

Une compagnie d'assurance de biens (automobile, immobilier, responsabilité civile) possède une application transactionnelle de production permet de gérer les polices (contrats) de ses clients ainsi que les sinistres (accidents) déclarés par ces clients.

1. Gestion des polices

Pour gérer les polices, les employés ou agents d'assurance peuvent effectuer les transactions suivantes :

- ◆ Créer, mettre à jour ou supprimer une police d'assurance
- ◆ Créer, mettre à jour ou supprimer un risque (pour une police donnée)
- ◆ Créer, mettre à jour ou supprimer des biens assurés (voiture, maison) sur un risque
- ◆ Chiffrer ou refuser le risque
- ◆ Valider ou refuser la police

On enregistre dans ces transactions un grand nombre d'informations, et notamment : date d'écriture (date de la transaction), date d'effet (date de début d'assurance), client (personne(s) privée(s), personne morale), opérateur (employé, agent : chiffrage, vérificateur : validation), risque (produit vendu par la compagnie d'assurance), couverture (description des biens assurés), police (numéro de police, « note » de la police ou du risque,...) , transaction (code transaction).

2. Gestion des sinistres

Pour gérer les sinistres déclarés par les clients, les employés ou agents d'assurance ont à leur disposition les transactions suivantes :

- ◆ Créer, mettre à jour ou supprimer une déclaration de sinistre
- ◆ Créer, mettre à jour ou supprimer une expertise
- ◆ Créer, mettre à jour ou supprimer des paiements
- ◆ Clore le sinistre

Ces transactions comportent notamment : date d'écriture (date de la transaction), date d'effet (date de déclaration), client, opérateur, risque, biens sinistrés, police, les tiers impliqués dans le sinistre, les montants financiers (limites, déjà payé, reste à payer, ...), code transaction.

3. Taille des bases de données

- ◆ Nombre de polices : 2 millions
- ◆ Moyenne de biens couverts par police : 10
- ◆ Nombre de transactions par an et par police : 12
- ◆ Nombre d'années : 3
- ◆ Taille d'une variable (clé ou indicateur) de table de faits : 8 octets
- ◆ Pourcentage de biens assurés donnant lieu à un sinistre par an : 5%
- ◆ Temps d'ouverture d'un sinistre : 1 an

Questions

A partir de cette application transactionnelle, on veut créer un entrepôt de données permettant de répondre aux questions suivantes :

- on ne s'intéresse qu'à la globalisation par mois des transactions.
- pour chaque bien assuré, on veut connaître le montant de la prime (somme annuelle payée par le client pour assurer le bien) associée au bien assuré, et le nombre de transactions du mois pour ce bien.
- On veut aussi l'« état » de la police pour en spécifier les phases particulières : police nouvellement créée, nouvellement modifiée, sinistre en cours, sinistre juste clos.
- On veut naturellement sortir des tableaux par client, agent ou employé, date d'effet, état, avec toutes les sommations possibles y compris par police et risque.
- De même on veut pouvoir sortir des tableaux de bord par sinistre avec le total payé dans le mois et le total reçu dans le mois pour ce sinistre.

Les tableaux de bord « sinistre » doivent pouvoir être édités par client, agent ou employé, date d'effet, état, avec toutes les sommations possibles y compris par police et risque. On veut pouvoir établir des tableaux de bord par client et bien assuré de l'activité sur le dossier (nombre de transactions, nombre de sinistres), du chiffre d'affaire, du taux de sinistres et du rendement (ratio versements/prime), et tous les totaux et sous totaux correspondants.

- On veut également déterminer la taille sur disque de l'ED.

On suivra la démarche suivante :

1. commencer par tracer quelques tableaux de bord à titre d'exemple de ce que peut éditer l'ED : quelques (de l'ordre de 5) tableaux à deux dimensions pour les polices et quelques uns pour les sinistres (toujours à deux dimensions). Tracer au moins un cube à trois dimensions.
2. faire le schéma en étoile d'un magasin de données « police » ne prenant pas en compte les sinistres. Tracer au moins un cube à trois dimensions.
3. de même, faire le schéma en étoile d'un magasin « sinistre »
4. faire un seul ED de ces deux magasins. Y a-t-il des dimensions conformes ? Quels tableaux de bord nouveaux peut-on alors éditer ?

I Première table de faits : deux dimensions et un indicateur

Commençons par créer une table de faits très simple, avec deux dimensions et un indicateur.

- 1) première dimension : le mois de « production » concaténé à l'année. On appelle mois de production le mois lors duquel le client a signé son contrat.
Exemple de valeur : « **1204** » pour décembre 2004.
- 2) deuxième dimension : le type de risque (produit). On suppose pour simplifier que la compagnie d'assurance vend seulement trois produits :
 - ◆ l'assurance automobile (type de produit **A**)
 - ◆ l'assurance habitation (type **H**)
 - ◆ l'assurance responsabilité civile (type **R**)
- 3) indicateur : le chiffre d'affaires du mois. Il s'agit de la somme des montants des contrats - pour un produit donné – signés dans le mois.

Début janvier 2005, un programme (ETL) charge dans cet entrepôt de données une couche composée de 3 enregistrements seulement. Exemple de la couche chargée début janvier 2005 :

Mois année	Type risque	CA du mois
1204	A	125 000 €
1204	H	28 000 €
1204	R	12 000 €

Figure 1

Pour toute l'année 2004, cette base de faits comporte 36 enregistrements (3 par mois * 12 mois)

Ils permettent déjà, par exemple, d'éditer 4 tableaux :

- évolution au cours de l'année du CA par type de risque (3),
- évolution au cours de l'année du CA tous risques confondus (1),

Exemple de tableau pour Type_risque = R (responsabilité civile) avec 1 seule dimension, le mois et l'indicateur :

Mois	CA
janvier	11000
février	12000
mars	13000
avril	11000
mai	14000
juin	15000
juillet	7000
août	2000
septembre	17000
octobre	9000
novembre	10000
décembre	12000

Figure 2

Exemple de tableau avec les deux dimensions, le mois et le type de risque, et 1 indicateur, le CA.

Mois	CA A	CA H	CA R
janvier	100000	20000	11000
février	80000	10000	12000
mars	120000	18000	13000
avril	110000	15000	11000
mai	70000	17000	14000
juin	130000	8000	15000
juillet	70000	10000	7000
août	30000	3000	2000
septembre	80000	18000	17000
octobre	90000	30000	9000
novembre	110000	25000	10000
décembre	125000	28000	12000

Figure 3

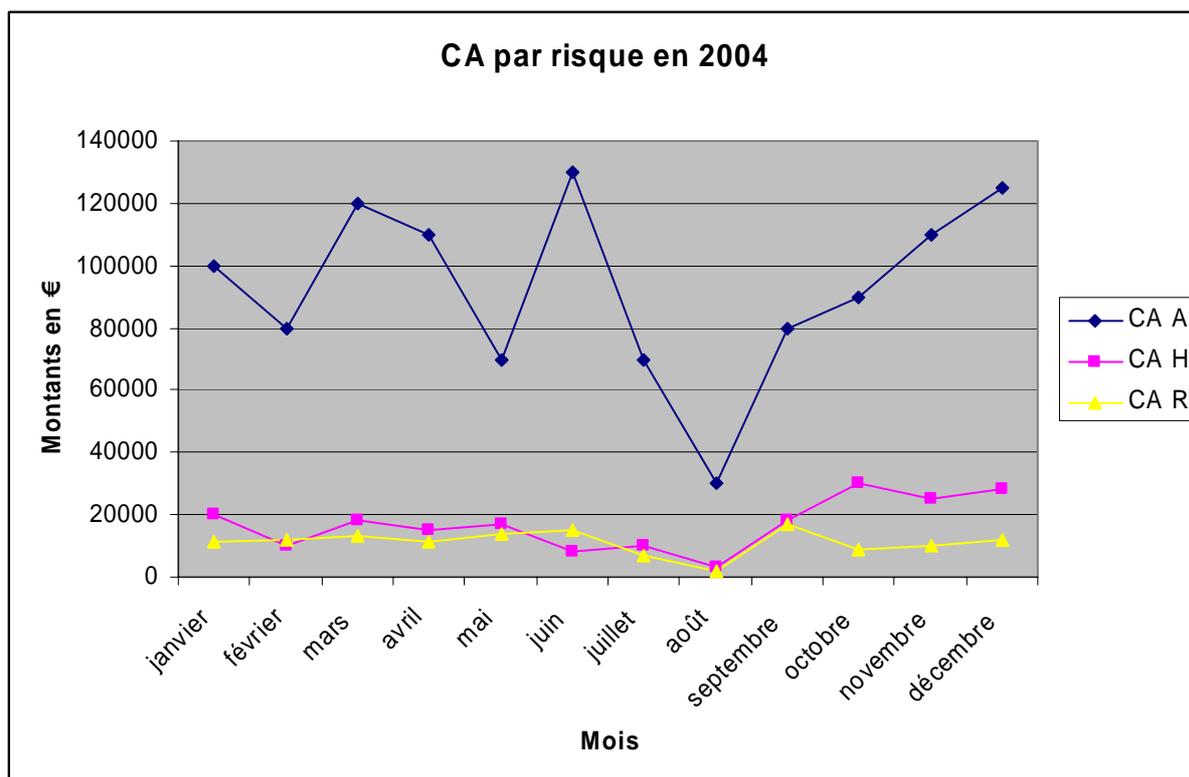


Figure 4

II Ajoutons une dimension, la note du risque.

Lorsqu'il chiffre le risque, l'agent lui donne une note de 1 à 3 en estimant la probabilité de coût pour l'entreprise, à partir d'un certain nombre de critères (classement du client, caractéristiques des biens assurés, etc...).

- 1 probabilité de coût élevé
- 2 probabilité de coût moyen
- 3 probabilité de coût faible

Les couches de novembre et de décembre 2004 deviennent :

Mois année	Type risque	Note	CA du mois
1204	A	1	10 000
1204	A	2	30 000
1204	A	3	5 000
1204	H	1	4 000
1204	H	2	20 000
1204	H	3	4 000
1204	R	1	1 000
1204	R	2	10 500
1204	R	3	500
1104	A	1	8000
1104	A	2	17 000
1104	A	3	2 000
1104	H	1	3 000
1104	H	2	16 000
1104	H	3	1 000
1104	R	1	2 000
1104	R	2	8 000
1104	R	3	200

Figure 5

Et le cube peut être représenté par le « cube » (3,3,2) ci-dessous.

Dans chaque élément du tableau, la valeur de l'indicateur CA.

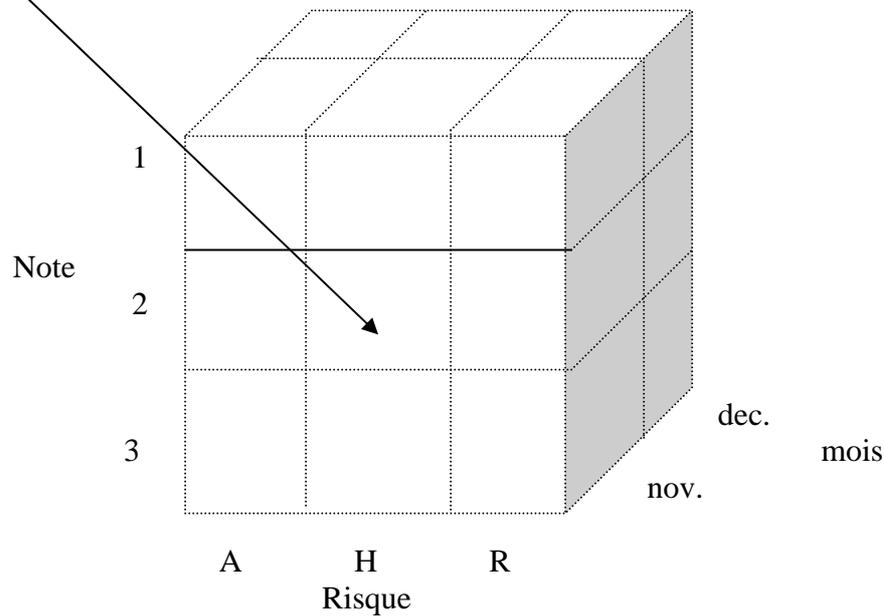


Figure 6

A ce stade, nous avons une base de faits du magasin « police » comportant les 4 variables suivantes :

Mois année	<i>dimension - élément de la clé multiple</i>
Type risque	<i>dimension - élément de la clé multiple</i>
Note	<i>dimension - élément de la clé multiple</i>
CA	<i>indicateur</i>

III Créons un deuxième magasin (datamart) « sinistre » avec pour commencer la table de faits suivante :

Mois année	<i>dimension - élément de la clé multiple</i>
Type risque	<i>dimension - élément de la clé multiple</i>
Paiement	<i>indicateur</i>

Paiement est le montant total des paiements effectués dans le mois considéré pour le risque donné. Sur 1 an, cette table de faits comporte 36 enregistrements (comme « police » page 4).

L'ED réunissant les deux magasins est le double schéma en étoile suivant :

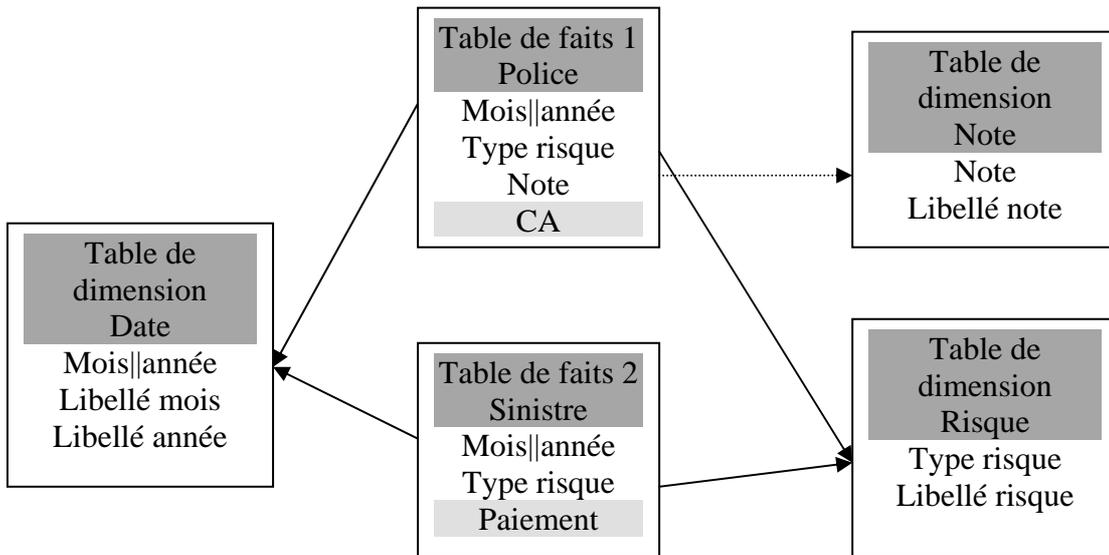


Figure 7

Cherchons par exemple à éditer le tableau ci-dessous avec 1 dimension, 2 indicateurs :

Risque responsabilité civile

Mois	CA	Paiement
janvier	11000	500
février	12000	400
mars	13000	600
avril	11000	800
mai	14000	800
juin	15000	500
juillet	7000	600
août	2000	300
septembre	17000	700
octobre	9000	300
novembre	10000	900
décembre	12000	1000

Figure 8

Le SQL permettant d'éditer le tableau ci-dessus est le suivant :

Select libellé_risque **From** Risque **Where** (risque.type_risque= »R ») ;

Select libellé_mois, CA, paiement **From** Date, Police, Sinistre
Where (risque.type_risque= »R ») **and** (police.type_risque=risque.type_risque)
and (risque.type_risque=sinistre.type_risque)
And (police.mois||année=date.mois||année) **and** (date.mois||année=sinistre.mois||année) ;

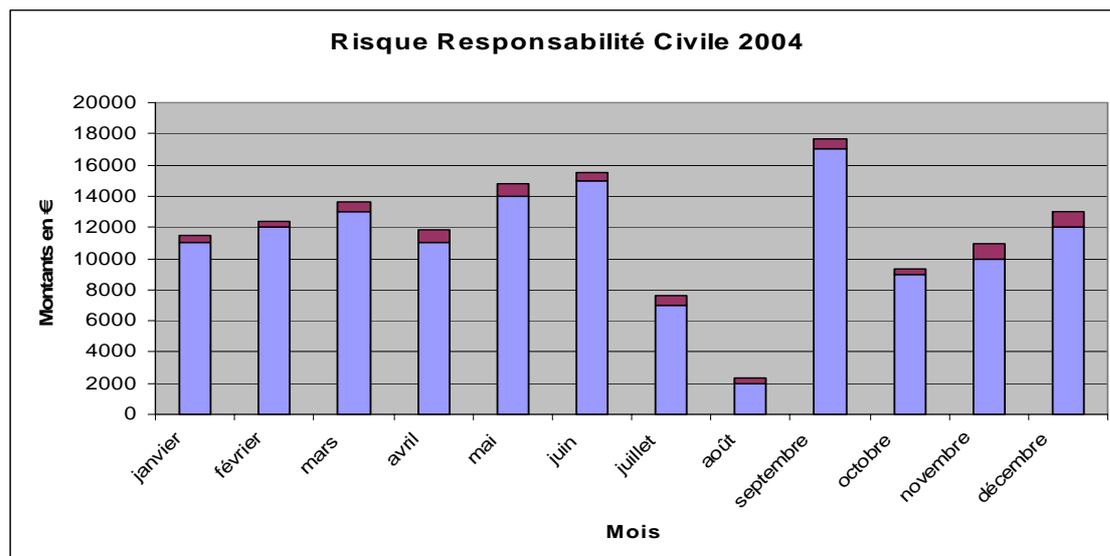


Figure 9

A partir du même ED, on peut aussi éditer le tableau suivant (ici limité à novembre et décembre), en ajoutant la note à partir de la table de faits « police » et en calculant le ratio paiement / CA du mois :

Mois année	Type risque	Note	CA du mois	Paiements	Ratio P/CA
1204	A	1	10 000	1000	0,10
1204	A	2	30000	2500	0,08
1204	A	3	5000	300	0,06
1204	H	1	4000	1000	**0,25**
1204	H	2	20000	1800	0,09
1204	H	3	4000	200	0,05
1204	R	1	1000	180	**0,18**
1204	R	2	10500	800	0,07
1204	R	3	500	20	--0,04--
1104	A	1	8000	700	0,09
1104	A	2	17000	2000	**0,12**
1104	A	3	2000	100	0,05
1104	H	1	3000	1200	**0,40**
1104	H	2	16000	1000	0,06
1104	H	3	1000	30	--0,03--
1104	R	1	2000	400	**0,20**
1104	R	2	8000	1000	**0,12**
1104	R	3	200	10	0,05

Figure 10

Les ratios > 0.10 sont entourés d'**, ceux < 0.05 sont entourés de –

Le tableau de bord des ratios pour décembre est :

Risque / note	1	2	3
A	0.10	0.08	0.06
H	0.25	0.09	0.05
R	0.18	0.07	0.04

Figure 11

Si la valeur 0.15 pour le ratio est le seuil de rentabilité d'un risque, la société d'assurance peut conclure au rejet des risques notés 1.

IV On peut maintenant écrire les tables de faits complètes.

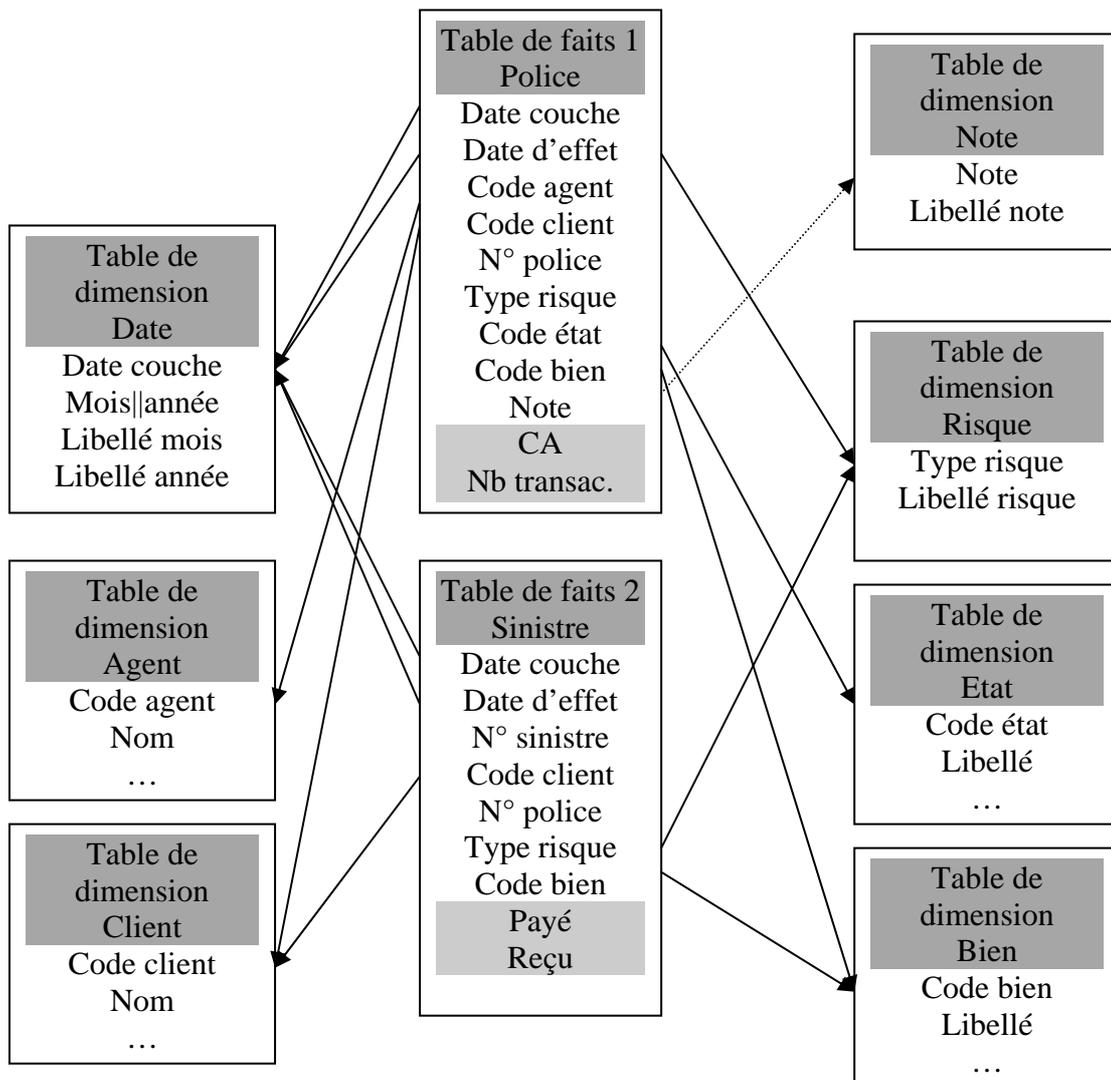


Figure 12

Commentaires :

La clé mois||année est remplacée par « date couche » afin d'utiliser le même format et la même table de dimension que pour la date d'effet, par exemple *jjmmaaaa*.

N° police et N° sinistre sont des dimensions dégénérées qui n'ont pour objet que de faire des regroupements sur la même police ou le même sinistre.

La table de dimension Etat est par exemple :

Code état	Libellé état
pnc	police nouvellement créée	
pnm	police nouvellement modifiée	
sec	sinistre en cours	
sjc	sinistre juste clos	

Figure 13

Quant à la table de dimension « bien », elle pourrait être :

Code bien	Libellé bien
a1	Premier véhicule	
a2	Deuxième véhicule	
.....	
hp	Habitation principale	
hs	Résidence secondaire	
.....	
rs	Responsabilité civile signataire	
rc	Responsabilité civile conjoint	

Figure 14

V Taille disque des tables de faits

S'il y a une seule police par client, et un seul agent par bien couvert, on obtient pour la table de faits « Police » :

Nombre d'enregistrements : 2 000 000 (nb polices) x 10 (biens) x 36 (mois) = 720 millions

Pour chaque enregistrement, 11 champs de 8 octets, soit :

$$720 \text{ millions} \times 11 \times 8 = (\text{environ}) 64 \text{ Giga -octets}$$

Pour la table de faits « Sinistre » :

Nombre d'enregistrements : 720 millions x 5% = 36 millions

Taille totale : 36 millions x 9 (champs) x 8 (octets) = (environ) 2.6 Giga-octets