Python HPC super calculateur

Cours Pratique de 5 jours - 35h
Réf : PYC - Prix 2024 : 2 970€ HT

Python est devenu en quelques années le langage de programmation privilégié de toutes les disciplines scientifiques. Bien qu'il soit interprété, ses librairies scientifiques sont particulièrement performantes car écrites dans des langages compilés, comme C/Cython et très bien parallélisées. Aujourd'hui la lenteur du langage n'est plus un frein et il fonctionne sur les plus puissants des supercalculateurs de la planète. Nous vous proposons d'apprendre les concepts de la programmation parallèle appliquée au HPC au travers des meilleures librairies Python utilisables sur ces environnements.

OBJECTIFS PÉDAGOGIQUES

À l'issue de la formation l'apprenant sera en mesure de :

Posséder une bonne compréhension des concepts des supercalculateurs et de leur programmation

Connaître les librairies Python adaptées au calcul sur HPC

Développer des algorithmes sur supercalculateur avec les librairies MPI4Py, Dask, Xarray, Dask+Scikit-Learn, PyTorch…

Exécuter des workflows avec Prefect

Visualiser des données volumineuses avec DataShader

MÉTHODES PÉDAGOGIQUES

Les travaux pratiques seront réalisés sur un supercalculateur (type Exaion).

LE PROGRAMME

dernière mise à jour : 05/2022

1) Découvrir les supercalculateurs

- Rétrospective du tout premier supercalculateur au plus puissant d'aujourd'hui.
- Qu'est-ce qu'un supercalculateur ?
- Principes et caractéristiques fondamentales : capacités de calcul, capacité réseau et capacité de stockage.
- Les différents classements : Top500, Green500, io500.
- Comment se programme un supercalculateur : les ordonnanceurs/gestionnaires de ressources : SLURM, PBS, ...
- Présentation du supercalculateur Exaion sur lequel nous travaillerons.

Travaux pratiques : Prise en main du supercalculateur Exaion : connexion, installation d'un environnement virtuel et exécution de premiers jobs avec Slurm.

2) Programmation MPI

- Présentation rapide des bases du calcul parallèle avec Python : multithreading, multiprocessing, GIL.
- Les concepts MPI et les différentes librairies disponibles.
- Les différentes primitives : send/receive, scatter/gather, broadcast/reduce, pools de process...

Travaux pratiques : Implémentation de différentes problématiques mettant en œuvre les principales primitives : traitement d'un lot d'images, calcul des décimales de PI...

PARTICIPANTS

Ingénieurs, développeurs, chercheurs, data scientistes, data analystes et toute personne ayant de forts besoins en capacité de calculs avec Python.

PRÉREQUIS

Pratique du langage Python, connaissance des librairies numpy et pandas.

COMPÉTENCES DU FORMATEUR

Les experts qui animent la formation sont des spécialistes des matières abordées. Ils ont été validés par nos équipes pédagogiques tant sur le plan des connaissances métiers que sur celui de la pédagogie, et ce pour chaque cours qu'ils enseignent. Ils ont au minimum cinq à dix années d'expérience dans leur domaine et occupent ou ont occupé des postes à responsabilité en entreprise.

MODALITÉS D'ÉVALUATION

Le formateur évalue la progression pédagogique du participant tout au long de la formation au moyen de QCM, mises en situation, travaux pratiques…
Le participant complète également un test de positionnement en amont et en aval pour valider les compétences acquises.

MOYENS PÉDAGOGIQUES ET TECHNIQUES

• Les moyens pédagogiques et les méthodes d'enseignement utilisés sont principalement : aides audiovisuelles, documentation et support de cours, exercices pratiques d'application et corrigés des exercices pour les stages pratiques, études de cas ou présentation de cas réels pour les séminaires de formation.
• À l'issue de chaque stage ou séminaire, ORSYS fournit aux participants un questionnaire d'évaluation du cours qui est ensuite analysé par nos équipes pédagogiques.
• Une feuille d'émargement par demi-journée de présence est fournie en fin de formation ainsi qu'une attestation de fin de formation si le stagiaire a bien assisté à la totalité de la session.

MODALITÉS ET DÉLAIS D'ACCÈS

L'inscription doit être finalisée 24 heures avant le début de la formation.

ACCESSIBILITÉ AUX PERSONNES HANDICAPÉES

Vous avez un besoin spécifique d'accessibilité ? Contactez Mme FOSSE, référente handicap, à l'adresse suivante psh-accueil@orsys.fr pour étudier au mieux votre demande et sa faisabilité.

3) Programmation MPI, les applications

- Présentation d'exemples d'applications MPI.

Travaux pratiques : Poursuite des travaux pratiques mettant en œuvre les principales primitives.

4) Dask et son écosystème

- Prise en main de dask : les concepts de base, dask array et dataframe.
- Les autres composants de dask : delayed, futures et bags.
- Dask sur HPC : Scheduler et workers, créer un cluster dask : Cluster MPI/Slurm...
- Panorama des différentes librairies de l'écosystème Dask.
- Manipuler des fichiers NetCDF avec XArray.

Travaux pratiques : Analyse de séries temporelles et climatiques, classifications et régressions avec Dask+Scikit-Learn, visualisation de données cartographiques.

5) Dask et les données volumineuses

- Visualiser des données volumineuses avec DataShader et Xarray.
- Créer des pipelines/workflows avec Prefect.
- DaskML : déployer vos algorithmes de machine learning sur HPC.

Travaux pratiques : Poursuite des travaux pratiques sur les analyses et les visualisations de données.

6) Calculer sur GPU

- Les concepts du calcul sur GPU avec Python : matériels, librairies.
- Fonctionnement d'un GPU.
- Dask sur GPU : Créer un cluster CUDA.
- Machine-learning avec PyTorchLightning et RapidsAI.

Travaux pratiques : Mise en œuvre de base avec les librairies PyCuda et Cupy. Manipulation de dataframes avec Dask-CUDF. Machine learning appliqué sur plusieurs nœuds de calculs et plusieurs GPU.

LES DATES

CLASSE À DISTANCE
2025 : 10 mars, 23 juin, 29 sept., 17 nov.

PARIS
2025 : 03 mars, 16 juin, 22 sept., 03 nov.