**实验二 使用 SSIS 创建项目和基本包**

**1. 实验环境**

 SQL Server 2019；

 先决条件：已安装 AdventureWorksDW2012 示例数据库

注：示例数据库可使用新教程中的版本

**2. 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。

**3. 实验内容**

 第一课问题：

①“平面文件连接管理器”在集成服务中的作用？

答：通过平面文件连接管理器，包可从平面文件中提取数据。 使用平面文件连接管理器，可以指定包从平面文件中提取数据时要应用的文件的名称与位置、区域设置与代码页以及文件格式，其中包括列分隔符。 另外，还可以为各个列手动指定数据类型；也可以使用“提供列类型建议”对话框，自动将提取出来的数据列映射到 数据类型。

②“OLE DB 连接管理器”在集成服务中的作用？

答：通过 OLE DB 连接管理器，包可以在任何 OLE DB 兼容的数据源中提取数据或加载数据。 使用 OLE DB 连接管理器，可以为连接指定服务器、身份验证方法和默认数据库。

③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？

答：1. 在步骤六中，"lookup"（查找）转换工具的作用是用于在数据流中查找引用数据的匹配项，并将匹配的引用数据的值添加到输入数据中。在这个具体场景中，“lookup”转换根据源文件中的CurrencyID和CurrencyDate列，与目标维度表DimCurrency和DimDate进行匹配，从而获取相应的CurrencyKey和DateKey。

1. 选择dbo.DimCurrency的原因是因为DimCurrency表是存储货币信息的维度表，包含CurrencyKey（货币主键）和CurrencyAlternateKey（如货币代码）的映射关系。通过查找转换，可以从源数据中的货币ID找到与之对应的CurrencyKey，从而将这个主键用于后续的数据分析或处理。
2. 除了lookup转换外，系统还提供了以下常用的转换功能：(1)数据转换: 将数据从一种数据类型转换为另一种类型，特别适合在源数据类型与目标数据类型不一致时使用。(2)合并: 将两个数据流按列进行合并，适用于当两个数据集有共同字段且需要合并时。(3)排序: 对数据流中的数据进行排序，可以根据一个或多个列进行升序或降序排序。(4)条件拆分 : 根据不同的条件将数据流拆分成多个输出，用于处理需要根据特定条件分流的数据。(5)派生列: 在数据流中添加新的计算列，或根据现有列对数据进行修改和转换。

 第二课问题：

①本课完成什么功能？

答：本课将修改第 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，并且不修改数据流，而只修改控制流。

② Foreach 循环容器起什么作用，该例中，什么控件被加入容器？

答：将该容器的枚举值映射为用户定义的包变量。 然后，该容器使用此变量来动态修改平面文件连接管理器的 ConnectionString 属性，并循环连接该文件夹中的每个平面文件。

③ 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？

答：平面文件连接管理器配置为静态加载单个文件。 若要启用平面文件连接管理器以重复加载文件，需要更改连接管理器的 ConnectionString 属性以使用用户定义的变量 User::varFileName，该变量包含要在运行时加载的文件的路径。

 第三课问题：

①日志文件的记录将起到什么作用？

答：对于日志文件的记录，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。

②日志文件的记录由哪些事件负责？

答：日志文件的记录由多个事件触发，常见的包括：

OnError：记录错误事件。

OnWarning：记录警告事件。

OnInformation：记录信息性事件，如任务开始和结束。

OnTaskFailed：记录任务失败事件。

③除了文本类型，系统还支持哪些其他格式的文件吗？

答：系统支持多种日志文件格式，包括：

文本文件

SQL Server日志

Windows事件日志

SQL Server Profiler

XML文件

 第四课问题：

①错误流重定位起到什么作用？

答：错误流重定向允许在数据流任务中处理错误数据。遇到处理失败的数据行时，系统可以将这些行重定向到错误输出流，而不是终止整个任务。这样可以处理并记录出错的行，避免包失败。

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？

答：Currency\_BAD.txt文件导入失败可能是由于文件中的数据格式不符合预期，比如数据类型不匹配。失败的级别通常是个别行或文件，而不是整个包。错误行会被定向到错误输出流中。

③ ErrorOutput.txt 文件中有哪些内容？

答：ErrorOutput.txt文件中记录了所有无法处理的数据行和对应的错误信息。它包含错误行的详细内容，以便后续分析和修复问题。