

TP ETL/Pentaho

Dépenses d'assurance maladie hors prestations hospitalières (données nationales)

Nous souhaitons analyser la base de données des remboursements mensuels effectués par le régime général de l'Assurance Maladie (hors prestations hospitalières) par type de prestations (soins et prestations en espèces), type d'exécutant (médecins par spécialité, chirurgiens-dentistes, auxiliaires médicaux, laboratoires d'analyse, pharmaciens, ...) et par type de prescripteurs. Les dépenses sont indiquées en montants remboursés et présentées au remboursement.

Nous nous sommes procurés ces données publiques sur le lien suivant :

<https://www.data.gouv.fr/fr/datasets/depenses-d-assurance-maladie-hors-prestations-hospitalieres-donnees-nationales/>

Nous avons deux sources de données, ayant des schémas conceptuels différents, sont :

- Source1 : un fichier Excel.
- Source2 : une base de données relationnelle, implantée sous MySQL.

Les deux sources contiennent le même type d'informations, avec des intitulés différents et des données non disjointes (il faut gérer les doublons) et qui se complètent.

La description des champs se trouve en Annexe.

I- Entrepôt de données

L'idée de ce projet est de concevoir, de créer et de peupler un entrepôt de données qui analyse les remboursements de la caisse d'assurance maladie : les taux de remboursement, les montants remboursés, les bases de remboursements par :

- Mois et année de remboursement de soins
- Mois et année des prestations/soins.
- La spécialité du prescripteur du soin : ex le médecin généraliste ou un spécialiste...
- La nature de l'assurance maladie : maladie, accident de travail, maternité...
- Nature de la prestation

Questions :

1. Concevoir le schéma conceptuel de l'entrepôt avec un modèle en étoile ou flocon : définir les dimensions, et la table des faits.
2. Créer l'entrepôt sous Oracle

3. Utiliser Pentaho pour mettre en place les transformations qui définissent les étapes d'extraction, de transformations et de chargement de données.

Le projet est à faire pendant la séance de cours, en trinôme à rendre à la fin de l'heure.

Vous devez rendre :

- 1- Un document word avec :
 - Le schéma de l'entrepôt
 - Le script de création de l'entrepôt sous Oracle.
 - Les captures d'écran des tables de l'entrepôt alimentées par Pentaho depuis les sources.
- 2- Les transformations Pentaho en .ktr qui ont permis d'alimenter les tables dimensions et la table de fait.

variable	libellé	description
act_coe	coefficient global	<p>Une partie des actes de soins est décrite par</p> <ul style="list-style-type: none"> • Une lettre clé qui varie selon les professions et catégories d'actes et qui a une valeur unitaire • Un coefficient qui reflète la complexité de l'acte, et qui, multiplié par la valeur unitaire de la lettre clé, donne le tarif de l'acte. <p>Ceci est notamment le cas des actes techniques médicaux et des actes d'auxiliaires médicaux (infirmier, kiné, ...)</p> <p>Par exemple, un pansement lourd exécuté par un infirmier est côté AMI 4 soit une base de remboursement de 3,15 (valeur conventionnelle AMI)*4=12,6 euros</p>
act_dnb	dénombrement	<p>Une partie des actes peut être dénombrée car codifiée avec un code qui permet de connaître directement la nature exacte et le tarif de l'acte (CCAM).</p> <p>C'est par exemple le cas des consultations,</p>
asu_nat	nature d'assurance	Risque au titre duquel est prise en charge la prestation: Maladie, Maternité, Accident du Travail-Maladie Professionnelle, Invalidité, Décès
cpl_cod	complément d'acte	Cette variable permet de toper les suppléments tarifés, ajouté au montant de l'acte lorsqu'il est réalisé dans un contexte spécifique (majoration de nuit, de dimanche ou jour férié, urgence)
dep_mon	dépassement d'honoraires	On parle de montant de dépassement lorsque les prix pratiqués par les professionnels de santé sont supérieurs aux tarifs de convention fixés pour chaque acte médical par la sécurité sociale. Les dépassements d'honoraires peuvent être pris en charge par les complémentaires santé. Cette notion ne concerne que les honoraires des médecins.
exe_spe	spécialité exécutant	Spécialité exercée par le médecin (généraliste, cardiologue, etc), nature d'activité des auxiliaires médicaux (infirmiers, masseurs-kinésithérapeutes)...
exe_spe1	regroupement spécialité exécutant	Spécialité regroupée par grande catégorie d'exécutant: médecins omnipraticiens, médecins spécialistes,... Cette notion ne concerne que les médecins.
exe_stj1	mode d'exercice exécutant	Statut du professionnel de santé sur sa manière d'exercer: libéral ou salarié
l_asu_nat	libellé assurance	
l_cpl_cod	libellé du complément d'acte	
l_exe_spe	libellé spécialité exécutant	
l_exe_spe1	libellé regroupement spécialité exécutant	
l_exe_stj1	libellé mode d'exercice exécutant	
l_pre_spe	libellé spécialité prescripteur	

—	l_pre_spe1	libelle regroupement spécialité prescripteur	
—	l_pre_stj1	libellé mode d'exercice prescripteur	
—	l_prs_nat	libellé prestation	
—	l_serie	libellé de la série Statistique Mensuelle	
—	pre_spe	spécialité prescripteur	Spécialité exercée par le médecin (généraliste, cardiologue, etc), nature d'activité des auxiliaires médicaux (infirmiers, masseurs-kinésithérapeutes)
—	pre_spe1	regroupement spécialité prescripteur	Spécialité regroupée par grande catégorie de prescripteur: médecins omnipraticiens, médecins spécialistes, chirurgiens dentiste, ...
—	pre_stj1	mode d'exercice prescripteur	Statut du professionnel de santé sur sa manière d'exercer: libéral ou salarié
—	prs_nat	prestation	Chaque prestation est codifiée par un code à 4 chiffres dans le SNIIRAM. Une prestation est un acte réalisé par professionnel de santé (médecin, auxiliaire, ...) ou un produit de santé (médicament, prothèse, ..) pris en charge par l'assurance maladie. Cela peut être également un versement de prestation en espèces (indemnités journalières dans le cas d'un arrêt maladie, rente d'accident du travail, pension d'invalidité)...
—	rec_mon	base de remboursement	La base de remboursement correspond au montant pris en considération par l'Assurance Maladie avant application du taux de remboursement. C'est le tarif conventionnel pour les Professionnels de Santé. Exemple : pour une consultation de médecin généraliste la base de remboursement est de 23€ et le taux de remboursement est de 70% (si vous avez un médecin traitant) sauf pour les personnes en affection longue durée 100%.
—	rem_date	mois et année de remboursement	Correspond au mois et à l'année auxquels l'organisme a remboursé la prestation.
—	rem_mon	montant remboursé	Montant que l'Assurance Maladie rembourse à l'assuré : (base de remboursement * taux de remboursement).

—	rem_tau	taux de remboursement	Le taux de remboursement de la sécurité sociale est le taux sur lequel cet organisme se base pour effectuer ses remboursements. Il s'applique sur le tarif conventionnel et varie en fonction de l'acte médical et du régime obligatoire de l'assuré social, et de la situation de l'assuré.
—	serie	série de la Statistique Mensuelle	Une série de la Statistique Mensuelle est un poste de dépenses obtenu par croisement de la prestation et de la spécialité exécutant. Pour ce qui concerne le médicament, un croisement est également opéré avec le taux de remboursement
—	sns_date	mois et année des soins	Mois et année auxquels le prestataire de soins a exécuté un acte ou délivré un produit: •C (consultation), V (visite à domicile) pour un médecin; •Médicaments pour un pharmacien; •etc. Pour les prestations en espèces : mois et année de début de la période de versement.