

数据仓库实验报告

|  |  |
| --- | --- |
| **实验名称：** | **实验二 使用 SSIS 创建项目和基本包** |
| **实验日期：** | **2021年10月5日星期二** |
| **实验地点：** | **宿舍** |
| **提交日期：** | **2021年10月8日星期五** |
|  | |
| **学号：** | **11920192203642** |
| **姓名：** | **袁佳哲** |
| **专业年级：** | **软工2019级** |
| **学年学期：** | **2021-2022学年第一学期** |

1. 实验环境

SQL Server 2019；

先决条件：已安装 AdventureWorksDW2012 示例数据库

1. 实验目的

（1）掌握使用 SSIS 设计器创建项目和基本 ETL 包的方法；

（2）掌握使用 SSIS 添加循环的方法；

（3）掌握使用 SSIS 添加日志记录的方法；

（4）掌握使用 SSIS 添加错误流重定向的方法。

1. 实验内容和步骤

问题及回答：

第一课问题：

1. “平面文件连接管理器”在集成服务中的作用？

平面文件连接管理器使包可以访问平面文件中的数据。 例如，平面文件源和目标可以使用平面文件连接管理器提取和加载数据

平面文件连接管理器只能访问一个文件。 若要引用多个文件，请使用多平面文件连接管理器，而不用平面文件连接管理器

1. “OLE DB 连接管理器”在集成服务中的作用？

OLEDB 连接管理器的作用是使用 OLEDB 提供程序将包连接到数据源。 例如，连接到 SQL Server 的 OLEDB 连接管理器可以使用适用于 SQL Server 的 Microsoft OLEDB 提供程序。

有若干 SQL Server Integration Services 任务和数据流组件使用 OLEDB 连接管理器。 例如，OLEDB 源和 OLEDB 目标使用此连接管理器来提取和加载数据。 执行 SQL 任务可以使用此连接管理器来连接到 SQL Server 数据库以运行查询。

OLEDB 连接管理器还用于在以使用 C++ 等语言的非托管代码编写的自定义任务中访问 OLEDB 数据源。

1. 在步骤 6 中，“lookup”转换工具的作用是什么，为什么选择 dbo.dimCurrency。除了“lookup”转换，系统还提供了哪些转换功能？

Lookup转换即查找转换。查找转换通过联接输入列中的数据和引用数据集中的列来执行查找。可以使用该查找在基于通用列的值的相关表中访问其他信息。引用数据集可以是缓存文件、现有的表或视图、新表或 SQL 查询的结果。查找转换使用 OLE DB 连接管理器或缓存连接管理器来连接到引用数据集。

因为我们要将指定输入列中的数据联接到引用数据集中的列来执行查找。我们在之前已经将平面文件列CurrencyID的目标列定义为 DimCurrency.CurrencyAlternateKey，因此这里需要选择dbo.dimCurrency。

除了“lookup”转换，系统还提供聚合转换、审核转换、数据转换、合并转换、Merge Join Transformation等

第二课问题：

1. 本课完成什么功能？

提取、转换和加载 (ETL) 过程通常从多个平面文件源中提取数据。 从多个源提取数据需要采用迭代控制流。 Microsoft Integration Services 可以轻松将迭代或循环添加到包。

本将修改第 1 课的示例 ETL 包以使用 Foreach 循环容器，并为该包设置用户定义的包变量。 然后，该变量用于循环访问示例文件夹中的匹配文件。

1. Foreach 循环容器起什么作用，该例中，什么控件被加入容器？

Foreach 循环容器定义包中的重复控制流。 循环的实现与编程语言中的 Foreach 循环结构类似。 在包中，通过使用 Foreach 枚举器启用循环。 Foreach 循环容器将重复指定枚举器的每个成员的控制流。

在该例中，“Extract Sample Currency Data”数据流任务被加入到 Foreach File in Folder Foreach 循环容器中 。

1. 变量 varFileName 什么时候被定义，变量的作用域是哪里，在哪个步骤被赋予什么值？

在运行时，Foreach 循环容器的每次迭代遍历示例数据文件夹中的 14 个平面文件。每次找到与指定的条件匹配的文件名时，Foreach 循环容器都将用该文件名填充用户定义的变量。 该变量又会更新平面文件连接管理器的 ConnectionString 属性，并与平面文件建立连接。

因此，变量的作用域是循环容器中的数据流任务，在每次循环时都会被赋值为指定文件夹中的文件名，如：Currency\_ARS.txt、Currency\_AUD.txr等

第三课问题：

1. 日志文件的记录将起到什么作用？

日志记录功能，可通过提供任务和容器事件跟踪监控包执行情况以及进行故障排除。

1. 日志文件的记录由哪些事件负责？

此例中日志文件的记录由“PipelineExecutionPlan” 和“PipelineExecutionTrees”事件负责

1. 除了文本类型，系统还支持哪些其他格式的文件吗？

除了文本类型，还支持以下格式：

SQL Server Profiler、Windows 事件日志、SQL Server、XML 文件

第四课问题：

1. 错误流重定位起到什么作用？

错误流重定向可以使用错误输出将组件配置为将失败的行重定向到其他处理路径，而不是允许转换失败。 使用单独的错误处理路径提供了更多选项。 例如，可以清除数据，然后重新处理失败的行。 或者，可以保存失败的行及其错误信息，以便以后进行验证和重新处理。

②步骤 2 中，Currency\_BAD.txt 文件为什么出现导入失败，失败的级别是个别文件还是整个包？

因为我们编辑“CurrencyID”列以包含错误的值，这会导致查找失败。 处理已损坏的文件时，查找失败将导致 Currency Key 查找转换失败，从而导致包的其余部分失败。

失败的级别是整个包

1. ErrorOutput.txt 文件中有哪些内容？

文件中，默认的列顺序为：AverageRate、CurrencyID、CurrencyDate、EndOfDateRate、ErrorCode、ErrorColumn、ErrorDescription

文件中的所有行都包含不匹配的 CurrencyID 值“BAD”、ErrorCode 值 - 1071607778、ErrorColumn 值 0 以及 ErrorDescription 值“在查找期间行没有生成任何匹配项”。 ErrorColumn 的值为 0，因为错误不是特定于列的，而是查找操作失败。

1. 实验总结(完成的工作、对实验的认识、遇到的问题及解决方法)

遇到的问题：

* + - 1. AdventureWorksDW2012下载错误，导致第1课1-6验证查询结果时出错。需要重新下载正确的文件
      2. 在第四课4-3中，点击编辑脚本并未自动打开Microsoft Visual Studio Tools for Applications (VSTA)，导致无法编辑脚本。解决方法是找到Lesson 4.dtsx文件并用文本编辑器打开，手动找到Input0\_ProcessInputRow函数并添加对应脚本语言