



XML: une réalité d'aujourd'hui

Quels bénéfices attendre de pureXML dans le serveur hybride DB2 9 ?

*Corinne Baragoin
Francis Arnaudès*

*Spécialistes IT: Gestion de l'information
18 Septembre 2006*



Technologie pureXML de DB2 9

- Qui utilise XML aujourd'hui et pourquoi?
 - Pourquoi XML?
 - Quelles industries?
 - Comment utilisent-elles XML?
- Comment DB2 9, serveur hybride intègre la technologie IBM pureXML?
 - Les atouts d'un stockage hybride ?
- Utiliser le serveur hybride DB2 9
 - Bénéfices au travers d'exemples d'utilisation



Pourquoi XML ?

- **Souplesse , Souplesse , Souplesse !**
 - Toute plate-forme, éditeur, OS, software, langage
 - XML est indépendant de l'éditeur et de la plate-forme

- ***XML est un modèle de données très souple : pour données structurées, semi-structurées, pour données sans schéma, XML est plus souple que les données relationnelles car il permet:***
 - répétition
 - choix
 - contenu mixte
 - nesting

- **Facile à étendre: définir autant de nouveaux tags/éléments qu'il en est besoin**

- **Facile à valider: par exemple pour vérifier la conformité avec un schéma - tout parser XML peut le "faire"!**

- **Facile à transformer des documents XML dans d'autres formats (HTML, etc.)**

- **XML est une "brique" pour la prochaine génération d'e-Infrastructure:**
 - Web Services, Grid computing, Semantic Web, ...

Qui utilise XML? Tout le monde!

Un passage imposé par standards et obligations légales

Santé

- **HL7** – *Gestion du Patient: diagnostics, traitements, prescriptions, etc.*
 - <http://www.hl7.org>

Banque et Marchés Financiers

- **IFX** - *Interactive Financial Exchange : trades, banque, transactions client, etc.*
 - <http://www.ifxforum.org>

Assurance

- **ACORD** – *Gestion des Polices – Indemnités, réclamations, etc.*
 - <http://www.acord.org>

Distribution

- **IXRetail** – *Inventaire, transactions client, et gestion des employés*
 - <http://www.nrf-arts.org>

Business to Business cross-industrie (y compris Gouvernement)

- **SWIFT** – *Transactions B2B – Factures, Commandes, statut inventaire, etc.*
 - <http://www.swift.com>

..... Et des centaines encore

Un exemple: Financial Data (FIXML)

▪ Achat de 1000 actions d'IBM .

8=FIX.4.2^9=251^35=D^49=AFUNDMGR^56=ABROKER^34=2
 ^52=20030615-01:14:49^11=12345^1=111111^63=0^64=2003
 0621^21=3^110=1000^111=50000^55=IBM^48=459200101^22=
 1^54=1^60=2003061501:14:4938=5000^40=1^44=15.75^15=USD
 ^59=0^10=127

Ancien protocole FIX

```
<FIXML >
  <NewOrdSingle  ClOrdID ="123456"
    Side ="2"
    TransactIm ="2003 -06 -15T01:14:49 -05:00"
    OrderType ="2"
    Price ="93.25"
    Acct ="26522154">
    <Header  Sent ="2001 -06 -21T01:31:28 -05:00"
      PosDup ="N"
      PosRsnd ="N"
      SeqNum ="521">
      <Sender  ID ="AFUNDMGR"/>
      <Target  ID ="ABROKER"/>
    </Header >
    <Instrument  Symbol ="IBM"
      ID ="459200101"
      IDSrc ="1"/>
    <OrderQuantity  Qty ="1000"  Cur ="USD"/>
  </NewOrdSingle >
</FIXML >
```

Nouveau protocole FIXML

- Extensible
- Développement et maintenance moins coûteux

Pourquoi XML?

Solution aux problématiques métier et IT d'aujourd'hui

- Intégration B2B et B2C
 - Mécanisme de transport indépendant de la plateforme

Ordres d'achat déclenchent des transactions transitant dans une architecture orientée Services

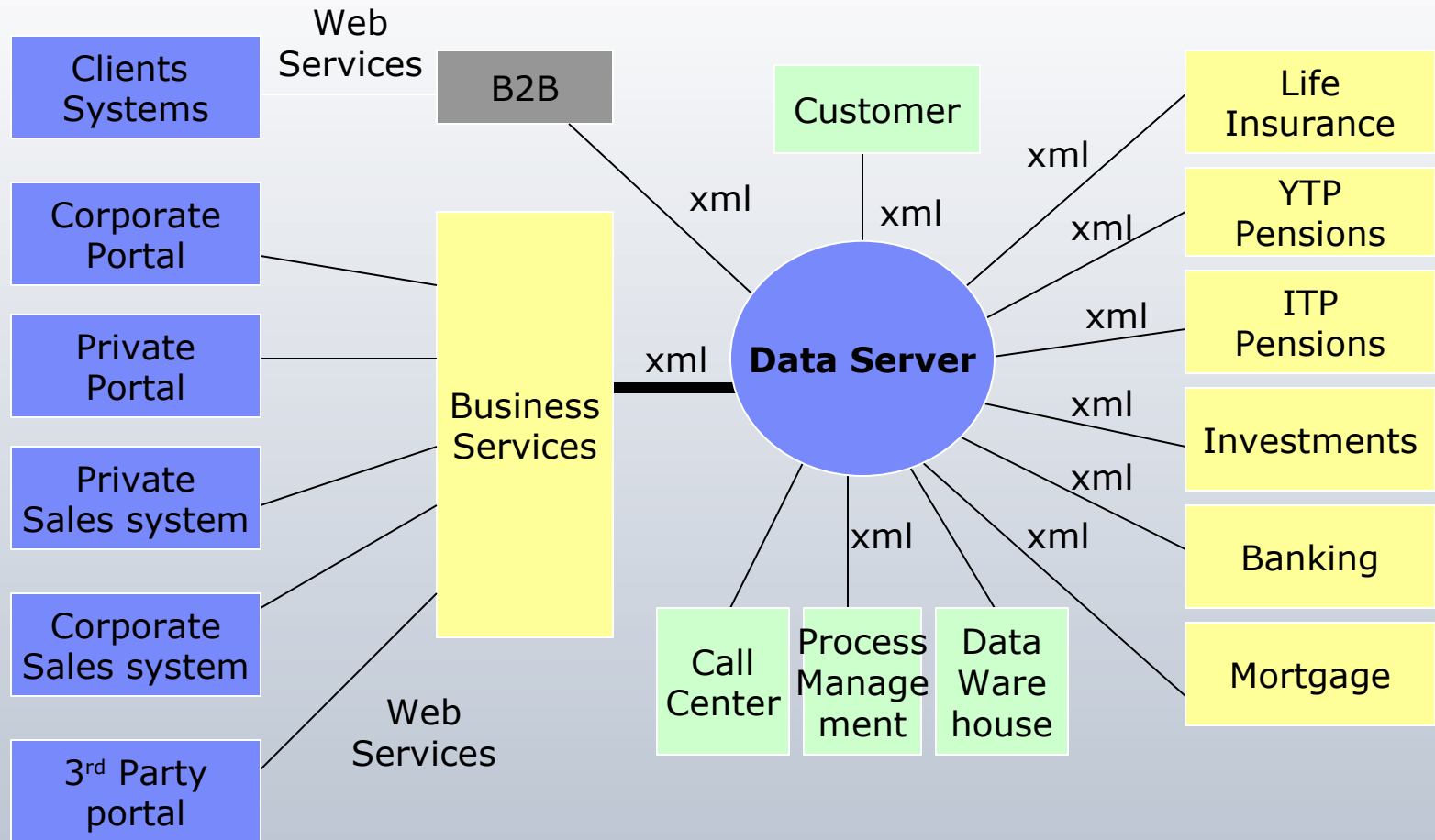
- Gestion de documents
 - Gouvernement et industries légales (Finances, Assurances...) nécessitent la signature électronique
 - Documents contiennent souvent des sous-documents, qui s'enrichissent dans le temps

Livres, chapitres et sous-chapitres



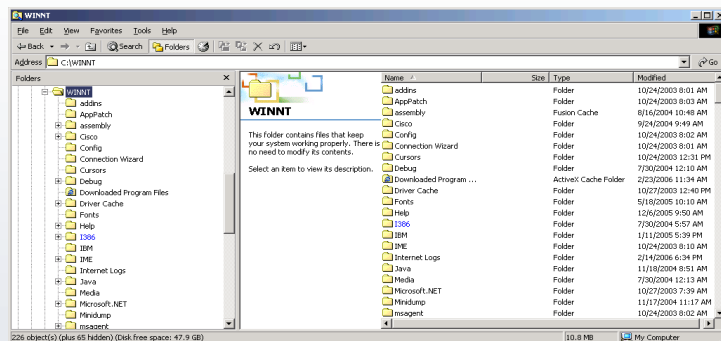
DB2 9 renforce les solutions SOA

XML est le lien

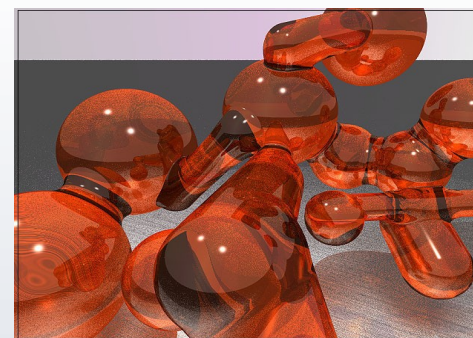


Où est XML?

XML peut être vu soit comme des données structurées soit comme un document de contenu



Dans des fichiers...



Dans LOBs...



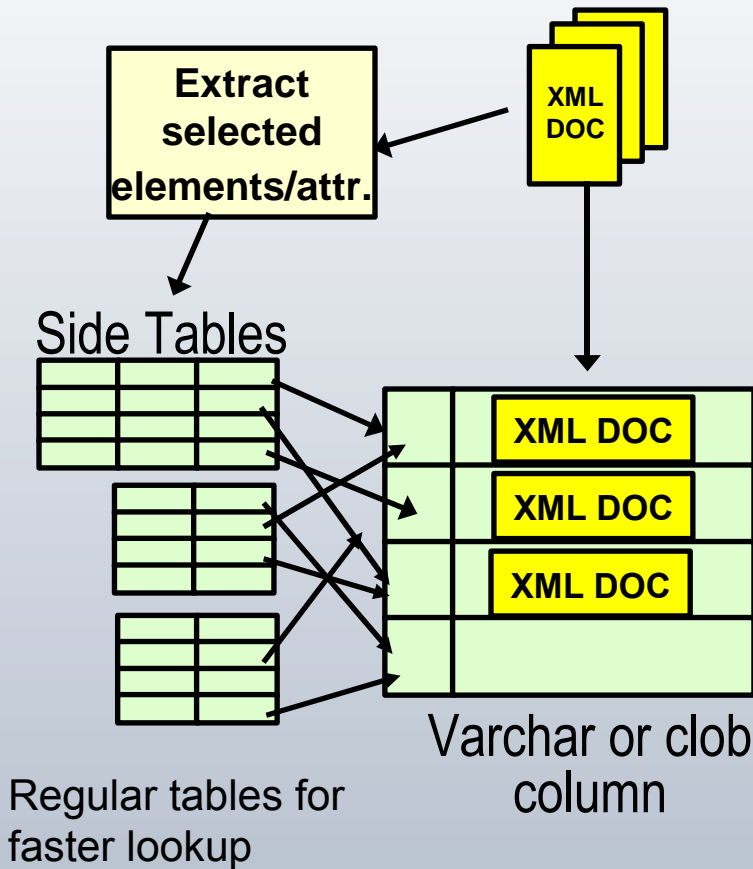
Décomposé dans des tables



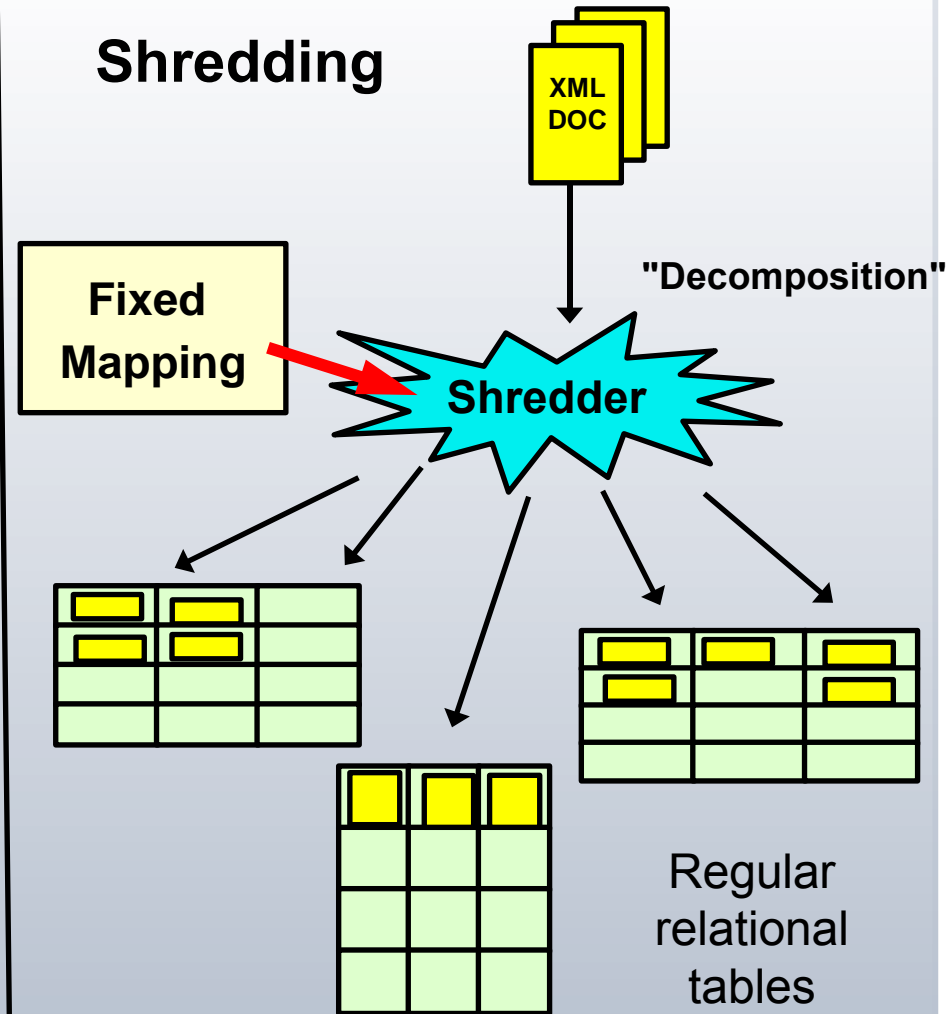
Dans une DB XML

XML-Enabled Databases: 2 options

CLOB/Varchar



Shredding



Technologie pureXML de DB2 9

- Qui utilise XML aujourd'hui et pourquoi?
 - Pourquoi XML?
 - Quelles industries?
 - Comment utilisent-elles XML?
- Comment DB2 9, serveur hybride intègre la technologie IBM pureXML?
 - Les atouts d'un stockage hybride ?
- Utiliser le serveur hybride DB2 9
 - Bénéfices au travers d'exemples d'utilisation



Bases de données XML

▪ *XML-enabled Databases*

- Le modèle de données est relationnel (non XML)
- Mapping entre le modèle de données XML et les données relationnelles
nécessite de connaître le modèle relationnel ou de stocker le fichier XML comme du texte
 - Un exemple DB2 XML Extender

▪ *Native XML Databases*

- Utilisent le modèle de données hiérarchique XML pour stocker et traiter XML en interne
- Pas de mapping, pas de stockage texte
- Format de stockage = Format de traitement
 - Un exemple: le serveur hybride DB2 9

Le mapping est complexe et fragile

- *The simplest rule for shredding XML is to create a new table for every element that can occur multiply within its parent element.*
- *Elements that occur exactly once within a parent element are placed in the same table as the parent.*
- *Optional elements are problematic.*

```

<? xml version="1.0" ?>
<purchaseOrder id="12345" secretKey="4x%$^">
  <customer id="A6789">
    <name>John Smith Co</name>
    <address>
      <street>1234 W. Main St</street>
      <city>Toledo</city>
      <state>OH</state>
      <zip>95141</zip>
    </address>
  </customer>
  <itemList>
    <item>
      <partNo>A54</partNo>
      <quantity>12</quantity>
    </item>
    <item>
      <partNo>985</partno>
      <quantity>1</quantity>
    </item>
  </itemList>
</purchaseOrder>
    
```

POid	CustId
987564331	A6789
....	...

POid	PartNo	Quantity
987564331	A54	12
987564331	985	1
....

Custid	Name	Street	City	State	Zip
A6789	John Smith Co	1234 W. Main St	Toledo	OH	95141
...

Shredding: et quand un schéma change...

Employees are now allowed to have multiple phone numbers..."

```
<DEPARTMENT deptid="15" deptname="Sales">
  <EMPLOYEE>
    <EMPNO>10</EMPNO>
    <FIRSTNAME>CHRISTINE</FIRSTNAME>
    <LASTNAME>SMITH</LASTNAME>
    <PHONE>408-463-4963</PHONE>
    <PHONE>415-010-1234</PHONE>
    <SALARY>52750.00</SALARY>
  </EMPLOYEE>
  <EMPLOYEE>
    <EMPNO>27</EMPNO>
    <FIRSTNAME>MICHAEL</FIRSTNAME>
    <LASTNAME>THOMPSON</LASTNAME>
    <PHONE>406-463-1234</PHONE>
    <SALARY>41250.00</SALARY>
  </EMPLOYEE>
</DEPARTMENT>
```

Requires:

- Normalization of existing data !
- Modification of the mapping
- Change of applications

Phone

EMPNO	PHONE
27	406-463-1234
10	415-010-1234
10	408-463-4963

Department

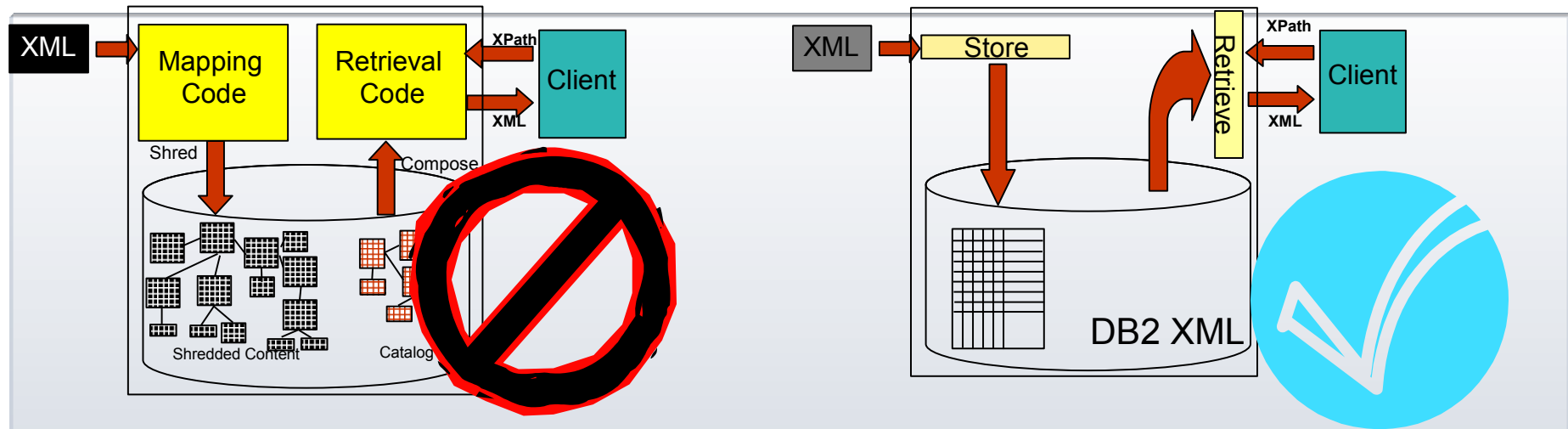
DEPTID	DEPTNAME
15	Sales

Costly!

Employee

DEPTID	EMPNO	FIRSTNAME	LASTNAME	PHONE	SALARY
15	27	MICHAEL	THOMPSON	406-463-1234	41250
15	10	CHRISTINE	SMITH	408-463-4963	52750

Bénéfices de DB2 9 pureXML



▪ Solution simplifiée et directe

- Pas de code de mapping à écrire et maintenir
- Pas de schema complexe à gérer et maintenir
- Pas de catalogue propriétaire
- Pas de parsing XPath et de recomposition du résultat

▪ Performance et souplesse accrues

▪ Moins de coûts de développement et de maintenance

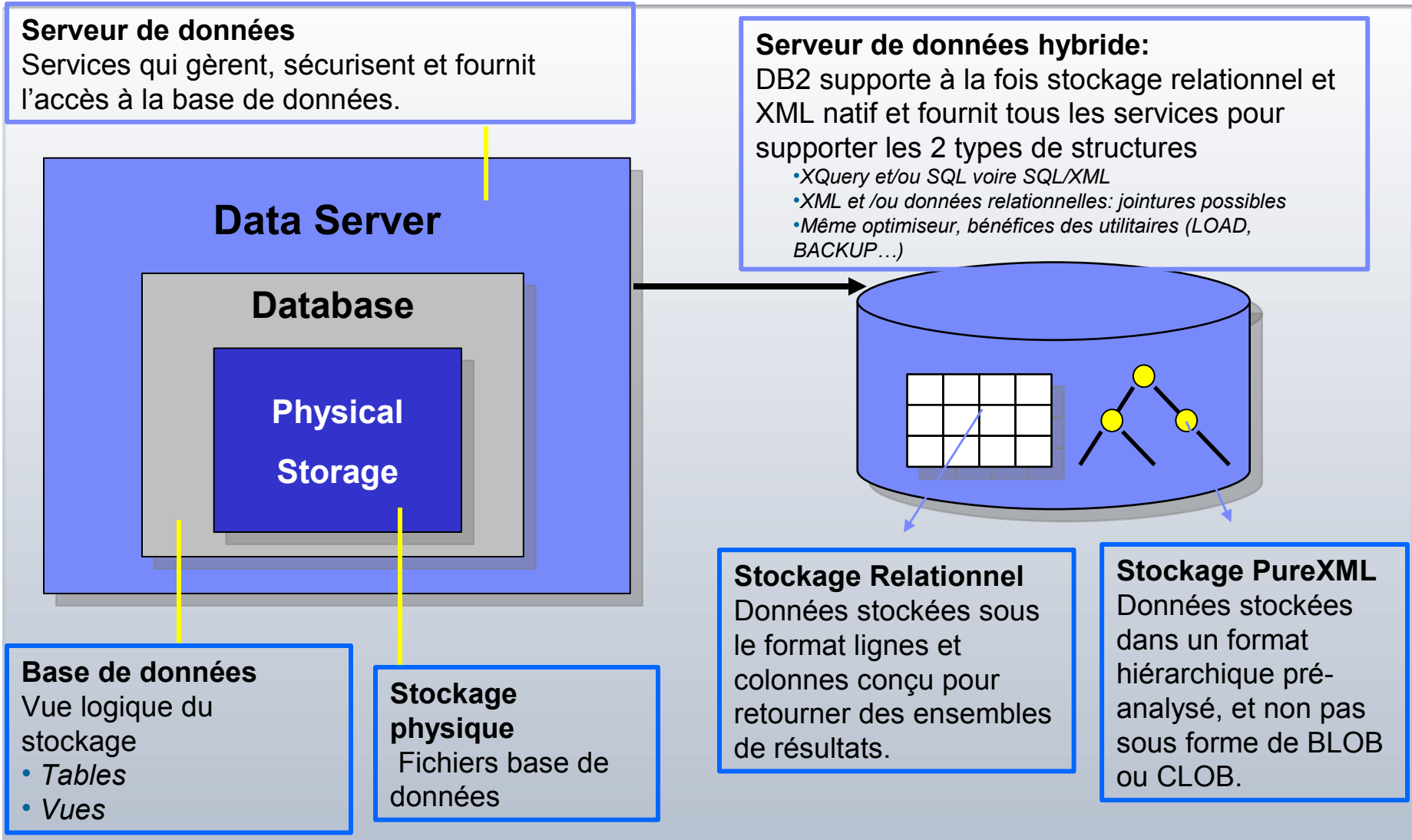
Garder XML sous XML

Chaque transformation (ex shredding) est coûteuse et peut entraîner des erreurs

XML peut être trop complexe à décomposer

Le format hiérarchique XML correspond mieux à des formes d'organisation de pensée humaine

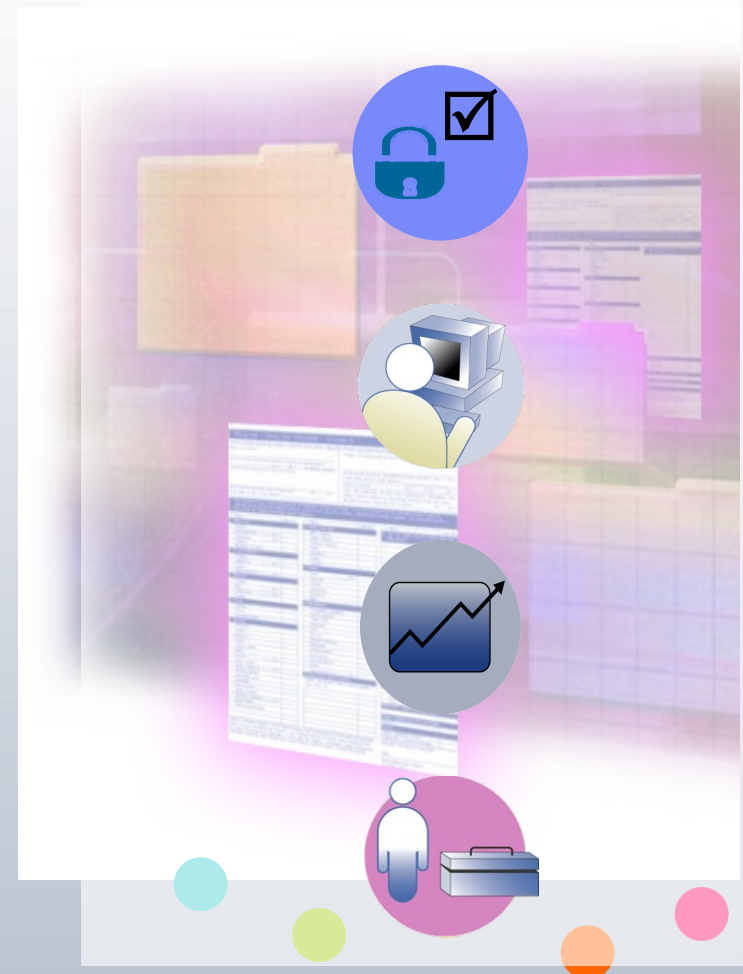
Serveur hybride DB2 9: à la fois stockage pureXML et relationnel



Données XML ont besoin de la maturité du relationnel

Complémentaire au traitement XML

- **Gérer de larges volumes de données XML est une question de base de données!**
 - *Efficient Search & Retrieval of XML*
 - *Persistency, Recovery, Transactions, ACID*
 - *Performance, Scalability*
 - *...all the same reasons as for relational data!*
- **Données XML ont besoin d'être protégées**
 - *Backup and recovery features to ensure continuity*
 - *Data is protected using database security*
- **Intégration**
 - *Integrate new XML data with existing relational data*
 - *Publish (relational) data as XML*
 - *Database support for web applications, SOA, web services (SOAP)*
- **Optimiser les investissements existants**
 - *Use existing technology infrastructure and skills to store and manage both relational and XML*



DB2 9 Serveur Hybride : les 2 mondes.... Et le meilleur des 2

Beaucoup de données dans des base relationnelles

Cela ne changera pas ...car les bases SQL sont rapides

Mais XML peut être un meilleur choix que le relationnel pour ...

- Données par nature hiérarchisées ou imbriquées
 - Exemple: Données médicales, Objets multi-valués, ...
- Ensemble de données avec des attributs partiellement renseignés
 - Exemple: FIXML, FpML, Profils clients
- Schémas qui évoluent continuellement
 - Exemple: Changements fréquents des services/produits/processus
- Schémas variables, de nombreux schémas
 - Exemple: Intégration de données, consolidation de sources de données variées
- Combiner données structurées et non structurées
 - Exemple: CM, Sciences de la Vie, News & Media

Problèmes résolus avec DB2 9 – Technologie pureXML

■ **Recherche de données**

- Besoin de rechercher tout composant au sein du document XML
- Besoin de retrouver très vite des ensembles de données



- Le stockage natif XML décrit chaque champ
- Combinaison de SQL et XQuery dans la requête (interface graphique)

■ **Shredding**

- Se débarrasser des contraintes liées au shredding



- Il n'y a plus de shredding avec DB2 9

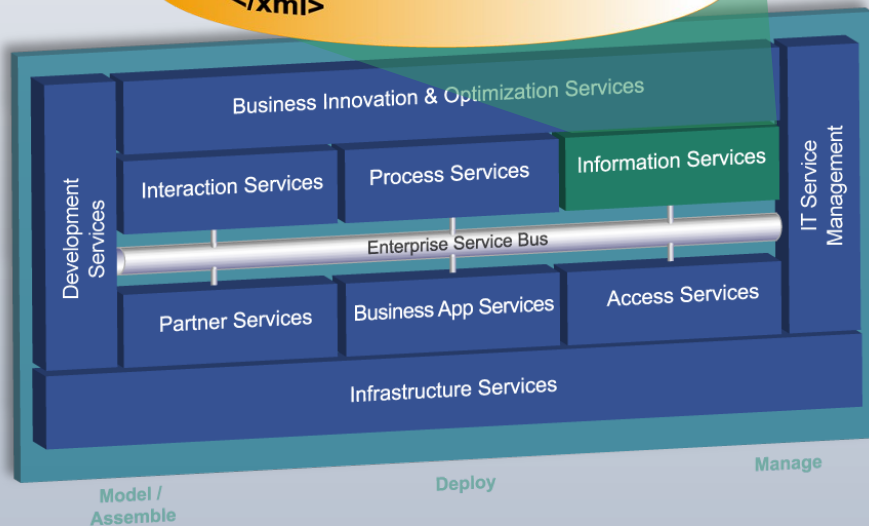
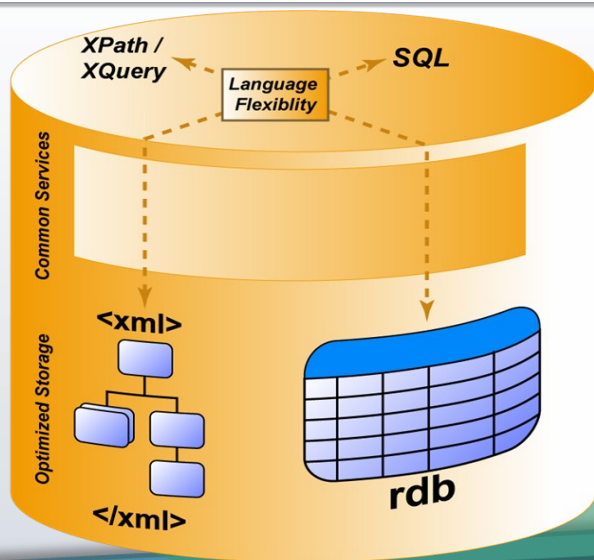
■ **Flexibilité**

- Pouvoir changer toute donnée à tout moment



- La fonction 'Schema évolution' traite les shemas multiples

DB2 9 pureXML, Serveur Hybride



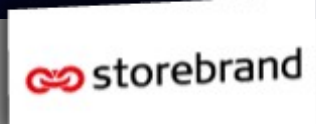
	Shred	CLOB	XML db	pureXML Hybrid
Information Fidelity	✗	✓	✓	✓
Integration	✓	✗	✗	✓
Schema Flexibility	✗	=	✓	✓
Performance/Scale	=	✗	✗	✓
Programming Models	✗	✗	=	✓
Manageability	✓	✗	✗	✓

Technologie pureXML de DB2 9

- Qui utilise XML aujourd'hui et pourquoi?
 - Pourquoi XML?
 - Quelles industries?
 - Comment utilisent-elles XML?
- Comment DB2 9, serveur hybride intègre la technologie IBM pureXML?
 - Les atouts d'un stockage hybride ?
- Utiliser le serveur hybride DB2 9
 - Bénéfices au travers d'exemples d'utilisation



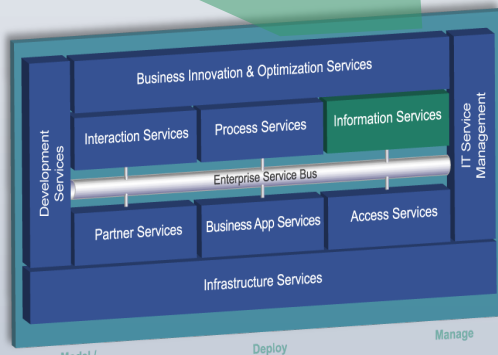
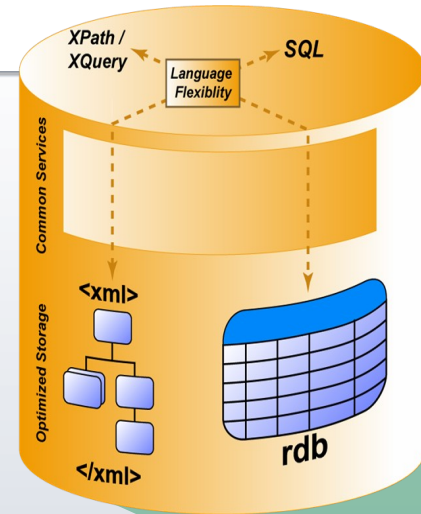
DB2 9 XML: Le retour des premiers tests client



Ce client basé en Norvège est un des leaders de l'assurance

Résultats proto-type utilisant solution SOA avec DB2 9

- Rapidité, accès aisé aux informations enrichies de produit & client
- Création plus rapide de produits personnalisés que clients demandent

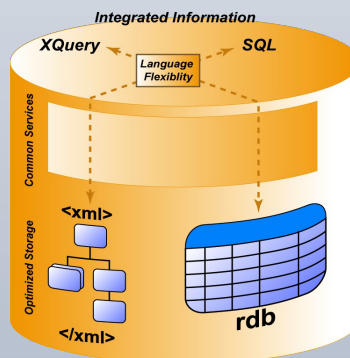


Facteurs de succès

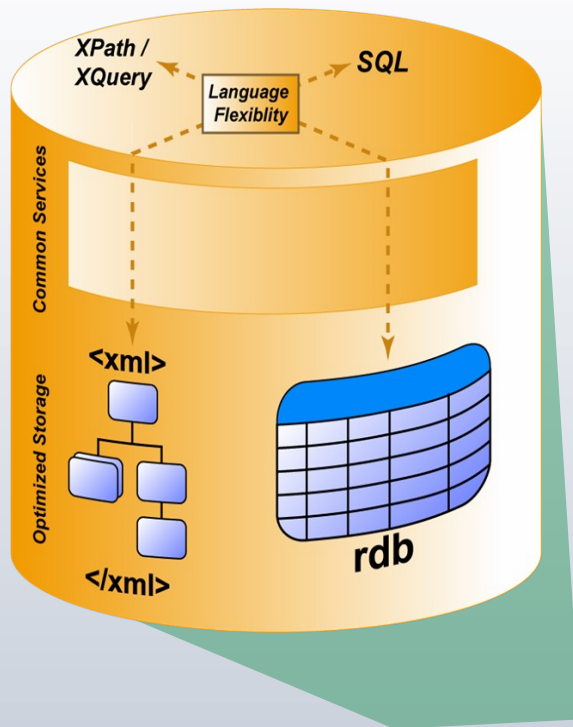
- Séparation des Process et Information
- Gestion intégrée de l'information XML
- Echanges entre processus sous format XML

Development and app. performance re: XML data	with relational data server	with DB2 9 hybrid data server
Development of search & retrieval business processes	CLOB: 8 hrs Shred: 2 hrs	30 min.
Add field to schema	1 week	5 min.
Relative lines of I/O code (65% reduction)	100	35
Queries	24 - 36 hrs	20 sec - 10 min
Query non-shredded XML element	1 week	½ day

Partenaires et solutions XML basées sur DB2 9

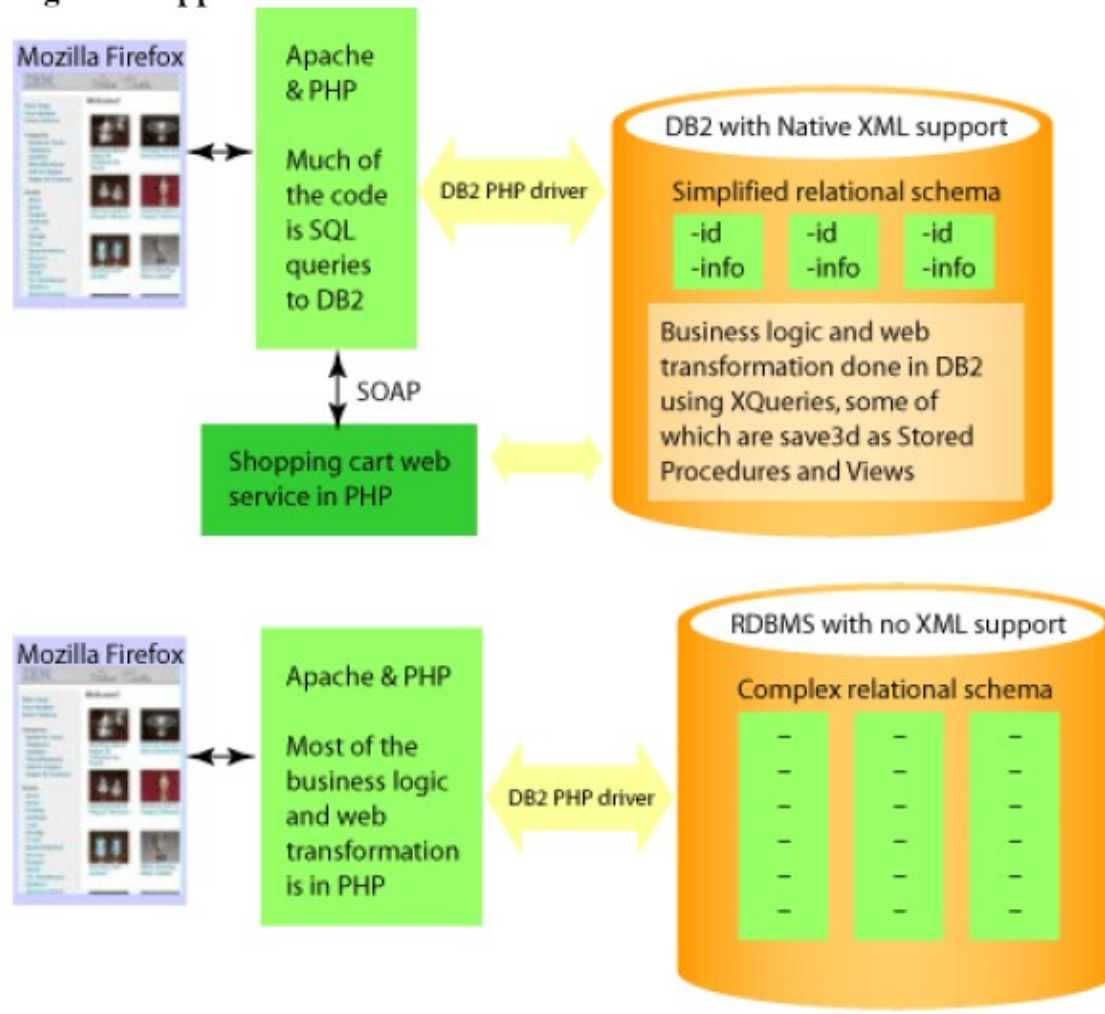


Exemple d'utilisation du serveur hybride – DB2 9



DB2 9 serveur Hybride	Serveur de Données Relationel	Serveur de Données natif XML
Description du modèle	✓	✓
Exemple de code: Chargement de données natives XML	✓	✓
Modification du modèle	✓	✓
Création d'une vue		✓
Création d'une procédure stockée		✓

Application XML Versus Relationnel



Création du modèle

Modèle natif XML

Product Table

- PID BIGINT
- INFO XML

Customer Table

- CID BIGINT
- INFO XML

Purchase Order

- POID BIGINT
- INFO XML

Modèle natif Relationnel

Product Table

- PID BIGINT
- NAME
- DETAILS
- PRICE WEIGHT
- CATEGORY
- BRAND
- SIZE
- DESCRIPTION
- CLOB

Image Table

- PID BIGINT
- TYPE
- ALIAS
- IMAGE
- VARCHAR

Customer Table

- CID BIGINT
- NAME
- COUNTRY
- STREET
- CITY
- PROVINCE
- ZIP
- PHONE
- INFO CLOB

Purchase Order

- POID BIGINT
- PODATE
- PORDER
CLOB

Chargement de données XML

```
$conn =db2_connect($dbname, $dbuser, $dbpass);  
$fileContents = file_get_contents("products/p1.xml");  
$dom = simplexml_load_string($fileContents);
```



Code commun

DB2 pureXML

```
$stmt =db2_prepare($conn, "INSERT INTO xmlproduct VALUES (?, ?);"  
db2_execute($stmt, array($prodID, $fileContents);
```

DB2 Relationnel

```
$prodID = (string) $dom["pid"];  
$prodName = (string) $dom->description->name;  
$prodDetails = (string) $dom->description->details;  
$prodPrice = (float) $dom->description->price;  
$images = array();
```



Shredding

```
....  
$stmt = db2_prepare($conn, "INSERT INTO sqlimages(Pid,Type,Alias,Location) VALUES  
(?, ?, ?, ?);"  
db2_execute($stmt, array($prodID, 'thumbnail', $prodImgAlias, $prodImgThumb));
```

•Demo

Modification du modèle de données

DB2 pureXML

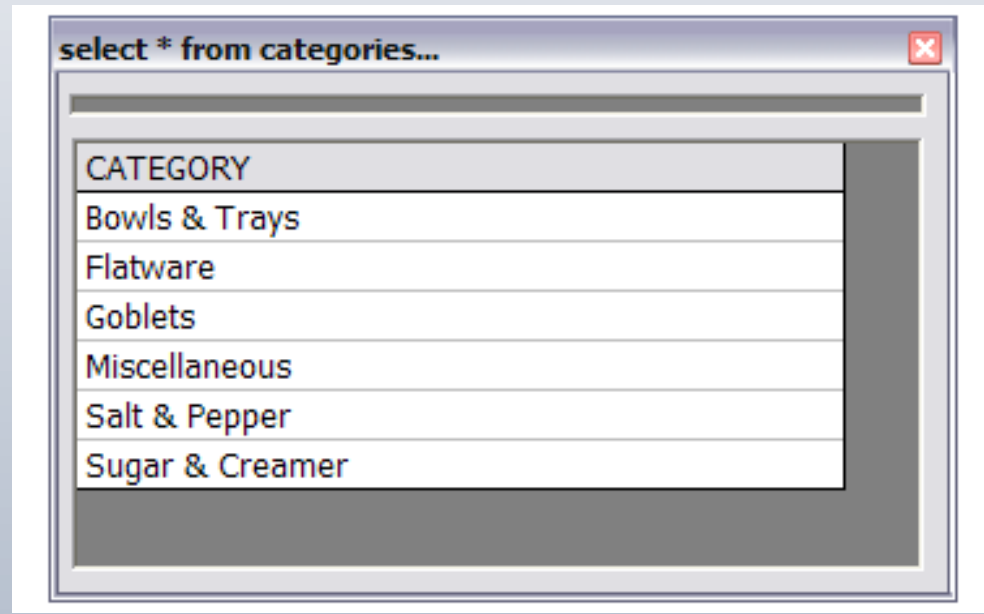
- Pas de modification
- Mise à jour des requêtes XQuery
- Peu ou pas de modification du code

DB2 Relationnel

- Rechargement des données pour intégrer les nouvelles colonnes
- Mise à jour des ordres INSERT
- Modification du code, par exemple le shredding

Intégration : Exemple de vue XML

```
CREATE VIEW Categories(Category) AS SELECT  
DISTINCT(XMLCAST(  
XMLQUERY('for $i in  
$t/product/description/category return $i'  
PASSING BY REF T.DESCRPTION AS "t"  
RETURNING SEQUENCE) AS  
VARCHAR(128)))  
FROM xmlproduct AS t;
```



The screenshot shows a window titled "select * from categories...". The window displays a table with a single column labeled "CATEGORY". The table contains the following rows:

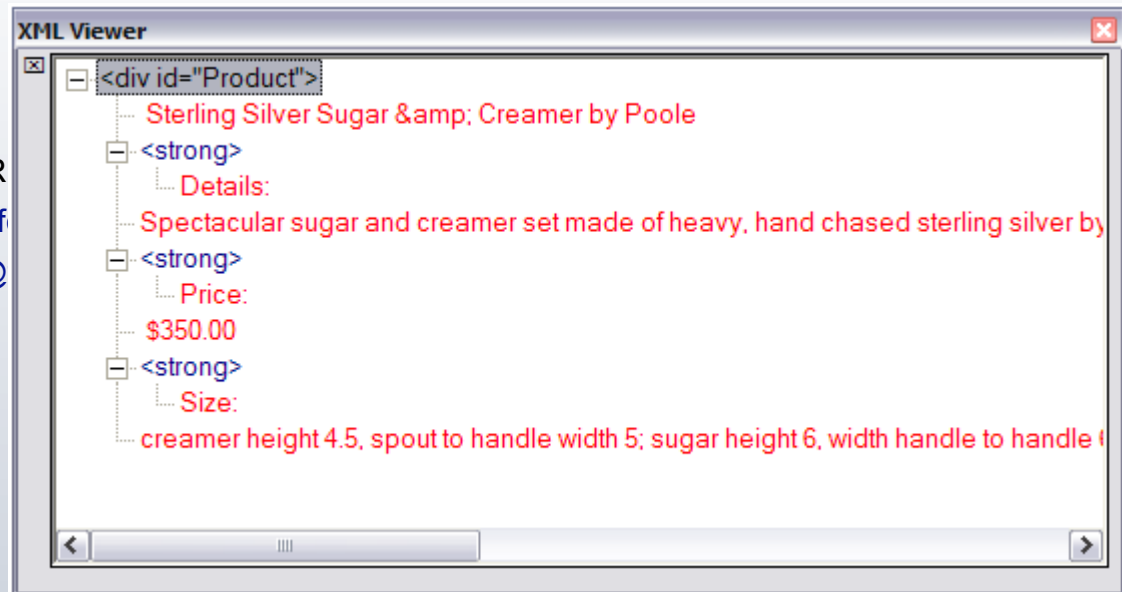
CATEGORY
Bowls & Trays
Flatware
Goblets
Miscellaneous
Salt & Pepper
Sugar & Creamer

Intégration : Exemple de Procédure Stockée

```

CREATE PROCEDURE getProduct(IN id VARCHAR(10))
DYNAMIC RESULT SETS 1
LANGUAGE SQL
BEGIN
  BEGIN
    DECLARE c_cur CURSOR WITH RETURN
      SELECT XMLSERIALIZE(XMLQUERY('f
let $thumb := $i/description/images/image[@
let $name := $i/description/name/text()
let $details := $i/description/details/text()
let $price := $i/description/price
return
  <div id="Product"> {$name}
    <strong>Details: </strong> {$details}
    <strong>Price: </strong> {$price/text()}
    <strong>Size: </strong> {$size/text()} {$size/@units/text()}
  </div>
' PASSING T.DESCRPTION AS "t" RETURNING SEQUENCE) AS CLOB(32K)) FROM xmlproduct
  T WHERE Pid = id;
    OPEN c_cur;
  END;
END

```



En Résumé – DB2 9 pureXML

Souplesse, intégration, simplicité pour le développement

- *Nouvelles techniques de stockages pureXML*
- *Nouvelles techniques d'indexation*
- *Nouveau langage XQuery*
- *Validation des documents XML (XML Schema)*
- *Nouveaux outils d'administration (Optimiseur, Import/Export, Explain, Haute Disponibilité, etc.)*
- *Support des API (Java, C/C++, .NET, PHP, etc.)*
- *...et plus*

DB2 9

**Solution
robuste et
éprouvée pour
le
développement
de nouvelles
applications**

DB2 9 XML : Ressources

- **Websites**
 - DB2 XML – <http://www.ibm.com/db2/xml>
- **White papers on developerworks:**
 - "Managing XML for Maximum Return"
 - "Use DB2 native XML with PHP "
 - "An Overview of Native XML Support in DB2"
 - "Comparing XML and relational storage – A best practices guide"
- **Redbook:**
 - « DB2 9: pureXML Overview and Fast Start »: SG24-7298
 - Remplissez le formulaire pour le recevoir !!!!
 - Notez si vous êtes intéressés par un Workshop technique/ POT sur DB2 9 XML