The Information Integration Platform

IIS Product and Technology Vision & Roadmap

Bob Zurek
Director, Advanced Technologies and Product Strategy
Information Integration Solutions

IBM Software Group
Agenda

- Introduction to the Information Integration Platform
- The Evolution of the IIS Platform: Hawk and Serrano
- A Glimpse into the Future
The IBM WebSphere Information Integration Platform
*Delivering information you can trust*

**Service Oriented Architecture**

- **Understand**
  - Discover, define, model, and govern information quality and structure

- **Cleanse**
  - Standardize, merge, and correct information

- **Transform**
  - Transform and enrich information

- **Federate**
  - Virtualize access to disparate information

**Integrated Metadata Management**

**Parallel Processing**

- Data → **Connect** → Content

- Access, publish, and replicate information
Connectivity

Developers

Connectivity

Testers

IT Admin

WebSphere. Information Integration Portfolio
Direct, native access to relevant sources bundled with each product in the portfolio

WebSphere. Information Integrator Replication Edition
Changed data and event-based publishing of data

WebSphere. Information Integrator Event Publisher Edition
Consolidation, synchronization, and distribution across disparate databases
Understanding

Data Analysts

Subject Matter Experts

Business Users

- Analyze data, and report and monitor based on integration and quality rules
- Perform structure-driven reporting and annotation
- Provide text analysis and semantic search
Cleansing

Data Analysts

Subject Matter Experts

Cleanse

Standardize source data fields
Correct, certify, and enrich data values
Match and link records together
Select the best values to merge into a final record
Transformation

- Architects
- Project Managers
- Developers
- Testers

IBM Information Integration Live! – Move Faster

High-volume, complex data transformation and movement

Embeddable, in-line validation and transformation of complex data types
Federation

Architects
IT Admin
Developers
Testers

WebSphere. Information Integrator
Classic Federation
SQL-based federation across mainframe
databases and files

WebSphere. Information Integrator
Standard Edition
SQL-based federation across databases, files,
services, and packaged applications

WebSphere. Information Integrator
Content Edition
Content-optimized federation across
repositories, files, and collaboration and workflow
systems
Platform Integration and SOA

- Deploys Services Codelessly In Minutes
- Enables Shared Data Integration Services
  - Shared Business Objects
  - Shared Transformations
  - Insulating Service Consumers From Complexity
- Provides Flexible Reuse
  - Logic Reuse Across Batch And Dynamic Access
  - Compliant With Enterprise Standards
- Provides A Fast, Resilient Foundation For Complex Data Integration Processing
  - Load Balancing
  - Parallel Processing
  - Clustering And Fault Tolerance
Information Integration Services

Providing a Roadmap to Project Success

- Center of Excellence for Data Integration
- Iterations® Methodology
- Architecture Design/Support
- Education and Mentoring
- Enterprise Data Models
- Virtual Consulting
- Certification
Agenda

- Introduction to the Information Integration Platform
- The Evolution of the IIS Platform: Hawk and Serrano
- A Glimpse into the Future
The Hawk & Serrano Combination

The Next Generation Of IBM’s Leading Information Integration Platform

- **Dramatically Simplifies Information Integration**
  - Unifies Multiple Aspects Of The Platform
  - Introduces Breakthrough Productivity Benefits

- **Enhances The Confidence, Quality And Precision Of Information**
  - Inspires Confidence In Information
  - Improves Information Quality And Precision

- **Addresses Growth & Sustainability Requirements Of Enterprises**
  - Expands Customer Value Through Information Integration Platform Synergies
  - Provides Access To More Sources And Types Of Information
  - Scales To Meet The Needs Of Business Scenarios Involving Any Volume
  - Protects The Privacy And Integrity Of Information
  - Provides An Open, Extensible Architecture
Product Roadmap

The Serrano Project

- New release of WebSphere Information Integrator products:
  - WebSphere Information Integrator Content Edition
  - WebSphere Information Integrator OmniFind Edition
- Introduction of new products:
  - Rational Data Architect

The Hawk Project

- New release of the WebSphere Data Integration Suite products and core technologies:
  - WebSphere DataStage
  - WebSphere QualityStage
- Introduction of new products:
  - WebSphere Information Analyzer (Project Sorcerer)
  - WebSphere Business Glossary
Expanded Capabilities and Product Integration

*Enhance Serrano And Hawk Beyond Original Plans*

- **Enrichment of the Information Integration Platform**
  - Extend Hawk Translations To Address The Broader Global Market
  - Extension of SOA (RTI) To Support II Federation
  - Connectivity Between The Suite And II Federation & Classic Federation
  - Integration Between The Suite And Event Publisher
  - DataStage / Classic Federation Native Z/Series Connectivity

- **Integration Across the Broader IBM Portfolio**
  - Tight Linkage Between Datastage TX & Websphere BI Message Broker
  - Unifying The Metadata Repositories From IBM & Ascential
  - Integration Of QualityStage With EAS For Postal Coverage
Additions to the Hawk Project Plan

*Feedback from Hawk Beta Preview & Enterprise Architecture Board*

- **Enterprise Data Integration Suite**
  - Added Web-based Interfaces
  - Added Resource Estimation And Measurement For PX
  - Added Teradata “Frictionless” Connector
  - Support For Silent Install
  - Expanded Enterprise Security

- **Websphere Qualitystage**
  - Expanded Global Postal Support
  - Added 20 Countries For Geolocation
The IBM WebSphere Information Integration Platform

Enhanced Capabilities with Hawk/Serrano

Service Oriented Architecture

Understand
Discover, define, model, and govern information quality and structure

Cleanse
Standardize, merge, and correct information

Transform
Transform and enrich information

Federate
Virtualize access to disparate information

Integrated Metadata Management

Parallel Processing

Data
Connect
Content

Access, publish, and replicate information
Information Virtualization Is The Enabler for SOA

Enhanced Capabilities with Hawk/Serrano

- **Information virtualization** - the ability to access, control and manage data from multiple disparate sources without knowing or caring where the data actually resides.

- **Database virtualization** - allows applications to access databases without being aware of the server that the databases are residing on at any given time (EII is about database virtualization).
Benefits Of Information Virtualization

Enhanced Capabilities with Hawk/Serrano

- More control and trust
  - Full support for cleansing, profiling, modeling, transformation, replication and other information services provides a complete interoperable portfolio of options for managing data

- Reducing cost
  - The structure, location or size of the databases and data files can change without impacting production applications

- Saving time
  - Reusable services can be created that transparently combine data from multiple sources in the background
The IBM WebSphere Information Integration Platform

Enhanced Capabilities with Hawk/Serrano

Service Oriented Architecture

Understand
Discover, define, model, and govern information quality and structure

Cleanse
Standardize, merge, and correct information

Transform
Transform and enrich information

Federate
Virtualize access to disparate information

Integrated Metadata Management

Parallel Processing

Data Connect Content
Access, publish, and replicate information
The IBM WebSphere Information Integration Platform

Enhanced Capabilities with Hawk/Serrano

Hawk/Serrano
Designed for the Enterprise

- Superior User Experience
  - Designed with Our Users
- Information Services Framework
  - Internal Services Architecture
- Enterprise Metadata Infrastructure
  - Scalable, Standards Based, Services Enabled
- Common Parallel Processing Engine
  - Integrated, Full Function, Expandable
- Frictionless Connectivity
  - Intuitive Configuration & Optimization
The IBM WebSphere Information Integration Platform

**Enhanced Capabilities with Hawk/Serrano**

- **Information Services Framework**
  - Common Services Infrastructure across Applications
  - Simplify Administration and Deployment Across the Enterprise
  - Easily Integrate with 3rd Party Operational and Management Systems
  - Open, Standard Access to IBM IIS Platform Services

- **Enterprise Metadata Infrastructure**
  - Open & Extensible
  - Highly Scalable
  - Enterprise-Class Management & Analysis Services
  - Geared for Business
The IBM WebSphere Information Integration Platform

**Enhanced Capabilities with Hawk/Serrano**

**Mozart User Experience**

- Rich Client
  - Information Analyzer
- Web Client
  - Administration
  - Operation
  - Reporting Console
  - Business Glossary

**Service-Oriented Architecture**

**Integrated Metadata Management**

**Parallel Processing**

**Common Services**
- Metadata
- Repository
- Product

**Other Services**
- Repository Management
- Persistence Management

**Repository**

**Dynamic Repository**

**Operational Repository**

**Rich Client**

**Web Client**
IBM Information Integration Live! – Move Faster

Understanding

Hawk/Serrano New Product Introductions

Architects

Data Analysts

Subject Matter Experts

Business Users

Rational
Data Architect
Structure- and data-driven data modeling and management

WebSphere
Information Analyzer
Analyze data, and report and monitor based on integration and quality rules

WebSphere
Business Glossary
Perform structure-driven reporting and annotation

WebSphere
Information Integrator OmniFind™ Edition
Provide text analysis and semantic search
IBM Rational Data Architect

Hawk/Serrano New Product Introduction

- **Gain understanding** with discovery of dependencies between disparate data sources
- **Simplify design** with the only data modeling tool for your relational and federated data structures
- **Reduce risk of change** with impact analysis, evaluation of alternative solutions, and synchronization of changes
IBM WebSphere Information Analyzer

Hawk/Serrano New Product Introduction

- Intuitive User Experience
  Maximizes Business Understanding

- Comprehensive Data Profiling, Analysis, And Quality Monitoring
  Ensures Confidence

- Shared Metadata For Data Discovery
  Reduces Risk In Integration Development

- Always Parallel Processing
  Accelerates Data Analysis
  Insuring On-time Project Delivery
IBM WebSphere Business Glossary

Hawk/Serrano New Product Introduction

- A web-based tool for authoring, managing & sharing business metadata
- Aligns the efforts of IT with the goals of the business
- Provides business context to information technology assets
- Establishes responsibility and accountability
OmniFind – UIMA Text Analytics Capabilities

_Hawk/Serrano Enhanced Capabilities from IBM Research_

Unstructured Information Management Architecture

- Provides Standard Interface For Adding **Text Analytics To Any Application**

- **Enables Interoperability** Of Analytics Solutions And Enterprise Applications

- Leverage Industry Specific And **Best Of Breed** Annotators From Partners
IBM WebSphere QualityStage

**Hawk/Serrano Enhanced Capabilities**

- Single design environment increases functionality and reduces development time
- Analysts use ‘data driven’ experience for dynamic tuning of match to improve results
- Service oriented provides ‘continuous’ quality delivers confidence in your data
- Always parallel architecture improves execution time
IBM WebSphere DataStage

**Hawk/Serrano Enhanced Capabilities**

- The Power Of Meta Data For **Increased Productivity**
  - Graphical Impact Analysis
  - Advanced Find
  - Job Diff

- Most Comprehensive & Easiest Transformation Capabilities
  **Simplify Complexity**

- **Enterprise-class** Deployment, Management, And Administration
IBM WebSphere DataStage TX v8.0

Recently Released – Improved Productivity & Flexibility

- **Accomplish More**
  - Quicker Debugging
  - Faster Execution
  - Higher Production Availability
  - Efficiency Gains In Resource Utilization

- **Deep Integration With Enterprise Environment**
  - Simpler Data Structures
  - Larger Data Volumes
  - Flexible Connectivity
Agenda

- Introduction to the Information Integration Platform
- The Evolution of the IIS Platform: Hawk and Serrano
- A Glimpse into the Future
A Glimpse into the Future

- **Complete Transition into the Mozart User Experience**
  - Integrate Design Experience into Mozart Framework

- **SOA Expansion: Integration with Business Processes**
  - Information Services Participation in Business Process Transactions

- **Expansion of Metadata Capabilities**
  - Tools for Technical Metadata Users
  - Federation of Repositories

- **Facilitating the Collaboration of Business and IT**

- **Leveraging the Strength of IBM Research**

- **Automation and Optimization**
Our Focus: Delivering Value to Customers

- **Powerful, Comprehensive, Single-vendor Platform**
  - Rapid Deployment, Ease Of Integration, And Asset Reusability
  - Ease Of Maintenance And Support, With Clear Accountability
  - Reduced Cost, Complexity, And Project Risk

- **Readiness For Growth And Change**
  - Adaptability To Growth In Volume And Velocity
  - Adaptability To A Changing Spectrum Of Data And Content

- **A Complete, Customer-focused Solution**
  - Integration Methodology And Best Practices
  - Architecture, Design And Implementation Services
  - Partnership With Customers For Accelerated Success
Up Close & Personal….

- **Product Spotlights**
- **Product Demo Room**
  - Monday 10:30 – 12:30  Key Largo
  - Monday 2:00 – 5:30   Key Largo
  - Tuesday 10:00 – 12:00 Key Largo
  - Tuesday 1:30 – 5:00   Andros A

- **Partner Pavilion and Meet the Engineers Reception**
  - Monday 5:30 – 7:30   Events Center B

- **Partner Pavilion and Roundtables**
  - Monday 7:00 – 7:45   Events Center B
  - Monday 12:30 – 2:00  Events Center B
  - Tuesday 7:00 – 7:45  Events Center B
  - Tuesday 12:00 – 1:30 Events Center B
Thank You