Anywhere integration with IBM InfoSphere DataStage V11.3

Highlights

- Scales for data of any size, regardless of volume and complexity

- Provides agile, reusable integration across diverse sources

- Helps users quickly respond to business changes

- Delivers tight integration for master data management (MDM) initiatives

- Enables self-service data integration for traditional and big data projects

- Delivers integration for cloud environments

- Provides enriched security

Built on a massively parallel processing (MPP) architecture, IBM® InfoSphere® DataStage® V11.3 is designed to help organizations transform and integrate large volumes of heterogeneous data. InfoSphere DataStage enables users to design jobs once and deploy anywhere, resulting in improved performance, greater integration agility and lower costs.

A new era of computing is unfolding, bringing with it an ever-accelerating explosion in the volume, variety and velocity of data. To achieve their business intelligence goals and gain a competitive edge, organizations must be able to integrate data from anywhere in their environment and transform it into trusted information. The ability to integrate information quickly and efficiently is crucial in this dynamic environment, even as those requirements continue to shift and change and data volumes increase.

InfoSphere DataStage V11.3 meets this need with robust and comprehensive "anywhere" information integration capabilities, helping organizations flexibly integrate data wherever it resides—whether on the mainframe, in big data sources or in the cloud. Flexible capabilities built into the software, including the ability to leverage common enterprise-wide business rules, enable users to prioritize and streamline tasks and quickly react to changing business conditions.

New features address complex data integration challenges

InfoSphere DataStage V11.3 offers powerful new capabilities for today's information-rich environments.
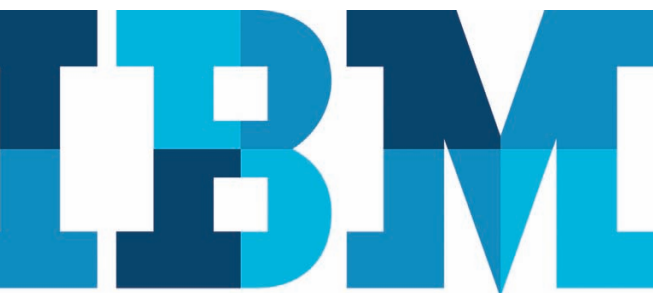
Data integration for cloud environments

InfoSphere DataStage V11.3 provides quick and easy data integration for cloud environments. When running on premises, it supports direct integration with the Amazon Simple Storage System (S3) to load data from and into the cloud. Once data is integrated within S3, it may be then be integrated with other cloud database technologies.

Additionally, InfoSphere DataStage V11.3 includes a new Hierarchical Stage (renamed and expanded from the "XML Stage") that supports interaction with REST application programming interfaces (APIs), enabling integration support for XML and JavaScript Object Notation (JSON) messages. The REST-based connectivity enables InfoSphere Information Server 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.

Tight integration for master data management projects

In V11.3, InfoSphere DataStage can directly integrate with InfoSphere MDM through an MDM Integration Stage, enabling users to easily employ InfoSphere DataStage to load data into and extract data out of InfoSphere MDM. By leveraging the MDM Integration Stage, customers can include MDM data within their data integration flows and can load domain data (such as customer, partner, supplier, product and other data) directly to an MDM system. Additionally, users may now benefit from peak scale and performance by leveraging bulk loading to achieve the best performance.

In addition to benefitting from the InfoSphere Information Server data transformation and delivery styles, the MDM Integration Stage can leverage InfoSphere Information Server data quality capabilities as well. For example, InfoSphere Information Server can standardize any data before it is loaded into the MDM system. By proactively addressing data quality in this way, projects benefit from improved matching accuracy and better support for 360-degree views of various entities.

Big data integration: A key to big data success

A fast and easy way to get data into and out of big data distributions is a must-have in today's fast-moving business climate. Thanks to an MPP architecture, InfoSphere DataStage can easily scale to meet the demanding data integration requirements of Hadoop environments. The InfoSphere DataStage "design once, run anywhere" approach allows users to easily move data integration tasks between a single machine and a cluster of low-commodity servers. Because job design is isolated from runtime, developers can focus on the business requirements at hand rather than coding explicitly for a given architecture.

InfoSphere DataStage always has been MPP-based, and it is the original integration platform to support extreme (big) data volumes—even before the term "big data" came into vogue. InfoSphere DataStage delivers some powerful big data–specific capabilities to ensure your big data projects proceed quickly and cost-efficiently. These include:

- **Balanced optimization for Hadoop:** This feature allows integration designers to build a job using the same design paradigm they use with traditional ETL and, when desired, run that logic through generated MapReduce. This eliminates the cost of retooling and training the team on MapReduce languages or other secondary toolsets that would only apply to this particular environment.
- **Big data job sequencing:** InfoSphere DataStage allows any Oozie-contained MapReduce job to be included in the job sequencer. Organizations can build workflows that load data to Hadoop, run a custom-developed MapReduce analytics program, and then load the analytical result to the data warehouse—all within a single graphical workflow construct. This provides end-to-end workflow across heterogeneous topologies executed in both InfoSphere Information Server and Hadoop.

- **Big data governance:** InfoSphere DataStage supports big data-related governance features such as impact analysis and data lineage on any integration points, enabling scalable analytics without sacrificing organizational insight.
- **IBM InfoSphere Streams integration:** For big data projects that focus on real-time analytical processing, IBM now offers direct data flow integration between InfoSphere DataStage and InfoSphere Streams. Organizations can use standard data integration conventions to gather and pass information to real-time analytical processes.
- **Big data accelerators:** These open source components are available on **ibm.com**/developerworks. They plug directly into the InfoSphere DataStage canvas and operate just like any other stage. Accelerators are available for MongoDB, Hive, Cassandra, HBase, Avro and more.

- **Expanded hierarchical data support:** Version 11.3 expands InfoSphere DataStage support to provide REST API support, allowing easy access to and integration of hierarchical data, such as XML and JSON messages.

Agile, self-service data provisioning and governance

InfoSphere Data Click, a feature now available as part of InfoSphere DataStage V11.3, helps speed up time-to-value, increases business agility and lower costs by shrinking the time required to complete tasks from days or weeks to minutes or hours. It provides broad connectivity, helping improve the timeliness of many different types of environments, including big data landing zones, data warehouses and cloud environments. InfoSphere Data Click provides both specific native DBMS connectivity for fast performance and near-universal connectivity for a very wide range of data sources through JDBC and ODBC (see Figure 1).
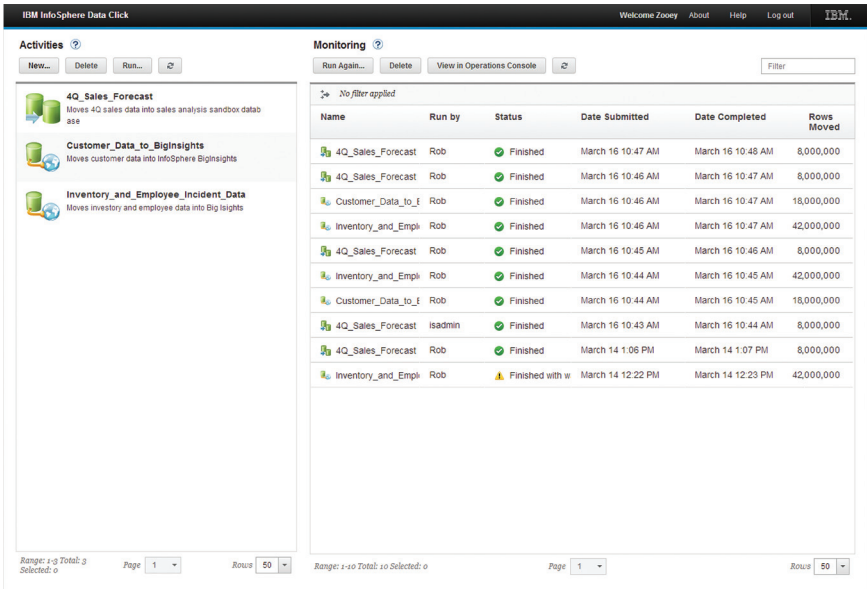

Figure 1. Users log directly into a streamlined InfoSphere Data Click UI. From here, they can quickly and easily create, edit, run and monitor activities.

Data can be sourced from or sent to a whole host of traditional and big data environments, including DB2, Oracle, EMC Greenplum, IBM Informix®, IBM PureData™ for Analytics (based on IBM Netezza®), Salesforce.com, Microsoft SQL Server, Teradata, Amazon S3 and more. InfoSphere Data Click also allows any user to easily provision data within their IBM InfoSphere BigInsights™ (that is, Hadoop) environment.

Enhanced security for InfoSphere Information Server

InfoSphere Information Server delivers common platform services (such as connectivity services, administration services, deployment services, etc.) to support its data integration, data quality and data governance capabilities. Along with cloud enhancements and additional support for massive data volumes, IBM incorporated additional security enhancements into InfoSphere Information Server V11.3, including:

- **Single sign-on (SSO):** Browser-based clients now support SSO, allowing customers who authenticate within one of the interfaces to more seamlessly interact across other user interfaces.
- **Secure Sockets Layer (SSL) communication:** InfoSphere Information Server employs SSL to provide communication security for all client interfaces.
- **Stronger encryption:** Version 11.3 delivers RSA-2048 and SHA-512 as default encryption mechanisms.
- **Cell sharing:** InfoSphere Information Server now fully leverages the IBM WebSphere® Application Server standard security domain. As such, it can be deployed into an existing cell managed by a secured deployment manager without disrupting the profiles and applications already deployed in that cell.

Optimized performance

InfoSphere DataStage V11.3 includes a number of enhancements to improve performance and reduce I/O and scratch space. For sort operations on bounded-length data, InfoSphere DataStage V11.3 can achieve up to 49 percent performance improvements and up to 92 percent scratch disk and I/O reduction.[1]

InfoSphere Information Server V11.3 now also has a lighter-weight services tier. Customers have the option to install either IBM WebSphere Application Server Network Deployment (WAS ND) or WebSphere Application Server Liberty Profile (WAS Liberty Profile). WAS ND provides a highly available services tier configurations, while WAS Liberty Profile helps decrease installation time, reduce feature pack cycle time and lower overall system resource usage costs.

A step further: InfoSphere Information Server for Data Integration

Together, InfoSphere DataStage and InfoSphere Information Server enable organizations to flexibly and robustly manage big data from new and emerging sources.

InfoSphere DataStage users with broader data integration requirements may also be interested in InfoSphere Information Server for Data Integration V11.3. This solution includes InfoSphere DataStage, along with additional, frequently requested data integration capabilities such as change data delivery, real-time data integration, data modeling, blueprinting, data discovery and governed metadata management. It also contains functionality to accelerate design time, creating source-to-target mappings and automatically generating jobs.

To meet other information integration requirements, IBM also offers InfoSphere Information Server for Data Quality, InfoSphere Information Governance Catalog and InfoSphere Information Server Enterprise Edition. The last of these packages is a comprehensive offering that includes all three sets of capabilities—data integration, data quality and information governance.

Beginning the big data journey with IBM Watson Foundations

IBM Watson™ Foundations, the IBM big data and analytics platform which includes InfoSphere Information Server and InfoSphere DataStage, can play an integral role in your big data journey. The components of Watson Foundations help you reduce the time and costs of projects, as well as achieve a rapid return on investment (ROI) by leveraging pre-integrated components. By building on those capabilities, you can start small with an initial use case and easily progress to others as you continue on your big data journey.

Within Watson Foundations, InfoSphere solutions provide a comprehensive information integration and governance platform that helps organizations:

- Adopt analytics based on a foundation of trusted, timely information
- Improve the efficiency of applications
- Secure enterprise data
- Consolidate and retire applications
- Build a single view
- Lower the cost of data

Why IBM?

As a critical element of IBM Watson Foundations, 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.

InfoSphere capabilities include:

- **Metadata, business glossary and policy management:** Define metadata, business terminology and governance policies with InfoSphere Information Governance Catalog.
- **Data integration:** Handle all integration requirements, including batch data transformation and movement (InfoSphere Information Server), real-time replication (InfoSphere Data Replication) and data federation (InfoSphere Federation Server).
- **Data quality:** Parse, standardize, validate and match enterprise data with InfoSphere Information Server for Data Quality.
- **MDM:** Act on a trusted view of your customers, products, suppliers, locations and accounts with InfoSphere MDM.
- **Data lifecycle management:** Manage the data lifecycle from test data creation through retirement and archiving with IBM InfoSphere Optim™.
- **Data security and privacy:** Continuously monitor data access and protect repositories from data breaches, and support compliance with IBM InfoSphere Guardium®. Help ensure that sensitive data is masked and protected with InfoSphere Optim.

For more information

To learn more about InfoSphere Information Server, InfoSphere Information Server for Data Integration or InfoSphere DataStage, please contact your IBM representative or IBM Business Partner, or visit the following websites:

- **ibm.com**/software/data/integration/info_server/data-integration
- **ibm.com**/software/data/integration/data/integration

If you are using InfoSphere DataStage V8.5 or below, and you would like to get a better understanding of the features and enhancements introduced as part of InfoSphere Information Server V8.7 and InfoSphere Information Server V9.1.2, please refer to these white papers:

- What's new in InfoSphere Information Server V8.7: http://bit.ly/1kD9HQe
- What's new in InfoSphere Information Server V9.1 and V9.1.2: http://bit.ly/1jWVatP

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

[1] IBM lab testing.

Please Recycle

IMD14372-USEN-02