25. Modèle relationnel et conception de bases de données

Niveau : Master 1 **Prérequis :** Licence Maths / Info

Introduction : traitement de données nombreuses et sensibles.

exemples : données médicales, étude de marché...

I Le modèle relationnel

On considère trois ensembles $\underline{Rel}$, $\underline{Att}$ et $\underline{Const}$ tels que $Rel \cap Att = \emptyset$, dont les éléments sont respectivement appelés relations, attributs et constantes.

A. La perspective sans noms

Définition 1 : Un tuple de données est un élément de $Const^k$, pour un entier $k \in \mathbb{N}$. k est appelé l'arité du tuple.

Définition 2 : Une table est un ensemble fini S de tuples de données de même arité. On note $\underline{\mathcal{T}}$ l'ensemble des tables.

Exemple 3 :

"Alice"	"guerrier"	13
"Bob"	"mage"	42

tuples → lignes ; } table

Définition 4 : Un schéma de base de données est une fonction partielle $S : Rel \to \mathbb{N}$ telle que $Dom(S)$ est fini. Pour tout $R \in Dom(S)$, on dit que $|S(R)|$ est l'arité de R dans S.

Définition 5 : Une base de données D de schéma S est une fonction $D : Dom(S) \to \mathcal{T}$ telle que pour tout $R \in Dom(S)$, l'arité de $D(R)$ est celle de $S(R)$.

On note $\underline{Inst}(S)$ l'ensemble des bases de données de schéma S.

Exemple 6 : S : | "Personnages" $\mapsto$ 3
"Équipement" $\mapsto$ 2

D("Personnages") = table de l'exemple 3.

DM 7 : Définition de la perspective avec noms et preuve de l'équivalence.

B. Les requêtes

DEV 1

Définition 8 : Une requête d'arité $k \geq 0$ sur un schéma S est une fonction $q : Inst(S) \longrightarrow \mathcal{P}_{fin}(Const^k)$.

⚠ cet objet est purement sémantique. Pour le rendre expressif (donc utile), on doit lui associer une syntaxe : SQL, FOL...

II. Algèbre relationnelle

A. Syntaxe

Définition 9 :

Les expressions relationnelles sur un schéma S et leurs arités sont définies inductivement comme suit :

Cas de base • $\forall a \in Const$, $\{a\}$ est une ER sur S d'arité 1.

• $\forall R \in Rel$ d'arité k sous S, R est une ER sur S d'arité k.

Sélection $\forall e$ ER sur S d'arité k et θ condition sur $[\![1,k]\!]$, $\sigma_\theta(e)$ est une ER d'arité k.

Projection $\forall e$ ER sur S d'arité k et $\alpha = (i_1, \dots, i_m)$ liste d'entiers de $[\![1,k]\!]$, $\pi_\alpha(e)$ est une ER d'arité m.

Produit Cartésien $\forall e_1, e_2$ ER sur S d'arités respectives k et m, $(e_1 \times e_2)$ est une ER d'arité $k+m$.

Union $\forall e_1, e_2$ ER sur S de même arité k, $(e_1 \cup e_2)$ est une ER d'arité k.

Différence $\forall e_1, e_2$ ER sur S de même arité k, $(e_1 - e_2)$ est une ER d'arité k.

B. Sémantique

Définition 10 : Étant donné un tuple $\bar{a} = (a_1, \dots, a_k) \in Const^k$, et une liste $\alpha = (i_1, \dots, i_m)$ d'entiers de $[\![1,k]\!]$, la projection $\pi_\alpha(\bar{a})$ est définie par l'égalité : $\pi_\alpha(\overline{a}) = (a_{i_1}, \dots, a_{i_m})$

Définition 11 : On définit inductivement la satisfaction de la condition Θ par le tuple $\overline{a}$ comme suit :

$\overline{a} \models i \doteq c$ si $a_i = c$ $(c \in Const)$

$\overline{a} \models i \doteq j$ si $a_i = a_j$

$\overline{a} \models \Theta \wedge \Theta'$ si $\overline{a} \models \Theta$ et $\overline{a} \models \Theta'$

$\overline{a} \models \neg\Theta$ si $\overline{a} \not\models \Theta$

Définition 12 : Soit D une base de données de schéma S et e une ER sur S. On définit inductivement la requête relationnelle $e(D)$:

si $e = R$ alors $e(D) = D(R)$

si $e = \{a\}$, alors $e(D) = \{a\}$ $(a \in Const)$

si $e = \sigma_\Theta(e_1)$, alors $e(D) = \{\overline{a} \mid \overline{a} \in e_1(D)$ et $\overline{a} \models \Theta\}$

si $e = \pi_\alpha(e_1)$, alors $e(D) = \{\pi_\alpha(\overline{a}) \mid \overline{a} \in e_1(D)\}$

si $e = (e_1 \times e_2)$, alors $e(D) = e_1(D) \times e_2(D)$

si $e = (e_1 \cup e_2)$, alors $e(D) = e_1(D) \cup e_2(D)$

si $e = e_1 - e_2$, alors $e(D) = e_1(D) \setminus e_2(D)$.

Définition 13 : La Θ-jointure de e_1 et e_2, deux ER, est définie comme

$$(e_1 \bowtie_\Theta e_2)(D) = \sigma_\Theta(e_1(D) \times e_2(D))$$

III Requêtes FOL ("first_order_logic")

Définition 14 : Étant donné un schéma S, les formules logiques du premier ordre sur S sont définies inductivement comme suit :

Cas de base : • $\forall a \in Const$, $x \in Var$, $y \in Var$, $x = a$ et $y = x$ sont des formules atomiques.

• $\forall u_1, \dots, u_k$ termes, et R un symbole de relation de S d'arité k, $R(u_1, \dots, u_k)$ est une formule atomique.

• $\forall \varphi_1, \varphi_2$ formules et $x \in Var$, $(\varphi_1 \wedge \varphi_2)$, $(\varphi_1 \vee \varphi_2)$, $\neg\varphi_1$, $\exists x.\varphi_1$, $\forall x.\varphi_1$ sont des formules.

Définition 15 : Sémantique des requêtes FOL.

$\boxed{\text{Théorème 16 : Les requêtes ER et FOL ont des expressivités équivalentes.}}$

DEV 2.