

IBM InfoSphere DataStage

IBM InfoSphere Advanced DataStage - Parallel Framework 11.5

3 jours (21h00) | ★★★★★ 4,6/5 | IBM-DATAV | Évaluation qualitative de fin de stage |
Formation délivrée en présentiel ou distanciel ⁽¹⁾

Formations Informatique > IBM > IBM InfoSphere DataStage



À l'issue de ce stage vous serez capable de :

- Décrire l'architecture parallèle et les rôles et éléments des fichiers de configuration de DataStage
- Décrire le processus d'exécution du runtime job
- Expliquer comment fonctionne le partitionnement et la collecte de données dans la structure parallèle
- Lister et sélectionner les algorithmes de partitionnement et de collecte
- Décrire les techniques d'optimisation de tri et la fonction tampon dans la structure parallèle
- Décrire les techniques d'optimisation de tampon
- Reconnaître et travailler les types de données et éléments de la structure parallèle
- Décrire le fonctionnement et l'utilisation de la propagation de la colonne Runtime (RCP)
- Créer des composants réutilisables d'emploi à l'aide de conteneurs partagés
- Décrire le fonctionnement et l'utilisation de l'optimisation équilibré
- Optimiser les travaux parallèles DataStage.

Niveau requis

Avoir suivi la formation IBM-DATFND "IBM InfoSphere DataStage – Essentials" ou avoir les connaissances équivalentes.

Public concerné

Développeurs DataStage expérimentés souhaitant approfondir leurs connaissances dans la mise en oeuvre de techniques avancées et la compréhension de l'architecture parallèle framework.

(1) Modalité et moyens pédagogique :

Formation délivrée en présentiel ou distanciel * (e-learning, classe virtuelle, présentiel à distance). Le formateur alterne entre méthodes ** démonstrative, interrogative et active (via des travaux pratiques et/ou des mises en situation). La validation des acquis peut se faire via des études de cas, des quiz et/ou une certification.

Les moyens pédagogiques mis en oeuvre (variables suivant les formations) sont : ordinateurs Mac ou PC (sauf pour les cours de l'offre Management), connexion internet fibre, tableau blanc ou paperboard, vidéoprojecteur ou écran tactile interactif (pour le distanciel). Environnements de formation installés sur les postes de travail ou en ligne. Supports de cours et exercices.

* Nous consulter pour la faisabilité en distanciel. ** Ratio variable selon le cours suivi.

Cette formation :

- Est animée par un consultant-formateur dont les compétences techniques, professionnelles et pédagogiques ont été validées par des diplômes et/ou testées et approuvées par l'éditeur et/ou par M2i Formation
- Bénéficie d'un suivi de son exécution par une feuille de présence émargée par demi-journée par les stagiaires et le formateur.

Programme

Introduction à l'architecture de cadres parallèles

- Décrire l'architecture de traitement en 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

Compilation et exécution de travaux

- Décrire les principales parties du fichier de configuration
- Décrire le processus de compilation et la SST que le processus de compilation génère
- Décrire le rôle et les principales parties du SCORE (Solution for Compliance in a Regulated Environment)
- Décrire le processus d'exécution du travail 3
- Partitionnement et collecte de données
- Comprendre comment fonctionne le partitionnement dans le cadre
- Affichage des partitions dans le SCORE
- Sélection d'algorithmes de partitionnement
- Générer des séquences de nombres (clés de substitution) dans un environnement partitionné et parallèle

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 travaux de 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 partition dans le cadre parallèle

Buffering dans les jobs parallèles

- Décrire le fonctionnement du tamponnage en parallèle
- Accorder des tampons dans des tâches parallèles
- Evitez les conflits de mémoire tampon

Types de données de cadre parallèle

- Décrire les ensembles de données virtuelles
- Décrire les schémas
- Décrire les mappages et les conversions de type de données
- Décrire la façon dont les données externes sont traitées
- Manipuler des "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 RCP
- Créer et utiliser des conteneurs partagés

Optimisation équilibrée

- Activer la fonctionnalité d'optimisation équilibrée dans Designer
- Décrire le flux de travail de Balanced Optimization

- Enumérer les différentes options de Balanced Optimization
- Traitement en mode push à une source de données
- Traitement en mode push vers une cible de données
- Optimiser un accès au système Hadoop HDFS
- Comprendre les limites des Balanced Optimizations

Modalités d'évaluation des acquis

- En cours de formation, par des études de cas ou des travaux pratiques
- Et, en fin de formation, par un questionnaire d'auto-évaluation ou une certification (M2i ou éditeur)