Publication date : 07/25/2024

Course : Python HPC supercomputer

Practical course - 5d - 35h00 - Ref. PYC
Price : 3070 CHF E.T.

4,5 / 5

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.

Teaching objectives

At the end of the training, the participant will be able to:

- A good understanding of supercomputer concepts and programming
- Python libraries for HPC computing
- Develop algorithms on supercomputers using MPI4Py, Dask, Xarray, Dask+Scikit-Learn, PyTorch libraries...
- Run workflows with Prefect
- Visualizing big data with DataShader

Intended audience

Engineers, developers, researchers, data scientists, data analysts and anyone with a strong need for Python computational capabilities.

Prerequisites

Python language skills, knowledge of numpy and pandas libraries.

Practical details

Teaching methods

Practical work will be carried out on a supercomputer (Exaion type).

Course schedule

PARTICIPANTS

Engineers, developers, researchers, data scientists, data analysts and anyone with a strong need for Python computational capabilities.

PREREQUISITES

Python language skills, knowledge of numpy and pandas libraries.

TRAINER QUALIFICATIONS

The experts leading the training are specialists in the covered subjects. They have been approved by our instructional teams for both their professional knowledge and their teaching ability, for each course they teach. They have at least five to ten years of experience in their field and hold (or have held) decision-making positions in companies.

ASSESSMENT TERMS

The trainer evaluates each participant's academic progress throughout the training using multiple choice, scenarios, hands-on work and more.
Participants also complete a placement test before and after the course to measure the skills they've developed.

1 Discover supercomputers

- From the very first supercomputer to today's most powerful.
- What is a supercomputer?
- Fundamental principles and features: computing capacity, network capacity and storage capacity.
- 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.

Hands-on work

Getting to grips with the Exaion supercomputer: connection, installation of a virtual environment and execution of first jobs with Slurm.

2 MPI programming

- Quick introduction to the basics of parallel computing with Python: multithreading, multiprocessing, GIL.
- MPI concepts and the different libraries available.
- The different primitives: send/receive, scatter/gather, broadcast/reduce, process pools...

Hands-on work

Implementation of various problems involving the main primitives: processing a batch of images, calculating PI decimals, etc.

3 MPI programming, applications

- MPI application examples.

Hands-on work

Continuation of practical work using the main primitives.

4 Dask and its ecosystem

- Getting to grips with dask: basic concepts, dask array and dataframe.
- Other dask components: delayed, futures and bags.
- Dask sur HPC : Scheduler et workers, créer un cluster dask : Cluster MPI/Slurm...
- Panorama des différentes librairies de l'écosystème Dask.
- Handling NetCDF files with XArray.

Hands-on work

Time series and climate analysis, classifications and regressions with Dask+Scikit-Learn, cartographic data visualization.

5 Dask and big data

- Visualize big data with DataShader and Xarray.
- Créer des pipelines/workflows avec Prefect.
- DaskML : déployer vos algorithmes de machine learning sur HPC.

Hands-on work

Continued practical work on data analysis and visualization.

TEACHING AIDS AND TECHNICAL RESOURCES

• The main teaching aids and instructional methods used in the training are audiovisual aids, documentation and course material, hands-on application exercises and corrected exercises for practical training courses, case studies and coverage of real cases for training seminars.
• At the end of each course or seminar, ORSYS provides participants with a course evaluation questionnaire that is analysed by our instructional teams.
• A check-in sheet for each half-day of attendance is provided at the end of the training, along with a course completion certificate if the trainee attended the entire session.

TERMS AND DEADLINES

Registration must be completed 24 hours before the start of the training.

ACCESSIBILITY FOR PEOPLE WITH DISABILITIES

Do you need special accessibility accommodations? Contact Mrs. Fosse, Disability Manager, at psh-accueil@orsys.fr to review your request and its feasibility.

6 GPU computing

- GPU computing concepts with Python: hardware, libraries.
- Fonctionnement d'un GPU.
- Dask sur GPU : Créer un cluster CUDA.
- Machine-learning avec PyTorchLightning et RapidsAI.

Hands-on work

Basic implementation with the PyCuda and Cupy libraries. Dataframe manipulation with Dask-CUDF. Machine learning applied to multiple compute nodes and GPUs.