

## 用于高速分析的数据仓库

*IBM DB2 with BLU Acceleration* 具备将信息转化为行动力的速度和灵活性



大数据不一定会成为大问题。事实上，大数据恰恰是汇集更多数据，快速分析并得出优化结果的良机。IBM 推出的新一代数据管理解决方案，可帮助企业更方便地分析由历史数据和当前数据构成的大型数据集，以更快的速度解答运营查询。

要在当今这个瞬息万变的业务环境中赢得竞争，企业必须具备审时度势、分析结果并迅速调整战略的能力，而这一切还要在有限的预算下完成。身处这个大数据时代，企业必须优化其数据仓库，为运营和分析系统提供支持，并在速度、简便性和经济性之间取得绝佳平衡。

IBM® DB2® 10.5 能帮助企业满足上述所有要求，而与管理、数据整合、安全和查询工具等有关的其他需求也可尽数解决。企业愈发为大数据所扰，DB2 可助其将问题化繁为简。

### 优化多工作负载场景中的性能

随着数据量的不断增加，用以查询数据库的最有效方法也在不断进化。有些交易查询只需相对有限的数量即可完成，但设想这样一个常见的零售场景：过去五年内，某种鞋款在某一特定地区共售出多少双？这一查询需要从多个方面进行分析，将涉及数百万条记录，可以在单一服务器上的一个数据集市（数据仓库的子集）中运行。

其他查询要更为复杂。某些运营查询中存在一定的实时或近实时元素，或者当中包含多个来源的数据。设想一项可能牵涉信用卡欺诈交易的数据库查询。要确定是否应拒绝这项交易，必须同时回答这样几个问题：所购商品是否为卡主经常购买之物？交易发生地是否在卡主居所附近？卡主的人口统计信息是否出现任何异常而需要对交易做进一步审查？

---

*“凭借使用 BLU Acceleration 的 DB2 10.5，我们的存储要求相较于未压缩数据表缩减 10 倍之多，查询性能也提高 10 倍以上。与竞争产品相比，DB2 10.5 with BLU Acceleration 占用的存储空间明显减少，其性能更是优于竞争产品 3 倍之多。”*

—欧洲电信服务提供商<sup>1</sup>

---

有多少种查询，几乎就有多少种数据仓库工作负载。DB2 10.5 能够支持多工作负载部署场景，其采用的先进技术有助于优化多类型工作负载的性能，此外，DB2 10.5 还能凭借新的加速技术，增强单一服务器场景中的报告和分析性能。

### 单服务器环境中的高性能分析工作负载

数据仓库查询技术领域的一项最新发展成果即内存处理。处理 Linux、UNIX 或 Microsoft Windows 服务器中的路径，将数据在内存与磁盘间相互转移，这些都会对查询性能产生巨大影响。不过，如果查询能在内存中完成，响应速度就会加快。这正是 DB2 10.5 with BLU Acceleration 的作用所在，不仅如此，它还能将超过内存大小的所有数据

挪到磁盘上，这使其具有比其他内存解决方案更进一步的优势。

DB2 10.5 with BLU Acceleration 完美适用于原始数据不超过 10 TB 的分析数据集市，它将这些新的行业技术与 IBM 成熟技术结合起来，用以提升运行性能，满足企业加快决策收益的需求。在 DB2 10.5 with BLU Acceleration 的运行效果方面，IBM 发现典型分析工作负载的性能有 8 至 25 倍的提高。<sup>2</sup>

为进一步提高查询速度，DB2 with BLU Acceleration 还应用了包括压缩数据、跳过查询不需要的数据、在多个处理器内核中分配查询在内的其他技术（参见图 1）。

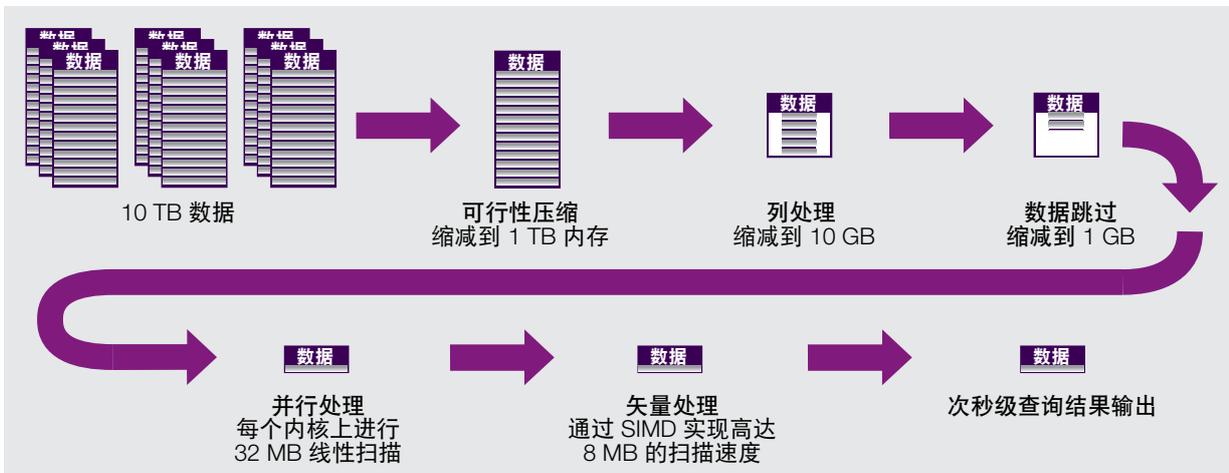


图 1: DB2 with BLU Acceleration 采用新的压缩和处理技术，为内存查询结果输出提速。

由于 DB2 with BLU Acceleration 能够在现今高端服务器领域的 32 核和 64 核处理器中并行处理，以前需要多条指令才能实现的操作如今只需一条指令便可完成，大大加快了处理速度。查询前，无需事先对编码数据进行解码或解压缩，进而节省处理时间和 I/O。DB2 10.5 with BLU Acceleration 以 DB2 10.1 中引进的现有压缩技术为构建基础，能够更加高效地处理列数据，进而减少存储空间需求，加快查询速度。据客户反映，相较于未压缩数据表，DB2 with BLU Acceleration 的压缩率要高出 10 倍。<sup>3</sup>

同时，DB2 with BLU Acceleration 也能简化构建商业智能查询的过程。省去了耗时的管理和调整工作，只需按动按钮即可轻松完成。采用该流程后，查询速度将相应提高，业务洞察力也能快速形成。

---

*“借助 BLU Acceleration，我们的数据表足足小了十倍。此外，我不需要创建索引或汇总，也不需要给数据分区，还省去了很多其他麻烦。考虑到我们所用的数据表类型多种多样，这个数字可达到 10 到 25 倍。”*

—Andrew Juarez，联合可口可乐瓶装公司 SAP Basis 和 DBA 负责人<sup>4</sup>

---

---

### DB2 10.5 实际效果解读：销售数据分析

通过销售数据分析查询，可以看出 DB2 with BLU Acceleration 以企业数据仓库卸载来提升分析性能的实力。解答销售趋势和人口统计等相关问题需要对多个系统的数据进行划分，然后快速、高效、准确地将其提供给分析和联机应用处理 (OLAP) 应用程序。利用 DB2 10.5，只需简单几步便可创建并加载 BLU Acceleration 内存数据集市，也无需创建或调整任何索引或汇总。多平台软件灵活性延展了整体接触面，而预构建模型包又可进一步提高部署速度。分析数据集市能够与业务同步创建，用以随时呈现及时、相关的销售数据、预测和洞察。

---

### 分配实时的运营分析工作负载

在需要从多个服务器调取数据或所需数据量超过 10 TB 的场景中，DB2 10.5 也能凭借 DB2 Database Partitioning Feature 提供应对复杂运营分析的绝佳扩展性能。

运营分析即应用于运营系统的分析。其特点是：需要对实时信息发起海量并发查询，同时调取数据仓库的查询数可达数百或数千项（或更多）。在这种场景中，有相当比例的查询（往往多达 80%）都只是关注于特定客户、账户、病患等相关数据的简单交互式查询。这些运营查询可能源自呼叫中心、移动销售应用程序、实时欺诈检测系统或其他支持实时决策制定的应用程序。



DB2 具有实时运营数据仓库功能，能够以较高频率、不间断地从运营系统中快速捕获精确数据。这些功能可帮助应用程序实时访问所需信息，以用户所需的时间、场合和方式提供深入洞察分析，进而为以事实为依据的业务决策提供支持，帮助提升企业绩效。

凭借 DB2 的先进数据分区功能，IT 用户可在大规模并行处理、对扩展性有所要求的场合中在多个服务器间分配数据。DB2 可以跨多个分区透明分割数据库，借助多台服务器的力量满足对大量信息的请求。SQL 语句会自动分解成在不同分区间并行运行的子请求，在最终结果内再重新整合起来。

从业务角度来看，当中的优势显而易见。无论大数据在企业中的发展有多快，其增长规模有多大，带来的多样性有多少，DB2 都能凭借出色的性能和扩展能力，提供可以与业务同步的查询结果。

---

#### DB2 10.5 实际效果解读：加快报告速度

借助 DB2 with BLU Acceleration 可增强传统数据仓库的性能，以更快的速度完成报告请求。无需对应用做任何更动，只要向 BLU 进行简单的数据表迁移，即可创建分析数据集市。利用 DB2 with BLU Acceleration，企业可优化现有数据仓库的商务智能 (BI) 工作负荷，为决策者提供必要信息，便于他们据此采取自信、迅捷的行动。

---

## 简化管理、提高生产力

借助当前植入 DB2 kernel 中的 BLU Acceleration 技术，DBA 需要花在例行维护及调整作业上的时间很少甚至为零。传统行数据表可直接迁移并转换为列数据表，众多其他管理活动现在也可实现自动化处理，这使得 DBA 能够专注于价值更高、更具创新性的工作。

除此之外，DB2 10.5 可将更大型工作负载的管理有效性和简便性提高到新的水平。例如，当可扩展性加快了多数据源的查询速度时，其他数据库处理活动也不能放缓。基于此，DB2 10.5 提供先进的工作负载管理功能，可帮助 DBA 建立并执行服务水平。现在，IT 部门可以优化来自不同用户和应用程序的查询，然后控制专用于每个流程的底层资源量，确保高优先级项目在需要时可获得处理资源。

内置工具有助于 DBA 改善以下性能：

- **持续数据摄取：**可同时跨多个处理器加载数据；在资源间实现动态切换，以最大化资源利用率；改善延迟（在对操作源执行实时分析，以便优化动态业务环境中的决策过程中，这是一项重要的影响因素）。
- **时间旅行查询：**用户可创建 SQL 查询，以追踪某个时间范围内的变更情况。随着数据保留时间的增长且相关性日益降低，可将其从数据库中删除。
- **先进的安全性：**DBA 可基于用户或组轻松确定具体人员对于任一数据表每一行和列进行读取和写入的访问权。这意味着，在允许用户获得相同数据的不同视图的前提下，DBA 仍然可以将安全或敏感数据屏蔽起来。
- **嵌入式分析功能：**这些工具可为广泛用户提供宝贵的商业智能。其中包括 Cubing Services for OLAP，借助该工具，无需从数据仓库中提取数据，即可进行多维数据分析，并且可向 IBM Cognos® Business Intelligence 软件提供数据。其他功能还包括针对 Microsoft PivotTable Service 的原生支持，这允许在执行 Microsoft Excel 应用程序作业的同时，进行即席分析。

## 聚焦经济性

IT 预算压力持续增大，削减成本势在必行。凭借内置智能、简便性及易用性，DB2 可缩减在 IT 管理、维护和调整方面的作业量，从而提供绝佳的经济性，为您节省大量运营成本。

DB2 具有可行性压缩和数据跳过功能，可帮助降低存储成本、节省 I/O 和内存资源，同时消除对索引的需求。通过使用以 DB2 with BLU Acceleration 部署的内存列存储方式，在 Linux、UNIX 和 Windows 服务器上运行的系统降低了磁盘存储需求。这些功能也有利于节省图像备份空间和备份所需的时间。与之前的 DB2 版本相比，DB2 10.5 需要更少的数据表和索引，系统因此可节省这些应用所需的

空间。另外，还涉及经济因素：DB2 的成本只有 Oracle 数据库的三分之一。<sup>5</sup> DB2 与 Oracle PL/SQL 的平均兼容性高达 98%。<sup>6</sup> 基于此强大的 SQL 兼容性，您无需重写使用 PL/SQL 的应用程序，即可将其迁移至 DB2 中，而且后续的学习曲线很短，您可以充分利用这一优势，更快地获得投资回报。

---

*DB2 的成本只有 Oracle 数据库的三分之一。<sup>5</sup> DB2 与 Oracle PL/SQL 的平均兼容性高达 98%。<sup>6</sup>*

---

## DB2 10.5: 前所未有的灵活性与强大性能的相结合

DB2 with BLU Acceleration 的最大优势在于，其整合了 IBM 研发实验室开发的整套创新技术，包括列、内存、先进的压缩算法、压缩、内存缓存、亲内核的并行操作以及单指令多数据集。上述技术融合在一起，可提供无与伦比的速度、简便性和资源节约，成就高性能的数据仓库。

现在，企业可针对多个工作负载优化其数据库软件，从而获得空前的灵活性，以适应并满足企业当今的业务需求：实现高速分析。

## 了解更多信息

如需了解有关 DB2 with BLU Acceleration 及其数据仓库功能的更多信息，请联系 IBM 代表或 IBM 业务合作伙伴，或者查阅以下资源：

- [ibm.com/db2/luw/](http://ibm.com/db2/luw/)
- 电子书：《Three must-have capabilities for today's warehouse environments》（当今仓库环境中的三个必备功能）：<http://ibm.co/111TYc0>
- 视频：BLU Acceleration in IBM DB2 概述：<http://ibm.co/11nebMY>
- 视频：紧跟业务发展速度的分析 – IBM DB2 with BLU Acceleration：<http://ibm.co/10lmxuC>



---

© Copyright IBM Corporation 2013

IBM Corporation  
Software Group  
Route 100  
Somers, NY 10589

美国印制  
2013 年 4 月

IBM、IBM 徽标、ibm.com、Cognos 以及 DB2 是国际商业机器公司的商标，已在全世界许多管辖地区注册。其他产品和服务名称可能是 IBM 或其他公司的商标。当前的 IBM 商标列表请参见网站的“版权和商标信息”版块：[ibm.com/legal/copytrade.shtml](http://ibm.com/legal/copytrade.shtml)

Linux 是 Linus Torvalds 在美国和/或其他国家/地区的注册商标。

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

UNIX 是 The Open Group 在美国和/或其他国家/地区的注册商标。

本文档包含截至发布之日的最新信息，IBM 可能随时更改。并非所有产品或服务在 IBM 开展业务的所有国家/地区均有提供。

本文所载信息按“原样”提供，不做任何明示或暗示的担保，包括对适销性、特定目的的适用性的任何担保，以及针对非侵权的任何担保或条件。IBM 根据产品交付协议中规定的条款和条件为产品提供担保。

关于 IBM 未来发展和意向的声明仅表示目标和意愿，可能随时更改或收回，恕不另行通知。



请回收再利用

---

<sup>1</sup> DB2 10.5 “早期使用计划”。

<sup>2</sup> 依据是 IBM 的内部样本分析工作负载（不含事务或 OLAP 工作负载）测试，测试中比较了 DB2 10.1 上基于行的数据表与 DB2 10.5 上基于列的数据表的查询访问性能。性能改进数据是工作负载中所有查询的累计量。根据工作负载、配置和条件的不同，个别结果会有所不同。

<sup>3</sup> DB2 10.5 “早期发布计划”是从客户端报告测试中衍生而来。根据工作负载、配置和条件（包括数据表大小和内容）的不同，个别结果会有所不同。

<sup>4</sup> DB2 10.5 “早期使用计划”。

<sup>5</sup> 依据是截至 2013 年 4 月 23 日有关 IBM DB2 Advanced Enterprise Server Edition 与 Oracle Database Enterprise Edition Software 公开可用的美国定价信息，后者包含可提供类似功能性的额外定价产品/功能（Oracle Real Application Cluster、Advanced Compression、Advanced Security、Partitioning licensing、Oracle Enterprise Manager: Database Diagnostics Pack、Oracle Enterprise Manager: Database Tuning Pack、Oracle Enterprise Manager: Lifecycle Management Pack、TimesTen In-Memory Database、Mobile Server、NoSQL Database Enterprise Edition）。对比价格不包含适用税费，并可随时进行更改，恕不另行通知。IBM：假定为 120 个处理器价值单位。Oracle：假定为 1.0 处理器乘数。二者均包含 Y1 维护/支持。

<sup>6</sup> 依据是 2011 年 9 月 28 日至 2012 年 3 月 7 日期间内部测试和报告的客户体验。