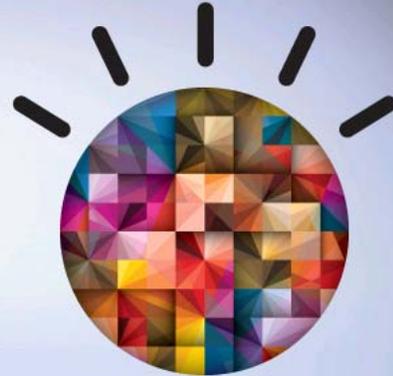




Bringing Big Data to the Enterprise



The value of InfoSphere BigInsights

IBM Silicon Valley Laboratory
May 2011

The value of IBM's InfoSphere BigInsights

You've probably heard about InfoSphere BigInsights, IBM's software platform that helps firms discover and analyze business insights hidden in large volumes of a diverse range of data – data that's often ignored or discarded because it's too impractical or difficult to process using traditional means. Examples of such data include log records, click streams, social media postings, news feeds, instant messages, emails, electronic sensor output, and even some transactional data. To help firms process these and other forms of data that are increasing in volume, variety, and velocity, IBM developed BigInsights, a platform that includes IBM's distribution of the open source Hadoop™ project from Apache™. You may be wondering what makes IBM's platform unique – in short, what value it provides. Let's take a closer look.

Built-in Analytics

BigInsights helps firms analyze large volumes of text-based information with its built-in text processing engine and library of annotators. Developers can quickly query and identify items of interest in documents and messages, such as persons, email addresses, street addresses, phone numbers, URLs, joint ventures, alliances, and more. Such capabilities can help firms understand the context and content of relevant business information hidden in unstructured text data.

Enterprise software integration

Integrating BigInsights – and its analysis of “Big Data” – with existing enterprise software is one of IBM's key initiatives. That's why BigInsights provides connectivity to popular data warehouse platforms, such as DB2 and Netezza.

In particular, BigInsights provides developers with JDBC connectivity to a remote Netezza data warehouse so they can retrieve information of interest and combine it with data managed by BigInsights as desired. In addition, BigInsights provides two sample DB2 user-defined functions that firms can

deploy so DB2 users can launch queries in BigInsights, join the output with DB2 data, and present the results to DB2 users and applications.

IBM's approach ensures that your “Big Data” and traditional enterprise data won't exist in isolation. Instead, you can use your existing enterprise software platforms to do what they do best, leverage BigInsights for analytical workloads that aren't appropriate or practical for these platforms, and broaden your business analysis capabilities in an integrated fashion.

Administrative and platform enhancements

IBM's Web-based management console provides a real-time view of your BigInsights environment. Through this console, you can start and stop nodes, inspect the status of MapReduce jobs, review log records, assess the overall health of your platform, start and stop optional components, navigate your distributed file system, and more.

In addition, BigInsights provides enhanced security by supporting LDAP authentication to its Web console.

LDAP and reverse proxy support help administrators restrict access to users with appropriate authorization.

Firms concerned about workload management can use a flexible job scheduling mechanism to fine tune resource allocation among long-running and short-running jobs. For example, the BigInsights scheduler can be directed to allocate maximum resources to small jobs to help ensure they complete quickly (and thereby fulfill *average response time* objectives). This job scheduling option is available in addition to Hadoop's first in/first out (FIFO) and "fair" scheduling approaches.

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.

Deploying mission-critical applications on a platform comprised of multiple open source projects requires firms to maintain deep in-house expertise of the code base to fix any bugs or shortcomings that may surface. While community-based support groups and forums can be very helpful, they're not obligated to fulfill any service-level agreements or respond to urgent inquiries. Hence, diagnosing and resolving platform problems – as well as upgrading the platform to maintain state-of-the-art technology – becomes each firm's responsibility. With BigInsights, firms avoid this complexity and operational risk.

Quick installation and deployment

To help firms 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 pre-requisites that may be incompatible with certain versions of other desired components. Furthermore, some open source offerings require code compilation and other post-installation efforts that can lengthen the time necessary to get up and running.

BigInsights' flexible installation tool 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.

If you're curious about the various technologies you'll find in BigInsights, refer to IBM's announcement letter and the BigInsights InfoCenter for details. However, very briefly, you'll find that the Enterprise Edition includes IBM's distribution of Hadoop, Pig, Hive, HBase, Jaql, Lucene, ZooKeeper, Flume, Oozie, Avro, Web Management Console, text analytics, and more.

IBM: Your source for Big Data software and expertise

IBM's depth and breadth of expertise spans a wide range of enterprise software, hardware, and services, enabling firms who partner with IBM to approach their Big Data projects with confidence and clarity. Visit <http://www.ibm.com/bigdata> today to learn what IBM can do for you.



© Copyright IBM Corporation, 2011
IBM
Armonk, NY
USA

All Rights Reserved.

Neither this documentation nor any part of it may be copied or reproduced in any form or by any means or translated into another language, without the prior consent of the IBM Corporation.

IBM and the IBM logo are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries or both.

Other company, product and service names may be trademarks or service marks of others.

The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law: INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

Any performance data contained herein was determined in a controlled environment. Therefore, the results obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurement may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

The information contained in this document is subject to change without any notice. IBM reserves the right to make any such changes without obligation to notify any person of such revision or changes. IBM makes no commitment to keep the information contained herein up to date.

References in this publication to IBM products or services do not imply that IBM intends to make them available in all countries in which IBM operates.

C. M. Saracco (saracco@us.ibm.com)