

L'utilisation d'hypercubes

Le passage aux comptes

Les programmes du passage aux comptes, c'est-à-dire les programmes qui permettent de passer des données du SIE aux estimations de la comptabilité nationale, font largement appel aux hypercubes. Ceux-ci peuvent être traités par des logiciels spécialisés mais ils peuvent également être traités dans le cadre de bases de données relationnelles par le langage SQL.

La logique du traitement du passage aux comptes sera présentée dans le cadre d'un exemple simplifié.

Supposons que l'économie soit constituée de 4 sous-secteurs d'activité AEGA01, AEGA02, AEGB01, AEGB02 et que nous cherchions à calculer P1, P2 et B1 à partir des variables suivantes :

R212 Achats matières premières
POURBOIR Pourboires
RC302 Production vendue
TRANSPRO Transport sur production

Les formules de calcul étant les suivantes :

$P1 = RC302 + POURBOIR - TRANSPRO$
 $P2 = R212 - TRANSPRO$
 $B1 = P1 - P2$

La première étape consiste à traduire ces égalités dans le cadre d'une matrice, c'est-à-dire d'un hypercube à deux dimensions, de la manière suivante :

		PCR	PCEA	PCEA
		P1	P2	B1
Achats matières premières	R212		1	-1
Pourboires	POURBOIR	1		1
Production vendue	RC302	1		1
Transport sur production	TRANSPRO	-1	-1	0

Les deux dimensions de ce tableau sont :

- l'opération de comptabilité nationale que nous désignerons par le code OP
- les éléments de calcul que nous appellerons variables exogènes et que nous considérerons appartenir à une dimension EX

Ce tableau possède même une troisième dimension : la dimension PC qui permet de déterminer si l'opération de comptabilité nationale correspond à une ressource ou à un emploi.

Le grand intérêt de ce tableau est de permettre une double lecture :

- En colonne, il montre comment est calculée une opération ;
- En ligne, il montre quel est l'impact d'une variable exogène sur les différentes opérations et en particulier les soldes tels que B1.

Ce tableau peut être représenté par un hypercube à trois dimensions. Dans une base de données relationnelle un hypercube correspond à une table, c'est-à-dire à un tableau où les colonnes correspondent aux variables et les lignes aux enregistrements, structurée de la manière suivante :

- Un hypercube ne comporte qu'une seule variable numérique, les autres sont des variables alphanumériques correspondant à des critères, c'est-à-dire aux dimensions du cube.

Ainsi, dans notre exemple, notre table de passage se présentera de la manière suivante :

OP	PC	EX	VAL
P2	PCEA	R212	1
B1	PCEA	R212	-1
P1	PCRP	POURBOIR	1
B1	PCEA	POURBOIR	1
P1	PCRP	RC302	1
B1	PCEA	RC302	1
P1	PCRP	TRANSPRO	-1
P2	PCEA	TRANSPRO	-1

Il ne reste plus qu'à présenter les variables exogènes sous la forme d'un hypercube. Celui-ci possèdera les dimensions suivantes :

- DA : l'année
- SI : le secteur institutionnel
- AE : le sous-secteur d'activité
- EX : la variable exogène

Supposons que les données exogènes se présentent sous la forme d'un hypercube à quatre dimensions, l'année (DA), le secteur institutionnel (SI), l'activité économique (AE) et le type de variable exogène (EX). Par exemple, cet hypercube se présente de la manière suivante :

DA	SI	AE	EX	VAL
DA2000	SIS11	AEGA01	RC302	200
DA2000	SIS11	AEGA02	RC302	300
DA2000	SIS11	AEGB01	RC302	400
DA2000	SIS11	AEGB02	RC302	500
DA2000	SIS11	AEGA01	R212	100
DA2000	SIS11	AEGA02	R212	150
DA2000	SIS11	AEGB01	R212	200
DA2000	SIS11	AEGB02	R212	250
DA2000	SIS11	AEGB01	POURBOIR	10
DA2000	SIS11	AEGB02	POURBOIR	20
DA2000	SIS11	AEGA01	TRANSPRO	10
DA2000	SIS11	AEGA02	TRANSPRO	30
DA2000	SIS11	AEGB01	TRANSPRO	20
DA2000	SIS11	AEGB02	TRANSPRO	30
DA2000	SIS14AA	AEGA01	RC302	100
DA2000	SIS14AA	AEGA02	RC302	150
DA2000	SIS14AA	AEGB01	RC302	200
DA2000	SIS14AA	AEGB02	RC302	250
DA2000	SIS14AA	AEGA01	R212	50
DA2000	SIS14AA	AEGA02	R212	75
DA2000	SIS14AA	AEGB01	R212	100
DA2000	SIS14AA	AEGB02	R212	125
DA2000	SIS14AA	AEGA01	POURBOIR	0
DA2000	SIS14AA	AEGA02	POURBOIR	0
DA2000	SIS14AA	AEGB01	POURBOIR	5
DA2000	SIS14AA	AEGB02	POURBOIR	10
DA2000	SIS14AA	AEGA01	TRANSPRO	5
DA2000	SIS14AA	AEGA02	TRANSPRO	15
DA2000	SIS14AA	AEGB01	TRANSPRO	10
DA2000	SIS14AA	AEGB02	TRANSPRO	15

Le programme de passage aux comptes est alors extrêmement simple : il consiste à réaliser à l'aide d'une requête SQL une jointure multipliant les deux hypercubes l'un par l'autre.

Si nous appelons TablePac l'hypercube correspondant à la table de passage et Exo l'hypercube des variables exogènes, la requête s'écrit de la manière suivante :

```
SELECT e.DA, e.SI, e.AE, e.EX, t.OP, t.PC, e.VAL*t.VAL as VAL
FROM TablePac t, Exo e
WHERE t.EX=e.EX
```

Les résultats de cette requête peuvent eux-mêmes être stockés dans une table qui comprend tous les résultats et qui montre comment ces résultats ont été obtenus. Par exemple, cette table peut être lue par un tableau croisé dynamique d'Excel qui présente l'avantage de pouvoir présenter facilement différents tableaux à partir de la même table de données. Ainsi, nous pouvons faire apparaître un premier tableau croisant les opérations, les secteurs institutionnels et les activités économiques :

OP	SI	A01	A02	B01	B02	Total
B1	SIS11	100	150	210	270	730
	SIS14AA	50	75	105	135	365
Somme B1		150	225	315	405	1 095
P1	SIS11	190	270	390	490	1 340
	SIS14AA	95	135	195	245	670
Somme P1		285	405	585	735	2 010
P2	SIS11	90	120	180	220	610
	SIS14AA	45	60	90	110	305
Somme P2		135	180	270	330	915

Mais il peut également être intéressant de montrer comment sont calculées les opérations à partir des variables exogènes. Ainsi le tableau suivant met en évidence le calcul de la production :

EX	A01	A02	B01	B02	Total
POURBOIR	0	0	15	30	45
RC302	300	450	600	750	2 100
TRANSPRO	-15	-45	-30	-45	-135
Total	285	405	585	735	2 010

Les produits matriciels

Il est possible de réaliser des produits de matrices de très grandes dimensions avec des requêtes SQL. En effet, faire un produit matriciel revient à faire des jointures par des requêtes SQL comme nous pouvons le montrer à partir d'un exemple simple.

Supposons donc une économie constituée de deux secteurs et de trois branches, nous disposons d'une matrice de passage secteurs-branches que nous nommons "Structure" et d'une matrice de production par secteurs "Secteurs". La matrice de passage secteurs-branches montre la répartition de la production de chaque secteur par activité économique et elle permet de calculer la production par branche lorsque l'on connaît la production par secteur. Ainsi, la matrice de production par branche s'obtient en faisant le produit de la matrice Structure par la matrice Secteurs. Par exemple :

Structure		
	S1	S2
B1	0,5	0,3
B2	0,2	0,6
B3	0,3	0,1

Secteurs	
	Prod
S1	100
S2	200

Branches	
	Prod
B1	110
B2	140
B3	50

Que fait le produit matriciel ?

- Il crée une nouvelle matrice ayant autant de lignes que la première matrice et de colonnes que la seconde. Dans notre exemple, il crée une matrice de trois lignes correspondant aux trois branches et une colonne correspondant à la valeur de la production.
- Pour calculer un élément se situant dans la matrice résultat à la ligne i et à la colonne j , il associe tous les éléments de la ligne i de la première matrice à tous les éléments de la colonne j de la deuxième matrice. Pour que cela soit possible, il est nécessaire que le nombre de colonnes de la première matrice soit égal au nombre de lignes de la deuxième. Dans notre exemple, cette condition est vérifiée parce que les lignes de la première matrice et les colonnes de la seconde correspondent toutes deux aux modalités du même critère, le secteur.
- L'association est faite dans l'ordre, c'est-à-dire qu'à chaque secteur de la première matrice est associé le même secteur de la deuxième matrice.
- Le calcul consiste à multiplier les éléments correspondants et à faire la somme des résultats.

Il est facile de reproduire ce processus avec une requête SQL, à condition, toutefois, d'avoir structuré les données en hypercubes. Ainsi, nous aurions pu saisir la même information en créant dans une base de données relationnelle une table *Structure* et une table *Secteurs* de la manière suivante :

Structure		
Secteur	Branche	Taux
S1	B1	0,5
S2	B1	0,3
S1	B2	0,2
S2	B2	0,6
S1	B3	0,3
S2	B3	0,1

Secteurs	
Secteur	Prod
S1	100
S2	200

On peut réaliser une jointure des deux tables par une requête SQL qui va multiplier les valeurs de la première table par celles de la deuxième, sous la condition que l'association ne portera que sur des données correspondant au même secteur. Cette requête peut s'écrire ainsi :

```

SELECT Structure.Secteur, Branche, Taux*Prod as Valeur
FROM Structure, Secteurs
WHERE Structure.Secteur=Secteurs.Secteur
;

```

Cette requête fait le produit cartésien des deux ensembles que constituent les deux tables en ne retenant que les combinaisons où le secteur est commun. On obtient alors la table *Branches* suivant :

Branches

Secteur	Branche	Valeur
S1	B1	50
S2	B1	60
S1	B2	20
S2	B2	120
S1	B3	30
S2	B3	20

Cette table peut être lue directement par un tableau croisé dynamique qui donnera le résultat en ne sélectionnant que le critère *Branche*. On peut également générer directement la table résultat par la requête SQL suivante :

```

SELECT Branche, SUM(Taux*Prod) as Valeur
FROM Structure, Secteurs
WHERE Structure.Secteur=Secteurs.Secteur
GROUP BY Branche
;

```

Ce qui donne :

Branche	Valeur
B1	110
B2	140
B3	50

On peut ainsi faire des produits de matrices de grande taille, par exemple de 700 x 700.

Notons, pour finir que les requêtes SQL sont plus puissantes que les produits matriciels et qu'elles en constituent en quelque sorte une extension. Il est, en effet, possible de travailler avec plus de deux critères. On peut ainsi ajouter des critères de taille et de catégorie juridique et obtenir par la même requête ce qui nécessiterait le recours à une multitude de produits matriciels.

Ce texte n'engage que son auteur : Francis Malherbe