



SSIS 专题二:

详解 SSIS 性能调整

详解 SSIS 性能调整

SQL Server 2005 采用 SSIS(SQL Server Integration Services)后, SQL Server 在数据库维护功能上有了一个很大的飞跃。微软 SQL Server 开发团队采用 SSIS 提供点击数据库维护、DTS 改善(数据库转换服务)和 SQL Server 2000 里的 SQLMaint。Database Maintenance Wizard 还用于建立最初的程序,但是在本地服务器上 SSIS 包作了大量的修正。和 SQLMaint 黑箱比较起来,新的界面更易于 DBA 们理解、并且也更安全。因此要求我们在更规则的基础上进行数据维护、需要我们全面提高性能、需要我们在提高性能方面花更多的精力。

SQL Server Integration Services 简化数据库维护

微软 SQL Server 开发团队采用 SSIS 简化数据库维护。提供 SQL Server 2000 中的 DTS 和 SQLMaint 最大的改善。数据库维护向导(Database Maintenance Wizard)还仍然在充当创建初始过程的角色,但现在我们对本地服务器上的 SSIS 包进行了修改。本章节中大家将会发现小小的维护对提高 SQL Server 性能发挥了重要作用。

- ❖ SQL Server Integration Services 简化数据库维护(一)
- ❖ SQL Server Integration Services 简化数据库维护(二)
- ❖ SQL Server Integration Services 简化数据库维护(三)

SSIS 中的错误信息

如果你 SQL Server Integration Services (SSIS)中收到了错误信息,你可能在 SQL Server 2005 客户机工具安装方面出了问题……

- ❖ SQL Server 2005 安装问题提示 SSIS 错误信息

SQL Server 2005 中调试 SSIS 包

在 SQL Server 2005 中调试 SSIS 包在寻找为你寻找问题原因提供了多种方法。本节中依照步骤分别介绍了两种方法：SSIS 包调试断点（breakpoints）和数据阅读器（data viewer）。这两种方法阐述了你的 SSIS 包内部工作的一些信息，以便于你对 SSIS 性能更好地进行性能调整。

❖ 在 SQL Server 中调试 SSIS 包（一）

❖ 在 SQL Server 中调试 SSIS 包（二）

调试 SSIS 中的 DTS 包

你是不是对是否将 DTS 包迁移到 SSIS 还在犹豫？作为一名 SQL Server DBA，你可能会经常修改 DTS 包，无论是修改逻辑上的错误、扩展 DTS 包性能或者是改变服务器、数据库或者是对象名称等等。文中你将了解各种在 SQL Server 2000 中的调试 DTS 包的方法。

❖ 在 SQL Server Integration Services 中运行 DTS 包

错误导入 Excel 数据到 SQL Server 2005

在将 Excel 表中的数据导入到 SQL Server 2005 时，是不是提示错误信息？如果有这种情况出现，你就应该对 Excel 表中的数据类型和 SSIS 进行匹配……

❖ 错误导入 Excel 数据到 SQL Server 2005

SQL Server Integration Services 简化数据库维护（一）

SQL Server 2005 采用 SSIS(SQL Server Integration Services)后，SQL Server 在数据库维护功能上有了一个很大的飞跃。微软 SQL Server 开发团队采用 SSIS 提供点击数据库维护、DTS 改善（数据库转换服务）和 SQL Server 2000 里的 SQLMaint。Database Maintenance Wizard 还用于建立最初的程序，但是在本地服务器上 SSIS 包作了大量的修正。和 SQLMaint 黑箱比较起来，新的界面更易于 DBA 们理解、并且也更安全。因此要求我们在更规则的基础上进行数据维护、需要我们全面提高性能、需要我们在提高性能方面花更多的精力。

虽然不能说 DBA 们应该放弃先进的维护方案，但他们也应该用 SSIS 的新功能简化维护程序。我赞赏创建这些方案的数据库管理员们，但是我也曾和那些没有将程序固定的 DBA 共事过。我相信 SSIS 界面提供了性能定期维护的基础。反过来，这项计划应确保整体 SQL Server 性能是一贯的高水平，和该公司匹配。

主要内容：

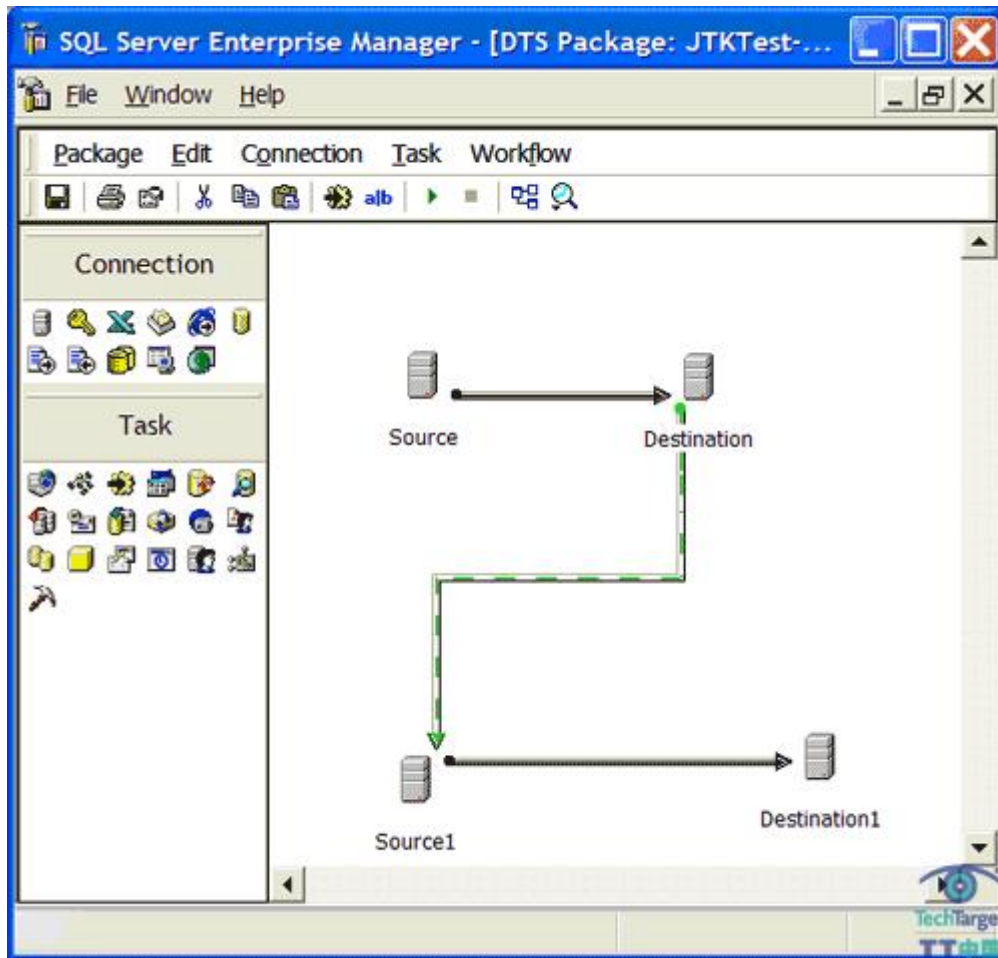
- DTS vs. SSIS：过去和现在
- SQL Server 2005 维护问题
- 其他 SSIS 资源

DTS vs. SSIS：过去和现在

说到数据库维护，微软一直保持了同样的功能，而且还提高了以下的技术。以下就是对这些变化的概括：

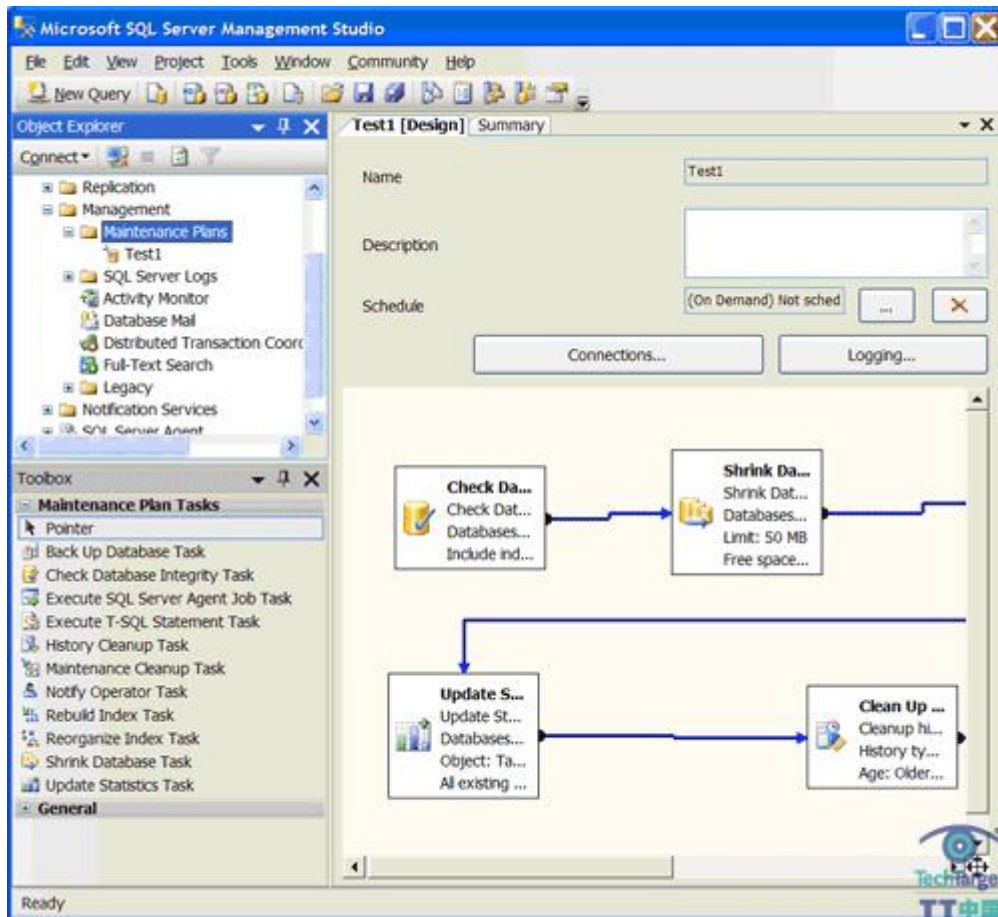
SQL Server 2000 Extraction, Transformation and Loading (ETL)

Data Transformation Services (DTS)



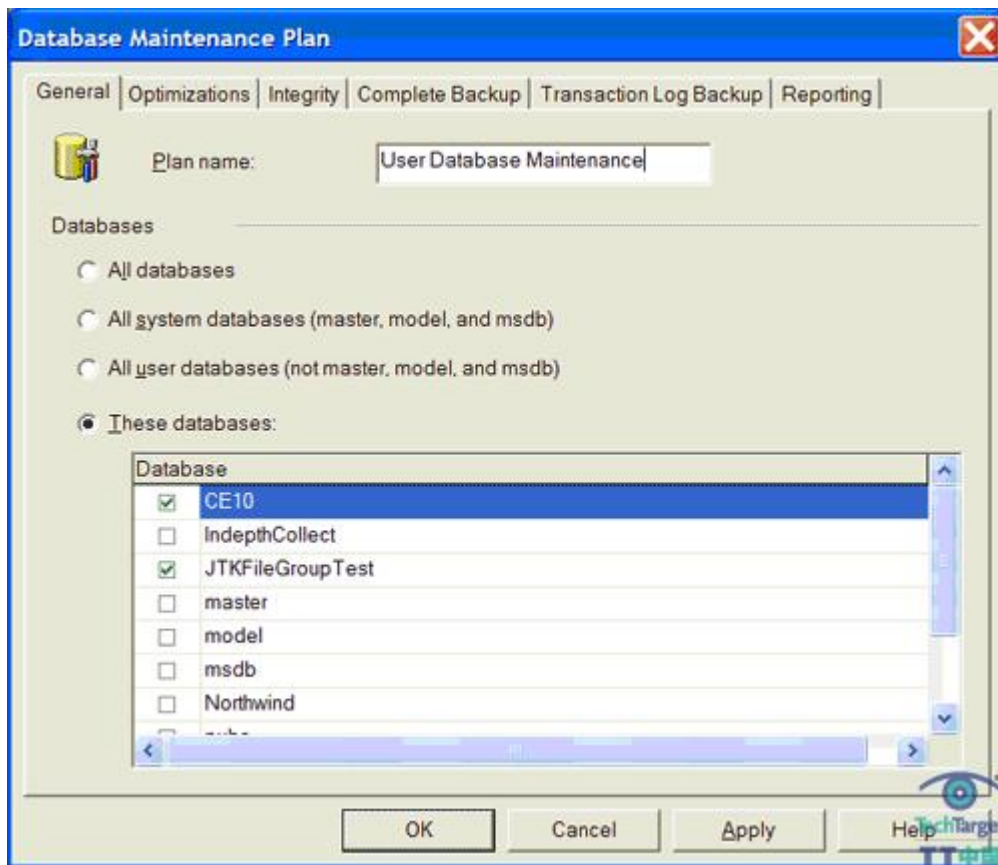
SQL Server 2005 Extraction, Transformation and Loading (ETL)

SQL Server Integration Services (SSIS)



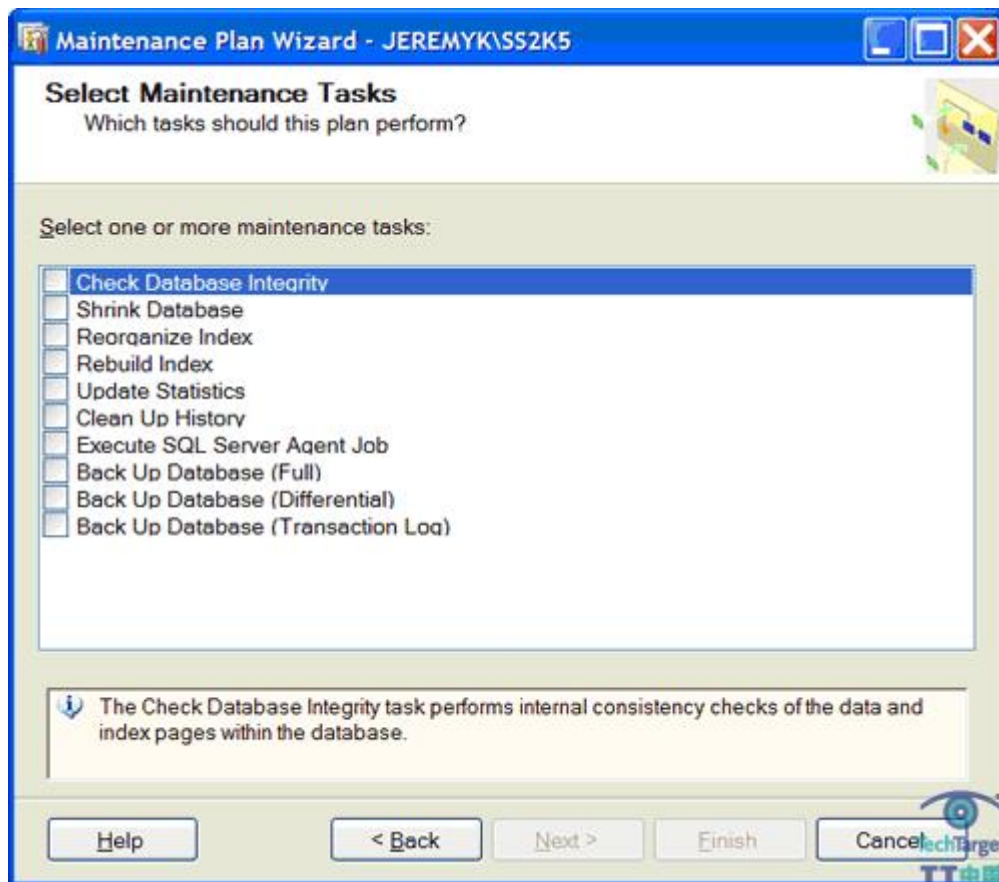
SQL Server 2000 点击式数据库维护

Database Maintenance Plans via SQLMaint



SQL Server 2005 点击式数据库维护

Management Studio's Maintenance Wizard



(作者: Jeremy Kadlec 译者: April 来源: TT 中国)

SQL Server Integration Services 简化数据库维护（二）

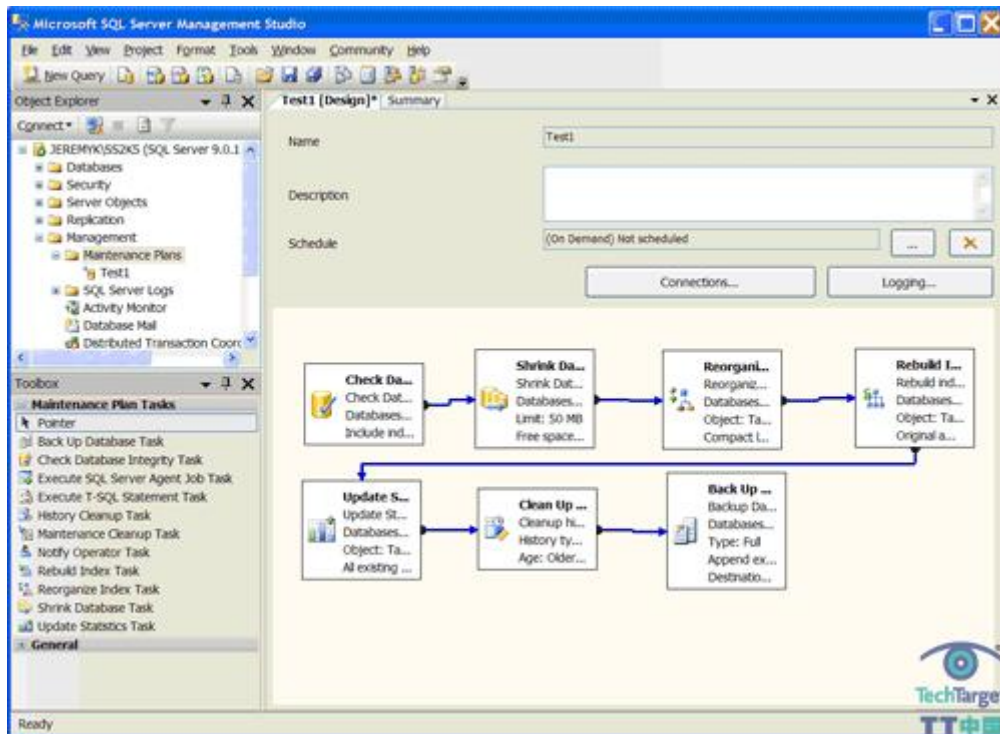
SQL Server 2005 维护问题

我怎么访问数据库维护功能？

- 1、打开 SQL Server 2005 Management Studio。
- 2、打开 Management 文件夹。
- 3、下拉至 Maintenance Plans 文件夹。
- 4、在 Maintenance Plans 文件夹上单击右键、选择 New Maintenance Plan。

基础界面是什么样的？

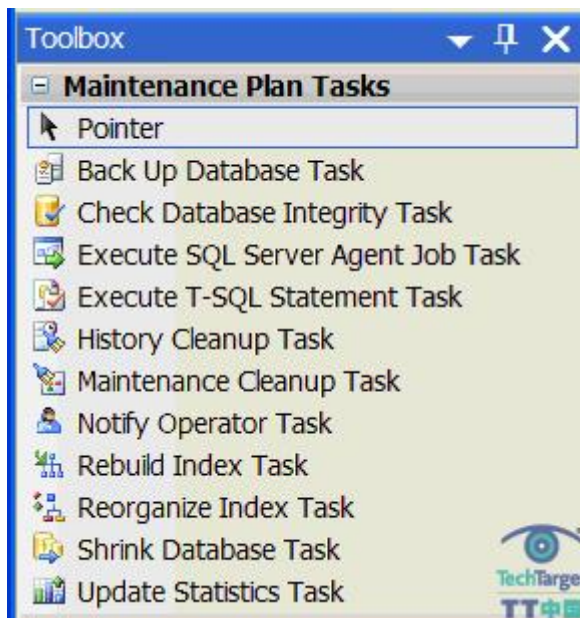
1. Management Studio Menus and Tool Bars - 最顶部
2. Object Explorer - 左上部分
3. Maintenance Plan Toolbox - 右下部分
4. General SSIS package information - 右上部分
5. SSIS workspace - 右下部分



SQL Server Management Studio -- Maintenance Plan

SSIS 数据库维护功能的定义是什么？

在界面的底部查询工具箱。你可以将这些项目从工具箱里拖进工作区。双击这些对象，你就可以对一对象进行设定，达到具体的维护要求。



Maintenance Plan Toolbox

能不能创建更多的 SSIS 数据库维护包？

可以，我们可以为不同的数据库、任务、周期等等设置 SSIS 包，这些 SSIS 包能够在同一个 SQL Server 上 100%独立于其他的 SSIS 包。

Job Schedule Properties - Test1-Schedule

Name: Jobs in Schedule:

Schedule type: ☒ Enabled

One-time occurrence

Date: Time:

Frequency

Occurs:

Recurs every: week(s) on

☐ Monday ☐ Wednesday ☐ Friday ☐ Saturday

☐ Tuesday ☐ Thursday ☒ Sunday

Daily frequency

☒ Occurs once at:

☐ Occurs every: hour(s) Starting at: Ending at:

Duration

Start date: ☐ End date: ☒ No end date

Summary

Description:

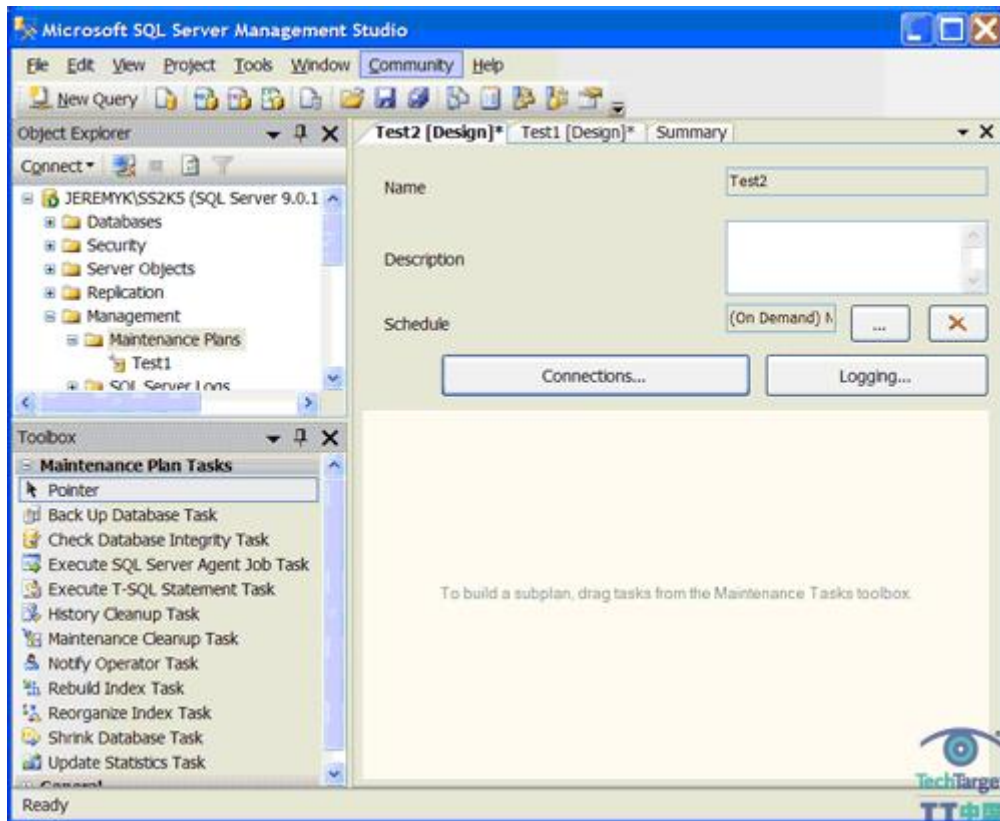
Job Schedule Properties

怎样预定 SSIS 数据库维护包？

在 Management Studio 里，对 Maintenance Plans 进行操作。双击需要预定具体方案。单击 Schedule 文本对面的椭圆形按钮，这时就会出现左边的屏，我们用它来设置 SQL Server Agent 用来管理 SSIS 包的进度表。

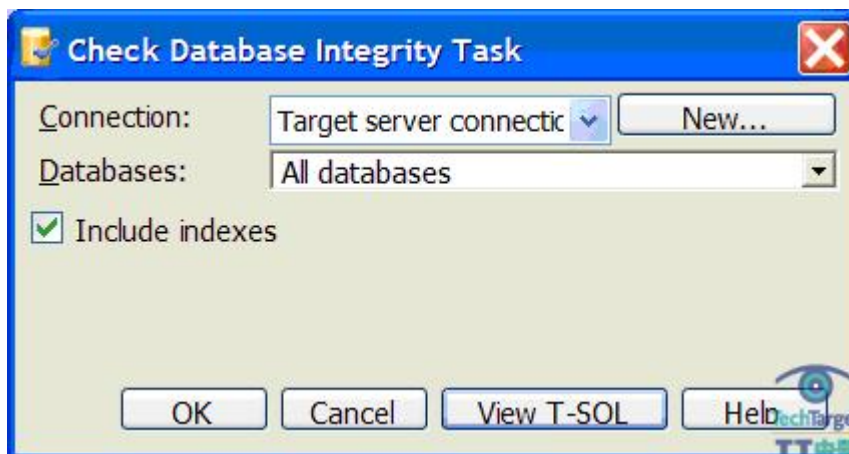
数据库维护 SSIS 包怎样设置？

在 SSIS 中会出现一个空白的工作区，你可以将工具箱任务迁到这个工作区中。



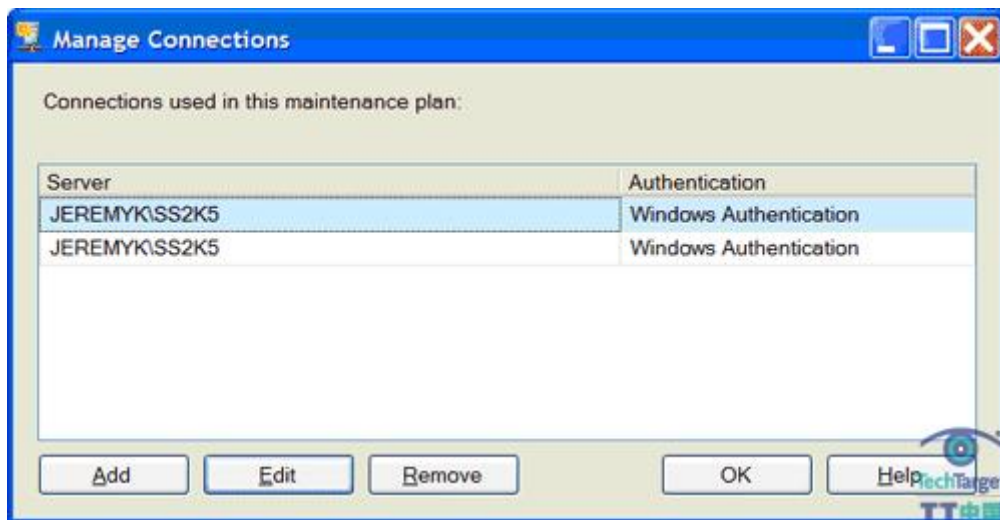
SSIS 工作区

任务拖进工作区之后，双击任务、通过下拉箱设置独立参数。你还可以点击 View T-SQL 按钮查看下面的命令，这样做有助于你更好地了解内在程序。



配置任务参数

用了其他服务器，那就有必要点击主要工作区里的 Connections 按钮创建这些连接对象。左边的 Manage Connections 界面可用于创建、编辑或删除到 SQL Server 的连接。



创建连接对象

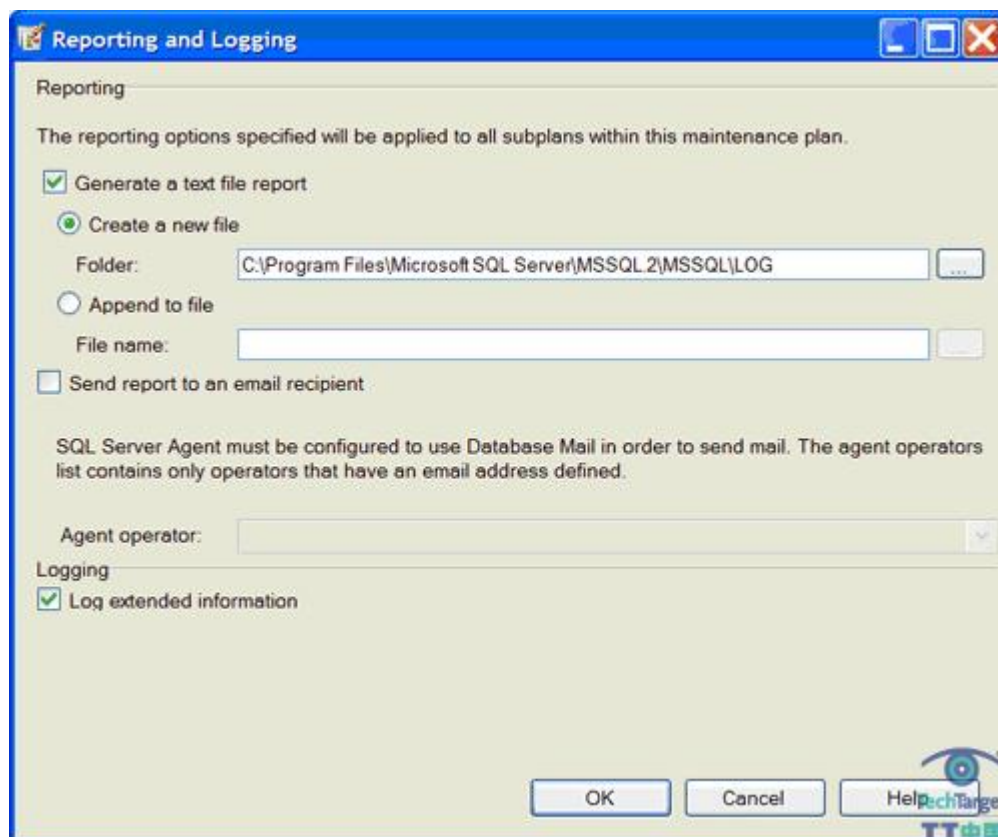
在预定 Maintenance Plan 之前的最后一个步骤就是配置 Reporting 和 Logging 界面。要访问这个界面，需点击主工作区中的 Logging 图标。

Reporting 和 Logging 界面可以用来启用日志记录到文本文件，还可以发电子邮件到全体技术维护团队。

(作者: Jeremy Kadlec 译者: April 来源: TT 中国)

SQL Server Integration Services 简化数据库维护（三）

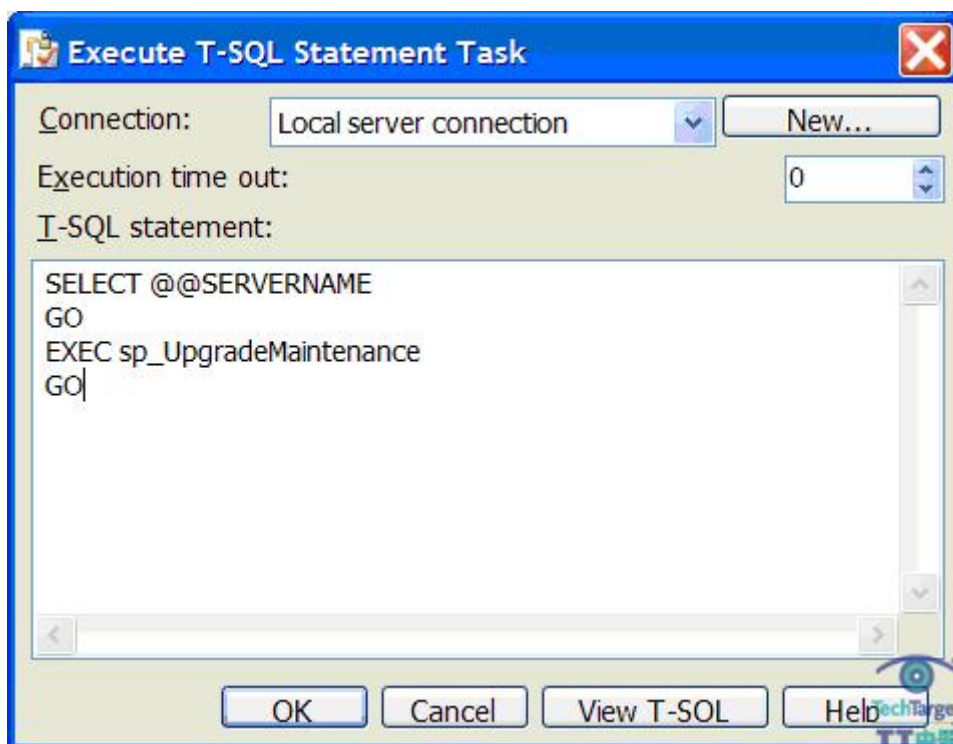
Reporting 和 Logging 界面可以用来启用日志记录到文本文件，并发电子邮件到全体技术维护团队。



配置 Reporting 和 Logging 界面

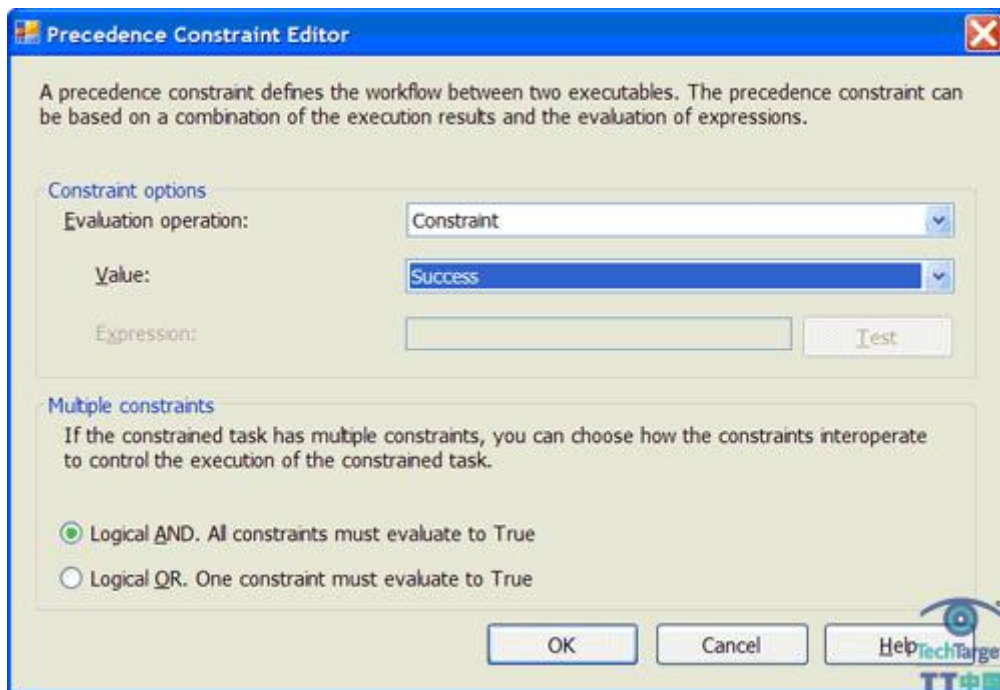
现有的 T-SQL 脚本怎么合并？

如果你有现存的 T-SQL 脚本或者是存储程序，你想将其合并进 SSIS 包内，你只需要在空间内进行 Simply drag and drop the Execute T-SQL Statement Task 操作。再将代码复制粘贴到该对象内并保存。



在 SSIS 包内合并现存脚本

该项任务完成后，就增加优先约束取保任务是按照现在的顺序调用的。我们可以在现存任务和新对象之间建立有限约束。在工作区间内双击优先约束对它进行设置，这时就会出现左边编辑器，此时你建立的关系就可能出现以下几种情况：成功、失败、完成，或者这种操作是建立在约束或表达式操作之上的。



保证按正确的顺序调用任务

是不是 SQL Server 2005 所有的版本都有 SSIS 数据库维护功能？

从本质上来说，只有标准版和企业版才支持 SSIS 和附属的数据库维护计划功能。Workgroup 和 Express 版本就没有这些功能，但是我们可以用以下的 T-SQL 命令进行维护。

是什么在操纵 SSIS 包？

SSIS 依赖于 SQL Server Integration Services Windows Service，而后者主要是负责处理所有的维修计划和自定义开发 SSIS 包。

其他 SSIS 资源

SSIS 引起了很多人的注意。目前我们可以查询下面这些书：

Wrox Press, Professional SQL Server 2005 Integration Services

Mann Publishing Group, The Rational Guide to Scripting SQL Server 2005
Integration Services Beta Preview

Sams Publishing, Microsoft SQL Server 2005 Integration Services

总结

只有 SSIS 才能进行数据库维护。从 SQL Server 2000 到 2005 性能方面的增强从 ETL 的角度来说相当有益。随着 SQL Server 2005 数据库维护工具 SSIS 的扩展和 point-and-click 性能大大改善有望减轻维护负担。

(作者: *Jeremy Kadlec* 译者: *April* 来源: *TT 中国*)

SQL Server 2005 安装问题提示 SSIS 错误信息

问：我目前用的是 SQL Server Integration Services (SSIS)。我进入 Business Intelligence Development Studio (BIDS) 试图创建新项目时，却收到了如下的错误信息：“不能下载' Microsoft.DataTransformationServices.Wizards 及其附件，系统找不到指定文件。”我在电脑上找到了我所需要的.dll 文件，它确实存在。我如何才能解决这个问题呢？

答：我发现在本地机器上安装 SQL Server 2005 会经常出现这样的问题。你没有详细说明你目前运行的 SQL Server 版本，所以我也不知道是不是在 SQL 2000 或 2008 上发生的。但如果你运行的是 SQL Server 2005，最常见的解决方法就是重新设置 SQL Server 客户机工具。通常你需要做的就是进入高级设置选项进行卸载然后重新设置客户机程序。

(作者: Mark Kromer 译者: April 来源: TT 中国)

在 SQL Server 中调试 SSIS 包（一）

SQL Server 2005 的新特性，SQL Server Integration Services (SSIS) 增加了一种企业级数据抽取、转换和下载 (ETL) 工具，丰富了不断增长的 SQL Server 工具套件。SSIS 取代了 DTS，这只是在 SQL Server 的两个版本中。除了有点慢、操作起来很灵活外，DTS 并没有在你遇到问题时提供任何简单的调试包的方法。在最近的 SQL Server 版本中有一些调试 SSIS 包的方法，这些方法可以帮你找到出现这些问题的根本原因。

SSIS 包嵌在 SQL Server Business Intelligence Development Studio (BIDS) 中，附带有一些 SQL Server 2005 客户机工具。事实上，BIDS 只是 Visual Studio 的另一个版本。实际上如果你在同一台机器上安装 Visual Studio 作为 BIDS，从这一点来看，你可以运行 Visual Studio 代替 BIDS。那么，这里的调试是什么意思？它的意思就是在你开发 SSIS 包时你就可以用 Visual Studio 了。

SSIS 包调试断点 (Breakpoint)

设置断点的能力就是调试过程之一。SSIS 可以让你设置两种不同的断点：一个在包、任务和存储器上，另一种就是在脚本对象里。设置好后，对 SSIS 包执行的操作就会在断点处停止，它可以让你观察处于暂停状态下的包。在断点的时候，我们要记住这些断点只能在控制流上的项目上进行设置，而不能在数据流任务上进行设置。

我们首先来看看在包、任务和容器上设置断点。

要在包上设置断点，你就要进入控制流 tab 键，在包的面上单击右键，选择 Edit Breakpoints。如果你想为任务或容器设置断点，就可以在你感兴趣的包或容器上单击右键，再选择 Edit Breakpoints。在这两种情况中，你都可以看到如图 1 所示的 Set Breakpoints 会话。

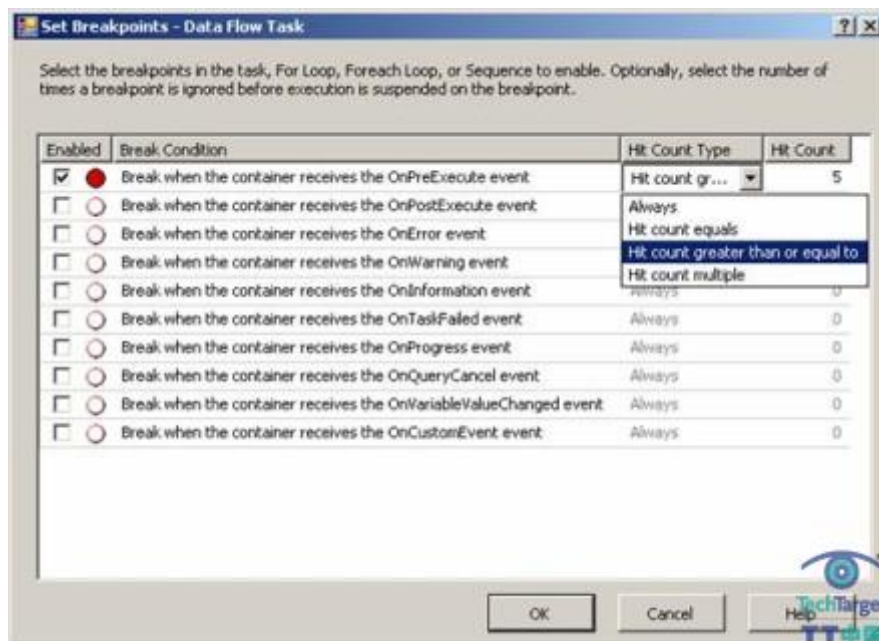


图 1: SSIS 包的 Set Breakpoint 会话

在该会话框中，你可以选择在这种条件下的 Break Condition 和 Hit Count。在该实例中，当收到 OnPreExecute 事件时，我就要暂停数据流任务。注意，Hit Count 有很多任务可供我们选择。在这个事例中，我们可以在任务收到 OnPreExecute 后再暂停，在任务运行之前 OnPreExecute 事件本来是触发的。你运行包的时候，OnPreExecute 事件就停止，并突出启动断点任务，如图 2 所示：

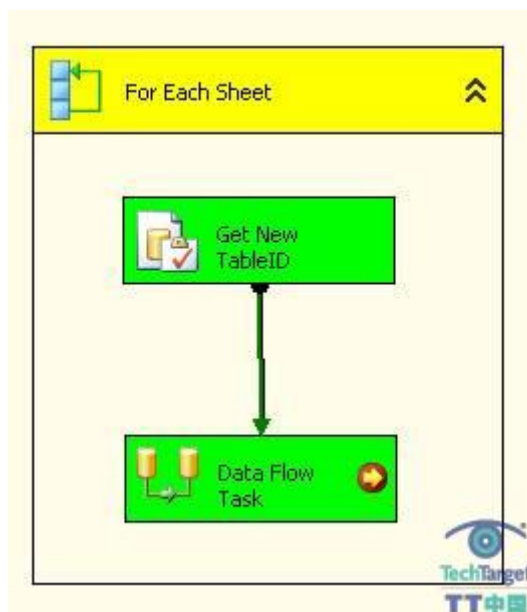


图 2：“数据流任务”上的断点已经启动

你可以看到，我们之前设置的断点已经启动，黄色的箭头指向活动的断点。这时，你就可以访问所有的本地变量、任意你曾配置的观察变量（watch variable）以及暂停之后表示包的状况的视图。

除了能够在包、容器和任务上设置断点之外，你还可以在脚本任务里对它们进行配置。有两种方法实现：你可以在脚本设计器里打开脚本或者单击右键选择 Breakpoint> Insert Breakpoint。或者，在编辑器左边的空白地方单击左键。和以前一样，遭遇断点之后 SSIS 包就会暂停。

(作者: Eric Johnson 译者: April 来源: TT 中国)

在 SQL Server 中调试 SSIS 包（二）

使用断点还是有局限性的。如果“错误逻辑（error logic）”调用了多次，你就可以设置启用它们；或者你可以在你只能看见在出现问题的地方有什么包时你也可以对它们进行设置。从 Integration Services 里对包进行编辑并运行之后，断点就不存在了。所以在启用产品包时我们不要担心对它造成的影响。

调试 SSIS 包的数据阅读器（data viewer）

所以你现在可以在控制流上设置断点。但是如果你要解决数据流内部的故障问题该怎么办呢？这时数据阅读器就派得上用场了。数据阅读器（有四种选择）可以在任意数据流任务上进行配置。通过点击右键（任务之间的一个箭头）选择数据阅读器，你就可以看见 Data Flow Path Editor 的数据阅读器部分。然后点“Add”、选择查看 Grid、Histogram、Scatter Plot 或 Column 图表。在图 3 中，你可以看到在 conditional split 的输出结果“Child Rows”上配置的 Grid Data Viewer。

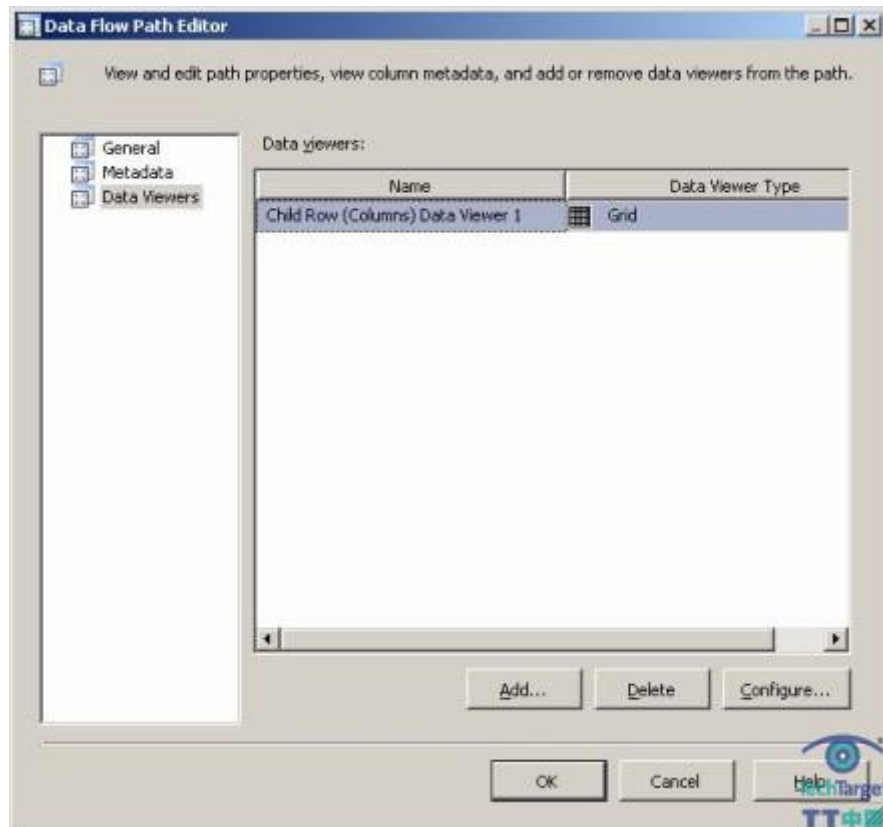


图 3: 在 conditional split 的输出结果“Child Rows”上配置的 Grid Data Viewer

你选择的这种数据阅读很大程度上取决于你查看的数据类型和你所要知道的东西。我只想知道 conditional split 任务执行之后的数据。你运行包的时候，它就在数据阅读器这个地方暂停，显示你选择的数据格式。图 4 表示我运行包时的 Grid Data Viewer。

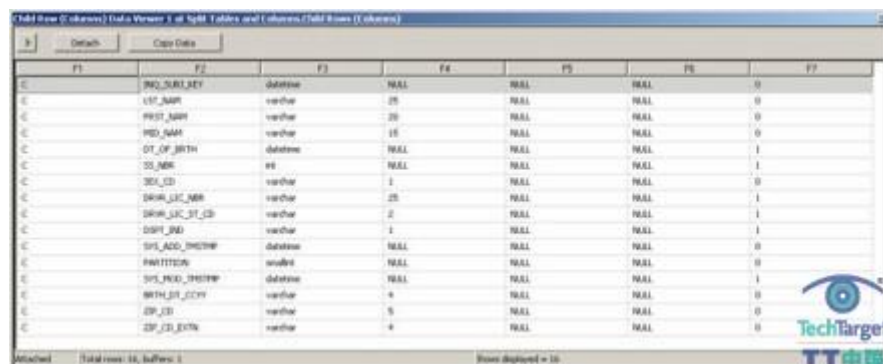


图 4: 在包执行过程中的 Grid Data Viewer

现在我可以检查数据，保证它们能够正常运行；或者我还可以查找有害数据，这些数据可能造成其他的问题。为了让这些包再次运行，我可以点击“Play”按钮或点击“Detach.”。如果我选择“Play”，SSIS 包就会继续运行。下次运行同样的任务时，例如在进行 loop 操作时，我可以给目前的数据获取另一个数据阅读器。如果我选择“Detach”，SSIS 包就会在不触及数据阅读器的情况下继续运行。所以如果同一个任务再次在 loop 中运行的时候，我就不会获取新数据阅读器。

SSIS 提供了几种新的调试方法。其中两种最简单的方法就是用断点和数据阅读器。这两种方法都展现了 SSIS 包内部运转的一些信息。数据阅读器和断点只是去掉了 SSIS 包里调试选项表面的东西。通过简短的归纳，你可能对 SSIS 包内在的东西有了充分的理解，甚至可以解决一些在包里可能遇到的问题。

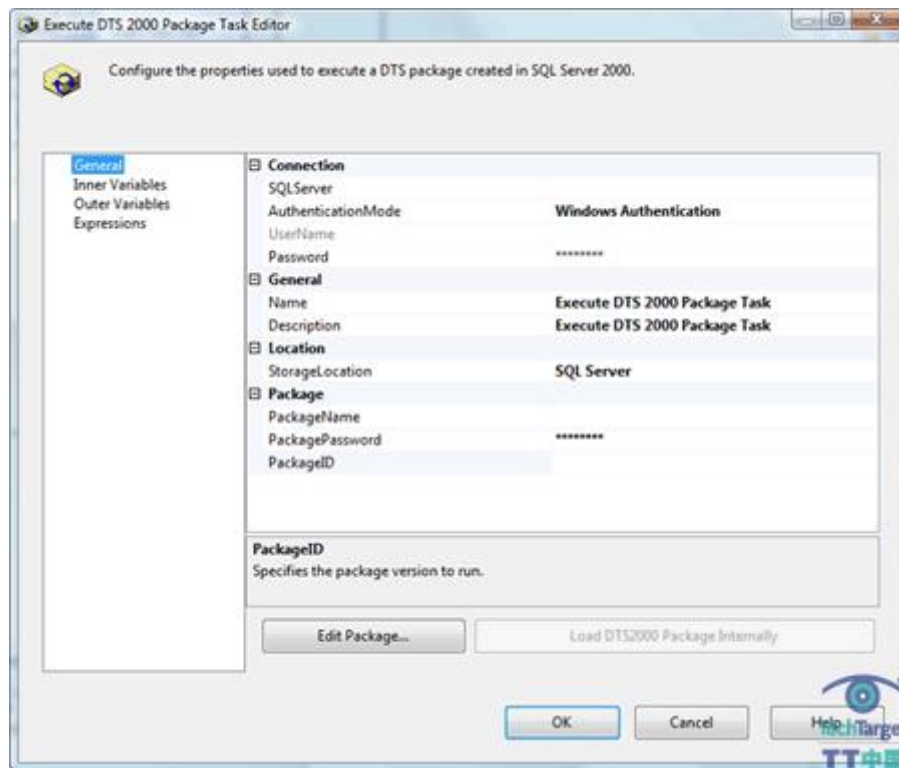
(作者: Eric Johnson 译者: April 来源: TT 中国)

在 SQL Server Integration Services 中运行 DTS 包

SQL Server Integration Services (SSIS) 是 SQL Server 中最新的 ETL 工具。它表示替换 Data Transformation Services (DTS) 易于使用并且十分灵活。既然是最新的、最好的，为什么你还想继续使用 DTS？答案很简单：SQL Server 2005 中引进 SSIS 之前，DTS 就将近有 7 年的时间。许多公司用了这个工具，并有些现存的 DTS 包很复杂，需要许多天、周、甚至是几个月的时间开发。如果你的 DTS 包开始运行，你为什么还要对它重新操作呢？

将 DTS 包转换成 SSIS 包有很多种方法，但是他们不会像预先设计的那样工作，你要花一些时间进行修正。用 DTS 包有什么意义？你是不是不得不放弃这些 DTS 包转而使用 SQL Server 2005 和 SSIS？显然不是。SQL Server Integration Services 对老版的 DTS 包来说完全是嵌入式的集成。

在开发 SSIS 包时，在控制流 (control flow) 方面你可以执行的任务为 Execute DTS 2000 Package Tasks。它可以让你用 SQL Server 2000 工具调用 DTS 包。如果你的 DTS 包设置在 SQL Server 7 中，那么你就没有别的办法了。用 Execute DTS 2000 Package Task 非常简单。将任务拖进控制流中，双击进行配置。打开的对话框如图一所示：



从这里你就可以指定是否将包存储在 SQL Server 上或文档里，提供对包的详细链接。此外，你可以对 DTS 包设置参数。这是其中的一种实现方法。首先来看对话框属性的 Inner Variable 部分，点击 New。这样，你可以对 DTS 包里的 Global Variables 指定一个值。

你还能增加 Outer Variables，即定义 SSIS 包里的变量。之后你还可以在 DTS 包脚本对象里使用这些变量。最后，你还可以让 SSIS 将内部的 DTS 包拖进 SSIS 包。你可以点击 Load DTS2000 Package Internally 按钮。这一操作就创建了 SSIS 包里的 DTS 包的复本，你也不再需要依赖源 SQL Server 或者文件夹了。记住如果你在内部进行的存储，你不能使用 SQL Server 2005 工具编辑或查看 DTS 包，那种方法之适合存储 DTS 包。

如果你已经设置好了许噢有的 Execute DTS 2000 Package Task 属性，那么它就成了你工作流的一部分。你可以进一步创建 DTS 包的 SQL Server Integration Service 功能、操纵你想操纵的数据、或者你也可以让 SSIS 包仅仅成为 DTS 包的一层外包装。将

Data Transformation Services 包转换为 SQL Server Integration Services 可以让你的工作更具有弹性，因为通常不值得你花那么长的时间完成转换工作。如果是这样的话，你只要知道用现存的方法从 SSIS 包里调用以前的 DTS 包的版本就可以了。

(作者: Eric Johnson 译者: April 来源: TT 中国)

错误导入 Excel 数据到 SQL Server 2005

问：我才刚刚开始学习 SQL Server 2005，我简直不敢相信 SSIS 和 DTS 比起来竟然这么难。我想在两个小时之内将一份 Excel 表里的数据导入进去（用 DTS 的话只需要两分钟），但是我仍然不断收到错误提示。由于不支持 DT_R8 和 DT_STR 数据类型之间的转换，所以不能插入列。但是我不知道什么是 DT_R8？

我想把 Excel 数据导入 SQL Server 中。我已经尝试过导入现存的表、并创建一个新表。我还尝试过改变所有的单元格。但是仍然无济于事。你可以帮我解决这个问题吗？

答：只要客户发送文件给我，我就会自通过 DTS 或 SSIS 包进行数据抽取。但是我会经常因为某些原因避开 Excel 表。如果是给我一份 csv 文件，一份 tab 界定的文件或者是 '|' 界定的文件，但不要选 Excel 文件。除 Excel 之外有很多都可以尝试。你可以列出一些能和它们匹配的数据类型。仔细检查你的 Excel 文件中的 这些数据类型——特别是不能导入的列。在单元格上点击右键、选择格式查看数据类型。你的目标表中的列必须和这些数据类型相匹配。如果它们不能够匹配，你就 考虑一下在插入之前用 SSIS Convert 任务改变数据类型。

(作者: Joe Toscano 译者: April 来源: TT 中国)