



---

## Highlights

- Easy-to-use GUI with drag-and-drop data connectors
  - Accelerated development through built-in transformation functions
  - Comprehensive and customizable data cleansing
  - Automated data survey and classification for improved governance
- 

# IBM BigInsights BigIntegrate and BigInsights BigQuality

*Information empowerment for your big data ecosystem*

Big data analytics has emerged as a key driver to deriving business value and other organizational advantages from exploding data volumes within the enterprise. While many savvy organizations are already using Hadoop for its unparalleled analytical opportunities and cost benefits, Hadoop itself is not designed for data integration. Successful data integration has its own unique requirements, including support for governance, metadata management, data quality and flexible data delivery styles.

Still, many essential data integration capabilities in the Hadoop stack are either immature or missing entirely, and have yet to be addressed in ongoing projects. More importantly, many organizations are realizing that Hadoop alone is not robust enough for successful data integration, and the skill and effort required to build a comprehensive solution is unsustainable.

To help organizations fully leverage the scale and promise of Hadoop, IBM has designed a powerful and cost-effective solution that supports critical projects and key analytics initiatives within the entire data ecosystem. Built using the IBM® InfoSphere® Information Server, IBM BigInsights® BigIntegrate and BigInsights BigQuality provide the end-to-end information integration and governance capabilities that allow organizations to:

- Understand data
- Cleanse, monitor, transform and deliver data
- Collaborate across the organization and bridge the gap between business and IT

By leveraging the flexible integration capabilities delivered by IBM BigInsights BigIntegrate and BigInsights BigQuality, organizations can ensure the information driving business and strategic initiatives—from systems of record to systems of insight—is trusted, consistent and governed in real time.





## **IBM BigInsights: Analytics on large data sets**

BigInsights helps organizations find patterns by looking at all the data, not just samples. To achieve analytics on very large data sets, BigInsights provides a full range of massively scalable machine-learning algorithms (formerly known as SystemML from IBM Research) that are programmable using the familiar R language. A sufficiently large data set is essential for building a successful statistical model. Better models enable more accurate predictive capabilities and ultimately lead to better results. As part of the BigInsights product family, BigInsights BigIntegrate and BigInsights BigQuality provide the means to assemble, transform, cleanse and make data available for downstream analysis.

## **BigInsights BigIntegrate and BigInsights BigQuality: Native data integration**

BigInsights BigIntegrate and BigInsights BigQuality enable organizations to integrate and transform any data with Hadoop, leveraging both existing and new data sources for big data initiatives.

An easy-to-use graphical interface helps organizations quickly transform information from across the enterprise. The feature-rich palette includes connectors to a very wide range of data sources including all major traditional databases on all platforms (distributed, IBM z/OS® and so on), file types, enterprise applications (including Oracle, Salesforce.com and SAP) and Hadoop. Developers make these data sources available through simple drag-and-drop capabilities, and adjust configuration parameters to maximize performance and flexibility.

BigInsights BigIntegrate and BigInsights BigQuality help improve performance and are highly scalable. Native connectivity to most commonly used data sources is available through specific interfaces, and hundreds of built-in transformation functions help accelerate development. As a result, many coding tasks can be completed in a fraction of the time compared to hand coding. In one study, a developer took two days to finish a task that took 30 days writing almost 2,000 lines of code without any proper documentation, lack of reuse and significant maintenance challenges.

Recent features in Hadoop use the same simple graphical design environment that IBM customers already have been using to build millions of integration applications. Within this environment, organizations can build new Hadoop-based, information-intensive applications without the need to retrain their development team on newly emerging languages that require manual hand coding and lack governance support.

## **Tight collaboration with fully integrated design tools**

An integrated design experience and common metadata framework help ensure cross-collaboration as all groups work from the same, consistent metadata. Through the graphical interface, end users can automate the design specification process and enrich metadata with a comprehensive, linked business view of technical assets. Integrated design tooling enables developers to build jobs that combine data integration and data quality processes in a single job artifact to achieve the highest level of efficiency. With a common design approach, organizations also have the flexibility to add quality assurance into the process, wherever it makes most sense.



In addition, BigInsights BigIntegrate and BigInsights BigQuality offers industry-leading integration and governance features. Data integration and transformation jobs can execute directly inside a Hadoop cluster, so powerful data connectivity, transformation, cleansing, enhancement and data delivery features are immediately available to run within the Hadoop platform. This approach helps solve some of the most complex data challenges since it minimizes data movement, thereby accelerating the time to deliver business value.

### **BigInsights BigQuality: Native data cleansing for Hadoop use cases**

BigInsights BigQuality enables organizations to cleanse data and monitor data quality continually, ensuring data becomes trusted information. Reliable data is essential to deriving business value from existing data resources, so organizations can improve the decision-making process and streamline business operations.

Comprehensive and customizable data cleansing functionality in batch and in real time automates source data investigation and data classification. This automation enables the data steward team to manage data assets effectively, and respond to business objectives with trusted data. Automation features also help process and manage the exploding amount of data now sent to Hadoop, scaling as required to keep up with volume.

Rapidly growing volumes of data also mean that as the community of data providers and consumers expands, so does the uncertainty over data sensitivity and how to comply with mandated regulatory requirements. BigInsights BigQuality helps to survey the various data sources, including Hadoop, and ensures appropriate data location and usage according to predefined policies. Furthermore, BigInsights BigQuality helps standardize and match records in accordance with customizable business rules. By establishing data quality metrics aligned with business objectives, organizations can quickly uncover quality issues and establish a remediation plan. These capabilities are key to maintaining an effective data governance program.

### **Why IBM?**

The new capabilities in BigInsights BigIntegrate and BigInsights BigQuality for Apache Hadoop expand the IBM Analytics portfolio of big data solutions, which spans software, services, research and hardware. The portfolio combines traditional data warehouse tools with current big data techniques and technologies—such as Hadoop, stream computing, data exploration, advanced analytics, enterprise integration and IBM Watson™ cognitive computing—to create an integrated solution that can bring big data and analytics into any organization.

### **For more information**

For more information about BigInsights BigIntegrate and BigInsights BigQuality and other data integration platform solutions, contact your IBM sales representative, or click here: [ibm.com/software/data/information-integration-governance](http://ibm.com/software/data/information-integration-governance)



---

© Copyright IBM Corporation 2015

IBM Analytics  
Route 100  
Somers, NY 10589

Produced in the United States of America  
September 2015

IBM, the IBM logo, ibm.com, BigInsights, IBM Watson, InfoSphere, and z/OS are trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the web at "Copyright and trademark information" at [ibm.com/legal/copytrade.shtml](http://ibm.com/legal/copytrade.shtml)

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.

The client is responsible for ensuring compliance with laws and regulations applicable to it. IBM does not provide legal advice or represent or warrant that its services or products will ensure that the client is in compliance with any law or regulation.



Please Recycle