IBM Software
Thought Leadership White Paper

June 2014

Bolstering data confidence with comprehensive information integration

IBM "anywhere integration" capabilities make data environments accessible and available for analysis

Robust integration capabilities for greater confidence in the big data era

The importance of gathering and gaining insight into data has never been greater, even as data sources increase in number. Organizations of all sizes must accommodate streams of structured and unstructured data; data from internal and third-party clouds; and even data as granular as departmental databases and user spreadsheets.

This puts a new spin on an old problem: the question of integration. Organizations need robust information integration capabilities that support their business requirements and create the utmost confidence in their data, and thus in the results of their decision making. They need to successfully and flexibly integrate data from any data source, all while applying governance and data-quality best practices. Only with that assurance of trust can firms be sure that critical projects and key analytics initiatives will succeed.

By using the end-to-end information integration capabilities of IBM® InfoSphere® Information Server, companies are able to better understand, cleanse, monitor, transform and deliver not only their data, but also data derived from external sources (see Figure 1). InfoSphere Information Server also helps organizations collaborate to establish or improve data governance—whether forming a business glossary or creating data governance rules and policies to bridge the gap between line-of-business and IT teams. These capabilities help firms ensure that the information that drives their business and strategic initiatives—from big data and point-of-impact analytics to master data management (MDM) and data warehousing (on-premises or in the cloud)—is trusted, consistent and conforms to governance policies.

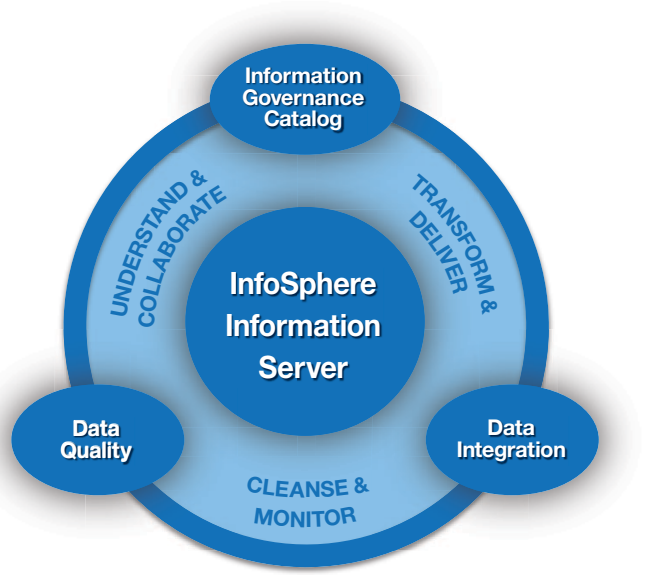

Figure 1. InfoSphere Information Server helps businesses make decisions confidently by enhancing trust in data, facilitating flexibility, supporting data accuracy and integrating data sources quickly.

Since its inception, InfoSphere Information Server has provided a massively parallel processing (MPP) platform, supporting data volumes of any size, regardless of complexity. It can deliver the flexibility (through extract, transform and load [ETL] or extract, load and transform [ELT]) performance, and scalability required for big data projects.

IBM InfoSphere Information Server V11.3 delivers the "anywhere" information integration capabilities that organizations need to address the increasing volume and complexity of data and data sources. Acknowledging the ongoing challenges

organizations face with agile integration, business-driven governance and sustainable data quality, IBM developed InfoSphere Information Server V11.3 to give them the sophisticated information integration capabilities necessary to thrive in today's information-rich environment.

InfoSphere Information Server V11.3 capabilities are available in four packages that help firms target key information challenges (see Figure 2):

- **InfoSphere Information Server for Data Integration:** Transform data in any format and deliver it to any system, supporting faster time to value and reduced risk for IT.
- **InfoSphere Information Server for Data Quality:** Establish and manage high-quality data, transforming a deluge of data into trusted information.
- **InfoSphere Information Governance Catalog:** Understand data and foster collaboration between IT and line-of-business teams to narrow the communication gap and create business-driven information integration.
- **InfoSphere Information Server Enterprise Edition:** Gain the capabilities of all three individual packages in one comprehensive package so firms can start information integration efforts in one area, and then expand efforts for further optimization.

This white paper discusses the data integration, data quality and catalog packages, as well as the ways InfoSphere Information Server V11.3 helps organizations address data integration needs in cloud environments (both private and public), gain better insight into big data, improve self-service data integration and accommodate tighter MDM integration. InfoSphere Information Server V11.3 also features performance and security enhancements.

Figure 2. InfoSphere Information Server Enterprise Edition consists of three packages.

The enhanced InfoSphere Information Server packages

Using InfoSphere Information Server packages, organizations can provide accurate, comprehensive information in near-real time to the systems and knowledge workers focused on strategic initiatives.

InfoSphere Information Server for Data Integration: Agile integration capabilities

InfoSphere Information Server for Data Integration delivers agile integration capabilities so that businesses can integrate data quickly and flexibly wherever it resides. Businesses can easily manage data integration for data warehouses, integrate big data, consolidate applications, deploy information in a private or public cloud or integrate on-premises data with cloud environments.

Data integration for cloud environments

In addition to its current support for private cloud deployments (either in conjunction with IBM PureApplication® System as a managed platform as a service, or in conjunction with IWD or SCO, for clients or partners that prefer using a self-service private cloud), InfoSphere Information Server V11.3 now supports public cloud environments and provides enhancements for integrating on-premises data with cloud environments. It also supports direct integration so users can load data into Amazon S3. After data is integrated within S3, it may be picked up by other cloud database technologies.

Building on that integration foundation, Version 11.3 includes integration with REST application programming interfaces (APIs), enabling support for XML and JSON messages. By delivering REST-based connectivity, InfoSphere Information Server is able to support distributed database-as-a-service (DBaaS) offerings, such as IBM Cloudant, as well as other on-premises and off-premises solutions that offer REST-based interaction.

Deeper integration for MDM projects

InfoSphere Information Server V11.3 delivers tighter integration with IBM InfoSphere Master Data Management (InfoSphere MDM) software. A new MDM Integration Stage enables users to easily load data into and extract data out of InfoSphere MDM. Users can now include MDM data within their data integration flows and load domain data (relating to customers, partners, suppliers, members, products, and other entities) directly to the MDM system.

Organizations can also leverage InfoSphere Information Server data quality capabilities to standardize data before loading it into InfoSphere MDM, which helps to increase matching accuracy and better support 360-degree views of entities. In addition, the MDM Integration Stage is supported by a metadata-driven design, which means metadata captured by the stage will align with the MDM data model, enabling multiple segments of data to be sent in a single request.

Enhanced agile, self-service integration with InfoSphere Data Click

IBM InfoSphere Data Click, a feature of InfoSphere Information Server, helps simplify the process of provisioning specific environments (such as analytics sandbox environments) with data. Traditionally the domain of highly skilled ETL engineers, InfoSphere Data Click enables novice or line-of-business users to easily retrieve data and populate new systems with a few clicks—and without additional IT involvement or coding (see Figure 3). InfoSphere Data Click helps speed time to value, increase business agility and lower costs by shrinking the time required to complete tasks from days or weeks to minutes or hours. To learn more about InfoSphere Data Click, visit http://ibm.co/1i3jQXb

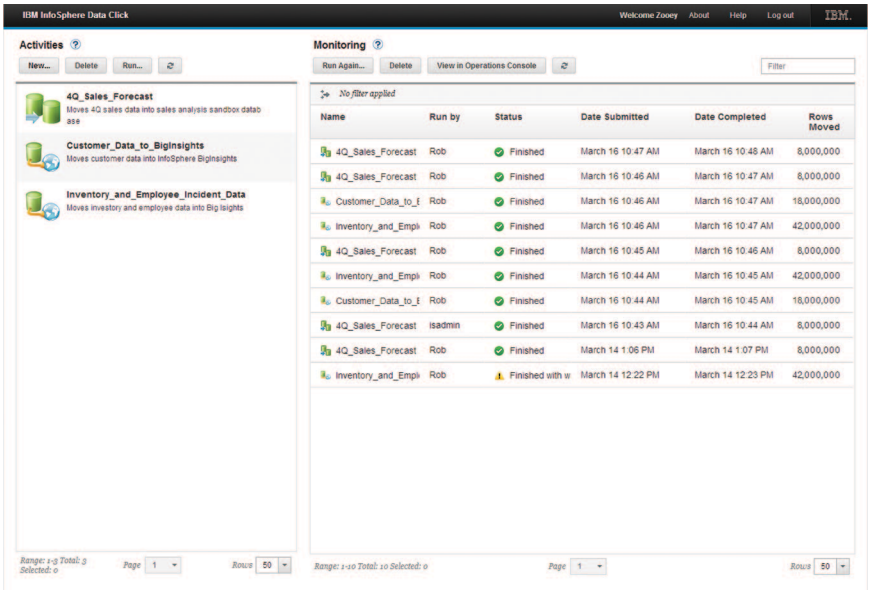

Figure 3. Users log directly into a streamlined InfoSphere Data Click UI. From here, they can quickly and easily create, edit, run and monitor activities.

New InfoSphere Data Click V11.3 capabilities include:

- **Cloud integration:** Integrate data directly to and from Amazon S3
- **Enhanced big data support:** Use native high-speed load to move data into IBM InfoSphere BigInsights™ and once data is in InfoSphere BigInsights, perform advanced analytics; access read data out of Hadoop distributions through Big SQL or through other JDBC-based methods, such as Hive and HBase
- **Catalog integration:** Search for information through the metadata available in InfoSphere Information Server Governance Catalog and launch directly into InfoSphere Data Click to act on integrating that data, for faster time to value
- **Enhanced usability:** InfoSphere Data Click now includes a new screen for creating and monitoring activities; a new web-based, streamlined method to create and run activities; and new links to the operations consoles to drill down into more monitoring detail
- **Relational database connector expansion:** Leverage JDBC, native Oracle and IBM DB2® connectivity to move data from more sources more quickly

Native Greenplum connectivity

A new Greenplum connector is available in V11.3 to help clients access Greenplum databases and perform operations. The connector can be used in data integration jobs to read or write data from Greenplum databases, or to look up data in the context of those jobs. Users can also import metadata from Greenplum databases using InfoSphere Metadata Asset Manager.

InfoSphere Information Server for Data Quality

InfoSphere Information Server for Data Quality provides capabilities to analyze, cleanse, monitor and manage data in both batch and real time. Two aspects define its usefulness: first, this package helps business and IT collaborate to align quantitative

metrics for business and governance objectives. Second, it helps users quickly and easily design, manage and monitor data quality in alignment with defined business policies.

To accelerate enterprises' ability to deliver high-quality, trusted data, InfoSphere Information Server for Data Quality incorporates the following new features:

- **Improved governance:** This package now incorporates the InfoSphere Information Governance Dashboard and its companion SQL views (a fully described query layer for key metadata). Expanded reports include analysis of critical business elements that lack a data steward. Organizations can also access IBM Cognos® Business Intelligence components to support deployment of an integrated data governance program for both IT and line-of-business users.
- **Matching stewardship:** Organizations can now find "clerical pairs" (records that are close matches but require review) in the Data Quality Console. Records are displayed in an intuitive format to facilitate investigation of the source data elements and related match scores.
- **Enhanced productivity:** Version 11.3 eliminates the need for a discrete metadata import activity by using IBM InfoSphere Information Analyzer to leverage the data connections and imported metadata from InfoSphere Metadata Asset Manager. It also enhances the Data Rule stage (for use within the job designer) with drag-and-drop functionality.
- **Optimized performance:** These improvements include new, faster algorithms for multi-column primary key detection, the ability to take full advantage of parallelism in column analysis, bulk-load write to the analysis database and optimized client/server communication to reduce network traffic.
- **Increased operational quality:** InfoSphere Information Server now provides pre-built data rules for operational metadata, so that operations teams can monitor how the data integration platform accommodates the service-level agreements and best practices of the organization.
- **Expanded connectivity:** InfoSphere Information Analyzer can now leverage native IBM PureData™ System for Analytics (formerly IBM Netezza®) connectivity mechanisms for data profiling and rule analysis.

InfoSphere Information Governance Catalog

InfoSphere Information Governance Catalog provides business-driven governance capabilities that help ensure information is correct, consistent and complete at the points where it directly impacts a business. Previously known as InfoSphere Business Information Exchange, it serves as a single location where users can find information sources that deliver business value. It offers a seamless user experience for creating, monitoring and publishing business-critical metadata that adds context and relevance to information in the "data lake." Business and IT users can collaborate and share insight about information sources, thereby speeding up the time to value for any project (including big data analytics, data warehousing, MDM and others) that requires well-governed, consistent and trusted data.

To improve enterprises' ability to govern data, InfoSphere Governance Catalog incorporates the following features:

- **Comprehensive data source search:** Users can search for any type of data source—tables, files, cloud structures and more—or business terms and policies, and then drill into the related data sources. This gives users the flexibility they need to find appropriate data sources quickly, explore the rich detail about those sources and confidently select the most relevant data for their purpose.
- **Integrated interface for business and technical metadata:** The governance catalog displays all technical and business metadata—including terms, policies and lineage, and queries—for enterprise data sources. By connecting the fully-featured business glossary and the metadata repository, users can better understand data's proper context.

- **Collections that support teaming:** Catalog users can create groups of metadata objects, known as collections, in order to better collaborate around sets of content with other users. These collections can be associated with business terms, stewards and labels to provide context for the group of assets. The creator of a collection can designate which users may view and edit the collection. Users who want to curate or refine the data sources in the collection can also launch directly into InfoSphere Data Click, which can copy the information from these data sources into the larger data lake.
- **New data lineage infrastructure:** Users get improved scalability and performance when using the new graphical Lineage Viewer and lineage composer service. Additionally, the UI now incorporates HTML 5 to enable access to lineage information from mobile devices.
- **Optimized operational metadata:** A new data flow model can efficiently process operational metadata from tens of thousands of data integration activities per day.

Key core enhancements in InfoSphere Information Server V11.3

Better big data, faster

InfoSphere Information Server V11.3 now includes InfoSphere BigInsights V3.0 to provide a fast, cost-effective and scalable approach to big data integration with Apache Hadoop.

Using the same simple design interface available in other data integration patterns, developers can now shift ETL and data integration workloads into a Hadoop infrastructure that is capable of simultaneously handling analytics and staging, preparation and transformation of data. InfoSphere Information Server manages the ingestion, mapping, metadata and quality processes, freeing data scientists and business users to begin exploring and analyzing data through the intuitive BigSheets interface.

A component of InfoSphere BigInsights, BigSheets allows users to visualize both traditional and new data types, such as JSON, that InfoSphere Information Server seeded into Hadoop.

Most important, organizations using InfoSphere Information Server with InfoSphere BigInsights gain the advantage of built-in data governance.

Support for public cloud environments

With Version 11.3, InfoSphere Information Server data quality, integration and governance functionality can be provisioned on the cloud in a matter of minutes. Thanks to the IBM holistic approach to cloud-hosted environments, InfoSphere Information Server Enterprise Hypervisor Edition can be combined with other software technologies to deliver a broad, end-to-end information integration solution. For more information, visit http://ibm.co/1vwL635

The cloud-hosted information integration platform is delivered through IBM PureApplication Service on SoftLayer®, which provides simple provisioning and configuration of IBM software within a public cloud environment. InfoSphere Information Server users can now extend their enterprise infrastructure into the cloud and provision data quickly, faster than traditional hardware acquisition and configuration cycles allow.

Enhanced security

Version 11.3 continues to meet the demand for strong security by delivering the following new features:

- InfoSphere Information Server browser-based clients now support single sign-on, so that customers who authenticate within one of the interfaces can seamlessly interact across other interfaces.
- SSL (Secure Sockets Layer) now provides communication security for all client interfaces.
- High-level RSA-2048 and SHA-512 security protocols provide the default encryption mechanisms.
- The IBM WebSphere® Application Server standard security domain provides cell-sharing, so that it can now be deployed into an existing cell managed by a secured deployment manager without disrupting existing profiles and applications.

New lighter-weight services tier

InfoSphere Information Server V11.3 users have the option to install either WebSphere Application Server Network Deployment (for highly available services tier configurations) or WebSphere Application Server Liberty Profile, a dynamic application server runtime environment for the services tier that helps decrease time during install and feature upgrades and lower overall system resource usage costs.

Why IBM?

As a critical element of IBM Watson™ Foundations, the IBM big data and analytics platform, InfoSphere Information Integration and Governance (IIG) provides market-leading functionality to handle the challenges of big data. InfoSphere IIG provides optimal scalability and performance for massive data volumes, agile and right-sized integration and governance for the increasing velocity of data, and support and protection for a wide variety of data types and big data systems. InfoSphere IIG helps make big data and analytics projects successful by delivering business users the confidence to act on insight.

For more information

To learn more about InfoSphere Information Server V11.3, please contact your IBM representative or IBM Business Partner, or visit the following website: **ibm.com**/software/data/integration/info_server

Additionally, IBM Global Financing can help you acquire the software capabilities that your business needs in the most cost-effective and strategic way possible. We'll partner with credit-qualified clients to customize a financing solution to suit your business and development goals, enable effective cash management, and improve your total cost of ownership. Fund your critical IT investment and propel your business forward with IBM Global Financing. For more information, visit: **ibm.com**/financing

IBM Corporation
Software Group
Route 100
Somers, NY 10589

Produced in the United States of America
June 2014

IMW14776-USEN-00