

DE LA RECHERCHE À L'INDUSTRIE



Les métiers du HPC au CEA - DIF

Thao LÊ
CEA - DIF

Journée Portes Ouvertes du Campus Teratec
2 octobre 2018

- 9 centres (Cadarache, Marcoule, Grenoble, Paris-Saclay, Cesta, DIF, Grammat, Ripault, Valduc)

- 15 942 techniciens, ingénieurs, chercheurs et collaborateurs en CDI
- 51 unités de recherche sous cotutelle du CEA et de partenaires académiques
- 64 accords cadres en vigueur avec les universités et écoles
- 762 dépôts de brevets prioritaires en 2017

- 27 Equipex (équipements d'excellence)
- 33 Labex (laboratoires d'excellence)
- 3 Idex (initiative d'excellence)

- 204 start-up technologiques depuis 1972 dans le secteur des technologies innovantes
- plus de 600 partenaires industriels
- 5 milliards d'euros de budget
- 424 projets européens en cours en 2017
- 30 pôles de compétitivité auxquels participe le CEA dont 17 où le CEA est administrateur

- 684 brevets déposés en 2017 - 4^e déposant en France
- Organisme de recherche public en 2017 le plus innovant en Europe

- Missions du CEA :
 - les énergies bas carbone (nucléaire et renouvelables)
 - la recherche technologique pour l'industrie
 - la recherche fondamentale en sciences de la matière et sciences de la vie
 - la défense et la sécurité

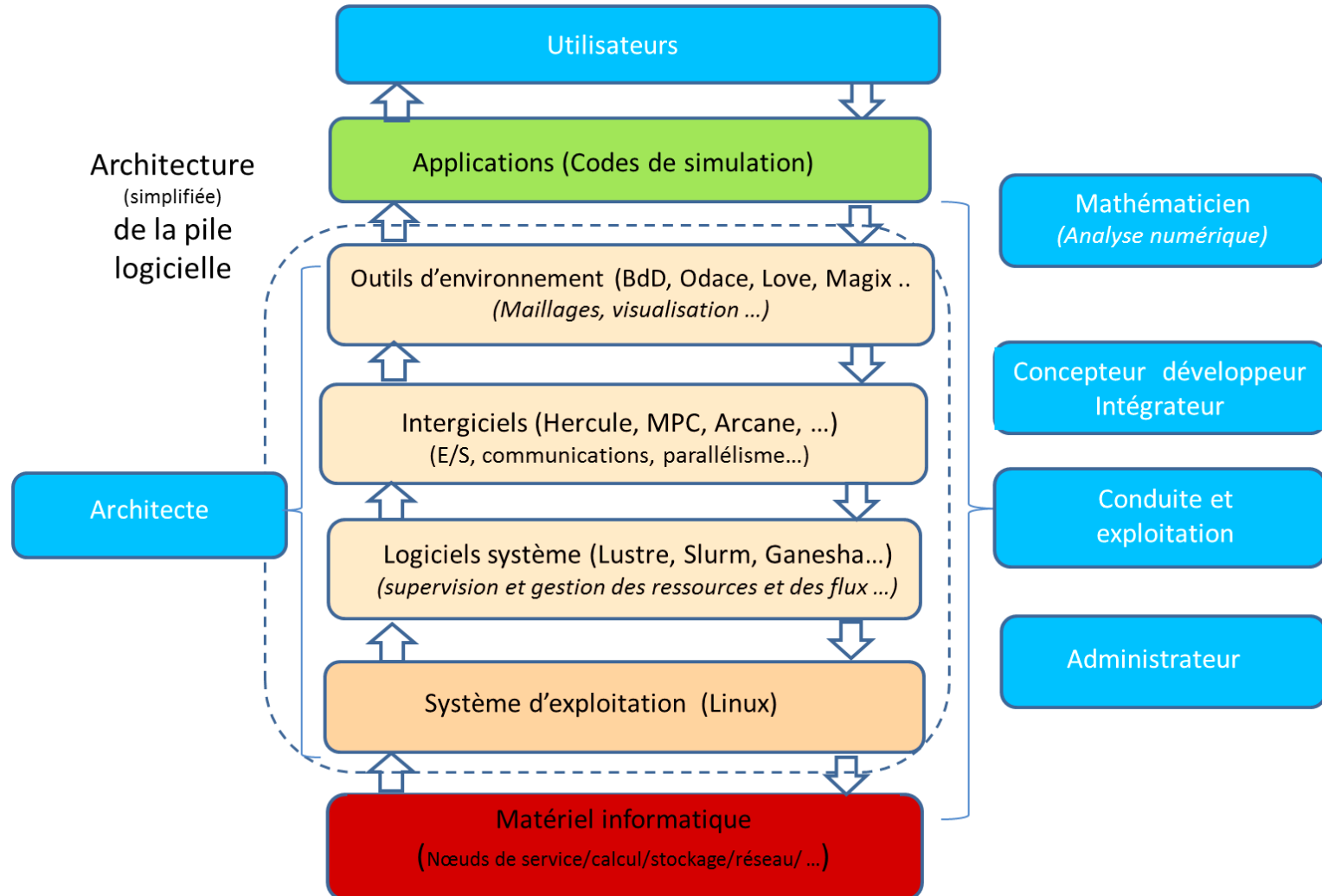
- 4 directions :
 - Direction de l'énergie nucléaire
 - Direction de la recherche technologique
 - Direction de la recherche fondamentale
 - Direction des applications militaires (Cesta, DIF, Grammat, Ripault, Valduc)
 - Répondre aux enjeux de la dissuasion nucléaire
 - Surveiller, analyser et intervenir pour la défense et la sécurité
 - Contribuer à l'excellence de la recherche et à la compétitivité de l'industrie

- Ce que l'on appelle simulation correspond à l'ensemble constitué par un modèle, les paramètres et contraintes, et les résultats obtenus
- Du phénomène physique ...
 - -> Modèles physiques théoriques
 - -> Formulations mathématiques des modèles physiques étudiés
 - -> Algorithmes numériques (discrétisation, maillage, solveurs...)



- ... à la simulation dans un grand centre de calcul

Pile logicielle simplifiée



■ Centre DIF

- Tera1000 Yoccoz d'Atos/Bull (2017) : 25 Pétaflops
- 14^e au Top500 en juin 2018

■ TGCC - Très Grand Centre de calcul du CEA (2010) avec le GENCI - Grand Équipement National de Calcul Intensif

■ 1) besoins de la recherche académique

- machine Curie d'Atos/Bull (2012) Bullx cluster : 1,8 Pétaflops
- machine Joliot-Curie de Bull (2018) Sequana X1000 : 9 Pétaflops
- 50% avec PRACE - Partnership for Advanced Computing in Europe
- 50% aux laboratoires de recherche français

■ 2) besoins de l'industrie avec le CCRT - Centre de Calcul Recherche et Technologie (2003)

- machine COBALT d'Atos/Bull (2016) 1,5 Pétaflops
- Airbus Safran Launchers, Areva, EDF, IFPEN, Ineris, IRSN, L'OREAL, SAFRAN Aero Boosters, SAFRAN Aircraft Engines, SAFRAN Helicopter Engines, Synchrotron Soleil, Thales, Thales Alenia Space, Valeo

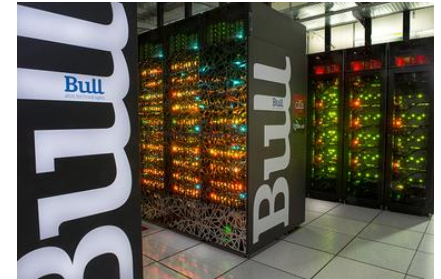
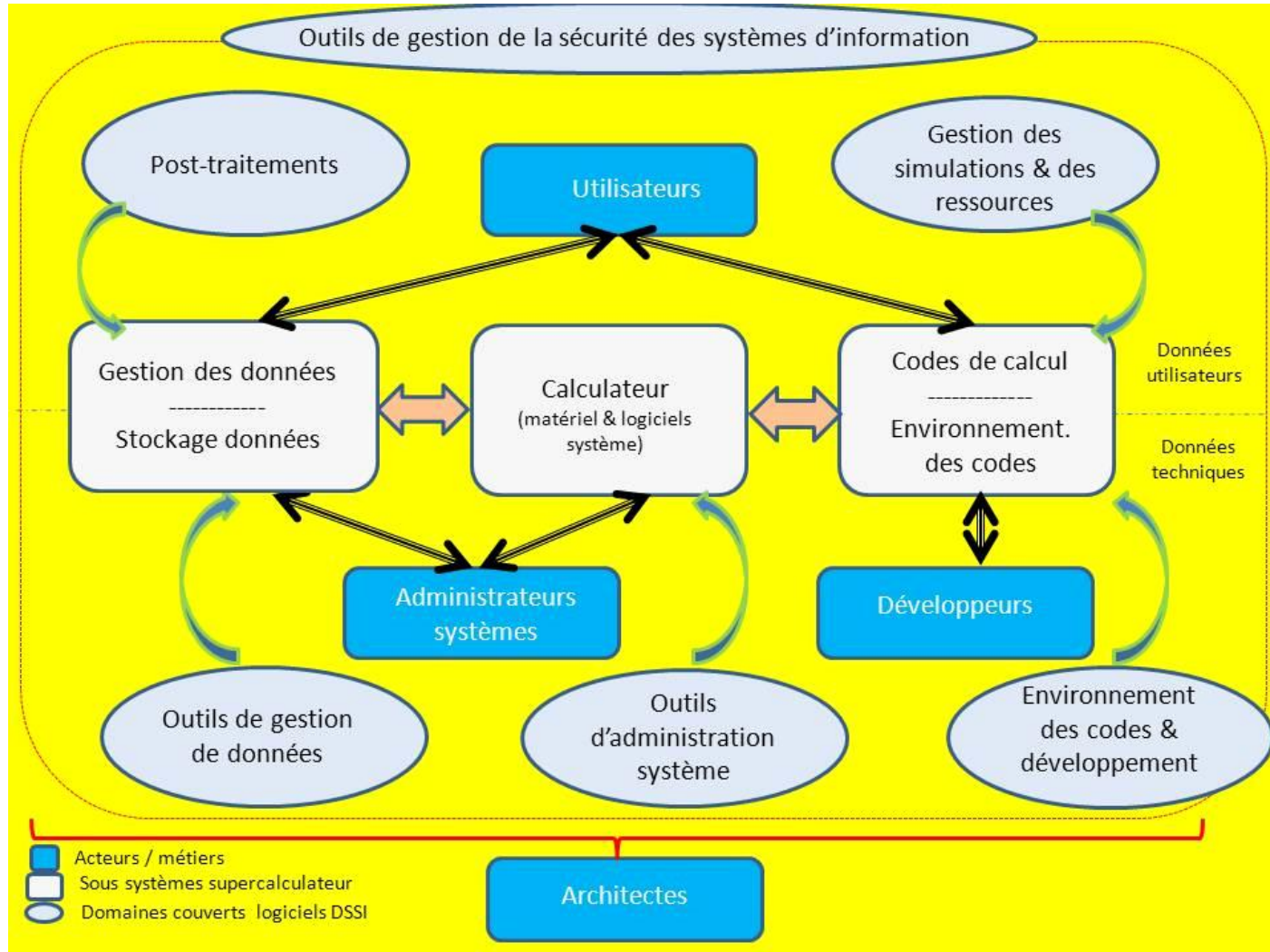
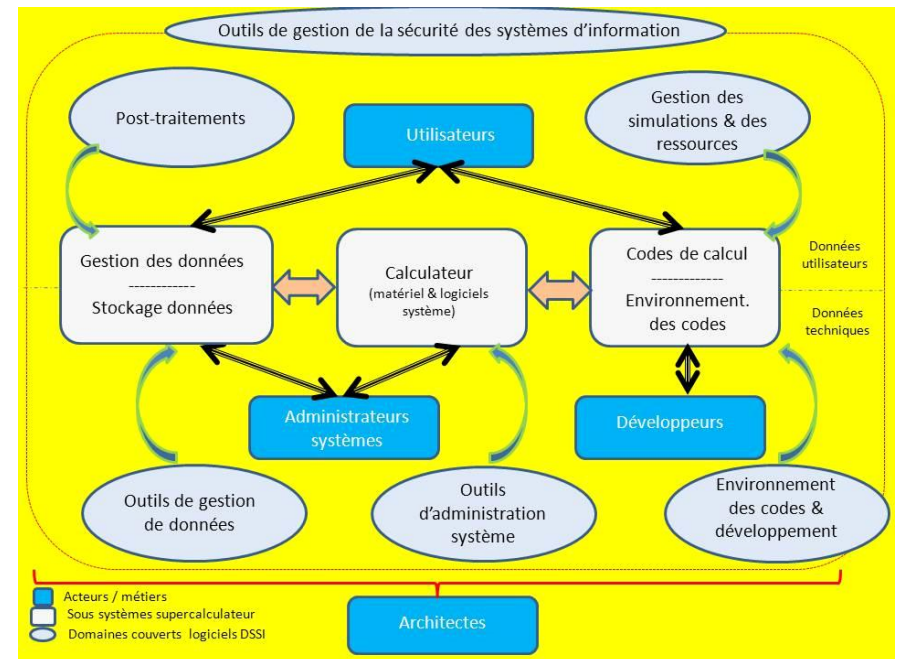


Schéma fonctionnel d'un centre de calcul



- 1. Gestion des simulations et des ressources
 - 2. Environnement des codes et développement
 - 3. Administration Système
 - 4. Gestion des données et du stockage
 - 5. Post traitements : tests, analyse, dépouillement, visualisation...
-
- Gestion de la sécurité des systèmes

- Utilisateur (physicien)
- Mathématicien (analyse numérique)
- Concepteur, développeur, intégrateur...
- Ingénieur d'exploitation
- Administrateur
- Responsable opérationnel
- Architecte



1. Gestion des simulations et des ressources

- Concepteur, développeur, intégrateur
 - Développement et maintenance des logiciels de maillage
 - Logiciel de système de protection / reprise et gestion du stockage des résultats

- Administrateur, responsable opérationnel
 - Logiciel de soumission des travaux pour les utilisateurs
 - Logiciel de supervision des travaux, équilibrage de charges, statistiques
 - Gestion des contraintes d'exécution pour les lancements d'exécution en parallèle

2. Environnement des codes

- **Mathématicien en analyse numérique**
 - Traduit les modèles en équations, définit les algorithmes pour les résoudre, met en point les codes numériques pour que ces algorithmes s'exécutent sur des calculateurs massivement parallèles
 - Valide les méthodes numériques : hydrodynamique, mécanique des fluides, photonique...

- **Concepteur, développeur, intégrateur**
 - Logiciel d'aide au développement des codes de physique et à l'optimisation de leur performance sur les grands calculateurs parallèles
 - Logiciel de délégations des E/S (abstraction du système de fichier) avec traitements possibles sur les données pendant l'exécution du code et partage avec d'autres codes
 - Outils de parallélisation ou d'aide à la parallélisation

3. Administration Système

- Administrateur, responsable d'exploitation de clusters
 - Mise à jour des systèmes d'exploitation et les logiciels
 - Développement et utilisation des outils d'installation, de configuration et d'administration
 - Analyse de la charge et la répartition des calculs, des E/S et du stockage

- Support applicatif
 - Gestion des utilisateurs
 - Gestion de l'utilisation des ressources et des clusters
 - Formation des utilisateurs : OpenMP, MPI vectorisé, profilage, outils de débogage...

- Sécurité des systèmes d'information

4. Gestion des données et du stockage

- Concepteur, développeur, intégrateur
 - Logiciel de stockage de données massif pour les architectures pétaflopiques futurs
 - Configuration des systèmes de fichiers parallèles sur un calculateur à architecture distribuée
 - Exploitation de grands volumes de données (Big Data)
 - Traitements intelligents des données (IA)

5. Post traitements

- Post traitements
 - Tests, analyse, dépouillement, visualisation...

- Outils de dépouillement
 - Logiciel de visualisation de données issues des codes de simulation
 - Logiciel d'exploration et d'analyse des données dans un contexte d'exploitation massive de grands volumes de données avec des services d'extraction, de combinaisons et de traitements mathématiques

- Outils de gestion et de stockage des données des codes

- L'Open Source permet
 - de garder la maîtrise des outils utilisés dans un grand centre de calcul
 - de permettre et faciliter le passage à l'échelle avec l'évolution du HPC
 - de provoquer un effet de levier grâce au retour des contributions externes et aux validations des autres contributeurs

- 34 +4 contributions aux logiciels Open Source dont la plupart à l'initiative du CEA

- ClusterShell, HP2P, MilkCheck, Pcooc, Shine, SSHPROXY, Modules TCL, KVSNS, Kernel Linux Module 9P, LCAP, Lustre, Mosshika, NFS-Ganesha, Phobos, RobinHood policy Engine, Space 9, Simund, Aujs, Bridge, Glost, Selfie, SnecPlayer, Arcane, Jchronoss, Frame, GMDS, MPC, WI4MPI, Hercule, Nabla, Malp, Love et Odace

- MIASM, IVRE, SYBIL et HAIRGAP

Exemples de métier

- Ingénieur Système, Développement et Administration
 - Contribution à l'ingénierie des système de calcul ou de stockage, au développement de logiciel libres, à l'administration du système, à la R&D et la veille technologique ainsi qu'au maintien en Condition Opérationnelle
 - Grande partie de l'activité consacrée au développement de fonctionnement de la pile logicielle dédiée au contrôle des supercalculateurs, à leur validations et déploiements

- Ingénieur Systèmes, Réseaux et Télécommunication
 - Chargé de l'administration et de l'exploitation des réseaux locaux ou distants, des réseaux très haut débit associés aux supercalculateurs
 - Une partie de l'activité consacrée à des études techniques pour les nouvelles architectures ou le renouvellement des équipements
 - Contribution à des projets de logiciels libres en relation les constructeurs leaders du domaine

- Ingénieur chercheur en calcul scientifique
 - Contribution à un grand projet de recherche et développement dans le domaine du traitement des données massives issues des codes parallèles
 - Contribution à des études exploratoires et innovantes du HPC pouvant donner lieu à l'encadrement de stagiaires et de doctorants

Activités sur le plateau CEA du campus

- Possibilité d'accueil de 35 jeunes chercheurs
- Avec 1/3 de doctorants qui participent à la journée des thésards du centre DIF
- 2 journées des stagiaires du DSSI

- Séminaires du DSSI



- Formations diverses :
 - Formation « ABINIT »
 - Formation CCRT « Initiation aux clusters du CCRT »
 - Formation CCRT « Outils de débogage et de profilage »
 - Formation CCRT « Calcul haute performance appliqué aux données génomiques »
 - Formation NVIDIA « Deep Learning »

- **InHP@CT** : Informatique Haute Performance @ Campus Teratec
 - Contexte de la conception et de l'exploitation d'ordinateurs pétaflopiques pour le HPC, présentation de méthodes pour les supports exécutifs, la programmation parallèle, hétérogène et hybride

- **SISMA** : Séminaire Informatique Scientifique et Mathématiques Appliquées
 - Contexte R&D de la simulation de phénomènes multi-physiques, avec des présentations de méthodes en mathématiques appliquées, en « algorithmique » et en traitement de très grands volumes de données.

- **LAMÉ** : LAngages et MÉthodes
 - Contexte de la conception de logiciels pour le HPC, présentation de méthodes en génie logiciel, modèles de programmation, langages et outils associés

■ Thèses en cours

- Caractéristiques et utilisation des mémoires hétérogènes multi-niveaux
- Gestion des threads spécialisés en contexte multi-modèles de programmation
- Gestion du compromis entre la performance et la précision du code de calcul
- Méthode ultra-faible pour la résolution numérique de l'équation du transfert des photons
- Méthodes d'apprentissage statistiques en analyse des signaux issus de calculateurs HPC
- Projet Européen BIG STORAGE - implémentation de storage blobs sur le cloud
- Stockage massif pour le HPC par utilisation de storage blobs
- Tension de surface en référentiel ALE pour des schémas colocalisés

■ Thèses futures

- Génération de polycubes dirigée à l'aide de champs d'orientation
- Génération et adaptation de maillages mixtes à l'aide de métriques orientées
- Partitionnement de maillages pour l'équilibrage de charge de simulations multi-physiques
- Schémas semi-Lagrangiens implicite pour l'interaction fluide-structure compressibles
- Support outillé pour le débogage de modèles métiers

Merci de votre attention !

Direction des Applications Militaires
Département Science de la Simulation et de l'Information

Commissariat à l'énergie atomique et aux énergies alternatives
Centre DAM Île-de-France – Bruyères-le-Châtel - 91297 Arpajon Cedex
T. +33 (0)1 69 26 40 00
Etablissement public à caractère industriel et commercial | RCS Paris B 775 685 019