

Mis à jour le 26/09/2024

S'inscrire

# Formation Apache Beam

2 jours (14 heures)

## Présentation

Vous cherchez à maîtriser une solution puissante et flexible pour traiter de grandes quantités de données en flux ou en batch ? Notre formation Apache Beam vous offre une introduction complète à ce framework open source, conçu pour unifier le traitement des données dans divers environnements.

[Apache Beam](#) permet de créer des pipelines de traitement de données portables, pouvant s'exécuter sur plusieurs moteurs comme Google Cloud Dataflow, Apache Spark et Flink.

Grâce à ses SDKs compatibles avec différents langages (Python, Java), il devient l'outil idéal pour les Data Engineers, Data Scientists et développeurs cherchant à traiter des volumes importants de données de manière scalable.

Durant cette formation, vous apprendrez à concevoir des pipelines efficaces, à gérer les sources et les puits de données, et à optimiser vos traitements en temps réel grâce à des techniques avancées telles que le fenêtrage et les déclencheurs.

Vous acquérez également les compétences nécessaires pour exécuter et déployer vos pipelines dans des environnements cloud, comme Google Cloud Platform.

Comme pour toutes, la formation Apache Beam vous sera présentée avec ses toutes [dernières nouveautés](#) (à la date de rédaction de l'article).

## Objectifs

- Comprendre les fondamentaux d'Apache Beam et sa place dans l'écosystème du traitement de données
- Savoir comparer Apache Beam à d'autres technologies
- Maîtriser l'architecture d'Apache Beam et ses différents composants
- Concevoir, structurer et exécuter des pipelines de traitement de données en flux et en batch

- Optimiser le traitement de données via le fenêtrage et la gestion des éléments tardifs

## Public visé

- Data Engineers
- Data Scientists
- Développeurs Big Data
- Architectes de données

## Pré-requis

- Connaissance des concepts fondamentaux en traitement de données et en data engineering
- Expérience avec un langage de programmation (idéalement Python ou Java)
- Compréhension des principes de bases de données et de data lakes

# PROGRAMME DE NOTRE FORMATION APACHE BEAM

## INTRODUCTION À APACHE BEAM

- Vue d'ensemble des technologies de traitement de données et positionnement d'Apache Beam
- Comparaison avec d'autres technologies comme Spark, Flink et Google Cloud Dataflow
- Installation et configuration de l'environnement Apache Beam

## ARCHITECTURE ET FONCTIONNALITÉS D'APACHE BEAM

- Compréhension de l'architecture et des composants clés de Beam
- Détail des SDKs disponibles et leur compatibilité avec les langages de programmation
- Présentation des différents runners (Direct, Dataflow, Flink, Spark)
- Back-ends de traitement distribué supportés

## MODÈLE DE PROGRAMMATION APACHE BEAM

- Flux de données dans un pipeline Beam et terminologies de base
- Création d'un pipeline simple de type WordCount pour illustrer le concept
- Analyse des transformations essentielles :
  - ParDo
  - GroupByKey
  - Windowing

## DÉVELOPPEMENT ET MISE EN ŒUVRE DE PIPELINES

- Écriture et structuration d'un pipeline Beam

- Utilisation de transformations complexes et création de transformations composites
- Gestion des sources et des puits de données : Lecture et écriture de données

## EXÉCUTION ET DÉPLOIEMENT DE PIPELINES

- Méthodes d'exécution de pipelines : locale, cloud et en cluster
- Configuration des pipelines pour différents environnements d'exécution
- Déploiement sur Google Cloud Platform avec Dataflow

## FENÊTRAGE, DÉCLENCHEURS ET GESTION DES ÉLÉMENTS TARDIFS

- Stratégies de fenêtrage :
  - Tumbling
  - Sliding
  - Session
  - Global Windows
- Utilisation des déclencheurs pour gérer les éléments tardifs
- Compréhension des Watermarks et leur impact sur le traitement des données en temps réel

## 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.