



IBM Software Group

# Le « Service Information »

Isabelle Claverie-Berge , Consultant Certifié Information Management , IBM France  
Michel Caussanel , Websphere Information Integration Team IBM France



© 2005 IBM Corporation

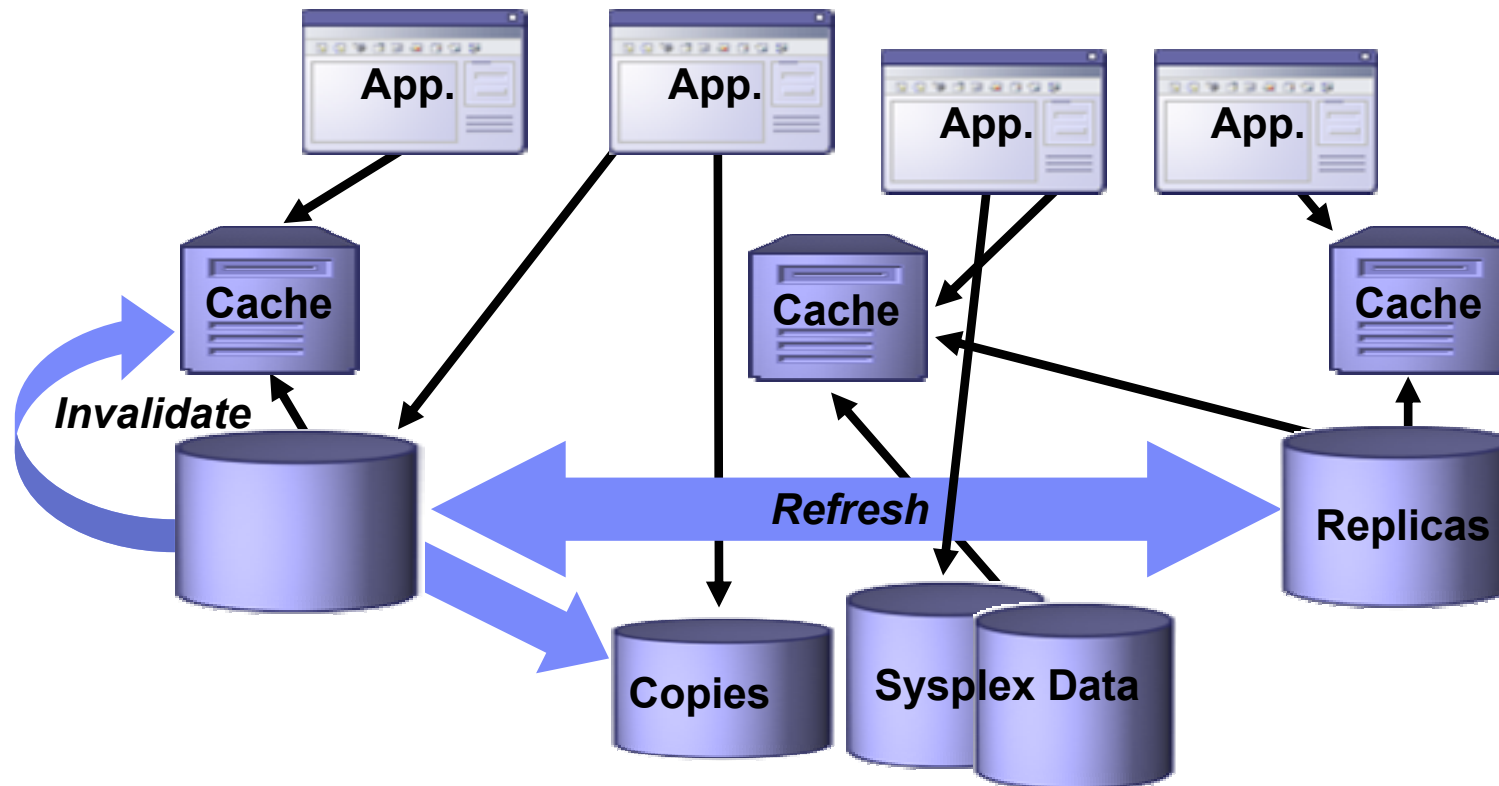
# Agenda

- Introduction : Pourquoi un « Service Information » dans une approche SOA
  - ▶ Le monde aujourd'hui
  - ▶ Le monde idéal
  - ▶ La réponse IBM
- L'infrastructure du « Service Information »
  - ▶ Recherche
  - ▶ Fédération
  - ▶ Transformation
  - ▶ Placement
  - ▶ Publication
- Conclusion

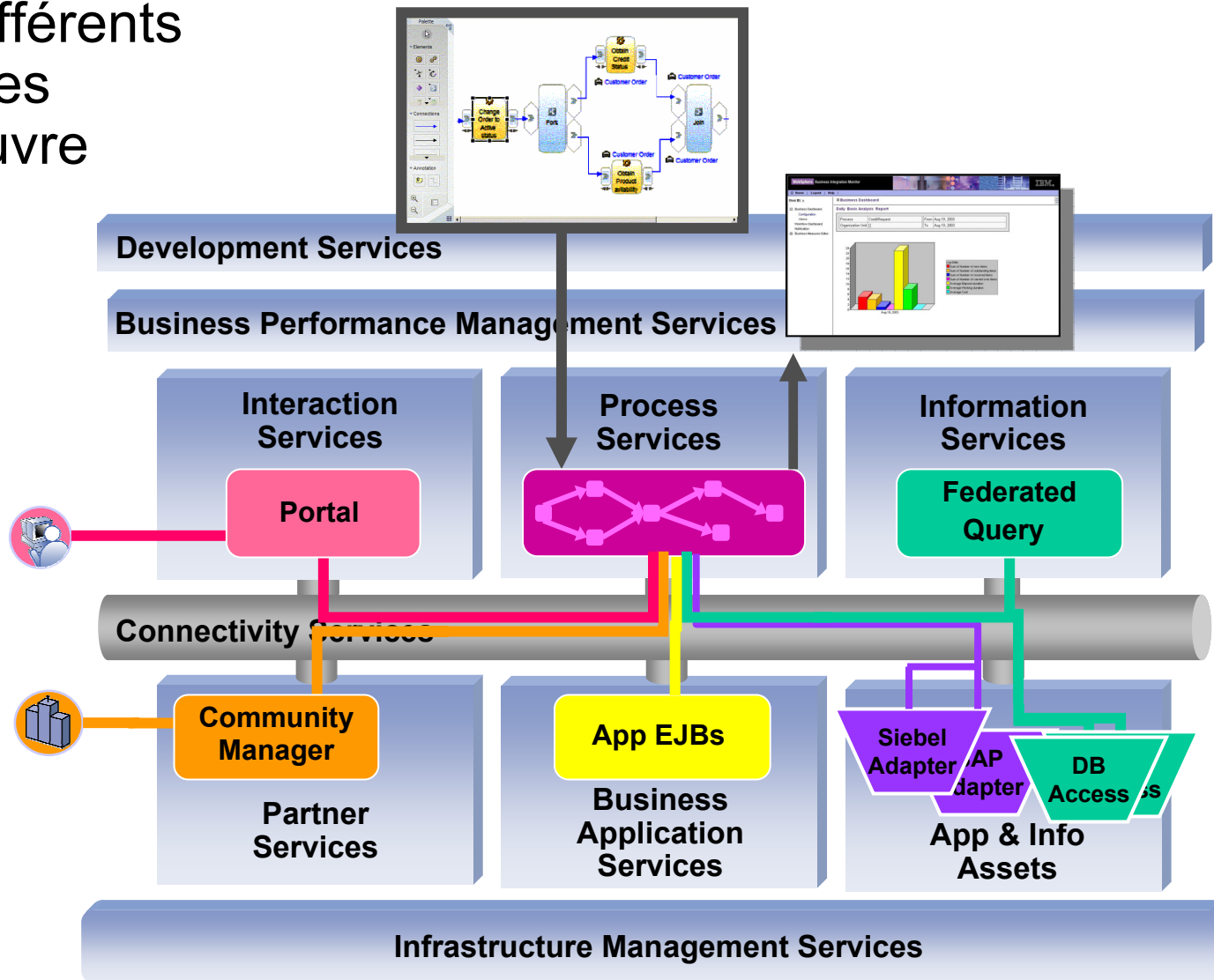


## Le monde de l'information aujourd'hui : complexe et coûteux

- Des informations hétérogènes et distribuées
- Des applications qui créent et maintiennent des caches et des réplicats
- Une prolifération de copies
- 30%- 50% des charges de conception sont consacrées à la gestion des duplications
- Aucune qualité garantie et pilotable sur les services attendus



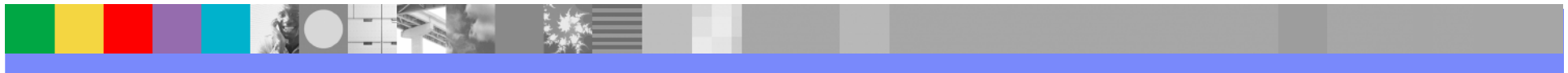
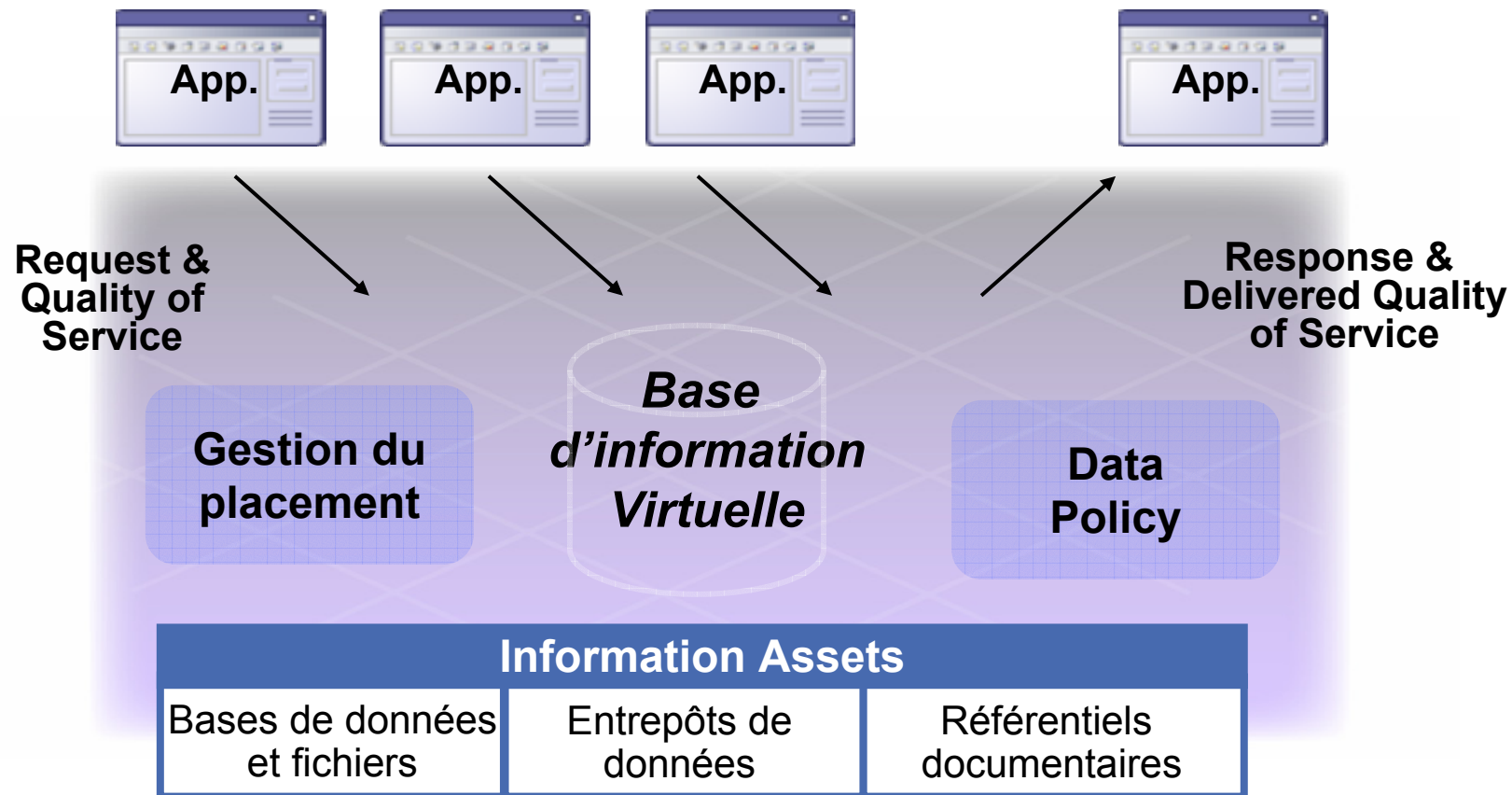
# Les différents services à l'oeuvre



# Le monde idéal

## ■ Base de données Virtuelle

- Politique de gestion des informations : “Cache” et répliquions mis en oeuvre en réponse à une demande de “Qualité de services”
- Gestion autonome du placement des informations
- Fédération des divers entrepôts pour utiliser les ressources existantes

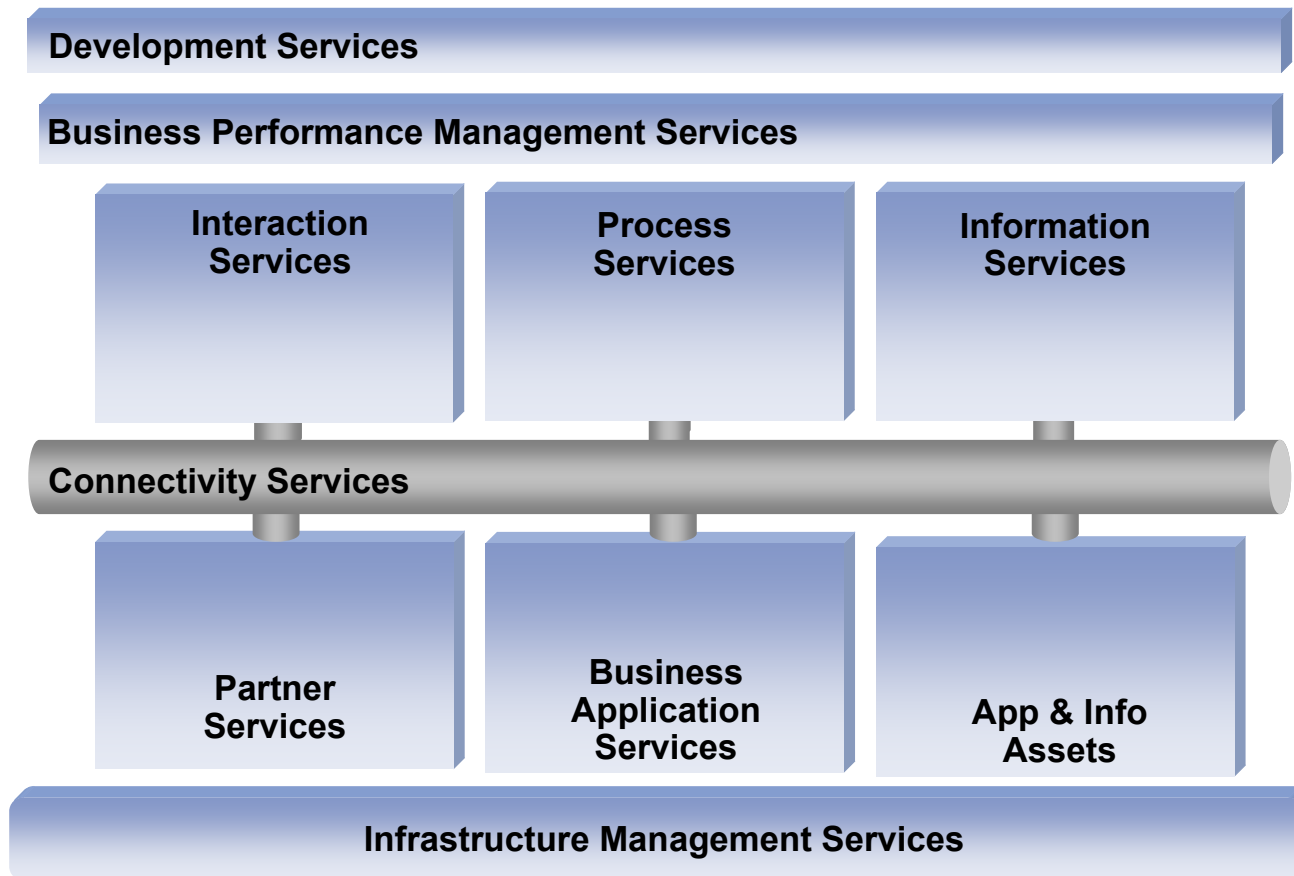


# Agenda

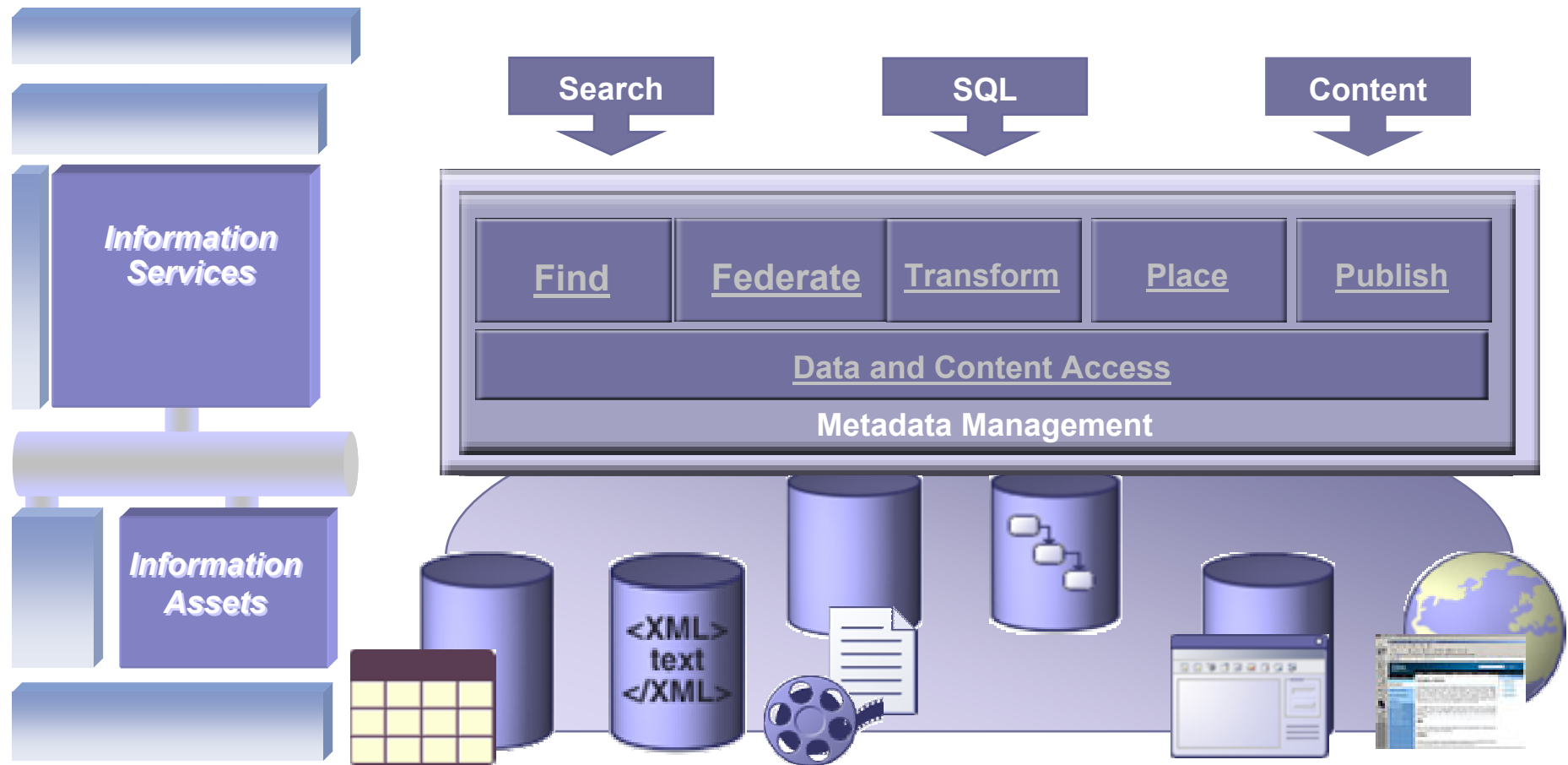
- Introduction : Pourquoi un « Service Information » dans une approche SOA
  - ▶ Le monde aujourd'hui
  - ▶ Le monde idéal
  - ▶ La réponse IBM
- L'infrastructure du « Service Information »
  - ▶ Recherche
  - ▶ Fédération
  - ▶ Transformation
  - ▶ Placement
  - ▶ Publication
- Conclusion



# Rappel : L'architecture SOA d'IBM



# Infrastructure de gestion de l'information



Conclusion





Collections | System | Security | Tools | Help

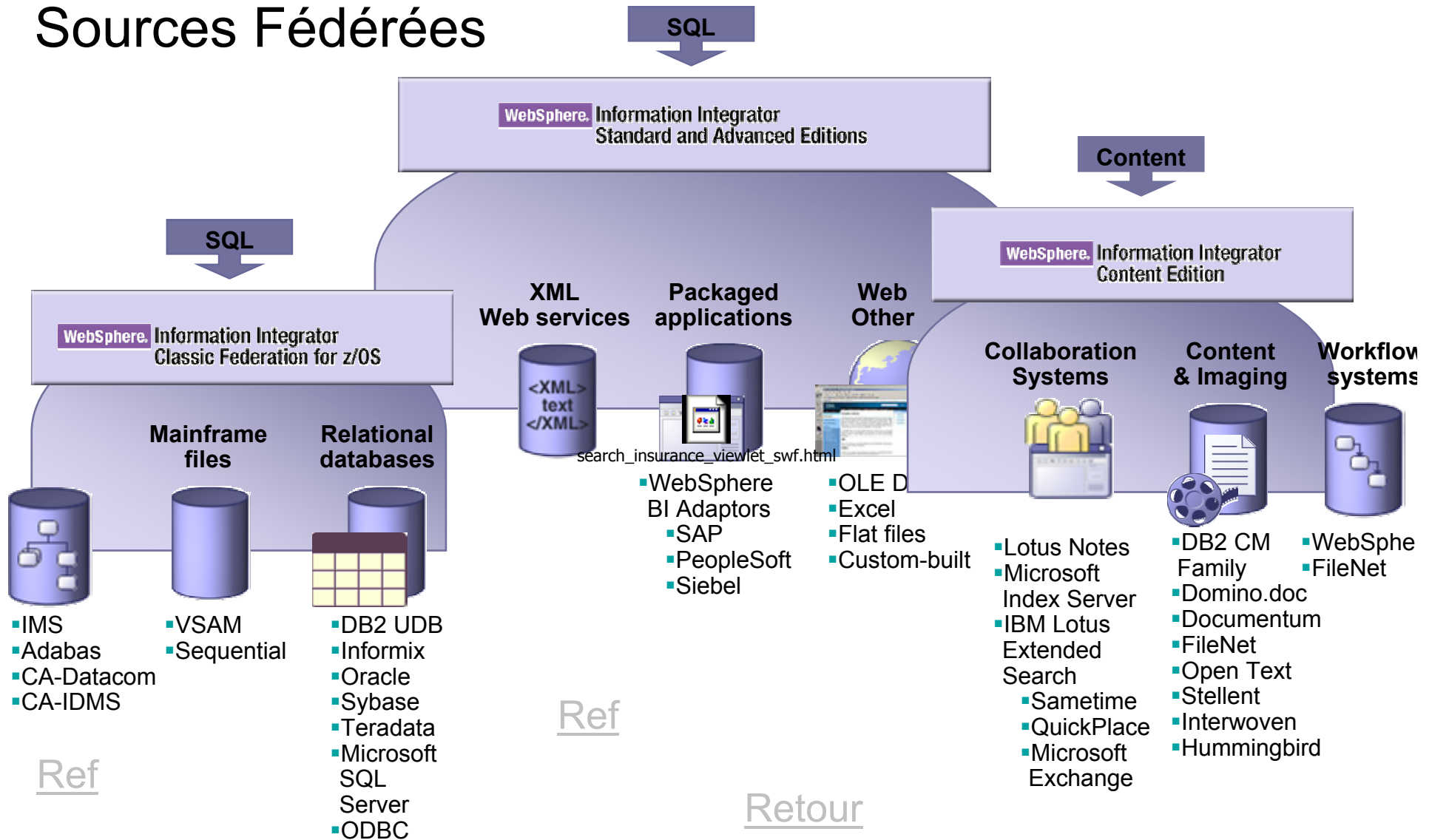
### Collections

New collection Create Wizard Last refreshed: September 30, 2003 at 7:10 AM EDT Refresh

Collection name	Search		Crawl	Parse	Index			
	1	2						
News	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>			
Finance	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>			
HR	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>			



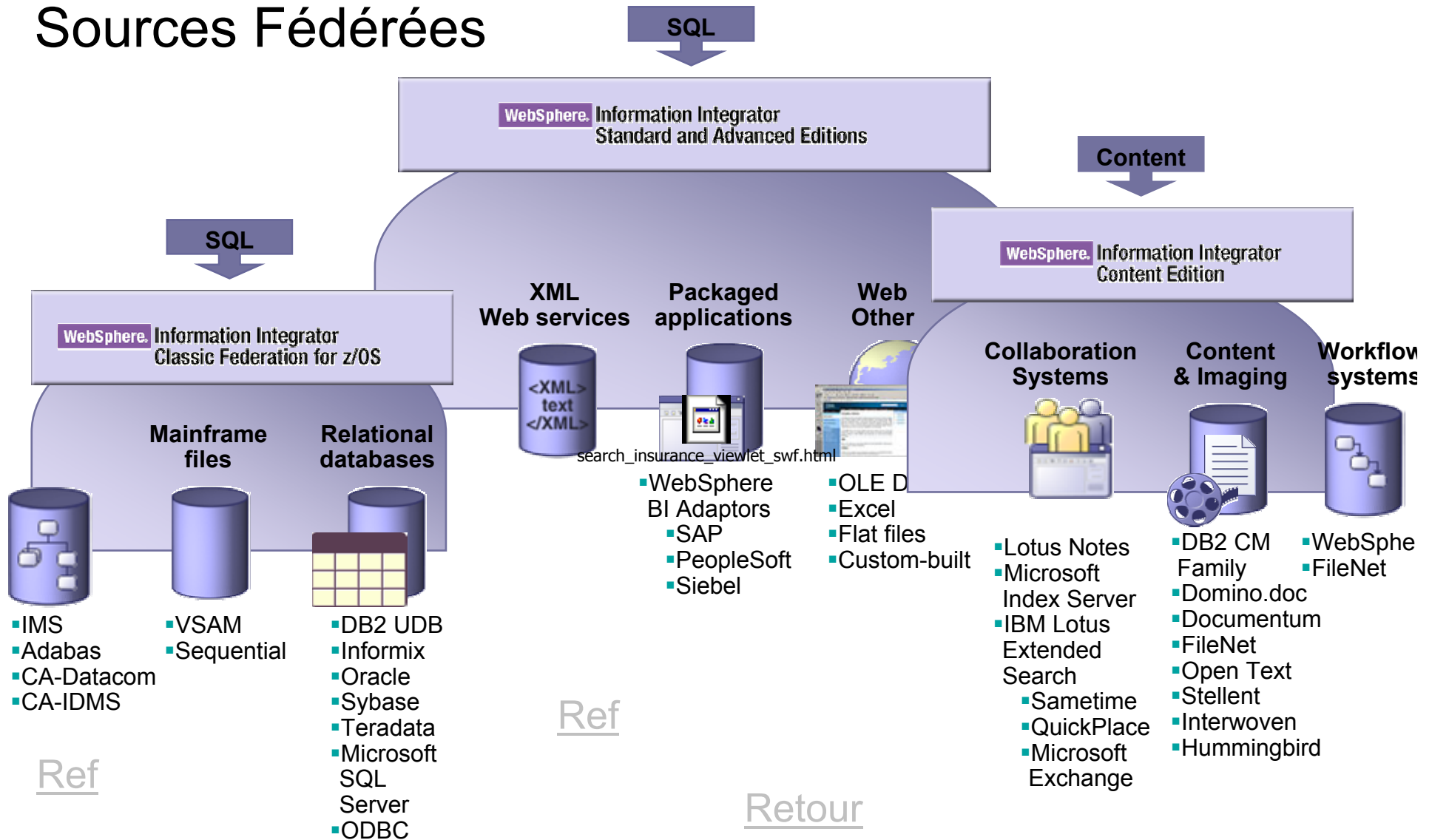
# Sources Fédérées



Plus partner tools and custom-built connectors extend access to more sources



# Sources Fédérées

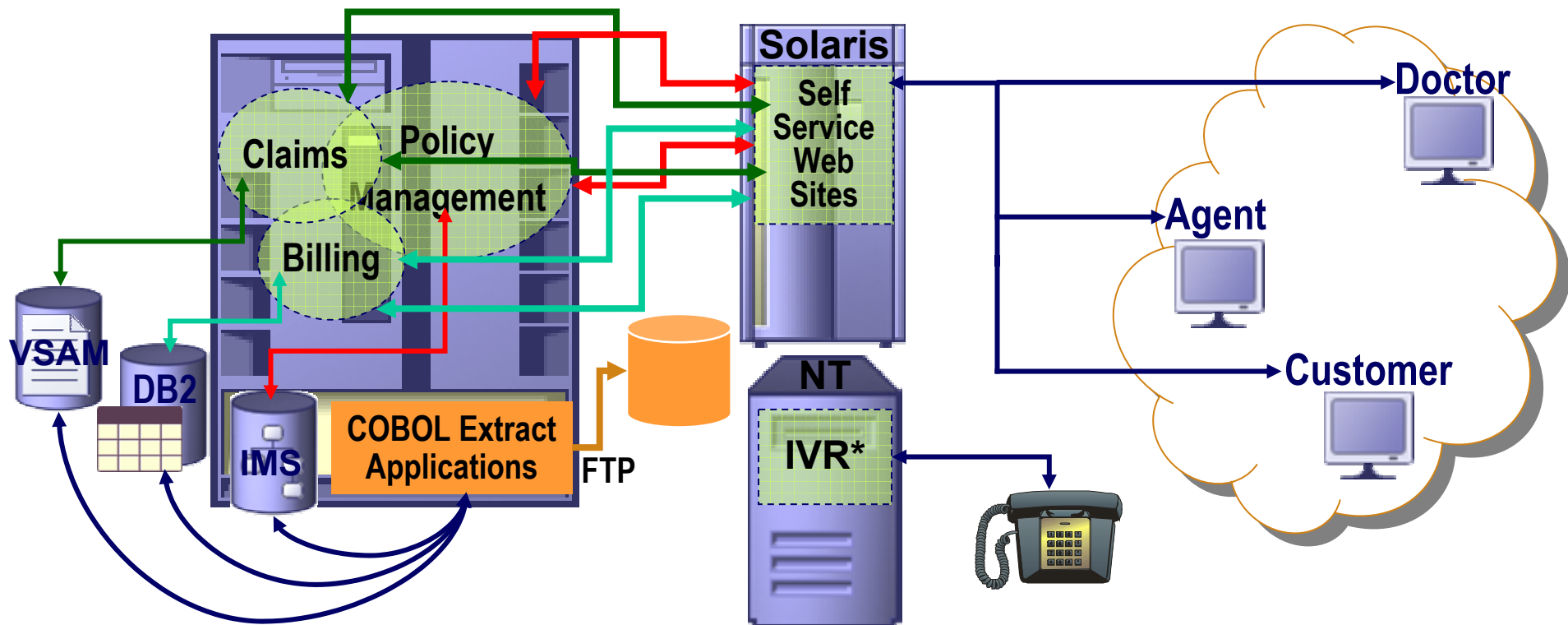


Plus partner tools and custom-built connectors extend access to more sources



# La solution Avant ....

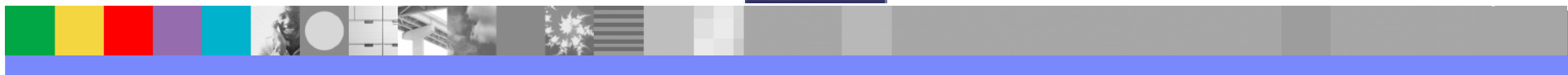
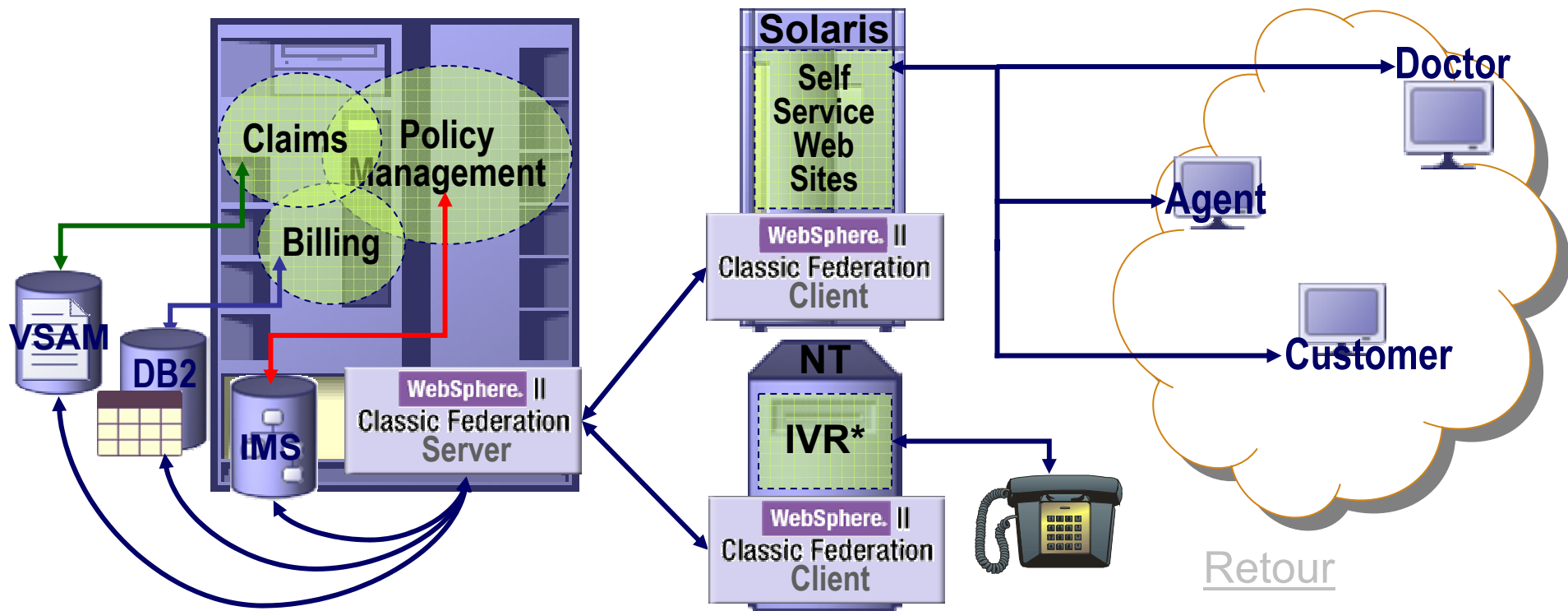
- ▶ **Solution a:** copie des données vers les environnements distribués
  - Coût estimé \$2M
  - Fraîcheur des données : rafraîchissement toutes les 30 heures
- ▶ **Solution b:** intégration des transactions IMS
  - Coût estimé 10,000 Heures/Homme par application



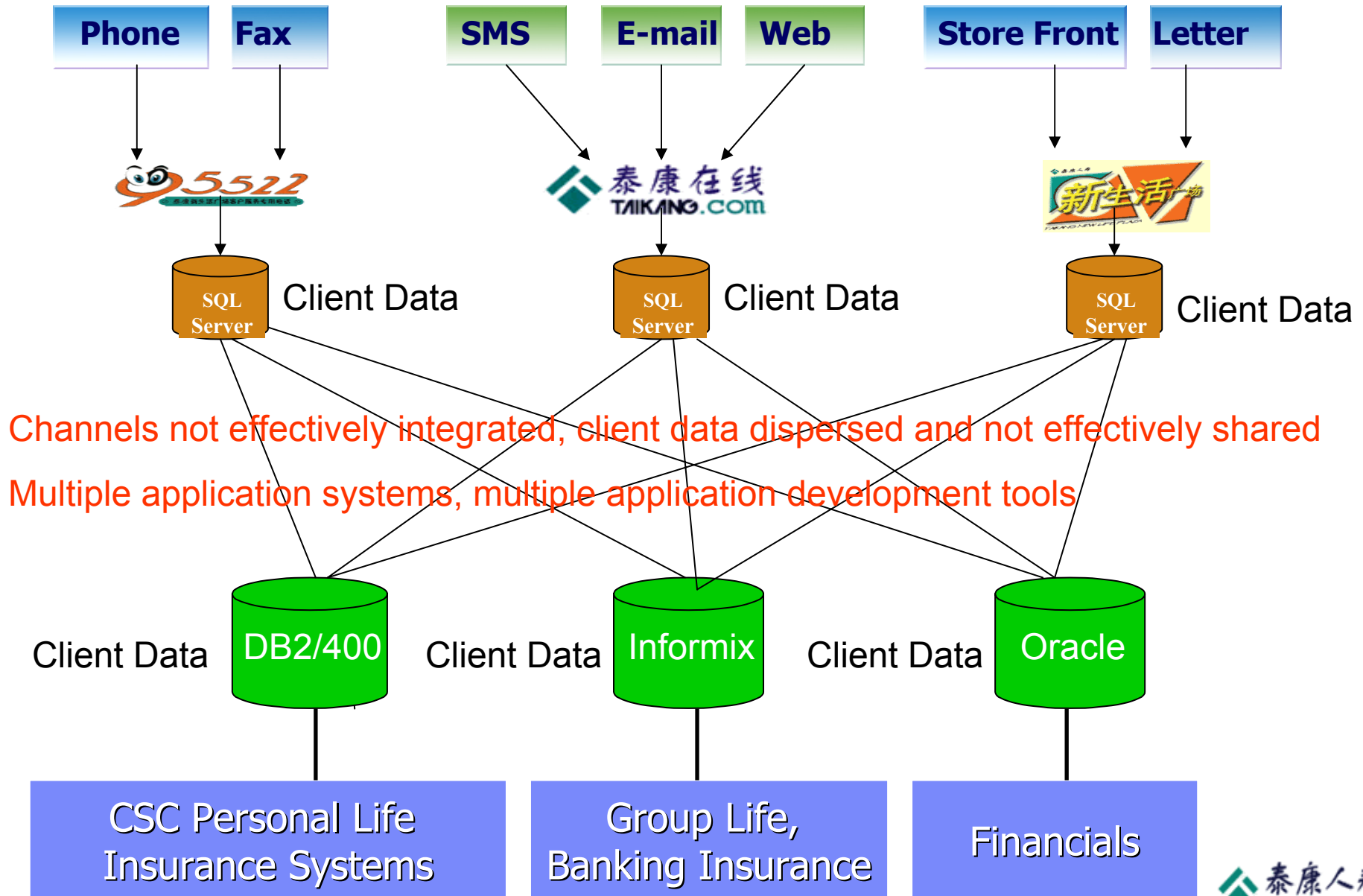
# La solution IBM solution -- *empower self-service environments*

**Provide up-to-the-minute policy, claims and accounting information**

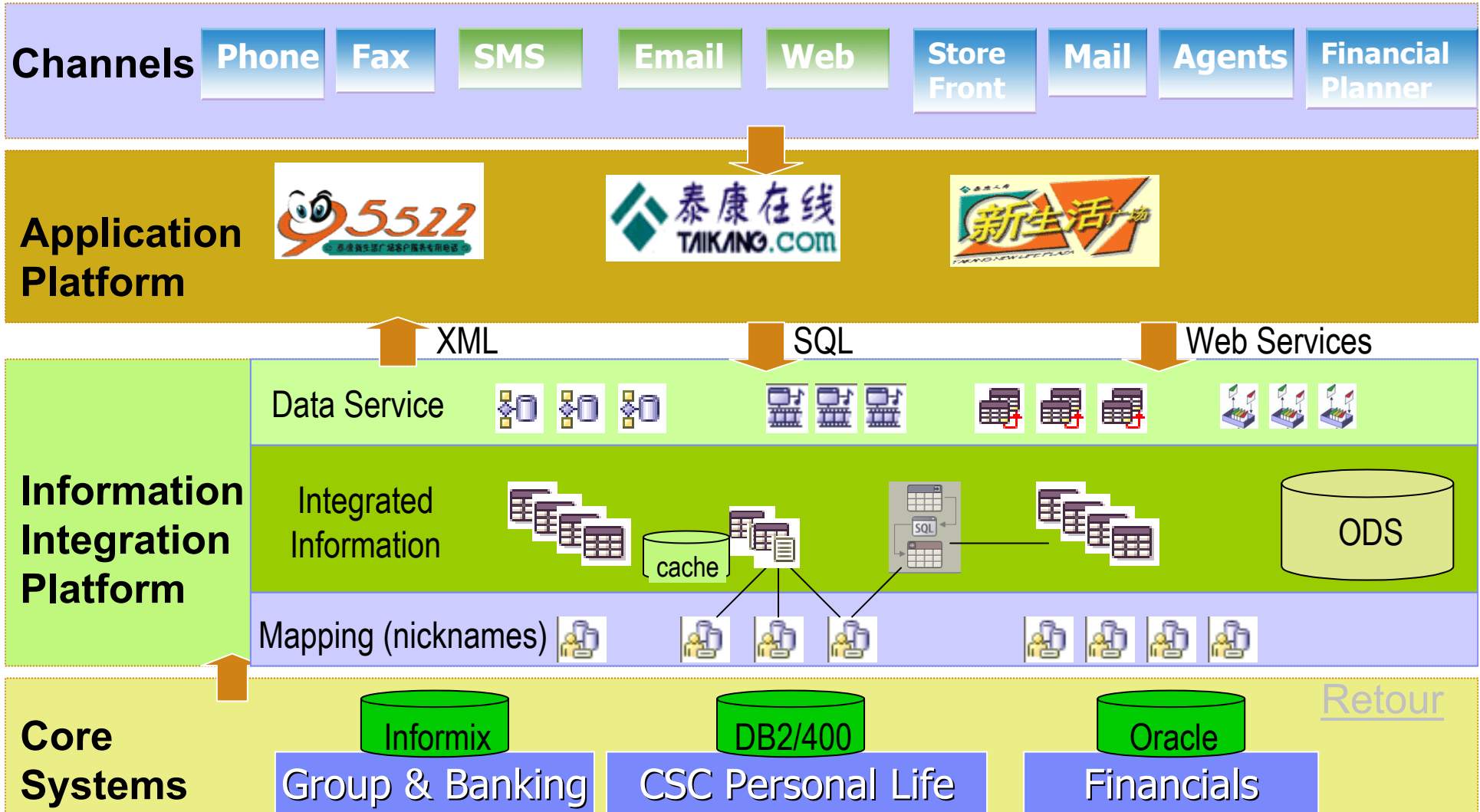
- ▶ Connect interactive voice response (IVR) system to IMS, VSAM & DB2
  - **\$250K versus \$2M**
- ▶ Connect operational data with self-service Web sites
  - **200 man-hours versus 10,000**



# Taikang Information Platform– Avant WS Information Integrator

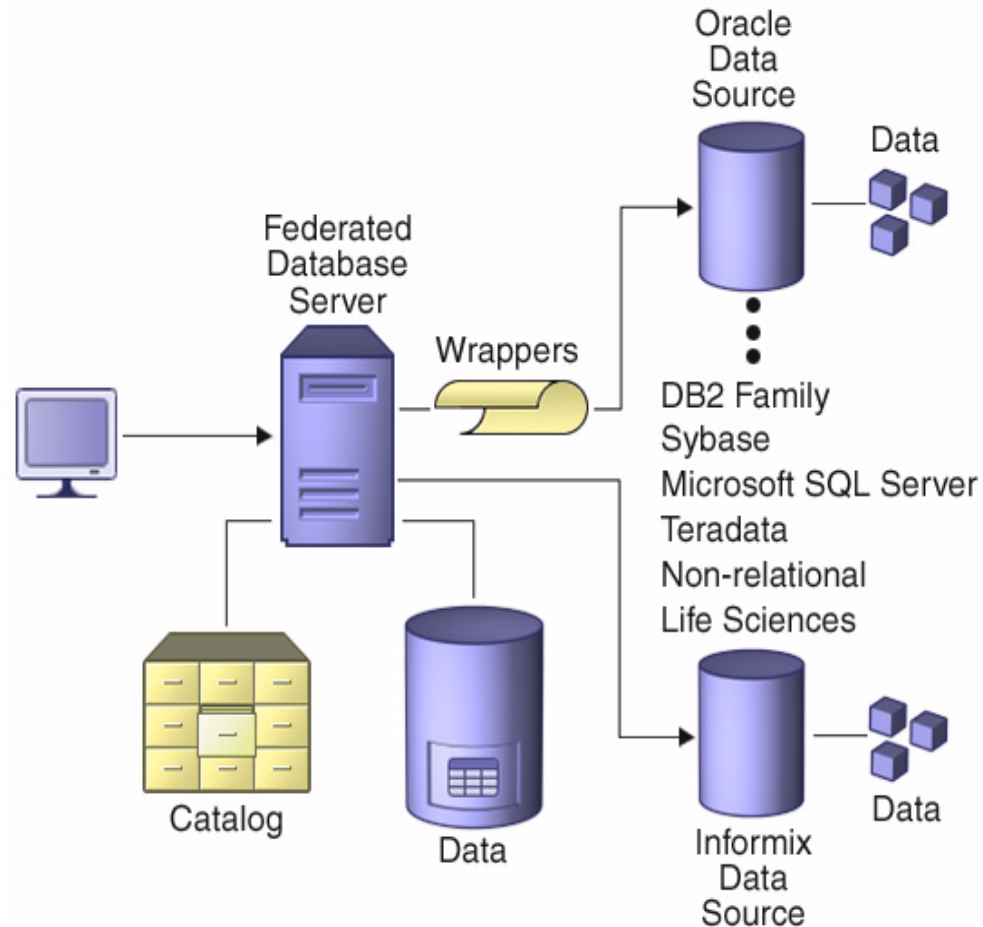


# Taikang Integrated Information Platform Architecture





# Architecture de fédération



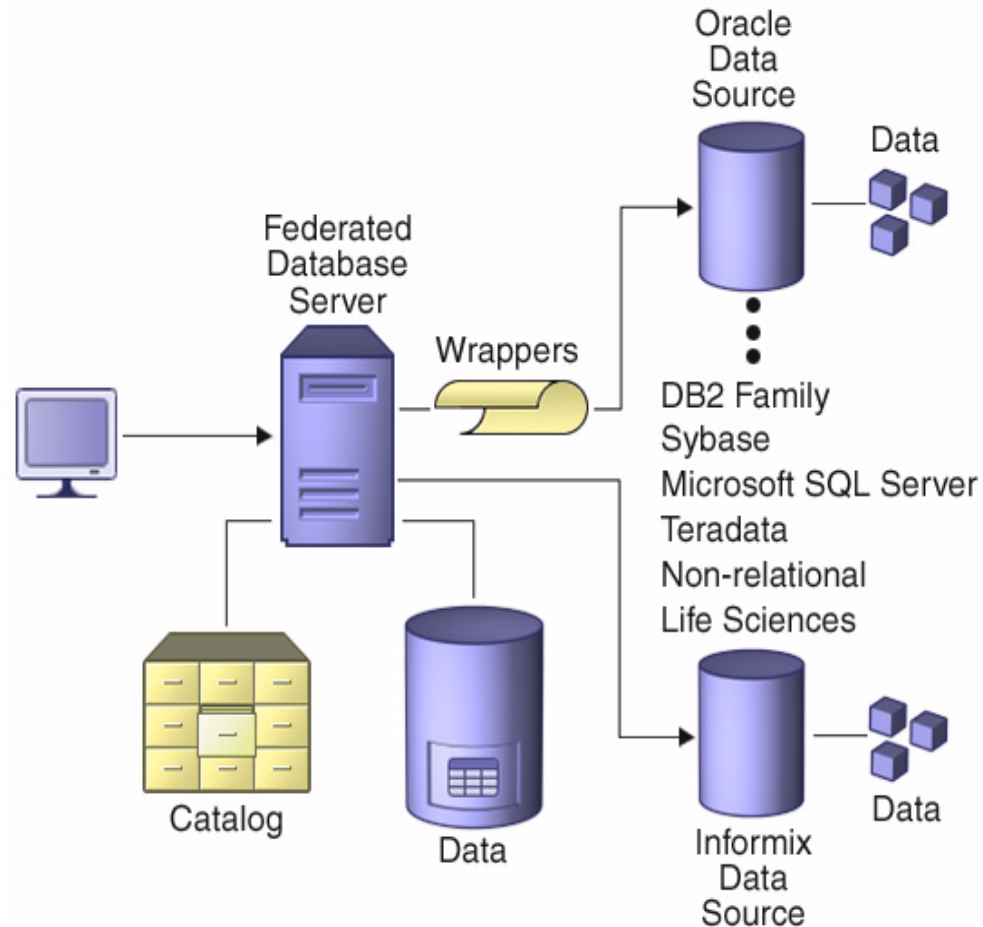
- **Transparence**
  - ▶ Le serveur est vu comme la seule source
- **Hétérogénéité**
  - ▶ Intègre les données de diverses sources : Relationnel, XML, messages, Web, fichier plat, non structuré...
- **Extensibilité**
  - ▶ [Wrapper Development Toolkit](#)
- **Large choix de fonctions**
  - ▶ Support complet des fonctions pour toutes les sources
  - ▶ Exploitation des fonctions aussi disponibles sur les sources
- **Autonomie**
  - ▶ Ne nécessite pas de modifications sur les sources, applications et systèmes.
- **Performance**
  - ▶ Optimisation des requêtes distribuées

Retour





# Architecture de fédération



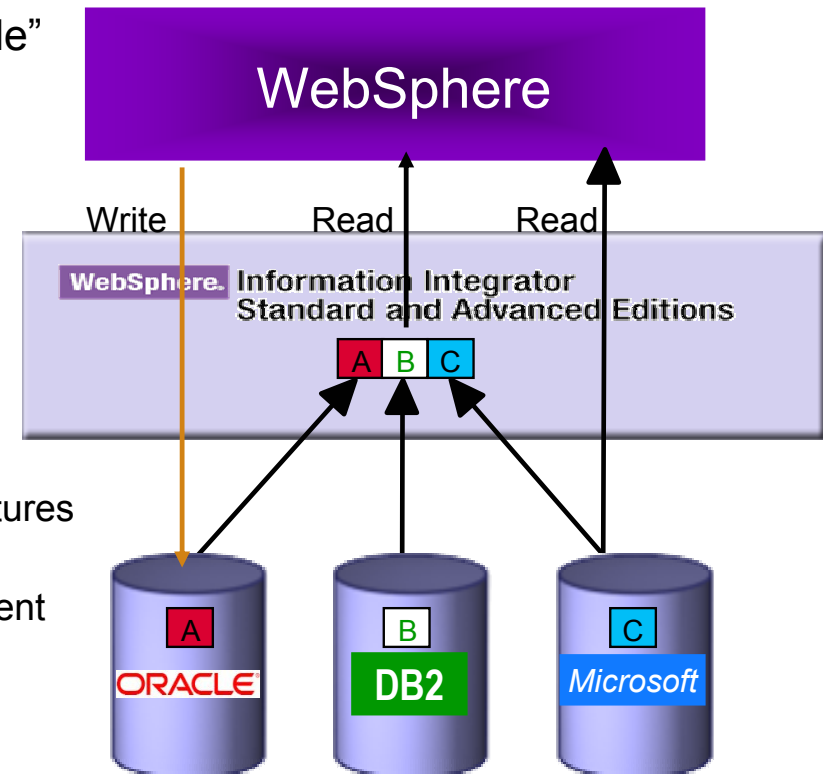
- **Transparence**
  - ▶ Le serveur est vu comme la seule source
- **Hétérogénéité**
  - ▶ Intègre les données de diverses sources : Relationnel, XML, messages, Web, fichier plat, non structuré...
- **Extensibilité**
  - ▶ [Wrapper Development Toolkit](#)
- **Large choix de fonctions**
  - ▶ Support complet des fonctions pour toutes les sources
  - ▶ Exploitation des fonctions aussi disponibles sur les sources
- **Autonomie**
  - ▶ Ne nécessite pas de modifications sur les sources, applications et systèmes.
- **Performance**
  - ▶ Optimisation des requêtes distribuées

Retour



# Placement des données : (1) Cache hétérogène

- Améliore les performances des requêtes et la disponibilité des données
- L'administrateur définit les “ Materialized Query Table”
  - Valeurs Pre-calculées ou fortement utilisées
  - N'importe quelles données issues de sources relationnelles
  - Le format du cache : relationnel ou XML
  - L'utilisation du cache peut être Implicite (routage automatique par l'optimiseur ) ou explicite
- Activation du cache par application
  - Si actif , les lectures se feront dans le cache, les écritures iront directement à la source
  - Sinon, lectures et écritures seront réalisées directement
- Le cache peut être géré :
  - Manuellement
  - Par réplication
  - Supporte de multiples stratégies de rafraîchissement



# Placement des données : (2) SQL Replication Architecture

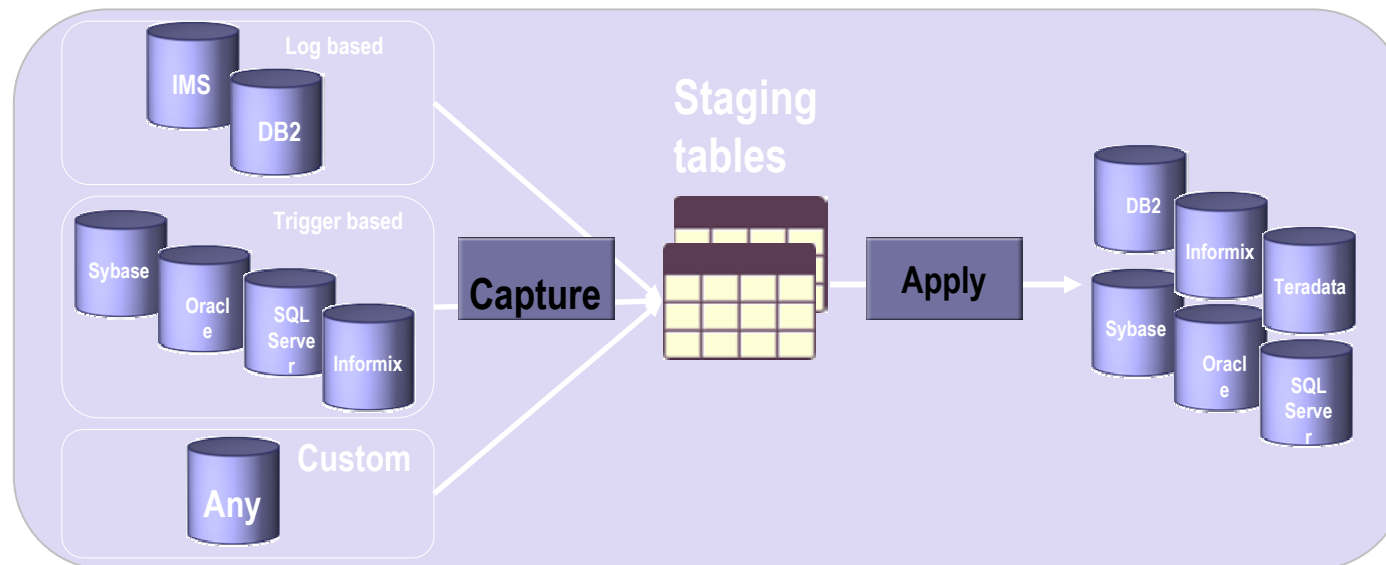
## *Flexible transformation and scheduling*

### Function

- ▶ Filter and transform, Apply by table or by transaction
- ▶ Choose latency by schedule, interval, event, or continuous
- ▶ Replicate point-to-point, for distribution, or for consolidation
- ▶ Maintain snapshots, simple copies, histories, or aggregates

### Usage

- ▶ Business intelligence
- ▶ Distribution and consolidation
- ▶ Application integration



## Placement des données : (3) Q Replication Architecture

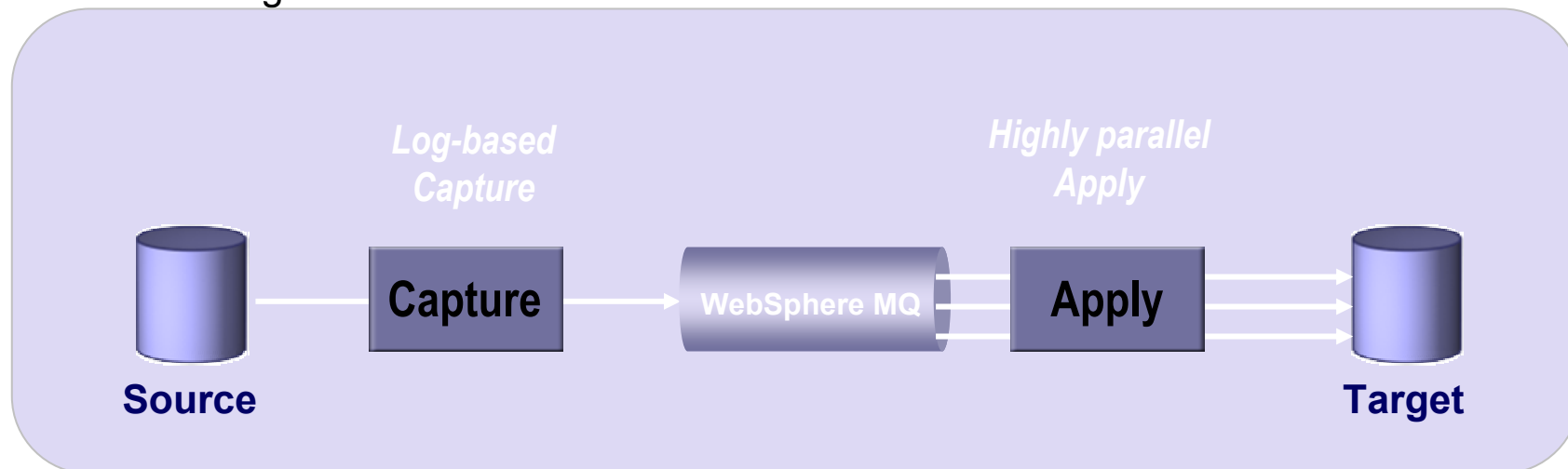
*New replication architecture for delivering extremely low latency replication for peer-to-peer environments*

### Function

- Replicate rows or transactions
- Filter and transform data
- Detect and resolve conflict
- Configure and monitor environment

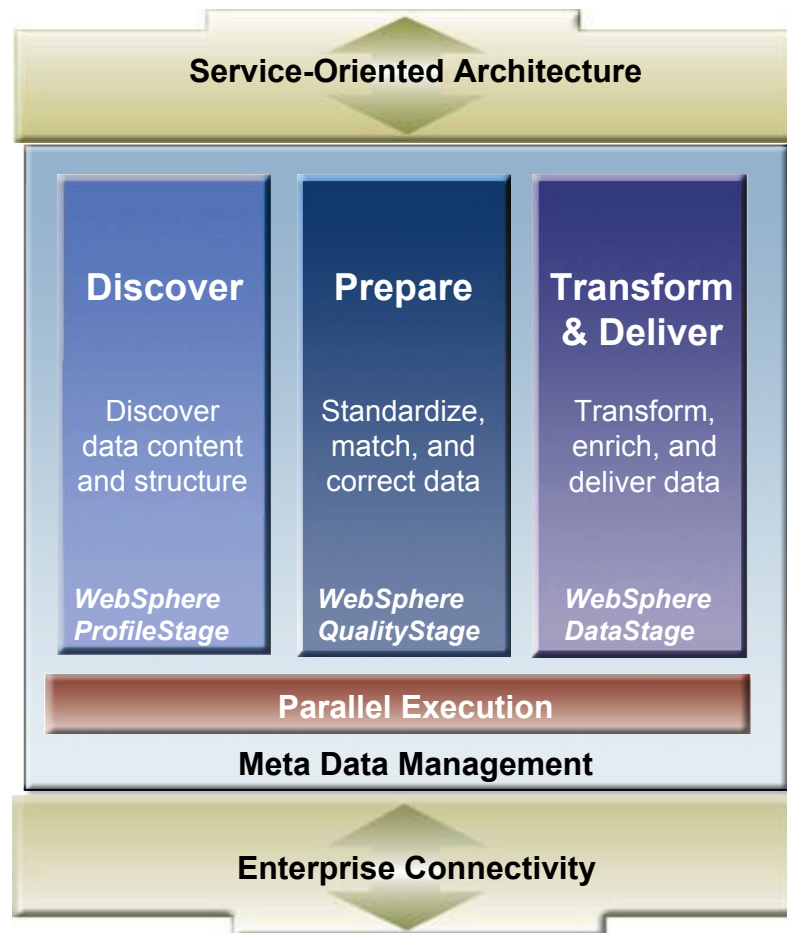
### Usage

- High availability
- Workload distribution
- Application integration



[Retour](#)

# Ascential Solution to Data Integration



## WebSphere Data Integration Suite

- Open, Service-Oriented Architecture Integrated Data Profiling & Data Quality
- Complex Data Transformation and Routing Addresses Operational, Transactional and Analytical environments
- High volume and high throughput capability
- Enterprise Meta Data Management
- Anytime, Anywhere Connectivity
- Supports major Industry Standards and platforms
- Additional products:  
WebSphere DataStage TX for message-based integration

Retour



# Event Publishing

## *Capture database changes as XML messages and publish them to WebSphere MQ*

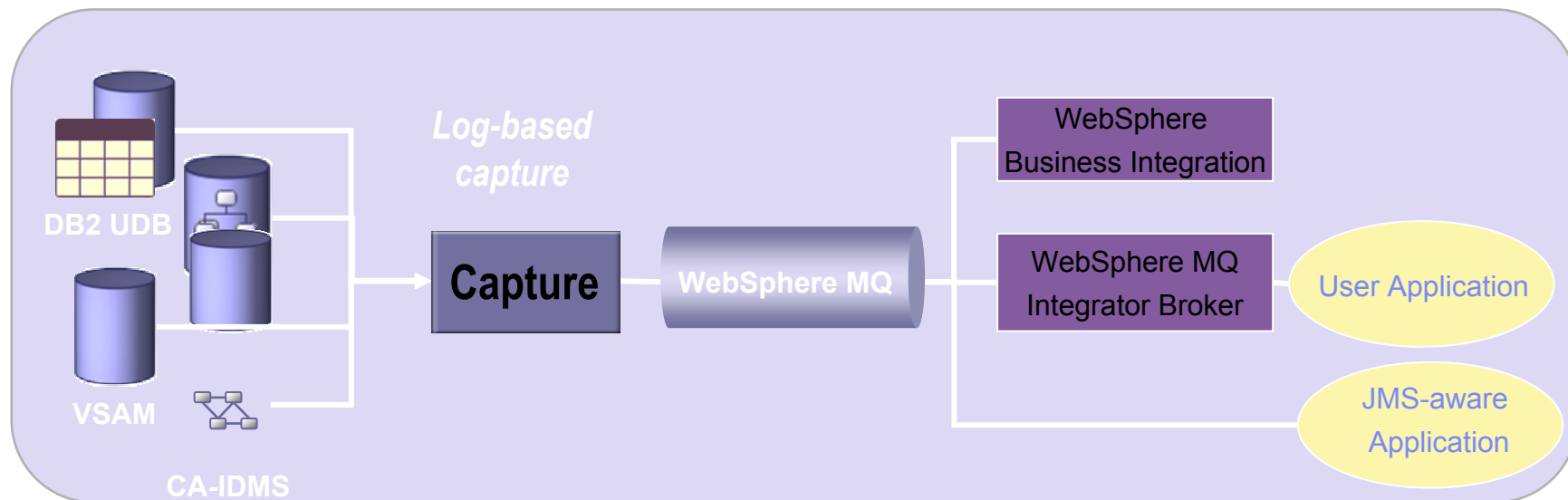
### Function

- Publish events to a message queue
- XML self-describing format
- Wizard-driven configuration

### Usage

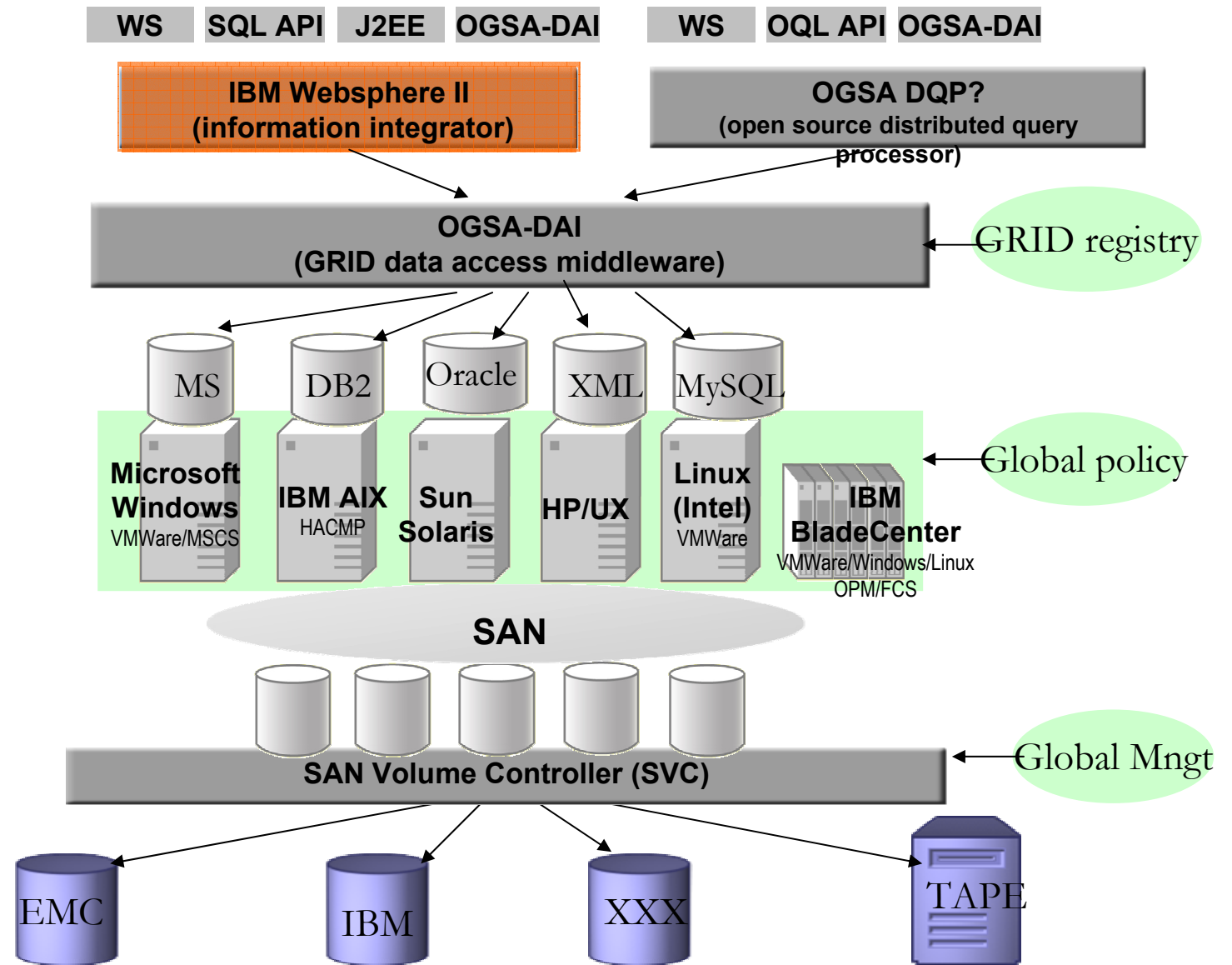
- Application to application messaging
- Initiate business processes
- Source for ETL tool

[Retour](#)



# Virtualization of data on the Grid

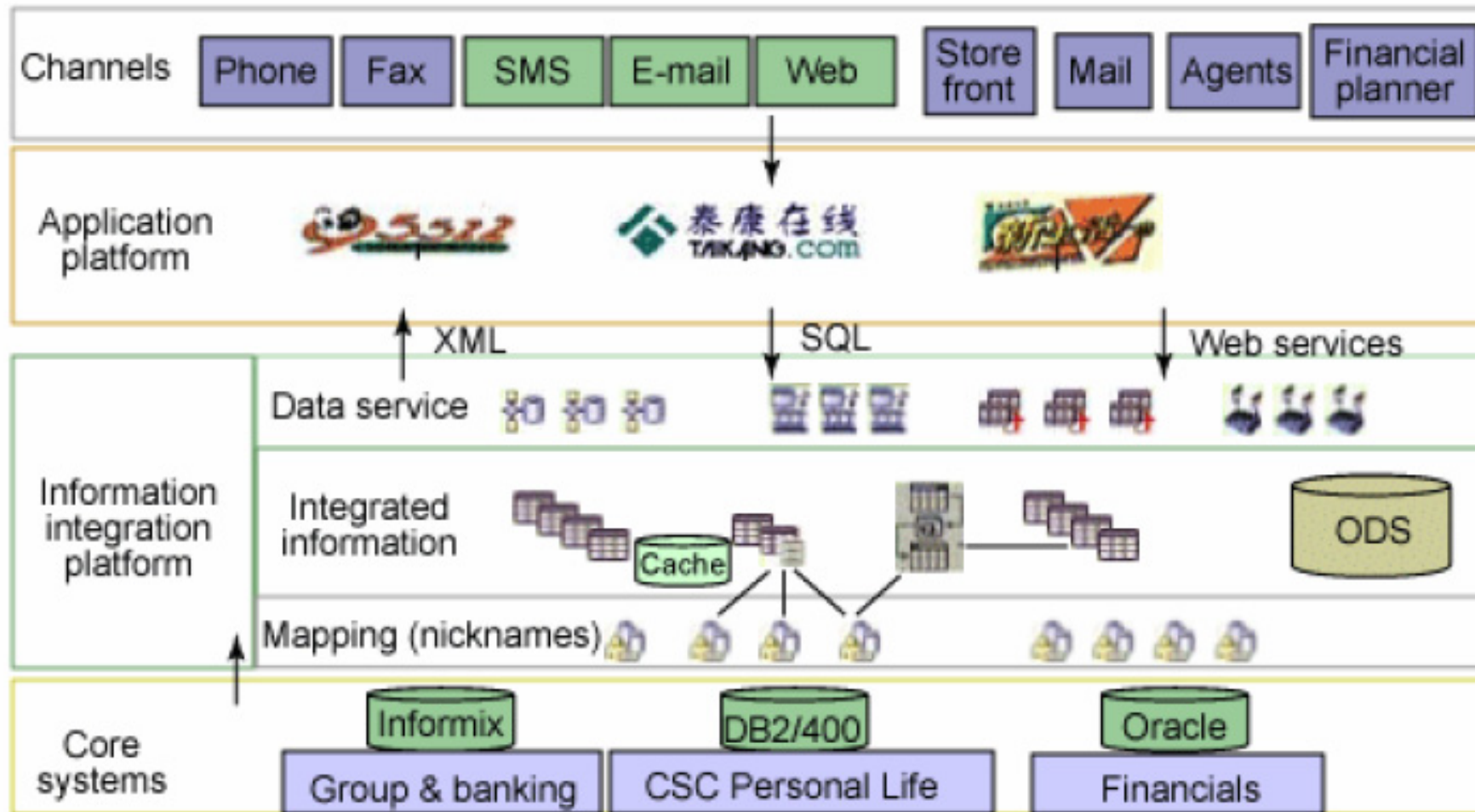
- Data Virtualization
- Schema integration Federation
- DataSource access Over SOA
- File system access (SAN FS)
- Disk Access



Retour



# En synthèse :





# Agenda

- Introduction : Pourquoi un « Service Information » dans une approche SOA
  - ▶ Le monde aujourd'hui
  - ▶ Le monde idéal
  - ▶ La réponse IBM
- L'infrastructure du « Service Information »
  - ▶ Recherche
  - ▶ Fédération
  - ▶ Transformation
  - ▶ Placement
  - ▶ Publication
- Conclusion



# Conclusion

- Dans une logique d'urbanisation, 2 questions doivent être posées :
  - ▶ Faut-il encore systématiquement dupliquer toutes les informations ?
  - ▶ L'accélération des échanges entre les systèmes d'information peut-elle encore longtemps faire l'impasse d'un contrôle des informations échangées ?
  
- Les technologies sollicitées dans cette logique de « Service Information » ne sont :
  - ▶ Ni triviales,
  - ▶ Ni systématiquement originales,
  
- Le « quasi temps-réel » rapproche les mondes et les hommes du transactionnel et du décisionnel. Ne nous trompons pas de challenge : le savoir-faire du monde décisionnel doit nous permettre de relever le défi de la flexibilité sans dégrader la qualité des informations manipulées
  
- Est-il pertinent de les regrouper ainsi pour identifier et « légitimer » un véritable « Service Information »

