



Extrait du Environnement iSeries

<https://xdocs400.com/spip.php?article387>

Identifier une rupture de numéros de séquences en SQL

- Les articles -



Date de mise en ligne : mardi 3 juillet 2007

Description :

Des enregistrements pour un regroupement donné comporte un numéro de séquence. Comment identifier les enregistrements qui ont une rupture dans ce numéro de séquence ?

Environnement iSeries

**Des enregistrements pour un regroupement donné comporte un numéro de séquence.
Comment identifier les enregistrements qui ont une rupture dans ce numéro de séquence ?
C'est possible grâce à une requête SQL, pas si compliquée que cela.**

Exemple utilisé

Une table TABLEA comporte les zones :

- ANNEE
- MOIS
- CODSOC
- CODMAT
- ZONED
- ZONEE
- ZONEF

La table stocke pour chaque année / mois, les codes sociétés, matricules et infos associées.

Problème posé

En théorie, pour une année, il ne doit pas y avoir de rupture dans le numéro de mois. un couple société / matricule ne peut pas être présent en août 2006 et en octobre 2006 mais pas en septembre 2006.

Le fichier est énorme. Comment rapidement identifier les cas anormaux pour une année donnée ?

Construisons une requête qui pour une clé donnée (ici code société + code matricule) compare : **la valeur maximum du numéro de séquence - la valeur minimum du numéro de séquence + 1** et le nombre d'éléments trouvés pour la clé.

Par exemple, pour un code société AAA et un matricule M01, si l'on a les enregistrements suivants :

- 2006, 03, AAA, M01, ZONE1, ZONE2
- 2006, 04, AAA, M01, ZONE1, ZONE2
- 2006, 05, AAA, M01, ZONE1, ZONE2

On ne détectera pas de rupture car $\max(\text{mois}) = 5$, $\min(\text{mois}) = 3$ et le nombre d'occurrence = 3 donc, $5 - 3 + 1 = 3$.

Ce couple société + matricule n'est donc pas sélectionné.

Par exemple, pour un code société AAA et un matricule M01, si l'on a les enregistrements suivants :

- 2006, 03, AAA, M01, ZONE1, ZONE2
- 2006, 05, AAA, M01, ZONE1, ZONE2

On détectera une rupture car $\max(\text{mois}) = 5$, $\min(\text{mois}) = 3$ et le nombre d'occurrence = 2 donc, $5 - 3 + 1 \neq 2$.

Ce couple société + matricule est donc sélectionné.

Une fois les couples société + matricules sélectionnés, une jointure est faite sur le même fichier, pour récupérer les valeurs d'informations associées.

La requête SQL pour une année donnée

```
WITH TMP AS (  
  SELECT  
    CODSOC , CODMAT, COUNT(*) , MIN(MOIS), MAX(MOIS)  
  FROM  
    TABLEA  
  WHERE ANNEE = 2006  
  GROUP BY CODSOC, CODMAT  
  HAVING MAX(MOIS) - MIN(MOIS) + 1 <> COUNT(*)  
)  
SELECT B.ANNEE, B.MOIS, A.CODSOC, A.CODMAT, B.ZONE1, B.ZONE2  
FROM TMP A  
INNER JOIN TABLEA B  
ON B.CODSOC = A.CODSOC  
AND B.CODMAT = A.CODMAT  
ORDER BY A.CODSOC, A.CODMAT, B.MOIS
```

La requête SQL sur tout le fichier, quelque soit l'année

Si on veut lancer cette requête sur l'ensemble du fichier, quelque soit l'année, il faut mettre en place une formule de calcul pour établir une séquence entre les mois (passage du mois 12 d'une année au mois 1 de l'année suivante à gérer).

- On va faire cela en établissant l'année minimum existant dans le fichier. Considérons 1995 dans notre cas.

Cela nous donne la requête suivante pour avoir la liste des codes sociétés + matricules ayant une rupture :

```
SELECT  
  CODSOC , CODMAT, COUNT(*) AS NOMBRE, MAX((ANNEE-1995)*12+MOIS)-MIN((ANNEE-1995)*12+MOIS)+1 AS AMPLITUDE  
FROM  
  TABLEA  
GROUP BY CODSOC, CODMAT  
HAVING MAX((ANNEE-1995)*12+MOIS)- MIN((ANNEE-1995)*12+MOIS)+1 <> COUNT(*)
```

Identifier une rupture de numéros de séquences en SQL

NOMBRE est le nombre de fois ou le matricule est trouvé. **AMPLITUDE** est le nombre de mois ou le matricule devrait être trouvé (différence entre l'année/mois le plus grand et le plus petit).

S'il n'y a pas rupture pour un matricule, $NOMBRE = AMPLITUDE$.