

Conception de Bases de Données

Dépendances fonctionnelles

Verónica Peralta

2021-2022

Plan

- ◆ **Règles informelles de conception BD**
- ◆ **Contraintes d'intégrité**
- ◆ **Dépendances fonctionnelles (DF)**
 - Définition de DF
 - Obtention de DF
- ◆ **Théorie des dépendances fonctionnelles**
 - Dérivation de DF
 - Fermeture d'un ensemble de d'attributs
 - Couverture minimale
- ◆ **Clés**



Critères de qualité de conception BD

- ◆ **Pendant la conception BD nous spécifions le schéma de la BD**
 - Comme toute activité de conception, la conception BD est une tâche subjective
- ◆ **4 règles informelles pour obtenir une BD de bonne qualité :**
 - Considérer la sémantique des attributs
 - Réduire les valeurs redondantes dans les tuples
 - Réduire les valeurs nulles dans les tuples
 - Ne pas générer des tuples erronés

Sémantique des attributs

videoid	acqDate	title	genre	length	rating
90987	05/02/2007	Star wars	Action	121	8.8
123	29/04/2001	Nikita	Drama	118	7.5
90988	25/03/2007	Star wars	Action	121	8.8
90989	05/04/2008	Star wars	Action	121	8.8
145	31/12/2005	Lady and the tramp	Comics	93	7.4
8034	05/04/2008	Lady and the tramp	Comics	93	7.4

- ◆ **Problème : Mélange d'attributs de sémantique différente**
 - Attributs concernant la vidéo avec des attributs d'un film.
- ◆ **Conséquence : Difficultés pour exprimer des requêtes**
 - Voir requêtes du TP 1

Sémantique des attributs

cup	winner	runnerUp	thirdPlace	fourthPlace	goalScore	host
1994	Brazil	Italy	Sweden	Bulgaria	6 - Oleg Salenko (Russia), Hristo Stoitchkov (Bulgaria)	United States
1998	France	Brazil	Croatia	Netherlands	6 - Davor Šuker (Croatia)	France
2002	Brazil	Germany	Turkey	South Korea	8 - Ronaldo (Brazil)	Japan, South Korea
2006	Italy	France	Germany	Portugal	5 - Miroslav Klose (Germany)	Germany
2010	Spain	Netherlands	Germany	Uruguay	5 - David Villa (Spain), Wesley Sneijder (Netherlands), Thomas Muller (Germany), Diego Forlan (Uruguay)	South Africa
2014	Germany	Argentina	Netherlands	Brazil	6 - James Rodriguez (Colombia)	Brazil
2018						Russia
2022						Qatar

- ◆ **Problème : Mélange de valeurs dans un même attribut**
- ◆ **Requêtes très complexes**

Quels joueurs ont été élus meilleurs buteurs dans plusieurs coupes du monde ?

Sémantique des attributs

◆ Règle 1 :

- Concevoir un schéma de relation de façon qu'il soit facile d'expliquer sa signification.
- Ne pas combiner des attributs de plusieurs objets dans une même relation.
- Ne pas combiner plusieurs valeurs dans un seul attribut.

◆ Intuition de la solution :

- Décomposer en plusieurs relations et plusieurs attributs
- Exemples :
 - VIDEOS (videoid, acqDate, title)
 - MOVIES (title, genre, length, rating)
 - A_TOP (cup, winner, runnerUp, thirdPlace, fourthPlace, goalScore, goalScorerName, goalScorerTeam)
 - A_CUPS (cup, host)

Exercice : Critiquer la relation

promo	date	creneau	idCours	nomCours	salle	enseignant
L3	25/10/2012	14:00	UE 5-4	Admin BD (cm)	333	Robichon
L3	25/10/2012	16:15	UE 5-4	Admin BD (cm)	333	Robichon
L3	24/10/2012	16:15	UE 5-1	Anglais (td)	333	Andre
L3, M1	25/10/2012	08:15	UE 6-5	IHM (cm)	Amphi 2	Antoine
L3	25/10/2012	10:30	UE 6-5	IHM (td)	202	Antoine
L3	26/10/2012	08:15	UE 5-2	Conception BD (tp)	103	Peralta
M1	22/10/2012	08:15	UE 1-6-1	Anglais (td)	109	Charpentier
M1	23/10/2012	08:15	UE 1-3-2	Optimisation BD (cm)	107	Marcel
M1	23/10/2012	10:30	UE 1-3-2	Optimisation BD (tp)	345	Khalil
M1, M2	25/10/2012	10:30	UE 1-4-1	Modélisation MD (cm)	112	Peralta

Chaque tuple de la relation représente l'affectation de ressources à une séance de cours

Réduction de valeurs redondantes

videold	acqDate	title	genre	length	rating
90987	05/02/2007	Star wars	Action	121	8.8
123	29/04/2001	Nikita	Drama	118	7.5
90988	25/03/2007	Star wars	Action	121	8.8
90989	05/04/2008	Star wars	Action	121	8.8
145	31/12/2005	Lady and the tramp	Comics	93	7.4
8034	05/04/2008	Lady and the tramp	Comics	93	7.4

◆ Problème : redondance

- À chaque fois qu'un film (title) apparaît, les valeurs pour le genre, length et rating se répètent.

◆ Observation :

- Le problème n'est pas que « Star wars » apparait 3 fois
 - S'il y a 3 vidéos de « Star wars » il faut le dire quelque part (c'est nécessaire)
- mais qu'il faut répéter aussi le genre, length et rating (pas nécessaire)

Réduction de valeurs redondantes

videold	acqDate	title	genre	length	rating
90987	05/02/2007	Star wars	Action	121	8.8
123	29/04/2001	Nikita	Drama	118	7.5
90988	25/03/2007	Star wars	Action	121	8.8
90989	05/04/2008	Star wars	Action	121	8.8
145	31/12/2005	Lady and the tramp	Comics	93	7.4
8034	05/04/2008	Lady and the tramp	Comics	93	7.4

♦ **La redondance cause des anomalies de mise à jour**

- **Modification** de la longueur du vidéo 90987 à **122** minutes
- **Insertion** du tuple <95673, 01/10/2011, Star wars, **Comedy**, 121, 8.8>
- **Suppression** du vidéo 123

Inconsistance
ou besoin de
plusieurs màj

Inconsistance

Perte des informations sur le film Nikita

Réduction de valeurs redondantes

◆ Règle 2 :

- Concevoir les schémas de relation de façon qu'il n'y ait pas d'anomalies d'insertion, d'élimination ou de modification dans les relations.
- S'il y a des anomalies, il faut les marquer clairement afin que les programmes de mise à jour de la BD puissent opérer correctement.

◆ Intuition de la solution : décomposer la relation

videold	acqDate	title
90987	05/02/2007	Star wars
123	29/04/2001	Nikita
90988	25/03/2007	Star wars
90989	05/04/2008	Star wars
145	31/12/2005	Lady and the tramp
8034	05/04/2008	Lady and the tramp

title	genre	length	rating
Star wars	Action	121	8.8
Nikita	Drama	118	7.5
Lady and the tramp	Comics	93	7.4

Exercice : Critiquer la relation

CONTRATS (Id, Date, IdClient, Nom, Prenom, Bonus, Malus, IdVehicule, Puissance, Marque, Type, Couleur, Risques)

Id	Date	IdC	Nom	Prenom	Bonus	Malus	IdV	Puiss.	Marque	Type	Couleur	Risques
1	02/01	022	Lopés	Daniel	60	0	055	5	Citroen	A	Rouge	Full
2	02/01	022	Lopés	Daniel	60	0	535	2	Citroen	A	Bleu	Full
3	02/01	015	Ba	Pierre	40	10	810	3	Peugeot	B	Gris	Tiers
4	02/01	134	Leroy	Anne	50	0	122	5	Renault	C	Blanc	Tiers
5	03/01	015	Ba	Pierre	40	10	910	3	Citroen	D	Rouge	Full
6	03/01	015	Ba	Pierre	40	10	811	3	Renault	D	Gris	Tiers
7	03/01	111	Li	Tao	40	20	410	5	Citroen	C	Gris	Tiers
8	04/01	246	Lopés	Anne	60	10	160	2	Peugeot	A	Noir	Tiers
9	31/01	246	Lopés	Anne	60	10	190	6	Mercedes	A	Vert	Tiers
10	31/01	022	Lopés	Daniel	60	0	122	5	Renault	C	Blanc	Full

Un client a un bonus et un malus. Il passe un contrat pour chacun de ses véhicules, et pour certains risques couverts.

Réduction de valeurs nulles

♦ Problèmes des valeurs nulles :

- Gaspillage d'espace
- Difficulté de compréhension, multiples interprétations
- Problème de calcul de fonctions agrégés (count, sum, avg)

studentId	telephone	comments
122	0612345678	dyslexia
123	0623456789	
124		
...		

Réduction de valeurs nulles

◆ Règle 3 :

- Dans la mesure du possible, éviter d'inclure dans une relation des attributs que puissent avoir des valeurs nulles.
- Si ce n'est pas possible, s'assurer de que les valeurs nulles existent seulement dans des cas exceptionnels (pas dans la plupart des tuples)

◆ Intuition de la solution :

- Stocker à part les attributs ayant des valeurs nulles

studentId	telephone
122	0612345678
123	0623456789
124	
...	


studentId	comments
122	dyslexia

Pas de génération de tuples erronés

videold	acqDate		acqDate	title	genre	length	rating
90987	05/02/2007		05/02/2007	Star wars	Action	121	8.8
123	29/04/2001		29/04/2001	Nikita	Drama	118	7.5
90988	25/03/2007		25/03/2007	Star wars	Action	121	8.8
90989	05/04/2008		05/04/2008	Star wars	Action	121	8.8
145	31/12/2005		31/12/2005	Lady and the tramp	Comics	93	7.4
8034	05/04/2008		05/04/2008	Lady and the tramp	Comics	93	7.4

- ◆ **Problème : une mauvaise décomposition peut introduire des tuples erronés lors des jointures**

Pas de génération de tuples erronés

videoid	acqDate		acqDate	title	genre	length	rating
90987	05/02/2007		05/02/2007	Star wars	Action	121	8.8
123	29/04/2001		29/04/2001	Nikita	Drama	118	7.5
90988	25/03/2007		25/03/2007	Star wars	Action	121	8.8
90989	05/04/2008		05/04/2008	Star wars	Action	121	8.8
145	31/12/2005		31/12/2005	Lady and the tramp	Comics	93	7.4
8034	05/04/2008		05/04/2008	Lady and the tramp	Comics	93	7.4

videoid	acqDate	title	genre	length	rating
90987	05/02/2007	Star wars	Action	121	8.8
123	29/04/2001	Nikita	Drama	118	7.5
90988	25/03/2007	Star wars	Action	121	8.8
90989	05/04/2008	Star wars	Action	121	8.8
90989	05/04/2008	Lady and the tramp	Comics	93	7.4
145	31/12/2005	Lady and the tramp	Comics	93	7.4
8034	05/04/2008	Star wars	Action	121	8.8
8034	05/04/2008	Lady and the tramp	Comics	93	7.4

Pas de génération de tuples erronés

◆ Règle 4 :

- Concevoir les schémas de relation de façon qu'ils puissent se joindre par des conditions d'égalité sur les attributs clé afin d'éviter la génération de tuples erronés

◆ Corolaire :

- Il faut concevoir plusieurs relations, mais ce n'est pas de n'importe quelle manière.
- L'intuition peut nous tromper.

Exercice : Critiquer la décomposition

CONTRATS (Id, Date, IdClient, Nom, Prenom, Bonus, Malus, IdVehicule, Puissance, Marque, Type, Couleur, Risques)

Id	Date	IdC	Nom	Prenom	Bonus	Malus	Type
1	02/01	022	Lopés	Daniel	60	0	A
2	02/01	022	Lopés	Daniel	60	0	A
3	02/01	015	Ba	Pierre	40	10	B
4	02/01	134	Leroy	Anne	50	0	C
5	03/01	015	Ba	Pierre	40	10	D
6	03/01	015	Ba	Pierre	40	10	D
7	03/01	111	Li	Tao	40	20	C
8	04/01	246	Lopés	Anne	60	10	A
9	31/01	246	Lopés	Anne	60	10	A
10	31/01	022	Lopés	Daniel	60	0	C

IdV	Puiss.	Marque	Type	Couleur	Risques
055	5	Citroen	A	Rouge	Full
535	2	Citroen	A	Bleu	Full
810	3	Peugeot	B	Gris	Tiers
122	5	Renault	C	Blanc	Tiers
810	3	Citroen	D	Rouge	Full
811	3	Renault	D	Gris	Tiers
410	5	Citroen	C	Gris	Tiers
160	2	Peugeot	A	Gris	Tiers
190	6	Mercedes	A	Vert	Tiers
573	5	Renault	C	Noir	Full

Exercice : Critiquer la décomposition

CONTRATS (Id, Date, IdClient, Nom, Prenom, Bonus, Malus, IdVehicule, Puissance, Marque, Type, Couleur, Risques)

Id	IdC	Nom	Prenom	Bonus	Malus	Id	Date	IdV	Puiss.	Marque	Type	Couleur	Risques
1	022	Lopés	Daniel	60	0	1	02/01	055	5	Citroen	A	Rouge	Full
2	022	Lopés	Daniel	60	0	2	02/01	535	2	Citroen	A	Bleu	Full
3	015	Ba	Pierre	40	10	3	02/01	810	3	Peugeot	B	Gris	Tiers
4	134	Leroy	Anne	50	0	4	02/01	122	5	Renault	C	Blanc	Tiers
5	015	Ba	Pierre	40	10	5	03/01	810	3	Citroen	D	Rouge	Full
6	015	Ba	Pierre	40	10	6	03/01	811	3	Renault	D	Gris	Tiers
7	111	Li	Tao	40	20	7	03/01	410	5	Citroen	C	Gris	Tiers
8	246	Lopés	Anne	60	10	8	04/01	160	2	Peugeot	A	Gris	Tiers
9	246	Lopés	Anne	60	10	9	31/01	190	6	Mercedes	A	Vert	Tiers
10	022	Lopés	Daniel	60	0	10	31/01	573	5	Renault	C	Noir	Full

Exercice : Critiquer la décomposition

CONTRATS (Id, Date, IdClient, Nom, Prenom, Bonus, Malus, IdVehicule, Puissance, Marque, Type, Couleur, Risques)

IdC	Nom	Prenom	Bonus	Malus	Id	Date	IdV	Puiss.	Marque	Type	Couleur	Risques
022	Lopés	Daniel	60	0	1	02/01	055	5	Citroen	A	Rouge	Full
015	Ba	Pierre	40	10	2	02/01	535	2	Citroen	A	Bleu	Full
134	Leroy	Anne	50	0	3	02/01	810	3	Peugeot	B	Gris	Tiers
111	Li	Tao	40	20	4	02/01	122	5	Renault	C	Blanc	Tiers
246	Lopés	Anne	60	10	5	03/01	810	3	Citroen	D	Rouge	Full
					6	03/01	811	3	Renault	D	Gris	Tiers
					7	03/01	410	5	Citroen	C	Gris	Tiers
					8	04/01	160	2	Peugeot	A	Gris	Tiers
					9	31/01	190	6	Mercedes	A	Vert	Tiers
					10	31/01	573	5	Renault	C	Noir	Full

Résumé

◆ Problèmes à éviter

- Anomalies d'insertion, d'élimination ou de modification de tuples à cause de la *redondance*
- Gaspillage d'espace et difficulté d'opération à cause de *valeurs nulles*
- Génération de données erronées à cause de mauvaises décompositions.

◆ Dans la suite nous verrons...

- La **formalisation des contraintes** qui visent à détecter et éviter ce type de problèmes

Plan

- ◆ Règles informelles de conception BD
- ◆ **Contraintes d'intégrité**
- ◆ **Dépendances fonctionnelles (DF)**
 - Définition de DF
 - Obtention de DF
- ◆ **Théorie des dépendances fonctionnelles**
 - Dérivation de DF
 - Fermeture d'un ensemble de d'attributs
 - Couverture minimale
- ◆ **Clés**



Problème de la redondance

videold	acqDate	title	genre	length	rating
90987	05/02/2007	Star wars	Action	121	8.8
123	29/04/2001	Nikita	Drama	118	7.5
90988	25/03/2007	Star wars	Action	121	8.8
90989	05/04/2008	Star wars	Action	121	8.8
145	31/12/2005	Lady and the tramp	Comics	93	7.4
8034	05/04/2008	Lady and the tramp	Comics	93	7.4

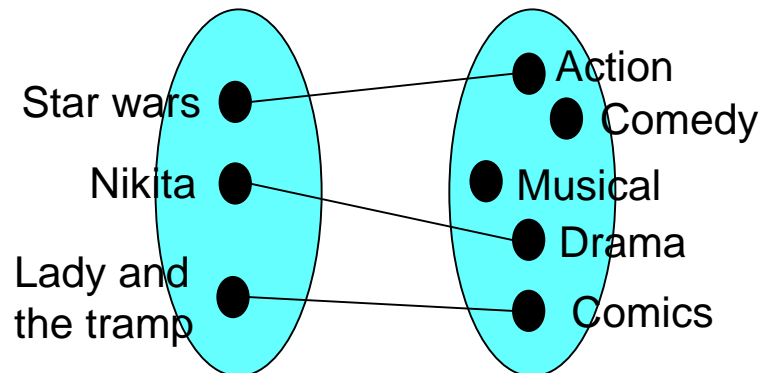
◆ Pour chaque vidéo d'un même film:

- Les valeurs de *genre* (idem pour *length* et *rating*) se répètent : **redondance**
- On pourrait se retrouver avec des valeurs différents de *genre*: **anomalies de mise à jour**
- Or, on sait que chaque film a un seul *genre*: **contrainte métier**

Besoin de contraintes

videoid	acqDate	title	genre	length	rating
90987	05/02/2007	Star wars	Action	121	8.8
123	29/04/2001	Nikita	Drama	118	7.5
90988	25/03/2007	Star wars	Action	121	8.8
90989	05/04/2008	Star wars	Action	121	8.8
145	31/12/2005	Lady and the tramp	Comics	93	7.4
8034	05/04/2008	Lady and the tramp	Comics	93	7.4

- ◆ Il y a une relation de **type fonctionnel** entre title et genre
 - Pour chaque valeur de title on doit avoir une seule valeur de genre



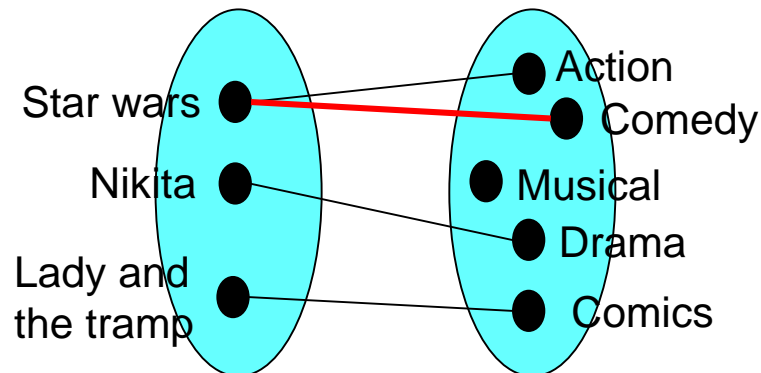
**C'est une
contrainte.**

On désire contrôler
qu'elle soit toujours
respectée

Besoin de contraintes

videoid	acqDate	title	genre	length	rating
90987	05/02/2007	Star wars	Action	121	8.8
123	29/04/2001	Nikita	Drama	118	7.5
90988	25/03/2007	Star wars	Action	121	8.8
90989	05/04/2008	Star wars	Comedy	121	8.8
145	31/12/2005	Lady and the tramp	Comics	93	7.4
8034	05/04/2008	Lady and the tramp	Comics	93	7.4

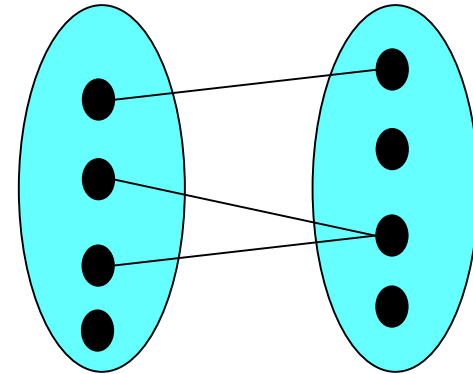
- ◆ Une **anomalie** représente une violation à la contrainte
 - Pour une valeur de title on a plusieurs valeurs de genre



Rappel : De la théorie des ensembles

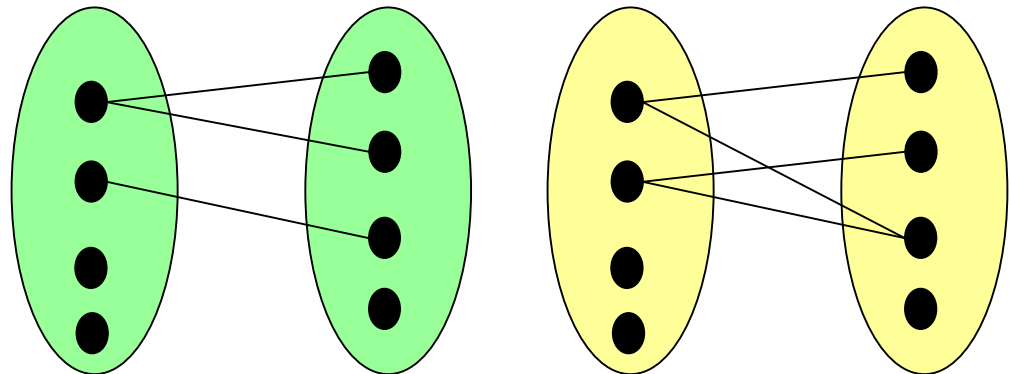
◆ Relations fonctionnelles :

- Exemples :
 - Etudiant – email @etu.univ-tours.fr
 - Voiture – matricule
 - Personne – résidence principale



◆ Relations pas fonctionnelles :

- Exemples :
 - Etudiant – téléphone
 - Etudiant – cours suivi
 - Film – acteur



Besoin de contraintes

- ♦ **On désire que le SGBD contrôle les contraintes :**
 - Qu'il empêche les mises à jour qui violent les contraintes,
 - évitant ainsi les anomalies d'insertion, suppression et modification
 - et laissant la BD dans un état cohérent
- ♦ **Pour cela, il faut, dès la conception, identifier toutes les contraintes à respecter**
- ♦ **Mais... quelles contraintes ?**

Contraintes d'intégrité : lesquelles ?

- ◆ Les **dépendances fonctionnelles** sont les contraintes les plus importantes d'un projet BD relationnelle
 - Une DF exprime une relation de type fonctionnelle entre deux ensembles d'attributs
 - Pour chaque valeur de *title* on doit avoir une seule valeur de *genre*
 - Dans d'autres mots, étant donnée un titre, on peut déterminer son genre
 - C'est comme une fonction : $genre = F(title)$
 - Il y a une dépendance entre les attributs :
 - *genre* dépend de *title*
 - *title* détermine *genre*
 - On la note : **title → genre**
- ◆ D'autres contraintes existent

Contraintes d'intégrité : lesquelles ?

◆ Exemples:

- Pour exprimer qu'une personne est associée à un et un seul siège et qu'un siège est associé à une et une seule personne:
 - $\text{personNb} \rightarrow \text{placeNb}$
 - $\text{placeNb} \rightarrow \text{personNb}$
- Pour exprimer qu'un film est identifié par son titre et année de sortie
 - $\text{title, year} \rightarrow \text{director, type, lenght, genre}$
- Pour exprimer qu'une personne est identifié par son ssn
 - $\text{ssn} \rightarrow \text{firstName, lastName, address}$

Plan

- ◆ Règles informelles de conception BD
- ◆ Contraintes d'intégrité
- ◆ **Dépendances fonctionnelles (DF)**
 - Définition de DF
 - Obtention de DF
- ◆ **Théorie des dépendances fonctionnelles**
 - Dérivation de DF
 - Fermeture d'un ensemble de d'attributs
 - Couverture minimale
- ◆ **Clés**



Dépendances fonctionnelles

◆ Définition :

- Soit une relation R
- Soit deux ensembles d'attributs $X, Y \subseteq \text{sort}(R)$

Il existe une dépendance fonctionnelle

$$X \rightarrow Y$$

si pour chaque paire de tuples t et u de toute instance $I(R)$,

$$\text{si } t(X) = u(X) \text{ alors } t(Y) = u(Y)$$

- ◆ Nous disons que X détermine fonctionnellement Y ou que Y dépend de X

Dépendances fonctionnelles

◆ Exemple

videold	acqDate	title	genre	length	rating
90987	05/02/2007	Star wars	Action	121	8.8
123	29/04/2001	Nikita	Drama	118	7.5
90988	25/03/2007	Star wars	Action	121	8.8
90989	05/04/2008	Star wars	Action	121	8.8
145	31/12/2005	Lady and the tramp	Comics	93	7.4
8034	05/04/2008	Lady and the tramp	Comics	93	7.4

- Cette instance satisfait la dépendance
 - title → genre, length, rating
- mais ne satisfait pas la dépendance
 - title → videold

Dépendances fonctionnelles

◆ Notation :

Soit un schéma $R(A,B,C,D,E,G)$

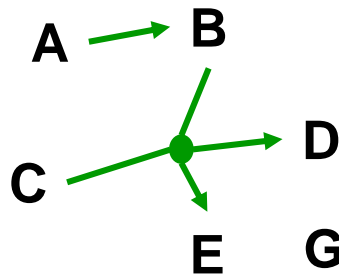
On peut exprimer un ensemble de DF de plusieurs façons :

- Notation d'ensemble :
 - $F = \{A \rightarrow B, BC \rightarrow DE\}$

- Sur le schéma :



- Graphiquement :



Exercice : Indiquer si l'instance satisfait les DF

EMPRUNTS (CodeEmprunt, DateEmprunt, IdAbonne, Nom, Prenom, Adresse, Cauton, MaxEmprunts, IdLivre, Titre, Editeur, Auteur)

Code	Date	IdA	Nom	Prenom	Adresse	Caut	MaxE	IdLivre	Titre	Editeur	Auteur
1	02/01	022	Lopés	Daniel	13 r Papin	100	5	431	Info App	A	Chen
2	02/01	022	Lopés	Daniel	13 r Papin	100	5	126	C++	A	Peters
3	02/01	015	Ba	Pierre	3 pl Jaurès	200	10	852	Espagne	B	Michel
4	02/01	134	Leroy	Anne	51 av Maunoury	40	2	668	Hist France	C	A.J.
5	03/01	015	Ba	Pierre	3 pl Jaurès	200	10	452	Barcelone	D	Torres
5	03/01	015	Ba	Pierre	3 pl Jaurès	200	10	452	Barcelone	D	Castillo
6	03/01	111	Li	Tao	7 r Monin	200	10	567	Algèbre	C	Smith
7	04/01	246	Lopés	Anne	13 r Papin	200	10	127	C++	A	Peters
8	04/01	997	Dupond	Jean	34 r Fénelon	20	1	863	Paris	B	Michel
101	31/01	246	Lopés	Anne	13 r Papin	200	10	127	C++	A	Peters
102	31/01	022	Lopés	Daniel	13 r Papin	100	5	567	Algèbre	C	Smith

Exercice : Indiquer si l'instance satisfait les DF

EMPRUNTS (CodeEmprunt, DateEmprunt, IdAbonne, Nom, Prenom, Adresse, Caution, MaxEmprunts, IdLivre, Titre, Editeur, Auteur)

- IdAbonne \rightarrow Nom, Prenom, Adresse
- Nom \rightarrow IdAbonne
- Nom, Prenom \rightarrow IdAbonne
- Adresse \rightarrow IdAbonne
- Nom, Adresse \rightarrow IdAbonne
- IdAbonne \rightarrow Caution
- Caution \rightarrow MaxEmprunts
- IdLivre \rightarrow Titre, Editeur
- Titre \rightarrow IdLivre
- IdLivre \rightarrow Auteur
- Auteur \rightarrow IdLivre
- IdAbonne \rightarrow CodeEmprunt
- CodeEmprunt \rightarrow IdAbonne
- IdLivre \rightarrow CodeEmprunt
- CodeEmprunt \rightarrow IdLivre
- CodeEmprunt \rightarrow DateEmprunt
- IdAbonne, IdLivre \rightarrow CodeEmprunt

Obtention de DF

♦ Observation importante :

- Les DF sont des propriétés de la sémantique des attributs
- Elles s'obtiennent de l'étude du monde réel pas de l'observation d'une instance

♦ Exemple :

- director → title ?

title	director	actor
The cameraman	Buster Keaton	Buster Keaton
Rear window	Hitchcock	James Steward
Rear window	Hitchcock	Grace Kelly
To be or not to be	Lubitsch	Carole Lombard
To be or not to be	Lubitsch	Jack Benny

Exemple : Trouver des attributs et DF

Une entreprise de location de véhicules souhaite concevoir une BD. On s'intéresse aux **véhicules**, identifiés par leurs **numéros d'immatriculation**, et à leur **marques**, **couleurs**, **modèles** et **années**. On s'intéresse également aux **clients**, identifiés par leurs **documents d'identité**, et à leurs **noms**, **adresses** et **téléphones**. Un **contrat de location** est identifié par un **numéro de contrat** ; il est réalisé à une **date**, et concerne un **client** et un **véhicule** ; on enregistre aussi la **période de location** (en jours) et le **prix**. L'entreprise ne loue pas le même véhicule au même client à une même date.

Exemple : Trouver des attributs et DF

Une entreprise de location de véhicules souhaite concevoir une BD. On s'intéresse aux **véhicules**, identifiés par leurs **numéros d'immatriculation**, et à leur **marques**, **couleurs**, **modèles** et **années**. On s'intéresse également aux **clients**, identifiés par leurs **documents d'identité**, et à leurs **noms**, **adresses** et **téléphones**. Un **contrat de location** est identifié par un **numéro de contrat** ; il est réalisé à une **date**, et concerne un **client** et un **véhicule** ; on enregistre aussi la **période de location** (en jours) et le **prix**. L'entreprise ne loue pas le même véhicule au même client à une même date.

- NumMat → marque, couleur, modele, annee
- DocId → nom, adresse, telephone
- NumContrat → date, DocId, NumMat, periode, prix
- Date, DocId, NumMat → NumContrat

Exercice : Déduire des DF

ENSEIGNEMENT (cours, jour, salle, enseignant, salaire, prime)

◆ Informations :

- Chaque cours n'a qu'un enseignant.
- Il peut avoir lieu plusieurs jours par semaine
- mais toujours dans la même salle.
- Chaque enseignant a un salaire de base
- et touche une prime pour chaque cours enseigné
- Le salaire dépend de l'expérience de l'enseignant
- mais les primes dépendent aussi des cours.

Exercice : Déduire des DF

RECETTES (id-recette, nom, type, livre, page, collection, editeur)

◆ Informations :

BD de recettes de cuisine d'un cuisinier amateur

- Chaque recette a un identifiant unique, un nom et un type
- Elle se trouve dans un livre, dans une seule page
- Chaque page d'un livre contient une seule recette
- Chaque livre appartient à une collection et a un éditeur
- Chaque collection est éditée par un même éditeur

Contrôle de DF

- ♦ **Les DF sont des contraintes que nous voulons imposer à la BD**
 - Base cohérente = BD qui respecte les contraintes d'intégrité
 - Toute modification qui viole la cohérence de la BD doit être **interdite**
 - Besoin de vérification pendant les mises à jour
 - Cette vérification peut être très chère
- ♦ **Exemple :**

Customer (accountId, lastName, firstName, street, city, state, zipCode)

zipCode → city, state

street, city, state → zipCode

 - Des modifications sur city, state ou zipCode doivent être **vérifiées** pour ne pas violer les contraintes !
- ♦ **Grand enjeu dans la conception BD :**
 - Trouver un moyen de rendre le contrôle de DF efficace et moins chère

Contrôle de DF

videold	acqDate	title	genre	length	rating
90987	05/02/2007	Star wars	Action	121	8.8
123	29/04/2001	Nikita	Drama	118	7.5
90988	25/03/2007	Star wars	Action	121	8.8
90989	05/04/2008	Star wars	Action	121	8.8
145	31/12/2005	Lady and the tramp	Comics	93	7.4
8034	05/04/2008	Lady and the tramp	Comics	93	7.4

♦ Observation : des DF peuvent être déduites à partir d'autres

Ex. videold → title
title → genre ➔ videold → genre

Comment savoir si une DF est valide ?

Comment trouver un ensemble « minimale » de DF ?

Résumé

- ◆ **Les DF représentent des contraintes que nous voulons imposer à la BD**
 - Elles s'obtiennent de l'étude du monde réel
- ◆ **Les DF doivent être contrôlés à chaque mise à jour**
 - Ce contrôle peut être très coûteux
 - Certains DF peuvent être déduites des autres (contrôle inutile)
- ◆ **Notre but :**
 - **Analyser** le monde réel pour trouver les DF le plus significatives
 - **Dériver** mécaniquement les autres
 - **Obtenir un ensemble minimal** de DF à contrôler

Plan

- ◆ Règles informelles de conception BD
- ◆ Contraintes d'intégrité
- ◆ Dépendances fonctionnelles (DF)
 - Définition de DF
 - Obtention de DF
- ◆ **Théorie des dépendances fonctionnelles**
 - Dérivation de DF
 - Fermeture d'un ensemble de d'attributs
 - Couverture minimale
- ◆ **Clés**



Inférence de DF

- ◆ Nous avons vu que certains DF peuvent être déduites à partir d'autres
- ◆ Question :
 - Etant donné un ensemble de DF déduites de l'étude d'un problème
 - Comment savoir si une DF $X \rightarrow A$ est valide aussi ?

Inférence de DF

◆ Implication sémantique :

- Soit F un ensemble de dépendances fonctionnelles sur R
- Soit $X \rightarrow Y$ une DF (pas forcément en F)
- Nous disons que

$$F \models X \rightarrow Y \text{ (} F \text{ implique } X \rightarrow Y \text{)}$$

si toute instance de relation sur R qui satisfait les dépendances dans F satisfait aussi $X \rightarrow Y$

◆ Raisonner avec les DF

- Nous avons des *règles d'inférence* capables de nous dire comment dériver une DF f à partir d'un ensemble de DF F
- Cet ensemble des règles d'inférence est connu par le nom de *Axiomes d'Armstrong*

Inférence de DF

♦ Axiomes d'Armstrong :

Soit F un ensemble de DF sur un ensemble d'attributs U, $X \subseteq U$, $Y \subseteq U$

– **Réflexivité** : Si $Y \subseteq X$ alors $X \rightarrow Y$

▪ Exemples :

- $AB \rightarrow AB$
- $AB \rightarrow A$
- $note \rightarrow acquis$

– **Augmentation** : Si $X \rightarrow Y$ alors $XZ \rightarrow YZ$

▪ Exemples :

- Si $AB \rightarrow C$ alors $ABD \rightarrow CD$
- Si $note \rightarrow acquis$ alors $UE note \rightarrow UE acquis$

– **Transitivité** : Si $X \rightarrow Y$, $Y \rightarrow Z$ alors $X \rightarrow Z$

▪ Exemples :

- Si $AB \rightarrow C$ et $C \rightarrow D$ alors $AB \rightarrow D$
- Si $UE etud \rightarrow note$ et $note \rightarrow acquis$ alors $UE etud \rightarrow acquis$

UE	etud	note	acquis
BD	1	12	oui
BD	2	8	non
BD	3	8	non
GL	1	10	oui
GL	2	10	oui
GL	3	12	oui

Dérivation de DF

◆ Système d'Armstrong

- Un ensemble de DF F implique fn s'il existe une suite de dépendances fonctionnelles

f_1, f_2, \dots, f_n

où pour tout $i = 1 \dots n$ soit f_i est dans F soit elle est impliqué par l'ensemble f_1, f_2, \dots, f_{i-1} en se servant des axiomes d'Armstrong

- ◆ La suite **f_1, f_2, \dots, f_{n-1}** est alors appelée une dérivation (ou encore une démonstration) de f à partir de F

Dérivation de DF

◆ Exemple :

- Soit $F = \{\text{etud} \rightarrow \text{nom}, \text{UE} \rightarrow \text{libelle}\}$
- $F \models \text{etud UE} \rightarrow \text{nom libelle} ?$

◆ Démonstration :

- $\text{UE} \rightarrow \text{libelle}$ est dans F
- $\text{etud UE} \rightarrow \text{etud libelle}$ par augmentation de (i)
- $\text{etud} \rightarrow \text{nom}$ est dans F
- $\text{etud libelle} \rightarrow \text{nom libelle}$ par augmentation de (iii)
- $\text{etud UE} \rightarrow \text{nom libelle}$ par transitivité de (ii) et (iv)

etud	nom	UE	libellé
1	Paul	1.1	BD
2	Marie	1.1	BD
3	Steve	1.1	BD
1	Paul	1.2	GL
2	Marie	1.2	GL
3	Steve	1.2	GL
4	Andrea	1.2	GL

Dérivation de DF

- ◆ **Le système d'Armstrong est correct et complet :**
 - Il n'engendre pas des dépendances fonctionnelles fausses
 - Étant donné un ensemble F de DF, le système d'Armstrong nous permet d'engendrer toutes les DF valides
- ◆ **Fermeture d'un ensemble de DF : F^+**
 - F^+ est l'ensemble de toutes les DF qui peuvent être inférés à partir de F
- ◆ **Comment savoir si une DF est valide ?**
 - Option 1 : Chercher une démonstration (séquence de règles d'inférence) : *pas toujours facile !*
 - Option 2 : Construire F^+ : *trop d'effort*
 - Option 3 : Trouver un algorithme efficace et mécanique : 😊

Fermeture d'un ensemble d'attributs

◆ Définition : X_+

- Soient:
 - U : un ensemble d'attributs
 - F : un ensemble de DF sur U
 - $X \subseteq U$
- La fermeture de X par rapport à F , notée X_+ , est l'ensemble d'attributs de U déterminés fonctionnellement par X

$$X_+ = \{ A \in U / F \models X \rightarrow A \}$$

◆ Un tel ensemble existe et est unique

◆ Théorème : $F \models X \rightarrow Y$ ssi $Y \subseteq X_+$

Fermeture d'un ensemble d'attributs

♦ Algorithme de calcul de X^+

```
X+ := X;  
repeat  
  for each  $Y \rightarrow Z$  in F do  
    if  $Y \subseteq X^+$  then  $X^+ := X^+ \cup Z$ ;  
while (there are changes in  $X^+$ );
```

♦ Observation :

- Cet algorithme peut être implémenté comme un algorithme de coloration du graphe de DF

Fermeture d'un ensemble d'attributs

◆ Exemple:

- VIDEOS (videold, acqDate, price, title, genre, length, rating)
- $F = \{ \text{videold} \rightarrow \text{acqDate title},$
 $\text{title} \rightarrow \text{genre length rating},$
 $\text{title acqDate} \rightarrow \text{price} \}$

- Quelles DF sont valides ?

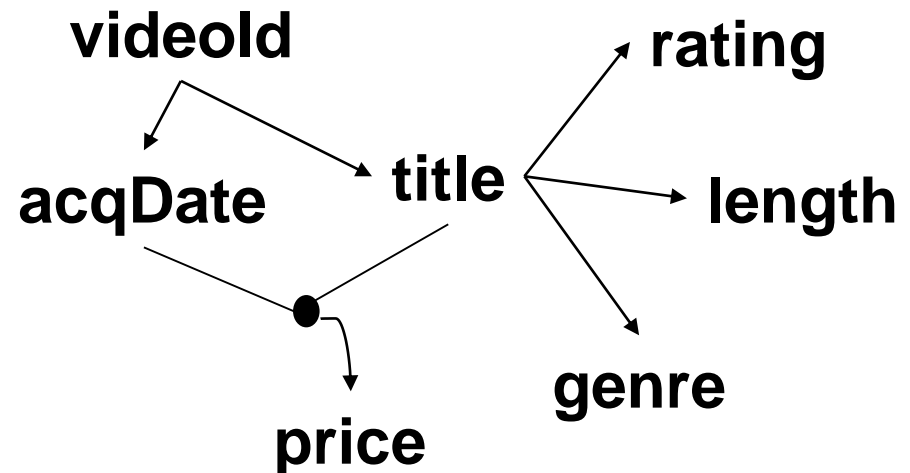
- $\text{videold} \rightarrow \text{rating}$
- $\text{title} \rightarrow \text{videold}$
- $\text{acqDate} \rightarrow \text{genre}$

- Réponses :

- $\text{videold}^+ = \{\text{videold}, \text{acqDate}, \text{price}, \text{title}, \text{genre}, \text{length}, \text{rating}\}$
- $\text{title}^+ = \{\text{title}, \text{genre}, \text{length}, \text{rating}\}$
- $\text{acqDate}^+ = \{\text{acqDate}\}$

- Observation : $(XY)^+ \neq X^+ \cup Y^+$

- $(\text{title acqDate})^+ = \{\text{title}, \text{acqDate}, \text{genre}, \text{length}, \text{rating}, \text{price}\}$



Exercice : Indiquer les DF valides

LOCATIONS (NumMat, marque, couleur, modele, annee, DocId, nom, adresse, telephone, NumContrat, date, periode, prix)

- $F = \{ \text{NumMat} \rightarrow \text{marque couleur modele annee},$
 $\text{DocId} \rightarrow \text{nom adresse telephone},$
 $\text{NumContrat} \rightarrow \text{date DocId NumMat periode prix},$
 $\text{date DocId NumMat} \rightarrow \text{NumContrat} \}$

Quelles-sont les DF valides ?

- $\text{NumContrat} \rightarrow \text{couleur}$
- $\text{marque} \rightarrow \text{DocId}$
- $\text{NumMat DocId} \rightarrow \text{date}$
- $\text{date DocId NumMat} \rightarrow \text{marque adresse}$

Équivalence d'ensembles de DF

◆ Équivalence entre des ensembles de DF : $F \equiv G$

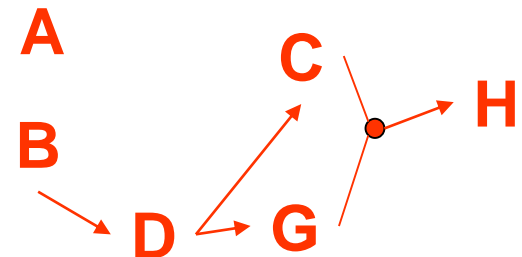
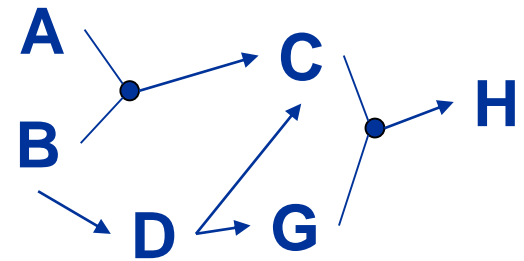
- Soient F et G deux ensembles de DF
- $F \equiv G$ (F est équivalent à G) si $F^+ = G^+$
 - Si les mêmes DF peuvent être inférés dans F et G

◆ Exemple :

- $R(A, B, C, D, G, H)$
- $F1 = \{AB \rightarrow C, B \rightarrow D, D \rightarrow GC, CG \rightarrow H\}$
- $F2 = \{B \rightarrow D, D \rightarrow G, D \rightarrow C, CG \rightarrow H\}$
- $F1 \equiv F2$
 - $F2$ est plus « simple » que $F1$.
 - $F1$ est plus « redondante » que $F2$.

◆ Objectif :

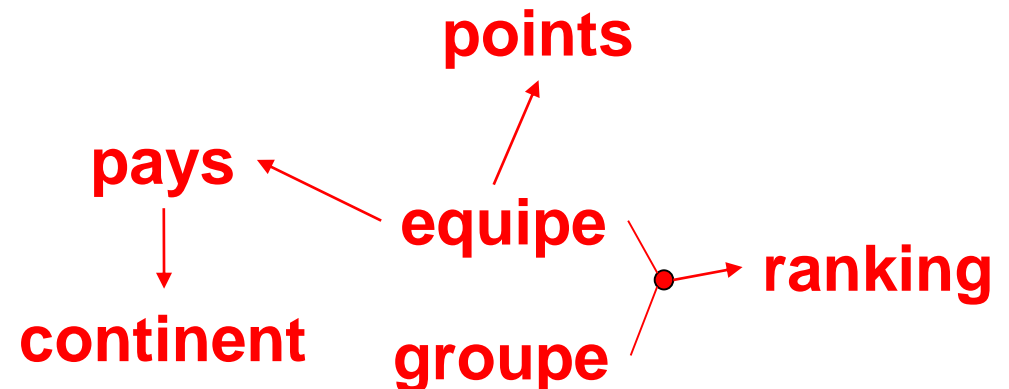
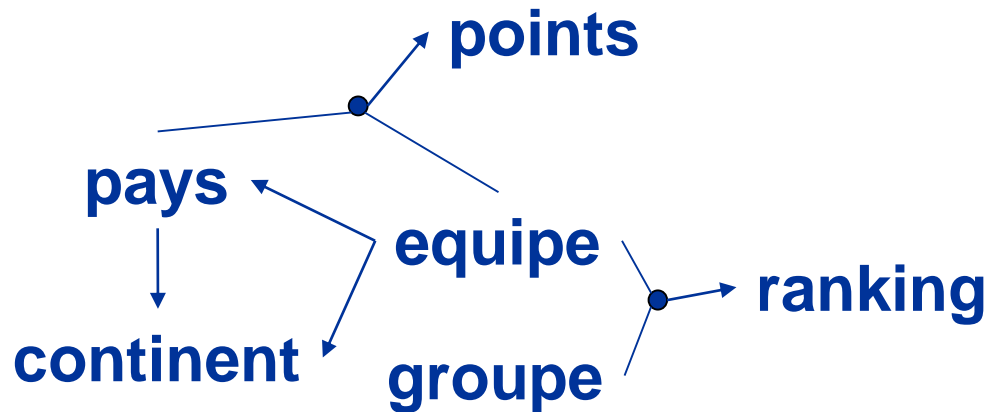
- Trouver un ensemble équivalent mais minimal



Exercice : Montrer l'équivalence

Ranking (groupe, ranking, equipe, pays, continent, points)

Ranking d'un équipe de foot lors de la phase de groupes d'une coupe.



Couverture minimale: A quoi ça sert ?

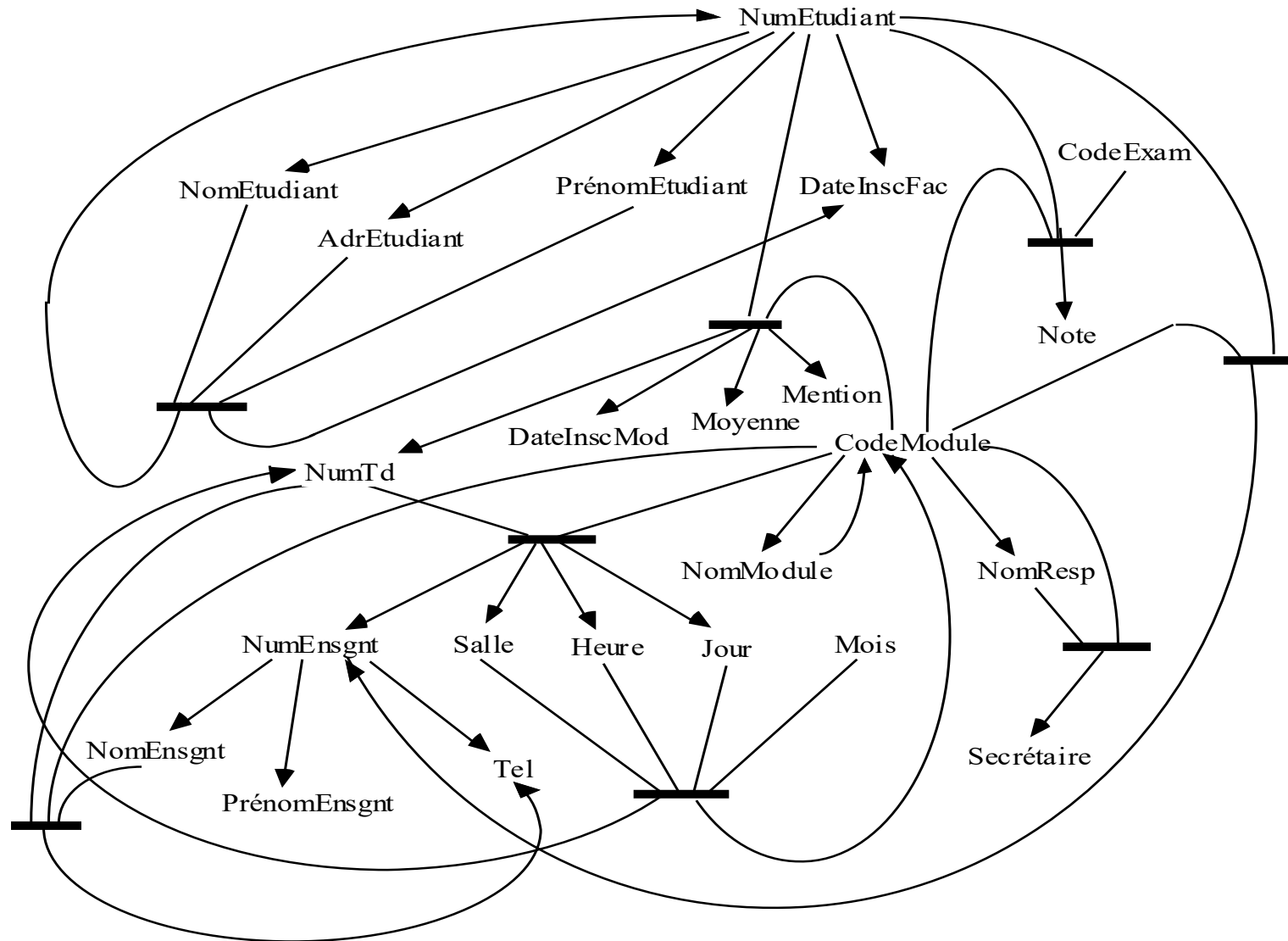
Intuition:

videold	acqDate	title	genre	length	rating
90987	05/02/2007	Star wars	Action	121	8.8
123	29/04/2001	Nikita	Drama	118	7.5
90988	25/03/2007	Star wars	Action	121	8.8
90989	05/04/2008	Star wars	Action	121	8.8
145	31/12/2005	Lady and the tramp	Comics	93	7.4
8034	05/04/2008	Lady and the tramp	Comics	93	7.4

videold	acqDate	title
90987	05/02/2007	Star wars
123	29/04/2001	Nikita
90988	25/03/2007	Star wars
90989	05/04/2008	Star wars
145	31/12/2005	Lady and the tramp
8034	05/04/2008	Lady and the tramp

title	genre	length	rating
Star wars	Action	121	8.8
Nikita	Drama	118	7.5
Lady and the tramp	Comics	93	7.4

Couverture minimale: A quoi ça sert ?



Couverture minimale

◆ Définition :

- Une couverture (ou réduction) minimale G est un ensemble de DF **équivalent** à F qui satisfait les propriétés suivantes :
 - **Réduction à droite** : Les DF de G ont un seul attribut à droite
 - **Réduction à gauche** : Les DF de G n'ont pas d'attribut inutile à gauche
 - **Réduction de redondances** : Les DF de G , ne peuvent pas être inférés à partir des autres

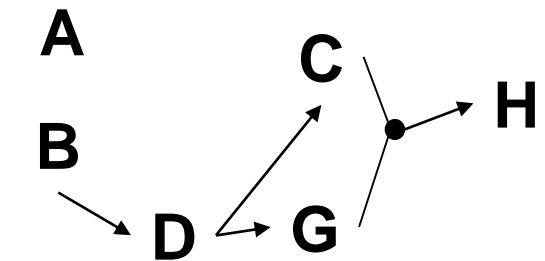
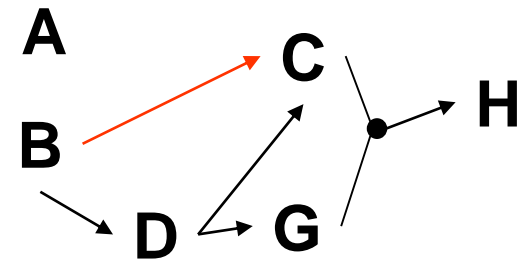
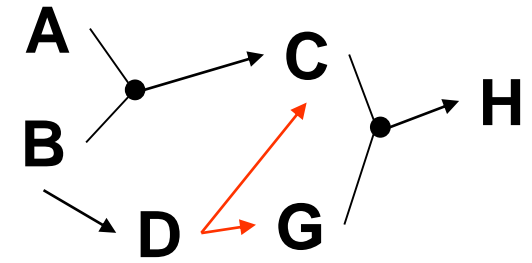
◆ Comment trouver une couverture minimale ?

- Pas 1 : Remplacer chaque DF $X \rightarrow A_1 \dots A_n$ par $X \rightarrow A_1 \dots X \rightarrow A_n$
 - Pas 2 : \forall DF $X \rightarrow A$, si $\exists X' \subset X$ tel que $X' \rightarrow A$, remplacer $X \rightarrow A$ par $X' \rightarrow A$
 - Pas 3 : \forall DF f , si elle peut être inférée à partir des autres, la supprimer
- Note: Suivant l'ordre dans lequel nous considérons les DF, nous pouvons avoir plusieurs réductions

Couverture minimale

◆ Exemple :

- $R(A, B, C, D, G, H)$
- $F = \{AB \rightarrow C, B \rightarrow D, D \rightarrow GC, CG \rightarrow H\}$
- Pas 1 : Réduction à droite:
 - remplacement de $D \rightarrow GC$ par $D \rightarrow G$ et $D \rightarrow C$
 - $F' = \{AB \rightarrow C, B \rightarrow D, D \rightarrow G, D \rightarrow C, CG \rightarrow H\}$
- Pas 2 : Réduction à gauche:
 - remplacement de $AB \rightarrow C$ par $B \rightarrow C$
 - $F' = \{B \rightarrow C, B \rightarrow D, D \rightarrow G, D \rightarrow C, CG \rightarrow H\}$
- Pas 3 : Réduction des redondances
 - $B \rightarrow C$ redondante
 - $F' = \{B \rightarrow D, D \rightarrow G, D \rightarrow C, CG \rightarrow H\}$



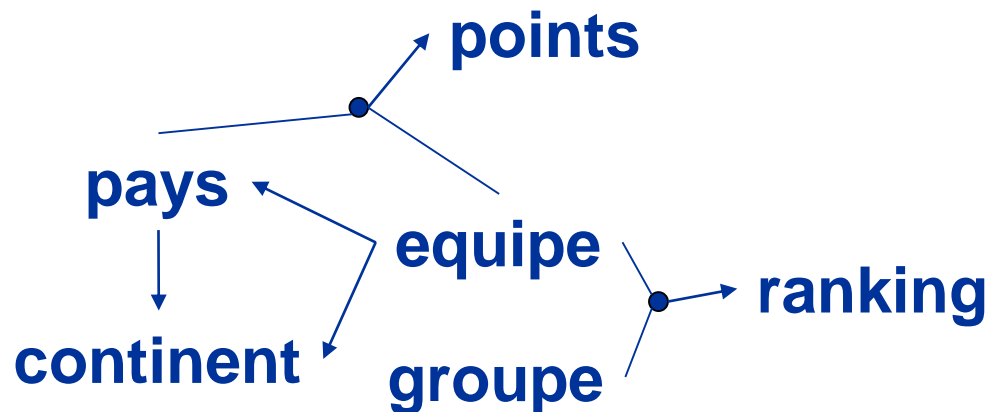
Exercice : Calculer une couverture minimale

Ranking (groupe, ranking, equipe, pays, continent, points)

Ranking d'un équipe de foot lors de la phase de groupes d'une coupe.

DF :

- (equipe, groupe) → ranking
- (equipe, pays) → points
- equipe → pays, continent
- pays → continent



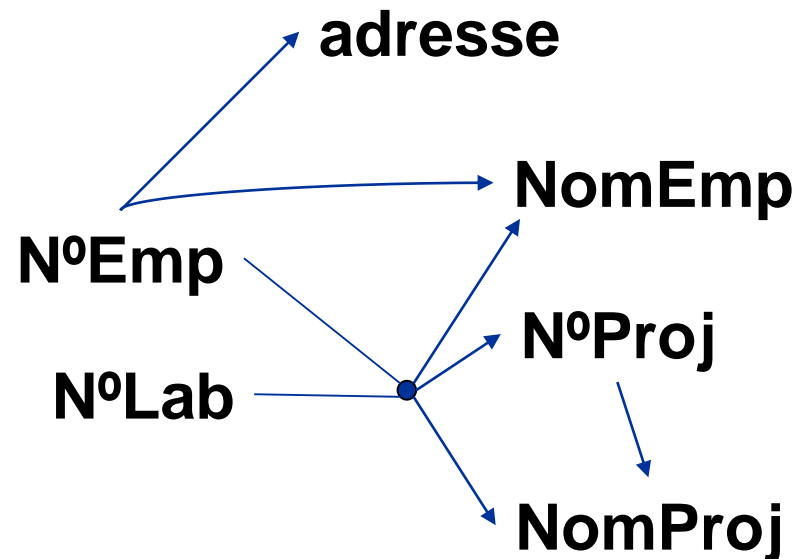
Exercice : Calculer une couverture minimale

Employé (N°Emp, N°Lab, N°Proj, NomEmp, NomProj, adresse)

Description d'un employé travaillant sur un projet d'un laboratoire.

DF :

- $(N^{\circ}Emp, N^{\circ}Lab) \rightarrow N^{\circ}Proj, NomProj, NomEmp$
- $N^{\circ}Emp \rightarrow NomEmp$
- $N^{\circ}Emp \rightarrow adresse$
- $N^{\circ}Proj \rightarrow NomProj$



Résumé

- ♦ **Grand enjeu dans la conception BD :**
 - Trouver un moyen de rendre la maintenance des DF efficace et moins chère
- ♦ **On peut dériver une couverture minimale des DF**
 - Ce qui réduit la complexité
 - Mais le contrôle est encore très couteux.
- ♦ **Avantage des clés sur les DFs :**
 - Les SGDB fournissent le contrôle des clés alors qu'ils ne fournissent pas le contrôle des DFs
- ♦ **Notre but : transformer les DF en clés !**

Plan

- ◆ Règles informelles de conception BD
- ◆ Contraintes d'intégrité
- ◆ Dépendances fonctionnelles (DF)
 - Définition de DF
 - Obtention de DF
- ◆ Théorie des dépendances fonctionnelles
 - Dérivation de DF
 - Fermeture d'un ensemble de d'attributs
 - Couverture minimale
- ◆ Clés



Clés

◆ Définition de clé basé sur les DF :

- Un ensemble d'attributs X est une clé pour une relation $R(A1...An)$ si :
 - 1 - Les attributs déterminent tous les autres attributs de la relation

$$X \rightarrow A1...An \in F+$$

En d'autres mots : deux tuples de R ne peuvent pas avoir les mêmes valeurs pour les attributs de X (définition classique de clé)

- 2 - Il n'existe pas un sous ensemble de X qui détermine tous les autres attributs de la relation

$$\neg \exists Y \subset X / Y \rightarrow A1...An \in F+$$

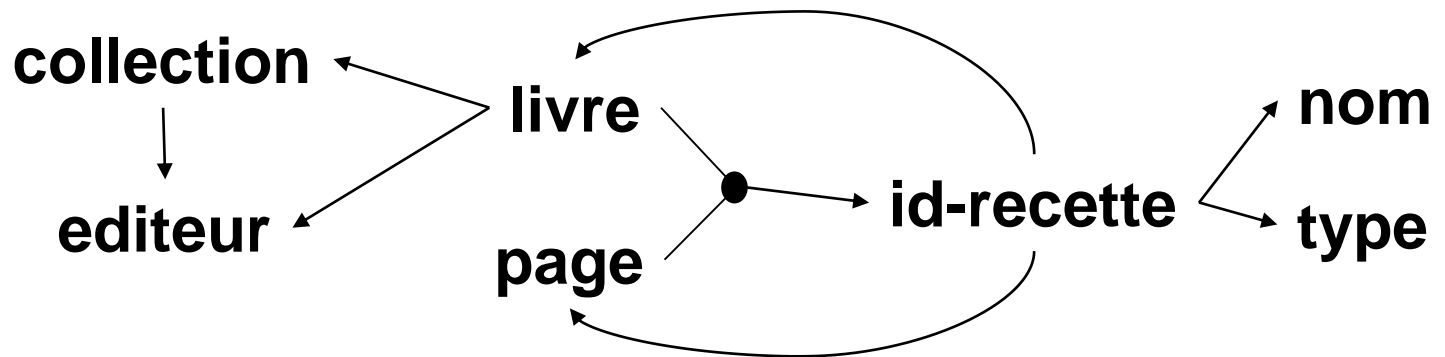
◆ Une relation peut avoir plusieurs clés

◆ Surclé: ensemble d'attributs qui contient une clé

Clés

◆ Exemple :

- RECETTES (id-recette, nom, type, livre, page, collection, editeur)



- **id-recette**⁺ = {id-recette, nom, type, livre, page, collection, editeur}
- **livre page**⁺ = {id-recette, nom, type, livre, page, collection, editeur}
- Y-a-t-il d'autres clés ?

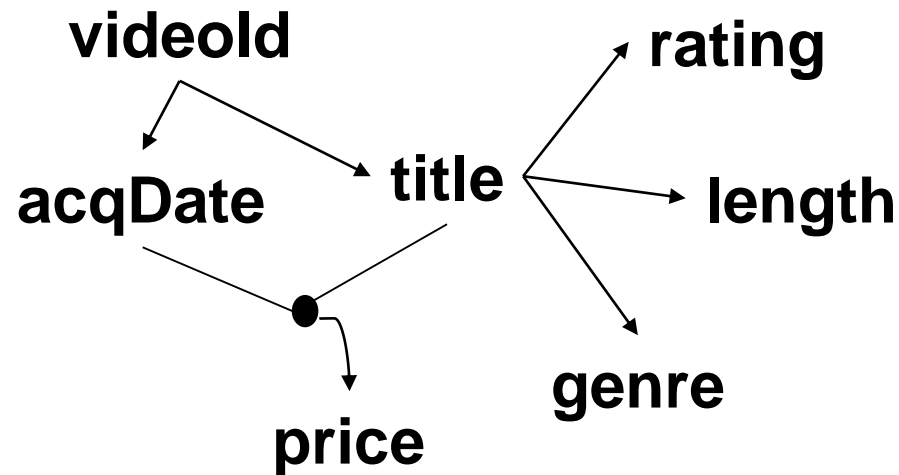
♦ Comment trouver les clés à partir des DF?

- Une méthode exhaustive consiste à calculer X^+ pour tout sous-ensemble X d'attributs
- Stratégie :
 - Commencer par des sous-ensembles d'un attribut, puis deux attributs et ainsi de suite
- Astuces :
 - Ne pas calculer les sous-ensembles qui contiennent déjà une clé.
Pourquoi ?
 - Les attributs qui ne sont à droite d'aucune DF appartiennent à toutes les clés. Pourquoi ?
 - Les attributs déterminés pour ceux qui ne sont à droite d'aucune DF n'appartiennent pas aux clés. Pourquoi ?
 - Les attributs qui ne sont à gauche d'aucune DF, mais qui sont à droite d'une DF n'appartiennent pas aux clés. Pourquoi ?

Clés

◆ Exemple:

- VIDEOS (videold, acqDate, price, title, genre, length, rating)

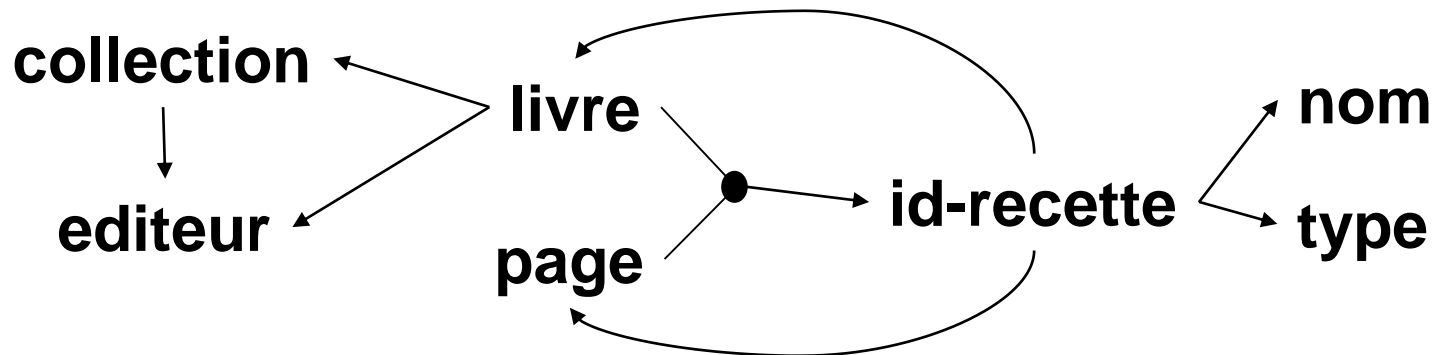


- Attributs qui ne sont à droite d'aucune DF : videold
 - ➔ videold appartient à toutes les clés
 - videold+ = {videold, acqDate, price, title, genre, length, rating}
 - ➔ videold est la seule clé de la relation

Clés

◆ Exemple :

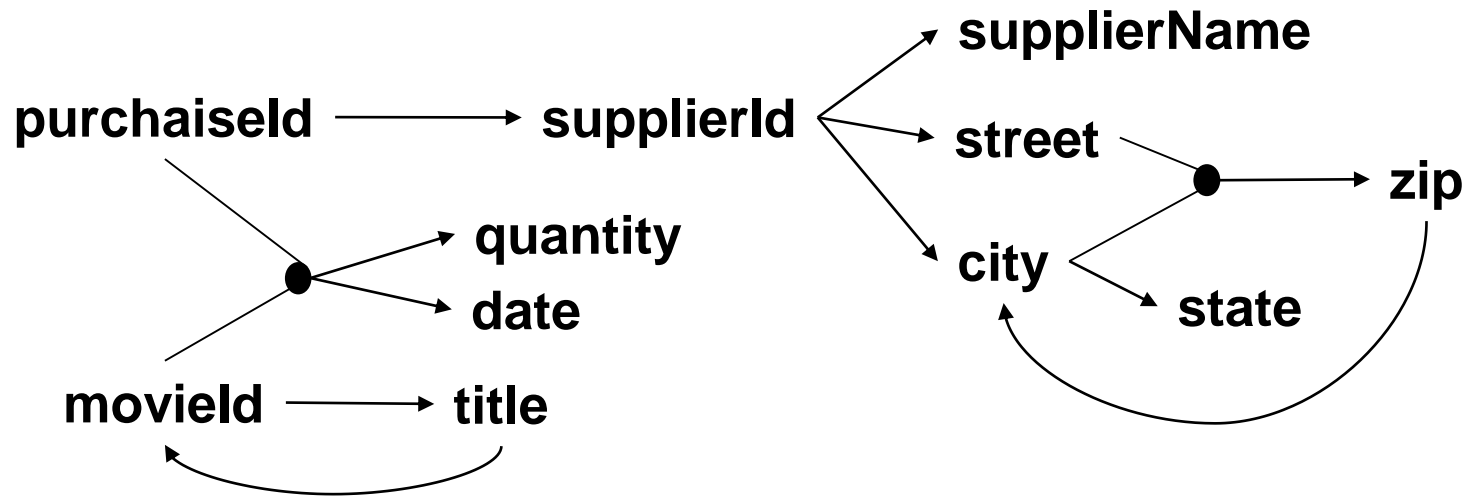
- RECETTES (id-recette, nom, type, livre, page, collection, editeur)



- Attributs qui ne sont à droite d'aucune DF : aucun. Pas de chance !
- Attributs qui sont à droite et pas à gauche des DF : {nom, type, editeur}
- On calcule les combinaisons de {id-recette, livre, page, collection} :
 - **id-recette+** livre+, page+, collection+
 - **(livre, page)+** (livre, collection)+, (page, collection)+

Exercice : Trouver toutes les clés

PURCHASEINFO (purchaseld, supplierId, supplierName, street, city, state, zip, movielfd, title, quantity, date)



Clés: $K1 = \{\text{purchaseld}, \text{movielfd}\}$,
 $K2 = \{\text{purchaseld}, \text{title}\}$

Résumé

- ♦ **Grande avantages des clés sur les DFs :**
 - Les SGDB fournissent le contrôle des clés alors qu'ils ne fournissent pas le contrôle des DFs
- ♦ **Notre but est donc de transformer les DF en clés !**
- ♦ **Technique de base :**
 - Identifier les tables avec des DFs qui ne sont pas des clés et les *décomposer* en tables plus petites
- ♦ **Dans la suite nous verrons...**
 - Des techniques de décomposition (appelé *normalisation*) et des critères pour obtenir des bonnes décompositions