# IBM InfoSphere Information Integration and Governance

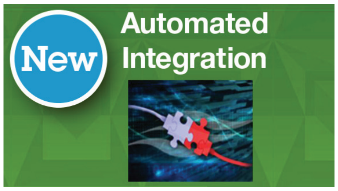
*Building confidence in big data*

## Highlights

- Innovations help organizations build confidence in their data to enable actions that leverage new insights.
- Automated integration accelerates information-intensive projects.
- Visual context for data accelerates understanding of the data and any associated issues.
- Agile governance applies just the right level of control, as appropriate to the data, the use case and the organization.

Big data has created a new era of opportunity for organizations of all types. It offers insights that can lead to solutions to some of the thorniest business challenges, and it opens the door to transformations that can make organizations more efficient, more effective and more creative. But the era of big data is the era of messy data. That is the big data paradox—larger volumes and variety of new sources are inherently complex, and that complexity can actually lower confidence. Successful organizations understand that it's no longer about making the data perfect; it's about making the data good enough for the task at hand.

Integration and governance must evolve and become more agile to handle the challenges of big data. IBM's latest innovations in integration and governance deliver context for big data, provide agile governance with automatic protection of big data, and automate ingestion of data from both traditional internal sources and new, external sources. These innovations are essential to build confidence in big data.



### InfoSphere Data Integration

Self-service access to a growing variety of big data in traditional, NoSQL and Hadoop souces

### Information Governance Dashboard

Immediate, visual context for critical decisions and actions understand big data to better leverage

### InfoSphere Privacy and Security

Find and protect sensitive big data single pointof security for traditional, NoSQL and big data

## The big data paradox: The rise in volume, variety and complexity can lower data confidence

Unanswered questions are reducing the confidence of business people in the data they need to run the business.

### Questions of Chief Marketing Officer

- Do I have the right context?
- Am I confident in the data used for this analyis?
- Is this data good enough to use?

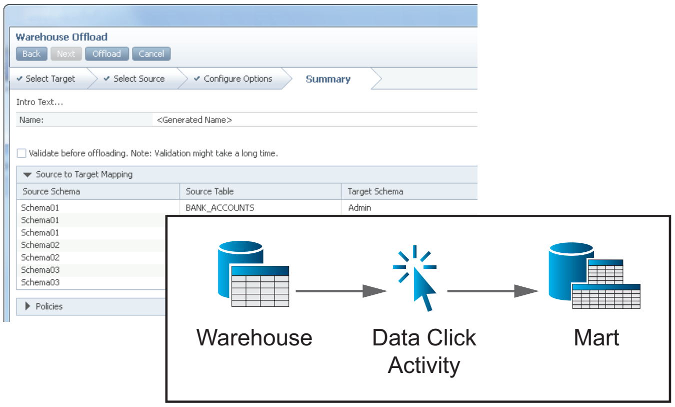
### Questions of Chief Data Officer

- Which data is correct?
- Am I protecting sensitive big data?
- Do I understand the risk of using this data “as is”?

To address the new challenges created by big data, IBM has announced innovations in the InfoSphere® Information Integration and Governance (IIG) portfolio to help organizations to answer critical questions, build confidence in their data, make well-founded decisions and take decisive actions to accelerate a whole range of information-intensive projects.

## Automated integration: Understand and integrate data more quickly

IBM innovations are making it easier than ever before for business users to get the information they need for their own analytical or operational projects.



**Warehouse Offload**

Back | Next | Offload | Cancel

✓ Select Target › ✓ Select Source › ✓ Configure Options › **Summary**

Intro Text...

Name: <Generated Name>

□ Validate before offloading. Note: Validation might take a long time.

▼ Source to Target Mapping

| Source Schema | Source Table | Target Schema |
|---|---|---|
| Schema01 | BANK_ACCOUNTS | Admin |
| Schema01 | | |
| Schema01 | | |
| Schema02 | | |
| Schema02 | | |
| Schema03 | | |
| Schema03 | | |

▶ Policies

*Figure 1*: InfoSphere Data Click—Self-service data provisioning

### Self-service integration

IBM InfoSphere Data Click, with the latest enhancements, enables ***self-service access*** to a growing variety of data in traditional, NoSQL and big data sources. Rather than building a queue of data requests to IT, business users can initiate data integration on their own, acquiring the data they need to move ahead with their projects.

### Through self service, data can now be sent to or sourced from a whole host of traditional and big data environments, including

IBM DB2®, Oracle, files, Greenplum, IBM Informix®, IBM PureData™ for Analytics (formerly known as Netezza®), salesforce.com, Microsoft® SQL Server, Teradata and more.

### Expanded integration of NoSQL and big data

Organizations implementing information integration projects can now efficiently leverage a whole range of traditional and big data types, including JSON, IBM InfoSphere BigInsights™ and multiple JDBC-accessible sources.

6x

more big data sources and targets supported[1]

### Real-time updates to Hadoop

Keeping data in Hadoop repositories up-to-date with the latest changes in source applications is now automated, because changed data from multiple sources can be replicated directly to Hadoop repositories.

### Certifications with Hadoop distributions

InfoSphere information integration solutions are now certified with key Hadoop distributions, including Cloudera (CDH 4.2), Hortonworks (HDP 1.2) and InfoSphere BigInsights (2.0).

### Automated analysis and validation of big data

To use big data effectively, analysts, data scientists and business users need to understand the content and quality of the data sources. The Hive ODBC driver provides native access to Hadoop-based file systems, for initial analysis as well as ongoing data validation.

### Accelerated delivery of big data to IBM System z®

IBM System z® is a frequent target for data workloads, whether they originate in traditional or from big data sources. Delivery to System z has now been accelerated by using native DB2 z/OS® partitioning capabilities to optimize data loads.

## Visual context: Seeing is believing

One reason for the lack of confidence in data—especially big data, with its high volume, velocity and variety—is the lack of a clear understanding of the data. The source of the data and its history are often unclear, and it is not clear how the data is measuring up to the organization's own data quality rules. It is also difficult to get a complete view of information governance policies as well as an indication of how the current data is performing against those policies. New InfoSphere IIG capabilities have been developed to address these causes of uncertainty and to increase user confidence in data.

### Information governance dashboard

A flexible dashboard clearly displays business-driven governance policies and rules as well as current and historical results from various sources—all tailored to meet your specific requirements. For example, users might be able to view data integration, data quality, master data management and data lifecycle-related metrics, all in a single view, with facilities for drilling down into further detail and taking appropriate action. In addition, InfoSphere provides capabilities to build specific dashboards and views as required.

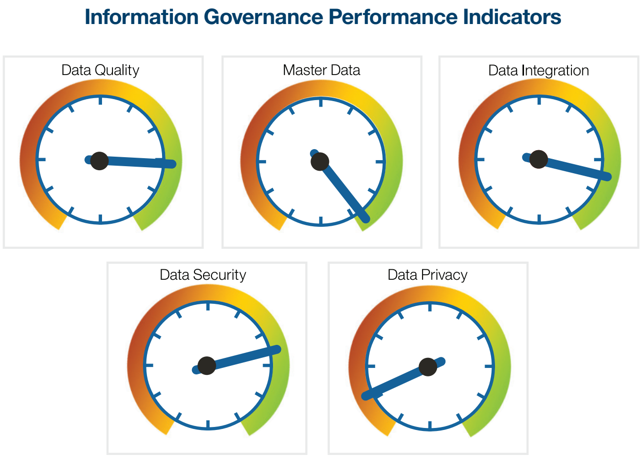


*Figure 2*: InfoSphere Governance Dashboard

### Rapid metadata ingestion

A distinguishing feature of the InfoSphere IIG platform is the shared metadata across multiple functional components. Now, regardless of the number or complexity of data sources, metadata can be quickly and easily imported from hundreds and thousands of big data sources. In the era of big data, more data means more metadata, and the performance and speed of discovery and profiling become critical. Once the metadata is imported, it becomes available for online viewing, so users can quickly understand the new data sources.

170x

Improvement in metadata import performance[2]

### Streamlined metadata classification

New integration between InfoSphere Business Information Exchange and InfoSphere Data Explorer extends IBM's strength in metadata management, making it easier for organizations to identify and classify metadata for big data sources.

## Agile governance: Providing appropriate governance for each big data use case

A common misconception is that data governance is a heavyweight process that needs to be applied consistently against all data, for all use cases, if at all. Fortunately, with the wide variety of information-intensive projects as well as the expanding array of available data, there is a much better approach: agile governance, with controls that are appropriate to the data, the use case and the organization. InfoSphere IIG now makes it easier than ever before to manage governance with the agility that today's data landscape requires.

### Activity monitoring and auditing in big data environments

IBM now provides agile privacy and security for sensitive data in both traditional environments and newer NoSQL platforms, including Cassandra, GreenPlum, Hortonworks and MongoDB.

### High performance security for big data

The new 64 bit architecture for high performance security provides data security at big data scale.

300%

Faster automated security[3]

### Masking of sensitive data for Hadoop

InfoSphere now allows masking of sensitive Personally Identifiable Information in multiple data stores that feed data into Hadoop projects. The data is fully usable for purposes such as analytics and testing, but the sensitive information remains anonymous for data privacy.

### Hadoop-based access to archived data

Business leaders can now leverage analytics to make informed decisions across structured data, including archived data, and unstructured data, as cold historical data is archived from traditional systems and loaded into Hadoop.

## Looking ahead: Statement of direction

Beyond the capabilities included in this announcement, IBM is planning to extend its support for automated integration, visual context and agile integration. Consider some of the additional capabilities included in this plan.

### Big Match

Matching master records is a compute-intensive process. With the growing volume of data and more repositories to integrate and match, this is becoming a significant challenge and bottleneck to exploiting big data. Without understanding master entities such as customers, products, and locations, how can you derive actionable and accurate insight from big data? Big Match capabilities are designed to marry the sophistication of the master data management (MDM) probabilistic matching engine with the raw computing power of Hadoop (InfoSphere BigInsights). By running the matching engine on Hadoop, MDM can match at high speed and in real time. This enables organizations to match data as they encounter it, and therefore derive insights from big data faster.

### Big Data Catalog

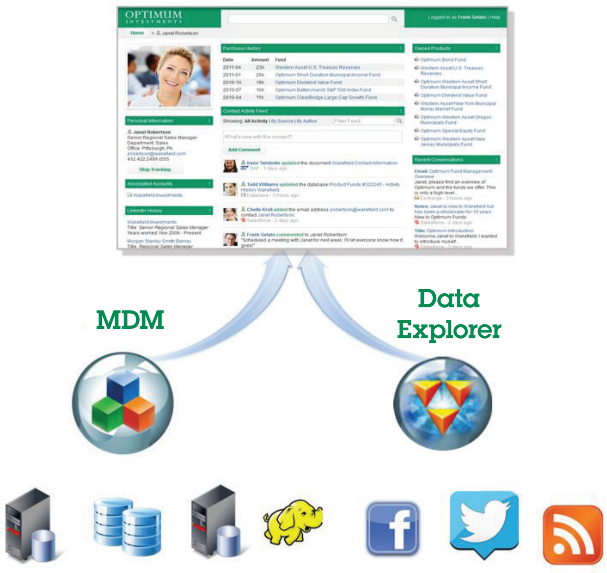
One of the hardest challenges of big data is simply finding the right data. Even when it's in one system, such as a Hadoop landing zone, the sheer volume and variety make it hard to find what you need.

A Big Data Catalog can make it easy for data users and scientists to 'shop for data.'  It ingests and stores metadata from every available source, classifies data (for example, its origin, lineage and potential value), and makes it easy to search and find via a user interface or SOA APIs. A Big Data Catalog is planned to provide structure to Hadoop landing zones, enabling users to search, find, and leverage big data more quickly.

### MDM for Big Data

Integration of InfoSphere MDM with InfoSphere Data Explorer is designed to provide the true and complete 360° view across structured and unstructured big data. InfoSphere Data Explorer provides the capabilities to extend MDM across unstructured big data with federated views and visualization of the complete master record.

Organizations today have fluid requirements that cannot all be addressed by a single MDM style, whether virtual or physical. Organizations need flexible and simultaneous support for multiple styles so they can rapidly onboard and master new big data sources, link them to master records, and as the data becomes more trusted, incorporate them into the single master data record. A unified InfoSphere MDM engine is planned to support implementations with virtual, physical and hybrid MDM styles, with high performance.



OPTIMUM INVESTMENTS

Logged in as Frank Gelato | Help

Home > Janet Robertson

Purchase History

| Date | Amount | Fund |
|---|---|---|
| 2011-04 | 23k | Western Asset U.S. Treasury Reserves |
| 2011-01 | 25k | Optimum Short Duration Municipal Income Fund |
| 2010-10 | 18k | Optimum Dividend Value Fund |
| 2010-07 | 10k | Optimum Batterymarch S&P 500 Index Fund |
| 2010-04 | 11k | Optimum ClearBridge Large Cap Growth Fund |

Personal Information

Janet Robertson
Senior Regional Sales Manager
Department: Sales
Office: Pittsburgh, PA
jrobertson@wakefield.com
412.422.2499 x555

Contact Activity Feed

Showing: All Activity | By Source | By Author

Irene Tambolin updated the document Wakefield Contact Information. SAP · 1 days ago

Todd Williams updated the database Product Funds #322245 - Activity History Wakefield. Database · 3 hours ago

Chelle Kroll added the email address jrobertson@wakefield.com to contact Janet Robertson. Salesforce · 2 days ago

Frank Gelato commented to Janet Robertson: "Scheduled a meeting with Janet for next week. I'll let everyone know how it goes!"

Owned Products

- Optimum Bond Fund
- Western Asset U.S. Treasury Reserves
- Optimum Western Asset Short Duration Municipal Income Fund
- Optimum Dividend Value Fund
- Western Asset New York Municipal Money Market Fund
- Optimum Western Asset Oregon Municipals Fund
- Optimum Special Equity Fund
- Optimum Western Asset New Jersey Municipals Fund

Associated Accounts

Wakefield Investments

LinkedIn History

Wakefield Investments
Title: *Senior Regional Sales Manager*
Years worked: Nov 2006 - Present
Morgan Stanley Smith Barney
Title: *Regional Sales Manager*

Recent Conversations

Email: Optimum Fund Management Overview
Janet, please find an overview of Optimum and the funds we offer. This is only a high level...
Exchange · 3 hours ago

Notes: Janet is new to Wakefield but has been a wholesaler for 10 years. New to Optimum Funds.
Salesforce · 2 days ago

Title: Optimum Introduction
Welcome Janet to Wakefield. I wanted to introduce myself...

*Figure 3*: The Complete 360° View—InfoSphere MDM + InfoSphere Data Explorer

## Delivering confidence for big data use cases

**The latest innovations help InfoSphere deliver critical confidence for key big data projects. For example:**

| | |
|---|---|
| For enhanced 360° view projects | New support for hybrid styles of master data management in a single system helps to tailor architectures and accelerate projects. |
| For data warehouse augmentation | New integration of NoSQL and big data plus auto-analysis and validation of big data streamlines warehouse projects. |
| For big data exploration | Rapid imports of metadata help users move quickly to visualize and understand new data sources. |
| For security/intelligence extension | Monitoring and auditing activity in big data environments provide critical protection. |
| For operations analysis | High-performance security for big data and masking of sensitive data in Hadoop files keep machine data protected for analysis. |

## About IBM InfoSphere Information Integration and Governance

Information Integration and Governance (IIG) capabilities bring together data from diverse sources for diverse targets, manage its quality and maintain master data for multiple domains. They secure and protect data, manage it across its lifecycle and facilitate information-based collaboration across business and technical teams. These broad capabilities help organizations increase the value of data for information-intensive projects like big data and analytics, application consolidation and retirement, security and compliance, 360° views and many others.

The new IIG capabilities are provided by products within the InfoSphere IIG portfolio, including InfoSphere Data Replication, InfoSphere Guardium, InfoSphere Information Server, InfoSphere Master Data Management and InfoSphere Optim.

## For more information

To learn more about IBM InfoSphere Information Integration and Governance, please contact your IBM representative, or visit: **ibm.com**/software/data/information-integration-governance

IBM Corporation
Software Group
Route 100
Somers, NY 10589

Produced in the United States of America
September 2013

This document is current as of the initial date of publication and may be changed by IBM at any time. Not all offerings are available in every country in which IBM operates.

THE INFORMATION IN THIS DOCUMENT IS PROVIDED "AS IS" WITHOUT ANY WARRANTY, EXPRESS OR IMPLIED, INCLUDING WITHOUT ANY WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND ANY WARRANTY OR CONDITION OF NON-INFRINGEMENT. IBM products are warranted according to the terms and conditions of the agreements under which they are provided.

IBM's statements regarding its plans, directions, and intent are subject to change or withdrawal without notice at IBM's sole discretion.

Information regarding potential future products is intended to outline our general product direction and it should not be relied on in making a purchasing decision.

The information mentioned regarding potential future products is not a commitment, promise, or legal obligation to deliver any material, code or functionality. Information about potential future products may not be incorporated into any contract. The development, release, and timing of any future features or functionality described for our products remains at our sole discretion.

Performance is based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput or performance that any user will experience will vary depending upon many factors, including considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve results similar to those stated here.

[1] Release-to-release functional comparison for InfoSphere Information Server with InfoSphere Data Click feature

[2] Source: IBM internal performance test results, comparing three test runs with the latest version of InfoSphere Information Server averaging 11.46 seconds vs 1,964 with the previous release

[3] Release-to-release comparison for InfoSphere Guardium, with more processing in parallel, processing in batches, improvements by a factor of at least 3X

♻ Please Recycle