

M164 – Contrôle des données importées dans une DB

F. Mauron

EPAI

10 mai 2024

- 1 Introduction
- 2 Critères de test pour contrôler les données importées
- 3 Commandes DML pour la vérification des données
- 4 Exemples

- 1 Introduction
- 2 Critères de test pour contrôler les données importées
- 3 Commandes DML pour la vérification des données
- 4 Exemples

Assurer l'exhaustivité et l'intégrité des données importées dans une base de données afin de garantir la qualité, la cohérence et la fiabilité des données stockées, ainsi que la prise de décision précise et fiable basée sur ces données.

Pourquoi est-ce important ?

- **Fiabilité des données** : Des données exhaustives et intégrées assurent que les informations stockées dans la base de données sont fiables et précises.
- **Précision des analyses** : Des données de haute qualité garantissent des analyses précises et fiables, ce qui permet de prendre des décisions éclairées et stratégiques basées sur les données.
- **Confiance des utilisateurs** : Des données fiables renforcent la confiance des utilisateurs dans le système et dans les informations qu'il fournit, ce qui favorise leur adoption et leur utilisation.
- **Éviter les erreurs et les incohérences** : La détection et la correction précoce des erreurs ou des incohérences dans les données évitent les conséquences néfastes sur les opérations commerciales et les prises de décision.
- **Conformité réglementaire** : Des données complètes et précises sont essentielles pour se conformer aux exigences réglementaires et aux normes de gouvernance des données.

Approches pour atteindre l'objectif

- **Contrôles avant importation** : Effectuez des contrôles approfondis sur les données avant leur importation pour détecter et corriger les erreurs éventuelles.
- **Validation des contraintes d'intégrité** : Assurez-vous que les contraintes d'intégrité de la base de données sont respectées, telles que les clés étrangères, les contraintes d'unicité, etc.
- **Vérification des types de données** : Assurez-vous que les données sont du bon type et qu'elles respectent les formats requis.
- **Comparaison avec les données sources** : Comparez les données importées avec les données sources pour garantir leur exactitude et leur cohérence.
- **Contrôles post-importation** : Effectuez des contrôles supplémentaires sur les données après leur importation pour identifier et corriger les erreurs éventuelles.
- **Documentation et journalisation** : Documentez les processus de contrôle des données et enregistrez les résultats des contrôles effectués pour une traçabilité et une audite ultérieures.

- 1 Introduction
- 2 Critères de test pour contrôler les données importées
- 3 Commandes DML pour la vérification des données
- 4 Exemples

Assurez-vous que toutes les colonnes requises sont présentes et remplies dans le jeu de données importé.

Vérifiez s'il y a des données manquantes dans les colonnes obligatoires ou si des lignes entières sont manquantes.

Recherche de lignes manquantes dans l'ensemble des données importées.

Vérifiez les contraintes de clé étrangère pour vous assurer que toutes les références vers d'autres tables sont valides.

Assurez-vous qu'il n'y a pas de doublons dans les colonnes où l'unicité est requise. Vérifiez si les valeurs respectent les contraintes de type de données (par exemple, les dates sont au bon format, les valeurs numériques sont dans les plages attendues, etc.).

Comparez les données importées avec les données sources pour détecter les écarts ou les incohérences.

Vérifiez si les totaux ou les sommes correspondent entre les données importées et les données sources.

- 1 Introduction
- 2 Critères de test pour contrôler les données importées
- 3 Commandes DML pour la vérification des données**
- 4 Exemples

Sélection de Données

Utilisation de la commande SELECT pour récupérer des informations spécifiques.

```
SELECT * FROM Table WHERE condition;
```

Filtrage des Données

Utilisation des clauses WHERE, GROUP BY et HAVING pour filtrer les données selon des critères spécifiques.

```
SELECT * FROM Table WHERE colonne = valeur;
```

Calcul des totaux ou des sommes

Utilisation des fonctions d'agrégation telles que SUM, COUNT, AVG, etc., pour calculer les totaux ou les moyennes sur les données.

```
SELECT SUM(colonne) FROM Table;
```

Élaboration de sommes de contrôle

Création de requêtes pour calculer des sommes de contrôle afin de vérifier l'intégrité des données. Comparaison des sommes de contrôle avec des valeurs fournies dans les données sources.

- 1 Introduction
- 2 Critères de test pour contrôler les données importées
- 3 Commandes DML pour la vérification des données
- 4 Exemples

Contraintes de clé étrangère (FK)

```
-- Vérifiez les références invalides dans la table Emprunts
SELECT *
FROM Emprunts
WHERE ID_Livre NOT IN (SELECT ID FROM Livres);
```

Contraintes d'unicité (UNIQUE)

```
-- Vérifiez s'il y a des doublons dans la colonne Email de la table Utilisateurs
SELECT Email, COUNT(*)
FROM Utilisateurs
GROUP BY Email
HAVING COUNT(*) > 1;
```

Type de données corrects

```
-- Vérifiez les valeurs non numériques dans la colonne Prix de la table Produits
SELECT *
FROM Produits
WHERE ISNUMERIC(Prix) = 0;
```

Format des données

```
-- Vérifiez si les dates ne sont pas au format attendu dans la colonne DateCommande
SELECT *
FROM Commandes
WHERE TRY_CAST(DateCommande AS DATE) IS NULL;
```


Vérification des relations métier

```
-- Vérifiez si les commandes ne sont pas attribuées à des clients inexistantes  
SELECT *  
FROM Commandes  
WHERE ID_Client NOT IN (SELECT ID FROM Clients);
```

Contrôle des totaux

```
-- Vérifiez si le total des ventes correspond au total attendu  
SELECT SUM(Montant) AS TotalVentes  
FROM Ventes;
```

Validation des données

```
-- Comparez les données importées avec les données sources dans la table Clients
SELECT *
FROM Clients AS C
WHERE NOT EXISTS (
    SELECT *
    FROM DonneesSources AS DS
    WHERE DS.ID = C.ID AND DS.Nom = C.Nom AND DS.Prenom = C.Prenom
);
```

Vérification des totaux

```
-- Comparez les totaux des ventes importées avec ceux des données sources
SELECT
    (SELECT SUM(Montant) FROM Ventes) AS TotalVentesImportees,
    (SELECT SUM(Montant) FROM DonneesSources) AS TotalVentesSources;
```