



Modélisation d'un schéma de données

Composé avec \LaTeX , révision 2837

1

Le socle des SI est souvent pourri !



2

Coût pour maintenir une donnée à jour ?



3

Coût d'une donnée pas à jour ?



4

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 (détournement d'attributs)

5

Plusieurs niveaux

- externe** vue (partielle ?)
pour utilisateurs et applications
- conceptuel** vue sémantique des données
pour les concepteurs du SI *schéma graphique*
- logique** organisation de la base
pour les développeurs *relations*
- physique** implémentation...
types, index, disques...
pour les administrateurs *tables*

6

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

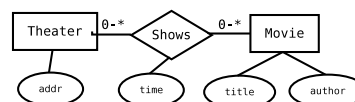
démarche bon sens et expérience

alternative UML (riche, complexe)

7

Modélisation E/A

- Entité objet identifié comme tel (**nom**)
- Association lien entre 2 entités (**verbe**), éventuellement réflexif
- Attribut propriété d'une entité ou association
- cardinalité** dénombrement des liens entre entité



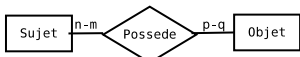
8

Cardinalités

nombre de liaisons entité-association

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

exemple générique



- un *sujet* possède entre *n* et *m* objets
- un *objet* est possédé par entre *p* et *q* sujets

attention côté inversé par rapport à UML !

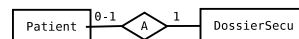
min/max	1	*
0	?	*
1	1	+

codage simplifié

cardinalités max d'une association entre des entités

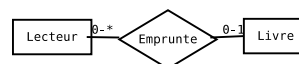
Cardinalité max 1-1

correspondance unique entre entités



Cardinalité max 1-n

lien hiérarchique, regroupement



Cardinalité max n-n

graphe complet d'associations



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...

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

TD Généalogie

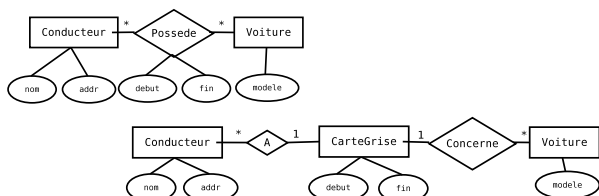
- objectif : représentation des informations de filiation
plus généralement analyses des familles...
extraction d'arbres généalogiques pour un individu
- individus : lieu et date de naissance, lieu et date de décès, parents
naturels, légal ou nourriciers, frères et sœur, demi-frères et
demi-sœurs...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

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 crédités : *Charles Chaplin, C. Chaplin, Charlie Chaplin*
- internautes pouvant donner leur avis et des notes sur les films

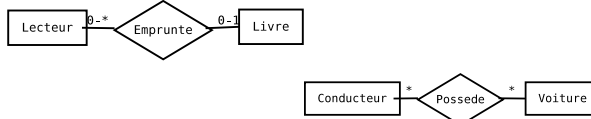
Solutions équivalentes...

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



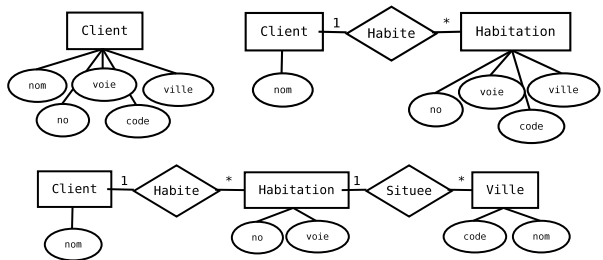
Subtilités...

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



Subtilités (suite)

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



Exemple de modélisation insuffisante

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

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

- logique** adresses postales imprimées sur trois lignes...
- problème** contenu des attributs non homogène (ville, code postal)
- conséquence** personnes à Moret ? requête sur trois attributs !

Table informellement reliées

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

name	addr	phone
IBM	Paris	12345
Oracle	La Défense	54321

Job
(descr VARCHAR (32),
start DATE,
duration INTERVAL,
cost MONEY);

descr	start	duration	cost
I.B.M. expert	2002-02-02	1 mon	\$10,000.00
IBM enginer	2003-03-03	2 mons	\$8,000.00
lbn 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

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

numéro de facture	Fac.FAFC0380440 (67 310.00 FHT) 30.07.98
10283-VT DU 21.10.85:	Fac.FA0714
9221424 13.11.89	Fac.FA0747 - 01.07.98;
70015048 13.11.92 (4 ELC)	Fac.9810060 19.10.98 (7428.96 FTTC);
.70015048 13.11.92 (4 ELC)	99003378 10.12.99 (4819 FHT);99003420 16.12.99 (4925 FHT)
Fac.13912 22.12.92 (9678.71)	3A012609TF 20/07/2000
FA95001529 29.03.95 (73081.32 TTC)::	3A012609TF 20.07.00
96125 du 22/01/1996	LVF0102185 06-04-2001
36201919 TF 18.07.96	F182 0539048 7/04/2003 (de 433.62 E)
Fac.IVRY051 12.02.97 (CHEQUE ARMINES)	Nos 1632 et 1633 du 30/12/2004
Fac.29710591 23.10.1997	521107 06/12/2005
LVF9909258 27.09.9	EN ATTENTE
	Fac. ?

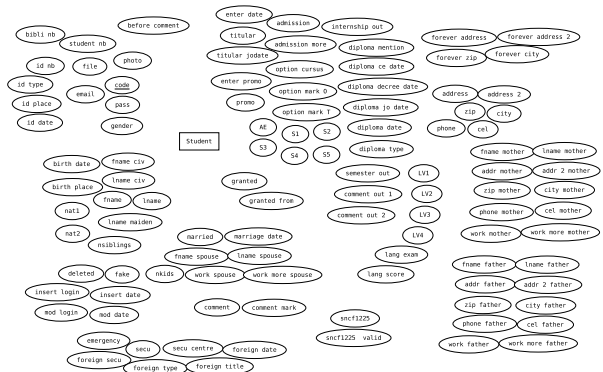
Problèmes multiples

- 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** précisions avec ou sans parenthèses, au milieu ou à la fin
mode de paiement, montant (monnaie ? HT/TTC ?), contenu

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

Relation STUDENT



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-blobs
10. runaway ORM

Nommage aléatoire

singulier/pluriel nom des relations

UserAccounts vs UserAccount deux écoles...
 faire un choix et s'y tenir!

style de nommage de même, être homogène

- camel** nomUtilisateur lastLogin
- souligné** nom_utilisateur last_login
- compressé** nmUtilst lstLogin lstLog
- super-compressé** nu ll
- hongrois** strNomUtilisateur boolLastLogin
- langue** userName *dans la table Utilisateur*

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 ?

comptabilité non orthogonalité des concepts

EAV : Entity-Attribute-Value

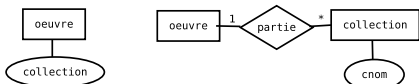
id	prop	val	
123	sexe	m	— exploitation pour une question ?
123	marié	oui	<i>quels sont les hommes mariés ?</i>
123	taille	173 cm	— attributs obligatoires ?
124	marié	non	— expression des contraintes ?
124	taille	1,60 m	

EAV 2.0 : E-blobs

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

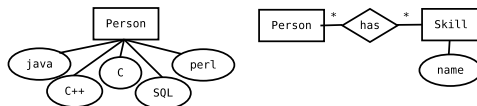
id	props
123	<sexe="m"><taille="173 cm"><marié="oui">
124	<marié="non"><taille="1,60 m">

Abstraction d'un attribut : entité périphérique



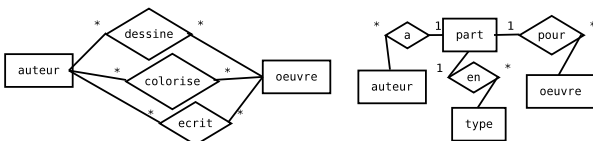
- nommage des valeurs, partage : liste vs champs libre
- utile pour les interfaces : menus PHP/Java
- coût faible : petit cache, gros espace et accès disque
- utiliser systématiquement si liste de valeurs textuelles ?
 à éviter pour dates, entiers, flottants

Abstraction d'attributs



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

Abstraction d'associations



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

Prévoir l'imprévisible !

souplesse nécessaire nouveaux usages, ou oublis

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...

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

- | | |
|--|--|
| <ul style="list-style-type: none"> 10 Cardinalité max n-n 11 TD Recettes de cuisine 12 TD Magasin 13 TD Généalogie 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 | <ul style="list-style-type: none"> 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 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 |
|--|--|