

# Python Big Data

**MANIPULER DE GRANDES QUANTITÉS DE DONNÉES**

LANGAGES GÉNÉRIQUES

Formation : Python Big Data

Référence : PYBD

Durée : 1 journée – 7 heures

Dates : voir calendrier INTER

Niveau : confirmé

Support de cours : français

Lieu : à distance, Paris, Bordeaux,  
Lyon ou sur site

Tarif : 1890€ HT

Date de mise à jour : 06/02/2025

## Objectifs de la formation

A l'issue de la formation, le stagiaire sera capable de :

- Traiter de grandes quantités de données avec Dask
- Faire du multiprocessing avec Dask
- Manipuler des données avec la librairie xarray
- Réaliser des représentations graphiques avec xarray

## Prérequis

Pour suivre efficacement cette formation, les stagiaires doivent avoir :

Une solide maîtrise de Python: Compréhension des structures de données (listes, dictionnaires, tableaux NumPy), des fonctions, des modules, des classes et des concepts d'orientations objet.

Une bonne connaissance des bibliothèques scientifiques Python: Notamment NumPy, Pandas et Matplotlib, pour la manipulation et la visualisation de données.

Des notions de base en algèbre linéaire et en statistiques: Ces connaissances sont essentielles pour comprendre les concepts sous-jacents aux algorithmes utilisés en analyse de données. Une compréhension des concepts de parallélisme et de concurrence: Bien que ce ne soit pas obligatoire, des connaissances de base sur ces sujets peuvent être utiles pour mieux appréhender les concepts de Dask

## Description

Cette formation est la suite logique de la formation Python Scientifique. Elle vous permettra de traiter, analyser et visualiser vos données les plus volumineuses avec les librairies Python Dask et xarray. Les notions liées à l'utilisation des librairies sont présentées de manière progressive. Afin que les participants s'approprient les notions présentées, plusieurs exercices pratiques sont proposés tout au long de la formation.

## Public concerné

Ingénieurs, développeurs, scientifiques, enseignants, etc.



### Nos Engagements Pédagogiques

- Des salles équipées de vidéoprojecteurs et de PC
- Des outils performants et adaptés aux formations en classe virtuelle
- 50% de cours théorique / 50% d'exercices pratiques
- Des formateurs qualifiés et expérimentés
- Chefs de projets, ingénieurs dans l'industrie




### Styrel: bien plus qu'une formation !

- Intégrateur en Informatique Industrielle depuis 30 ans
- Une vision terrain indépendante de tout constructeur
- Des solutions complètes, du logiciel au matériel, en passant par le service



### En situation de handicap ?

Contactez notre référent handicap :

 +33 1 69 88 85 29 handicap@ame.ametrargroup.com



STYREL

AMETRA GROUP

10<sup>3</sup>

# Programme de la formation

MANIPULER DE GRANDES QUANTITÉS DE DONNÉES

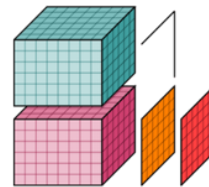
LANGAGES GÉNÉRIQUES

## ➤ Dask (4h)

- Introduction
- Exploiter Dask Array
- Exploiter Dask DataFrame
- Profiter de l'architecture multi-coeurs des ordinateurs pour accélérer l'analyse de données

## ➤ Xarray (3h)

- Introduction
- Créer des DataArrays
- Manipuler les DataArrays
- Utiliser les « Fonctions universelles »
- Représentation graphique
- Créer des DataSets
- Paralléliser des calculs



xarray



## Évaluation et suivi de la formation

### Notre centre de formation

N° d'enregistrement Formation Continue :  
11 91 02 737 91

- A l'issue de chaque demi-journée de formation, une feuille d'émargement est signée, à la fois par le stagiaire et le formateur.
- Des exercices pratiques de programmation en cohérence avec les objectifs de ce programme, permettent de vérifier l'acquisition des compétences tout au long de la session.
- Une attestation de fin de formation est remise au stagiaire lui permettant de faire valoir l'acquisition de ses nouvelles compétences.