

IBM InfoSphere BigInsights Enterprise Edition

经济高效地存储、管理大数据并从中获得洞察

要点

- 在企业就绪的平台上提供大数据分析
 - 集成非IBM解决方案，以及流行的IBM产品，如IBM® Netezza®设备、IBM DB2® 数据库软件、IBM InfoSphere® Streams和IBM InfoSphere DataStage®
 - 支持结构化、半结构化和非结构化数据，提供最高的灵活性
-

许多公司都看到其业务所生成的信息在种类、速度和数量方面呈现大幅度的增长。各种组织都在努力管理大量的各种传统结构化数据，以及半结构化或非结构化数据——各种未开发的大型数据集定义了一个新的信息类别：大数据。组织看到了深入洞察的巨大潜力，它可以推动快速实现清晰且细致的决策，但他们需要的数据管理和分析工具与以往任何时候相比，都要在全新的水平上有效才行。

IBM InfoSphere BigInsights Enterprise Edition让组织能够创建新的解决方案，以经济高效的方式将大量的复杂数据转化为洞察。此企业就绪分析平台结合了Apache Hadoop与独特的IBM创新，提供大规模向外扩展式的数据处理和分析，并具有内置的弹性和容错功能。

底线：企业最终可以得到大量未开发的数据，并以高效的、优化的和可扩展的方式深入挖掘这些数据，从中获得有价值的洞察。



将大数据带进企业

InfoSphere BigInsights采用开源Hadoop, 并增加了必要的企业级功能和集成, 可帮助企业满足各种关键业务需求。组织可以在经济高效的服务器硬件集群上运行大型分布式分析作业。这种基础架构充分利用了Hadoop的MapReduce框架, 通过跨多个节点拆分数据, 并在一个大规模并行环境中协调数据的处理工作, 从而处理非常大型的数据集。一旦将原始数据存储在分布式集群上, 就可以高效地处理数据的查询和分析工作, 并且在读取时提供数据格式的动态解释。

InfoSphere BigInsights提供了一个已经过全面测试和集成的解决方案, 结合了领先技术的优势与成熟的企业就绪特性。管理员首先使用一个GUI驱动的安装工具, 这样可快速启动和运行该解决方案。引导式安装让管理员可以指定要安装哪些可选组件以及如何配置平台。系统实时报告安装进度, 还有一个内置的健康检查, 其目的是自动验证安装是否成功。这些高级安装特性最大限度地减少了安装和调试所需的时间, 让管理员可以专注于其他重要项目。

一旦安装完该解决方案, InfoSphere BigInsights就可以提供有助于简化工作负载管理和系统管理的企业级特性(参见图1)。例如, 强大的作业管理特性为组织提供其InfoSphere BigInsights作业的所有方面的细粒度控制。技术人员可以很轻松地引导作业的创建、提交和取消工作; 他们也可以随时了解工作负载的进度, 集成的作业状态显示、日志和计数器提供了有关配置、任务、尝试和其他重要信息的大量细节。此外, InfoSphere BigInsights提供了广泛的管理特性, 包括Hadoop Distributed File System (HDFS)和MapReduce管理、集群和服务器管理、查看HDFS文件内容的能力, 以及角色特定的视图。



图1: 利用InfoSphere BigInsights探索任务, 并链接到Web控制台功能。

立即试驾: 下载InfoSphere BigInsights Basic Edition

IBM为开发人员、合作伙伴和愿意尝试该技术的任何人提供了InfoSphere BigInsights Basic Edition, 这是免费的入门级版本的InfoSphere BigInsights。可以在您所选择的供应商的云中下载并运行BigInsights Basic Edition。有关更多信息, 请访问:

ibm.com/software/data/infosphere/biginsights/basic.html

基本版的用户可以选择购买24x7 Elite Support。您可以将使用InfoSphere BigInsights Basic Edition开发的解决方案无缝地部署到InfoSphere BigInsights Enterprise Edition中。

为大数据分析增加安全性

涉及安全性时，企业有着非常严格的要求。InfoSphere BigInsights提供一些复杂的选项，帮助确保数据的安全性和私密性。

身份验证

管理员可以为InfoSphere BigInsights控制台选择平面文件、轻量级目录访问协议(LDAP)或不使用身份验证。如果使用LDAP身份验证，InfoSphere BigInsights安装程序将与LDAP凭据存储进行通信，从而完成身份验证。然后，管理员可以基于角色成员关系来提供InfoSphere BigInsights控制台的访问，因此能轻松地设置用户组的访问权限。

授权

InfoSphere BigInsights提供四个级别的用户授权，称为以下角色：System Administrator、Data Administrator、Application Administrator和Non-Administrative User。用户对数据和特性的访问权限取决于为该用户所分配的角色。

最大限度提高性能，简化作业处理

IBM添加了一些功能，有助于提高性能，让InfoSphere BigInsights更灵活并且可兼容企业环境。

BigInsights Scheduler实现工作流分配

并非所有工作负载都具有相同的优先级，BigInsights Scheduler为MapReduce作业提供了一个自适应的工作流分配模式，可以根据用户选择的策略来优化处理工作。此调度程序是Hadoop Fair Scheduler的扩展，其目的是随着时间的推移，保证所有作业都会得到公平的集群资源份额。

BigIndex实现大规模索引

BigIndex让基于Hadoop的索引成为BigInsights中的一个本地功能，从而有助于让该索引更简单。在Apache Lucene的基础上，BigIndex为大数据提供低延迟的全文搜索功能。使用BigIndex模块作为工作流的一部分，可以构建、扫描和查询索引。BigIndex也支持其他复杂的功能，如分布式索引和分面搜索，这反过来又为自定义应用程序开发和搜索技术选择提供了极高的灵活性。

Adaptive MapReduce实现作业加速

在InfoSphere BigInsights上运行的作业往往最终会创建多个小任务，它们使用了与其大小不成比例的大量系统资源。为了解决这个问题，IBM发明了一种称为Adaptive MapReduce的技术，通过改变MapReduce任务的处理方式来加快小任务的执行，这种方法不需要改变作业创建方式。Adaptive MapReduce对于MapReduce操作和Hadoop API操作而言是透明的。

使用分析加速器深入了解大数据

InfoSphere BigInsights包括了丰富多样的分析工具和功能。由于它是开箱即用的，所以组织很快就能发现其数据中的模式，并且可以构建功能强大的自定义分析应用程序，以交付针对特定业务需求而定制的成果和洞察。

BigSheets

BigSheets代表了数据分析工具的一次革命，它是一个基于浏览器的工具，让业务用户能够浏览在BigInsights集群中存储的数据并创建分析查询，无需编写任何代码(参见图2)。内置的分析宏可满足常见的数据浏览需求，进一步提高了数据的可访问性。BigSheets可以帮助业务用户：

- 将来自基于Web的存储库的GB、TB或PB级非结构化数据进行集成
- 收集源于用户定义的种子URL的大量非结构化Web数据
- 使用文本分析提取Web数据并提高其质量
- 在特定的、用户定义的上下文中浏览和可视化数据

BigSheets将大数据分析带给企业的业务用户，为他们提供执行临时分析(并开发自己的洞察)的工具，而无需IT支持。

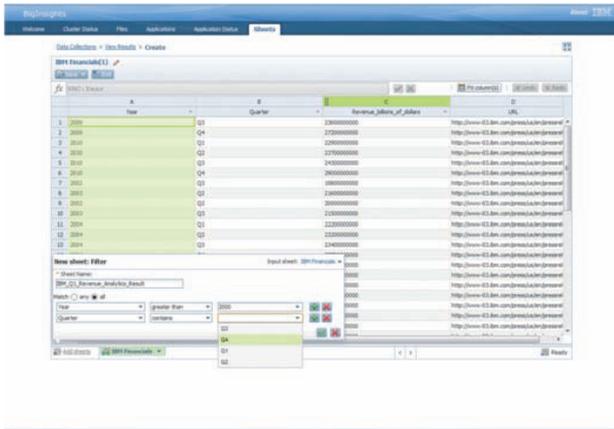


图 2: BigSheets 基于浏览器的工具为用户提供了对非结构化数据的快速访问，并能快速创建查询，而无需编写任何代码。

高级文本分析加速器

BigInsights包括由IBM开发的强大文本分析引擎，IBM Watson™在Jeopardy!问答比赛电视节目上使用该引擎击败了比赛中的两位最佳参赛者。使用一个全面的规则库(或他们自己的自定义规则)，开发人员可以快速查询和识别在文档和消息中感兴趣的项目，包括人、电子邮件地址、街道地址、电话号码、URL、合资、联盟等。该文本分析引擎支持英语、荷兰语/佛兰芒语、法语、德语、意大利语、葡萄牙语、西班牙语、日语和中文。

Jaql

由IBM开发并贡献给开源社区的一个强大的高级声明性查询语言，Jaql提供了处理结构化和非结构化数据的能力。它有一个类似SQL的接口，让熟悉SQL语言的开发人员可以很轻松地学习该语言，并且它有助于简化与关系型数据库的集成。InfoSphere BigInsights 自带了预构建的Jaql模块：IBM包括适用于Lucene索引、IBM Netezza数据仓库设备系列、HBase(Hadoop数据库)的Jaql模块，并且其工作流包括InfoSphere BigInsights内置的文本分析功能。

将大数据集成到现有的信息架构中

大数据技术在企业信息供应链中可以发挥重要的作用，但只有将它们与现有系统深入、紧密地集成时才可以做到。IBM已认识到这一点，所以InfoSphere BigInsights Enterprise Edition包括了适用于IBM Netezza数据仓库设备系列、IBM DB2、IBM InfoSphere Warehouse以及IBM Smart Analytics System的高速连接器。在与DB2或IBM仓库产品配合使用时，这些高速连接器有助于简化并加速数据处理任务。

InfoSphere BigInsights Enterprise Edition还自带了一个标准的Java Database Connectivity (JDBC)连接器，让组织能够快速集成多种数据和信息系统，包括Oracle、Microsoft SQL Server、MySQL和Teradata。

InfoSphere DataStage工具包括一个连接器，允许您在DataStage ETL作业中充分利用BigInsights数据。IBM InfoSphere Streams包括一个连接器，允许最终用户读取BigInsights文件系统中的内容并向该系统中写入数据。

InfoSphere BigInsights还支持各种开放的开发标准和Apache Nutch的Web搜索软件，可以抓取位于企业内外的非结构化数据内容。

获得企业级的支持

由于开源软件的本质，因此这种软件不包括技术支持，并且它可能带有不适合某些组织要求的法律条款和条件。与此相反，随InfoSphere BigInsights Enterprise Edition一起交付的是标准的IBM软件许可和支持协议。组织可以根据熟悉的许可条款部署它，这有助于最大限度地减少不确定性和风险，同时组织可放心地获得24×7支持服务、教育和遍布全球的专业服务机构的支持。

IBM在全面、集成和企业就绪的解决方案中将开源技术带给组织。IBM通过将大数据技术集成到现有的信息供应链中，帮助企业简化并加快大数据技术的引入。此外，组织知道IBM可在一个解决方案中提供他们所需的性能和可靠性，而该解决方案将成为其决策制定流程中不可分割的一个部分。

硬件要求和操作系统支持

Intel x86服务器，64位，至少4 GB内存和30 GB磁盘存储

Red Hat Enterprise Linux 5.3、5.4、5.5、5.6、6.0

SUSE Linux Enterprise 11、11 SP1

注：有关操作系统版本支持的最新信息，请参阅产品文档：

ibm.com/common/ssi/cgi-bin/ssialias?infotype=AN&subtype=CA&htmlfid=897/ENUS211-501&appname=USN



更多信息

如需了解有关InfoSphere BigInsights Enterprise Edition的更多信息, 请联系IBM销售代表, 或者访问:

ibm.com/software/data/infosphere/biginsights/enterprise.html

此外, IBM Global Financing可帮助您以最经济高效的战略性方式获得您的业务所需的IT解决方案。我们将与信用合格的客户展开合作, 定制一个IT融资解决方案来满足您的业务目标, 实现有效的现金管理, 以及改善您的总体拥有成本。IBM Global Financing是您进行关键IT投资和向前推进您业务的最智慧选择。有关更多信息, 请访问: ibm.com/financing

© 版权所有IBM Corporation 2012

IBM Corporation Software Group Route 100
Somers, NY 10589 U.S.A.

在中国印刷
2013年1月

IBM、IBM徽标、ibm.com、DataStage、DB2和InfoSphere是国际商业机器公司在全球多个司法管辖区内注册的商标。其他产品和服务名称可能是IBM或其他公司的商标。有关IBM商标的最新列表, 请访问ibm.com/legal/copytrade.shtml的“Copyright and trademark information”部分。

Intel是Intel Corporation或其子公司在美国和/或其他国家的注册商标。

Linux是Linus Torvalds在美国和/或其他国家的注册商标。

Microsoft是Microsoft Corporation在美国和/或其他国家的商标。

Netezza和Netezza Performance Server是IBM子公司Netezza Corporation的商标或注册商标。

本文包含截至出版之日的最新信息, IBM可能随时更改这些信息。不是所有产品都可用于IBM运营的每个国家/地区。

本文中的信息“按原样”提供, 不含任何明示或暗示的担保, 包括但不限于适用性、特定用途的适用性, 以及有关非侵权性的任何担保或条件。IBM产品的担保依据的是它们所遵循的协议中的条款和条件。



请回收利用