



亮点

- 即时访问关键业务洞察力
 - 快速完成数据的加载、分析及可视化
 - 上传多个来源的数据并与 R 集成
 - 利用数据库内预测性分析算法提升洞察力
 - 将内部数据仓库环境扩展到云端
 - 通过与 IBM Cloudant 原生集成，分析 JSON 数据
-

IBM dashDB

专为分析而构建的云数据仓库即服务

IBM® dashDB™ 是一种反应敏捷的全运维管理云数据仓库，能够利用集成分析工具迅速地向用户提供分析结果。借助 dashDB 独特的数据仓库内分析，R 预测性建模及商业智能工具，用户可以更加自由地分析数据，更加快速地获得精准的洞察力。

dashDB 的启动和运行非常简便，同时也支持在 IBM Bluemix™ 中快速部署。现在，对于 1GB 以内的数据，您可以免费测试或试用 dashDB，此后，您只需每月支付 50 美元便可享受 20GB 的数据存储。由于您手头的数据量会不断增加，而用户也需要专用环境，所以我们会为您提供具有多 TB 容量的大规模实例。大规模并行处理 (MPP) 有助于加快查询速度并扩大数据集规模。

不论组织规模如何，IBM dashDB 均可提供简便易用的数据仓库一体机即服务，同时实现云端的敏捷性和可扩展性。借助 IBM Bluemix 提供的各种开发者服务及补充性服务，或是您青睐的本地工具，您可以快速地开发部署多个分析应用。

dashDB 之前的数据仓库

过去，构建数据仓库是一项十分费力的工作。我们先要确定数据仓库软件，再为其分配硬件和存储空间，并确保分配工作的均衡性。确定数据仓库的物理组成之后，我们需要构建物理系统，以及支持相应计划的逻辑数据模型。如果我们要扩展数据仓库（这是必然的，因为收集的数据会不断增长，而且在此基础上会构建新应用），就必须配置新的处理功能、存储空间和软件。



换言之，我们每次希望改变数据仓库时，与之相关的整个流程都会产生一系列的风险。是否维持了数据处理环节之间的适当平衡？是否采购了合适的硬件？整个系统内的数据分配是否正确？硬件是否已过期并已不能与市场上的最新技术兼容？软件是否需要更新？所以说，不论处于发展数据仓库的哪个阶段，您都会面临一系列的潜在问题。此外，您还必须按容量付费，即便是在还未使用的情况下。由此可见，几乎所有风险都由您和您所在的组织承担。

运用数据仓库一体机

数据仓库一体机能帮助您扫除多重障碍，轻松构建自己的数据仓库。这些系统可提供预先配置的集成数据仓库，进一步提升分析性能。在此基础上，您只需要选择合适的模式和规模，然后在启用并运行之后，将所需的数据加载至数据仓库。

尽管硬件和软件的分配均已事先完成，但在到达资源限制时，您仍需采购新系统，以确保一体机的正常使用。随着数据量、新应用和新用户的不断增长，在动态扩展期间实施新数据仓库可能带来较高的成本压力。在此情况下，您不仅需要兼顾新数据仓库的升级、补丁修复、维护和整体硬件折旧，还需要处理有关传统数据仓库的大量维护工作。

不过，数据仓库一体机依旧是组织决策管理系统的重要组成部分，但是否存在某种互补技术能够有效地破除数据仓库的发展瓶颈？不妨选择 IBM dashDB。

云技术优势

由于云计算本身具有可扩展性及按需分配的特质，您可以借助这些优势快速地启动并运行数据仓库，同时快速完成云端的部署工作。此外，您不需要进行任何基础架构投资，即可获得切实的业务敏捷性支持。您可以按己所需购买产品及服务。这一切都由您掌控。

由于 dashDB 是一种全运维管理云服务，您无需顾及其日常后端维护工作。此类维护工作不仅仅局限于一些可能适用或不适用的修补程序包（这些其实是无关紧要的，因为在阅读相应文档之前，您无法获悉修补程序包是否存在），同时还包括可能适用或不适用的版本。借助 dashDB，我们的专业工程师和开发人员能够不断构建新功能，并确保这些功能与产品之间的兼容和集成，而从用户角度来看，这也是一个水到渠成的过程。

利用 dashDB 保护数据

从设计到部署，我们从各个环节完成了 dashDB 的优化，旨在为您提供全方位的数据保护。dashDB 包含有多个安全防护层，比如静态数据和动态数据自动加密、IBM InfoSphere® Guardium® 数据活动监控、高级数据库访问控制和部署固化。

采用最佳设计及开发实践方案是确保数据安全的前提。就 dashDB 的开发而言，我们运用了多种实践方案，比如风险评估、威胁建模，以及运用 IBM AppScan® 实现的静态和动态代码分析。

全心专注业务，轻松管理数据仓库，全面提升发展速度

云技术的简便性是关键，具体体现在：dashDB 是一种全运维管理服务；云技术本身具有多方面的优势，比如自动完成补丁修复、版本控制，实现快速云部署，确保业务敏捷性等。通过应用 dashDB 即服务，您可以根据使用情况，采用即用即付的方式来购买产品。这种方式完全不同于购买完整的数据仓库。借助 dashDB，您可以根据数据的增长情况来购买附加容量，这种方式不仅简便，而且易于控制。

事实上，此款解决方案的计费方式很简单，它属于一种服务，而不是硬件集群或一体机，因而能够满足您的业务增长需求。应用此解决方案后，即便在云端需求增加的情况下，您也无需购买、安装和测试新的软件和硬件。

内置高性能结合内存技术，实现快速响应

dashDB 与其他基于云的数据仓库有哪些区别？很多。dashDB 的核心是 IBM BLU Acceleration 技术。

IBM BLU Acceleration 是一项内存数据库技术，能够打破内存解决方案的常见瓶颈，为用户提供一流的数据库性能。由于 dashDB 基于 BLU 构建，因而能够保持 BLU 的所有优势：

- **高级处理**：dashDB 不需要整个数据集来配合内存需求，同时确保快速完成数据处理，因为它采用一系列已获专利的算法，能够实现敏捷的内存数据处理。
- **数据预取**：dashDB 支持数据预取，能够提前预见所需数据并自动作出调整，以便将必要的放入 CPU 内或其附近。
- **无需解压**：dashDB 能够在保持数据顺序的情况下，对压缩数据执行多种操作，包括数据合并、谓词计算等，从而显著提升数据处理速度。
- **数据忽略**：处理大数据时，您很可能不需要利用数据仓库中的*所有信息*来响应某一查询请求。dashDB 采用 BLU Acceleration 技术，能够针对特定的查询请求，自动确定不合格的分析数据。因而，dashDB 会在查询期间忽略大量的不相关数据，为您节省宝贵的时间和资源。

MPP 功能有助于更快速地处理查询请求和海量数据集

dashDB 的 MPP 功能具备标准 dashDB 服务的优势，且在此基础上进一步提升了敏捷性和可扩展性，以便用户能够处理更大型的数据集。MPP 架构是一种以并行方式运作的服务器网络集群，能够更快地为用户提供查询结果。在 dashDB MPP 集群中，多台服务器会同时处理同一查询请求，而每台服务器的查询也可在所有处理器之间并行处理。

在标准架构中，并行处理只会在处理器级别上发生。在拥有 MPP 架构的情况下，系统会对查询请求进行拆分，以便各自拥有本地存储空间和计算容量的多个服务器能够同时处理不同的数据片段。这种协作方式极大地提高了处理查询请求的效率，同时也大幅降低了输入/输出要求。在 MPP 中，处理查询的各个独立的服务器采用 BLU 动态内存列式存储技术，相比传统行式存储数据库，这有助于最大程度地降低输入/输出要求，并实现处理速度的大幅提升。

对于 MPP 架构，每为网络集群添加一个新的服务器，就能实现整体性能的提升。举例来说，如果使用单个服务器的标准架构需要一小时来完成查询处理，那么仅利用 4 个服务器的 MPP 集群只需约 15 分钟就能完成此项任务。在此基础上再添加一个服务器，也就是总共 5 个服务器，我们就能将处理查询的时间缩减为 12 分钟，而如果有 6 个服务器，处理时间就会减为 10 分钟，以此类推。由此可见，借助 dashDB MPP，实现扩展就如同向集群添加服务器一般简单。

专为分析构建，帮助用户深入理解数据和业务

由于组织要处理的数据量呈大幅增长趋势，内存处理技术正以前所未有的速度在快速发展。随着硬件和内存进一步迈向商品化，我们能够更多数据放入内存并加以处理。

过去，组织必须等待接收数据仓库提供的分析报告；但借助内存计算领域取得的新进展，这种情况已不复存在。现如今，我们能够实时提供决策结果。由于返回结果的速度更快，分析师就能够根据自身需要对所获得的结果进行更加深入的分析。

目前，数据仓库已成为支持数据分析的主要数据存储方式。因此，更可行的方式是将数据放入数据仓库分析，而非从中取出数据，在其他位置加以分析。

提供数据库内分析支持，实现效率和性能的提升

dashDB 具备 BLU Acceleration 的性能优势，同时与 IBM Netezza® Analytics 相集成，可提供全面的数据库内高级分析支持。同样的技术也应用于 IBM Netezza 一体机和 IBM PureData™ for Analytics 系统。

换言之，您可以通过 dashDB 获得多种 *直接嵌入到数据库中的* 预测性建模算法。您可以根据自身需求，随时使用这些算法。

dashDB 可提供的算法包括：

- 线性回归
- 决策树聚类
- K 均值聚类
- 与 Esri 兼容的地理空间扩展

通过直接在存储数据的数据库中运行这些分析工具，您的组织将从中获得巨大的效益。您无需提取数据，将其发送至其他位置，待加载之后再进行处理，您只需在数据的原始存放位置（数据仓库），直接应用分析工具即可。

与 R、IBM Watson Analytics 等高级工具兼容

R 是一种开源编程语言，其开发主要面向高级数据分析及图形可视化处理。R 语言可用于分析来自多个不同数据源的数据，比如外部文件或数据库等。通过数据附带的 R 运行时，dashDB 能够集成 R 语言完成预测性建模。Web 控制台可用于加载数据，并在数分钟内执行分析。数据分析可能涉及 SQL、BI 工具，或 R 脚本和模型。利用 dashDB 和开源 R 语言，您可以有多种多样的分析选择。

IBM dashDB 包含有 RStudio，这是一种完全集成的 R 语言开发环境，可提供基于 R 语言的快速预测性分析。RStudio 可提供多种功能支持，包括 R 语言代码完成功能、面向 R 语言包的集成帮助、文件管理功能等。如需安装其他 R 语言包，您可以直接在 RStudio 内轻松完成。

我们在开发 dashDB 时，重点考虑了大型的商业智能生态系统。因此，dashDB 服务能够直接与多种 IBM 核心技术进行协作，其中包括 IBM Watson™ Analytics、IBM Cognos® BI、IBM DataWorks 等等，但它的优势并不仅限于此。dashDB 还可以与 IBM 业务合作伙伴、BI 工具集进行协作，比如 Looker、Aginity Workbench、Tableau 等，不胜枚举。

IBM 合作伙伴生态系统的发展十分迅速，您可以将多款第三方工具连接到 dashDB 服务。例如：

- 连接至 IBM InfoSphere Data Architect 之后，可设计和部署您的数据库模式
- 连接至 Esri ArcGIS 之后，可执行地理空间分析，并实现发布与数据之间的映射
- 连接至 IBM Cognos 服务器之后，可针对数据运行 Cognos 报告。
- 连接至 Tableau、Microstrategy 或 Microsoft Excel 等基于 SQL 的工具之后，可操控或分析您的数据
- 连接至需要分析数据库的 Bluemix™ 应用
- 连接至 Aginity Workbench，可将 Netezza 数据模型和数据迁移至 dashDB

用例

组织能够通过多种方式来利用 dashDB 的优势。根据我们对目前 dashDB 活跃客户的观察，我们总结了以下四类关键用例：

1. 扩充现有数据仓库 – 混合模式

据 90% 以上的客户反映，他们有计划扩充现有数据仓库¹。dashDB 的设计能够很好地满足此种使用需求。借助 dashDB，组织能够将内部数据仓库扩展至云端。由于此平台可灵活伸缩、按需进行扩展，您只需根据当前的需求购买容量即可。拥有 dashDB 之后，您可以按需购买实用程序，而不会受制于任何硬性要求。

2. 分析 NoSQL 数据

第二类关键用例有关 IBM 核心技术与 Cloudant® 和 dashDB 之间的集成。您可以轻松地将 Cloudant 内的 JSON 数据同步到 dashDB 内的结构化数据。然后，您就可以运用传统 BI 工具和数据仓库的常用分析工具。由于 dashDB 已嵌入数据库内预测算法，使用 Cloudant 的客户便可轻松无忧地对 JSON 文档存储进行分析。

3. 数据科学，数据存储

“数据科学，数据存储”是一种有关数据库内分析的概念，下面我们将针对这一概念介绍第三类关键用例。对于有统计需求的用户和数据科学家来说，dashDB 可以提供一系列强大的预测分析算法。如前所述，dashDB 已嵌入 R 运行时和 RStudio。就统计学家和数据挖掘人员而言，R 语言已广泛应用于开发统计软件和数据分析工具。

4. 独立数据仓库即服务

dashDB 在很大程度上会用作云端的独立数据仓库。不论是为获取 Cloudant 数据而构建的小型数据集市、基本测试环境或开发环境，还是位于云端的完整企业数据仓库，dashDB 均可全天候地为您和您的组织提供服务。

dashDB 入门介绍

IBM dashDB 是一种全运维管理数据仓库服务，您无需购买、安装任何硬件或软件。您只需访问 www.dashdb.com，完成账户注册，即可免费试用。如果您需要进一步了解 dashDB，或确定此产品是否能够满足您组织的要求，欢迎您访问 dashdb.com 与我们联系，我们会竭诚为您解答，或与您的技术团队开展深入探讨。

关于 IBM dashDB 解决方案

IBM 在数据仓库、信息管理和业务分析方面提供了最完善的软件、硬件和解决方案组合，可帮助客户实现信息资产的价值最大化并从中发现新洞察力，以便更快速地作出更佳决策，同时优化业务成效。

有关更多信息

如欲了解有关 dashDB 的更多信息，请联系您的 IBM 代表或 IBM 业务合作伙伴，或访问以下网站：www.dashdb.com

此外，IBM 全球融资部可帮助您以最具成本效益及战略性的方式获得贵企业所需的 IT 解决方案。我们将与可信的客户合作，以便定制一款适于贵企业业务目标的 IT 财务解决方案，确保实现高效现金管理并降低您的总体拥有成本。IBM 全球融资部是您规划关键 IT 投资并推动企业发展的明智之选。

有关更多信息，敬请访问 ibm.com/financing



© Copyright IBM Corporation 2015

IBM Corporation
New Orchard Road
Armonk, NY 10504

美国印刷
2015 年 7 月

IBM、IBM 徽标、ibm.com、AppScan、Bluemix、Cloudant、Cognos、dashDB、Guardium、IBM PureData、IBM Watson、InfoSphere 及 Netezza 是 International Business Machines Corporation 在世界各地司法辖区的注册商标。其他产品和服务名称可能是 IBM 或其他公司的商标。Web 站点 www.ibm.com/legal/copytrade.shtml 上的“Copyright and trademark information”部分中包含了 IBM 商标的最新列表。

Netezza 是 IBM 的子公司之一 IBM International Group B.V. 的注册商标。

Microsoft、Windows、Windows NT 及 Windows 徽标是 Microsoft Corporation 在美国和/或其他国家或地区的商标。

本文档截至最初公布日期为最新版本，IBM 可随时对其进行修改。IBM 并不一定在开展业务的所有国家或地区提供所有这些产品或服务。

本文档内的信息“按现状”提供，不附有任何种类的（无论是明示的还是默示的）保证，包括不附有任何关于适销性、适用于某种特定用途的保证以及不侵权的保证或条件。IBM 产品根据其提供时所依据的协议的条款和条件获得保证。

客户应负责确保与适用的法律和法规的合规性。IBM 并不提供法律建议，亦不声明或保证其服务或产品可确保符合任何法律或法规。

1 Predicts 2014: Why You Should Modernize Your Information Infrastructure. 2013 年 11 月 28 日. Gartner.



请回收利用