

IBM InfoSphere DataStage v11.7 - Les bases

Créer, optimiser et administrer des jobs parallèles pour des flux ETL performants.

 Présentiel ou en classe à distance



4 jours (28 h)

Réf.: IBMKM304

Formation officielle



Idéal en
Distanciel

IBM InfoSphere DataStage est un outil ETL (Extract, Transform, Load) puissant permettant de gérer et d'orchestrer des flux de données complexes. Conçu pour des environnements à forte volumétrie, il offre un moteur de traitement parallèle avancé, garantissant performance et scalabilité. Grâce à ses fonctionnalités comme la gestion des métadonnées, l'intégration avec les bases de données relationnelles et le support des connecteurs multiples, DataStage est une solution de choix pour les entreprises souhaitant exploiter efficacement leurs données.

Se former à IBM DataStage v11.7 est essentiel pour les professionnels de la data souhaitant optimiser les processus d'intégration et de transformation des données. La maîtrise de cet outil permet de construire des pipelines de données robustes, d'améliorer la qualité des données et de répondre aux enjeux de la gouvernance et de la conformité.

A qui s'adresse cette formation ?



Pour qui

- Chefs de projets et développeurs ETL responsables des extractions et transformations des données en utilisant DataStage



Prérequis

- Bonnes connaissances de l'utilisation du système d'exploitation Windows
- Être familier avec les techniques d'accès aux bases de données

Programme

1 - Introduction à DataStage

- Lister et décrire les utilisations de DataStage
- Lister et décrire les clients DataStage
- Décrire le Workflow DataStage
- Décrire les deux types de parallélisme possibles avec les tâches en parallèle dans DataStage

2 - Déploiement

- Identifier les composants Information Server qui doivent être installés
- Décrire en quoi consiste un déploiement
- Décrire les différentes options de déploiement du domaine
- Décrire le processus d'installation
- Démarrer Information Server

3 - Administration DataStage

- Ouvrir la console Web
- Créer des nouveaux utilisateurs et des groupes
- Assigner des rôles
- Suite et composant aux utilisateurs et aux groupes
- Donner les accès DataStage aux utilisateurs
- Connexion à DataStage en tant qu'administrateur
- Ajouter un utilisateur DataStage dans l'onglet Permission et spécifier le rôle de l'utilisateur
- Spécifier les projets par défaut
- Lister et décrire les variables d'environnement importantes

4 - Travailler avec les métadonnées

- Connexion à DataStage
- Se déplacer dans DataStage Designer
- Importer et exporter les objets DataStage dans un fichier
- Importer une définition de table pour un fichier séquentiel

5 - Créer des tâches parallèles

- Concevoir des tâches parallèles dans DataStage
- Définir le paramètre de la tâche
- Utiliser Row Generator, Peek et Annotation dans une tâche
- Compiler les tâches
- Activer les tâches
- Surveiller les journaux des tâches
- Créer un jeu de paramètres et l'utiliser dans une tâche

6 - Accéder aux données séquentielles

- Comprendre les étapes pour accéder aux différents types de fichiers de données
- Lire et écrire des séquences de fichiers via Sequential File Stage
- Lire et écrire des fichiers d'ensemble de données via les étapes DataStage
- Créer des liens de rejet
- Travailler avec les valeurs NULL dans les fichiers séquentiels
- Lire des fichiers séquentiels multiples via des modèles de fichiers
- Utiliser des lecteurs multiples

7 - Partitionner et collecter

- Décrire l'architecture du processus parallèle
- Décrire le pipeline et la partition du parallélisme
- Lister et décrire le partitionnement et les algorithmes de rassemblement
- Décrire les fichiers de configuration
- Décrire le processus de compilation des tâches parallèles
- Expliquer OSH

8 - Combiner les données

- Combiner les données via les étapes Lookup
- Définir les "range lookups"
- Combiner les données via la fusion
- Combiner les données via la jointure
- Combiner les données via Funnel

9 - Étape de traitement de groupe

- Trier les données
- Combiner l'étape d'agrégation
- Combiner les données et supprimer les doublons

10 - Étape de la transformation

- Utiliser l'étape de transformation dans les tâches parallèles
- Définir les contraintes
- Définir les dérivations
- Utiliser les variables
- Créer un jeu de paramètres et utiliser ses paramètres dans les contraintes et les dérivations

11 - Fonctions Repository

- Améliorer la recherche simple
- Améliorer la recherche avancée

- Améliorer l'analyse d'impact
- Comparer les différences entre deux définitions de table
- Comparer les différences entre deux tâches

12 - Travailler avec des données relationnelles

- Importer les définitions de tables pour les tables relationnelles
- Créer les connexions de données
- Utiliser ODBC et le connecteur DB2 dans une tâche
- Utiliser SQL Builder pour définir SQL SELECT et INSERT
- Utiliser les liens d'entrée multiples dans le connecteur pour mettre à jour les tables multiples dans une transaction simple
- Créer des liens de rejets à partir du Connecteur pour capturer les lignes avec les erreurs SQL

13 - Contrôler les tâches

- Utiliser le séquenceur de tâches DataStage pour créer une tâche qui contrôle une séquence de tâches
- Utiliser les liens du séquenceur et les étapes pour contrôler la séquence d'un jeu de tâches
- Utiliser les déclencheurs et les étapes pour contrôler les conditions selon lesquelles les tâches sont activées
- Transmettre des informations dans les paramètres des tâches à partir du master de contrôle des tâches
- Prise en main des erreurs et des exceptions



Les objectifs de la formation

- Comprendre comment combiner les données en utilisant les instructions Lookup, Join et Merge
- Savoir transformer, combiner, filtrer et trier des données à l'aide de travaux (jobs)
- Être en mesure de construire des agrégats de données
- Maîtriser la création de rapports avec des travaux (jobs)
- Être en mesure de construire des agrégats de données



Evaluation

- Cette formation fait l'objet d'une évaluation formative.



Les points forts de la formation

- Les apports théoriques sont très largement complétés par des phases de mise en pratique qui amènent les participants à réaliser de nombreuses manipulations.
- Des consultants expérimentés partagent leur savoir-faire avec les participants.
- La qualité d'une formation officielle IBM (support de cours numérique en anglais).