

利用 IBM InfoSphere Warehouse V10 提供性能更高且成本更低的解决方案

IBM 红皮书 - 解决方案指南

及时访问精确信息是致力于更好地服务客户、赢得竞争和促进创新的企业的关键。IBM® Information Management 提供一个全面的数据仓库解决方案（图 1），因此组织可以作为其运营和策略业务应用程序的一部分集中、精确且安全地分析和交付信息。



图 1. 全面的数据仓库解决方案

IBM InfoSphere® Warehouse V10 提供一系列超越传统仓库功能的强大功能。这个全面的平台将 IBM DB2® 数据库的优势和动态数据仓库基础结构相集成，能够处理传统商业智能 (BI) 工作负载以及更高的运营业务需求。此外，InfoSphere Warehouse Advanced Enterprise Edition 交付一组增强的数据库性能、管理和设计工具。这些工具帮助公司维护和提高仓库价值，而且还帮助减少维护这些复杂环境的总成本。

您是否知道？

随着整个星球实现了更透彻的感知和度量、更全面的互联互通和更深入的智能分析，数字信息总量和种类（结构化和非结构化）急剧增长。仅仅是社交媒体，就涉及 **TB** 级别新数据。成功的关键是从数据中获取洞察力并利用其获取业务商机。利用 **IBM InfoSphere Warehouse**，您将具有此能力。

业务价值

通过高级数据仓储技术，**IBM** 帮助组织从几乎任何类型的数据中抽取洞察力。**IBM** 帮助在恰当时间在适合的环境中交付正确信息，从而使业务主管能够快速制订正确决策。**IBM** 高级仓储解决方案可集成数据仓储和业务分析，帮助定义组织的中心业务概念以及支持这些概念所需的数据。这些解决方案允许组织从各种企业和源系统捕获数据更改，而这些是传统的 **BI** 和数据仓储解决方案过去无法访问的数据。

因此，**IT** 组织可以更好地针对可操作信息支持业务需求。此信息不仅是原始数据，这些数据还受智能支持，可帮助人员采取行动以及制订合理业务决策。

InfoSphere Warehouse V10 基于 **DB2 10**，包含一组新的高级功能来支持实时运营分析，使组织能够在业务事件发生时主动、及时且明智地制订决策。**InfoSphere Warehouse V10** 提供以下收益：

- 快速且精确的决策制订和转向时间
 - 商业智能，持续向仓库提供数据
 - 针对决策制订者和专业分析人员的商业智能和分析工具
- 提高的成本效率
 - 高级存储技术
 - 帮助支持在线恢复丢失数据的高级恢复解决方案
- 高性能
 - 星型模式优化交付，提供更快的响应时间，交付三倍 **BI** 工作负载性能
 - 高可用性运营访问，同时提供分析
- 提高了团队生产效率
 - 内置时间旅行查询，支持更快的历史和趋势分析查询
 - 行和列访问控制，支持多个租赁运营仓库
 - 基本双时态支持，提高开发者和数据库管理员 (**DBA**) 生产效率
- 访问和分析广泛的信息
 - 呼叫中心注释、电子邮件和博客中的非结构化信息
 - 数据库、电子表格和其他数据源中的结构化信息

解决方案概述

企业必须应对挑战并致力于实现按需应变的洞察力访问。然而，这些挑战在捕获和装入运营数据方面存在瓶颈，使企业无力以及时方式作出反应。而且，处理繁重的工作负载和分析处理的复杂查询所涉及的其他资源和规划也产生性能挑战。

为解决挑战，企业可以将更智慧的客户群作为目标，并与之就个别的需求和期望进行沟通，同时从当前业务场景中挖掘新的市场商机。他们可以发现并投资甚至最小的趋势，获取通常只有通过更灵活且更小的动态业务才能实现的竞争优势。他们可以检测到能够显著改变和影响企业的收入、开支和增长的小行为模式。最重要的是，企业可以围绕着数据推动的洞察力构建竞争策略，最后，生成优异的业务成果。

InfoSphere Warehouse 基于 DB2 for Linux、UNIX 和 Windows 数据服务器。利用其庞大的可伸缩、不共享体系结构，DB2 可提供高性能以进行关系和基本 XML 数据的混合工作负载查询处理。此类高级功能（例如，数据库和表分区、压缩、多维集群 (MDC)、具体化查询表 (MQT) 和 OLAP 功能）使 DB2 成为一个功能强大的运营仓储引擎（图 2）。

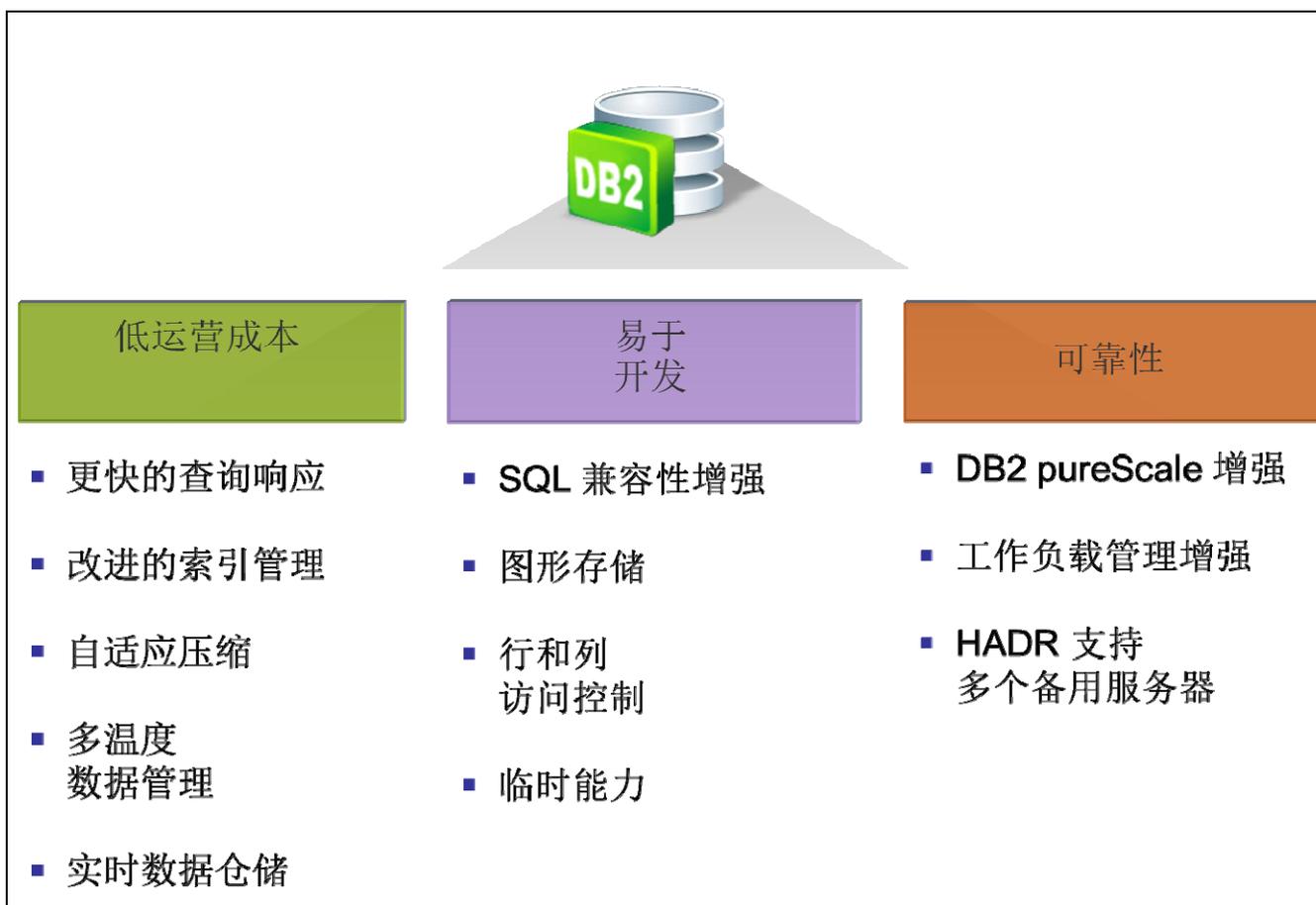


图 2. 以 DB2 为基石构建

InfoSphere Warehouse 提供高级功能用于数据库分区，例如，IT 用户可以具有多种方式针对大规模并行性和线性可伸缩性在服务器上分发数据。DB2 的无共享体系结构帮助确保性能不会随着仓库的增长而下降。而且，因为 InfoSphere Warehouse 可在多个维上物理集群数据、按价值范围排序数据以及将 I/O 限于相关数据分区，因此可帮助减少解决多个查询所需的工作。

InfoSphere Warehouse 在多个分区之间透明拆分数据库，并且使用多个服务器的能力来满足大信息量请求。SQL 语句将自动分解为在每个数据库分区上并行运行的子请求。然后合并子请求的结果以提供最终结果。

IBM InfoSphere Warehouse 包含以下丰富的功能部件和功能：

- **表分区**提供易于滚入和滚出的表数据、灵活的索引配置以及有效的查询处理。表分区通过允许在个别数据分区上执行管理任务，增强表级别管理的灵活性。这些任务包括拆分和重新连接数据分区、备份和恢复个别数据分区，以及重新组织个别索引。可以将耗时的维护操作分解为一系列较小的操作，以达到简化目的。例如，在将数据分区放置在单独的表空间时，备份操作可以逐个处理数据分区。
- 通过使用 **持续数据获取**，您可以透明地将数据从外部源装入 **InfoSphere Warehouse** 数据库而无需停机，而且能够行实时业务分析和决策制订。
- **时间旅行查询**集成到 **DB2 10** 和 **InfoSphere Warehouse** 中，以提供更简单且更快速的基于时间（基于历史趋势）的分析应用程序。增加 **zigzag** 连接帮助大幅减少复杂多维业务查询的时间。增强的查询连接和优化器增强帮助提高其他分析查询的查询性能以及减少更多索引的需求。
- **自适应压缩**还可帮助减少存储成本和提高性能，尤其是大型 I/O 绑定的仓库应用程序和查询工作负载。数据行压缩能够节省存储空间，以及帮助减少磁盘访问时间。同时，将压缩存储的页面，而这将进一步增强磁盘压缩。另外，由于压缩数据，可以在数据库缓冲池中高速缓存更多行，改善了查询响应时间，而且 **DBA** 不再需要频繁执行 **REORG** 操作。
- 新的 **行和列访问控制**提供简单且灵活的规则以及角色定义，用于管理和控制数据访问，帮助增强安全性和简化应用程序开发。这些安全性功能提供一组强大且灵活的规则和访问空间，以管理和帮助数据访问，减少安全性风险。
- **多维集群**提供一种灵活的方法来持续和自动集群多个维中的表数据。此类型的集群减少所需的 I/O 总量。此外，可帮助减少数据库维护活动需求，例如，重新组织。
- **InfoSphere Warehouse 工作负载管理**功能支持实时业务洞察力交付，且不影响性能。对于传统服务器，混合工作负载的强度可能阻碍了向大量用户和应用程序交付信息。而通过 **InfoSphere Warehouse** 提供的高级工作负载管理，**DBA** 可以建立和增强用户的服务级别。他们可以划分不同用户和应用程序的查询优先级，然后控制专用于这些过程的底层资源的数量。
- **InfoSphere Replication Server** 技术包含在 **InfoSphere Warehouse** 的所有修订版中。关注于提供主动/主动可用性的组织可以在源和目标 **DB2 for Linux、UNIX 和 Windows** 数据服务器对之间使用双向 **Q** 复制。

- **嵌入式分析功能**在数据仓库中交付一组成熟且易于使用的工具。这些工具向众多用户提供宝贵的商业智能。**Cubing Services for OLAP** 功能部件支持多维数据分析，而无需从仓库抽取数据。**InfoSphere Warehouse** 包含 **Microsoft PivotTable Service** 的基本支持，支持标准电子表格报告的特别分析或交付，而且同时在 **Microsoft Excel** 应用程序中执行所有工作。此外，**Cubing Services** 立方体是 **IBM Cognos®** 平台一流数据提供商。**Cognos** 客户机和应用程序的整个套件可使用这些功能强大的基于仓库的数据立方体。**InfoSphere Warehouse** 提供嵌入式数据挖掘、建模和计分功能。利用这些功能，业务用户可以实时处理当前数据并交付分析，帮助他们快速发现收入商机。
- 通过使用 **IBM Cognos Business Intelligence**，业务用户可以评估一组丰富的 **BI** 功能而不会产生前期开销。业务用户可以简单访问数据仓库中的数据。利用报告和分析功能的帮助，他们可以根据需要选择合适的方式、时间和场所交付相关信息。通过使用基于 **Web** 的用户界面、企业级别面向服务的体系结构 (**SOA**) 基础以及访问任意数据源的能力，业务用户可以简单开发和部署有关仓库中数据资产的报告。通过组合 **Warehouse Packs (Advanced Editions 中可用)**，**Cognos Business Intelligence** 提供一个快捷方式来部署仓库报告以及从数据中快速获取价值和洞察力。
- **IBM InfoSphere Optim Database Administrator** 帮助组织在无中断情况下管理数据库和数据库更改，简化位置变化和数据库迁移场景。内置的分析和迁移功能可确保迁移所有相关项目，有助于避免应用程序停运。此外，通过确保索引得到更新而支持出众的性能，通过确保迁移特权而促进可用性。**InfoSphere Warehouse** 还包含 **InfoSphere Optim Performance Manager**，后者提供可立即使用的性能监控和管理，帮助提高服务质量以及避免影响业务运营。其直观的基于 **Web** 的用户界面提供可随处使用的监控、警报和潜在性能瓶颈诊断。

InfoSphere Warehouse 提供一组工具，帮助简化数据仓库以及分析开发和部署。利用这些界面，用户可以设计仓库并填充数据结构。他们还可以通过公共界面执行分析和**管理数据挖掘和多维立方体**。

- **Design Studio** 提供一个图形用户界面 (**GUI**)，因此，架构设计师可以设计、建模、反向工程和验证物理数据库模式。**Design Studio** 基于 **IBM InfoSphere Data Architect** 软件，并且可以从各种源导入和导出模型，包括 **CA ERwin**。通过使用 **SQL Warehousing** 工具，**DBA** 可以准备和填充数据挖掘、多维分析和嵌入式分析所需的数据仓库结构。通过使用 **Design Studio** 构建数据流、控制流和变换，并将这些部署在仓库中。
- **IBM InfoSphere Optim™ Development Studio** 软件可帮助提高 **Java** 数据访问的开发效率，以及促进交叉系统开发和迁移。它支持 **DB2**、**Oracle** 数据库和 **IBM Informix®** 软件的开发。**SQL** 大纲功能通过快速隔离所有 **SQL** 以供复审来促进开发者和 **DBA** 协作，并通过关联 **SQL** 和源代码、数据库对象和 **ALTER** 请求支持影响分析。

解决方案体系结构

InfoSphere Warehouse Advanced Edition 将成功、经济高效的数据仓库解决方案所需的所有组件集中在一起。组件从创建抽取、变换和装入 (**ETL**) 操作、**OLAP** 和数据挖掘所需的开发工具一直到用于了解市场的 **BI** 工具。**InfoSphere Warehouse Advanced Edition** 还提供管理后端策略、促进业务一致性以及获取最佳数据仓库以及相连接的应用程序的性能所需的工具。

InfoSphere Warehouse 的中心是 **DB2** 关系数据库引擎，它提供一个用户数据存储库和基础结构来支持在数据上执行的多个运营操作。与 **IBM WebSphere® Application Server** 中托管的 **InfoSphere Warehouse** 应用程序结合在一起，这些元素组合构成 **InfoSphere Warehouse** 解决方案的运行时组件。多个客户机产品、**Data Studio**、**Design Studio** 和 **Web** 浏览器提供了支持这些运行时元素所需的开发工具和管理组件。图 3 显示 **InfoSphere Warehouse** 的完整功能组件体系结构。

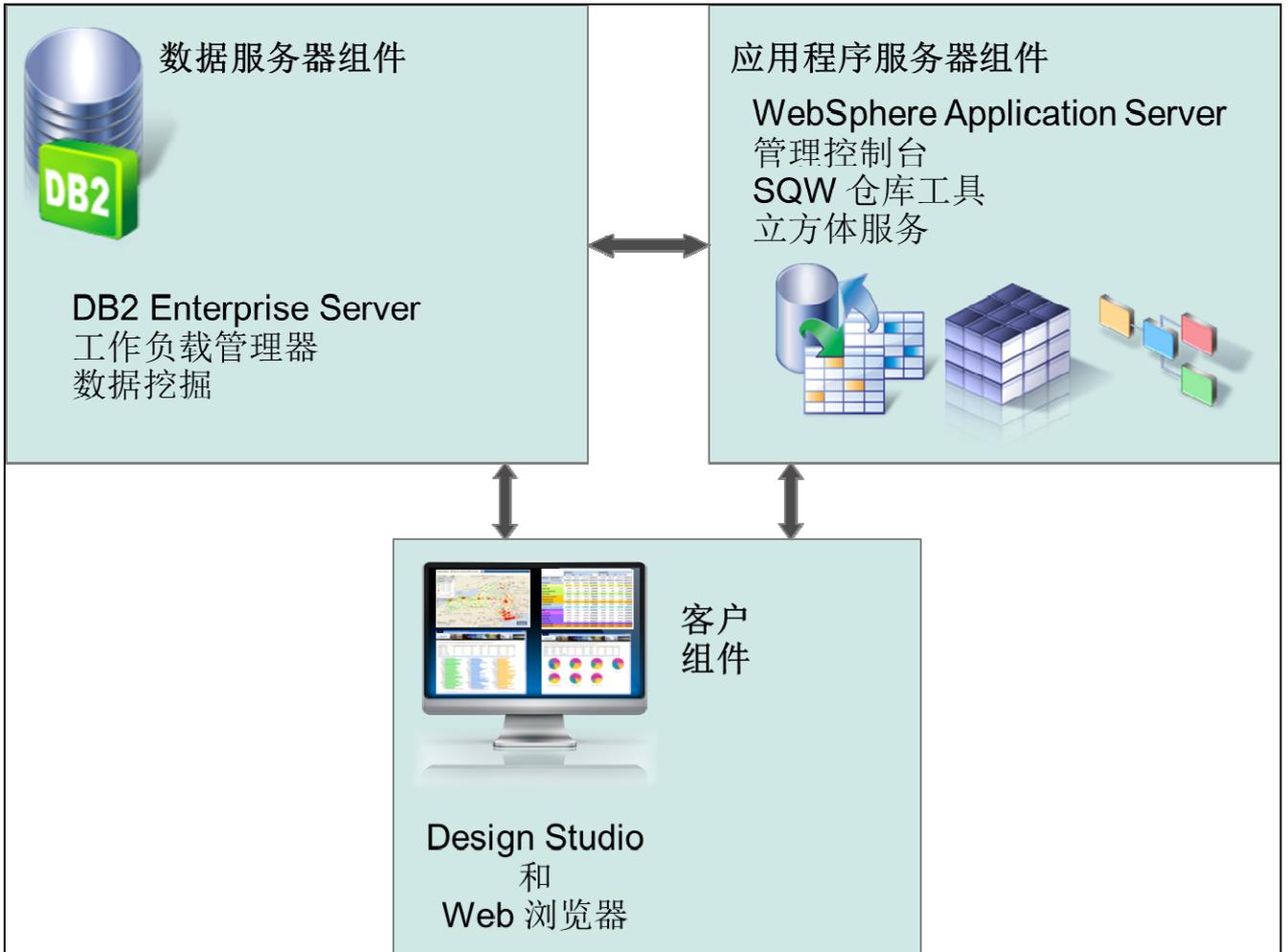


图 3. InfoSphere Warehouse 的功能组件体系结构

数据仓库核心的 DB2 关系数据库的实例可配置为单个或多个分区数据库，这些数据库安装在一个硬件服务器上或多个硬件服务器上。这一 DB2 灵活性大大提高了仓库数据的主存储库性能，这一存储库通常被称为*执行数据库*。

除了主数据存储库和执行数据库，相同的 DB2 实例还托管另外两个数据库。这些更小的数据库包含 InfoSphere Warehouse runtime 和 Cognos BI 服务器应用程序所需的元数据。

InfoSphere 数据仓库应用程序服务器组件由执行多个重要功能的企业 Java 应用程序组成,其中包括完整解决方案中的以下功能:

- InfoSphere Warehouse 解决方案管理控制台
- 存储、运行和管理 ETL 过程的能力
- 存储和管理立方体服务
- 存储和管理数据挖掘服务

InfoSphere Warehouse 管理控制台具有基于 Web 的界面,允许任何浏览器用于运行时环境功能元素的配置和管理。因此,可使用单个浏览器来处理所有生产、测试和开发环境。

SQL Warehousing 应用程序针对执行数据库中的数据执行 ETL 操作。这些操作包括使用 Design Studio 创建的控制流和数据流,这是 SQL Warehousing 开发工具。通过 Design Studio,可以针对实际数据库测试和调试 SQL Warehousing 数据和控制流,然后分组到仓库 SQL Warehousing 应用程序。然后,通过管理控制台将这些 SQL Warehousing 应用程序部署到 SQL Warehousing 运行时元素中。

OLAP 立方体服务服务器(简称为立方体服务器)是立方体进程的运行时元素。此立方体服务器是独立的 Java 进程,其托管各种立方体、接收入局连接和查询请求、处理请求、构造结果集,并将它们返回至调用应用程序。此 Java 进程独立于安装的 WebSphere Application Server 运行,但是需要位于相同的物理服务器上,因此可通过管理控制台应用程序进行管理。

通过使用 Design Studio 开发工具可在立方体服务器中实施立方体。在成功实施立方体模型后,可将其部署至 InfoSphere Warehouse 服务器。

使用方案

构成核心 InfoSphere Warehouse 实施的组件可分为四个安装类别:

- “数据服务器”组件涵盖主 DB2 平台,在 IBM AIX®、HP-UX、Solaris、各种 Linux 实施和 Windows 上受支持。
- “应用程序服务器”组件涵盖 IBM WebSphere Application Server,这是仓库产品组的一部分。
- “客户机”组件涵盖所有命令行和基于 GUI 的平台,通常可安装在用户的个人计算机或笔记本中。
- “文档”组件涵盖产品组文档的在线和 PDF 版本。

涵盖这些类别的 InfoSphere Warehouse 组件可安装在一系列拓扑中的硬件平台上。InfoSphere Warehouse 体系结构具有以下常见拓扑（图 4）：

- 单层体系结构通常用于开发、测试和培训环境。所有主要组件（包括客户机）都安装在单个硬件平台上。
- 双层体系结构也主要用于开发和测试环境。但是，通过配备适当的服务器和存储器，此拓扑可用于较小的仓库实施。
- 在三层体系结构中，客户机组件、DB2 组件和 WebSphere Application Server 组件安装在单一的硬件系统中。您可以在生产系统上使用此拓扑。

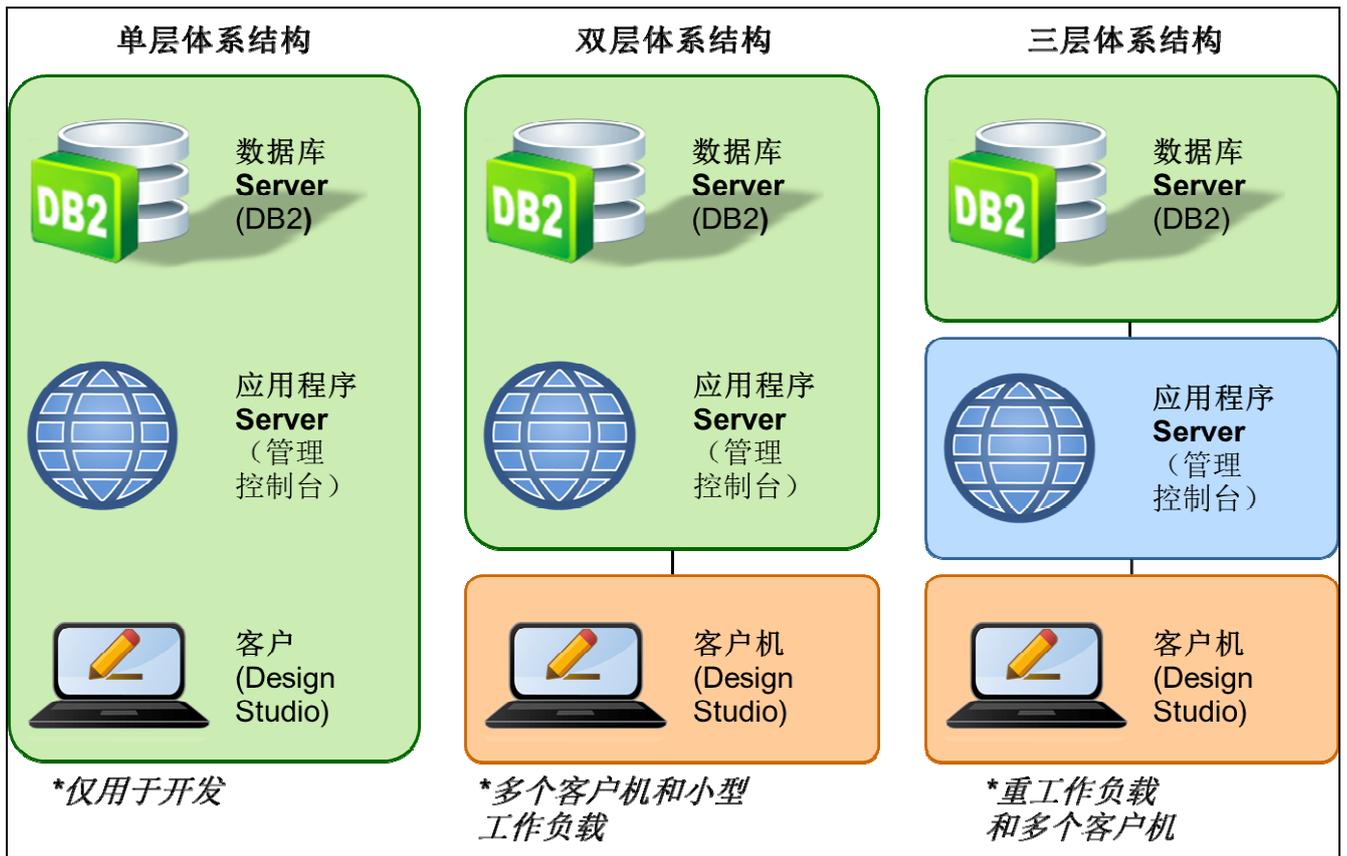


图 4. InfoSphere Warehouse 体系结构的三个常见拓扑

图 5 阐述多个物理硬件平台上的三层实施，其中数据库服务器具有一个管理数据库节点和多个数据节点。

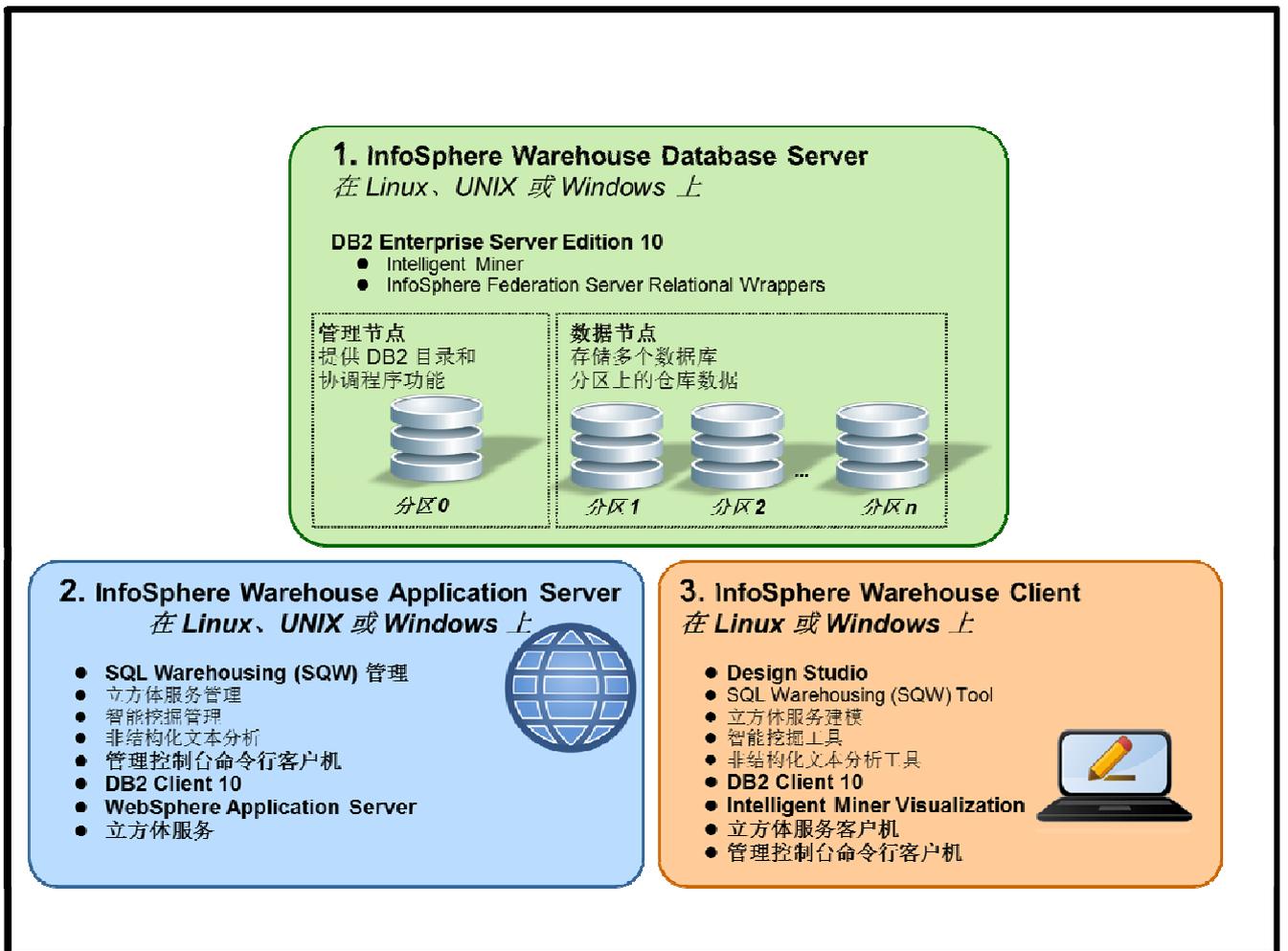


图 5. 多层物理服务器上的三层解决方案

受支持的平台

InfoSphere Warehouse 是一个产品套件，将 DB2 Enterprise Edition 的强度与 IBM 的数据仓储基础结构相组合。InfoSphere Warehouse 具有一个基于组件的体系结构，包括数据服务器组件组、应用程序服务器组件组和客户机组件组。在典型生产环境中，在不同的计算机上安装其中每个组件组以创建全面的仓储解决方案。

InfoSphere Warehouse 的系统需求会不定期进行更新。要获取最新信息，请参阅 InfoSphere Warehouse 产品页面：

<http://www.ibm.com/software/data/infosphere/warehouse/sysreqs.html>

订购信息

InfoSphere Warehouse 产品范围从企业级数据仓库发行版一直到专业版本：

- InfoSphere Warehouse V10.1 Advanced Enterprise Edition
- InfoSphere Warehouse V10.1 Enterprise Edition
- InfoSphere Warehouse V10.1 Advanced Departmental Edition
- InfoSphere Warehouse V10.1 Departmental Edition
- InfoSphere Warehouse V10.1 Developer Edition

要了解订购信息，请联系您的 IBM 代表或 IBM 业务合作伙伴。另请参阅 IBM InfoSphere Warehouse V10.1 销售手册：

<http://ibm.co/XkAEgz>

相关信息

要了解更多信息，请参阅以下文档：

- *Solving Operational Business Intelligence with InfoSphere Warehouse Advanced Edition*, SG24-8031
<http://www.redbooks.ibm.com/abstracts/sg248031.html>
- *InfoSphere Warehouse: A Robust Infrastructure for Business Intelligence*, SG24-7813
<http://www.redbooks.ibm.com/abstracts/sg247813.html>
- IBM InfoSphere Warehouse 信息中心：
<http://bit.ly/SC2IWU>
- Workload Management (WLM) Tutorial
<http://ibm.co/RimG9Z>
- Best Practices Workload Management
<http://www.ibm.com/developerworks/data/bestpractices/workloadmanagement>
- IBM InfoSphere Warehouse V10.1 销售手册：
<http://ibm.co/XkAEgz>

声明

本信息是为在美国提供的产品和服务编写的。

IBM 可能在其他国家或地区不提供本文中讨论的产品、服务或功能特性。有关您当前所在区域的产品和服务的信息，请向您当地的 IBM 代表咨询。任何对 IBM 产品、程序或服务的引用并非意在明示或暗示只能使用 IBM 的产品、程序或服务。只要不侵犯 IBM 的知识产权，任何同等功能的产品、程序或服务，都可以代替 IBM 产品、程序或服务。但是，评估和验证任何非 IBM 产品、程序或服务，则由用户自行负责。IBM 公司可能已拥有或正在申请与本文档内容有关的各项专利。提供本文档并未授予用户使用这些专利的任何许可。您可以用书面方式将许可查询寄往：

IBM Director of Licensing, IBM Corporation, North Castle Drive, Armonk, NY 10504-1785 U.S.A.

本条款不适用英国或任何这样的条款与当地法律不一致的国家或地区：INTERNATIONAL BUSINESS MACHINES

CORPORATION“按现状”提供本出版物，不附有任何种类的（无论是明示的还是暗含的）保证，包括但不限于暗含的有关非侵权、适销和适用于某种特定用途的保证。某些国家或地区在某些交易中不允许免除明示或暗含的保证。因此本条款可能不适用于您。本信息中可能包含技术方面不够准确的地方或印刷错误。此处的信息将定期更改；这些更改将编入本资料的新版本中。IBM 可以随时对本资料中描述的产品和/或程序进行改进和/或更改，而不另行通知。

本信息中对非 IBM Web 站点的任何引用都只是为了方便起见才提供的，不以任何方式充当对那些 Web 站点的保证。那些 Web 站点中的资料不是 IBM 产品资料的一部分，使用那些 Web 站点带来的风险将由您自行承担。IBM 可以按它认为适当的任何方式使用或分发您所提供的任何信息而无须对您承担任何责任。涉及非 IBM 产品的信息可从这些产品的供应商、其出版说明或其他可公开获得的资料中获取。IBM 没有对这些产品进行测试，也无法确认其性能的精确性、兼容性或任何其他关于非 IBM 产品的声明。有关非 IBM 产品性能的问题应当向这些产品的供应商提出。本信息包含在日常业务操作中使用的数据和报告的示例。为了尽可能完整地说明这些示例，示例中可能会包括个人、公司、品牌和产品的名称。所有这些名字都是虚构的，若现实生活中实际业务企业使用的名字和地址与此相似，纯属巧合。

此处包含的任何性能数据都是在受控环境中测得的。因此，在其他操作环境中获得的数据可能会有明显的不同。有些测量可能是在开发级的系统上进行的，因此不保证与一般可用系统上进行的测量结果相同。此外，有些测量是通过推算而估计的，实际结果可能有所不同。本文档的用户应当验证其特定环境的适用数据。

版权许可：

本信息包括源语言形式的样本应用程序，这些样本说明不同操作平台上的编程方法。如果是为按照在编写样本程序的操作平台上的应用程序编程接口（API）进行应用程序的开发、使用、经销或分发为目的，您可以任何形式对这些样本程序进行复制、修改、分发，而无须向 IBM 付费。这些示例并未在所有条件下作全面测试。因此，IBM 不能担保或暗示这些程序的可靠性、可维护性或功能。

© Copyright International Business Machines Corporation 2012. All rights reserved.
Note to U.S. Government Users Restricted Rights - Use, duplication or disclosure restricted by
GSA ADP Schedule Contract with IBM Corp.

本文于 2012 年 11 月 8 日创建或更新。

请通过以下任一方式将您的意见发送给我们：

- 使用位于以下地址的在线**联系我们**审阅表单：
ibm.com/redbooks
- 通过电子邮件将您的意见发送至：
redbook@us.ibm.com
- 将您的意见邮寄至：
IBM Corporation, International Technical Support Organization
Dept. HYTD Mail Station P099
2455 South Road
Poughkeepsie, NY 12601-5400 U.S.A.

本文档可通过以下地址在线访问：ibm.com/redbooks/abstracts/tips0932.html

商标

IBM、IBM 徽标和 ibm.com 是 International Business Machines Corporation 在美国和/或其他国家或地区的商标或注册商标。这些术语和其他 IBM 已注册商标的术语在本信息中首次出现时都使用适当的符号(® 或 ™) 标记，以表示在本信息发布时由 IBM 在美国注册或拥有的普通法商标。这些商标也可能是其他国家或地区的注册商标或普通法商标。在以下 Web 站点上提供 IBM 商标的最新列表：ibm.com/legal/copytrade.shtml

以下术语是 International Business Machines Corporation 在美国和/或其他国家或地区的商标。

AIX®
Cognos®
DB2®
IBM®
Informix®
InfoSphere®
Optim™
Redbooks (徽标)®
WebSphere®

以下术语是其他公司的商标：

Linux 是 Linus Torvalds 在美国和/或其他国家或地区的商标。

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

Java 和所有基于 Java 的商标和徽标是 Oracle 和/或其子公司的商标或注册商标。

UNIX 是 The Open Group 在美国和其他国家或地区的注册商标。

其他公司、产品或服务名称可能是其他公司的商标或服务标记。