



# Microsoft 大数据

解决方案简介

## 目录

简介 .....	2
Microsoft 大数据解决方案 .....	3
主要优势 .....	3
随时随地获得深入的洞察力 .....	3
与全世界的数据连接 .....	4
随时随地处理任意大小的任何数据 .....	4
回馈 Hadoop 社区 .....	4
其他信息 .....	4

## 客户面临的主要挑战

- **了解数据爆炸：**企业需要借助适当的工具来了解由降低的硬件成本和复杂的数据源带来的海量数据。
- **了解更多类型的数据：**企业需要同时分析关系型和非关系型数据。捕获的数据中 85% 以上都是非结构化数据。
- **实时分析数据：**新的数据源（例如 Twitter、Facebook 和 LinkedIn 之类的社交媒体网站）会实时生成海量数据，而这些数据无法通过简单的批量处理进行有效地分析。
- **实现简化部署和管理：**企业需要更加简单顺畅的部署和设置体验。理想情况下，他们希望使用包含 Hadoop 相关项目的少量安装文件，而不是自己从项目中进行选择。

## 简介

如今，企业正竭尽全力地从他们捕获的海量数据中获取业务洞察力。这些数据包括大量非结构化数据，例如文件、图片、视频、博客、点击流和地理空间数据。对企业来说，主要的挑战是在无需设置复杂的分布式存储和计算集群的条件下，如何同时有效地处理结构化和非结构化数据。

企业正在寻求一种可以将内部和外部的数据及服务结合起来的有效方法。他们希望从 Twitter、Facebook 和 LinkedIn 等社交媒体网站中挖掘数据；同时还希望根据这些捕获的数据做出更加及时有效的决定。为实现此目的，企业需要实时分析其数据，而非简单地依赖于批量处理。

新出现的一些技术（例如 Apache Hadoop）使客户无需付出高昂费用即可存储和分析海量的非结构化数据。此外，新出现的云计算和新标准（例如开放数据协议 (OData)）可让企业比以往任何时候都更加轻松地共享与合并数据。如今，许多供应商都提供了 Hadoop 部署，但其中大部分都是在位于中央 IT 范围之外的孤立存储区运营的，尚不适用于企业级应用。

Microsoft 早在大数据流行之前就开始对之加以利用了。例如 Microsoft® Bing™ 会分析超过 100 PB 的数据以提供高品质的搜索结果。企业可以利用 Microsoft 大数据解决方案从自己的结构化和非结构化数据中施展业务洞察力。Microsoft 的大数据解决方案可让任何人通过 Microsoft Office 和 SharePoint® 等熟悉的工具从他们所有的数据中施展可执行的洞察力；还能通过连接到公用的全球数据和服务，让客户发现新价值。

## Microsoft 大数据解决方案

Microsoft 的愿景是让所有用户都能从几乎任何数据中获得业务洞察力，包括以前隐藏在非结构化数据中的洞察力。为实现此愿景，Microsoft 制定了全面的大数据策略以解决以下关键的客户要求：

- **灵活的数据管理层**，可支持所有数据类型，包括结构化、半结构化和非结构化的静态或动态数据。
- **扩充层**，用于发现、转换、共享和监管数据。
- **令人瞩目的 BI 工具套件**，可帮助用户通过分析获得洞察力。
- **更深入的洞察力**，将企业的数据和来自外部源的数据及服务相结合。（以下图片展示了完整的大数据解决方案。）



为帮助企业快速采用其大数据解决方案，Microsoft 将在 Microsoft Windows® Azure™ 平台上提供基于云端的 Hadoop 服务，同时在 Windows Server® 上提供基于本地的 Hadoop 版本。

Microsoft 新的 Hadoop 版本使得客户可以从几乎任何规模的结构化和非结构化数据中获得业务洞察力并应用新型数据，而无需考虑其具体位置。Hadoop 的丰富洞察可以与 Microsoft Business Intelligence (BI) 平台无缝结合，使客户能够借助熟悉的工具（例如 Microsoft Office 和 SharePoint 等）以及公用的数据和服务来丰富他们的模型。

Microsoft 的大数据解决方案还通过简单的部署以及与 Active Directory® 和 System Center 等组件的集成，为 Hadoop 提供了 Microsoft Windows® 的易用性和可管理性。凭借 Windows Azure 上基于 Hadoop 的服务，Microsoft 为其大数据解决方案在云端提供了灵活性。

## 主要优势

- 使用熟悉的 Microsoft Office 和 BI 工具从任何数据中**随时随地获得深入的洞察力**。
- **与全世界的数据连接，从而发现隐藏的价值**。通过将内部与公用的数据和服务（包括社交媒体网站）相结合来实现这一目标。
- 通过支持任何数据的现代数据管理平台**随时随地处理任意大小的任何数据**，该平台不仅具有 Windows 和 SQL Server 的易用性，同时具有云的弹性和可扩展性。

## 随时随地获得深入的洞察力

Microsoft 的大数据解决方案可让客户用熟悉的 BI 工具从他们的结构化和非结构化数据中获得可执行的洞察力。值得一提的是，Microsoft 的解决方案可让客户：

- **使用熟悉的工具分析 Hadoop 数据**：Microsoft 可让用户利用 Excel 的 Hive 组件在熟悉的 Microsoft Excel® 环境中与 Hadoop 中的非结构化数据进行交互并加以分析。
- **通过任何数据获得深入的洞察力**：企业可以用熟悉的 BI 工具（例如 Microsoft SQL Server® Analysis Services (SSAS)、PowerPivot 和 Power View）通过 Hive Open Database Connectivity (ODBC) Driver 来分析 Hadoop 中的非结构化数据。企业还可以用 SQL Server 2012 上的 PowerPivot 和 Power View 对关系型数据采用自助服务的 BI 产品。
- **通过简化的编程驱动洞察力**：Microsoft 通过与 .NET 和新的 JavaScript 库集成简化了 Hadoop 的编程。开发人员可以在 JavaScript 中使用新的 JavaScript 库来轻松编写 MapReduce 程序，然后通过简单的浏览器来部署他们的 JavaScript 代码。

## 与全世界的数据连接

Microsoft 的大数据解决方案将数据和模型与公用的数据和服务（包括 Twitter、Facebook 和 LinkedIn 等社交媒体网站）相结合，从而能够实现突破性的发现。这使得客户能够使用 Windows Azure Marketplace 上的应用程序和挖掘算法来发现隐藏的模式。

- **通过 Windows Azure Marketplace 进行共享和协作：** Microsoft 大数据解决方案可让客户通过 Windows Azure Marketplace 共享数据并发现新的洞察力，Windows Azure Marketplace 可通过开放数据协议 (OData) 展露数百种来自 Microsoft 和第三方的应用程序和数据挖掘算法。
- **与社交媒体集成：** Microsoft 的大数据解决方案可让客户通过来自社交媒体网站（例如 Twitter 和 Facebook）的公用数据来扩展他们的分析。Microsoft 的一款代号为“Social Analytics”的基于云的项目允许企业将社交媒体信息与业务应用程序相集成。
- **借助 Hadoop 执行高级分析：** Microsoft 大数据解决方案支持传统的 BI 以及高级分析（例如数据挖掘和图形挖掘），从而可让客户从他们所有的数据中发现新价值。Hive ODBC Driver 可让客户使用 SQL Server 数据挖掘工具执行预测分析。Microsoft 还将支持 Mahout 等其他高级分析工具，以及使用 C++、C#、Python、Ruby 和 Pearl 编写的挖掘算法。

## 随时随地处理任意大小的任何数据

通过现代数据管理平台，Microsoft 可让客户无缝存储和处理所有类型的数据，包括结构化、非结构化和实时数据。它为 Hadoop 提供了 Windows 和 SQL Server 的易用性，通过 Hadoop 扩展了数据仓库，还提供了大数据在云端的弹性和可扩展性。

- **为 Hadoop 提供 Windows 和 SQL Server 的易用性和可管理性：** 通过智能化打包，Microsoft 使得在 Windows 上安装和设置 Hadoop 更加简单；安装完毕只需不到三小时，而无需花费数周的时间。Windows Azure 平台上基于 Hadoop 的服务进一步简化了部署，可让客户在 10 分钟内在云端部署 Hadoop 集群。通过与 Microsoft System Center 集成，可让 IT 人员像管理其他任何 IT 资产一样轻松管

理他们的 Hadoop 集群。与 Active Directory 的集成可让 IT 人员使用基于企业的安全策略来保护他们的 Hadoop 集群。

- **通过 Hadoop 扩展数据仓库：** 两个用于连接 SQL Server 和 SQL Server 并行数据仓库的新型双向 Hadoop 连接器可让客户轻松地将 Hadoop 与 Microsoft 企业数据仓库及 BI 解决方案相集成，以便从结构化和非结构化数据中获取更加深入的业务洞察力。
- **获取云端的弹性和可扩展性：** Microsoft 提供了两种在 Windows Server 上部署 Hadoop 的选项：在基于云的环境中或在本地部署。Windows Azure 平台是一种基于云的服务，可在 Microsoft 的云平台上提供灵活的千兆级分析。对于希望将数据保留在其数据中心的客户，Microsoft 将提供 Windows Server 上基于 Hadoop 的版本。

## 回馈 Hadoop 社区

与其他的竞争对手不同，Microsoft 通过回馈 Hadoop 社区扩展了访问权限。Microsoft 致力于提倡改变 Apache Hadoop，以拓宽 Hadoop 社区的使用范围。通过与 Hortonworks 通力合作，Microsoft 已向 Apache 提交了首份提议，并且正在拟定有关适用于 Hadoop 的新 JavaScript 库（由 Microsoft 开发）以及 Hive ODBC Driver 的深度提议。

## 其他信息

要了解有关 Microsoft 大数据解决方案的详情，请访问 [www.microsoft.com/bigdata](http://www.microsoft.com/bigdata)。

要下载 Microsoft SQL Server Connector for Apache Hadoop，请访问 <http://www.microsoft.com/download/en/details.aspx?id=27584>。