

Discipline: 2^{ème} Ing GL

Matière : Entrepôt de données (ED)

Enseignants : D.Kanzari

TD 3 : Modélisation Multidimensionnelle & MDX

Exercice :

Soit R le schéma relationnel suivant permettant de stocker les informations relatives à la projection des films cinématographiques :

Film (IdF, titre, année, durée, genreFilm, NumRéalisateur);

Salle_Projection (IdS, Nom, capacité, IdCinema)

Projection (IdF, IdS, IdD, horaire, prix, nbr_projection)

Questions:

A partir du schéma R on souhaite modéliser un schéma multidimensionnel S orienté métier qui permet de dégager le nombre de films projetés par salle de projection, par genre de film, par date, (.., etc.), ainsi que toutes les opérations agrégatives sur les chiffres d'affaires des projections.

- 1) Proposer un modèle en étoile (hiérarchisé) correspondant à S.
- 2) Proposer un schéma en flocon de S qui tient compte des informations relatives au réalisateur du film (nom, prénom, âge, adresse) et au cinéma de projection (nom, adresse, ancienneté).
- 3) A partir cube C, indiquez les opérations OLAP (RollUp, DrillDown, Slice, Dice) ainsi que des opérations agrégatives (sum, avg, count) qu'il faut appliquer afin d'obtenir les informations suivantes :
 - Le nombre de projections par salle de projection mensuellement.
 - Le nombre de films projetés par jour et par genre.
 - Le nombre de films projetés par genre en 2016.
 - Le chiffre d'affaire des films par salle de projection en Decembre 2017.
- 4) Ecrire les requêtes MDX qui retournent les informations précédentes.

Il chiffe le moyen par réalisateur entre 2016 et 2017.
Le nombre de films projetés par genre de film et nom de cinéma.
et annuellement.
et genre de film.

Travaux Dirigés 2: Modélisation d'un ED

Exercice 1 :

Soit la base de données (opérationnelle) suivante :

STATION (NumSta, NomSta, Altitude, Région)

HOTEL (NumHot, NomHot, NumStat, Catégorie)

CHAMBRE (NumHot, NumCh, NbLits)

RESERVATION (NumCli, NumHot, NumCh, DateDeb, DateFin, NbPers)

CLIENT (NumCli, NomCli, AdrCli, TelCli)

Questions

1. Dessiner le schéma conceptuel en étoile
2. Proposer un schéma en flocon

q- Bat o Tm hystat ?? // flocon. Belu qm Sen de;

Exercice 2 :

Une banque distribue une carte de paiement « carte de crédit » à ses clients. Elle décide de réaliser un Datawarehouse (DW) afin de faire le suivi des paiements effectués avec la carte tels que:

- a. Voyages en avion,
- b. Locations de voiture,
- c. Hôtellerie.

Elle veut faire un suivi indépendant de chacun des paiements a, b ou c, mais aussi avoir la possibilité d'un suivi global.

- a. A chaque déplacement en avion, la compagnie aérienne lui envoie un fichier contenant les éléments suivants:

- identification de la carte de paiement,
- coordonnées du client
- coordonnées de la compagnie aérienne;
- ville de départ
- ville d'arrivée,
- n° du vol,

- date du vol,
- n° du billet,
- classe du siège,
- distance parcourue,
- date d'achat
- et prix payé.

b. Les loueurs de véhicule transmettent après chaque location:

- identification de la carte de paiement,
- coordonnées du client,
- coordonnées de la société de location de véhicules,
- catégorie du véhicule,
- date de début de location,
- date de fin de location,
- nombre de jours,
- distance parcourue,
- date de réservation
- et prix payé.

c. L'hôtel transmet à chaque séjour:

- identification de la carte de paiement,
- coordonnées du client et de l'hôtel,
- catégorie de chambre,
- date de début de séjour,
- date de fin de séjour,
- nombre de nuitées,
- date de réservation,
- prix de l'hébergement
- et prix de la restauration.

1. Un premier DW ne concerne que les déplacements en avion.

- Etablir les dimensions et les indicateurs
- Proposez une hiérarchie pour ces dimensions.
- Etablir le modèle dimensionnel du DW.

instance: 91
Ce DW doit permettre de répondre aux questions suivantes : quel est le chiffre d'affaires (CA) par client, par date de voyage (et par mois, trimestre et année), par compagnie aérienne, par ville de destination?

2. De même, établir deux autres modèles dimensionnels, l'un pour les locations de voiture, l'autre pour l'hôtellerie.

Dans le cas de la location de voiture, on souhaite éditer le CA, le nombre de jours de location, et le kilométrage pour chaque: client, date de réservation, ville, loueur, et catégorie de véhicule.

Dans le cas de l'hôtellerie, on veut des tableaux de bord par client, hôtel, ville, date de début de séjour, catégorie de chambre, faisant apparaître le nombre de nuitées, le prix total payé.

3. On envisage regrouper ces trois DW en un seul central, afin de répondre aux questions supplémentaires suivantes :

- Quel est le CA total induit par un déplacement en avion ?
- Quelle est la durée du séjour ?
- Quel est le CA en location de voiture ? En hôtellerie ?

On désire pouvoir éditer les détails de CA par période de temps et par client, ville de destination, ville de location (si différente), ville d'hébergement (si différente), compagnie aérienne, loueur et hôtelier, et faire tous les regroupements utiles.

Proposer un modèle dimensionnel du DW central.

Exercice 3 :

On souhaite créer un entrepôt de données pour un opérateur de téléphonie dont le système opérationnel est décrit comme suit :

Le système a une base de données pour la gestion de ses clients. Il propose plusieurs types d'abonnements à ses clients. Les clients peuvent obtenir plusieurs numéros de téléphone auprès de leur opérateur. Pour chacun de ces numéros, l'opérateur enregistrera les informations nécessaires pour établir une facture mensuelle détaillée. On enregistre le nom, prénom, l'adresse, le type d'abonnement ainsi que les numéros de téléphone du client (on fera une facture par numéro). Le système sauvegarde aussi les prix des communications définies par le numéro appelé, la date et l'heure de l'appel ainsi que la durée de l'appel. Le prix d'un appel téléphonique varie selon le type d'abonnement, l'heure de la communication et la géo localisation du numéro appelé (on mentionne pour chaque appel le tarif appliqué).

Le service marketing souhaite étudier l'évolution de l'utilisation des différents types d'abonnements par les clients en fonction du temps, de l'âge, du genre et de la localisation géographique.

Le service vente veut étudier les différents facteurs influençant le chiffre d'affaire des ventes des abonnements (tels que la localisation géographique, le genre de l'abonné, l'âge de l'abonné, etc.).

Le service technique envisage analyser les réclamations des abonnées (indisponibilité des services, interférences des appels, etc.).

Exercice 4 :

On envisage réaliser un entrepôt de données afin d'analyser l'évolution des salaires des employés selon l'âge et le niveau d'étude et la situation géographique des établissements (privés ou étatiques) d'embauche.

L'analyse selon l'âge peut se faire par année ou par décade (tranches de 10 années à partir de 18 ans et jusqu'à 70ans).

L'analyse du niveau d'étude peut se faire par le niveau d'enseignement atteint en fin d'études (primaire, secondaire, supérieur) ou par le dernier diplôme obtenu (certificat de fin d'étude primaire, Bac, Licence, Ingénieur, Master).

L'analyse de la situation géographique peut se faire par ville, département, région, état ou pays.

Matière :Entrepôt de données	Date : 11/06/2021
Code matière: 7410.....	Durée:1h30.....
Enseignant Responsable: Kanzari Dalel.....	Nombre de pages :2.....
Classe : Ing-GL-A2.....	Documents autorisés :Non.....

Examen Principal

Exercice :

Une agence immobilière, à plusieurs filiales, cherche à créer un entrepôt de données afin d'analyser son système de gestion des biens immobiliers mis à sa disposition et de prendre de nouvelles orientations stratégiques tels que : extension territoriale à de nouvelles régions, fermeture ou transfert de filiales mal implantées, formation des dirigeants des filiales, extension des types de logements, etc.


On dispose des informations (extraites du système opérationnel) suivantes :

- Pour chaque logement on fournit : l'adresse, le nom du propriétaire, le type (maison/appartement), le nombre de pièces, la surface habitable, l'état de l'habitation (neuf, bon état, très bon état, à rénover), l'objectif de gestion (vente, location), le prix de mise en vente ou de location mensuelle, la date de disponibilité, la ville et la région du bien. Chaque propriété peut avoir un garage. Un garage est caractérisé par un type (box, emplacements numérotés, etc.) et dans certain cas peut avoir une adresse différente de celle de la propriété.
- Un propriétaire, qui est caractérisé par un identifiant unique, un nom, un prénom, un genre, une catégorie d'âge (jeune, adulte, âgé), un numéro de téléphone et une adresse, peut mettre en location ou en vente un de ses logements auprès de l'agence.
- Un logement à vendre (resp. à louer) est acheté (resp. loué) par un seul acheteur(resp. locataire).
- Un acheteur (resp. un locataire) est identifié par un numéro unique, un nom, un prénom, un genre, une catégorie d'âge (jeune, adulte, âgé), un numéro de téléphone et une adresse.
- Pour chaque transaction de vente (resp.de location) :
 - L'agence touche une commission qui correspond à un pourcentage du prix de la transaction (qui est composé d'une valeur fixe (vf) à laquelle on additionne 3% du montant de la transaction).

- On enregistre la date de la transaction, les identifiants du propriétaire et de l'acheteur (resp. locataire), le montant de la transaction, la commission et le nombre de vente (resp. location).

- 1) Proposer un modèle en étoile (hiérarchisé) correspondant à S.
- 2) Proposer un schéma en flocon de S (hiérarchisé).
- 3) Soit le cube OLAP C une représentation multidimensionnelle de S. A partir de C, indiquez les opérations OLAP (RollUp, DrillDown, Slice, Dice) ainsi que des opérations agrégatives (sum, avg, count, max, min) qu'il faut appliquer afin d'obtenir les informations suivantes :
 - Les nombres des transactions de ventes par catégorie d'âge des acheteurs en 2019.
 - Le nombre mensuel des locations par genre des locataires et par type de logement.
 - La meilleure commission de vente en décembre 2020.
 - Les montants de location par ville et par nombre de pièces.
 - La moyenne des montants de vente par type de logement entre juin et août 2020 et par région.
 - Le nombre de ventes par catégorie d'âge des acheteurs en 2020.
 - Les montants des ventes par type de garage et par région de la filiale.
 - Le nombre saisonnier de location par catégorie d'âge des locataires et par état de logement.

B.travail

Institut Supérieur des Sciences Appliquées et de Technologie de Sousse 	Département : Informatique Année universitaire : 2015 - 2016 Matière : Entrepôt de Données Filières : 2 ^{ème} ING -GL Enseignant : D. Kanzari Date : 28 Mai 2016 Nombre de pages : 02 Durée : 1h 30 min
Examen Session Principale	

Exercice 1:

- Quelles sont les principaux types de moteurs OLAP.
- Expliquer l'utilité des termes suivants: granularité, schéma hiérarchique, mesures et attributs d'analyse.
- Quelle est la différence entre set "{ }" et tuple "(")" dans une requête MDX. Donner un exemple.
- Exprimer la syntaxe ainsi que le rôle des opérations suivantes : DICE et DRILL UP.
- Expliquer la différence entre les méthodes supervisées et non supervisées dans la restitution des informations. Citer des exemples de méthodes.

Exercice 2:

Soit R le schéma relationnel d'un système opérationnel défini par:

Candidat (Num_C, CIN, nom, prenom, age, genre, adresse, tel, num_auto_ecole)

Lieu_examen (Num_L, ville, code-postal, région, pays)

Date_examen (Num_D, jour, mois, année)

Inspecteur (Num_I, nom, prenom, adresse, age, genre, num_direction)

Examen_permis (Num_C, Num_L, Num_D, Num_I, note_examen, nbr_fautes, frais_examen, nbre_examen)

Les règles de gestion de R sont:

- La clé de chaque relation est soulignée
- L'attribut num_auto_ecole est l'identifiant de l'auto-école du candidat. Une auto-école est définie par (nom, id_directeur, adresse, catégorie_formation.)
- L'attribut num_direction est le numéro de la direction à laquelle appartient l'inspecteur. Une direction est représentée par (nom, désignation, nom_directeur, adresse.)
- L'attribut note_examen désigne la note attribuée par l'inspecteur au candidat.
- L'attribut nbr_fautes indique le nombre de fautes que le candidat a commis lors de l'examen.

- Un candidat peut passer plusieurs examens de permis de conduire pour des dates différentes avec des inspecteurs différents.

Questions:

A partir du schéma R on souhaite modéliser un schéma multidimensionnel S orienté métier qui permet de dégager les notes moyennes de même que les frais et les nombres de fautes dans le temps, par candidat, par inspecteur et par lieu d'examen ainsi que toutes les opérations agrégatives possibles.

- 1) Proposer un modèle en étoile (hiérarchisé) correspondant à S.
- 2) Proposer un schéma en flocon de S (hiérarchisé) qui tient compte des informations relatives à l'auto-école du candidat et à la direction de l'inspecteur.
- 3) Soit le cube OLAP C une représentation multidimensionnelle de S. A partir de C, indiquez les opérations OLAP (RollUp, DrillDown, Slice, Dice) ainsi que des opérations agrégatives (sum, avg, count, max, min) qu'il faut appliquer afin d'obtenir les informations suivantes :
 - Les frais totaux des examens de permis de conduire par âge du candidat en 2013.
 - La meilleur note d'examen en decembre 2014.
 - Le nombre de fautes en moyenne par genre de candidat et par région d'examen.
 - Le nombre d'examens par genre d'inspecteur en mars 2015 à Tunis.
- 4) Ecrire les requêtes MDX qui permettent de répondre aux requêtes précédentes.

Bon Travail

Travaux Dirigés n°1 : Modélisation d'un ED

Exercice 1 :

Une entreprise commerciale de vente de matériels informatique à plusieurs filiales internationales souhaite rassembler dans un Entrepôt de Données (ED) des informations sur les ventes du jour afin de dresser des tableaux de bord sur les chiffres d'affaires réalisés.

* L'entreprise dispose d'un système d'information complexe, constitué des éléments suivants:

- des applications et bases de données éparées et hétérogènes sur les produits qu'elle vend,
- des applications et bases de données variées sur les clients,
- des Bases de données sur les personnels de l'entreprise,
- des Bases de données sur les filiales de l'entreprise.

L'ED à modéliser doit pouvoir fournir les chiffres d'affaires des ventes d'un produit, par date, par client, et par vendeur, ainsi que toutes les opérations d'agrégation (sum, avg, min, max, etc) possibles sur les chiffres d'affaires.

Nous disposons des informations suivantes :

- Une vente correspond à un produit et un seul, (A - IP)
- Un produit est défini par: Refp, nom, description, type, famille.
- Une vente est effectuée par l'un des vendeurs du service de vente spécialisé dans le produit.
- Un client est défini par : RefC, nom, prénom, genre, adresse, catégorie.
- Un vendeur est défini par : RefV, nom, prénom, genre, adresse, filiale.

Questions :

- Proposer des indicateurs d'analyse pour ce système de gestion
- Proposer des axes d'analyse pour ce système.
- Donner la table principale de l'entrepôt ainsi que les tables dimensions relatives.
- Proposez une hiérarchie pour chaque dimension.
- Proposer un schéma en étoile dimensionnel puis en flocon de l'ED.

Exercice 2 :

Un institut universitaire cherche à étudier les facteurs influant sur la réussite de ses étudiants aux examens. Pour cela elle décide de construire un ED.

Elle souhaite pouvoir répondre aux questions suivantes:

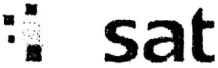
- Quel est le nombre de réussites aux examens par cours, pour l'année 2016?
- Quel est le nombre de réussites aux examens d'un cours obligatoire, pour l'année 2016?
- Quel est le nombre de réussites aux examens par genre (féminin, masculin), pour l'année 2016?
- Combien d'étudiants ayant un âge de 22 ans ont réussi leurs examens de base de données relationnelle?
- Quel est le nombre de réussites aux examens pendant le semestre d'hiver 2015?

Pour cela elle dispose des données suivantes:

Pour chaque examen passé, on connaît l'âge et le genre de l'étudiant, le nom du cours (les cours peuvent être regroupés en cours obligatoire et cours à option), la date de l'examen, la note obtenue et si l'examen est réussi ou non.

Questions :

- Donner la table principale de l'entrepôt ainsi que les tables dimensions relatives.
- Proposez une hiérarchie pour ces dimensions.
- Tracer le schéma en étoile dimensionnel puis en flocon de l'ED.

<p>Institut Supérieur des Sciences Appliquées et de Technologie de Sousse</p> <p></p>	<p>Département : Informatique Année universitaire : 2017 - 2018 Matière : Entrepôt de Données Filières : 2^{ème} ING -GL Enseignant : D. Kanzari Date : 01 Juin 2018 Nombre de pages : 01 Durée : 1h 30 min</p>
<p align="center">Examen Session Principale</p>	

Exercice :

Une société de distribution de produits alimentaires à plusieurs filiales approvisionne des magasins en divers types de ses produits, en effectuant au plus une livraison par jour et par magasin par un de ces vendeurs. Un bon de livraison doit contenir : le numéro de référence, la date de livraison, la référence du magasin, le numéro du vendeur, et pour chaque type de produit livré sa référence et la quantité livrée (le nombre d'articles). La société souhaite analyser ces informations stockées dans ses bases de gestion afin d'améliorer le service de distribution et mieux répondre aux souhaits de ses clients.

Les analyses se font suivant plusieurs axes, et à plusieurs niveaux, en analysant les mouvements des produits par jour et par mois, par ville et par région, par vendeur, par fournisseur et par catégorie de produit.

On supposera qu'un fournisseur peut fournir au distributeur des produits dans plusieurs catégories et qu'une catégorie de produit peut être fournie par plusieurs fournisseurs.

Questions :

1. Proposer un schéma S en étoile du Datawarehouse
2. Proposer un schéma en flocon.
3. Soit le cube OLAP C une représentation multidimensionnelle de S. A partir de C, indiquez les opérations OLAP (RollUp, DrillDown, Slice, Dice) ainsi que des opérations agrégatives (sum, avg, count, max, min) qu'il faut appliquer afin d'obtenir les informations suivantes:
 - La quantité livrée par ville entre 2016 et 2017.
 - Le produit le plus livré en Juin 2017.
 - La quantité livrée par catégorie de produit et par magasin.
 - La montant moyen des livraisons par fournisseur et par catégorie de produit en 2017.
 - Le montant total mensuel des livraisons par vendeur et par type de produit.
4. Ecrire les requêtes MDX qui permettent de répondre aux requêtes précédentes.

Bon Travail