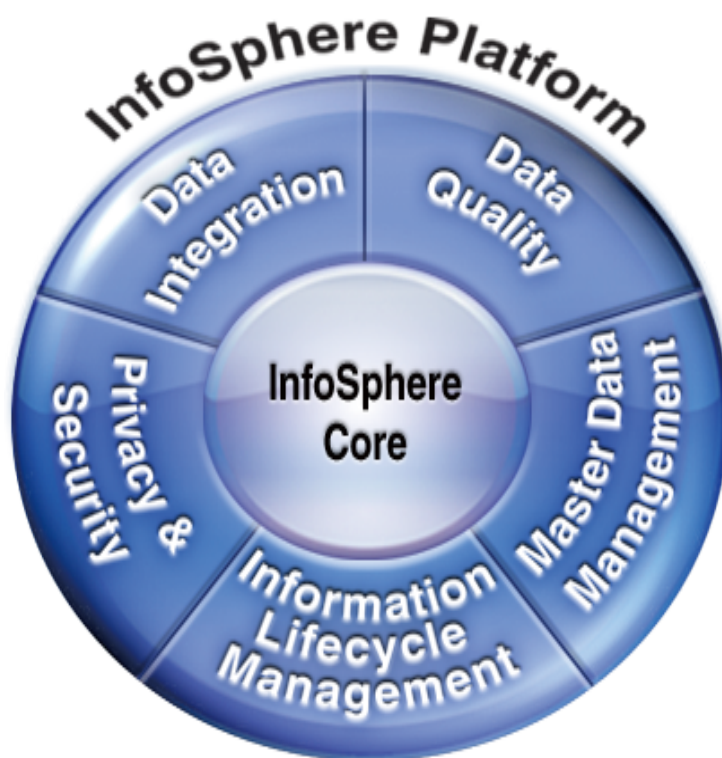


IBM InfoSphere 信息整合平台

制作可依赖的信息



无论是金融业、医疗业、制造业、零售业，甚至政府机关或公共事业等，各行各业的信息使用量均相当庞大。随着数据量日益剧增，企业决策比以往更需要具备强大的信息管理和掌控能力，协助解决关键问题。若要提升企业竞争优势，必须随时掌握企业内的信息，并透过这些信息获得新的企业价值。除非这些信息兼顾实时、安全及完整一致等特点，否则这些决策便会存在高风险。企业数据通常分散在各部门的数据区内，而且不提供分享，因此无法得知哪些才是最新最正确的数据。若要有效运用企业收集和建立的数据，必须学会如何管理、整合、分析及标准化。企业信息必须完整一致并且准确。有了信息整合方案和技术，就能获取各种来源数据，并于统一整理后适时分发到公司内外部各个目的位置。信息整合可协助公司执行多项重要任务，包括将多重来源数据加载仓储、合并应用程序以及连接各部门信息。无论企业信息存在何处，只要将资料视为单一资源，就能迅速处理信息、降低停机时间、减少客户服务问题，并在分发信息时将效能影响降至最低。与此同时，每天有越来越多的数据(例如社交媒体、各种传感器检测数据、手机GPS信号等)产生出来，这些数据被业界称为大数据。大数据蕴含着丰富的有价值的商业信息，具有大容量(Volume)、高速度(velocity)和多样性(variety)的特点。IBM大数据平台提供先进的大数据分析技术，提高客户从大数据中挖掘商业洞察的能力。



IBM Information Management下的InfoSphere产品线提供一整套的信息整合解决方案可复制，整合并转换数据，并且让您一览整合、完整、一致和标准化的企业数据。通过提供可信赖的信息，帮助企业制定及时正确的决策。InfoSphere BigData产品通过对大数据volumes, variety和velocity三方面的良好支持，为新型的流数据(比如社交媒体数据，Web点击流数据等)提供方便的Map/Reduce计算和分析平台。

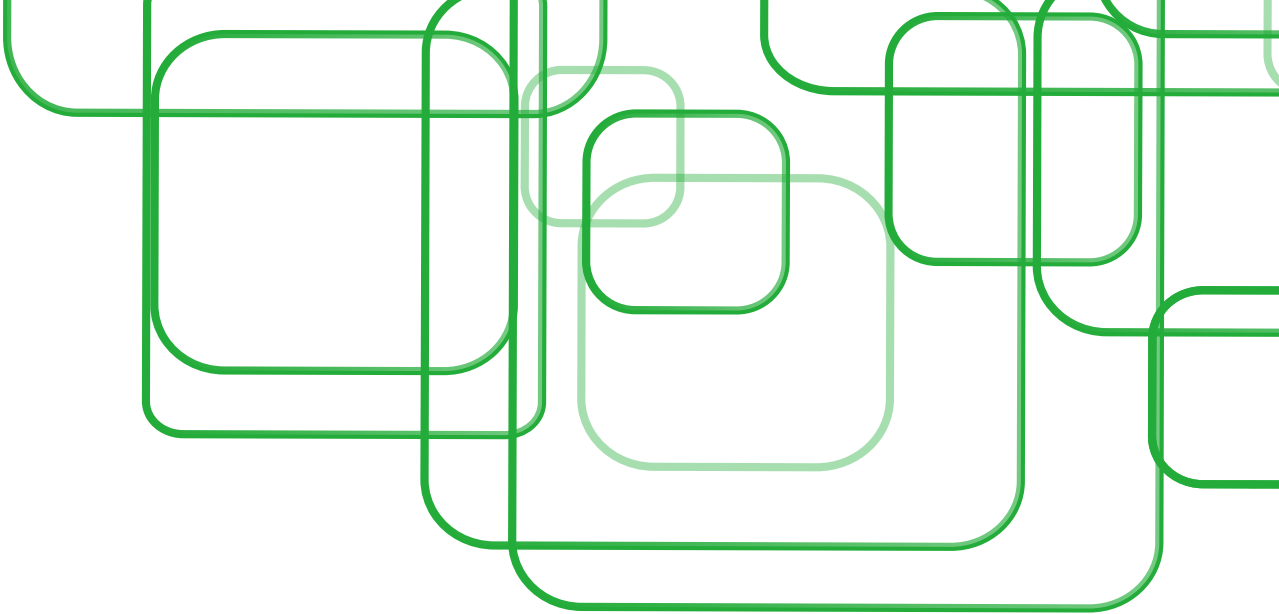
资源:

信息集成与主数据管理

<http://www-01.ibm.com/software/data/infosphere/>

大数据

<http://bigdatauniversity.com/>



产品名录

信息集成

IBM InfoSphere Information Server介绍	1
IBM InfoSphere Blueprint Director产品介绍	2
IBM InfoSphere Business Glossary产品介绍	4
IBM InfoSphere Discovery产品介绍	6
IBM InfoSphere Information Analyzer产品介绍	8
IBM InfoSphere FastTrack产品介绍	10
IBM InfoSphere Metadata Workbench产品介绍	12
IBM InfoSphere QualityStage产品介绍	14
IBM InfoSphere DataStage产品介绍	16
IBM InfoSphere Information Services Director产品介绍	18

数据复制

IBM InfoSphere Data Replication产品Q复制介绍	20
IBM InfoSphere Data Replication产品CDC介绍	22

主数据管理

IBM InfoSphere Master Data Management产品介绍	24
---	----

大数据

IBM InfoSphere BigInsights产品介绍	26
IBM InfoSphere Streams产品介绍	28

IBM InfoSphere Information Server介绍

IBM InfoSphere Information Server是业界领先的企业级信息集成平台。它能够帮助客户理解异构系统中的各种复杂信息，并且通过清洗和转换生成一致、完整的可信赖信息，最后将可信赖信息以各种方式交付给各种业务系统(如业务分析系统、数据仓库、主数据管理、应用程序整合和迁移等)，从而帮助客户从已有信息中获得更多价值。

功能介绍

InfoSphere Information Server由许多产品模块组成，用户可以单独订购产品模块或套件组件，用于解决各种类型的业务问题。各个产品模块可以一起提供给用户一个集成的端到端解决方案，帮助用户完成信息集成的各个阶段工作。并且各个产品模块使用了统一的元数据存储机制，这使得不同角色间(如业务人员和IT人员)的协作更加容易，从而在整体上提高信息集成的效率。InfoSphere Information Server提供了以下四个核心的集成功能：

- 了解数据

InfoSphere Information Server 可以帮助用户自动发现和定义信息内容及结构并对其进行建模，以及了解和分析信息的含义、关系和谱系。

- 清洗信息

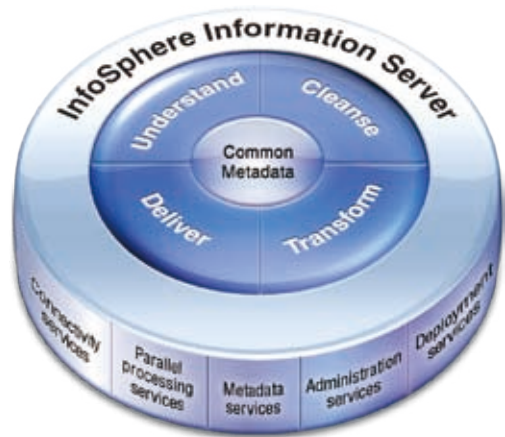
InfoSphere Information Server通过对数据进行标准化、校验、匹配和合并操作，确保数据的质量和一致性。

- 将数据变换为信息

InfoSphere Information Server能够帮助您变换并充实信息，确保它处于适当的上下文中以投入新的使用。数以百计预构建的变换功能可用来组合、重构和聚集信息。InfoSphere Information Server 还提供大量复杂的数据变换和移动功能，这些功能可用于独立的“抽取-转换-转载”(ETL)场景，或用于业务流程中的实时数据处理。

- 交付信息

InfoSphere Information Server能够将信息以服务的方式提供给需要它的人员、业务流程或者应用程序。它也可以通过使用基于时间或基于事件的处理来传递信息、或将大批量信息从一个位置移动到另一个位置。



客户价值

- 为企业信息架构提供全面的统一的基础，可以根据数据量和处理需求扩展
- 确保可审计的数据质量，这是在整个企业中提供可信信息的基础
- 元数据驱动的集成，这会大大提高集成和丰富信息的生产力和灵活性
- 一致的可重用的信息服务，以及应用程序服务和流程服务，这是企业必备的
- 通过可靠的符合行业标准的解决方案和专门技术加快产生价值的速度
- 最广泛、最深入的连接能力，可以访问各种数据源中的信息：结构化、非结构化、大型机和应用程序

产品模块



Business Glossary



Discovery



Information Analyzer



FastTrack



Blueprint Director



Metadata Workbench



DataStage



QualityStage



Information Service Director

IBM InfoSphere Blueprint Director产品介绍

InfoSphere™ Blueprint Director实现信息系统构架解决方案的一致性和联通性。它将概要和详细设计文档连接到您的整体信息架构解决方案,让所有团队成员了解项目的进展。

基本功能

- 创建项目的蓝图, 让您可以将蓝图和元素相关联。例如, 您可以关联InfoSphere Metadata Workbench中所引用的元数据。同样, 您可以使用以下连接数据模型的元素:
连接器(由InfoSphere Architect定义)
ETL(由InfoSphere DataStage定义)
度量 and 维度模型(由Cognos Framework Manager产生)
网址, 或任何文件和其相关的应用程序
- 蓝图元素连接元数据的存储库, 文件, 并通过URL访问资产
- 使用业务术语表的条款和术语概念模型, 创建和标记蓝图元素
- 提供预建的模板, 如商业智能(BI)和数据仓储领域定义的蓝图

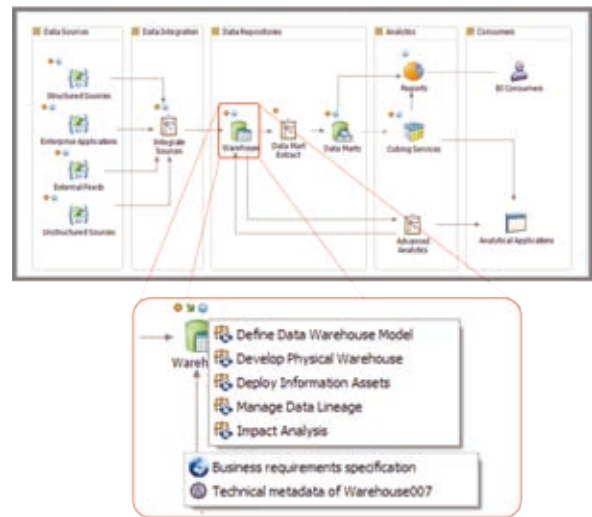
平台支持

- Windows

主要特点

- 基于模板定义项目愿景/建筑(参考架构)
- 方便快捷地在界面中管理各种方法和状态
- 在界面中采取行动, 可直接浏览到相应的工具并开始工作

产品展示图



- 创建集成项目的整体视图, 由团队定义和管理终端到终端的信息流
- 连接蓝图元素, 提高可预见性, 确保项目成功。蓝图元素如下:
参考架构
可重复使用的最佳实践和方法论
业务和技术构件
- 提供控制和洞察力的信息路线图, 实现一个合作项目生命周期的演进(愿景, 执行, 完成)的管理

重要用途

在各种各样的数据整合项目中实现高效的管理任务

- 里程碑及时间轴管理
- 集成任务管理
- 集成计划管理

IBM InfoSphere Blueprint Director 典型案例

项目背景

某一石油化工集团，为了迎接新形势下行业改革与发展的挑战，提出了“基于市场做出更准确的决策，改善用户满意度，为客户提供更好的服务”等目标，并确立了数据整合及分析改善建设任务。

由于项目比较庞大，涉及多个业务部门的协作，项目整体的状态和子项目工作(包括细节任务，进度，最佳实践分享等)如何有效关联成为挑战。

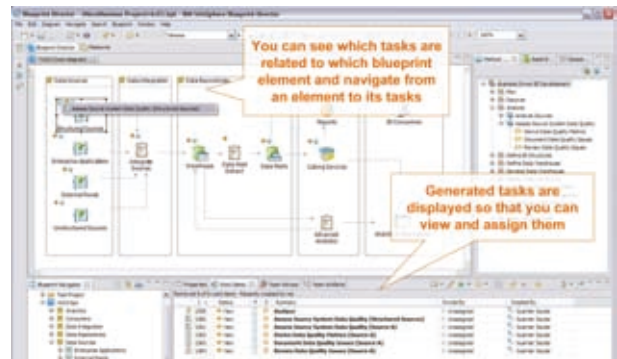
项目挑战

- 项目比较庞大，涉及多个业务部门
- 缺乏共享机制，各个业务部门信息沟通不畅，项目进度协调较为困难
- 项目在实施过程中，会有需求的变化，因此项目的模块需要调整。跟踪管理项目实施过程中的设计变化非常必要
- 该项目中涉及多个模块，模块中涉及多个任务。由于任务较多，有些任务有关联性，因此集中管理这些任务非常困难

解决方案

- 公司高层非常重视，由多位领导亲自挂帅，分别组建了项目领导小组，项目管理小组，应用功能组，数据准备组，技术开发组以推动项目的顺利实施。
- 引入InfoSphere Blueprint Director管理各个项目的进度状态，协调各个任务的优先级关系。
- 各子项目负责人定期及时更新Blueprint Director中相关任务的状态。

Blueprint任务管理



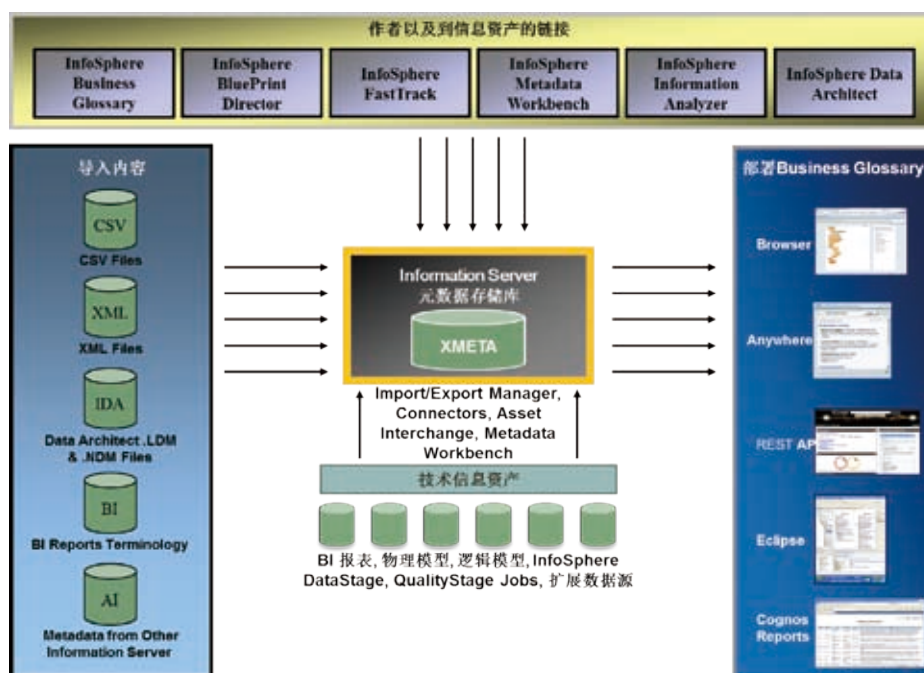
方案收益

- 针对各个项目模块，各个业务部门之间的沟通变得顺畅。
- 将项目，模块，任务整合在一个系统里管理，方便了项目参与者对整体项目进度的了解。
- 节省了项目管理的成本，缩短了项目实施的周期。

IBM InfoSphere Business Glossary产品介绍

InfoSphere™ Business Glossary是IBM®的一款基于Web的元数据管理工具，它支持用户创建、管理和共享企业词汇表和分类系统。

Business Glossary架构



基本功能

- 提供通用的可消费的词汇，以便整个企业的清晰交流
- 提供业务术语和其所分配的IT资产之间的链接，以便了解IT上下文和业务连接点
- 为企业各角色提供一个协作环境，以便丰富词汇、维护词汇
- 给业务术语项及其IT资产分配数据管理员，以便分配和发行领域知识
- 提供可定制的、可发行的业务术语集(包括所有相关的业务属性)，以便整个企业的知识转移及重复利用

平台支持

- AIX
- HP-UX
- Solaris
- Windows
- Red Hat on AMD/Intel/System z
- SUSE on AMD/Intel/System z

产品模块

- InfoSphere Business Glossary Browser建立和管理业务词汇和关系。
- InfoSphere Business Glossary Anywhere轻量级桌面客户端，保持当前应用的同时实现词汇表的快速检索。
- REST API web service在自定义应用、门户网站、mashups和工作流系统中集成词汇表内容。
- Business Glossary Eclipse plugin – IDA使数据建模工作与业务目标保持一致。
- Business Glossary Eclipse plugin – UML确保IT项目都是由业务目标驱动的。

项目背景

某公司信息办公室需要改善部门间通讯并且建立一个数据治理方案，在整个企业间建立标准一致的业务术语。这种治理方案有助于发现数据存储整合机会，执行影响分析变化，并为商业智能及数据管理工作发现最佳数据源。

项目挑战

- 在企业范围内定义标准规范的业务术语
企业各部门对同一个业务术语理解差异造成商业智能报表数据不一致及数据来源错误问题。利用InfoSphere Business Glossary定义标准业务术语库，为企业个角色提供协作环境。
- 如何利用标准业务术语设计数据模型及管理业务和数据模型的关系。
模型设计人员在InfoSphere Data Architect中可以将业务术语下载到本地，并通过托拽等简单操作，生成数据模型同时建立业务术语与数据模型间分配关系。最后通过简单的指令将此分配关系上传到InfoSphere Information Server。
- 随时快速检索企业元数据
InfoSphere Business Glossary Anywhere是轻量级的桌面应用程序，点击任何基于文本的应用程序上的文字触发热键，通过该工具检索企业元数据视图。

解决方案

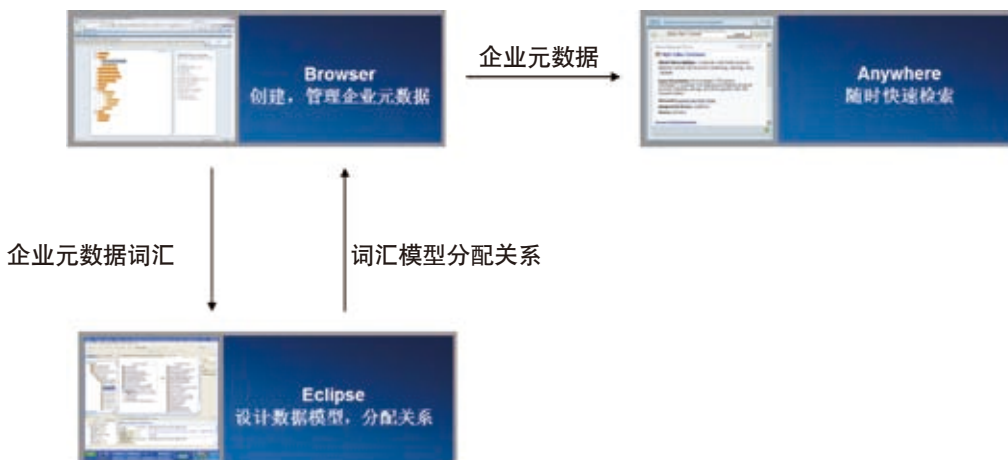
IBM® InfoSphere®解决方案在整个企业范围内定义了标准一致的业务术语并提供便捷的元数据和业务术语访问方式。利用InfoSphere Business Glossary在整个企业范围内定义业务术语。之后，根据分而治之的原则，各项目组将业务术语分配给物理数据实体(表或列等)，再通过简单的命令行指令将此分配关系上传至InfoSphere Information Server。Business Glossary Eclipse plugin使得设计人员可以在InfoSphere Data Architect中通过托拽等操作轻松根据业务术语定义设计数据模型及其物理实现，并建立业务术语和物理实现之间的分配关系。这种分配关系，可以用来在整个企业范围内确立数据整合机会并且支持变更影响分析。这种企业元数据视图还可以在Business Glossary Anywhere中检索和查看，点击任何基于文本的应用程序上的文字都可以显示该词汇相关的业务术语及元数据信息。

方案收益

有助于系统整合和精简应用组合以降低开发和维护成本。

项目的目标是要改善支持服务。一致的用户界面和数据质量是至关重要的。InfoSphere技术解决方案使团队可以巩固数据存储，提高数据质量，建立更好的系统。更快地获得技术支持，工作人员可以帮助客户减少停机时间，在全球范围内降低成本和提高生产率。

部署架构



IBM InfoSphere Discovery产品介绍

InfoSphere Discovery提供了全方位的自动化分析过程，代表了新一代的启发式数据分析技术，能够以高出10倍的效率完成分析过程，帮助用户节约成本。比如，自动化单数据源主外键发现、业务对象发现，跨数据源的重叠分析，匹配密钥的发现，原型设计，测试数据整合以及自动发现业务转换规则等。

基本功能

- **单一源分析:** 检查数据质量和详细地理解每一个源的结构和格式，主外键自动发现分析，数据对象分析。
- 实现一致的数据源间数据比较的基线。
- **重叠和关键数据元素(CDE)分析:** 跨数据源发现冗余字段和唯一字段。提供不同字段之间的重叠数据值，特定字段之间的重叠。
- **创建统一数据模型(USB):** 接收重叠分析的结果作为输入，帮助确定哪些数据将被合并到数据移植、主数据管理或数据仓库等。
- **匹配键分析:** 自动化分析确定用于在多个数据源中排列记录的匹配键属性。
- **冲突分析:** 跨数据源的值发生冲突时，自动生成基于统计分析分配的信任规则
- **转换/业务规则发现:** 通过分析两个数据源的数据值和模式自动发现复杂的跨多数据源转换和业务规则。

数据库支持

- Datacom*
- IBM® DB2® for LUW
- IBM® DB2® for z/OS®
- IBM® Informix®
- Netezza®
- IMS™*
- Microsoft™ SQL Server
- Optim™ Archive Files
- Oracle
- Teradata
- VSAM* on z/OS®
- Sybase
- CA IDMS

关键应用

- **数据增长管理**
自动化发现数据的完整性和业务对象
- **数据整合，集成和迁移**
自动发现数据源之间的转换和业务逻辑
跨数据源的原型设计
- **资料隐私**
发现隐藏的敏感数据



什么是独特的?
分析数据值和模式，并产生可操作的结果
发现数据源之间的内在复杂关系
在同行业中最大的自动化水平的专利方法

- Cross-Profiler
- Unified Schema Builder
- Transformation Analyzer

客户价值

Discovery 能够加速许多 IT 项目的部署，其中包括:

- 数据存档
- 应用报废
- 应用合并
- 主数据管理
- 敏感数据发现和隐蔽
- 数据沿袭发现和记录
- 数据仓库

平台支持

- Windows xp, Windows 7
- Windows Server 2003, 2008

IBM InfoSphere Discovery产品介绍

典型案例

项目背景

某International Truck Manufacturer希望整合多个财务系统，建立唯一的主数据管理系统共有来自14家客户，9家供应商和6种材料数据源需要被整合。数据的复杂性决定了，数据的手动分析会给项目带来风险，该公司希望项目实现以下目标：

- 改善客户体验
- 减少金融系统的数量
- 为多个大型迁移项目，一次性改善地址数据的质量

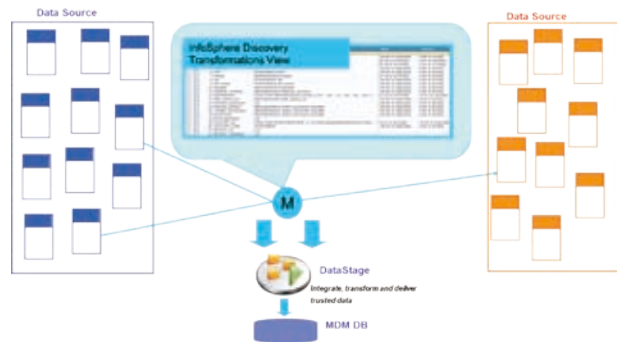
项目挑战

- 至少需要6个月，即使只针对每两个系统作交叉分析
- 数据和文件往往存在不正确，质量参差不齐
- 每个分析的属性涉及3-7业务和数据分析
- 手动跨系统的组合查询容易出错
- 针对跨系统分析的QA流程，几乎不存在

解决方案

- 部署Discovery,执行跨系统的数据分析和数据映射，替代原有的多数手动分析，减少项目风险
- 加速信息转型项目的部署
 - 提高精确度，可预测和可重用的能力
 - 加速数据分析工作，最多可以提高10倍的效率
 - 减少SME参与的时间和精力

部署架构



方案收益

- 12个月的分析工作只用1个月完成
- 增加项目的可预见性，降低风险
- 在对业务规则和关系的假设作决定之前，可以先对数百万行的数据进行测试和验证
- 一个人可以在一个小时内回答，曾经需要三个或更多的人，几天来回答的问题
- 项目只需要更少的SME，并且花费更少的时间用于SME讨论和作决定

IBM InfoSphere Information Analyzer产品介绍

InfoSphere™ Information Analyzer是IBM®提供的一款数据质量分析工具。它提供了数据质量评估，数据质量监控以及灵活的数据规则设计和分析等功能。它能帮助您更快地交付可信任的信息。

基本功能

- **全面的数据分析:** 从各个角度为您提供数据的分析图表；允许用户随时记录所有被发现的异常数据特征；帮助用户定位非法数据。
- **实时的数据查验:** 允许用户实时地查验数据分析结果集中的任意记录。
- **集成的规则分析:** 提供了可持续的开发，部署和评估关键数据规则的能力；支持规则，记录以及数据源等不同粒度的校验能力。
- **基线分析:** 通过对比同一份数据在不同时期的分析结果，可以帮助用户方便地找出在基准版本之后发生的数据结构或内容变化。
- **数据处理量的高可扩展性**
- **统一的元数据管理:** 分析结果可以在所有Information Server家族的产品之间方便地自动共享。
- **广泛的数据源支持**
- **广泛的报告格式支持:** 支持多达80余种报告类型。可视化的分析结果、趋势及指标信息可以帮助您快速高效地理解您的数据。

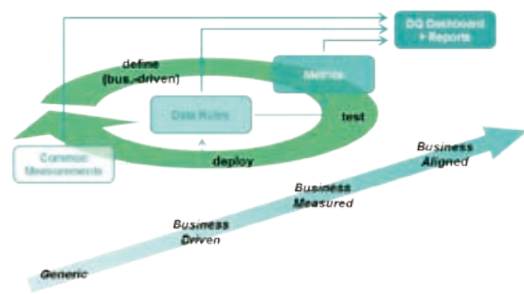
平台支持

- AIX
- HP-UX
- Solari
- Windows
- Red Hat on AMD/Intel/System z
- SUSE on AMD/Intel/System z

数据源支持

- IBM DB2®
- IBM Informix®
- Oracle
- Microsoft® SQL Server
- Sybase
- Teradata
- Microsoft Access
- 其他数据源(例如，文本文件等)

关键模块



分析: 起始于业务需求，关注于业务价值，结束于业务均衡。

- **标准评估**
为您数据源的结构、内容和质量建立一个全面、整体的认知。
- **数据规则**
通过定制并不断地调整自定义数据质量规则来对您的数据进行更深入的质量验证，趋势预测和模式分析。
- **报告指标**
通过对分析结果的鉴别、评估以及异常管理来限制数据质量的恶化，从而降低风险。

客户价值

- **您对您的数据有信心吗？结构、内容、质量？**
IA为您提供了数据分析、数据评估以及数据监控等全方位服务。您可以从结果报告中轻松的找出所有异常信息，并可实时进行数据查验。
- **您对多样的异种数据源以及数以G计的庞大数据记录集的数据规则分析感到迷茫吗？**
您可以在IA中随意定制基于您业务需求的数据分析规则(集)，并可方便地重用于各个数据源以及其他Information Server产品中，同时Information Server产品独特的可扩展并行架构可以为您提供海量数据下的性能保证。
- **您的生产系统中缺乏相应的数据校验和审计流程吗？**
IA提供了持续不断的数据审计能力，您可以在任何需要的环节使用它。
- **您想将您的数据审计结果应用到其他的生产程序中吗？**
IA支持广泛的审计报告输出格式，您可以方便地将它共享给其他角色。

IBM InfoSphere Information Analyzer产品介绍

典型案例

项目背景

2013年，某零售巨头多年来第一次发生亏损。该企业通过一系列的研究与自我反思，认为其主要是由以下几方面原因引起的。

- 缺乏对库存商品销售数据的有效洞察和糟糕的商品推销以及预测应用系统问题，致使其无法及时地调整出货品种及推销策略来改善情况。
- 过长的生产交易期以及已有的大量生产合同制约，公司历史遗留的供应链等问题，致使其便已经深刻了解到这些问题，也不能迅速地改变他们的产品线结构。
- 其他诸如PeopleSoft财务系统问题和当前不合理的数据仓库设计等问题，致使其无法快速高效地整合统一所有资源。

项目挑战

- 信息太杂太乱，不知道哪些是重要的
无法及时找到需求点来改善供应链
无法通过客户行为分析来指导卖场活动
无法分析有价值的非结构化数据
- 事实资料版本太多，真相难以提取
难以理顺客户、产品以及合作伙伴的关系
缺乏透明统一的数据记录规则
- 缺乏可信任的信息
不完备的，过时的，不正确的数据泛滥
难以理解并管理信息的使用方式
- 缺乏数据敏捷性
难以利用创新的机遇
因死板的系统和变化的需求而导致的不断增加的成本开销

解决方案

Information Analyzer是IBM推出的一款数据质量分析工具。它能帮助您方便快捷地全面理解您的数据。它提供了数据质量评估，数据质量监控以及可定制的数据规则设计和分析功能。这些强大的能力能让您轻易的掌控整个企业的杂乱数据，并极大地加速您的信息整合工程。因此本产品可以很好地满足该企业的需求

并加速实现其转型目标。

在IBM为企业制定的全套解决方案中，Information Analyzer主要应用于以下3个关键阶段：

- 源数据分析
全面分析理解该企业各个数据库系统的数据质量，找出异常数据，为随后的数据清洗、转换、整合提供有力的支持。
- 业务洞察及趋势预测
通过对清洗过的历史数据进行高效的比对，找出数据变化趋势，分析客户行为，从而及时调整相应的推销策略。
- 各个环节之间的数据质量审计
自动按需审计各个接口的数据质量，从而保证最低的脏数据率和最高的数据兼容性。

部署架构



方案收益

- 通过列分析、主/外键分析、域交叉分析等手段自动分析您的数据，快速完成整体分析并生成数据的结构、内容和质量报告。
- 通过基线分析便捷地找出数据的变化部分，以此来研究业务行为的变化趋势。
- 通过定制的数据规则分析，筛选或预测出有业务价值的数形态。
- 降低数据分析校验环节所需要的时间使业务决策更及时加速整个数据转换，整合过程
- 与Information Server其他产品相互协作，保证数据的唯一性，准确性和完备性。

IBM InfoSphere FastTrack产品介绍

IBM InfoSphere FastTrack可加速业务需求向数据集成项目的转换。它在数据集成项目中可以帮助业务分析员、数据建模者和开发者实现有效协作，从而加速各种用户角色、产品、布局和工作流程之间的协作开发。

基本功能

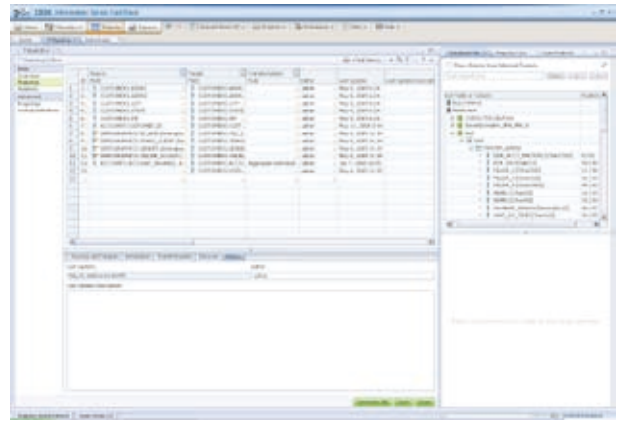
- 可实现跨越用户角色、产品和地区的协作开发。
- 易于使用的导入/导出功能允许业务用户利用熟悉的 Microsoft Excel界面创建新的规范。
- 使用可自定义的发现算法，实现“源到目标”的列的智能匹配。
- 根据业务需求自动生成ETL作业(InfoSphere DataStage)。
- 利用已发布的 InfoSphere Information Analyzer中的分析结果来定义了大多数完整的业务规范。
- 作为映射和文档流程的一部分，创建 InfoSphere Business Glossary术语及其关系，并链接到相应的物理数据库列中。
- 为跟踪和审计数据整合项目的规范提供集中化的位置。

客户价值

在传统的系统集成项目中，用户会碰到以下问题：

- 开发时间没有主要用在编程或设计上，而是用在了不同角色之间的来回沟通上。项目无法跟上进度是因为耗费了过多的时间来澄清业务需求，和解释创建源到目标映射所使用的方式。通常，时间耗费在澄清术语、修改不完整和不清楚的规范时所需的来回通信上。
- 低效的手动协作流程。在典型的工作流程中，数据建模者对信息源和目标进行建模，然后将这些结构发送给业务分析员。业务分析员通过创建电子表格来手动定义源到目标的映射，并且将其打印出来发给开发者。这些劳动密集型的流程缺乏自动化和审计跟踪功能，会使项目速度迟缓，增加了产生错误的可能性，失去了创建可重用项目组件(这种组件可存储在公共存储库中，在下游过程中由其他成员使用。例如表定义，字段定义等)的机会。
- InfoSphere FastTrack能帮助业务分析员、数据建模者和开发者实现有效协作，从而加速各用户角色、产品、布局和工作流程之间的协作开发。

产品展示



- 使用InfoSphere FastTrack，用户不需要从各种项目文档之间复制和粘贴用于设计、复审和分派规范的信息。InfoSphere FastTrack能帮助管理这些信息(例如，映射中源和目标之间的数据转换)，使开发者能够将重点关注到更复杂的逻辑实现上。InfoSphere FastTrack 将各种源中的设计元素变换为便于其他人理解和利用的详细信息。这些信息都被链接在一起，可在IBM InfoSphere Information Server 的元数据存储库中访问，而不是以电子表格形式存储在文件系统中。
- InfoSphere FastTrack有助于将业务需求转换为业务应用程序。通过利用元数据集成，InfoSphere FastTrack能够发现表列关系、将列与业务词汇表中的术语链接，并生成用于复杂数据转换的DataStage初始作业。源到目标映射可包含用于定义如何构建应用程序的数据值变换逻辑(作为规范的一部分)。

IBM InfoSphere FastTrack产品介绍

典型案例

项目背景

某全球性的食品公司，它的数千名员工分布在世界各地的数十个工厂和办公场所中。公司在全球的每个工厂和办公场所都有自己的IT系统和业务流程。这样带来了以下的问题：各个工厂和办公场所不能进行有效的沟通，并且与各地的客户和供应商也缺乏有效的沟通。这个已经变成了一个制约公司发展的一个致命因素。因此，该公司开始了一个长期的SAP项目致力于整合全球所有的业务流程到一个核心的SAP平台，从而实现对全球业务的有效管理。

在进行该SAP项目过程中，公司需要一个全面的数据迁移和集成方案，安全精确的把分布在全球个工厂和办公场所的应用数据转换到SAP平台。

项目挑战

● 复杂的IT架构带来数据迁移的难度

全球各个工厂和办公场所拥有独立的IT架构和业务流程，很多应用都运行在不同的平台和不同的数据库系统中，每个应用可能用了不同的数据模型(不同的数据表、字段类型、长度等等)。如何有效的从这些应用中迁移数据，将是一个问题。

● 令人担心的数据质量问题

由于录入人员的失误，各地独立的应用自身可能有数据质量问题。在加上需要整合所有的数据，可能会带来多种的数据质量问题。如何确保整合后数据的一致性和准确性，将是一个问题。

● 全球各办公场所协同开发的项目，有效的协同工作将是个问题

数据迁移的开发人员需要与各地的IT人员、SAP平台的业务设计人员和数据建模者共同完成这个项目，如何有效的协同工作将是一个问题。

解决方案

该公司选择IBM InfoSphere Information Server数据整合平台来实现这个数据迁移项目。其中，使用Information Analyzer来分析各地应用的数据质量，并且根据分析结果执行了数据清洗计划。使用QualityStage进行数据清洗，使用DataStage以及SAP Pack开发ETL作业。使用MetaData Workbench建立统一的数据集成视图。使用FastTrack以通用的方法定义业务需求促进了业务分析人员、数据建模人员和开发人员的协同工作，并且根据这些定义的需求直接生成了对应的DataStage ETL初始作业，加快整个数据迁移项目的进程。

方案收益

通过该方案，该公司获得了以下能力：

- 提前完成数据迁移项目，成功地迁移和整合所有的数据到统一的SAP平台。
- 通过提高数据准确度改进了业务流程。
- 确保了新业务平台的数据质量和有效管理。
- 合并标准化的业务角色从而全面简化与不同地域的员工、客户、和供应商的沟通和协作。

IBM InfoSphere Metadata Workbench产品介绍

IBM InfoSphere Metadata Workbench是InfoSphere Information Server的关键组件，提供了基于Web的方式查看由IBM InfoSphere Information Server所生成、使用并导入的元数据。InfoSphere Metadata Workbench可帮助业务和IT用户探索和管理这些元数据资产，提供关于数据流、数据沿袭以及更改信息所产生的影响报告。用户使用InfoSphere Metadata Workbench能发现和分析元数据存储库中的资产信息资产之间的关系，了解和管理整个企业的数据流。

基本功能

- 浏览,查找元数据资产
- 管理, 查询元数据存储库
- 浏览元数据模型
- 导出、删除和保存业务衍生资产
- 创建数据衍生报告、业务衍生报告和影响分析报告
- 设置扩展数据衍生
- 管理“Manage Lineage”实用程序

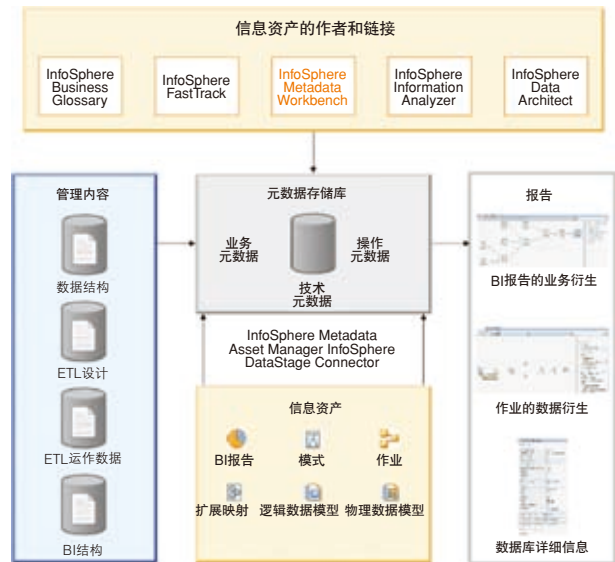
平台支持

- AIX
- HP-UX
- Solaris
- Windows
- Red Hat on AMD/Intel/System z
- SUSE on AMD/Intel/System z

数据支持

- Cognos
- MicroStrategy
- Business Objects *
- IBM Rational Data Architect *
- CA ERwin *
- User Information *
- Categories and Terms *
- Database ODBC
- Other 3rd party product metadata includes limited import of physical schemas, BI reports and BI OLAP models. IBM DB2 for LUW

体系架构



客户价值

- 为企业信息集成平台，提供360度全方位的元数据视图
- 提供清晰，完整的数据血统分析，帮助用户理解数据的准确来源和去向
- 洞察数据来源,提供数据是如何通过ETL处理,质量规则应用和其他关键应用的细节
- 追踪问题数据的关键来源,了解哪些系统都受到了影响,并建立可信的分析结果
- 帮助用户理解数据库表，列，ETL job等等技术元数据的业务意义

IBM InfoSphere Metadata Workbench产品介绍

典型案例

项目背景

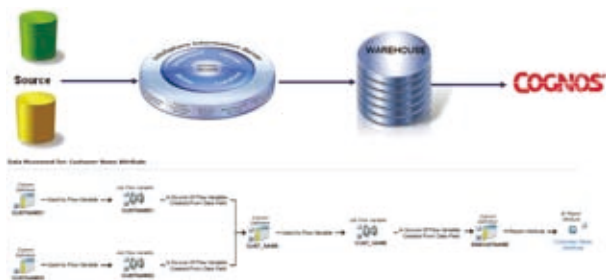
随着业务的快速扩展，某大型集团公司并购了数家海外公司，其全球化进程的到很大提高。

集团公司为了快速应对全球客户需求的挑战，需要整合各个分公司已有的不同数据仓库，商业智能系统和业务支撑系统，从而建立起一个统一的信息集成平台。建立一套标准化的指标及报表用以对公司业绩和客户进行统一的分析和评估。

项目挑战

- 改进，增强已存ETL流程，提高性能以应对日益增长的数据量
- 查找，减少冗长的处理流程和数据表
- 需要为整个工程的软件开发和部署定义一个统一的集成流程
- 开发人员缺少元数据集成的经验
- 需要寻找使用一个经过验证的、良好定义的方法快速迁移信息资产
- 需要跨分公司去分析销售和客户服务数据
- 流水线化报表流程，以满足高水平的报表和数据来源分析。
- 整合各个分公司和分部已经使用的不同系统和工具

数据流



解决方案

使用InfoSphere Information Server平台系列软件建立起统一的信息集成平台，优化数据集成基础设施的架构部署和管理。

- 针对各种不同类型的数据库和已存的支撑系统，使用InfoSphere DataStage按需集成，包括实时数据集成和对大数据集的高性能处理。
- 使用InfoSphere Business Glossary建立起一个针对整个集团业务元数据的标准化词汇和分类系统。帮助非业务人员清晰的理解业务数据。
- 使用InfoSphere Metadata Workbench对整个数据集成平台建立了一个面向集团的统一的元数据视图。当ETL设计变动时，其影响分析功能有助于用户快速的进行定位和执行评估。其数据血缘分析功能，能帮助用户理解报表数据的来源，增强报表数据的可信性。

使用IBM Cognos Business Intelligence软件扩展了客户的报告功能，提供标准化的报告和业绩分析，包括统计分析和策略支持。同时还支持使用分数卡、仪表盘、报告和实时监测等多种手段来监控和测量业务业绩。

使用IBM InfoSphere Warehouse 作为高性能数据仓库。

方案收益

- 建立统一的，360度全范围的元数据视图
- 帮助业务和技术人员对已有的数据和业务进行快速和容易的辨别
- 大大缩短ETL工程开发周期，提高后期维护效率
- 丰富的报表功能快速灵活的反应了业务的变化和趋势，帮助决策支持

IBM InfoSphere QualityStage产品介绍

InfoSphere™ QualityStage是IBM®的一款数据清洗工具。它可以显著改善企业数据质量，从而确保您的系统为整个企业内的业务用户交付准确、完整的信息。

基本功能

- 通过数据调研可以揭示数据中的趋势和潜在异常，验证和发现数据字段中的业务规则，识别无效值或默认值，在上下文中全面了解面向业务的数据语义。
- 标准化数据，将不同格式的数据转换成统一的标准的格式，识别并改正无效值，标准化拼写和缩写格式。
- 利用业界领先的基于统计的概率性匹配算法识别一个或多个数据源之间的重复数据，建立不同记录间的数据联系。
- 根据重复数据交叉填充“最佳的”可用数据，保留“最佳”的信息创建一个“合并的”记录。
- 直观易用的图形化用户界面，与DataStage无缝整合。
- 提供多种扩展模块，其中AVI可以支持对240个国家的地址数据进行验证。

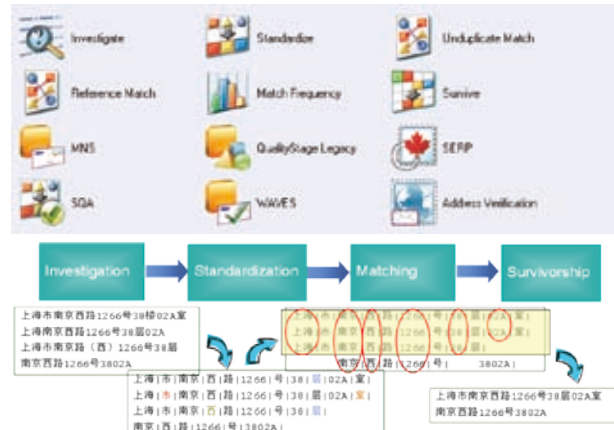
平台支持

- AIX
- Solaris
- Red Hat on AMD/Intel/System z
- SUSE on AMD/Intel/System z
- HP-UX
- Windows

应用场景

- 数据整合业务中质量问题
- 商业智能(BI)以及数据仓库项目的实现
- 主数据整合以及管理
- 数据治理
- 解决业务系统中的各种数据质量问题
- 企业级应用，包括ERP,CRM, SCM等

关键组件和清洗流程



- Investigate调研源数据
- Standardize标准化数据
- Unduplicate Match/ Reference Match对具有相同属性的数据进行分组
- Survive根据分组数据，留存最佳数据
- 地址校验模块，确保企业地址数据可邮递

数据连接支持

RDBMS DB2 (on Z, I, P or X series) Oracle Informix (IDS and XPS) MySQL Netezza Progress RedBrick SQL Server Sybase (ASE & IQ) Teradata HP NeoView Universe UnData Greenplum PostresSQL And more....	Standards & Real Time WebSphere MQ JMS Java Distributed Transactions XML & XSL-T Web Services (SOAP) Enterprise Java Beans (EJB) EDI EBXML FIX SWIFT HIPAA CDC / Replication DB2 (on Z, I, P, X series) Oracle SQL Server Sybase Informix IMS VSAM ADABAS IDMS General Access Sequential File Complex Flat File File / Data Sets Named Pipe FTP External Command Call	Legacy ADABAS VSAM IMS IDMS Datacom/DB <u>3rd party adapters:</u> Allbase/SQL C-ISAM D-ISAM DS Mumps Encosbe FOCUS ImageSQL Infoman KSAM M204 MS Analysis Nomad NonStopSQL RMS S2000 And many more.... Bold / Italics indicates Additional charge item...
--	---	---

IBM InfoSphere QualityStage产品介绍

典型案例

项目背景

某银行正在受到客户满意度低，市场反应速度慢，为了维护不同的系统数据需要花费高昂的成本等由于数据质量差而造成的困扰，为了基于市场做出更准确的决策，改善用户满意度，为客户提供更好的服务，降低维护成本等，正寻求建立一个完整一致的主数据管理系统，但数据分散在不同的系统和部门中，数据整合和数据清洗成为不可逾越的障碍。

项目挑战

- 客户信息分散在大量系统和部门中
- 不完整的数据，不一致的数据格式分散在不同的系统中
- 大量的有用信息被埋藏在自由格式的数据中
- 不同系统中的数据缺乏联系
- 大量的重复客户数据
- 很难理解和反应客户的需求
- 谁是我们的客户？我们给了客户最好的服务了吗？我们真正满足客户的需求了吗？我们是否正在错失一些机会？

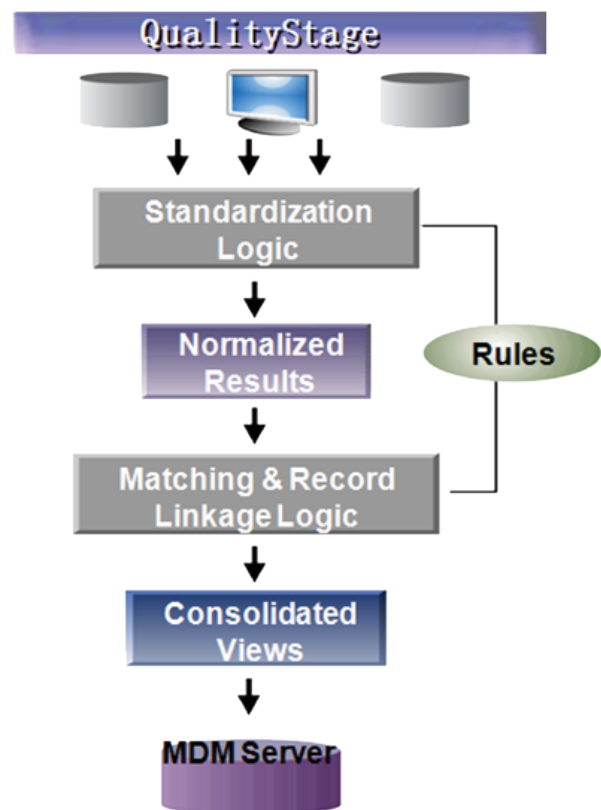
方案收益

- 改善了用户的满意度，显著地降低了成本。
- 销售和市场部门可以利用客户的完整视图，将其它产品推销给客户。
- 加强了数据治理和合规流程。
- 为客户提供新的服务变得容易。
- 更好的客户营销和洞察客户需求。

解决方案

- 建立客户数据治理委员会
- 利用QualityStage清洗标准化客户数据
- 利用QualityStage的先进匹配算法去除重复客户数据
- 利用SOA建立以消费者为中心的枢纽
- 为公司职员提供唯一的精确的客户视图
- 利用IBM Information Server和MDM Server能够将不同系统中的数据标准化，清洗，迁移到一个单一系统中

部署架构



- 灵活的解析和标准化能力
- 处理复杂的自由格式的数据
- 比竞争对手高出2%-4%的匹配成功率
- 强大透明的并行处理能力

IBM InfoSphere DataStage产品介绍

作为IBM®信息集成平台的核心产品模块，IBM® InfoSphere® DataStage®提供了一个强大的图形化框架，使用它用户可以设计和运行用于数据抽取、转换和加载的高性能的并行作业。

基本功能

- 基于几百个内置的转换功能对数据流提供无需编码的可视化设计界面
- 支持以团队为基础的开发和协作
- 能够对来自广泛来源的数据源提供数据集成
- 使用元数据驱动开发完成对数据的ETL功能
- 数据集成的设计无需关注数据量的大小和时间限制
- 支持SMP、Cluster、GRID和MPP

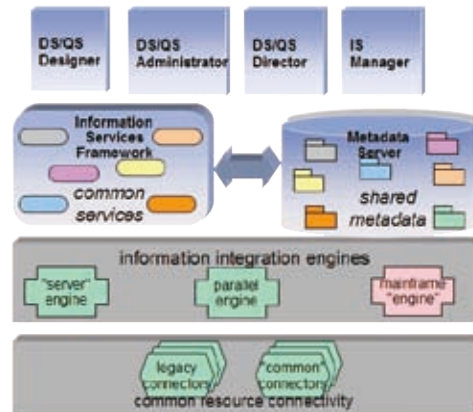
平台支持

- AIX
- Solaris
- HP-UX
- Windows
- Red Hat on AMD/Intel/System z
- SUSE on AMD/Intel/System z

数据连接支持

RDBMS DB2 (on Z, I, P or X series) Oracle Informix (IDS and XPS) MySQL Netezza Progress RedBrick SQL Server Sybase (ASE & IQ) Teradata HP NeoView Universe UnData Greenplum PostresSQL And more....	Standards & Real Time WebSphere MQ JMS Java Distributed Transactions XML & XSL-T Web Services (SOAP) Enterprise Java Beans (EJB) EDI EBXML FIX SWIFT HIPAA CDC / Replication DB2 (on Z, I, P, X series) Oracle SQL Server Sybase Informix IMS VSAM ADABAS IDMS General Access Sequential File Complex Flat File File / Data Sets Named Pipe FTP External Command Call	Legacy ADABAS VSAM IMS IDMS Datacomp/DB 3rd party adapters: Allbase/SQL C-ISAM D-ISAM DS Mumps Enscribe FOCUS ImageSQL Infoman KSAM M204 MS Analysis Nomad NonStopSQL RMS S2000 And many more.... Bold / Italics indicates Additional charge item....
--	---	--

关键组件和框架图



- 可扩展的平台—使用多处理器硬件平台的并行处理功能管理数据
- 全面的源和目标支持—在一个作业中几乎能集成无限制的异构数据源和目标
- 实时数据集成支持—使用相同的转换规则，在集成批量数据的相同平台上实时捕获消息或抽取数据
- 易于使用—最大限度提升构建、部署、更新和管理数据集成基础架构的速度、灵活性和有效性
- 大数据支持—强大的并行引擎，保证对海量数据处理的高性能

客户价值

- **集成信息:** 最小化客户投资成本
DataStage平台可以连接任何应用程序、数据和内容，无论这些源是结构化的还是非结构化的，大型机的还是分布式的。同时实现企业开发团队协作开发，向用户提供统一系统开发平台。
- **转换信息:** 加快客户价值实现时间
无需编码的开发方式、整合程序的可视化设计实现更快的作业开发。内置多种转换功能降低常见任务的复杂性，与使用单独工具和引擎开发所用时间相比减少高达40%的开发时间。
- **交付信息:** 提供可信、及时、准确的信息
数据交付的多样性和灵活性帮助企业持续应对与日俱增的业务需求挑战，同时负载均衡提供用户所需的高可用性和并行引擎极大的缩短了数据处理时间，实现信息的及时交付。帮助组织和企业从其系统中的各种复杂、异构的信息中获得更多价值。

IBM InfoSphere DataStage产品介绍

典型案例

项目背景

某大型国有银行已经初步建成了一套比较完善的银行业务系统，包括核心业务系统和信贷业务系统、以及为满足管理需要的各类管理信息系统。建设了用于各类报表统计和数据查询系统，并为各管理系统提供数据服务的数据集成平台，提高了数据的共享程度。但随着业务的快速发展和管理的不断提升，对现有的数据集成平台在数据整合能力、稳定性、性能、数据质量的探查和和分析能力等方面提出了更高的要求。

项目挑战

1. 数据整合是银行数据仓库建设重要的环节，但目前该银行的数据来源非常广泛，包括近10个子系统和来自各个部门的T级别总量的数据文件，几十个庞大数据库，以及现有的应用程序等等。对于整合各数据源和目标非常复杂，当前的应用整合的方式不但开发效率低，而且不易于维护。
2. 由于银行内各个生产系统是在不同时期，由不同的开发商开发完成的，在建立之初缺乏统一的工具和标准，因此在数据仓库建立的过程中需要做各种数据的转换和复杂的数据处理。
3. 银行希望能统一多个独立部门的相关数据，创建组织整体业绩和利润的综合视图。但是由于数据量比较庞大，数据仓库中报表数据的构建比较费时，报表生成比较迟缓。如何高效快速的处理海量数据到数据仓库，为报表系统及时提供数据对于银行而言，这个突破性的结果仍然相当困难。

解决方案

- 通过部署IBM InfoSphere DataStage平台系统，实现了各个信息系统、数据库之间的数据交换集成，完成了对该银行各个部门各种数据源中数据的抽取分发，处理流转，质量提升等需求。由ETL作业提高了数据关系管理和工作效率。真正实现了此银行开发团队协作开发，统一了系统开发平台。
- 使用DataStage内置的上百个数据转换功能，极大的提高了数据处理的效率，实现了各种复杂的数据转换场景。

- 采用IBM DataStage集群系统设计，充分利用多余的硬件资源，最小化该银行的硬件成本，并改变效率低下的局面。将动态资源分配等优势和功能体现到了该系统上，平衡了总体工作负载。从根本上改变了管理数据中心的方式，大大的缩短了海量数据的处理时间，提供了具有巨大性能优势的分析型数据生产能力实现了对数据仓库报表数据的及时构建，为进一步发掘核心业务数据的价值提供了巨大帮助。

部署架构



方案收益

- 通过部署IBM DataStage系统，真正实现了此银行开发团队协作开发，统一了系统开发平台，缩减了大量投资成本。
- 使用IBM DataStage内置的上百个数据转换功能，实现了各种复杂的数据转换场景，极大的提高了数据处理的效率。
- 采用IBM DataStage集群系统设计，降低当前多架构数据中心的复杂性，最小化该银行的硬件成本，并改变效率低下的局面。将动态资源分配等优势和功能体现到了该系统上，平衡了总体工作负载，从根本上改变了管理数据中心的方式。
- 通过部署DataStage集群环境，大大的缩短了海量数据的处理时间，对数据仓库报表数据的构建起到了至关重要的作用。为Cognos报表系统及时提供了数据，使得之前的4天延迟缩短到8个小时。

IBM InfoSphere Information Services Director 产品介绍

InfoSphere Information Services Director(简称ISD)可以将InfoSphere Information Server数据集成平台中的数据整合过程以SOA的方式发布成可重用的服务。

基本功能

- 自动发布可重用的服务
- 支持多种协议
- 提供负载均衡和容错功能
- 创建批处理和实时处理的服务

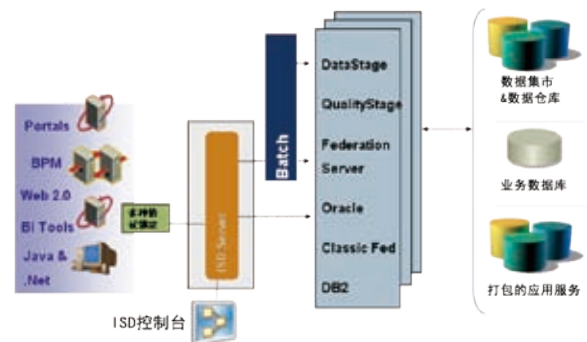
服务提供程序

- InfoSphere DataStage
- InfoSphere QualityStage
- IBM DB2
- Oracle
- IBM InfoSphere Federation Server
- IBM InfoSphere Classic Federation Server for z/OS
- InfoSphere Master Data Management Server

支持协议

- EJB
- SOAP OVER HTTP
- TEXT OVER JMS
- SOAP OVER JMS
- REST
- REST 2.0
- RSS
- TEXT OVER HTTP

工作方式



ISD的主要特性:

- 快速开发及发布SOA服务: 通过GUI界面, 可在数分钟内将数据整合逻辑发布为一个SOA信息服务, 无需任何手工编码。
- 集成的元数据管理: 集成了InfoSphere Information Server的元数据存储库, 实现SOA信息服务元数据的统一存储和管理。
- 灵活的部署: ISD支持多种协议和数据格式。

客户价值

- 交付一致、准确的业务数据和可重用的信息服务。
- 无需编码, 在数分钟内就可以构建并发布一个数据整合服务。
- 支持多种协议和数据格式, 能为各种类型的应用提供可重用的信息服务。
- 可将服务定义为业务对象, 部署在本地应用服务器中, 对服务消费者完全隐藏实现的复杂性。
- 提供一个支持容错、负载均衡和并行执行的灵活环境, 实现高可用性。
- 基于纯J2EE 架构, ISD提供了一种安全且具有灵活性的框架。它可用于托管数据集成服务并将请求分派给IBM InfoSphere Information Server、WebSphere Process Server 或其他的应用程序和数据库。
- 提供信息服务的元数据管理, 通过关联IBM InfoSphere Information Server的底层元数据基础架构, 实现了对数据整合服务进行集中管理。
- ISD为数据整合提供一个SOA环境, 可方便利用诸如WS-I、OASIS、W3C和Java Community Process之类的组织的开放标准行为。

项目背景

某集团公司已建立了ERP、CRM、HR、库存、采购和销售等信息系统。由于这些系统是在不同时期为特定的一个或多个在地理上分散的部门建立的，所以在这些信息系统之间存在信息冗余，数据格式不一致，连通性差和信息服务难以集中管理等问题。这在一定程度上形成了信息孤岛。为提高集团的竞争力，企业信息系统必须解决上述问题，使集团的任何部门无论何时何地都能个性化定制所需的服务，从而得到准确的、一致的和可信赖的业务数据。

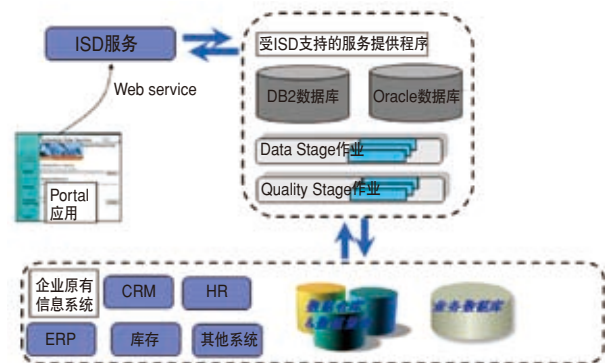
项目挑战

- 如何集中管理和部署可重用的SOA服务
- 如何从各个不同的信息系统中，集成有效信息并提供准确的、一致的和可信赖的业务信息
- 如何应对项目需求变更快，开发周期短
- 集团缺乏熟练的IT开发测试人员
- 实时定制调用业务数据
- 如何处理大数据量与容错
- 系统安全性

解决方案

- 使用InfoSphere DataStage 和InfoSphere QualityStage完成集团内各信息系统间数据信息的抽取、转换、加载、清洗等工作，从而确保向集团提供准确的、一致的和可信赖的业务信息。
- 根据集团业务需求，使用ISD构建基于DataStage, QualityStage, Oracle, DB2等开发的应用服务，并将它们发布在SOA中作为可重用的服务。
- 使用IBM WebSphere Portal从ISD发布的服务中获取相关业务信息，从而使客户可以定制个性化的Web内容。

部署架构



方案收益

- 集团的所有信息得到了有效的整合集成，消除了信息孤岛。
- 集团各部门都可以得到准确、一致的和可信赖的业务信息。
- 集团IT部门可以根据企业业务需求做出快速响应，集中管理和部署可重用的SOA服务，以便让新业务快速上线。同时也无需大量IT专业人才来开发和维护所需信息系统。既缩短了开发周期，也节省IT成本，降低了企业技术难度。从而可以把更多的精力放在业务上。
- 集团的信息可以被个性化定制和实时获取。
- 新系统具备高可靠性和安全性，且有良好的客户环境。

IBM InfoSphere Data Replication产品Q复制介绍

InfoSphere™ Data Replication是IBM®的一款数据库实时复制产品，包含SQL复制、Q复制和CDC三大技术框架。其中的Q复制为数据分发、迁移、整合和同步提供了高吞吐的强健方案，保证了业务的高可用性和连续性。

产品特点

- 基于日志的变更数据捕获(Change Data Capture)技术可以实时复制任务关键型数据事件，同时又不会影响系统性能。
- 低延迟、高吞吐量的数据复制，同时保证事务的完整性和一致性。
- 采用专业消息传递软件IBM WebSphere MQ保证消息的高速可靠传输。
- 易于使用的图形化用户界面(Replication Center)，支持便捷的数据复制配置。
- 全面的监视功能(Q Replication Dashboard)提高复制环境的可视化水平，加速问题解决。
- 跨硬件平台和操作系统的实时数据复制。
- 支持单向、双向、多向配置及其组合。

平台支持

- IBM z/OS
- AIX
- Linux
- Windows
- HP-UX
- Solaris

数据源支持

- IBM DB2 for z/OS
- IBM DB2 for LUW
- IBM InfoSphere Warehouse
- Oracle Database
- IBM Informix Dynamic Server
- Sybase Adaptive Server
- Microsoft SQL Server
- Teradata

关键组件



- Q Capture程序: 从源端数据库日志捕获数据变化。支持IBM DB2数据库、数据仓库和Oracle数据库。
- MQ组件: 传输并保存数据变化。Q复制中使用WebSphere MQ保障数据的永久性和可信传输。
- Q Apply程序: 在目标端应用数据变化。
- 配置工具: 支持图形化用户界面Replication Center，以及命令行工具ASNCLP。
- 监控工具: 可使用图形化用户界面Replication Center观察Q复制状态及数据传输性能，也可使用基于WEB的InfoSphere Q Replication Dashboard产品全面监控Q复制。为方便管理人员随时随地观察Q复制动态，InfoSphere Q Replication Dashboard目前已支持手机等移动设备。

客户价值

- **高可用性**——双向复制功能支持关键业务应用的计划内和计划外切换及回切，减小业务停顿时间。
- **负载均衡**——以事务为单位进行复制，保证目标端数据一致性和可用性，为平衡部署工作负载提供了可能。
- **实时报表**——借助基于日志的变更数据捕获技术，把对交易服务器的影响降低到最小，获得的准实时的业务数据拷贝可用于快速的报表生成和业务决策。
- **信息集成**——自动将数据从总部分发到多个分支机构，或从多个分支机构合并到总部，为企业提供一份集成的业务数据，且对应用透明，降低了应用开发工作量。
- **快速反应**——支持实时事件推送，帮助企业迅速对市场变化做出响应。

IBM InfoSphere Data Replication产品Q复制介绍

典型案例

项目背景

北京某政府金融单位主要负责国际收支、对外债权债务的统计和检测，按规定发布相关信息，承担国家跨境资金流动检测的相关工作，并需要根据外汇市场变化研究提出外汇管理体制变革和防范国际收支风险、促进国际收支平衡的政策建议，以及向央行提供制定人民币汇率政策的建议和依据。因此该单位需要有一套完备的数据系统实现数据的实时整合，从而为数据实时分析或数据报告的生成提供有力的保障。

项目挑战

目前该单位已部署了一系列以IBM DB2企业版为数据库的WEB应用信息系统。但是为保证数据的安全性，该单位将数据操作进行了分类——来自授权银行或企业的内网用户连接仅供内部使用的DB2数据库进行事务操作，而外网用户通过连接单独的供外部使用的DB2数据库进行数据查询。

因为现有的系统结构会导致信息更新不及时，影响数据报告的准确性，最终导致决策失误，因此单位想要实现以下几个目标：

- 同步内网DB2数据库和外网DB2数据库间的数据
- 数据同步近实时，延时在可接受范围内
- 数据复制要保证数据安全
- 数据复制不能影响业务系统的性能

解决方案

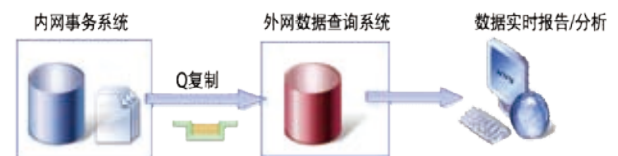
InfoSphere Data Replication软件是IBM的一款数据库实时复制产品。其中的Q复制技术自2004年发布以来，因其低延时、高吞吐、安全可信的强健特性，得到了业内的一致认可。且因为其基于日志的变更数据捕获技术在执行数据复制任务的同时不会影

响系统性能，能够很好地满足该单位的需求并实现项目目标。

因此IBM基于InfoSphere Data Replication软件为该单位制定了一套合理的数据复制方案。该方案最终得到了单位的认可和采用。

- 使用4台运行IBM AIX V6系统的IBM Power 595服务器支持现有的用于数据管理的外汇交易管理系统。
- 在Power 595服务器上部署InfoSphere Data Replication软件Q复制。
- 在Power 595服务器上安装IBM WebSphere MQ软件用于传输在内网DB2数据库和外网DB2数据库间的数据。
- 使用Q复制配置工具在内网DB2数据库和外网DB2数据库之间创建单向Q复制。
- 使用Q复制监控工具实时监控数据复制的状态和性能。

部署架构



方案收益

通过该数据复制方案的实施，该单位具备了在内网数据库和外网数据库之间安全、可信地进行数据转移的能力，并且进一步促进了企业流程改进，降低了外部访问可能带来的网络安全威胁。很好地满足了该单位对数据安全性的要求。更重要的是，该数据实时方案保证了数据报告或数据分析时，数据的及时性和完整性，从而为决策的准确性提供了帮助。

DB2复制 版本10新特性

- 自动复制IBM DB2表定义改动，如新建表，删除表
- 自动复制IBM DB2列定义改动，如新建列，修改列数据类型
- 为支持GDPS Active/Active Sites双活方案所做的各项功能增强
- 移动版Q复制监控工具

IBM InfoSphere Data Replication产品CDC介绍

InfoSphere™ Data Replication是IBM®的一款数据库实时复制产品，包含SQL复制、Q复制和CDC三大技术框架。其中Change Data Capture(CDC)在异构数据库复制和与IBM产品集成方面(Information Server, MDM, Netezza), 具有很强优势。

产品特点

- 采用基于日志的变更数据捕获技术，对关键业务系统及其性能影响很小。
- 仅发送变更数据到目标端，网络影响小，数据吞吐量高，实时性强。
- 简单易用的图形化用户界面(Management Console)，配置快捷，监控全面，支持拖拽，自动映射，瓶颈分析等功能。
- 支持异构数据库，跨硬件平台及操作系统。
- 支持单向、双向、一到多、多到一、级联、多线程等多种拓扑结构及其组合。
- 支持自适应，审计，汇总等多种复制模式。
- 与Information Server直接集成。

平台支持

- IBM z/OS
- IBM i/OS
- AIX
- HP-UX
- Solaris
- Linux
- MS-Windows

数据源支持

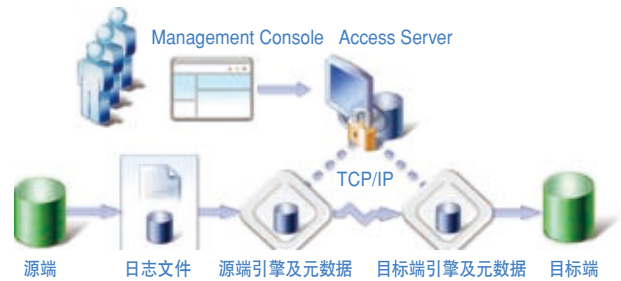
源端/目标端同时支持

- DB2 for z/OS
- Oracle
- DB2 UDB for LUW
- Sybase
- DB2 i
- MS SQL Server
- Informix
- Solid DB

仅目标端支持

- Information Server
- Teradata
- Cognos Now!
- MySQL
- Netezza
- Greenplum

关键组件



- **Management Console(MC):** 图形化界面工具，用于管理和配置复制信息，并对复制状态进行实时监控，获取CDC运行信息。
- **Access Server(AS):** 用户配置及安全管理工具，支持图形化界面(V6.5之后集成在MC中)及命令行。
- **日志文件:** 源端数据库日志文件，包括(Journal Log、Redo/Archive log)。
- **源端引擎:** 读取源端数据库日志捕获变更数据，经过行列过滤，字符编码转换后由TCP/IP发送给目标端。
- **目标端引擎:** 接收源端发送的变更数据，经过数值转换，字符编码转换，冲突检测后应用到目标数据库。
- **元数据:** 存储MC中配置的复制信息和当前复制进行状态。

客户价值

- 轻松实现跨平台异构数据库的数据迁移，并能确保迁移过程中关键业务系统的可用性与新系统中数据的完整一致性。
- 及时交付变更数据，支持主数据管理、动态数据仓库和ETL流程优化，提升业务速度，优化商业决策。
- 降低企业级数据集成和分发的的工作量，确保数据在企业范围内的实时一致。
- 提供高可用数据及灾备能力，方便系统随时升级维护并确保数据安全。
- 与IBM InfoSphere DataStage直接集成，确保事务完整性，向ETL流程提供实时数据源，生成实时报表。
- 支持面向服务的架构(SOA)，将实时数据事务打包为XML文档，并与消息传递中间件(如WebSphere MQ)相互传递。
- 支持数据库集群架构，对于Oracle数据库支持RAC架构，对于DB2数据库支持DPF功能。

IBM InfoSphere Data Replication产品CDC介绍 典型案例

场景一 动态数据仓库

项目挑战

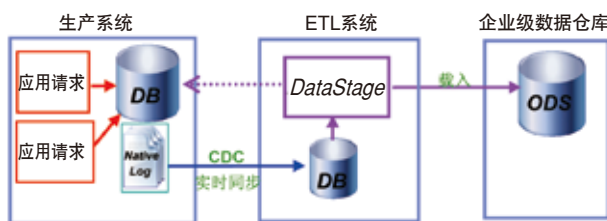
- 采用批量抽取方式对生产系统数据进行ETL，使得数据仓库中的数据非最新。
- 对生产系统的访问请求在进行数据批量抽取时将被不响应。
- 生成报表和处理用户特殊请求时需要访问生产系统，占用其系统资源。

解决方案

- 采用CDC实时同步生产系统变更数据到ETL系统。

方案受益

- 可将生产系统的变更数据通过ETL实时更新至数据仓库，以满足不同业务应用。
- 对生产系统的访问请求不再受影响。



场景二 数据迁移

项目挑战

- 传统迁移过程需要宕机，对于金融行业损失巨大。
- 不同平台不同数据库之间做数据迁移需要编程实现。

解决方案

- 采用CDC进行跨平台数据迁移。

方案受益

- 有效降低生产系统宕机时间，迁移过程无需编程支持。
- 提供安全机制确保事务正常回滚，支持大事务。



场景三 数据的整合与分发

项目挑战

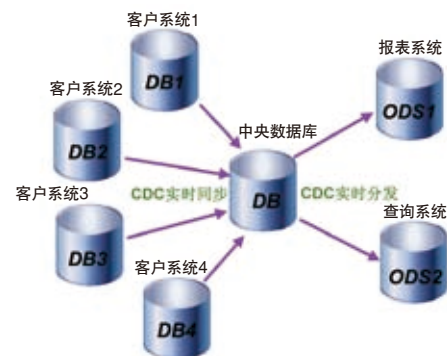
- 整合多个生产系统中分布在不同数据库中的用户数据。
- 为降低系统负载，需要将整合的用户数据分发到不同系统进行应用。

解决方案

- 采用CDC对不同系统数据进行数据库级，表级，甚至行级整合。
- 采用CDC对整合数据进行实时分发，并确保数据的完整一致。

方案受益

- 用户数据能实时整合并被分发至报表系统和查询系统，既能保证数据的实时性与完整性，又可降低中央数据库负载，有效控制中央数据库的数据增长。



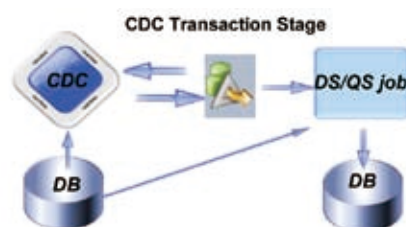
场景四 与DataStage直接集成

CDC 6.5新功能

- CDC通过DS8.5特有的CDC Transaction Stage与DS直接集成。

方案受益

- 可为DS提供跨多种数据库平台的实时数据。
- 利用CDC标签机制保证流入DS数据的可靠与一致。
- 实时性远好于通过Staging Table，文件及MQ方式与其集成。



IBM InfoSphere Master Data Management产品介绍

InfoSphere™ Master Data Management是IBM®的一款主数据管理工具。它为业务数据提供高质量，统一的数据视图。同时它也是一个高可靠性，高性能的主数据引擎保证业务的连续性和平滑性

基本功能

- 基于模糊的匹配算法，充分考虑数据的容错性
- 基于确定性匹配算法，支持数据转化规则的多样性
- 完善的匹配算法调优工具，通过直观的图形帮助开发人员生成最优算法
- 丰富的中文算法支持，支持简体中文，繁体中文的匹配
- 业务流程管理，保证主数据完整性，一致性
- 高性能的匹配与搜索引擎
- 基于配置的用户界面生成
- 丰富的系统集成接口

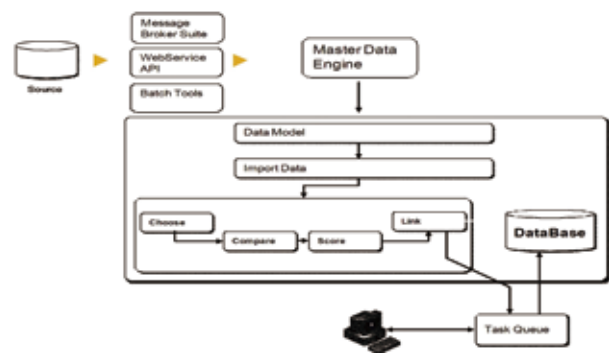
平台支持

- Windows
- Linux
- AIX
- Solaris

数据库支持

- IBM DB2 for LUW
- IBM Informix
- Oracle
- Microsoft SQL Server

关键组件



- Message Broker, Web Service API和其他的数据导入工具负责将不同数据源的数据信息导入到主数据服务器中(Master Data Engine)。
- 主数据服务器中(Master Data Engine)是主数据的核心引擎，进行主数据的匹配，评分与关联等操作。
- 数据管理员操作管理任务队列(Task Queue)中存放的任务,进行主数据的人工处理。

客户价值

- 从不同的业务信息系统中提取，转换，合并生成统一的业务数据视图为客户提供可靠的，高质量的关键性业务数据。
- 提供高可靠性和高性能的主数据引擎保证业务的连续性和平滑性。
- 为数据仓库，商业智能分析系统提供统一的高质量数据，帮助客户做出正确的商业分析。

项目背景

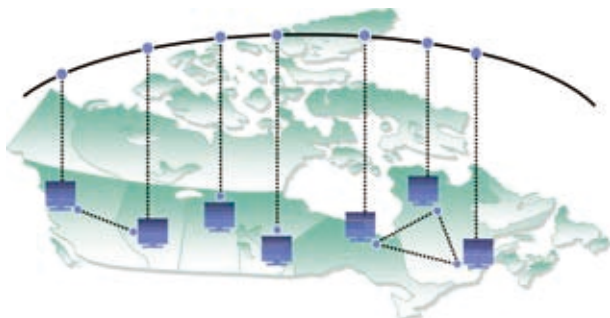


Canada Health Infoway (Infoway)成立于2001年,是一个独立的非盈利组织,其成员包括加拿大14个联邦、省市和地区的医疗部副部长。作为一个战略投资者, Infoway与联邦、省、市、管辖区和地区医疗机构以及其他医疗组织和信息技术供应商和提供商合作实施并复用了兼容的医疗信息系统,以支持更安全、更高效的医疗系统。客户端注册表是Infoway的九大目标项目领域之一, Infoway将其视为用于达成以下最终目标的“蓝图”的关键因素:使每个省份和地区(包括当地居民)都能从新医疗信息系统中受益(特别是,到2010年,使50%的加拿大公民能够访问安全的电子健康档案)。

项目挑战

- 如何集中管理和部署可重用的SOA服务
- 如何从各个不同的信息系统中,集成有信息并提供准确的、一致的和可信赖的业务信息
- 如何应对项目需求变更更快,开发周期短
- 集团缺乏熟练的IT开发测试人员
- 实时定制调用业务数据
- 如何处理大数据量与容错
- 系统安全性

部署架构



解决方案

IBM Initiate软件提供的注册式的主数据管理系统作为加拿大iEHR解决方案的核心组件来提供客户注册功能。

2006年, Infoway发布了一份意向书,称IBM Initiate软件满足Infoway蓝图(有关系统集成以实现Infoway EHR策略的指南)所列出的有关客户端注册表解决方案的所有严格需求。但是,各省仍自行制定了有关是否采用IBM解决方案的决策。选择IBM EMPI的省份之所以这么做的原因在于: IBM算法非常准确,而且在不同临床医疗数据系统中实施该软件的过程非常快速简便。IBM的行业领先EMPI可以根据需求快速准确地链接数据,从而以虚拟方式集成不同系统和应用程序中的数据。而且,它还可以帮助组织将任意数量的数据转化为所必需的清晰完整的医疗记录,以便更加经济有效、更负责、更好地与病人进行交互。还可以通过区域性实施IBM Initiate软件来实现中央客户端注册表标准化。

方案收益

八个省使用IBM Initiate MDM解决方案来提供省级EMPI

- British Columbia(2004), Alberta(2005), Manitoba(2006), Saskatchewan(2006), Ontario (2007), Newfoundland(2003), Nova Scotia(2008), New Brunswick(2009)
同时其中两个省还使用IBM Initiate MDM解决方案来提供省级医疗提供商的注册管理。
- Nova Scotia (2008), Alberta (2007)
八个省中的六个使用EMPI来做实时的病人识别
八个省的中六个将省级的EMPI和PACS进行了集成

客户收益:

- 不列颠哥伦比亚省提高了数据质量
- 艾伯塔省统一了病人视图
- 萨斯喀彻温省对服务点进行了增强
- 安大略省缩短了等待时间
- 纽芬兰省实现了流程标准化

IBM InfoSphere BigInsights产品介绍

InfoSphere™ BigInsights是IBM®的大数据平台的核心产品之一，它是一个以Hadoop为基础的对海量数据存储、管理和分析的企业级平台。

产品特点

- 增强Hadoop等开源组件的企业级特性，加强数据的安全性和可靠性。
- 对集群、文件、作业以及各组件进行实时的监控 (web console)。
- 基于Web的可视化数据分析工具(BigSheets)和文本分析组件(Text Analytics)。
- 为开发者提供基于Eclipse的大数据应用程序开发套件(Eclipse Tooling)。

平台支持

- Red Hat Enterprise Linux (RHEL)
- SUSE Linux Enterprise Server (SLES)
- Power Linux

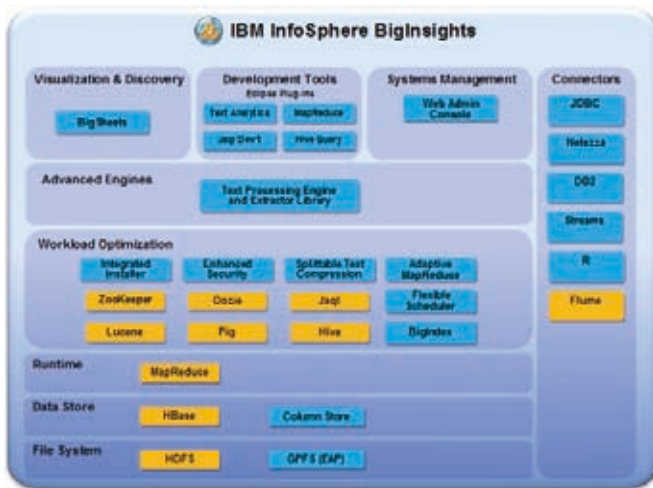
数据源支持

- IBM DB2
- IBM InfoSphere Warehouse
- IBM InfoSphere Data Stage
- IBM Netezza
- IBM InfoSphere Streams

客户价值

- 对Apache Hadoop进行了优化，提高作业的执行效率
- 与企业级的应用及文件系统集成，可以与多种数据源连接
- 行业应用加速器和数据分析加速器缩短了实施时间
- 可视化的用户界面便于用户从海量数据中挖掘更多的商业价值

关键组件



- 用户接口: 满足系统管理员、终端用户和应用开发人员的业务需求
- 高级引擎、负载优化: 整合开源项目和IBM的产品
- 运行环境: 支持MapReduce
- 数据源和接口: 既可以从传统的关系型数据库中读取数据，也可以和Flume等开源项目连接
- 存储层: 数据可以同时存放在企业级的分布式文件系统 (GPFS)和支持Hadoop的文件系统或数据库中

IBM InfoSphere BigInsights产品介绍

典型案例

项目背景

丹麦维斯塔斯(Vestas)风力系统公司主要从事开发、制造、销售、维护用于风力发电的电力系统。其产品范围包括陆地和海上风力发电机组。该公司已经安装了五大洲65个国家的43000风力涡轮机，其全球员工总数超过22000人，并且已经在12个国家建立生产基地。目前正在中国、西班牙和美国开辟新的生产基地。

项目挑战

维斯塔斯风力系统公司提供的风力发电机组作为替代能源解决方案凭借其客户满意度和富有竞争力的价格在行业内享有一席之地。

风力涡轮机的前期投资需要上亿美元，使用年限在20-30年，其安装位置和操作的地点会对单位机组的发电量以及产品运作周期产生巨大的影响。安装地点的选择需要同时考虑大量因素，包括该地点的温度、降水、风速、湿度和大气压力。根据工程师的经验，实际需要处理的数据总量大概为6PB。然而，

- 现有的数据处理程序不支持一次性处理大规模的数据
- 计算出一个解决方案需要花费几个星期的时间

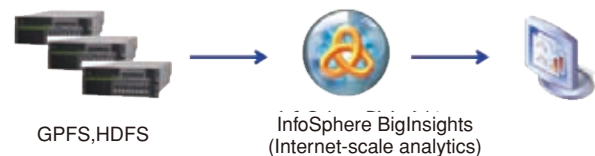
解决方案

维斯塔斯使用InfoSphere BigInsights管理分析天气和风力涡轮机安置地的数据信息，为最终的决策方案提供更多有价值的信息。InfoSphere Biginsights企业版的特性包括：

- 用BigInsights自带的高性能企业级分布式文件系统GPFS-SNC替代HDFS
- 用JAQL进行数据处理

- JAQL对JDBC的支持，实现了BigInsights与任何兼容JDBC的数据存储之间的集成
- 可以与DB2/RDBMS、Netezza、InfoSphere Warehouse进行双向连接，并且能够和ETL集成
- 以图表展示数据，允许用户可视化的查看非结构化和非传统的网络数据
- 提供作业和工作流程的管理
- 支持文本分析和建立大规模的索引
- 24 * 7的企业级支持

部署架构



方案收益

该方案的实施将计算时间从原有的3个星期降至3天，大大提高了该公司的效率，降低了成本，从而帮助其在市场中赢得更多的收益。

维斯塔斯是第一个IBM大数据解决方案的成功案例，也是第一个投资基于Hadoop的解决方案的大型集团公司。

IBM InfoSphere Streams产品介绍

InfoSphere™ Streams是一个支持开发和执行数据流应用程序的平台。Streams能够持续快速地分析实时产生的海量数据，帮助企业提高洞察力和决策速度。

产品特点

- 能够在大规模的集群环境中并行、高性能地处理流数据
- 在已搭建好的集群环境中自动部署流处理应用
- 不需要重启现有环境下对原有应用进行动态扩展
- 安全、可审核的执行环境
- 提供IBM流处理语言(SPL),使得终端用户可以直接对流数据进行操作

平台支持

- Linux (32-bit or 64-bit)
- IBM POWER7® (64-bit)

数据源支持

- IBM DB2 for LUW
- Oracle Database
- IBM Informix
- IBM Netezza Twinfin
- IBM solidDB
- Microsoft SQL Server
- MYSQL

关键组件



- 数据源: 支持多种类型的流数据作为输入
- Streams运行环境: 支持联合分析, 即同时进行数据流的分析和合并; 支持对流的循环操作, 即可以对一个流进行多种类型的分析操作
- 显示界面: 输出结果既可以以可视化的方式输出, 也可以以通知的形式告知接收系统
- 开发环境: 开发者可以使用IDE开发Streams的应用

客户价值

- 低延迟、实时响应 – 尤其适用于对响应时间有较高要求的应用, 例如欺诈检测、网络管理
- 解决数据量大, 存储成本高的问题 – 直接对数据进行分析, 不需要存储; 对有价值的数据进行深入分析
- 可以跨多个数据流进行分析

IBM InfoSphere Streams产品介绍

典型案例

项目背景

安大略理工大学(University of Ontario Institute of Technology)是加拿大安大略省的公立研究型大学。现有12个学院，3所附属大学学院和其他职业学校，超过23000名本科生和5000名研究生。该学校某研究所希望能够改进现有的同步辐射实验的能力。

项目挑战

同步加速器是能够产生高亮度X射线束的大型电子存储环。科学家利用光束在原子水平上探测材料，获得材料结构特性的更详细的信息。

现有的同步实验使用X射线衍射(XRD)扫描，获得原子级的数据。一次X射线衍射扫描包括数千计的连续的数字图像，每个实验通常需要多次扫描。然而，

- 现有的分析软件只能依次对每个X衍射射线图像单独处理
- 数据处理过程往往需要数天甚至是数周的时间

项目挑战

InfoSphere Streams是IBM大数据平台的核心产品。它是一个高性能的计算平台，允许用户根据其项目需求开发应用程序，以提高数据采集、分析的效率

- 用IBM流处理语言(SPL)实现InfoSphere Streams的应用程序
- Streams将快速删除XRD图像中的背景，提取出相关原子数据为后续分析使用
- 将应用程序部署在一个四核服务器上

部署架构



方案收益

通过实施该解决方案，UWO将数据处理时间降低至几分钟，大大节省了实验时间，使得原有实验方法得到了改进和提高。

UWO使用的数据处理方法有助于增加其项目的商业投资价值，也可以在全球科研机构的范围内被广泛应用。

IBM®