

您的潜力，我们的动力

Microsoft
微软(中国)有限公司

SQL2005企业级ETL平台特性与组件

—— Lesson 2

金立钢

BI咨询顾问
北京迈思奇科技有限公司

介绍

- 北京迈思奇科技有限公司 (Minesage Co.,Ltd.)
 - 核心业务: 面向各行业提供商业智能解决方案;
 - 提供商业智能领域内的咨询和培训。
 - 美国微软公司指定的数据挖掘解决方案提供商。

咨询部经理

咨询培训: 企业应用; 产品发布;

项目实施: 行业解决方案; BI应用系统;

5年的数据分析工作经验

www.minesage.com

ljin@minesage.com

ljin.bloggerhome.com

您的潜力. 我们的动力

Microsoft
微软(中国)有限公司

收听本次课程需具备的条件

- 对ETL有一定的了解
- 了解SQL Server 2005的开发环境

Level 200

本次课程内容包括

- Integration Services新增组件
 - 查找
 - 模糊查找
 - 模糊分组
- Integration Services其他任务及转换介绍

案例

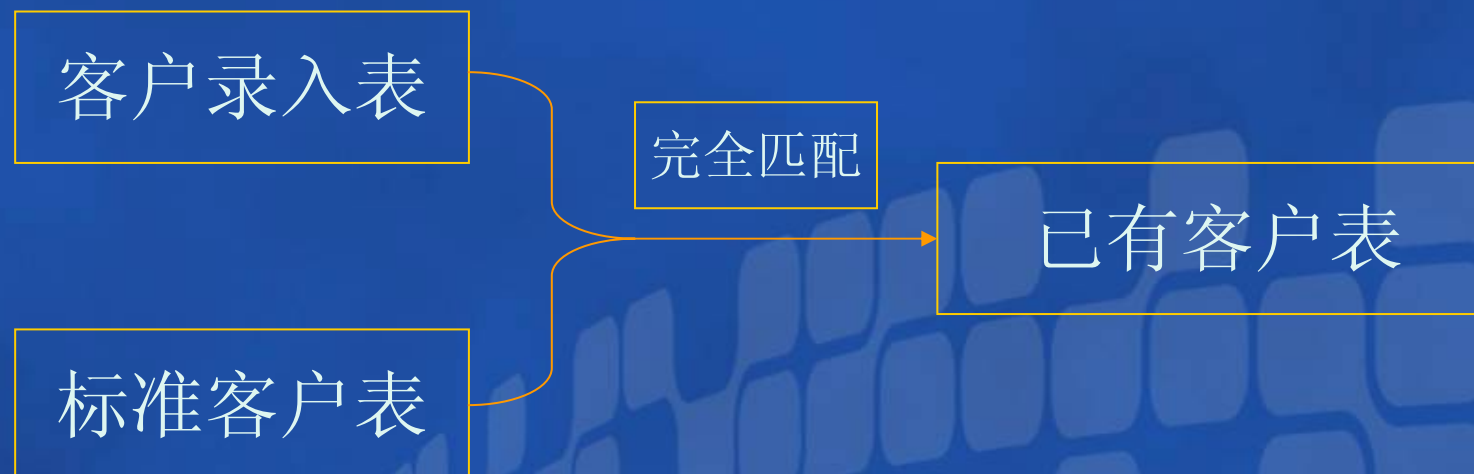
一个公司做了一次长达一年的客户问卷调查，调查结束后，公司想知道参加这次调查的客户哪些是老客户，哪些是新客户？

First Name
Last Name
AddressLine1
City
State
Country Region

查找

您的潜力. 我们的动力

Microsoft
微软(中国)有限公司



您的潜力，我们的动力

Microsoft[®]
微软(中国)有限公司

DEMO

查找

有问题的数据

FirstName	LastName	AddressLine1	City	State	CountryRegion
Jon	Yang	3761 N. 14th St	Rockhampton	Queensland	Australia
E.	Huang	2243 W St	Seaford	Victoria	NULL
Reuben	Tores	5844 Linden	Hobart	NULL	NULL
Chris	Z.	1825 Village Place	NR	NS Wales	Australia
Pannarat	Pattanapitakkul	1 Microsoft Way	Redmond	Washington	USA
pannarat	pattanapitakkul	Microsoft Way	Redmond	Washington	USA
Pannarat	Paddanapitakul	1 Microsoft Way	Redmond	Washington	NULL
Pan	Pattanapitakkul	1 Microsoft Way	Redmond	NULL	USA
pannarat	pattanapitakkul	One microsoft way	redmond	washington	u. s. a.
Panarat	Pattanapittakul	One Microsoft Wy.	Redmond	Wash	USA
Göktuğ	Oğuz	111 Broadway Avenue	Seattle	Washington	USA
Gökktuğ	Oğuz	111 Broadway	Seattle	WA	NULL
goktug	ogguz	111 broadway ave	seattle	NULL	NULL
Goktuğ	Oğuz	111 Broadway Ave	Seattle	WA	USA
Göktuğ	Oğuz	111 Broadway Ave.	NULL	Washington	USA

数据可能是有问题的

- 拼写错误
- 截断
- 缺少或插入的标记
- 空字段
- 意外的缩略语
- 其他不规则问题

模糊查找

您的潜力, 我们的动力

Microsoft
微软(中国)有限公司



标准客户录入表 (相似匹配)

CustomerID	FirstName...	LastName ...	AddressLine1 (Input)	City (Input)	State (Input)	CountryRegion (Input)
11000	Jon	Yang	3761 N. 14th St	Rockhampton	Queensland	Australia
11001	E.	Huang	2243 W St	Seaford	Victoria	NULL
11002	Reuben	Tores	5844 Linden	Hobart	NULL	NULL

您的潜力，我们的动力

Microsoft[®]
微软(中国)有限公司

DEMO

模糊查找

新客户录入表

FirstName	LastName	AddressLine1	City	State	CountryRegion
Chris	Z.	1825 Village Place	NR	NS Wales	Australia
Pannarat	Pattanapitakkul	1 Microsoft Way	Redmond	Washington	USA
pannarat	pattanapitakkul	Microsoft Way	Redmond	Washington	USA
Pannarat	Paddanapitakul	1 Microsoft Way	Redmond	Washington	NULL
Pan	Pattanapitakkul	1 Microsoft Way	Redmond	NULL	USA
pannarat	pattanapitakkul	One microsoft way	redmond	washington	u. s. a.
Panarat	Pattanapittakul	One Microsoft Wy.	Redmond	Wash	USA
Göktuğ	Oğuz	111 Broadway Avenue	Seattle	Washington	USA
Gökktuğ	Oğuz	111 Broadway	Seattle	WA	NULL
goktug	ogguz	111 broadway ave	seattle	NULL	NULL
Goktuğ	Oğuz	111 Broadway Ave	Seattle	WA	USA
Göktuğ	Oğuz	111 Broadway Ave.	NULL	Washington	USA

模糊分组

您的潜力. 我们的动力

Microsoft
微软(中国)有限公司



您的潜力. 我们的动力

Microsoft
微软(中国)有限公司

唯一客户表

FirstName	LastName	AddressLine1	City	State	CountryRegion
Chris	Z.	1825 Village P...	NR	NS Wales	Australia
Pannarat	Pattanapitakkul	1 Microsoft Way	Redmond	Washington	USA
Goktuğ	Oğuz	111 Broadway Ave	Seattle	WA	USA

您的潜力，我们的动力

Microsoft[®]
微软(中国)有限公司

DEMO

模糊分组

Integration Services 其他任务 及转换介绍

- 控制流
 - 容器
 - 任务
- 数据流
 - 源
 - 转换
 - 目标

容器

您的潜力. 我们的动力

Microsoft
微软(中国)有限公司

- For 循环
- For each 循环
- 序列容器

您的潜力，我们的动力

Microsoft[®]
微软(中国)有限公司

DEMO

序列容器

任务

您的潜力. 我们的动力

Microsoft
微软(中国)有限公司

- 数据流任务
- 数据准备任务
- 工作流任务
- SQL Server 任务
- 脚本任务
- Analysis Services 任务
- 维护任务

您的潜力. 我们的动力

Microsoft
微软(中国)有限公司

数据准备任务

- 文件系统任务
- FTP 任务
- Web 服务任务
- XML 任务

您的潜力. 我们的动力

Microsoft
微软(中国)有限公司

workflow 任务

- 执行包任务
- 执行 DTS 2000 包任务
- 执行进程任务
- 消息队列任务
- 发送邮件任务

SQL Server 任务

- 大容量插入任务
- 执行 SQL 任务
- 传输数据库任务
- 传输错误消息任务
- 传输作业任务
- 传输登录名任务
- 传输主存储过程
- 传输 SQL Server 对象任务

您的潜力. 我们的动力

Microsoft
微软(中国)有限公司

脚本任务

- ActiveX 脚本任务
- 脚本任务

您的潜力，我们的动力

Microsoft[®]
微软(中国)有限公司

DEMO

脚本任务

您的潜力. 我们的动力

Microsoft
微软(中国)有限公司

Analysis Services 任务

- Analysis Services 处理任务
- Analysis Services 执行 DDL
- 数据挖掘查询任务

维护任务

- “备份数据库”任务
- 检查数据库完整性
- 执行 SQL Server 代理作业
- 执行 T-SQL 语句任务
- “清除历史记录”任务
- “通知操作员”任务
- “重新生成索引”任务
- “重新组织索引”任务
- 收缩数据库任务
- “更新统计信息”任务

您的潜力. 我们的动力

Microsoft
微软(中国)有限公司

数据流——源

- Data Reader 源
- Excel 源
- 平面文件源
- OLE DB 源
- 原始文件源
- 脚本组件
- XML 源

您的潜力. 我们的动力

Microsoft
微软(中国)有限公司

数据流——转换

- 商业智能转换
- 行转换
- 行集转换
- 拆分和联接转换

您的潜力. 我们的动力

Microsoft
微软(中国)有限公司

商业智能转换

- 模糊分组转换
- 模糊查找转换
- 字词提取转换
- 字词查找转换
- 数据挖掘查询转换

您的潜力. 我们的动力

Microsoft
微软(中国)有限公司

行转换

- 字符映射表转换
- 复制列转换
- 数据转换
- 派生列转换

行集转换

- 聚合转换
- 排序转换
- 百分比抽样转换
- 行抽样转换
- 透视转换
- 逆透视转换

透视转换

下面的关系图显示对 Product 列透视数据之前的数据集。

用途	SetKey	PivotKey	PivotedVal
列名	Cust	Product	Qty
数据记录	Kate	Ham	2
	Kate	Soda	6
	Kate	Milk	1
	Kate	Beer	12
	Fred	Milk	3
	Fred	Beer	24
	Fred	Chips	2

下面的关系图显示对 Product 列透视数据之后的数据集。

源列	Cust	Qty	Qty	Qty	Qty	Qty
透视键值		Ham	Soda	Milk	Beer	Chips
列名	Cust	HAM	Soda	Milk	Beer	Chips
数据记录	Kate	2	6	1	12	
	Fred			3	24	2

您的潜力. 我们的动力

Microsoft
微软(中国)有限公司

拆分和联接转换

- 条件性拆分转换
- 多播转换
- Union All

数据流——目标


- 数据挖掘模型定型目标
- DataReader 目标
- 维度处理目标
- Excel 目标
- 平面文件目标
- OLE DB 目标
- 分区处理目标
- 原始文件目标
- 记录集目标
- 脚本组件
- SQL Server Mobile 目标
- SQL Server 目标

获取更多MSDN资源

- **MSDN中文网站**
<http://www.microsoft.com/china/msdn>
- **MSDN中文网络广播**
<http://www.msdnwebcast.com.cn>
- **MSDN Flash**
<http://www.microsoft.com/china/newsletter/case/msdn.aspx>
- **MSDN开发中心**
<http://www.microsoft.com/china/msdn/DeveloperCenter/default.msp>



Question & Answer

如需提出问题，请单击“提问”按钮并在随后显示的浮动面板中输入问题内容。一旦完成问题输入后，请单击“提问”按钮。

 **问题和解答 (无问题)** ▲ ×

在此会议中尚未解答任何问题。

要向演示者提问，请在此处键入问

提问(A)

删除(D)

问题管理器(Q)

您的潜力，我们的动力

Microsoft®
微软(中国)有限公司

Microsoft®

msdn


MSDN Webcasts

您的潜力. 我们的动力

Microsoft
微软(中国)有限公司

客户录入表

FirstName	LastName	AddressLine1	City	State	CountryRegion
Jon	Yang	3761 N. 14th St	Rockhampton	Queensland	Australia
E.	Huang	2243 W St	Seaford	Victoria	NULL
Reuben	Tores	5844 Linden	Hobart	NULL	NULL
Chris	Z.	1825 Village Place	NR	NS Wales	Australia
Pannarat	Pattanapitakkul	1 Microsoft Way	Redmond	Washington	USA
pannarat	pattanapitakkul	Microsoft Way	Redmond	Washington	USA
Pannarat	Paddanapitakul	1 Microsoft Way	Redmond	Washington	NULL
Pan	Pattanapitakkul	1 Microsoft Way	Redmond	NULL	USA
pannarat	pattanapitakkul	One microsoft way	redmond	washington	u. s. a.
Panarat	Pattanapittakul	One Microsoft Wy.	Redmond	Wash	USA
Göktuğ	Oğuz	111 Broadway Avenue	Seattle	Washington	USA
Gökktuğ	Oğuz	111 Broadway	Seattle	WA	NULL
goktug	ogguz	111 broadway ave	seattle	NULL	NULL
Goktuğ	Oğuz	111 Broadway Ave	Seattle	WA	USA
Göktuğ	Oğuz	111 Broadway Ave.	NULL	Washington	USA

[返回](#)

您的潜力. 我们的动力

Microsoft
微软(中国)有限公司

标准客户表

CustomerID	FirstName	LastName	AddressLine1	City	State	CountryRegion
11000	Jon	Yang	3761 N. 14th St	Rockhampton	Queensland	Australia
11020	Jordan	King	7156 Rose Dr.	Metchosin	British Columbia	Canada
11024	Russell	Xie	8734 Oxford Place	Concord	California	United States
11031	Theresa	Ramos	6465 Detroit Ave.	Matraville	New South Wales	Australia
11030	Bethany	Yuan	636 Vine Hill Way	Cloverdale	South Australia	Australia
11093	Aimee	He	1833 Olympic Drive	Wollongong	New South Wales	Australia
11041	Amanda	Carter	5826 Escobar	Glendale	California	United States
11032	Denise	Stone	626 Bentley Street	Melbourne	Victoria	Australia
11058	Marc	Diaz	6645 Sinaloa	Findon	South Australia	Australia
11057	Carl	Andersen	6930 Lake Nadine Place	Lane Cove	New South Wales	Australia
11098	Mallory	Rubio	6452 Harris Circle	Rhodes	New South Wales	Australia
11078	Gina	Martin	1349 Sol St.	Cliffside	British Columbia	Canada
11100	Latasha	Navarro	6954 Ranch Road	Cloverdale	South Australia	Australia
11043	Nathan	Simmons	1170 Shaw Rd	Glendale	California	United States
11102	Julia	Nelson	2196 Coat Ct.	Sydney	New South Wales	Australia
11103	Cassie	Chande	9448 San Marino Ct.	Darlinghurst	New South Wales	Australia

[返回](#)

您的潜力. 我们的动力

Microsoft
微软(中国)有限公司

标准客户录入表 (完全匹配)

CustomerID	FirstName	LastName	AddressLine1	City	State	CountryRegion
11000	Jon	Yang	3761 N. 14th St	Rockhampton	Queensland	Australia

[返回](#)