

TP3 Entrepôts de Données

Master 1 MIAGE

Flux de données et Entrepôt en constellation

Le Système d'information Opérationnel

On dispose d'une base de données transactionnelle (SIO) servant à la gestion des stocks de médicaments dans un hôpital. Cette gestion des stocks est modélisée sous la forme de l'ensemble des déplacements de médicaments qui existent entre les services.

Pour obtenir un médicament, une personne fait la demande d'une certaine quantité de ce médicament, et on enregistre la date de la demande et le lieu où ce médicament est attendu. On ne mémorise pas le nom de la personne qui a fait la demande mais la localisation permet de retrouver quel est le service demandeur, ce qui est suffisant pour ce système d'information. Soit la pharmacie dispose de cette quantité et la livre directement à la localisation voulue (ou éventuellement une quantité plus grande), soit elle en livre une quantité inférieure. A défaut de médicament dans la pharmacie, il est aussi possible que ce médicament soit stocké dans un autre lieu « source » et soit livré de ce lieu source vers le lieu voulu. Dans ce système d'information, la consommation d'un médicament par un patient est représentée comme le déplacement du médicament vers la localisation « Patient ». Il en est de même pour le recyclage des médicaments usagés (déplacement vers la localisation « recyclage »).

Le MCD de la base de données se trouve à la fin du sujet, et le fichier « donneesSIO.sql » contient le script de création des tables et d'insertion de quelques données.

Modélisation dimensionnelle

On s'intéresse d'une part à la consommation des médicaments, et d'autre part aux ruptures de stock, c'est à dire aux demandes non satisfaites dans un délai raisonnable.

- **Consommation des médicaments** : on veut pouvoir faire des requêtes sur l'entrepôt pour étudier la consommation (poids, volume, quantité) des médicaments par date (ou mois ou trimestre, ...), par service ou pôle, par médicament, par catégorie ou par dénomination pharmacologique (DCI)¹. Il est aussi très important de savoir, dans cette consommation, dans quelle proportions des médicaments n'ont pas été réellement utilisés par les patients mais sont partis au recyclage (en général parce que la date de péremption a été dépassée mais on ne dispose pas suffisamment d'informations pour connaître la cause exacte).
- **Rupture de stock** : on veut analyser les demandes non satisfaites afin de pouvoir améliorer la gestion des stocks et résoudre le plus possible ces problèmes de rupture de stock. On considérera qu'une demande n'est pas satisfaite si le médicament n'est pas arrivé au service

¹ La DCI, ou dénomination commune internationale, n'est pas connue dans le SIO, mais on dispose de fichiers de données permettant de trouver la DCI associée à chaque médicament.

au plus tard le lendemain de la demande. Comme pour la consommation des médicaments, on s'intéressera en particulier à une analyse par date, par service ou pôle, par médicament, catégorie ou dénomination pharmacologique (DCI) ou laboratoire.

Questions de modélisation :

1. Proposer un modèle dimensionnel (schéma en étoile avec une table de faits pour la consommation et une table de faits pour les ruptures de stock et des tables de dimensions) capable de répondre aux besoins des utilisateurs. Représenter, sur un diagramme conceptuel, les données de l'entrepôt sous la forme du schéma en constellation proposé. Enoncer les hiérarchies pertinentes pour chaque dimension.
2. Créer en SQL les tables correspondant aux faits et dimensions modélisés au point précédent.

Ces tables vont être alimentées grâce à l'ETL Talend.

Question concernant l'alimentation des tables :

Définir des jobs Talend permettant d'alimenter les tables de dimensions, la table de faits « consommation des médicaments » et la table de faits « rupture de stock ».

Concernant la DCI, on dispose des fichiers « CIS_bdpm.txt » et « CIS_COMPO_bdpm.txt » fournis par la base de données publique des médicaments.

Question d'exploitation des données :

Posez les requêtes SQL permettant de répondre aux questions suivantes :

1. Pour chaque médicament, quelle est la quantité totale consommée par les patients et celle mise à la poubelle.
2. Quelles sont les 3 services où il y a le plus de gaspillage (en quantité).
3. Combien y-a-t-il eu de ruptures de stock, par laboratoire. Ici, on peut s'intéresser à la quantité de médicaments qui n'ont pas été fournis ainsi qu'au nombre de demandes insatisfaites (ou partiellement satisfaites).
4. La quantité moyenne de Paracétamol consommée par jour, par service, les deux.
5. Quel est le jour du mois d'octobre où il y a eu le plus de demandes non satisfaites.
6. Quelle est la répartition par tiers des catégories de médicament selon leur consommation.

Utilisez Talend pour extraire au format CSV toutes les consommations du pôle « Enfant » pour le mois d'octobre 2015.



