

Introduction aux bases de données et ETL

2 j (14 heures)

Ref : BDDI

Public

Toute personne appelée à travailler sur un projet décisionnel d'entreprise (développeur, responsable de projet décisionnel, etc.)

Pré-requis

Maîtriser les bases de l'informatique bureautique (Excel, etc.)

Moyens pédagogiques

Formation réalisée en présentiel ou à distance selon la formule retenue
Exposés, cas pratiques, synthèse, assistance post-formation pendant trois mois
Un poste par stagiaire, vidéoprojecteur, support de cours fourni à chaque stagiaire

Modalités de suivi et d'évaluation

Feuille de présence émargée par demi-journée par les stagiaires et le formateur
Exercices de mise en pratique ou quiz de connaissances tout au long de la formation permettant de mesurer la progression des stagiaires
Questionnaire d'évaluation de la satisfaction en fin de stage
Auto-évaluation des acquis de la formation par les stagiaires
Attestation de fin de formation

Les sociétés disposent aujourd'hui d'une quantité d'information de plus en plus importante. De plus, les systèmes d'information sont de plus en plus complets et le parc logiciel devient de plus en plus dense. Or, il est parfois capital de réussir à croiser des données hétérogènes et multi-sources, de les réconcilier et de les stocker dans un entrepôt de données centralisé. C'est le rôle des ETL (Extract Transform Load) et des bases de données décisionnelles. Pour assurer un pilotage complet de l'entreprise, ces données doivent être analysées avec soin. L'informatique décisionnelle étudie des données provenant de différentes sources pour en restituer un résultat clair et concis. Cette formation permet aux participants de découvrir la mise en place d'un système décisionnel avec toutes ses étapes.

Objectifs

- Comprendre les modèles de bases de données
- Comprendre l'architecture SGBD (Système de Gestion de Base de Données) et ETL
- Faire un tour d'horizon des SGBD
- Explorer le marché des ETL
- Apprendre à utiliser un ETL avec un exemple pratique

Programme détaillé

INTRODUCTION AUX BASES DE DONNEES

Qu'est-ce qu'une base de données ?

Les éditeurs de bases de données du marché (Oracle, Microsoft...)

Les différents types de bases de données (en ligne, colonnaire, vectorielle....)

LES CONCEPTS DE BASE

ETL compilant en SQL

ETL compilant en C / JAVA

Types de données

Orchestration / Gouvernance

LES JOINTURES

Les jointures internes

Les jointures externes

Avec et sans rejets

LA NORMALISATION

1ère Forme Normale

2ème Forme Normale

3ème Forme Normale

Forme normale de BOYCE-CODD

Normalisation / Dé-normalisation : quel impact sur les performances, les volumétries ?

LES CONCEPTS AVANCÉS

Le suivi de l'exploitation

Savoir choisir entre script et ETL

Choisir son ETL

Optimisation : ELT / Bulk

Historisation : SCD

Bonnes pratiques de développement (Design, nommage, découpage technique / fonctionnel)

INTRODUCTION AUX ETL

Qu'est-ce qu'un ETL ?

Comparatif avec un script SQL ?

L'architecture fonctionnelle

Les différents ETL du marché

Les liens avec les SGBD

ALGORITHMIE

La décomposition en éléments simples

Les boucles

La gestion des flux de données

Orchestration

Normalisation / Dé-normalisation

Routines

LA QUALITÉ DES DONNÉES

La gestion des rejets

Les contraintes d'intégrité des données

Les formats des données

LES MODALITÉS DE CHARGEMENT

Insert / Update

Update / Insert

Delete / Insert

Truncate / Insert

Orchestration / Ordonnancement

LA GESTION DES ERREURS

Les erreurs techniques

Les erreurs fonctionnelles

Monitoring
