

# Intégration de données

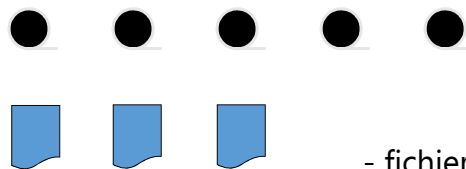
M.-S. HACID  
UCBL, LIRIS CNRS UMR 5205

07/09/2023

1

*TTI est à la fin d'une longue et complexe évolution mais  
continue d'évoluer*

1960

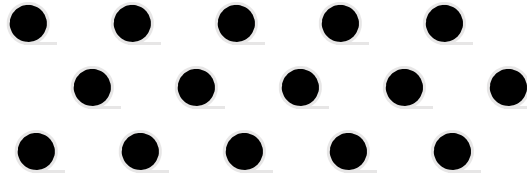


- fichiers
- programmes (Cobol)
- rapports

07/09/2023

2

1965



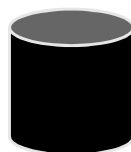
Beaucoup de fichiers!!!

- complexité
- maintenance
- développement
- synchronisation
- matériel

07/09/2023

3

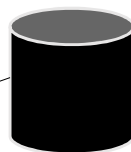
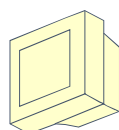
1970



SSAD  
SGBD

Base de données – “une source de  
données unique  
pour tous les traitements”

1975

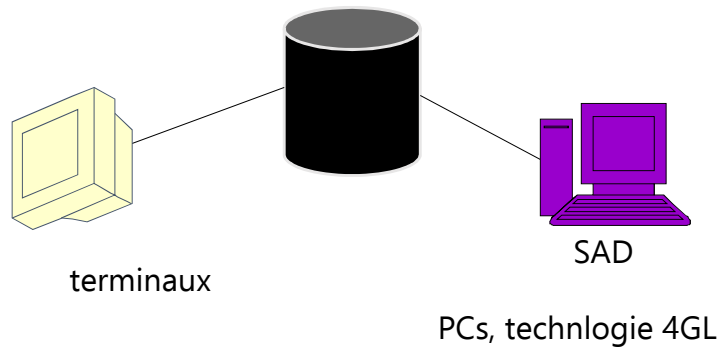


OLTP

07/09/2023

4

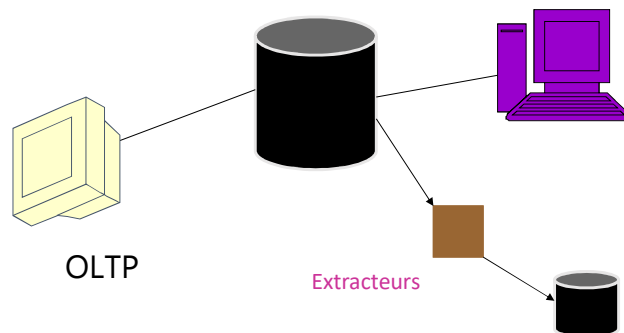
1980



07/09/2023

5

1985



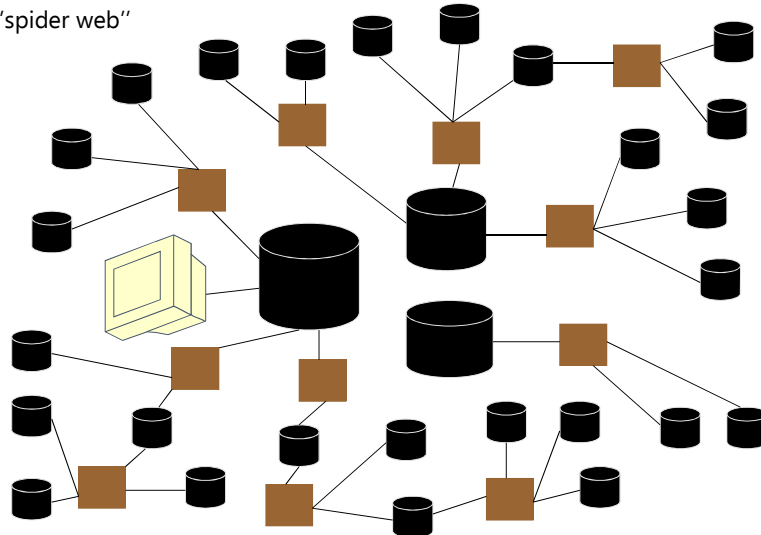
Pourquoi des extracteurs?

- Performance
- Contrôle

07/09/2023

6

1990 "spider web"



07/09/2023

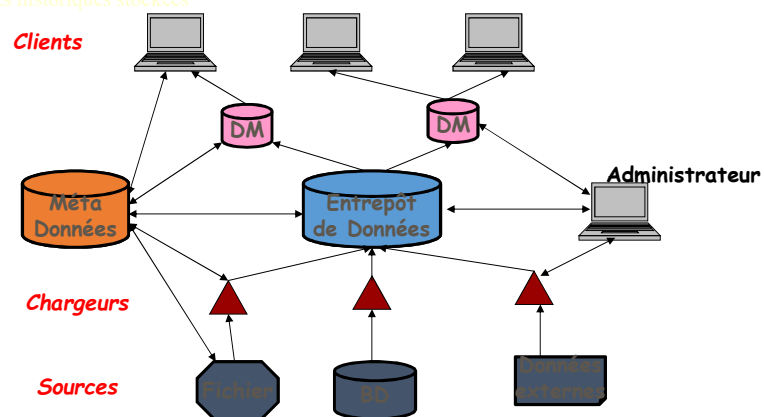
7

## Passer des données à l'information

Les applications n'étaient pas construites dans un but d'intégration.

Pas assez de données historiques stockées

"Comparer les activités d'un compte d'une année sur l'autre pendant les cinq dernières années"



07/09/2023

8

## Langage de règles (Datalog)

07/09/2023

9

### Faits

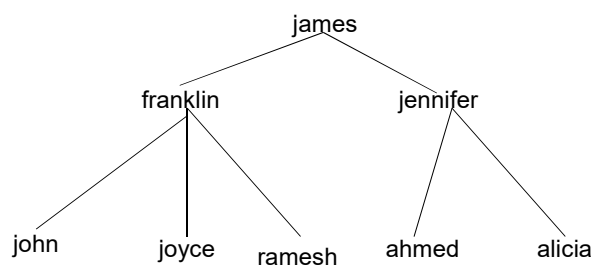
supervise(franklin, john).  
 supervise(franklin, ramesh).  
 supervise(franklin, joyce).  
 supervise(jennifer, alicia).  
 supervise(jennifer, ahmed).  
 supervise(james, franklin).  
 supervise(james, jennifer).  
 .....

### Règles

supérieur(X, Y):-supervise(X, Y).  
 supérieur(X, Y):-supervise(X, Z), supérieur(Z, Y).  
 subordonné(X, Y):-supérieur(Y, X).

### Requêtes

supérieur(james, Y)?  
 supérieur(james, joyce)?



07/09/2023

10

## Interprétation des règles

**Théorie de la preuve:** dans cette interprétation, les faits et les règles sont considérés comme des assertions vraies, ou axiomes. Les faits sont des axiomes clos et les règles sont appelées des axiomes déductifs. Les axiomes déductifs peuvent être utilisés pour construire des preuves qui dérivent de nouveaux faits à partir de faits existants.

Exemple:

- |   |           |
|---|-----------|
| 1. supérieur(X, Y):-supervise(X, Y).                  | (règle 1) |
| 2. supérieur(X, Y):-supervise(X, Z), supérieur(Z, Y). | (règle 2) |

- |  |                                 |
|--|---------------------------------|
| 3. supervise(jennifer, ahmed).   | (donné)                         |
| 4. supervise(james, jennifer).   | (donné)                         |
| 5. supérieur(jennifer, ahmed).   | (appliquer la règle 1 à 3)      |
| 6. supérieur(james, ahmed):-supervise(james, jennifer), supérieur(jennifer, ahmed) | (appliquer la règle 2 à 4 et 5) |

—————→ supérieur(james, ahmed)

L'interprétation par la théorie de la preuve donne une procédure ou approche **computationnelle** pour calculer une réponse à une requête Datalog.

07/09/2023

11

**Théorie des modèles:** Etant donné un domaine fini ou infini de constantes, on assigne à un prédicat toute combinaison possible de valeurs comme arguments. Ensuite on détermine si le prédicat est vrai ou faux. En général, il suffit de spécifier les combinaisons d'arguments qui rendent le prédicats vrai, et conclure que toutes les autres combinaisons rendent le prédicat faux.

07/09/2023

12

**Exemple:***Règles*

supérieur(X, Y) :- supervise(X, Y).  
 supérieur(X, Y) :- supervise(X, Z), supérieur(Z, Y).

*Interprétation***Faits dérivés****Faits connus**

supervise(franklin, john) est **vrai**.  
 supervise(franklin, ramesh) est **vrai**.  
 supervise(franklin, joyce) est **vrai**.  
 supervise(jennifer, alicia) est **vrai**.  
 supervise(jennifer, ahmed) est **vrai**.  
 supervise(james, franklin) est **vrai**.  
 supervise(james, jennifer) est **vrai**.  
 supervise(X, Y) est **faux** pour toutes les  
 autres combinaisons (X, Y).

supérieur(franklin, john) est **vrai**.  
 supérieur(franklin, ramesh) est **vrai**.  
 supérieur(franklin, joyce) est **vrai**.  
 supérieur(jennifer, alicia) est **vrai**.  
 supérieur(jennifer ahmed) est **vrai**.  
 supérieur(james, franklin) est **vrai**.  
 supérieur(james, jennifer) est **vrai**.  
 supérieur(james, john) est **vrai**.  
 supérieur(james, ramesh) est **vrai**.  
 supérieur(james, joyce) est **vrai**.  
 supérieur(james, alicia) est **vrai**.  
 supérieur(james, ahmed) est **vrai**.  
 supérieur(X, Y) est **faux** pour toutes les  
 autres combinaisons (X, Y).

Une interprétation est un **modèle** pour un ensemble spécifique de règles, si de telles règles sont toujours vraies sous une telle interprétation; c'est-à-dire, pour toutes valeurs assignées aux variables dans les règles, la tête des règles est vraie quand on substitue les valeurs de vérité assignées aux prédicats dans le corps des règles par l'interprétation.

Si supervise(a, b) et supérieur(b, c) sont tous les deux vrais sous une interprétation, mais que supervise(a, c) n'est pas vrai, l'interprétation ne peut pas être un modèle pour la règle réursive:

supérieur(X, Y) :- supervise(X, Z), supérieur(Z, Y).

## Mécanismes d'inférence de base pour les programmes logiques

### *Bottom-up ou chaînage avant ou résolution bottom-up*

Le moteur d'inférence part des faits et applique les règles pour générer de nouveaux faits. Lorsque des fait sont générés, ils sont confrontés au prédicat but de la requête pour tester le matching. Le terme chaînage avant indique que l'inférence va des faits vers le but.

Exemple

```
supervise(franklin, john).
supervise(franklin, ramesh).
supervise(franklin, joyce).
supervise(jennifer, alicia).
supervise(jennifer, ahmed).
supervise(james, franklin).
supervise(james, jennifer).
.....
```

```
supérieur(X, Y):-supervise(X, Y).
supérieur(X, Y):-supervise(X, Z), supérieur(Z, Y).
subordonné(X, Y):-supérieur(Y, X).
```

```
supérieur(james, Y)?
supérieur(franklin, john).
supérieur(franklin, ramesh).
supérieur(franklin, joyce).
supérieur(jennifer, alicia).
supérieur(jennifer, ahmed).
supérieur(james, franklin).
supérieur(james, jennifer).
supérieur(james, john).
supérieur(james, ramesh).
supérieur(james, joyce).
supérieur(james, alicia).
supérieur(james, ahmed).
```

Y=franklin

### *Top-down ou chaînage arrière ou résolution top-down*

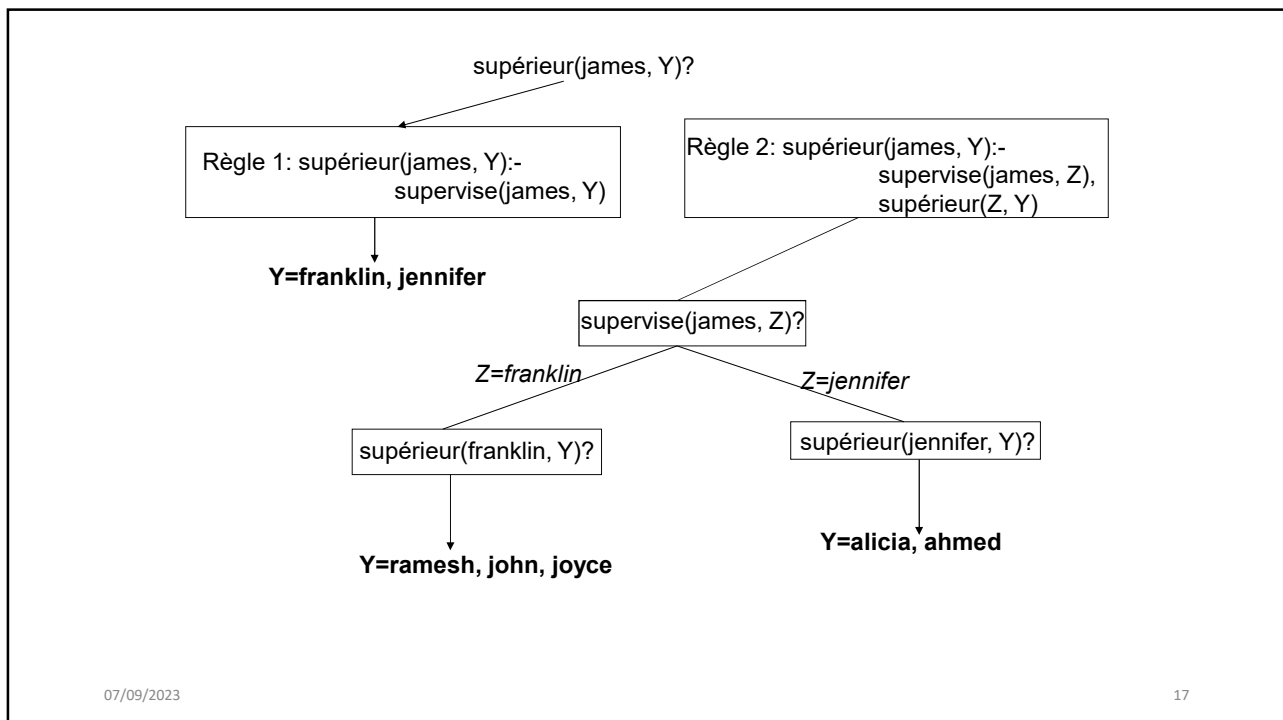
Cette technique part du but et tente de trouver une substitution de variables qui conduit à un fait valide dans la base. Le terme chaînage arrière indique que l'inférence se fait en partant du but en essayant de déterminer les faits qui pourront satisfaire le but.

Exemple

supérieur(james, Y)?

supérieur(james, Y):-supervise(james, Y).





## Qu'est-ce que l'intégration de données ?

Contrairement aux SGBDR traditionnels, les nouveaux services doivent pouvoir

- ☐ partager des données entre plusieurs organisations
- ☐ intégrer les données de manière souple et efficace

### Intégration des données :

- ☐ Ensemble de techniques permettant de construire des systèmes conçus pour le partage et l'intégration flexibles de données entre plusieurs fournisseurs de données autonomes.

## Pourquoi l'intégration des données ? (1/2)

- Avec des aspects tels que les **normalisations** et les compromis dans les **choix de conception**, différentes personnes conçoivent différents schémas pour les mêmes données
- Parfois, les besoins sont également différents
  - toutes les personnes n'ont pas besoin de tous les attributs
- Parfois, les gens veulent partager leurs données
  - les collaborateurs
  - chercheurs qui souhaitent publier des données pour d'autres usages

## Pourquoi l'intégration des données ? (2/2)

### ❑ Sur le web

- De nombreux sites Web affichant des demandes d'emploi, des offres d'hôtels ou de vols, des informations sur les films, etc.
- Pour rester au courant des nouvelles informations et des nouveaux besoins, vous devrez peut-être les consulter toutes,
- Mais aujourd'hui, il existe des sites web qui vous permettent d'accéder à toutes les informations.
- Par exemple, TripAdvisor vous aide à voir le prix du même hôtel sur le site de l'hôtel, hotels.com, booking.com, expedia, ....

### ❑ Mais ce type d'intégration de données présente des difficultés aussi

## Les défis de l'intégration de données

### L'interrogation

- ☐ Offrir un accès uniforme à un ensemble de sources de données autonomes et hétérogènes
- ☐ Interroger des sources de données disparates, parfois les mettre à jour

## Les défis de l'intégration de données

### Nombre de sources

- ☐ Défi même pour 10 ou 20 sources de données
- ☐ Défi amplifié pour des centaines de sources, par exemple à l'échelle du Web

## Les défis de l'intégration de données

### Hétérogénéité

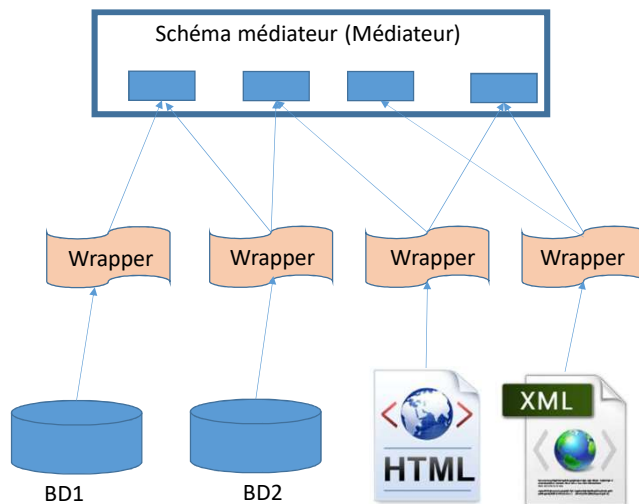
- ☐ Les sources de données ont été créées indépendamment les unes des autres
- ☐ Bases de données, fichiers, html
- ☐ Schémas et références différents
- ☐ Certaines sont structurées d'autres non

## Les défis de l'intégration de données

### L'autonomie

- ☐ Les sources peuvent ne pas appartenir à la même entité administrative
- ☐ Peuvent ne pas avoir un accès complet aux données
- ☐ Il peut y avoir des problèmes de protection de la vie privée
- ☐ Les sources peuvent modifier leurs formats et leurs modes d'accès à tout moment sans en avertir

## Architecture virtuelle d'intégration de données



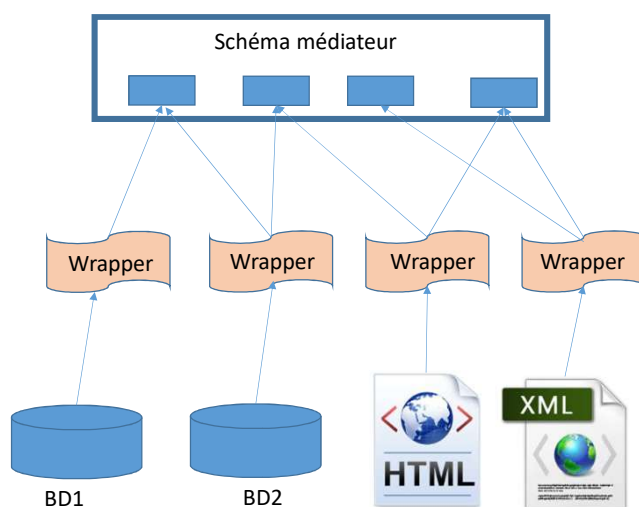
### Trois composantes

- ☐ Sources de données
- ☐ Wrappers
- ☐ Schéma médiateur

07/09/2023

25

## Architecture virtuelle d'intégration de données



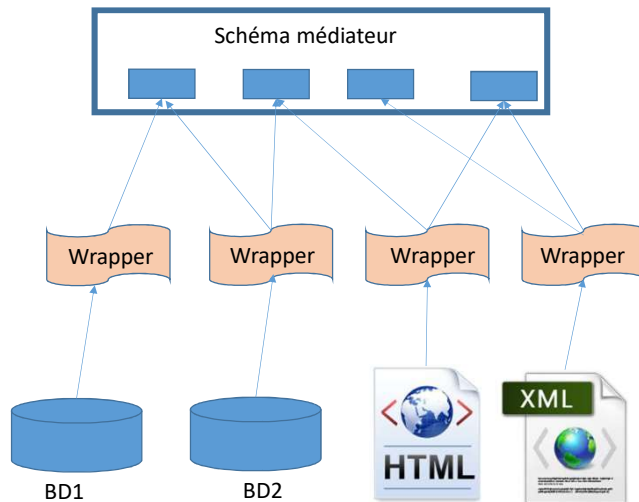
### Les sources de données

- ☐ Peuvent être n'importe quel modèle de données comme les bases de données relationnelles avec interface SQL
- ☐ XML avec l'interface Xquery
- ☐ HTML

07/09/2023

26

## Architecture virtuelle d'intégration de données



### Wrappers

- ☐ programmes qui envoient des requêtes à une source de données
- ☐ reçoit des réponses
- ☐ applique quelques transformations de base

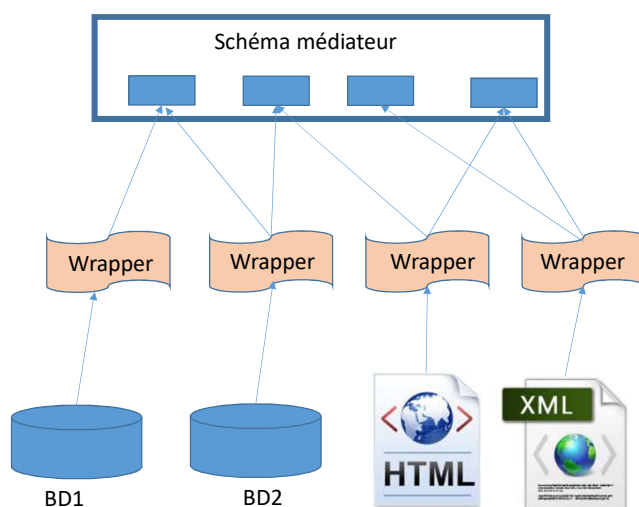
### Par exemple à un formulaire sur le Web

- ☐ traduit la requête en une requête **http** avec une **url**
- ☐ lorsque la réponse est retournée sous la forme d'un fichier html, extraire les tuples

07/09/2023

27

## Architecture virtuelle d'intégration de données



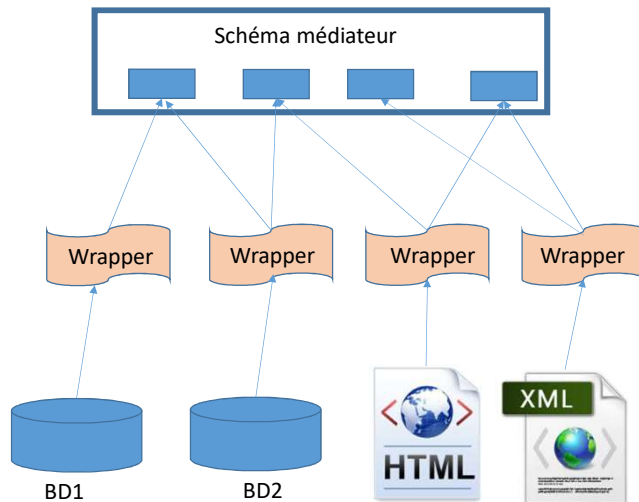
### Schéma médiateur (Médiateur)

- ☐ construit pour l'application d'intégration de données
- ☐ ne contient que les aspects/choses qui sont pertinents
- ☐ peut ne pas contenir tous les attributs
- ☐ ne stocke aucune données
- ☐ schéma logique pour spécifier des requêtes par les utilisateurs

07/09/2023

28

## Descriptions des sources



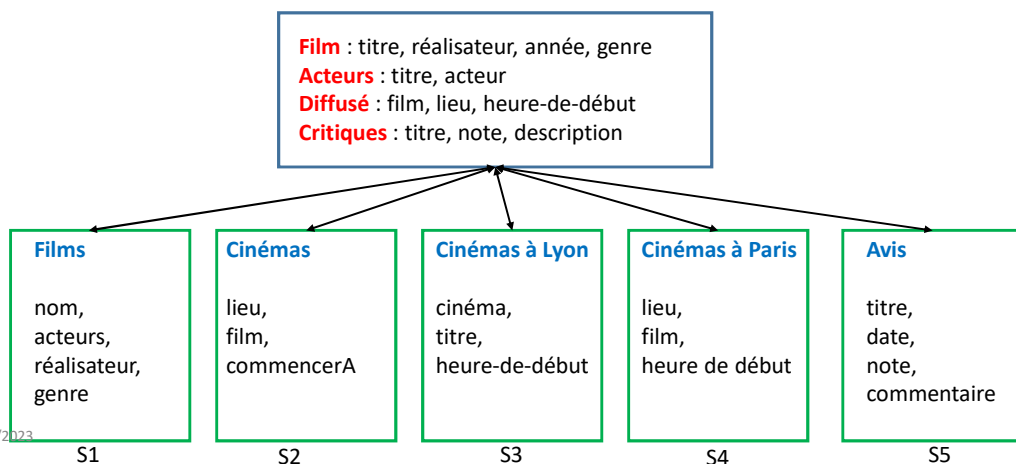
- ❑ Spécifier la propriété des sources que le système doit connaître
- ❑ Les principaux composants sont les correspondances sémantiques
  - relie le schéma des sources aux attributs du schéma médiateur
- ❑ spécifier de manière déclarative
- ❑ entre les sources de données et le schéma médiateur
  - pas entre deux sources
- ❑ spécifier également
  - si les sources sont complètes ou non
  - accès restreints

07/09/2023

29

## Exemple

- ❑ la source **S2** peut ne pas contenir tous les horaires de cinéma de l'ensemble du pays
- ❑ la source **S3** peut être connue pour contenir tous les horaires des cinémas à Lyon
- ❑ pour obtenir une réponse de la source **S1**, il faut une entrée pour au moins un de ses attributs



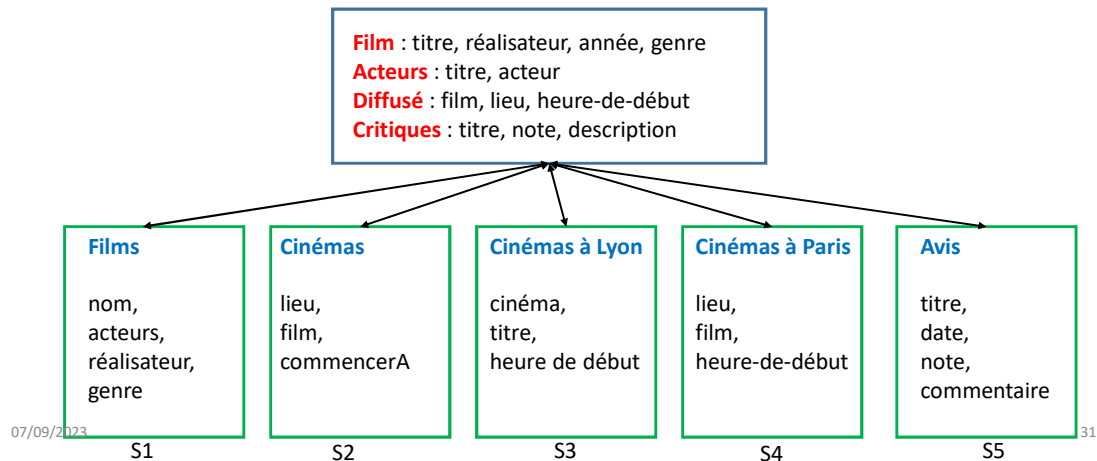
07/09/2023

30

### Exemple : Requêter le médiateur

```
SELECT titre, heure-de-début
FROM Film, Diffusé
WHERE Film.titre = Diffusé.film AND lieu= "Lyon" AND réalisateur="Woody Allen"
```

horaires de projection des films, réalisés par Woody Allen, à Lyon



### Exemple : Reformulation de requête

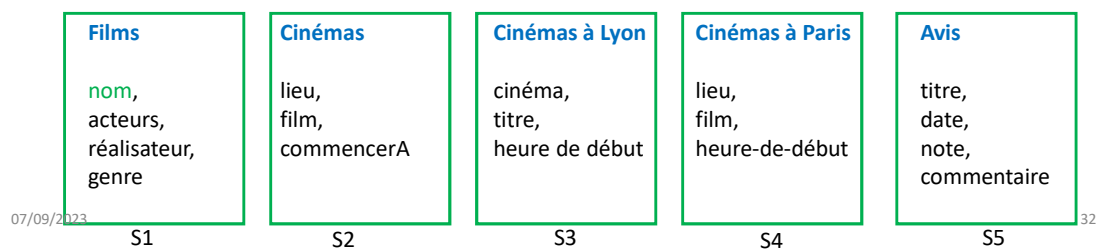
```
SELECT titre, heure-de-début
FROM Film, Diffusé
WHERE Film.titre = Diffusé.film AND lieu= "Lyon" AND réalisateur="Woody Allen"
```

horaires de projection des films, réalisés par Woody Allen, à Lyon

**Film** : titre, réalisateur, année, genre  
**Acteurs** : titre, acteur  
**Diffusé** : film, lieu, heure-de-début  
**Critiques** : titre, note, description

Les films peuvent être obtenus à partir de la source **S1**

➤ mais l'attribut **titre** doit être renommé en attribut **nom**





### Exemple : Reformulation de requête

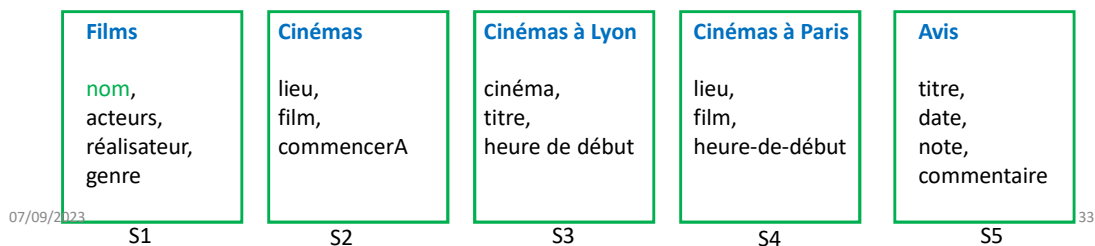
**SELECT** titre, heure-de-début  
**FROM** Film, Diffusé  
**WHERE** Film.titre = Diffusé.film **AND** lieu= "Lyon" **AND** réalisateur="Woody Allen"

horaires de projection des films, réalisés par Woody Allen, à Lyon

**Film** : titre, réalisateur, année, genre  
**Acteurs** : titre, acteur  
**Diffusé** : film, lieu, heure-de-début  
**Critiques** : titre, note, description

Les tuples pour Diffusé peuvent être obtenus soit à partir de S2 soit à partir de S3.

➤ Puisque cette dernière est complète pour les diffusions à Lyon, nous la préférons à S2



### Exemple : Reformulation de requête

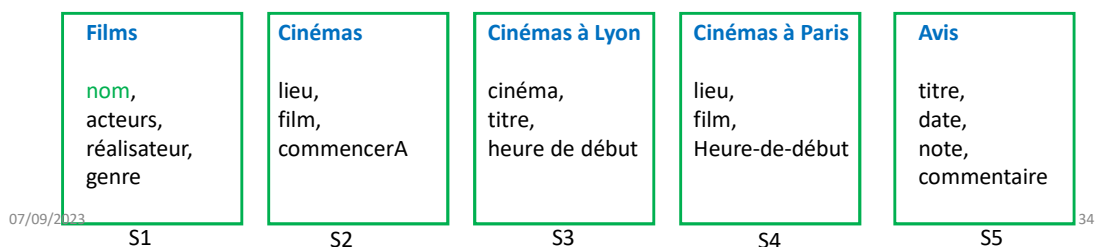
**SELECT** titre, heure-de-début  
**FROM** Film, Diffusé  
**WHERE** Film.titre = Diffusé.film **AND** lieu= "Lyon" **AND** réalisateur="Woody Allen"

horaires de projection des films, réalisés par Woody Allen, à Lyon

**Film** : titre, réalisateur, année, genre  
**Acteurs** : titre, acteur  
**Diffusé** : film, lieu, heure-de-début  
**Critiques** : titre, note, description

La source S3 nécessite le titre d'un film comme entrée

- mais ce titre n'est pas spécifié dans la requête
- le plan de la requête doit d'abord accéder à la source S1
- puis transmettre les titres de films renvoyés par S1 à la source S3



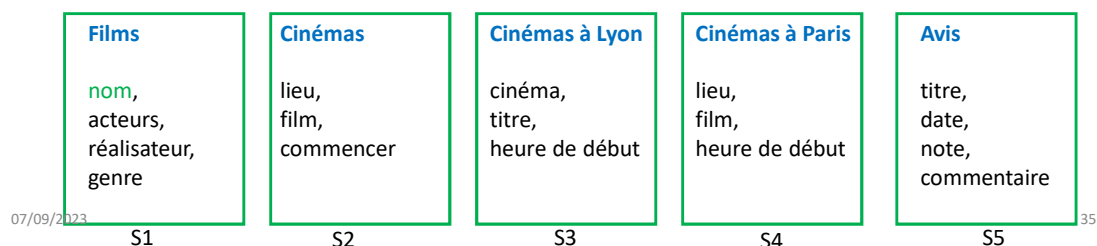
## Exemple : Reformulation de requête

**SELECT** titre, heure-de-début  
**FROM** Film, Diffusé  
**WHERE** Film.titre = Diffusé.film **AND** lieu= "Lyon" **AND** réalisateur="Woody Allen"

horaires de projection des films, réalisés par Woody Allen, à Lyon

**Film** : titre, réalisateur, année, genre  
**Acteurs** : titre, acteur  
**Diffusé** : film, lieu, heure-de-début  
**Critiques** : titre, note, description

- ❑ Options concernant le plan de requête logique :
  - accès à **S1**, **S3**
  - possibilité d'accéder à **S1** puis à **S2** également (éventuellement incomplète)
- ❑ Optimisation de la requête
  - comme dans un système de base de données traditionnel
  - prendre un plan logique produire un plan physique



## La mise en correspondance des schémas doit gérer les divergences entre le schéma source et le schéma du médiateur

Noms des relations et des attributs

- "description" dans le schéma médiateur (**SM**) identique à « commentaire » dans **S4**
- "nom" des acteurs dans **SM** et dans **S3** (CinémasALyon) n'est pas le même

**Film** : titre, réalisateur, année, genre  
**Acteurs** : titre, nom  
**Diffusé** : film, lieu, heure-de-début  
**Critiques** : titre, note, description

**S1:**  
**Acteur** (IDA, prénom, nom, nationalité, annéeDenaissance)  
**Film** (IDF, titre), **JoueDans**(IDA, IDF)  
**DétailsFilm** (IDF, réalisateurF, genre, année)

**S4:**  
**Avis**(titre, date, note, commentaire)

**S2:**  
**Cinémas**(lieu, film, heureprojection)

**S3:**  
**CinémasALyon** (nom, titre, débutprojection)

**S5:**  
**GenresFilms**(titre, genre)

**S7:**  
**AnnéesFilms**(titre, année)

**S6:**  
**RéalisateurFilm**(titre, réal)

## La mise en correspondance des schémas doit gérer les divergences entre le schéma source et le schéma du médiateur

### Organisation tabulaire

- Dans **SM**, **Acteurs** stocke le titre du film et le nom de l'acteur.
- Dans **S1**, une jointure est nécessaire et doit être spécifiée par la fonction d'association (mapping)

**Film** : titre, réalisateur, année, genre  
**Acteurs** : titre, nom  
**Diffusé** : film, lieu, heure-de-début  
**Critiques** : titre, note, description

**S1:**  
**Acteur** (IDA, prénom, nom, nationalité, annéeDenaissance)  
**Film** (IDF, titre), **JoueDans**(IDA, IDF)  
**DétailsFilm** (IDF, réalisateurF, genre, année)

**S4:**  
**Avis**(titre, date, note, commentaire)

**S2:**  
**Cinémas**(lieu, film, heureprojection)

**S5:**  
**GenresFilms**(titre, genre)

**S3:**  
**CinémasALyon** (nom, titre, débutprojection)

**S7:**  
**AnnéesFilms**(titre, année)

**S6:**  
**RéalisateurFilm**(titre, réal)

07/09/2023

37

## La mise en correspondance des schémas doit gérer les divergences entre le schéma source et le schéma du médiateur

### Couverture du domaine

- la couverture et le niveau de détail peuvent varier
- **S1** stocke plus d'informations sur les acteurs que **SM**

**Film** : titre, réalisateur, année, genre  
**Acteurs** : titre, nom  
**Diffusé** : film, lieu, heure-de-début  
**Critiques** : titre, note, description

**S1:**  
**Acteur** (IDA, prénom, nom, nationalité, annéeDenaissance)  
**Film** (IDF, titre), **JoueDans**(IDA, IDF)  
**DétailsFilm** (IDF, réalisateurF, genre, année)

**S4:**  
**Avis**(titre, date, note, commentaire)

**S2:**  
**Cinémas**(lieu, film, heureprojection)

**S5:**  
**GenresFilms**(titre, genre)

**S3:**  
**CinémasALyon** (nom, titre, débutprojection)

**S7:**  
**AnnéesFilms**(titre, année)

**S6:**  
**RéalisateurFilm**(titre, réal)

07/09/2023

38

## La mise en correspondance des schémas doit gérer les divergences entre le schéma source et le schéma du médiateur

### Variations au niveau des données

- Une lettre comme **note** ou un chiffre entre 0 et 4
- **S1** stocke les noms des acteurs dans deux colonnes, **SM** les stocke dans une seule colonne

**Film** : titre, réalisateur, année, genre  
**Acteurs** : titre, nom  
**Diffusé** : film, lieu, heure-de-début  
**Critiques** : titre, note, description

S1:  
**Acteur** (IDA, prénom, nom, nationalité, annéeDenaissance)  
**Film** (IDF, titre), **JoueDans**(IDA, IDF)  
**DétailsFilm** (IDF, réalisateurF, genre, année)

S4:  
**Avis**(titre, date, **note**, commentaire)

S2:  
**Cinémas**(lieu, film, heureprojection)

S5:  
**GenresFilms**(titre, genre)

S3:  
**CinémasALyon** (nom, titre, débutprojection)

S7:  
**AnnéesFilms**(titre, année)

S6:  
**RéalisateurFilm**(titre, réal)

07/09/2023

39

## Trois propriétés souhaitées pour les langages de mise en correspondance des schémas (langages de mapping)

### Flexibilité

- différences importantes entre des schémas disparates
- les langages doivent être très souples
- doivent pouvoir exprimer une grande variété de relations entre les schémas

07/09/2023

40

## Trois propriétés souhaitées pour les langages de mise en correspondance des schémas (langages de mapping)

### Reformulation efficace

- L'objectif est d'utiliser les correspondances (mappings) pour reformuler des requêtes
- On devrait pouvoir développer des algorithmes de reformulation dont les propriétés sont bien comprises et qui sont efficaces dans la pratique
- Compromis flexibilité-expressivité (car avec les langages plus expressifs il n'est pas toujours facile de raisonner)

## Trois propriétés souhaitées pour les langages de mise en correspondance des schémas (langages de mapping)

### Mise à jour facile

- pour qu'un formalisme soit utile en pratique, il doit être facile d'ajouter et de supprimer des sources
- Si l'ajout d'une nouvelle source de données nécessite potentiellement d'inspecter toutes les autres sources, le système résultant sera difficile à gérer pour un grand nombre de sources

## Trois langages de mise en correspondance des schémas

1. Global-as-View (GAV)
2. Local-as-View (LAV)
3. Global-Local-as-View (GLAV)

## Global-as-View (GAV)

- ❑ L'approche **GAV** définit le médiateur comme un ensemble de vues sur les sources de données.
  - Médiateur = Schéma global
  - Une approche intuitive
- ❑ Schéma du médiateur G
  - $G_i$  = une relation dans G
  - $X_i$  désigne les attributs dans  $G_i$
- ❑ Schéma source  $S_1, S_2, \dots, S_n$

## Global-as-View (GAV)

□ Un schéma de correspondance GAV M est un ensemble d'expressions de la forme :

- $G_i(X_i) \supseteq Q(S_1, S_2, \dots, S_n)$ 
  - ✓ hypothèse du monde ouvert
  - ✓ les instances calculées pour **SM** sont supposées être incomplètes
- ou,  $G_i(X_i) = Q(S_1, S_2, \dots, S_n)$ 
  - ✓ hypothèse du monde fermé
  - ✓ les instances calculées pour **SM** sont supposées être complètes

07/09/2023

45

## Exemple de GAV

- $\text{Film}(\text{titre}, \text{réalisateur}, \text{année}, \text{genre}) \supseteq S1.\text{Film}(\text{IDF}, \text{titre}), S1.\text{DétailsFilm}(\text{IDF}, \text{réalisateur}, \text{genre}, \text{année})$
- $\text{Film}(\text{titre}, \text{réalisateur}, \text{année}, \text{genre}) \supseteq S5.\text{GenresFilms}(\text{titre}, \text{genre}), S6.\text{RéaliseursFilm}(\text{titre}, \text{réalisateur}), S7.\text{AnnéesFilms}(\text{titre}, \text{année})$
- $\text{Diffusé}(\text{film}, \text{lieu}, \text{heure-de-début}) \supseteq S2.\text{Cinemas}(\text{lieu}, \text{film}, \text{heure-de-début})$
- $\text{Diffusé}(\text{film}, \text{lieu}, \text{heure-de-début}) \supseteq S3.\text{CinemasALyon}(\text{lieu}, \text{film}, \text{heure-de-début})$

**Film** : titre, réalisateur, année, genre  
**Acteurs** : titre, nom  
**Diffusé** : film, lieu, heure-de-début  
**Critiques** : titre, note, description

S1:  
**Acteur** (IDA, prénom, nom, nationalité, annéeDenaissance)  
**Film** (IDF, titre), **JoueDans**(IDA, IDF)  
**DétailsFilm** (IDF, réalisateurF, genre, année)

S4:  
**Avis**(titre, date, note, commentaire)

S2:  
**Cinémas**(lieu, film, heureprojection)

S3:  
**CinémasALyon**(nom, titre, débutprojection)

S5:  
**GenresFilms**(titre, genre)

S7:  
**AnnéesFilms**(titre, année)

S6:  
**RéalisateurFilm**(titre, réal)

07/09/2023

46

## Discussions : GAV

- Supposons que nous ayons une source de données **S8** qui stocke des paires (acteur, réalisateur) qui ont travaillé ensemble sur des films.
  - La seule façon de modéliser cette source dans GAV est d'utiliser les deux descriptions suivantes qui utilisent NULL :
- Acteurs(**NULL**, acteur)  $\supseteq$  S8(acteur, réalisateur)
  - Film(**NULL**, réalisateur, **NULL**, **NULL**)  $\supseteq$  S8(acteur, réalisateur)
  - Ces descriptions créent des tuples dans le schéma qui incluent des **NULL** dans toutes les colonnes à l'exception d'une seule

S8:  
**TravaillerEnsemble**(acteur, réalisateur)

**Film** : titre, réalisateur, année, genre  
**Acteurs** : titre, nom  
**Diffusé** : film, lieu, heure-de-début  
**Critiques** : titre, note, description

07/09/2023

47

- Si la source **S8** contient les tuples (Robert, Alain) et (Paco, Copola), alors les tuples calculés pour SM seraient :
  - Acteurs(NULL, Robert), Acteurs(NULL, Paco)
  - Film(NULL, Alain, NULL, NULL), Film(NULL, Copola, NULL, NULL)
- Supposons maintenant que nous ayons la requête suivante qui recrée S8 :  
 Q(acteur, réalisateur) :- Acteurs(titre, acteur), Film(titre, réalisateur, genre, année)
- Nous ne pourrions pas récupérer les tuples de **S8** car les descriptions sources ont perdu la relation entre l'acteur et le réalisateur.

S8:  
**TravaillerEnsemble**(acteur, réalisateur)

**Film** : titre, réalisateur, année, genre  
**Acteurs** : titre, nom  
**Diffusé** : film, lieu, heure-de-début  
**Critiques** : titre, note, description

07/09/2023

48



## Local-as-View (LAV)

- décrit chaque source de données aussi précisément que possible et indépendamment de toute autre source de données.
  - approche opposée à GAV
- Schéma médiateur (SM) G
- Schéma de la source S1, S2, ..., Sn
  - Xi désigne les attributs dans Si

## LAV (Définition)

- Un mapping LAV M est un ensemble d'expressions de la forme:
  - $S_i(X_i) \subseteq Q_i(G)$ 
    - ✓ hypothèse du monde ouvert
  - ou,  $S_i(X_i) = Q_i(G)$ 
    - ✓ hypothèse du monde fermé
    - ✓ mais complétude pour les données sources et pas pour le médiateur

## Exemple de LAV

- S5. **GenresFilms**(titre, genre)  $\subseteq$  **Film**(titre, réalisateur, année, genre)
- S6. **RéalisateurFilm**(titre, réal)  $\subseteq$  **Film**(titre, réal, année, genre)
- S7. **AnnéesFilms**(titre, année)  $\subseteq$  **Film**(titre, réalisateur, année, genre)
- S8. **TravaillerEnsemble**(acteur, réalisateur)  $\subseteq$  **Film**(titre, réalisateur, année, genre), **Acteurs**(titre, acteur)

On peut également spécifier des contraintes sur le contenu

- S9(titre, année, "comédie")  $\subseteq$  **Film**(titre, réalisateur, année "comédie"), année  $\geq 1970$

**Film** : titre, réalisateur, année, genre

**Acteurs** : titre, nom

**Diffusé** : film, lieu, heure-de-début

**Critiques** : titre, note, description

S1:

**Acteur** (IDA, prénom, nom, nationalité, annéeDenaissance)

**Film** (IDF, titre), **JoueDans**(IDA, IDF)

**DétailsFilm** (IDF, réalisateurF, genre, année)

S4:

**Avis**(titre, date, note, commentaire)

S2:

**Cinémas**(lieu, film, heureprojection)

S3:

**CinémasALyon**(nom, titre, débutprojection)

S5:

**GenresFilms**(titre, genre)

S7:

**AnnéesFilms**(titre, année)

S6:

**RéalisateurFilm**(titre, réal)

07/09/2023

51

## Quelques ressources

- ❑ <https://www.elastic.io/wp-content/uploads/2019/09/ebook-Data-Integration-Best-Practices-mobile-final-small.pdf>
- ❑ <http://www.punjabiversity.ac.in/Pages/Images/elearn/BasicsofDataIntegration.pdf>
- ❑ <https://research.cs.wisc.edu/dibook/>
- ❑ [https://cours.etsmtl.ca/mti820/public\\_docs/acetates/MTI820-Acetates-ETL\\_1pp.pdf](https://cours.etsmtl.ca/mti820/public_docs/acetates/MTI820-Acetates-ETL_1pp.pdf)
- ❑ <https://www.diag.uniroma1.it/~lenzerin/homepagine/talks/TutorialPODS02.pdf>

07/09/2023

52

*C'est tout pour aujourd'hui*