

运用逻辑数据仓库改进数据仓库，获得毫无束缚的敏捷性

解决方案

逻辑数据仓库

行业

适用于所有行业

官方网站

www.denodo.com

产品概述

Denodo 平台可供广泛访问企业、大数据和云数据源中的结构化与非结构化数据，并有批量和实时两种模式，充分满足数据密集型企业对分析型和操作型应用场景的性能需求，所需时间较传统数据集成工具也大幅缩短。

商业离不开具有可操作性的智能。多年以来，这种智能通常由商业智能工具来提供，方法就是从数据仓库中提取数据。业务相关者往往会期望所有相关的数据都能够复制进数据仓库中，即使有时候需要一天的时间才能等来最新数据。

随着近期技术的飞速发展，如大数据和云分析等，业务分析人员也以为在获取商业智能方面遇到的限制将变少，但遗憾的是，现实情况恰恰与之相反。

问题就在于，有些较新的数据源，如社交媒体平台数据、事务进程数据或有关机器运转的原始数据，其格式往往不适合存放在传统的数据仓库中。这些数据确实可以存储进去，也可以重新进行格式化，但成本非常高昂。

此外，如果这些数据重新格式化后存储在数据仓库中，数据量将迅速增长，成本也会继续大幅增加。与之相对的，公司会将这些数据存储在价格更低廉的云端存储系统中，如 Hadoop，但还是会有问题：不是所有数据都同时存在于数据仓库中，因此制作报表时就没办法使用全部数据。

逻辑数据仓库则是这个难题的最佳解决方案：所有数据都留在原位，但在显示上是位于同一位置，这样不管什么时候，所有数据都可以用来制作报表。

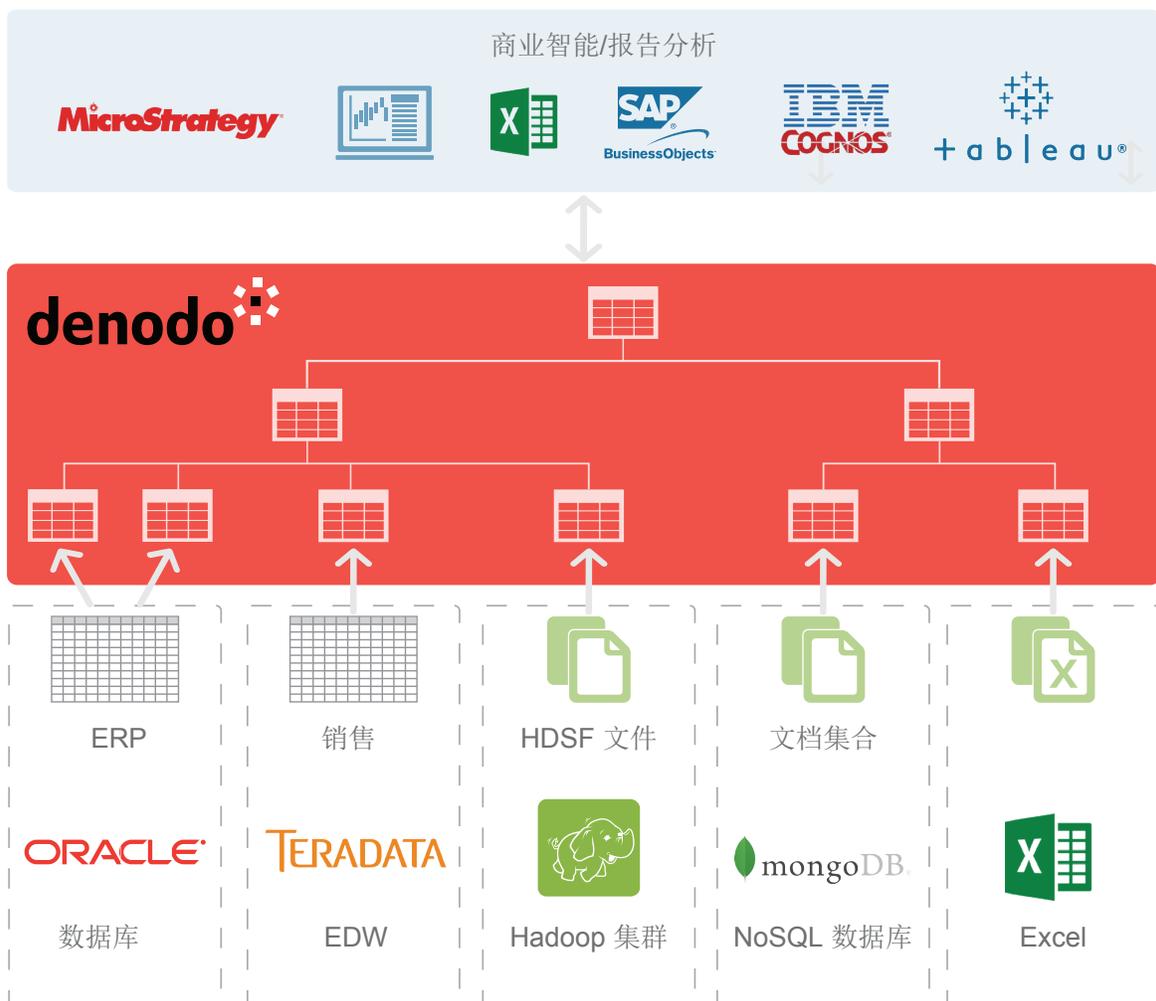
什么是逻辑数据仓库？

这个想法最先由 Gartner 提出，并这样描述逻辑数据仓库：位于传统数据仓库之上的“语义层”。语义的重要意义在于，两个系统只有在使用同一种语言时才能共享数据。将语义置于单独的一层中，由该层充当解释器，就能够与任何数据源进行对话。要访问任何一个数据集，业务分析人员需要首先与语义层对话，语义层再与数据源对话。

数据仓库研究所 (TDWI) 高级总监 Philip Russom 将逻辑数据仓库的定义简化为：“数据仓库架构的逻辑层或虚拟层，其下则集成了架构的物理层”。在这个定义中，“逻辑或虚拟”层对应 Gartner 的“语义”层，“物理层”对应“传统数据仓库”。

在这两种定义中，数据使用者均通过位于各个数据源之上的层访问数据，汇总到一起后制作报表。请注意，逻辑数据仓库不包含实际数据；而仅仅包含了用于访问每个数据源的智能。

Denodo 平台运用数据虚拟化技术，让公司能够建立起逻辑数据仓库，可以实时连接各种数据源，无论这些数据源位于本地还是在云端。



逻辑数据仓库：六大场景

Denodo 平台兼具灵活性和强大功能，让公司能够建立起广泛的逻辑数据仓库场景。以下是六个最常见的应用场景：



1. 虚拟数据集市

这是逻辑数据仓库最简单的配置，让业务分析人员只管访问所需的任何数据，不必操心数据访问的技术细节。虚拟数据集市集成了多个原本离散的数据源，并使之在呈现给用户和报告工具时都位于同一位置。

2. 数据仓库，扩展包含了主数据

使用逻辑数据仓库还可以将数据仓库中的数据与主数据管理系统中的数据相结合。主数据管理系统是客户或产品存储的“黄金记录”，因此当业务分析人员从数据仓库中提取数据时，就会与主数据实时同步，在提高客户服务敏捷性的同时大大减少代价高昂的错误。

3. 数据仓库，扩展包含了云数据

不同云环境（如 Salesforce.com）中的数据未必都存储在数据仓库中，因此会极大地妨碍商业智能。在这一场景中，逻辑数据仓库能够无缝实时访问云端存储的几乎所有数据，并与数据仓库中存储的数据完全集成，以实现强大的商业智能方案。

4. 多个数据仓库，合而为一

在此场景中，逻辑数据仓库可以集成两个或更多现有的数据仓库，并使之在显示上合而为一。这是一种强大的应用场景，由于公司间的合并，许多公司将被迫容纳两个或更多的数据仓库；并需要将所有数据从彼此隔离的仓库迁移到单一的整体数据仓库中，进行极其耗时耗资的物理集成。

在迁移过程中，例如从本地数据仓库迁移到云端数据仓库（如 Redshift 或 Snowflake），逻辑数据仓库可以解决用户需要访问两个数据源的复杂问题，从而实现平滑过渡。事实上，有了逻辑数据仓库，迁移过程甚至都不会引起用户注意。

5. 数据仓库历史数据卸载

如今，由于成本低廉，许多公司都在用云端存储器集群（如 Hadoop）来存储大量历史数据。但出于制作报表的目的，这类数据随后就会与数据仓库中的数据相分离。在这种场景下，逻辑数据仓库能够再一次实时将来自两个系统的数据相结合，这样您就可以对所有数据运行查询，而不会影响到业务流程。

6. 数据仓库扩展

类似地，公司往往希望将常用数据存储在手边的数据仓库中，将很少使用的数据放到云存储中。逻辑数据仓库让公司能够完全自助选择数据存储位置，而不会妨碍实时商业智能。

性能

在本篇简介中，我们提到过逻辑数据仓库能够提供对各种数据源的实时访问。那么在处理查询方面，逻辑数据仓库与传统数据仓库又有哪些区别？

逻辑数据仓库和传统数据仓库在查询处理性能上几乎没有差别。在一项传统数据仓库和逻辑数据仓库之间的性能对比测试中，传统数据仓库在查询“按客户列出总销售额”时，用了 20.9 秒返回了 199 万行结果，而逻辑数据仓库解决方案只多用了半秒时间。

逻辑数据仓库的业务优势

- 实时访问所有企业数据，以制作业务报表。
- 提升追加销售和交叉销售机会。
- 提高操作速度和客户服务能力。
- 支持在不影响业务连续性的情况下进行数据迁移。
- 在数据访问方面保持与 IT 的独立性。

客户案例

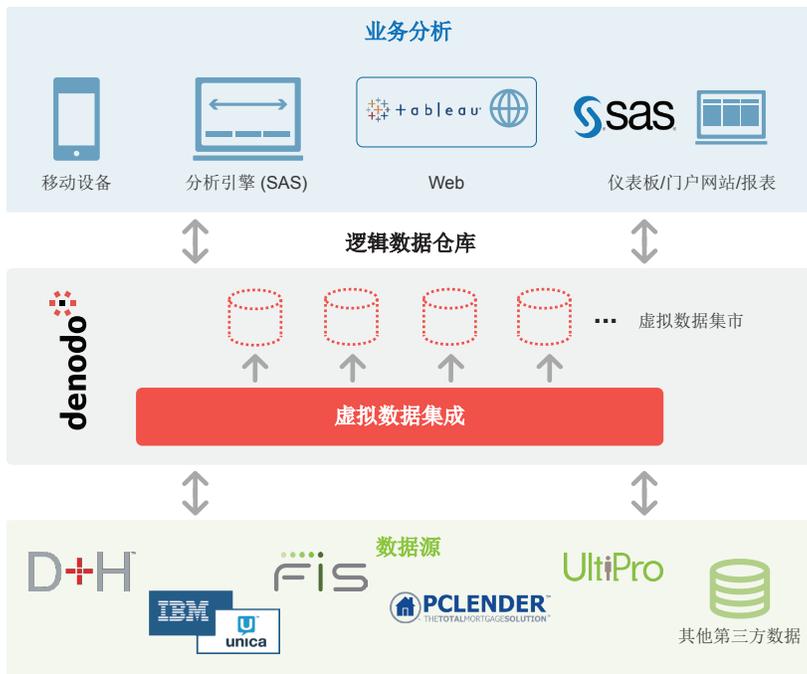
本节我们将介绍四家已成功利用逻辑数据仓库强大功能的公司客户案例。

客户案例：Seacoast Bank

Seacoast Banking Corporation of Florida 是佛罗里达州最大型的社区银行之一。近期，该银行深深感受到分别维护后台运营、数据仓库和贷款发放等功能系统所带来的影响，而一系列并购动作又进一步增加了复杂性。

Seacoast Banking 中来自核心银行、网络银行、风险以及其他小组的用户都需要向 IT 团队提出请求，才能够获得静态报表进行操作和分析。这种临时性的手动报表制作过程效率低下又耗时。Seacoast 希望实施一种现代化的商业智能工具，能够帮助用户进行数据切片和分块，满足分析性报表的制作需求。

解决方案



Seacoast 借助 Denodo 平台，利用数据虚拟化建立起逻辑数据仓库。逻辑数据仓库可无缝集成云端和本地信息系统中的运营数据，还提供分析和报表工具（如 SAS 和 Tableau）的聚合视图。随着逻辑数据仓库的建立，如今 Seacoast 已经能够提供企业范围内的自助式商业智能，并且可使用现代化的报表工具制作交互式趋势报表。

更重要的是，现在 Seacoast 已经能够对新数据进行快速访问、统一并建模，以运用于多个业务部门，支持银行近乎实时地处理信贷管理、风险防控、内部运营和《银行保密法》要求。Seacoast 现在能够快速提取数据，对于该行而言这已经是一项创造了价值的重要改进。

Seacoast Bank 副总裁兼业务技术和数据管理总监 Mark Blanchette 称，“在支持我们的业务用户通过自助式报表获取有价值的信息方面，Denodo 的数据虚拟化技术发挥了至关重要的作用。Denodo 平台的强大功能显著加快了 Seacoast Bank 开展业务的速度。”

优势

- 逻辑数据仓库大幅提高了 Seacoast 银行业务运营团队（如存款和贷款业务部门）及时、准确做出决策的能力。
- 相较于传统的提取、转换和加载 (ETL) 解决方案，逻辑数据仓库能够在不到一半的时间内集成数据，支持 Seacoast 满足企业内多个业务部门的运营和分析需求。
- 逻辑数据仓库帮助 Seacoast 将静态报表的制作时间由最长三天缩短到两小时，快速为贷款、存款、欺诈、信贷和风险管理业务部门的业务用户提供交互式自助报表。
- 在逻辑数据仓库的技术支持下，如今 Seacoast 已经可以实时执行关键的业务操作，如贷款处理等。

客户案例：VIZIENT

Vizient, Inc. 是美国最大型的会员制医疗保健绩效改善公司，主营提供创新性的数据驱动解决方案、专业支持和协作机会，从而改善患者治疗效果并降低成本。Vizient 与退伍军人健康管理局 (VHA)、大学健康系统联盟、Novation、MedAssets Spend, Clinical Resource Management 和 Sg2 进行联盟。

由于近期的一系列收购动作，财务数据、供应商销售数据和会员销售数据分别存储在各种不同的系统和数据集中，因此 Vizient 急需在这些离散数据影响到业务连续性之前将数据进行集成。此外，Vizient 还需要整合此类数据，以便获得高度的敏捷性，把握住新身份所带来的新机会。

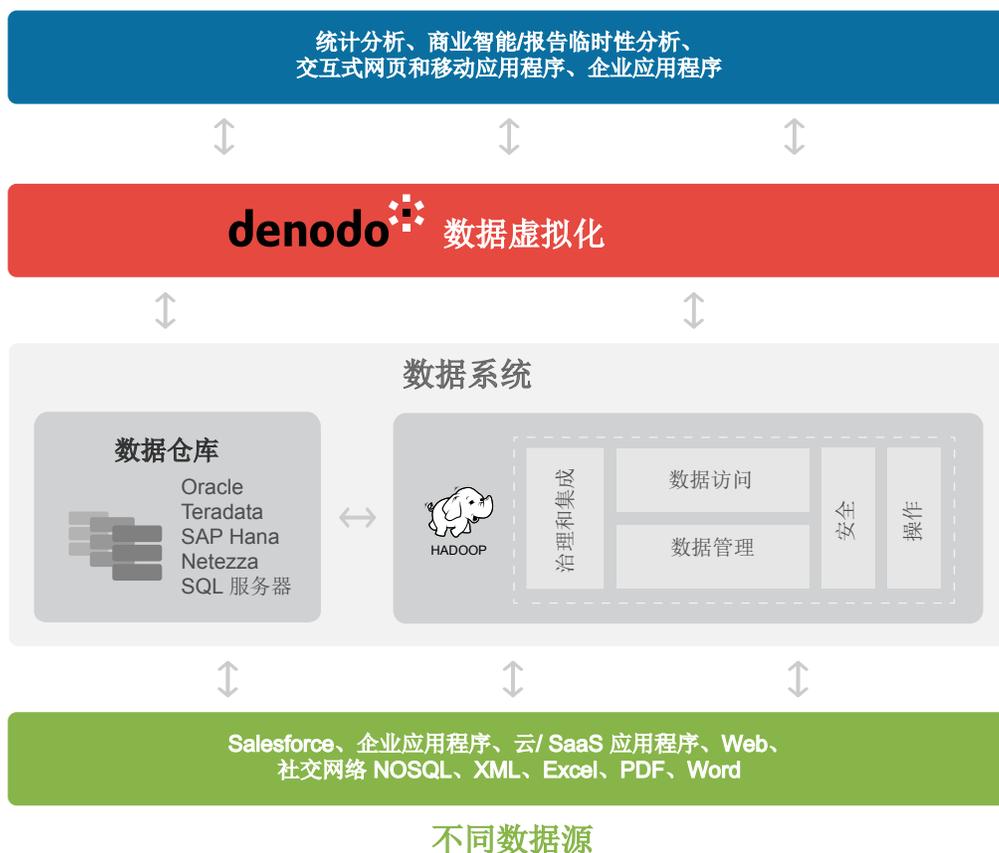
解决方案

Vizient 借助 Denodo 平台，通过逻辑数据仓库配置建立了一系列协同运作的逻辑数据集市，将所有数据关联在一起。

优势

在逻辑数据仓库的支持下，Vizient 能够：

- 将各个原有机构中离散的会计和财务数据集市统一到共享的虚拟库中。
- 为 Vizient 各机构的所有客户提供统一的供应商销售数据视图。
- 整合 Vizient 所有机构的会员支出和供应商销售数据，识别出可提高合同利用率的机会。

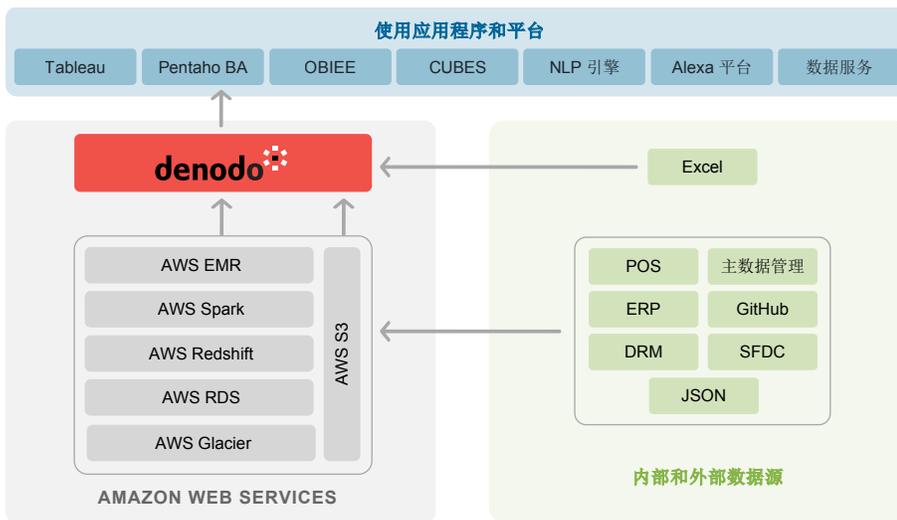


客户案例：Logitech

Logitech 成立于瑞士，是一家全球性的个人计算机和平板电脑配件提供商。EMEA 总部位于瑞士洛桑，美国总部位于加州纽瓦克。该公司主营个人电脑周边设备等产品的开发和销售，包括键盘、鼠标、轨迹球、麦克风、游戏控制器和网络摄像头。公司还在欧洲、亚洲和美洲设有办事处，并在美洲、欧洲、中东、非洲和亚太地区开展销售和营销活动。

Logitech 多年来致力于开发和提供数据服务，并使用本地部署系统进行分析。但为业务用户预配数据服务这种方式被动、耗时且效率低下。该公司的现代化产品和服务（如安全视频分析和智能家居设备）需要的是预测分析、实时数据分析和认知科学。为了实现这些功能，并且能够在合适的时间向业务用户提供合适的服务，Logitech 想将 IT 运营迁移到云端。云技术使得 IT 组织能够重新定义数据服务的生产和交付方式。

解决方案



Logitech 选择了 Amazon AWS 上托管的 Denodo 平台，来建立逻辑数据仓库。在这个全新架构中，来自不同本地数据源和其他第三方云端数据源（如 DRM、MDM、ERP、POS、Github 和 Salesforce）的数据都会加载到 Amazon S3 中。

Denodo 还集成了直接来自本地 Excel 文件和机器生成的数据、社交媒体数据、互联网数据以及其他来源的数据。

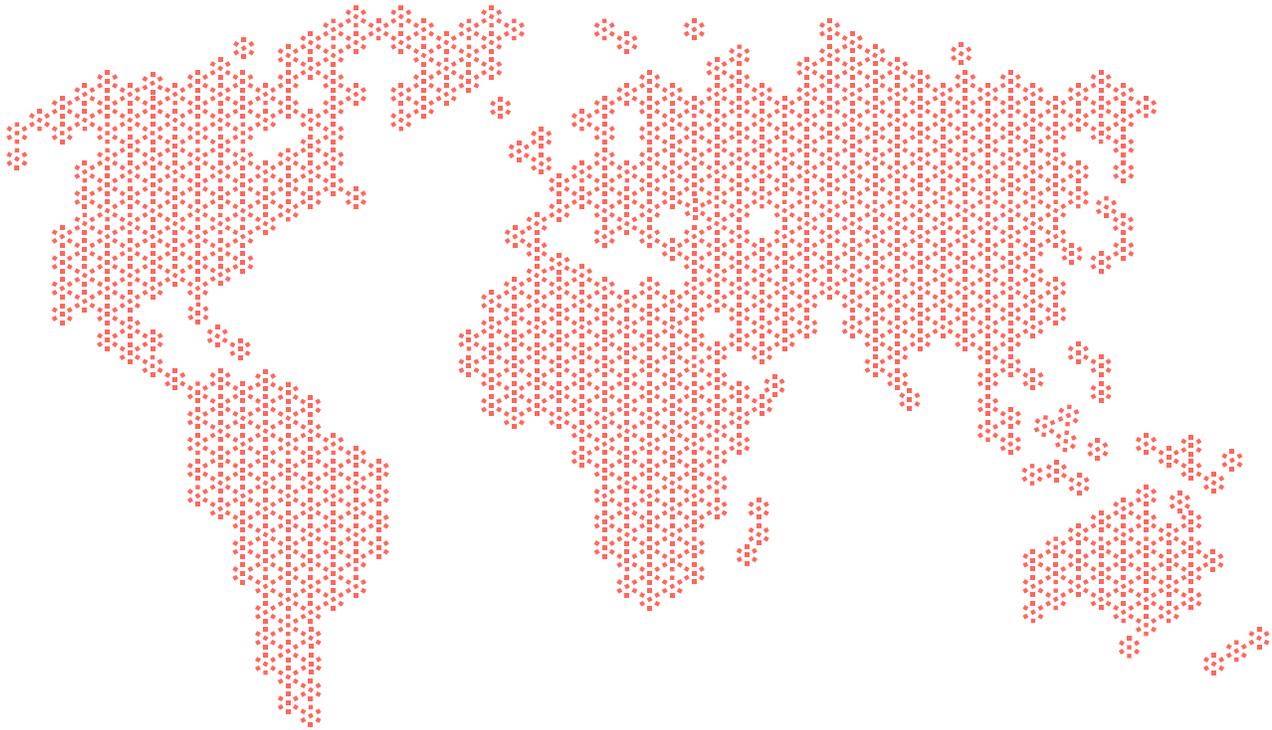
创建单一的一致数据存储后，Denodo 可向 Tableau、Pentaho BA 和 Web 服务等分析和报告应用程序提供数据。在 Logitech 基础架构中，Denodo 已成为单一可信来源，能够满足整个使用层的需求。

“我们部署了 Denodo 平台，以便将数据分析平台迁移到云，Denodo 数据虚拟化在这个过程中发挥了关键作用。” Logitech 大数据和分析负责人 Avinash Deshpande 表示。

优势

借助 Denodo 平台打造的逻辑数据仓库，Logitech 得以：

- 实施向云端的实时迁移，并将对业务运营的影响降到最低。
- 在必要的治理结构支持下，为快速创新创造共生环境。
- 达到卓越服务水平的同时降低成本。
- 帮助业务用户以便捷的自助方式使用信息。
- 快速形成原型设计，减少运营费用。
- 大幅加快数据科学和分析工作。



Denodo Technologies 是数据虚拟化领域的领导者，可为最广泛的企业、云、大数据和非结构化数据源提供敏捷、高性能的数据集成、数据抽象化和实时数据服务，而成本仅为传统方法的一半。Denodo 遍布各大行业的客户都明显提升了业务敏捷性和投资回报率。

官方网站 www.denodo.com | 电子邮件 info.cn@denodo.com | 社区网站 community.denodo.com



Denodo 平台 8.0 数据手册

现今的数据管理环境越来越复杂，因为数据分布在多个结构各异的数据系统（数据仓库、列式数据库、MPP 引擎、专用数据存储、SaaS/云应用程序等），而这些数据系统又可能分布在多个位置（本地、云等）。因此，很难以统一格式向业务应用展示数据，也很难确保治理策略和规则在整个数据交付链中得到实施。

逻辑数据结构是针对统一数据交付平台提出的一个构想，这种平台可以帮助业务使用者简化对多个数据系统的访问，隐藏复杂性，以方便业务使用的格式展示数据，同时保证数据按照预定义的语义和治理规则传递，从而解决上述主要问题。

数据虚拟化是实现逻辑数据结构构想的关键技术。数据虚拟化是唯一一种专为分布式架构设计的数据集成方法，在混合分布式架构的多个异构系统之上提供逻辑数据访问层。如今，数据管理生态系统在本质上是分布式的，因此逻辑数据访问层（例如数据虚拟化）是最佳方案。

Denodo 平台 8.0 是实现逻辑数据结构构想的重要一步。它具有的功能可以跨越多个地理位置分散的异构系统，以最合适每个使用者的格式，更快地将控制数据交付至业务应用。

Denodo 平台 8.0 为核心数据虚拟化应用场景提供了更强大的支持：逻辑分析架构、逻辑数据仓库和数据服务 API。但是，它较传统数据虚拟化方案更胜一筹，可以更好地支持新型用户和新型应用场景，例如数据科学和机器学习 (ML) 计划。

此外，这是平台即服务 (PaaS) 云战略的一大进步，能够从集中式 Web 控制台自动管理云基础架构。

Denodo 平台 8.0 有哪些主要功能？

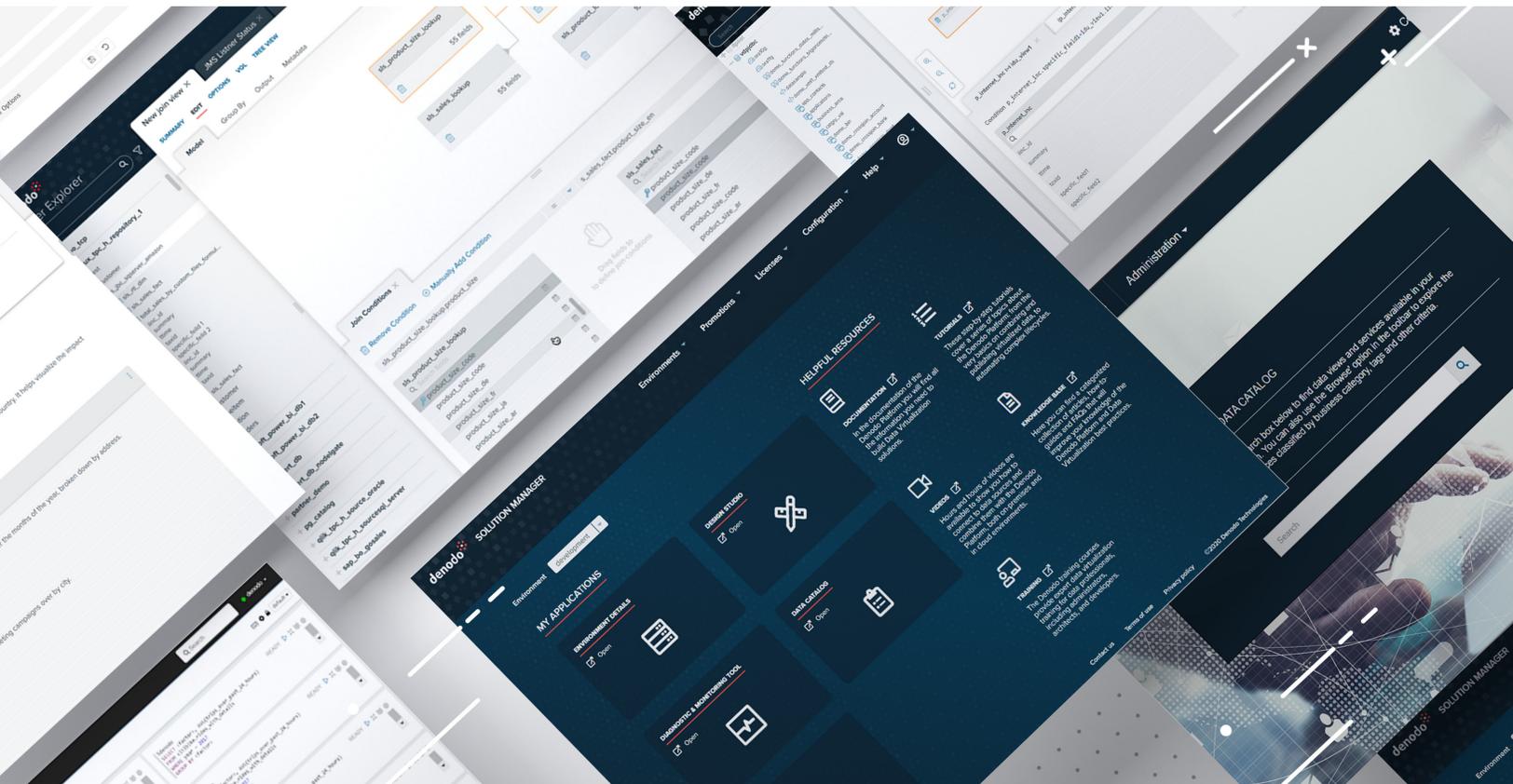
Denodo 平台 8.0 为所有 Denodo 平台组件提供完全基于 Web 的界面，强化统一的用户体验，大大简化 Denodo 平台在混合本地/云端部署中的管理。

全新 Denodo 中央 Web 控制台集成在解决方案管理器中，让本地和云端所有 Denodo 平台环境中的所有用户通过单一入口点即可访问所有 Denodo 工具。通过 Kerberos、SAML、OpenID 和 OAuth 支持 SSO，可与所有 Denodo 平台工具无缝连接。同时，工具之间的集成也更加紧密（例如，诊断和监视工具现已集成到 Solution Manager 中）。

开发人员现在可以使用基于 Web 的全新 Studio 工具开发视图和数据服务（也可以继续使用桌面版本，Denodo 平台 8.0 仍然提供支持）。这个新工具的设计原则是保持并增强 Denodo 平台向来拥有的易用性。

在性能方面，Denodo 平台 8.0 推出一个新概念“智能查询加速”，可以加快逻辑数据仓库/数据结构架构中的查询执行速度。在特定类型的多个查询中，事实和维度表的部分聚合（称为“摘要”）通常会联结在一起，对其进行预先计算，用于加快未来查询的执行速度。由于这些性能增强是由 Denodo 平台在后台进行处理，因此这项技术可显著提高性能，并有助于在自助服务方案中构建即席查询。这扩展了 Denodo 平台在分布式查询优化中的领导地位，增强了其他功能的优势，例如高级查询下推、MPP 加速和先进的缓存机制。

Denodo 8.0 最重要的功能之一是云中的自动化基础架构管理，可以自动执行与安装、配置、部署和升级 Denodo 平台群集有关的所有任务。这项功能最初面向 AWS（后续即将支持 Azure），通过新的 Solution Manager 提供。Solution Manager 现在提供基于 Web 的 UI，可用于定义和配置集群，给出 TLS 配置、负载平衡、自动缩放等方面的用户首选项。用户使用这套 UI 定义集群后，只需按下“开始”，集群就会自动安装和创建。这套 UI 还集成有监控功能。



Denodo 平台 8.0 通过灵活的交付选项（REST、SOAP、OData、用于文档的 OpenAPI）提供对数据服务的高级支持，能够以多种格式（XML、JSON、HTML、RSS）公开数据，并且支持最新的安全协议（OAuth、JSON Web 令牌、SAML、Kerberos、HTTPS、HTTP 基本摘要身份验证或 WS-Security）。在这一方面，Denodo 平台 8.0 还增加了对增长最为迅速的数据服务标准之一 GraphQL 的支持。GraphQL 是 API 的一种查询语言，可以简化多个 REST 端点的查询。在 Denodo 平台 8.0 中，它可以利用 Denodo 高级查询优化引擎，无需任何代码即可实现简化，并且性能大大提升。

为了进一步方便数据科学家通过 Denodo 平台访问数据，Denodo 平台 8.0 推出一个基于 Apache Zeppelin 的新工具 Denodo Notebook。Denodo Notebook 与包括 SSO 支持在内的 Denodo 安全系统完全集成，因此，数据科学家使用 Denodo Notebook 时，Denodo 平台层定义的所有安全和数据治理策略都会得到实施。

对于企业用户，Denodo 数据目录采用了一项新设计，与其他 Denodo 平台工具统一外观，以实现更加整体化的体验。现在目录增加了新功能，例如自动推荐、通过背书和评论等方式加强协作以及增强的分析和搜索功能（智能排序）。

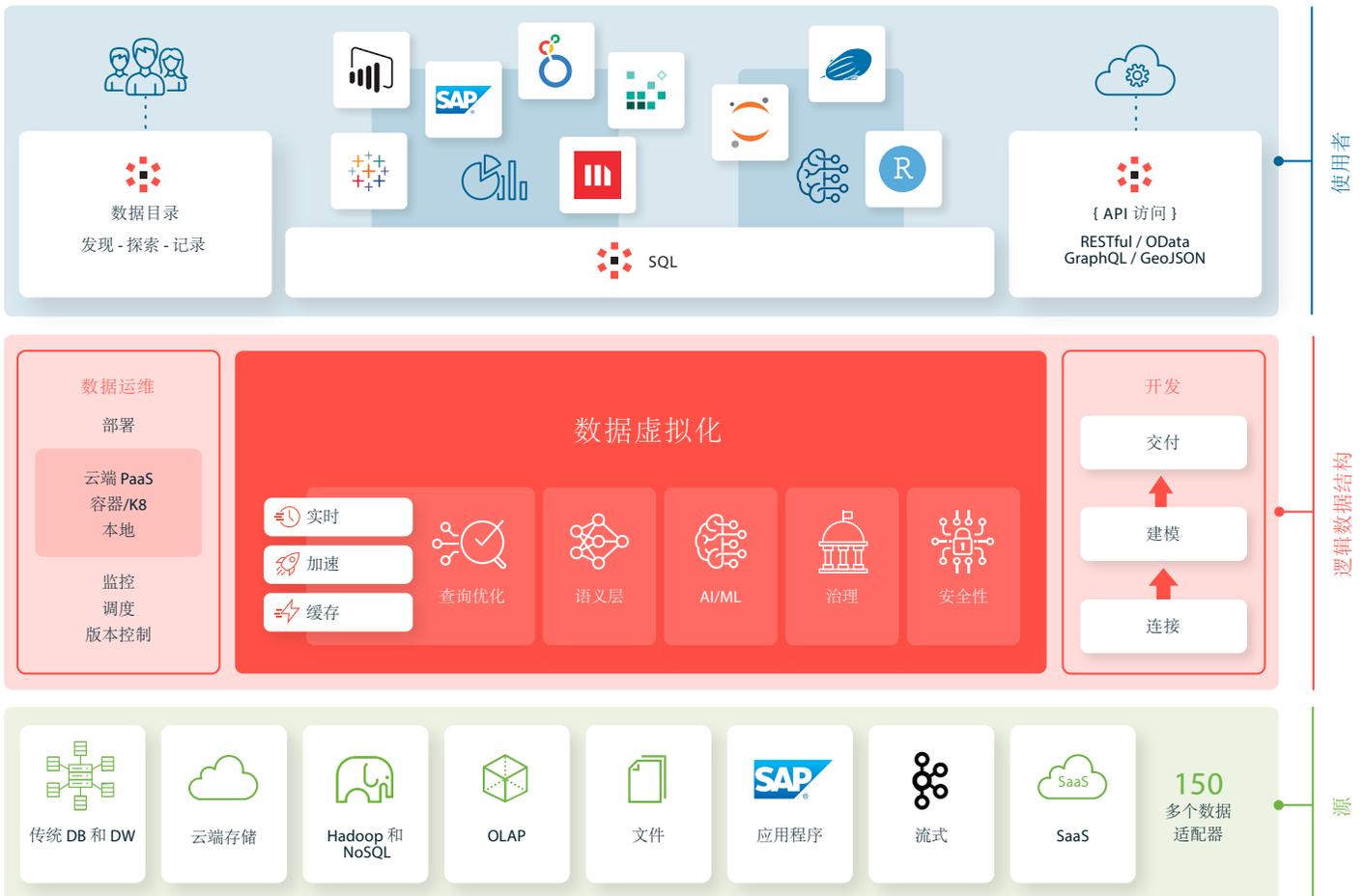


图 1 Denodo 平台 8.0 架构

Denodo 平台 8.0 的主要功能可总结如下：



面向所有支持 SSO 的 Denodo 工具提供完全基于 Web 的界面：涵盖所有工具的基于 Web 的整体化体验。



GraphQL 支持：零代码创建 GraphQL 数据 API，利用 Denodo 查询优化器获得一流性能。



面向开发人员提供基于 Web 的 Design Studio 工具，在数据服务开发过程的所有步骤中均可轻松使用。



“Apache Zeppelin for Denodo” Notebook：数据科学家可以构建将查询、代码和文本整合于一体的叙述，协助进行数据分析，帮助他们解释工作并与同事共享。



面向分析的智能查询加速：预先计算部分聚合（“摘要”），加快未来的查询速度。Denodo 平台 8.0 为所有数据源和使用者提供这种加速机制。



重新设计的数据目录，具有自动推荐以及增强协作、分析和搜索功能



面向云端的自动化基础架构管理：PaaS 支持，包括群集配置（TLS、负载平衡、自动缩放等）、启动/停止控件、自动安装更新以及集成监控。

商业效益

经济实惠

可靠集成数据的同时，所需时间、成本、基础架构投资和刚性均只有提取、转换和加载 (ETL) 流程等传统数据集成方法的一小部分。

更快产生价值

更快提供上下文相关的可靠信息，产生具有可操作性的见解。敏捷的企业更有可能成为行业领导者。

不受位置影响

本地、云端或边缘，任何位置均可部署，不影响规模或数据治理功能。

方便业务使用

消除业务中现代数据生态系统的复杂性（多来源、多格式、分布式、异构、多样化）。以正确的格式公开数据，使用各类用户和应用所需的命名约定，成本几近于零。快速适应需求变化。

企业级

支持多业务部门、多应用场景、多角色和数以千计的用户。

数字市场

建立数字市场，帮助分析人员社区快速找到和使用信息资产，这在自助式分析时代至关重要。

Denodo 平台 8.0 功能手册

数据源

关系数据库

- 通用 (JDBC)
- IBM DB2 (JDBC): 8、9、10、11、12 for LUW, 9、10 for z/OS、AS400
- 多层 Denodo 部署 (JDBC): 5.5、6.0、7.0、8.0
- Apache Derby (JDBC): 10
- Informix (JDBC): 7、12
- MS SQL*Server (JDBC、ODBC): 2000、2005、2008、2008R2、2012、2014、2016、2017
- MySQL (JDBC): 4、5
- Oracle (JDBC): 8i、9i、10g、11g、12c、18c、19c
- Oracle E-Business Suite (JDBC): 12
- PostgreSQL (JDBC): 8、9、10、11

Sybase Adaptive Server Enterprise (JDBC): 12、15

- MS Access (ODBC)

内存数据库

- SAP HANA (JDBC): 1
- Oracle TimesTen (JDBC): 11g
- Oracle 12c In-Memory

并行数据库和设备

- GreenPlum (JDBC): 4.2
- HP Vertica (JDBC): 7、8
- Oracle Exadata (JDBC): X5-2
- ParAccel 8.0.2 (使用 ParAccel 2.5.0.0 JDBC3g 和 SSL 驱动程序)
- Netezza (JDBC): 4.6、5.0、6.0、7.0
- SybaseIQ (JDBC) 12.x、15.x
- Teradata (JDBC): 12、13、14、15
- Yellowbrick

云数据仓库 /RDBMS

- Amazon Redshift (JDBC)
- Amazon Athena (JDBC)
- Amazon Aurora (JDBC)
- Amazon DynamoDB
- Amazon RDS
- Azure Cosmos DB
- Azure SQL Database
- Azure Synapse Analytics
- Delta Lake
- Google Cloud SQL
- Google BigQuery (JDBC)
- MongoDB Atlas
- Snowflake (JDBC)

大数据

- Apache Hive (JDBC): 0.12、1.1.0、1.1.0 for Cloudera、1.2.1 for Hortonworks、2.0.0
- Impala (JDBC): 2.3
- Spark SQL (JDBC): 1.5、1.6
- Presto (JDBC)
- Databricks 2.x

NoSQL

- MongoDB
- Cassandra

多维源

- SAP BW (BAPI/XMLA): 3.x
- SAP BI 7.x (BAPI): 7.x
- Mondrian (XMLA): 3.x
- IBM Cognos TM1
- MS SQL Server Analysis Services 200x
- Essbase (XMLA): 9、11

数据湖存储

- S3
- Azure Data Lake Storage
- Azure Data Lake Storage Gen 2
- Azure Blob Storage
- Google Cloud Storage
- Parquet (分布式文件系统连接器)
- Avro

Web 服务

- SOAP
- REST (XML、RSS、ATOM、JSON)

平面和二进制文件

- CSV、竖线分隔、正则表达式解析
- MS Excel xls 97-2003
- MS Excel xlsx 2007 或以上版本
- MS Access
- XML
- JSON
- SAS 文件 (SAS7BDAT)
- 所有文件均可在本地访问或通过 FTP/SFTP/FTPS 在远程文件系统中访问，提供清晰、压缩和 / 或加密格式。

索引和非结构化内容

- CMS、文件系统、pdf、word、文本、电子邮件服务器、知识库、索引
- Elastic Search 6.4、6.7

采用简化 OAuth 安全协议的云、Saas、Web 源

- Amazon
- Google
- Google Sheets
- Facebook
- LinkedIn
- MS Azure 数据湖
- MS Sharepoint (使用 OData 连接器)
- MS Dynamics 365 Business Central / Customer Engagement
- Marketo
- ServiceNow
- Salesforce (SOQL)
- Twitter, 通过采用简化 OAuth 集成 (1.0、1.0a 和 2.0) 的 API
- Workday

Active Directory 作为源或利用安全功能

- LDAP v3
- Microsoft Active Directory 2003、2008

MS Queues 作为数据源和交付

- MQSeries
- SonicMQ
- ActiveMQ
- Tibco EMS

语义存储库

- 三元组存储 /RDF 中的语义存储库，通过 SPARQL 端点访问。

打包应用程序

- SAP ERP/ECC (BAPIS 和表格)
- Oracle E-Business Suite 12
- Siebel
- SAS (SAS JDBC 驱动程序): 7 或以上版本

面向定制连接器的 Denodo SDK

- CouchDB
- Lotus Domino

大型机

- IMS
 - IBM IMS 原生驱动程序: 8、9
 - IMS Universal 驱动程序: 11

分层数据库

- Adabas (SOA 网关和 Denodo 的 SOAP 连接器): 5、6

以下数据源已使用 JDBC 和 ODBC 驱动程序、WS/SOAP 和 WS/REST 以及 DenodoConnect 适配器 (在此并未逐一罗列) 在 Denodo 中成功完成测试:

- Apache Solr
- Kafka Messages
- Hadoop HBase
- Hadoop HCatalog
- Hadoop HDFS
- IBM BigInsights
- Pivotal HAWQ

发布选项

- 通过 JDBC、ODBC 以及 ADO.NET 进行基于 SQL 的访问
- Web 服务
 - SOAP
 - REST
 - OData
 - Open API (也称 Swagger)
 - GraphQL
- Oauth、OAuth 2.0 (JWT)
- SAML
- SSL
- WS-Security
- 用于消息队列的 JMS 侦听器

- 用于批处理和精简版 ETL 的 Denodo 调度程序

数据目录

- 用于业务用户数据发现和探索的 Web-UI
- 业务类别和业务标签
- 全面的元数据和实际数据搜索功能
- 用于定制数据集的查询向导
- 导出为 CSV、Excel 和 Tableau 数据提取
- 协作功能（保存和共享）
- 在最终发布供全球使用之前，导出到共享沙盒以进行 IT 审核
- 自动个性化推荐和推荐数据集的快捷方式
- 背书数据集、注释、警告等。
- 使用情况统计：何人在何时通过何种方式使用哪些数据
- 分析信息
- 可对结果进行智能排序的智能搜索

性能优化

- 面向分析的智能查询加速
 - 聚合感知、摘要
- 通过集成大规模并行处理 (MPP) 实现查询加速和缓存
- 完全和部分聚合以及联结下推，甚至在联合视图中也能实现
- 支持替代数据源
- 即时数据移动以便优化
- 基于成本的优化（数据统计、数据源索引、数据源执行模型和参数、网络传输率）
- 在联合视图中也可下推选择 / 投影 / 联结 / 分组操作
- 多重联结策略
- 简化分区联合（分区修剪）
- 等等
- 多模式缓存：完全、部分、增量或全部刷新，基于事件或计划，视图级别配置，针对 SaaS 源的增量查询

缓存和数据移动选项

- Amazon Athena
- Azure SQL
- Azure SQL 数据仓库
- Azure Synapse Analytics
- Amazon Redshift
- Databricks 2.x
- Delta Lake
- IBM DB2（8、9、10、11 面向 LUW，9、10、11 面向 z/OS）
- Hive 2.0.0
- Impala
- MS SQL Server (2000、2005、2008、2008R2、2012、2014、2016、2017)
- MySQL（4 和 5）
- Netezza（6 和 7）
- Oracle（8i、9i、10g、11g、12c、12c in-memory、18c、19c）
- Oracle TimesTen 11g

- PostgreSQL（9 和 10）
- Presto
- SAP Hana
- Snowflake
- Spark（1.5、1.6 和 2.x）
- Teradata（12、13、14、15 和 16）
- Vertica（7 和 9）
- Yellowbrick
- 可配置的“通用”适配器，用于其他具有 JDBC 驱动程序数据库

数据管道

- 远程表（通过 UI 或存储过程创建）
- Denodo 调度程序

大规模并行处理选项

- Impala
- Presto
- Spark 1.5、1.6、2.x
- Databricks 2.x

数据治理

- 数据源刷新、变更影响分析、依赖关系树、完整数据沿袭
- Denodo Governance Bridge：与 IBM Information Governance Catalog API 集成，
- 将包括沿袭信息的元数据发布到数据治理工具

安全性

动态数据 - 安全通道

- 使用 SSL/TLS
- 客户端 - Denodo 和 Denodo - 源
- 适用于所有协议（JDBC、ODBC、DO.NET 和 WS）

静态数据 - 安全存储

- 缓存：第三方数据库。可利用自有加密机制
- 交换到磁盘：序列化后临时存储在可配置的文件夹中，文件夹可由操作系统加密

加密 / 解密

- 支持文件和 Web 服务的自定义解密
- 与 RDBM 加密透明集成
- 加密元数据导入 / 导出

基于用户和角色，包括与 AD/LDAP 集成

- 行级别和列级别授权
- 屏蔽
- 用于特定安全约束的自定义策略以及与外部策略服务器的集成

身份验证

- 基于原生和 LDAP/Active Directory 的 Kerberos 和 Windows SSO 支持
- Base64
- Kerberos
- NTLM
- Oauth、OAuth 2.0 (JWT)

- SAML
- 双因素身份验证（通过支持的身份提供者：Okta、Duo 等）
- SSL
- WS-Security
- 传递会话凭据
- 利用现有的源特权

数据建模

- Design Studio：基于 Web 的开发工作室，可用于数据建模（仍提供桌面版本）
- 自底向上和自顶向下（通过界面视图）
- Denodo Model Bridge：集成第三方建模工具
 - ER/Studio Data Architect
 - ERwin Data Modeler
 - IBM InfoSphere Data Architect
 - SAP PowerDesigner

数据质量

- 用于验证、清理、丰富、标准化、匹配和合并数据的转换库、过滤和匹配功能以及质量规则
- 可通过自定义功能扩展
- 集成外部 DQ 工具

监控

- 集成在 Solution Manager 中的 Denodo 诊断和监控工具 (DMT)、Denodo Query Monitor、Denodo Monitor Reports：通过 Denodo 工具提供所有监控信息，用于实时和历史分析。可使用监控信息创建历史仪表盘
- 也可以通过 SNMP 和 JMX 标准进行监控。因此可与大多数领先的系统管理软件包（例如 HP OpenView、Nagios、Zenoss、Osmius、IBM Tivoli 和 Microsoft WinRM）进行交互操作

操作

- Solutions Manager 可自动执行操作和升级任务
 - 集中管理并向客户端分发更新
 - 集中管理许可证密钥
 - 定义升级修订及其依赖项，并在不停机的情况下将其部署到生产群集中
 - 集中管理数据源属性和日志
 - REST API 可自动执行 DevOps 工具（例如 jenkins）的任务
- 集成云基础架构管理 (AWS)
 - 创建和管理集群：定义 EC2 实例类型、EC2 实例数量等。
 - 创建负载均衡器和自动调整组。
 - 安装并启动 Denodo 服务器。
 - 不停机的情况下，在具有多个服务器的群集上安装 Denodo 更新
 - 在 Denodo 服务器中启用 SSL。

- 具有版本控制集成的多用户开发
 - Subversion
 - Microsoft TFS
 - Git
- 资源管理器通过可为每个应用优化资源利用率的方式限制资源并分配给每个会话、角色或用户
 - 更改资源优先级
 - 实施限制超时或限制行数
 - 添加每分钟/每天/每月的每日配额
例如每天仅限 50 次查询

部署模式

- 本地、私有云、公有云
 - 基本的单一服务器配置
 - 具有负载均衡的 HA 群集（主动 - 被动和主动 - 主动）
 - 共享或分布式本地缓存
 - 地理位置分散的服务器环境
 - 对等或多层的多个 Denodo 实例
 - 通过 Docker 提供容器化支持
- 公有云
 - Denodo Platform for AWS
 - Denodo Platform for Azure
 - Denodo Platform for GCP
 - 在 AWS 和 Azure 中均支持自动调整

用户界面

- 中央 Web 控制台：集成对所有支持 SSO（Kerberos、SAML、OpenID 和 OAuth）的 Denodo UI 的访问
- Solution Manager 基于 Web 的 UI：管理员可使用集中式 UI 管理部署和升级，包括自动管理云基础架构（AWS）
- Design Studio：基于 Web 的 Development Studio，拖放式低代码开发人员工作室，适合数据导向的开发人员，例如数据工程师、高级用户和大众集成者；只需几下点击即可发布数据服务。
- Desktop Dev.Studio（VDP 管理工具）
- 数据目录基于 Web 的 UI：易于使用的界面，面向业务导向型用户，例如数据管理员、数据分析师和大众分析师。
- 诊断和监控工具：基于 Web 的集中式 UI，用于数据工程师和管理员进行监控、审核和故障排除。

操作系统

- Microsoft Windows（32 位和 64 位平台）：Windows Server 2019、Windows Server 2016、Windows

- Server 2012、Windows Server 2008、Windows 10、Windows 8.1 和 Windows 7
- Linux（32 位和 64 位平台）：Red Hat Enterprise Linux (RHEL) 6 和 7、Oracle Linux 6 和 7、Ubuntu 12.04 LTS 和 14.04 LTS、CentOS 6 和 7
- UNIX（64 位平台）：Sun Solaris
- 任何兼容 Java 11 或以上版本的操作系统

最低硬件要求

- 处理器：Intel Xeon 四核或同等产品。高负载方案或复杂计算场景可能需要 8 个或以上核心。
- 物理内存 (RAM)：16 GB 内存，这样 Denodo 服务器可分配最多 8 GB 运行时堆空间。
- 磁盘空间：最低：5 GB，推荐：100 GB。Denodo 仅需大约 1 GB 磁盘空间。如果将高速缓存安装在同一服务器上，则需要更多磁盘空间。

