

## 打破“添加更多节点”的循环

*IBM Netezza* 数据仓库设备可用于  
扩展数据基础架构



## 目录

- 2 时间价值
- 4 灵活性
- 4 长期维护成本
- 5 性能

执行复杂分析的能力正在成为新的核心竞争能力。分析应用包括客户和利润分析、欺诈检测和风险分析。通过数据仓库可以近实时收集、合并和分析非常大量的客户数据；部署此技术的组织可以提供更好的客户服务，树立竞争优势。

在当今充满挑战和快节奏的经济环境中，时间极其重要：组织需要立即获得答案，不能再等待深度分析查询运行数小时甚至数天。分析通常是重复的；当分析人员搜索数据中深藏的线索时，一条查询的答案可能会产生一连串更多的问题。传统的分析方法使用数据样本，其缺陷显而易见——被分析的数据量限于一小部分数据，洞察力受到削弱，无法预见到可能引发重要的新营销活动或关注新型欺诈团伙的趋势。

一直以来，Teradata 与其他数据库供应商的差异在于，其产品特别为在线分析处理 (OLAP) 运营而设计，而不像其他的用户产品，最初设计用于在线事务处理 (OLTP) 工作负载，然后大幅调整系统来改进用于数据仓库工作负载的性能。Teradata 的架构差异现在基本上已经消失，因为大规模并行数据库和数据仓库已经成为行业标准。

IBM Netezza 数据仓库设备通过提供简便易用，具有行业领先的超高性价比的平台，改变了整个数据仓库和分析环境。该数据仓库设备在高级分析中开拓出新的领域，能够以惊人的速度解决重大的处理挑战，没有障碍或妥协。事

实证明，IBM Netezza 数据仓库设备的速度、灵活性、简便性和可扩展性，对于使用现有系统时遇到性能或存储障碍的 Teradata 客户很具吸引力，因为他们短短几天便可将新的应用程序和工作负载轻松地转移到 IBM Netezza 数据仓库设备。考虑到 IBM Netezza 数据仓库设备的价值生成时间，许多 Teradata 客户已将其部分或全部数据和分析应用程序转移到 IBM Netezza 数据仓库设备。

客户转移到 IBM Netezza 数据仓库设备的原因：

- 价值生成时间；
- 随着新的工作负载而发展的能力和灵活性；
- 降低其长期维护成本；
- 提高查询性能——特别是对于数据量不断增长时的复杂分析和即席环境。

### 价值生成时间

无论是建立新的业务或应用程序、为即将举行的直接营销活动购买数据，还是创建供业务用户运行即席分析的沙盒，系统的安装、连接、运行以及数据加载速度越快，就能越早开始赚钱或省钱。

虽然 Teradata 宣称其系统的设置比其他数据仓库系统更简便、更快，但在开始加载数据之前需要进行大量的前期规划和设置。我们的客户报告，Teradata 在上线之前需要数周的时间来创建数据库，然后经过几个月的系统调整才可达到合格的性能。这些工作由 Teradata 客户自己提供支持，其中某些客户甚至必须在维度模型中创建整个数据集的副本才可实现其 SLA。此维度模型可能包含许多星型结构，以及多份含有不同数据层级的数据表。<sup>1</sup>使用此方法需要处理的问题是，应用程序开发人员和业务用户必须完全了解这些表格和集合中实际包含的数据，否则他们可能会根据旧的数据快照或少量歪曲的数据样本做出业务决策。

此外，如果用户误选了错误的“表格”，其查询的运行时间可能会延长 10 倍。这种工作不仅耗时，成本昂贵，也没有什么业务价值，通常需要对加入仓库的每个新业务或应用程序重复执行。

Teradata 政府服务部的 Thomas Glowacki 认为：“Teradata 对其数据结构使用第三范式，是针对构成决策支持和分析处理基础的即席查询的最优架构。”<sup>2</sup>以第三范式管理数据会导致数据库管理系统为大多数查询连接多个表格。而连接——导致大量数据在系统中四处移动——可能是基于 MPP 的数据库崩溃的祸根，因为它们可能严重影响查询性能。Teradata 对这种根本性弱点的解决方法是，建议其客户通过数据库管理员来创建和维护表格中类型繁多的“连接索引”。在 Teradata 数据库中，如果连接索引是使用不同的列加入到表格中，表格具有 3、4、5 个或更多连接索引的现象很普遍。这些连接索引需要了解系统中目前及未来执行的查询、报告和分析。连接索引是实现价值生成时间的障碍——这些数据库结构要耗费技术专家的宝贵时间，Teradata 对这些结构的需求可能使其数据库系统无法支持真正的即席查询。而且，当您尝试添加更多应用程序或数据时，Teradata 数据库管理员需要不断调整（这还算幸运的）或扩展系统。总而言之，Teradata 仓库部署速度慢、维护和调整成本高、难以立即为最终客户提供洞见。

IBM Netezza 数据仓库设备平台的简便性可创造市场上最快的价值生成时间。使用 IBM Netezza 数据仓库设备，无需索引，不用担心空间分配，也不需要调整数据库参数。这意味着，IBM Netezza 数据仓库设备系统一经交付并插入网络，您只需创建一次表格、加载数据和运行报告。新的工作负载可以轻松纳入系统，不需要重新处理所有索引、集合以及为现有工作部署的其他结构。

由于所有这些前期规划和额外的工作，Teradata 数据仓库开始运行的时间比 IBM Netezza 数据仓库设备要迟数倍。Overstock.com 是 Teradata 的参考客户之一，他们在 180 天后才开始在 Teradata 上完全实施其新的数据仓库。<sup>3</sup>为了减少麻烦，Teradata 采用分阶段的方法，在 60 天和 120 天后“释放”部分数据。

与此相比，Con-way Freight 使用 IBM Netezza 数据仓库设备，初始设置不到 48 小时，然后在两天之内，就可以加载数据，用户也开始访问系统。在概念验证系统到位的情况下签订购买订单后四天，Con-way Freight 内部用户就可以使用新的 BI 项目。从购买之日起三个星期，IBM Netezza 数据仓库设备便全面投入使用。<sup>4</sup>

对于部署 Teradata 的组织，项目前 30 天通常是检查其数据和查询，以确定如何创建数据库来确保查询有效地运行。<sup>5</sup>如果在批处理和现成报表的时代，这或许可以接受，但在当今时效性强的分析时代，则明显落伍了。

<sup>1</sup> 会议 1263，合作伙伴，2011 Teradata User Group - Aggregate Join Indices & Dimensional models delivering extraordinary performance (具有非凡性能的集合连接索引和维度模型) - 作者: APP Consulting and Independence Blue Cross 的 Jose M Borja 和 Lana Batyaykina

<sup>2</sup> Thomas Glowacki (隶属 Teradata 政府服务部) 向俄亥俄州医疗补助研究计划委员会提交的书面声明，2006 年 10 月 4 日

<sup>3</sup> Teradata 客户故事 - Overstock.com

<sup>4</sup> Con-way Freight 案例研究

<sup>5</sup> Teradata Accelerate for Data Warehousing

## 数据仓库服务时间表

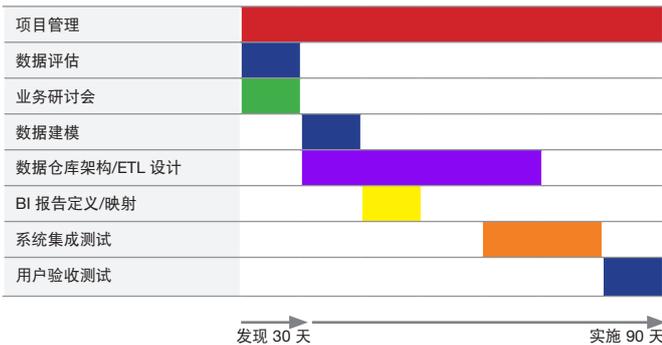


图 1: Teradata 的数据仓库规划和部署时间表。(资料来源: Teradata Accelerate for Data Warehousing)

IBM Netezza 数据仓库设备的客户不需要 30 天来规划其数据结构，只需要复制其现有表格的表格定义（架构）、加载数据，然后便开始运行其查询/报告。在 Teradata 可能需要数周规划时间和数月部署时间的领域，IBM Netezza 数据仓库设备可以提供非常快的价值生成时间。

## 灵活性

在创建数据仓库之后，您很快就会意识到，数据库将不断增长，而且通常是非常快速、急剧地增长。这种增长一般来自新数据、新应用程序、新用户，更可能是这三因素的组合。若不扩展系统，数据量或性能最终会达到组织现有系统的临界点。该环境中新增应用程序后，通常需要添加未计划的新节点/服务器，还要在已经拥挤不堪的计算机中心添加更多物理硬件。这样的结果不是增加业务价值，而是削弱价值，无论是时间还是成本都无法承受。

在这些情况下，IBM Netezza 数据仓库设备为客户提供经济、高性能、部署简便的方案，用以替代扩展 Teradata 系统。许多 Teradata 客户选择 IBM Netezza 数据仓库设备作为其下一代数据仓库和 BI 系统平台。这些客户将

IBM Netezza 数据仓库设备部署成连接其核心企业数据仓库 (EDW) 部署“轮毂”的高性能“轮辐”，或者部署成全新、完全独立的企业数据仓库平台。这样客户在短短数天内便可新增数据和应用程序，而不需要数月时间，可以立即收获新的洞见，提升利润。

## 长期维护成本

上一节提到 Teradata 的长期维护问题；下面多花点时间来探讨一下。分析人员在进行深入的思想研究时，需要对他/她的问题立即做出反应——任何延迟都会打断其思路，结果往往失去很多机会。如果答案与预期不一样，分析人员会提出不同的问题。意外的答案可能会触发其他问题——都可以对所研究的问题产生新的洞见和了解。这种思想的速度研究的现实是，分析人员无法预先知道数据或分析的方向，也不知道他们将要根据提示问哪些问题。

如何针对此类工作负载调整数据库？无法调整。因为您无法预先确定具有创意的分析人员可能提出哪些问题，无法为每个可能的查询创建索引。如果您找到一些重要的数据，并且想重复运行同一查询，数据库管理员需要事后做出反应，并创建所需的索引。《Teradata 杂志》中一篇标题为“Peak Performance”（尖峰性能）的文章表示，Teradata 数据库管理员需要“调整优先级”，<sup>7</sup>同时指出，他们的“数据仓库经理和数据库管理员通常缺乏调整现有应用程序所需的时间、培训或资源。”<sup>8</sup>

IBM Netezza 数据仓库设备革命化的设计提供卓越的性价比，无需编制索引或进行调整。作为用于高速分析的专用设备，优势不仅在于其具有功能最强大、最昂贵的组件，更在于其恰当组合的组件能协同提供最大的性能。现场可编程门阵列 (FPGA) 是 IBM Netezza 数据仓库设备独有的功能，在许多情况下都能提供远超预期的性能。而且，作为一个使用简便的设备，系统可开箱即用直接提供显著的效果。每当执行查询时，优化的硬件可以立即产生分析查询性能，不需要调整和创建索引。

<sup>7</sup> Peak Performance (尖峰性能) : Teradata 杂志 (2009 年 6 月)

<sup>8</sup> Peak Performance (尖峰性能) : Teradata 杂志 (2009 年 6 月)

## 性能

业务数据中的重要洞见往往发现于“边缘”——与平均值的两项或更多标准偏差。这意味着要分析所有数据。创建少量数据样本似乎是维护查询性能的捷径，但可能无法了解样本之外的数据中的真实信息。也就是说，您的洞见只是来自样本数据，而非整个数据集。

IBM Netezza 数据仓库设备架构采用基本的计算机科学原理：在处理大型数据集时，除非绝对必要，否则不会移动数据。IBM Netezza 数据仓库设备完全遵循此原理，它利用称为现场可编程门阵列 (FPGA) 的商品组件尽早滤掉数据流中无用的数据，并尽快从磁盘中获取数据流。这个数据消除过程靠近数据源，可突破 I/O 瓶颈，避免 CPU、内存和网络等下游组件处理不必要的信息，因而对系统性能产生数量级的巨大效应。<sup>9</sup>

根据我们的经验，随着数据和用户量的增长，Teradata 客户无法实现 IT 与业务部门之间商定的服务水平。为了保持合理的性能水平，IT 部门陷入一种进退两难的困境：购买更多节点、限制用户，或者通过创建样本、汇总和剪切旧数据将企业的查询限于一部分数据。后者会使企业失去其宝贵的资产之一：历史数据。

IBM Netezza 数据仓库设备为客户提供经济、高性能、部署简便的 Teradata 替代方案。许多 Teradata 客户选择 IBM Netezza 数据仓库设备作为其下一代数据仓库和 BI 系统的平台。这些客户充分利用 IBM Netezza 数据仓库设备平台行业领先的价值生成时间，实现效益的速度比任何其他平台都快。

IBM Netezza 数据仓库设备特别设计为可以对超大量的数据执行超快的分析处理。IBM Netezza 数据仓库设备无需创建索引和持续调整，也没有磁盘传输 I/O 速率慢、网络传输速率慢和低效缓存而引起的技术瓶颈。

加入今天的分析革命——从被动转为主动，从报告历史趋势转为根据预测性分析和优化来预测未来结果。希望先于竞争对手寻求、找到和利用机会的主动型企业，利用 IBM Netezza 数据仓库设备平台高性能、低成本的先进分析功能来支持其业务战略。IBM Netezza 数据仓库设备这种强大的高性能、可扩展、简便易用的先进分析技术可以挖掘您的业务潜力，提供持续的竞争优势。

## 关于 IBM Netezza 数据仓库设备

IBM Netezza 数据仓库设备是对数据仓储和高级分析的革命，将数据库、服务器和存储设备集成到一个易于管理的设备中，几乎不需要设置和长期管理，同时提供更快、更一致的分析性能。IBM Netezza 数据仓库设备系列将所有分析活动合并到设备中，准确找到数据所在的位置，可以极大地简化业务分析，提供行业领先的性能。请访问 [ibm.com/software/data/netezza](http://ibm.com/software/data/netezza)，了解我们的数据仓库设备系列如何消除每个步骤的复杂性，帮助您为组织创造真正的业务价值。有关最新的数据仓库和高级分析博客、视频等，请访问：[thinking.netezza.com](http://thinking.netezza.com)。

<sup>9</sup> Netezza 架构白皮书

## 关于 IBM 数据仓储和分析解决方案

IBM 提供最广泛、最全面的数据仓储、信息管理及业务分析软件、硬件和解决方案的组合，帮助客户实现其信息资产的最大价值，发现新的洞见以做出更好、更快的决策，优化其业务成果。

## 了解更多信息

如需了解有关 IBM 数据仓储和分析解决方案的更多信息，请联系 IBM 销售代表或 IBM 业务合作伙伴，或访问：  
[ibm.com/software/data/netezza](http://ibm.com/software/data/netezza)



---

© IBM 公司版权所有 2011

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

美国印制  
2011 年 12 月

IBM、IBM 徽标、ibm.com 和 Netezza 是 International Business Machines Corporation 在美国和/或其他国家/地区的商标或注册商标。如果这些商标及其他 IBM 商标是首次以商标标志 (® 或 ™) 出现在信息中，则这些标志指该信息发布时 IBM 在美国的注册商标或普通法商标。此类商标还可为 IBM 在其他国家/地区的注册商标或普通法商标。当前的 IBM 商标列表请见网站的“版权和商标信息”版块：  
[ibm.com/legal/copytrade.shtml](http://ibm.com/legal/copytrade.shtml)

其他公司、产品和服务名称可能为其他公司的商标或服务标识。



请回收再利用