



Datastage

目录

1、 IBM 公司简介	3
2、 ETL 工具—DATASTAGE 产品特性.....	4
功效强大的解决方案	4
完善的信息来源	4
实时数据采集	4
先进的维护与开发功能	4
完善的开发环境	5
迅速上手	6
强大的预设功能	6
重复使用、版本管理与分享	6
以事件为基础的调度与监控功能	6
扩充性最高的平台	7
开放式的延伸环境	7
弹性的并行处理功能	7
分割与自动重新分割	7
广泛的并行处理硬件支持	8
3、 DATASTAGE 产品优势.....	9

1、IBM 中国公司简介

IBM，即国际商业机器公司，1911年创立于美国，是全球最大的信息技术和业务解决方案公司，业务遍及170多个国家和地区。2007年，IBM公司的全球营业收入达到988亿美元。

在过去的九十多年里，世界经济不断发展，现代科学日新月异，IBM始终以超前的技术、出色的管理和独树一帜的产品领导着全球信息工业的发展，保证了世界范围内几乎所有行业用户对信息处理的全方位需求。

IBM与中国的业务关系源远流长。早在1934年，IBM公司就为北京协和医院安装了第一台商用处理机。随着中国改革开放的不断深入，IBM在华业务日益扩大。80年代中后期，IBM先后在北京、上海设立了办事处。1992年IBM在北京正式宣布成立国际商业机器中国有限公司。到目前为止，IBM在中国的办事机构进一步扩展至哈尔滨、沈阳、深圳、南京、杭州、成都、西安、武汉、福州、重庆、长沙、昆明、乌鲁木齐、济南、天津、郑州、合肥、南昌、南宁、宁波、石家庄、太原、苏州等26个城市，从而进一步扩大了在华业务覆盖面。伴随着IBM在中国的发展，IBM中国员工队伍不断壮大，目前已达到14000人。除此之外，IBM还成立了10家合资和独资公司，分别负责制造、软件开发、服务和租赁的业务。

IBM非常注重对技术研发的投入。1995年，IBM在中国成立了中国研究中心（2006年更名为IBM中国研究院），是IBM全球八大研究中心之一，现有200多位中国的计算机专家。随后在1999年又率先在中国成立了软件开发中心，现有3000多位中国软件工程师专攻整合中间件，数据库，Linux等领域的产品开发。

二十多年来，IBM的各类信息系统已成为中国金融、电信、冶金、石化、交通、商品流通、政府和教育等许多重要业务领域中最可靠的信息技术手段。IBM的客户遍及中国经济的各条战线。与此同时，IBM在多个重要领域占据着领先的市场份额，包括：服务器、存储、服务、软件等。

对于IBM在中国的出色表现和突出贡献，媒体给予了IBM十分的肯定。IBM先后被评为“中国最受尊敬企业”、“中国最受尊敬的外商投资企业”、“中国最具有价值的品牌”、“中国最佳雇主”等。2004年，IBM中国公司被《财富》杂志中文版评选为“中国最受赞赏的公司”，并荣居榜首。2005至2007年，IBM连续三次被中国社会工作协会企业公民工作委员会授予“中国优秀企业公民”荣誉称号。

2、ETL 工具—DataStage 产品特性

功效强大的解决方案

DataStage 可搜集、整合与转换大量的数据，不管是简单或极为复杂的数据结构，DataStage 都可以胜任。DataStage 不仅可管理实时存取的数据，还可以处理每天、每周、每个月定期更新的大量数据。

完善的信息来源

DataStage 可通过一个 job 提供各种数据来源系统与目标系统的支持，包括：

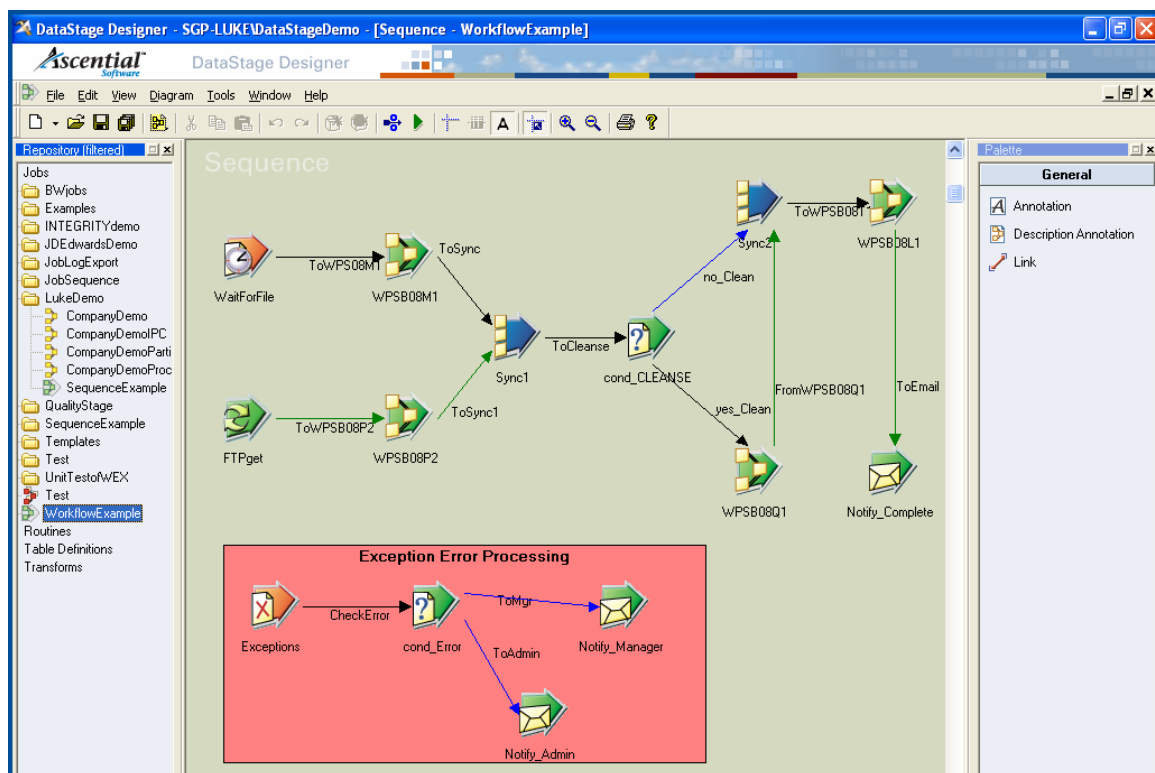
- 文本文件
- 复杂的 XML 数据结构
- SAP、PeopleSoft 等 ERP 系统
- 各种数据库的数据，包括已分区的数据库，如 Oracle、DB2 EE/EEE/ESE (含/不含 DPF)、Informix、Sybase、Teradata、SQL Server 以及使用 ODBC 存取的数据。
- WEB Service
- SAS

实时数据采集

DataStage 可以在整合了大量数据的同一个平台上实时采集或传送数据。如果客户有实时数据交换的需要，可以触发系统预先定义的 DataStage 的数据采集功能，自动进行数据采集、转换并收集到数据仓库。

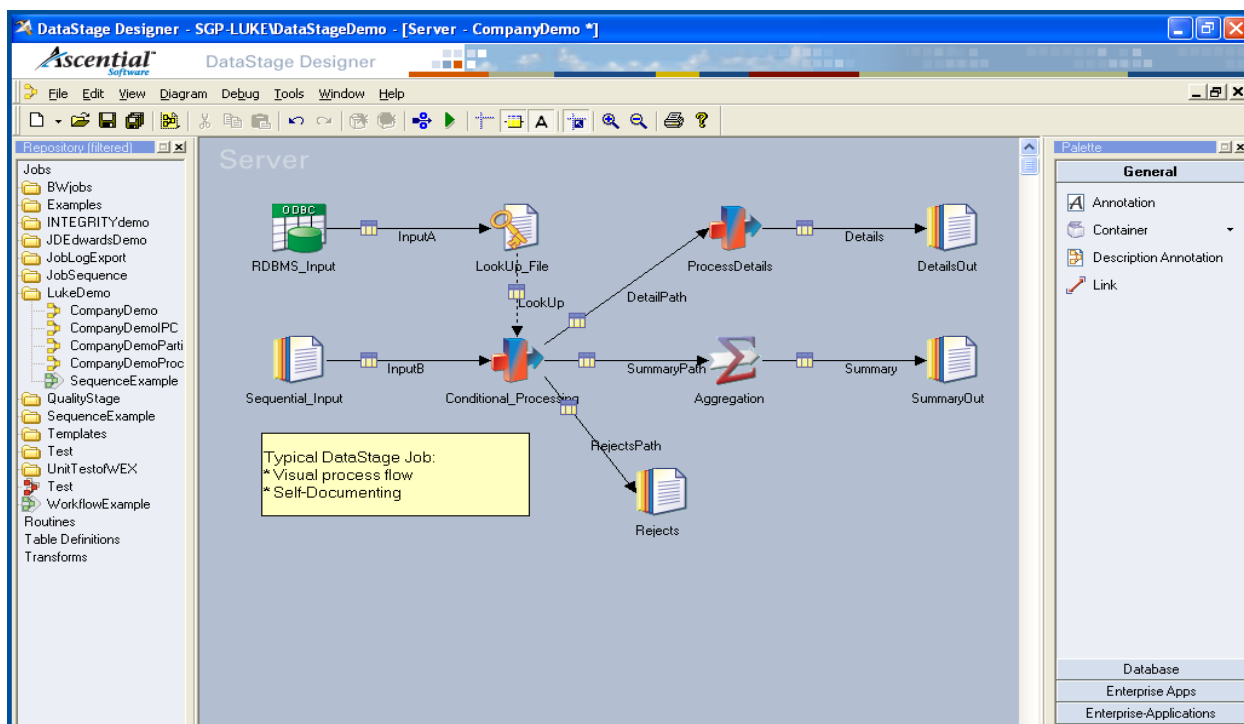
先进的维护与开发功能

DataStage 强大的架构让开发人员在建立、部署、更新与管理数据整合基础结构时享有最快的速度、最大的灵活性以及最高的效率。DataStage 的功能可提高产能、缩短学习周期、简化管理流程并让企业充分利用开发资源，以缩短数据整合应用程序的开发与维护周期。有了 DataStage，企业便可花很少时间在整合的开发上，并专注在如何制定出可获得最高报酬率的策略。



完善的开发环境

“随心所欲地工作”可说是 DataStage 设计的最佳写照。开发人员通过数据流模式的程序设计功能建立一连串的数据流程图。完整的图形接口允许开发人员使用 GUI 拖放功能来描绘出企业内部的数据流。多样性的描述语言、强大的调试排错功能以及允许使用外部程序代码的开放式应用程序设计接口 (API) 都可让开发人员在设计时更得心应手。



迅速上手

当您建立第一个工作时，DataStage 会提供向导似的 Intelligent Assistants 小帮手。Intelligent Assistants 中提供创建渐变维的工作(支持类型 1、2 与 3)之类的范例。工作模板以及预设的组件也可帮助您更快速地开发。

强大的预设功能

DataStage 提供业界中最广泛的数据整合开发环境，它的链接库提供了 400 多种内建函数与程序，供开发人员选取使用。

重复使用、版本管理与分享

DataStage 提供可重复使用现有数据整合商业逻辑的功能，因此可以缩短开发周期。此功能以容器的概念为基础，在同一个容器内建立的工作和元数据都可以被其它工作分享与使用。版本管理功能则可支持其它开发人员或 DataStage 服务器的开发、测试与部署作业。

以事件为基础的调度与监控功能

系统管理员可以使用内建的日历或指令行调度功能来排定 DataStage 工作。您也可以直接使用各种企业日程管理工具来调用 DataStage。您可以通过整合到现有基础结构的 GUI、文本与 XML 来取得详细的工作资料，以协助您检查问题、调整行程与监控作业。

扩充性最高的平台

DataStage 提供可处理大数据量的高性能处理程序，帮助企业解决大型业务的问题。藉由多处理器平台的平行处理功能，**DataStage** 企业版便可处理日益增加的数据量并减少处理的窗口。**DataStage** 不仅缩短了处理时间，并加速了大数据量的整合作业。使用者也不需要为同步执行的应用程序编写程序代码，大幅提升了开发人员的产能。编写程序代码通常需要由许多专家合力完成，是一项高成本的工作。现在有了 **DataStage**，使用者只需要使用顺序逻辑便可完成开发，而部署的设定会自动新增您要的并行处理级别。

开放式的延伸环境

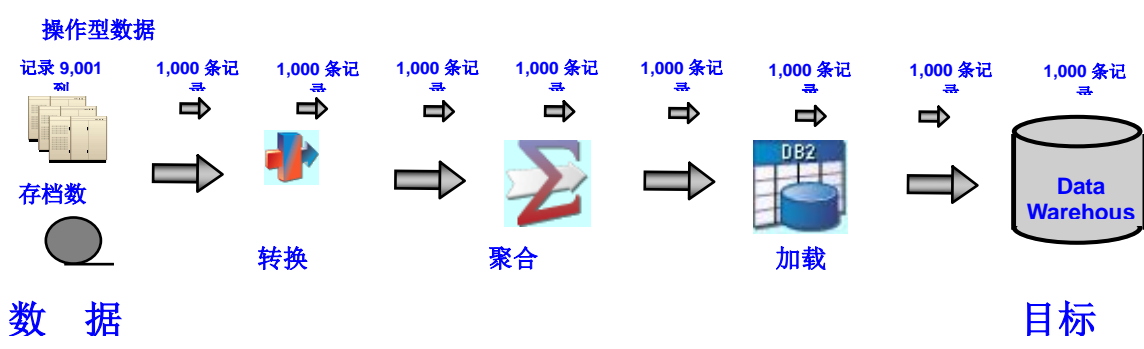
DataStage 企业版是强大的开放式环境，它不仅支持 **Ascential ProfileStage™** 与 **Ascential QualityStage™**等 **Ascential** 整合产品，还支持 **SAS** 等合作厂商的产品。此外，**DataStage** 还支持自定义程序代码，允许企业重复使用自行开发的程序代码，并同步处理所有数据，而不受到数据量的限制。

弹性的并行处理功能

灵活的配置文件允许使用者无须变更应用程序的程序代码即可定义并行处理的级别。因此，若企业需要提高并行级别，只需要对配置文件做一个变更，便可让使用者在早上同步调用 2 个应用程序、在下午同步调用 32 个应用程序并在晚上同步调用 128 个应用程序。

分割与自动重新分割

DataStage 的并行处理技术使用分区处理技巧，将大型的整合工作切割为分区(“分割并行处理”)并将这些分区同时传送给所有处理器(“管道并行处理”)。管道与分区并行处理的组合可提供完整的并行处理功能(效能提升与处理器数量成正比)，并让硬件成为性能的唯一决定性因素。不过下游的处理可能需要以不同方式分区数据。例如原本根据客户姓氏转换的作业在传送到下游的处理程序时可能会变成必须根据客户的信用卡号码将数据加载到仓库(详细信息请参阅以下说明的并行数据库接口)。透过自动重新分区数据的功能，使用者便可根据下游处理数据分割的需求重新分区处理流程上的数据，而不需要将数据储存在磁盘上。



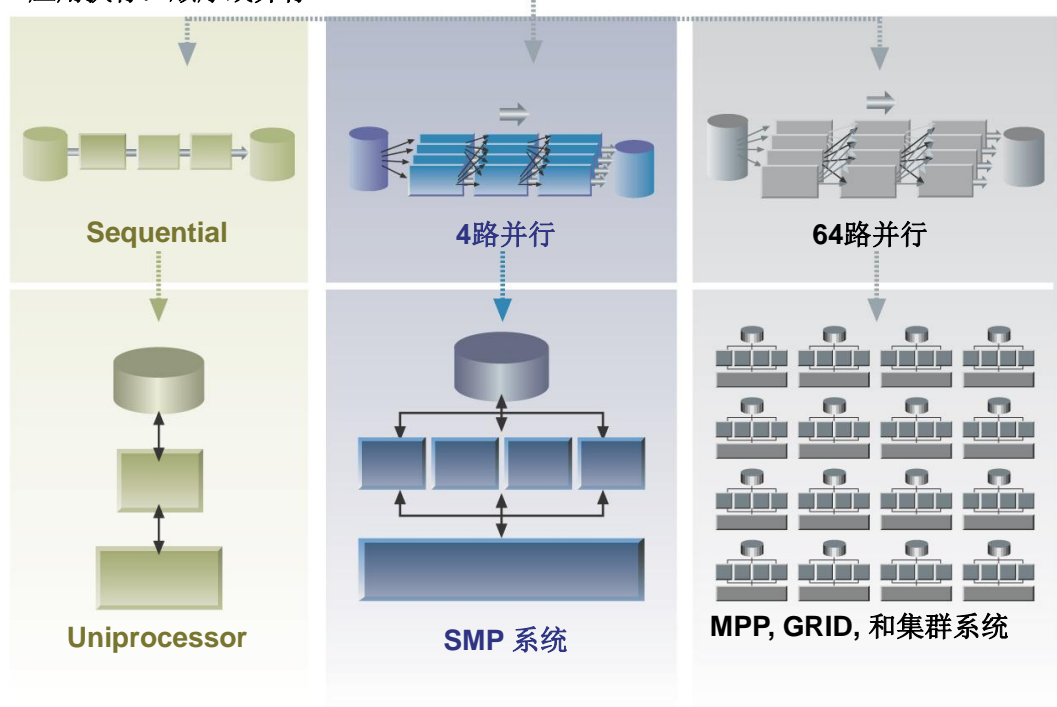
广泛的并行处理硬件支持

DataStage 可从 SMP 与 SMP 集群扩充到拥有数百个处理器的 MPP。重要的整合应用程序可随企业发展的规模同步扩充。将来数据仓库需要进行数据量扩展，以及应用扩展时，数据采集工具在这方面的能力很重要。

集合应用：一个数据流视图



应用执行：顺序或并行



3、DataStage 产品优势

在目前同类产品市场上，同为高端产品的 IBM DataStage 和 Informatica PowerCenter 对比情况如下：

DataStage 在同类产品中独有的特点

1、SOA 面向企业服务架构

- 端到端元数据管理
- 任意数据源以及各种应用软件包的完全连接
- 无限扩展的处理能力，高性能处理大规模的数据
- 实时整合服务(RTI)，使用网络服务(.NET; Java; EJB)发送服务(JMS)，实现 on-demand
- 与现有 IT 投资完全兼容的协同操作，全面增强数据整合能力(WebSphere; TIBCO; SeeBeyond)，充分利用原有 IT 投资。
- 企业用户在任何时间、任何地方使用任何服务，关键在于让用户整合其全部资源。

2、并行处理能力

- 支持所有硬件配置；

SMP，特别是 4-64 个处理器、大型 MPP 系统、具有上百个处理器的集群环境。

- 只需增加处理器或节点，无需改变应用，无需手工编码并行处理或重新编译就可以升级。现有应用与数据整合处理和并或并行；
- 外部配置文件指定硬件配置和资源；
- 支持并行 RDBMS，包括 IBM DB2 UDB, Oracle, Teradata;
- 设计顺序进行，开发平行；
- 无限的可升级性。

3、实时数据整合 (Realtime Integration Services)

唯一提供与交易中间件的按需求整合服务。实时集成服务 (RTI 服务)，使用网络服务 (.NET 和 Java)、Enterprise Java Bean(EJB)或 Java 信息发送服务 (JMS)完成实时数据集成，实现按需访问关键性商业信息；和现有 IT 投资完全兼容

的协同操作性（包括 IBM WebSphere，TIBCO 和 SeeBeyond 集成平台），以全面增强数据整合能力；增强了数据的统一性，降低了项目开发的时间和成本。

4、DataStage TX

- 自动处理和整合事务型和操作型系统的数据
- 根据消息内容自动决定路由，消除开销大的、耗时的手工编程
- 在面向服务的体系结构中协同工作，充分发挥现有和未来 IT 增值作用
- 兼容控制数据传输和整合的行业及政府标准、规则（EDI，HIPAA，SWIFT），比如银行之间的数据处理。

Datastage 优势	具体阐述	为用户带来的商业价值
数据整合套件	数据轮廓，数据质量，元数据，以及 ETL/数据转换，所有的产品都来自于同一家软件厂商。在实时性，大批量数据处理，和业务在线处理方面提供连续的产品系列。	单一的软件提供商和统一、连续的技术投资缩短了项目的开发周期。可重复使用的企业应用服务满足了数据整合的需求。
可参照的开放企业架构	提供领先的开放企业架构（playbooks/blueprints）和为 ISV 伙伴、用户提供数据整合指导技术论文。	客户化，改善用户 IT 投资，允许快速实施数据整合，和减少时间、资源，及人力。
方法论和最佳实践	不管小型项目，还是大型企业级项目，都可利用可重复使用的菜单方式应用进行数据整合，涵盖了数据轮廓，数据质量，元数据管理，ETL/数据转换。	快速发布预定义角色和各种功能—减少成本，避免风险，和为你的团队积累经验。
可证实的优秀的软件功能和性能	基于全球超过 3000 个客户的广泛客户基础，在所有的产业领域，和全世界不同的地方，为包括很多大型的数据中心环境提供最佳软件功能和性能。	为客户成功的数据整合实施带来价值，并具有高扩展性。
广泛深入的数据分析	为数据内容和数据架构提供完整的视图，提高用户 ROI，并使数据质量贯穿在整个企业中。	易于理解，评估和管理数据质量，基于完整性，有效性，准确性，和相关性。
数据轮廓集成在数据整合套件中	在数据整合生命周期里缩减重复工作，增加准确性。	可衡量的企业 ROI，潜在商业价值，和节约成本。

表现性和可扩展性	基于大容量的分析数据和数据源以减少数据冗余。	减少成本，提高生产力，缩减时间周期。
提高数据质量和数据控制范围	调查，标准化，和增强数据范围，对输入的数据易于跟踪调试，支持二字节的全局区域业务稽核。	完整性，准确性，和适时性。
优秀的数据匹配引擎	使用概率统计提高 2—4% 的匹配率。很小的匹配出错几率。可适用在任何数据实体上。调节出错匹配率以满足用户的需求。	可衡量的 ROI，增加商业价值，和节约成本。
基于面向服务的架构	在面向服务的架构中采用离散的匹配和标准化的服务	为企业提供统一质量规则集
优异的性能和扩展性	同竞争对手相比，总体处理速度快 5 倍。 利用统一规则集，用户根据需要进行批处理或实时交易。	扩展性使得更多得数据质量问题被快速的处理。
转换	在产品的范围内能够作的更多。复杂的转换处理能力，不需要手工编码方式完成。数据转换提供了强大的支持，可从自上而下，或自下而上的方式实现设计。	可衡量的 ROI，快速开发使得成本降低，相对比于手工编码或其他方式。
快速，大规模的处理	为大规模处理提供广泛的功能，包括： <ul style="list-style-type: none"> ● 合并（merging），排序（sorting），联接（joining），过滤（filtering），转换（transforming）组件 ● 对于文件系统和关联表采用平行 I/O ● 自动的数据管道和数据分区分配 ● GRID, SMP, MPP 支持 	满足用户服务协议。优化资源投资。
平台互联性和互操作性	不用花费时间在那些不能支持应用数据增长，变化的整合需求，相联需要从源到端的整合工具上，Ascential 公司的企业整合组件具有很好的平台互联性和互操作性	在所有的平台整合可靠性方面，处理所有的数据。
业务信息的传递和管理工作	赋予创立，管理，和分析标准信息传递和维护数据管理的能力。	一致的和达成跨业务和 IT 的数据标准。使得数据具有自完善的能力，为用户提供更好的决策和数据所属。

跨工具的影响分析和数据依赖	对于跨多种工具的数据整合和数据依赖，提供数据影响分析和集成能力。	对于不断更新的业务需求，用户的业务能够快速的发展，并且具有更大的适用性。
不依赖于硬件的全方向交换处理	提供跨多种工具的元数据共享和再使用的集成能力	对于所有的数据整合项目，加强了数据的连续性和数据的可靠性。
基于模块的可扩充性	提供图形化的模块处理，管理新生成的或遗留的元数据，不需要手工编码。	在广泛的范围内提供再使用，管理和分析元数据能力。
面向服务的数据整合	RTI 容许数据整合工作被发布在共享的业务目标中进行使用，不需要了解 Ascential 的整个数据整合组件，就可以使用用户的需求。	提高数据的连续性，可使用性，和可靠性。缩减开发的成本，易于再使用的维护。
快速服务实施	通过使用 RTI，可以实施快速的服务，不需要手工编码。	对市场需求相应快速和简化的技术复杂性。
灵活的捆绑	RTI 采用的服务可以支持多种捆绑，包括 JMS，EJB，以及基于 Web 的服务。	易于再使用和采用多样性的需求和标准。
已被验证的高可靠性基础	RTI 提供企业级的数据整合服务架构，具有无限的扩展，高可靠，和容错的功能。	对任何的数据容量或加载进行性能的优化，降低风险和高产出。

其他因素对比

对比因素		Datastage
原厂商分析	国际行业领导；国内必须有原厂商的分支机构	同行业中国际原厂商唯一目前在国有全资子公司；全球和中国的业务稳健增长
代理商开发使用经验	代理商有成功实施大型数据整合的经验	大型金融机构的成功实施案例如 BOC；交通银行；民生银行；光大银行；
客户群体对比	全球有广泛的同类用户	面向全球大型企业，提供全面的数据整合产品技术；可支持到数据仓库
策略伙伴对比		IBM 全球代理产品；SAP；Peoplesoft OEM；
本地服务分析	原厂商能够提供本地的服务支持	总代理提供第一线的服务，原厂商随时提供国际水平的技术支持，提供 7x24 小时的电子化服务
中国用户分析		2003 年用户超过 50 家

金融行业		5 大银行总行中有 3 家； 中国银行（2004 年新客户）交通银行（2004 年新客户） 成功的为金融界客户提供服务。在光大银行、民生银行；交通银行等支持客户获得好评。
电信行业		仅以中国移动为例，中国移动首选产品；在全国 31 个点中已经有 20 个省份安装使用：上海、江苏、浙江、山东、辽宁、吉林、安徽、广东、四川、北京、重庆.....
其他企业		上海通用汽车已经成功上线，并获得客户好评
保险行业		平安保险
本地服务分析		<p>中国大部分客户优选 Ascential 的主要原因：</p> <ul style="list-style-type: none"> ● 确保项目成功是关键因素 ● 产品技术标准化与功能 ● 厂商稳定性与长期合作能力 ● 本地服务能力，原厂商支持能力 ● 产品价格 ● 国际领先厂商 ● 产品本地应用成熟度 <p><u>客户群体的共同认可程度。大多数用户的选择表明了共同认可，技术成熟；服务有保证。</u></p>