

Concepteurs : Paule-Annick DAVOINE, Ahlame DOUZAL, Didier DONSEZ

Durée : 2 heures

Remarques : Calculatrice et tout document autorisé

Conseil : Lire le sujet jusqu'au bout.

Rendre les réponses aux 3 problèmes sur des copies séparées

Problème 1 (6 points): Système d'information décisionnel épidémiologique

Le Ministère de la Santé et du Bien-Etre de Grolang vous sous-traite la réalisation d'un entrepôt de données pour réaliser des études épidémiologiques pour détecter et prévenir les catastrophes sanitaires (légionellose, grippe, VIH, décès pour cause de canicule, ...). Les bases de production de cet entrepôt sont les systèmes d'information des centres de sécurité sociale, des hôpitaux, des services d'urgence (SMUR, SAMU, Pompier, Police Secours, des morgues, ...)

Le schéma de l'entrepôt est constitué des tables suivantes (les clés primaires sont soulignées)

Date(CléDate, Année, Mois, JourDeMois, JourDeSemaine, TrancheHoraire, DrapeauVacances)

Géographie(CléGéo, Région, Département, District, Ville, Quartier, zone-intervention_sanitaire, zone_intervention_securite)

Personne(CléPersonne, MoisNaissance, AnnéeNaissance, MoisDécès, AnnéeDécès, NbPersonneFoyer, CléDémó)

Demographie(CléDémó, MoisNaissance, AnnéeNaissance, MoisDécès, AnnéeDécès, RevenuAssuré, RevenuFoyer, CatégorieSocioProfessionnel, SousCatégorieSocioProfessionnel, DomaineActivité)

ActeurSanitaire(CléActeur, Type, Spécialité, SousSpécialité)

Intervention(CléDate, CléPersonne, CléActeur, CléPathologie, CléDémó, CléGéo, NombreDeJoursDArrêtDeTravail, DrapeauDécès, DrapeauInvalidant).

Pathologie(CléPathologie, DésignationNormalisé, Spécialité, SousSpécialité, TauxDIncapacité, DuréeTraitement, Chronicité, DrapeauMaladieProfessionnelle)

Rétro-Conception

Q1: Quelle est la table de fait dans cet entrepôt ?. Justifiez en 1 ligne!

Q2: Il y a t'il une mini-dimension ?. Pourquoi est elle nécessaire ? Justifiez en 3 lignes!

Dimensionnement

Q3: Donnez le nombre de faits présents dans la table de fait.

Nombre de personnes	60 Millions
Nombre d'interventions par acteur et par jour (Un acteur travaille 365 jours par an)	10
Nombre d'acteur	500000
Nombre d'années	10
Taille des clés	4 octets
Taille des attributs numériques	4 octets
Taille des attributs booléens (comme les drapeaux !)	1 octet

Donnez la taille d'un enregistrement de la table de fait ?

Donnez la taille (en Octets) de stockage de la table de fait.

Q4 : Est il peu judicieux d'utiliser les index bitmap sur les fichiers (stockage physique des

tables) dont les enregistrements ont une taille variable (justifiez en 6 lignes)

Rapports

Q5: Donnez le nombre de cas de gripes par région et par semaine et sa moyenne glissante sur les 3 semaines précédents.

Problème 2 (4 points): Analyse décisionnelle d'un Entrepôt de Données sur l'assurance maladie

On considère le tableau suivant généré à partir d'un entrepôt sur l'assurance maladie donnant la description des assurés de la caisse d'assurance maladie par leur Age, leur catégorie socio-professionnelle (Cat-So-Pr), la situation économique de leur foyer d'appartenance (Stab-Eco), s'ils bénéficient ou pas d'une caisse complémentaire (Caisse-Com), le nombre de praticiens différents visités durant l'année (Nb-Praticien), le montant total des dépenses (Mt-Dep) ainsi que le montant total des remboursements (Mt-Rem).

Assurés	Age	Cat-So-Pr	Stab-Eco	Caisse-Com	Nb-Praticien	Mt-Dep	Mt-Rem
P1	25	O	I	0	4	200	100
P2	32	C	MS	0	8	650	200
P3	58	C	S	1	12	800	700
P4	62	R	S	1	20	1500	1200
P5	75	R	I	1	6	1000	350
P6	84	R	MS	1	6	1900	900
P7	43	O	I	0	4	280	120

Type et domaine des attributs :

- *Attributs quantitatifs*

Age : entier sur [0-150]

Nb-Praticien : Entier

Mt-Dep : Réel

Mt-Rem : Réel

- *Attributs qualitatifs ordonnés*

Stab-Eco : I (Instable), MS (Moyennement Stable), S (Stable) avec $I < MS < S$

- *Attributs qualitatifs non ordonnés*

Cat-So-Pr : O (Ouvrier), C(Cadre), R (Retraité)

- *Attributs binaires*

Caisse-Com : 1 (l'assuré bénéficie d'une assurance complémentaire partenaire de la caisse d'assurance maladie)

0 (l'assuré ne bénéficie pas d'une assurance complémentaire ou celle-ci n'est pas partenaire de la caisse d'assurance maladie).

La caisse d'assurance maladie considère comme dépense élevée toute dépense annuelle supérieure (au sens strict) à 800 euros. Par ailleurs, les experts en assurance maladie considèrent qu'un pourcentage de remboursement supérieur à 70%, la catégorie socio-professionnelle des assurés ainsi que la situation économique de leur foyer d'appartenance ont un effet sur l'augmentation des dépenses.

Q1 : Après avoir procédé aux codages adéquats, appliquez un arbre binaires de décision afin d'extraire les principales règles expliquant les dépenses élevées en fonction de la catégorie socio-professionnelle, de la stabilité économique des assurés ainsi que du pourcentage des remboursements. L'hypothèse énoncée ci-dessus est-elle vraie ? Justifiez votre réponse.