » (MVA) orienté sur la modélisation des données, les tech-

niques mathématiques et algorithmiques ;
 © le deuxième porte sur la recherche, avec pour pilier l'équipe « Machine Learning & Massive Data Analysis » créée en 2012 autour de 12 chercheurs au sein du Centre de mathématiques et de leurs applications qui a fusionné début 2020 avec l'unité CognaG pour former le Centre Borelli ;
 © enfin, le troisième axe est celui de l'innovation et du partage d'expérience au bénéfice de la dynamique de startup et de petites et moyennes entreprises du plateau de Saclay. Cette chaire permettra l'émergence de nouveaux métiers et de nouveaux outils dans le domaine de la Data Science pour répondre à des enjeux stratégiques, technologiques ou économiques majeurs de notre société.

CHAPITRE 8

LE CALCUL HAUTE PERFORMANCE : UN TRAVAIL D'ÉQUIPES

Le programme Simulation ne peut atteindre ses objectifs sans les compétences et la motivation des hommes et des femmes qui le mettent en œuvre. Dans sa dimension « calcul et simulation », il repose sur un travail d'équipes aux compétences complémentaires alliant les mathématiques à l'informatique.

Les métiers du HPC au CEA/DAM

— Tous les métiers du calcul à haute performance sont représentés au CEA/DAM. Dans un environnement technologique en constante évolution, les équipes doivent s'adapter et renouveler leurs compétences.

Le volet numérique du programme Simulation, qui a conduit à l'adoption, par le CEA/DAM, du calcul en mode parallèle et à l'utilisation de logiciels issus de la communauté internationale de « l'**open source** », a conduit à des évolutions importantes des métiers des ingénieurs-chercheurs en calcul scientifique. S'y ajoutent la participation du CEA/DAM, en co-design avec Atos/Bull, au renouveau d'une filière industrielle européenne de supercalculateurs, ainsi que l'implication croissante dans de nombreux projets de R&D sur le **HPC** financés soit par les pouvoirs publics français dans le cadre de la participation au pôle de compétitivité **SYS-TEM@TIC Paris-Région**, soit par la Commission européenne dans le cadre du programme **Horizon 2020**.

Les équipes du CEA/DAM sont impliquées dans tous les métiers du calcul scientifique :

- de la conception d'architecture de supercalculateurs et de centres de calcul, jusqu'à leur mise en service et leur exploitation. Cette maîtrise de l'ensemble du cycle de vie d'un supercalculateur procure aux salariés concernés d'exceptionnels savoir-faire, connaissances et compétences dans leur métier ;
- comme éditeurs de logiciels spécifiques à l'environnement du calcul à haute performance (architecte/concepteur/développeur/intégrateur) dédié aux grands systèmes informatiques. Les équipes concernées interviennent dans toutes les phases de conception, développement et maintien en conditions opérationnelles de logiciels du calcul scientifique.

Ils opèrent en tant qu'architecte/concepteur, administrateur, développeur de logiciels scientifiques, mathématicien en analyse numérique ou en traitement des données, ainsi que comme responsable opérationnel de centre de calcul, responsable de la sécurité informatique, expert en infrastructures de centre de données, pilote d'installations, responsable du support applicatif...

Les métiers d'architecte et concepteur sont dédiés aussi bien au niveau système (centre de calcul) que sous-systèmes (calculateurs, stockage, réseaux, infrastructures/utilités techniques), associés à l'ensemble des logiciels utiles au fonctionnement du centre de calcul et à son exploitation (pile logicielle système, traitement et gestion des données des codes de simulation dans un contexte HPC, sécurité informatique...).

Les évolutions permanentes des technologies ainsi que les objectifs du programme Simulation (voir article « De nouveaux enjeux pour la prochaine décennie ») conduisent à adapter régulièrement les métiers et les compétences. En relation avec la Direction des relations humaines, les responsables des équipes concernées s'appuient sur la gestion prévisionnelle des emplois et compétences (GPEC) pour adapter et faire évoluer les compétences des salariés aux besoins opérationnels et préparer l'avenir. La formation professionnelle représente un des leviers essentiels à l'adaptation des compétences, aussi bien dans les domaines scientifique et technique que dans la conduite de projets. Pour y répondre, un programme pluriannuel de formation est mis en place.

Nicolas Legoff découpe le domaine de calcul

Préparer les données pour un code de calcul demande de développer des logiciels adaptés. Un travail qui nécessite des compétences informatiques, mais aussi le goût d'aider les utilisateurs.

— Pour la plupart des codes de simulation, il est nécessaire de découper le domaine de calcul en une multitude d'éléments simples. Cette étape, appelée « génération du **maillage** », s'effectue à l'aide de logiciels dédiés. « Cela fait partie des nombreux outils que nous développons nécessaires pour mener à bien les calculs », indique Nicolas Legoff, qui effectue cette tâche au CEA/DAM au sein d'une petite équipe. Au quotidien, cela consiste à intégrer des logiciels existants, à les adapter ou à en développer afin que les utilisateurs génèrent eux-mêmes ces éléments simples.

R&D et support aux utilisateurs

« C'est intéressant car nous répondons précisément aux besoins, estime Nicolas Legoff. Nous avons une activité de R&D et de prospective. Parallèlement, nous interagissons avec les utilisateurs. Ceux-ci demandent régulièrement de nouvelles fonctionnalités. Mais ils ont souvent du mal à préciser leurs besoins ou prendre en compte a priori ce qu'il est réellement possible de réaliser. C'est à nous de leur proposer des solutions pertinentes. »

Il existe de nombreuses méthodes pour générer les maillages, mais elles ne donnent pas toujours le résultat idéal. L'une des tâches de l'équipe est donc de faire progresser ces méthodes, mais également d'automatiser le processus. « Aujourd'hui, nos produits sont complexes à utiliser », reconnaît-il. Concrètement,

générer du maillage consiste à découper un modèle géométrique en un modèle topologique. Mais de nombreux choix sont possibles, et celui fait au départ peut contraindre fortement la suite des travaux et empêcher de mailler correctement. Faire les bons choix tout au long du processus est donc crucial.

Coopérer pour avancer

Les résultats de leurs travaux sont publiés dans les revues professionnelles, partagés lors de colloques internationaux ou de collaborations industrielles et académiques. « Cela peut être utilisé dans d'autres sciences : les besoins de maillage sont largement partagés », souligne le jeune ingénieur. Actuellement, parmi les pistes étudiées figurent la génération semi-automatique et l'aide à la prise de décisions : « Si on ne peut pas totalement automatiser la génération d'un maillage, il est possible de permettre à l'utilisateur de débloquer nos algorithmes lorsqu'ils échouent, ou alors guider l'utilisateur pour que sa tâche soit plus facile. »

Pour ce métier, il faut bien sûr des compétences en développement logiciel, mais pas seulement. « Il faut aimer être en support, indique Nicolas Legoff. Et cela nécessite de la curiosité : il reste plein de choses à faire. Il faut notamment que nous arrivions à rendre nos outils plus simples. Mais simplifier, c'est complexe ! »

NOUVELLES TECHNOLOGIES : EXASCALE, INTELLIGENCE ARTIFICIELLE, RENFORCER ET ÉLARGIR LES COMPÉTENCES

Pour entrer dans l'ère du calcul **exaflopique**, le CEA/DAM poursuit sa participation active au co-design avec Atos/Bull. Des ruptures technologiques sont nécessaires pour y parvenir,

notamment pour améliorer l'efficacité énergétique, réguler les flux et gérer les volumes considérables de données produites par des simulations de plus en plus précises de phénomènes multiphysiques et multidimensionnels. Le programme de recherche en place vise aussi à renforcer la maîtrise des éléments essentiels de la chaîne de valeur du HPC : architecture

système, réseau d'interconnexion, système de refroidissement, **processeur**, procédés industriels de réalisation et d'installation...

L'évolution des architectures de calculateurs, l'objectif de l'**exascale** et l'exploitation encore plus poussée du **parallélisme**¹ constituent un véritable défi pour les codes de calcul et les logiciels scientifiques du CEA/DAM. Un renforcement des profils dans le domaine des mathématiques appliquées, de l'informatique scientifique ou du génie logiciel est donc requis.

La participation du CEA/DAM au projet européen de conception de microprocesseurs (**EPI**), élément stratégique de la chaîne de valeur du HPC, fait apparaître un besoin nouveau d'expertise (amélioration des outils de compilation, rétro-ingénierie et analyse des vulnérabilités

1_Sur les systèmes dits « généralistes », on observe jusqu'à 3 niveaux de parallélisme : la parallélisation inter-nœuds ou intra nœuds à mémoire distribuée, la parallélisation intra-nœud à mémoire partagée et les opérations vectorielles. Les nouvelles générations de supercalculateurs de classe exaflopique disposeront de plusieurs millions de cœurs de calcul.

liées aux microarchitectures). La connaissance fine des caractéristiques des éléments de calcul permettra de concevoir les **intergiciels** nécessaires au développement de codes efficaces sur différentes cibles matérielles.

Le calcul à haute performance entre dans une nouvelle ère de changement. Il s'étend et intègre de nouveaux domaines dont l'utilisation en simulation numérique est porteuse de progrès : **intelligence artificielle**, environnements d'exécution pour les volumes massifs de données (big data)...

La valorisation des données, qu'elles soient issues des simulations, des instruments ou des expériences, est encore à améliorer. Le développement de l'usage au CEA/DAM de ces technologies nécessitera la mise en place de compétences adaptées, avec des métiers de type :

- ⊙ architecte appliqué aux données : collecte, fusion et architecture de plateforme pour la capitalisation/valorisation des données ;
- ⊙ mathématicien appliqué en sciences des données.

Enseignement dans des écoles d'ingénieurs et les universités (ENSIIE, Centrale-Supélec, École polytechnique, UVSQ...)

Plusieurs dizaines d'experts du CEA/DAM dispensent des cours sur le calcul haute performance ou les mathématiques appliquées dans plus d'une demi-douzaine d'écoles d'ingénieurs, pour des classes de 2^e et de 3^e années, ou dans les masters universitaires. Dans le domaine du HPC, l'enseignement est centré sur les architectures et leur constitution, l'ensemble des logiciels nécessaires au fonctionnement, les systèmes de fichiers et le stockage objet, les réseaux d'administration et d'interconnexion (topologie), les modèles spécifiques de programmation : le parallélisme à base de threads, les multiples modèles de programmation parallèle, la programmation scientifique (résolution numérique, méthodes de **discrétisation**, visualisation des données, etc.). L'enseignement ainsi dispensé permet à ces établissements de bénéficier d'un savoir-faire exceptionnel issu des meilleurs experts, et en retour de favoriser l'arrivée au sein des équipes du CEA/DAM de stagiaires/doctorants alimentant le vivier de recrutement.



Élodie Ardoin, ingénieure réseau

Après mes études d'ingénieure en électronique et traitement du signal, je souhaitais partir à l'étranger, raconte-t-elle. Or, le Département analyse, surveillance, environnement du CEA/DAM proposait des postes en volontariat international en entreprise (VIE).

— « Je suis ainsi partie deux ans en Bolivie pour aider à la mise en place d'un réseau infrasonique destiné à la surveillance des essais nucléaires. »

Concevoir, évaluer, tester

Puis, saisissant l'occasion de se spécialiser, elle intègre les équipes informatiques en 2009 à un poste d'architecte des réseaux d'entreprise, avant d'évoluer comme architecte réseau pour les grands calculateurs. « Le laboratoire réseau conçoit de nouvelles solutions pour les réseaux informatiques, décrit Élodie Ardoin. Cela consiste à évaluer les technologies disponibles sur le marché ou à l'horizon 2022, afin de vérifier si elles répondent à nos critères. Puis j'imagine des solutions de test et d'évaluation de ces solutions, et je prévois les plannings de test et de déploiement, en collaboration avec d'autres équipes. Enfin, je m'intéresse à la configuration logicielle, en analysant attentivement la documentation technique des fabricants, en ayant des discussions théoriques avec eux sur leurs solutions, et en menant des tests. On peut alors décider si la solution correspond à nos besoins. »

Une fois l'architecture choisie, son travail n'est pas fini : il faut maintenant intégrer cette solution dans l'environnement de production du CEA/DAM. Il s'agit

notamment de superviser son déploiement : préparation du financement, commande et réception du matériel, organisation du travail de ceux qui installent et configurent... Un métier très varié, d'autant qu'il faut parallèlement suivre la R&D du domaine, donner un feedback au constructeur, l'assister sur la spécification de nouveaux produits, l'aider à mettre en place des campagnes de tests... « Ce que j'aime dans mon activité, c'est l'aspect innovation, développement de nouveaux produits, souligne l'ingénieure. Les défis technologiques sont passionnants. De plus, au CEA, on est poussés à trouver des idées nouvelles pour résoudre plein de problèmes. La R&D sur les réseaux est très rare en France, ce qu'on fait au CEA est stimulant. »

Aider l'innovation française

Élodie Ardoin apprécie également de travailler avec le constructeur Atos, qui fournit des supercalculateurs et les trois dernières générations de machines Tera. « Collaborer avec eux permet d'aider l'innovation française et européenne, souligne-t-elle. Atos est en compétition avec les mastodontes du secteur, essentiellement américains. Or, il est important de soutenir une entreprise française et européenne, pour garder notre indépendance technologique. »

ANTICIPER LES RUPTURES DE L'INFORMATIQUE QUANTIQUE

L'informatique quantique, comme son nom l'indique, est à la croisée entre informatique et physique quantique. Cette technologie révolutionnaire mais encore balbutiante se trouve aujourd'hui à une étape équivalente à celle de l'industrie informatique au milieu des années cinquante. Néanmoins, le CEA/DAM, en relation avec ses partenaires du **CCRT** (voir encadré « Se préparer à l'informatique quantique au CCRT »), s'est déjà engagé dans une phase d'exploration qui concerne aujourd'hui les spécialistes en physique plutôt que ceux en informatique ou en mathématiques appliquées.

CONCLUSION

Pour faire face aux évolutions architecturales et technologiques du HPC, aux enjeux futurs du programme Simulation et au renouvellement des équipes dans ce domaine, le CEA/DAM a établi un réseau de collaborations et de partenariats avec un grand nombre d'écoles d'ingénieurs et d'universités. Ce réseau (voir encadré « Enseignement dans des écoles d'ingénieurs et les universités (ENSIIE, Centrale-Supélec, École polytechnique, UVSQ...) ») doit permettre de constituer le vivier de talents qui alimentera les prochaines générations d'ingénieurs-chercheurs du CEA/DAM. ■



Cédric Chevalier optimise les performances des simulations numériques

Améliorer les performances des calculs sur les supercalculateurs : c'est le métier de Cédric Chevalier. Il s'attaque aux problèmes informatiques les plus complexes tels que le partitionnement de graphes, tout en restant proche des utilisateurs.

— « Dans les machines Tera, le calcul se fait en parallèle sur des mémoires presque autonomes les unes des autres, explique l'ingénieur-chercheur. Pour que le calcul soit efficace, il faut très peu de communication entre ces entités. Mon travail consiste donc à développer les outils pour répartir au mieux les calculs afin que chaque entité soit le plus autonome possible, et ainsi utiliser au mieux la machine. »

Aspect artisanal

Cédric Chevalier est un spécialiste de la partition de graphe, puisqu'il a fait sa thèse sur ce sujet à l'Inria de Bordeaux. Il préparait alors les données des calculs du Cesta (le Centre d'études scientifiques et techniques d'Aquitaine du CEA/DAM à Bordeaux) pour Tera1. Il n'était donc pas trop dépaysé lorsqu'il est ensuite arrivé à Bruyères-le-Châtel !

Pour répartir ses calculs, Cédric Chevalier dispose d'outils, les partitionneurs de graphes. « Mais le pilotage de ces outils reste très manuel, il faut effectuer de multiples réglages, note-t-il. Leur utilisation dans un contexte industriel n'est pas triviale, il reste un côté artisanal. » Les partitionneurs de graphes développés dans le monde académique ne conviennent que partiellement au contexte des simulations du CEA/DAM. C'est pourquoi il reste nécessaire de les adapter, ou d'en développer de nouveaux. C'est l'objet de plusieurs thèses conduites au CEA/DAM.

Problèmes concrets

Cependant, partitionner les graphes reste un problème considéré comme difficile en informatique. « Trouver la meilleure solution nécessiterait de toutes les parcourir, explique Cédric Chevalier. C'est impossible, et on développe donc des méthodes pour trouver rapidement une solution acceptable, même si ce n'est pas la meilleure. Pour cela, on adapte

les recherches du monde entier à la résolution de problèmes concrets, c'est passionnant ! Nous coopérons avec les universités et les organismes de recherche, comme les National Laboratories du département de l'Énergie américain. »

Taille humaine

Ce mélange de recherche et d'ingénierie plaît particulièrement à Cédric Chevalier. Lui qui a mené des études d'ingénieur suivies d'une thèse retrouve ces deux aspects complémentaires. « On n'apporte pas seulement la preuve de concept, on prend le temps de réécrire les algorithmes pour qu'ils soient les plus performants possibles, souligne l'ingénieur-chercheur. On est en lien avec la production, et proches des utilisateurs. » Ces outils sont régulièrement remis en question, car les machines évoluent. Étonnamment, cette évolution conduit à mieux appréhender les problèmes traités. « Aujourd'hui, on comprend mieux pourquoi une méthode fonctionne et pas une autre, alors qu'à la base, nos choix étaient fondés sur l'intuition », observe Cédric Chevalier. Il apprécie également de travailler dans un organisme de la taille du CEA/DAM : suffisamment gros pour disposer de beaux moyens de travailler, et néanmoins à l'échelle humaine. « On a une vision assez globale de ce que l'on fait, alors qu'aux États-Unis, où j'ai passé deux ans et demi, ils sont plus spécialisés et voient moins l'impact de leurs algorithmes sur les codes de calcul. » Et à l'inverse de l'industrie, où les questions de temps et de budget sont primordiales, « on ne fait pas de compromis pour des raisons autres que techniques, on applique ce que l'on pense être le mieux. Ce sont des projets longs, dans lesquels on n'évite pas les difficultés : on les attaque de front. » Patience et persévérance sont indispensables, mais la satisfaction retirée est à la hauteur !

Glossaire

A

- ABINIT** Logiciel sous licence libre publié en open source utilisé pour la recherche en science des matériaux. ABINIT permet d'obtenir les propriétés locales d'un matériau en résolvant de manière la plus exacte possible les lois de la mécanique quantique. Il est développé dans le cadre d'un projet collaboratif auquel le CEA/DAM contribue activement.
- Action malveillante** Action dont le but est de modifier le flux de commandes d'un système d'information pour lui faire exécuter des commandes choisies par l'intrus.
- Alien** Bibliothèque qui fournit des interfaces unifiées pour l'utilisation de solveurs linéaires open source comme Hypre ou ceux de Trilinos.
- AMR** (Adaptative Mesh Refinement) : algorithme de raffinement adaptatif de maillage.
- API** (Application Programming Interface) : ensemble de fonctions, procédures, services mis à disposition par une bibliothèque logicielle, un système d'exploitation ou un composant logiciel pour programmer une application.
- Architecture cloisonnée** Architecture interdisant, par défaut, toute communication entre les différents éléments ; seules les communications nécessaires entre les différents composants sont autorisées.
- ASIC** (Application specific integrated circuit ; en français, circuit intégré propre à une application) : circuit intégré spécialisé. En général, il regroupe sur la même puce un grand nombre de fonctionnalités uniques ou sur mesure.
- ASMP-A** Missile air-sol moyenne portée – amélioré ; l'ASMP-A emporte la tête nucléaire aéroportée (TNA).
- AVIESAN** Alliance d'organismes français de recherche publique pour les sciences de la vie et de la santé.

B

- Barrière** En programmation parallèle, action consistant à synchroniser toutes les entités présentes dans un modèle de programmation. Ainsi, la barrière se termine quand tous les flots sont arrivés à cette barrière.
- Boîte de calcul** Domaine spatio-temporel sur lequel s'effectue la simulation numérique par ordinateur.
- Bombe logicielle** Dispositif programmé pour déclencher une action (virus, vers...) sur un événement particulier (date, commande exécutée...).
- BXI** Bull eXascale Interconnect : réseau d'interconnexion entre les processeurs d'un supercalculateur. Ce système est déployé sur TERA1000 depuis 2017.
- Bit** Grandeur de base en informatique. Un bit est un élément pouvant être égal à 0 ou à 1 (deux valeurs possibles donc). Un fichier est un ensemble de bits. Un ensemble de bits forme ce qu'on appelle un mot binaire (un octet fait 8 bits, unité la plus petite généralement utilisée pour former des mots).

C

- Byte/Flop** Unité de bande passante mémoire.
- CAD** Collecteur analyseur de données (pour la santé). L'objectif de ce centre national d'analyse des données est de traiter, analyser et exploiter les données générées dans le cadre de parcours de soin ou de projets de recherche.
- Calcul *ab initio*** Calcul des propriétés d'un système à partir des principes fondamentaux de la physique, sans connaissance expérimentale *a priori* du système considéré et donc sans paramètre empirique.
- CAO** (Conception Assistée par Ordinateur) : désigne l'ensemble des logiciels et des techniques de modélisation permettant de concevoir, tester à l'aide de simulations numériques sur ordinateur et fabriquer des produits comme un moteur ou un avion ou des outillages pour les fabriquer.

CCRT	Centre de Calcul Recherche et Technologie du CEA, hébergé au TGCC. Le CCRT est ouvert à la communauté civile de la recherche et de l'industrie.
Cesimat	(Centre d'Expertise en Simulation des Matériaux) : consortium constitué pour répondre à un appel d'offres « Calcul Intensif et Simulation Numérique », lancé dans le cadre des Plans d'Investissement d'Avenir (PIA) - comprenant le CEA, Michelin, L'Oréal, Safran, Faurecia, Kitware, Paratools et l'École Centrale de Nantes
Champ de vecteurs	Notion mathématique utilisée en physique pour modéliser par exemple la vitesse et la direction d'un fluide en mouvement dans l'espace, ou la valeur et la direction d'une force, comme la force magnétique ou gravitationnelle, qui évoluent d'un point à son point voisin. En termes mathématiques, un champ de vecteurs ou champ vectoriel est une fonction qui associe un vecteur à chaque point d'un espace euclidien ou plus généralement d'une variété différentielle (voir WIKIPEDIA).
CNRS	Centre national de la recherche scientifique
Client / serveur	Logiciel où un processus, dit client, est situé sur le poste de travail d'un ingénieur et communique avec un ou plusieurs processus, dits serveur(s), situé(s) par exemple sur un supercalculateur. De façon simplifiée, le processus client sert à visualiser et interagir avec les données disponibles côté serveur(s) sur le supercalculateur.
Cœur/cœurs de calcul	Un cœur de calcul est une unité d'exécution capable d'effectuer des calculs arithmétiques et logiques simples tels que l'addition de nombres réels ou la gestion de l'exécution d'une instruction. C'est donc un élément, sous-composant d'un processeur moderne, constitué d'un ensemble de circuits capables d'exécuter des programmes de façon autonome.
Complexité	Pour un algorithme, le nombre d'opérations ou de quantité de mémoire nécessaires à sa réalisation (fonction de la taille des données du problème, évaluée souvent dans le « pire des cas », parfois « en moyenne » suivant une mesure appropriée).
Compromission	Action de rendre un système d'information non sûr après qu'une action malveillante a été conduite et a pu réussir.
Convergence	Propriété mathématique d'un schéma numérique assurant que l'écart (au sens d'une norme mathématique) entre la solution approchée et la solution exacte de l'équation représentée par le schéma numérique tend vers 0 lorsque le pas de discrétisation tend vers 0 (ou lorsque chacun des pas globaux associés aux différentes directions tend vers 0).
CPU	(Computing Processing Unit, unité centrale de traitement ou encore processeur) : composant électronique d'un ordinateur qui exécute les programmes.
dB	(Décibel) : unité de grandeur définie comme dix fois le logarithme décimal du rapport entre deux puissances. On utilise une échelle logarithmique car la sensibilité humaine suit en général une loi de puissance entre puissance du signal et stimulus perçu.
DDR	(Double Data Rate) : technique d'accélération des composants d'ordinateur, s'applique principalement aux mémoires.
Défense en profondeur	Technique permettant d'utiliser plusieurs techniques de sécurité pour réduire le risque qu'un composant particulier de sécurité compromis ou défaillant impacte la sécurité globale de l'ensemble du système d'information.
Dépouillement	Action d'extraire, visualiser et analyser des informations significatives à partir de données, généralement de grandes tailles, produites par un code de simulation.
DFT	(Density fonctional theory ou théorie de la fonctionnelle densité) : proposée par Hohenberg et Kohn en 1964, la DFT est une théorie exacte qui exprime l'énergie comme une fonctionnelle de la seule densité.
Discrétisation	C'est la transposition d'un état continu (fonction, modèle, variable, équation) en un équivalent discret. Ce procédé constitue en général une étape préliminaire à la résolution numérique d'un problème ou sa programmation sur machine.
DMFT	(Dynamical mean field theory ou champ moyen dynamique) : méthode de prise en compte des interactions électroniques locales en champ moyen dynamique dans le cadre de la théorie de la fonctionnelle de la densité.
DOE	Department of Energy aux États-Unis.

D

DRAM	(Dynamic Random Access Memory) : mémoire de type DDR utilisée pour réaliser la mémoire centrale sur l'ensemble des calculateurs ou des stations de travail.
DSL	(Domain Specific Language) : langage de programmation dédié, non généraliste, dont les spécifications sont ajustées à un domaine d'application précis pour en exprimer plus efficacement les particularités.
Échanges de messages	Des calculs parallèles ou distribués décomposent un problème complexe en sous-problèmes – les unités de traitement en charge des sous-problèmes ont besoin de partager des informations pour assurer une résolution cohérente globale ; le modèle de programmation associé le plus courant est à base d'échanges de messages (véhiculant des informations et/ou des données).
ECR	(Exascale Computing Research) : collaboration entre le CEA, l'Université Versailles Saint-Quentin et INTEL qui se matérialise par un laboratoire de recherche commun sur les thèmes du calcul parallèle, de l'efficacité énergétique, de l'optimisation de performance pour le HPC et l'analyse de données.
EE-HPC	(Energy Efficient HPC (Working Group)) : groupe de travail international d'initiative américaine qui étudie et encourage la mise en œuvre de mesures d'économie d'énergie et la conception éco-énergétique dans le calcul haute performance (HPC), et stimule le partage d'idées et d'expériences en la matière.
Étude paramétrée/paramétrique	Désigne une expérimentation numérique où une simulation numérique est exécutée un grand nombre de fois en faisant varier des paramètres de la simulation. Ceci peut avoir pour but de dimensionner correctement certains paramètres ou de mesurer la sensibilité à la variation d'un paramètre par exemple.
EPI	(European processor initiative) : projet financé par l'union européenne afin de produire un processeur basse consommation performant, destiné notamment au HPC et à l'automobile.
Epure	(Expérience de Physique Utilisant la Radiographie Éclair) L'installation EPURE située au CEA-Valduc est un élément clé du programme Simulation mis en œuvre par le CEA-DAM. Mise en service en 2014, elle est dédiée aux expériences hydrodynamiques qui permettent de caractériser, avec la plus grande précision, l'état et le comportement des matériaux dans les conditions rencontrées dans la phase de fonctionnement pré-nucléaire des armes.
Équation intégrale	Équation dont l'une des indéterminées est une intégrale.
EuroHPC	Entreprise commune (Joint Undertaking) entre la Commission européenne, des pays participants et des industriels, afin de développer le HPC en Europe (infrastructures, technologies, applications).
European code of conduct for data centres	Ce code de conduite a été créé en 2008 en réponse à l'augmentation de la consommation d'énergie dans les centres de données et à la nécessité de réduire les impacts environnementaux, économiques et de sécurité d'approvisionnement en énergie. Ce code de conduite est une initiative volontaire visant à rassembler les parties prenantes intéressées autour des <i>datacenters</i> : fabricants, fournisseurs, consultants, services publics.
Exaflopique	Néologisme toléré, se dit d'un ordinateur dont la puissance de calcul atteint l'exaflops.
Exaflops	Milliard de milliards d'opérations sur des nombres réels par seconde (10^{18}).
Exascale	Anglicisme toléré, se dit d'un ordinateur capable de réaliser 1018 opérations par seconde, soit à l'échelle de l'exaflops.
Flop	Opération arithmétique en virgule flottante.
Flops ou flop/s	(Floating Point operations per second) : unité de mesure de la performance d'un système informatique représentant le nombre d'opérations en virgule flottante par seconde. 1 PFlop/s correspond à 10^{15} opérations flottantes par secondes. Les ordinateurs les plus rapides actuels sont dans le domaine de la centaine de pétaflops (soit 10^{17} flops).
France Génomique	L'infrastructure France Génomique rassemble la majorité des plateformes de séquençage et de bio-informatique françaises (du CEA, du CNRS, de l'INRA et de l'INSERM) au sein d'un consortium coordonné par le CEA.
Furtivité	Caractéristique d'un objet ayant une surface équivalente radar (SER) réduite afin d'être moins détectable ou identifiable par un radar.

E

F

G	Genci	(Grand équipement national de calcul intensif) : société civile de droit français dédiée au calcul intensif et chargée de mettre à disposition des moyens de calcul et de traitement de données massives performants pour promouvoir la simulation numérique et le calcul intensif auprès de la recherche académique, des industriels.
	Gigaflops ou gigaflop/s	Milliard d'opérations sur des nombres réels par seconde (10 ⁹).
	GPU	(Graphics Processing Unit) : ensemble de processeurs spécialisé dans le calcul des images. Leur utilisation dans les jeux vidéo a stimulé leur développement. Ils possèdent essentiellement des fonctions de calcul, ne peuvent pas fonctionner de façon autonome et doivent être pilotés par un processeur ordinaire (CPU).
	Grappes de serveurs	Regroupement de plusieurs ordinateurs indépendants interconnectés.
H	HDD	(hard disk drive) : mémoire de masse à disque tournant magnétique utilisée principalement dans les ordinateurs.
	Horizon 2020	Horizon 2020 ou H2020 est le 8 ^e programme-cadre européen pour la recherche et le développement (PCRD) qui s'est étendu sur la période 2014-2020.
	HPC	(High Performance Computing ou calcul haute performance) : désigne les activités de calculs réalisées sur un supercalculateur, en particulier pour la simulation numérique.
	HPDA	(High Performance Data Analysis) : technique utilisant le HPC et permettant d'analyser rapidement de grands volumes de données afin d'en déduire des informations.
	HPSS	(High Performance Storage System (origine IBM, 1992)) : système de gestion de stockage « hiérarchique », pouvant déplacer des fichiers volumineux et de grandes quantités de données sur des réseaux comportant différents types et niveaux de ressources (supercalculateurs, disques, bandes...).
	Hypre	Bibliothèque de routines pour résolution (parallèle) de systèmes linéaires. Hypre est soutenu par le Lawrence Livermore National Laboratory.
I	IA/intelligence artificielle	Ensemble de théories et techniques mises en œuvre en vue de réaliser des machines capables d'effectuer des simulations de raisonnements logiques ou de l'apprentissage automatique. La notion a vu le jour en 1950 grâce au mathématicien anglais Alan Turing.
	IFPEN	IFP Énergies Nouvelles.
	Inconnue	En algèbre, élément constitutif d'une question de même nature qu'une équation – façon symbolique de désigner une quantité recherchée, à évaluer ou calculer. L'inconnue permet de décrire une propriété vérifiée par une ou plusieurs valeurs qui prendraient la place de cette inconnue, ces valeurs étant souvent des nombres.
	Inria	Institut national de recherche en informatique et automatique.
	Intercode	Dans une chaîne de calcul complexe enchaînant plusieurs étapes à base de codes de simulations différents, les intercodes désignent des fichiers de données (résultats intermédiaires) produits par un code, pour usage par le code suivant de la chaîne.
	Iso-contour	Ou ligne de niveau : sous-ensemble de l'espace de définition d'une fonction mathématique sur lequel elle prend une valeur donnée constante.
K	KNL	(Intel Knight Landing) : processeur manycore fabriqué par Intel. Le processeur de la machine Tera1000-2.
L	Langage C++	Langage de programmation créé dans les années 1980, par extension du langage C, en y ajoutant des mécanismes de programmation objet (organisation des programmes par classes – modules – regroupant les traitements associés à une catégorie d'objet ; ex : les transformations agissant sur un polygone).
	Langage C#	C# (C « sharp » en anglais, avec un clin d'œil : le « dièse » # évoque le ++ de C++) est un langage de programmation orientée objet. Créé par Microsoft au début des années 2000. Il est dérivé du C++ et très proche du Java.
	Latence	Délai entre le moment où un signal entre et le moment où il ressort d'un système.
	Linpack	Test de performance qui mesure le temps nécessaire à un ordinateur pour résoudre un système d'équations linéaires dense de n équations à n inconnues.

Linux	Système d'exploitation open source de type Unix, créé en 1991 par Linus Torvalds, ayant donné lieu à de très nombreuses variantes et extensions, et désormais très répandu sur tous types de systèmes, jusqu'aux supercalculateurs.
LMJ	Mis en service en 2014 au CEA-CESTA, le Laser Mégajoule ou LMJ est un élément clé du programme Simulation mis en œuvre par le CEA-DAM. Il permet de valider les modèles fondamentaux décrivant la physique du fonctionnement des armes nucléaires et de réaliser des expériences permettant de tester l'enchaînement et l'imbrication de ces modèles. Il est dimensionné pour délivrer, sur une cible millimétrique et en quelques nanosecondes, une énergie lumineuse supérieure à 1 mégajoule.
Logiciels libres ou open source	Logiciels dont l'utilisation, l'étude, la modification et la duplication par autrui en vue de sa diffusion sont permises selon les critères établis par l'Open source initiative.
Loi de Moore	À l'origine, loi empirique exprimée en 1965 dans le magazine Electronics par Gordon E. Moore, ré-évaluée en 1975, postulant un doublement du nombre de transistors sur les circuits semi-conducteurs tous les 2 ans – souvent simplifiée abusivement en « la performance double régulièrement à coût constant ». Conjecture toutefois vérifiée depuis lors, approchant désormais des limites physiques de la miniaturisation, et à nuancer car d'autres considérations pilotent les coûts et les performances effectifs des composants électroniques.
LUSTRE	Lustre est un système de fichiers distribué libre, généralement utilisé pour de très grandes grappes de serveurs. Le nom est la réunion de Linux et cluster. L'objectif du projet est de fournir un système de fichiers distribué capable de fonctionner sur plusieurs centaines de nœuds, avec une capacité d'un pétaoctet, sans altérer la vitesse ni la sécurité de l'ensemble. Lustre est distribué sous licence GPL.
Machine learning	Apprentissage machine : partie de l'intelligence artificielle utilisant des méthodes mathématiques et statistiques pour donner à des programmes la capacité d'« apprendre », s'entraînant à partir de certaines données, puis pouvant en tirer des associations sur de nouvelles données, voire s'améliorer au passage.
Machine virtuelle	Une machine virtuelle est un programme informatique qui se comporte comme un système ou ordinateur réel, à l'intérieur d'un autre système ou ordinateur. La machine virtuelle isole son contenu et ses comportements du reste du système d'accueil. Elle peut ainsi servir à sécuriser des systèmes ; à faire fonctionner des systèmes en dehors de leur environnement d'origine ; à faire cohabiter différents environnements sur un ordinateur unique ; à optimiser le partage d'un ordinateur entre plusieurs usages ayant chacun leur machine virtuelle.
Maillage	Partition du domaine de calcul en sous domaines appelées mailles afin d'y résoudre les équations de la physique par approximation numérique.
Maillage AMR	Maillage où chaque élément peut lui-même être découpé en une sous-grille.
Maillage non structuré	Maillage dont les éléments sont quelconques et où aucune structure n'apparaît. C'est par exemple le cas des maillages simpliciaux composés de triangles en 2D et de tétraèdres en 3D.
Maillage structuré	Maillage correspondant à une grille où les éléments sont des quadrilatères (2D) ou des hexaèdres (3D) (isomorphe à une matrice 2D ou 3D).
Maillage structuré par blocs	Ensemble de blocs formant un maillage grossier où chaque bloc correspond à un maillage structuré.
Maille	Élément d'espace issu du découpage du domaine de calcul sur lequel est effectuée la simulation numérique.
Manycore	Type d'architecture de processeur basée sur un très grand nombre de cœurs par puce physique (« pluricœurs »).
MapReduce	Modèle de traitement distribué, et implémentation associée, pour des données de grande taille ou « big data » (inventé par Google). MapReduce comporte deux étapes pour produire des correspondances clés-valeurs (utiles notamment aux recherches de type web) ; Map découpe un problème, le découpe en sous-problèmes, et le délègue à d'autres unités (étape « descendante ») ; Reduce (étape « remontante ») agrège et consolide les données de plusieurs sous-problèmes et renvoie un résultat global.

M

MCDRAM	(Multi-Channel DRAM) : mémoire rapide directement raccordée sur les processeurs KNL.		
MESRI	Ministère de l'enseignement supérieur, de la recherche et de l'innovation.		
Métadonnées	Donnée servant à définir ou décrire une autre donnée, par exemple pour une donnée de type image ou photo, les métadonnées peuvent contenir la date et la localisation du lieu de production ou d'enregistrement.		
Metis	Util open source de partitionnement de graphes, de maillages de type éléments finis.		
Middleware/intergiciel	(Ou logiciels médiateurs) : ensemble de logiciels, regroupés au sein de bibliothèques, permettant de développer efficacement des programmes ; ces couches logicielles sont intermédiaires entre le logiciel système et le code de calcul.		
Modèle de données	Un modèle de données définit un principe d'organisation et de représentation de données ; dans le cadre du HPC, de manière à décrire et partager et exploiter de manière cohérente des données complexes entre codes de calcul, de visualisation, de post-traitement (maillages, champs de vitesses, particules...); un modèle précède la mise en place de formats et structures de données l'implémentant en mémoire ou dans des fichiers.		
Monte-Carlo	Méthodes de calcul approché d'une valeur numérique en utilisant des procédés aléatoires, c'est-à-dire des techniques probabilistes, par utilisation de données « tirées au hasard ». Le nom de ces méthodes fait allusion aux jeux de hasard pratiqués au casino de Monte-Carlo.		
MPI	(Message Passing Interface) : MPI est un modèle de programmation à mémoire distribuée reposant sur une interface standardisée. À travers des appels de fonction, il est alors possible d'écrire une application parallèle où chaque entité sera isolée et pourra communiquer avec les autres à travers de messages contenant des données.		
N	Nabla	Le langage Nabla est développé au CEA/DAM, il s'agit d'un DSL ; Nabla permet l'écriture de schémas numériques.	
	NabLab	L'outil Nablab fournit un environnement de programmation pour le langage spécifique Nabla.	
	NIC	(Network Interface Controller/Connect) : ensemble de composants électroniques qui assurent l'interface entre l'équipement, la machine dans laquelle il est monté, avec les machines connectées sur le même réseau.	
	NNSA	National nuclear security administration (États-Unis), agence fédérale dépendante du DOE en charge de la dissuasion américaine.	
	Nœud/nœuds de calcul	Un nœud de calcul, véritable brique de base du supercalculateur, regroupe un assemblage de processeurs partageant la même vision de la mémoire. Il est alors possible de partager des données entre les différents cœurs d'un même nœud.	
	Non-régression	Propriété d'un logiciel qui signifie qu'il n'y a pas de défaut introduit à l'occasion de corrections de bogues ou de n'importe quel changement établi dans le logiciel comme l'ajout de nouvelles fonctionnalités, la modification de fonctionnalités existantes ou la modification d'un composant externe au logiciel lui-même.	
	O	Octet	Unité de stockage de données. 1 octet correspond à 8 bits et permet de stocker un caractère. Un nombre réel double précision est stocké sous 8 octets selon la norme IEEE. Un octet se traduit par byte en anglais. Pour exprimer des débits on utilise le petit b pour désigner les débits en bits et B pour désigner les débits en bytes (1 GB/s = 8 Gb/s).
		Téraoctet, pétaoctet	Multipléte de 8 bits.
		OpenMP	Modèle de programmation à mémoire partagée basée sur des directives de compilation et une interface standardisée. Certaines directives permettent ainsi de rendre des parties d'une application parallèle en créant des entités qui partageront la même vision de la mémoire et, par conséquent, de certaines données.
		Open Cascade	Open CASCADE est un environnement logiciel Open Source pour la conception et la modélisation 3D, développé à partir des années 1990. Il est composé d'une vaste bibliothèque d'objets et d'outils annexes pour le développement d'applications spécifiques.

PaDaWAn	(Parallel Data Workflow for Analysis) : infrastructure logicielle accélérant les échanges de données entre applications d'un workflow.
Parallélisation/ parallélisme	Méthode d'agrégation d'entités élémentaires exécutant simultanément des tâches différentes, pour obtenir un système ayant des performances ou des capacités supérieures à la performance ou à la capacité d'une seule entité, et la plus proche possible de la somme des capacités unitaires.
Partitionnement	Découpage d'un graphe ou d'un maillage en sous-ensembles adaptés par exemple à un calcul parallèle - répartissant des parties du maillage sur différents processeurs ou unités de calcul
PCOCC	(Private Cloud On a Compute Cluster) : logiciel open source développé au CEA permettant aux utilisateurs des calculateurs d'y déployer des piles logicielles personnalisées grâce à des machines virtuelles ou des conteneurs.
Pétaflops ou pflops ou pflop/s	Million de milliard d'opérations sur des nombres réels par seconde (Pflops) (10^{15}).
PFMG25	(Plan France Médecine Génomique 2025) : plan piloté par AVIESAN et soutenu par l'État qui vise à positionner la France dans le peloton de tête des pays engagés dans la révolution de la médecine génomique. Cette stratégie conduira à une évolution de la façon de diagnostiquer, prévenir, et soigner les patients.
Plugin ou plug-in	(Appelé aussi module externe ou plugiciel ou add-in ou add-on) : extension, composant qui complète un logiciel hôte pour lui apporter de nouvelles fonctionnalités.
Potentiel d'interaction atomique	Modèle d'énergie potentielle servant à décrire l'interaction entre atomes. En dehors des gaz rares, les atomes à température ambiante s'associent spontanément pour former des molécules ou des cristaux. La force qui attire ces atomes entre eux est conservative, et on peut donc lui associer une énergie potentielle, qu'on appelle souvent « potentiel » par abus de langage.
Prace	(Partnership for Advanced Computing in Europe) : créée au printemps 2010 sous la forme d'une association internationale sans but lucratif (AISBL) de droit belge par les représentants de 26 pays européens, l'association PRACE a pour objectif de mettre en place une infrastructure de supercalculateurs en Europe et de la mettre à disposition des chercheurs académiques et industriels.
Processeur	Composant de calcul ou traitement présent dans de nombreux dispositifs électroniques, exécutant les instructions machine des programmes informatiques. L'invention du transistor, en 1948, permet la miniaturisation des composants électroniques, qui se poursuit encore aujourd'hui. Un processeur construit en un seul circuit intégré est un microprocesseur. Les processeurs multi cœurs (multicores) récents comportent maintenant plusieurs sous-unités ou « cœurs » dans un seul circuit intégré.
Profilage	Analyse de comportement d'un système (établissement de son « profil ») à partir de la collecte de données de fonctionnement, en général en vue de son optimisation (par exemple à partir de compteurs de temps de calcul pour évaluer la performance d'une simulation).
Protection/ reprise	Un code de simulation complexe a en général besoin de temps d'exécution excédant la durée d'une session système et donc s'organiser en succession de phases de calcul (par exemple tranches de un jour) ; le code doit donc sauvegarder son état intermédiaire, afin de s'arrêter et redémarrer à partir de cet état – c'est le principe des protections ; suivies de reprises ; ce principe sert aussi à sécuriser régulièrement l'état atteint par un calcul, de manière à ne pas avoir à le reprendre à partir de zéro en cas de panne ou plantage.
PUE	(Power Usage Effectivness) : ratio qui qualifie l'efficacité énergétique d'un centre de calcul.
Python	Langage de programmation interprété qui permet d'écrire rapidement des programmes scientifiques à l'aide de nombreux modules disponibles.

Q	QLM	(Quantum Learning Machine) : émulateur de calcul quantique développé par Atos, capable d'exécuter des programmes exprimés en primitives ou portes quantiques, indépendamment de toute implantation physique ; permet de préparer des applications en anticipation de machine réellement quantique, ou en frontal d'une telle machine ; une machine QLM est implantée au CCRT.
	RAID	(Redundant Array of Inexpensive Disks) : ensemble de techniques de virtualisation de stockage permettant de répartir sur plusieurs disques durs les données par assurer une meilleure sécurité des données et/ou un plus grand débit de lecture/écriture.
R	Rang	Le rang d'une entité parallèle est un identifiant unique permettant de reconnaître cette entité parmi l'ensemble des autres flots parallèle disponibles dans l'application. Dans la plupart des modèles de programmation (comme MPI et OpenMP), le rang est un entier commençant à 0.
	RENATER	Réseau National de télécommunications pour la Technologie, l'Enseignement et la Recherche.
S	Schéma numérique	Système de relations liant des valeurs de fonctions inconnues en certains points proches les uns des autres. Un schéma numérique est une formulation algébrique d'un problème discret conçu typiquement à l'aide de la méthode des différences finies, qui est une technique permettant de rechercher des solutions approchées d'équations aux dérivées partielles par discrétisation des opérateurs de dérivation.
	Scripting	Terme indiquant que l'on peut utiliser les fonctionnalités d'un logiciel ou d'une bibliothèque informatique via l'écriture d'un script textuel où ces fonctionnalités sont enchaînées et paramétrées selon les besoins de la personne écrivant le script ; en général un script est écrit dans un langage interprété et non compilé.
	Scotch	Outil de partitionnement de maillages et de graphes, open source, d'origine INRIA Bordeaux, LABRI.
	Sécurité périmétrique	Sécurité qui ne prend en compte que la frontière entre un système d'information et les éléments extérieurs à celui-ci.
	SER	Surface ou Signature Équivalente Radar permettant d'évaluer la furtivité d'un objet.
	SIMD	(Single Instruction on Multiple Data) : instruction s'appliquant simultanément à plusieurs données pour produire plusieurs résultats.
	SLURM	logiciel open source de gestion et de soumission des travaux informatiques sur un supercalculateur parallèle
	SNLE-NG	Sous-marin nucléaire lanceur d'engins de nouvelle génération (classe le Triomphant).
	Somme de contrôle	(checksum en anglais) : courte séquence de données numériques, signature d'un bloc de données plus important, permettant de vérifier, avec une très haute probabilité, que l'intégrité du bloc initial a été préservée lors d'une opération (stockage, déplacement, copie...).
	Sous-domaine	Définit une partie d'un maillage et les données associées aux éléments de maillage de cette partie.
	SSD	(Solid-State Disk) : « disque » utilisant de la mémoire non volatile au lieu de d'un support magnétique (plateaux).
	SSL	(Secure Sockets Layer) : protocole de sécurisation des échanges de données par un réseau informatique, par exemple Internet. Il est remplacé aujourd'hui par TLS.
Système d'équations linéaires	Un système d'équations linéaires est un ensemble d'équations linéaires portant sur les mêmes inconnues. Si n est le nombre d'inconnues et m le nombre d'équations, il peut s'écrire sous la forme matricielle $A X = B$ où A est une matrice de dimension n x m, X contient les n inconnues, et B le second membre connu de dimension m.	
SYSTEM@TIC/ Systematic Paris-Région	Pôle de compétitivité d'Île-de-France créé en 2005, consacré aux systèmes complexes et rassemblant plus de 750 acteurs innovants. Identifié comme « pôle à vocation internationale », sur le thème de la conception, de la réalisation et de la maîtrise des systèmes complexes.	

Tape ou bande	Dispositif de stockage de l'information sous forme d'une bande magnétique.	T
Téraflops ou téraflop/s	Mille milliards d'opérations sur des nombres réels par seconde (10^{12}).	
Teratec	Association créée en 2005, l'Association Teratec comporte en 2020 plus de 80 membres, industriels utilisateurs, fournisseurs de technologies et services, établissements d'enseignement et de recherche. Teratec a pour objectifs la maîtrise des technologies numériques et leur diffusion dans l'ensemble de l'industrie, en fédérant l'ensemble des acteurs concernés et de promouvoir et d'accroître l'attractivité du domaine en favorisant le développement économique.	
TGCC	Très Grand Centre de calcul du CEA opéré par les équipes du CEA/DAM à Bruyères-le-Châtel. Ouvert à la communauté scientifique européenne, le TGCC est un centre de calcul construit pour accueillir le supercalculateur français dans le cadre de PRACE.	
TLS	(Transport Layer Security) : protocole de sécurisation des échanges par réseau informatique, notamment par Internet. A remplacé SSL en éliminant les vulnérabilités de ce protocole.	
TNA	Tête nucléaire aéroportée. La TNA est emportée par le missile ASMP-A et équipe les forces aériennes stratégiques.	
TNO	Tête Nucléaire Océanique. La TNO est emportée par le missile M51 installé à bord des SNLE de nouvelle génération (SNLE-NG) et équipe la force océanique stratégique.	
Top500	Classement mondial des supercalculateurs reposant sur les performances du linpack.	
Trilinos	Collection de bibliothèques de logiciels open source, comportant des blocs de construction pour le développement d'applications scientifiques. Trilinos a été développé aux Sandia National Laboratories.	
UML	Unified modeling language	U
Unité de staging	Unité de calcul utilisée pour stocker temporairement des données en mémoire.	
Vectorisation	La vectorisation dans le cadre du calcul consiste à appliquer une opération simultanément sur plusieurs données semblables, au lieu d'une seule, augmentant la performance (forme de parallélisme de bas niveau). Il existe des processeurs, ou des extensions de processeurs, spécialement câblés pour appliquer ainsi une opération sur un « vecteur » - tableau - de plusieurs données à la fois.	V
Ver	En informatique, logiciel malveillant auto répliquatif conçu pour se propager en utilisant un réseau informatique. Le ver peut se propager sans avoir besoin de se lier à un programme exécutable, contrairement au virus.	
Verrou	Mécanisme de synchronisation très utilisé en programmation parallèle. Même s'il existe beaucoup de variantes, le principe de base repose sur un objet qui ne peut appartenir qu'à une seule entité à la fois. Ainsi, en protégeant des parties de code sur la possession d'un verrou, il est possible de garantir que cette partie sera exécutée par une seule entité à la fois.	
Virus	Automate auto répliquatif (malveillant ou non) conçu pour se propager en s'insérant dans des logiciels légitimes.	
VPU	(Vector Processing Unit ou processeur vectoriel) : processeur qui possède différentes fonctionnalités d'architecture lui permettant de profiter du parallélisme inhérent à l'usage de tableaux, de matrices et ainsi d'améliorer l'exécution de programmes les utilisant massivement. Il peut ainsi travailler sur un vecteur entier en une instruction par exemple.	
Workflow	Représentation sous forme de flux (enchaînement de traitements sur des données) d'une série de tâches à réaliser.	

Liens utiles

— **CEA/DAM**

<http://www-dam.cea.fr/>
<http://www-dam.cea.fr/dam/rejoignez-nous/>

— **ATOS**

<https://atos.net/en/solutions/high-performance-computing-hpc>

— **GENCI**

<http://www.genci.fr/fr>

— **Complexe de calcul du CEA**

<http://www-hpc.cea.fr/fr/complexe>

— **CCRT**

<http://www-ccrt.cea.fr/>
<http://www-hpc.cea.fr/fr/complexe/ccrt.htm>

— **TGCC**

<http://www-hpc.cea.fr/fr/complexe/tgcc.htm>

— **TERATEC**

<http://www.teratec.eu/technopole/Cesimat.html>

— **PRACE (Partnership for Advanced Computing in Europe)**

<http://www.prace-ri.eu/>

— **EUROHPC**

<http://eurohpc.eu/>
<https://ec.europa.eu/digital-single-market/en/eurohpc-joint-undertaking>

— **Horizon 2020**

<https://ec.europa.eu/programmes/horizon2020/en/h2020-sections-projects>

— **European technology platform for high performance computing (ETP 4 HPC)**

<http://www.etp4hpc.eu/>
et pages reliées :
<http://www.etp4hpc.eu/membership/1/319-bull-atos-technologies.html>
<http://www.etp4hpc.eu/missionvision.html>

— **European Code of Conduct for Data Centres**

<https://e3p.jrc.ec.europa.eu/communities/data-centres-code-conduct>

— **EE-HPC (Energy Efficient High Performance Computing Working Group)**

<https://eehpcwg.llnl.gov/>

— **EPI (European Processor Initiative)**

<https://www.european-processor-initiative.eu/>
<https://ec.europa.eu/digital-single-market/en/news/european-processor-initiative-consortium-develop-europes-microprocessors-future-supercomputers>

— **Pôle de compétitivité SYSTEM@TIC Paris-Région**

<https://SYSTEM@TIC-paris-region.org/fr/>

— **FAME2**

<https://SYSTEM@TIC-paris-region.org/fr/projet/fame2/>

— **Exascale**

<http://www.exascale-computing.eu/>

— **TOP500**

<https://www.top500.org/lists/top500/>

— **RENATER**

<http://www.renater.fr>

— **AVIESAN (Alliance pour les sciences de la vie et de la santé)**

<https://www.aviesan.fr/>

— **NEUROSPIN**

http://joliot.cea.fr/drf/joliot/Pages/Entites_de_recherche/NeuroSpin.aspx

— **France génomique**

<https://www.france-genomique.org/spip/>

— **NNSA**

<https://www.energy.gov/nnsa/national-nuclear-security-administration>

— **JUELICH**

https://www.fz-juelich.de/portal/EN/Home/home_node.html

— **RIKEN**

<https://www.riken.jp/en/>

— **ARCANE**

<http://www-hpc.cea.fr/fr/red/docs/Arcane.pdf>

— **BULL SEQUANA**

<https://atos.net/fr/produits/calcul-haute-performance-hpc/supercalculateurs-bull-sequana-x/gamme-bullsequana-x1000>

— **QUANTIQUE**

<https://atos.net/en/insights-and-innovation/quantum-computing/atos-quantum>

— **Chaire industrielle Data analytics and machine learning CEA-ATOS-ENS Paris Saclay**

https://atos.net/fr/2016/communiqués-de-presse_2016_10_20/lens-paris-saclay-atos-et-le-cea-creent-la-chaire-industrial-data-analytics-machine-learning-pour-lemergence-dune-filiere-francaise-data-science
<https://ens-paris-saclay.fr/actualité/mathilde-mougeot-titulaire-de-la-chaire-industrial-sata-anlytics-machine-learning/>

— **ABINIT**

<https://www.abinit.org/>

— **Python**

<https://www.python.org/>

— **VTK - the Visualization Toolkit**

<https://www.vtk.org/>

— **Paraview, an open-source, multi-platform data analysis and visualization application**

<https://www.paraview.org/>

— **MPC**

<http://mpc.hpcframework.com>

— **MPI**

<https://www.mpi-forum.org/docs/>
<http://www.idris.fr/formations/mpl/>

— **SHINKEN**

<http://www.shinken-monitoring.org>

— **GRAFANA**

<https://grafana.com>

— **Plateforme d'aide au développement**

<https://github.com/cea-hpc/pcoc>
<https://github.com/cea-hpc/selFle>

— **Université Versailles Saint-Quentin en Yvelines**

<http://www.uvsq.fr/>

— **Master Calcul Hautes Performances et Simulation (CHPS)**

<http://www.chps.uvsq.fr/>

— **École Polytechnique**

<https://www.polytechnique.edu/>

— **Cours de programmation parallèle de l'École Polytechnique**

<https://moodle.polytechnique.fr/enrol/index.php?id=7303>

— **ENSIIE**

<https://www.ensiie.fr/>

https://www.ensiie.fr/formation_ingenieur-ensiie/

— **MASTER MATHÉMATIQUES, VISION, APPRENTISSAGE**

<https://www.master-mva.com/>

— **LUSTRE**

www.lustre.org

— **ROBINHOOD**

<https://github.com/cea-hpc/robinhood>

— **SHINE**

<https://github.com/cea-hpc/shine>

— **DATA CENTER RESOURCES**

<https://datacenterresources.com>

— **FENIX**

<https://fenix-ri.eu>

Tera

au fil du temps



Commissariat à l'énergie atomique et aux énergies alternatives Tera - La saga des supercalculateurs

Image de couverture

Laëtitia et Nicolas, administrateurs système,
examinent l'état physique de nœuds de calcul

Directrice de la publication

Laurence Bonnet,
Directrice scientifique du CEA/DAM

Pilote scientifique

Pierre Leca

Ont contribué à ce numéro

Daniel Bouche, Hervé Jourden,
Jacques-Charles Lafoucrière, Daniel Vanderhaegen, Didier Juvin,
Pierre-Franck Piserchia, Hervé Lozach, François Robin, Christine Ménaché,
Daniel Aguilera, Gilles Wiber, Frédéric Combeau, Pierre Barnabé,
Jean-Marc Ducos, Patrick Carribault, Julien Jaeger, Marc Perache,
Gilles Gropellier, Cédric Chevalier, Benoit Lelandais, Marie-Pierre Oudot,
Nicolas Legoff, Franck Ledoux, Olivier Bressand, Jérôme Dubois,
Agnès Grenouille, Gwenolé Harel, Jacques-Bernard Lekien, Julien Capul,
Sébastien Morais, Catherine Cherfils, Laurent Colombet, Cédric Augonnet,
David Goudin, Matthieu Kuhn, Philippe Deniel, Dominique Martinet,
Élodie Ardoin, Jean-Christophe Weill, Guillaume Colin De Verdière,
Max-André Darizcuren, Matthieu Hautreux, Frédéric Feyel, Pierre Leca,
Vincenzo Salvetti, Jean-Philippe Nominé

**Remerciements à Mme Cécile Michaut pour son regard professionnel
qui nous a permis d'améliorer l'ouvrage et d'y introduire de l'humain au travers des témoignages
de salariés qui sont au cœur de la saga des supercalculateurs du CEA/DAM.**

Création

EFIL / www.efil.fr

Correction

Stylience / www.stylience.fr

Impression

Gibert et Clarey (37), brochure imprimée sur papier ecogéré

ISSN 2115-7057

Diffusion

Maryline Brusy CEA – DAM, Direction scientifique
Bruyères-le-Châtel, F-91297 Arpajon Cedex, France

Courriel

maryline.brusy@cea.fr

L'ouvrage est consultable à l'adresse www-dam.cea.fr
Dépôt légal à parution

Tera

La saga des
supercalculateurs



CEA - Direction des applications militaires

www.cea-dam.fr

—
Septembre 2021