

# IBM InfoSphere BigInsights Enterprise Edition

*For turning complex, internet-scale information into insight, cost-effectively.*

## Highlights

- Enterprise-ready platform meets big data requirements
- Integrates with popular IBM offerings such as Netezza and DB2
- Supports structured, semi-structured, and unstructured data for maximum flexibility
- Allows data models to be introduced on the fly
- Supports massive scale-out
- Runs on commonly available hardware

## Enables a new class of solutions to gain insight from massive volumes of raw data

Many companies are seeing dramatic growth in the variety, velocity, and volume of information being generated by their businesses. Organizations are struggling with how to manage these vast and diverse data sets that includes traditional structured data as well as semi-structured or unstructured data types including sensor data, web pages, web log files, click-streams, AVI files, search indexes, text messages, and email, to name a few. These large, untapped data sets define a new category of information—big data. This data offers tremendous potential for deep insights that drive faster, clearer, and more nuanced decision-making. Companies need an approach that allows this information to be effectively understood and analyzed.

IBM® InfoSphere™ BigInsights Enterprise Edition enables new solutions that turn large, complex volumes of data into insight, cost effectively. InfoSphere BigInsights delivers an enterprise-ready big data solution by combining Apache Hadoop, including the MapReduce framework and the Hadoop Distributed File Systems (HDFS), with unique technologies and capabilities from across IBM.

## Use the power of distributed computing to process big data

How can businesses process tremendous amounts of raw data in an efficient and timely manner to gain actionable insights? InfoSphere BigInsights allows organizations to run large-scale, distributed analytics jobs on clusters of cost-effective server hardware. This infrastructure can be leveraged to tackle very large data sets by breaking up the data into "chunks" and coordinating the processing of the data across a massively parallel environment. Once the raw data has been stored across the nodes of a distributed cluster, queries and analysis of the data can be handled efficiently, with dynamic interpretation of the data format at read time.

The bottom line: Businesses can finally get their arms around massive amounts of untapped data, and mine that data for valuable insights in a more efficient, optimized, and scalable way.



## Bringing big data to the enterprise

InfoSphere BigInsights delivers a Hadoop-based offering with the enterprise-class functionality and integration necessary to ensure client success.

### ***Designed for complex data management and analysis:***

- Provides a solution that allows organizations to extract insight from a wide range of data, including data that's often ignored or discarded.
- Makes it possible to load and make sense of various types of data in their native formats, without needing to pre-define and map data to a single schema or structure.
- Leverages automatic parallelization and resource optimization to scale economically.

### ***Quick installation and deployment:***

To help enterprises get off to a quick start with their big data projects, BigInsights provides a pre-tested and pre-configured platform that features a compelling collection of popular open source and IBM technologies. Instead of iteratively downloading, configuring, and testing the individual open source projects for a comprehensive software platform, firms can invoke BigInsights' integrated, GUI-driven installation tool and enjoy a working platform within minutes. This saves considerable time, particularly since many open source projects have specific software prerequisites that may be incompatible with certain versions of other necessary components. Furthermore, some open source offerings require code compilation and other post-installation efforts that can lengthen the time necessary to deploy a working environment.

BigInsights' flexible installation tools allows you to specify which optional components to install and how to configure your platform. Details about the progress of a BigInsights installation are reported in real-time, and a built-in "health check" automatically verifies the success of the installation.

### ***Enterprise-class management and security features:***

BigInsights provides a complete solutions that allows organizations to enjoy the benefits of using leading-edge technologies with the usability that make it an enterprise-ready offering, including:

- Administrative features including Hadoop HDFS and MapReduce administration, cluster and server management, ability to view HDFS file content, and role-specific views.

- Robust job management features including: job creation, submission, cancellation capabilities, job status display and auto/manual refresh, and job details including job configuration info, task and attempt info, logs and counters.
- Security features that include authentication using LDAP or flat files, role based authorization, and reverse proxy features.

### ***Built-in analytics that extends the value of Hadoop:***

BigInsights can help firms analyze and process large volumes of text-based information with its built-in text processing engine and library of annotators. With BigInsights, developers can quickly query and identify items frequently of interest in documents and messages, including persons, email addresses, street addresses, phone numbers, URLs, joint ventures, alliances, and more.

IBM has also developed Jaql, a powerful query language that has been contributed to the open source community. Jaql is a high level declarative language that provides the capability to process both structured and non traditional data. Its SQL-like interface is well suited for quick ramp-up of developers familiar with SQL languages, and makes it easier to integrate with relational databases. Jaql is also highly extensible via modules, and IBM is leveraging this capability to include useful pre-built Jaql modules in BigInsights to enable integration, e.g. with text analytics, and Netezza warehouses.

### ***Reduced operational risk:***

As you might expect, IBM's standard software licensing and support agreements apply to BigInsights. This eliminates legal concerns that some firms have about the terms and conditions of certain open source projects and minimizes the operational risk of deploying software that doesn't include a technical support contact.

### ***Integration with existing systems:***

InfoSphere BigInsights supports open development standards and works with existing systems, including IBM Netezza, InfoSphere Warehouse, IBM Smart Analytics System and IBM DB2 for Linux, UNIX, and Windows.

### ***Free Download:***

#### ***InfoSphere BigInsights Basic Edition***

For developer, partners, and anyone who wants to test drive the technology, IBM offers InfoSphere BigInsights Basic Edition, a no-cost, entry-level edition of InfoSphere BigInsights. Users have the option to purchase 24x7 Elite Support. Solutions developed using InfoSphere BigInsights Basic Edition can be seamlessly deployed using InfoSphere BigInsights Enterprise Edition.



## Hardware requirements & operating system support

- Intel ® x86 servers, 64-bit, with a minimum of 4 GB memory and 30 GB of disk storage
- Red Hat Enterprise Linux® (RHEL) 5.3, 5.4, or 5.5
- SUSE Linux Enterprise (SLE) 11

Note: Refer to product documentation for up-to-date operating system version support.

## Why IBM?

Customers choose IBM for big data because they know they are getting a solution that is:

- **Comprehensive:** IBM's customer tested platform for big data can help you use *ALL* your data as a strategic asset.
- **Integrated:** IBM simplifies and accelerates the introduction of big data technologies to your enterprise by enabling integration with your information supply chain including the data warehouse, database, business process management solutions, and information integration solutions.
- **Enterprise-ready:** Has the performance and reliability that you expect and need from a solution that will become integral to your business' decision-making processes. Supplemented by 24x7 support offerings, education and a worldwide professional services organization.
- **Open source based:** Brings the best of big data open source technologies to your enterprises with the necessary support and refinements. In keeping with our history, we contribute back to the community to expand the breadth and usability of these technologies.

## For more information

Interested in InfoSphere BigInsights? Contact your IBM sales representative today and visit:

[www.ibm.com/software/data/infosphere/biginsights](http://www.ibm.com/software/data/infosphere/biginsights)

© Copyright IBM Corporation 2011  
IBM Corporation  
Software Group  
Route 100  
Somers, NY 10589 U.S.A.  
Produced in the United States of America  
June 2011  
All Rights Reserved

IBM, the IBM logo, ibm.com and InfoSphere are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries, or both. If these and other IBM trademarked terms are marked on their first occurrence in this information with a trademark symbol (® or ™), these symbols indicate U.S. registered or common law trademarks owned by IBM at the time this information was published. Such trademarks may also be registered or common law trademarks in other countries. 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)  
Other product, company or service names may be trademarks or service marks of others.

