

IBM InfoSphere Advanced DataStage - Parallel Framework v11.5

Techniques avancées de développement de tâches parallèles dans DataStage

Cours Pratique de 3 jours - 21h

Réf : IB7 - Prix 2024 : nous consulter

Avec cette formation, vous explorerez des techniques avancées de développement de tâches parallèles dans DataStage. Vous approfondirez votre compréhension de l'architecture DataStage, en mettant l'accent sur les environnements de développement et d'exécution. Grâce à cette connaissance approfondie, vous serez en mesure de concevoir des tâches parallèles robustes, moins sujettes aux erreurs, facilement réutilisables et optimisées pour des performances améliorées.

OBJECTIFS PÉDAGOGIQUES

À l'issue de la formation l'apprenant sera en mesure de :

Comprendre l'architecture de traitement parallèle ainsi que les environnements de développement et d'exécution

Connaître les processus de compilation et d'exécution des travaux

Optimiser la phase de tri et de mise en mémoire tampon dans le cas de travaux en parallèle

Comprendre comment utiliser les types de données liés au cadre parallèle

Créer des composants de travaux réutilisables

Générer des étapes personnalisées intégrant des fonctions de transformation

Traiter des données XML dans les travaux DataStage

Concevoir une tâche qui traite une base de données de schéma en étoile

MÉTHODES PÉDAGOGIQUES

Animation de la formation en français. Support de cours officiel au format numérique et en anglais. Bonne compréhension de l'anglais à l'écrit.

CERTIFICATION

Cette formation prépare à la certification IBM Certified Solution Developer - InfoSphere DataStage v11.3.

LE PROGRAMME

dernière mise à jour : 01/2024

1) Introduction à l'architecture du Parallel Framework

- Décrire l'architecture de traitement parallèle.
- Présentation du pipeline et des partitions.
- Décrire le rôle du fichier de configuration.
- Conception d'une tâche de création de tests de données.

2) Compilation et exécution des tâches

- Décrire les principales parties du fichier de configuration.
- Le processus de compilation.
- L'OSH généré par le processus de compilation.
- Rôle et principales parties du Score.

PARTICIPANTS

Développeurs expérimentés sur DataStage.

PRÉREQUIS

Avoir suivi la formation IBM InfoSphere DataStage Essentials ou posséder les connaissances équivalentes. Avoir au moins un an d'expérience dans le développement de tâches parallèles avec DataStage.

COMPÉTENCES DU FORMATEUR

Les experts qui animent la formation sont des spécialistes des matières abordées. Ils sont agréés par l'éditeur et sont certifiés sur le cours. Ils ont aussi été validés par nos équipes pédagogiques tant sur le plan des connaissances métiers que sur celui de la pédagogie, et ce pour chaque cours qu'ils enseignent. Ils ont au minimum trois à dix années d'expérience dans leur domaine et occupent ou ont occupé des postes à responsabilité en entreprise.

MODALITÉS D'ÉVALUATION

Évaluation des compétences visées en amont de la formation.

Évaluation par le participant, à l'issue de la formation, des compétences acquises durant la formation.

Validation par le formateur des acquis du participant en précisant les outils utilisés : QCM, mises en situation...

À l'issue de chaque stage, ITTCERT fournit aux participants un questionnaire d'évaluation du cours qui est ensuite analysé par nos équipes pédagogiques. Les participants réalisent aussi une évaluation officielle de l'éditeur. Une feuille d'émargement par demi-journée de présence est fournie en fin de formation ainsi qu'une attestation de fin de formation si le participant a bien assisté à la totalité de la session.

MOYENS PÉDAGOGIQUES ET TECHNIQUES

Les ressources pédagogiques utilisées sont les supports et les travaux pratiques officiels de l'éditeur.

MODALITÉS ET DÉLAIS D'ACCÈS

L'inscription doit être finalisée 24 heures avant le début de la formation.

ACCESSIBILITÉ AUX PERSONNES HANDICAPÉES

Vous avez un besoin spécifique d'accessibilité ? Contactez Mme FOSSE, référente handicap, à l'adresse suivante psh-accueil@orsys.fr pour étudier au mieux votre demande et sa faisabilité.

3) Partitionnement et collecte de données

- Comprendre le fonctionnement du partitionnement dans le Framework.
- Afficher des partitions dans le Score.
- Sélectionner des algorithmes de partitionnement.
- Générer des séquences de nombres (clés de substitution) dans un environnement parallèle partitionné.

4) Tri des données

- Trier les données.
- Rechercher les tris insérés dans le Score.
- Réduire le nombre de tris insérés.
- Optimiser l'emploi des jointures.
- Utiliser les étapes de tri pour déterminer la dernière ligne d'un groupe.
- Décrire la clé de tri et la clé logique de partitionnement.

5) Gestion des buffers en exécution parallèle

- Fonctionnement des buffers.
- Optimisation des buffers.

6) Types de données

- Décrire les Virtual Data Sets.
- Décrire les schémas.
- Conversion de types.
- Gestion des données externes.
- Gérer les valeurs nulles.
- Travailler avec des données complexes

7) Composants réutilisables

- Créer un schéma de fichier.
- Lecture séquentielle d'un fichier à l'aide d'un schéma.
- Utiliser Runtime Column Propagation (RCP).
- Créer et utiliser des conteneurs.

8) Optimisation équilibrée

- Activer la fonctionnalité d'optimisation équilibrée dans Designer.
- Décrire les processus d'optimisation.
- Répertorier les différentes options d'optimisation équilibrée.

LES DATES

Nous contacter