Session B11

Information Integration Patterns

Gil Lee
Information Integration

IBM Data Management Technical Conference

Anaheim, CA   **Sept 9 - 13, 2002**

B11   **Information Integration Scenarios**
Gil Lee, Information Integration Marketing Manager, IBM

Companies see integration as their number 1 strategic software initiative according to a recent survey of CIOs. This session starts with an introduction to business integration and the role of information integration. A discussion follows on customer business problems that require integration. Next is a review of common application patterns which can be used to implement information integration solutions including customer examples. Concluding the session is an outlook on where IBM is headed with information integration.
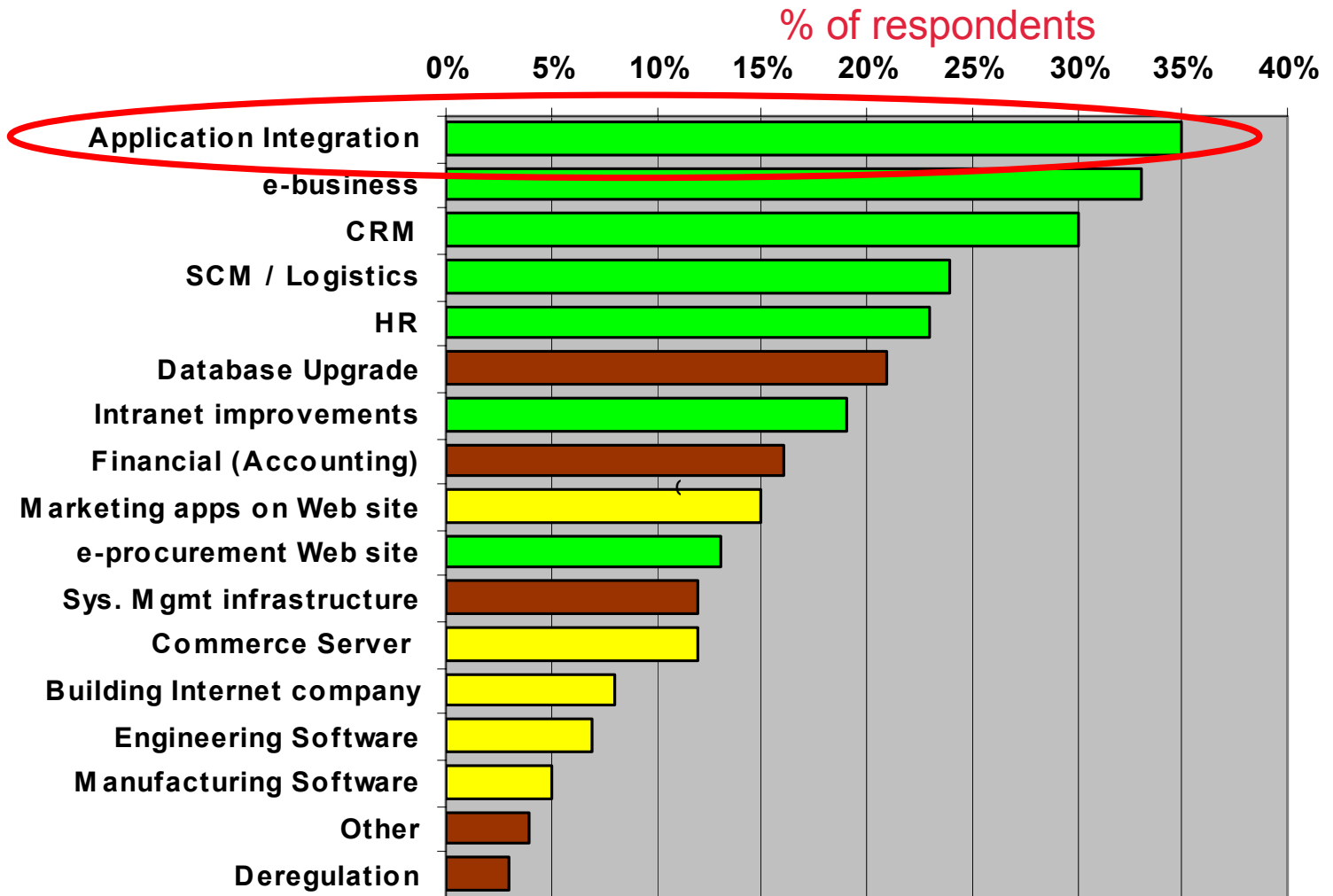
Agenda

- Market view of integration
- Customer requirements
- Integration patterns
- Future
- Q&A

What the CIO's are saying

Top strategic software platform project over the next year

% of respondents

Category	
Application Integration	~35%
e-business	~33%
CRM	~30%
SCM / Logistics	~24%
HR	~23%
Database Upgrade	~21%
Intranet improvements	~19%
Financial (Accounting)	~16%
Marketing apps on Web site	~15%
e-procurement Web site	~13%
Sys. Mgmt infrastructure	~12%
Commerce Server	~12%
Building Internet company	~8%
Engineering Software	~7%
Manufacturing Software	~5%
Other	~4%
Deregulation	~3%

Axis: 0% 5% 10% 15% 20% 25% 30% 35% 40%

Source: CIO Magazine Survey, May 2001
Note: Multiple responses permitted

What the Analysts are Saying

The rapidly changing economic climate is driving the need for improved access to information, flexible analytical capabilities, and formal information inventories.

Enterprises lacking integrated analytical data stores (or heterogeneous data access) will have difficultly altering strategies without potentially creating corporate imbalance.
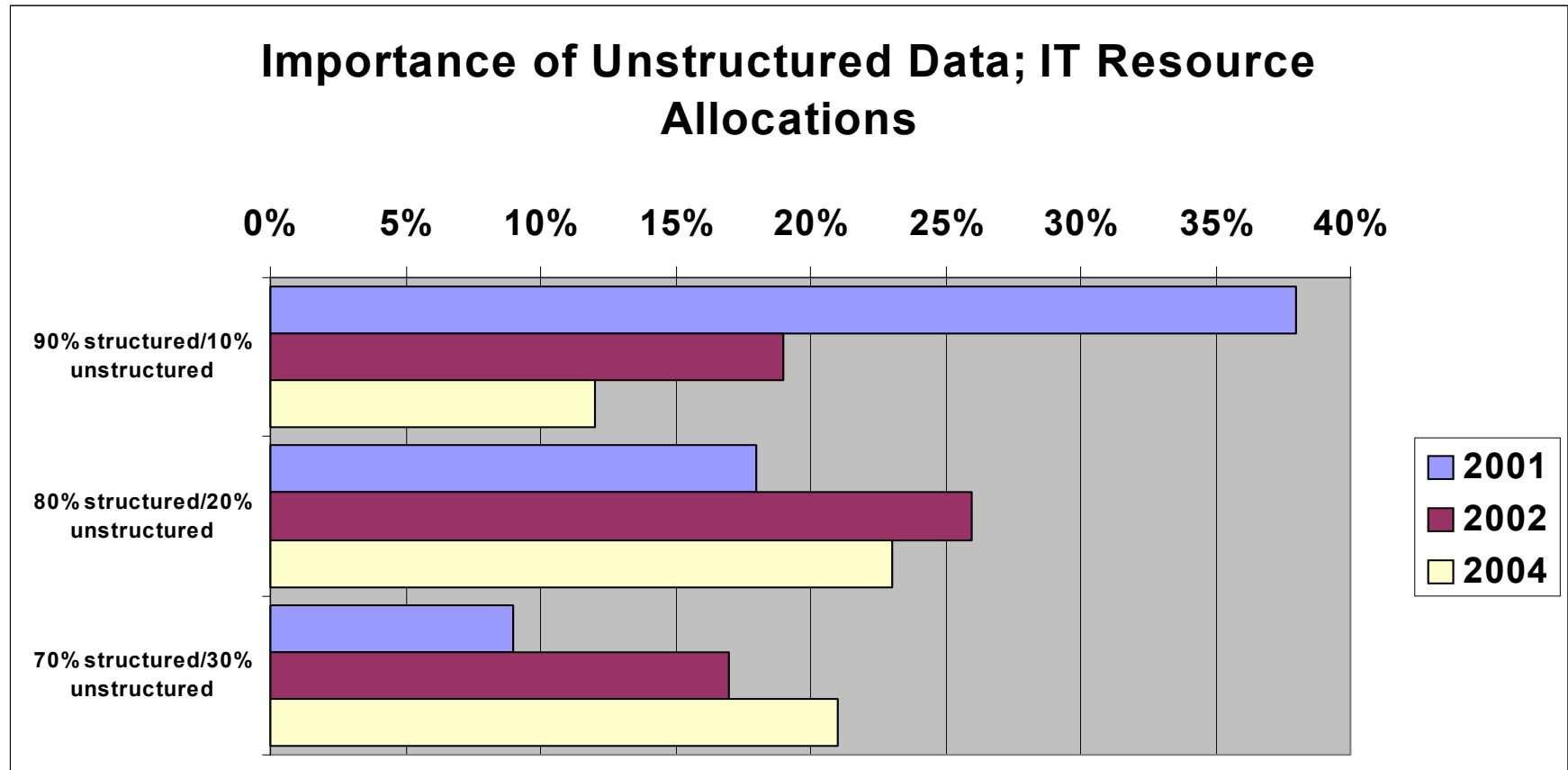
Furthermore, successfully analyzing indeterminate business events requires analytic horsepower and functionality beyond most so-called "business intelligence" tools.

During 2002/03, enterprises with experience and technology to monitor business events for altered/emerging trends, as well as the management agility to exploit them, will become industry leaders. An enterprise's capacity to express its information value (and externalize it) will improve its attractiveness to partners and suitors.

Bottom Line: Organizations must improve their ability to assemble, analyze, and account for their information assets.

Meta Group Metafax 9/28/2001

What the Analysts are Saying

Importance of Unstructured Data; IT Resource Allocations

Legend: 2001, 2002, 2004

Categories:
- 90% structured/10% unstructured
- 80% structured/20% unstructured
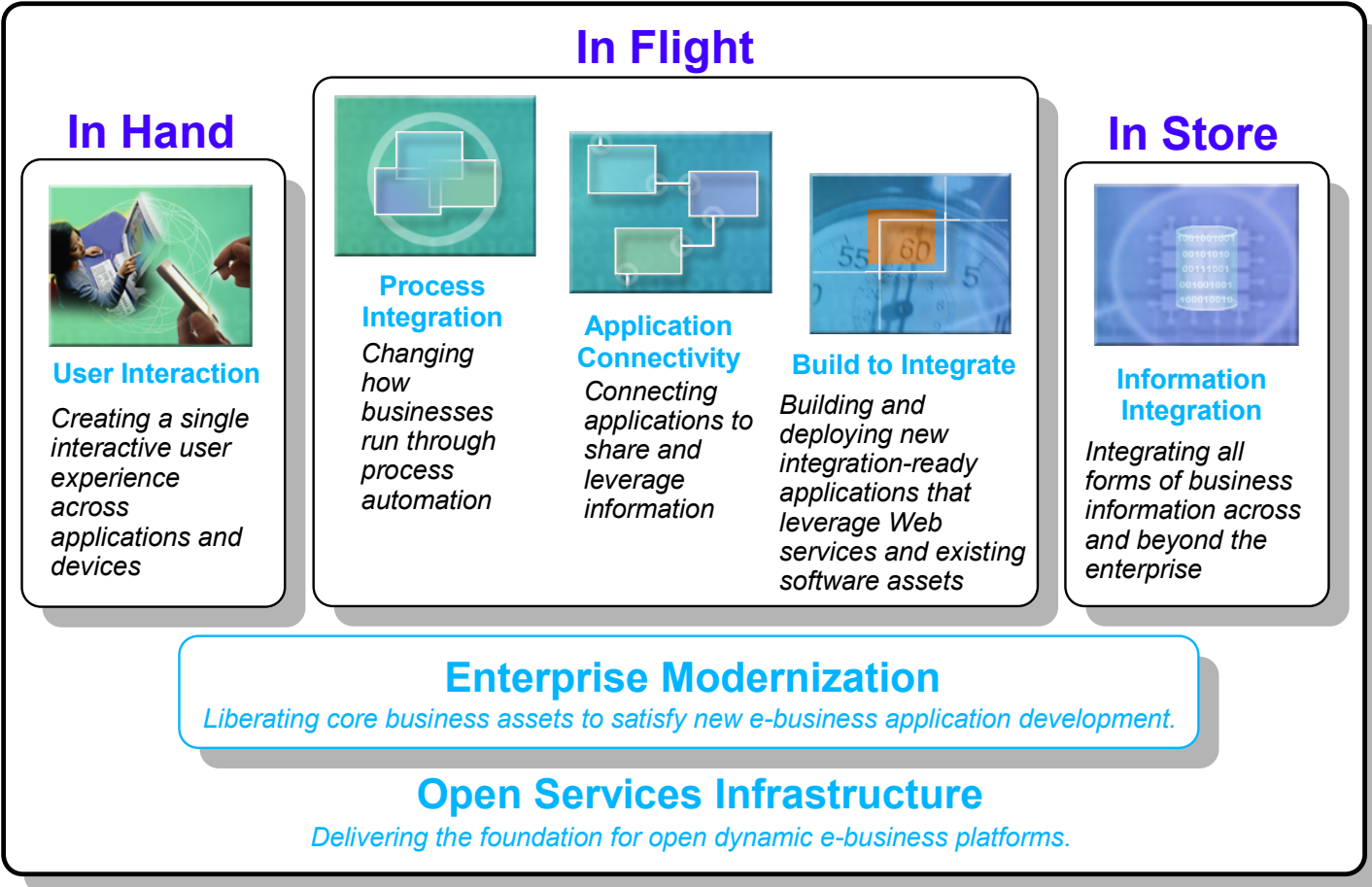- 70% structured/30% unstructured

"We are in the midst of profound changes in thinking about data management. The trend is for fewer organizations to have such an extreme disparity in resource allocation. Management of unstructured data is moving out of the niche role it occupies today and will command a larger share of most IT departments' time, energy, and money."
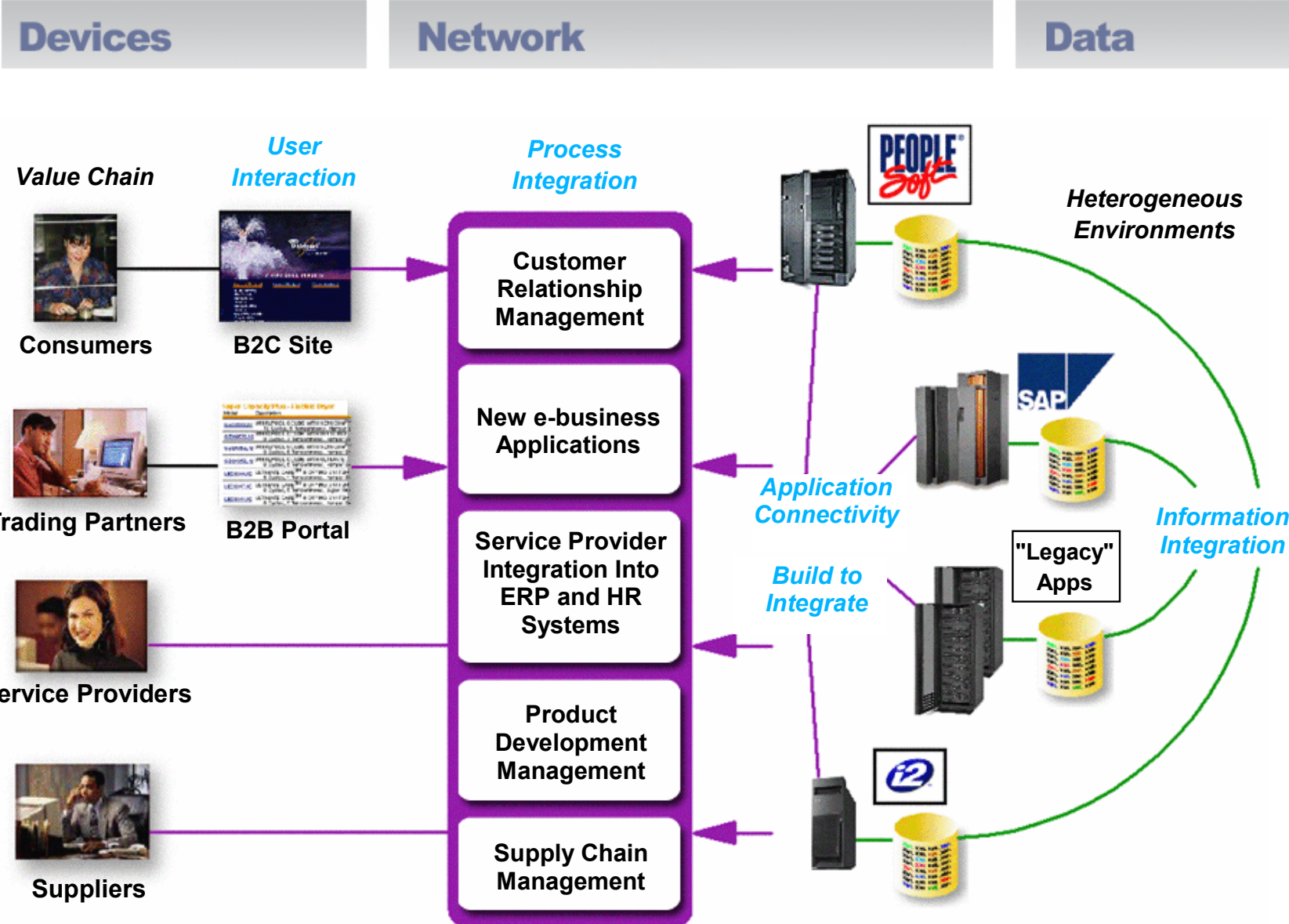
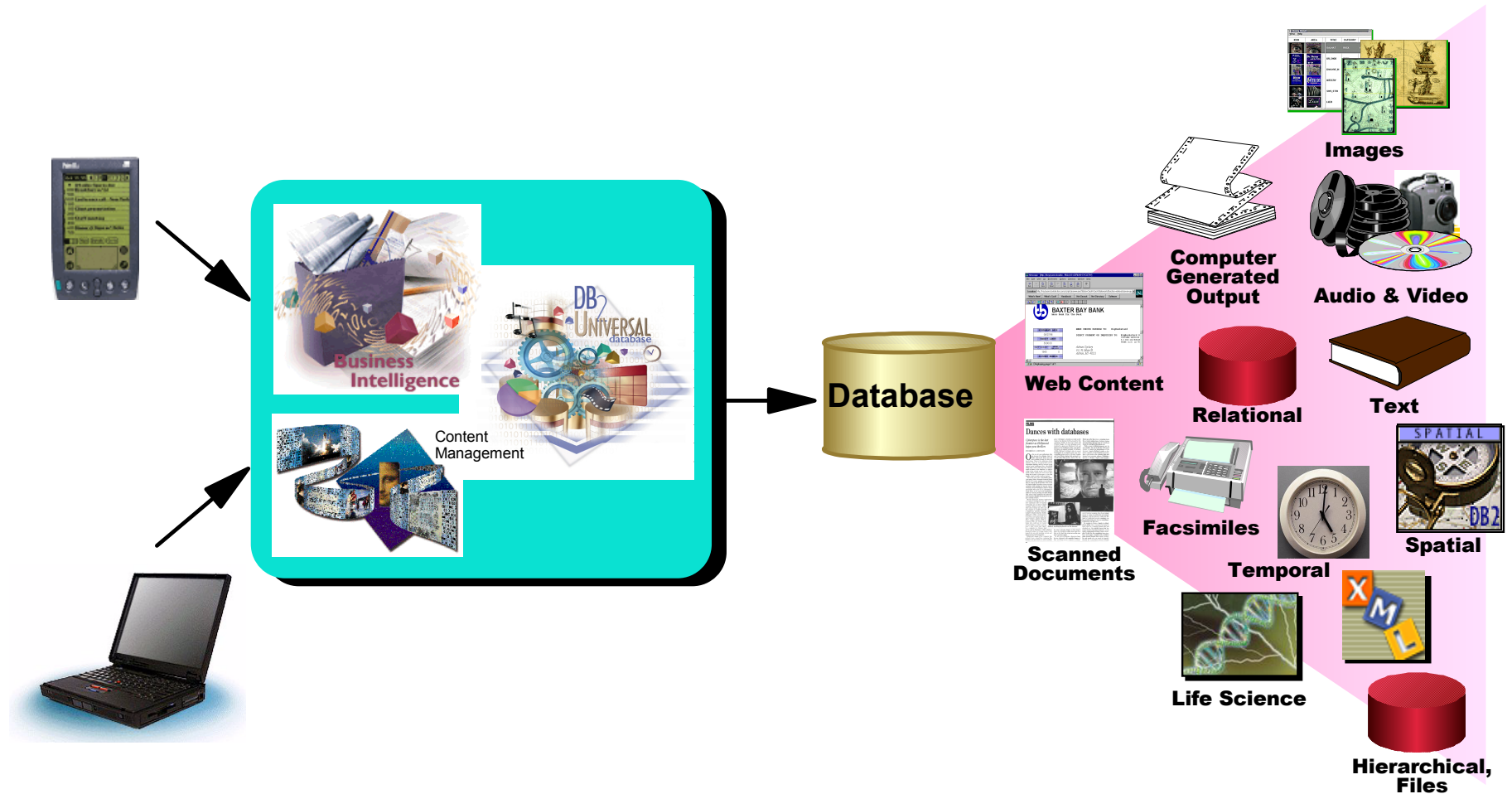-- Grant Laing, Sr Analyst, Intellor Group

The Integration Story...

Integration is a multidimensional business issue....

In Flight

In Hand

User Interaction

Creating a single interactive user experience across applications and devices

Process Integration

Changing how businesses run through process automation

Application Connectivity

Connecting applications to share and leverage information

Build to Integrate

Building and deploying new integration-ready applications that leverage Web services and existing software assets

In Store

Information Integration

Integrating all forms of business information across and beyond the enterprise

Enterprise Modernization

Liberating core business assets to satisfy new e-business application development.

Open Services Infrastructure

Delivering the foundation for open dynamic e-business platforms.

Different Types of Integration Required

Devices | Network | Data

Value Chain

User Interaction

Process Integration

Heterogeneous Environments

Consumers — B2C Site → **Customer Relationship Management** ← PEOPLESoft

Trading Partners — B2B Portal → **New e-business Applications** ← SAP

Application Connectivity

Service Providers → **Service Provider Integration Into ERP and HR Systems** ← *Build to Integrate* — "Legacy" Apps

Information Integration

Product Development Management

Suppliers → **Supply Chain Management** ← i2

Information Integration
Requirements

Images

Computer Generated Output

Audio & Video

Web Content

Relational

Text

Database

Facsimiles

Spatial

Scanned Documents

Temporal

Life Science

Hierarchical, Files

Business Intelligence

Content Management

IBM

Scenario: B2B and B2C

- Challenge
  - Become the partner of choice for deliveries resulting from B2B and B2C transactions.

- Solution
  - Extract shipping data directly from their customers orders and shipping databases.
  - Quickly return delivery status information to customers
  - Use XML and WebServices for process integration and messaging across enterprises.

- Value
  - Improved customer satisfaction by shortening the delivery process
  - Increase cash flow and that of its customers
  - Reduce operational cost.

- Technical Requirements
  - Integration/access to structured and unstructured data (including application data, e.g., SAP)
  - Application and process integration (messaging, workflow)
  - Repository to store XML messages

Scenario: Analytics

- Challenge
  - Enhance effectiveness of telesales

- Solution
  - Record sales conversation and store in text format
  - Combine data and text mining techniques over sales transaction data and sales conversation data.

- Value
  - Improve sales offerings and strategies by analyzing patterns that would not be visible had the structured and unstructured data remained isolated.

- Technical Requirements
  - Integration of structured and unstructured data
  - Application of mining techniques over the integrated data

Scenario: Layered Architecture

- Challenge
  - One comprehensive set of Business Services that can be driven from internal applications and from interfaces that operate within the portals and web pages of customers and agents
  - Improved integration and reduced coupling between Business Services and back end systems
- Solution
  - XML based presentation services
  - Layered services based architecture providing abstraction at each layer
    - Business logic, presentation services, Integration services for data access, workflow and messaging
- Value
  - Improved customer service
  - Reduced costs based on reuse of Business Services
  - Lower maintenance costs and improved flexibility
- Technical Requirements
  - Access to structured and unstructured data and core insurance applications
  - Interfaces to workflow, messaging and replication

What is the Trend?

- **There is a significant acceleration of the need to tightly integrate business processes and the underlying data and applications to serve needs around**
  - ► faster response times
  - ► improved business effectiveness
  - ► new business models.

- **This is not done purely around relational data**

- **This is being done**
  - ► around XML and the web
  - ► with focus on integration of structured and unstructured data
  - ► support for analytics (sophisticated search, mining, etc.)
  - ► infrastructure for enterprise application integration (messaging, workflow, etc.)

Information Integration
Patterns

Images

Computer Generated Output

Audio & Video

Web Content

Relational

Text

Database

Facsimiles

Temporal

Spatial

Scanned Documents

Life Science

Hierarchical, Files

Business Intelligence

Content Management

DB2 Universal database

IBM®
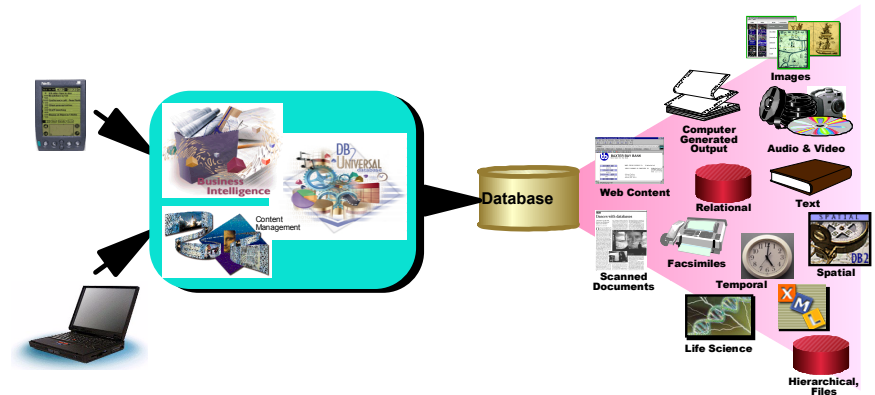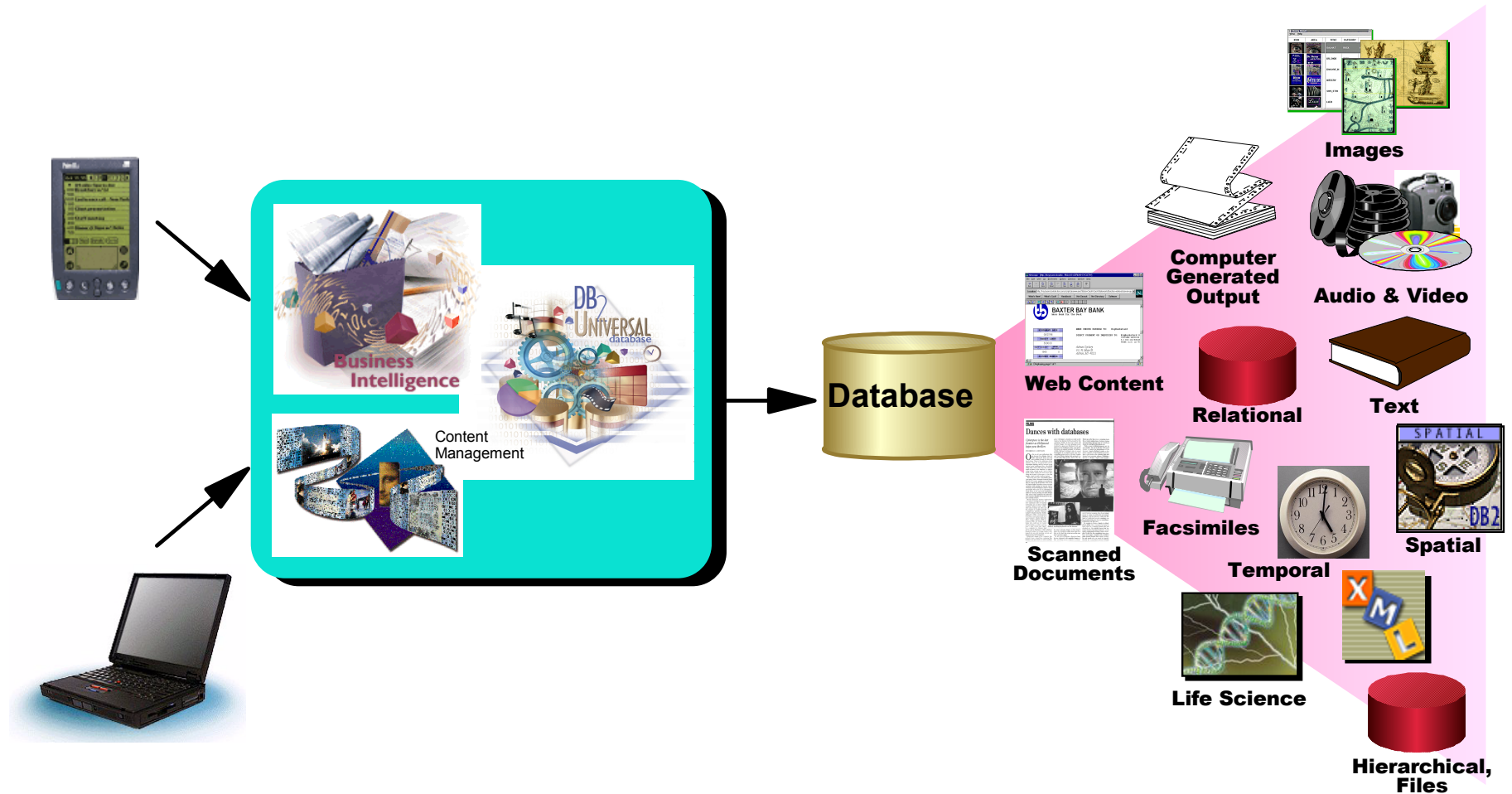
Information integration consists of seven high-level sets of function

Advanced search and access

ETML function

Replication function

Metadata

Federation engine

Development environment

Wrapper   **Wrapper**   **Wrapper**   **Wrapper**

IBM ®
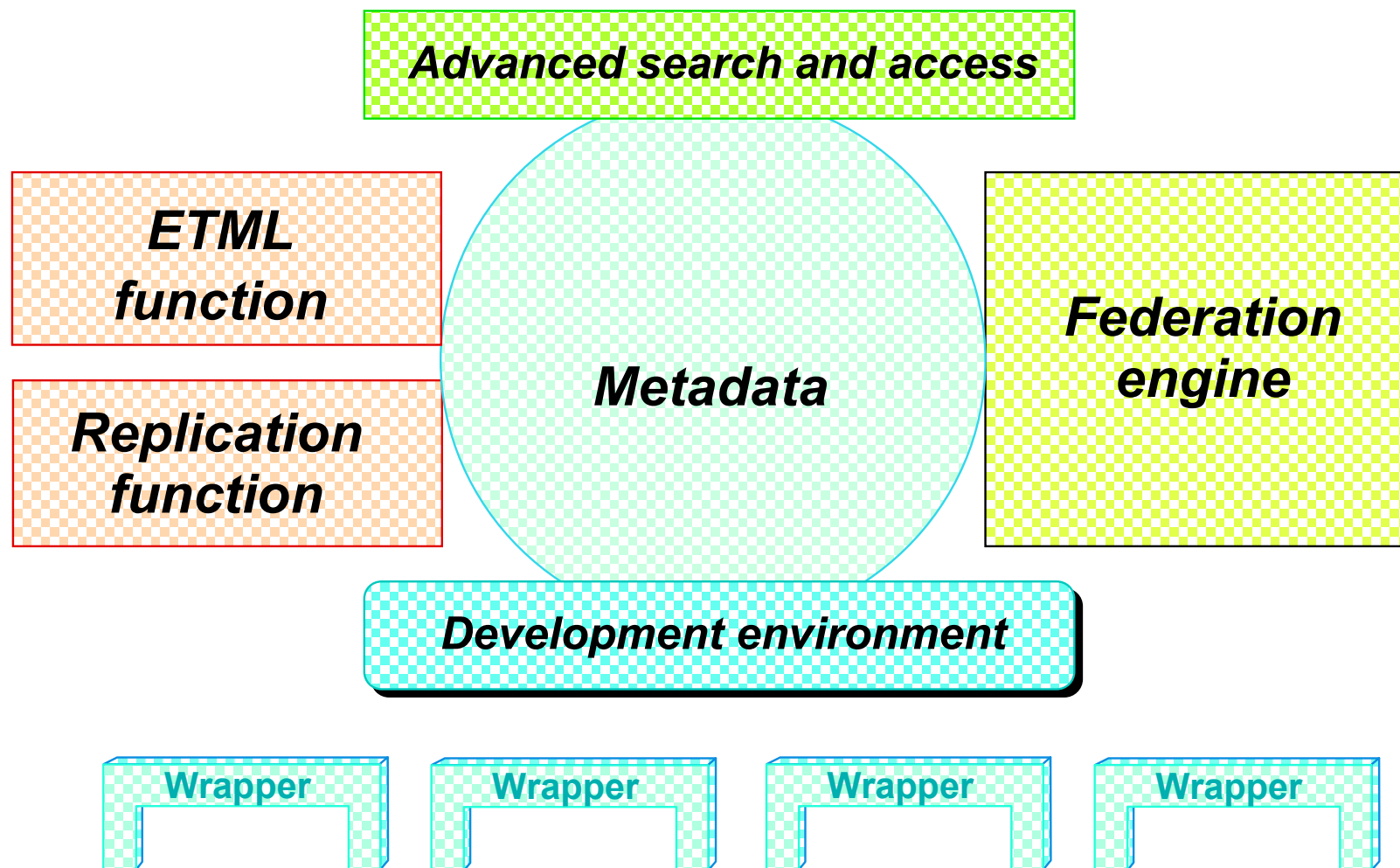
Integration Patterns Approach

- Most modern integration needs are a complex mix of both application and information integration
  - ► Joint scenarios needed in many cases
  - ► Technical and non-technical criteria for when each type of integration should be used
- Simple but interesting integration patterns that each focus on one or two integration aspects
  - ► Based on known customer pain points
- General technical patterns rather than specific business problems
  - ► But each pattern can be applied to specific industry or even customer scenarios
  - ► Patterns can be combined

Patterns

1. Beyond the data warehouse

2. Supporting the distributed enterprise

3. Integrated view of real-time customer information

4. Creating a fully integrated environment

1 - Beyond the data warehouse

Client

Advanced search & access

BI Tool

Information integration

ETML / Replication function

Metadata

Federation engine

RDBMS Data Mart

RDBMS Data Mart

RDBMS Data Warehouse

RDBMS Database

Wrapper **DBMS** Database

Wrapper **Applicatio**

Wrapper **DBMS** Content

RDBMS Data Mart

Data warehouse and data marts

Existing Operational Systems
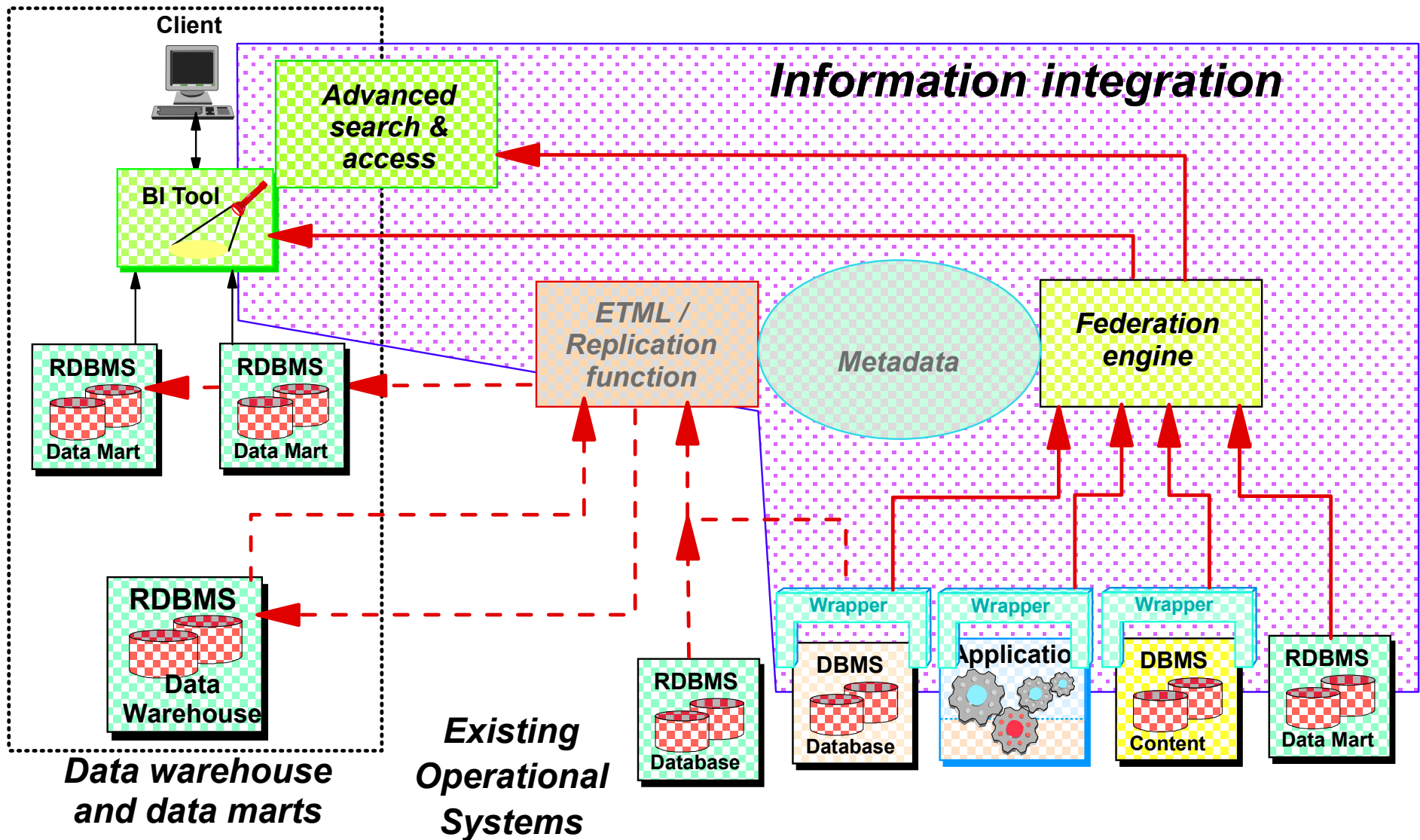
1 - Beyond the data warehouse

- Premise:
  - ▶ Every large company has multiple, growing data warehouses. The further requirement to add external, often unstructured data to the warehouse is creating additional burdens.
- Approach:
  - ▶ Federated access to trusted data, existing data marts, reference data and unstructured data from BI tools reduces the need to physically copy data to one place.
  - ▶ Added value through analysis of unstructured data
- Functionality used:
  - ▶ Federation of unstructured and structured data, wrappers
  - ▶ Advanced search and access

1 - Beyond the data warehouse
Usage cases/scenarios

- Analytics
  - Extending reach, scalability of query/reporting tools when querying several warehouses/marts
- Virtual Data Marting
  - Constructing project-specific or domain-specific marts from underlying warehouses and marts
    - Financial reporting, Performance metrics
  - Caching/updating strategies for frequently-hit tables
- Data Warehouse overlay/supplementals
  - XML, WebServices, message queueing infrastructure to accelerate integration
  - Neutral representation of data for partners, clients
  - Supplement structured data with unstructured
    - Example: IMS, VSAM, flat file, spreadsheet, industry-specific, etc.

Customer Scenario - Aventis

- The Company
  - A world leader in pharmaceuticals
  - Headquarters in Strasbourg France
  - 92,500 employees
- The Goal: Improve drug discovery process
  - Improve drug research efficiency by providing access to data where ever it's stored
  - Encourage interdisciplinary cooperation between chemists and biologists
- The Challenge
  - Enabling access to external genomics and proteomics data sources and to databases at company locations in Paris, Frankfurt, New Jersey and Arizona

Aventis Data Access

"Find a compound with structure similar to this one, and with the following assay results"

Frankfurt

Paris
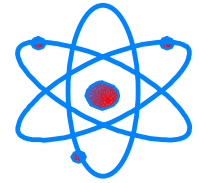
Bridgewater
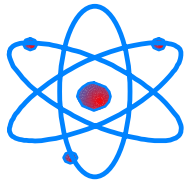
Tucson

Industry and Gov't sources of proteomic, genomic data*
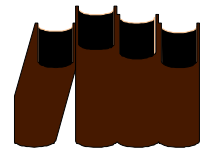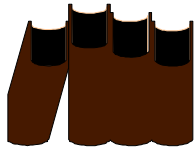
* Future

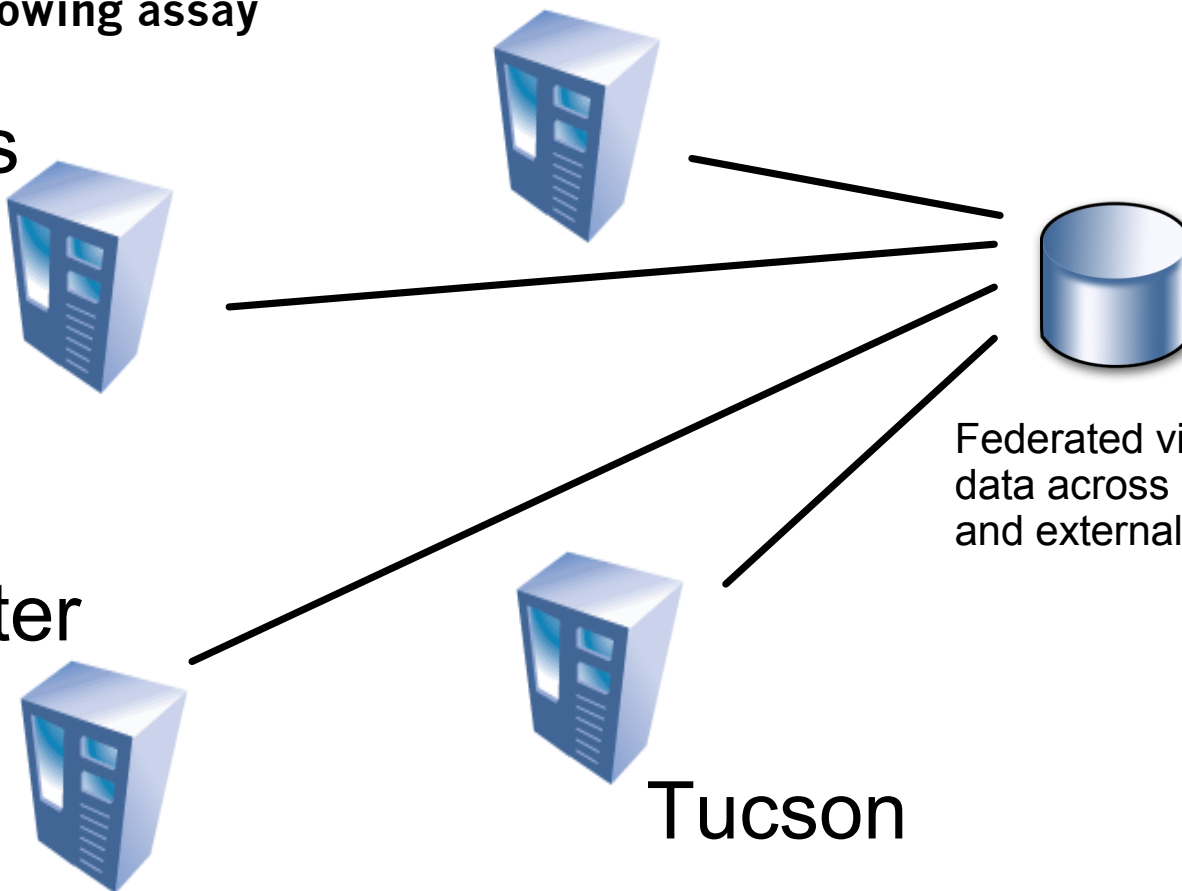# Aventis Data Access - Logical View

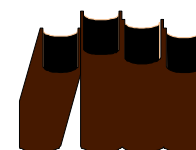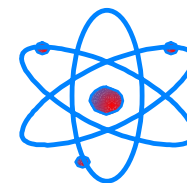"Find a compound with structure similar to this one, and with the following assay results"

Frankfurt

Paris

Bridgewater

Tucson

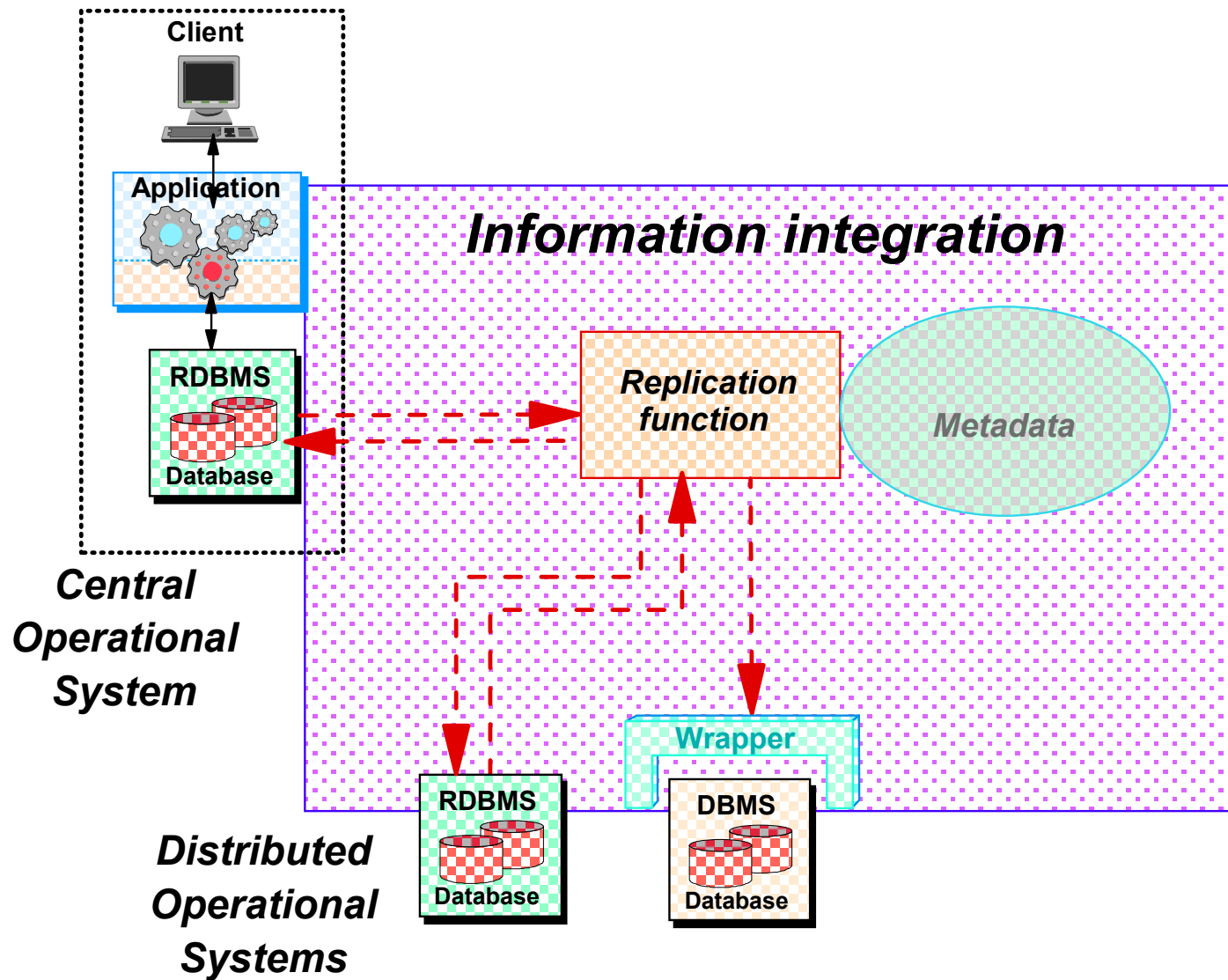Federated view to data across internal and external sites

Aventis

- The Solution
  - ► DB2 Relational Connect enables federated access to multiple, heterogeneous platforms, in this case Oracle. DB2 Life Sciences Data Connect accesses Life Sciences-specifc sources such as those with protein and genome information
  - ► The federated view over both relational and non-relational data makes it simple for end users to run powerful queries that join data across both
  - ► DB2 on pSeries servers at all 4 sites. DB2 client was installed on all Oracle databases
- The Results
  - ► The solution is expected to increase research productivity leading to drug innovation

2 - Supporting the distributed enterprise

Client

Application

RDBMS

Database

Central Operational System

Distributed Operational Systems

Information integration

Replication function

Metadata

Wrapper

RDBMS

Database

DBMS

Database

IBM.

2 - Supporting the distributed enterprise

- Premise:
  - Many companies have data subsetted and distributed across branch offices, remote users, etc. and need to have such data synchronized with a central system. Data flows in both directions, but (usually) at different times
- Approach:
  - Replication between central and distributed databases, e.g., from centralized at start of day and from distributed at end of day
- Functionality used:
  - Replication to/from relational systems
 - Wrappers to include non-relational systems

Customer Scenario - s.Oliver

- The Company
  - International fashion house
  - 21 showrooms across Europe with the latest men's, women's and children's apparel

- The Goal: Increase sales
  - Improve inventory visibility to sales personnel
  - Improve the efficiency of its order processing - enable production to act immediately on new order information
  - Put near realtime data on the showroom floors and replicate new showroom orders to ERP system automatically

- The Challenges
  - Needed near realtime product and sales information for remote order-entry and management information system
  - Couldn't get data where they wanted it, which made it hard to keep abreast of demand in the ultra-competitive global fashion industry

Multiple Replication Processes

Showrooms

Headquarters

Netfinity

Inventory

Order

Replication

Replication

Order

ERP
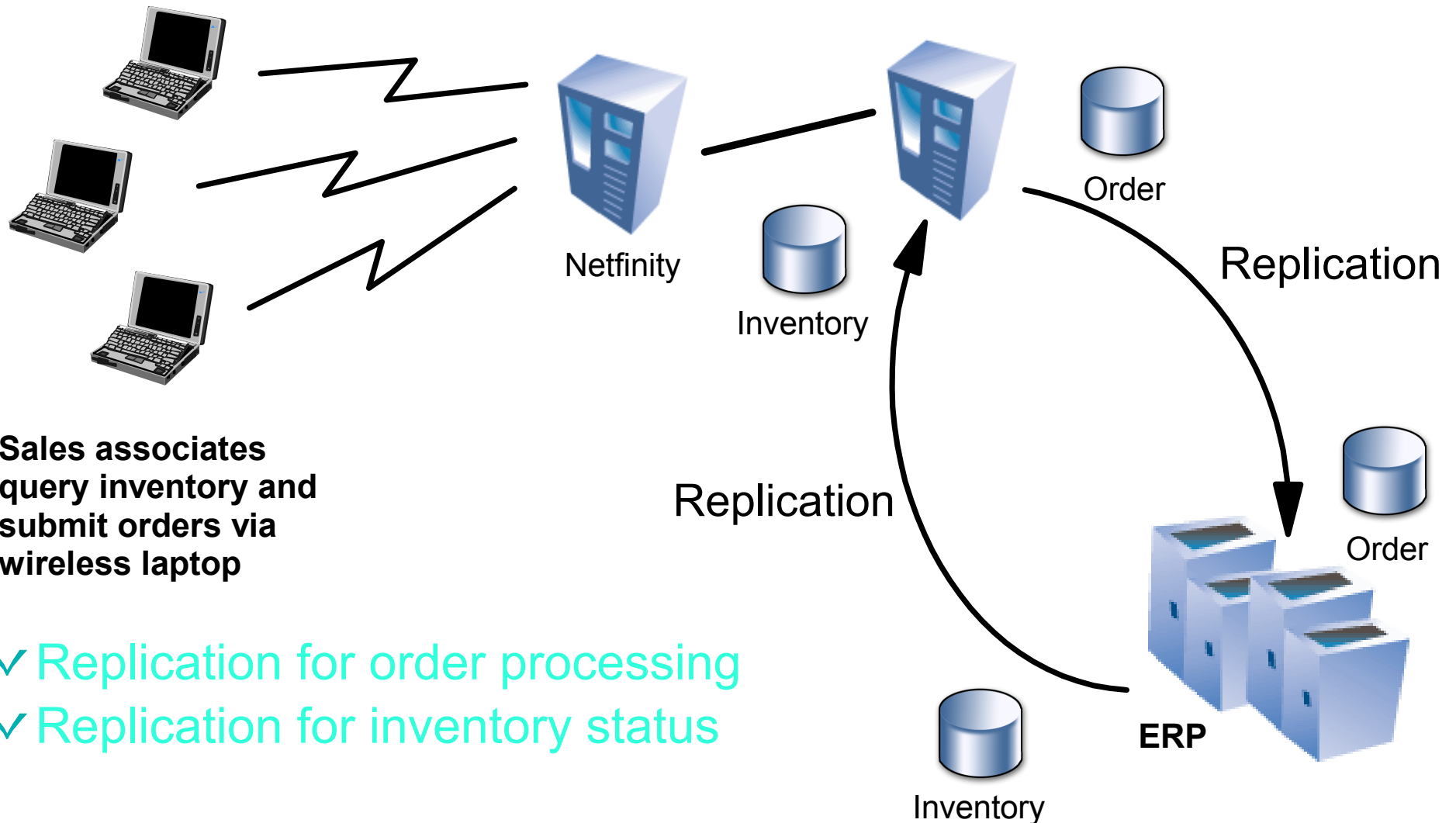
Inventory

Sales associates query inventory and submit orders via wireless laptop

✓ Replication for order processing
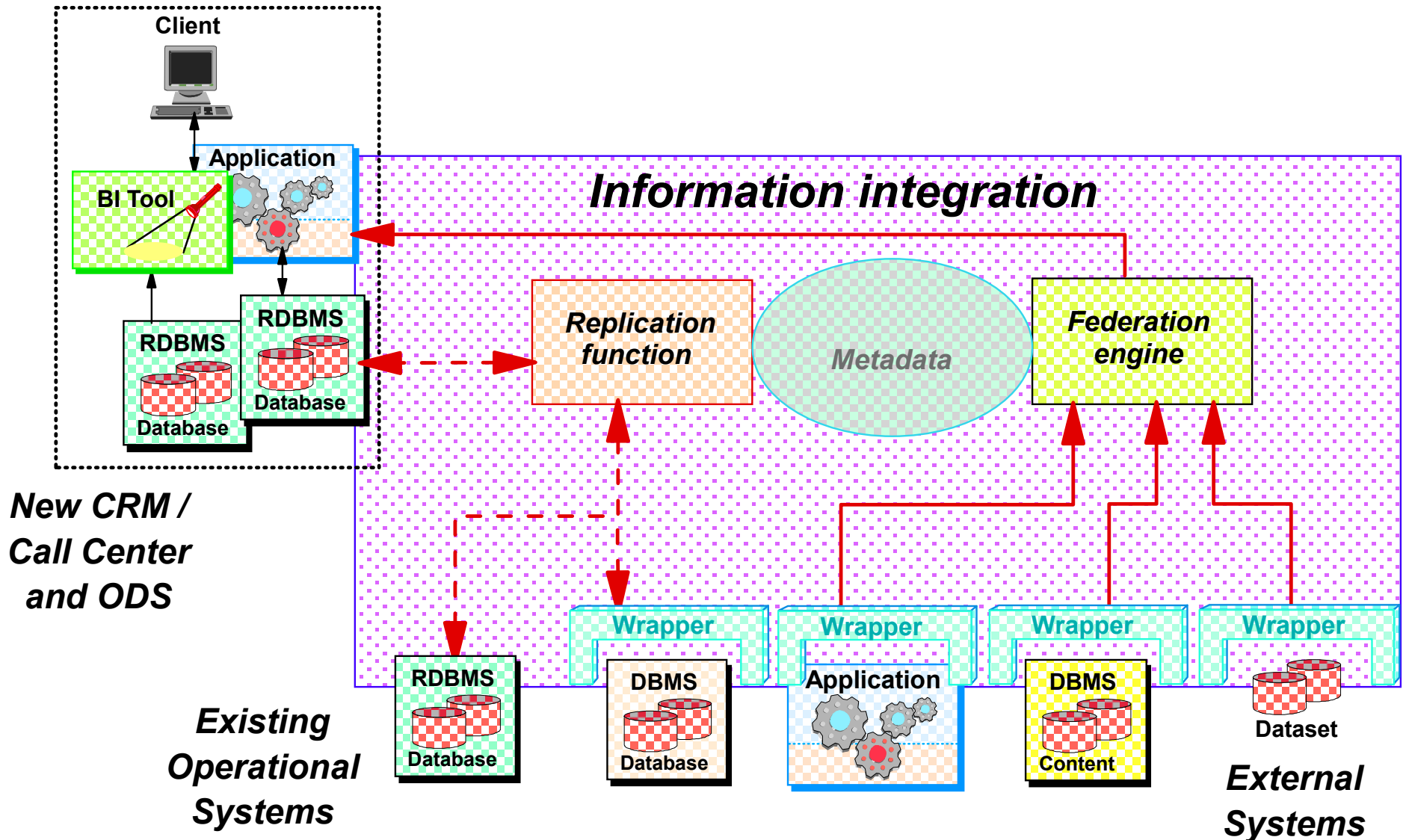✓ Replication for inventory status

s.Oliver

- The Solution
  - ► Showroom floor ordering application enabled through DB2 replication solution and high-availabililty e-business infrastructure
  - ► Sales associates use notebook computers to look up collection inventory and place orders. DB2 DataPropagator replicates the latest orders from the staging servers to the production AS/400 server and the latest inventory status from the AS/400 server to the p-Series servers

- The Results
  - ► Projected sales increase of 9.1M USD
  - ► Savings of 40 employee-hours per week by eliminating redundant data-entry tasks
  - ► Realtime initiation of manufacturing fulfillment processes compared to previous 1-day delay
  - ► Improved order accuracy and customer satisfaction

3 - Integrated view of real-time customer info

Client

Application

BI Tool

RDBMS
Database

RDBMS
Database

**New CRM /
Call Center
and ODS**

Information integration

Replication
function

Metadata

Federation
engine

Wrapper

Wrapper

Wrapper

Wrapper

**Existing
Operational
Systems**

RDBMS
Database

DBMS
Database

Application

DBMS
Content

Dataset

**External
Systems**

IBM

3 - Integrated view of real-time customer info

- Premise:
  - ► CRM, call centers, etc. require a common, consistent view of customer / client information. There is an increasing need for a near real-time view of at least some of this information. Such information is usually spread throughout many diverse and often conflicting sources throughout the organization.
- Approach:
  - ► Creation and ongoing maintenance of an operational data store (ODS) containing basic customer data - with bi-directional data flows. Federated access to real-time status info.
- Functionality used:
  - ► Federated access to relational and non-relational structured data via wrappers
  - ► Replication of data to/from the ODS

3 - Integrated view of real-time customer info Usage cases/scenarios

- Portals
  - Complement portal infrastructure with caching (performance, scalability, availability)
  - Simplify application development by reducing the access logic/complexity in application
  - Simple read/write semantics to underlying RDBs and well-behaved apps (referential integrity)
- Custom development
  - Use XML, WebServices, message queueing infrastructure to accelerate integration
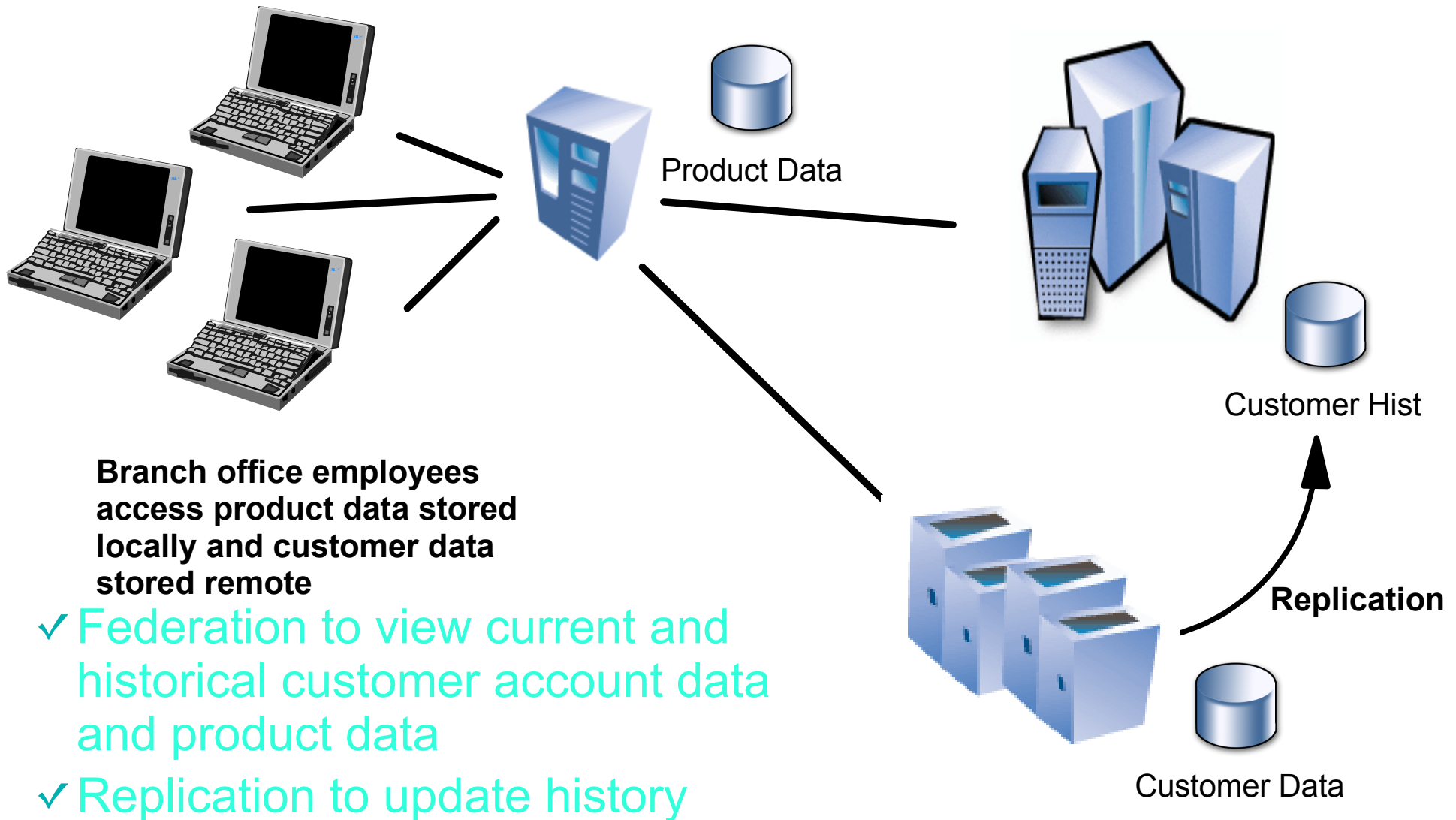
Customer Scenario - Banque Nagelmackers

- **The Company**
  - ► Oldest and one of most prestigious banks in Belgium
  - ► 80 branches, 600 employees
  - ► Known for its personalized service
- **The Goal: Re-engineer IT systems**
  - ► Improved customer service
  - ► Reduced need for IT support
  - ► Robust environment for new application development to improve operations
- **The Challenges**
  - ► Enabling transparent enterprise wide data access so business users can easily access the information they need, when they need it, without IT support

Federated Access & Replication

Branch Offices

Central Server

Product Data

Customer Hist

Replication

Customer Data

Branch office employees access product data stored locally and customer data stored remote

✓ Federation to view current and historical customer account data and product data

✓ Replication to update history

Banque Nagelmackers

- **The Solution**
  - ► DB2 DataJoiner enables easy access to data scattered on multiple, heterogeneous platfroms. It ensures data location transparency, thus making access simple for end users
  - ► A Netfinity server at the bank's headquarters at Brussels runs DB2 UDB and DataJoiner. Customer account information is stored on a NT server and updated automatically to the host S/390 system into DB2, using DataJoiner and capture/apply capabilities
- **The Results**
  - ► Easy data access for end users, while ensuring that consistent data is consolidated on the mainframe and is available on demand to all the branches

"With this three-tier client/server architecture, we simplify data access for end users, while ensuring that consistent data is consolidated on the mainframe and is available on demand to all the branches," Didier Roquet explains. "Without location transparency it would be very difficult to develop new applications or improve performance. We evaluated other solutions from other database management system providers, but IBM gave us an integrated solution with DB2 and DB2 DataJoiner, which simplified development and speeded up our time-to-market."

4 - Creating a fully integrated environment
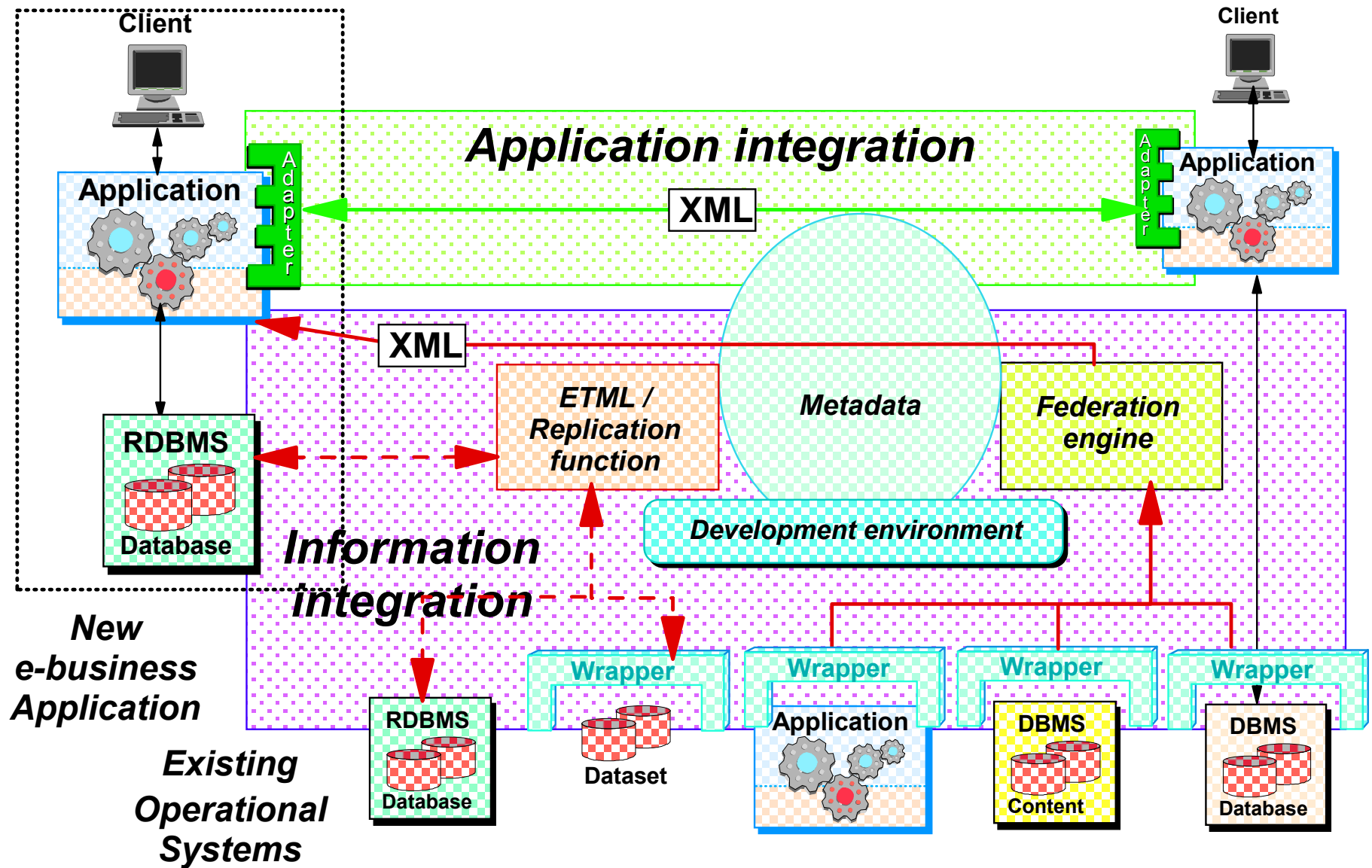
4 - Creating a fully integrated environment

- Premise:
  - ► With the introduction of e-business, applications must be more tightly integrated.
    - – Web-based order-entry system needs read access to various sources such as catalogs, preferences, etc.
    - – Order must be constructed under process control, requiring synchronization of actions over a number of applications, such as billing, inventory, etc.
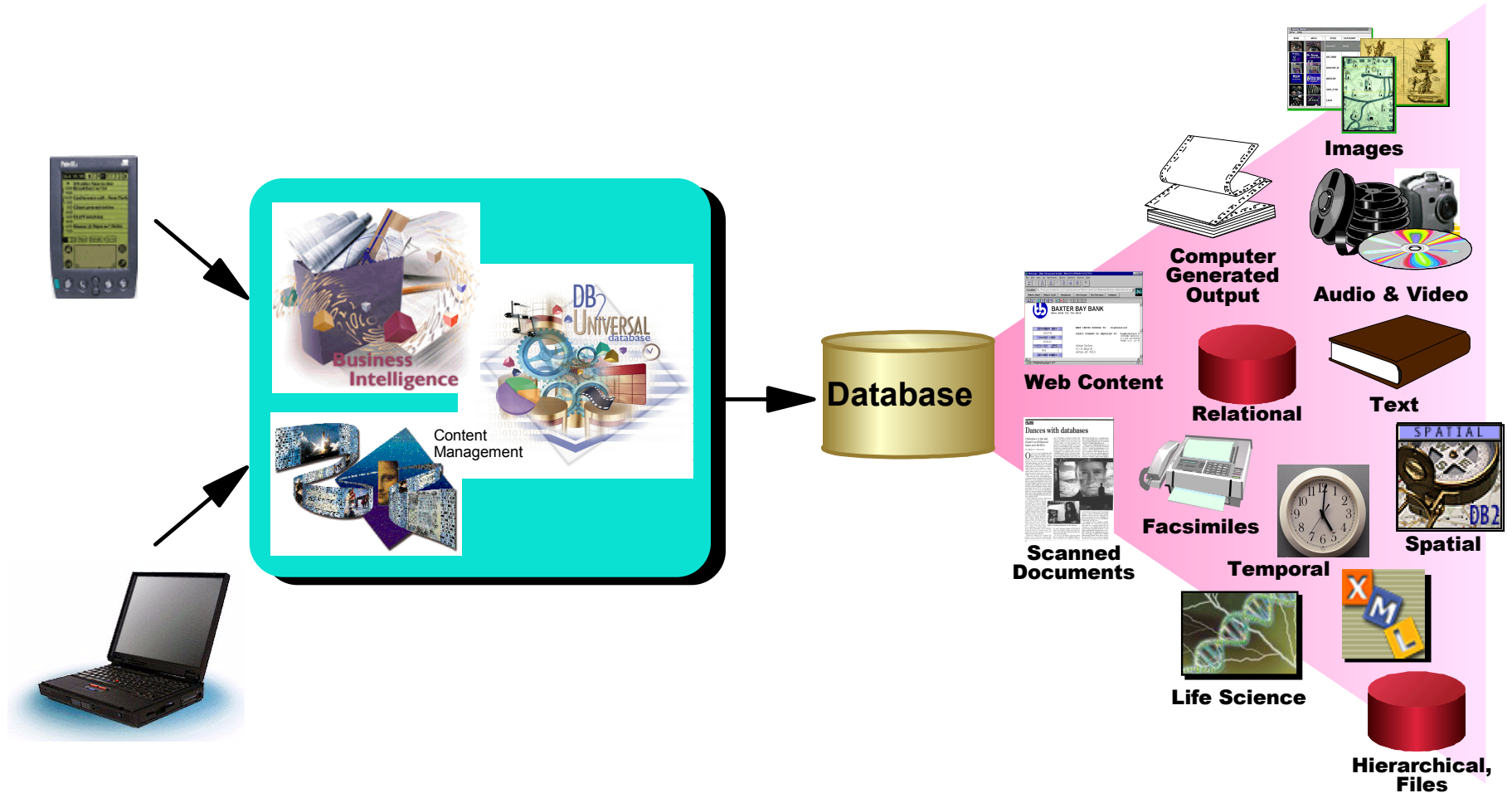- Approach:
  - ► Federation allows the app to read existing data in real time. Use of XML query expected to increase
  - ► XML messaging is used under process control to update the required data via its owning app
- Functionality used:
  - ► Federation of structured data, XML
  - ► May include some bidirectional federation
  - ► Replication / ETML for reference, non-time-critical data
  - ► Application integration - messaging

Information Integration
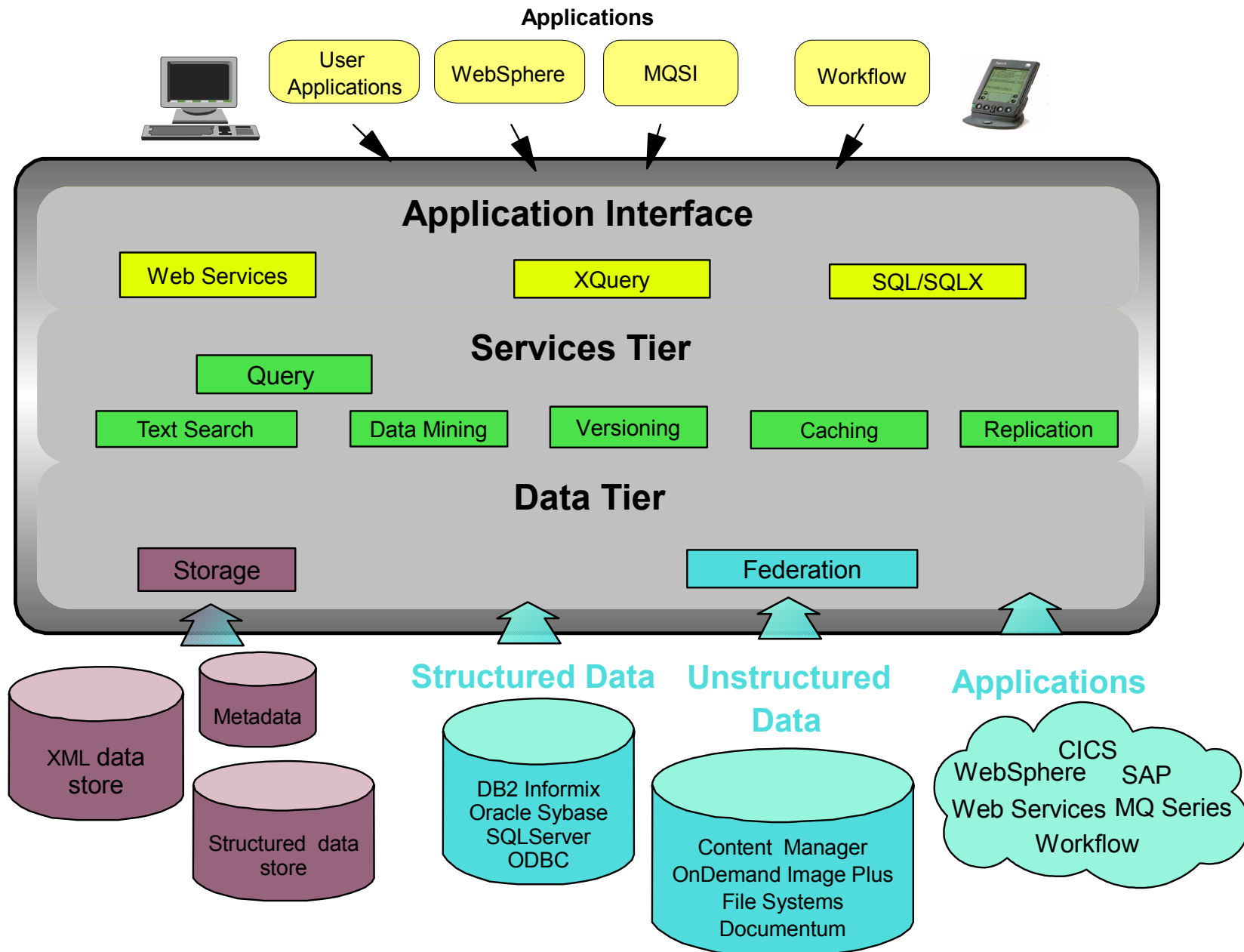## *The Future*

Images

Computer Generated Output

Audio & Video

Web Content

Relational

Text

Business Intelligence

Content Management

Database

Facsimiles

Temporal

Spatial

Scanned Documents

Life Science

Hierarchical, Files

What is Needed for Information Integration?

- Middleware platform that partners and customers can exploit for integrating information:
  - Integration of unstructured, structured data and semi-structured data
    - federation
    - data placement - replication and caching
  - Advanced integration services, search, and analytics
    - meta data management
    - mining
  - Robust infrastructure to complement application and process integration
    - messaging
    - workflow
  - XML support
    - integrated data store for XML and relational data
    - XQuery
    - web services, SOAP, UDDI

Information Integration Platform

Applications

User Applications · WebSphere · MQSI · Workflow

Application Interface

Web Services · XQuery · SQL/SQLX

Services Tier

Query

Text Search · Data Mining · Versioning · Caching · Replication

Data Tier

Storage · Federation

XML data store

Metadata

Structured data store

Structured Data

DB2 Informix Oracle Sybase SQLServer ODBC

Unstructured Data

Content Manager OnDemand Image Plus File Systems Documentum

Applications

CICS WebSphere SAP Web Services MQ Series Workflow

For more information:
http://www7b.boulder.ibm.com/dmdd/library/techarticle/0206roth/0206roth.html

IBM®

Thank You!