

# 利用 **IBM Netezza** 的高级分析功能

设备方法如何简化高级分析的使用



---

## 目录

- 2 简介
  - 3 IBM Netezza Analytics 能为我做什么?
  - 4 我们如何使用 IBM Netezza Analytics?
    - 数据准备和变换
    - 创建并发布模型
    - 模型计分
    - 算法开发
  - 7 摘要
  - 7 关于 Netezza (一家 IBM 公司)
- 

## 简介

带有 IBM Netezza Analytics 的 IBM Netezza 数据仓库设备将突破性的数据仓库设备融入高性能、可扩展的分析，可处理大量数据，从而以比传统解决方案快数个数量级的速度解决复杂问题，将分析发挥至极致。IBM Netezza Analytics 是开放式、灵活的高级嵌入式分析平台，随每一个 IBM Netezza 设备交付，该设备支持数据库内分析的开发和部署，以实现颠覆性结果。

IBM 通过将分析嵌入设备并使其成为设备不可或缺的一部分，简化分析部署，同时充分利用了 IBM Netezza 的底层体系结构高处理吞吐量的优势。

利用 IBM Netezza Analytics，数据准备、数据挖掘、预测性建模和优化的并行分析可以利用 IBM Netezza 的非对称大规模并行处理 (AMPP) 体系结构针对大量数据的高级分析实现高吞吐量。IBM Netezza 数据仓库设备可轻松扩展，以创建自身功能强大的高级分析，并将其嵌入设备。分析应用程序、虚拟化工具和商业智能工具可通过各种编程方法（例如，SQL、Java、MapReduce、Python、R、C、C++ 和 Fortran）来利用并行高级分析，从而交付功能强大、富于洞察力的分析。综合性高级分析环境更加便于按您自己的方式对其加以利用 - 使用您首选的工具进行高级分析的即席分析、原型制作以及生产部署。

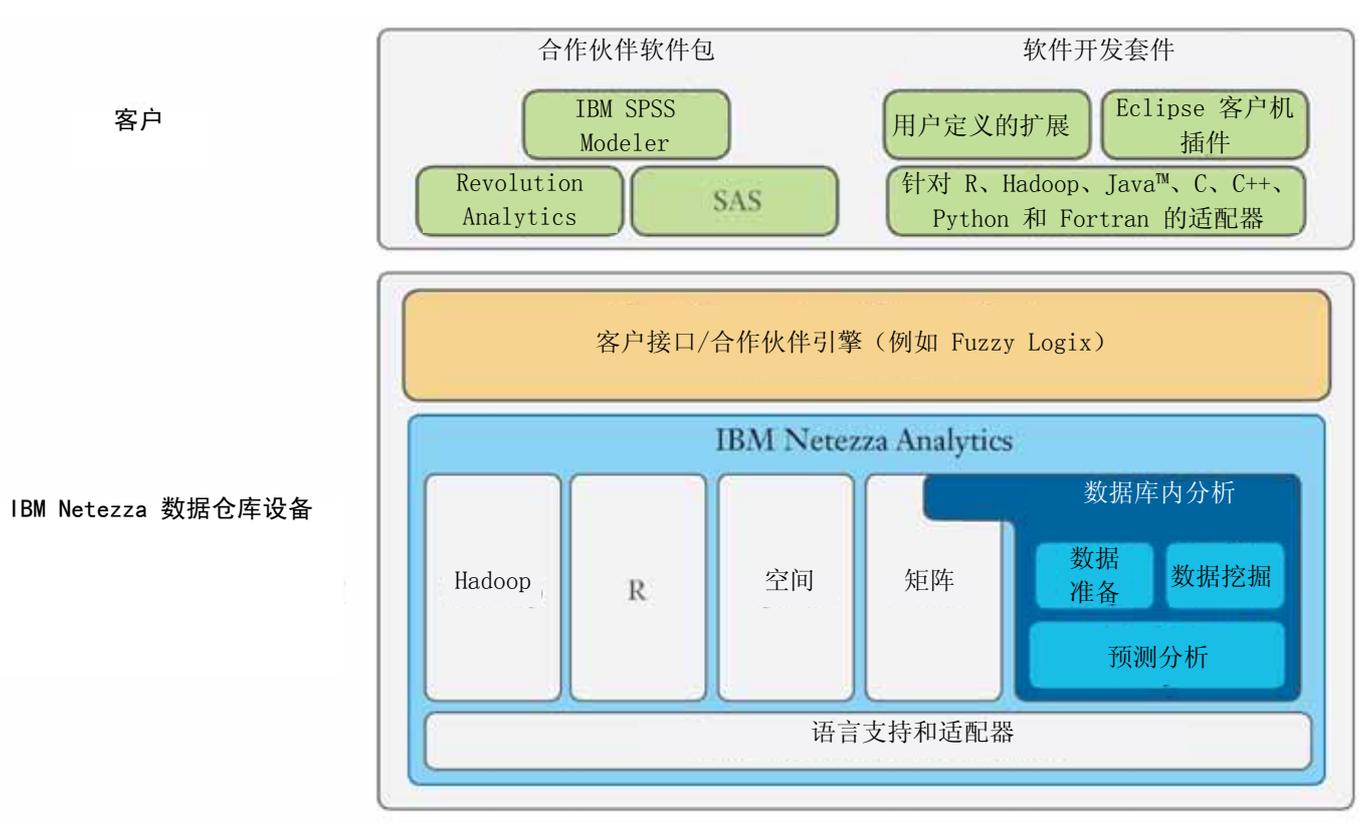


图 1: IBM Netezza Analytics 体系结构

### IBM Netezza Analytics 能为我做什么？

IBM Netezza Analytics 利用大量数据和并行处理的融合来支持高性能、可扩展的分析应用程序。IBM Netezza 设备的非对称大规模并行处理 (Asymmetric Massively Parallel Processing™, AMPP™) 使得分析得益于数据和任务并行性，

从而快速的得到富于洞察力的成果。通过将分析与我们独一无二的 AMPP 体系结构紧密集成，IBM 突破了技术限制，达到更快的表现性能，并且能够处理先前不可能处理的数据。这一数据量和计算复杂性的增加使统计人员、建模人员和开发人员得以在正确的时间使用正确的工具来发现高价值的商机。

目前，从数据仓库复制数据以进行分析处理，导致开发周期延长，软件维护成本增加的情况屡见不鲜。这包括程序在数据仓库与其他平台间抽取、转换并装入数据。带有 IBM Netezza Analytics 的 IBM Netezza 数据仓库设备在数据仓库内部将分析处理移到数据旁来简化该过程，从而消除或者显著减少这些不产生效益的步骤。通过简化环境，分析人员可以专注于附加值更高的活动，例如建模、模拟和优化，从而为其组织带来显著价值。

IBM Netezza Analytics 整合了来自 IBM、合作伙伴和开放式源代码社区的丰富的高级分析集合。应用程序开发者和建模人员可以针对其独特的状况使用正确的高级分析组合来设计并开发综合性解决方案，从而解决其特殊挑战。各公司可以创建自己的高性能分析或者利用分析来更快速地开发分析应用程序，从而将高性能的并行分析作为起点来创建其自己的“独门秘笈”（或竞争优势）。

例如，在客户间拥有大量事务和高度可变性的 CRM 中，客户细分通常用于识别高利润客户，而预测建模可用于确定如何将低利润客户转变为高利润客户。建模和优化相结合，以确定提供给客户的最佳设计、消息传递和促销，从而提升响应率。数据挖掘和预测建模可用于识别并维持忠诚客户，并找到通过适当奖励增加客户忠诚度的方法。

## 我们如何使用 IBM Netezza Analytics?

IBM Netezza Analytics 是高级嵌入式分析平台，专为开展严格的分析而构建。IBM Netezza Analytics 用于：

- 构建并部署高级分析应用程序
- 通过可视化或商业智能工具利用并行分析
- 尤其针对大量数据或计算密集型问题执行即席分析

可视化和商业智能工具通过 SQL 进行目标查询，以利用 IBM Netezza Analytics 中的分析。但是，应用程序开发者和建模人员需要更大的灵活性来开发应用程序或即席分析，包括：

- 数据准备和变换
- 模型创建和发布
- 模型记分
- 算法开发

### 数据准备和变换

执行数据准备和变换活动的分析人员可以使用 IBM Netezza Analytics 工具和实用程序在过程中提供帮助。此类工具包括：

- 说明性统计信息
- 数据清理
- 数据变换
- 功能部件选择和规格缩减
- 统计测试和实用程序

需要数据准备和变换时，IBM 可通过高性能软件包加速该过程，同时压缩可存储在该设备上的数 PB 的用户数据。

### 创建并发布分析模型

创建分析模型通常是迭代过程。用于构建模型的数据可由为模型准备数据的建模人员（数据挖掘机、统计人员或量化分析人员）准备并进行变换。建模人员可从执行某些主管请求的即席分析开始。一旦结果通过复审，那么建模人员就可以创建改进，而不仅仅是最终生产部署的原型。建模人员通常在此过程中开发多个模型，并评估和比较各模型间的性能，然后选择用于最终部署的模型。在迭代开发周期的每个阶段，建模人员可以根据不断演进的需求和性能目标使用不同的开发工具和分析。

IBM Netezza Analytics 可为该过程的每个阶段提供灵活性。建模人员可以使用任一受支持的开发环境（如 Eclipse、R GUI）来创建其模型，或直接使用：

- IBM 的 Eclipse 插件
- IBM 的 R GUI 插件
- 合作伙伴的集成开发环境（IDE）
- 合作伙伴的分析开发环境（ADE）来创建分析 workflow
- 任何 IBM Netezza 支持的编程语言 - SQL、Java、MapReduce、Python、R、Fortran、C 或 C++

模型可使用以下任何高级分析的组合，包括：

- **开放式源代码分析** - R CRAN 软件包，可通过 IBM Netezza Analytics 轻松安装
- **合作伙伴分析** - 业内经验证的的分析，可在应用程序开发中加以利用
- **数据库内分析** - 我们的并行、高性能数据库内分析，用于数据准备、数据挖掘、预测性分析和地理空间分析

一旦模型用于预测将来成果，即可通过将优化与预测性建模结合起来将该模型用于实现最优决策的自动化。通过嵌入优化，该应用程序可解决棘手、复杂的业务难题，例如：

- 最小化运输成本，同时最小化库存
- 最小化资本投资，同时最小化整个产品服务组合的风险
- 最大化收入，同时最大化客户忠诚度
- 最小化市场营销预算，同时最大化线索生成

解决这些难题的同时，应考虑各种现实世界的约束，包括预算、人员配备、客户偏好和行为、公司偏好等。

### 模型记分

一旦构建完模型，即可将其应用于通常与预测模型相关联的大量数据。该过程称为记分。记分的示例包括将模型应用于数百万条记录，以：

- 检测欺诈
- 计算信用分数
- 对客户帐户进行重新定价
- 预测将来客户购买
- 决定将大量市场营销促销信息发送给谁
- 建议是否应增加/减少信用额度
- 总结是否应抛售存在拖欠行为的帐户
- 基于现场支持电话确定产品保修策略
- 确定适当的客户促销，以增加销售额

### 算法开发

算法开发者拥有众多开发工具选择，包括针对基于 GUI 的开发和编程语言方法的集成开发环境 (IDE)。IBM 为 Eclipse (开放式源代码 IDE) 提供插件。此外，IBM 合作伙伴可提供 IDE 以加速算法开发周期。IBM 支持 R GUI，用于基于 R 语言开发算法。IBM 支持多种用于算法开发的编程语言，包括 Java、MapReduce、Python、R、Fortran、C 和 C++，以及用户定义的扩展。用户定义的扩展包括：

- **UDF** - 用户定义的函数，处理一行
- **UDA** - 用户定义的聚集，处理一组行
- **UDTF** - 用户定义的表函数，返回任何大小的表
- **UDAP** - 用户定义的分析过程，提供完整的灵活性，以执行一个或多个处理步骤
- **存储的过程** - 用户将函数、聚集和其他逻辑分组为多个过程

开发并测试算法后，可在设备上注册算法，该设备使算法可与任何其他数据库内分析一样，通过支持的编程语言进行访问。

## 总结

带有 IBM Netezza Analytics 的 IBM Netezza 数据仓库设备是功能强大且可扩展的高级分析平台，可简化高级分析的开发、部署和使用，同时达到无与伦比的性能和可扩展性。

IBM Netezza、合作伙伴、开放式源代码以及您自己的高级分析可通过 SQL、Java、MapReduce、Python、R、Fortran、C 或 C++ 进行组合与利用，从而利用 IBM Netezza 数据仓库设备的高性能并行体系结构为您的业务带来关键性的成果。

## 关于 IBM 数据仓储和分析解决方案

IBM 提供最广泛且最全面的数据仓储、信息管理和业务分析软件、硬件和解决方案的组合，以帮助最大程度发挥信息资产的价值，并发现新的洞察力，以更好更快地制定决策，优化业务成果。所有 IBM 的数据仓储和分析解决方案均旨在简化并加速交付业务分析洞察力。

IBM 的产品服务组合所包含的数据仓库设备将数据库、服务器和存储器集成到易于管理的单一设备中，该设备只需最低限度的设置和日常管理，并提供更快且更一致的分析性能。IBM 还提供预先构建、预先集成且针对工作负载优化的数据仓储和分析平台，以及数据仓库软件，以实现运营智能化。这些产品通过针对大数据和新型的分析工作负载的额外支持得到了增强，包括对大量传输中数据进行连续且快速的分析。

## 了解更多信息

要了解有关 IBM 数据仓储和分析解决方案的更多信息，请联系 IBM 销售代表或者访问：

[ibm.com/software/data/infosphere/data-warehousing/](http://ibm.com/software/data/infosphere/data-warehousing/)



---

国际商业机器中国有限公司  
北京市朝阳区工体北路甲 2 号  
盈科中心 IBM 大厦 25 层  
邮编:100027

IBM 官方网站:

**ibm.com**

IBM、IBM 徽标、ibm.com 和 Netezza 是 International Business Machines Corporation 在美国和/或其他国家或地区的商标。如果这些名称和其他 IBM 已注册为商标的名称在本信息中首次出现时使用商标符号 (® 或 ™) 加以标记, 那么表明这些符号在本信息发布时已经是由 IBM 根据美国联邦法律注册或根据普通法注册的商标。这些商标也可能是在其他国家或地区的注册商标或普通法商标。以下 Web 站点上的“Copyright and trademark information”部分中包含了 IBM 商标的最新列表:

[ibm.com/legal/copytrade.shtml](http://ibm.com/legal/copytrade.shtml)

Microsoft 和 SQL Server 是 Microsoft Corporation 在美国和/或其他国家或地区的商标。

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。

© Copyright IBM Corporation 2011  
All Rights Reserved.



请回收再利用