

Modèles de Données et Intégration

Georges GARDARIN
Laboratoire PRISM, UVSQ
Versailles, France

© G. Gardarin

Plan

- Modèle patrimoniaux
- Modèle relationnel
- Modèle objet-relationnel
- Modèle XML
- Intégration et Médiation
- Et la sémantique ?

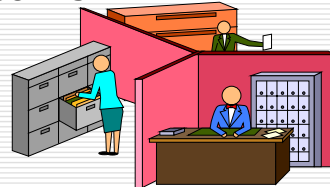
1. Réseau et Hiérarchique

➤ La première génération :

- BD = Ensemble de fichiers reliés par des pointeurs
- Modèles hiérarchique et réseau
- Langages navigationnels (Codasyl DML, IMS DL1)

➤ Dates clés

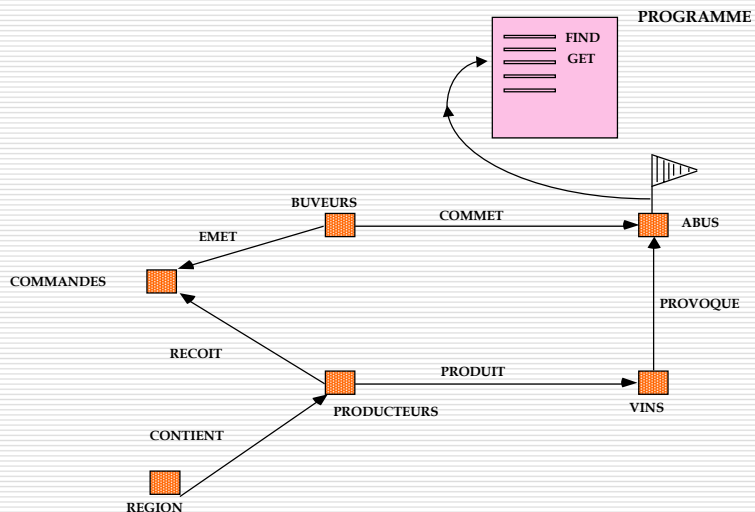
- 1966 : IDS.I, IMS.I
- 1971 : RECOMMANDATIONS DU CODASYL
- 1978 : CODASYL REVISITE



3

© G. Gardarin

Modèle Réseau : Exemple



4

© G. Gardarin

Avantages et Inconvénients

- Modélisation assez riche
- Performance des accès
- Langage procédural intégré à la programmation
- Très proche du physique
- Manque d'indépendance programme-données
- Pas de langage assertionnel
- Evolution de schéma difficile

2. Le Relationnel

- 2e GENERATION DE SGBD
- LANGAGE SQL NON PROCEDURAL
 - 1970 : THE RELATIONAL MODEL (CODD- ACM)
 - 1974 : DEBUT DES PROJETS R et Ingres
 - 1980 : COMMERCIALISATION SQL/DS, Oracle et Ingres
 - 1986 : NORMALISATION DE SQL1
 - 1989 : EXTENSION SQL1 (INTEGRITE)
 - 1992 : NORMALISATION DE SQL2 (COMPLETUDE)

Le modèle relationnel

- Modèle inventé par Codd (IBM 70)
- Simplicité de la représentation tabulaire
- Puissance du formalisme relation + logique

BUVEURS		
NB	Nom	Adresse
24	Pierre	Clermont
25	Jean	Riom
26	Paul	Thiers

- Représentation des données
 - tables à 2 dimensions
 - composées de lignes et colonnes

VINS		
NV	Cru	Millésime
1	Chablis	1996
2	Volnay	1978
3	Médoc	1984

ABUS		
NB	NV	Quantité
24	1	10
24	2	15
25	2	12

7

© G. Gardarin

L'algèbre relationnelle

- Ensembles d'opérations de base sur les tables
- Restriction
 - sélection des lignes selon un critère
 - $\sigma_{\text{cru} = \text{"Volnay"}}(\text{VINS})$
- Permet de représenter toute question de SQL
 - SELECT ... FROM ... WHERE
 - sous forme d'une expression d'opérations de l'algèbre
- Projection
 - sélection de colonnes par nom
 - $\Pi_{\text{cru, millésime}}(\text{VINS})$
- Jointure
 - composition sur colonnes comparables
 - $\text{VINS} \bowtie \text{ABUS}$
- Extensible avec calculs et agrégats
- Union, Intersection, Différence
 - opérations ensemblistes

8

© G. Gardarin

Le langage SQL

➤ La recherche

```
SELECT [DISTINCT]<COLONNE>+  
FROM <TABLE>+  
WHERE <CRITERE>+  
GROUP BY <COLONNE>+
```

➤ Les mises à jour

- INSERT INTO <TABLE>
TUPLES
- UPDATE <TABLE>
SET <COLONNE> = <expression>
WHERE <CRITERE>+
- DELETE <TABLE>
WHERE <CRITERE>+

Avantages et Inconvénients

➤ Un modèle simple et facile

- Table, Lignes et Colonnes

➤ Langage assertionnel puissant

- SQL2

➤ Une théorie bien assise

- Relation, Logique

➤ Le standard des données

- Depuis 1986

➤ Gestion et décisionnel

➤ De nombreux produits

➤ Couplage LP un peu difficile

- Impédance mismatch

➤ Ne supporte pas les données complexes

- Table en 1NF

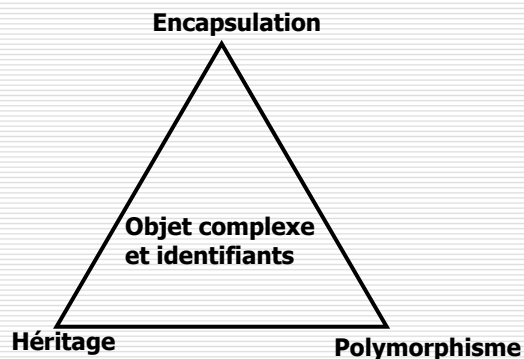
➤ Peu de sémantique

- Contraintes

3. Objet-Relationnel

- 1990 : The Object-Oriented Manifesto
 - 1991 : The Third Generation Manifesto
 - 1996 : L'objet-relationnel s'impose dans l'industrie
 - 1999 : Normalisation du langage SQL3
-
- **Fonctionnalités :**
 - extension du relationnel avec les concepts de l'objet
 - support de types de données abstraits extensibles
 - extension de SQL pour le support des objets
 - intégration des règles (triggers)

Concepts fondamentaux



L'apport des modèles objets

➤ Identité d'objets

- introduction de pointeurs invariants
- possibilité de chaînage d'objets

➤ Encapsulation des données

- possibilité d'isoler les données par des opérations
- facilite l'évolution des structures de données

➤ Héritage d'opérations et de structures

- facilite la réutilisation des types de données
- permet l'adaptation à son application

➤ Possibilité d'opérations abstraites (polymorphisme)

- simplifie la vie du développeur







13

© G. Gardarin

Extension des tables relationnelles

➤ Peuvent contenir des attributs de type complexe

➤ Peuvent être des tables d'objets

Incident Table			
Policy#	Incident#	Report	Photo
24	134		
24	219		
24	037		

14

© G. Gardarin

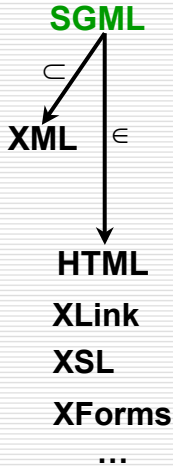
SQL3 et l'objet-relationnel

- Standard ISO accepté depuis 1999
 - SQL3 aussi appelé SQL99
- Objectif :
 - étendre le standard SQL2 (adopté par l'ISO en 1992) avec des fonctionnalités objets et autres
- Forme :
 - un ensemble de composants dont Fondation, PSM, CLI, ...
- Implémentation progressive dans les SGBD
 - DB2, Informix, Oracle, Versant, SQL Server ...

Avantages et Inconvénients

- | | |
|--|---|
| ➤ Un modèle riche dérivé du génie logiciel et des BD | ➤ Un modèle complexe |
| ➤ Garde les avantages du relationnel | ➤ La mariage de la carpe et du lapin |
| ➤ Un modèle très complet | ➤ Un langage de requêtes peu élégant et complexe (SQL3) |
| ➤ Un modèle extensible | ➤ Les types complexes ne sont pas standardisés |
| ➤ Un standard de requêtes <ul style="list-style-type: none">• SQL3 | |

4. XML: Rappels



```

<Vin>
  <Cru>Volnay</Cru>
  <Commentaire>Un des plus fameux</Commentaire>
  <Region>
    <Nom>Bourgogne</Nom>
    <Pays>France</Pays>
  </Region>
  <Millésime>1995</Millésime>
  <Millésime>1996</Millésime>
  <Prix Unité="Euro">10.5</Prix>
</Vin>

```

- Éléments balisés et attributs sont la bases**
- Les éléments peuvent être imbriqués et répétitifs**
- Une grammaire des balises peut être imposée (DTD, Schema)**
- Le vocabulaire des balises peut être issu d'un référentiel**

Modèle de données XML

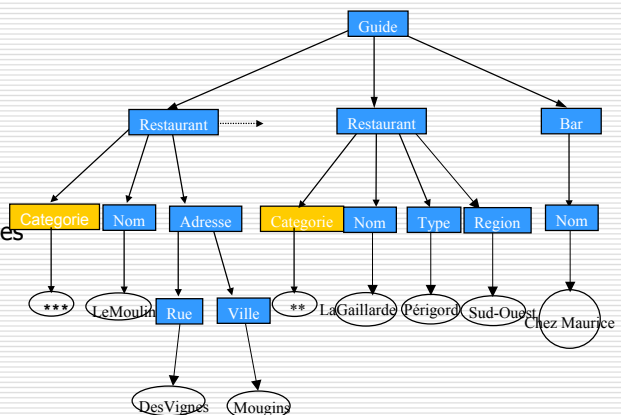
➤ Graphe à nœuds étiquetés

- Nœuds internes
 - balises et attributs
- Nœuds feuilles
 - données
- Arcs
 - Imbrications, hyperliens

➤ Schéma XML

- Types simples variés
- Types complexes extensibles
 - séquence, choix, tas
- Contraintes (cardinalités, patterns)

➤ Le schéma est optionnel



Langage d'interrogation pour XML

➤ SQL/XML

- Extension de SQL avec XPath

➤ XQuery

- Proposition du W3C

➤ Caractéristiques essentielles

- Documents et collections
- Langage fonctionnel
- Expressions de chemins XPath
- Expressions FLWR
- Tris et agrégats
- Expression conditionnelles
- Fonctions externes

```
for $r in
  doc("http://example.fr/Guide.xml")/Restaurant
```

```
where $r/@categorie = "****"
```

```
return
```

```
<Resultat>
```

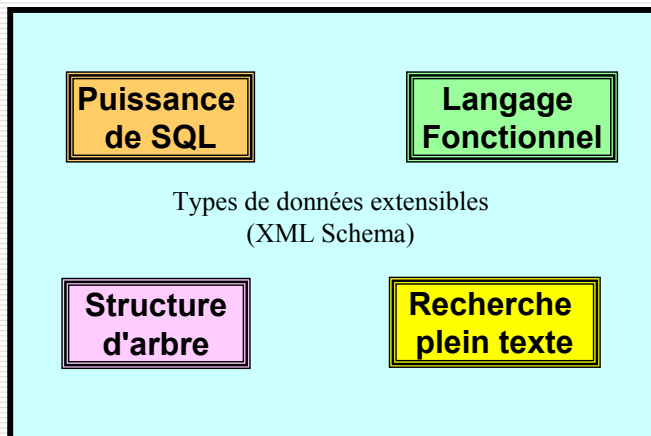
```
<Nom> {$R/Nom} </Nom>
```

```
<Adresse> {$R/Adresse}
```

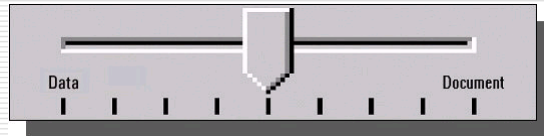
```
</Adresse>
```

```
</Resultat>
```

La puissance de XQuery



Avantages : XML est partout !



- Structuration forte
- Tables en bases relationnelles
- Mises à jour en place
- Intégrité référentielle
- SQL
- Indexation exacte
- OLTP et OLAP
- Grandes bases
- Structuration faible
- Objets complexes
- Versions
- Liens hypertexte
- Recherche plein texte
- Indexation approchée
- Documentaire
- Très grandes bases

21

© G. Gardarin

Requêtes sur structure et contenu

➤ Recherche sur mots et phrases dans des éléments

- Trouvez tous les livres avec la phrase "volcan d'Auvergne" et le mot "pouzzolane" dans le résumé.

➤ Recherche via thésaurus

- Trouvez tous les livres avec des mots phonétiquement similaires à "pouzzolane".
- Vérifier si les articles de Georges Hebrail incluent des travaux sur les règles associatives et d'autres méthodes de fouille de données.

➤ Requêtes avec classement par pertinence

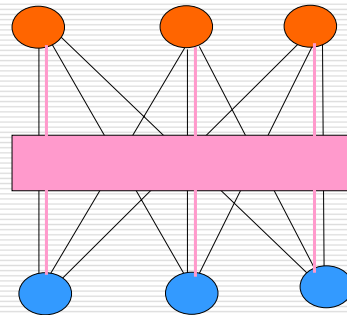
- Trouvez les trois meilleurs livre sur la fouille de texte.

22

© G. Gardarin

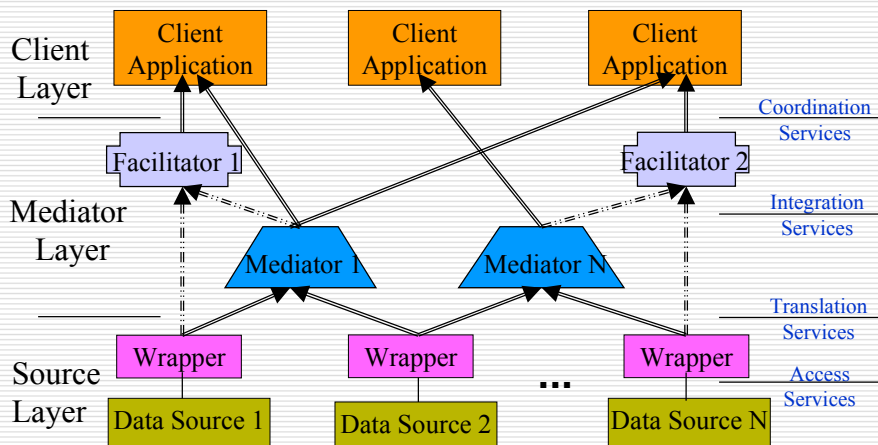
5. Intégration: Médiation d'information

- L'intégration d'information nécessite un modèle pivot (global, fédérateur, d'échange)
- Le relationnel à d'abord été utilisé
- XML est conçu pour cela
- XQuery est le langage de requêtes associé



$n+m$ versus $n*m$

Architecture de Référence (I3)



Un exemple de médiateur : XLive

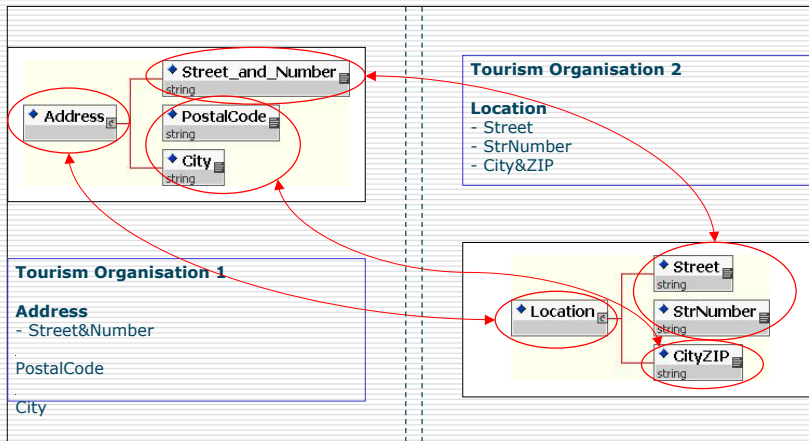
- Un médiateur développé à PRiSM comme véhicule de recherche
- Version industrielle dans une start-up
 - e-XMLMedia/XQuark Mediator
 - Open Source à www.xquark.org
- Version plus légère universitaire
 - XQuery limité
- Ciblage de thèmes de recherche
 - P2P
 - Médiation sémantique

Médiation et Web Sémantique

- Qu'et-ce-que le Web Sémantique ?
 - Un nouveau Web bien organisé avec la signification des informations exploitable par les machines
- Le rôle de XML est central
 - Toutes les données seront codées en XML
 - Toutes les métadonnées seront aussi en XML
 - Schema : définition de types de pages
 - RDF : description des ressources
 - OWL : définition d'ontologies
- Les moteurs de requêtes doivent être capables d'interroger toutes les données de manière intégrée

Oui mais la sémantique ?

➤ Nécessité de réconcilier les informations



27

© G. Gardarin

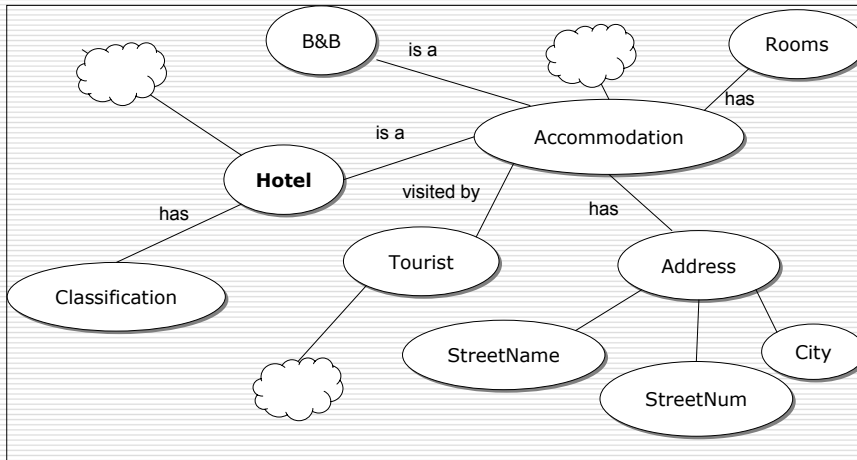
Utilisation d'ontologies

- Un ensemble de définitions de termes et de relations entre termes visant à décrire et représenter un domaine de connaissances.
- Dans le but de partager des informations sur un domaine
 - un domaine est un thème de connaissance spécifique comme le tourisme, la médecine, l'immobilier, la réparation automobile, la finance, ...
- Les définitions de concepts et relations inter-concepts d'une ontologie doivent être utilisables par un programme
 - exprimée en XML, par exemple avec OWL (recommandation W3C)
- Une ontologie code de la connaissance sur un domaine mais aussi inter-domaines
 - Les ontologies peuvent être étendues, modifiées, reliées, ...

28

© G. Gardarin

Exemple d'ontologie



29

© G. Gardarin

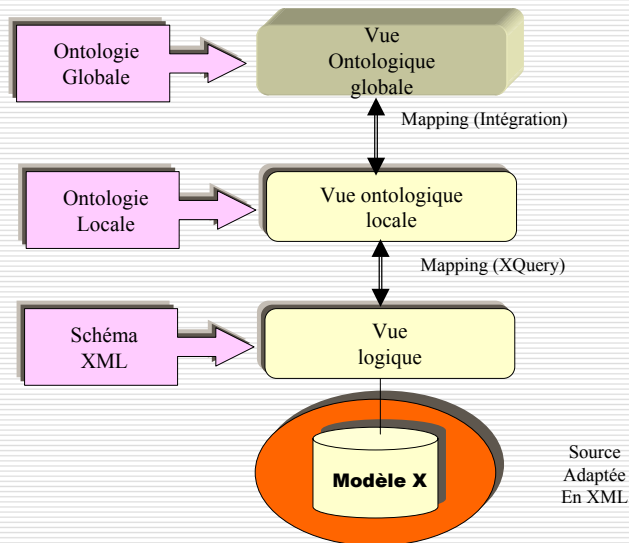
OWL

- Langage standard pour définir des ontologies
- Basé sur RDF schema
 - Modèle objet exprimé en RDF
 - RDF est un langage d'annotation
 - ressource – propriété → valeur (ou ressource)
- Étend les constructions de base pour apporter :
 - Plus d'interopérabilité (e.g., équivalences)
 - Plus de raisonnements (e.g., logique de description)
 - Plus d'évolution (e.g., intégration d'ontologies)
- Une hiérarchie de langages
 - XML → RDF → RDFS → OWL

30

© G. Gardarin

La médiation sémantique



31

© G. Gardarin

6. Conclusion

- Les modèles de données sont multiples et variés
- XML est LE standard émergent pour l'intégration de données
- Les Web services apportent la couche syntaxique pour l'interopérabilité applicative
- La médiation XML propose une technologie d'intégration de données hétérogènes basée sur XQuery (ou SQL/XML)
- Le Web Sémantique devrait apporter la couche sémantique nécessaire pour intégrer les données selon des ontologies
- XQuery vise à devenir le langage de requêtes universel pour les BD XML et le Web (Sémantique) ...

32

© G. Gardarin