



DB2 Information Management Software



IBM DB2 Data Warehouse Edition

软件功能说明与特点

ON 按需应变的业务

目录

Content

DB2 DWE 软件功能说明及特点	2
数据仓库海量关系型数据库 DB2 ESE 与数据分区模块 DPF	3
高扩展性	3
高性能	3
高可靠性	3
易管理性	3
DB2 调度复杂即席查询的查询管理器 DB2 Query Patroller	8
DB2 多维分析工具 DB2 Cube Views	9
WebSphere Information Integrator	12
联邦数据服务器 (Federated Data Server)	12
针对异构关系型数据库的复制服务器 (Replication Server)	13
数据仓库管理工具 DB2 Warehouse Manager	14
数据挖掘 DB2 Intelligent Miner	15
DB2 Intelligent Miner Scoring	16
DB2 Intelligent Miner Modeling	17
DB2 Intelligent Miner Visualization	17
前端展示应用开发平台 DB2 AlphaBlox	18

DB2 DWE 软件功能说明及特点

DB2 Data Warehouse Edition (DB2 数据仓库版)

一个强大的商业智能平台——扩展了 IBM DB2 UDB 用于构建数据仓库和分析应用的商业智能特性。包括联邦数据访问，数据分区，在线分析处理，数据挖掘，ETL，工作负载管理等功能。整合了 IBM DB2 UDB 的商业智能特性，是面向客户和合作伙伴，构建企业按需应变架构，实施实时商业智能的新一代商业智能解决方案。

主要产品包括如下：

产品:	功能模块
1.数据仓库存储的海量关系型数据库 DB2 UDB ESE	DB2 UDB ESE 企业版,数据仓库引擎
2.数据分区模块 DB2 UDB DPF	DB2 UDB DPF 模块,提供分区支持, 支持多机并行
3.调度复杂即席查询的查询管理器 Query Patroller	基于成本的查询负载管理工具, 提高数据仓库吞吐量
4.多维分析工具 DB2 Cube Views	1.OLAP 元数据交换工具 2.模型级物化查询表生成器, 加速 OLAP 多维分析
5.实时数据仓库及信息集成器 IBM Websphere INFORMATION INTEGRATOR	1.关系数据库联接 Relation Connect 2.非关系数据库联接 Life Sencie Connect
6.ETL 工具 DB2 Warehouse Manager	1.用于数据抽取、清洗、加工和装载的 ETL 工具 DB2 Warehouse Manager Transformer; 2 用于分布数据仓库 ETL 过程的数据仓库代理 Agent。3. 数据仓库的开发、监控以及 Meta Data 管理工具 DB2 Warehouse Center; 4 数据仓库信息目录 Information Catalog; 5.报表生成工具 QMF
7.DB2 Intelligent Miner Scoring	数据挖掘评分工具 在数据仓库中直接实现(实时)评分
8.DB2 Intelligent Miner Modeler	数据挖掘建模工具,从数据仓库中发现规律(模型)
9.DB2 Intelligent Miner Visualization	数据挖掘模型图示化工具
10.IBM Office Connect Analytical/Enterprise Web Edition	Excel 多维分析插件, 在 Excel 中实现查询、报表、多维分析
11. DB2 AlphaBlox	提供开发分析应用的平台, 构建客户自己的分析应用

数据仓库海量关系型数据库

DB2 ESE 与数据分区模块 DPF



高扩展性

在单 UNIX 主机 CPU 扩展方面，TPC-C (OLTP) 和 TPC-H (Ad-Hoc Query)，显示了 DB2 在扩展性方面的能力。在 SMP 环境中，经 SUN 在 64 CPU E1000 上的测量，可达到 90% 的扩展性 (详细信息参见 www.tpc.org)。

在集群技术方面，采用 Share Nothing 的 MPP 体系结构，每个节点独享各自的硬盘空间，各个节点间通过网络交换数据。可伸缩性强，最多可以扩充到 1000 个节点。国外有 512 节点的实例。每个数据库中单个表的大小可以达到 512GB*1000=500TB。在全球范围有许多 TB 级的数据仓库。美国电信公司 Sprint PCS 的实时数据仓库由 55 个 DB2 节点服务器组成，每个节点 4CPU，运行 50TB 实时数据仓库。

高性能

DB2 的基于成本优化技术已经有 27 年历史，积累了大量独有专利技术。针对任意的查询，DB2 独特的查询重写功能将所有 SQL 语句改写成为语义上完全相同，但是可以使用优化器所有优化功能的语句，优化器再自动选择最优的查询路径完成查询。此功能特别适用于以图形化界面生成的数据仓库应用。针对 OLAP 应用做了相当多的优化。强大的基于成本的优化，独特的星型连接算法、动态位图索引、OLAP 算子等。

高可靠性

支持 Cluster、Standby 等双机热备份、联机快速备份、快速加载数据和快速备份数据恢复。提供双日志功能，具有表空间级的备份和恢复功能，可以从整个数据库备份集中选择需要的单元进行恢复。提供在线重组织表，在线重组织索引，在线配置主要参数的能力，在线创建、删除和修改缓冲池的能力。

易管理性

IBM DB2 是一个 SMART 的数据库，即自我管理及资源调度的智能数据库系统。DB2 由数据库系统自己进行错误分析、可能的解决方案建议及自我修复的功能，提供部分核心配置参数进行自动配置，而不需数据库管理员进行指定和修改。

技术特点:

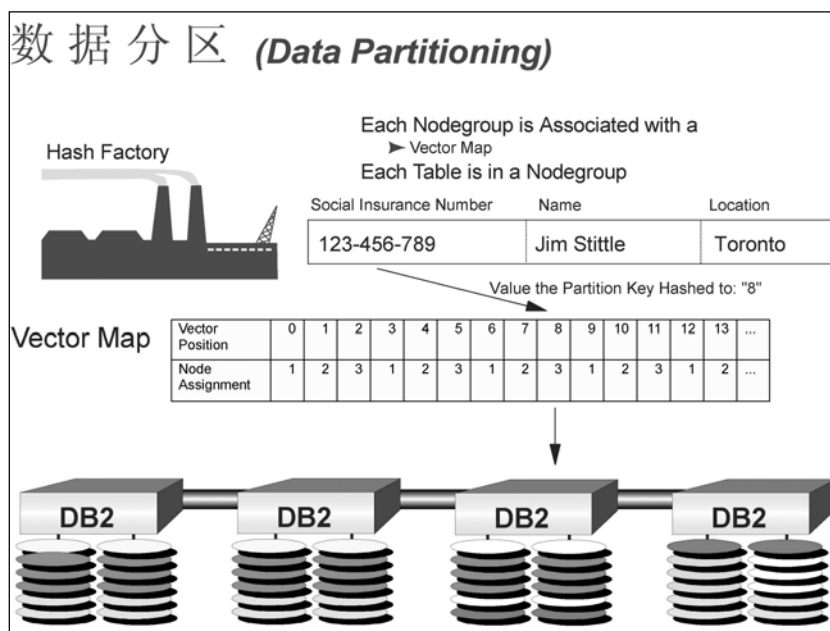
数据分区技术:

- DB2 Data Partitioning Option (数据分区技术):

DB2 UDB 提供了先进的“哈希 (HASH) 算法”映射数据库的每一条记录到特定的数据库分区中。“哈希算法”使用表中的一列(或一组列)作为分区关键字，得到 0 至 4095 的数值。分区图定义了为 4096 个值中的每一个值分配的特定的数据库分区。

DB2 UDB为数据存储提供了灵活的拓扑结构以达到高性能及高并行。其中每个数据库由一些数据库分区组成，每个数据库分区实际上是数据库的一个子集，它包含自己的用户数据，索引，交易日志及配置文件。在数据库中，管理员需要定义节点组(Node Group)——数据库分区所分布的节点集合。节点组

能够跨越为该数据库设置的数据库分区的一部分或全部。在节点组中，还要定义表空间，以说明用来存储表数据及索引的容器(Container)(文件或设备)。在数据库分区中，如果为每个表空间定义多个容器，则数据库管理系统可以利用 I/O 的并行机制提高性能。



图一.数据分区分布图

DB2 UDB数据库分区的体系结构具有很多优势：一张数据库表被分布在多个数据库分区上，因此一张大规模数据库表可以大到 TB 级。

DB2 UDB 在数据定义语言(DDL)，数据操作 SQL，以及运行时都引用了分区的模式。其分区方法还可以看做为装载平衡的工具(通过修改分区关键字及分区图，各分区中的记录数可以调整)。DB2 UDB

优化器利用分区知识来估价不同操作的耗费，从而为每个 SQL 语句选择最优的执行策略。

数据的分布通过对分区关键字进行哈希算法完成，分区图中提供了每条记录的存放位置。

如果在初次分布数据之后，出现了数据存放不均的现象，DB2 UDB 能够自动分析并更正。DB2 UDB 可以通过修改分区的分布自动创建一个新的分区图

来平均分布当前不均的数据。其中涉及到的数据记录自动移到它新被分到的数据分区。

对于不断增长的数据库,我们可以增加分区(同时增加处理能力),修改分区图来包含这些新的数据库分区,而后系统能够自动的重新分布数据,以达到新的平衡。DB2 UDB提供了这一功能,使得系统具有非常好的扩展性。

处理能力较强的数据库分区可以存放较多的数据,从而在一切不共享的配置下可以充分利用各节点的处理能力使其负载均衡。DB2 UDB可以用来按比例将更多的数据分布在具有更强处理能力的数据库分区上。

应用可以调用API找到记录的存放位置,然后将交易送到记录所在的节点。该API也可以直接被交易处理应用来调用,如IBM CICS, Encina,将交易送到适当的节点而提高性能。

- **Multi-Dimension Cluster (多维群集技术):**

为了提高对一张大表的访问效率,IBM DB2提供一种新的索引技术—MDC (多维群集) 技术。

多维群集 (MDC) 提供了一个出色的方法,可灵活、连续、自动群集多维数据。这将会极大地提高查询的性能,而且大幅度降低了数据维护操作的费用,例如重组织,以及插入、更新和删除操作过程中的索引维护操作。多维群集主要用于在线事务处理 (OLTP) 环境、数据仓库和大型数据库环境中。

首先,MDC采用“BLOCK”来进行索引的组织,一个BLOCK会包含很多条传统索引机制所采用的“行”记录,因而大大的提高的索引的粒度。使得索引的定位变得更快。

利用MDC,可以使一个表在物理上同时群集在多个主键 (或维度) 上。利用群集索引,在表中插入和更新记录时,DB2以索引中主键的顺序在物理数据存储页面中维护着数据的实际顺序。群集索引大量提高了大范围查询的性能,在良好的群集的帮助下,由于当页面存储连续,完成预读取的性能将非常高,从而提高数据查询的效率。

这些查询不仅只访问包含有正确的维度数值的记录的页面,并且这些符合条件的页面将会根据范围进行分组。而且,尽管具有一个群集索引的数据表可以随着表空间的填充,解除群集,但MDC表能够自动连续维护所有维度上的群集,从而不必为了恢复数据的物理顺序而重新组织表。

在一张大表上定义一个MDC后,原有的索引都仍然有效,换句话说,我们可以在一个已经存在的大数据表上,随时增加MDC,以提高对它的访问效率。

- **UNION ALL View:**

IBM DB2 UDB支持在多个小型数据库表上增加一个UNION ALL VIEW,从而建立一个逻辑上的大表。如果由于硬件等原因,使得对一张大数据表的存储处理变得困难时,我们可以支持将数据分布在多个较小型的数据表中,然后使用UNION ALL VIEW技术来实现一个逻辑大表的组织和访问。透过UNION ALL View,用户可以透明地对View中的多个较小规模的表实现UPDATE、DELETE、INSERT、SELECT操作。

并行技术

DB2 UDB无论在SMP还是在MPP环境下,甚至在SMP节点组成的MPP环境下,都可以通过完善

的协同处理和事务控制技术保证处理的并行、完整和一致性，充分发挥其并行处理能力。

查询执行时被透明地分开后并行执行(称作节点间查询并行性: Intra-query parallelism)，过去需要数小时的查询现在只需几分钟就可以完成。过去不能执行的查询现在不仅可行，而且还能从中获益。

在单台SMP环境下处理并行(称作节点间查询并行性: Intra-query parallelism)，DB2 UDB采用的是吸管模型(Straw model)。此时被执行的SQL相当于杯子中的水，而每个CPU相当于一根吸水的吸管，这样被执行的SQL很快就被CPU“吸干”了，同时可以使用到多I/O的并行进行数据的存取操作，此种并行模式称作节点内并行 (Intra-Partition parallelism)。在MPP环境下或者多SMP组成的集群环境下处理并行，则相当于把杯子中的水先智能地分配给多个小水杯(参与SQL执行的各个节点机)，这样小水杯中的水就分别被每个节点机“吸干”了，此种并行模式称作节点间并行 (Inter-Partition parallelism)。有了好的模型仅仅是开了个好头，DB2 UDB中融入的最先进的技术才是成功的关键。

- 并行优化: 从很多系统中抽取非常多的数据，这会耗费大量的时间，如果效率不高，还会浪费大量宝贵的处理能力。从串行数据库中抽取数据有很多方法，它们没有必要象并行数据库那样运作。DB2 UDB有一个查询优化器，是由IBM研究机构开发的，它是专为提高并行抽取数据的效率而设计的。这样就可获得高品质的查询性能，特别是对特大型数据库。

- 全面并行(Parallel Everything): DB2首先把数据分配到数据库中的多个分区或子集中，这些数据库位于多个MPP节点机或SMP服务器内。接着，DB2自动创建一个并行处理访问计划。数据扫描、合并、分类、负载平衡、表格重组、数据调用、创建索引、索引访问、备份与恢复等一系列工作都是在所有不同的节点里同时完成的。DB2 UDB以并行方式执行全部数据库功能，这包括全部SQL语句(Select、Insert、Update和Delete)、实用程序(backup, restore, reorg, load)和数据存取方法(连接、表扫描和索引扫描)等，而且无需任何额外的编程。这不仅提供了更好的性能和可伸缩性，而且也提供了更佳的管理性——有能力利用全部处理机去执行数据库管理任务。进一步说，DB2 UDB既可用于联机事务处理(OLTP)，又可用于决策支持查询工作。
- 管理工具在并行环境同样适用: Governor帮助您控制每个用户及应用程序的资源利用率，可自动调整查询的优先级。从而，在线平衡负载，简化系统管理，减少关机时间。通过并行在线备份功能可显著减少备份及恢复所需的时间。
- 并行环境下功能不受任何限制: DB2 UDB对多媒体数据的支持、支持的客户端平台、支持的应用开发接口和开发工具以及动态位图索引等多维分析功能、对WEB和Java的支持均不受限制。
- 多用户并发控制: 并发控制是通过行级封锁、查询结果游标处理以及层次隔离等手段进行维护的，层次隔离包括游标稳定性、读稳定性、可重复读和未提交读等方式。

- **DB2 UDB的并行查询处理:** DB2 UDB中的并行查询处理是通过并行执行任务实现的,每个任务仅在自身分区的数据上工作。例如,在6个单处理器的集群环境中,表的扫描被分成在6个单独数据库分区(Partition)上执行的6个单独的表扫描。在一个4 CPU的集群中,一般情形下会有4个数据库分区,这样,表的扫描在所有服务器上并发执行。SMP节点内的查询则被分成单独的执行任务利用SMP的多处理器并行执行。结果被汇总后传递给用户,在这样的环境中,用户无需了解有关数据库分区及并行处理的情形。从最终用户的角度来看,尽管DB2 UDB的性能大大的提高了,但它与传统数据库(非集群)的使用

方法是一样的。缺省情况下, DB2 UDB使用TCP/IP在集群系统的分区间进行通讯,信息及数据通过高效缓存机制进行交换,大大减少了通讯的需要。

- **并行备份和恢复:** 能够用并行方式把数据库或表空间备份到多台设备上或者反过来从多台设备上恢复。用并行方式执行备份/恢复大大地减少所花费的时间,从而满足关键性维护任务的需求。备份和恢复实用程序能以并行方式使用若干设备,这样就减少了这些实用程序所耗费的运行时间。



DB2 调度复杂即席查询的查询管理器

DB2 Query Patroller



DB2 Query Patroller 是一种强大的查询管理系统，数据库管理员可以使用该系统按照以下方式对 DB2 数据仓库前瞻性地动态控制查询流程，针对不同大小的查询定义单独查询类别，以便更好地在各查询间共享系统资源，以及防止较小的查询被困在较大查询之后。

- 将特定用户提交的查询排列为高优先级，以便使这些查询更快速地运行
- 自动暂停大型查询，以便能够取消它们或安排它们在非峰值时间运行
- 跟踪并取消失控查询

Query Patroller 可使数据库管理员调整数据库查询工作负载，以便使小型查询和高优先级查询能够立即运行，并使数据仓库的系统资源得到有效利用。此外，还可收集和分析已完成的查询信息，以便确定查询、高使用量用户、频繁使用的表和索引间的趋势。

管理员可使用 Query Patroller 来：

- 建立系统级和用户级资源使用策略
- 通过取消和重新安排可能影响数据库性能的查询，监控和关联系统使用情况
- 生成有助于确定数据库使用趋势的报告，例如正在访问哪些对象，哪些个人或用户组产生的工作负载最大

查询提交者可使用 Query Patroller 来：

- 监控他们已提交的查询
- 保存查询结果，以备未来进行检索和重用，从而有效避免了重复性的查询提交
- 设置各种参数，以定制查询提交，例如是否在查询完成时接收电子邮件通知

DB2 Query Patroller 可单独提供，也可包含在 DB2 Data Warehouse Enterprise Edition 中。

DB2 多维分析工具 DB2 Cube Views

DB2 Cube Views 是 DB2 UDB 中的最新一代 OLAP 支持，它能够使关系数据库成为管理及部署企业多维分析应用的 OLAP 服务器平台。凭借 DB2 Cube Views，数据仓库设计人员能够提供可更快速部署、更容易管理并能够在各种分析应用程序 (无论使用哪些特定的 OLAP 工具和技术) 间提高性能的 OLAP 解决方案。

DB2 Cube Views 功能部件是如何做到这一点呢？首先，它允许 DB2 支持 OLAP。如果没有多维数据视图，关系数据库对于不熟悉数据库结构和内容的人来说似乎是一组令人畏的带有一些列的表。DB2 Cube Views 元数据使您能够更完整地记录 DB2 数据库中的底层结构。

通常，仓库或数据集市是特别为维分析设计的。在这种情况下，数据库通常由被组织为一个或多个星型 (或雪花型) 模式的表组成。简单的星型模式包含一个事实表，事实表周围有一些维表。例如，事实表可以记录公司各个产品和分店每日销售数据。维表 (或许是 PRODUCTS 和 STORES 表) 可能包含有关各个产品和分店的详细信息，并且与事实表相连接。

元数据对象

DB2 Cube Views 元数据尤其擅长捕获星型或雪花型模式中固有的结构 (请参阅图 1)。该元数据通过一组元数据对象提供关系数据的“空间视图”。元数据对象被划分为多层，从简单对象 (如 Attribute，简单的 Attribute 对表列建模) 到更复杂的对象 (如 Join、Hierarchy、Dimension 和 Cube Model)。完整定义的多维数据模型对象通常符合星型 (或雪花型) 模式，并对一组具有公共维度的度量建模。然后，多维数据模型就封装其它 DB2 Cube Views 元数据对象 (如 Dimension 和 Measure)，这一点都不奇怪。

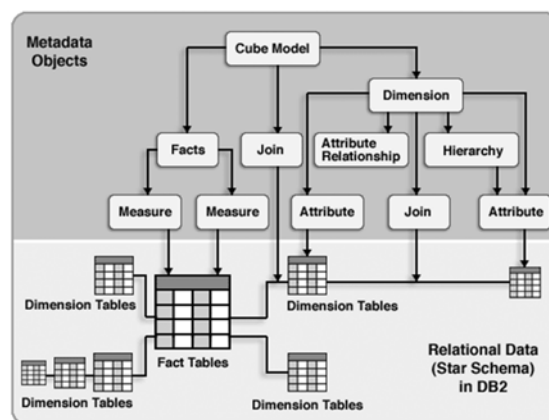


图 1. 引用关系星型模式的 DB2 Cube Views 多维数据模型

简而言之，DB2 Cube Views 元数据使您能够捕获 DB2 数据库的多维结构和设计。

DB2 Cube Views 元数据除了使 DB2 支持 OLAP 外，至少还有两种使用方法：

- DB2 Cube Views 功能部件利用元数据来优化命中 DB2 数据库的 SQL 查询。
- 位于 DB2 之上的产品和应用程序可以使用元数据。

这两种用法的简要介绍：

将 DB2 Cube Views 元数据用于优化

DB2 Cube Views 通过创建合适的实例化的查询表

(materialized query table, MQT) 来加速 SQL 查询，这些 MQT 中保存了预先聚集的数据。MQT 也称为 (自动) 汇总表。如果可以，DB2 优化器会将 SQL 查询重新路由给 MQT (请参阅图 2)。MQT 通常比底层的基本表 (MQT 构建于这些基本表之上) 小好几个数量级。因而，对于同一个 SQL 查询，将该查询重新路由给 MQT 这种查询方法通常比用查询基本表的方法快得多。

DB2 Cube Views 有一个基于元数据和用户输入的优化顾问程序，它推荐一组合适的 MQT。

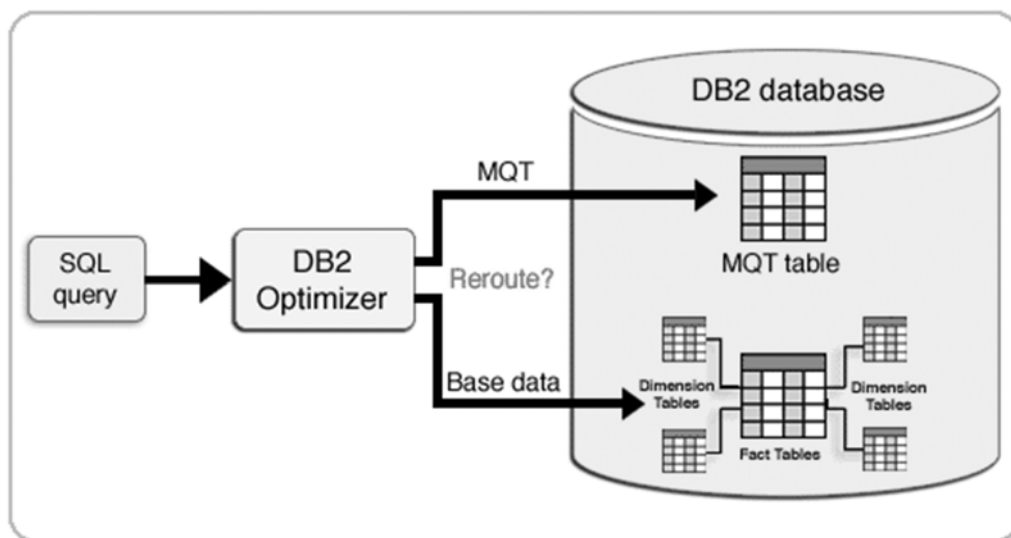


图 2. DB2 优化器在适当时将查询路由到 MQT

元数据使用者

多维数据视图元数据还有其它好处。商业智能工具可以从 DB2 读取元数据来了解 DB2 表的多维结构。然后，根据元数据，这些工具提供处理 DB2 数据的查询和报告解决方案。

使用 DB2 Cube Views 元数据的工具和应用程序一般分为两类：

- 第一类应用程序“咬住”DB2 Cube Views。它们通常通过使用元数据来构造 SQL 查询，以读取元数据并直接用它对 DB2 数据库进行分析。大

多数情况下，这些工具对 DB2 Cube Views 多维数据模型和多维数据对象进行查询和报告。

- 第二类工具使多维数据视图元数据从 DB2 流入自己的工具，然后根据自己的元数据查询 DB2。这些工具必须能够将 DB2 Cube Views 元数据映射到自己的元数据。这一映射的复杂程度取决于元数据的相似程度 (或相异程度)。

元数据生产者

并非所有对 DB2 Cube Views 元数据感兴趣的应用程序都读取元数据，这一点也很重要。还有一类工具和应用程序产生元数据并将它推入 DB2。维元数据通常来自数据库设计、抽取-转换-装入 (extract-transform-load, ETL) 和其它数据库管理工具。通过使维元数据流入 DB2, 这些工具可以利用先前提到的这些优化和管理好处。显然，为了读取、创建、修改和删除元数据对象，工具和应用程序需要一种与 DB2 中多维数据视图元数据交互的工具，这个工具就是 DB2 Cube Views。

DB2 Cube Views 组件

- 多维元数据对象

用户可以创建元数据对象集，以便在维度上对关系数据和 OLAP 结构进行建模。DB2 Cube Views 保存了用户能够在 DB2 目录中创建的所有元数据对象。

- OLAP Center

凭借 OLAP Center，用户可以创建、处理、导入或导出多维数据集模型、多维数据集，以及可用于

OLAP 工具的其它元数据对象。OLAP Center 提供了易用的向导和窗口来帮助用户使用元数据。例如，Optimization Advisor 可分析元数据，并建议如何构建对 OLAP 式 SQL 查询的聚合数据加以保存和编写索引的汇总表。

- OLAP Center 管理界面
- DB2 Office Connect Analytic Edition

DB2 Office Connect Analytic Edition 是一种可查询 DB2 中 OLAP 数据的易用型电子表格外接工具。凭借 DB2 Office Connect Analytic Edition, 用户可以连接到 DB2 数据库，选择 DB2 Cube Views 多维数据集，并探索 Microsoft Excel 中的数据。

- 多维服务

DB2 Cube Views 提供了面向 OLAP 工具和应用程序开发者且基于 SQL 和 XML 的应用程序编程界面 (API)。通过 CLI、ODBC 或 JDBC 连接，或者通过将嵌入式 SQI 应用到 DB2，应用程序和工具可使用单个存储过程来创建、修改及检索元数据对象。

- 样本数据

DB2 Cube Views 还提供了样本应用程序和数据库来帮助用户学习。用户可以在 DB2 目录与 OLAP 工具之间交换元数据对象。

- DB2mdapiclient 实用程序

这种 DB2mdapiclient 实用程序是作为对面向多维服务的应用程序进行编码的样本源代码提供的。

WebSphere Information Integrator



由于当前信息技术的飞速发展，各企业的业务环境日益复杂，各种业务数据不断的增加，使得企业中各种数据类型繁多，数据源也十分分散。用户越来越希望能够对整个企业不同数据源及数据类型的数据进行统一的相关联的管理。但是目前很多企业受各方面因素的约束，分散的且不同厂商的数据源很难复制或集中到单一的数据库上。因此，这就使用户越来越渴望能够获得一个可以整合企业中各个分布式的数据源以方便访问多样化数据的解决方案。在这种需求之下，IBM公司提供了IBM Information Integrator 系列产品，能够在整个企业范围内访问

关系型、结构化和非结构化数据。它能提供战略上的数据集成架构，以帮助用户去访问、处理以及整合异构的、分布式的实时数据。IBM Information Integrator 包括联邦数据服务器 (Federated Data Server) 和复制服务器 (Replication Server)，用于整合异构的实时数据。联邦数据服务器 (Federated Data Server) 利用 SQL 或产生 SQL 的工具 (整合的开发环境、报表、分析工具) 访问、整合及处理分布式的和异构的数据。该产品主要适用于数据源为各类关系型数据库及其他如XML、Web或内容数据源。



IBM Information Integrator 核心内容包括:

联邦数据服务器 (Federated Data Server)

联邦数据服务器是一种特殊类型的分布式数据库管

理系统。一个联邦数据服务器包括一个作为服务器的DB2实例，另外还可以有一个或多个不同的数据源对应不同厂商的关系型数据库，以及访问数据库和数据源的客户端 (用户和应用)。利用联邦数据服务器，用户可以在一个SQL语句中向不同的数据源

发送分布式请求，从而实现集成异构数据库的功能。DB2 联邦数据服务器的强大性在于：

- 联合本地表和远程数据源，就象所有的数据都在本地一样操作。
- 利用数据源的处理功能，向数据源发送分布式请求。
- 在一个联邦数据服务器上处理分布式请求的各个部分，弥补数据源上的 SQL 限制。

联邦数据服务器支持的数据源包括：

关系型数据库：IBM DB2，IBM Informix Dynamic Server，IBM Informix Extended Parallel Server，Microsoft SQL Server，Oracle，Sybase SQL Server，Sybase Adaptive Server Enterprises，Teradata，以及 ODBC 数据源。

非关系型数据源：IBM Websphere MQ Message queues，Web services，Microsoft Access，Microsoft Excel spreadsheets，flat 文件，XML 文档，LDAP 目录，和 Entrez，Blast，HMMer，BioRS，Documentum，IBM Lotus Extended Search 能够访问的数据源。IBM Lotus Extended Search 提供访问多个数据存储，包括 Domino，IBM DB2 Information Integrator for Content sources (比如 IBM Content Manager，IBM Content Manager OnDemand，和 IBM ImagePlus)，关系型数据库 (IBM DB2，Oracle，Sybase，Microsoft SQL Server，Microsoft Access)，Lotus Domino.Doc，Microsoft Index Server，Microsoft Site Server，Microsoft Exchange，和超过 18 个搜索网站。

- 针对开发：

一个开发工具箱，提供开发访问其它的数据源。

应用程序能够查询或检索整合后的数据源就象它们在单一的数据库上

- 查询操作使用标准 SQL。
- 联邦数据服务器使用基于成本的分布式查询优化器选择最佳的访问路径以获取更好的查询性能。
- 用户可以在联邦数据服务器上定义数据缓存以提高查询效率。
- 应用程序能够访问传统数据库或 Web 应用客户端。

针对异构关系型数据库的复制服务器 (Replication Server)

客户可以在异构关系型数据资源之间复制数据

IBM DB2，IBM Informix，Microsoft，Oracle，和 Sybase 都是可以作为复制源或复制目标。Teradata 可以作为复制目标。

用户可以配置一个多样性的拓扑环境

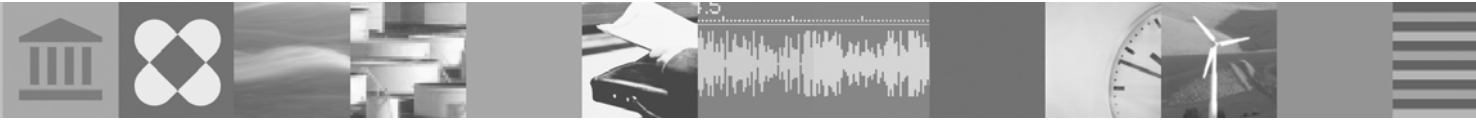
- 复制服务器支持分布式 (从一个数据库到多个数据库的数据迁移) 和集中式 (从多个数据库到一个数据库的数据迁移)。
- 数据迁移可以同时使用标准的 SQL 表达式或存储过程。
- 数据迁移可以定时的自动执行，或在一定的时间间隔执行，或连续执行，或者由事件触发执行。

数据仓库管理工具 DB2 Warehouse Manager

DB2 Warehouse Manager 提供了用于构建、管理及访问数据仓库的强大工具。Warehouse Manager 具有以下特性：

- 开发简单。只需要用简单的点击、拖拉等 Windows 常用操作，系统会自动生成标准的 SQL 语句，当然，也可以自己定义 SQL。一般情况下不需要编写程序，但提供接口为满足特殊需要，用户也可以自己用 VB、C/C++ 等写数据处理程序。
- 维护简单。由于基本不需要编程实现，因此可以适应于变化的需求。当需求发生变化时，只需要改变数据仓库中的定义即可。
- 自动化的执行。通过调度(schedule)，数据仓库可以自动完成所有的数据抽取、整理、转换，以及分发等功能。
- 支持众多的关系数据库，以及非关系数据库。从个人机到 NT 服务器、Unix 服务器以及 IBM 的主机系统。
- 提供分布式数据处理功能。利用 Agent 技术，可以以分布式处理来自多个数据源的数据。这一点对于企业级数据仓库来说非常重要，因为它可以提供良好的可扩展性。
- 元数据管理。DataGuide 提供方便的手段管理数据仓库的元数据，用户可以用各种方式查看它们。同时，IBM 正在致力于制定数据仓库元数据的国际标准(以 Warehouse Manager 的 TAG 文件为基础)，以实现不同数据仓库厂商之间的数据交换。
- 众多的第三方工具。如 DataStage、Vality、ETI 等可以和 Warehouse Manager 非常好的集成在一起，为数据处理提供增强的功能。
- DB2 UDB 是业界公认的性能优良、开放、具有出色可扩展性的数据库。Warehouse Manager 和 DB2 UDB 密切协同工作，为数据仓库的良好运作提供保证。这也是为什么世界上许多大型跨国公司，如花旗银行、联邦快运等都将数据仓库建立在 IBM 方案之上。

数据挖掘 DB2 Intelligent Miner



DWE 中的 IBM DB2 Intelligent Miner 主要是以下产品的集合:

- Intelligent Miner Scoring (IM Scoring)。
- Intelligent Miner Modeling (IM Modeling)。
- Intelligent Miner Visualization (IM Visualization)。

这些产品支持快速实现 IM 分析, 此分析是嵌入在商业智能 (BI)、电子商务或传统联机事务处理 (OLTP) 应用程序中的。它们作为独立定价的产品提供。它们还在 IBM DB2 Data Warehouse Edition 中提供。

在 DB2 中, 专业人员可以交互地从命令中心或命令行处理器发出 SQL 语句。他们还可以从这些接口之一启动 Intelligent Miner 函数。

Intelligent Miner V8.2 提供基于 Excel 电子表格的加载演示。此演示阐明了 SQL API 的功能以及如何使用它。有了这个演示, 您可以用 IM 函数为原型开发一个“概念证明”而无需处理 SQL API 的复杂情况。

为了辅助应用程序开发者, Intelligent Miner V8.2 提供了 WebSphere Studio Plug-in。该插件包含集

成在 WebSphere Studio Plug-in 环境中的图形向导和编辑器。通过这些工具, 应用程序开发者可以图形化方式为挖掘任务建模, 生成 SQL 将 IM SQL 功能嵌入他们的商业应用程序中。

根据经验、个人偏好、挖掘技能和待解决问题的复杂度, 您可以在以下开发挖掘解决方案的方法中进行选择:

- 使用样本和教程开始并进行改编使它们适应您自己的商业问题。
- 使用 WebSphere Studio Plug-in 提供的图形界面来定义挖掘过程。生成代码并将它集成到您自己的商业应用中。
- 对普通挖掘任务使用简单挖掘过程来执行典型的挖掘任务。
- 使用命令行脚本生成器 idmmkSQL 作为起始点来编写 Scoring 语句。
- 在 SQL 脚本中或者从任何 JDBC、CLI、ODBC 或 SQLJ 应用程序中使用功能强大的低级别 SQL/MM API。

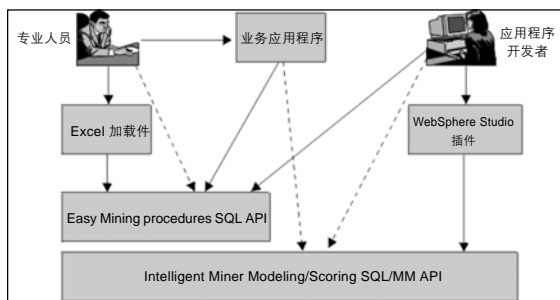


图 1. 在您的商业环境中使用 IM 产品

图 1 显示了如何在商业环境中使用 IM 产品的典型应用场合:

商业应用程序的应用程序开发者使用开发者工具将 IM SQL 功能集成到应用程序中。

专业人员从商业应用程序中使用此挖掘功能。

DB2 Intelligent Miner Scoring

IM Scoring 扩展了数据库功能，并使用户能够实时部署数据挖掘分析系统。通过数据挖掘所产生的分析数据，为企业和消费客户提供更加个性化的服务，或通过不断进行模型改进以响应特定情况，用户可以在商业智能和运营应用程序中采用它们，以便为企业及提供更好的服务。

IM Scoring 使应用程序能够将 PMML 模型应用到大型数据库、数据库子集或单行或情况中。应用程序使用 SQL API，此 API 由用户定义的函数 (UDF) 和用户定义的方法 (UDM) 组成，以执行记分操作。PMML 模型可能由某个 Intelligent Miner 产品创建

或通过其它支持 PMML 模型的应用程序和工具来创建。IM Scoring 的特点如下:

- DB2 Intelligent Miner Scoring 提供了如同数据库扩展的无与伦比的计分技术: DB2 Extender 和 Oracle cartridge。
- 通过简单的 SQL 编程界面和标准开发界面，DB2 Intelligent Miner Scoring 提供了无与伦比的可扩展性和强大功能。
- DB2 Intelligent Miner Scoring 在关系数据库中将挖掘模型另存为 XML 对象。实施基于数据挖掘的新行业标准。
- DB2 Intelligent Miner Scoring 根据数据挖掘模型中表示的一系列预定标准对记录进行计分——对这些记录的主题进行分段、分类和分级。
- DB2 Intelligent Miner Scoring 支持所有可提供数据计分功能的 DB2 Intelligent Miner。例如决策树、基于中心和基于分发的群集、多项式回归及神经网络。通过使挖掘模型和计分逻辑与应用程序隔离开来，您可以在趋势改变或获得其它信息时不断改进模型——无需中断应用程序。

用户可以在各种环境下执行 DB2 Intelligent Miner Scoring。例如 DB2 Extenders 与 Oracle cartridge、通过 DB2 DataJoiner 对联合数据的 DB2 的存取。

IM Scoring 包括 IM Scoring Java Bean，使您可以在给定 PMML 模型的情况下对 Java 应用程序中的单个数据记录计分。这可以用于将计分集成到电子商务应用中，例如用于客户关系管理 (CRM) 系统的实时计分。

DB2 Intelligent Miner Modeling

IM Modeling 提供 IM Modeling 技术，例如 DB2 extender。它使得 SQL 应用程序能够基于 DB2 Universal Database(TM) V8.1 或 V8.2 SQL 访问的数据来开发分析模型。DB2 IM Modeling 的特性如下：

- DB2 Intelligent Miner Modeling 为以下建模操作提供了 DB2 Extender。
- 关联发现。应用示例包括发现购物篮分析中的产品关联、电子商务站点的站点访问模式或购买的金融产品组合。
- 人口群集。应用示例包括市场细分、商店背景描述及购买行为模式。
- 分类树。应用示例包括基于理想结果的客户背景描述，例如购买倾向、预计的花费水平，以及在某段时间内客户减少的可能性。
- DB2 Intelligent Miner Modeling 是 DB2 数据库的高级 SQL 扩展，它能够使建模功能嵌入到商业应用程序中。
- DB2 Intelligent Miner Modeling 支持以符合业界新分析模型标准“预测模型标记语言 (PMML) V2.0”的形式开发数据挖掘模型。
- 当发现新关系时，DB2 Intelligent Miner Scoring 可将这些新关系实时应用到新数据。
- 数据挖掘模型分析可通过基于 Java 的结果浏览器 DB2 Intelligent Miner Visualizer 获得。通过 DB2 Intelligent Miner Modeling，即使非专家级人员也能够查看及评估数据挖掘建模流程的结果。

DB2 Intelligent Miner Visualization

IM Visualization 提供以下 JAVA 观测器来显示数据建模结果以供分析：

- 关联观测器
- 分类观测器
- 群集观测器
- 回归观测器 (仅对 IM Modeling 模型)

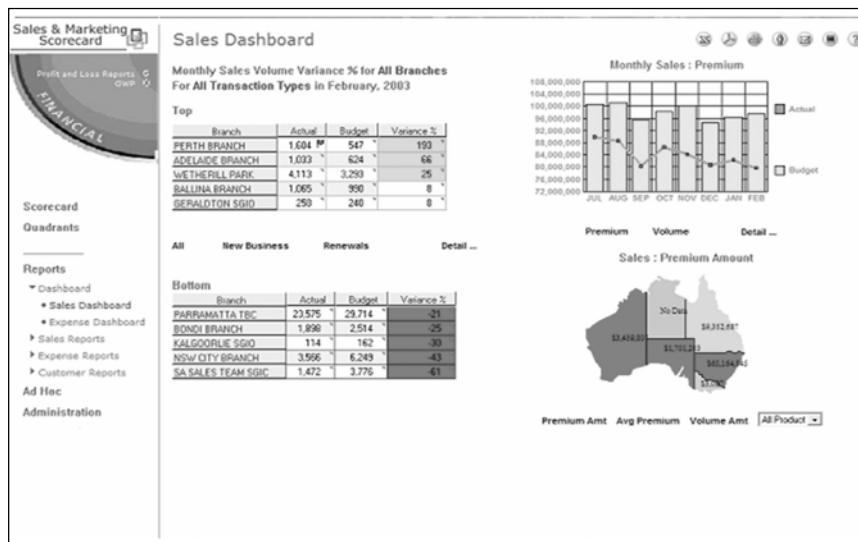
可使用 Intelligent Miner Visualizer 使符合 PMML 的挖掘模型可视化。应用程序可调用这些观测器来显示模型结果，或者您可将观测器部署为 Web 浏览器中的小应用程序以供现成分发。可通过使用 IM Modeling 或其它通过使用 PMML 模型支持互操作性的应用程序和工具来开发模型，或者可将 Intelligent Miner for Data 的模型作为 PMML 模型导出。模型作为平面文件存储，或者您可以直接从数据库中将挖掘模型可视化。数据库中的模型已经通过 IM Modeling 创建或者已经使用 IM Scoring 的导入功能导入数据库中。

PMML 标准的焦点是计分的互操作性。如果 PMML 模型包含 IM Visualization 扩展 (例如，IM Modeling 产生的分布统计或质量信息)，则使用 IM Visualization 能够实现最好的结果。

IM Visualization V8.2 提供以下增强功能：

- 成果图
- 字段重要性
- 质量分析

前端展示应用开发平台 DB2 AlphaBlox



DB2 AlphaBlox 是一种业界领先的平台，它可实现应用程序中嵌入的集成分析的快速汇编和广泛部署。其拥有基于 J2EE (Java 2 平台、企业版) 标准——开发基于 Web 的企业应用程序的业界标准——的开放式、可扩展架构。通过在无需复杂编程的情况下自动处理应用程序行为的众多细节，其简化了企业应用程序开发。

DB2 AlphaBlox 提供了各种模块化、可再用的 Blox 组件，以及应用程序框架、功能强大的编程模型和各种用于对分析应用程序进行汇编的开发工具。针对其运行时环境，DB2 AlphaBlox 充分利用了标准

J2EE 应用程序服务器。DB2 AlphaBlox 可安装在领先的商用 J2EE 应用程序服务器上，例如 IBM WebSphere 等。

当开发具有嵌入式 DB2 AlphaBlox 功能的应用程序时，您能够利用由基本 J2EE 应用程序服务器提供的众多功能，包括增强的性能、安全性及个性化。与应用程序服务器环境的集成可使应用程序生成器充分利用 DB2 AlphaBlox 来提供与构建、部署及执行分析应用程序相关的基本功能，同时依靠应用程序服务器来提供强大可靠的管理和部署服务。

组件

DB2 AlphaBlox 提供了可满足集成分析应用程序设计要求的广泛 Blox 库，以便实现最高的可用性。这些组件包括：

数据存取 Blox，其通过用户界面与相应数据源之间的连接管理数据存取。由于 DB2 AlphaBlox 直接存取您数据库中的数据，因此充分利用其功能的应用程序将遵守任何安全特性或内置到您数据库中的限制。DB2 AlphaBlox 显示了由多维数据库引擎提供的所有分析功能（例如：分级、推导出的计算、排序、高级过滤、百分位数、十分位数、变异数、标准差、关联、趋势分析、统计功能，及其它高级计算）。此外，DB2 AlphaBlox 还可使最终用户和应用程序开发人员创建定制的计算所得成员。

DataBlox 还提供了以 XML 格式返回数据的 API。这开启了通向可扩展性的大门，充分利用 DB2 AlphaBlox 的应用程序将与企业应用程序相集成。它还实现了到支持 XML 的客户端（包括蜂窝电话、寻呼机和 PDA）的数据传输。应用程序开发人员还能够显示 Web 服务中的数据和 / 或构建定制的用户界面。

由 DB2 AlphaBlox 提供的用户界面 Blox 具有强大的功能和高交互性特点，它完全可以定制，以提高您应用程序的可用性。这些用户界面元素采用 DHTML 技术来提供丰富的用户体验，包括菜单栏、右键菜单，以及瘦客户端中的定制布局（无需 Java、ActiveX 或其它浏览器插件）。

窗体元素 Blox 提供了多个非常适用于开发定制分析应用程序的窗体元素 Blox。所有窗体元素 Blox 不仅保持了窗体元素的当前状态，从而使开发人员无需编写额外代码，而且还与 Java beans 等其它组件相链接，包括能够以最少的编码提供最常需要的功能的数据存取 Blox 和用户界面 Blox。

商业逻辑 Blox 提供了有助于将动态复杂商业逻辑结合到集成分析应用程序中的商业逻辑。

分析基础设施 Blox 包含大量定制、个性化及协同功能。应用程序开发人员能够定制他们的业务部门应用程序，并通过 Blox 属性、Blox JavaScript/Java API 及应用程序和用户定制属性，针对每个单独用户对界面进行个性化。

DB2 AlphaBlox 编程模型

DB2 AlphaBlox 支持标准的 J2EE 应用程序开发模型，从而提供了可实现应用程序交付的全面开发范例。通过将每个组件显示为 Java bean 并通过丰富的 Java API 集实现对这些 bean 的访问，它为应用程序开发人员提供了定制用户界面和添加自己商业及应用程序逻辑的灵活性。

凭借 DB2 AlphaBlox 标记库，JSP 编码器不必知道隐含在这些组件中的低级技术细节，它们只需知道各自部分的语法和函数。这使没有 Java 经验的页创作者能够利用一流的创作工具在企业内部网和外部网上无缝结合分析系统。每个 Blox 均具有可利用标记在 JSP 页中轻松设定为定制值的全面属性集。



© International Business Machines Corporation 2005
国际商业机器中国有限公司

北京总公司

北京朝阳区工体北路甲二号
盈科中心 IBM 大厦 25 层
邮政编码: 100027
电话: (010)65391188
传真: (010)65391688

上海分公司

上海市淮海中路 333 号
瑞安广场 10 楼
邮政编码: 200021
电话: (021)63262288
传真: (021)63261177

广州分公司

广州市天河北路 183 号
大都会广场 18-20 层
邮政编码: 510620
电话: (020)87553828
传真: (020)87550182

沈阳分公司

沈阳市沈河区青年大街 219 号
华新国际大厦 19 层
邮政编码: 110015
电话: (024)23962288
传真: (024)23961040

武汉分公司

武汉市汉口建设大道 700 号
武汉香格里拉大饭店 302 室
邮政编码: 430015
电话: (027)85805588
传真: (027)85800088

深圳分公司

深圳市深南中路 333 号
信兴广场地王商业大厦
38 层 3805, 3806
邮政编码: 518008
电话: (0755)82462193
传真: (0755)82462186

南京分公司

南京市新街口街金陵饭店
世界贸易中心 16 楼
邮政编码: 210005
电话: (025)84716677
传真: (025)84729054

成都分公司

成都市人民南路 2 段 18 号
川信大厦 27 层
邮政编码: 610016
电话: (028)86199888
传真: (028)86199500

西安分公司

西安市高新区科技路 48 号
创业广场 B 座 1202 室
邮政编码: 710075
电话: (029)88316868
传真: (029)88323777

杭州分公司

杭州市杭大路 15 号
嘉华国际商务中心 1506 室
邮政编码: 310007
电话: (0571)28896988
传真: (0571)28891128

昆明办事处

昆明市洪化桥 20 号
海逸酒店 512, 513 室
邮政编码: 650031
电话: (0871)5388555
传真: (0871)5380199

福州办事处

福州市五四路 73 号
福建外贸中心酒店 9925 室
邮政编码: 350001
电话: (0591)87523388-9925/9938
(0591)87600122
传真: (0591)87541814

重庆办事处

重庆市渝中区邹容路 68 号
大都会商厦 21 楼 2105 房
邮政编码: 400010
电话: (023)63830503
传真: (023)63830513

长沙办事处

长沙市解放东路 380 号
华天大酒店贵宾楼 1008 室
邮政编码: 410001
电话: (0731)4169188
传真: (0731)4116845

乌鲁木齐办事处

乌鲁木齐市东风路 1 号
海德酒店 17 楼 B 座
邮政编码: 830002
电话: (0991)2338911
传真: (0991)2831805

哈尔滨办事处

哈尔滨市道里区友谊路 555 号
哈尔滨香格里拉大饭店 4 层
邮政编码: 150018
电话: (0451)87606688
传真: (0451)84899988