

M2

Option Master 2 :
Calculs et traitements de données
haute performance

High Performance Computing

Savoir programmer efficacement

- Plusieurs cœurs/processeurs de calcul
- Des accélérateurs de calcul (par ex. des cartes graphiques)
- Des supercalculateurs constitués d'ordinateurs en réseaux
- Chacun des cœurs individuels, aussi 😊

Pourquoi ?

- Parce que tous les ordinateurs sont parallèles !
- La grande majorité des programmes n'exploitent même pas 1% des performances du matériel



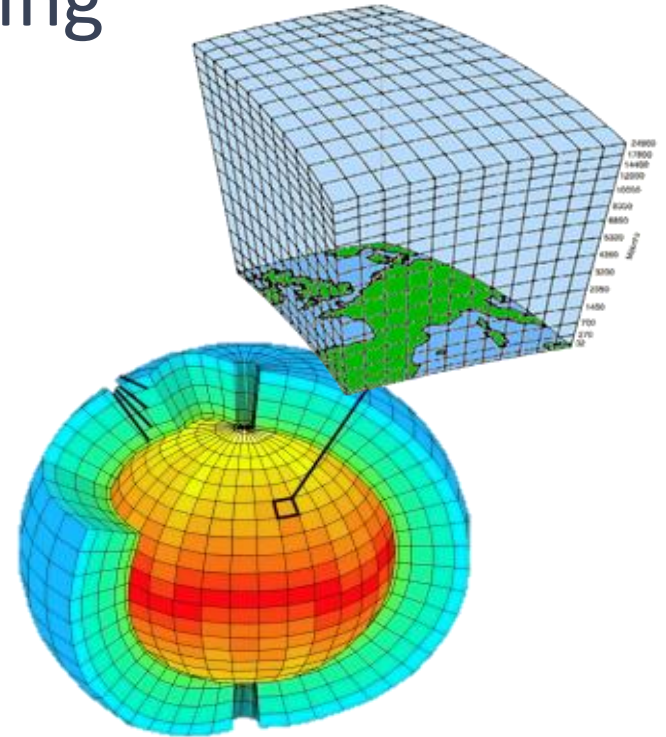
High Performance Computing

La simulation numérique est devenue un pilier incontournable de la science

- Simulation de phénomènes impossibles à observer en laboratoire
- Physique, Chimie, Mécanique, Sismologie, Climatologie, Astrophysique, Biologie, etc. (même Finance)

On utilise de puissants ordinateurs

- Pour calculer plus vite
- Pour davantage de précision
- Pour que le problème « tienne » en mémoire !



High Performance Data Analytics

Les scientifiques et les industriels sont submergés par des masses de données dont il faut savoir extraire, analyser et visualiser des phénomènes clés

- Santé, épidémiologie
- Energie, écologie
- Sécurité
- Commerce

Ces données proviennent

- De capteurs/sondes, objets connectés
- De grands équipements (télescopes, accélérateurs de particules, etc.)
- De transactions informatiques
- De la simulation !

Convergence HPC - HPDA

Analyser efficacement de gros volumes de données et raisonner à partir des données extraites

- Extraction, filtrage
- Algorithmes de classification et d'apprentissage
- (Réinjection dans la simulation)
- Visualisation

Axe de recherche et développement prioritaire au niveau international

Objectifs de la formation

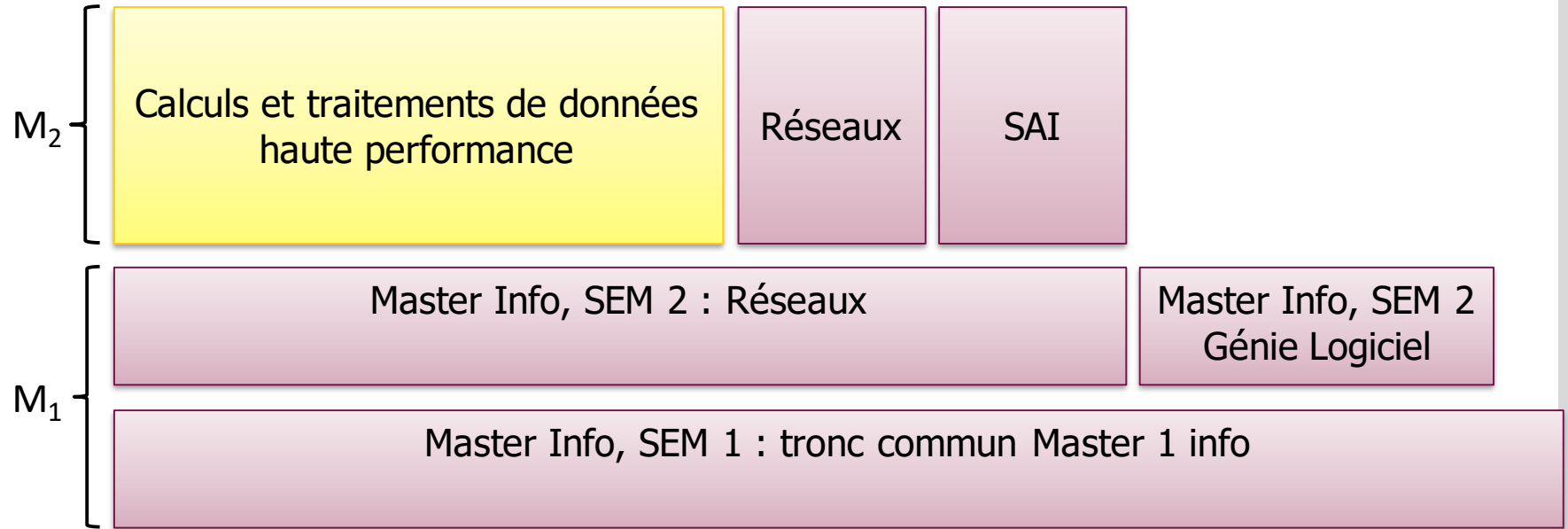
Former aux méthodologies pour la convergence entre les deux disciplines

- Calcul Haute Performance (HPC)
- Analyse et traitement des données massives (HPDA)

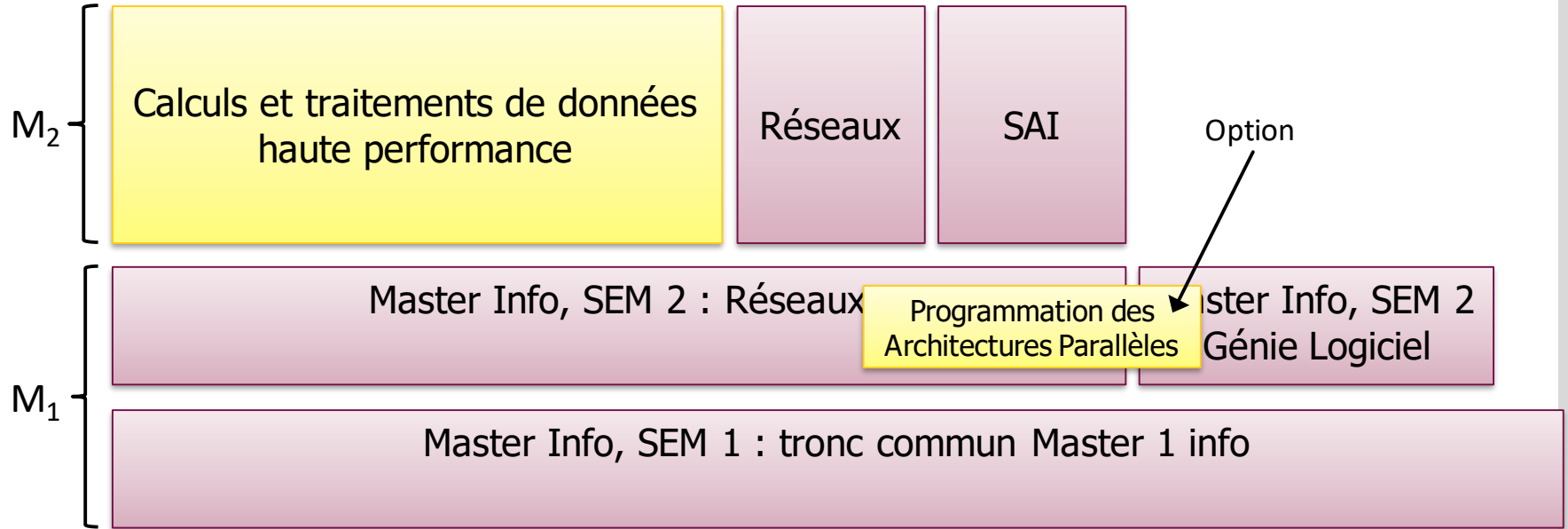
Afin d'acquérir des compétences à chaque étage des piles matérielle, logicielle, algorithmique et applicatives

- Algorithmique parallèle, Parallélisation sur les plateformes convergées hétérogènes modernes, Optimisation et performance des codes, Visualisation
- Méthodes et algorithmes pour le traitement et l'analyse de données massives, Apprentissage
- Illustration sur des applications académiques, sociétales et industrielles

En pratique



En pratique



En pratique

Mutualisation

- Enseirb
- Master informatique de l'Université de Bordeaux
- Matmeca

Stages, débouchés

- Dassault, IBM, Thales, Total
- CEA, EDF, IFP, Airbus
- Banques
- E-commerce (Cdiscount, ...)
- Facebook, Google
- CapGemini, ATOS

Organisation en UV

Un tronc commun de 3 UV obligatoires avec comme fil rouge la complémentarité entre les deux domaines étudiés

- Architectures matérielles, langages, modèles de programmation, optimisation pour les architectures parallèles et distribuées
- Algorithmique parallèle, méthodes numériques / non numériques / statistiques pour l'analyse et l'extraction d'informations dans des grandes masses de données
- Visualisation et IO pour le calcul et le traitement des données massives

Trois ½ UV (modules) à choisir parmi :

- 2 modules applicatifs axés sur les outils numériques et leurs applications à des cas d'étude de simulation numérique
- 2 modules applicatifs axés sur les outils numériques et non numériques pour le data analytics (apprentissage et deep learning, cas d'études en IA)

Organisation en UV

Un tronc commun de 3 UV obligatoires

- Architectures matérielles, langages, modèles de programmation, optimisation pour les architectures parallèles et distribuées
 - > Multicoeurs, hiérarchies mémoire, accélérateurs
 - > OpenMP, MPI, OpenACC, programmation par tâches, langage dédié au traitement statistique des données (R)
 - > MapReduce
- Algorithmique parallèle, méthodes numériques / non numériques / statistiques pour l'analyse / l'extraction d'informations dans des grandes masses de données
 - > Algo du HPC: parallélisation, ordonnancement, structures de données pour le parallélisme (octrees, graphes, tenseurs)
 - > Algo du HPDA: apprentissage, réseaux de neurones, classification et filtrage (nettoyage, stats, visu), réseaux bayésiens, estimateurs
- Visualisation et I/O pour le calcul et le traitement des données massives
 - > Paraview, VisIt, Approches in-situ
 - > YARN, Spark, Hadoop
 - > Tensorflow, PyTorch, BigDL
 - > I/O, Lustre

Organisation en UV

Trois ½ UV (modules) à choisir parmi :

- 2 modules applicatifs axés sur les outils numériques et leurs applications à des cas d'étude de simulation numérique
 - > 1 module purement numérique
 - > 1 module In-Situ HPC - BigData
- 2 modules applicatifs axés sur les outils numériques et non numériques pour le data analytics (apprentissage et deep learning, cas d'études en IA)
 - > 1 étude complète d'une application telle que Plantnet / Vision par ordinateur
 - > 1 étude orientée analyse de données (OSInt: open-source intelligence, Facebook, Google)

Offre de formation en France

- HPC – BigData (Sem 9, ENSEEIHT)
- Calcul intensif et données massives (ENSIIE, Evry)
- Calcul intensif et sciences des données (Université Côte d'Azur, en partenariat avec Mines Paris)

- Calcul Haute Performance Simulation (Université Versailles Saint Quentin)

- Valorisation des données massives (ENSEEIHT – INSA Toulouse)
- BigData (Polytech Annecy-Chambéry)

M2

Option Master 2 : Calculs et traitements de données haute performance

raymond.namyst@u-bordeaux.fr