



### Modélisation d'un schéma de données

Composé avec EFGA, version 2388

1

### Coût pour maintenir une donnée à jour ?



3

### Modélisation d'un schéma de données

**objectif** produire un ensemble de tables

- capturer la complexité du monde réel
- rester aussi simple que possible
- adapter aux besoins effectifs...
- engagement à long terme, socle du SI !

**réalité** des schémas de données

- souvent erronés (codages, redondance, hypothèses implicites)
- mal maintenues par les utilisateurs (déroulement d'attributs)

5

### Modélisation Entité/Association (E/A)

**méthode graphique** pour fixer les idées

approche **ascendante**, regroupement de parties

simple, intuitive, peu de concepts

notation MCD de la méthode Merise

**mais** quelques manques

démarche non déterministe : pas d'unicité du modèle

pas de contraintes d'intégrité

pas de procédure / méthode

**démarche bon sens et expérience**

**alternative** UML (riche, complexe)

7

### Le socle des SI est souvent pourri !



2

### Coût d'une donnée pas à jour ?



4

### Plusieurs niveaux

**externe** vue (partielle ?)

pour utilisateurs et applications

**conceptuel** vue sémantique des données

pour les concepteurs du SI

**logique** organisation de la base

pour les développeurs

**physique** implémentation...

types, index, disques...

pour les administrateurs

*schéma graphique*

*relations*

*tables*

6

### Modélisation E/A



objet identifié comme tel (nom)

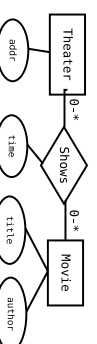


lien entre 2 entités (verbe), éventuellement réflexif



propriété d'une entité ou association

**cardinalité** dénombrement des liens entre entité



8

## Cardinalités

nombre de liaisons entité-association

format min-max min vaut 0 ou 1, max vaut 1 ou \*

exemple générique



– un *sujet possédé* entre  $n$  et  $m$  objets

– un *objet est possédé par* entre  $p$  et  $q$  sujets

attention côté inverse par rapport à UML !

codage simplifié

min/max	1	*
0	?	*
1	1	+

cardinalités max d'une association entre des entités

9

## TD Recettes de cuisine

- les aliments et leur caractéristiques
- coût économique, valeur énergétique, glucides... vitamines...
- des recettes de cuisine
- leur composition en aliments
- leur réalisation, en temps...
- des plats, leur décomposition en recettes...
- classement des recettes
- entrée, plat, dessert, viande, poisson, fromage, pays d'origine...

11

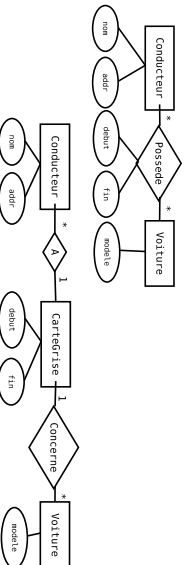
## TD Généalogie

- Objectif : représentation des informations de filiation plus généralement analyses des familles...
- extraction d'arbres généalogiques pour un individu
- individus : lieu et date de naissance, lieu et date de décès, parents naturels, légal ou nourriciers, frères et sœur, demi-frères et demi-sœurs...
- familles : mariages, enfants, dissolution
- liaisons : enfants, témoins, prêtre ou officier d'états civils, parrainages, mariages, participation à une réunion de famille...

13

## Solutions équivalentes...

- promotion d'une association en entité
- qu'est-ce qui à le droit d'être une entité ?



15

## Cardinalité max 1-1

correspondance unique entre entités



## Cardinalité max 1-n

lien hiérarchique, regroupement



## Cardinalité max n-n

graphe complet d'associations



10

## TD Magasin

- boutique sur le web
- l'internaute est identifié par un cookie
- différents produits onéreux sont proposés à l'internaute
- il peut les ajouter à son panier
- il peut passer commande, auquel cas il doit s'identifier
- une adresse de facturation, plusieurs adresses de livraison
- l'état de la commande peut être suivie en ligne

12

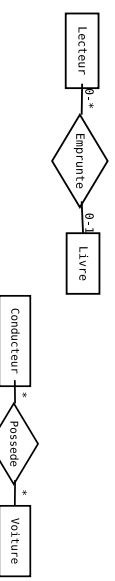
## TD Internet Movie Data Base (IMDB)

- objectif : base d'informations sur les films et leurs variantes
- filmographies de tous les artistes
- notation et commentaires des internautes
- œuvres cinématographiques
- avec des versions de films, de bandes sons, de sous-titres...
- informations techniques éventuelles
- artistes impliqué dans les œuvres
- metteur en scène, acteurs et leurs rôles, scénaristes...
- noms crédifiés : *Charles Chaplin*, *C. Chaplin*, *Charlie Chaplin*
- internautes pouvant donner leur avis et des notes sur les films

14

## Subtilités...

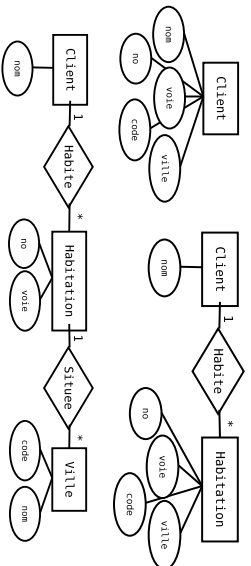
- un livre a plusieurs lecteur dans le temps
- où est le livre (dernier emprunt) vs historique des emprunts



16

## Subtilités (suite)

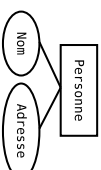
- ajout d'une notion d'habitation, de ville, de voie...
- contraintes structurées sur les données, mises à jour distinctes



17

## Exemple de modélisation insuffisante

Personne



(nom VARCHAR (32) ,  
 adr1 VARCHAR (32) ,  
 adr2 VARCHAR (32) ,  
 adr3 VARCHAR (32) ) ;

nom	adr1	adr2	adr3
Calvin	12 rue Verneux	77 250 Moret/Luding	
Hobbes	Building A	14 rue Verneux	77 Moret sur Loing
Susy	Mairie 77250 Moret		
Moe	15 rue Moret	77210 Verneux	

logique adresses postales imprimées sur trois lignes...

problème contenu des attributs non homogène (ville, code postal)

conséquence personnes à l'over ? requête sur trois attributs !

18

## Table informellement reliées

Company  
 (name VARCHAR(16) ,  
 addr VARCHAR(32) ,  
 phone VARCHAR(16) ) ;

name	addr	phone
IBM	Paris	12345
Oracle	La Défense	54321

descr	start	duration	cost
IBM, export	2002-02-02	1 mon	\$10,000.00
IBM enginer	2003-03-03	2 mois	\$8,000.00
Ibm cook	2004-04-04	1 day	\$250.00
oracle seer	2001-01-01	1 day	\$600.00

- le nom du job référence implicitement une entreprise
- le nom de l'entreprise n'est pas homogène

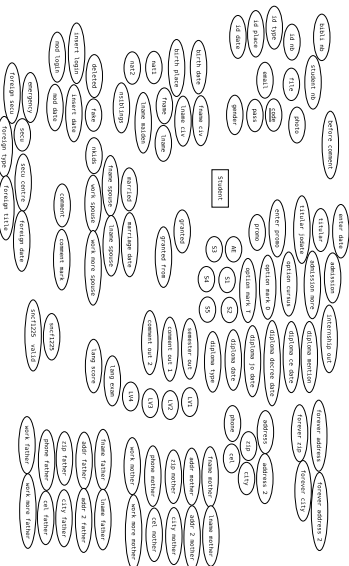
19

## Problèmes multiples

- séparateurs , ; de DU du et / -, espaces, lignes
- factures simple ou multiples (jusqu'à 3)
- préfix Fac. FAC. no Nos, erreurs, inhomogène
- numéro absent, ou Fac. ? Fac en attente EN ATTENTE
- dates absente, 20\_03\_70 20/03/70 20\_03\_1970 20\_03\_1970\_20\_03\_1970, erreur
- misc prédictions avec ou sans parenthèses, au milieu ou à la fin
- mode de paiement, montant (monnaie ? HT/TTTC ?), contenu

21

## Relation STUDENT



23

## Modélisation insuffisante (suite)

- application de suivi du matériel (quoi, où, quand, garanties...)
- attribut texte n\_facture (numéro de facture) table machine pas de contrainte sur le champs, 400 lignes (tout petit !)
- factures classées par date... indispensable

numero de facture	date	description
10283V7TU21.10.85;		
9221484.13.11.89		
70015048.13.11.92.14.EI.CJ		
20015048.13.11.92.14.EI.CJ		
Fac:13912.22.1232.0808.9711		
FA95601529.2501358.073681.02.TT01::		
961725.04.22/01/1996		
36201919 TF.18.07.96		
Fac:LVRY/251.12.02.97.CHELOUE.RABINE(S)		
Fac:29710691.23.10.1997		
Fac:7		
Fac:FA09147.01.07.96;		
Fac:3910066.19.10.96.07426.96.FTTO;		
96003378.10.12.99.0419.FHTJ.96003420.16.12.99.0425.FHTJ		
340128091TF.20.07.00		
340128091TF.20.07.00		
LVF7010162.06.04.2001		
FL162.053968.70402003.06.433.62.EI		
Nos.16324#.1633.04.30/22/2004		
521107.06/12/2005		
EN.ATTENTE		
Fac:7		
LVF9602659.27.09.9		

20

## Sous modélisation

- relations STUDENT PROF USER
- données communes code, mot de passe, coordonnées, email...
- description non homogène de certains concepts COUNTRY pour PROF, pas pour les autres ADDR2 pour STUDENT, pas pour PROF

22

## 10 manières de casser vos données (Joshua Berkus, 2008)

1. one big spreadsheet
2. random naming
3. no keys
4. no foreign keys
5. no constraints
6. non atomic fields
7. polymorphic fields
8. magic number 2010-02-30
9. EAV and E-libs
10. runaway ORM

24

### Nommage aléatoire

**singulier/pluriel** nom des relations

`USERACCOUNTS` vs `USERACCOUNT`

faire un choix et s'y tenir !

deux écoles...

**style de nommage** de même, être homogène

`camel` `nomUtilisateur lastLogin`

`souligné` `nom_utilisateur last_login`

`compressé` `mUliSt lStLogin lStLog`

`super-compressé` `nu ll`

`hongrois` `strNomUtilisateur boolLastLogin`

`langue` `userName dans la table Utilisateur`

25

### Attributs multi-dimensionnels

**nom** *Fabien Coelho, F Coelho, Fabien L. J Coelho, Coelho Fabien*

séparation nom et prénom ? identification du nom ?

**état** *a admin, u user, i inactif*

quels sont les administrateurs inactifs ?

**compatibilité** non orthogonale des concepts

26

### EAV : Entity-Attribute-Value

id	prop	val
123	sexe	m
123	marlé	oui
123	taille	173 cm
124	marlé	non
124	taille	1,60 m

– exploitation pour une question ?  
*quels sont les hommes marlés ?*  
 – attributs obligatoires ?  
 – expression des contraintes ?

27

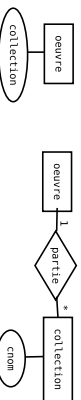
### EAV 2.0 : E-Blocks

Sérialisation : HTML, XML, json, java...

id	props
123	<sexe="m"><taille="173 cm"><marlé="oui">
124	<marlé="non"><taille="1,60 m">

28

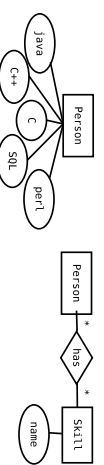
### Abstraction d'un attribut : entité périphérique



- nommage des valeurs, partage : liste vs champs libre
- utile pour les interfaces : menus PHP/Java
- coût faible : petit cache, gros espace et accès disque
- utiliser systématiquement si liste de valeurs textuelles ?
- à éviter pour dates, entiers, flottants

29

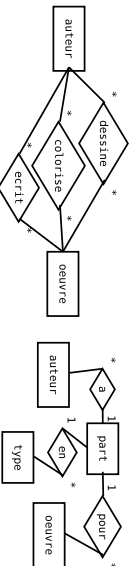
### Abstraction d'attributs



- collection de booléens des compétences
- attributs transformés en valeurs, boolean → association
- facilement extensible

30

### Abstraction d'associations



- ensemble d'association transformé en entité avec type
- étoile de cardinalité 1 sur une entité

31

### Prévoir l'imprévisible ? !

**souplesse nécessaire** nouveaux usages, ou oubliés

**attribut(s) libre(s)** à côté d'attributs contraints ?

*remarques, divers, commentaires, notes, compléments...*

problème CNIL si nominatif

**application** prévoir vérification de la **version** du schéma

- table `VERSION` contenant un simple entier
- détecte si modification ultérieure...

32

## Conclusion

- outils très simplifiés (par rapport au modèle complet)
- entité** objets, concepts nommés
- association** liens entre objets, verbes
- cardinalité** des associations
- attribut** données des entités et des associations
- l'important est le modèle représentant le réel
- l'outil n'est qu'un outil

33

## List of Slides

- 1 Modélisation d'un schéma de données
- 2 Le socle des SI est souvent pourri!
- 3 Coût pour maintenir une donnée à jour ?
- 4 Coût d'une donnée pas à jour ?
- 5 Modélisation d'un schéma de données
- 6 Plusieurs niveaux
- 7 Modélisation Entité/Association (E/A)
- 8 Modélisation E/A
- 9 Cardinalités
- 10 Cardinalité max 1-1
- 10 Cardinalité max 1-n

- 10 Cardinalité max n-n
- 11 TD Recettes de cuisine
- 12 TD Magasin
- 13 TD Génétique
- 14 TD Internet Movie Data Base (IMDB)
- 15 Solutions équivalentes...
- 16 Subtilités...
- 17 Subtilités (suite)
- 18 Exemple de modélisation insuffisante
- 19 Table informellement reliées
- 20 Modélisation insuffisante (suite)
- 21 Problèmes multiples
- 22 Sous modélisation

- 23 Relation STUDENT
- 24 10 manières de casser vos données (Joshua Berkus, 2008)
- 25 Nommage aléatoire
- 26 Attributs multi-dimensionnels
- 27 EAV : Entity-Attribute-Value
- 28 EAV 2.0 : E-blobs
- 29 Abstraction d'un attribut : entité périphérique
- 30 Abstraction d'attributs
- 31 Abstraction d'associations
- 32 Prévoir l'imprévisible ?!
- 33 Conclusion