IBM InfoSphere Information Server
What’s new in Version 8.5

<your name, title>
IBM InfoSphere Information Server is the market-leading data integration platform

- Complete data lineage insight
- Scalable to manage exploding data volumes
- Operational integration 24x7
- Aligns business and IT objectives
- Connects to existing data sources
Three Key Marketing Themes for Information Server 8.5

<table>
<thead>
<tr>
<th>Improve time to value</th>
<th>Address expanding enterprise requirements</th>
<th>Accelerate data integration projects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Achieve new levels of data integration to address a rapidly growing number of formats and back-end systems</td>
<td>• Revolutionary XML handling capabilities • Enhanced transformation handling stages and lifecycle management of ELT-generated logic • Advanced metadata flows and automation across tools • Reliable, real-time trickle feeds to ETL process to alleviate batch windows</td>
<td>• Blueprint Director is a new capability to further strengthen the alignment of business and IT allowing you to govern your information projects • Simplified installation, configuration, and patching tools • Direct tooling integration with standard source code control systems • Automatic generation of mapping specifications from existing job designs to accelerate analyst collaboration • Enhance collaboration between business and IT through data validation, publishing of data quality information and viewing lineage from your data discovery effort</td>
</tr>
<tr>
<td>Address expanding enterprise requirements</td>
<td>Meet the most demanding availability, scalability and performance needs of large enterprises</td>
<td>• High availability and clustering support • Improved performance for large development teams • Improved end-to-end XA support • New grid toolkit to enable simplified scaling to grids</td>
</tr>
</tbody>
</table>
Blueprint Director

The GPS for your Information Project

- Create and manage actionable blueprints of your information project
- Accelerate information centric projects by leveraging templates with methods – e.g. warehousing, master data – and customize your best practices
- Establish business-driven development by
  - aligning business & IT views
  - creating a consistent end-to-end design from business requirements
- Navigate through your information project by managing the information roadmap & its evolution over time

Information project leadership team
(stakeholders, bus. Analysts, stewards, specialty architects, etc.)
Overview of the interface

Palette free form “sketching” elements

Diagram for a blueprint
- Method browser (displaying method content)
- Asset browser (browsing metadata repository)
- Glossary explorer (showing glossary tree view)

- Outline (zoom in/out view)
- Blueprint explorer (shows tree view of the elements in the blueprint)

Context specific property view
Expose quality rules to business users

Allow LOB to review & validate the quality of data relative to key business elements (e.g. product)

Meet acceptable compliance levels and data hygiene requirements

Features:

- View business rules associated with a term
- View the percentage of data that meets the business rule acceptance level
- View valid values, value ranges (min/max), and summary of data domain contents

Empower the Business User in Data Quality Assessment Programs
Business Glossary – What’s New in 8.5?

Extend the Collaboration between Business and IT

- Expose business terms to data modelers
- Ensure data modeling is driven by business objectives and thru collaboration with LOB
- Minimize communication latency and implement projects faster

Features:
- Browse and search business terms
- View term definitions, data stewards, term status (candidate, accepted, deprecated, etc.) and other properties
- Name model elements based on terms
- Assign (classify) model elements to terms and publish to Business Glossary

InfoSphere Data Architect showing Business Glossary terms in the left navigation tree, with drag and drop capability to the modeling canvas.
Business Glossary – What’s New in 8.5?

Increase the Productivity of Business Intelligence Projects

- Jump-start glossary content from your BI platform
- Share BI and reporting terminology across the organization
- Build a glossary that is aligned with your BI projects to rapidly improve the productivity and communication among your BI users

Feature:
- Populate glossary categories and terms from various Business Intelligence models
Business Glossary – What’s New in 8.5?

**Empower the “Casual” Business User**

- Establish a common understanding among all users, including casual every day users who don’t need any technical information
  - Customer service representatives
  - Human resources
  - Branch tellers
  - Claim agents

**Feature:**
- The “Basic” user view displays business terms and definitions, without listing the relative underlying technical assets
Gain Visibility to a More Complete View of Information Assets

- Get a complete picture and understanding of technical data elements with insights into their
  - Domain values
  - Relationships to other sources
  - Associated business rules

Features:
- View data relationships and insights discovered by InfoSphere Discovery
  - Database tables and columns
  - Database analysis and inferred database key information
- View data rule descriptions, policies, stewards, and rule expressions defined in Information Analyzer
Metadata Workbench – What’s New in 8.5?

*Extend the Data Lineage Coverage*

- Enrich data lineage flows by including
  - Designed data transformation rules
  - Discovered and detected legacy lineage
    - *Among various data sources*
    - *Between two databases connected through homegrown solutions*

**Features:**
- Display FastTrack source to target mappings in data lineage
- Display discovered relationships between data sources and targets generated by Discovery
Information Analyzer – What’s New in v8.5?

Continuously Manage & Monitor Data Quality

- Enhance collaboration between business and IT for data validation
  - Quickly associate or lookup terminology for data or rules
  - Identify valid values with terms

- Leverage analytical results in broader business context
  - CLI and REST API for publishing results to external files or building custom dashboards, applications, or reports

- Expanded native data source support
  - Added native Oracle connectivity
  - Enhanced ODBC, DB2, and Teradata connectivity

- Exception Management capabilities for managing source to target transformations
  - Review exceptions and recommend remediation
Enhanced Collaboration with Business Glossary and Metadata Workbench

- Fast access to Business Glossary Terms
- Connect tables and columns to Terms via right-click menu options
- Display linked Terms and associated properties as part of Data Classification
Validity and Value Information in Business Glossary and Metadata Workbench

Available information includes:
- Method of validation used
- Percentage of valid data
- Min/Max values
- Validation range
- Reference table information

```
Range Validation —
- Minimum Value — 600
- Maximum Value — 700
- Compliance — 44%
```

```
Require Unique Values — False
Total Number of Rows — 221
Number of Complete Values — 221
Number of Valid Values — 98
Number of Empty Values — 0
Number of Null Values — 0
Number of Distinct Values — 221
Number of Distinct Patterns — 0
Number of Distinct Formats — 1
Inferred Data Type — INT16
Inferred Format — 999
Domain Type —
Inferred Length — 3
Inferred Scale — 0
Inferred Precision — 3
```
Flexible Processing and Result Publication
REST API / CLI

- Provide external access to Information Analyzer results, rules, and project information
- Utilize data as needed for external or downstream processing (such as custom reports)
  - Incorporate into scripts, customer dashboards, or applications
  - Results generated as XML files for flexibility

```
GET columnAnalysis/results
```

![Diagram showing XML output and XSLT transformations to HTML and CSV reports]

© 2010 IBM Corporation
InfoSphere FastTrack v8.5 – What’s New in 8.5?

Business-driven Data Integration & Automation

- Leverage prior knowledge to achieve consistency
  - Re-engineer ETL job designs to jump-start specifications
  - Reuse discovered and established rules for consistent application
  - Create specifications from existing InfoSphere DataStage jobs

- Ensure compliance to organization standards
  - Establish template parameters for mapping specifications within project
  - Review specifications through configurable reporting
  - Track changes to mapping specifications through audit history

- Enhanced import of .CSV files
  - Map multiple columns to reduce effort
Reverse Engineering of DataStage Jobs

- Starting point for new work or as documentation of the finalized job specification
  - Create specifications including the sources, targets, and identifiable transformation rules
Auditing of Mapping Specifications

- Provides audit history and traceability by tracking changes over time: who + when + what + why was it changed
- At project level or mapping specification level
- Option to enable audit history
Data Architect – What’s New?

Enhancements to improve ROI of warehousing and BI initiatives

- Dimensional modeling to enhance data warehouse and BI ROI
  - Data design and business analysis integration to reduce development time of data warehouse and BI initiatives
    - Dimensional model aligned to InfoSphere Warehouse OLAP solution
    - Deliver consistent information to business users with Cognos BI integration
  - Single source for enterprise models
    - Enriched dimension notations and diagramming options
    - Multi-layer design with logical and physical dimensional models
    - Ability to align changes in business and IT using forward synchronization
    - Faster impact analysis

- Enhanced productivity and visualization
  - Incorporate IBM ILOG technology for data model visualization
  - New comparison filter options to streamline change management between models and databases
New! InfoSphere Data Architect Integration
Support for Cognos and InfoSphere Data Warehouse

Data model of record for the enterprise
- Design warehouse optimized for BI
- Ensure business requirements are met
- Deliver accurate reports

Fast time to value for warehouse design
- Auto discover warehouse schemas
- Determine problems quickly across warehouse/BI applications
- Jumpstart warehouse/BI applications

InfoSphere Foundation Tools
- Industry Data Models
- InfoSphere Business Glossary
- InfoSphere Discovery
- Information Analyzer
- Metadata Workbench

Rational & WebSphere
- Rational Software Architect
- Rational ClearCase
- WebSphere Business Modeler

InfoSphere Optim Tools
- Optim Development Studio
- Optim Data Administrator
- Optim Designer
- Optim Manager

InfoSphere Data Architect
- Logical Models
- Physical Models
- Dimensional Models
- Business Glossary Models

InfoSphere Warehouse

Cognos

New! InfoSphere Data Architect Integration
Support for Cognos and InfoSphere Data Warehouse

Data model of record for the enterprise
- Design warehouse optimized for BI
- Ensure business requirements are met
- Deliver accurate reports

Fast time to value for warehouse design
- Auto discover warehouse schemas
- Determine problems quickly across warehouse/BI applications
- Jumpstart warehouse/BI applications
Enhance ROI for Warehousing and BI solutions

Data modeling support across the enterprise

Source Systems → InfoSphere Warehouse

Third Normal Form (3NF) ER Model → Normalized Enterprise Warehouse Model

InfoSphere Data Architect

- Logical Models
- Physical Models
- Dimensional Models
- Business Glossary Models

Data Marts → OLAP

InfoSphere Cubing Services, Cognos Cubes

Cognos BI reports

InfoSphere Warehouse

Denormalized Dimensional Model
InfoSphere DataStage and QualityStage

- Increased Availability
  - Improved support and levels of HA
- SCCS integration
  - Integration with Enterprise Source Control providers via Eclipse Team Plugins
- Performance (Design and Runtime)
  - Improved Use Experience and runtime performance
- Balanced Optimization
  - Deliver seamless integration with DataStage GUI and improved metadata lineage
- Transformer Enhancements
  - Looping, improved NULL handling
- Vertical Pivot
  - Extend current Horizontal Parallel pivot
- z/OS File Stage
  - Deliver improved native MVS File processing support on Linux for System Z
- XML PACK (post vNext)
  - Deliver new features and improved performance / volume support
- Standardization and Match Specification reports
- Match Designer Enhancements
Increased Availability

- Cluster support for Application Server Tier
  - WAS 6.1 and 7.0 ND
- Cluster support for Repository Tier
  - DB2, Oracle, SQL Server
- Improved Failover support for Engine Tier
- Delivering higher levels of Availability for Production and Development environments
- Delivering Horizontal and Vertical scaling / load balancing of the Domain and DB Tiers
SCCS Integration

- Leveraging the Eclipse Platform for Team Development
- Integration with code-management (SCCS) providers supported through Eclipse Plugins
  - Support for ClearCase, CVS
    - Other providers configurable via Eclipse Team Plugins
- Functions to interact with the SCCS invoked from the Information Server Manager
Balanced Optimization

- DataStage Balanced Optimization seamless integration with DataStage Designer
- Supported dependency relationship
  - Optimized job – depends_on - Root Job
  - Improved support for Impact Analysis, Where Used and Lineage
  - Find all optimizations of a root job
  - Find the root job for any optimized job
- Improved Optimization Logging
- Improved job development lifecycle support for Optimized Jobs
  - create, delete, edit, rename, move folder, compile, deploy, export/import
General Enhancements

- Design Time Performance
  - 40% Performance improvement in Job Open, Save, Compile etc.
- Improvements in startup time for jobs
- Runtime Performance
- New Utility, String and Date/Time functions
  - eg: IsValidTime, NthWeekdayFromDate, DecimalToTimestamp
Transformer Enhancements

- Deliver Looping in the Transformer
- Allow multiple output rows to be produced from a single input row
- Support for End of Data Flag to support Key Break Logic
- New Input Cache
  - `SaveInputRecord()`
  - `GetSavedInputRecord()`
- New System variables, function
  - `@ITERATION`, Loop Count
  - `@EOD`, End of data flag for last row
  - `LastRowInGroup(InputColumn)`, automate change detection
- Stage, Loop Variables support for Nullability
- More options for Null Handling
Vertical Pivot

- Enhanced *Pivot* stage to support - Vertical Pivoting
  - mapping multiple input rows with a common key, to a single output row containing multiple columns
z/OS File stage (already enabled for V8.1)

- New Native Support for Accessing Mainframe files from Distributed platforms and Linux for System Z
  - Through a new stage called z/OS File stage
  - VSAM files - KSDS, ESDS, RRDS
  - Sequential files - QSAM, Sequential read of BDAM/BSAM, PDS members, GDG files.

- Initial release
  - Read/Write for Sequential files and read only for VSAM.
  - Fixed and variable-length records
  - single or multi record type format files will be supported

- Leveraging InfoSphere Classic Federation
zOS File Stage Solution

Information Service Framework

DS Client

DataStage Server

DataStage Server

zOS File Stage

zOS File Stage Solution

Information Service Framework

Metadata Repository

PX Engine

zOS Read OP

z/OS Write OP

TCP

Classic Fed Server

Flat Files

z/OS
XML Pack (post vNext)

- Deliver improved environment for mapping complex XML documents from one schema to another in single or multiple steps
- Support for Schemas (XSD XML Schema 1.0, WSDL 1.1)
- Deliver Support for complex XML Transformations with need for shredding the XML
  - Hierarchical Join, Relational join, Filter, Switch, Sort, Union, Regroup, RowToColumns, ColumnsToRows, Aggregate, Distinct

- Support for invoking Web Services as a transformation step
- Support for multiple input and output links, including reference, reject
- Shall support partitioning, multi threaded and stream processing of large XML documents
- Performance and volume improvements
  - reduced memory requirements, increased throughput
  - Processing 163mb complex XML
    - XML PACK 2.0 391 seconds
    - XML PACK 3.0 51 seconds
- Unlimited Document Size
**Standardization Quality Assessment**

- **Summary reporting**: to understand how well Standardization is performing.
- **Sections**:
  - Summary results for records assessed
  - Percentage of records that populate each individual output category to determining anomalies
  - Composition of standardized output patterns
Match Specification Report

- Introduced a match specification report to capture the details of the match specification
- Includes a glossary and all the pertinent details for the specification and passes
- Enables the Business to review Match Rules more effectively
Other QualityStage Enhancements

- Standardization
  - 30% Performance Improvement to deliver significant gains in runtime and system resource usage
  - Added support for specifying a tokenizer to handle mixed international datasets in a single job.

- New RuleSets
  - Argentina, Brazil, Chile, Peru, Mexico, Netherlands, India, Traditional Chinese and Japanese Kana
  - Delivering a Product ruleset accelerator

- Improved Match Designer
  - Added support in the Match Designer test environment setup for using DS connection objects
  - Added support for parameterizing match cutoffs
  - Reordering of blocking and match commands,
  - Extended support for all types of reference matches and improved interval match comparison types
Enhanced Match Designer

• Improved Match Weight Comparison
• More Control, Data, Statistical details delivering improved Analysis
Platform Support

New
- Red Hat Enterprise Linux 5,6 (64 bit)
- SUSE Linux Enterprise Server 9,10 (64 bit)
- Windows Server 2008 64bit (32-bit app)
- AIX 5.3, 6.1 (64 bit)
- Solaris 9,10 (64 bit)
- HP-UX Itanium (64 bit)
- Red Hat Enterprise Linux for System Z (64 bit)
- SUSE Linux Enterprise Server for System Z (64 bit)
- Windows Server 2003 (32-bit)
- Red Hat Enterprise Linux 5, 6 (as 32-bit app)
- SUSE Linux Enterprise Server 9, 10 (as 32-bit app)

New
- Clients – Windows XP, Vista and 7 (32 & 64 bit)

New
- Repository – DB2 9/5, 9.7, Oracle 10g, 11g, SQL Server 2005, 2008
Connection Migration Tool

- Modifies jobs that use legacy plug-in/operator stages to use newer Connectors
- Migrates all compatible stages
- GUI and command line (batch) modes
- Server and Parallel jobs
- Backup, clone or replace jobs
- Jobs are annotated with information about the migration
Certifications Delivered

New database connectivity enhancements include support for -

- IBM DB2 LUW 9.7, 9.8 (Pure Scale)
- IBM DB2/Z v10
- IBM WebSphere MQ Series 7.0
- Informix IDS v11.7
- Teradata 13
- Oracle 11gR2
- Sybase ASE v15.5
- Sybase IQ v15.2
- Netezza 5
- DataDirect ODBC v6.0 SP2 (including new Bulk Loading for SQL Server from DataStage on Linux/Unix)
- DB2/Z Bulk Load/UnLoad Stage – enables FTP data movement and invocation of native DB2/Z utilities

Add-ons:

- z/OS File Stage – new component that enables more direct/easy access to native mainframe files.
- Optimized Integration with Changed Data Capture (CDC) and Changed Data Delivery (CDD)
- Packs all (for SAP R/3 and BW, Oracle eBusiness/Siebel/PeopleSoft/Hyperion Essbase/JD Edwards), updated for use with Information Server/DataStage v8.5 with new functionality for salesforce.com bulk loading.
Thank You
BACKUP
InfoSphere Business Glossary

Your Business-Driven Entry Point to All Information Centric Projects

- Entry point to Master Data, Data Warehousing, App Consolidation & Migrations

- Start from a business concept (customer) & drill into a complete set of relevant information:
  - Business terms & business rules
  - Data Models & BI structures
  - Structured & Unstructured data
  - ETL processes
  - Business lineage

- Access glossary content
  - Directly from any Windows application
  - Via open APIs - REST, Eclipse
  - From software platforms – InfoSphere Information Server, Cognos BI, Rational...

- Populate glossary with rich industry standard terms and definitions
InfoSphere Metadata Workbench
Support Information Governance and Compliance Reporting Initiatives

- Visualize and trace information flows to support audit and compliance reporting
- Provide visibility to trusted sources of information
  - Where is the data coming from?
  - Who has modified it? When and why?
  - Where is the information used?
- Build trust and confidence in data by viewing how the data flows through Information Server or 3rd party tools
- Assess the impact of information changes and minimize project risks
- Promote reuse and collaboration among technical teams
IBM InfoSphere Data Architect

Model, relate and standardize diverse and distributed data assets

Requirements

- Design and manage enterprise models
- Enforce model conformance to enterprise standards
- Leverage industry data models for best practices
- Reverse engineer and identify existing warehouse schemas

Benefits

- Speed design activities
- Populate Business Glossary from model terms
- Validate models for enterprise conformance
- Optimize existing investments