

Cours

“Bases de données”

3° année (MISI)

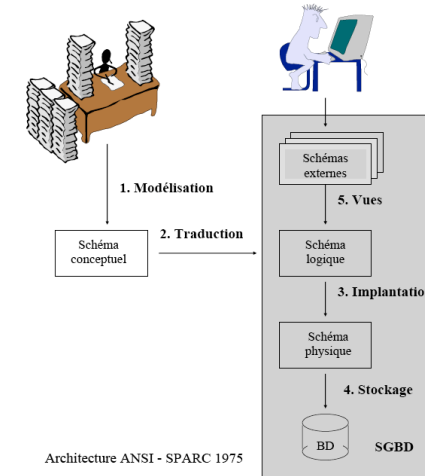
Antoine Cornuéjols

www.lri.fr/~antoine

antoine.cornuejols@agroparistech.fr

Introduction

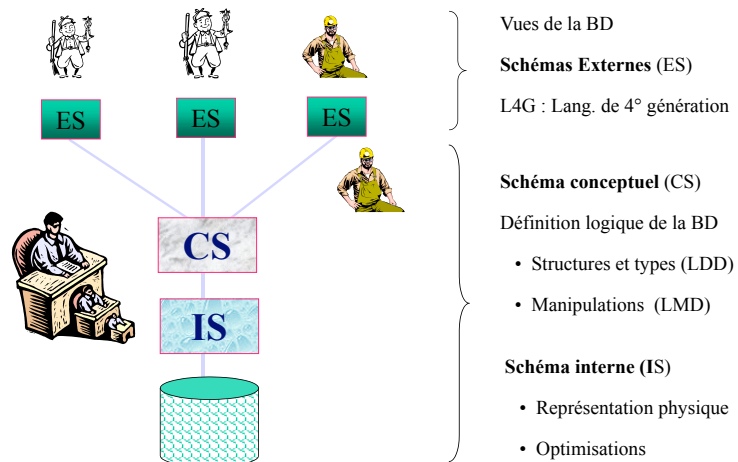
Conception d'une base de données



2

Introduction

Architecture ANSI-SPARC



3

Introduction

Conception d'une application pour une BD

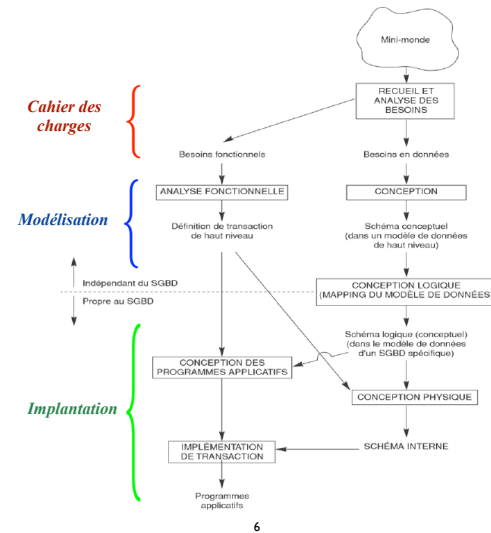
1. *Décider de la structure logique*
2. (Décider de la structure physique)
3. *Écrire des programmes d'application en utilisant la structure logique*
4. Le SGBD se charge de transcrire les commandes du LMD en instructions appropriées à la représentation physique

4

Conception d'une base de données

1. Un modèle conceptuel :
le **modèle entité-association**
2. Le **modèle relationnel**
 - 2.1. Définition
 - 2.2. Règles de traduction
 - 2.3. Dépendances entre données et formes normales

Le modèle entité-association



Le modèle entité-association

Introduction

Avantages

- **Simple**
- Suffisamment puissant pour représenter des **structures relationnelles**
- **Représentation graphique** facilitant sa compréhension

Mais

- **Ne propose que des structures : pas d'opérations**
- Peut mener à certaines ambiguïtés pour des schémas complexes

Le modèle entité-association

Un mauvais exemple

Titre	Année	nomMES	PrénomMES	AnnéeNaiss
Alien	1979	Scott	Ridley	1943
Vertigo	1958	Hitchcock	Alfred	1899
Psychose	1960	Hitchcock	Alfred	1899
Kagemusha	1980	Kurosawa	Akira	1910
Volte-face	1997	Woo	John	1946
Pulp Fiction	1995	Tarantino	Quentin	1963
Titanic	1997	Cameron	James	1954
Sacrifice	1986	Taskovski	Andrei	1932

Redondances (duplications)

- Erreurs possibles lors de l'**insertion**
- Erreurs possibles lors de la **mise-à-jour**
 - Si l'on modifie l'année de naissance d'Hitchcock pour Vertigo ...
- Si **destruction** d'un film, on risque de supprimer le metteur en scène

Le modèle entité-association

La bonne méthode

- Représenter individuellement les *films* et les *réalisateurs*
 - Une modification de l'un n'entraîne pas une modification de l'autre
- Définir une méthode d'**identification** des *films* ou des *réalisateurs*
 - La même information est représentée une seule fois
- Préserver le **lien** entre *films* et *réalisateurs*, mais sans introduire de redondance

9

Le modèle entité-association

La bonne méthode

Titre	Année
Alien	1979
Vertigo	1958
Psychose	1960
Kagemusha	1980
Volte-face	1997
Pulp Fiction	1995
Titanic	1997
Sacrifice	1986

Table des films

Id	nomMES	PrénomMES	AnnéeNaiss
1	Scott	Ridley	1943
2	Hitchcock	Alfred	1899
3	Kurosawa	Akira	1910
4	Woo	John	1946
5	Tarantino	Quentin	1963
6	Cameron	James	1954
7	Tarkovski	Andrei	1932

Table des réalisateurs

👉 Pas de redondance

10

Le modèle entité-association

La bonne méthode

Titre	Année	idMES
Alien	1979	1
Vertigo	1958	2
Psychose	1960	3
Kagemusha	1980	4
Volte-face	1997	5
Pulp Fiction	1995	6
Titanic	1997	7
Sacrifice	1986	8

Table des films

Id	nomMES	PrénomMES	AnnéeNaiss
1	Scott	Ridley	1943
2	Hitchcock	Alfred	1899
3	Hitchcock	Alfred	1899
4	Kurosawa	Akira	
5	Woo	John	
6	Tarantino	Quent	
7	Cameron	James	
8	Taskovski	Andre	

Table des réalis

La modélisation avec un graphique Entité/Association offre une méthode simple pour arriver à ce résultat, et ce même dans des cas beaucoup plus complexes.

[Rigaud, poly, p.19]

- Plus d'anomalies

- d'insertion
- de mise à jour
- de destruction

Conservation de l'information sans redondance

11

Le modèle conceptuel des données

Concepts de base

Schéma E/A \equiv *abstraction* d'un domaine d'étude

- Entité** : objet concret ou abstrait

- ayant une existence propre
- fonction des besoins de modélisation

Ex: Assuré, Film, Contrat

- Association** : lien sémantique entre entités

- fonction des besoins de modélisation

Ex: Réaliser (entre réalisateur et film)

12

Le modèle conceptuel des données

Concepts de base

👤 **Propriété** : donnée élémentaire

- 👤 ayant un sens
- 👤 pouvant être utilisée de manière autonome

Ex : NomAssuré, AnnéeSouscription, NbContrats

- 👤 Servent à décrire les entités et les associations
- 👤 = *Attributs* ou *colonnes*
- 👤 prennent des valeurs appelées *occurrences* de la propriété

Propriété	Occurrences
NomAssuré	Dupont
	Dupond
	MaRtin
AnnéeSouscription	1988
	1989

13

Le modèle conceptuel des données

Remarques

- 👤 La modélisation conceptuelle est totalement indépendante de tout choix d'implantation.
 - Partie la plus stable d'une application
 - Se concentrer sur l'essentiel : *Que veut-on stocker dans la base ?*
- 👤 Le modèle E/A a été conçu en 1976 et est à la base de la plupart des méthodes de conception.
- 👤 La syntaxe est souvent celle d'UML
 - En France, on utilise parfois la syntaxe MERISE (quasi équivalente)

14

Le modèle conceptuel des données

Les types d'entités

👤 Le **type d'une entité** est composé de :

- 👤 son *nom*
- 👤 la *liste de ses attributs* avec, —optionnellement— le domaine dans lequel l'attribut prend ses valeurs : les entiers, les chaînes de caractères :
- 👤 l'indication du (ou des) attribut(s) permettant d'identifier l'entité : ils constituent la *clé*.

- 👤 Une entité *e* est une *instance* de son type *E*.
- 👤 Un ensemble d'entités $\{e_1, e_2, \dots, e_n\}$ instances d'un même type *E* est une *extension* de *E*.

👤 **Définition (clé)**

- 👤 Soit *E* un type d'entité et *A* l'ensemble des attributs de *E*. Une **clé** de *E* est un *sous-ensemble minimal* de *A* permettant d'identifier de manière unique une entité parmi n'importe quelle extension de *E*.

15

Le modèle conceptuel des données

Les clés : exemples

👤 Exemple :

👤 **Internaute** : plusieurs attributs

- email
- nom
- prénom
- région

👤 **email** : clé naturelle car unique et discriminante pour chaque internaute

👤 **nom** : impossible car plusieurs internautes peuvent avoir le même nom

👤 **(nom, prénom)** : clé possible, mais peut poser des problèmes de performance et complique les manipulations par SQL

16

Le modèle conceptuel des données

Les clés : caractéristiques souhaitables

🔑 Quand plusieurs clés possibles pour un même ensemble d'entités

- 🔑 On en choisit une comme *clé primaire*
- 🔑 et les autres comme *clés secondaires*

🔑 Bonne clé primaire :

- 🔑 sa valeur est connue pour toute entité
- 🔑 on ne doit jamais avoir besoin de la modifier
- 🔑 sa taille de stockage doit être la plus petite possible (pour des raisons de performance)

🔑 Si difficile à trouver

- 🔑 créer un identifiant « abstrait », indépendant de tout autre attribut

17

Le modèle conceptuel des données

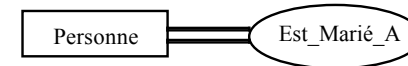
Caractéristiques d'une association

Dimension d'une association : son arité

🔑 Exemple : L'association **Ecrit_Par** est de dimension deux (binaire)



🔑 Exemple : L'association **Est_Marié_A** est de dimension deux (binaire réflexive)



🔑 Exemple : L'association **A_Vendu_A** est de dimension trois (ternaire)



18

Le modèle conceptuel des données

Caractéristiques d'une association

Type de liaison inter-entités : trois types de liaisons

🔑 **Liaison de type 1 à 1** : À toute occurrence de X correspond une et une seule occurrence de Y et réciproquement



Dans le mariage judéo-chrétien, un homme est marié à une seule femme et réciproquement.

🔑 **Liaison de type 1 à plusieurs (1 à n)** : À toute occurrence de X correspond une ou plusieurs occurrences de Y et une seule de X.



Un client a établi un ou plusieurs contrat(s) mais un contrat a été établi avec un seul client.

19

Le modèle conceptuel des données

Caractéristiques d'une association

Type de liaison inter-entités : trois types de liaisons

🔑 **Liaison de type plusieurs à plusieurs** : À toute occurrence de X correspond une ou plusieurs occurrences de Y et réciproquement



Un client peut commander plusieurs produits et un produit peut être commandé par plusieurs clients.

20

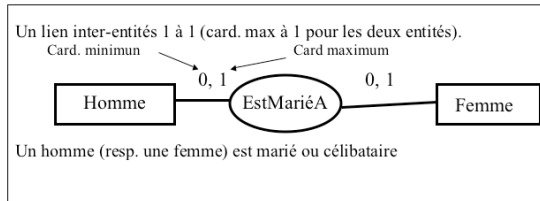
Le modèle conceptuel des données

Caractéristiques d'une association

Cardinalités (minimales et maximales)

- **Cardinalité minimum** : nombre minimum d'occurrences d'une entité X dans l'association considérée.
- **Cardinalité maximum** : nombre maximum d'occurrences d'une entité X dans l'association considérée.

Exemples :



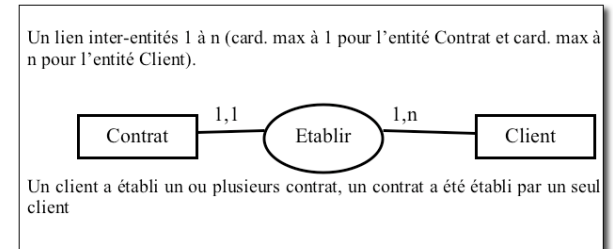
21

Le modèle conceptuel des données

Caractéristiques d'une association

Cardinalités (minimales et maximales)

Exemples :



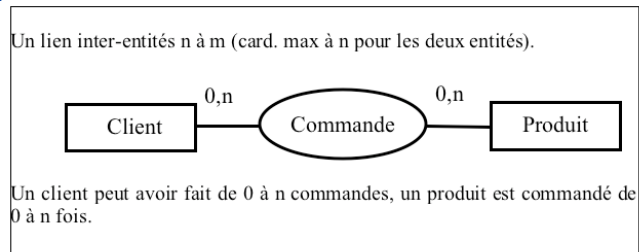
22

Le modèle conceptuel des données

Caractéristiques d'une association

Cardinalités (minimales et maximales)

Exemples :



23

Le modèle conceptuel des données

Cardinalités

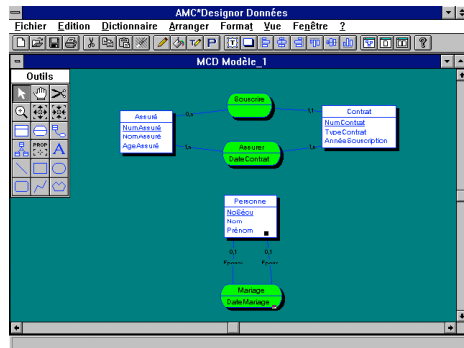
Les valeurs de cardinalités sont en général 0, 1 ou N :

- La cardinalité **minimum à 0** veut dire que certaines occurrences de l'entité X ne sont pas impliquées dans une occurrence de l'association.
- La cardinalité **minimum à 1** veut dire qu'une occurrence de l'entité X ne peut exister sans participer à une occurrence de l'association.
- La cardinalité **maximum à 1** veut dire que toute occurrence de l'entité X ne peut participer au plus qu'à une occurrence de l'association.
- La cardinalité **maximum à N** veut dire qu'une occurrence de l'entité X peut être impliquée dans un maximum de N occurrences de l'association.

24

Le modèle conceptuel des données

Concepts de base

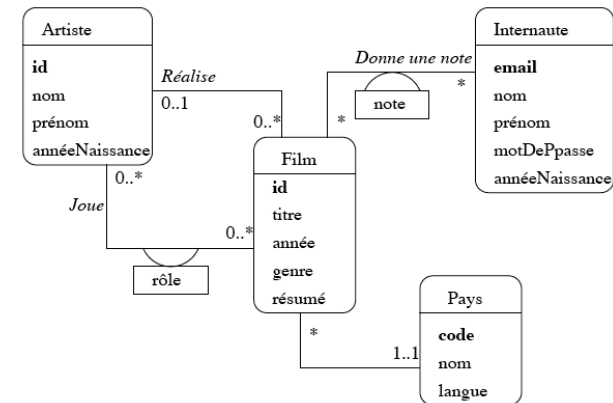


- Plusieurs associations peuvent exister entre les mêmes entités
- Une association peut être *réflexive* (relie une entité à elle-même)

25

Le modèle conceptuel des données

Concepts de base



26

Le modèle conceptuel des données

Concepts de base

Chaque **entité** est caractérisée par un *ensemble d'attributs*, parmi lesquels un ou plusieurs forment *l'identifiant unique* (en gras)

Les **associations** sont caractérisées par des *cardinalités* (minimale .. maximale).

- «0..*» sur le lien *Réalise* (côté entité *Film*)
 - un artiste peut réaliser plusieurs films ou aucun
- «0..1» sur le lien *Réalise* (côté entité *Artiste*)
 - un film ne peut être réalisé que par au plus un artiste

27

Le modèle conceptuel des données

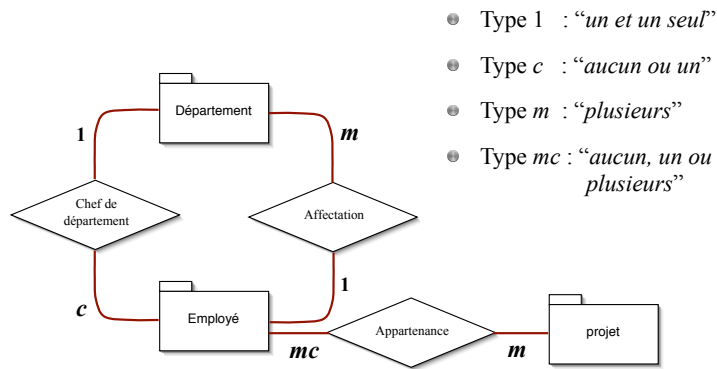
Les types d'associations

- Association *simple* (type 1)
 - Bijection entre E1 et E2
- Association *conditionnelle* (type c)
 - À chaque entité de E1 correspond *au plus une* entité de E2
- Association *multiple* (type m)
 - À chaque entité de E1 correspond *plusieurs* entités de E2
- Association *multiple conditionnelle* (type mc)
 - À chaque entité de E1 correspond *aucune, une, ou plusieurs* entité(s) de E2

28

Le modèle conceptuel des données

Les types d'associations



29

Le modèle conceptuel des données

Généralisation – spécialisation

Généralisation :

- Processus d'abstraction consistant à généraliser les entités, et les ensembles d'entités, en un seul ensemble ascendant.

Spécialisation :

- Les sous-ensembles d'entités dans une hiérarchie de généralisation résultent du processus de spécialisation.

- Les sous-ensembles peuvent avoir une intersection
 - Ex : Employé : appartenant à club de photo et club de sport
- Les sous-ensembles ont une intersection et sont exhaustifs
 - Ex : appartenant forcément à club de photo et/ou club de sport et/ou club d'échec
- Sous-ensembles mutuellement disjoints
- Sous-ensembles mutuellement disjoints et exhaustifs

30

Le modèle conceptuel des données

Clé d'une association

- La clé d'une association (binaire) entre un type d'entité $E1$ et un type d'entité $E2$ est le couple constitué de la clé $c1$ de $E1$ et la clé $c2$ de $E2$.

Mais cela peut poser des problèmes, d'où des aménagements possibles.



31

Le modèle conceptuel des données

Entités faibles

Définition (entité forte)

- Entité qui, disposant de son propre identifiant, peut être considérée isolément.
- Ex : Assuré (N° de SS)

Définition (entité faible)

- Entité ne pouvant exister qu'en étroite association avec une autre.
- Ex : salle de cinéma, ayant_droit d'un employé

32

Le modèle conceptuel des données

Avantages et inconvénients du modèle E/A

🧠 Avantages :

1. Il n'y a que trois concepts : *entités*, *associations* et *attributs*
2. Représentation graphique intuitive (même si beaucoup de conventions)
3. Permet de modéliser rapidement des structures pas trop complexes

🧠 Limites :

1. **Pauvreté**
 1. Difficile d'exprimer des contraintes d'intégrité
 2. Difficile d'exprimer des structures complexes
2. La conception en schémas reste en partie matière de **bon sens** et d'**expérience**
3. Ne propose **pas d'opérations sur les données**

33

Le modèle conceptuel des données

Exemple : un mini-monde de l'entreprise

1. **La société est organisée en services.** Chaque service a un nom (`nomSce`) et un numéro unique (`noSce`) et est dirigé par un employé unique (`noSsDir`). La date (`dateDebDir`) à laquelle l'employé a commencé à diriger le service est comptabilisée. Un service peut avoir plusieurs emplacements.
2. Un **service** contrôle un certain nombre de projets, chacun d'entre eux ayant un nom (`nomProjet`), un numéro (`numeroProjet`) et un emplacement unique (`emplProjet`).
3. Le nom de chaque **employé**, son numéro de Sécurité sociale (`noSs`), son adresse, son salaire, son sexe et sa date de naissance sont mémorisés. Un employé est affecté à un service (`noSce`), mais peut travailler sur plusieurs projets qui ne sont pas forcément contrôlés par le même service. Le nombre d'heures hebdomadaires travaillées par chaque employé par projet est comptabilisé. Le supérieur immédiat (`noSsSuper`) de chaque employé est lui aussi mémorisé.
4. Les **ayants droit** de chaque employé (`noSsEmpl`) doivent être indiqués pour des raisons d'assurance. Leur prénom, leur sexe, leur date de naissance et leur lien de parenté avec l'employé sont mémorisés.

34

Conception d'une base de données

1. Un modèle conceptuel :
le **modèle entité- association**
2. Le **modèle relationnel**
 - 2.1. Définition
 - 2.2. Règles de traduction
 - 2.3. Dépendances entre données et formes normales

Le modèle relationnel

Rappels

🧠 Un **modèle de données** définit un mode de représentation de l'information selon trois composantes :

- 🧠 Des **structures de données**
- 🧠 Des **contraintes** qui permettent de spécifier les règles que doit respecter une base de données
- 🧠 Des **opérations** pour manipuler les données : interrogations et mises à jour.

- **Langage de définition de données** (DDL)
 - Structures de données + contraintes → Schéma d'une base de données
- **Langage de manipulation des données** (DML)
 - Opérations → SQL

🧠 Le modèle relationnel offre une **totale indépendance** entre

- 🧠 la représentation logique
- 🧠 et l'implantation physique

36

Le modèle relationnel

Définitions

- Il n'existe qu'une seule structure : la **relation**
- Une relation peut se représenter par une **table**

Domaine de valeurs

- Ensemble d'instances d'un type élémentaire
- Ex : les entiers, les réels, les chaînes de caract.

Attributs

- Nomment les **colonnes** d'une relation
- Un attribut est toujours associé à un domaine

Titre	Année	Genre
Alien	1979	S-F
Vertigo	1958	Suspense
Volte-face	1997	Thriller
Pulp Fiction	1995	Policier

Schéma de relation

- Un nom suivi de la liste des attributs (associé à son domaine)
- Syntaxe : $R(A_1 : D_1, A_2 : D_2, \dots, A_n : D_n)$
- Arité** : nombre d'attributs de la relation

37

Le modèle relationnel

Définitions

Relation (ou instance d'une relation R)

- Sous-ensemble fini du produit cartésien des domaines des attributs de R
- Une relation est donc un ensemble
 - L'ordre des lignes n'a pas d'importance
 - Pas deux lignes identiques
 - Pas de "case vide" dans la table

Clé d'une relation

- Le plus petit sous-ensemble des attributs permettant d'identifier chaque ligne de manière unique

Tuple (ou n-uplet)

- Liste de n valeurs (v_1, \dots, v_n) où chaque valeur v_i est la valeur d'un attribut A_i de domaine D_i
- Ex : ('Cyrano', 1992, 'Rappeneau')

Titre	Année	Genre
Alien	1979	S-F
Vertigo	1958	Suspense
Volte-face	1997	Thriller
Pulp Fiction	1995	Policier

38

Le modèle relationnel

Définitions

Base de données (ou instance de base de données)

- Ensemble fini (d'instances) de relations

Schéma d'une base de données

- Ensemble des schémas de relations de cette base

Le choix des relations (et schémas de relations) est essentiel

- Il détermine les qualités de la base
 - Performances, exactitude, exhaustivité, disponibilité des informations, ...
- La théorie des BD relationnelles définit ce qu'est un bon schéma et propose des outils formels pour y parvenir
 - Contraintes d'intégrité
 - Normalisation

39

Le modèle relationnel

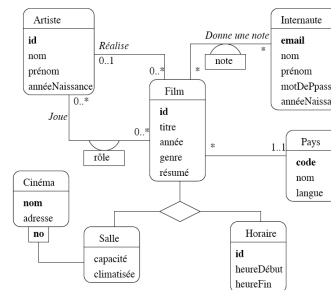
Passage d'un schéma E/A à un schéma relationnel

- Passage d'un modèle à **deux structures** (entités + associations) à un modèle à **une structure** (relations)

Règle 1 : entités

Pour chaque entité du schéma E/A

- On crée une relation de même nom que l'entité
- Chaque propriété de l'entité, y compris l'identifiant, devient un attribut de la relation (une colonne)
- Les attributs de l'identifiant constituent la clé de la relation



Film (**id**Film, titre, année, genre, résumé)
 Artiste (**id**Artiste, nom, prénom, annéeNaissance)
 Internaute (email, nom, prénom, région)
 Pays (code, nom, langue)

40

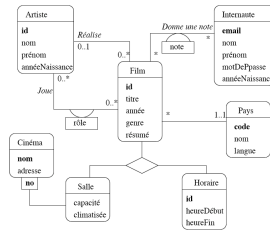
Le modèle relationnel

Passage d'un schéma E/A à un schéma relationnel

Règle 2 : associations de un à plusieurs

Soit une association de un à plusieurs entre *A* et *B*.

- 1- On crée les relations R_A et R_B correspondant respectivement aux entités *A* et *B*
- 2- L'identifiant de *B* devient un attribut de R_A .



Film (*idFilm*, titre, année, genre, résumé)

Artiste (*idArtiste*, nom, prénom, annéeNaissance)

Internaute (*email*, nom, prénom, région)

Pays (*code*, nom, langue)

...

41

Film (*idFilm*, titre, année, genre, résumé, *idMES*)

Artiste (*idArtiste*, nom, prénom, annéeNaissance)

Internaute (*email*, nom, prénom, région)

Pays (*code*, nom, langue)

...

Le modèle relationnel

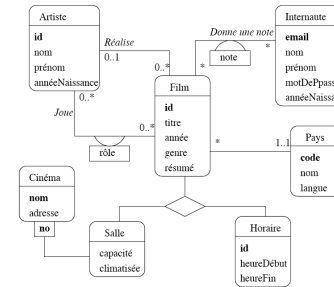
Passage d'un schéma E/A à un schéma relationnel

Règle 3 : associations avec type entité faible

Une entité faible est toujours identifiée par rapport à une autre entité (e.g. salle de cinéma).

Il s'agit d'une association "un à plusieurs".

➔ Même règle de passage : on utilise un mécanisme de clé étrangère pour référencer l'entité forte dans l'entité faible. La clé étrangère est une partie de l'identifiant de l'entité faible.



Cinéma (*nomCinéma*, numéro, rue, ville)

Salle (*nomCinema*, no, capacité)

42

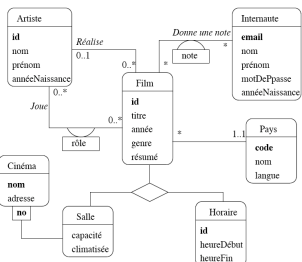
Le modèle relationnel

Passage d'un schéma E/A à un schéma relationnel

Règle 4 : associations binaires de plusieurs à plusieurs

Soit une association binaire de *n-m* entre *A* et *B*.

- 1- On crée les relations R_A et R_B correspondant respectivement aux entités *A* et *B*
- 2- On crée une relation $R_{A,B}$ pour l'association
- 3- La clé de R_A et la clé de R_B deviennent des attributs de $R_{A,B}$
- 4- La clé de cette relation est la concaténation des clés des relations R_A et R_B
- 5- Les propriétés de l'association deviennent des attributs de $R_{A,B}$.



43

Film (*idFilm*, titre, année, genre, résumé, *idMES*, *codePays*)

Artiste (*idArtiste*, nom, prénom, annéeNaissance)

Internaute (*email*, nom, prénom, région)

Role (*idFilm*, *idActeur*, nomRôle)

Notation (*email*, *idFilm*, note)

Le modèle relationnel

Retour sur le choix des identifiants

Il est préférable, en général, de choisir un identifiant "neutre" qui ne soit pas une propriété de l'entité

- Chaque valeur de l'identifiant doit caractériser de manière unique une occurrence
 - Ex : **Titre** pour la relation *Film* ou **nom** pour la relation *Acteur* ne sont pas de bons choix

Si on utilise un ensemble de propriétés comme identifiant, la référence à une occurrence est très lourde

- Ex : si on prenait pour clé de *Cinéma* l'identifiant (**nom, rue, ville**)

L'identifiant sert de référence externe et ne doit jamais être modifiable (il faudrait répercuter les mises à jour)

- Ex : **Titre** pour la relation *Film* ou **nom** pour la relation *Acteur* ne sont pas de bons choix

44

Le modèle relationnel

Dépendances fonctionnelles

Soient $R(A_1, A_2, \dots, A_n)$ un schéma de relation et X et Y deux sous-ensembles de (A_1, A_2, \dots, A_n)

● Dépendance fonctionnelle

- Y dépend fonctionnellement de X ssi à toute valeur de X correspond une valeur unique de Y .

● Détermination

- On dit que $X \rightarrow Y$ (X détermine Y) si pour toute extension r de R , pour tout tuple t_1 et t_2 de r , on a :

$$R[X](t_1) = R[X](t_2) \Rightarrow R[Y](t_1) = R[Y](t_2)$$

45

Le modèle relationnel

Règles d'inférence sur les DF

- **Réflexivité** : Y inclus dans $X \Rightarrow X \rightarrow Y$
- **Augmentation** : $X \rightarrow Y \Rightarrow XZ \rightarrow YZ$
- **Transitivité** : $X \rightarrow Y$ et $Y \rightarrow Z \Rightarrow X \rightarrow Z$
- **Union** : $X \rightarrow Y$ et $X \rightarrow Z \Rightarrow X \rightarrow YZ$
- **Pseudo-transitivité** : $X \rightarrow Y$ et $WY \rightarrow Z \Rightarrow XW \rightarrow Z$
- **Décomposition** : $X \rightarrow Y$ et Z inclus dans $Y \Rightarrow X \rightarrow Z$

● Dépendance fonctionnelle élémentaire

- Une DF de la forme $X \rightarrow A$ où A est un attribut unique non inclus dans X et où il n'existe pas de X' inclus dans X tel que $X' \rightarrow A$.

46

Le modèle relationnel

Fermeture et couverture minimale

● Fermeture transitive

- La fermeture transitive F^+ d'un ensemble F de DF est constituée d'elle-même et des DF déduites par transitivité.

● Couverture minimale

- La couverture minimale de F notée $Min(F)$ est telle que :
 - Minimalité : aucune DF de $Min(F)$ n'est redondante (i.e. $F^+(Min(F)-f) \neq F^+(Min(F))$)
 - Exhaustivité : $(Min(F))^+ = F^+$

Exemple :

$F = \{NV \rightarrow Type, Type \rightarrow Marque, Type \rightarrow Puissance, NV \rightarrow Couleur\}$

$F^+ = \{NV \rightarrow Marque, NV \rightarrow Puissance\} \cup F$

$Min(F) = F$

47

Le modèle relationnel

Les formes normales

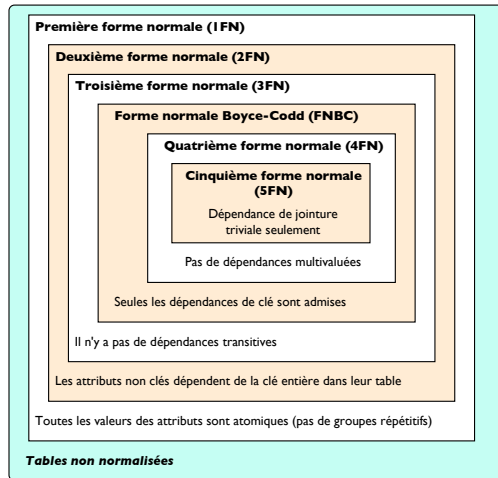
Objectif : détecter et étudier les dépendances à l'intérieur des tables pour en éliminer les informations redondantes et les anomalies qui en résultent.

Un **attribut** dans une table est **redondant** lorsque ses valeurs peuvent être éliminées de cette table sans perte d'information.

48

Le modèle relationnel

Les formes normales



49

Le modèle relationnel

Les contraintes d'intégrité structurelles



50

Le modèle relationnel

Dénormalisation du modèle logique

- 🗑 Suppression de relations
 - 🗑 XXX
- 🗑 Introduction de redondances
 - 🗑 XXX

[Rigaux, p41-42]

Il faut systématiquement comparer le gain attendu avec le risque encouru.

51

Langage de manipulation des données

1. **L'algèbre relationnelle**

- 1.1. Opérateurs ensemblistes
- 1.2. Opérateurs relationnels
- 1.3. Expression de requêtes

2. **Le langage SQL**