

# Bases de données

## Cours 1

### Introduction et Schémas relationnels

Marie Pelleau & Laurent Tichit  
[marie.pelleau@univ-cotedazur.fr](mailto:marie.pelleau@univ-cotedazur.fr),  
[laurent.tichit@univ-cotedazur.fr](mailto:laurent.tichit@univ-cotedazur.fr)

25 octobre 2022

- 1 Introduction
  - Organisation
  - Références
  - Motivation
  - Historique
  - SGBD
- 2 Modèle relationnel
- 3 Algèbre relationnelle

# Organisation

- Cours et TP
  - 6 séances de cours (le mardi de 15h30 à 17h30 en Amphi Géologie)
  - 6 matinées de TP (le mercredi matin en salles PV 315 et PV 316)
- Évaluation
  - un « mini-projet » (rendu mi-décembre)
  - un examen en janvier

## Quelques références pour ce cours :

- **Database Management Systems**, Ramakrishnan, Gehrke
- **Fundamentals of Database Systems**, Elmasri, Navathe
- **Bases de données : concepts, utilisation et développement**, Hainaut
- **Database Design and Relational Theory : Normal Forms and All That Jazz**, 2nd Edition, C.J. Date
- **Use the index, Luke, A Guide to Database Performance for Developers**, <https://use-the-index-luke.com/>
- **Joe Celko's SQL for Smarties, Advanced SQL Programming**, Celko, 2014
- Polycopié de Philippe Rigaux, **Cours de bases de données**, [http://www.info.univ-angers.fr/~gh/Pluripass/Db/coursBD\\_Rigaux.pdf](http://www.info.univ-angers.fr/~gh/Pluripass/Db/coursBD_Rigaux.pdf)
- ...

- Énorme quantité de données à notre disposition (et connue sur nous).
- But : exploiter des ensembles de données complexes.
- Besoin : outils efficaces pour gérer et extraire informations.
- **Base de données** : collection de données (exemples : activités d'une ou plusieurs sociétés, gestion d'une université).
- SGBD (Système de Gestion de Base de Données) : logiciel qui permet de gérer une BD partagée par plusieurs utilisateurs.

# Gérer des données

- Alternative aux SGBD : utiliser des fichiers et écrire du code spécifique pour les gérer.
- But du cours : donner une introduction aux SGBD en insistant sur les façons de
  - **concevoir** une BD
  - **utiliser** un SGBD efficacement.

Détails plus techniques :

- Conception de bases de données (schémas entités-associations).
- Langage SQL.
- Gestion des transactions.

# Exemples de bases de données

- De nombreuses situations concrètes peuvent être modalisées par une base de données :
  - Achats au supermarché
  - Achats à l'aide d'une carte bancaire
  - Visite à une bibliothèque de la ville
  - Livre de recettes
  - SMS/MMS sur un smartphone
  - <https://www.imdb.com/>
- Les informations ne seront pas rangées en vrac. Exemple :
  - **True Grit** est un film américain de Joel Coen et Ethan Coen (2010)
  - Acteurs principaux : Hailee Steinfeld (Mattie Ross), Jeff Bridges (Rooster Cogburn), Matt Damon (LaBoeuf)
- But : regrouper les informations de façon efficace

# Historique

- Charles Bachman (General Electric) début années 1960  $\leadsto$  modèle réseau.  
Prix Turing 1973.
- Fin des années 1960 : IBM crée le SGDB Information Management System (IMS)  $\leadsto$  modèle hiérarchique.  
Utilisé par Saturn V (programme lunaire), encore utilisé pour des distributeurs automatiques de billets.
- 1970 : Edgar Codd (IBM) propose nouveau système pour représenter les données  $\leadsto$  **modèle relationnel**.  
Développement rapide de SGBD utilisant modèle relationnel et recherches théoriques.  
Prix Turing 1981.
- Années 1980 : consolidation de la position dominante du modèle relationnel. Langage de requête SQL standard (ANSI et ISO).
- Années 2000 : remise en cause de la domination du modèle relationnel avec le développement des grandes entreprises Internet (NoSQL).

## 1 Introduction

## 2 Modèle relationnel

- Concepts
- Domaines et types
- Instance du schéma relationnel
- Contraintes
- Clés étrangères

## 3 Algèbre relationnelle

# Concepts du modèle relationnel

- Base de données : collection de **relations**.
- Informellement, **relation** = **table**. Exemple : Classification classique des espèces animales.

Nom	Eucaryote	Multicellulaire	Propriété
bactéries	faux	faux	
archées	faux	faux	
protistes	vrai	faux	
champignons	vrai	vrai	décompose
végétaux	vrai	vrai	photosynthétise
animaux	vrai	vrai	ingère

- Ligne de la table :  **$n$ -uplet** (tuple).
- $n$ -uplet : objet que l'on veut gérer dans la BD (espèce animale, étudiant, voiture, ...)
- Objets du même type sur les différentes lignes (on ne mélange pas étudiants et animaux !)
- colonne : propriété des objets, appelée **attribut**.

# Domaines

- **Domaine** : ensemble de valeurs **atomiques** (indivisibles) pour un attribut donné.
- Mauvais exemple de relation : pas d'atomicité.

---

espèce

---

bactéries : procaryotes unicellulaires

---

archées : procaryotes unicellulaires

---

protistes : eucaryotes unicellulaires

---

champignons : eucaryotes multicellulaires qui décomposent

---

végétaux : eucaryotes multicellulaires qui photosynthétisent

---

animaux : eucaryotes multicellulaires qui ingèrent

---

- Spécifier domaine : donner un type de données dans lequel prendre les valeurs.

## Exemples de domaines

- Numéros de téléphone : l'ensemble des numéros de téléphones à 10 chiffres valables en France.
- Numéros de sécurité sociale : l'ensemble des numéros de sécurité sociale à 15 chiffres valides.
- Notes : valeurs possibles des notes d'un cours ; chacune doit être un nombre réel (à virgule) entre 0 et 20.
- Âge des employés : âge possible des employés d'une entreprise ; chacun doit être un entier entre 16 et 70 ans.
- Département de l'université : l'ensemble des noms des départements de l'université, par exemple, mathématiques, informatique, physique...

# Type de données

- Définitions du transparent précédent : définitions logiques.
- Spécifier aussi types de données.
- Exemples :
  - Numéros de téléphone : chaîne de 10 caractères, chaque caractère est un chiffre (décimal).
  - Âge : nombre entier entre 16 et 70.
  - Eucaryote (pour espèce animale) : booléen (V/F).

# Types de données numériques

Type	Représente
SMALLINT	Nombre entier [-32768, 32767]
INT	Nombre entier [-2147483648, 2147483647]
BIGINT	Nombre entier [-9223372036854775808, 9223372036854775807]
SERIAL	Nombre entier incrémenté automatiquement (le SGBD gère la numérotation)
DOUBLE	Nombre réel [-1.7976931348623157E+308, 1.7976931348623157E+308]
FLOAT	Nombre réel [-3.402823466E+38, 3.402823466E+38]
NUMERIC(n,d)	Nombre à virgule de n chiffres dont d décimales

- Nombres entiers positifs seulement : UNSIGNED
  - Les valeurs possibles sont décalées (ex : de 0 à 65535 pour SMALLINT)

# Types de données autres

Type	Représente
VARCHAR(n)	Chaîne d'au plus n caractères
CHAR(n)	Chaîne d'exactly n caractères
TEXT	Chaîne de taille non limitée
DATE	Date au format 'aaaa-mm-jj'
TIME	Heure au format 'hh:mm:ss'
BIT(n)	Vecteurs de n bits

- Chaînes de caractères :
  - VARCHAR si le nombre de caractères peut varier
  - CHAR s'il est fixe (Exemple : immatriculation d'un véhicule)
  - TEXT si le texte peut être très long (plusieurs centaines de caractères)

# En pratique

- Schéma relationnel  $R(A_1, A_2, \dots, A_n)$ 
  - liste d'attributs  $A_1, A_2, \dots, A_n$
  - chaque attribut  $A_i$  a pour domaine  $D_i$ .
  - $n$  : degré ou arité de  $R$ .
- Exemple : relation de degré 4 qui stocke des informations sur les films.
  - `FILM(No, Titre, DateSortie, Pays)`
  - Ou avec les types de données :  
`FILM(No : SMALLINT, Titre : VARCHAR(50), DateSortie :  
DATE, Pays : CHAR(3))`

# Instance du schéma relationnel

- Une **instance**  $r$  du schéma relationnel  $R(A_1, A_2, \dots, A_n)$  est un ensemble de  $n$ -uplets (lignes).
- Chaque  $n$ -uplet  $t = (v_1, v_2, \dots, v_n)$  est une liste ordonnée de  $n$  valeurs où  $v_i$  est élément du domaine de  $A_i$  ou NULL.
- Cette valeur NULL signifie « inconnue » ou « non applicable ».
- Exemple : une instance de schéma relationnel :

## ETUDIANT

Nom	Prénom	Numéro	Tél_fixe	Adresse	Tél_portable
Leroy	Richard	210236	0123456789	2 Rue Papu Rennes	0610111213
Calzati	Giovanni	210282	0132456798	3 Avenue Valrose Nice	0610111111
Dumas	Céline	214781	NULL	4 Rue Barla Nice	0701010101
Kachour	Élise	236230	0217212121	NULL	NULL

## Instance de schéma relationnel (2)

- On peut utiliser le vocabulaire de la théorie des ensembles.
- Une instance de relation  $r$  du schéma  $R(A_1, \dots, A_n)$  est

$$r \subseteq D_1 \times D_2 \times \dots \times D_n,$$

où  $D_i$  est le domaine de l'attribut  $A_i$  pour tout  $i$ .

- Conséquence : si  $|D| = \text{cardinal de } D$ , alors le nombre total de  $n$ -uplets (lignes) possibles est  $|D_1| \cdot |D_2| \cdot \dots \cdot |D_n|$ .

**Cardinal** : nombre total de  $n$ -uplets (lignes) de la relation.

- $r$  ensemble de  $n$ -uplets : **pas de répétition, deux  $n$ -uplets (lignes) sont toujours différents.**

## Instance de schéma relationnel (3)

- Deux attributs peuvent avoir le **même domaine**, mais des interprétations différentes.  
Exemple : `TRAJET(Ville_Dép, Ville_Arr)`, alors les deux attributs ont le même domaine (celui des noms de Ville).
- Le choix de l'ordre des colonnes n'est pas important, mais on le conserve par commodité.
- L'ordre des  $n$ -uplets (lignes) n'a pas d'importance.

# Contraintes du domaine relationnel

- On ajoute des contraintes sur le schéma relationnel pour pouvoir le manipuler convenablement ou qu'il réponde à nos objectifs.
- **Contraintes de domaine** : les composantes de chaque  $n$ -uplet doivent appartenir au domaine spécifié au départ.
- **Contraintes de clés**. Rappel : les  $n$ -uplets (lignes) sont deux-à-deux distincts.
  - il existe un sous-ensemble d'attributs sur lesquels deux  $n$ -uplets de la relation diffèrent toujours. Cet ensemble est appelé **super-clé**.
  - Une **clé candidate** est une **super-clé** minimale.
  - Parmi les clés candidates, on en choisit une et on l'appelle **clé primaire**. On souligne le ou les attributs clé primaire.

# Contraintes de clés

- Exemple : ESPÈCE

Nom	Eucaryote	Multicellulaire	Propriété
bactéries	faux	faux	NULL
archées	faux	faux	NULL
protistes	vrai	faux	NULL
champignons	vrai	vrai	décompose
végétaux	vrai	vrai	photosynthétise
animaux	vrai	vrai	ingère

Schéma relationnel :

ESPÈCE(Nom, Eucaryote, Multicellulaire, Propriété)

## Contraintes de clés (suite)

- Exemple : DOCUMENT

<u>Id</u>	<u>Langue</u>	Description
rapport1	français	
rapport1	anglais	Le rapport 1 en anglais
rapport2	français	Le rapport 2 en français
rapport3	anglais	Le rapport 3 en anglais

- Schéma relationnel :  
DOCUMENT(Id, Langue, Description)
- La clé primaire est **composée** du couple (Id, Langue).

# Contraintes du modèle relationnel (suite)

- Contraintes de **valeur NULL**.
  - On spécifie à l'avance si une valeur NULL est admissible ou non.
  - Exemple : toute espèce doit avoir un nom valide ; l'attribut Nom est NOT NULL.
- Contraintes d'**intégrité**.
  - 1 La clé primaire ne peut pas être NULL.
  - 2 Contraintes d'intégrité référentielles :
    - Règles spécifiées entre deux relations pour compatibilité.
    - Informellement : un  $n$ -uplet d'une relation qui fait référence à une autre relation doit faire référence à un  $n$ -uplet existant.
    - Exemple : CLIENT(No\_client, Nom, Adresse)  
ACHAT(No\_produit, No\_client, Date, Qte)  
Un  $n$ -uplet de la relation ACHAT doit contenir un No\_client qui existe dans CLIENT.

# Clés étrangères

- Définition précise : soient  $R_1$ ,  $R_2$  deux schémas relationnels. Un ensemble d'attributs CE de  $R_1$  qui fait référence à la relation  $R_2$  est **clé étrangère** si les règles suivantes sont vérifiées :
  - Les attributs de CE ont le(s) même(s) domaine(s) que les attributs de clé primaire CP de  $R_2$
  - Une valeur de CE pour un  $n$ -uplet de  $R_1$  est soit une valeur existante de CP dans  $R_2$ , soit NULL.
- **Clé étrangère = clé primaire d'une autre relation.**
- On souligne en pointillés les clés étrangères.

CLIENT(No\_client, Nom, Adresse)

ACHAT(No\_produit, No\_client, Date, Qte)

## Exemple

No	Nom	AnnéeNaiss
1	Jeff Bridges	1949
2	Ethan Coen	1954
3	Joel Coen	1957
4	John Goodman	1952
5	Henry Hathaway	1898
6	Bong Joon-ho	1969
7	Hailee Steinfeld	1996
8	John Wayne	1907

PERSONNE

Rôle	ActeurNo	FilmNo
Walter Sobchak	4	1005
The Dude	1	1005
Rooster Cogburn	8	1001
Rooster Cogburn	1	1020
Mattie Ross	7	1020

RÔLE

RéalNo	FilmNo
3	1005
2	1020
3	1020
5	1001
6	1030

RÉALISATION

FilmNo	Titre	Année
1001	True Grit	1969
1005	The Big Lebowski	1998
1020	True Grit	2010
1030	Parasite	2019

FILM

# Schéma relationnel de l'exemple précédent

- Schéma relationnel obtenu :

PERSONNE(No, Nom, AnnéeNaiss)

RÔLE(Rôle, ActeurNo, FilmNo)

RÉALISATION(RéalNo, FilmNo)

FILM(FilmNo, Titre, Année)

- Remarques :

- Dans RÉALISATION, la clé primaire est le couple (RéalNo, FilmNo)
- Dans RÔLE, FilmNo est partie de clé primaire et clé étrangère, on souligne avec deux traits.

## 1 Introduction

## 2 Modèle relationnel

## 3 Algèbre relationnelle

- Présentation
- Opération - sélection
- Opération - projection
- Opération - renommage
- Opérations ensemblistes - union, intersection, différence
- Opérations ensemblistes - produit cartésien
- Jointure
- Jointure externe
- Combiner ces opérations

- Une relation peut être vue comme une table dont les colonnes sont les **attributs**. Les lignes de la table sont appelées  **$n$ -uplets** en français, **tuples** en anglais.
- Algèbre relationnelle : définir des opérations sur les relations.
- Résultats des opérations : des relations.
- Peut paraître assez abstrait (au premier abord), mais constitue la base des requêtes SQL (dans un cours prochain).

# Sélection – définition et exemples de bases

- **Sélection** : choisir un sous-ensemble des  $n$ -uplets (lignes) qui vérifient la condition de sélection. Notation :  $\sigma$
- Sélection peut être vue comme filtrage :
  - Séparation des  $n$ -uplets en 2 paquets : ceux qui vérifient la condition (conservés) et les autres (rejetés).
- Exemples :
  - Films dont le titre est *True Grit* :  $\sigma_{\text{Titre}='True Grit'}(\text{FILM})$

FilmNo	Titre	Année
1001	True Grit	1969
1020	True Grit	2010

$\sigma_{\text{Titre}='True Grit'}(\text{FILM})$

- Films sortis avant 2000 :  $\sigma_{\text{Année} < 2000}(\text{FILM})$

FilmNo	Titre	Année
1001	True Grit	1969
1005	The Big Lebowski	1998

$\sigma_{\text{Année} < 2000}(\text{FILM})$

# Sélection – généralités

- **Notation : sélection sur une relation R :**

$\sigma_{\langle \text{condition de sélection} \rangle} (R)$

- Condition formée d'un nombre quelconque de clauses de la forme  
 $\langle \text{nom attribut} \rangle \langle \text{op. comparaison} \rangle \langle \text{valeur constante} \rangle$   
 ou  $\langle \text{nom attribut} \rangle \langle \text{op. comparaison} \rangle \langle \text{nom attribut} \rangle$   
 où  $\langle \text{op. comparaison} \rangle$  est un des opérateurs  $=, <, \leq, >, \geq, \neq$ .
- Les clauses peuvent être reliées par les opérateurs booléens **et**, **ou**, **non** (négation  $\neg$ ).
- Exemple : sélectionner les personnes dont le numéro est 1 ou nées dans les années 1950 :

No	Nom	AnnéeNaiss
1	Jeff Bridges	1949
2	Ethan Coen	1954
3	Joel Coen	1957
4	John Goodman	1952

$\sigma_{\text{No} = 1 \text{ ou } (\text{AnnéeNaiss} \geq 1950 \text{ et } \text{AnnéeNaiss} \leq 1959)} (\text{PERSONNE})$

# Sélection – propriétés

## Propriétés

- Le résultat d'une sélection à partir d'une relation  $R$  est une relation de même degré (nombre d'attributs) que  $R$ .
- Le cardinal (nombre de  $n$ -uplets ou lignes) d'une sélection à partir d'une relation  $R$  est inférieur ou égal au cardinal de  $R$ .
- L'opération de sélection est commutative :

$$\sigma_{\langle \text{cond1} \rangle} (\sigma_{\langle \text{cond2} \rangle} (R)) = \sigma_{\langle \text{cond2} \rangle} (\sigma_{\langle \text{cond1} \rangle} (R)).$$

En effet,  $\sigma_{\langle \text{cond1} \rangle} (\sigma_{\langle \text{cond2} \rangle} (R)) = \sigma_{\langle \text{cond1} \rangle} \text{ et } \langle \text{cond2} \rangle (R)$ .

## Projection – définition et exemples de bases

- **Projection** : sélectionner certaines colonnes de la table (et jeter les autres). Notation :  $\pi$ .
- Projection utilisée si seuls certains attributs (colonnes) nous intéressent.
- Exemples : titres, années des films, puis titres des films

Titre	Année
True Grit	1969
The Big Lebowski	1998
True Grit	2010
Parasite	2019

$\pi_{\text{Titre,Année}}(\text{FILM})$

Titre
True Grit
The Big Lebowski
Parasite

$\pi_{\text{Titre}}(\text{FILM})$

- Les répétitions sont éliminées.

## Projection – cas général

- Notation : projection pour une relation  $R$  :  $\pi_{\langle \text{liste d'attributs} \rangle} (R)$
- Il est possible que des  $n$ -uplets (lignes) identiques apparaissent. Mais la **projection supprime les  $n$ -uplets identiques**.
- Le résultat de l'opération de projection est donc un ensemble d'uplets distincts deux à deux (et donc une relation valide).

### Propriétés

- Le cardinal (nombre de  $n$ -uplets ou lignes) d'une projection à partir d'une relation  $R$  est inférieur ou égal au cardinal de  $R$ .
- La projection est idempotente :

$$\pi_{\langle \text{cond} \rangle} (\pi_{\langle \text{cond} \rangle} (R)) = \pi_{\langle \text{cond} \rangle} (R).$$

## Renommage – définition

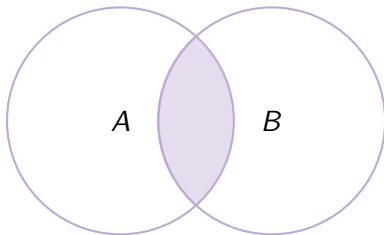
- Le renommage consiste à changer le nom d'une colonne (un attribut).
- Exemple : renommer l'attribut Année en Sortie

FilmNo	Titre	Sortie
1001	True Grit	1969
1005	The Big Lebowski	1998
1020	True Grit	2010
1030	Parasite	2019

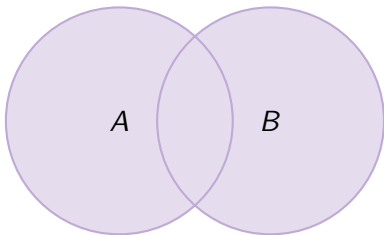
$\rho_{\text{Année/Sortie}}(\text{FILM})$

- On peut utiliser les opérations standard de la théorie des ensembles pour manipuler les relations.

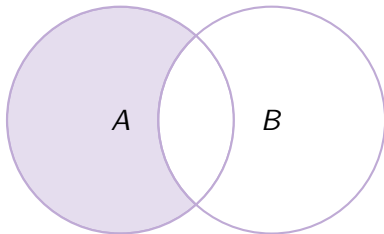
$$A \cap B$$



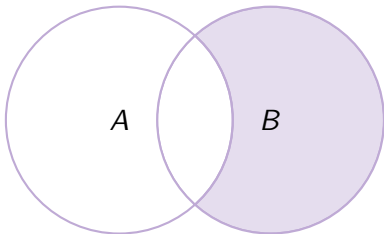
$$A \cup B$$



$$A - B$$



$$B - A$$



# Union, intersection, différence

- Les relations  $R(A_1, \dots, A_n)$  et  $S(B_1, \dots, B_m)$  sont dites **compatibles** si les deux conditions suivantes sont vérifiées :
  - elles ont même degré (même nombre d'attributs, ou encore de colonnes, ou encore  $n = m$ )
  - pour tout  $i$ ,  $1 \leq i \leq n$ ,  $A_i$  et  $B_i$  ont même domaine.
- Pour deux relations  $R$  et  $S$  compatibles,
  - l'**union** de  $R$  et  $S$ , notée  $R \cup S$ , est la relation qui contient les  $n$ -uplets qui sont dans  $R$ , dans  $S$ , et dans les deux à la fois. Les répétitions sont éliminées.
  - L'**intersection** de  $R$  et  $S$ , notée  $R \cap S$ , est la relation qui contient les  $n$ -uplets qui sont dans  $R$  et dans  $S$ .
  - La **différence** de  $R$  et  $S$ , notée  $R - S$  est la relation qui contient les  $n$ -uplets qui sont dans  $R$  et pas dans  $S$ .
- Par convention, les noms des attributs sont ceux de la 1<sup>ère</sup> relation. On peut toujours faire un renommage au besoin.

# Union, intersection, différence : exemples

FilmNo	Titre	Année
1001	True Grit	1969
1005	The Big Lebowski	1998
1020	True Grit	2010
1030	Parasite	2019

FILM1

FilmNo	Titre	Année
1001	True Grit	1969
1011	The African Queen	1951
1025	Birdman	2014

FILM2

Les relations FILM1 et FILM2 sont compatibles.

FilmNo	Titre	Année
1001	True Grit	1969

FILM1  $\cap$  FILM2

FilmNo	Titre	Année
1005	The Big Lebowski	1998
1020	True Grit	2010
1030	Parasite	2019

FILM1 - FILM2

FilmNo	Titre	Année
1001	True Grit	1969
1005	The Big Lebowski	1998
1020	True Grit	2010
1030	Parasite	2019
1011	The African Queen	1951
1025	Birdman	2014

FILM1  $\cup$  FILM2

FilmNo	Titre	Année
1011	The African Queen	1951
1025	Birdman	2014

FILM2 - FILM1

# Union, intersection, différence : exemples

FilmNo	Titre	Année
1001	True Grit	1969
1005	The Big Lebowski	1998
1020	True Grit	2010
1030	Parasite	2019

FILM

No	Nom	AnnéeNaiss
1	Jeff Bridges	1949
2	Ethan Coen	1954
3	Joel Coen	1957
4	John Goodman	1952

PERSONNE

Les relations FILM et PERSONNE sont compatibles.

FilmNo	Titre	Année
--------	-------	-------

FILM  $\cap$  PERSONNE

FilmNo	Titre	Année
1001	True Grit	1969
1005	The Big Lebowski	1998
1020	True Grit	2010
1030	Parasite	2019

FILM - PERSONNE

FilmNo	Titre	Année
1001	True Grit	1969
1005	The Big Lebowski	1998
1020	True Grit	2010
1030	Parasite	2019
1	Jeff Bridges	1949
2	Ethan Coen	1954
3	Joel Coen	1957
4	John Goodman	1952

FILM  $\cup$  PERSONNE

No	Nom	AnnéeNaiss
1	Jeff Bridges	1949
2	Ethan Coen	1954
3	Joel Coen	1957
4	John Goodman	1952

PERSONNE - FILM

# Produit cartésien

- R et S deux relations (pas de compatibilité requise).
- Le **produit cartésien** de R et S, noté  $R \times S$  est la relation dont les n-uplets (lignes) sont obtenus en combinant tous les n-uplets de R avec ceux de S.
- Exemple :  $FILM \times RÔLE$

FILM  $\times$  RÔLE

FILM.FilmNo	Titre	Année	Rôle	ActeurNo	RÔLE.FilmNo
1001	True Grit	1969	Walter Sobchak	4	1005
1001	True Grit	1969	The Dude	1	1005
1001	True Grit	1969	Rooster Cogburn	8	1001
1001	True Grit	1969	Rooster Cogburn	1	1020
1001	True Grit	1969	Mattie Ross	7	1020
1005	The Big Lebowski	1998	Walter Sobchak	4	1005
1005	The Big Lebowski	1998	The Dude	1	1005
1005	The Big Lebowski	1998	Rooster Cogburn	8	1001
1005	The Big Lebowski	1998	Rooster Cogburn	1	1020
1005	The Big Lebowski	1998	Mattie Ross	7	1020
1020	True Grit	2010	Walter Sobchak	4	1005
1020	True Grit	2010	The Dude	1	1005
1020	True Grit	2010	Rooster Cogburn	8	1001
1020	True Grit	2010	Rooster Cogburn	1	1020
1020	True Grit	2010	Mattie Ross	7	1020
1030	Parasite	2019	Walter Sobchak	4	1005
1030	Parasite	2019	The Dude	1	1005
1030	Parasite	2019	Rooster Cogburn	8	1001
1030	Parasite	2019	Rooster Cogburn	1	1020
1030	Parasite	2019	Mattie Ross	7	1020

## Produit cartésien – cas général

- $R(A_1, A_2, \dots, A_n)$  et  $S(B_1, B_2, \dots, B_m)$  deux relations.
- Alors  $R \times S$  relation de degré  $n + m$  et d'attributs

$$(A_1, A_2, \dots, A_n, B_1, B_2, \dots, B_m).$$

- $\text{Cardinal}(R \times S) = \text{Cardinal}(R) \times \text{Cardinal}(S)$
- Quel intérêt ? Seul, peu pertinent. Plus utile suivi d'une sélection pour faire correspondre les attributs.

## Sélection sur le produit cartésien

FILM.FilmNo	Titre	Année	Rôle	ActeurNo	RÔLE.FilmNo
1001	True Grit	1969	Rooster Cogburn	8	1001
1005	The Big Lebowski	1998	Walter Sobchak	4	1005
1005	The Big Lebowski	1998	The Dude	1	1005
1020	True Grit	2010	Rooster Cogburn	1	1020
1020	True Grit	2010	Mattie Ross	7	1020

$$\sigma_{\text{FILM.NoFilm} = \text{RÔLE.NoFilm}}(\text{FILM} \times \text{RÔLE})$$

- $\rightsquigarrow$  Films avec les rôles correspondants.
- Produit cartésien puis sélection : opération très courante, peut être vue comme une **jointure**.

# Jointure : définition

- Jointure, dénotée par  $\bowtie$ , utilisée pour combiner des  $n$ -uplets apparentés de deux relations en un  $n$ -uplet « plus long ».
- Pour obtenir l'ensemble des films avec les rôles.

FILM.FilmNo	Titre	Année	Rôle	ActeurNo	RÔLE.FilmNo
1001	True Grit	1969	Rooster Cogburn	8	1001
1005	The Big Lebowski	1998	Walter Sobchak	4	1005
1005	The Big Lebowski	1998	The Dude	1	1005
1020	True Grit	2010	Rooster Cogburn	1	1020
1020	True Grit	2010	Mattie Ross	7	1020

FILM  $\bowtie$  FILM.FilmNo=RÔLE.FilmNo RÔLE

## Jointure : cas général

- Forme générale d'une jointure sur deux relations  $R(A_1, A_2, \dots, A_n)$  et  $S(B_1, B_2, \dots, B_m)$  :

$$R \bowtie_{\langle \text{condition jointure} \rangle} S$$

- Résultat de cette opération : relation à  $n + m$  attributs

$$(A_1, A_2, \dots, A_n, B_1, B_2, \dots, B_m).$$

- Les n-uplets de la jointure sont formés d'uplets de  $R$  concaténés avec des n-uplets de  $S$  pourvu qu'ils vérifient la condition de jointure.
- La condition de jointure est spécifiée sur les attributs de  $R$  et de  $S$ .

## $\theta$ -jointure

Une condition de jointure générale est de la forme  
<condition> et <condition> et ... et <condition>  
où <condition> est de la forme

$$A_i \theta B_j,$$

où  $A_i$  (attribut de  $R$ ) et  $B_j$  (attribut de  $S$ ) ont le même domaine, et  $\theta$  est une des comparaisons  $=, <, \leq, >, \geq, \neq$ .

Une jointure avec une telle condition générale est appelée  $\theta$ -jointure.

# Équijointure et jointure naturelle

- Les jointures avec égalité (=) sont appelées **équijointures**.
- Le résultat d'une équijointure comporte deux attributs identiques, l'un provenant de la relation R1, l'autre provenant de R2.
- $\rightsquigarrow$  **jointure naturelle, notation \***
  - il faut que les deux attributs aient le même nom.
  - Exemple : combiner FILM et RÔLE pour obtenir aussi le nom du film correspondant à chaque rôle.

**La jointure naturelle supprime l'un des deux attributs dupliqués.**

FilmNo	Titre	Année	Rôle	ActeurNo
1001	True Grit	1969	Rooster Cogburn	8
1005	The Big Lebowski	1998	Walter Sobchak	4
1005	The Big Lebowski	1998	The Dude	1
1020	True Grit	2010	Rooster Cogburn	1
1020	True Grit	2010	Mattie Ross	7

FILM \* RÔLE

# Jointure externe gauche

- Dans la jointure naturelle précédente, le film *Parasite* n'apparaît pas, parce qu'aucun rôle de ce film n'est renseigné dans la relation RÔLE.
- **Jointure externe gauche** de R et S :
  - uplets de R et de S qui vérifient le critère de jointure sur leurs noms d'attributs communs (i.e. jointure naturelle),
  - ainsi que les uplets de R qui n'ont pas d'uplets correspondants dans S.
- Notation :  $R \bowtie S$ .
- Exemple :  $FILM \bowtie RÔLE$

FilmNo	Titre	Année	Rôle	ActeurNo
1001	True Grit	1969	Rooster Cogburn	8
1005	The Big Lebowski	1998	Walter Sobchak	4
1005	The Big Lebowski	1998	The Dude	1
1020	True Grit	2010	Rooster Cogburn	1
1020	True Grit	2010	Mattie Ross	7
1030	Parasite	2019	NULL	NULL

$FILM \bowtie RÔLE$

# Jointures externes droite et totale

- **Jointure externe droite** de R et S (notation  $R \bowtie S$ ) :
  - uplets de R et de S qui vérifient le critère de jointure sur leurs noms d'attributs communs (i.e. jointure naturelle),
  - ainsi que les uplets de S qui n'ont pas d'uplets correspondants dans R.
- **Jointure externe totale** de R et S (notation  $R \Join S$ ) :
  - uplets de R et de S qui vérifient le critère de jointure sur leurs noms d'attributs communs (i.e. jointure naturelle),
  - ainsi que les uplets de R qui n'ont pas d'uplets correspondants dans S.
  - ainsi que les uplets de S qui n'ont pas d'uplets correspondants dans R.

# Jointures externe totale

- Définissons une relation ANCIENS =  $\sigma_{\text{AnnéeNaiss} < 1955}$  (PERSONNE) et renommons l'attribut RéalNo en No pour préparer la jointure.

No	Nom	AnnéeNaiss
1	Jeff Bridges	1949
2	Ethan Coen	1954
4	John Goodman	1952
5	Henry Hathaway	1898
8	John Wayne	1907

ANCIENS =  $\sigma_{\text{AnnéeNaiss} < 1955}$  (PERSONNE)

No	FilmNo
3	1005
2	1020
3	1020
5	1001
6	1030

RÉAL =  $\rho_{\text{RéalNo/No}}$  (RÉALISATION)

- Jointure externe totale :

No	Nom	AnnéeNaiss	FilmNo
1	Jeff Bridges	1949	NULL
2	Ethan Coen	1954	1020
3	NULL	NULL	1005
3	NULL	NULL	1020
4	John Goodman	1952	NULL
5	Henry Hathaway	1898	1001
6	NULL	NULL	1030
8	John Wayne	1907	NULL

# Combiner toutes ces opérations

- On peut combiner toutes les opérations pour obtenir les informations qui nous intéressent.
- Tous les rôles de Jeff Bridges avec les titres correspondants :

$\pi_{\text{Titre}, \text{Rôle}} \left( (\text{RÔLE} * \text{FILM}) * \rho_{\text{No}/\text{ActeurNo}} \left( \sigma_{\text{Nom}=' \text{Jeff Bridges}' } (\text{PERSONNE}) \right) \right)$

Titre	Rôle
The Big Lebowski	The Dude
True Grit	Rooster Cogburn

# À suivre

