

概要：数据迁移和数据集成 项目的投资回报计算

白皮书



本档含有 Informatica Corporation 的保密、专有信息和商业秘密信息（“机密信息”），未经 Informatica 的事先书面同意，不得进行拷贝、散发、复印或以任何其它方式复制。

尽管我们尽最大努力确保本档中信息的准确性和完整性，但仍可能存在一些排版错误或技术误差。如因使用本档所含信息而造成任何损失，Informatica 概不负责。本档中包含的信息随时可能更改，恕不另行通知。

Informatica 自行决定将这些材料中讨论的产品属性纳入其任何软件产品的发布或升级中，并自行决定任何此类发布或升级的时间安排。

受下列一项或多项美国专利保护：6,032,158；5,794,246；6,014,670；6,339,775；6,044,374；6,208,990；6,208,990；6,850,947；6,895,471；或受下列正在申请的美国专利保护：09/644,280；10/966,046；10/727,700。

此版本已于 2006 年 11 月发布

目录

简介	2
Informatica Data Explorer 的主要优势	2
背景	3
为什么了解数据很困难?	3
示例项目	5
项目 1: 数据仓库	5
项目 2: ERP 实施	5
项目 3: 数据整合	5
ROI 方案	6
降低直接项目成本	6
早期项目实施的商业价值	7
降低项目超支成本	8
提升数据质量的价值	9
Informatica Data Explorer 的成本	10
总投资回报率	10
总结	11

简介

行业经验表明，数据迁移和数据集成项目很容易出现所有 IS 项目中常见的相同挑战及问题。这些项目时间紧、预算超支、需要在质量与期限之间进行权衡，也可能会彻底失败。项目经理每天都在面对这些难题，几乎没有选择的余地和好的解决方案。

Informatica 为当今复杂的企业环境提供了全面的数据集成解决方案。通过处理最广泛范围内的数据源，Informatica 解决方案可帮助各种规模的公司和政府机构随时随地访问、发现、清洗、集成和传输所需数据，从而为企业实现整体数据价值。

数据探查是数据集成生命周期中的基本步骤。从头开始可以显著地缩短计划和执行数据集成策略所花的时间。企业必须充分了解数据的内容、质量和结构之后，才能在数据仓库、CRM、ERP 或业务分析应用程序中集成或使用这些数据。Informatica Data Explorer（以前称为 Similarity Systems Evoke AXIO）是一款全面的数据研究、发现和映射产品，数据分析师可借助它来规划复杂数据迁移和数据集成项目。通过自动化和流程驱动数据探查，Informatica Data Explorer 为所有企业数据提供完整且极其准确的图片。

Informatica Data Explorer 的价值在于它能够快速分析多源系统，并帮助用户确定迁移或数据集成项目的规模和复杂度。数据分析师可借助此软件快速发现隐藏的数据质量问题、差异以及数据源内的一致性和不兼容问题。其结果是在中央公开资料库中生成一个精确的探查，可用于加速新应用程序、数据库和数据质量计划的设计和 实施。

Informatica Data Explorer 的主要优势

- 提供准确的源系统知识
- 提升企业数据质量和精确度
- 实现准确的数据迁移、集成和整合
- 有助于加速多个不同数据源的集成
- 降低数据管理项目的风险
- 最大程度地减少企业应用程序项目的超支
- 提高数据管理项目的效率
- 提高合并与收购数据集成的成功率
- 降低成本

背景

根据 Standish Group 一份题为“迁移难题”的报告显示，不管最初规划得多么仔细，在金额超过 100 万美元的数据迁移项目中，仍然有 4/5 (83%) 的项目面临延期、成本超支或完全失败的风险。虽然这些项目成本增加或浪费非常明显，但是比起延迟或取消新业务系统或数据仓库的实施而损失的商业价值而言，仍然相形见绌。为了最大程度实现商业价值，数据迁移项目的经理和规划人员必须立即解决这些超支和失败的潜在原因。他们不仅要确保准时完成项目，而且要通过提前完成项目来努力提升商业价值。

根据 Standish Group 的报告显示，数据迁移项目超支和失败的主要原因之一就是企业在进行数据迁移之前对源数据缺乏了解。此次调查对象涉及了美国 4 个城市内一系列重点小组中的 IT 主管人员。据这些主管人员反映，在数据迁移项目中，他们将 60%-80% 的工作精力放在了尝试了解和映射源数据上。他们同时也指出，这是一个永无尽头的重复过程。此次调查的结论是，如果能够更有效、更准确地了解源数据，就可以准时甚至是提前实现本次项目的商业价值。

通过对此问题进行更深入地研究发现，大多数迁移项目和数据集成方案都依赖外部信息来了解数据。这些文档、源计划、现有数据模型和工作经验等大部分信息通常都已过时、不正确或缺失信息。如果此类信息无效，则可能要进行多次重复工作来挖掘新信息并验证其是否真正正确（例如新信息确实代表了源数据）。在此方案中，多达 50% 的项目总体人工预算可能浪费在手动、过时数据分析和诊断技术上。同时，项目的整体成功率可能会因为缺少对源数据的了解而受到严重威胁。

为什么了解数据很困难？

- 手动数据探查单调、速度慢、费力而且容易出错
- 元数据文件可能会缺失、不完整或严重过时
- 可能无法找到遗留系统的源代码
- 数据元素之间的关系并非总是非常明显
- 假定关系和相关性可能出错
- 数据库不是静止的；可能会随时间的推移而被损坏

对多个数据源进行分析需要花很长时间—平均每个属性通常要花费 3 到 5 个小时，稍微复杂一点的将达到 10 个小时或更长。举个极端的例子：一家大型保险公司花了 2 个月的时间才完成对 6 个属性的完整分析。按行业标准来看，平均每个属性花费 3 到 5 个小时尚可接受，这与按照业务规则提取算法估计的时间大致一致。¹

但是，企业可通过更为直接的方法来了解源数据。Informatica Data Explorer 可用来彻底地分析实际源数据，而非依赖于不准确的元数据和文件。此解决方案可自动识别有问题的数据和元数据，使企业能够随时识别整个企业数据库中不一致、冗余以及不准确的数据。借助数据探查解决方案，企业可快速了解数据，从而更轻松地完成数据迁移项目或数据集成方案。

Informatica Data Explorer 支持在详细检查源数据过程中使用六阶段流程方法，提供对数据内容、结构、质量和完整性的全面理解，然后在此理解基础上生成来源到目标的属性映射。数据探查和映射流程可快速生成正确的数据模型和一组来源到目标的转换映射，并消除在使用手动方法时导致的无法预料且耗时的重复周期。

此解决方案通过以下方式帮助降低项目成本、减少项目风险、提升商业价值并实现更佳效果：

- 降低直接项目成本，如劳动力和资源
- 通过早期项目实施提高项目价值
- 避免与超支或注销相关联的意外成本和延迟效益
- 在不延期的前提下提升迁移或转换数据的数据质量

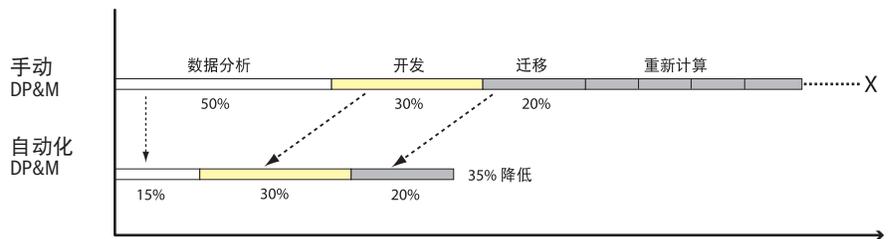


图 1: 加速 ROI

上图显示了如何使用 Informatica Data Explorer 通过削减数据分析所需的时间来立即缩短项目时间线。重复工作以及后续阶段中的延迟减少，导致进一步节省时间。

¹ Terry Moriarty, “实现业务规则自动化”, 数据库编程和设计, 1997 年 10 月: 74

示例项目

在本文档中，针对三个不同大小的项目来确定这四个优势分别对总投资回报率 (ROI) 的相关贡献。计划因素、估算以及下面显示的项目结果由大量客户的实际经验组成，不能反映出任何在特定客户站点获得的总体结果。

项目 1：数据仓库

某保险公司将数据从遗留应用程序迁移至数据仓库和多个数据集市。初始评估确定对遗留数据进行全面分析需耗时一年甚至更久，而且在完成以后可能不会产生有用的结果。此项目暂停实施。后来公司引入了 Informatica Data Explorer，从而得以在 6 周内完成数据探查和映射流程，并在 10 周内完成整个项目。所有工作通过内部项目团队完成；此领域中总成本通常为每小时 95 美元。

项目 2：ERP 实施

某制造公司经历了从多个遗留 VSAM 应用程序到 ERP 系统的迁移。准确完整地将人力资源数据（包括往年利润和大量的历史制造数据）迁移到新系统至关重要。项目由内部人员和外部顾问合作完成。大多数工作由内部人员完成，平均每小时花费 75 美元的成本。

项目 3：数据整合

某大型国际运输公司需要对最新获得的数据进行整合。预计分析每个属性耗时 4 个小时并不算久，但是由于项目的大小可能无法得到充分测试。公司最初采用 Informatica Data Explorer 来验证初始数据分析的准确性，后来又用于开发精确数据探查和数据映射。项目团队包括内部人员和来自某主要系统集成商的顾问；将这两部分人员都考虑在内的总成本为每小时 100 美元。

ROI 方案

对于这 3 个示例项目，投资回报率 (ROI) 分 4 个独立方案来演示：

- 降低直接项目成本，如劳动力和资源
- 通过早期项目实施提高项目价值
- 避免与超支或注销相关联的意外成本和延迟效益
- 在不延期的前提下提升迁移或转换数据的数据质量

虽然针对每个示例项目列举了不同的方案，但是在建立与使用数据探查解决方案相关的预期 ROI 时，大多数项目都将侧重于这些方案的子集。

ROI 方案在人员计算过程中采用了术语“全工时评量法”或“FTE”。FTE 可能只是代表某个项目的兼职人员，或可能由多名员工组成，他们具有项目所需的各种不同技能。

降低直接项目成本

一个正确规划的项目包括：分配用于探查数据资源的资源，设计目标系统和映射规范，以及执行提取、清洗和转换流程。如果在探查和映射阶段做得不够彻底而不能发现问题，导致需要在最后出现这些问题时加以解决，将会增加成本。相对于手动估算每个属性平均需要花费 3 到 5 个小时，Informatica 客户指出数据探查和映射的最长时间为平均每个属性 15 分钟。

表 1 显示了使用手动方法与使用 Informatica Data Explorer 方法时，3 个示例项目在数据探查和映射 (DP&M) 阶段投入的直接项目成本差别。Standish Group 报告显示，在平均金额为 100 万美元的数据迁移项目中，手动数据分析将花费整个项目预算的 50%。在更大的项目中，时间、成本和资源限制导致其中许多项目无法对源数据进行 100% 的分析。在表 1 的方案中，每个项目只对源数据进行 75% 的手动分析，但事实上对数据进行的分析只要不足 100% 就将导致项目风险增加和数据质量下降。同时，此方案将 Informatica Data Explorer 作为实施、培训和启动时间的应急措施，使数据探查和映射所需的全部工作加倍。具有丰富经验的团队开展后续项目就不需要加大总投入。

	公式	数据仓库	ERP 实施	数据整合
A. 初始项目预算	输入值	\$1,200,000	\$3,000,000	\$12,000,000
B. 每小时每 FTE 总成本	输入值	\$95	\$75	\$100
C. 属性数量	输入值	1,750	4,675	16,200
手动分析				
D. 每个属性耗时数	输入值	5.0	4.0	3.0
E. 总小时数 (75% 分析) *	$C \times D \times .75$	6,563	14,025	36,450
F. 手动 DP&M** 的成本	$B \times E$	\$623,485	\$1,051,875	\$3,645,000
G. DP&M** 占总预算的百分比	F / A	52%	35%	30%
使用 Informatica Data Explorer 时				
H. 每个属性耗时数	输入值	0.25	0.25	0.25
I. 总小时数 (100% 分析) ***	$C \times D \times .2$	875	2,338	8,100
J. 使用 Informatica Data Explorer 时的 DP&M** 成本	$B \times I$	\$83,125	\$175,350	\$810,000
K. 使用 Informatica Data Explorer 时的项目预算	$A - F + J$	\$659,640	\$2,123,475	\$9,165,000
L. DP&M** 占修正后预算的百分比	J / K	13%	8%	9%
直接项目成本减少金额				
M. 净减少额	$F - J$	\$540,360	\$876,525	\$2,835,000

表 1: 直接项目成本减少金额

*借助手动方法只能进行 75% 的分析 ** 数据探查和映射
*** 以系数 2 针对首个项目的培训和实施进行调整

早期项目实施的商业价值

通过使用 Informatica Data Explorer 可实现的一项重要投资回报是早期项目实施带来的商业价值。如果能估算出项目的商业价值，就很容易计算出这部分价值。在表 2 中，通过将 3 个示例项目每个月商业价值分别与月数相乘，来显示使用 Informatica Data Explorer 加速项目实施带来的总体商业价值。假定每个月工作了 175 个小时。

	公式	数据仓库	ERP 实施	数据整合
A. FTE 数	输入值	3	5	10
B. 项目每月商业价值	输入值	\$60,000	\$90,000	\$200,000
手动分析				
C. 总小时数 (75% 分析)	表 1.E	6,563	14,025	36,450
D. 花费时间 (月) *	$C / A / 175$	12.5	16.0	21.0
使用 Informatica Data Explorer 时				
E. 总小时数 (100% 分析)	表 1.I	875	2,338	8,100
F. 花费时间 (月) *	$C / A / 175$	1.7	2.7	4.6
直接项目成本减少金额				
G. 节约的时间数 (月)	$D - F$	10.8	13.3	16.4
H. 净价值	$B \times G$	\$648,000	\$1,197,000	\$3,280,000

表 2: 早期项目实施的商业价值

* 每月工作 175 个小时

降低项目超支成本

经事实证明，传统手工数据探查和映射方法是一个重复性很高的流程。元数据、副本、文件以及某些情况下的物理数据都是以手动方式进行分析。在此分析基础上，会生成许多假设和转换规范并对它们进行编码，同时数据被提取、转换和载入至新系统。在超过 80% 的数据迁移项目中，此流程并不是一开始就正常运作—将会不时涉及到编码、加载和扩充操作。结果，整个流程不断重复，直到项目取得成功或被取消。

然而，企业无法事先预测出具体的重复次数以及流程的完成时间。使用 Informatica Data Explorer 的关键好处之一就是可以对数据探查和映射流程进行预测。而且，由于数据 100% 支持生成的转换规范，因此项目很有可能在第一次实施就获得成功，而不会有无尽的重复工作。

据 Standish Group (“CHAOS’ 98: 摘要”，1998) 研究表明，在通过手动方法进行数据探查和映射的数据迁移项目中，83% 的项目不是超出预算就是彻底失败。全部项目调查显示，49% 的项目将超支，34% 的项目将会失败。平均项目延迟达 79%，项目成本超支达 69%。第二个百分比在表 3 中用来计算 3 个示例项目在没有使用 Informatica Data Explorer 等自动数据探查和映射产品时的预期超支成本。此方案同样假设在使用 Informatica Data Explorer 时，这些超支成本将会降低 85%，这一点已在许多数据探查和映射项目中得到证实。虽然一些项目可能根本就没有超支，但是好的商业规划应该将它作为整个规划的一部分考虑在内。

还有其它一些与项目相关的成本节约，视特定项目而定。假设这是一个更大型 IT 项目的一部分，例如 SAP 实施，数据分析的延迟将导致整个项目延迟。在更大型的 IT 项目中，企业可能会让 SAP 人员和软件/硬件在一旁等候数据。另一项成本是，特别是对迁移/整合来说，如果企业维护旧系统的时间比最初计划的更长，将会承担额外成本（例如为将要废弃的大型机上的 MIPS 支付费用）。为便于计算，我们没有考虑这类成本。

	公式	数据仓库	ERP 实施	数据整合
A. 手动 DP&M* 的总预算成本	表 1. F	\$623,485	\$1,051,875	\$3,645,000
B. 超支补贴	A x .69	\$430,205	\$725,794	\$2,515,050
项目超支成本减少金额				
C. 净减少额	B x .85	\$365,674	\$616,925	\$2,137,793

表 3: 项目超支成本减少金额

* 数据探查和映射

提升数据质量的價值

当 IS 项目面临紧张的期限时，为保证项目按时完成，质量通常会大打折扣。Informatica Data Explorer 在迁移过程中提供了大量信息，可以作为受数据质量管理平台 Informatica Data Quality 驱动的持续数据质量方案的起点。虽然很容易展示 Informatica Data Explorer 如何让项目经理在按时或提前完成任务的同时确保目标数据库中的数据准确且具有很高质量，但是要量化这一优势就困难得多。商业用户必须自行衡量拥有准确数据的价值。表 4 中使用的传统方法表明，在 12 个月内，数据质量改善后每个月创造的价值至少为项目商业价值的 20%。试想一下，企业必须至少花费同等数目的费用和时间才能恢复并更正低数据质量。

	公式	数据仓库	ERP 实施	数据整合
A. 项目每月商业价值	表 2. B	\$60,000	\$90,000	\$200,000
提升数据质量的價值	输入值	\$60,000	\$90,000	\$200,000
B. 净价值	$A \times 0.2 \times 12$	\$144,000	\$216,000	\$480,000

表 4：提升数据质量带来的价值

Informatica Data Explorer 的成本

表 5 显示了针对示例项目购买并实施 Informatica Data Explorer 的全部费用。此成本已纳入后续方案的 ROI 计算中。费用基于 Informatica 于 2006 年 11 月公布的价目表。价目表如有变动恕不另行通知，并且不代表报单价或要约卖出价。

	公式	数据仓库	ERP 实施	数据整合
A. Informatica Data Explorer	输入值	\$150,000	\$180,000	\$225,000
B. 源数据导入器 (VSAM 或关系)	输入值	\$0	\$20,000	\$20,000
C. 一年维护	A 的 20%	\$30,000	\$36,000	\$45,000
D. 培训和咨询	输入值	\$25,000	\$64,000	\$104,000
Informatica Data Explorer 的成本				
F. 总成本	A 至 D 总计	\$205,000	\$300,000	\$394,000

表 5: Informatica Data Explorer 的成本

总投资回报率

表 6 显示了计算 3 个示例项目的总投资回报率的一个方法。总投资回报率是减去直接项目成本，减去项目超支成本、早期实施的商业价值和提升的数据质量的价值的总和。无论在什么情况下，Informatica Data Explorer 的成本均会远远低于其带来的收益。

	公式	数据仓库	ERP 实施	数据整合
A. 直接项目成本减少金额	表 1, M	\$540,360	\$876,525	\$2,835,000
B. 早期项目实施的价值	表 2, H	\$648,000	\$1,197,000	\$3,280,000
C. 项目超支成本减少金额	表 3, C	\$365,674	\$616,925	\$2,137,793
D. 提升数据质量带来的价值	表 4, B	\$144,000	\$216,000	\$480,000
总价值				
E. 总价值	A 至 D 总计	\$1,698,034	\$2,906,450	\$8,732,793
F. Informatica Data Explorer 的成本	表 5, F	\$205,000	\$300,000	\$394,000
ROI 占 IDE 投资的百分比	$100 (E - F) / F$	728%	869%	2116%

表 6: 总体 ROI

总结

在初始阶段使用了 Informatica Data Explorer 的数据迁移和数据集成项目通常比最初的预期时间提前完成，同时也比用其它方法快得多。优点包括降低直接项目成本，提升商业价值，降低项目超支和失败的风险以及提升数据最终的质量。使用 Informatica Data Explorer 获得的经济利益将是早期软件投资的许多倍。

INFORMATICA®

www.informatica.com.cn

北京办事处

地址：北京市朝阳区建国门外大街乙 12 号
LG 双子座大厦东塔 19 层 1906 室

邮编：100022

电话：86-10-5879 3366

传真：86-10-5879 3130

上海办事处

地址：上海市世纪大道 1090 号
斯米克大厦 14 楼

邮编：200120

电话：86-21-5835 0596

传真：86-21-5835 6037

广州办事处

地址：天河区体育东路 138 号
金利来数码网络大厦 3 层 321 室

邮编：510620

电话：86-20-2886 0646

传真：86-20-3878 1801