Extending your business intelligence investment with DB2 Information Integrator

November 2003

business on demand software

Nelson Mattos

Distinguished Engineer and Director, Information Integration

Agenda

It's here now
The next generation of
data integration software

- Information Integration & on demand computing

- Introducing DB2 Information Integrator V8.1

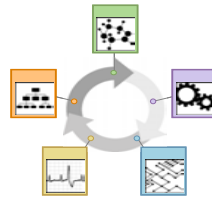- Partner Ecosystem

- Usage Scenarios

- Roadmap

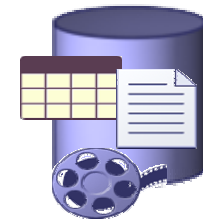- Summary

Defining Business Integration

The efficient and flexible combination of resources to optimize operations across and beyond the enterprise

People + **Processes** + **Information**

Portal – personalized information	Workflow management	Federation & data placement
Collaboration technology	B2B connectivity	(replication, ETL, caching)
Adaptable workplace	Messaging infrastructure (EAI)	XML (store, query, web services)
Consistent rule-based	e-Business transactions	Meta-data management
experience across devices	Service-oriented architecture	Text Search and analytics

Can I create business value from my existing IT systems? ➔ **People, Process, Information**

Can users react in real-time to the most recent information? ➔ **People, Information**

Are business operations fully integrated for optimal efficiency? ➔ **Process, Information**

# Information Integration is Key to Business Integration Solutions

"It is impossible in most cases to separate data integration from application or BP integration because all integration has a data component. For this reason the DBMS must play a large role in an overall integration solution."

Charles Garry, META Group, 2002

People **+** **Processes** **+** **Information**

Data federation is the only way customers will be able to integrate petabytes of information.

**Teri Palanca,
Giga Information Group, 2002**
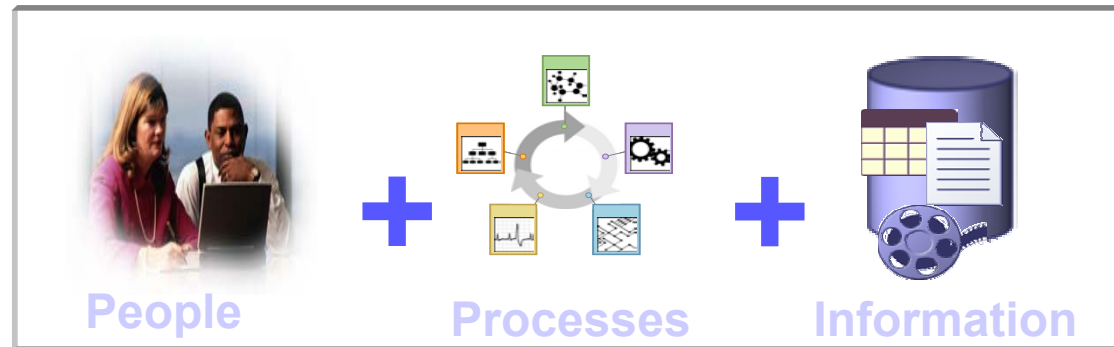
At least 30% of all new e-business applications face the problem of integrating multiple data sources.

Giga Information Group, 2000

Data Challenges

Variety, Velocity, and Volume

- New composite applications need data from multiple sources
  - Consumers expect holistic, personalized, and value-added content
  - Relational, XML, packaged applications, content repositories, file systems all contain critical business information

- Increasing emphasis on current data
  - Real-time analytics
  - Business activity monitoring

- Petabytes will be the measure of available online data
  - All client interactions are important (e.g. instant messages, audio records, Web traffic,…)
  - Internet and intranet content

The world produces 250MB of information every year for every man, woman and child on earth.

85% of digitized information unstructured.

Complementary Information Integration Approaches

- **Consolidate data for local access**
  - ▶ Access performance or availability requirements demand centralized data.
  - ▶ Currency requirements demand point-in-time consistency, e.g. close of business
  - ▶ Complex transformation is required to achieve semantically consistent data
  - ▶ Production applications, data warehouses, operational data stores
  - ▶ Typically managed by ETL (Extract, Transform, and Load) or replication technologies

- **Integrated access to distributed sources**
  - ▶ Access performance and load on source systems can be traded for overall lower cost implementation
  - ▶ Currency requirements demand a fresh copy of the data
  - ▶ Data security, licensing restrictions, or industry regulations restrict data movement
  - ▶ Combining mixed format data, e.g. customer ODS with related contract documents or images
  - ▶ Query requires real-time data, e.g. stock quote, on-hand inventory
  - ▶ Represents an emerging category of technology – Enterprise Information Integration (EII)

What is Enterprise Information Integration?

Enterprise Information Integration refers to a category of middleware which lets applications access diverse and distributed data as if it were a single source, whether or not it is.

It enables the integration of structured and unstructured data

▸ to provide real-time read and write access,

▸ to transform data for business analysis and data interchange, and

▸ to manage data placement for performance, currency, and availability.

EII Components

Integrating diverse business information across and beyond the enterprise

Federation Placement Transformation
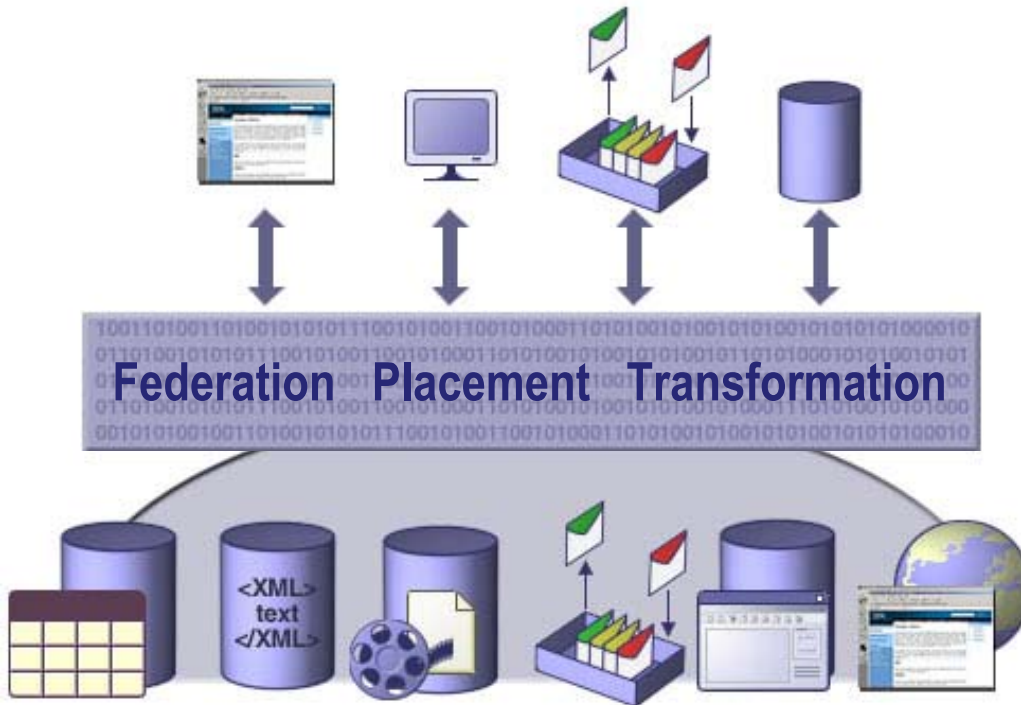
- Data federation
  - ▸ Standard programming models
    - • SQL, XQuery, Search
  - ▸ Read/write access across diverse data and content sources
  - ▸ Extensible access

- Data placement
  - ▸ Caching and replication over heterogeneous information

- Data transformation
  - ▸ Leveraging SQL and XML standards
  - ▸ Advanced search and mining
  - ▸ Metadata management and interchange

Analyst Views

- Redmonk
  - ▶ James Governor, an analyst at RedMonk, calls **federated data management a "crucial technology"...."**Offering the ability to pull together far flung information in a centralized location and at manageable expense, this approach will become **a no-brainer for those in need of real-time access to real-time data**," he said.

    Dow Jones International News, May 21, 2003
    THE BOTTOM LINE: Software Heralds The Utd States Of Data

- Gartner Group
  - ▶ Ted Friedman, a principal analyst at Gartner, agrees. **"The potential impact of this technology could be very large, especially given the size of the vendors involved,"** he said....Friedman said the federated approach will be most relevant where there are a limited number of data sources, few users and when real-time data are needed.

    Dow Jones International News, May 21, 2003
    THE BOTTOM LINE: Software Heralds The Utd States Of Data

- Robert Francis Group
  - ▶ RFG believes IT executives who apply EII technologies effectively and appropriately will gain a new ability to **produce substantive results more quickly, with lower costs and less training and difficulty than required by alternate approaches.**

    Evan Bauer, June 17, 2003

    Enterprise Information Integration – An Idea Whose Time May Have Finally Come (Part 1)

Agenda

It's here now
The next generation of
data integration software

- Information Integration & on demand computing

- Introducing DB2 Information Integrator V8.1

- Partner Ecosystem

- Usage Scenarios

- Roadmap

- Summary

DB2 Information Integrator 8.1

A Federated Data Server – Query distributed data as if it were a single source

- **Define integrated view across diverse and distributed data**
  - Wide range of data and content sources
  - Extensible to virtually any data source
- **Query as if a single source**
  - Use standard SQL query and SQL expressions
  - Include text semantics in the search
  - Surface specialized functions into SQL
  - Leverage query optimization and caching
- **Compose XML documents**
  - Combine diverse sources
  - Validate against DTDs or schema
- **Publish results to a message queue**
  - Familiar DB programming model
- **Single source, relational updates**
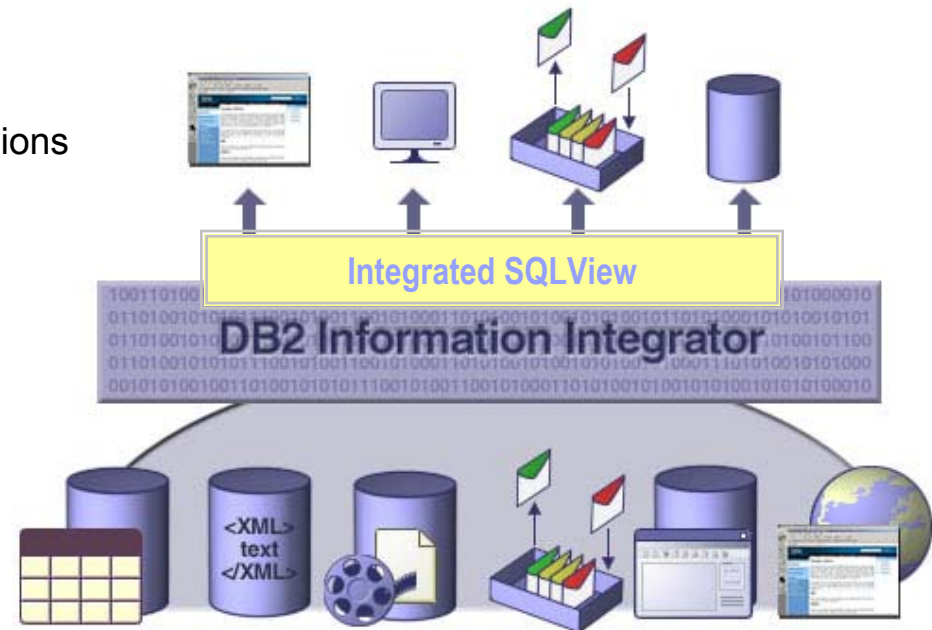
DB2, Informix, Oracle, SQL Server, Sybase, Teradata, OLE DB, ODBC, Excel, XML, message queues, Web services, flat files, document repositories, content repositories, LDAP directories, WWW, email databases, and more.

Integrated SQLView

DB2 Information Integrator

```
100110100                                        101000010
01101001010                                      01010011000
01101001010                                      00100101100
011010010101110010100110101001010100110000010100
```
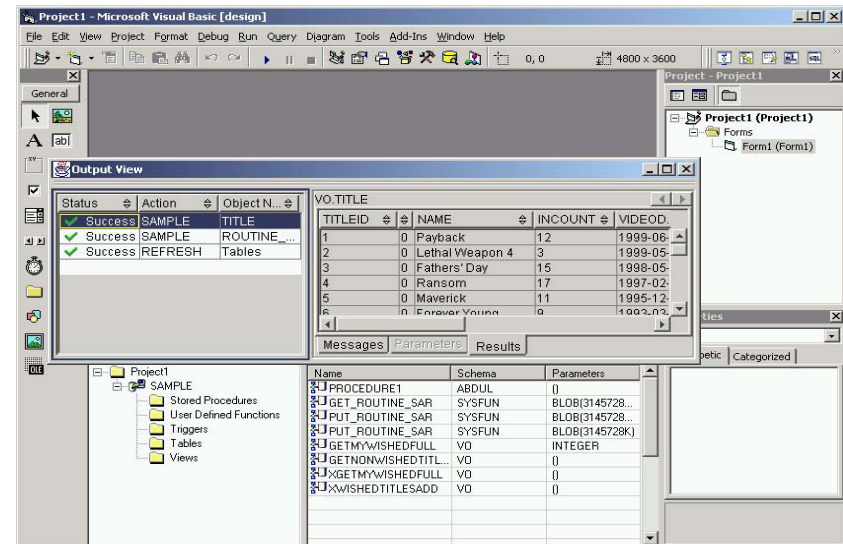
Speed Application Development

Developer productivity for both worlds: Java and .NET

WebSphere Studio
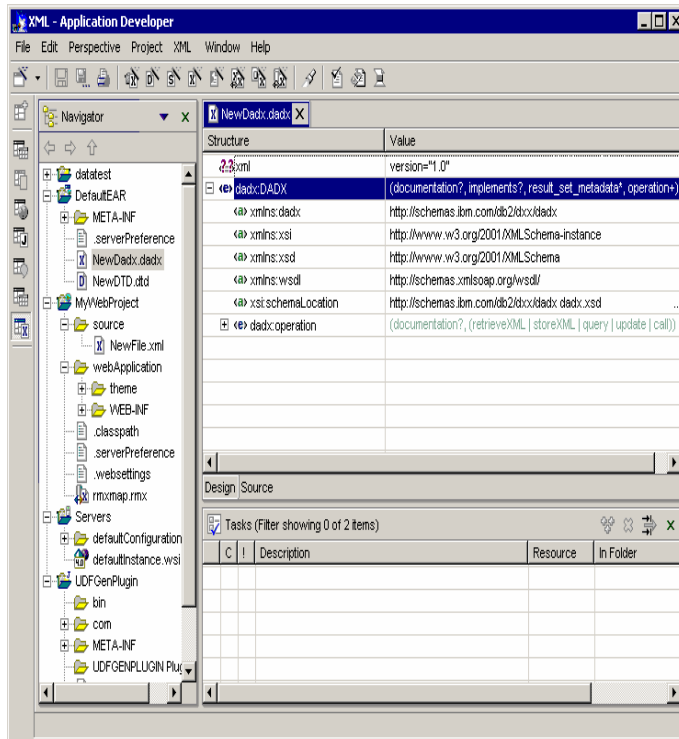
### Microsoft Visual Studio

- Reduce hand coding 40%-65%
- Maintain performance
- Ideal for portal and analytical applications

DB2 Information Integrator 8.1

A Replication Server – Manage placement for performance and availability

- **Distribute data among relational databases**
  - ▸ DB2, Informix, Microsoft, Oracle, Sybase, Teradata
- **Support flexible topologies**
  - ▸ Distribution: One to many
  - ▸ Consolidation: Many to one
  - ▸ Column and row subsetting
  - ▸ Multiple flavors, e.g. history tables, key changes
- **Match data movement modes to usage requirements**
  - ▸ Table-at-a-time for warehouse loading during batch window
  - ▸ Transaction-consistent for online data
- **Choose latency characteristics**
  - ▸ Scheduled, interval-based, event-driven, continuous
- **Apply transformations in-line**
  - ▸ Standard SQL expressions or stored procedure execution.

ORACLE → DB2

DB2 Information Integrator

Microsoft   Microsoft
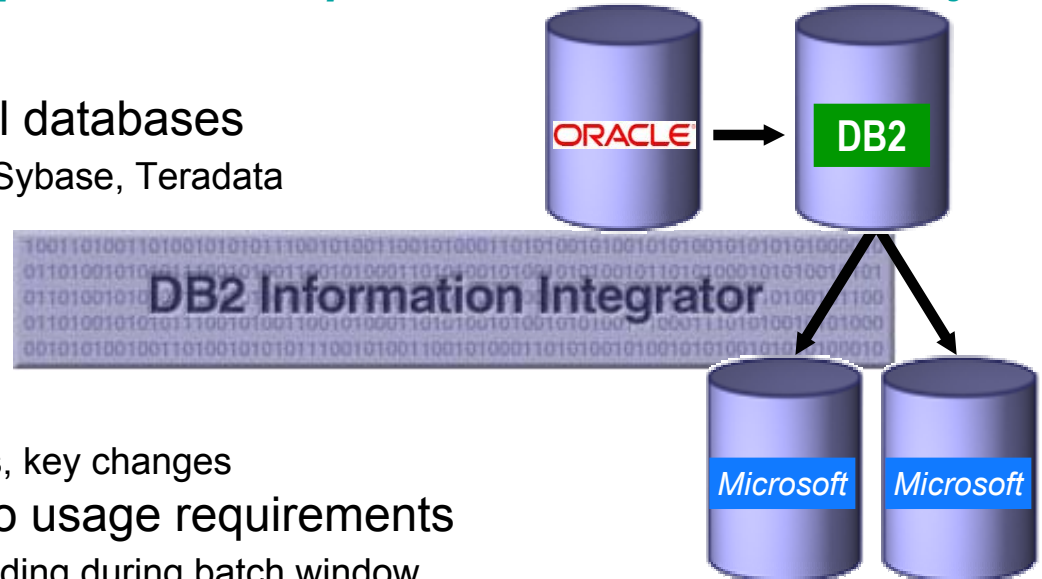
IBM Acquires CrossAccess

- Extending leadership of the enterprise information integration market
  - Key to IBM's on demand initiative

- Technology
  - eXadas Data Integrator (XDI)
    - Real-time SQL read/write access to mainframe
    - Metadata driven, no mainframe coding required
    - Supports joins and views across data sources
    - ODBC and JDBC compliant connectors
  - eXadas Synchronization Manager (XSync)
    - Change capture from mainframe and distributed sources
    - Non-intrusive capture for mainframe sources plus Oracle and SQL Server
    - 100% recoverable

- Significant client value
  - Extends current IT investments (on mainframe databases)
  - Speeds time-to-value for integration projects
  - Reduces dependency on scarce mainframe skills
  - Simplifies transactions and maintenance, increases confidence
  - Enables worldwide coverage through IBM global channel

DB2 Information Integrator
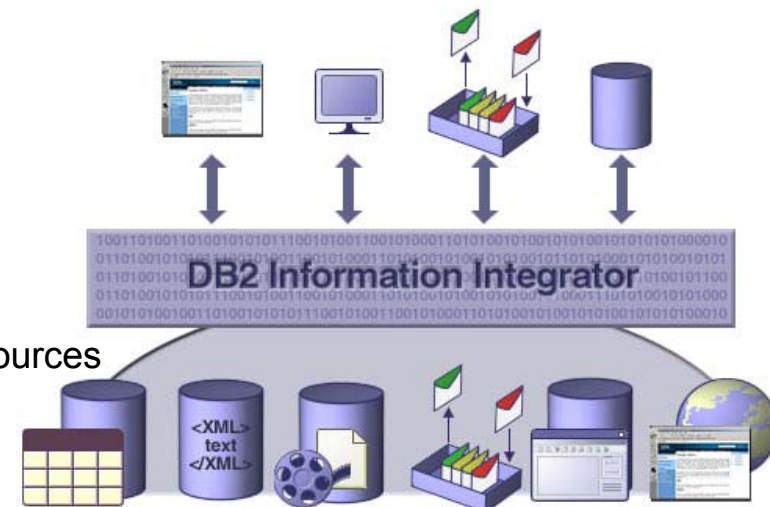
*Extends reach to IMS, VSAM, CA-IDMS, CA-Datacom/DB, Adabas*

Value to Clients

- Extend current investments
  - ▶ Work within your existing infrastructure
  - ▶ Consolidate data or access distributed data as if it were a single data source
  - ▶ Combine existing data and content assets in new ways
  - ▶ Use familiar SQL programming model and existing tools
  - ▶ Build on a standards-based, strategic integration platform
- Speed time to value for composite applications
  - ▶ Reduce hand-coding 40%-65%
  - ▶ Reduce skill requirements
  - ▶ Reduce development time by half
- Control costs
  - ▶ Reduce payroll costs
  - ▶ Reduce need to rip and replace
  - ▶ Reduce need to manage redundant data

Administration Demo

Four easy steps

- Create wrapper
- Create server
- Create user mapping
- Create nickname
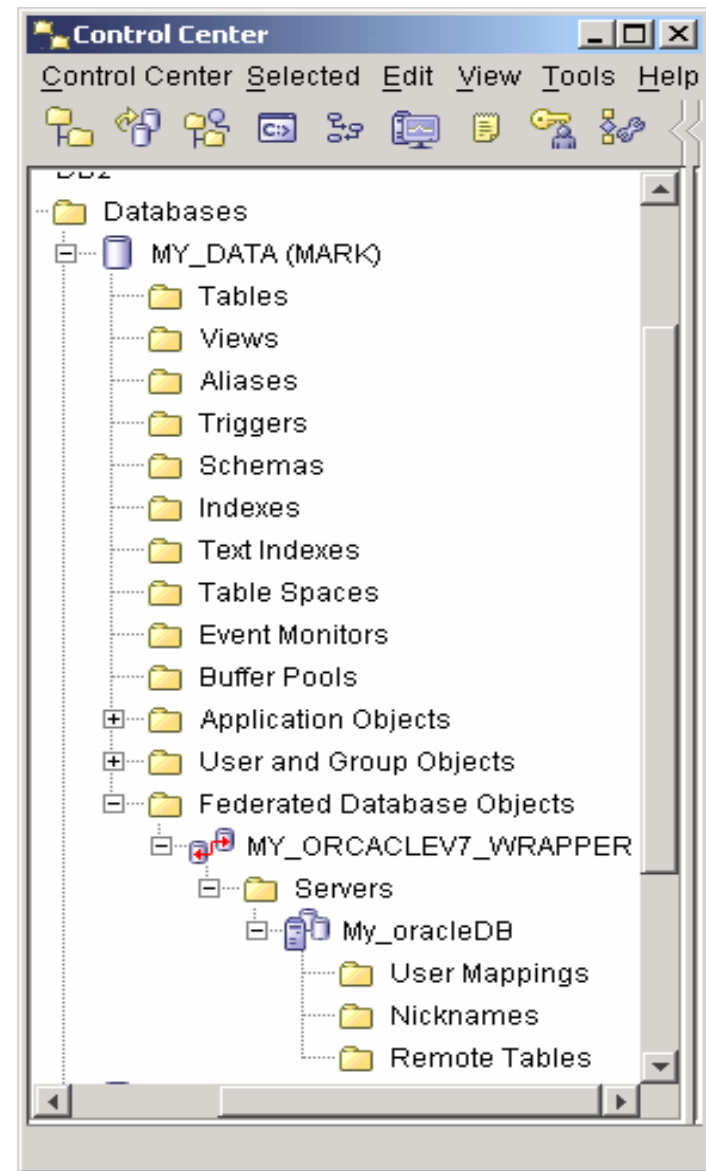
Agenda

It's here now
The next generation of
data integration software

- Information Integration & on demand computing

- Introducing DB2 Information Integrator V8.1

- Partner Ecosystem

- Usage Scenarios
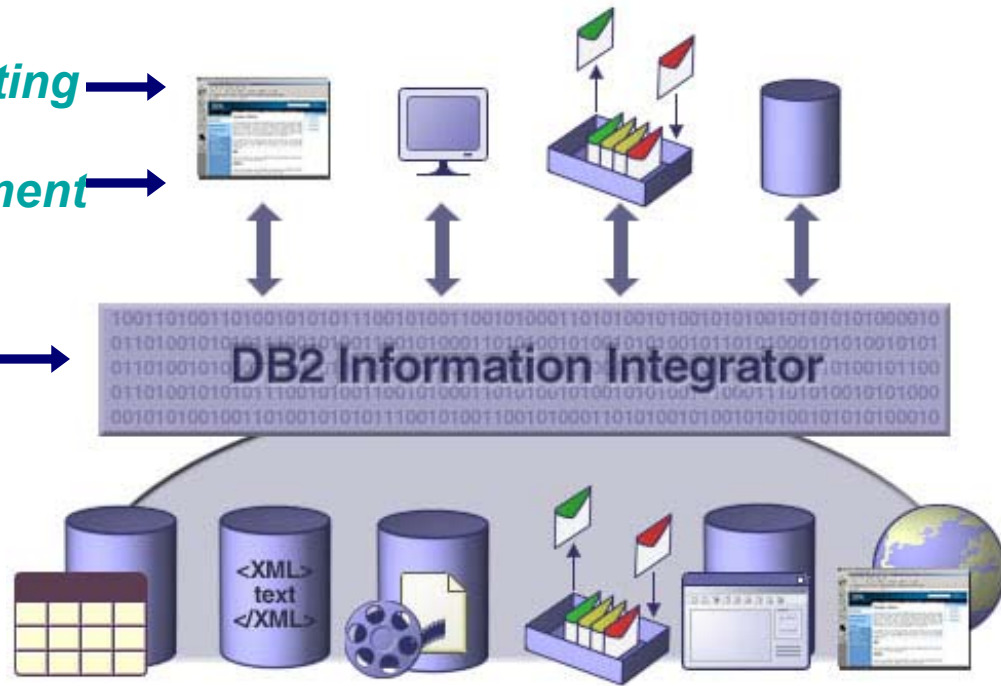
- Roadmap

- Summary

World-Class Partner Ecosystem

Enhance analysis and reporting →

Enable application development →

Enrich function →

DB2 Information Integrator

Extend reach →

Expedite deployment ←→

<XML> text </XML>

World-Class Partner Ecosystem

"..exceptionally fast and efficient federated querying"

"..completely transparent.."

"..opens new opportunities.."

"..allows users access to important data that was often forgotten or used infrequently.."

"..providing integrated on-demand analysis and execution.."

"..an open, extensible, and consistent information integration framework.."

COGNOS

crystal decisions

BUSINESS OBJECTS

ACXIOM

BRIO SOFTWARE

siperian

ANSI

INFORMATICA

SAS

Ascential Software

NEON SYSTEMS, INC.

Striva

MicroStrategy
Best In Business Intelligence

iWay Software

ATTUNITY

symphony RPM

Agenda

It's here now
The next generation of
data integration software

- Information Integration & on demand computing

- Introducing DB2 Information Integrator V8.1

- Partner Ecosystem

- Usage Scenarios

- Roadmap

- Summary

Enhancing Business Intelligence

Chinese Ministry of Railways

- Warehouse Maintenance
  - ▸ Licensing agency replicates POS transaction detail from DB2 UDB to DB2, Oracle, or Microsoft SQL Server reporting databases

- Extended Reporting
  - ▸ Extend reporting with real-time data, unstructured content, or cross-warehouse views

- Research and Development
  - ▸ Complex query over diverse and distributed data

- Business Activity Management
  - ▸ Access and integrate KPIs, planned targets, historical trends, and current values

Enhancing Business Intelligence

ITERGO

JOHN DEERE

for internal use only

Business Portals Business Integration
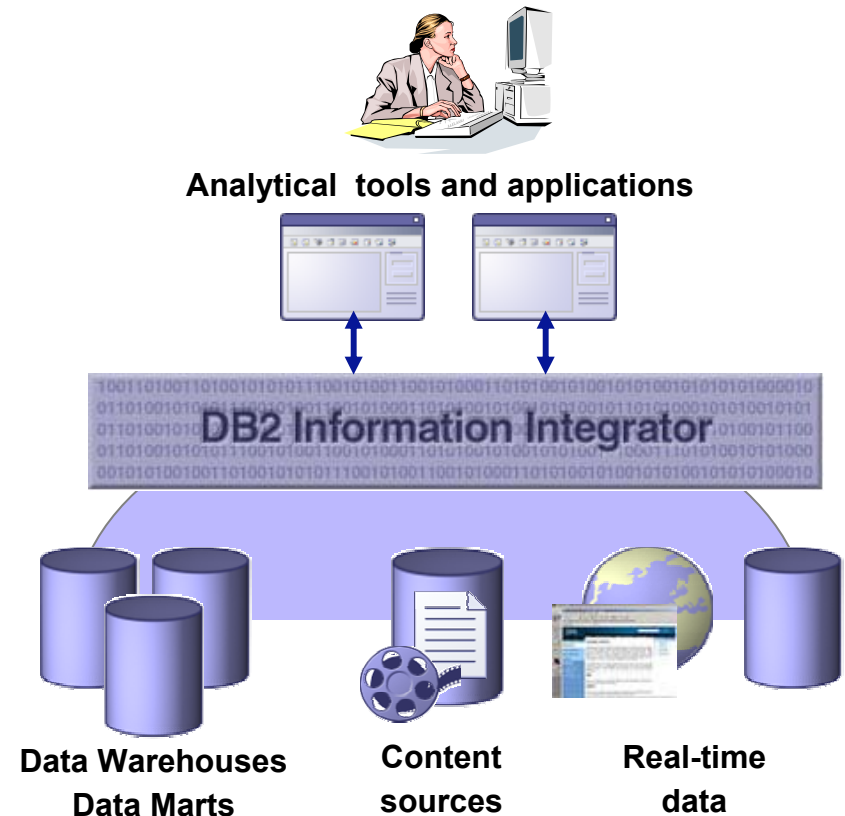
WebSphere

Foundation & Tools

- Customer Data Integration
  - ▸ Integrate customer operational data stores or registries with related detail

- Mergers and Acquisitions
  - ▸ Create integrated views across business units

- Portal Development
  - ▸ Access and integrated diverse sources for Web deployment

Extending the Data Warehouse

Customer Value

▸ Better decisions from more comprehensive & more up-to-date information

▸ More responsive to business needs that require access to new data sources

▸ Increased ROI from existing data warehouses

Analytical tools and applications

DB2 Information Integrator

Data Warehouses
Data Marts

Content sources

Real-time data

Considerations for Extending the Data Warehouse

- The majority of the data resides in the data mart/warehouse
  - ▶ Even better if the data mart and the federated engine are co-located
- The data required from the remote store is focused
  - ▶ Retrieved through a single request
  - ▶ Relatively small number of records (may contain significant amount of data)
  - ▶ Preferably through an interaction for which the remote store is optimised
- The models of the data mart and remote store can be related
  - ▶ Analysis is required – ad hoc access cannot be allowed
  - ▶ Key matching may be required as part of the federation
- Performance vs. storage and management costs
  - ▶ Network performance and availability
  - ▶ Costs of storage and copy management

Ministry of Railways (China)
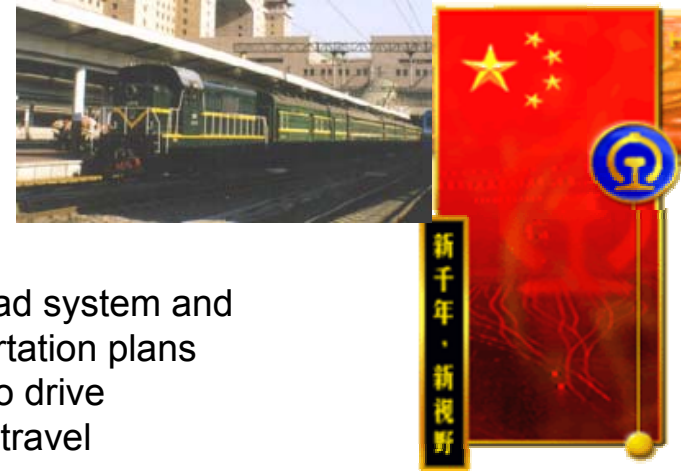
Vision

- ▶ Ministry of railways manages the Chinese the national railroad system and is responsible for defining and implementing railway transportation plans
- ▶ Want to provide an organization–wide intelligence platform to drive decision making that will increase freight and passenger rail travel

Challenge

- ▶ Data needed for effective decision-making is spread across multiple databases
- ▶ Begin solution rollout with limited skill and resource

Solution

- ▶ Give business managers access to summarized historical data in a DB2 based warehouse with federated access to real-time DB2 and Oracle freight transportation operational data and passenger data in Sybase and SQl Server.

Value

- ▶ Improve the quality of railway transportation management
- ▶ Respond more quickly to market demand by improving freight forecasting
- ▶ Optimize carrying resource assignment
- ▶ *"With its real time federation capability DB2 Information Integrator significantly enhances the power of the data warehouse supporting our intelligent transportation management systems and other enterprise applications. The XML and unstructured data support of this product will help us fulfill the reporting requirements to the states administrative bureaus."*

Chief General Engineer, Ministry of Railways -China Republic Contact

Crystal Decisions

Vision

▶ As a world-leading information infrastructure company, Crystal Decisions helps businesses make better decisions by bringing together their people and their information.

Challenge

▶ Improve response time for complex queries over distributed heterogeneous data sources

Solution

▶ Provides transparent, globally optimized access to heterogeneous, distributed data. Crystal Reports accesses the distributed data as if it were a single database. Response time improvement of up to 98% seen in house.

Value

▶ *"Users of Crystal Reports and Crystal Enterprise, with DB2 Information Integrator, can … discover new ways to meet the information needs of their organization."*

- Janet Wood, Vice President of Business Development, Crystal Decisions.

▶ *"DB2 Information Integrator provides Crystal Reports with exceptionally fast and efficient federated querying capability."*

- Trevor Smith, Program Manager, Business Development Group, Crystal Decisions

MicroStrategy Validation

- MicroStrategy clients need the ability to perform analysis across multiple databases
- Full MicroStrategy test suite executed against DB2 Information Integrator
    - 1000s of reports
    - Across DB2 UDB and Oracle sources
- MicroStrategy queries typically well-optimized by DB2 Information Integrator
    - MicroStrategy is tuned to DB2 UDB
    - DB2 optimization of sub-queries is excellent

"The results … demonstrate that implementing a combination of MicroStrategy's BI Platform™ and IBM's DB2 Information Integrator enables cross-database analysis through a unified interface. Most significantly, the installation and configuration … is simple enough to make this an extremely viable and desirable BI solution for enterprises with multiple data sources."
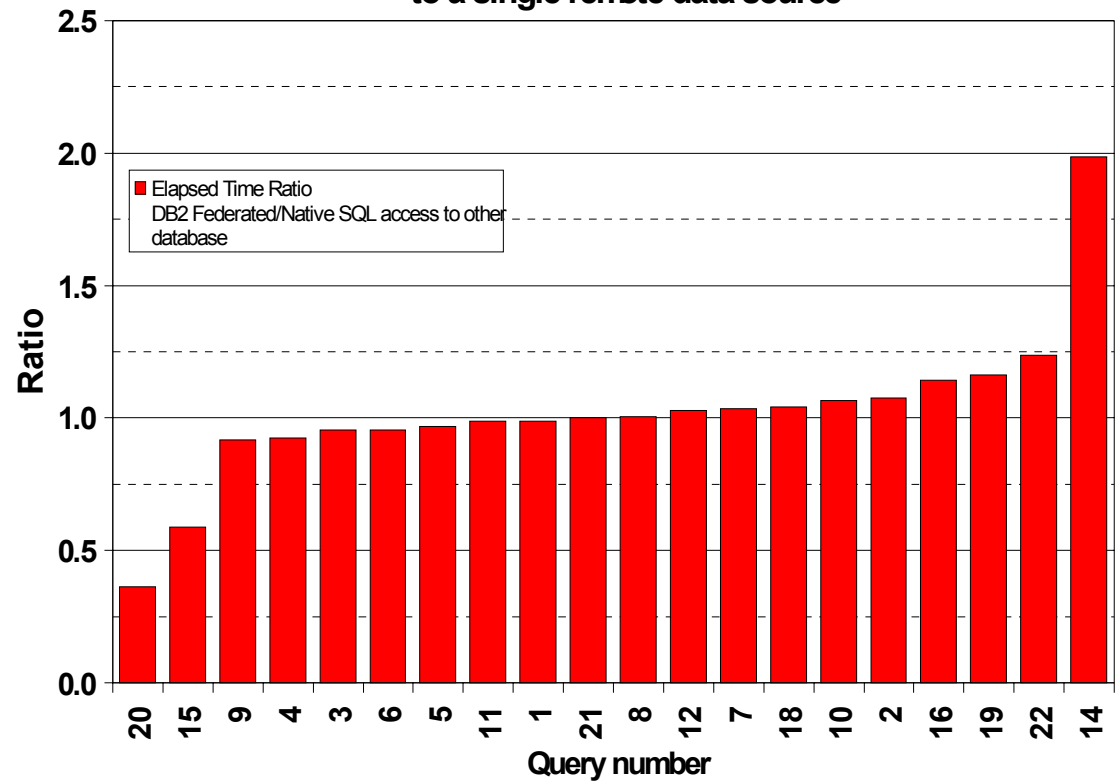
MicroStrategy®
Best In Business Intelligence™

IBM Performance Results

- **TPC-H using DB2 Information Integrator vs native client**

  - ▶ Majority of queries achieved comparable performance, thus pushdown is working well

  - ▶ Performance deterioration points us to areas for focus

- **DataJoiner comparison**

  - ▶ 80% of queries were faster

  - ▶ Half of those were twice as fast

**Ratio of query elapsed time:
DB2 Federated V8 vs. Native SQL interface
to a single remote data source**

Legend: ■ Elapsed Time Ratio
DB2 Federated/Native SQL access to other database

Y-axis: Ratio (0.0 to 2.5)
X-axis: Query number (20, 15, 9, 4, 3, 6, 5, 11, 1, 21, 8, 12, 7, 18, 10, 2, 16, 19, 22, 14)
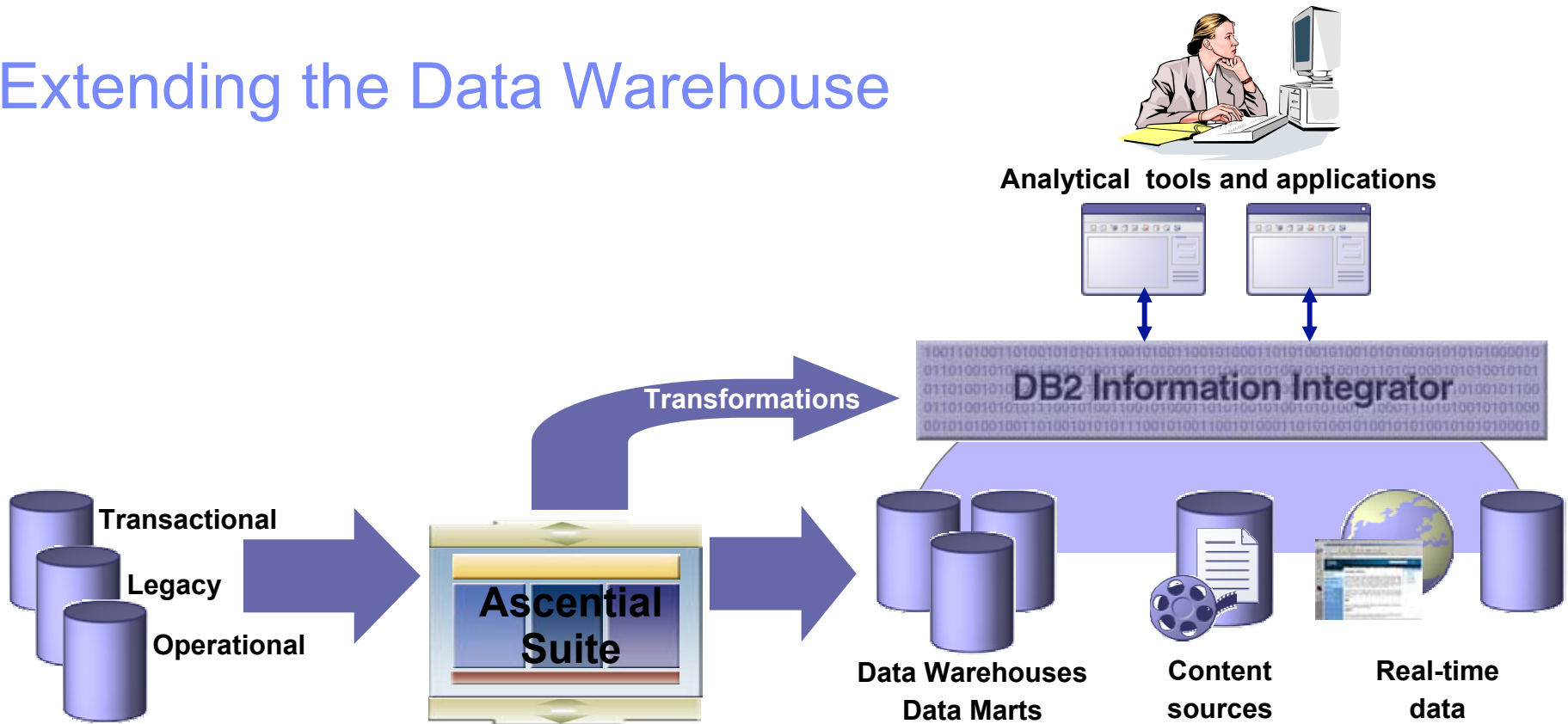
IBM and Ascential Software

- Standard SQL API to access any data
  - ▶ Mainframe and distributed
  - ▶ Structured and unstructured
  - ▶ Proprietary and public
- Unmatched performance and scalability
  - ▶ Query optimization
  - ▶ Caching
  - ▶ Data movement
- Consistent, reusable transformation library
  - ▶ Reuse ETL logic in real time
  - ▶ In-flight data quality, enrichment, and transformation
- Rapid application development
  - ▶ Encapsulate and reuse data transformation
  - ▶ Use pre-packaged libraries of transforms and visual tools
  - ▶ Integrate into applications using standard integrated development environment
  - ▶ End-to-end meta data management for accelerated information discovery and ease of management

A comprehensive information integration platform
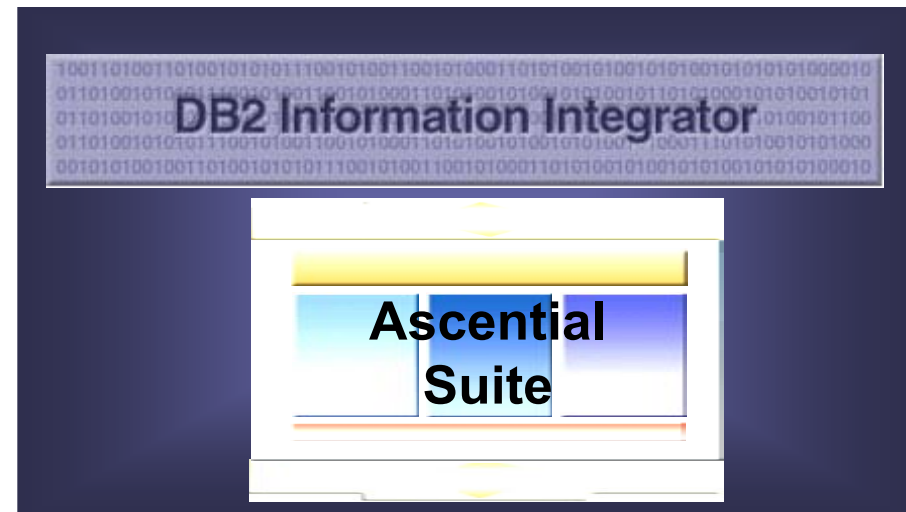
Extending the Data Warehouse

Analytical tools and applications

Transformations

DB2 Information Integrator

Transactional

Legacy

Operational

Ascential Suite

Data Warehouses Data Marts

Content sources

Real-time data

Customer Value

▶ Better decisions from more comprehensive & more up-to-date information

▶ More responsive to business needs that require access to new data sources

▶ Pre-packaged library & custom transform rules without hand-coding

▶ Integrated profiling, standardization, matching & survivorship

Business Activity Monitoring

Customer Value

- Access real-time data across disparate data stores including ODS, event logs, transactional database,…
- Correlate real-time data with historical data in warehouses or planned targets in spreadsheets
- Use standard analytical, portal, and dashboard tools for analysis and graphical, easy to consume display

DB2 Information Integrator

| Transactions | Event Logs | Master ODS | Data Warehouses Data Marts |
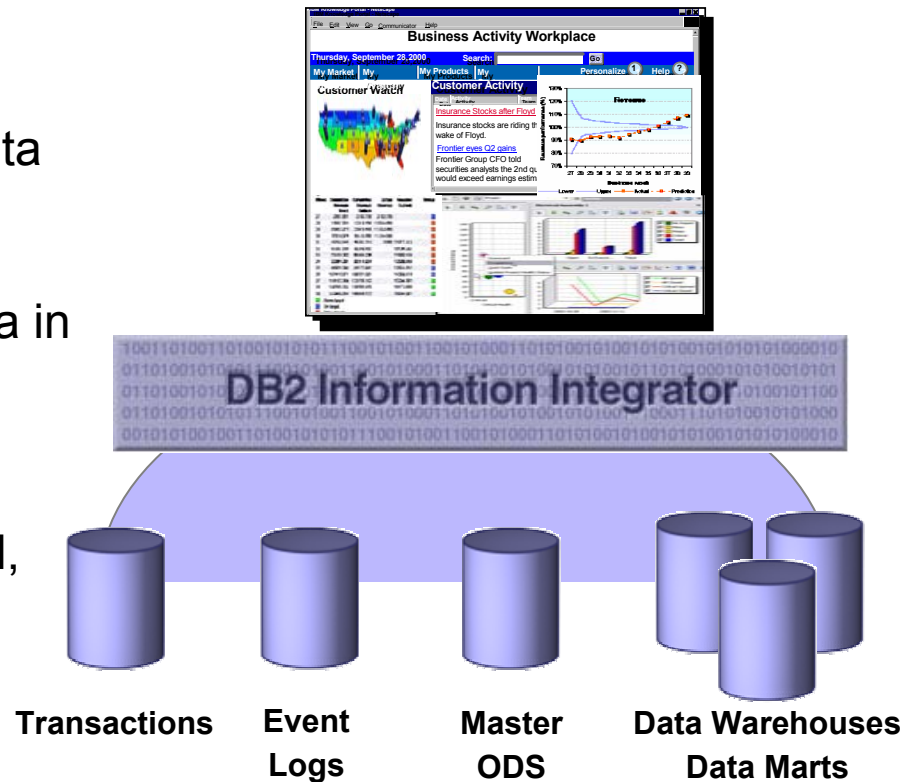
SymphonyRPM

Vision

▶ Empower all levels of your organization to make the time-critical, high-value decisions that turn corporate goals into reality.

Challenge

▶ Materialize federated result sets across an assortment of enterprise data marts

▶ Acceptable real-time performance

▶ Replace poor performing federation tool

▶ Cost of federation technology

Customer Value

▶ Produced the federated result sets faster by a factor
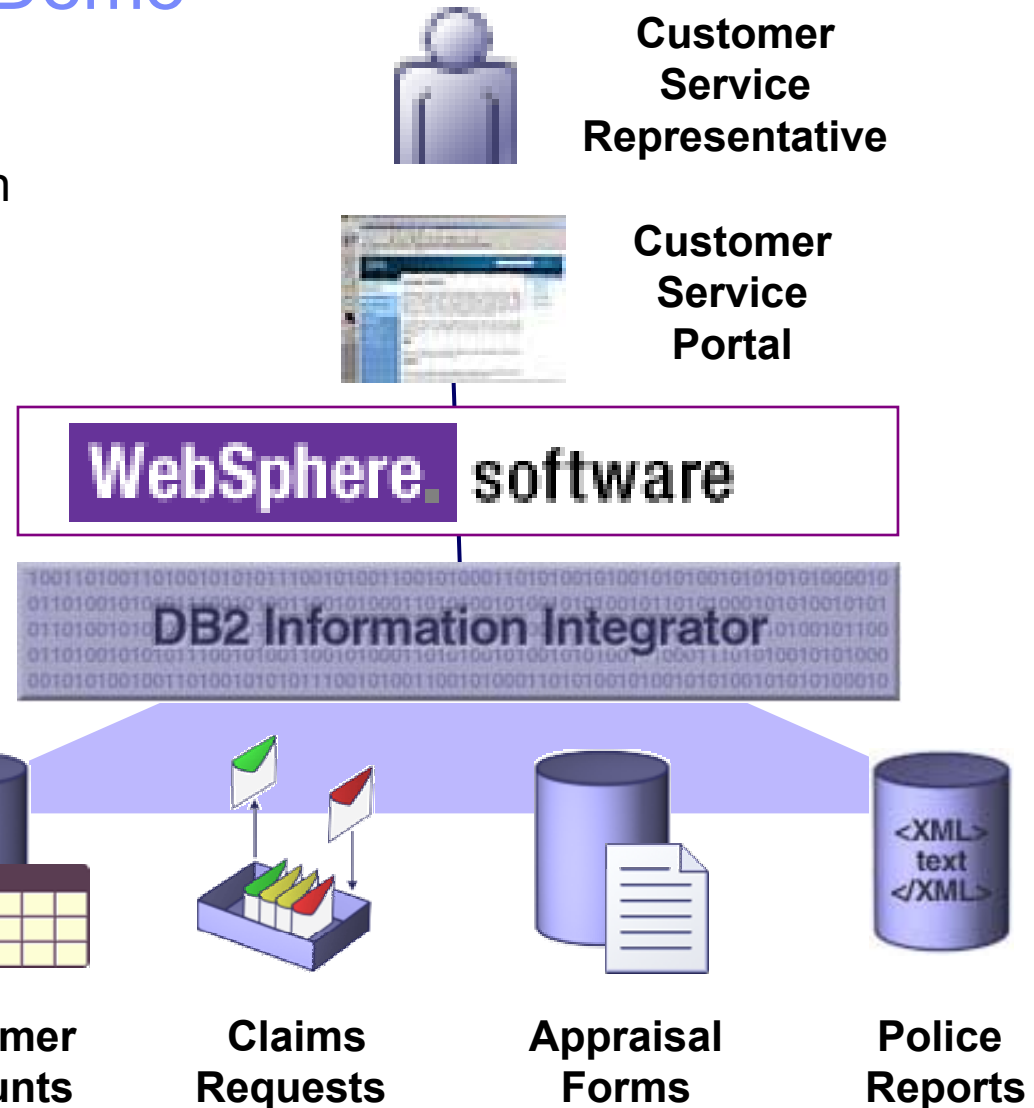
▶ Reduce the cost of the federation technology

WebSphere Synergy

- Plug into a services oriented architecture
  - All the power of EII through Web service
  - Extend reach and transformation richness with Web services
- Improve developer productivity for composite applications
  - Reduce lines or code, development time, and skill needs
  - Give applications access to all the relevant data sources
- Reduce cost and complexity of complex business processes
  - Simplify development for composite business objects
  - Leverage set-base relational processing
- Enable business activity management
  - Correlate real-time and historical data
- Increase return on messaging system deployment
  - Analyze queue traffic with standard analytical tools
  - Make messaging infrastructure more easily accessible by DB community

Business Portals

Business Integration

WebSphere®

Foundation & Tools

DB2 Information Integrator
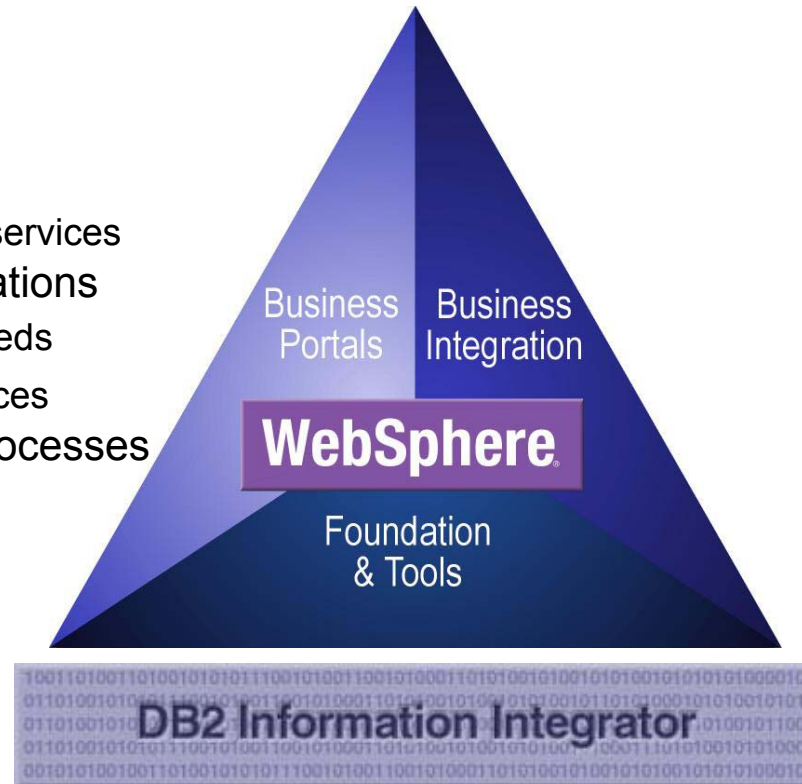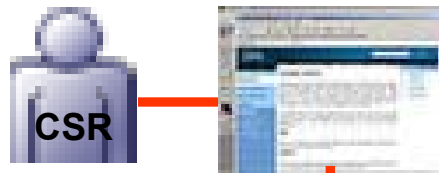
Cotton-wood Insurance Demo

- Give service personnel integrated access to relevant claims information
  - ▶ Find claims with no incident report
  - ▶ Find policy information from name
  - ▶ Find customer record plus related claims and incident reports
  - ▶ Retrieve full claims documents
  - ▶ Open new claims
- Explore integration of WebSphere and DB2 Information Integrator

Customer Service Representative

Customer Service Portal

WebSphere. software

DB2 Information Integrator

Customer Accounts

Claims Requests

Appraisal Forms

Police Reports

`<XML> text </XML>`

Cotton-wood Insurance Demo…

CSR

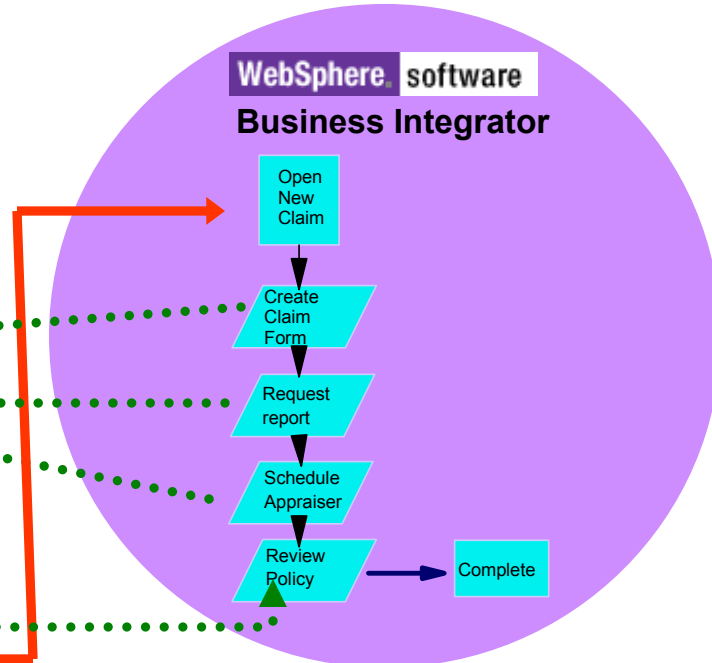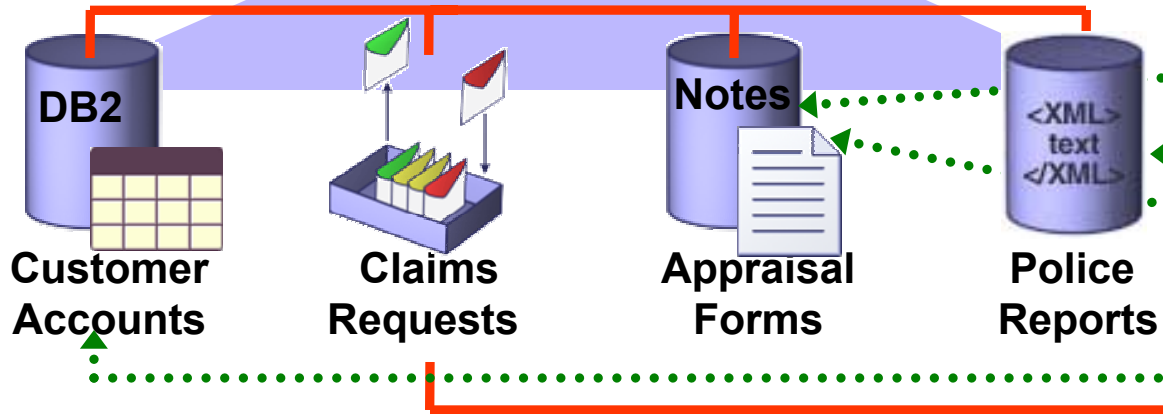**WebSphere. software**

DB2 Information Integrator

- **Opening a new claim at customer request**
  - ▸ Information collected by CSR and submitted by application
    - • MQ message written to queue
    - • MQ message presence initiates a new collaboration
      - − WBI process starts for new claims work flow
      - − Data acquired and/or updated by collaboration steps
        - − Same data access by portal via DB2 II

WebSphere. software
Business Integrator

DB2

Customer Accounts

Claims Requests

Notes

Appraisal Forms

<XML> text </XML>

Police Reports

Open New Claim

Create Claim Form

Request report

Schedule Appraiser

Review Policy → Complete

Agenda

It's here now
The next generation of
data integration software

- Information Integration & on demand computing

- Introducing DB2 Information Integrator V8.1

- Partner Ecosystem

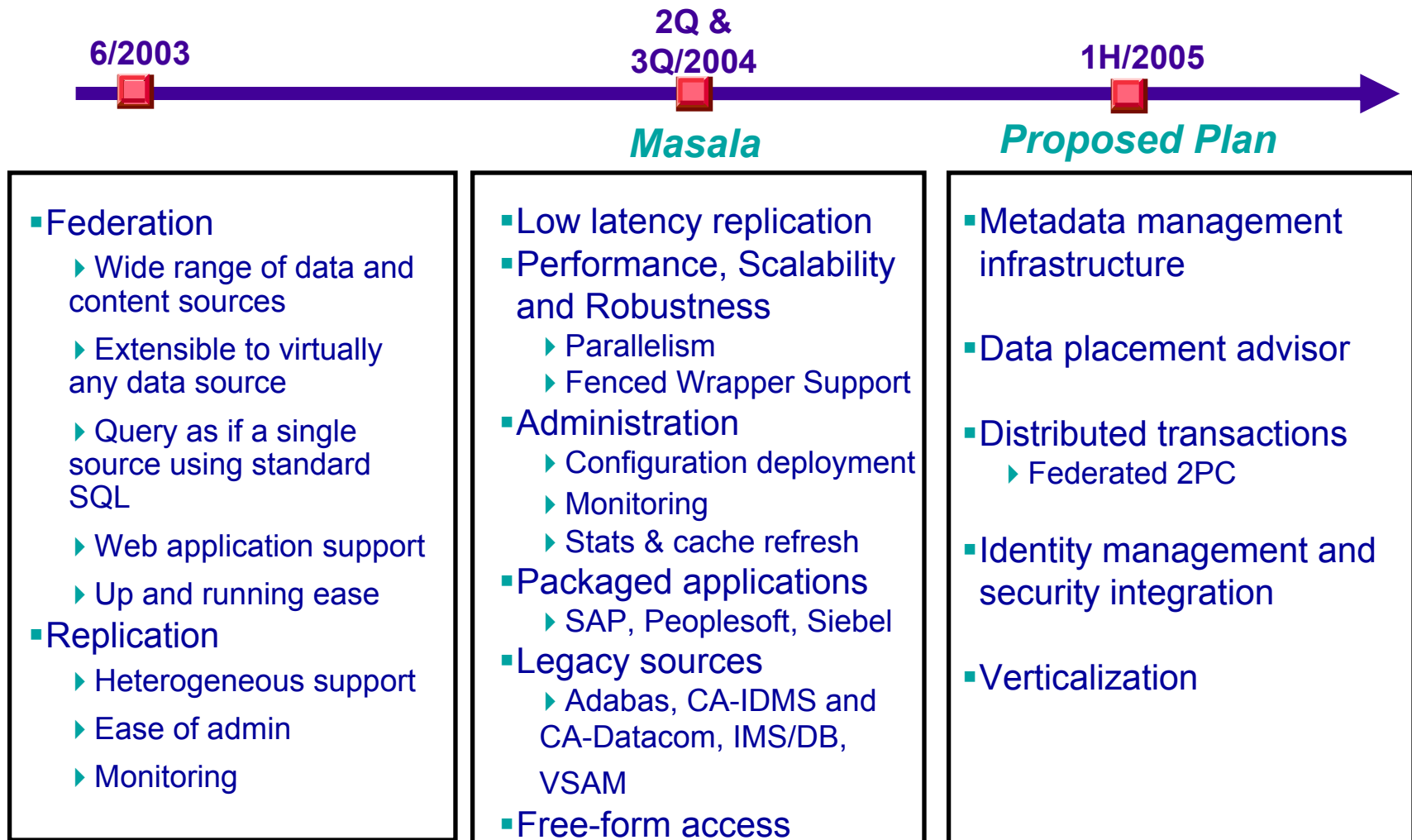- Usage Scenarios

- Roadmap

- Summary

Project Masala

Masala is a rich, flavorful mixture of spices used in Indian cooking

- Symbolizes the harmony of the different technologies required in building an information management infrastructure as well as the diversity of data and content making up information assets throughout an enterprise.

  ▶ Driving ROI
    - Expanding the flexibility, speed, and applicability of the infrastructure.
    - Enriching administrative capabilities
  ▶ Expanding data reach
    - Package applications
    - Complex Web services
    - Legacy sources on mainframe
  ▶ Free-form access
    - Highly relevant enterprise search
    - Easy administration

Technology Roadmap

6/2003 **2Q & 3Q/2004** **1H/2005**

Masala *Proposed Plan*

|		
▪Federation	▪Low latency replication	▪Metadata management infrastructure
▸ Wide range of data and content sources	▪Performance, Scalability and Robustness	
▸ Extensible to virtually any data source	▸ Parallelism	▪Data placement advisor
	▸ Fenced Wrapper Support	
▸ Query as if a single source using standard SQL	▪Administration	▪Distributed transactions
	▸ Configuration deployment	▸ Federated 2PC
▸ Web application support	▸ Monitoring	
▸ Up and running ease	▸ Stats & cache refresh	▪Identity management and security integration
▪Replication	▪Packaged applications	
▸ Heterogeneous support	▸ SAP, Peoplesoft, Siebel	
▸ Ease of admin	▪Legacy sources	▪Verticalization
▸ Monitoring	▸ Adabas, CA-IDMS and CA-Datacom, IMS/DB, VSAM	
	▪Free-form access	

IBM Confidential

Agenda

It's here now
The next generation of data integration software

- Information Integration & on demand computing

- Introducing DB2 Information Integrator V8.1

- Partner Ecosystem

- Usage Scenarios

- Roadmap

- Summary

Driving Visibility - 190 Articles Worldwide YTD
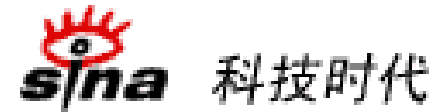
Analyst Support

Meta Group

> ▸ IBM's **big advantage will be its reliance on its own optimization logic, "which is second to none," and its ability to handle structured and unstructured data**, Laney said.

<div align="right">

January, 1 2003

CRN, <u>IBM: Xperanto Rollout to Start In Early 2003</u>

</div>

- Aberdeen

> ▸ **IBM differentiates its EII offering especially by its open standards support, its scalability and performance, its range of data sources supported, and its "deployability"** due to IBM's services strength.

<div align="right">

Wayne Kernochan, May 21, 2003

DB2 Information Integrator: Scope, Power, Services

</div>

- Bloor Research

> ▸ Microsoft notionally embraces the concept but it has done little to implement it, while Oracle's approach runs directly counter to federalism, with the company espousing consolidation (centralization) instead. Thus IBM has to market federation on its own. That shouldn't stop it: **the idea nicely aligns with its On-Demand Computing concept and it will make sense to a lot of users (as well as me).**

<div align="right">

Phil Howard, May 21, 2003

IT-Director.com: Information Integrator Goes Live

</div>

Summary

- Information integration addresses a key pain point for customers

- It is fundamental to On Demand computing

- DB2 Information Integrator leads the industry in Enterprise Information Integration
  - ▶ Structured and unstructured data
  - ▶ Scalability and performance
  - ▶ Integration with partner tools

- DB2 Information Integrator complements existing business intelligence investments
  - ▶ Feed warehouse incrementally for more current data
  - ▶ Extend current BI tools beyond the single warehouse database
  - ▶ Access and correlate real-time events with planned target and historical performance for BAM
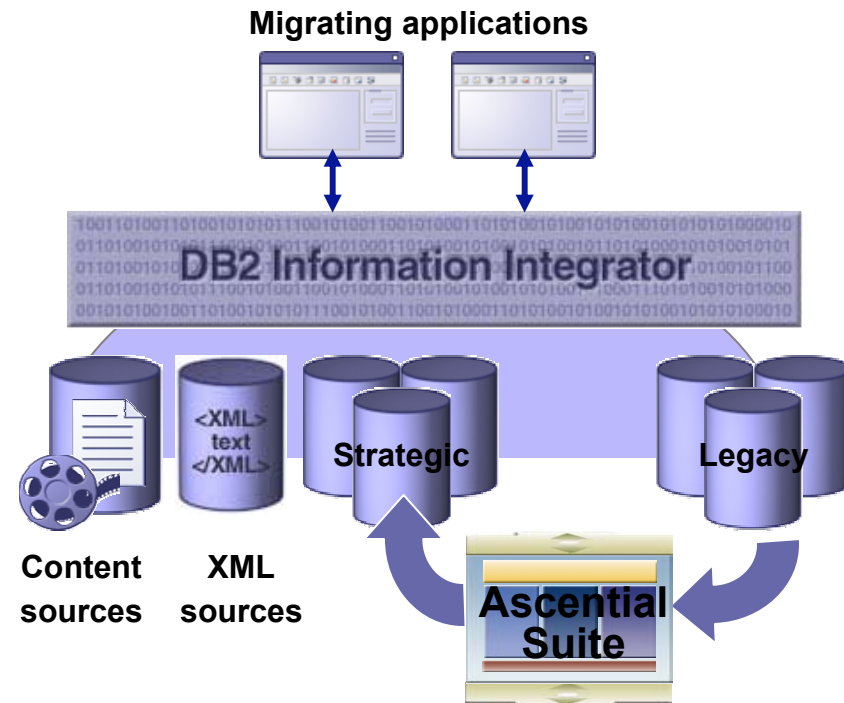 - ▶ Facilitate and speed portal development and deployment

Scenario Backup

Legacy Data Coexistence

Customer Value

▶ Ability to access & integrate legacy & other data sources through a single API

▶ Flexibility to incrementally migrate legacy data to strategic database with minimal impact to new applications

▶ Minimize application skills needed to work with legacy & various data sources thereby reducing application development and maintenance costs
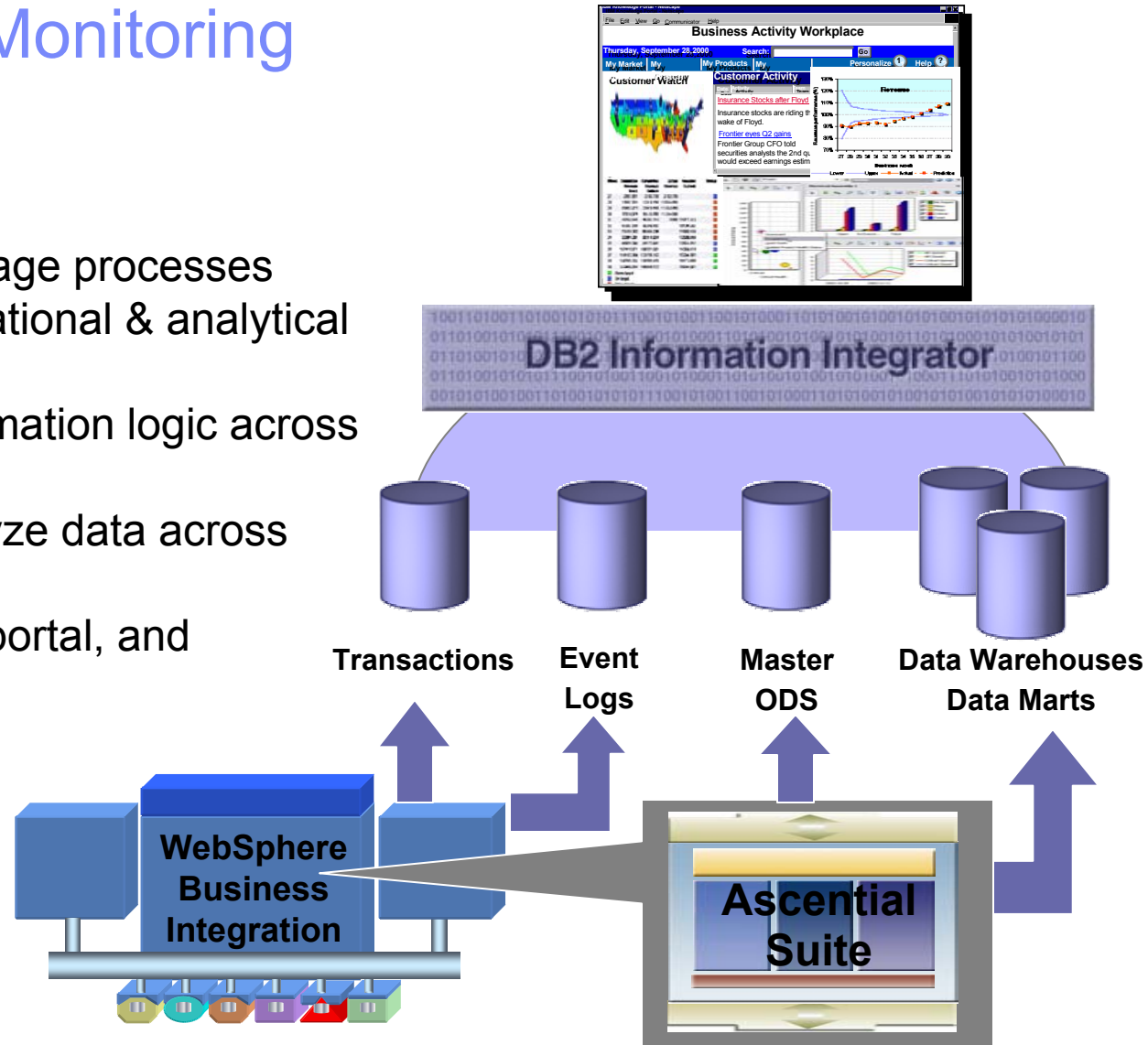
▶ Filter underlying data changes from developers

Migrating applications

DB2 Information Integrator

Strategic

Legacy

Content sources

XML sources

Ascential Suite

Ascential handles:
✓Data profiling
✓Data conversions
✓Data quality
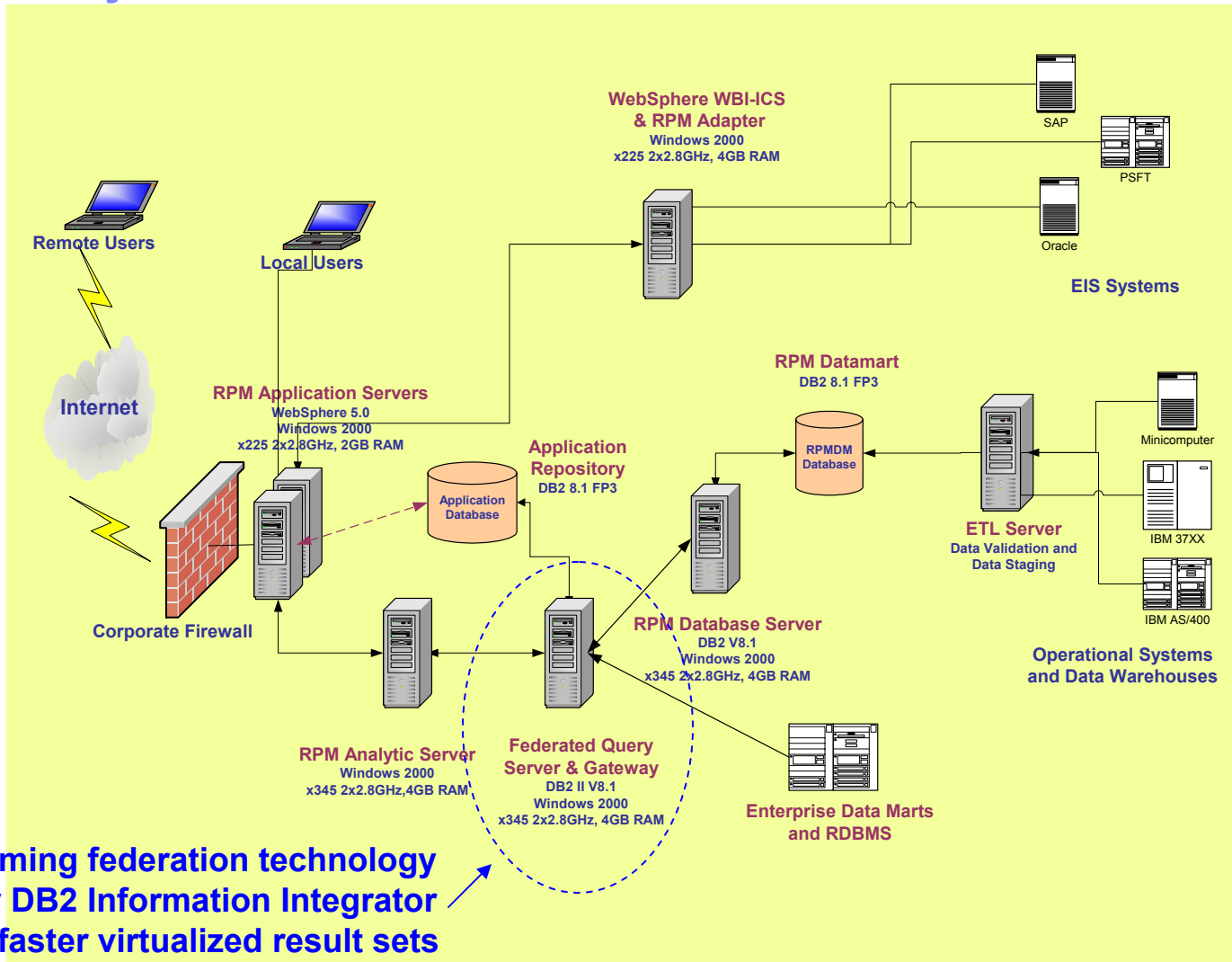✓Bulk data transforms & load

Business Activity Monitoring

Customer Value

- Model, monitor, and manage processes
- Transform data into operational & analytical stores in real time
- Apply consistent transformation logic across all systems
- Access, match, and analyze data across disparate systems
- Use standard analytical, portal, and dashboard tools

DB2 Information Integrator

Transactions   **Event Logs**   **Master ODS**   **Data Warehouses Data Marts**

WebSphere Business Integration

Ascential Suite

SymphonyRPM Scenario

WebSphere WBI-ICS & RPM Adapter
Windows 2000
x225 2x2.8GHz, 4GB RAM

SAP

PSFT

Oracle

EIS Systems

Remote Users

Local Users

Internet

RPM Application Servers
WebSphere 5.0
Windows 2000
x225 2x2.8GHz, 2GB RAM

Application Repository
DB2 8.1 FP3

Application Database

RPM Datamart
DB2 8.1 FP3

RPMDM Database

Minicomputer

ETL Server
Data Validation and Data Staging

IBM 37XX

IBM AS/400

Corporate Firewall

RPM Database Server
DB2 V8.1
Windows 2000
x345 2x2.8GHz, 4GB RAM

Operational Systems and Data Warehouses

RPM Analytic Server
Windows 2000
x345 2x2.8GHz,4GB RAM

Federated Query Server & Gateway
DB2 II V8.1
Windows 2000
x345 2x2.8GHz, 4GB RAM

Enterprise Data Marts and RDBMS

Poor performing federation technology replaced by DB2 Information Integrator To support faster virtualized result sets

Masala Backup

Replication Architecture Extension

- Low latency, high throughput replica management
  - ▶ Reference data management
  - ▶ High availability
- Database event publishing
  - ▶ Application integration scenarios

	Control				Control	
Source	**Capture**			**Apply**		**Target**

Log based

WebSphere MQ
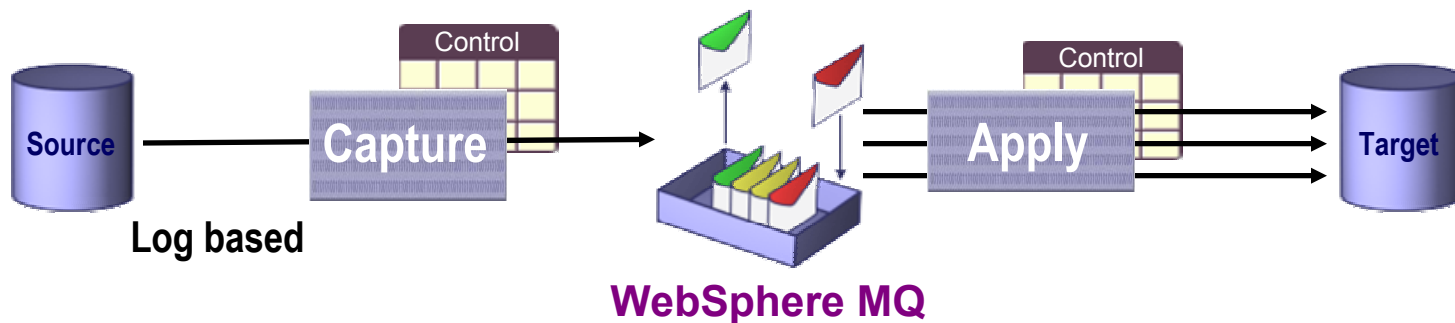
- Each message represents a transaction
- XML format option for publishing
- Highly parallel apply process
- Differentiated conflict detection and resolution
- Staged availability of heterogeneous support

Accessing Packaged Applications

Map resulting WBI BO's to relational schemas to support set oriented querying across packaged applications
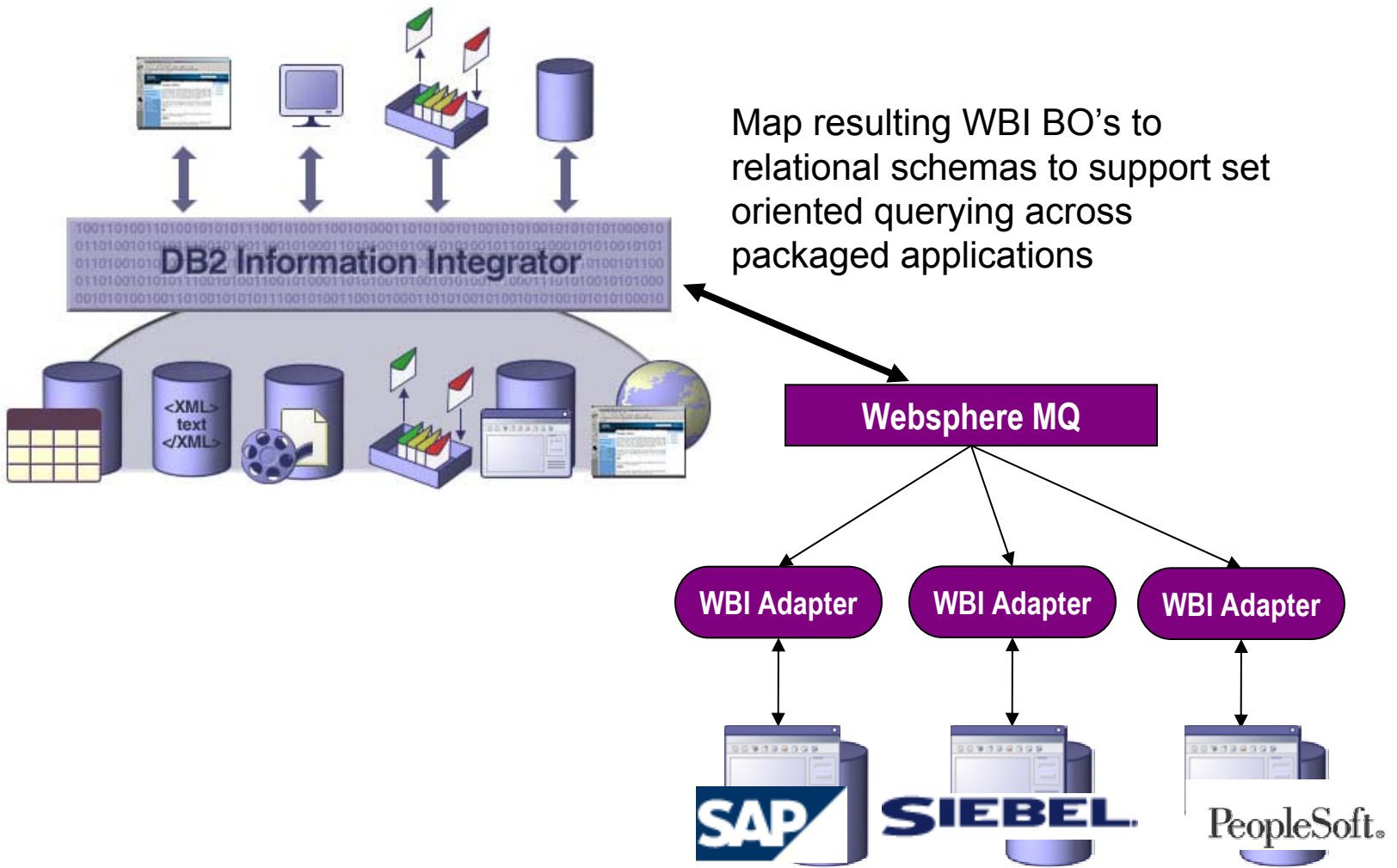
Parallel Federated Queries

- Inter-partition query parallelism (MPP)
  - ▸ Enables parallel join processing of nicknames with local partitioned tables
  - ▸ Enables parallel processing of remote data
- Intra-partition query parallelism (SMP)
  - ▸ Initially limited to local data access only
  - ▸ Will enable SMP cluster if run in DB2 UDB Data Partitioning Feature
- Wrappers can be run in fenced mode, where the fenced process can maintain the persistent connections.
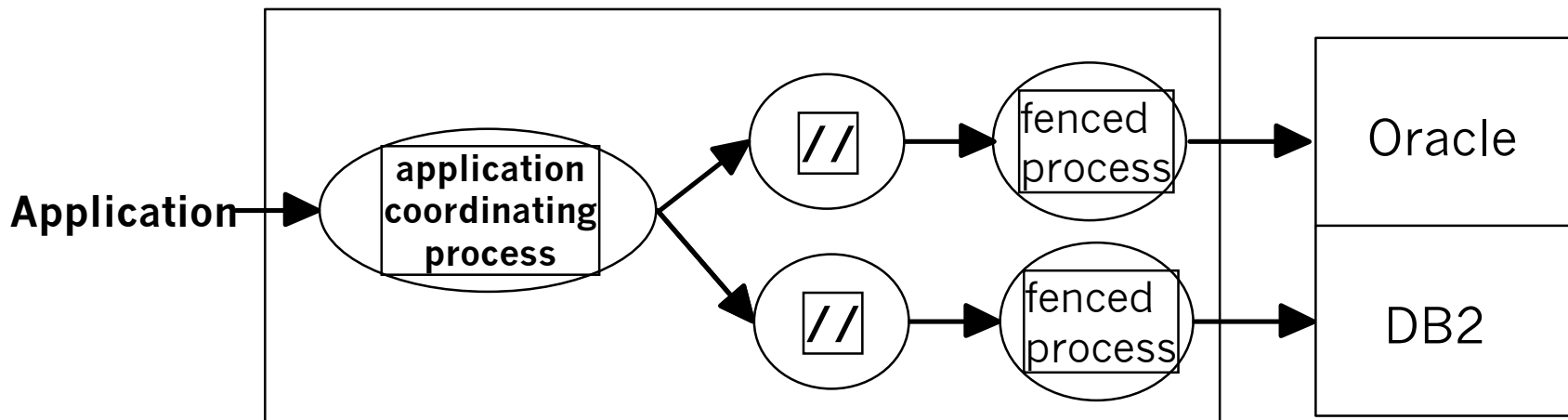
Administration Enhancements

- **Create nickname wizard**
  - ▶ Guides you through the process of setting up a nickname
  - ▶ Selects/creates wrapper, selects/creates server, selects/creates nickname
- **Create cache table wizard**
  - ▶ Guides you through the process of setting up a cache table.
  - ▶ Selects/creates a nickname, creates the cache table and sets up replication
- **Export**
  - ▶ Capture control center actions to a script, edit the script, and deploy the script to the Task Center
- **Monitoring**
  - ▶ Monitor a set of nicknames to detect and report schema changes that may affect the nickname definition
- **Update statistics**
  - ▶ Schedule a task to refresh nickname statistics
- **Server discovery**

Cache Table Wizard

Guided through setup steps

MQT subset defined

Tablespace properties defined

Create Cache Table Wizard

1. Data source
2. Predicate
3. Replication Settings
4. Summary

Predicate

Define a cache for the nickname. Specify the WHERE clause for the cache and the table space information. A materialized query table (MQT) is created when you create a cache.

Cache table name: ANHHONG.N_EMPLOYEE

MQT schema: ANHHONG

MQT name: N_EMPLOYEE_MQT2

WHERE clause

region = 'NORTH_AMERICA'

SQL Assist ...

Cache table space

Estimate required size

○ Use an existing table space

● Create a new table space   N_EMPLOYEE__TS8

Container name N_EMPLOYEE__TS8   Size 10   Unit MB

◀ Back   Next ▶   Finish   Cancel

Metadata Strategy

- Manage metadata
  - ▸ A common repository for metadata assets
  - ▸ Recognize, store, visualize and search relationships between assets
  - ▸ Change management and impact analysis
- Construct software artifacts
  - ▸ Generate new code (Java, DDL, …)
  - ▸ Extract data models and logic flow from legacy apps
  - ▸ Generate artifacts (views, transformations, flows) to integrate data and apps
- Provide runtime context for search and queries
  - ▸ Data cleansing
  - ▸ Exploit ontologies, classification, and inference engines for semantic search

Discover, explore, and manage metadata assets

GUIs
Browse, navigate, visualize, search, invoke services

API layer
Persistent Data Model, Search, Navigate, Insert update delete

Enterprise System Catalog

Artifact Repository

Persistent Repository

Discovery

Mapping

Artifact Generation

Import Export

Find, correlate, collaborate, modify, iterate

lock update, version, group

Search, relate

XSLT

Views

etc

Categorize

Import XML xsd, xslt, wsdl

Extract relationships

Register dependencies

<XML> text </XML>

Administer, detect, notify, report

Export artifacts and register dependencies

Masala technology on w3.ibm.com

A very large corporate Intranet…

- About 25 Million unique URLs
- About 10,000 websites
- Average home page visits per business day - 0.75 Million
- Average html/page views per business day - 2.2 Million
- Average total hits per business day - 12.5 Million
- 7 - 10 Million unique documents
- 20 K per document
- Search speed w/ Cache: 340-450 msec/query

…and growing

Search Demo

w3.ibm.com Search Configuration