

Mis à jour le 11/04/2024

S'inscrire

# Formation Apache Arrow

2 jours (14 heures)

## Présentation

Traitez efficacement vos données en mémoire avec notre formation Apache Arrow pour permettre d'avoir des échanges fluides entre vos différents frameworks.

À l'issue de cette formation, vous allez apprendre à utiliser Apache Arrow pour manipuler des données massives de manière efficace, à intégrer plusieurs outils de traitement de données et aussi à effectuer des opérations analytiques complexes.

De plus, cet outil vous enseignera à maîtriser plusieurs concepts avancés de gestion de mémoire, d'optimisation des performances et de parallélisme des données à grande échelle.

Par ailleurs, cette technologie vous garantit la portabilité des données et elle simplifie le développement de logiciels.

Comme dans toutes nos formations, celle-ci vous présentera la toute dernière version d'Apache Arrow (à la date de rédaction de l'article : [Apache Arrow 15](#)).

## Objectifs

- Explorer l'architecture interne d'Apache Arrow
- Développer et tester des fonctionnalités avec Apache Arrow
- Intégrer les pratiques d'intégration continue et de packaging
- Savoir effectuer des opérations analytiques

## Public visé

- Développeurs
- Ingénieurs
- Data Analyst

## Pré-requis

- Connaissance de base en programmation (Python, Java)
- Compréhension des concepts de traitement de données et de manipulation de fichiers

## PROGRAMME DE NOTRE FORMATION APACHE ARROW

### INTRODUCTION À APACHE ARROW

- Présentation d'Apache Arrow et son écosystème
- Importance d'Apache Arrow dans le Big Data et l'analyse de données
- Concepts clés : colonnes, tableaux, schémas, et types de données
- Avantages de l'utilisation d'Apache Arrow pour le traitement de données en mémoire
- Installation d'Apache Arrow et configuration de l'environnement de développement

### ARCHITECTURE ET CONCEPTION

- Vue d'ensemble de l'architecture interne d'Apache Arrow
- Comprendre la représentation en mémoire et le format de données columnar
- La bibliothèque Arrow et ses différents langages de programmation (C++, Python, Java, etc.)
- Exploration des interfaces IPC (Inter-Process Communication) et Flight (gRPC)
- Utilisation des datasets et des RecordBatches

### CONTRIBUER À APACHE ARROW

- Guide pour les nouveaux contributeurs : Comment commencer
- Processus pour signaler des bugs et suggérer des fonctionnalités
- Réaliser et soumettre votre première Pull Request (PR)
- Bonnes pratiques pour le travail collaboratif avec Git et GitHub
- Cycle de vie d'une pull request et revue de code

### DÉVELOPPEMENT ET TESTS

- Configuration de l'environnement de développement pour différentes langues
- Compilation des bibliothèques Arrow : étapes et dépannage
- Écrire, exécuter et déboguer des tests unitaires et d'intégration
- Appliquer les conventions de style de codage d'Apache Arrow
- Utilisation des outils de développement quotidiens comme Archery

### INTÉGRATION CONTINUE ET PACKAGING

- Présentation de l'intégration continue dans le projet Arrow

- Exécution de builds Docker pour la validation des changements
- Utilisation de Crossbow pour le packaging et les tests
- Comprendre les outils et les processus de build automatisés
- Résolution des problèmes courants lors de l'intégration continue

## TUTORIELS ET RESSOURCES SUPPLÉMENTAIRES

- Tutoriels pratiques pour l'utilisation d'Arrow en Python et R
- Accéder à des ressources complémentaires pour approfondir vos connaissances
- Aide avec la documentation : comment contribuer et améliorer les documents
- Exemples d'utilisation avancée d'Apache Arrow dans des cas réels
- Discussion sur les dernières mises à jour et les fonctionnalités à venir

## Sociétés concernées

Cette formation s'adresse à la fois aux particuliers ainsi qu'aux entreprises, petites ou grandes, souhaitant former ses équipes à une nouvelle technologie informatique avancée ou bien à acquérir des connaissances métiers spécifiques ou des méthodes modernes.

## Positionnement à l'entrée en formation

Le positionnement à l'entrée en formation respecte les critères qualité Qualiopi. Dès son inscription définitive, l'apprenant reçoit un questionnaire d'auto-évaluation nous permettant d'apprécier son niveau estimé sur différents types de technologies, ses attentes et objectifs personnels quant à la formation à venir, dans les limites imposées par le format sélectionné. Ce questionnaire nous permet également d'anticiper certaines difficultés de connexion ou de sécurité interne en entreprise (intraentreprise ou classe virtuelle) qui pourraient être problématiques pour le suivi et le bon déroulement de la session de formation.

## 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.