

新增功能指南

Sybase® ETL 4.9

文档 ID: DC00960-01-0490-01

最新修改日期: 2009 年 9 月

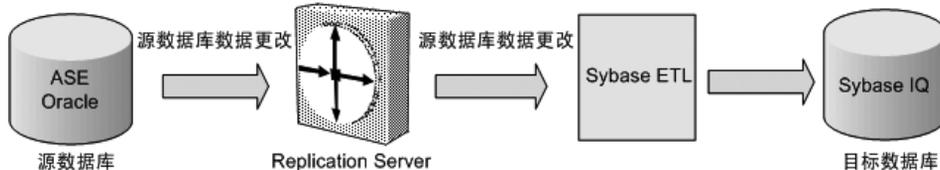
本指南介绍了 Sybase® ETL 4.9 中的新增功能。

主题	页码
将 ETL 与 Sybase Replication Server 结合使用可以实现增强的增量装载	1
ETL 计划程序支持	3
为运行时事件配置警报	4
转换组件增强功能	5
添加的 SQL Executor 组件	5
添加的 Copy Splitter 组件	6
增强的 Data Splitter JavaScript 组件	6
事务性增强	7
对连接共享的支持	9
DB Staging 组件增强	9
Data Sink 组件增强	10
并发执行多个项目	10
支持的存储库	11

将 ETL 与 Sybase Replication Server 结合使用可以实现增强的增量装载

现在, 您可以将 ETL 与 Sybase Replication Server® 结合使用来装载 Replication Server 从 Sybase Adaptive Server® Enterprise (ASE) 和 Oracle 数据库捕获到 Sybase IQ 中的数据更改。此解决方案支持源表和目标表之间的不同架构, 以及由 Sybase ETL 执行的数据转换。

图 1: 使用 Sybase Replication Server 执行增量装载



Sybase ETL 4.9 包括新组件 CDC Provider Sybase Replication Server，该组件支持基于 Sybase Replication Server 的增量装载解决方案，进而可增强性能和可用性，以及支持其他用例。

在 4.9 之前的版本中，增量装载受 DB Data Provider Index Load 组件支持，该组件仅根据升序索引属性值来捕获修改的数据。

此外，您现在可以捕获所有的数据更改，包括源数据库中的插入、更新和删除活动，而无需依靠或修改源表模式，也不会妨碍源数据库的性能。为确保一致性，可以启用 ETL 项目级事务性功能。

源数据库中的所有数据更改都由 Sybase Replication Server 进行捕获。ETL 将从 Replication Server 收到数据更改，对其进行转换，然后将其装载到 Sybase IQ 中。

使用 CDC Provider Sybase Replication Server 组件，可以自动执行配置 Sybase Replication Server 的过程，以创建和删除复制定义、复制连接、函数字符串和复制订阅，或标记源数据库中要为所有的 ETL 任务复制的表。

可以使用 ETL 增量装载解决方案执行以下操作：

- 从 Oracle 和 Sybase Adaptive Server Enterprise 数据库提取数据更改。
- 将数据装载到目标 Sybase IQ 数据库或文件。
- 将数据更改从源表传输到具有不同表模式的目标表。
- 对从 Replication Server 接收的不同的已更改数据集使用不同的转换规则。
- 检索旧值，然后应用转换规则，以便对旧数据与已更改的数据进行比较。

请参见《Sybase ETL 4.9 用户指南》中的第 5 章“组件”。

ETL 计划程序支持

Sybase ETL 4.9 通过提供其自有的 ETL 计划程序，将其计划程序功能扩展到所有的可用平台。通过 ETL 计划程序，您可以使用从 ETL Development GUI 中选择的引擎来创建、列出、编辑、删除、执行和终止调度的任务。

早期版本的 Sybase ETL Development 直接使用 Windows 任务计划程序。如果正在从 Sybase ETL Development 4.8 或更早版本进行升级并要迁移计划的任務，则必须使用 Runtime Manager 导入现有任务。

此外，ETL 计划程序还可以同时运行任务的多个实例。对于已运行的计划任务，可以指定以下任何一个选项：

- 并发执行新任务 – 启动新计划执行以及当前进程。
- 按顺序执行新任务 – 等待当前进程完毕后启动下一个计划执行。
- 不执行新任务 – 继续处理当前任务，忽略启动新计划执行的请求。
- 执行新任务之前取消正在运行的任务 – 停止当前进程并立即启动新计划执行。

ETL 计划程序遵循基于文件的配置。所有的配置都随每个网格引擎存储在独立的平面文件中。文件锁用于确保仅为每个主机启动一项 ETL Schedule 服务。Sybase 建议不要手动修改配置文件。使用 ETL Development GUI 中的 ETL Runtime Manager 执行任何计划程序配置。

请参见《Sybase ETL 4.9 用户指南》中的第 4 章“高级概念和工具”。

为运行时事件配置警报

在 Sybase ETL 4.9 中，可以在运行时事件（如项目或作业启动、完成或错误）出现时设置警报。出现以下任何事件时，可以配置 ETL 以发送电子邮件警报：

注意 可以指定预订 ETL 事件通知的多个电子邮件地址。

- 作业 – 在作业启动、成功完成或意外完成且返回错误时发出通知。
- 作业启动 – 作业启动时出现，且包含作业名称，运行作业的网格引擎，以及作业启动的时间。
- 作业完成 – 在作业正常完成时出现，且包含作业名称、运行作业的网格引擎、作业的开始和结束时间，以及所有项目的列表。
- 作业错误 – 作业完成且返回错误时出现，且包含如作业名称、网格引擎名称、启动时间、失败的时间以及错误消息等详细信息。
- 项目 – 项目启动、成功完成，或意外结束且返回错误时出现。
- 项目启动 – 项目启动时出现，且包含如项目名称、启动项目的网格引擎，和项目启动时间等详细信息。
- 项目完成 – 在项目成功结束时出现，且包含如项目名称、运行项目的网格引擎名称、项目启动和结束时间，以及成功装载的行数等信息。
- 项目错误 – 在项目意外结束且返回错误时出现。邮件中包含如项目名称、运行项目的网格引擎名称、项目启动和失败时间、成功装载的行数和错误消息等详细信息。

注意 在同一网格引擎上运行的多个网格节点进程共享同一个警报服务配置文件和警报历史记录日志文件。

使用 ETL Development GUI 中的 Alert Manager 创建、编辑和删除警报定义。可以针对项目模拟以及作业和项目执行的任何事件类型创建多个警报。但是，警报名称必须是唯一的。

可以在 **Log File Inspector** 中以及 **ETL Web** 接口上查看事件警报历史记录。映射警报时，还可以设置过滤条件。例如，可以定义过滤器，以便仅检查要触发警报的特定项目名称。

Sybase ETL 4.9 还支持将所有的警报历史记录存储在单独的日志文件中，以便在故障排除期间起到帮助作用。

Sybase 建议在一台计算机上设置电子邮件警报通知。然后，您便可以跨网络上运行的不同网格引擎手动复制警报配置文件。

请参见《**Sybase ETL 4.9 用户指南**》中的第 4 章“高级概念和工具”。

转换组件增强功能

本节说明了对转换组件执行的增强操作。

添加的 SQL Executor 组件

ETL 4.9 包括新的转换组件 **SQL Executor**，使用该组件可以针对数据库服务器执行自定义 **SQL** 语句或多个语句。**SQL Executor** 是不含输入端口或输入端口的独立组件，可以放置在与其它组件分离的项目中，或包含一个或多个 **SQL Executor** 组件的单个项目中。

此外，**SQL Executor** 组件还允许您在单个事务中使用多个 **SQL** 语句将数据从多个输入数据流装载到 **Sybase IQ** 数据库中。这会使得所有 **ETL** 用户的灵活性达到最高程度。

使用 **SQL Executor** 组件可以执行以下操作：

- 允许其他项目将数据读取或写入到相同数据库中。
- 使用 **SQL** 语句将数据从源表装载到 **IQ** 支持的文本文件格式。
- 使用 **Load Table** 命令，在单个事务中将数据从文本文件装载到目标 **IQ** 数据库中。

SQL Executor 组件附带了 pre-SQL 和 post-SQL 语句参数。该组件可以逐行执行 SQL 脚本，然后检查 SQL 脚本在目标数据库中是否在语法构成上正确无误。如果脚本成功运行或失败，您还可以定义要执行的不同自定义 SQL 脚本。

请参见《Sybase ETL 4.9 用户指南》中的第 5 章“组件”。

添加的 Copy Splitter 组件

Sybase ETL 4.9 包括新的 Copy Splitter 组件，该组件可以将输入数据无条件复制到每个输出端口。通过 Copy Splitter 组件，可以将数据写入每个输出端口，而不需要计算输出端口条件。缺省情况下，该组件具有两个输出端口。

使用 Copy Splitter 组件可以不必调用 JavaScript，而且可以降低条件计算成本，进而提高性能。

请参见《Sybase ETL 4.9 用户指南》中的第 5 章“组件”。

增强的 Data Splitter JavaScript 组件

Data Splitter JavaScript 组件已经过增强，允许您计算并定义互斥输出端口条件，进而显著提高性能，不仅能消除内存分配系统调用，而且能提供有效的内存管理方法。

在 Sybase ETL 4.9 之前的版本中，Data Splitter JavaScript 只能针对每个输出端口的条件来计算输入记录，进而会导致潜在瓶颈和内存封锁。此外，它未提供使用任何其他有效逻辑构造的灵活性。

您现在可以使用 Data Splitter JavaScript 组件来定义互斥的输出端口条件，进而显著减少条件计算工作量。每个输入记录均与 0 或 1 个输出端口表达式匹配。具有匹配条件的第一个输出端口将接收输入记录。不会对后续端口条件进行计算。如果表达式不匹配，则会转到下一个端口来计算表达式，直到匹配项出现后才在匹配端口中提交数据。

如果存在某些条件优于其他条件的数据趋势，则可以指定自定义计算序列，在开始时列出更高的概率条件。当输出端口数增加时，这种趋势会特别有用，因为它可以帮助避免计算不必要的条件。

请参见《Sybase ETL 4.9 用户指南》中的第 5 章“组件”。

事务性增强

Sybase ETL 4.9 在写入操作结束时在执行或模拟作业和项目的多个组件中扩展数据提交和回退支持。系统会在成功执行的写入操作结束时提交数据，对于未成功执行的写入操作将进行回退。此外，**Multi-Project** 和 **Synchronizer** 作业组件现在提供对包含项目的更多的控制，允许您在处理完成后立即提交中间作业。如果 **Multi-Project** 组件的项目下游失败，则已提交的项目始终不受影响；仅会对下游项目进行回退。

在 4.9 之前的版本中，无法跨多个组件或项目使用事务。此外，您无法在目标组件中回退完整的更改集。这是因为，事务性仅在组件和批处理级别上受支持。

现在，您可以跨多个项目使用事务性，进而允许您执行或应用全有或全无的数据库更新和装载策略。您还可以设置作业或项目属性，以便对支持事务的所有这些组件启用提交和回退功能。如果不选择“**Propagate Rollback**”选项，则项目或作业不会强制对成功的组件执行事务回退（如果一个或多个组件失败）。每个失败的组件都会回退其各自的事务。

您还可以选择提交还是回退事务组件在项目模拟期间完成的任何操作。可以在执行终止时模拟成功或失败条件，并将项目重置为其初始状态。这将清除所有的端口缓冲区，释放临时表，并关闭数据库连接和临时文件。您还可以在作业或项目中对所有组件禁用新的事务性功能。

使用改进事务性功能，可以：

- 执行从源数据库到 Sybase IQ 的增量数据装载。如果成功，将插入新行并将修改的行更新到 Sybase IQ 中。如果执行失败，则对 IQ 中的目标表和暂存表所做的所有修改都将作为单个事务的一部分进行回退。
- 将多个表装载到 Sybase IQ 中。如果成功，将在 IQ 中的所有表中装载来自源数据库中的数据的新信息，如果执行失败，则将回退对所有目标表进行的修改。
- 在单个事务中从目标 IQ 数据库中删除多个行。
- 通过 Load Table 加载 IQ 作为单个事务的一部分。
- 在项目执行结束时，从目标表回退源组件预处理和后处理 SQL 数据。
- 使用 Text Data Sink 组件将数据写入本地文本文件时，模拟 DBMS 样式的事务。
- 在单个分布式事务中提交多项目作业的所有项目。
- 模拟事务项目，并在模拟过程中的任何步骤提交或回退。
- 停止事务作业并确保写入作业中的事务资源的数据进行回退。
- 使用迁移向导创建作业，为所有项目和目标 DB 组件定义事务性。

所有数据库更改（包括数据库事务中涵盖的 pre-SQL 和 post-SQL 语句随项目中的所有其他事务组件一起进行回退（如果项目执行失败）。您还可以在项目模拟期间手动回退所有未提交的任务。

但是，在执行可能需要回退的任何 pre-SQL 命令之前，必须执行数据定义语言 (DDL) SQL 命令（create、drop 和 alter）。这是因为 DDL SQL 命令会导致数据库服务器执行隐式提交事务。例如，如果将数据插入到表中，并执行 CREATE TABLE 命令，则插入的数据无法回退，因为它已经由数据库隐式提交。

请参见《Sybase ETL 4.9 用户指南》中的第 5 章“组件”。

对连接共享的支持

ETL 4.9 包括对所有以下目标组件的连接共享的支持：

- DB Bulk Load Sybase IQ
- DB Data Sink Insert
- DB Data Sink Update
- DB Data Sink Delete

借助这项新功能，可以使用多个数据库组件来访问相同的源表或目标表，而不会出现任何锁定表失败错误。这项新功能允许各个组件与定义了相同的连接和数据库参数的其他目标组件共享到数据库的单个连接。在 4.9 之前的版本中，当两个或更多组件尝试同时访问同一个表时，项目执行会失败。现在，使用这项新的连接共享功能，您可以同时从多个源表提取数据并将其上载到单个目标表中。此外，还可以借助于该功能执行到 Sybase IQ 数据库的批量装载，进而获得更好的性能。

连接共享功能仅在同一个项目的各个组件之间受支持。存在于不同项目但在同一个作业内的组件不能共享连接。此外，您不能将连接共享功能与现有的 ETL 多写入器功能结合使用。

注意 定义相同的数据库接口和登录信息，但具有不同数据库选项的组件不能共享连接，并且会在执行或模拟项目时生成错误。

请参见《Sybase ETL 4.9 用户指南》中的第 5 章“组件”。

DB Staging 组件增强

使用 Sybase IQ 作为暂存平台时，Sybase ETL 4.9 允许您使用 Load Table 语句将数据批量装载到暂存表中，进而使性能更快。在早期版本中，DB Staging 组件可以使用 INSERT 语句仅执行基于行的操作，以便将传入数据装载到目标数据库中。

DB Staging 组件已经过增强，可以：

- 在运行时根据输入端口结构自动创建表。项目完成处理后，可以删除在运行时创建的表。
- 在创建输出之前，执行其他 SQL 脚本以修改或更新 staging 表。现在，在将转换流中的所有数据装载到与输入端口关联的表之后，且在检索查询结果集之前，可以执行 SQL 转换脚本。
- 确保在成功执行的写入操作结束时提交数据，对于未成功执行的写入操作将进行回退。

在对将 ODBC 作为接口的 Sybase IQ 15.0 staging 数据库启用客户端装载时，您不需要指定“Load Stage Path”。IQ 服务器将自动使用其缺省 LOAD TABLE 语句将记录从位于远程主机上的文件添加到 Sybase IQ 表中。

请参见《Sybase ETL 4.9 用户指南》中的第 5 章“组件”。

Data Sink 组件增强

DB Data Sink Insert、DB Data Sink Update 和 DB Data Sink Delete 组件已经过增强，可以使用 Load Table 语句将数据批量装载到 Sybase IQ 数据库的目标表中。

请参见《Sybase ETL 4.9 用户指南》中的第 5 章“组件”。

并发执行多个项目

Sybase ETL 4.9 支持并行执行项目，进而可以提高性能，特别是在使用单个 ETL 服务器计划多个项目时尤为如此。在早期版本中，只能针对每个网格引擎对每个作业运行一个项目，这会导致可用资源利用不足。现在，您可以在单个或多个网格引擎上并发执行多个 ETL 项目。多个 ETL 用户还可以在同一个远程网格引擎上并发执行项目。

必须编辑位于安装文件夹的 etc 子目录中的 Default.ini 文件，以指定要同时执行的最大项目数。缺省情况下，一次可执行的最大项目数为 10。但是，如果将最大项目值 (MAXPROJECTS) 设置为 0 或负数，则对可在网络引擎上并发执行的项目数没有任何限制。

请参见《Sybase ETL 4.9 用户指南》中的第 6 章“Sybase ETL 服务器”。

支持的存储库

Sybase ETL 4.9 仅支持 SQL Anywhere® 11 作为其存储数据库。如果使用的是早期版本的 ETL 中的任何其他存储库，则必须将现有 ETL 存储库迁移到 SQL Anywhere 11。

请参见《Sybase ETL 安装指南》的第 4 章“升级”中的“从现有存储库迁移到 SQL Anywhere”。

