Rappel

Trois couches : Externe, conceptuelle, interne

L'utilisateur ne voit que la vue de la base qui lui a été attribuée (schéma externe relationnel).

Il spécifie ce qu'il veut via un langage de requête déclaratif (SQL).

Une fonction défini le passage schéma externe → schéma conceptuel

Une compilation effectue le passage conceptuel → interne

Chacune des couches peut être modifiée indépendamment des autres.

Les performances dépendent de la qualité de la compilation → restriction des possibilité de l'utilisateur : SQL.

COMMANDES(num,cnom,pnom,quantité)
FOURNITURE(pnom,fnom,prix)

REQUÊTE: *Nom, prix et fournisseurs des Produits commandés par Jean*

```
SELECT pnom,prix,fnom
FROM   FOURNITURE
WHERE  pnom IN (SELECT pnom
                FROM   COMMANDES
                WHERE  cnom = 'Jean')
```

ou

```
SELECT FOURNITURE.pnom,prix,fnom
FROM   FOURNITURE,COMMANDES
WHERE  FOURNITURE.pnom = COMMANDES.pnom
AND    cnom = ``Jean''
```


Quelle requête est la meilleure?

Aucune!!!

On verra que, en théorie, le compilateur devrait être capable de générer le meilleur plan d'évaluation indépendamment de la façon dont la requête est exprimée.

Le but est que l'utilisateur n'ait pas à se soucier de ce genre de "détail".

La semaine passée

Présentation du modèle relationnel

SQL

Aujourd'hui et la semaine prochaine

Les fondements du relationnel:

Le calcul relationnel: CALC

L'algèbre relationnelle: ALG

Le théorème d'équivalence: ALG=CAL=SQL

Intérêt de CALC : étude théorique des propriétés de SQL

Intérêt de ALG : facile à implanter, permet de pré-compiler SQL, l'optimisation se fait à ce niveau.

Relationnel: rappel

DOMAINE , ATTRIBUTS

SCHÉMA de relation

Film(titre:string, année:int, genre:string)

SCHÉMA de base de donnée

BD = { Film, Cinéma, Acteur }

TUPLE d'une relation

INSTANCE d'un schéma de relation

INSTANCE d'un schéma

Requêtes Conjonctives

Sous ensemble de SQL.

Les plus utilisées en pratique.

Celles qu'on sait très bien optimiser.

On va voir:

- Déclaratif: CQ $\subsetneq$ FO (pour faire de la belle théorie)
- Procédural : SRPJ $\subsetneq$ ALG (pour implanter, optimiser etc.)
- Utilisateur : SFW $\subsetneq$ SQL (ce qu'on utilise en pratique)

Rappel sur le relationnel : vue ensembliste

- En théorie

  Pas deux lignes identiques

  Tous les attributs d'une ligne sont connus

  L'ordre des lignes est à priori sans importance

- En pratique

  On n'élimine pas les duplicata sauf si explicitement demandé

  Certains attributs peuvent ne pas être connus

  L'ordre des lignes d'un résultat peut être imposé

L'algèbre SRPJ

Manipule des relations (tables) via leur nom et celui de leurs attributs.

Salle(nom,horaire,titre)
Film(titre,réalisateur,acteur)

Contient quatre opérations:

- σ: sélection
- π: projection
- $\bowtie$: jointure
- ρ: renommage

σ, sélection, opérateur unaire.

$$\sigma_F(R)$$

R est un nom de relation

F est un prédicat de filtre : A = Cst ou A = B, où A et B sont des noms d'attribut.

R

A	B
1	2
4	2
6	6
7	7
1	7
1	6

$\sigma_{A=1}\ (R)$

A	B
1	2
1	7
1	6

Salle(nom,horaire,titre)
Film(titre,réalisateur,acteur)

REQUÊTE : *tout sur le film "Vertigo"*

$$\sigma_{\text{titre="Vertigo"}}(\text{Film})$$

REQUÊTE : *les films qui ont un acteur qui est aussi le réalisateur*

$$\sigma_{\text{réalisateur=acteur}}(\text{Film})$$

π, projection, opérateur unaire.

$$\pi_\Delta(R)$$

R est un nom de relation

Δ est un ensemble de noms d'attribut, par exemple { acteur, titre }

R

A	B
1	2
4	2
6	6
7	7
1	7
1	6

$\pi_A(R)$

A
1
7
6
4

Salle(nom,horaire,titre)
Film(titre,réalisateur,acteur)

REQUÊTE : *La liste des salles de cinéma*

$$\pi_{\text{nom}}(\text{Salle})$$

REQUÊTE : *les acteurs du film "Vertigo"*

$$\pi_{\text{acteur}}(\sigma_{\text{titre="Vertigo"}}(\text{Film}))$$

$\bowtie$, jointure, opérateur binaire.

$$R_1 \bowtie R_2$$

R_1 et R_2 sont des noms de relation

R

A	B
1	2
4	2
6	6
7	9
1	7
1	6

S

B	C
2	3
2	5
9	1
8	8

$R \bowtie S$

A	B	C
1	2	3
1	2	5
4	2	3
4	2	5
7	9	1

Salle(nom,horaire,titre)
Film(titre,réalisateur,acteur)

REQUÊTE : *Les salles où l'on joue un film avec Stewart*

$$\pi_{\text{nom}}(\sigma_{\text{acteur="Stewart"}}(\text{Film} \bowtie \text{Salle}))$$

ou

$$\pi_{\text{nom}}(\sigma_{\text{acteur="Stewart"}}(\text{Film}) \bowtie \text{Salle})$$

ρ, renommage, opérateur unaire.

$$\rho_{A \to A'}(R)$$

R est un nom de relation

A et A' sont des noms d'attribut

R

A	B
1	2
4	2
6	6
7	7
1	7
1	6

$\rho_{A \to A'}\ (R)$

A'	B
1	2
4	2
6	6
7	7
1	7
1	6

Permet de changer le nom de certains attributs afin de faire (ou ne pas faire) des jointures.

SRPJ contient quatre opérations:

- σ: sélectionne des lignes suivant un critère donné.
- π: sélectionne des colonnes
- $\bowtie$: fait un "pont" entre deux tables
- ρ: renommage

Il est très facile à compiler.

Exemples

Salle(nom,horaire,titre) Film(titre,réalisateur,acteur)
Produit(producteur,titre)

Quels sont les acteurs qui produisent un film dans lequel ils jouent ?

$$\pi_{\text{acteur}}(\sigma_{\text{producteur=acteur}}(\text{Produit} \bowtie \text{Film}))$$

ou

$$\pi_{\text{acteur}}(\rho_{\text{producteur}\to\text{acteur}}(\text{Produit}) \bowtie \text{Film})$$

Quels sont les acteurs qui produisent un film qu'ils ont réalisé?

$$\pi_{\text{acteur}}((\rho_{\text{réalisateur}\to\text{acteur}}\ (q) \bowtie \pi_{\text{acteur}}(\text{Film})))$$

où $q \equiv \pi_{\text{réalisateur}}(\rho_{\text{producteur}\to\text{réalisateur}}(\text{Produit}) \bowtie \text{Film})$

Les salles où l'on joue un film avec Stewart

$$\pi_{\text{nom}}(\sigma_{\text{acteur="Stewart"}}(\text{Film} \bowtie \text{Salle}))$$

ou $\pi_{\text{nom}}(\sigma_{\text{acteur="Stewart"}}(\text{Film}) \bowtie \text{Salle})$

Quels sont les acteurs qui produisent un film dans lequel ils jouent?

$$\pi_{\text{acteur}}(\sigma_{\text{producteur=acteur}}(\text{Produit} \bowtie \text{Film}))$$

ou $\pi_{\text{acteur}}(\rho_{\text{producteur}\to\text{acteur}}(\text{Produit}) \bowtie \text{Film})$

Une même requête peut s'évaluer de plusieurs manières différentes. On verra qu'un SGBD possède un module appelé l'optimiseur qui choisira le plan d'évaluation qu'il considère comme le meilleur.

Savez vous faire?

- Quels acteurs jouent dans tous les films de Welles ?
- Qui produit tous les films de Kurosawa ?
- Quels films ne passent dans aucune salle ?
- Qui produit un film qui ne passe dans aucune salle ?
- Quels producteurs voient tous les films de Kurosawa ?

CQ

Langage déclaratif (il faudra le compiler)

C'est facile d'y exprimer ce que l'on veux.

Manipule des noms de relation sans se préoccuper du nom des attributs mais uniquement de leur ordre.

Film(x,y,z)

Grâce au schéma de relation on sait que x est in titre, y est un réalisateur et z un acteur.

L' arité d'une relation est son nombre d'attributs (de colonnes).

soit un schéma de base τ.

Définition de CQ ensemble des formules bien formées définies par:

termes: x,y,z… (variables)

"Bergman", 12, … (constantes)

clauses : t=t', t<t', t>t', R($t_1, \cdots, t_k$) où R $\in \tau$ d'arité k

clôture par $\wedge$ et $\exists x$

exemple : $\exists x$ R(x,y) $\wedge$ $\exists x \exists y$ (R'(x,y,z) $\wedge x = 12$)

Une occurrence d'une variable x est dite liée si elle apparaît dans un bloc où x est quantifiée.

Une occurrence d'une variable x est dite libre si elle n'est pas quantifiée.

exemple:

∃ x R(x , y)

R(x , x) ∧ (∃ x R(x , y))

∃ y (R(x , x) ∧ (∃ x R(x , y)))

∃ x (R(x , x) ∧ (∃ x R(x , y)))

Soit φ une formule de CQ. Une variable x est dite libre dans φ si elle a une occurrence libre.

Soit $x_1 \cdots x_k$ les variable libres de φ.

φ défini alors une requête d'arité k

"L'ensemble des valeurs de $x_1, \cdots, x_k$ qui rendent la formule φ vraie"

Étant donné une base de donnée **I**, son domaine actif, noté adom(**I**) est l'ensemble des valeurs apparaissant dans **I**

On défini de même le domaine actif d'une requête q, adom(q) comme l'ensemble des constantes apparaissant dans q.

L'union des deux: adom(**I**,q).

Une valuation $f_{I,q}$ d'un ensemble V de variables est une fonction de V dans adom(**I**,q).

Sémantique de CQ

Fixons un schéma de base de donnée τ. Soit **I** une instance de base de donnée sur τ. Soit f une valuation.

I $\models_f$ R($x_1, \cdots, x_k$) si $(f(x_1), \cdots, f(x_k))$ est un tuple de R dans **I**

I $\models_f \psi_1 \wedge \psi_2$ si **I** $\models_f \psi_1$ et **I** $\models_f \psi_2$

I $\models_f \exists x \psi$ si il existe c dans adom(I,ψ) tel que **I** $\models_{f'} \psi$ où $f' = f \cup x \rightarrow c$.

Soit q une formule de CQ avec k variables libres $x_1, \cdots, x_k$, φ défini la requête suivante:

q(**I**) = $\{(c_1, \cdots, c_k) \in$ adom(**I**,φ) $\mid$ **I** $\models_f \varphi$ où $f(x_1) = c_1 \cdots f(x_k) = c_k\}$.

Exemples

Salle(nom,horaire,titre)

Film(titre,réalisateur,acteur)

Tout sur le film "Vertigo"

Film(x,y,z) $\wedge$ x="Vertigo"

$\sigma_{\text{titre="Vertigo"}}$(Film)

Les films qui ont un acteur qui est aussi le réalisateur

Film(x,y,z) $\wedge$ y=z

$\sigma_{\text{réalisateur=acteur}}$(Film)

Salle(nom,horaire,titre)
Film(titre,réalisateur,acteur)

La liste des salles de cinéma

$$\exists\, y, z \;\; \text{Salle(x,y,z)}$$

$$\pi_{\text{nom}}(\text{Salle})$$

Les acteurs du film "Vertigo"

$$\exists\, x, y \;\; (\text{Film(x,y,z)} \;\wedge\; \text{x="Vertigo"})$$

$$\pi_{\text{acteur}}(\sigma_{\text{titre="Vertigo"}}(\text{Film}))$$

Salle(nom,horaire,titre)
Produit(producteur,titre)
Film(titre,réalisateur,acteur)

Les salles où l'on joue un film avec Stewart

$$\exists\, x, y, z, v \;\; (\text{Film(x,y,z)} \;\wedge\; \text{Salle(u,v,x)} \;\wedge\; \text{z="Stewart"})$$

$$\pi_{\text{nom}}(\sigma_{\text{acteur="Stewart"}}(\text{Film} \bowtie \text{Salle}))$$

ou $\pi_{\text{nom}}(\sigma_{\text{acteur="Stewart"}}(\text{Film}) \bowtie \text{Salle})$

Quels sont les acteurs qui produisent un film dans lequel ils jouent?

$$\exists\, x, y \;\; (\text{Film(x,y,z)} \;\wedge\; \text{Produit(z,x)})$$

$$\pi_{\text{acteur}}(\sigma_{\text{producteur=acteur}}(\text{Produit} \bowtie \text{Film}))$$

ou $\pi_{\text{acteur}}(\rho_{\text{producteur}\to\text{acteur}}(\text{Produit}) \bowtie \text{Film})$

Salle(nom,horaire,titre)
Produit(producteur,titre)
Film(titre,réalisateur,acteur)

Quels sont les acteurs qui produisent un film qu'ils ont réalisé?

$$\exists\, x, z \; (\text{Film(x,y,z)} \;\wedge\; \text{Produit(y,x)} \;\wedge\; \exists\, u, v \;\; \text{Film(u,v,y)})$$

$$\pi_{\text{acteur}}((\rho_{\text{réalisateur}\to\text{acteur}}\,(q) \bowtie \pi_{\text{acteur}}(\text{Film})))$$

où $q \equiv \pi_{\text{réalisateur}}(\pi_{\text{réalisateur}}(\rho_{\text{producteur}\to\text{réalisateur}}(\text{Produit}) \bowtie \text{Film}))$

Que veut dire la requête de CQ: $x = x$?

A priori: adom(**I**)

Comme cela n'a pas beaucoup de sens en pratique, on élimine les requêtes de ce type.

Une requête q de CQ est dite sûre si chacune de ses variables libres apparaît dans une clause de q avec un nom de relation.

$$\text{R(x, y)} \;\wedge\; \text{x=z} \;\wedge\; \exists\, u, v \;\; \text{u=v}$$

$\Rightarrow$ associe un nom d'attribut à chaque variable.

Maintenant on appellera SCQ les formules sûres de CQ.

Forme normale

Toute formule q de CQ peut s'écrire sous la forme:

$$\exists\, y_1 \cdots y_n \;\; \bigwedge_i \theta_i(\bar{x}, \bar{y})$$

où $\bar{x}$ est l'ensemble des variables libres de q

et pour tout i, θ_i est une clause.

Savez vous faire?

- Quels acteurs jouent dans tous les films de Welles ?
- Qui produit tous les films de Kurosawa ?
- Quels films ne passent dans aucune salle ?
- Qui produit un film qui ne passe dans aucune salle ?
- Quels producteurs voient tous les films de Kurosawa ?

SFW

C'est SQL sans

GROUP-BY/HAVING

OR, UNION, EXCEPT

requêtes imbriquées: IN, NOT IN, EXISTS NOT EXISTS

Mais on garde

SELECT, FROM, WHERE, AND

Théorème d'équivalence:
SFW=SCQ=SRPJ

Le passage SFW $\to$ SRPJ correspond à la compilation de la requête. On reverra cela en détail plus tard.

Exemples

Salle(nom,horaire,titre)   Film(titre,réalisateur,acteur)

Tout sur le film "Vertigo"

Film(x,y,z) $\wedge$ x="Vertigo" $\sigma_{\text{titre="Vertigo"}}(\text{Film})$

```
SELECT *
FROM Film
WHERE titre="Vertigo"
```

Les films qui ont un acteur qui est aussi le réalisateur

Film(x,y,z) $\wedge$ y=z $\sigma_{\text{réalisateur=acteur}}(\text{Film})$

```
SELECT *
FROM Film
WHERE acteur=réalisateur
```


Salle(nom,horaire,titre)   Film(titre,réalisateur,acteur)

La liste des salles de cinéma

$\exists\, y, z$ Salle(x,y,z) $\pi_{\text{nom}}(\text{Salle})$

```
SELECT nom
FROM Salle
```

Les acteurs du film "Vertigo"

$\exists\, x, y$ (Film(x,y,z) $\wedge$ x="Vertigo") $\pi_{\text{acteur}}(\sigma_{\text{titre="Vertigo"}}(\text{Film}))$

```
SELECT acteur
FROM Film
WHERE titre="Vertigo"
```


Salle(nom,horaire,titre)   Film(titre,réalisateur,acteur)

Les salles où l'on joue un film avec Stewart

$\exists\, x, y, z, v$ (Film(x,y,z) $\wedge$ Salle(u,v,x) $\wedge$ z="Stewart")

$\pi_{\text{nom}}(\sigma_{\text{acteur="Stewart"}}(\text{Film} \bowtie \text{Salle}))$

```
SELECT nom
FROM Salle,Film
WHERE acteur="Stewart"
AND   Film.titre=Salle.titre
```


Salle(nom,horaire,titre)   Film(titre,réalisateur,acteur)
Produit(producteur,titre)

Quels sont les acteurs qui produisent un film dans lequel ils jouent?

$\exists\, x, y$ (Film(x,y,z) $\wedge$ Produit(z,x))

$\pi_{\text{acteur}}(\sigma_{\text{producteur=acteur}}(\text{Produit} \bowtie \text{Film}))$

```
SELECT acteur
FROM Film,Produit
WHERE Produit.titre=Film.titre
AND   producteur=acteur
```


Salle(nom,horaire,titre)   Film(titre,réalisateur,acteur)
Produit(producteur,titre)

Quels sont les acteurs qui produisent un film qu'ils ont réalisé?

$\exists\, x, z$ (Film(x,y,z) $\wedge$ Produit(y,x) $\wedge$ $\exists\, u, v$ Film(u,v,y))

$\pi_{\text{acteur}}((\rho_{\text{réalisateur}\to\text{acteur}}(q) \bowtie \pi_{\text{acteur}}(\text{Film})))$

où $q \equiv \rho_{\text{producteur}\to\text{réalisateur}}(\text{Produit}) \bowtie \text{Film}$

```
SELECT F2.acteur
FROM Film F1, Produit, Film F2
WHERE F1.titre=Produit.titre
AND   réalisateur=producteur
AND   F2.acteur=producteur
```


Exercices

Pour chacune des requêtes de la feuille de TD2, voir si elle est conjonctive ou non.

Si oui l'exprimer dans SCQ et SRPJ.