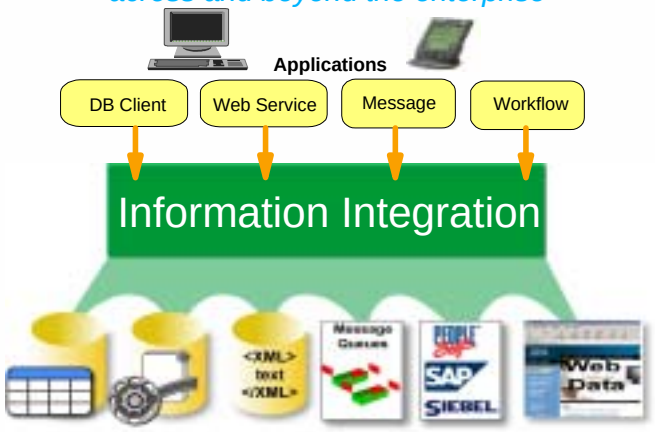
Information Integration

Building a strategic information infrastructure

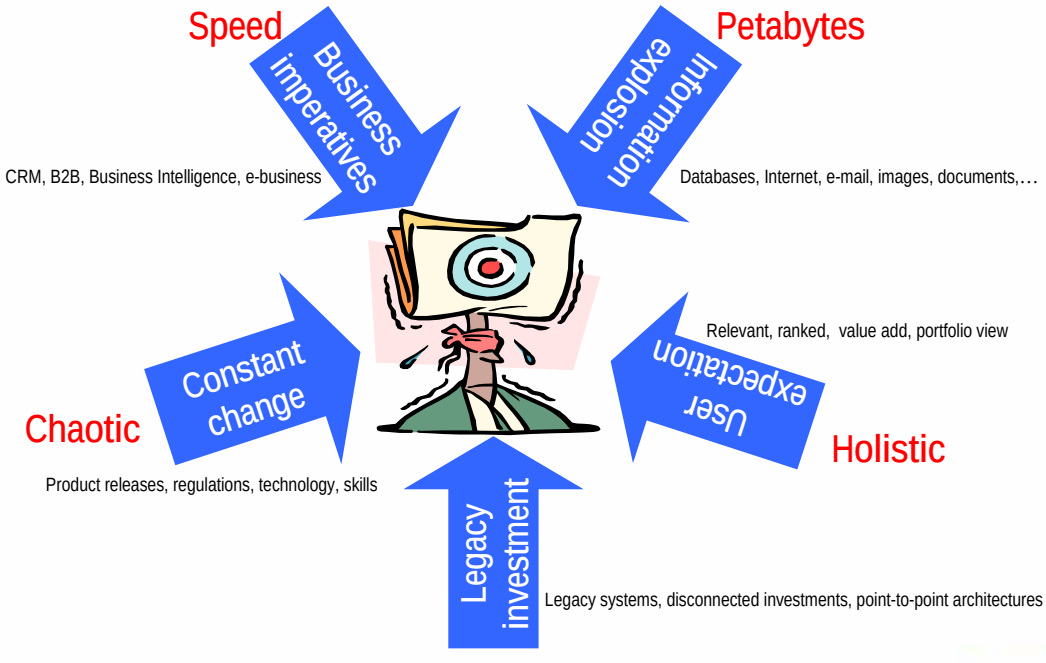
Nelson Mattos, DE and Director, Information Integration

Database Renaissance - Tokyo, October, 2002

Integrating diverse business information across and beyond the enterprise

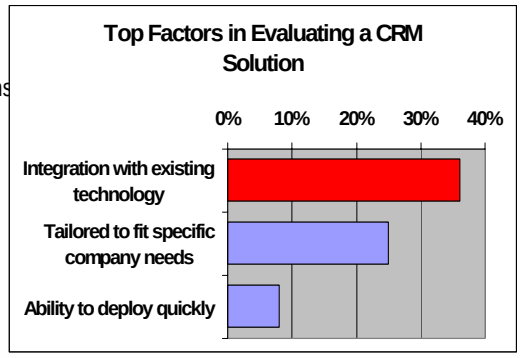


IT Challenges



Business depends on integration

- CRM
  - ►Integrating all customer information cross-enterprise with purchased demographics to personalize all customer interactions for improved customer loyalty
- E-business
  - ►Delivering new services over the Web integrated with core business processes
- Portal
  - ►Enhancing value of Web site with integrated and targeted information to increase site stickiness.
- B2B
  - ► Integrating manufacturing availability and shipping capability with order processing for fastest delivery
- Business Intelligence
  - ►Integrating cross-divisional sales information with demographics to mine buying patterns and increase revenues
- Consolidation
  - ►Reducing the risk and cost of executing mergers and acquisitions through effective integration of incompatible systems and infrastructures
- Operational efficiency
  - ►Integrating process steps eliminating unnecessary work and delay



- Biggest implementation barriers to date:
  - ►End user adoption
  - ►Inability to adequately leverage legacy data
- Top candidates for future CRM expenditures:
  - ►Mobile/wireless , and voice access
  - ►Multi-channel integration capabilities

-- June, 2001 study by YouCentric & WebSurveyor of 250 business execs responsible for e-business and CRMS

Integration Spending Priorities

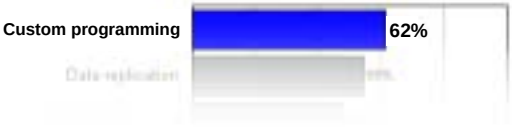
IT Spending Priorities for CIOs

Source: CIO Magazine, March 1, 2002

Requirement for customized integration programming is growing

Source: IBM Customer Surveys, 2001, 2002

Q. What technologies are you currently using to achieve integration?



Q. Which of the following integration approaches does your company use to link your e-business applications?

It is generally estimated that for each $1 spent for a packaged application, customers spend on average $5 to $9 on the labor for integration

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 difficulty altering strategies without potentially creating corporate imbalance.

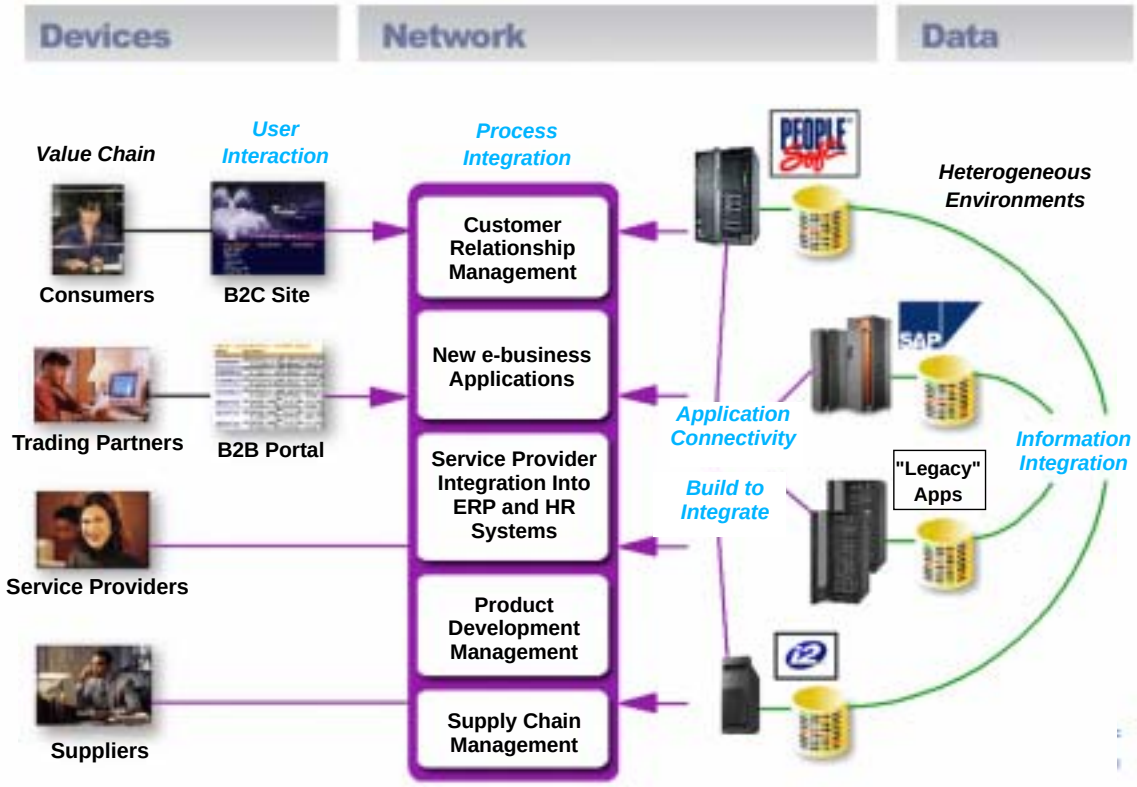
Furthermore, successfully analyzing indeterminate business events requires analytic horsepower and functionality beyond most so-called "business intelligence" tools.

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.

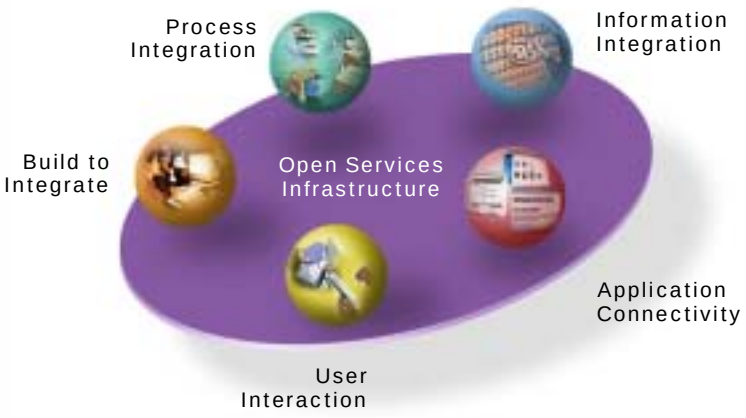
Meta Group Metafax
9/28/2001

Different Types of Integration Required



What is Information Integration?

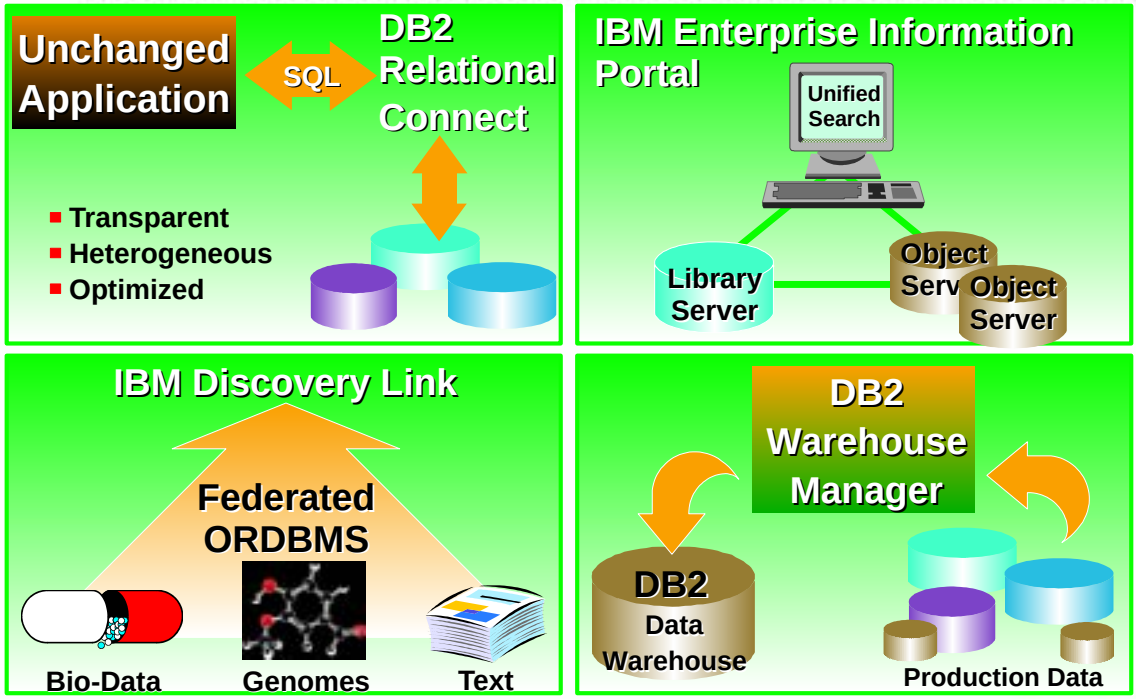
Information integration refers to a category of middleware which lets applications access data as though it were in a single database, whether or not it is. It enables the integration of data and content sources to provide real-time read and write access, to transform data for business analysis and data interchange, and to manage data placement for performance, currency, and availability.



Analysts on Information Integration

- New Enterprise Information Integration (EII) technology will fuel a $7.5 billion market in leveraging Internet information by 2003.
  - Aberdeen, "Enterprise Information Integration: The New Way To Leverage e-Information, May 2002
- "The key to cutting administrative and new application development costs is to implement a common software-infrastructure architecture - and EII infrastructure lies at the core of such an architecture. EII, which puts a one-to-many database "veneer" on existing legacy data sources, has the potential to provide the same business benefits on the back end as a Web services can on the front end - the administrative cost savings, the programmer productivity, and the leveraging of competitive-advantage information to deliver new ties to customers and suppliers.
  - Aberdeen, press release New Enterprise Information Integration Sector to Fuel $7.5 Billion Market by 2003 May 2002
- For clients integrating portals, commerce servers or customer applications to two or more relatively complex and disparate data sources, an Internet Data Integration solution should be seriously considered as a way of not only reducing initial effort, but also creating a leveraged reusable data access component.
  - Giga Information Group, Mike Gilpin, Emerging Internet Data Integration Solutions, Nov 2000

Information Integration Offerings



IBM

Aventis

Vision

- A leader in the discovery and development of innovative pharmaceutical products dedicated to improving life through the discovery and development of innovative products.

Challenge

- Increase drug research efficiency and encourage interdisciplinary cooperation between chemists and biologists. Scientific users require integrated view of chemical & biological information stored in distributed Oracle sources, as well as external non-relational sources.

Solution

- DiscoveryLink provides federated access to Oracle databases and external sources such as genomics and proteomics data across four worldwide locations.
- Sophisticated scientific mining algorithms

Business Value

- Increased research productivity leading to drug innovation and reduced time-to-market

Competitive Value

- " DiscoveryLink allows us to access and mine the physical data in a way never before possible, significantly speeding up the drug discovery and development process." -- Peter Loupos, Global

Aventis

IBM

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

- Relational Connect 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.

Business Value

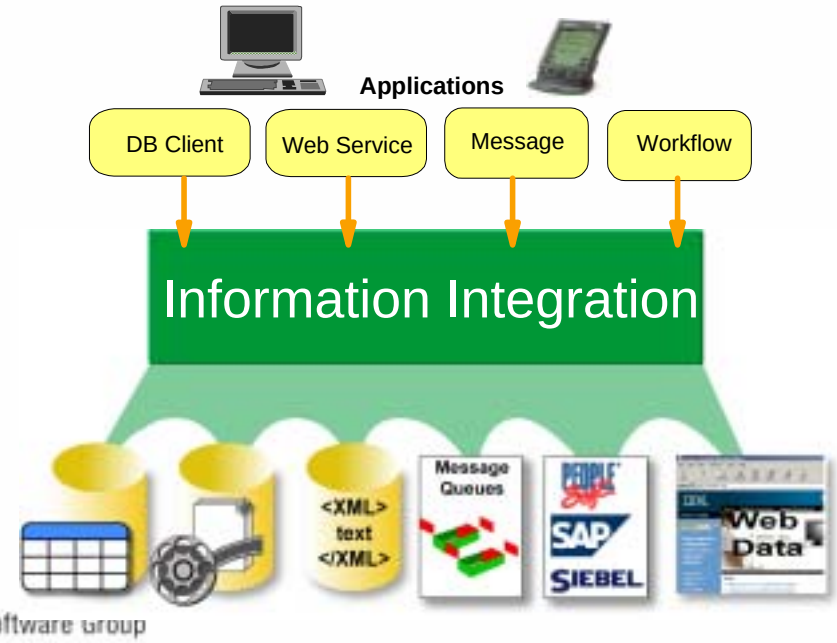
- "Users can provide coherent reporting accessing non-DB2 data sources and discover new ways to meet the information needs of their organizations. And the more information business analysts can incorporate into their views of their company's activities, the more effectively they can steer their companies in the direction of higher profits." - Janet Wood, vice president of business development, Crystal Decisions.

Competitive Value

- "DB2 Relational Connect provides Crystal Reports with the fastest federated querying capability on the market today." -Trevor Smith, Program Manager, Business Development Group, Crystal Decisions

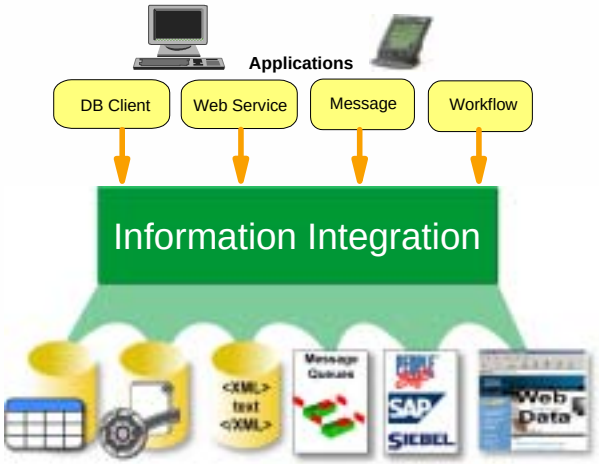
IBM Vision

Integrating diverse business information across and beyond the enterprise



Information Integration

Integrating diverse business information across and beyond the enterprise

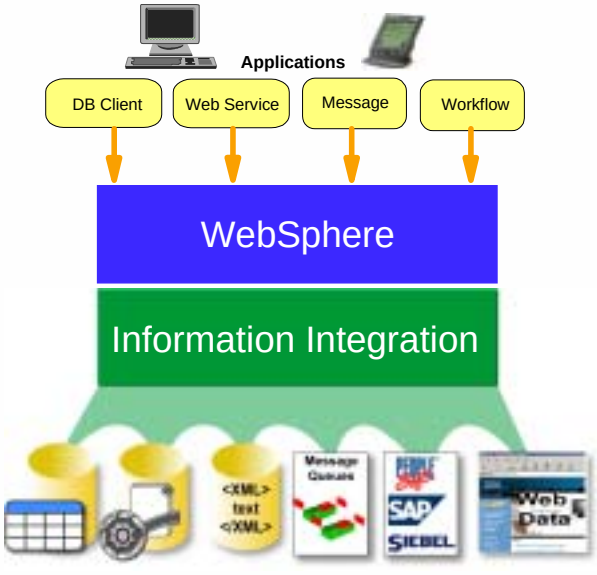


- Data federation
  - Read/write access across diverse data and content sources
  - Database flavor (SQL)
  - Content flavor (CM API)
  - XML flavor (XML API)
- Data placement
  - Caching and Replication over heterogeneous information
- Data transformation
  - SQL, XML
  - Advanced search and mining
  - Metadata management

IBM

Information Integration

Integrating diverse business information across and beyond the enterprise



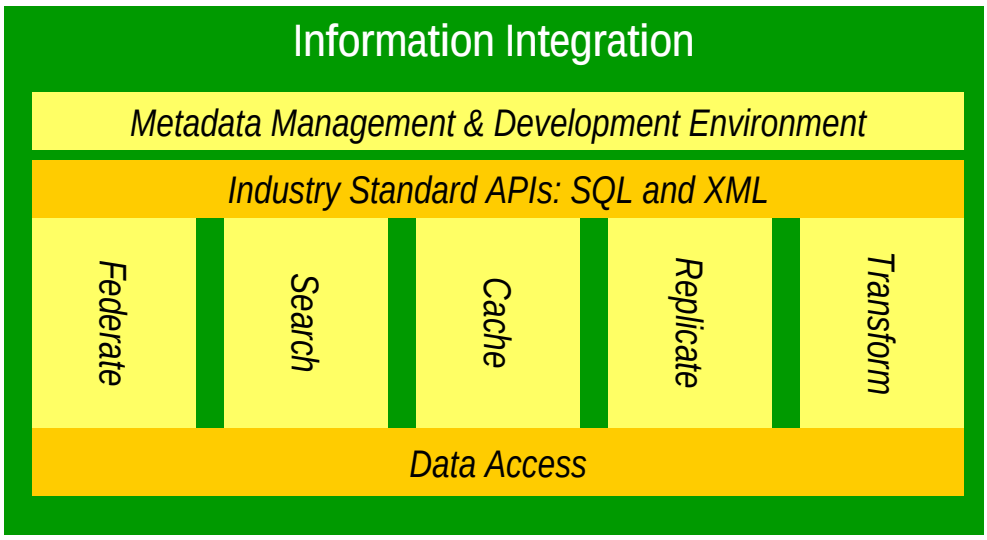
- Data federation
  - Read/write access across diverse data and content sources
  - Database flavor (SQL)
  - Content flavor (CM API)
  - XML flavor (XML API)
- Data placement
  - Caching and Replication over heterogeneous information
- Data transformation
  - SQL, XML
  - Advanced search and mining
  - Metadata management
- Part of a complete business integration solution
  - Best data store for XML data - DB2
  - WebSphere business integration
  - Open platform based on industry standards

IBM

IBM Information Integration

- Integrate data and content without moving the data or changing the platform
  - ►Access diverse and distributed data as though it were in a single database, whether or not it is
- Make more progress, more quickly, at a lower cost
  - ►Reduce coding requirements that integrate diverse data tenfold or more
  - ►Deploy current skills over a greater range of project requirements
  - ►Complete integration projects faster
- With more freedom to do it the way you d like to
  - ►Centralize or cache information for availability or performance needs
  - ►Manage distributed access to data that must remain in place
- With more confidence that it will work
  - ►Rely on 25 years of query optimization research and development for a scalable infrastructure
  - ►Integrate with IBM WebSphere integration technologies for a complete business integration infrastructure
- With better protection of your current and future IT investments
  - ►Base investments on industry standards
  - ►Reduce the need to rip and replace systems to make it work together
  - ►Create a strategic, reusable, and open information integration platform

Core Technologies



Flexible access with standard APIs

- SQL
  - Familiar language with widely deployed skills
  - Rich analytical capabilities
  - Traditional database clients
  - Extensions for XML data (SQL/XML)
- XML
  - Emerging standard for interchange
  - *XQuery - XML Query Language*
    - *Based on a formal algebra*
    - *IBM is co-submitter of XML Query specification (http://www.w3.org/TR/xquery/)*
  - *Exploit unique features within XML data model - hierarchy, sequence*
  - Web services

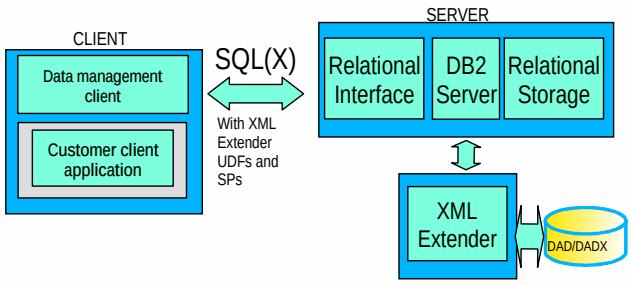
ZapThink

Financial Services Sector to spend $8.3 Billion (US) on XML and Web Services by 2005

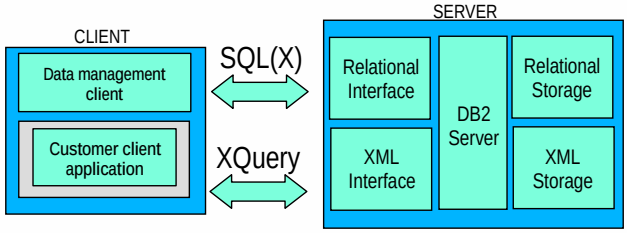
--ZapThink, March 2002

XML Technology

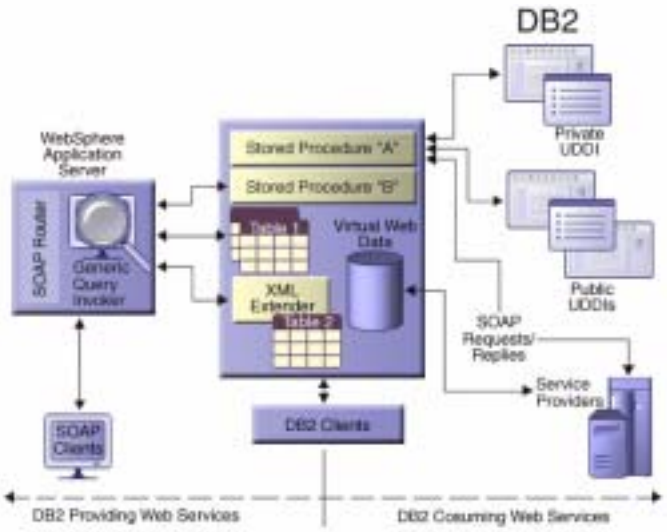
- Object-relational implementation
  - Store, retrieve, compose, decompose, validate, extract, transform
  - Storage options
    - Store intact
    - Store as a collection of columns



- Hybrid XML-relational store
  - SQL or XQuery
  - XML specific storage, query, indexing, privileges, transformation, schema, interfaces, search
  - DB2 engine core attributes: scalability, availability, reliability, manageability

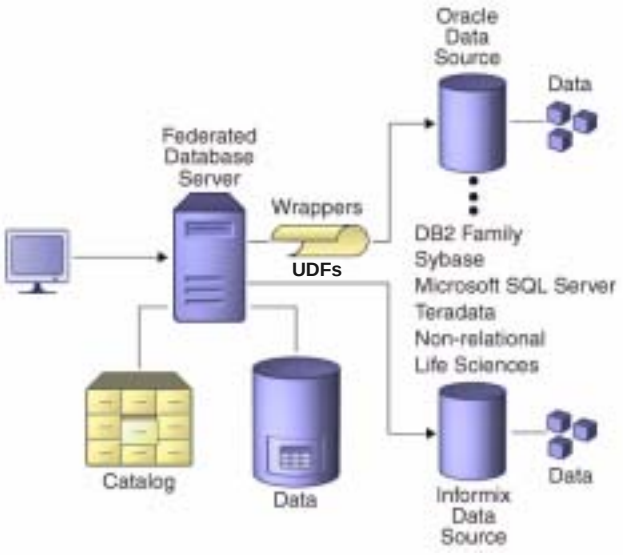


Web Services



- Provider support
  - ►Access resources from Web clients
- Consumer support
  - ►Extend reach of database to nontraditional, real-time data sources
- UDDI
  - ►Catalog Web services for public of private use
- XML Registry
  - ►Manage XML artifacts such as XML schemas, style sheets, DTDs c

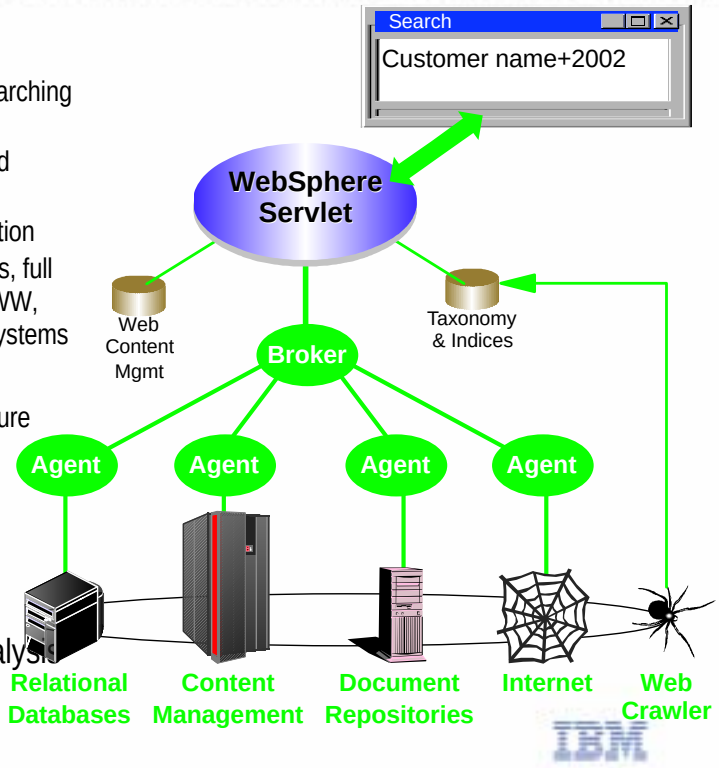
Federation



- Transparency
  - ►Appears to be one source
- Heterogeneity
  - ►Integrates data from diverse sources
  - ►Structured, XML, unstructured, messages, Web, c
- Function and Extensibility
  - ►SQL/XQuery plus backend specific functions
- Autonomy
  - ►Non-disruptive to data sources
- Performance
  - ►Distributed optimization

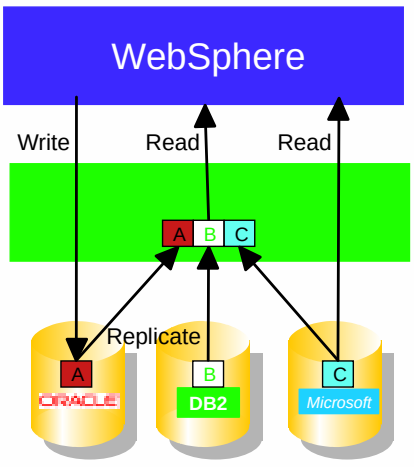
Extended Search: Unstructured Access

- IBM Extended Search
  - ►Brokered search architecture for searching thousands of existing data sources
  - ►Results are aggregated, ranked, and returned in a single hit list
  - ►Easily embeddable into any application
  - ►Lotus databases, document systems, full text indexes, e-mail, directories, WWW, syndicated content, relational, file systems
- Combined with text mining
  - ►Summarization, categorization, feature extraction, c
- Combined with federation
  - ►Generated search arguments
  - ►Sophisticated ordering
  - ►XML document generation
- Institute for Search and Text Analysis
  - ►Major investment area



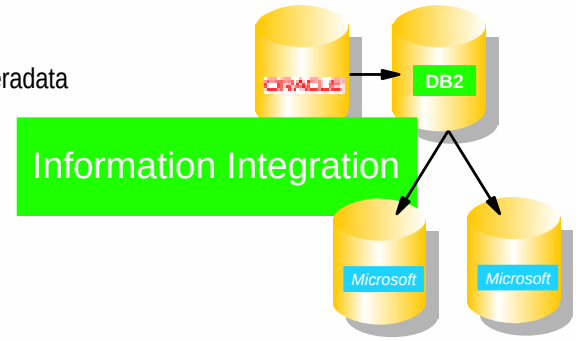
Multi-tiered Caching

- Improve query performance and availability
- Administrator defines Materialized Query Table
  - ►Precomputed or frequently used values
  - ►Any data from the federated system
  - ►Application indicates ability to use cache
  - ►Implicit or explicit use
- Developer enables cache use
  - ►If enabled, reads are handled from the cache, writes passed through to the source
  - ►If not, reads and writes passed through to source
- Cache refresh managed:
  - ►Manually
  - ►By replication
  - ►Various refresh strategies under design
- Flexible caching topologies supported



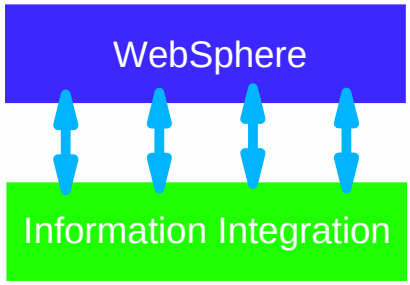
Replication

- Applications
  - Warehouse and ODS applications
  - Consolidation and distribution
  - Application integration
  - Availability management
- Heterogeneous replication
  - DB2, Oracle, Sybase, Informix, Microsoft, Teradata
- Table-based or transaction-consistent
- Point-in-time or continuous operation
- Embedded transformation
- Key investments
  - Transaction-based replication over MQ
  - Publish changes over MQ



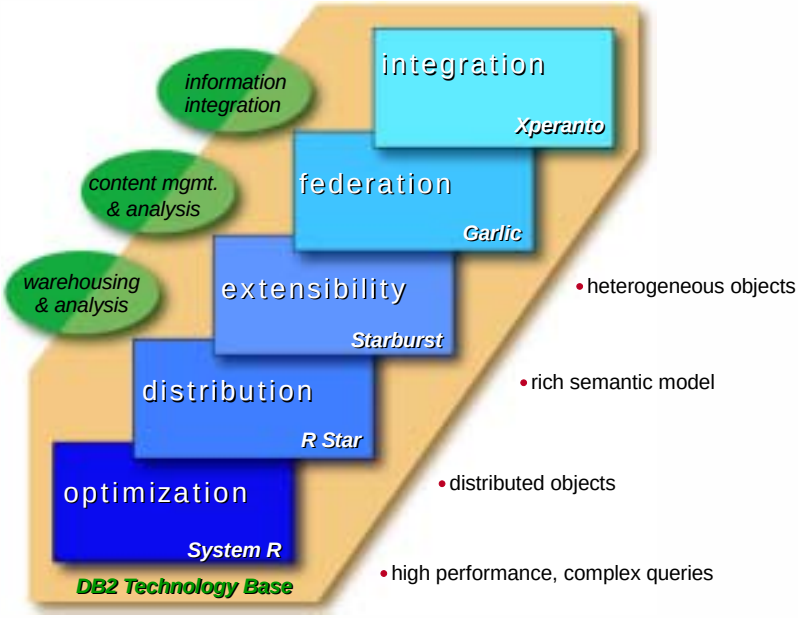
WebSphere Integration Points

- MQSeries
  - Publish, Read (nondestructive), Retrieve (destructive)
  - Listen for events
- Workflow
  - Listen for events
  - Microflow integration
- Crossworlds
  - Exploiting CW connectors for federation
  - CW exploiting federation
  - Combining set oriented and event oriented for scalability
- Portal
  - Platform for integrating data and content
- Tooling
  - Single set of tools across all integration offerings
- ...



Data Management Technology Evolution

- Twenty-five years of research and development

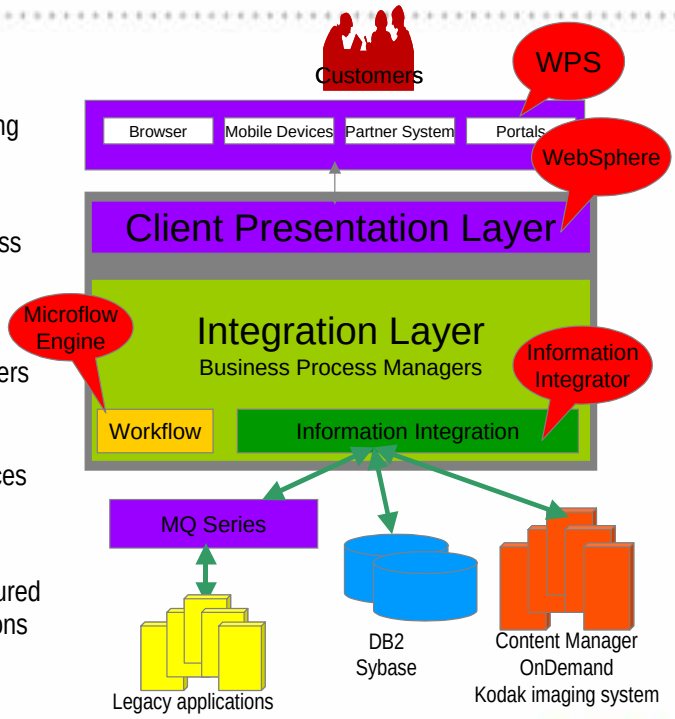


Xperanto Demo

- Raise awareness of the information integration challenges in the industry
- Position the importance of key technologies in addressing those challenges:
  - ►XML and XQuery ? lingua franca for interchange
  - ►Web services ? extends the reach of the database to nontraditional data
  - ►Federation ? key for data volume, rate of change, and diversity
  - ►Text search ? core technology for XML-based world
- Refresh coming this fall
- http://www.ibm.com/software/data/developer/demos/xperanto/

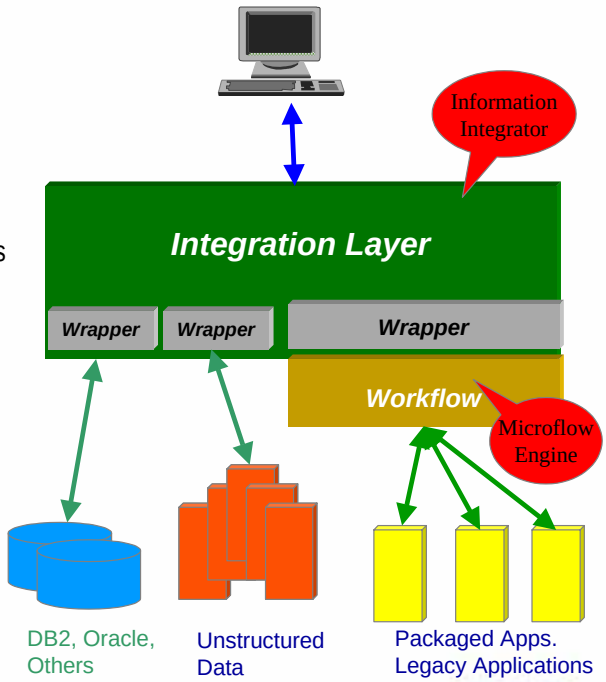
Insurance

- Business Problem
  - Operations efficiency and improving customer loyalty
- 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
- Technical Requirements
  - Access to structured and unstructured data and core insurance applications
  - Interfaces to workflow, messaging and replication



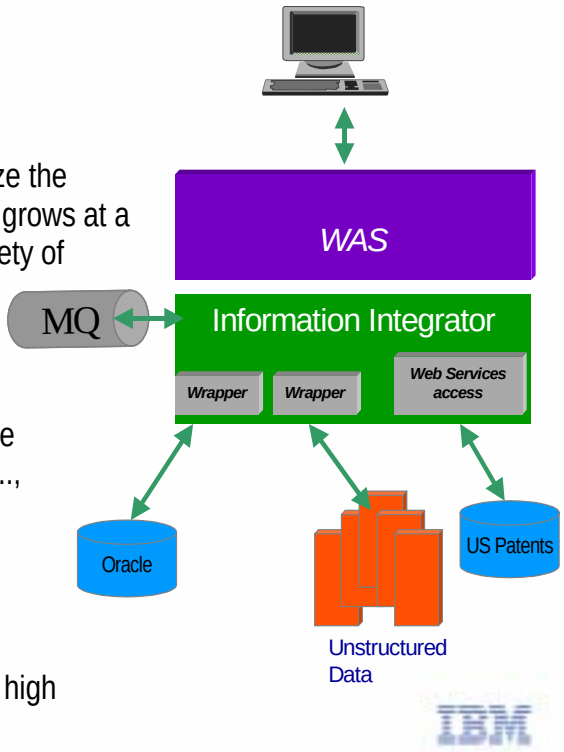
Manufacturing

- Business Problem
  - Consolidation
- Challenge
  - Improved integration of islands of information and back end systems
  - Reduced impact to existing infrastructure
  - Increase flexibility of access and analyses of existing information
- Technical Requirements
  - Access to structured and unstructured data
  - Access to applications (legacy, app packages)
  - Integration (wrapping) of workflow and messaging



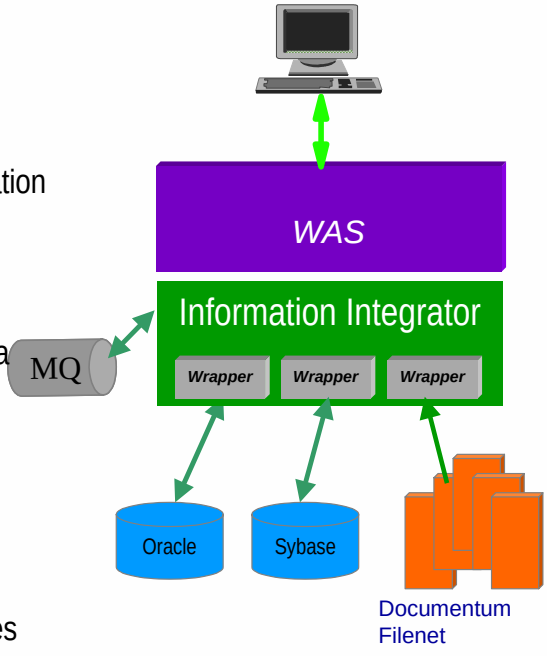
Science

- Business Problem
  - Operations efficiency
- Challenge
  - Provide researchers the ability to analyze the results of their experimental data, which grows at a phenomenal rate and comes from a variety of sources and formats
- Technical Requirements
  - Read/write federation to Oracle systems
  - Federation to modeling output files: some semi-structured, some unstructured (e.g.., Gaussian)
  - MQ Integration
  - Publish XML data
  - XML support for XSD Schema
  - Fast Replication for data placement and high availability



Financial Services

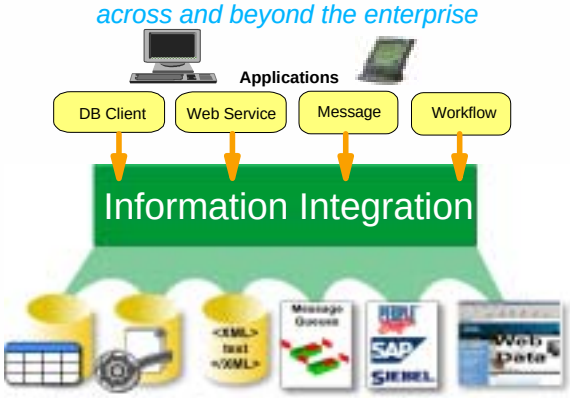
- Business Problem
  - Integrating the virtual enterprise
- Challenge
  - Develop corporate information integration architecture
  - Huge amount of information
  - 100s of GBs in Sybase (FPRS)
  - legacy applications platforms and data
  - Oracle/Sybase/Informix databases
  - structured and unstructured data
  - Documentum/Filenet
  - and more..
- Technical Requirements
  - Access to heterogeneous data sources
  - XML transformations and repository for XML artifacts



Summary

- Information integration
  - ►a key component of IBM ŝ business integration infrastructure
- Information integration enables integration of data and content sources for
  - ►Real-time read and write access
  - ►Transformation for business analysis and data interchange
  - ►Data placement for performance, currency, and availability
- IBM is well positioned to lead the market
  - ►25 years R&D in query optimization
  - ►Complete portfolio of integration offerings
  - ►Customer experience and credibility

Integrating diverse business information across and beyond the enterprise



IBM