

Pôle Sciences, Technologies et Numérique (STN)

Fiche Formation Mastère

Intitulé Formation : MASTÈRE SPECIALISÉ CALCUL HAUTES PERFORMANCES / DONNÉES MASSIVES /

Fait l'objet d'un test ?

Oui

Non

Sélection sur dossier et entretien

PÔLE SCIENCES, TECHNOLOGIES ET NUMÉRIQUE (STN)

Domaine : Sciences et Technologies

Mention : INFORMATIQUE

Spécialité(s) : Calculs Hautes Performances – Données Massives – Intelligence Artificielle

Présentation

La puissance des supercalculateurs est une clé de l'innovation et de la compétitivité. Avec des technologies comme le big data analytics, l'intelligence artificielle ou le machine learning dont l'exploitation nécessite toujours plus de puissance de calcul, le calcul hautes performances (HPC), support des développements de l'Intelligence Artificielle (IA), devient un enjeu économique et politique majeur.

La simulation à haute performance est devenue un outil essentiel de la recherche scientifique, technologique et industrielle. Ces technologies constituent des outils en matière de développement, tant pour le Sénégal, qui a fait du numérique un facteur de développement à travers le projet Sénégal Numérique, que pour l'Afrique.

Ce mastère a pour objectif de permettre aux étudiants diplômés bac+4

et bac+5 ainsi qu'aux professionnels ayant déjà de solides connaissances en mathématiques et en algorithmique d'acquérir les connaissances, les méthodes et les techniques particulières requises permettant de maîtriser deux évolutions technologiques majeures : l'utilisation des ordinateurs à haute performance et la simulation numérique.

La formation couvre les domaines scientifiques suivants :

- Les architectures hautes performances
- Les modèles de programmation parallèle permettant d'exploiter au mieux les architectures hautes performances et développer des logiciels efficaces pour la simulation
- Les techniques logicielles permettant de traiter un très grand volume de calculs ou d'informations de manière simultanée via des algorithmes spécialisés
- La gestion des flux de données

2. Cibles

- Titulaires d'un diplôme d'ingénieur (habilité par la CTI)
- Titulaires d'un diplôme de master (M2)
- Titulaires de diplômes étrangers équivalents justifiant d'études d'un Master 1 et de 3 années d'exp. prof. ou justifiant d'études d'un Master 1 par dérogation.

3. Matières dominantes

- Optimisation en grandes dimensions
- Machine learning
- Programmation parallèle
- Python
- Management de projet et application transversale (Scrum)
- Hadoop et Spark
- Open Data
- Simulation numérique
- Systèmes à base de connaissances
- Apprentissage par renforcement

4. Organisation

La formation est structurée en 2 semestres :

- Le premier semestre est composé d'un ensemble d'unités d'enseignements (UE). À chaque UE est affectée une valeur en crédits. L'auditeur choisit les unités d'enseignement de son choix parmi la liste décrite (voir fiche technique plus bas) ;
- Le deuxième semestre comporte un stage obligatoire qui permet une mise en application des connaissances acquises. Cette expérience professionnelle est obligatoire et contrôlée avant la délivrance du diplôme.
- Le master est délivré à tout auditeur remplissant les conditions suivantes :
- Avoir validé les unités d'enseignements choisies et posséder les 20 crédits associés ;
- Avoir validé les formations Open Data Camp et posséder les 10 crédits associés ;
- Avoir validé 30 crédits au titre de l'expérience professionnelle acquise au cours d'un stage de 6 mois.

5. Objectifs

- Acquérir un savoir-faire pluridisciplinaire
- alliant maîtrise des techniques de programmation de l'informatique haute performance, maîtrise des techniques de modélisation et de simulation, avec une forte expertise en parallélisme (matériel, logiciel, numérique) et en calcul distribué
- Approfondir ses connaissances en mathématiques appliquées
- en couvrant les aspects avancés de la statistique jusqu'aux techniques d'apprentissage adaptées au traitement des Big Data
- Acquérir une expérience de participation productive à la conception et à la mise en œuvre d'applications répondant aux besoins réels des entreprises, par les travaux pratiques et le projet d'intégration
- Se familiariser avec le contexte de l'intelligence artificielle
- dans ses dimensions technologique et administrative
- Approfondir ses connaissances
- sur les modes de gestion des projets
- développer sa capacité de travail en équipe
- de même que son expression orale et écrite, de façon à assurer une communication efficace
- Travailler efficacement au sein d'une équipe multidisciplinaire
- développer le sens de l'analyse de problèmes et la recherche de solutions appropriées, se sensibiliser au contexte de l'industrie, ses enjeux socio-économiques et éthiques.

Pôle Sciences, Technologies et Numérique (STN)

6. Finalités – Débouchés

- Poursuite des études en thèse (recherche)
- Débouchés professionnels :
- Data Scientist
- Data Analyst
- Architecte Big Data
- Administrateur de bases de données
- Responsable projet informatique

7. Coûts de la formation

Le coût total de la formation s'élève à 1 300.000 F CFA, payables selon les modalités suivantes :

- 1ere tranche (payable à l'inscription) : 500.000 F CFA
- 2e tranche : 400.000 F CFA (payable 3 mois après l'inscription)
- 3e tranche : 400.000 F CFA (payable 6 mois après l'inscription)

8. Stages

- Stage de fin de formation
 - Durée : 5 à 6 mois (30 ECTS)
 - Objectif : permettre en pratique les connaissances acquises en étant en situation de spécialiste en Big Data. Réalisation technique, Rapport + soutenance orale.

9. Open Data Camp

- Projet en Data Science organisé par la société GAINDE2000
 - Durée 30H (1,5 UE = 3 ECTS)
- Projet en Data Science organisé par la société BAAMTU
 - Durée 30H (1,5 UE = 3 ECTS)
- Projet en Data Science organisé par la société ATOS
 - Durée 40H (2 UE = 4 ECTS)

Durant une semaine, porteurs de projets et (ré)utilisateurs de données se retrouveront pour poursuivre ou proposer des initiatives élaborées à partir de données ouvertes, afin d'analyser un sujet, de développer un service ou de résoudre un problème. L'objectif est de faire communiquer les étudiants, professeurs et professionnels de la donnée autour d'un problème unique permettant de faire converger à la fois différentes compétences et surtout différentes visions de ce que l'on peut faire avec les données.

10. Informations complémentaires

Les intervenants du parcours viennent de milieux académiques et d'entreprises pour offrir une formation professionnelle complète.

Les UE sont des modules de formation capitalisables qui peuvent être suivis individuellement et qui pourront être combinés dans des parcours libres, certifiants ou diplômants.

Unités d'Enseignement (UE)	Crédits	Eléments constitutifs (EC)
SEMESTRE 1 – Mastère CHP/DM/IA		
Optimisation en grandes dimensions	4	<p>Très souvent, les problèmes d'optimisation sont des problèmes d'optimisation discrète où les variables modélisent des entités indivisibles : un nombre de voitures ou d'avions à construire par exemple, ou bien des décisions à prendre, choix d'un coup ou d'une stratégie dans un jeu. Un problème d'optimisation discrète revient à trouver dans un ensemble discret (les solutions réalisables), la meilleure solution relativement à une fonction objective donnée. Les problèmes d'optimisation discrète sont, la plupart du temps, difficiles et il est important de savoir les résoudre efficacement.</p> <p>Compétences attendues : Avoir des notions de complexité des algorithmes, connaître les outils efficaces pour résoudre les problèmes d'optimisation discrète de façon exacte ou approchée, savoir modéliser un problème et utiliser un solver (commercial ou libre)</p>
Machine learning	4	<p>Le machine Learning, apprentissage machine en français, regroupe un ensemble de méthodes mathématiques et algorithmiques permettant d'apprendre une fonction à partir d'exemples. Ces méthodes sont devenues aujourd'hui incontournables pour l'exploitation des bases de données numériques, de plus en plus nombreuses. La session proposera à la fois des cours théoriques sur les enjeux mathématiques de ces méthodes et des travaux pratiques qui permettront d'utiliser les méthodes présentées sur des problématiques de régression ou de classification en utilisant le logiciel R.</p> <p>Compétences attendues :</p> <ol style="list-style-type: none"> 1. Introduction au machine learning, à l'apprentissage supervisé et aux modèles prédictifs. 2. Modèles de régression pénalisée. Critères de performances. Sélection de modèles. 3. Modèles de classification. Courbes ROC. 4. Modèles d'ensemble. Arbres de classification et de régression. Bagging. Random Forest. 5. Agrégation de modèles. Boosting



Pôle Sciences, Technologies et Numérique (STN)

<p>Apprentissage renforcement</p> <p style="text-align: right;">par</p>	<p style="text-align: center;">4</p>	<p>L'apprentissage par renforcement est utilisé dans plusieurs applications: robotique, gestion de ressources, vol d'hélicoptères, chimie. Cette méthode a été appliquée avec succès à des problèmes variés, tels que le contrôle robotique, le pendule inversé, la planification de tâches, les télécommunications, le backgammon et les échecs.</p> <p>L'apprentissage par renforcement consiste à apprendre par interaction avec l'environnement et, en observant le résultat de certaines actions. Il permet à des machines de déterminer automatiquement le comportement idéal dans un contexte spécifique, afin de maximiser ses performances. Pour cela, un simple retour des résultats est nécessaire pour apprendre comment les machines doivent agir. Ceci est appelé le signal de renforcement.</p> <p>Cela imite la manière fondamentale dont les humains et les animaux apprennent. En tant qu'êtres humains, nous pouvons effectuer des actions et observer leurs résultats sur notre environnement. Connue sous le nom de « cause à effet », c'est sans doute la clé de la construction de notre connaissance tout au long de notre vie.</p> <p>Le cours portera sur des aspects théoriques de l'apprentissage par renforcement (A/R), les algorithmes de bandit pour le compromis exploration-exploitation, et la programmation dynamique avec approximation (PDA), dans le cadre des processus de décision markoviens (PDM).</p> <p>Compétences attendues : Être capable de présenter les bases de l'A/R discret et de la programmation dynamique. Savoir bien distinguer la programmation dynamique, l'apprentissage par renforcement direct, l'approches acteur-critique, l'apprentissage par renforcement indirect.</p>
<p>Simulation numérique</p>	<p style="text-align: center;">4</p>	<p>Ce cours présente les bases des méthodes de simulation utilisées en statistique notamment en statistique bayésienne, en particulier les méthodes de calcul de maximisation et d'intégration en dimension élevée qui sont nécessaires pour traiter les modèles complexes utilisés dans les domaines tels que l'économétrie, la finance, la génétique, l'écologie ou la physique.</p> <p>Compétences attendues : Méthodes de Monte Carlo, méthodes de Quasi Monte Carlo, méthodes MCMC (Algorithme de Metropolis-Hastings, le recuit simulé...), Bootstrap, Inférence bayésienne.</p>



Pôle Sciences, Technologies et Numérique (STN)

Mise en pratique sous Hadoop et Spark	4	Hadoop logiciel/framework utilisé pour répondre aux problématiques « big data ». Un des enjeux concerne le traitement de grandes quantités de données. Ce traitement ne peut être réalisé avec les paradigmes classiques de traitement de données et nécessite l'utilisation de plateformes distribuées de calcul.
Programmation Parallèle	4	Présentation des différentes briques fonctionnelles intervenant lors du développement d'un code de calcul scientifique, de son exploitation en phase de production, et enfin lors de la phase d'extraction et de présentation des résultats. Mise en pratique des bonnes pratiques de développement, de production et de post-processing, tout en découvrant les contraintes spécifiques du calcul scientifique hautes performances. Connaissances : Être capable de structurer un projet de code scientifique, de lancer ou de rejoindre un développement collaboratif, de déployer le code sur un super-calculateur et d'extraire et visualiser les données résultantes. Une introduction à la visualisation scientifique sera également proposée. Compétences attendues : Outils de développement, python comme glue, portage / déploiement sur super calculateur, debug / profiling, stockage de résultats, post-processing, extraction de données, visualisation de données.
Mise en pratique sous Python	2	Ce module sera consacré à l'utilisation du langage Python comme « glue » entre différents langages de plus bas niveau, et les élèves seront amenés à écrire des interfaces et à échanger des données depuis et vers Python.
Management de projet et application transversale	4	Maitriser la méthode SCRUM qui est une méthode de gestion de projet agile. Comprendre les fondements de l'agilité pour bien comprendre et appliquer correctement les méthodes qui en découlent. Connaître les différentes méthodes qui gravitent autour de l'agilité et de SCRUM. Objectifs : Connaître autant les concepts que savoir les appliquer justement. Travail en groupe encadré par des professionnels. À travers la réalisation d'un cahier des charges et du développement d'un pro-prototype, mise en oeuvre des notions acquises dans le module. Connaissances : Gestion de projet Agile - Scrum Compétences attendues : Capacité à travailler en équipe en mode projet.
Total	30	

Unités d'Enseignement (UE)	Crédits	Eléments constitutifs (EC)
SEMESTRE 2 – Mastère CHP/DM/IA		
Open Data Camp	10	<ul style="list-style-type: none"> • Projet en Data Science organisé par la société GAINDE2000 • Durée 30H (1,5 UE = 3 ECTS) • Projet en Data Science organisé par la société BAAMTU • Durée 30H (1,5 UE = 3 ECTS) • Projet en Data Science organisé par la société ATOS • Durée 40H (2 UE = 4 ECTS)
Stage	4	<p>Le machine Learning, apprentissage machine en français, regroupe un ensemble de méthodes mathématiques et algorithmiques permettant d'apprendre une fonction à partir d'exemples. Ces méthodes sont devenues aujourd'hui incontournables pour l'exploitation des bases de données numériques, de plus en plus nombreuses. La session proposera à la fois des cours théoriques sur les enjeux mathématiques de ces méthodes et des travaux pratiques qui permettront d'utiliser les méthodes présentées sur des problématiques de régression ou de classification en utilisant le logiciel R.</p> <p>Compétences attendues :</p> <ol style="list-style-type: none"> 1. Introduction au machine learning, à l'apprentissage supervisé et aux modèles prédictifs. 2. Modèles de régression pénalisée. Critères de performances. Sélection de modèles. 3. Modèles de classification. Courbes ROC. 4. Modèles d'ensemble. Arbres de classification et de régression. Bagging. Random Forest. 5. Agrégation de modèles. Boosting
Total	40	