



Le socle des SI est souvent pourri !

Modélisation d'un schéma de données

Fabien Coelho, Claire Medrala

Mines Paris – PSL

Décembre 2023

Modélisation

Intro

E/A

Exercices

Exemples

Conclusion



Modélisation d'un schéma de données

Objectifs

Modélisation d'un schéma de données

4 niveaux

Modélisation

Intro

E/A

Exercices

Exemples

Conclusion

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)

Modélisation

Intro

E/A

Exercices

Exemples

Conclusion

- réel** complexité du monde pour un domaine thématique
- 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

Réel			
Modèle conceptuel	<ul style="list-style-type: none"> • Indépendant du modèle de données • Indépendant du SGBD 		
Modèle logique	<ul style="list-style-type: none"> • Dépendant du modèle de données • Indépendant du SGBD 	Relationnel	Objet XML
Modèle Physique	<ul style="list-style-type: none"> • Dépendant du modèle de données • Dépendant du SGBD 	<ul style="list-style-type: none"> • Organisation physique des données • Structures de stockage des données • Structures accélératrices (index) 	

B. Nguyen - INRIA



Modélisation Entité/Association (E/A)

- Modélisation
- Intro
- E/A
- Exercices
- Exemples
- Conclusion

origine Peter Chen 1976
The Entity-Relationship Model, Toward a Unified View of Data
 ACM Transactions on Database Systems (TODS)

- avantages** méthode graphique pour fixer les idées
- simple, intuitive, peu de concepts
 - approche **ascendante**, regroupement de parties
 - démarche : **bon sens** et **expérience**
 - notation MCD de la méthode Merise

- inconvénients** quelques manques
- démarche non déterministe : pas d'unicité du modèle
 - pas de contraintes d'intégrité
 - pas de procédure / méthode

alternative UML *Unified Modeling Language* : riche, complexe

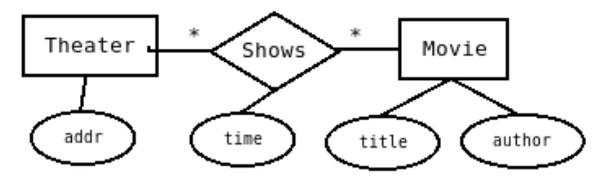
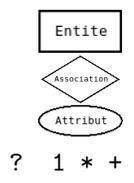


Modélisation entité-association (E/A)

4 concepts

- Modélisation
- Intro
- E/A
- Exercices
- Exemples
- Conclusion

- entité** objet identifié comme tel (**nom**)
- association** lien entre 2 entités (**verbe**)
- attribut** propriété d'une entité ou association
- cardinalité** dénombrement des liens entre entité



Cardinalités

- Modélisation
- Intro
- E/A
- Exercices
- Exemples
- Conclusion

nombre de liaisons entité-association



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

notation simplifiée

? 1 * +

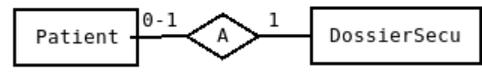
min/max	1	*
0	?	*
1	1	+



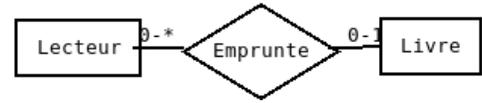
Cardinalités maximales

- Modélisation
- Intro
- E/A
- Exercices
- Exemples
- Conclusion

1:1 entités équivalentes



1:n ou n:1 lien hiérarchique, regroupement



n:n liaisons multiples





TD

Recettes de cuisine



TD

Magasin

Modélisation

Intro

E/A

Exercices

Exemples

Conclusion

- 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...

9 / 29

Modélisation

Intro

E/A

Exercices

Exemples

Conclusion

- 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

10 / 29



TD

Généalogie



TD

Internet Movie Data Base (IMDB)

Modélisation

Intro

E/A

Exercices

Exemples

Conclusion

- 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...

11 / 29

Modélisation

Intro

E/A

Exercices

Exemples

Conclusion

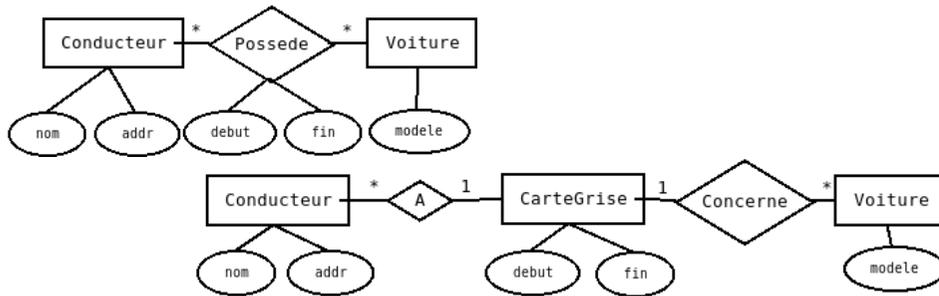
- objectif : base d'informations sur les films et leurs variantes
filmographies de tous les artistes
notation et commentaires des internautes
- œuvres cinématographiques
avec des versions de films, de bandes sons, de sous-titres...
informations techniques éventuelles
- artistes impliqué dans les œuvres
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

12 / 29



Solutions équivalentes...

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

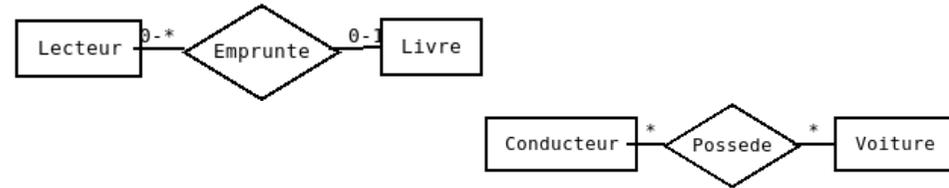


13 / 29



Subtilités...

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

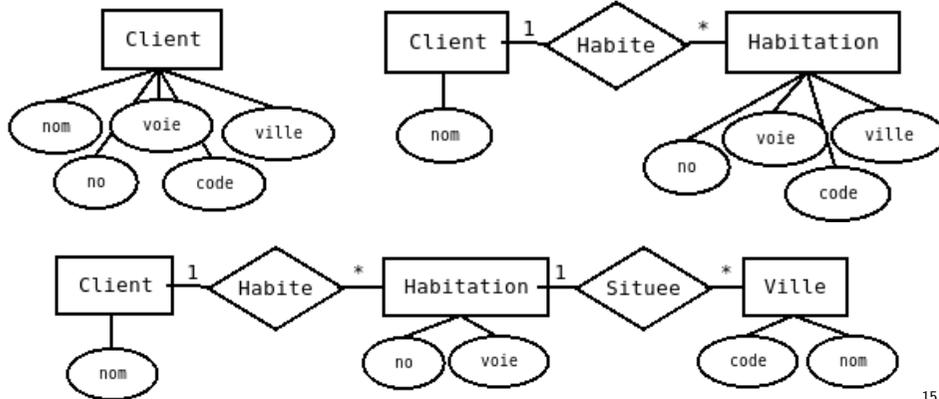


14 / 29



Subtilités (suite)

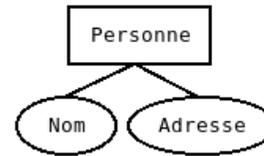
- ajout d'une notion d'habitation, de ville, de voie...
- contraintes structurelles sur les données, mises à jour distinctes



15 / 29



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	
Susy	Mairie 77250 Moret		77 Moret sur Loing
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 !

16 / 29



Table informellement reliées

Modélisation
Intro
E/A
Exercices
Exemples
Conclusion

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 FLOAT);
```

descr	start	duration	cost
I.B.M. expert	2002-02-02	1 mon	10000
IBM engineer	2003-03-03	2 mons	8000
Ibm cook	2004-04-04	1 day	250
oracle seer	2001-01-01	1 day	600

- le nom du job référence **implicite**ment une entreprise
- le nom de l'entreprise n'est pas homogène



Modélisation insuffisante (suite)

Modélisation
Intro
E/A
Exercices
Exemples
Conclusion

- 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
10283-VT DU 21.10.85;
9221424 13.11.89
70015048 13.11.92 (4 ELC)
.70015048 13.11.92 (4 ELC)
Fac.13912 22.12.92 (9678.71)
FA95001529 29.03.95 (73081.32 TTC); ;
96125 du 22/01/1996
36201919 TF 18.07.96
Fac.IVRY/351 12.02.97 (CHEQUE ARMINES)
Fac 29710591 23.10.1997
LVF9909258 27.09.9

Fac.FAC980440 (67 310.00 FHT) 30.07.98
Fac.FA0714
Fac.FA0747 - 01.07.98;
Fac.9810060 19.10.98 (7428.96 FTTC);
99003378 10.12.99 (4819 FHT),99003420 16.12.99 (4925 FHT)
3A012609TF 20/07/2000
3A012609TF 20.07.00
LVF0102165 06-04-2001
F182 0539048 7/04/2003 (de 433,62 E)
Nos 1632 et 1633 du 30/12/2004
521107 06/12/2005
EN ATTENTE
Fac.?
```



Problèmes multiples

Modélisation
Intro
E/A
Exercices
Exemples
Conclusion

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

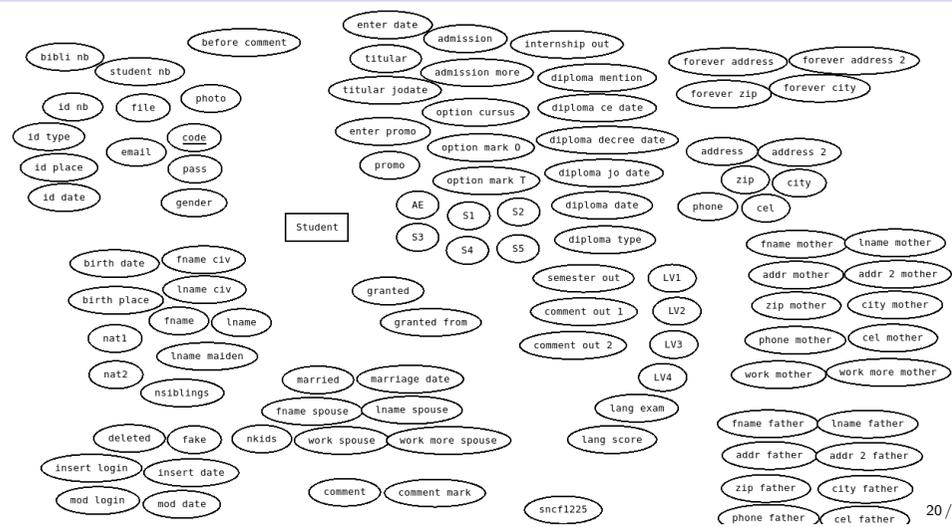
dates absente, 20.03.70 20/03/70 20.03.1970 20-03-1970, erreur

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



Relation STUDENT

Modélisation
Intro
E/A
Exercices
Exemples
Conclusion





Nommage aléatoire

- Modélisation
- Intro
- E/A
- Exercices
- Exemples
- Conclusion

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** nmUtlst lstLogin lstLog
- compressé²** nu ll
- hongrois** strNomUtilisateur boolLastLogin
- langue** userName *dans la table Utilisateur*



Attributs multi-dimensionnels

- Modélisation
- Intro
- E/A
- Exercices
- Exemples
- Conclusion

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

- Modélisation
- Intro
- E/A
- Exercices
- Exemples
- Conclusion

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

- exploitation pour une question ?
quels sont les hommes mariés ?
- attributs obligatoires ?
- expression des contraintes ?



EAV 2.0 : E-blobs

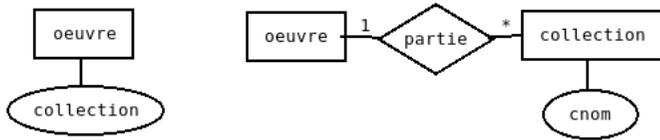
- Modélisation
- Intro
- E/A
- Exercices
- Exemples
- Conclusion

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

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



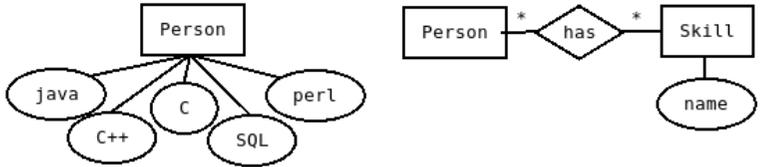
Abstraction d'un attribut : entité périphérique



- 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



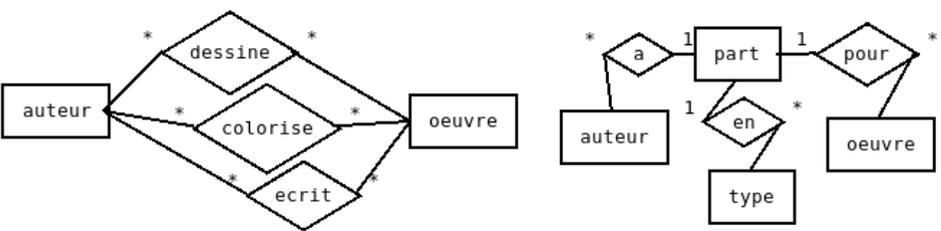
Abstraction d'attributs



- 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

Modélisation

Intro

E/A

Exercices

Exemples

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