

Mis à jour le 16/05/2023

S'inscrire

Formation Dask : La mise à l'échelle des outils Python

2 jours (14 heures)

Présentation

La formation [Dask](#) vous permettra de traiter un set de données volumineux et d'apprendre la programmation parallèle en Python. Dask transforme le code Python des machines multicoeurs locales vers des grands clusters distribués dans le cloud.

Lors de cet enseignement, vous pourrez acquérir les connaissances et compétences nécessaires pour interagir avec Dask. Le logiciel vous propose une interface utilisateur familière en reflétant les API de diverses bibliothèques de l'écosystème tels que PyData ou Pandas.

Dask possède un planificateur de tâches exécutant des graphes de tâches parallèlement. Par conséquent, les charges de travail personnalisées et les collections de haut niveau seront alimentées par vous grâce aux collections de bas niveau.

Comme toutes nos formations, celle-ci vous sera présentée avec la toute dernière version stable en date ainsi que l'ensemble de ses nouveautés ([2023.1.0](#) sortie le 13 janvier 2023 à la date de l'article).

Objectifs

- Savoir utiliser les DataFrames
- Faciliter la mise à l'échelle des bibliothèques PyData
- Mettre à l'échelle différents types de codes
- Comprendre comment déployer localement des tâches à l'aide de plusieurs processus Python

Public visé

- Data Analyst
- Data Miners

- Développeurs

Pré-requis

- Savoir programmer en Python
- Connaissances en traitement et manipulation de données

du Programme de la formation Dask

Introduction

- Présentation de Dask
- Installation et configuration
- Les meilleures pratiques

Fondamentaux

- Déploiement de clusters de tâches
- Prise en charge d'un cadre de tâches en temps réel
- Déploiement de Dask
 - Créer des tableaux Dask
 - Chevauchement des calculs
 - Conception interne
 - Tableaux clairsemés
- Créer et stocker des DataFrames Dask
 - DataFrame et Parquet
 - DataFrame et SQL
 - Indexation dans les DataFrames
 - Conception interne

Internes

- Dask Interne
- Comprendre les coûts associés au code
- Planification
- Travailler avec des graphiques de tâches
- Débogage et performances

Gestion de l'environnement

- Maintenir des environnements cohérents
- Installations temporaires
- Envoyer des fichiers directement aux nœuds de calcul
- Redéfinition des objets Dask

- Ordonnanceurs mono-machine

Sélection du backend de collecte

- Changer la bibliothèque principale par défaut
- Définir un nouveau backend de collection
- Définir une méthode distincte pour chacune routine de création répartissable

Référence

- API
- Interface de ligne de commande
- Directives de développement
- Directives du responsable
- Journal de modifications
- Configuration

Xarray avec dask arrays

- Lancement du client pour fournir un dashboard
- Ouverture d'un dataset
- Exécuter des opérations Xarray standard
- Parallélisation automatique et workflows personnalisés
- Calcul parallèle et calcul continu
- Conservation des données en mémoire

Sociétés concernées

Cette formation s'adresse aux entreprises, petites ou grandes, souhaitant former ses équipes à une nouvelle technologie informatique avancée.

Méthodes pédagogiques

Stage Pratique : 60% Pratique, 40% Théorie. Support de la formation distribué au format numérique à tous les participants.

Organisation

Le cours alterne les apports théoriques du formateur soutenus par des exemples et des séances de réflexions, et de travail en groupe.

Validation

À la fin de la session, un questionnaire à choix multiples permet de vérifier l'acquisition correcte des compétences.

Sanction

Une attestation sera remise à chaque stagiaire qui aura suivi la totalité de la formation.