



HRG 评估：

IBM dashDB

IBM dashDB 是一系列基于云、用于分析和交易处理的数据仓库和数据库解决方案。IBM 的 dashDB 可借助内存内列式处理以及数据库内分析技术生成基于数据的洞察力，包括 R 统计语言、RStudio 开发环境、Apache Spark 分析操作环境、Netezza 的数据库内分析和分析库。客户可以运行“假设”场景和即席查询，以研究趋势、提供分析、获取竞争优势。dashDB 具有几种不同的版本：dashDB for Analytics、dashDB Local 和 dashDB for Transactions。IBM dashDB for Analytics 是完全托管的数据仓库公有云服务，由 IBM 进行管理和维护，确保客户始终处于最新的发布水平。IBM dashDB Local 是客户管理的数据仓库私有云服务，其在客户的基础架构或 SoftLayer、AWS、Azure 等托管服务之上运行。IBM dashDB Local 为客户提供数据仓库部署控制和灵活性。IBM dashDB for Transactions 是完全托管的操作数据库公有云服务，不在本评估范围之内。

挑战

这个世界的芯片和传感器日趋增多，它们跟踪和测量交易和交互，产生越来越多的数据。IT 支持的业务环境发展迅速，因此，交付的基础架构必须流畅、高度可扩展，且必须能够支持迅速发展和同化的应用。挑战之处在于，交付的系统和服务必须支持端到端供应链交互、商业智能及分析，同时还应降低成本，提高客户满意度，提高市场占有率和相关利益方价值。

将自定义产品作为交易或进行持续的客户互动正成为一种业务需求。新一代客户体验应用要求对分布式和异构数据源（包括结构化、非结构化、结构不良、大容量和高速数据）进行近乎实时的集成和分析。管理和分析该数据并将其转换为可执行信息的能力是实现云计算、大数据和物联网真正优势的关键。

组织转型

大数据、社交和情感分析正在改变客户使用（结构化和非结构化）数据的方式。若要实现成功的业务演变，关键在于获得选择权和灵活性，识别并收集有用的数据，同时避免前一代“所谓”简化方法所带来的高成本和复杂性。利用数据影响客户与企业之间正在发生的事务性互动结果具有革新性。引入高度相关的附加产品后，可通过客户互动时的实时分析让数据变得有用，这将有助于实现升级销售。移动和 Web 平台上用于改善客户体验的数据大多数都是非结构化或结构不良的数据，其来源包括 Facebook、Twitter、LinkedIn、Pinterest 和 Instagram 等。

时至今日，无论任何类型或位置的数据，都能进行收集和集成。用户不再受限于数据的物理或虚拟位置、应用来源，甚至数据的类型（结构化数据或非结构化数据）也不再是限制因素。企业需要根据以各种格式收集的整个企业内外近乎实时的数据流进行运营分析和决策制定。组织若要应对不断演变的机遇和挑战，就必须能够集成和分析交易系统中活动的实时运营数据、传统数据仓库和数据集市中的历史数据以及非结构化高速数据。

应用最初是指业务逻辑的自动化，而今也必须不断变化，才能使企业实现演变。HRG 相信，随着时间的推移，应用和分析工具将逐渐融合，成为新一代应用或洞察力系统，能够通过实时发现更改业务流程。

dashDB

IBM dashDB 包含一系列预配置的数据仓库解决方案，既可用作公有云上的完全托管服务，也可用作私有云中的客户托管服务，在客户现有的 IT 基础架构上使用。通过同时使用 dashDB 托管服务和 dashDB Local，客户可以在二者之间轻松移动数据和工作负载。由于具有托管灵活性以及动态移动数据和工作负载的能力，因此可以为设有远程办事处的 dashDB Local 客户或远程研究人员提供价值，也可以为存在充分理解的功能需求的部门用户提供价值，如生产前工程原型设计和概念测试。

dashDB 托管服务和 dashDB Local 共享同样的外观和通用的代码库，其包含被迁移至新 dashDB 代码库的关键 Netezza 数据库内分析功能和分析库。IBM dashDB 和 dashDB Local 可提供按需型数据仓库以及来自 IBM DB2 with BLU Acceleration 的内存内列式数据处理。

dashDB 和 dashDB Local 都提供内存内列式存储技术和 MPP 感知查询引擎、查询规划、查询优化、压缩列式数据上的动态内存处理、并行向量处理以及数据忽略（来自 IBM DB2 with BLU Acceleration），并针对高度并行的多核系统进行了优化。借助大型数据集的 MPP 功能，客户可以从单节点系统快速扩展至企业级 MPP 数据仓库集群，在一段时期的高峰需求过去后，又可以回缩至单个 SMP 节点。借助 dashDB 分析引擎，SQL 可与 DB2、Netezza 和 Oracle 兼容，进而能够降低迁移难度，还可利用 ODBC、JDBC、SQL PL 和 PL/SQL 获得泛型编程支持。

dashDB for Analytics (托管服务)

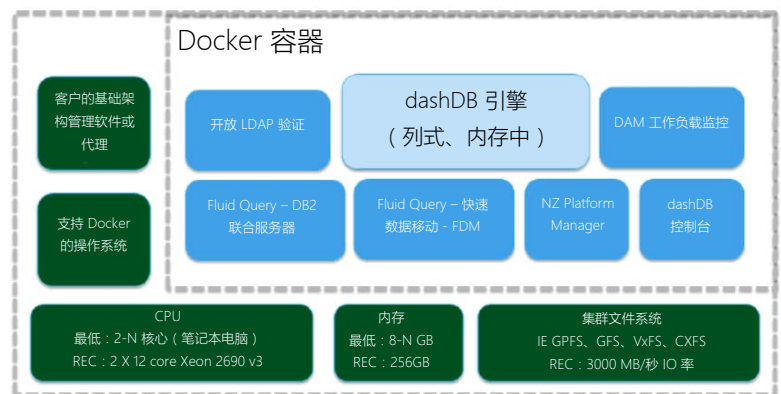
dashDB for Analytics 托管服务是托管于公有云上的按需型数据仓库服务，可通过运行于 IBM SoftLayer 之上的 IBM Bluemix 使用，也可用于 Amazon Web Services (AWS) 上。dashDB 托管云服务能够与 IBM Cloudant 和其他可通过 Bluemix 使用的 IBM 云服务集成，进而有助于快速拼接基于云的解决方案。例如，深层数据分析借助具备集成工具的 dashDB 启用，如 Watson、Cognos 和 SPSS。通过与 Aspera 集成，客户可以在 WAN 上快速加载数据。最后，IBM DataWorks Lift 快速将数据从传统数据中心 IT 来源移动至云端，以便借助 dashDB 的数据库内分析开展分析。IBM dashDB for Analytics 托管服务可作为数据源用于通过 Fluid Query 进行的联合查询，后者来自于 IBM PureData System for Analytics 及其前身 Netezza。（有关 Fluid Query 的更多信息，请参阅下面的 dashDB Local 部分。）

使用 dashDB 托管服务的客户可以采用自动模式发现和映射数据，通过 IBM Cloudant 访问 JSON 数据，特别是 Web、移动或物联网高速数据，随后将从 Cloudant 导入的 JSON 数据直接加载并同步至 dashDB。这样，来自传感器（物联网和销售点设备）或社交媒体平台的数据可用于通过同步复制填充 dashDB 表格。很多面向客户的新一代应用将从中受益，包括将该数据用于升级销售，以及当互动发生时为客户提供建议。IBM 和 Apple 近期宣布，Cloudant 将作为全新的下一代移动应用的数据层。现在，Cloudant 中捕捉和存储的 JSON 数据可用于 dashDB，以便为全新的下一代移动应用运行各种分析，如报告、数据可视化或地理空间分析等。

dashDB Local

dashDB Local 是软件定义环境 (SDE) 的客户托管数据仓库服务，如私有云、虚拟私有云以及其他支持 Docker 容器技术的基础架构。由于使用 Docker 进行快速部署，客户可以快速获得内存内数据处理、数据库内分析等功能，以及可扩展的 MPP 架构以及 Oracle 和 Netezza 兼容性。dashDB Local 打包于 Docker 容器（见图）之中，该容器可运行于 Linux、Mac OS 或 Microsoft Windows 操作环境中，或在 SoftLayer、AWS 或 Microsoft Azure 等私人托管服务之上运行。IBM 表示，客户反馈借助 dashDB Local，20 分钟内就可为 MPP 集群启动并运行数据仓库，10 分钟内可为单节点 SMP 系统启动并运行数据仓库。

架构：客户云或内部数据中心上的主机系统



dashDB Local 含有作为嵌入式分析运行时环境的 Apache Spark；借此，用户可以利用 dashDB Local 的数据库内分析功能执行 SQL 和 Spark 查询。dashDB Local 支持 MPP Spark 集群成员与数据仓库集群成员的共置设置，因而，成员之间可以通过共享内存快速交换数据。数据科学家可以将 dashDB 中的结构化数据作为存储层，从而利用 Scala 和 Python 在 Spark 上开发和部署应用。IBM 表示即将推出 dashDB 托管服务的 Spark 集成。

借助源自 Netezza、嵌入在 dashDB Local 中的 Fluid Query 联合和数据移动功能，可以跨越物理和虚拟数据存储传送和移动数据。近期，HRG 将会看到 IBM 交付的动态下推，将其作为与 Spark 共同使用的一项功能，从而将过滤动态下推至数据源引擎，以支持 dashDB Local 到 dashDB Local 的联合查询、dashDB 到 dashDB Local 的联合查询以及 dashDB 到 Hadoop 的查询联合。所有这些都意味着散布于多个数据源的更快的数据分析。未来，HRG 预计将看到对 Oracle、Teradata、Microsoft 和一般 JDBC 数据源的第三方查询联合的支持。

安全

IBM dashDB 托管服务借助 IBM Security Directory Server LDAP 进行用户认证，借助 IBM Security Guardium Data Activity Monitor 进行自动加密、数据库活动监控和访问控制。部分版本的 dashDB 托管服务所支持的 Guardium 尚不可用于 dashDB Local。dashDB 托管服务在数据库中存储数据，使用行业标准 AES 256 对称加密自动保存加密备份镜像。dashDB 会自动加密静态数据，采用 BLU 压缩在已压缩数据上执行分析的方式分析已加密数据。只要以明码传输的未加密端口 50000 被禁用，就可使用 SSL port 50001 加密传输中数据。

dashDB 已使用最佳实践开发和测试，包括风险评估、威胁建模以及静态和动态节点分析。dashDB 数据库活动处于监测之中，监测项目包括访问人员、连接人员、执行的动作以及做出的更改。该信息通过 dashDB 控制台向客户提供。控制台支持通过 dashDB 进行 HTTPS 加密通信，有助于控制数据访问者以及被访问数据的类型。

MPP

借助大型数据集的 dashDB MPP 功能，客户可以根据需要从单节点轻松扩展至企业级 MPP 数据仓库集群，在一段时期的高峰负载过去后，又可以回缩至单个 SMP 节点。在 MPP 架构内，数据可以采用本机 BLU 列式矢量格式在查询引擎内的各个服务器（节点）之间交换。当前，dashDB Local MPP 向外扩展的最大限度为 24 节点（2016 年 8 月）。HRG 预计，只要客户需求出现，IBM 将轻松解决这一问题。

高度可用性

如果发生节点或服务器故障，dashDB Local 可提供高可用性功能；因此，当故障发生时，数据仓库可以继续活动。实现该功能的基础是具备自动重启和故障转移的脉动信号机制。脉动信号可以检测出宕机的节点、分区、Web 控制台或 IP 地址，随后集群管理器将采取恰当的措施。此外，客户也可以提供冗余的集群文件系统存储，以期获得更可靠的存储管理。

向外扩展和向内扩展

dashDB 以数种方式提供简单的可扩展性。CPU 和内存资源被添加至特定集群，以便向上扩展。新节点被添加至集群，以便向外扩展。高峰工作负载一旦完成，资源（CPU、内存、节点）便可返回池中，以便向内扩展。这些操作要求客户停止并通过添加的其他资源重启系统上的服务。这一功能目前尚未实现完全自动化，但 HRG 预计 IBM 未来将提供完全自动化的功能。企业的工作负载计算资源需求差异很大。由于 dashDB Local 具有灵活的扩展功能，因此，当业务工作负载需求因季节性需求变化出现扩展或缩减时，它可以成本高效地满足关键的 SLA 指标。与传统的行存储数据库相比，dashDB 借助动态内存内列式数据存储技术最大程度地减少了 I/O，显著提高了速度。

数据库内分析

由于使用 R、RStudio 和 Netezza 数据库内分析进行 dashDB 数据库内分析，因此可在数据所在地进行查询。此外，IBM 以及合作伙伴也提供其他分析工具，包括 Cognos、ESRI 和 Aginity。dashDB 旨在与 Watson Analytics 一同获得认知智能。正如上文 dashDB Local 部分所述，由于 dashDB 与 Spark 的集成，数据科学家可以借助 Spark 集群成员和 dashDB 集群成员之间的紧密集成，将 dashDB 用作以 Scala 或 Python 编写的应用的数据层。

用例：医疗保健与个性化医疗

由于使用远程监控设备将病患数据收集并传输至医疗专业人员，医疗大数据不断增长。借助遥测数据，医护人员可以监控并跟踪个人，由此在情况恶化前早期介入。利用遥测医疗设备，可以更及时地为老年人和农村病患提供医护。使用这些设备后，可以通过降低再入院率减少医护成本，也可降低相关的运输费用。由于更易获得医护人员的帮助，而且医护人员的响应性也有所提升，因此实现了显著的优势。现在，很多病患都使用在线聊天、Web 门户或电子邮件与医生、护士和医护人员沟通。这种在线沟通进一步增加了 EHR（电子健康记录）量和其他病患数据。

最终，dashDB Local 平台可以实施分析，以便匹配提供医护和服务过程中整个完整设置的病患健康记录；由此，医生、医护团队成员甚至病患自己都可以查看个体病患医护需求的完整视图。随着人们越来越多地使用移动智能手机和远程监控技术管理和报告实时的健康信息和数据，医疗管理和交付机制将持续发展。

医疗保健提供商从运营、病患、网站和其他来源生成越来越多的数据。通过记录该数据并进行分析，可以提供重要的可执行信息，从而为病患、提供商和支付方提供重大优势。借助内存内处理和数据库内分析功能，IBM dashDB 可以在临床或远程设置中首次出现症状时实现疾病的有效诊断和治疗。通过实时分析，dashDB 有助于发现全球性健康威胁以及逐渐随之出现的流行病。此外，该解决方案还可用于将最有效的干预、控制和处理方法通知第一响应团队，从而避免巨大的灾难和不必要的人员伤亡。

借助 dashDB 和 dashDB Local，研究人员可以利用内存内处理和数据库内分析，将某个基因组与整个人类基因组的基础水平映射或同等代理进行比较，以便分析异常或差异。这将有助于研究人员发现和解析疾病特定的遗传标记，从而构建个体/个人化的预防和治疗性医疗应用。安全、隐私、灾难恢复、可用性以及其他新问题都必须得到解决，才能为履行个性化医疗承诺创造条件。不管怎样，dashDB 的数据加密和强大的访问控制对于这些问题的解决大有帮助。

用例：石油和天然气

石油和天然气信息技术若要产生作用，就必须摄取、集成、分析和管理来自油井和地面设施上安装的成千上万传感器的数据，以便持续地实时监控所有的设备、油井和环境条件。现在，借助 dashDB Local 的内存内处理和数据库内分析功能，石油和天然气专家可以将当前条件和数据与所有已知和类似的状态相比较，从而快速做出正确的决策，产生与物流、供应商、营销和交易相关的可执行洞察力，进而改善和支持从石油到交付的整个能源生产和交付供应链的管理。借助使用 R 的预测性数据库内分析，同时将当前和实时数据与历史数据相结合，石油和天然气供应商可以预测和消除潜在的设备 and 油井故障，避免人为失误。

用例：金融服务

金融服务公司若要缩短产品和服务创新周期，就必须部署和管理核心基础架构系统，同时还需应对有限的时间、资金和技术资源。对于分析、规划和决策制定而言，基于最佳可用数据的优质信息至关重要；在此，dashDB 能起到重要作用。在整个金融服务领域，实时、快速响应竞争和市场变化的能力已然不可或缺。

金融服务公司的高管、总监和管理人员都知道，如果能够在无需遭受潜在复杂性的前提下利用所有的企业和市场数据，将有助于获得优势。时至今日，没有任何类型、位置和格式的数据无法进行集成。公司性质决定其必然不断发展并积累数据。如果公司能够正确地解释历史业务和市场数据，并在最短的运行时间内将该知识用于分析当前业务和市场数据，就将成为市场领导者。随着公司从业务运营、网站、营销计划、销售和其他来源收集的数据日趋增多，如果能够记录此类信息，并借助 dashDB 提供的内存内处理和数据库内分析功能，通过数据分析在业务中充分利用此类信息，就能获得重要而独特的竞争优势。

结论及建议

若要实现云计算、大数据和万物互联，解决方案就必须能够组织和理解目前可用的不断增长的数据。对于传统数据仓库和商务智能而言，多年来一直都采用数据集成。对于业务成功而言，能够结合来自多个来源的实时信息的新应用变得至关重要。若要过渡到更加“数字化和社交化”的业务模式，发展成为数字化、社交化和基于云的实体，企业就需要正确的数据和数据管理工具，以支持当今客户面临的 B2B 应用演变。

因此，组织已开始通过实施聚合和超聚合系统来精简和简化 IT 基础架构，以提高能源效率，降低 IT 的物理基础架构需求。这些组织正在迁移至混合云实施，以期降低客户和员工使用信息和应用的难度，同时提高生产率，降低成本。随着时间的推移，很多组织将几乎完全迁移至 Cloud。IBM dashDB 可帮助组织转变自身，以期更好地利用 Cloud、大数据和物联网，准备好迎接瞬息万变的未来。dashDB 基于云的托管服务可提供简易性，让客户在云端完成实施，而由 IBM 负责应用和系统的维护。如果您想要充分利用竞争日趋激烈、瞬息万变的全球业务环境，IBM dashDB 和 dashDB Local 值得认真考虑。

Harvard Research Group 是一家信息技术市场研究和咨询公司。该公司为计算机硬件、软件和服务供应商及用户提供高度集中的市场研究和咨询服务。如欲了解更多信息，请联系 Harvard Research Group：

Harvard Research Group

Harvard, MA 01451 USA

电话：(978) 456-3939

电子邮箱：hrg@hrgresearch.com

<http://www.hrgresearch.com>