



Bases de données

Master 1 Humanités numériques
2021-2022

Jérôme Darmont

<https://eric.univ-lyon2.fr/jdarmont/>

Actualité du cours



https://eric.univ-lyon2.fr/jdarmont/?page_id=3139



<https://eric.univ-lyon2.fr/jdarmont/?feed=rss2>



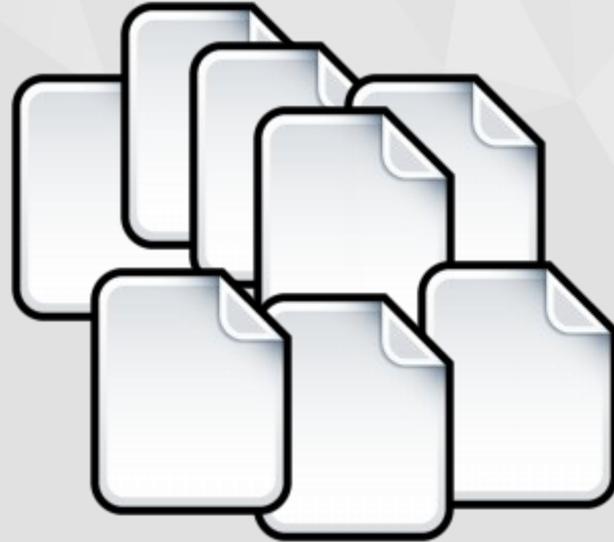
https://twitter.com/darmont_lyon2 #hnbd

Définition

Base de données (BD) : Collection de données
cohérentes et **structurées**

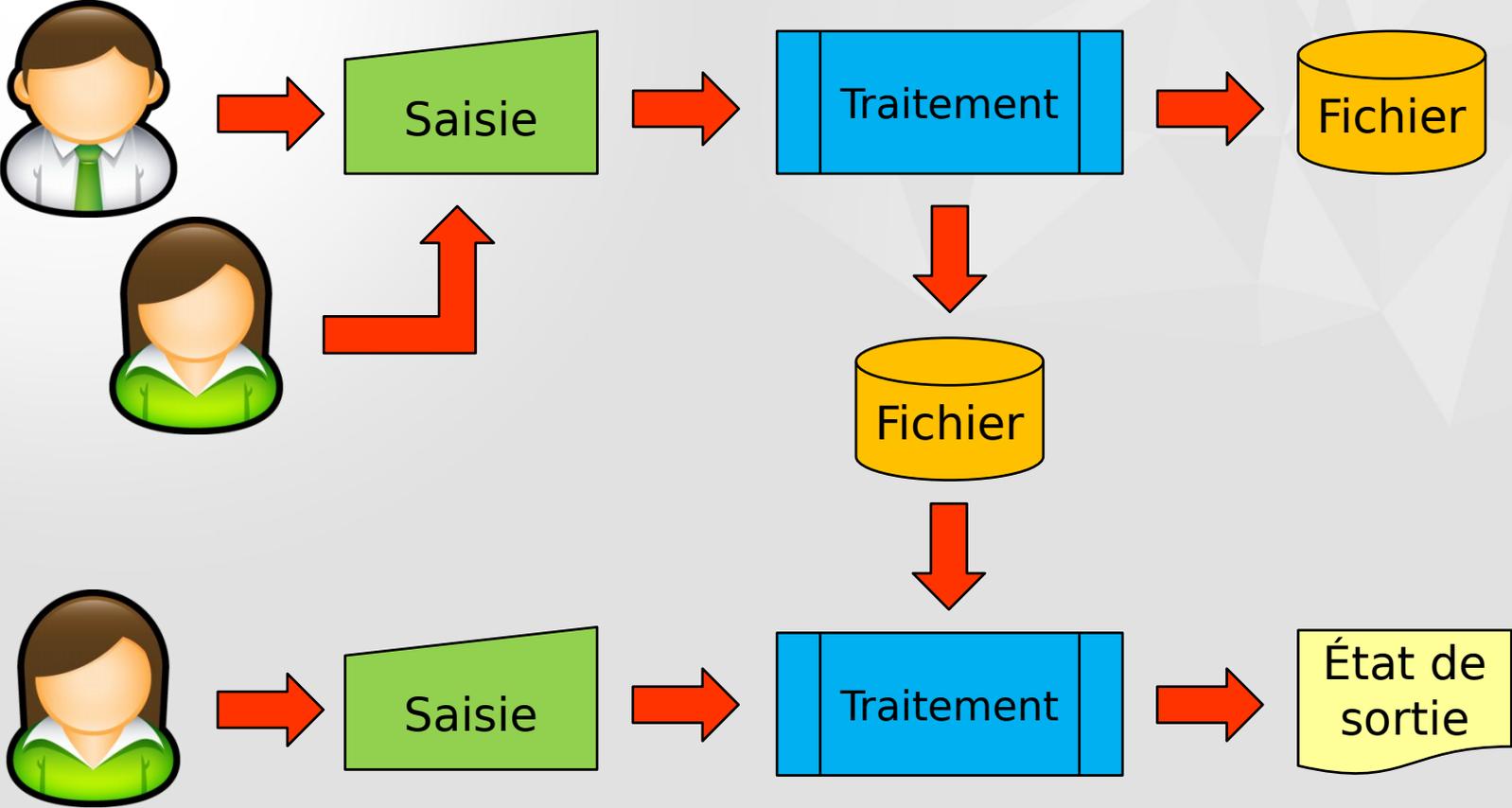


Base de données

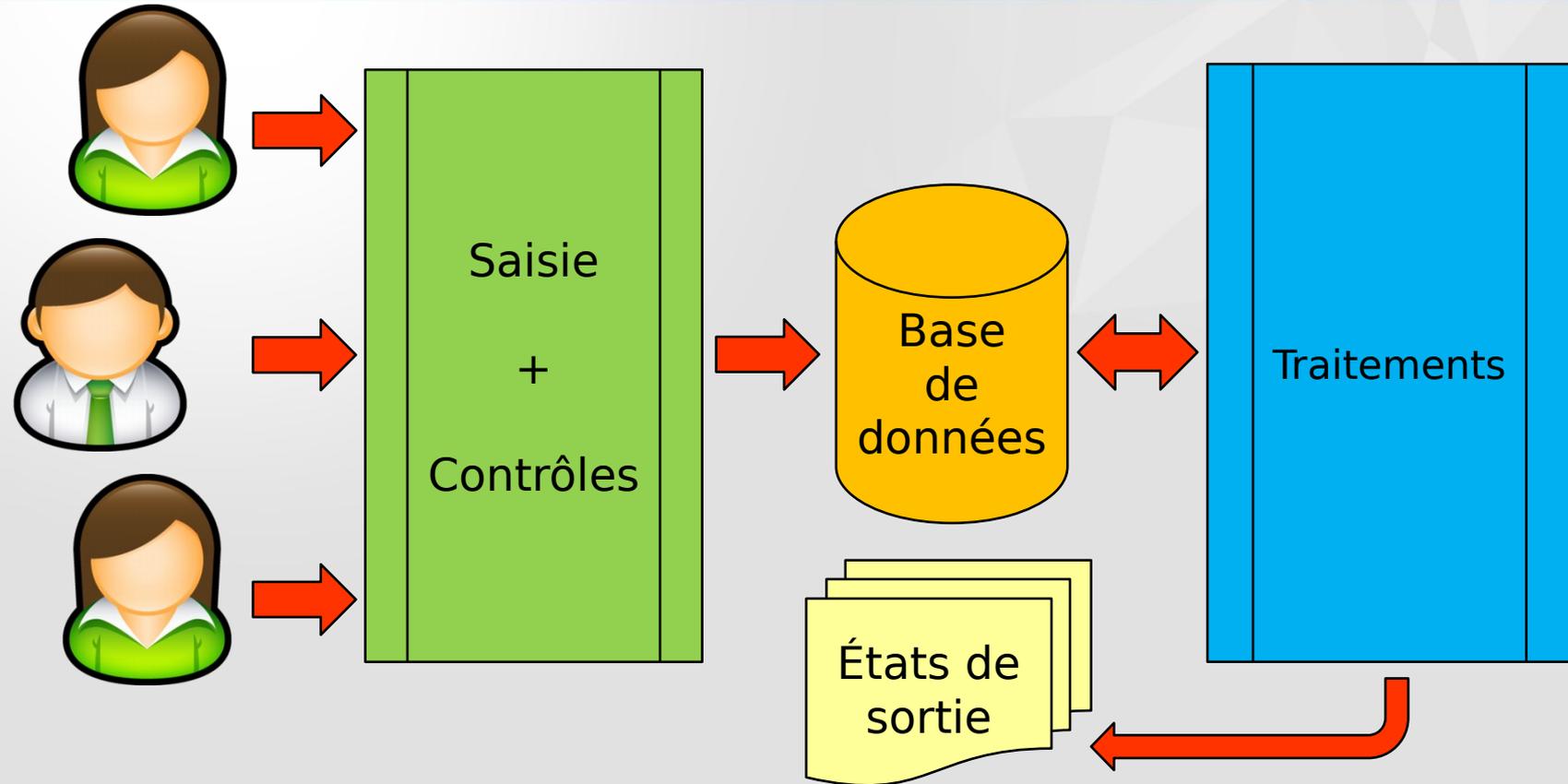


Fichiers

Organisation en fichiers



Organisation en BD



Avantages de l'organisation en BD

- ▶ Uniformisation de la saisie
- ▶ Standardisation des traitements
- ▶ Contrôle de la validité des données
- ▶ Partage de données entre plusieurs traitements

Propriétés de l'organisation en BD

- ▶ Usage **multiple** des données
- ▶ **Accès** facile, rapide, protégé, souple, puissant
- ▶ **Coût réduit** de stockage, de mise à jour et de saisie
- ▶ Disponibilité, exactitude, **cohérence** et protection des données ; **non redondance**
- ▶ **Évolution** aisée et protection de l'investissement de programmation
- ▶ **Indépendance** des données et des programmes
- ▶ Conception *a priori*

Qu'est-ce qu'un SGBD ?

- ▶ **Système de Gestion de Bases de Données :** Logiciel(s) assurant structuration, stockage, maintenance, mise à jour et consultation des données d'une BD
- ▶ **Exemples**
 - SGBD « bureautiques » : Access, Base, Filemaker, Paradox...
 - SGBD serveurs : Oracle, DB2, SQL Server, PostgreSQL, MySQL, MariaDB...

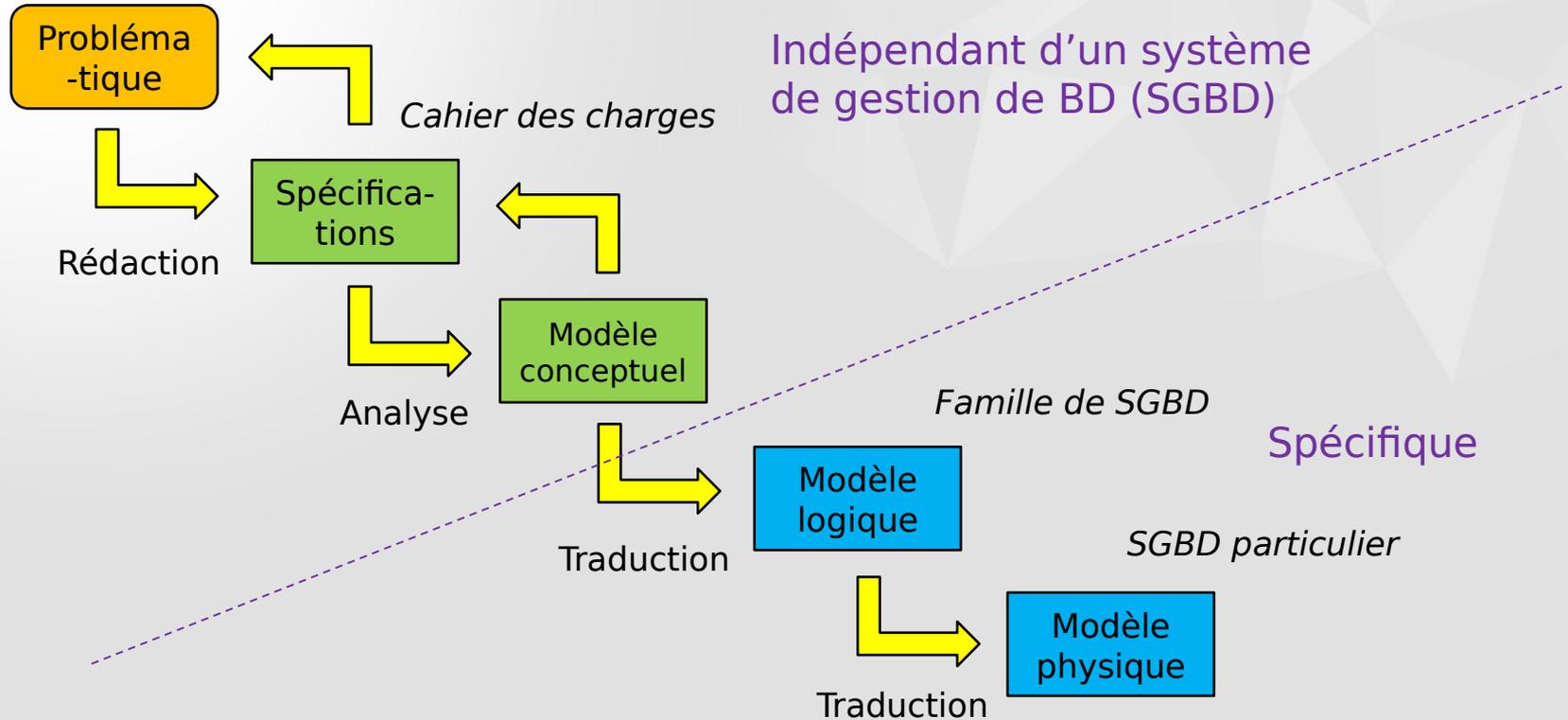
Objectifs des SGBD

- ▶ Indépendance physique
- ▶ Indépendance logique
- ▶ Manipulation facile des données
- ▶ Administration facile des données
- ▶ Efficacité des accès aux données
- ▶ Redondance contrôlée des données
- ▶ Cohérence des données (contraintes d'intégrité)
- ▶ Partage des données
- ▶ Sécurité des données

Fonctions des SGBD

- ▶ Description des données : Langage de Définition de Données (LDD)
 - ▶ Recherche des données
 - ▶ Mise à jour des données
 - ▶ Transformation des données
 - ▶ Contrôle de l'intégrité des données
 - ▶ Gestion de transactions et sécurité
- } Langage de Manipulation de Données (LMD)
- } Langage de Contrôle de Données (LCD)

Processus de conception d'une BD



Plan du cours

▶ Séances 1 et 2 : Conception de bases de données



PublisherID	Publisher	PubAddress
03-420322	Random House	123 4th Street, New York
04-773303	Wiley and Sons	45 Lincoln Blvd, Chicago
05-409323	O'Reilly Press	77 Boston Ave, Cambridge
03-392088	City Lights Books	59 Market, San Francisco

AuthorID	AuthorName	AuthorDOB
1452023310	Heidi Pitlorick	18-Feb-53
392-48-9985	Joe Blow	14-Mar-15
454-22-4512	Gale Hemmings	12-Sep-10
863-59-1254	Hannah Arendt	17-Mar-08

ISBN	AuthorID	PublisherID	Date	Title
1-34532-402-1	345-28-2939	03-4472622	1990	Cold Fusion for Dummies
1-38463-989-1	352-48-9895	04-773303	1985	Maneats and Blow-Tops
2-35921-499-4	454-22-4912	03-4899223	1982	Fluid Dynamics of Aqueducts
1-38278-292-4	463-59-1254	03-392088	1987	Bases, Bases, & Phenomena

Modèle relationnel

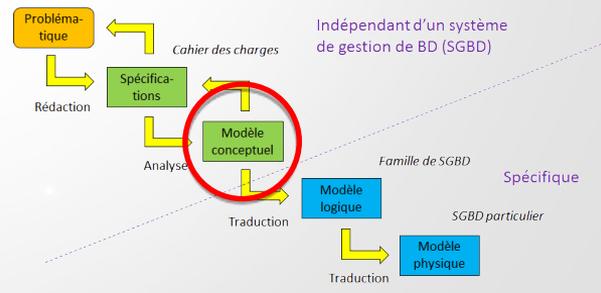


▶ Séances 3 et 4 : Interrogation de bases de données



Partie 1

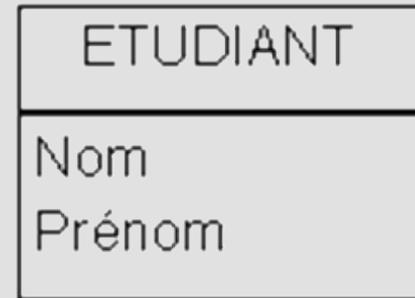
Modélisation conceptuelle



Modèle conceptuel UML

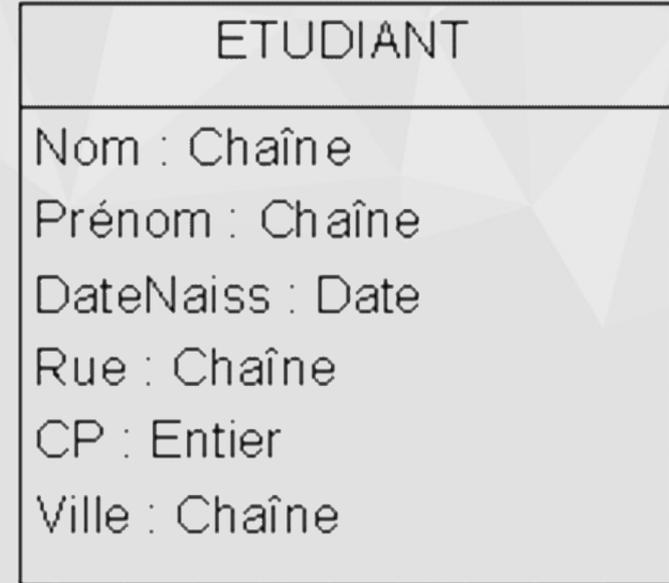
- ▶ **Standard** de l'Object Management Group
- ▶ Ensemble de **formalismes graphiques**
- ▶ **Diagramme de classes**

- ▶ **Classe** : Groupe d'entités du monde réel ayant les mêmes caractéristiques et le même comportement
ex. ETUDIANT
- ▶ **Attribut** : Propriété de la classe
ex. Nom et Prénom de l'étudiant·e
- ▶ **Représentation graphique** :



► Type d'attribut :

- Nombre entier (**Entier**)
- Nombre réel (**Réel**)
- Chaîne de caractères (**Chaîne**)
- Date (**Date**)



Instances

- ▶ Objets (individus) de la classe ETUDIANT = les étudiant·es

<u>Nom</u>	<u>Prénom</u>	<u>DateNaiss</u>	<u>Etc.</u>
Dupont	Albertine	01/06/1993	...
West	James	03/09/1994	...
Martin	Marie	05/06/1995	...
Abidi	Rachid	15/11/1995	...
Titgoutte	Justine	28/02/1996	...
Dupont	Noémie	18/09/1995	...
Dupont	Albert	23/05/1990	...

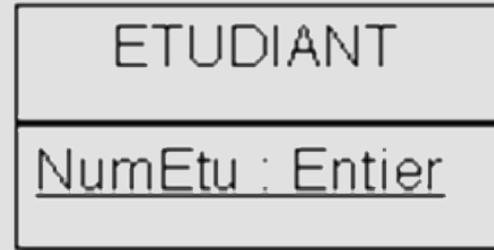
Problème : Comment distinguer les Dupont ?

Identifiant (1/2)

- ▶ Solution : Ajouter un attribut **numéro d'étudiant** !

<u>NumEtu</u>	<u>Nom</u>	<u>Prénom</u>	<u>DateNaiss</u>
1110	Dupont	Albertine	01/06/1993
2002	West	James	03/09/1994
3333	Martin	Marie	05/06/1995
4042	Durand	Rachid	05/11/1995
5552	Titgoutte	Justine	28/02/1996
6789	Dupont	Noémie	18/09/1995
7000	Dupont	Albert	23/05/1990

- ▶ Le numéro d'étudiant est un attribut **identifiant**.
- ▶ Un identifiant caractérise **de façon unique** les instances d'une classe.
- ▶ **Convention graphique :**
NB : Ne pas confondre avec les attributs de classe UML dont c'est la notation usuelle



- ▶ **Association** : liaison perçue entre des classes
ex. Les étudiant·es passent des épreuves.



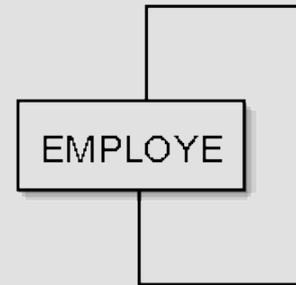
- ▶ Les classes ETUDIANT et EPREUVE peuvent être qualifiées de **participantes** à l'association PASSER.
- ▶ **Degré** ou **arité** d'une association : nombre de classes participantes.
En général : **associations binaires** (de degré 2).

- ▶ Une classe peut être associée à elle-même, chaque instance pouvant jouer plusieurs rôles dans l'association.

ex. Employés et supérieurs hiérarchiques

- ▶ **Rôle** : fonction de chaque classe participante (+).

+EST SUPERIEUR



HIERARCHIE

+EST SUBALTERNE

► **Définition** : Indicateur qui montre combien d'instances de la classe considérée peuvent être liées à une instance de l'autre classe participant à l'association

- 1 Un et un seul
- 0..1 Zéro ou un
- 0..* ou * Zéro ou plus
- 1..* Un ou plus
- M..N De M à N (M, N entiers)
ex. 4..10 (de 4 à 10)

- ▶ ex. Un·e étudiant·e possède une et une seule carte Izly. Cette dernière n'est possédée que par un·e seul·e étudiant·e.



Lire « Un·e étudiant.e possède **multiplicité (1)** carte Izly ».

- ▶ ex. Une épreuve relève d'une et une seule matière.
Une matière peut donner lieu à plusieurs épreuves.



NB : La multiplicité un à plusieurs (1..*) peut aussi être zéro à plusieurs (0..* ou *).

- ▶ ex. Un·e étudiant·e peut appartenir ou non à un groupe de TD. Un groupe de TD réunit plusieurs étudiant·es.



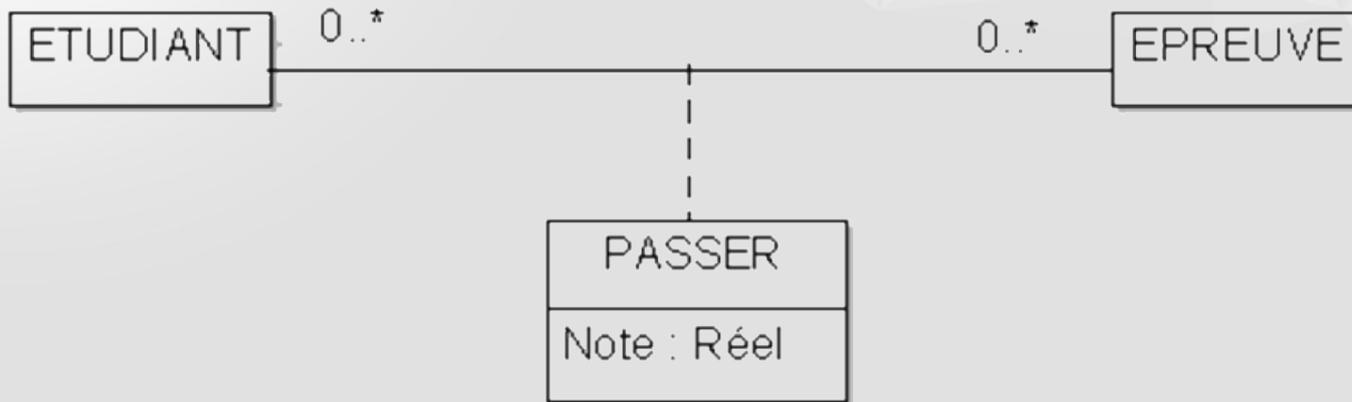
NB : La multiplicité un à plusieurs (1..*) peut aussi être zéro à plusieurs (0..* ou *).

- ▶ ex. Un·e étudiant·e peut passer plusieurs épreuves. Une épreuve peut être passée par plusieurs étudiant·es.



NB : Les multiplicités un à plusieurs (1..*) peuvent aussi être zéro à plusieurs (0..* ou *).

- ▶ Il est possible de caractériser une association par des attributs.
ex. Un·e étudiant·e qui passe une épreuve obtient une note.



NB : Une classe-association demeure une association.

Exemple : Spécifications (1/2)

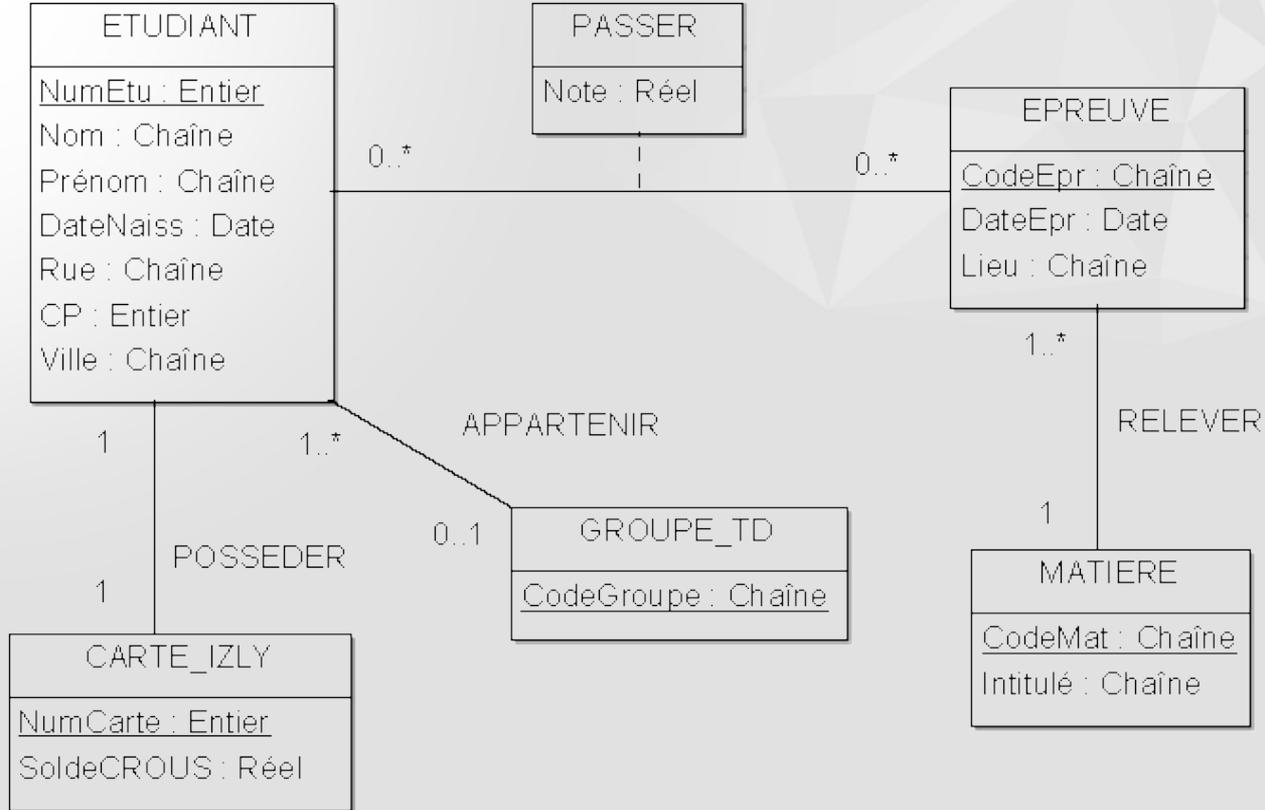
- ▶ Les étudiant·es sont caractérisé·es par un numéro unique, leur nom, prénom, date de naissance, rue, code postal et ville.
- ▶ Les étudiant·es possèdent une carte Izly caractérisée par un numéro unique et un solde d'argent utilisable au CROUS.
- ▶ Selon qu'ils ou elles sont dispensé·es ou non d'assiduité, les étudiant·es appartiennent à un groupe de TD caractérisé par un code unique.

Exemple : Spécifications (2/2)

- ▶ Les étudiant-es **passent** **des** épreuves et obtiennent une **note** pour chacune.
- ▶ Les **épreuves** sont caractérisées par un **code** **unique**, ainsi que la **date** et le **lieu** auxquels elles se déroulent.
- ▶ Chaque épreuve **relève** d'**une** matière **unique** (mais une matière donnée peut donner lieu à **plusieurs** épreuves).
- ▶ Les **matières** sont caractérisées par un **code** **unique** et un **intitulé**.

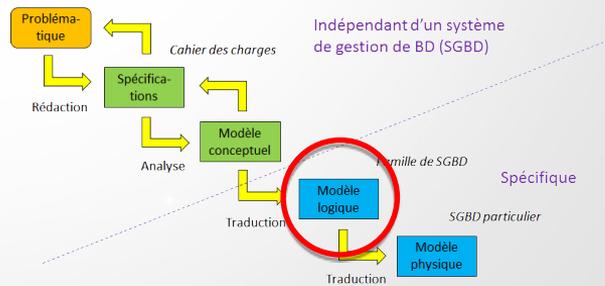
1. Identifier les **classes**
2. Identifier les **associations** entre les classes
3. Identifier les **attributs** de chaque classe et de chaque classe-association
4. Identifier et souligner l'**identifiant** de chaque classe
5. Évaluer les **multiplicités** des associations

Exemple : Diagramme de classes



Partie 2

Modélisation logique



Modèle logique relationnel

- ▶ **Modèle** associé aux SGBD relationnels
(ex. Oracle, SQL Server, DB2, MariaDB, Access...)
- ▶ **Objectifs du modèle relationnel**
 - Indépendance physique
 - Traitement du problème de redondance des données
 - Langages non procéduraux (faciles à utiliser)
 - Devenir un standard

Relations et attributs



ISBN	Publisher	PubAddress	AuthorID	AuthorName	AuthorCity	Title	Date
03-447-1822	Random House	121 4th Street, New York	140-26-2930	Marie Perle	14-Aug-92		
04-712001	Wiley and Sons	45 Lincoln Street, Chicago	39-24-8965	Joel Sizer	14-Mar-16		
03-4850223	O'Reilly Press	77 Boston Ave, Cambridge	454-24-4012	Betsy Hammonds	12-Sep-70		
02-302088	Ch. Light Books	89 Market, San Francisco	100-59-1251	Harrah Avenue	12-Mar-08		
1-34433-432-1	445-26-2930	03-4472822	1893	Cold Fusion for Dummies			
1-3842-895-1	392-48-9985	04-712001	1895	Murcine and Straw Tong			
2-39521-499-4	454-24-4012	03-4892223	1852	Fluid Dynamics of Aqueducts			
1-38278-293-4	463-05-1254	02-392088	1987	Reads, Baskets & Revolution			

Modèle relationnel

- ▶ Une relation R est un ensemble d'attributs $\{A_1, A_2, \dots, A_n\}$.
ex. La relation EPREUVE est l'ensemble des attributs $\{\text{CodeEpr}, \text{DateEpr}, \text{Lieu}\}$.
- ▶ Chaque attribut A_i prend ses valeurs dans un domaine $\text{dom}(A_i)$.
ex. $\text{Note} \in [0, 20]$
 $\text{Lieu} \in \{\text{'Amphi Say'}, \text{'Amphi Aubrac'}, \text{'Salle D101'}, \dots\}$



Publisher	Author
Random House	Jane Austen
Wiley and Sons	John Doe
O'Reilly Press	John Doe
Ch. Light Books	John Doe

► Notation d'une relation : $R (A_1, A_2, \dots, A_n)$

ex. EPREUVE (CodeEpr, DateEpr, Lieu)

► Un **n-uplet** t est un ensemble de valeurs $t = \langle V_1, V_2, \dots, V_n \rangle$ où $V_i \in \text{dom}(A_i)$ ou bien V_i est la valeur nulle (**NULL**).

ex. $\langle \text{'InfoS2'}, \text{'30-06-2016'}, \text{'Amphi Aubrac'} \rangle$ est un n-uplet de la relation EPREUVE.

Contraintes d'intégrité (1/2)

PubID	PubIsoter	PubAddress
03-447-2822	Random House	121 4th Street, New York
04-773063	Wiley and Sons	45 Lincoln Blvd, Chicago
03-4850223	O'Reilly Press	77 Boston Ave, Cambridge
03-302088	Ch. Light Books	89 Market, San Francisco

AuteurID	AuteurNom	AuteurDate
140-26-2930	Hale Gelasco	14-Aug-92
392-24-8960	Joel Bow	14-Mar-16
404-24-4012	Baily Hammings	12-Sep-70
100-59-1254	Hannah Akerst	12-Mar-08

Modèle relationnel

- ▶ **Clé primaire** : Ensemble d'attributs dont les valeurs permettent de distinguer les n-uplets les uns des autres.
ex. CodeEpr est clé primaire de la relation EPREUVE.
- ▶ **Clé étrangère** : Attribut qui est clé primaire d'une autre relation.
ex. Connaître la matière dont relève chaque épreuve
→ ajout de l'attribut CodeMat à la relation EPREUVE

Contraintes d'intégrité (2/2)



PubID	Publisher	PubAddress
03-447-2822	Random House	121 4th Street, New York
04-7733051	Wiley and Sons	45 Lincoln Street, Chicago
03-4858223	O'Reilly Press	77 Boston Ave, Cambridge
03-3020886	Ch. Light Books	89 Market, San Francisco

AuthorID	AuthorName	AuthorCity
140-26-2930	Hale Gessesse	14-Aug-92
392-48-8985	Joel Bow	14-Mar-16
454-24-4012	Billy Hammings	12-Sep-70
192-58-1251	Harrah Aeneas	12-Mar-08

BookID	AuthorID	BookID	BookID	BookID
1-34433-432-1	140-26-2930	03-4472822	1893	Cold Fusion for Dummies
1-3882-895-1	392-48-8985	04-7733051	1895	Murcine and Straw Timg
2-39521-489-4	454-24-4012	03-4899223	1892	Fluid Dynamics of Aquabucts
1-38278-293-4	463-05-1254	03-3920886	1987	Reads, Baskets & Revolution

- ▶ Notations : Clés primaires soulignées, clés étrangères postfixées par le caractère #.
ex. EPREUVE (CodeEpr, DateEpr, Lieu, CodeMat#)
- ▶ Contraintes de domaine : Les attributs doivent respecter une condition logique.
ex. Note ≥ 0 ET Note ≤ 20

Contraintes d'intégrité en pratique

Title	Author	Publisher	PubAddress	ISBN	AuthorID	PublisherID	EditionYear
03-447-2822	Random House	121 4th Street, New York					
04-7173903	Wiley and Sons	45 Lincoln Park, Chicago					
03-4850223	O'Reilly Press	77 Boston Ave, Cambridge					
02-3020886	Ch. Light Books	99 Market, San Francisco					

Modèle relationnel

EPREUVE

<u>CodeEpr</u>	DateEpr	Lieu	Codemat#
ECOS101	15/01/2016	Amphi Aubrac	ECO
ECOS102	16/01/2016	Amphi Aubrac	ECO
GESS201	25/05/2016	Salle 201	GES
INFOS101	20/01/2016	Salle 101	INFO

MATIERE

<u>CodeMat</u>	Intitulé
ECO	Économie
GES	Gestion
INFO	Informatique

Traduction UML-relationnel (1/4)

- ▶ Chaque **classe** devient une **relation**.
- ▶ Les **attributs** de la classe deviennent **attributs** de la relation.
- ▶ L'**identifiant** de la classe devient **clé primaire** de la relation.

ex. ETUDIANT (NumEtu, Nom, Prénom, DateNaiss, Rue, CP, Ville)

Traduction UML-relationnel (2/4)

- ▶ Chaque **association 1-1** est prise en compte en incluant la clé primaire d'une des relations participante comme **clé étrangère** dans l'autre relation.

ex. CARTE_IZLY (NumCarte, SoldeCROUS)

ETUDIANT (NumEtu, Nom, Prénom, DateNaiss, Rue, CP, Ville, **NumCarte#**)

Traduction UML-relational (3/4)

- ▶ Chaque **association 1-N** est prise en compte en incluant la clé primaire de la relation dont la **multiplicité maximale est 1** comme **clé étrangère** dans l'autre relation participante.

ex. EPREUVE (CodeEpr, DateEpr, Lieu, **CodeMat#**)

MATIERE (CodeMat, Intitulé)

Traduction UML-relational (4/4)

- ▶ Chaque **association M-N** est prise en compte en créant une nouvelle relation dont la clé primaire est la **concaténation des clés primaires** des relations participantes. Les attributs de la classe-association sont insérés dans cette nouvelle relation si nécessaire.

ex. PASSER (NumEtu#, CodeEpr#, Note)

Exemple : Modèle logique relationnel

PubID	Publisher	PubAddress
03-4472822	Randon House	121 4th Street, New York
04-4773901	Wiley and Sons	45 Lincoln Blvd, Chicago
03-4850223	O'Reilly Press	77 Boston Ave, Cambridge
02-3020886	Ch. Light Books	99 Market, San Francisco

AuthorID	AuthorName	PubID	PubTitle	PubDate
190-26-2930	Male Galesse	14-Aug-92		
39-24-8965	Joia Bow	14-Mar-16		
454-24-4012	Baily Hammonds	12-Sep-70		
190-26-2930	Male Galesse	14-Aug-92		

PubID	AuthorID	PubID	PubTitle	PubDate
1-34433-432-1	145-26-2930	03-4472822	1993	Cold Fusion for Dummies
1-38462-895-1	39-24-8965	04-7733903	1995	Mechanics and Draw: Trig
2-39921-499-4	454-24-4012	03-4892223	1992	Fluid Dynamics of Aqueducts
1-32278-293-4	463-05-1254	02-3920886	1987	Reads, Baskets & Revolution

Modèle relationnel

CARTE_IZLY (NumCarte, SoldeCROUS)

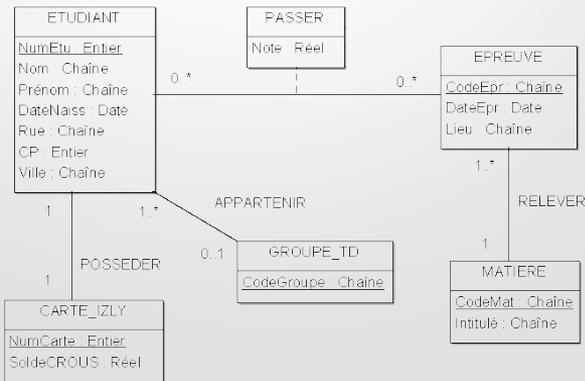
GROUPE_TD (CodeGroupe)

ETUDIANT (NumEtu, Nom, Prénom, DateNaiss, Rue, CP, Ville, NumCarte#, CodeGroupe#)

MATIERE (CodeMat, Intitulé)

EPREUVE (CodeEpr, DateEpr, Lieu, CodeMat#)

PASSER (NumEtu#, CodeEpr#, Note)



Traduction d'une association M-N

PubID	Publisher	PubAddress	AuthorID	AuthorName	AuthorCity	Title	Price
034473822	Random House	121 4th Street, New York	1340-26-2930	Hale Gessesse	14-Aug-92		
047729021	Wiley and Sons	45 Lincoln Street, Chicago	392-48-8985	Joel Broy	14-Mar-16		
034850223	O'Reilly Press	77 Boston Ave, Cambridge	454-24-4012	Billy Hammings	12-Sep-70		
023020804	Ch. Light Books	99 Market, San Francisco	160-59-1251	Hannah Jensen	12-Mar-08		

PubID	AuthorID	BookID	Book Title	Price
1-34433-432-1	145-26-2930	03-4472822	1993 Cold Fusion for Dummies	
1-34432-895-1	392-48-8985	04-7733063	1995 Microsoft and Java: The	
2-39521-489-4	454-24-4012	03-4892223	1992 Fluid Dynamics of Aqueducts	
1-38278-293-4	463-05-1254	02-3920886	1987 Reads, Baskets & Revolution	

Modèle relationnel

ETUDIANT

<u>NumEtu</u>	Nom	Prénom
1110	Dupont	Albertine
2002	West	James

EPREUVE

<u>CodeEpr</u>	DateEpr	Lieu
ECOS101	15/01/2020	Say
ECOS102	16/01/2020	Say
GESS201	25/05/2020	D201
INFOS101	20/01/2020	D101

PASSER (table « pont »)

NumEtu#	CodeEpr#	Note
1110	INFOS101	15,5
2002	ECOS101	8,5
2002	ECOS101	13,0
1110	GESS201	14,0
2002	GESS201	14,5

Traduction d'une association M-N

PubID	Publisher	PubAddress	AuthorID	AuthorName	AuthorCity	Title	Price
03-447-1822	Random House	121 4th Street, New York	1345-26-2930	Male Debasise	14-Aug-92		
04-7129031	Wiley and Sons	45 Lincoln Blvd, Chicago	392-48-8985	Joel Blou	14-Mar-16		
03-4850223	O'Reilly Press	77 Boston Ave, Cambridge	454-24-4012	Billy Hammings	12-Sep-70		
02-3020884	Ch. Light Books	99 Market, San Francisco	160-59-1251	Harshad Arora	12-Mar-08		

Modèle relationnel

ETUDIANT

<u>NumEtu</u>	Nom	Prénom
1110	Dupont	Albertine
2002	West	James

EPREUVE

<u>CodeEpr</u>	DateEpr	Lieu
ECOS101	15/01/2020	Say
ECOS102	16/01/2020	Say
GESS201	25/05/2020	D201
INFOS101	20/01/2020	D101

PASSER (table « pont »)

NumEtu#	CodeEpr#	Note
1110	INFOS101	15,5
2002	ECOS101	8,5
2002	ECOS102	13,0
1110	GESS201	14,0
2002	GESS201	14,5

Problème de la redondance

PubID	Publisher	PubAddress	AuthorID	AuthorName	AuthorAddress	Title	Date
034473822	Random House	121 4th Street, New York	1910-26-2930	Harel Gellase	392-48-8985	Cold Fusion for Dummies	14-Aug-92
047733951	Wiley and Sons	45 Lincoln Street, Chicago	454-24-4012	Sally Hammings	192-59-1254	Murkine and Straw-Ting	14-Mar-93
034858223	O'Reilly Press	77 Boston Ave, Cambridge	03-48-9923	1952	Fluid Dynamics of Aqueducts	12-Sep-70	
02-30-0288	Ch. Light Books	99 Market, San Francisco	02-39-2088	1987	Reads, Baskets & Revolution	12-Mar-08	

Modèle relationnel

- ▶ Lorsque l'on effectue directement une modélisation logique ex. Soit la relation PASSER_EPREUVE.

<u>NumEtu</u>	<u>Note</u>	<u>CodeEpr</u>	<u>Lieu</u>
1110	15,5	INFOS101	Amphi Aubrac
1110	14,0	ECOS101	Amphi Aubrac
2002	13,0	ECOS102	Salle D201
3333	10,5	INFOS101	Amphi Aubrac

Cette relation présente différentes anomalies.

Anomalies liées à la redondance

PubID	Publisher	PubAddress
034473822	Randon House	121 4th Street, New York
044733901	Wiley and Sons	45 Lincoln Blvd, Chicago
034850223	O'Reilly Press	77 Boston Ave, Cambridge
023020886	Ch. Light Books	99 Market, San Francisco

AuthorID	AuthorName	EditorName	EditorDate
1340-20-2030	Hale Glesse	14-Aug-92	
262-48-8965	Jon Blow	14-May-16	
454-24-4012	Billy Hammings	12-Sep-70	
160-58-1251	Harrah Avenue	12-Mar-08	

ECOS	AuthorID	PubID	ECOS I	ECOS II	ECOS III
1-34433-482-1	145-28-2538	03-4472822	1893	Cold Fusion for Dummies	
1-38482-895-1	192-48-8985	04-7733901	1895	Microwave and Straw Tieg	
2-39921-489-4	454-22-4012	03-4892223	1852	Fluid Dynamics of Aqueducts	
1-38278-293-4	463-05-1254	02-3920886	1987	Reads, Baskets & Revolution	

Modèle relationne

- ▶ Anomalies de modification : Si l'on souhaite mettre à jour le lieu d'une épreuve, il faut le faire pour tous les n-uplets concernés.
- ▶ Anomalies d'insertion : Pour ajouter une nouvelle épreuve, il faut obligatoirement fournir des valeurs pour **NumEtu** et **Note**.
- ▶ Anomalies de suppression
ex. La suppression de l'étudiant n° **2002** fait perdre toutes les informations concernant l'épreuve **ECOS102**.

Éviter la redondance

► Pourquoi ?

- Suppression des problèmes de mise à jour
- Minimisation de l'espace de stockage

► Comment ?

- Dans le modèle conceptuel, ne spécifier que des attributs **non décomposables** (première forme normale).
ex. Une adresse doit être décomposée en rue, code postal, ville...
- C'est tout !

Partie 3

Modélisation physique

TP sur

