Bases de Données avancées

Algèbre relationnel, Dépendances Fonctionnelles et Formes Normales

Travaux dirigés

Exercice 1 : Algèbre relationnel

Soit la base de donnée sur les schémas de relations suivants, représentée ci-dessous :
— Commandes = {Num, Cnom, Pnom, Qte} ;
— Fournisseurs = {Fnom, Status, Ville} ;
— Produits = {Pnom, Fnom, Prix}.

commandes	Num	Cnom	Pnom	Qte
	1535	Jean	Cornas	6
	1854	Jean	Bordeaux	20
	1254	Paul	Chablis	20
	1259	Paul	Chablis	25
	1596	Paul	Cornas	12

fournisseurs	Fnom	Status	Ville
	Vini	SARL	Dijon
	BonVin	SA	Dijon
	Chapoutier	SA	Valence
	SaV	Association	Antraïgues

produits	Pnom	Fnom	Prix
	Cornas	BonVin	20
	Cornas	Chapoutier	18
	Bordeaux	Vini	8.2
	Boudes	Vini	4.3
	Bordeaux	Chapoutier	18.5
	Chapoutier	Chapoutier	5.1
	Chablis	Chapoutier	5

Donner une expression en algèbre relationnelle des requêtes ci-après.
Exemple *les villes des fournisseurs de Cornas* :

$$\pi_{\text{Ville}}(\text{Fournisseurs} \bowtie \sigma_{\text{Pnom}=\text{`Cornas'}}(\text{Produits}))$$

1. Donner toutes les commandes.
2. Donner les noms des produits commandés.
3. Donner les noms des produits commandés par Jean.
4. Donner les noms de fournisseurs de Bordeaux ou de Cornas vendus à un prix inférieur à 10€.
5. Donner les noms des produits dont le nom est le même que le nom d'un fournisseur.
6. Donner le nom, le prix et les fournisseurs potentiels des produits commandés par Jean.
7. Donner les paires de fournisseurs qui habitent dans la même ville. Idem sans doublons, c-à-d, on retourne soit $\{(\text{Vini}, \text{BonVin})\}$ soit $\{(\text{BonVin}, \text{Vini})\}$ mais pas les deux.
8. Donner les noms des produits qui coûtent plus de 15€ ou qui sont commandés par Jean.
9. Donner les noms des produits qui n'ont pas été commandés.
10. Donner les noms des produits commandés au moins une fois en quantité supérieure à 10 et dont le prix est inférieur à 15€ chez au moins un fournisseur.
11. Donner les noms des produits qui sont fournis par tous les fournisseurs.
12. Donner les noms des produits les plus chers.

Exercice 2 : DF, DJ et cahier des charges

Une chaîne de magasins fournit des véhicules dans ses différentes agences. Sa Base de Données gère les véhicules (identifiés par V) qui sont d'un type donné (identifié par T), les agences (identifiées par A) et les commerciaux (identifiés par C). Exprimez sous forme de dépendances les contraintes suivantes du cahier des charges.

1. (Dépendances Fonctionnelles)
   1. Chaque agence est basée dans une ville (*Ville*)
   2. Les Agences proposent différents types de véhicules, pour un prix (*Prix*) qui varie selon les années.
   3. Chaque année civile (*Annee*), les commerciaux peuvent changer la liste des agences avec lesquelles ils collaborent. Les mutations ne sont jamais faites en cours d'année.
   4. Pour chaque type de véhicule, on retient la date de la dernière formation (*DateFormation*) qu'un commercial a suivi
   5. Lorsqu'un commercial effectue une vente d'un véhicule, on retiendra le prix effectif (*prix_vente*) et la date de la transaction *Date_Vente*. Chaque véhicule ne peut bien sûr être vendu qu'une seule fois - et pas un seul commercial.
2. (Dépendances de Jointure) On considère maintenant plus précisément une relation $R(ATC)$ définie uniquement sur ces attributs. Celle-ci pourrait représenter différentes cahiers des charges ; donnez les contraintes qui découlent des versions suivantes.
   1. Chaque agence rattache un ensemble de commerciaux. Elle fournit des types de véhicules.
   2. Chaque agence rattache un ensemble de commerciaux. Elle possède des types de véhicules. Les commerciaux sont habilités à intervenir sur des types de véhicules.
   3. Les commerciaux sont affectés à des agences, mais pour s'occuper de certains types de véhicules seulement. Dans deux agences, ils peuvent être affectés à des types de véhicules différents. Ces affectations sont décidées par la hierarchie et stockées dans la base.
3. Pour les cahiers de charges de la question 2, faites une esquisse de diagramme E/A pour représenter la situation. Pour quel cahier des charges le schéma de relation $R(ATC)$ est-il une bonne représentation relationnelle ?

Exercice 3 : Propagation des DF

Soit r une instance de R qui satisfait à la dépendance $R : X \to Y$ (soit $r \models X \to Y$) et s une instance quelconque sur un schéma S. Pour chaque expression ci-dessous, indiquer en le justifiant si elle est vraie.

1. $\sigma_C(r) \models X \to Y$
2. $\pi_W(r) \models X \to Y$ (en supposant $XY \subseteq W$)
3. $r \bowtie s \models X \to Y$
4. En supposant $R = S$:
   — $r \cup s \models X \to Y$
   — $r \setminus s \models X \to Y$
5. En supposant $R \cap S = \emptyset$: $r \times s \models X \to Y$

Exercice 4 : Démonstration de la transitivié chez les DF

En vous appuyant sur la définition de la satisfaction d'une DF par une relation r, démontrez la propriété de transitivé des DF.

Aide : considérez une relation r quelconque qui satisfait $X \to Y$ et $Y \to Z$, puis vérifiez alors que r satisfait $X \to Z$.

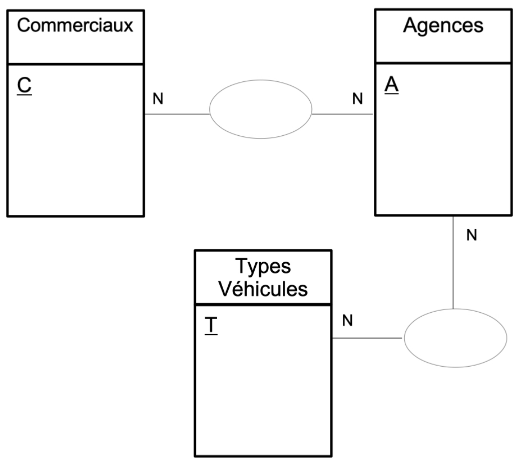
Corrections

Solution de l'exercice 1

1. Commandes
2. $\pi_{\text{Pnom}}(\text{Commandes})$
3. $\pi_{\text{Pnom}}(\sigma_{\text{Cnom='Jean'}}(\text{Commandes})$
4. $\pi_{\text{Fnom}}(\sigma_{(\text{Pnom='Cornas'} \vee \text{Pnom='Bordeaux'}) \wedge \text{Prix} \leq 10}(\text{Produits})$
5. Une réponse proposée est $\pi_{Pnom}(\sigma_{Pnom=Fnom}(Produits))$ elle est erronée car elle est limitée aux produits vendus par des fournisseurs de même nom. $\pi_{\text{Pnom}}(\text{Produits}) \cap \rho_{\text{Fnom/Pnom}}(\pi_{\text{Fnom}}(\text{Fournisseurs}))$
6. $\pi_{\text{Pnom,Prix,Fnom}}(\text{Produits} \bowtie \sigma_{\text{Cnom='Jean'}}(\text{Commandes}))$
7. $\sigma_{\text{Fnom}<\text{Fnom}'}(\text{Fournisseurs} \bowtie \rho_{\text{Fnom/Fnom}',\text{Status/Status}'}(\text{Fournisseurs}))$
8. $\pi_{\text{Pnom}}(\sigma_{\text{Cnom='Jean'}}(\text{Commandes})) \cup \pi_{\text{Pnom}}(\sigma_{\text{Prix} \geq 15}(\text{Produits}))$
9. $\pi_{\text{Pnom}}(\text{Produits}) \setminus \pi_{\text{Pnom}}(\text{Commandes})$
10. $\pi_{\text{Pnom}}(\sigma_{\text{Prix} \leq 15}(\text{Produits})) \cap \pi_{\text{Pnom}}(\sigma_{\text{Qte} \geq 10}(\text{Commandes})$
11. $\pi_{\text{Pnom}}(\text{Produits} \div \text{Fournisseurs})$ ou sans division explicite
    $\pi_{\text{Pnom}}(\text{Produits}) \setminus \pi_{\text{Pnom}}(((\pi_{\text{Pnom}}(\text{Produits})) \times (\pi_{\text{Fnom}}(\text{Fournisseurs}))) \setminus \pi_{\text{Pnom,Fnom}}(\text{Produits}))$
12. $\pi_{\text{Pnom}}(\text{Produits}) \setminus \pi_{\text{Pnom}}(\sigma_{\text{Prix}<\text{Prix}'}(\text{Produits} \times \rho_{(\text{Pnom/PNom}',\text{Fnom/FNom}',\text{Prix/Prix}')}(\text{Produits})))$

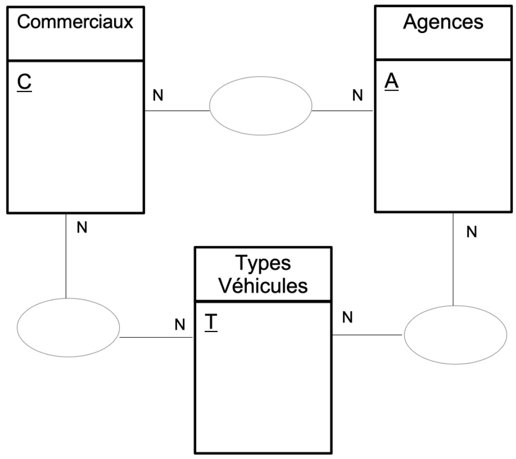
Solution de l'exercice 2

1. 1. $A \rightarrow Ville$
   2. $A, T, Annee \rightarrow Prix$
   3. Aucune !
   4. $C, T \rightarrow DateFormation$
   5. $V \rightarrow C, PrixVente, DateVente$. On pourra d'une façon encore plus précise utiliser un attribut *VVendu* pour gérer un sous-ensemble des voitures qui sont effectivement vendues, car les autres ne participent pas à cette DF. Cela donnerait : $VVendu \rightarrow C, PrixVente, DateVente$
2. 1. On voit qu'il n'y a pas de lien direct entre les commerciaux et les types de véhicules. Un commercial sera en relation avec un type de véhicule par le simple fait que son agence fournit ce véhicule. C'est typiquement une dépendance multivaluée : $A \rightarrow\rightarrow C|T$ qu'on peut écrire comme une dépendance de jointure de taille 2 : $\bowtie [AC, AT]$.



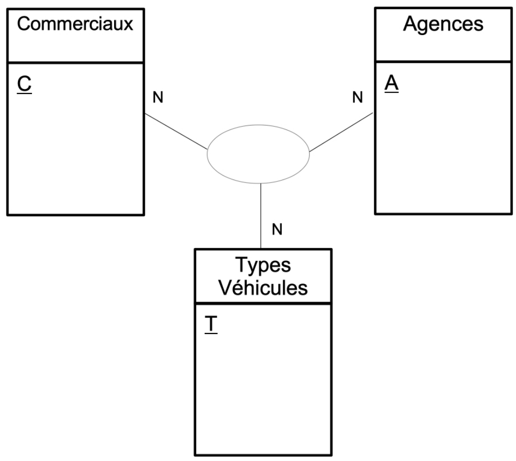
La traduction en relationnel ici n'est pas conforme à ce qu'on aurait dû faire : une relation pour chaque association soit 2 relations. La dépendance multivaluée montre cette mauvaise conception.

2. Ici chacun des trois liens est pertinent et de découle pas simplement des deux autres ; il n'y a pas de dépendance multivaluée. En revanche, il n'y a pas de lien particulier qui associe les trois attributs ; leur association est le résultat des 3 liens de taille 2. En d'autres termes, "Si un vendeur est habilité à vendre le type de véhicule commercialisé par une agence où il est affecté (on a bien les 3 liens), alors ce vendeur vend effectivement ce type de véhicule dans cette agence (triplet conséquence). Ceci s'exprime par la dépendance de jointure : $\bowtie [AC, AT, CT]$.



Là encore, le schéma de relation R devrait être remplacé par trois schémas de relation, un par association. la DJ montre bien cette mauvaise conception.

3. En revanche ici, les triplets sont bien intentionnellement créés et conservés dans la base ; on ne peut les retrouver par une jointure, il faut les stocker pour savoir lesquels sont réellement existant dans l'application. IL n'y a pas de dépendance de jointure.



La relation R est la bonne représentation en relationnel de ce cahier des charges.

Solution de l'exercice 3

Puisque $r \models X \rightarrow Y$, il n'y a aucun couple de tuple "contre-exemple" de cette DF dans r : $\nexists t_1, t_2 \in r | t_1[X] = t_2[X] \cap t_1[Y] = t_2[Y]$. Dans la suite, il suffira de vérifier si de nouvelles combinaisons sur XY sont générées par la requête, susceptibles de générer un contre-exemple.

1. vraie car comme $\sigma_C(r) \subseteq r$ il ne peut pas y avoir plus de contre-exemples de $X \rightarrow Y$ dans $\sigma_C(r)$ que dans r qui n'en contient aucun ;

2. vraie car $\pi_{XY}(\pi_W(r)) = \pi_{XY}(r)$. Donc on ne crée pas de nouvelle combinaison XY.
3. vraie car $\pi_{XY}(r \bowtie s) \subseteq \pi_{XY}(r)$. Donc on ne crée pas de nouvelles combinaisons XY.
   - faux en général, même si $s \models X \rightarrow Y$ car on crée de nouveaux couples de tuples à partir des tuples de r et de s. On peut générer des contre-exemple de la DF.
   - vraie, car $\pi_{XY}(r \setminus s) \subseteq \pi_{XY}(r)$ donc on ne crée pas de nouvelles combinaisons XY.
4. vraie, $\pi_{XY}(r \times s) \subseteq \pi_{XY}(r)$ donc on ne crée pas de nouvelles combinaisons XY.

Solution de l'exercice 4

Soit r une relation sur R telle que $r \models X \rightarrow Y$ et $r \models Y \rightarrow Z$.

Soient t_1 et t_2 deuc tuples de r, tels que $\underline{t_1[X] = t_2[X]}$. Puisque $r \models X \rightarrow Y$, on a $t_1[Y] = t_2[Y]$. Et puisque $r \models Y \rightarrow Z$, alors $\underline{t_1[Z] = t_2[Z]}$. CQFD.