

Bases de données (organisation générale)

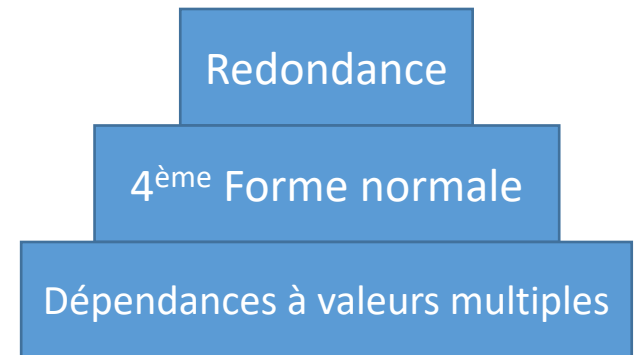
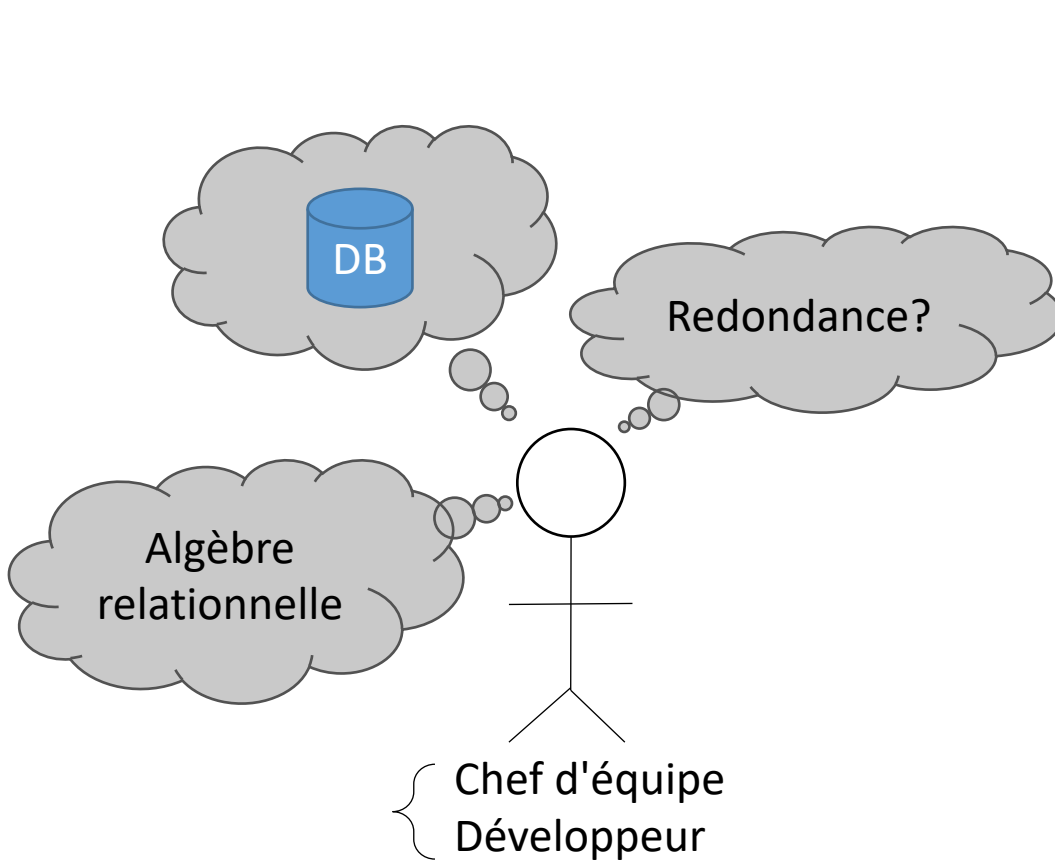
Répétition 5

La 4FN,

l'algèbre relationnelle étendue,

introduction à SQL

La quatrième forme normale: pourquoi?



Rappel sur les Dépendances à valeurs multiples

- Une relation r de schéma R satisfait une dépendance à valeurs multiples (DVM) $X \twoheadrightarrow Y$ si pour toute paire de tuples $t, s \in r$ tels que $t[X] = s[X]$, il existe deux tuples $u, v \in r$ tels que :

1. $u[X] = v[X] = t[X] = s[X]$
2. $u[Y] = t[Y]$ et $u[R - X - Y] = s[R - X - Y]$
3. $v[Y] = s[Y]$ et $v[R - X - Y] = t[R - X - Y]$

- Soumis aux règles suivantes :

- Si $Y \subseteq X$, alors $X \twoheadrightarrow Y$
- Si $X \twoheadrightarrow Y$, alors $X \twoheadrightarrow (R - XY)$
- Si $X \rightarrow Y$, alors $X \twoheadrightarrow Y$
- Si $X \twoheadrightarrow Y$ et $X \twoheadrightarrow Z$, alors $X \twoheadrightarrow YZ$
- Si $X \twoheadrightarrow Y$ et $X \twoheadrightarrow Z$, alors $X \twoheadrightarrow Y \cap Z$
- Si $X \twoheadrightarrow Y$ et $X \twoheadrightarrow Z$, alors $X \twoheadrightarrow Y \setminus Z$

(réflexivité)

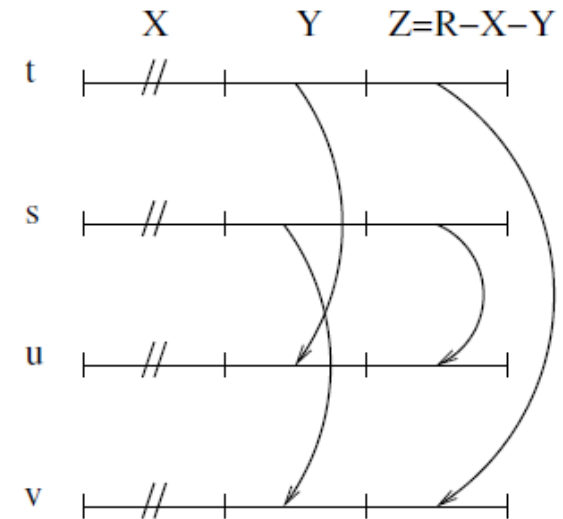
(complémentation pour DVM)

(reproduction)

(union pour DVM)

(intersection pour DVM)

(différence pour DVM)



Exercice 1

- Quelles sont les dépendances à valeurs multiples satisfaites par r ?

A	B	C	D
a_1	b_1	c_1	d_1
a_2	b_2	c_1	d_1
a_1	b_1	c_1	d_2
a_1	b_2	c_2	d_2
a_1	b_2	c_2	d_1

Exercice 1

- Quelles sont les dépendances à valeurs multiples satisfaites par r ?

A	B	C	D
a_1	b_1	c_1	d_1
a_2	b_2	c_1	d_1
a_1	b_1	c_1	d_2
a_1	b_2	c_2	d_2
a_1	b_2	c_2	d_1

- Recherche des dépendances fonctionnelles (reproduction)
- $A \rightarrow B$? Non, ligne 1 et 4
- $A \rightarrow D$? Non, ligne 1 et 3
- $B \rightarrow C$? Non, ligne 2 et 4
- $C \rightarrow A$? Non, ligne 1 et 2
- $C \rightarrow D$? Non, ligne 1 et 3
- $D \rightarrow B$? Non, ligne 1 et 2
- $A \rightarrow C$? Non, ligne 1 et 4
- $B \rightarrow A$? Non, ligne 2 et 4
- $B \rightarrow D$? Non, ligne 1 et 3
- $C \rightarrow B$? Non, ligne 1 et 2
- $D \rightarrow A$? Non, ligne 1 et 2
- $D \rightarrow C$? Non, ligne 3 et 4₅

Exercice 1

- Quelles sont les dépendances à valeurs multiples satisfaites par r ?

A	B	C	D
a_1	b_1	c_1	d_1
a_2	b_2	c_1	d_1
a_1	b_1	c_1	d_2
a_1	b_2	c_2	d_2
a_1	b_2	c_2	d_1

- Recherche des dépendances fonctionnelles (reproduction)

$AB \rightarrow C$? OK

$AB \rightarrow D$? Non, ligne 1 et 3

$AC \rightarrow B$? OK

$AC \rightarrow D$? Non, ligne 1 et 3

$AD \rightarrow B$? Non, ligne 1 et 5

$AD \rightarrow C$? Non, ligne 1 et 5

$BC \rightarrow A$? OK

$BC \rightarrow D$? Non, ligne 1 et 3

$BD \rightarrow A$? Non, ligne 2 et 5

$BD \rightarrow C$? Non, ligne 2 et 5

$CD \rightarrow A$? Non, ligne 1 et 2

$CD \rightarrow B$? Non, ligne 1 et 2

$ABC \rightarrow D$? Non, ligne 1 et 3

$ABD \rightarrow C$? OK

$ACD \rightarrow B$? OK

$BCD \rightarrow A$? OK

$\Rightarrow AB \twoheadrightarrow C, AC \twoheadrightarrow B, BC \twoheadrightarrow A, ABD \twoheadrightarrow C, ACD \twoheadrightarrow B, BCD \twoheadrightarrow A$

Mais les 3 dernières sont triviales, tout comme $ABC \twoheadrightarrow D$

Exercice 1

- Quelles sont les dépendances à valeurs multiples satisfaites par r ?

A	B	C	D
a_1	b_1	c_1	d_1
a_2	b_2	c_1	d_1
a_1	b_1	c_1	d_2
a_1	b_2	c_2	d_2
a_1	b_2	c_2	d_1

- Recherche directe des DVM
- $A \twoheadrightarrow B$? Non, ligne 3 et 4
- $A \twoheadrightarrow D$? OK
- $B \twoheadrightarrow C$? Non, ligne 2 et 4
- $B \twoheadrightarrow D$? Non, ligne 2 et 4
- $C \twoheadrightarrow A$? Non, ligne 2 et 3
- $C \twoheadrightarrow D$? Non, ligne 2 et 3
- $D \twoheadrightarrow B$? Non, ligne 1 et 2
- $A \twoheadrightarrow C$? Non, ligne 3 et 4
- $B \twoheadrightarrow A$? Non, ligne 2 et 4
- $C \twoheadrightarrow B$? Non, ligne 1 et 2
- $D \twoheadrightarrow A$? Non, ligne 1 et 2
- $D \twoheadrightarrow C$? Non, ligne 3 et 4

Exercice 1

- Quelles sont les dépendances à valeurs multiples satisfaites par r ?

A	B	C	D
a_1	b_1	c_1	d_1
a_2	b_2	c_1	d_1
a_1	b_1	c_1	d_2
a_1	b_2	c_2	d_2
a_1	b_2	c_2	d_1

- Recherche directe des DVMs

$AB \twoheadrightarrow C$? OK (DF)

$AB \twoheadrightarrow D$? OK (Compl.)

$AC \twoheadrightarrow B$? OK (DF)

$AC \twoheadrightarrow D$? OK (Compl.)

$AD \twoheadrightarrow B$? Non, ligne 1 et 5

$AD \twoheadrightarrow C$? Non, ligne 1 et 5

$BC \twoheadrightarrow A$? OK (DF)

$BC \twoheadrightarrow D$? OK (Compl.)

$BD \twoheadrightarrow A$? Non, ligne 2 et 5

$BD \twoheadrightarrow C$? Non, ligne 2 et 5

$CD \twoheadrightarrow A$? Non, ligne 1 et 2

$CD \twoheadrightarrow B$? Non, ligne 1 et 2

Exercice 1

- Note:
 - Remarquez que la décomposition n'est pas une des propriétés des DVM.
 - J'ai bien $A \twoheadrightarrow BC$ (par complémentation avec $A \twoheadrightarrow D$), mais je n'ai ni $A \twoheadrightarrow B$, ni $A \twoheadrightarrow C$.
 - Dans ce cas, pourquoi n'ai-je pas calculé, par exemple, $B \twoheadrightarrow AC$?
 - Parce que je sais que $B \twoheadrightarrow D$ n'est pas satisfait, donc, par complémentation, $B \twoheadrightarrow AC$ ne sera pas satisfait non plus.
- Quelles sont les dépendances à valeurs multiples satisfaites par r ?
 - Triviales:
 - $A \twoheadrightarrow A, B \twoheadrightarrow B, \dots$
 - $AB \twoheadrightarrow A, BC \twoheadrightarrow B, \dots$
 - $ABC \twoheadrightarrow D, ABD \twoheadrightarrow C, \dots, AB \twoheadrightarrow CD, \dots$
 - Non-triviales:
 - $A \twoheadrightarrow D, A \twoheadrightarrow BC,$
 - $AB \twoheadrightarrow C, AB \twoheadrightarrow D, AC \twoheadrightarrow B, AC \twoheadrightarrow D, BC \twoheadrightarrow D, BC \twoheadrightarrow A.$

Exercice 2

- Trouver une relation r de schéma $R(A, B, C)$ contenant au moins un tuple et où la dépendance $A \twoheadrightarrow B$ est satisfaite.
- Exemple de réponses acceptables:

A	B	C
a_1	b_1	c_1

A	B	C
a_1	b_1	c_1
a_2	b_1	c_1

A	B	C
a_1	b_1	c_1
a_1	b_1	c_2

A	B	C
a_1	b_1	c_1
a_1	b_2	c_1
a_1	b_1	c_2
a_1	b_2	c_2
a_2	b_3	c_4

Exercice 3

Soit un schéma $R(A, B, C, D)$ satisfaisant $F = \{A \twoheadrightarrow BC, C \rightarrow A, B \rightarrow CD\}$.

- i. Ce schéma est-il en 4FN ?
- ii. Sinon, appliquez l'algorithme de décomposition vu au cours.

Rappel: Un schéma de relation est en 4FN si pour toute dépendance $X \twoheadrightarrow Y$,

- soit $Y \subseteq X$,
- soit $XY = R$,
- soit X est une super-clé de R .

Exercice 3

Soit un schéma $R(A, B, C, D)$ satisfaisant

$$F = \{A \twoheadrightarrow BC, C \rightarrow A, B \rightarrow CD\}.$$

i. Ce schéma est-il en 4FN ?

Quelles sont mes DVMs?

1. Celles fournies dans F , ici $A \twoheadrightarrow BC$
2. Celles que j'obtiens par reproduction, ici $C \twoheadrightarrow A, B \twoheadrightarrow CD$ (mais aussi $B \twoheadrightarrow C, B \twoheadrightarrow D$ (décomposition DF), $B \twoheadrightarrow A$ (transitivité), $B \twoheadrightarrow CA$, (pseudo-union) et toutes les combinaisons possibles avec B dans la partie de gauche.)
3. Celles que j'obtiens par complémentation, ici $A \twoheadrightarrow D$ et $C \twoheadrightarrow BD$
4. Les triviales ($A \twoheadrightarrow BCD$ ou $BD \twoheadrightarrow B$) qui ne poseront pas de problèmes pour la 4FN.

J'ai donc

$$D = \{A \twoheadrightarrow BC, A \twoheadrightarrow D, B \twoheadrightarrow CD, \dots, B \twoheadrightarrow A, C \twoheadrightarrow A, C \twoheadrightarrow BD\}$$

Quelles sont mes clés?

Le calcul de la fermeture de F^+ m'apprend que la seule clé est B .

Exercice 3

Schéma $R = \{A, B, C, D\}$

$$F = \{A \twoheadrightarrow BC, C \rightarrow A, B \rightarrow CD\}$$

$$D = \{A \twoheadrightarrow BC, A \twoheadrightarrow D, C \twoheadrightarrow A, C \twoheadrightarrow BD, B \twoheadrightarrow CD, B \twoheadrightarrow A, \dots\}$$

- Ce schéma est-il en 4FN?
 - Non, car j'ai par exemple $C \twoheadrightarrow A$, mais $A \notin C$, $AC \neq R$ et C n'est pas une super-clé.
- Sinon, appliquez l'algorithme de décomposition vu au cours.
 - $C \twoheadrightarrow A$ pose problème en 4FN
 - Décomposition en $R_1(A, C)$ et $R_2(B, C, D)$
 - $R_1(A, C)$ avec $D = \{C \twoheadrightarrow A\}$ (OK)
 - $R_2(B, C, D)$ avec $D = \{C \twoheadrightarrow BD, B \twoheadrightarrow CD, \dots\}$ (OK)

– On décompose R en $R_1 = XY$ et $R_2 = X(R - XY)$ (sans perte vu le critère).

– On applique l'algorithme à : $R_1, \pi_{R_1}(D), R_2, \pi_{R_2}(D)$

Rappel sur les opérateurs relationnels étendus.

- Retrait des doublons

$$r : \begin{array}{c|c|c} A & B & C \\ \hline a & b & c \\ d & a & f \\ d & a & f \end{array}$$
$$\Delta(r) : \begin{array}{c|c|c} A & B & C \\ \hline a & b & c \\ d & a & f \end{array}$$

- Tri

$$r :$$

ID_ET	ID_COURS	COTE
1	INFO0009	18
3	INFO0009	8
3	INFO0016	7
2	INFO0016	17

$$\tau_{ID_ET, COTE}(r) :$$

ID_ET	ID_COURS	COTE
1	INFO0009	18
2	INFO0016	17
3	INFO0016	7
3	INFO0009	8

Rappel sur les opérateurs relationnels étendus.

- Opérations sur les groupements

r :

ID_ET	ID_COURS	COTE
1	<i>INFO0009</i>	18
3	<i>INFO0009</i>	8
3	<i>INFO0016</i>	7
2	<i>INFO0016</i>	17

$\gamma_{ID_COURS, AVG(COTE) \rightarrow M_COURS}(r)$:

ID_COURS	M_COURS
<i>INFO0009</i>	13
<i>INFO0016</i>	12

Exercice 4

La base de données d'un site de vente en ligne est constituée des relations suivantes.

- *client*(E-MAIL, NOM, PRENOM, ADRESSE) contenant la liste des clients ;
- *article*(NO_ARTICLE, LIBELLE, TYPE, PRIX_ACHAT) contenant la liste des articles ;
- *commande*(NO_COMMANDE, #E-MAIL, DATE_COMMANDE) contenant la liste des commandes ;
- *detail*(NO_COMMANDE, NO_ARTICLE, QUANTITE, PRIX_VENTE) contenant la liste des articles commandés dans le cadre d'une commande ;

Exprimer les requêtes suivantes en algèbre relationnelle étendue.

- Rechercher, pour chaque commande, le nombre d'articles différents commandés.
- Rechercher, pour chaque commande, le bénéfice réalisé.
- Rechercher le nom et le prénom de la personne qui a commandé le plus d'articles en une seule commande.
- Rechercher le nom et le prénom des clients qui ont commandé au moins une fois (et potentiellement dans des commandes différentes), l'ensemble des articles disponibles.

Exercice 4

- a) Rechercher, pour chaque commande, le nombre d'articles différents commandés.

L'information se trouve dans:

- *detail*(NO_COMMANDE, NO_ARTICLE, QUANTITE, PRIX_VENTE)

Je vais grouper l'information par numéro de commande, puis compter, pour chaque commande, le nombre de tuples (ce qui correspond au nombre d'articles différents commandés).

$\gamma_{NO_COMMANDE, COUNT(NO_ARTICLE) \rightarrow NB_ARTICLES}(detail)$

Exercice 4

b) Rechercher, pour chaque commande, le bénéfice réalisé.

L'information se trouve dans :

- *detail*(NO COMMANDE, NO ARTICLE, QUANTITE, PRIX_VENTE), où je récupère les articles commandés par commande, ainsi que leur prix de vente.
- *article*(NO ARTICLE, LIBELLE, TYPE, PRIX_ACHAT), où je récupère le prix d'achat pour un article donné.

$\gamma_{NO_COMMANDE, SUM(QUANTITE * (PRIX_VENTE - PRIX_ACHAT)) \rightarrow BENEFICE}$ (

$\pi_{NO_ARTICLE, PRIX_ACHAT}(article) \bowtie detail$)

Exercice 4

- c) Rechercher le nom et le prénom de la personne qui a commandé le plus d'articles en une seule commande.

L'information se trouve dans :

- *detail*(NO_COMMANDE, NO_ARTICLE, QUANTITE, PRIX_VENTE), où je récupère la quantité des articles commandés.
- *commande*(NO_COMMANDE, #E-MAIL, DATE_COMMANDE), pour obtenir l'adresse e-mail de la personne qui a passé la commande.
- *client*(E-MAIL, NOM, PRENOM, ADRESSE), pour obtenir le nom et le prénom de la personne.

$\pi_{NOM,PRENOM}(\$

$\tau_{NB_ARTICLES}(\gamma_{NO_COMMANDE, SUM(QUANTITE) \rightarrow NB_ARTICLES}(detail)$
 $\bowtie \pi_{NO_COMMANDE, EMAIL}(commande)$
 $\bowtie \pi_{E-MAIL, NOM, PRENOM}(client)))$

Le premier (ou dernier) tuple retourné est la solution.

Exercice 4

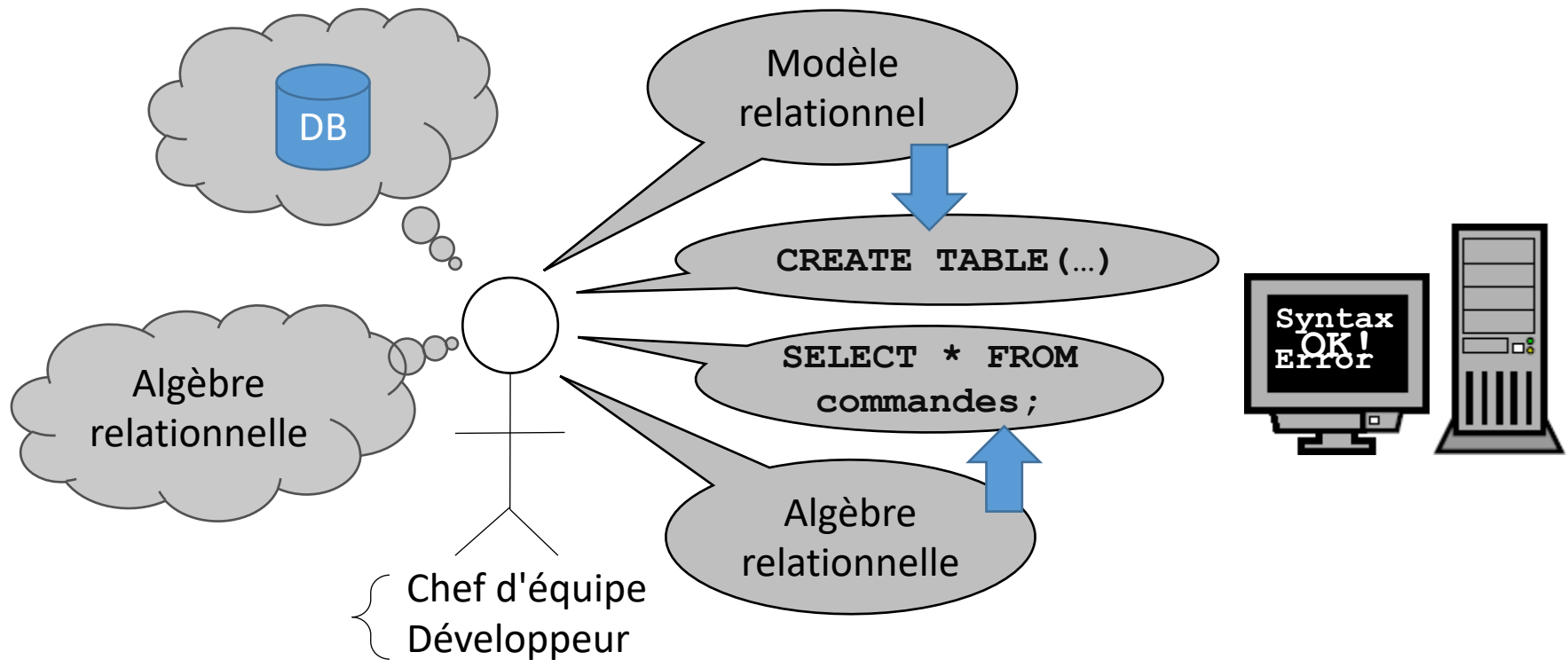
- d) Rechercher le nom et le prénom des clients qui ont commandé au moins une fois (et potentiellement dans des commandes différentes), l'ensemble des articles disponibles.

L'information se trouve dans:

- *detail*(NO_COMMANDE, NO_ARTICLE, QUANTITE, PRIX_VENTE), pour obtenir les articles liés à une commande.
- *article*(NO_ARTICLE, LIBELLE, TYPE, PRIX_ACHAT), qui contient la liste de tous les articles.
- *commande*(NO_COMMANDE, #E-MAIL, DATE_COMMANDE), pour récupérer l'adresse e-mail correspondant à une commande.
- *client*(E-MAIL, NOM, PRENOM, ADRESSE), pour récupérer le nom et le prénom

$$\pi_{NOM,PRENOM}(\pi_{E-MAIL}(\Delta(\pi_{E-MAIL,NO_ARTICLE}(\pi_{NO_COMMANDE,NO_ARTICLE}(detail) \bowtie \pi_{NO_COMMANDE,E-MAIL}(commande)))) \div \pi_{NO_ARTICLE}(article)) \bowtie \pi_{E-MAIL,NOM,PRENOM}(client))$$

Le langage SQL : Pourquoi?



Requêtes SQL

```
SELECT < [DISTINCT]? * |(Attr)+ >  
FROM < (table|(table [NATURAL|LEFT|RIGHT|OUTER]? JOIN table))+ >  
(WHERE < (condition)+ >)?  
(GROUP BY < (Attr)+ >  
(HAVING < (condition sur le groupe)+ >)?)?  
(ORDER BY < (Attr [ASC|DESC]?)+ >)?  
(LIMIT < nombre de tuples >)?  
;
```

1. Inclusion des tables (FROM)
2. Sélection (WHERE; optionnel)
3. Groupement (GROUP BY; optionnel)
4. Sélection sur les groupes (HAVING; optionnel; seulement après un GROUP BY)
5. Tri (ORDER BY; optionnel; peut entrer en conflit avec GROUP BY)
6. Limitation du nombre de tuples (LIMIT; optionnel)
7. Projection (SELECT)

Algèbre relationnelle (étendue) \Rightarrow SQL

- $commandes \Rightarrow$ SELECT * FROM commandes;
- $commandes \times articles \Rightarrow$
SELECT *
FROM commandes, articles;
- $commandes \bowtie articles \Rightarrow$
SELECT *
FROM commandes NATURAL JOIN articles;
- $\sigma_{PRIX > 10}(articles) \Rightarrow$
SELECT *
FROM articles
WHERE PRIX > 10;

Algèbre relationnelle (étendue) \Rightarrow SQL

- $\Pi_{NOM,PRENOM}(personne) \Rightarrow$
SELECT NOM, PRENOM
FROM personne;
- $\Delta(\Pi_{NOM,PRENOM}(personne)) \Rightarrow$
SELECT DISTINCT NOM, PRENOM
FROM personne;
- $\tau_{Age}(personne) \Rightarrow$
SELECT *
FROM personne
ORDER BY Age;

Algèbre relationnelle (étendue) \Rightarrow SQL

- $\gamma_{ID_CLIENT, COUNT(ID_ARTICLE) \rightarrow NB_ARTICLE}(articles) \Rightarrow$
SELECT ID_CLIENT, COUNT(ID_ARTICLE) AS NB_ARTICLE
FROM articles
GROUP BY ID_CLIENT;
- $\sigma_{COMPTE > 3}(\gamma_{ID, SUM(NB) \rightarrow COMPTE}(ventes)) \Rightarrow$
SELECT ID, SUM(NB) AS COMPTE
FROM ventes
GROUP BY ID
HAVING COMPTE > 3;

Exercice 5

- Ecrivez les requêtes suivantes en langage SQL:

a) $\Pi_{NOM,PRENOM}(\sigma_{Age>18}(Personne))$

b) $\Delta(\Pi_{NO_ARTICLE}(commande \bowtie article))$

c) $\Pi_{ISBN}(\sigma_{Code>10}(ouvrage) \bowtie exemplaire)$

d) $\tau_{BENEFICE}(\gamma_{NO_COMMANDE,SUM(VENTE-ACHAT)\rightarrow BENEFICE}(commande))$

e) $\sigma_{M_COURS>10}(\gamma_{ID_COURS,AVG(COTE)\rightarrow M_COURS}(\sigma_{ID_COURS \text{ commence } (notes))})$
par "INFO"

f) $\Pi_{NOM,PRENOM}(\sigma_{CP=4000}(personne) \bowtie \sigma_{Univ=ULiège}(colloque))$

Exercice 5

- Ecrivez les requêtes suivantes en langage SQL:

$$a) \underbrace{\Pi_{NOM,PRENOM}}_{\text{Projection}} \left(\underbrace{\sigma_{Age > 18}}_{\text{Filtrage}} \left(\underbrace{Personne}_{\text{Table}} \right) \right)$$

```
SELECT NOM, PRENOM  
FROM Personne  
WHERE Age > 18;
```

Exercice 5

- Ecrivez les requêtes suivantes en langage SQL:


$$b) \Delta(\underbrace{\Pi_{NO_ARTICLE}(commande \bowtie article)}_{\text{}})$$

```
SELECT DISTINCT NO_ARTICLE  
FROM commande NATURAL JOIN article  
;
```

Exercice 5

- Ecrivez les requêtes suivantes en langage SQL:

c) $\Pi_{ISBN}(\sigma_{Code > 10}(\text{ouvrage}) \bowtie \text{exemplaire})$




```
SELECT ISBN
FROM   exemplaire NATURAL JOIN
      (SELECT *
       FROM ouvrage
       WHERE Code > 10) AS t1
;
```

Exercice 5

- Ecrivez les requêtes suivantes en langage SQL:

d) $\tau_{BENEFICE}(\gamma_{NO_COMMANDE}, \text{SUM}(VENTE-ACHAT) \rightarrow BENEFICE(\text{commande}))$



```
SELECT NO_COMMANDE, SUM(VENTE-ACHAT) AS BENEFICE
FROM commande
GROUP BY NO_COMMANDE
ORDER BY BENEFICE
;
```

Exercice 5

- Ecrivez les requêtes suivantes en langage SQL:


e) $\sigma_{M_COURS > 10}(\gamma_{ID_COURS, AVG(COTE) \rightarrow M_COURS}(\sigma_{ID_COURS \text{ commence par "INFO"}}(notes)))$

```
SELECT  ID_COURS, AVG(COTE) AS M_COURS
FROM    notes
WHERE   ID_COURS LIKE "INFO*"
GROUP  BY ID_COURS
HAVING M_COURS > 10
;
```

Exercice 5

- Ecrivez les requêtes suivantes en langage SQL:

f) $\Pi_{NOM,PRENOM}(\sigma_{CP=4000}(personne) \bowtie \sigma_{Univ=ULiège}(colloque))$



```
SELECT NOM, PRENOM
FROM
    (SELECT *
     FROM personne
     WHERE CP = 4000) AS t1
NATURAL JOIN
    (SELECT *
     FROM colloque
     WHERE Univ LIKE "ULiège") AS t2
;
```