

Formation IBM InfoSphere Advanced DataStage - Parallel Framework v11.5

Objectifs de la formation InfoSphere v11.5

Cette formation InfoSphere Advanced DataStage - Parallel Framework v11.5 est consacrée aux techniques avancées de développement de jobs en parallèle dans DataStage v11.5. Vous approfondissez votre connaissance de l'architecture de DataStage, ainsi que de ses environnements de développement et d'exécution. Vous pourrez ainsi créer des tâches en parallèle robustes, plus fiables, réutilisables et optimisées pour de meilleures performances.

À l'issue de cette formation, les participants auront acquis les connaissances et les compétences nécessaires pour :

- Décrire l'architecture parallèle
- Décrire les rôles et éléments des fichiers de configuration de DataStage
- Décrire le processus d'exécution du runtime job
- Décrire 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
- Décrire la fonction tampon dans la structure parallèle
- Décrire les techniques d'optimisation de tampon
- Décrire et travailler les types de données et éléments de la structure parallèle

Capital Social: 50000 DT **MF:** 1425253/M/A/M/000 **RC:** B91211472015

Tél. / Fax.: +216 73362 100 **Email:** contact@itgate-training.com **Web:** www.itgate-training.com

Adresse : 12 Rue Abdelkadeur Daghri - Hammam Sousse 4011 – Tunisie

- 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

À qui s'adresse cette formation ?

Public :

Ce cours est destiné aux développeurs DataStage expérimentés souhaitant perfectionner leurs techniques de jobs et comprendre l'architecture du framework Parallel.

Prérequis :

Pour suivre ce cours, il est nécessaire d'avoir suivi la formation InfoSphere DataStage Essentials et d'avoir au minimum une année d'expérience en développement de jobs parallèles avec DataStage.

Contenu du cours InfoSphere v11.5

Introduction à l'architecture du framework Parallel

Comprendre l'architecture de Parallel

Comprendre le pipeline et la partition

Comprendre le rôle du fichier de configuration

Créer un job de test de données robuste

Compilation et d'exécution de jobs

Connaître les parties principales du fichier de configuration

Connaître le processus de compilation et l'OSH

Capital Social: 50000 DT **MF:** 1425253/M/A/M/000 **RC:** B91211472015

Tél. / Fax.: +216 73362 100 **Email:** contact@itgate-training.com **Web:** www.itgate-training.com

Adresse : 12 Rue Abdelkadeur Daghri - Hammam Sousse 4011 – Tunisie

Connaître le rôle et les parties principales du Score

Connaître le processus d'exécution des tâches

Le partitionnement et la collecte des données

Comprendre le partitionnement dans le framework

Voir le partitionnement dans Score

Sélectionner les algorithmes de partitionnement

Générer des séquences de nombres (clés de substitution) dans un environnement parallèle partitionné

Trier les données

Trier les données dans Parallel

Trouver des types de tris dans le Score

Réduire le nombre de tris

Optimiser les tâches Fork-Join

Utiliser la fonction Sort stages pour déterminer la dernière ligne d'un groupe

Décrire les clés de tri et la logique de la clé de partition dans Parallel

Les buffers dans les tâches en parallèle

Comprendre le fonctionnement des buffers dans les jobs en parallèle

Régler les buffers

Éviter les restrictions de buffers

Les types de données dans Parallel

Connaître les ensembles de données virtuelles

Connaître les schémas

Connaître le mapping et les conversions de types de données

Comprendre le traitement des données externes

Capital Social: 50000 DT **MF:** 1425253/M/A/M/000 **RC:** B91211472015

Tél. / Fax.: +216 73362 100 **Email:** contact@itgate-training.com **Web:** www.itgate-training.com

Adresse : 12 Rue Abdelkadeur Daghrrir - Hammam Sousse 4011 – Tunisie

Gérer les valeurs nulles

Travailler avec des données complexes

Les composants réutilisables

Créer un fichier Schéma

Lire un fichier séquentiel en utilisant un schéma

Connaître Runtime Column Propagation (RCP)

Activer et désactiver RCP

Créer et utiliser des containers partagés

L'optimisation et équilibrage

Activer la fonction Balanced Optimization dans Designer

Connaître le déroulement de l'optimisation équilibrée

Connaître les différentes options de la fonction d'optimisation

Pousser le stage processing vers une source de données

Pousser le stage processing vers une cible de données

Optimiser ses tâches à l'aide d'Hadoop HDFS

Comprendre les limites de l'optimisation équilibrée.