



ORGANISME DE FORMATION AUX TECHNOLOGIES ET METIERS DE L'INFORMATIQUE

Formation IBM InfoSphere Advanced DataStage - Parallel Framework v11.5

N° ACTIVITÉ : 11 92 18558 92

TÉLÉPHONE : 01 85 77 07 07

E-MAIL : inscription@hubformation.com

Ce cours est conçu pour présenter les techniques avancées de développement de travaux parallèles dans DataStage v11.5. Dans ce cours, vous développerez une compréhension plus approfondie de l'architecture DataStage, y compris une compréhension plus approfondie des environnements de développement DataStage et d'exécution. Cela vous permettra de concevoir des tâches parallèles robustes, moins sujettes aux erreurs, réutilisables et optimisées pour de meilleures performances.

Référence	KM404G
Durée	3 jours (21h)
Tarif	2 385 €HT

Objectifs

- | Compiler et exécuter des travaux
- | Partitionner et collecter des données
- | Trier des données
- | Mettre en mémoire tampon dans les travaux parallèles
- | Identifier les types de données d'infrastructure parallèles
- | Créer des composants réutilisables
- | Activer la fonctionnalité d'optimisation équilibrée dans Designer

Public

| Développeurs DataStage expérimentés recherchant une formation aux techniques de travail plus avancées de DataStage et cherchant à comprendre l'architecture de la structure parallèle

Prérequis

- | cours IBM InfoSphere DataStage Essentials ou équivalent
- | au moins un an d'expérience dans le développement de travaux parallèles à l'aide de DataStage

Programme de la formation

Introduction à l'architecture du framework parallèle

- | Décrire l'architecture de traitement parallèle
- | Décrire le parallélisme des pipelines et des partitions
- | Décrire le rôle du fichier de configuration
- | Concevoir un travail qui crée des données de test robustes

Compiler et exécuter des travaux

- | Décrire les parties principales du fichier de configuration
- | Décrire le processus de compilation et le SST généré par le processus de compilation
- | Décrivez le rôle et les principales parties de la partition
- | Décrire le processus d'exécution du travail

Partitionner et collecter des données

- | Comprendre le fonctionnement du partitionnement dans la structure
- | Affichage des partitionneurs dans la partition
- | Sélection d'algorithmes de partitionnement
- | Générer des séquences de nombres (clés de substitution) dans un environnement

PROCHAINES SESSIONS

Pour connaître les prochaines dates ou organiser un intra-entreprise, contactez-nous, nous vous répondrons sous 72 heures.

parallèle partitionné

Tri des données

- | Trier les données dans le cadre parallèle
- | Trouver des tris insérés dans le score
- | Réduire le nombre de tris insérés
- | Optimiser les tâches Fork-Join
- | Utilisez les étapes de tri pour déterminer la dernière ligne d'un groupe
- | Décrire la clé de tri et la logique de la clé de partitionnement dans le cadre parallèle

Mise en mémoire tampon dans les travaux parallèles

- | Décrire le fonctionnement de la mise en mémoire tampon dans les travaux parallèles
- | Ajuster les tampons dans les travaux parallèles
- | Eviter les conflits de tampon

Types de données d'infrastructure parallèles

- | Décrire les ensembles de données virtuels
- | Décrire les schémas
- | Décrire les mappages et les conversions de types de données
- | Décrivez comment les données externes sont traitées
- | Gérer les nulls
- | Travailler avec des données complexes

Composants réutilisables

- | Créer un fichier de schéma
- | Lire un fichier séquentiel à l'aide d'un schéma
- | Décrire la propagation de colonne d'exécution (RCP)
- | Activer et désactiver le RCP
- | Créer et utiliser des conteneurs partagés

Optimisation équilibrée

- | Activer la fonctionnalité d'optimisation équilibrée dans Designer
- | Décrire le flux de travail d'optimisation équilibrée
- | Répertoirez les différentes options d'optimisation équilibrée.
- | Traitement en phase poussée vers une source de données
- | Traitement du stade poussé vers une cible de données
- | Optimiser une tâche en accédant au système de fichiers Hadoop HDFS
- | Comprendre les limites des optimisations équilibrées

Méthode pédagogique

Chaque participant travaille sur un poste informatique qui lui est dédié. Un support de cours lui est remis soit en début soit en fin de cours. La théorie est complétée par des cas pratiques ou exercices corrigés et discutés avec le formateur. Le formateur projette une présentation pour animer la formation et reste disponible pour répondre à toutes les questions.

Méthode d'évaluation

Tout au long de la formation, les exercices et mises en situation permettent de valider et contrôler les acquis du stagiaire. En fin de formation, le stagiaire complète un QCM d'auto-évaluation.

Accessibilité

Les sessions de formation se déroulent sur des sites différents selon les villes ou les dates, merci de nous contacter pour vérifier l'accessibilité aux personnes à mobilité réduite.
Pour tout besoin spécifique (vue, audition...), veuillez nous contacter au 01 85 77 07 07.