



Informatica® PowerCenter
10.5

高级工作流指南

Informatica PowerCenter 高级工作流指南
10.5
2020 年 3 月

© 版权所有 Informatica LLC 2001, 2021

本软件和文档仅根据包含使用与披露限制的单独许可协议提供。未事先征得 Informatica LLC 同意，不得以任何形式、通过任何手段（电子、影印、录制或其他手段）复制或传播本文档的任何部分。

Informatica、Informatica 标志、PowerCenter 和 PowerExchange 是 Informatica LLC 在美国和世界其他许多司法管辖区的商标或注册商标。欲获得 Informatica 商标的最新列表，请访问 <https://www.informatica.com/trademarks.html>。其他公司和产品名称可能是其各自所有者的商业名称或商标。

美国政府权利交付给美国政府客户的程序、软件、数据库及相关文档和技术数据是指适用的联邦采购条例和政府机构特定补充条例中定义的"商业计算机软件"或"商业技术数据"。因此，使用、复制、披露、修改和改编应遵循适用的政府合同中规定的限制和许可条款、政府合同条款的适用范围以及 FAR 52.227-19 商用计算机软件许可中规定的额外权利。

本软件和/或文档中的若干部分受第三方版权约束。所需的第三方声明随产品一起提供。

本文档中的信息如有更改，恕不另行通知。如发现本文档中有什么问题，请通过以下电子邮件地址向我们报告：infa_documentation@informatica.com。

Informatica 产品根据对应协议的条款和条件进行担保。INFORMATICA 按"原样"提供本文档中的信息，无任何明示或暗示的担保，包括但不限于任何适销性和特定用途适用性担保，也没有任何非侵权担保或条件。

发布日期: 2021-05-03

目录

前言	15
Informatica 资源	15
Informatica Network	15
Informatica 知识库	15
Informatica 文档	15
Informatica 产品可用性矩阵	16
Informatica Velocity	16
Informatica Marketplace	16
Informatica 全球客户支持部门	16
 第 1 章：了解管道分区	17
了解管道分区概览	17
分区属性	18
分区点	18
分区数	18
分区类型	20
动态分区	20
配置动态分区	21
动态分区的规则和准则	21
结合使用动态分区和分区类型	21
配置分区级别的属性	22
缓存分区	22
分区管道中的映射变量	22
分区规则	23
编辑对象的分区约束	23
PowerExchange 的分区约束	24
配置分区	24
将分区点添加到管道	25
配置分区点	25
分区点节点	27
非分区点节点	28
 第 2 章：分区点	29
分区点概览	29
添加和删除分区点	30
添加和删除分区点的规则和准则	30
对关系源进行分区	32
输入 SQL 查询	32
输入筛选条件	32
对文件源进行分区	33

对文件源进行分区的规则和准则.	33
使用一个线程读取文件源.	34
使用多个线程读取文件源.	34
配置以实现文件分区.	34
对关系目标进行分区.	38
数据库兼容性.	38
对文件目标进行分区.	38
配置连接设置.	38
配置文件属性.	39
对自定义转换进行分区.	40
使用多个分区.	41
创建分区点.	41
使用线程.	41
对联接器转换进行分区.	42
对已排序联接器转换进行分区.	42
使用已排序平面文件.	43
使用已排序关系数据.	44
使用排序器转换.	45
使用分区优化已排序联接器转换.	46
对查找转换进行分区.	46
缓存分区查找转换.	47
对管道查找转换缓存进行分区.	47
对序列生成器转换进行分区.	48
对排序器转换进行分区.	48
配置排序器转换工作目录.	49
对 XML 生成器转换进行分区.	49
转换的限制.	49
数值函数的限制.	49
第 3 章：分区类型.	50
分区类型概览.	50
在管道中设置分区类型.	51
设置分区类型.	51
数据库分区的分区类型.	53
对数据库源进行分区.	53
目标数据库分区.	55
哈希自动键分区类型.	56
哈希用户键分区类型.	56
键范围分区类型.	57
添加分区键.	58
添加键范围.	59
传递分区类型.	60
循环分区类型.	61

第 4 章：下推优化.....	62
下推优化概览.....	62
下推优化类型.....	63
运行源端下推优化会话.....	63
运行目标端下推优化会话.....	63
运行完整下推优化会话.....	63
活动和空闲数据库.....	64
使用数据库.....	64
使用 ODBC 连接下推优化到数据库.....	65
比较集成服务的输出和数据库.....	65
IBM DB2 的规则和准则.....	66
Netezza 的规则和准则.....	66
Teradata 的规则和准则.....	66
Vertica 的规则和准则.....	66
Microsoft Azure SQL 数据仓库的规则和指南.....	66
下推兼容性.....	67
连接兼容性.....	67
日期兼容性.....	69
运算符兼容性.....	70
变量兼容性.....	70
适用于云数据仓库应用程序的函数.....	70
适用于数据仓库应用程序的函数.....	76
适用于企业版应用程序的函数.....	80
适用于关系数据库的函数.....	80
错误处理、日志记录和恢复.....	83
错误处理.....	83
日志记录.....	83
恢复.....	84
使用渐变维度.....	84
使用序列和视图.....	84
序列.....	84
视图.....	85
孤立序列和视图故障排除.....	86
使用 \$\$PushdownConfig 映射参数.....	87
为会话配置下推优化.....	89
下推选项.....	89
分区.....	89
目标加载规则.....	91
查看下推组.....	91
第 5 章：下推优化和转换.....	93
下推优化和转换概览.....	93

常规下推限制.	94
汇总器转换.	94
表达式转换.	95
筛选器转换.	96
联接器转换.	97
查找转换.	98
未连接的查找转换.	99
具有 SQL 替代的查找转换.	99
路由器转换.	100
序列生成器转换.	100
排序器转换.	101
源限定符转换.	102
具有 SQL 替代的源限定符转换.	103
目标.	104
联合转换.	105
更新策略转换.	105
第 6 章：实时处理.	107
实时处理概览.	107
了解实时数据.	108
消息和消息队列.	108
Web 服务消息.	109
来自 PowerExchange CDC 源的更改数据.	109
配置实时会话.	110
终止条件.	111
空闲时间.	111
消息计数.	111
读取器时间限制.	111
刷新延迟.	112
提交类型.	112
消息恢复.	113
前提条件.	113
启用消息恢复的步骤.	113
恢复文件.	114
针对 JMS 和 WebSphere MQ 源的消息恢复.	114
针对 SAP IDoc、TIBCO 和 webMethods 源的消息恢复.	115
消息恢复.	115
会话恢复数据刷新.	115
恢复表.	116
PM_REC_STATE 表.	116
消息处理.	116
消息恢复.	116
恢复队列和恢复主题.	117

消息处理.	117
消息恢复.	117
恢复忽略列表.	117
停止实时会话.	118
重新启动和恢复实时会话.	118
重新启动实时会话.	118
恢复实时会话.	118
重新启动和恢复命令.	119
实时会话的规则和准则.	119
消息恢复的规则和准则.	119
实时处理示例.	120
PowerCenter 实时产品.	121
 第 7 章：提交点.	 123
提交点概览.	123
基于目标的提交.	123
基于源的提交.	124
确定提交源.	124
从基于源的提交切换到基于目标的提交.	125
用户定义的提交.	127
回滚事务.	128
了解事务控制.	130
转换范围.	130
了解事务控制单元.	132
使用事务控制的规则和准则.	133
按事务创建目标文件.	133
设置提交属性.	133
 第 8 章：行错误日志记录.	 135
行错误日志记录概览.	135
错误日志代码页.	136
了解错误日志表.	136
PMERR_DATA.	136
PMERR_MSG.	138
PMERR_SESS.	139
PMERR_TRANS.	139
了解错误日志文件.	140
配置错误日志选项.	142
 第 9 章： workflow 恢复.	 144
workflow 恢复概览.	144
运行状态.	145
workflow 运行状态.	145

会话运行状态.	145
目标恢复表.	145
恢复选项.	148
挂起工作流.	149
配置挂起电子邮件.	149
配置工作流恢复.	149
恢复已停止、已中止和已终止的工作流.	150
恢复挂起的工作流.	150
配置任务恢复.	150
任务恢复策略.	151
自动恢复终止的任务.	152
继续进行会话.	152
使用可重复数据.	153
源可重复性.	154
转换可重复性.	154
配置映射以进行恢复.	154
恢复工作流和任务的步骤.	157
恢复工作流.	157
恢复会话.	157
从会话中恢复工作流.	158
会话恢复的规则和准则.	158
将恢复配置为从上一检查点继续.	158
不可恢复的工作流或任务.	158
 第 10 章： 停止和中止.	 160
停止和中止概览.	160
错误类型.	161
阈值错误.	161
致命错误.	161
会话失败的集成服务处理.	161
停止或中止工作流.	162
停止或中止任务.	162
停止或中止的步骤.	163
 第 11 章： 并发工作流.	 164
并发工作流概览.	164
配置唯一的工作流实例.	164
按实例名称恢复工作流实例.	165
运行同一实例名称的并发实例的规则和准则.	165
配置同一名称的并发工作流.	165
运行并发 Web 服务工作流.	165
配置同一名称的工作流实例.	166
恢复相同名称的工作流实例.	166

运行同一实例名称的并发实例的规则和准则.	166
使用参数和变量.	166
访问运行实例名称或运行 ID.	167
配置并发工作流的步骤.	167
启动和停止并发工作流.	168
从 Workflow Designer 启动工作流实例.	168
启动一个并发工作流.	168
从命令行启动并发工作流.	168
停止或中止并发工作流.	169
监视并发工作流.	169
查看会话和工作流日志.	169
唯一的工作流实例的日志文件.	170
同一名称的工作流实例的日志文件.	170
并发工作流的规则和准则.	170
 第 12 章： 网格处理.	 172
网格处理概览.	172
在网格上运行工作流.	173
在网格上运行会话.	173
使用分区组.	174
不根据资源要求形成分区组.	174
根据资源要求形成分区组.	175
创建分区组的规则和准则.	175
使用缓存.	175
网格连接和恢复.	176
将工作流或会话配置为在网格上运行.	176
将工作流或会话配置为在网格上运行的规则和准则.	176
 第 13 章： 负载均衡器.	 178
负载均衡器概览.	178
向工作流分配服务级别.	178
向任务分配资源.	179
 第 14 章： 工作流变量.	 181
工作流变量概览.	181
预定义工作流变量.	182
使用表达式中的预定义工作流变量.	184
评估工作流中的条件.	185
评估工作流中的任务状态.	185
在工作流中评估上一个任务状态.	185
用户定义的工作流变量.	186
工作流变量起始值和当前值.	187
数据类型默认值.	187

创建用户定义的工作流变量.	187
使用工作集变量.	189
持久性工作集变量.	189
替代初始值.	189
使用工作集变量的规则和准则.	189
在工作集中分配变量值.	189
在工作集之间传递变量值.	190
配置变量分配.	190
 第 15 章：会话中的参数和变量.	192
使用会话参数.	192
更改会话日志名称.	194
更改目标文件和目录.	195
更改文件中的源参数.	195
更改连接参数.	195
获取运行时信息.	196
创建文件参数和数据库连接参数的规则和准则.	196
会话中的映射参数和变量.	196
在会话中分配参数和变量值.	197
在会话之间传递参数和变量值.	197
配置变量分配.	198
 第 16 章：形参文件.	199
参数文件概览.	199
参数和变量类型.	200
在何处使用参数和变量.	200
替代参数文件中的连接属性.	207
参数文件结构.	208
参数文件部分.	208
注释.	209
空值.	210
示例参数文件.	210
配置参数文件名称和位置.	210
将参数文件与工作流或会话结合使用.	211
将参数文件与 pmcmd 结合使用.	212
参数文件示例.	213
创建参数文件的准则.	213
参数和参数文件故障排除.	214
参数和参数文件的提示.	215
 第 17 章：FastExport.	216
使用 FastExport 概览.	216
步骤 1. 创建 FastExport 连接.	217

验证代码页映射文件.	218
步骤 2.更改读取器.	218
步骤 3.更改源连接.	219
步骤 4.替代控制文件（可选）.	219
使用 FastExport 的规则和准则.	220
第 18 章：外部加载.	221
外部加载概览.	221
开始之前.	221
外部加载器行为.	222
将数据加载到命名管道.	222
将数据暂存到平面文件.	222
使用外部加载器对会话进行分区.	222
加载到 IBM DB2.	223
IBM DB2 EE 外部加载器.	223
IBM DB2 EEE 外部加载器.	223
IBM DB2 EEE 外部加载器的规则和准则.	224
设置操作模式.	224
配置授权、特权和权限.	224
配置 IBM DB2 EE 外部加载器属性.	225
配置 IBM DB2 EEE 外部加载器属性.	226
加载到 Oracle.	228
Oracle 外部加载器的规则和准则.	228
将多字节数据加载到 Oracle.	229
配置 Oracle 外部加载器属性.	229
加载到 Sybase IQ.	230
Sybase IQ 外部加载器的规则和准则.	230
将多字节数据加载到 Sybase IQ.	230
配置 Sybase IQ 外部加载器属性.	230
加载到 Teradata.	231
Teradata 外部加载器的规则和准则.	232
替代控制文件.	232
在控制文件中创建用户变量.	232
配置 Teradata MultiLoad 外部加载器属性.	233
配置 Teradata TPump 外部加载器属性.	235
配置 Teradata FastLoad 外部加载器属性.	237
在会话中配置外部加载.	238
将会话配置为写入文件.	239
配置文件属性.	239
选择外部加载器连接.	239
对外部加载进行故障排除.	240

第 19 章：FTP.....	241
FTP 概览.....	241
使用 FTP 的规则和准则.....	241
SFTP.....	242
集成服务行为.....	242
将 FTP 与源文件结合使用.....	242
将 FTP 与目标文件结合使用.....	243
在会话中配置 FTP.....	243
在会话中配置 SFTP.....	243
选择 FTP 连接.....	244
配置源文件属性.....	244
配置目标文件属性.....	245
第 20 章：会话缓存.....	248
会话缓存概览.....	248
高速缓存.....	249
缓存文件.....	250
缓存文件的命名约定.....	250
缓存文件目录.....	252
配置缓存大小.....	252
计算缓存大小.....	252
自动缓存大小.....	253
配置数值缓存大小.....	253
配置缓存大小的步骤.....	253
缓存分区.....	254
为缓存分区配置缓存大小.....	255
汇总器缓存.....	255
增量汇总.....	255
为汇总器转换配置缓存大小.....	255
汇总器缓存故障排除.....	256
联接器缓存.....	256
1:n 分区.....	257
n:n 分区.....	257
为联接器转换配置缓存大小.....	257
联接器缓存故障排除.....	258
查找缓存.....	259
共享缓存.....	259
为查找转换配置缓存大小.....	259
等级缓存.....	260
为等级转换配置缓存大小.....	260
排序器缓存.....	261
为排序器转换配置缓存大小.....	261

XML 目标缓存.	262
为 XML 目标配置缓存大小.	262
优化缓存大小.	262
 第 21 章：增量汇总.	 264
增量汇总概览.	264
增量汇总的集成服务处理.	265
重新初始化汇总文件.	265
移动或删除汇总文件.	265
查找索引和数据文件.	266
增量汇总的分区准则.	266
为增量汇总做准备.	267
配置映射.	267
配置会话.	267
 第 22 章：会话日志界面.	 268
会话日志界面概览.	268
实现会话日志界面.	268
集成服务和会话日志界面.	268
实现会话日志界面的规则和准则.	269
会话日志界面中的函数.	269
INFA_InitSessionLog.	269
INFA_OutputSessionLogMsg.	270
INFA_OutputSessionLogFatalMsg.	271
INFA_EndSessionLog.	271
INFA_AbnormalSessionTermination.	272
会话日志界面示例.	272
构建外部会话日志库.	272
使用外部会话日志库.	273
 第 23 章：了解缓冲区内存.	 274
了解缓冲区内存概览.	274
自动缓冲区内存设置.	275
将会话配置对象用于内存配置.	275
配置缓冲区内存.	275
配置会话缓存内存.	275
会话缓存限制.	276
配置会话缓存的自动内存设置.	276
 第 24 章：高精度数据.	 278
高精度数据概览.	278
长整型.	278
小数.	279

索引.....	280
---------	-----

前言

要了解下推优化、管道分区和网格处理等高级工作流概念，请阅读《PowerCenter® 高级工作流指南》。您还可以创建会话并查看会话日志中的信息。您可以添加转换逻辑和将管道分区来优化性能。

Informatica 资源

Informatica 通过 Informatica Network 和其他在线门户为您提供一系列产品资源。使用这些资源，可以充分利用 Informatica 产品和解决方案，并向其他 Informatica 用户和主题专家学习。

Informatica Network

在 Informatica Network 中可以获得许多资源，包括 Informatica 知识库和 Informatica 全球客户支持。要进入 Informatica Network，请访问 <https://network.informatica.com>。

作为 Informatica Network 成员，您可以选择以下服务：

- 在知识库中搜索产品资源。
- 查看产品可用性信息。
- 创建并检查您的支持案例。
- 查找当地的 Informatica 用户组网络并与您的伙伴进行协作。

Informatica 知识库

使用 Informatica 知识库可查找产品资源，例如操作方法文章、最佳实践、视频教程以及常见问题的答案。

要搜索知识库，请访问 <https://search.informatica.com>。如果您对知识库有任何疑问、意见或建议，请与 Informatica 知识库团队联系，电子邮件地址为 KB_Feedback@informatica.com。

Informatica 文档

使用 Informatica 文档门户可浏览大量当前与最近产品版本的文档库。要浏览文档门户，请访问 <https://docs.informatica.com>。

如果您对产品文档有任何疑问、意见或建议，请与 Informatica 文档团队联系，电子邮件地址为 infa_documentation@informatica.com。

Informatica 产品可用性矩阵

产品可用性矩阵 (PAM) 指明了产品版本支持的操作系统版本、数据库以及数据源和目标的类型。您可以在以下网址中浏览 Informatica PAM:

<https://network.informatica.com/community/informatica-network/product-availability-matrices>。

Informatica Velocity

Informatica Velocity 是由 Informatica 专业服务根据数百个数据管理项目的实际经验所开发出来的，其中汇集了大量使用技巧和最佳实践。Informatica Velocity 代表了 Informatica 顾问的集体知识，这些顾问与世界各地的组织合作，共同计划、开发、部署和维护成功的数据管理解决方案。

您可以在以下网址中找到 Informatica Velocity 资源：<http://velocity.informatica.com>。如果您对 Informatica Velocity 有任何疑问、意见或建议，请通过 ips@informatica.com 与 Informatica 专业服务联系。

Informatica Marketplace

Informatica Marketplace 是一个论坛，该论坛中提供的解决方案可扩展和增强您的 Informatica 实施。利用 Informatica 开发人员和合作伙伴在 Marketplace 中提供的数以百计的解决方案，可提高您的工作效率并加快项目实施时间。您可以在以下网址中找到 Informatica Marketplace：<https://marketplace.informatica.com>。

Informatica 全球客户支持部门

您可以通过电话或 Informatica Network 与全球支持中心联系。

要查找您当地的 Informatica 全球客户支持部门电话号码，请访问 Informatica 网站，链接为：<https://www.informatica.com/services-and-training/customer-success-services/contact-us.html>。

要在 Informatica Network 上查找联机支持资源，请访问 <https://network.informatica.com>，然后选择 eSupport 选项。

第 1 章

了解管道分区

本章包括以下主题：

- [了解管道分区概览, 17](#)
- [分区属性, 18](#)
- [动态分区, 20](#)
- [缓存分区, 22](#)
- [分区管道中的映射变量, 22](#)
- [分区规则, 23](#)
- [配置分区, 24](#)

了解管道分区概览

为每个希望集成服务运行的映射创建会话。每个映射包含一个或多个管道。一个管道由一个源限定符和从该源限定符接收数据的所有转换与目标组成。集成服务运行会话时，通过分区管道和并行执行每个分区的提取、转换和加载操作，可以实现更高性能。

分区是在单个读取器、转换或写入器线程中执行的管道阶段。任何管道阶段中的分区数等于该阶段中的线程数。默认情况下，集成服务会在每个管道阶段中都创建一个分区。

如果拥有“分区”选项，可以为单个管道阶段配置多个分区。可以配置可控制主线程为管道创建的读取器、转换和写入器线程数的分区信息。可以配置集成服务从源中读取数据、将数据行分发给每个转换以及将数据写入到目标的方式。可以配置要使用的源和目标连接的数量。

完成以下任务以配置会话的分区：

- 设置分区属性，包括分区点、分区数量和分区类型。
- 可以启用集成服务以在运行时设置分区。启用动态分区后，集成服务会基于诸如源数据库分区或网格中的节点数等因素增加会话分区的数量。
- 在配置会话以进行分区后，可以配置每个转换的内存要求和缓存目录。
- 集成服务会计算目标加载顺序组中每个分区的映射变量。可以使用映射中的变量函数来设置变量值。
- 在一个管道中创建多个分区时，Workflow Manager 会验证集成服务是否会维护使用多个分区的会话中的数据一致性。在会话中编辑对象属性时，可能会影响分区并导致会话失败。
- 在会话属性中添加或编辑分区点。更改分区点时，可以定义分区类型以及添加或删除分区。

分区属性

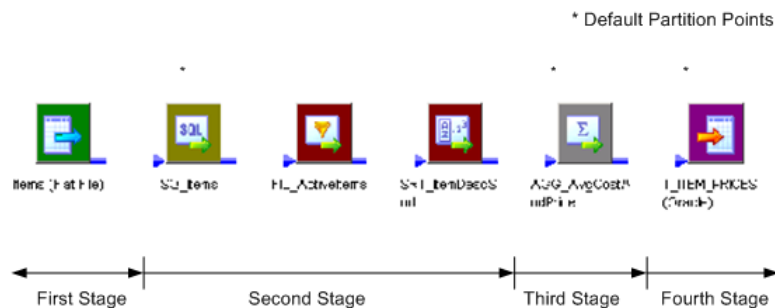
可以设置以下属性来分区管道：

- **分区点。**分区点标记了线程边界并将管道划分为不同阶段。集成服务会在分区点重新分发数据行。
- **分区数。**分区是在单个线程中执行的管道阶段。如果购买了分区点，则可以在任何分区点设置分区数。添加分区时，可以增加处理线程的数量，从而提高会话性能。
- **分区类型。**集成服务会在每个分区点创建默认分区类型。如果拥有“分区”选项，可以更改分区类型。分区类型可控制集成服务在分区点的分区之间分发数据的方式。

分区点

默认情况下，集成服务会在管道中的各种转换处设置分区点。分区点标记了线程边界并将管道划分为不同阶段。阶段是任意两个分区点之间的一段管道。在某个转换处设置分区点时，新管道阶段包括该转换。

下图显示了具有一个管道的映射的默认分区点和管道阶段：



添加一个分区点时，会为管道阶段数量加一。同样，删除一个分区点时，会为阶段数量减一。分区点标记了管道中的点，集成服务可以在这些位置在分区之间重新分发数据。

例如，如果将分区点置于筛选器转换处并定义多个分区，则集成服务可以在筛选器转换处理数据之前在分区之间重新分发数据行。在此分区点设置的分区类型可控制集成服务将数据行传递至每个分区的方式。

分区数

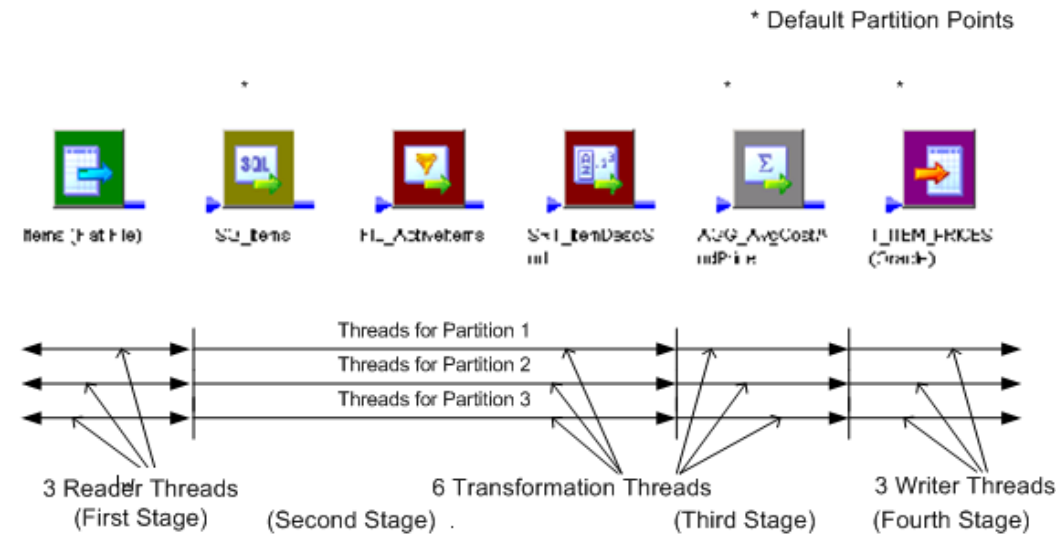
处理每个管道阶段的线程数取决于分区数。分区是在单个读取器、转换或写入器线程中执行的管道阶段。任何管道阶段中的分区数等于该阶段中的线程数。

在管道中的任何分区点最多可以定义 64 个分区。在任何分区点增加或减少分区数时，Workflow Manager 都会在管道中的所有分区点增加或减少分区数。分区数仍在整个管道中保持一致。如果在任何分区点定义三个分区，则 Workflow Manager 会在管道中的所有其他分区点均创建三个分区。在某些情况下，必须将管道中的分区数设置为一。

增加分区或分区点数目可以增加线程的数量。因此，增加分区或分区点数也会增加节点上的负载。如果节点包含足够 CPU 带宽，则并发处理会话中的数据行可以提高会话性能。但是，如果在会话中创建处理大量数据的大量分区或分区点，则可能会使系统过载。

创建的分区数等于与源或目标的连接数。如果管道包含关系源或目标，则源限定符或目标实例处的分区数等于与数据库的连接数。如果管道包含文件源，则可以配置会话以读取具有一个线程或多个线程的源。

下图显示了包含三个分区的映射中的线程：



例如，跨映射定义 3 个分区时，主线程会在每个管道阶段创建 3 个线程，总共 12 个线程。

集成服务会并发运行分区线程。运行具有多个分区的会话时，线程运行情况如下：

1. 读取器线程会并发运行以从源中提取数据。
2. 转换线程会在每个转换阶段中并发运行以处理数据。集成服务会在分区点的分区之间重新分发数据。
3. 写入器线程会并发运行以将数据写入目标。

分区多输入组转换

主线程会为目标加载顺序组中的每个管道创建读取器和转换线程。目标加载顺序组在其包含具有多输入组的转换时拥有多个管道。

将多个管道连接到多输入组转换时，集成服务会维持转换线程或创建新转换线程，具体取决于多输入组转换是否是分区点：

- **多输入组转换不存在分区点。**当多输入组转换不存在分区点时，集成服务会一次为阶段中的多输入组转换和所有下游转换处理一个线程。
- **多输入组转换存在分区点。**当多输入组转换存在分区点时，集成服务会创建新管道阶段并为每个分区处理具有一个线程的阶段。集成服务会为每个分区都创建一个转换线程，无论转换包含的输出组数量是多少。

分区类型

配置管道的分区信息时，必须在管道中的每个分区点定义分区类型。分区类型可确定 PowerCenter 集成服务在分区点之间重新分发数据的方式。

PowerCenter 集成服务会在每个分区点创建默认分区类型。如果拥有“分区”选项，可以更改分区类型。分区类型可控制 PowerCenter 集成服务在分区点的分区之间分发数据的方式。可以在管道中的不同点创建不同分区类型。

可以在 Workflow Manager 中定义以下分区类型：

- **数据库分区。**PowerCenter 集成服务会查询 IBM DB2 或 Oracle 数据库系统以查找表分区信息。它会从数据库中对应的节点读取已分区数据。可以在多节点表空间中结合使用数据库分区与 Oracle 或 IBM DB2 源实例。可以结合使用数据库分区与 DB2 目标。
- **哈希自动键。**PowerCenter 集成服务使用哈希函数来对各分区中的数据行进行分组。PowerCenter 集成服务基于分区键对数据进行分组。PowerCenter 集成服务将所有已分组或已排序端口用作复合分区键。在等级、排序器和未排序转换时可能需要使用哈希自动键分区。
- **哈希用户键。**PowerCenter 集成服务使用哈希函数来对各分区中的数据行进行分组。定义要生成分区键的端口数。
- **键范围。**使用键范围分区，PowerCenter 集成服务会基于您定义为分区键的一个端口或一组端口分发数据行。对于每个端口，都可以定义值范围。PowerCenter 集成服务使用键和范围将行发送至相应分区。当管道中的源和目标已按键范围进行分区时，使用键范围分区。
- **传递。**在传递分区中，PowerCenter 集成服务会处理数据，而不会在各分区中重新分发。单个分区中的所有行在错过传递分区点后将保留在该分区。希望创建其他管道阶段以提高性能但不希望更改各分区之间的数据分发时，选择传递分区。
- **循环。**PowerCenter 集成服务会将数据块分发给一个或多个分区。使用循环分区以便每个分区均可基于块大小和数量处理行。

动态分区

如果数据量增长或添加更多 CPU，可能需要调整分区，因此会话运行时不会增加。使用动态分区时，可以配置分区信息，因此集成服务可确定要在运行时创建的分区数量。

集成服务会基于诸如源数据库分区或网格中的节点数等因素增加运行时的会话分区数量。

如果阶段中的任何转换都不支持分区，或分区配置不支持动态分区，则集成服务无法增加管道中的分区数量。数据通过一个分区传递。

完成以下任务以增加使用动态分区的会话分区数量：

- **设置分区。**集成服务会根据您选择的分区方法增加分区数量。
- **设置动态分区的会话属性。**可以设置标识源和目标文件名和目录的会话属性。会话使用会话属性来为其在运行时创建的每个分区创建分区级别属性。
- **配置分区类型。**可以使用会话属性“映射”选项卡上的“分区”视图编辑分区点和分区类型。

注意：请勿为包含手动分区的会话配置动态分区。如果将动态分区设置为除“禁用”之外的值并手动分区会话，则该会话无效。

配置动态分区

您在会话属性的“配置对象”选项卡上配置动态分区。使用以下其中一种方法配置动态分区：

- **禁用。**请勿使用动态分区。在“映射”选项卡上定义分区数量。
- **基于分区数量。**将分区数设置为您在“分区数量”属性中定义的数量。请使用 `$DynamicPartitionCount` 会话参数，或输入一个大于 1 的数字。
- **基于网格中的节点数量。**将分区数设置为运行会话的网格中的节点数量。如果您为不在网格上运行的会话配置此选项，会话将在一个分区内运行，并在会话日志中记录相关消息。
- **基于源分区。**使用数据库分区信息确定分区数量。分区数量是源的最大分区数。对于使用复合分区的 Oracle 源，分区数量为该源的子分区数最大值。
- **基于 CPU 数量。**将分区数量设置为等于准备会话的节点上的 CPU 数量。如果您将会话配置为在网格上运行，动态分区功能会将分区数量设置为准备会话的节点上的 CPU 数量乘以网格中的节点数量。

相关主题：

- [“数据库分区的分区类型” 页面上 53](#)

动态分区的规则和准则

使用动态分区时，请遵循以下规则和准则：

- 动态分区为每个分区使用相同连接。
- 无法结合使用动态分区与 XML 源和目标。
- 无法结合使用动态分区与调试器。
- 如果启用动态分区，使用 SFTP 的会话将失败。
- 如果将动态分区设置为除“禁用”之外的值并在“映射”选项卡上手动分区会话，则会使该会话失效。
- 如果使用除 `$DynamicPartitionCount` 之外的参数设置分区数量，会话将失败。
- 以下动态分区配置会使会话使用一个分区来运行：
 - 替代汇总器、联接器、查找或等级转换的默认缓存目录。默认值为 `$PMCacheDir` 时，集成服务会分区转换缓存目录。
 - 替代排序器转换默认工作目录。默认值为 `$PMTempDir` 时，集成服务会分区排序器转换工作目录。
 - 可以结合使用无限制范围的数字和日期键与键范围分区类型。
 - 使用除数字或日期之外的数据类型作为键范围分区中的键。
 - 使用键范围关系目标分区。
 - 创建用户定义的 SQL 语句或用户定义的源筛选器。
 - 将动态分区设置为网格中的节点数量，并且会话不在网格上运行。
 - 使用传递关系源分区。
 - 结合使用动态分区和应用程序源限定符。
 - 结合使用 SDK 或 PowerConnect 源和目标与动态分区。

结合使用动态分区和分区类型

以下规则适用于结合使用动态分区与不同分区类型：

- **传递分区。**如果在分区点更改分区数量，则每个管道阶段中的分区数量都会更改。如果结合使用传递分区与关系源，则会话在阶段中的一个分区中运行。

- **键范围分区。**必须定义封闭范围的数字或日期键才能使用动态分区。这些键必须为数值或日期数据类型。动态分区不会使用关系目标上的键范围分区增加分区数量。
- **数据库分区。**使用数据库分区时，集成服务会基于源数据库分区创建会话分区。结合使用数据库分区与 Oracle 和 IBM DB2 源。
- **哈希自动键、哈希用户键或循环。**使用哈希用户键、哈希自动键和循环分区类型分发使用动态分区的行。希望集成服务按分组将行分发给分区时，使用哈希用户键和哈希自动键分区。希望 PowerCenter 集成服务将数据块分发给一个或多个分区时，使用循环分区。

配置分区级别的属性

使用动态分区时，集成服务会为其在运行时创建的每个分区定义分区级别属性。它会根据您在会话属性中定义的会话级别属性名称命名文件和目录属性。

例如，您将会话拒绝文件名定义为 `accting_detail.bad`。集成服务在运行时创建分区时，会为每个分区创建拒绝文件，如 `accting_detail1.bad`、`accting_detail2.bad`、`accting_detail3.bad`。

缓存分区

创建具有多个分区的会话时，集成服务会将缓存分区用于汇总器、联接器、查找、等级或排序器转换。集成服务分区缓存时，会为每个分区创建单独的缓存并将已配置的缓存大小分配给每个分区。集成服务在每个缓存中存储不同数据，每个缓存仅包含该分区所需的行。因此，集成服务需要总缓存内存的一部分用于每个分区。

在配置会话以进行分区后，可以在会话属性的“映射”选项卡上的“转换”视图中配置每个转换的内存要求和缓存目录。要配置内存要求，请计算转换的总要求并按分区数据进行划分。要提高性能，可以为每个分区配置单独的目录。

下表介绍了集成服务为每个适用转换使用缓存分区时的情况：

转换	说明
汇总器转换	在使用汇总器转换的会话中创建多个分区。无需在汇总器转换时设置分区点。
联接器转换	在联接器转换时创建分区点。
查找转换	在查找转换时创建哈希自动键分区点。
等级转换	在使用等级转换的会话中创建多个分区。无需在等级转换时设置分区点。
排序器转换	在使用排序器转换的会话中创建多个分区。无需在排序器转换时设置分区点。

分区管道中的映射变量

在使用映射变量的目标加载顺序组中指定多个分区时，集成服务会分别计算每个分区中映射变量的值。集成服务使用以下进程计算变量值：

1. 它会根据映射中使用的变量函数分别更新每个分区中的当前变量值。

- 2. 在加载目标加载顺序组中的所有目标后，集成服务会根据变量的汇总类型将每个分区中的当前值合并到单个最终值中。
- 3. 如果会话中有多个目标加载顺序组，则一个目标加载顺序组中映射变量的最终当前值会变为下一个目标加载顺序组中的当前值。
- 4. 集成服务完成加载最后一个目标加载顺序组时，会将变量的最终当前值保存至存储库。

使用映射中的以下变量函数之一来设置变量值：

- SetCountVariable
- SetMaxVariable
- SetMinVariable

下表介绍了集成服务跨分区计算变量值的方法：

变量函数	跨分区变量值计算
SetCountVariable	集成服务计算所有分区中的最终计数值。
SetMaxVariable	集成服务比较每个分区的最终变量值并保存最大值。
SetMinVariable	集成服务比较每个分区的最终变量值并保存最小值。

注意：对于管道中的每个映射变量，只能使用一次变量函数。集成服务会处理在映射中遇到的变量函数。对于每次会话运行，集成服务在映射中遇到变量函数的顺序可能各不相同。当您在映射中多次使用相同的变量函数时，这可能导致不一致的结果。

分区规则

如果集成服务在其处理已分区数据时维护数据一致性，则可以在管道中创建多个分区。创建会话时，Workflow Manager 会验证每个管道以进行分区。

编辑对象的分区约束

编辑对象属性时，可能会影响您在会话中创建多个分区或运行具有多个分区的现有会话的功能。

创建会话之前

创建会话时，Workflow Manager 会检查映射属性。映射会动态获取对快捷方式的更改，但不会获取对可重用对象的更改，如可重用转换和 Mapplet。因此，如果在保存映射后和创建会话前在 Designer 中编辑可重用对象，则必须打开该映射并重新保存以使 Workflow Manager 识别对象的更改。

创建具有多个分区的会话之后

如果在创建具有多个分区的会话后编辑映射，Workflow Manager 不会使会话失效，即使更改违反分区规则。集成服务会在下次会话运行时使会话失败，除非编辑会话以使其不再违反分区规则。

对映射的以下更改可能导致会话失败：

- 删除为分区点的转换。
- 添加为默认分区点的转换。

- 移动为其他管道分区点的转换。
- 使用以下任一方式更改为分区点的转换：
 - 现有分区类型无效。
 - 转换不再支持多个分区。
 - 转换不再是有效分区点。
- 在创建具有多个分区的管道后，禁用分区或在转换中更改单个节点和网格之间的分区。
- 在创建具有多个分区的管道后，切换连接器转换的主源和详细源。

PowerExchange 的分区约束

可为 PowerExchange® 和 PowerExchange Client for PowerCenter 指定多个分区。但是，还有其他限制。有关这些产品的详细信息，请参阅产品文档。

配置分区

创建或编辑会话时，可以在映射中更改每个管道的分区。如果映射包含多个管道，可以在某些管道中指定多个分区，并在其他管道中指定单个分区。可以使用会话属性“映射”选项卡上的“分区”视图更新分区信息。您可以在 Workflow Designer 中配置不可重用会话的分区，在任务开发程序中配置可重用会话的分区。

在会话属性的“分区”视图上添加、删除或编辑分区点。如果添加键范围分区点，可以定义每个范围中的键。

下表列出了“映射”选项卡上“分区”视图的配置选项：

分区视图选项	说明
添加分区点	单击可添加新分区点。添加分区点时，转换名称会在分区点节点下显示。
删除分区点	单击可删除选定分区点。 无法删除某些分区点。
编辑分区点	单击可编辑选定分区点。此时会打开“编辑分区点”对话框。
键范围	显示分区点的键和键范围，具体取决于分区类型。 对于键范围分区，指定键范围。 对于哈希用户键分区，此字段显示分区键。 Workflow Manager 不对其他分区类型显示此区域。
编辑键	单击可为键范围或哈希用户键分区添加或移除分区键。无法为哈希自动键、循环或传递分区创建分区键。

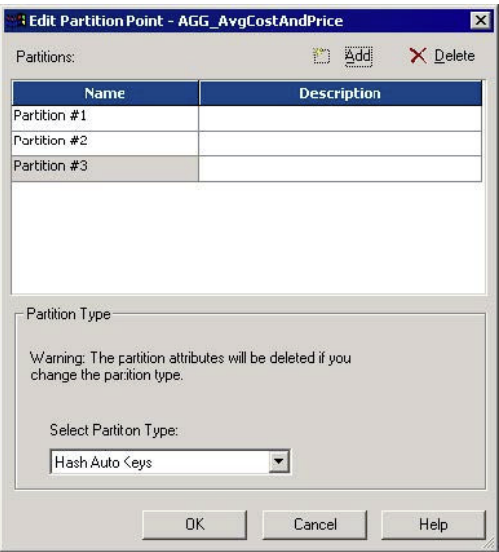
将分区点添加到管道

在会话属性的“映射”选项卡中添加分区点。

1. 在“映射”选项的“分区”视图上，选择尚不是分区点的转换，并单击“添加分区点”按钮。

提示: 可以从非分区点节点中选择转换。

下图显示了“编辑分区点”对话框，您可以在此对话框中向管道添加分区点：



转换显示在会话属性“映射”选项卡上“分区”视图中的分区点节点中。

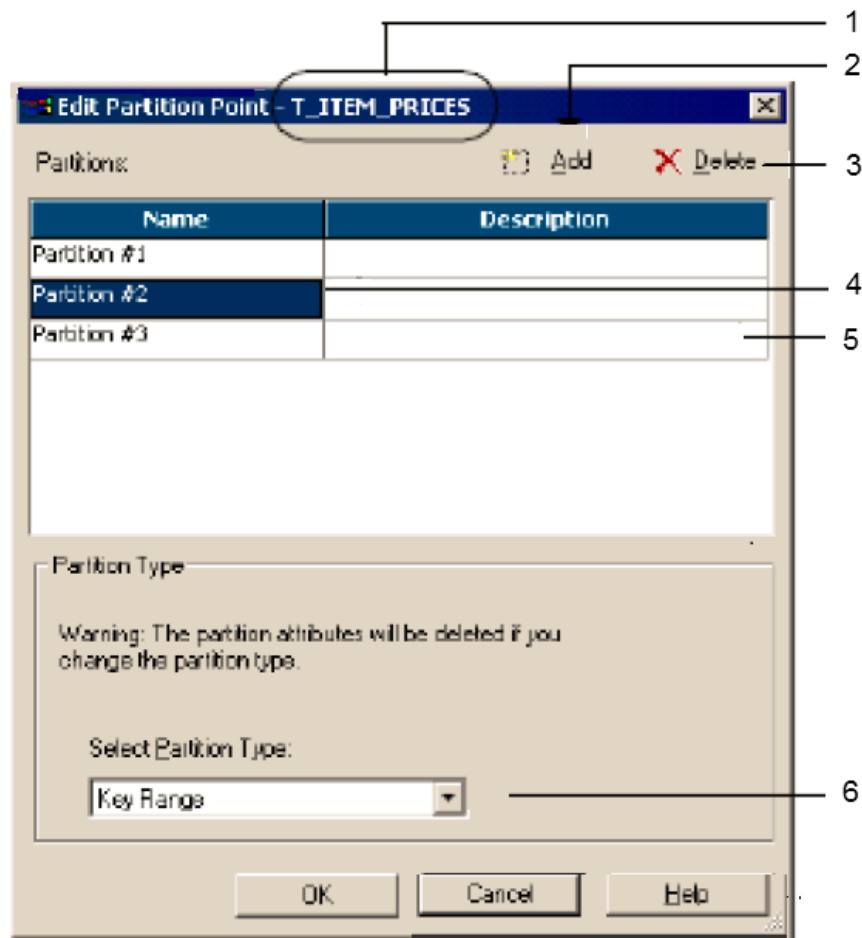
2. 为分区点选择分区类型或接受默认值。
3. 单击“确定”。

配置分区点

编辑或添加分区点时，可以执行以下任务：

- 在该分区点指定分区类型。
- 添加和删除分区。
- 为每个分区输入说明。

下图显示了可用于编辑分区点的配置选项：



1. 选定的分区点
2. 添加分区
3. 删除分区
4. 选择分区
5. 输入分区说明
6. 指定分区类型。

下表介绍了分区点的配置选项：

分区选项	说明
选择分区类型	更改分区类型。
分区名称	从此对话框中选择单个分区以进行配置。

分区选项	说明
添加分区	添加分区。在任何分区点最多可以添加 64 个分区。分区数仍必须在整个管道中保持一致。因此，如果在一个分区点定义三个分区，则 Workflow Manager 会在管道中的所有分区点均定义三个分区。
删除分区	删除选定分区。每个分区点必须至少包含一个分区。
说明	输入当前分区的可选说明。

可以为您创建的每个分区输入说明。要输入说明，请在“编辑分区点”对话框中选择分区，然后在“说明”字段中输入说明。

分区点节点

分区点节点显示具有转换图标的映射。分区点节点会列出树中的分区点。选择要配置其属性的分区点。

在分区点节点中，可以为映射中的每个管道配置以下选项：

- 添加和删除分区点。
- 在每个分区点指定分区类型。
- 添加和删除分区。
- 为每个分区输入说明。
- 为某些分区类型添加键和键范围。

下表介绍了分区点节点：

分区点节点	说明
添加分区点	单击可将新分区点添加到转换列表。
删除分区点	单击可删除当前分区点。无法删除某些分区点。
编辑分区点	单击可编辑当前分区点。
编辑键	单击可为键范围或哈希用户键分区添加、移除或编辑键。此按钮不可用于自动哈希、循环或传递分区。

编辑分区点

“编辑分区点”对话框可允许您添加和删除分区以及选择分区类型。

下表介绍了“编辑分区点”对话框中的选项：

编辑分区点选项	说明
“添加”按钮	单击可添加分区。最多可以添加 64 个分区。
“删除”按钮	单击可删除选定分区。
名称	分区数量。

编辑分区点选项	说明
说明	为当前分区输入说明。
选择分区类型	从列表选择一个分区类型。

编辑分区键

在任何分区点指定键范围或哈希用户键分区时，必须将一个或多个端口指定为分区键。单击“编辑键”以显示“编辑分区键”对话框。

可以将一个或多个端口指定为分区键。要重新排列构成键的端口的顺序，请在“选定端口”列表中选择某个端口，然后单击向上箭头或向下箭头。

非分区点节点

非分区点节点在图标化视图中显示映射对象。“分区点”节点将在树中列出非分区点。可以选择非分区点并添加分区（如果需要）。

第 2 章

分区点

本章包括以下主题：

- [分区点概览, 29](#)
- [添加和删除分区点, 30](#)
- [对关系源进行分区, 32](#)
- [对文件源进行分区, 33](#)
- [对关系目标进行分区, 38](#)
- [对文件目标进行分区, 38](#)
- [对自定义转换进行分区, 40](#)
- [对联接器转换进行分区, 42](#)
- [对查找转换进行分区, 46](#)
- [对序列生成器转换进行分区, 48](#)
- [对排序器转换进行分区, 48](#)
- [对 XML 生成器转换进行分区, 49](#)
- [转换的限制, 49](#)

分区点概览

分区点可标记管道中线程之间的边界。集成服务会在分区点重新分发数据行。您可以添加分区点以增加转换线程的数量并提高会话性能。

配置会话以对某个源数据库进行读取时，集成服务将为每个分区创建与该源数据库的单独连接和 SQL 查询。可以自定义或替代 SQL 查询。

配置会话以将数据加载到关系目标时，集成服务将为目标实例上的每个分区创建与目标数据库的单独连接。为该目标配置拒绝文件名和目录。集成服务为每个目标分区创建一个拒绝文件。

可以配置会话以使用一个线程或多个线程来读取源文件。必须为读取该文件的所有分区选择相同的连接类型。

配置会话以写入文件目标时，可以将目标输出写入每个分区的单独文件，或写入包含目标输出的所有分区的合并文件。可以为每个目标分区配置连接设置和文件属性。

在转换中创建分区点时，Workflow Manager 将设置默认的分区类型。可以根据转换类型更改分区类型。

添加和删除分区点

分区点标记了管道中的线程边界并将管道划分为不同阶段。

添加分区点时，将增加转换线程的数量，这可提高会话性能。集成服务可在分区点上重新分发数据行，这也可提高会话性能。创建会话时，Workflow Manager 将在管道中的每个转换中创建一个分区点。

可以根据下列管道中的转换或目标实例保留或删除分区点：

源限定符转换

控制集成服务如何从源提取数据并将其传递到源限定符。不能删除此分区点。

规范器转换

控制集成服务如何从源提取数据并将其传递到源限定符。不能删除此分区点。

等级转换

确保集成服务在将行发送到转换之前对其进行正确分组。如果管道中仅包含一个分区，或者集成服务在组中所有行进入转换之前将其传递到单个分区，则可以删除这些分区点。

未排序汇总器转换

确保集成服务在将行发送到转换之前对其进行正确分组。如果管道中仅包含一个分区，或者集成服务在组中所有行进入转换之前将其传递到单个分区，则可以删除这些分区点。

目标实例

控制写入器如何将数据传递到目标。不能删除此分区点。

多个输入组转换

将 Workflow Manager 配置为使用一个线程处理每个分区时，或将下游多个输入组自定义转换配置为使用一个线程处理每个分区时，Workflow Manager 将在多个输入组转换中创建一个分区点。

例如，Workflow Manager 可以在已排序联接器转换中创建一个分区点。将联接器转换连接到配置为每个分区使用一个线程的下游自定义转换时，Workflow Manager 将创建一个分区点。

这可确保集成服务在每个分区需要一个线程的自定义转换中使用一个线程来处理每个分区。不能删除此分区点。

添加和删除分区点的规则和准则

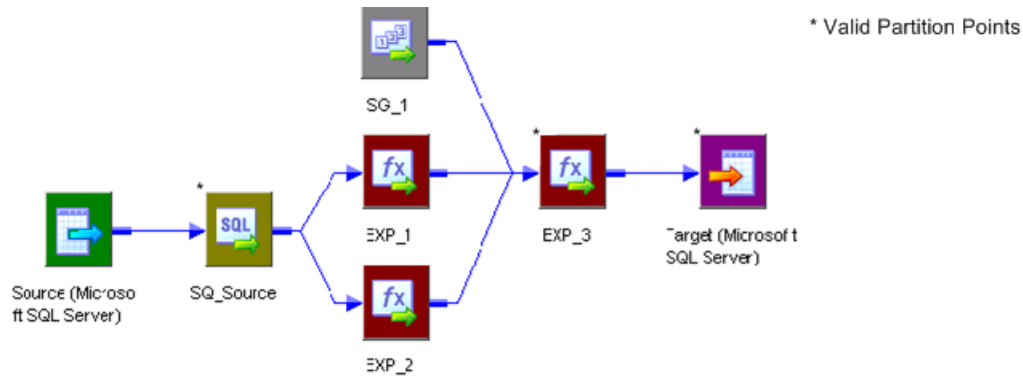
添加和删除分区点时适用以下规则和准则：

- 不能在源实例上创建分区点。
- 不能在序列生成器转换或未连接的转换中创建分区点。
- 只要没有分区点从多个管道阶段接收输入，即可在任何其他转换中添加分区点。
- 不能在源限定符转换、适用于 COBOL 源的规范器转换或目标实例中删除分区点。
- 不能在配置为每个分区使用一个线程的多个输入组自定义转换中删除分区点。
- 不能在配置为每个分区使用一个线程的多个输入组自定义转换的上游多个输入组转换中删除分区点。
- 以下分区类型对动态分区存在限制：
 - 传递。使用动态分区时，如果更改了某个分区点上的分区数量，则每个管道阶段中的分区数量都将发生更改。
 - 键范围。要将键范围与动态分区结合使用，必须定义已结束的数字范围或日期键。如果使用开放式的范围，会话将在一个分区内运行。

根据以下规则，可以在管道中的其他转换中添加和删除分区点：

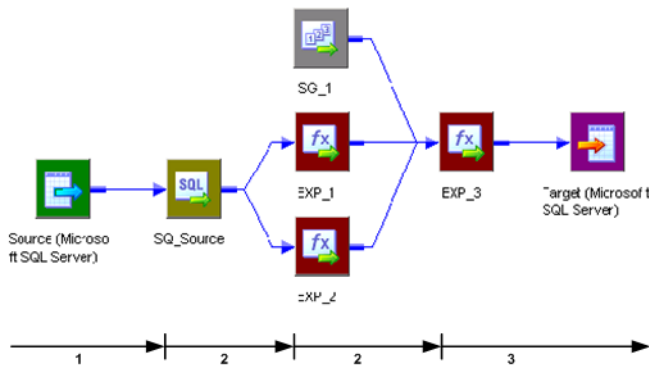
- 不能在源实例上创建分区点。
- 不能在序列生成器转换或未连接的转换中创建分区点。
- 只要没有分区点从多个管道阶段接收输入，即可在任何其他转换中添加分区点。

下图显示了映射中的有效分区点：



在此映射中，默认情况下，Workflow Manager 在源限定符和目标实例中创建分区点。可以在表达式转换 EXP_3 中额外放置一个分区点。

如果在 EXP_3 中放置一个分区点并定义一个分区，则主线程将创建以下线程：



1. 读取器线程。
2. 转换线程。
3. 写入器线程。

在此情况下，每个分区点仅从一个管道阶段接收数据，因此 EXP_3 是有效的分区点。

以下转换不是有效的分区点：

转换	原因
源代码	源实例。
SG_1	序列生成器转换。
EXP_1 和 EXP_2	如果可以在 EXP_1 或 EXP_2 上放置一个分区点，则可以创建一个额外的管道阶段，该阶段可将源限定符中的数据处理到 EXP_1 或 EXP_2 中。在此情况下，EXP_3 会从两个管道阶段接收数据，而这是不允许的。

对关系源进行分区

运行对关系源或应用程序源进行分区的会话时，集成服务将为每个分区创建与源数据库的单独连接。然后为每个分区创建 SQL 查询。可以在“映射”选项卡上的“转换”视图输入筛选条件，以便为每个源分区自定义查询。也可以使用“映射”选项卡上的“转换”视图来替代每个源分区的 SQL 查询。

注意：创建自定义 SQL 查询以读取数据库表以及设置数据库分区时，集成服务将还原到传递分区并在会话日志中显示一条消息。

输入 SQL 查询

如果要自定义 SQL 查询中的 SELECT 语句，可以输入 SQL 替代。在“映射”选项卡的“转换”视图上输入的 SQL 语句可替代配置源限定符转换时在 Designer 中设置的任何自定义 SQL 查询。

该 SQL 查询还可替代为源分区输入的任何键范围和筛选条件。因此，如果还输入键范围和源筛选器，集成服务会使用 SQL 查询替代来提取源数据。

如果创建的键包含空值，可以通过创建另一个分区并输入用于提取空值的 SQL 查询或筛选器来提取空值。

要为每个分区输入 SQL 查询，请单击“SQL 查询”字段中的“浏览”按钮。在“SQL 编辑器”对话框中输入查询，然后单击“确定”。

如果在配置源限定符转换时已在 Designer 中输入了 SQL 查询，则该查询会显示在每个分区的“SQL 查询”字段中。要替代此查询，请单击“SQL 查询”字段中的“浏览”按钮，在“SQL 编辑器”对话框中修订查询，然后单击“确定”。

输入筛选条件

如果在关系源限定符上指定键范围分区，则可以输入其他筛选条件。执行此操作时，集成服务将生成一个 WHERE 子句，其中包含在会话属性中输入的筛选条件。

在“映射”选项卡的“转换”视图上输入的筛选条件可替代配置源限定符转换时在 Designer 中设置的任何筛选条件。

如果使用键范围分区，可将筛选条件与键范围配合使用。例如，您要根据客户 ID 选择数据，但不希望提取美国以外客户的信息。定义以下键范围：

CUSTOMER_ID	Start Range	End Range
Partition #1		135000
Partition #2	135000	

如果了解到美国以外客户的 ID 均在特定的分区范围内，可以在该分区中输入一个筛选器以将其排除。因此，为第二个分区输入以下筛选条件：

```
CUSTOMERS.COUNTRY = 'USA'
```

会话运行时，会话日志中将显示以下两个分区的查询：

```
READER_1_1_1> RR_4010 SQ instance [SQ_CUSTOMERS] SQL Query [SELECT CUSTOMERS.CUSTOMER_ID,  
CUSTOMERS.COMPANY, CUSTOMERS.LAST_NAME FROM CUSTOMERS WHERE CUSTOMER.CUSTOMER ID < 135000]
```

```
[...]
```

```
READER_1_1_2> RR_4010 SQ instance [SQ_CUSTOMERS] SQL Query [SELECT CUSTOMERS.CUSTOMER_ID,  
CUSTOMERS.COMPANY, CUSTOMERS.LAST_NAME FROM CUSTOMERS WHERE CUSTOMERS.COUNTRY = 'USA' AND 135000 <=  
CUSTOMERS.CUSTOMER_ID]
```

要输入筛选条件，请单击“源筛选器”字段中的“浏览”按钮。在“SQL 编辑器”对话框中输入筛选条件，然后单击“确定”。

如果在配置源限定符转换时已在 Designer 中输入了筛选条件，则该查询会显示在每个分区的“源筛选器”字段中。要替代此筛选器，请单击“源筛选器”字段中的“浏览”按钮，在“SQL 编辑器”对话框中更改筛选条件，然后单击“确定”。

对文件源进行分区

当会话使用文件源时，可以对其进行配置以使用一个线程或多个线程来读取源。将会话配置为使用一个线程进行读取时，集成服务将创建一个与文件源的连接，而将会话配置为使用多个线程进行读取时，集成服务将创建多个与文件源的并发连接。

使用以下类型的已分区文件源：

- **平面文件。**可以将会话配置为读取平面文件、XML 或 COBOL 源文件。
- **命令。**可以将会话配置为使用操作系统命令生成源数据行或文件列表。

连接到文件源时，必须为所有分区选择相同的连接类型。可以选择不同的连接对象，只要每个对象属于相同的类型即可。

要为平面文件源指定单线程或多线程读取，请为分区 2 至 n 配置源文件名属性。要配置为单线程读取，请在分区 2 至 n 中传递空数据。要配置为多线程读取，请将分区 2 至 n 的源文件名留空。

对文件源进行分区的规则和准则

配置具有多个分区的文件源会话时，请遵循以下规则和准则：

- 在源限定符上使用传递分区。
- 对平面文件或 COBOL 源使用单线程读取或多线程读取。
- 对 XML 源使用单线程读取。
- 如果源文件为非磁盘文件（如 FTP 文件或 WebSphere MQ 源），则无法使用多线程读取。
- 如果使用转换敏感代码页，请在满足以下条件时使用多线程读取：
 - 该文件具有固定宽度。
 - 该文件不换行。
 - 未在源定义中启用用户定义的转换状态。
- 要从三个平面文件中并发读取数据，必须在源限定符上指定三个分区。接受默认的分区类型，传递。

- 如果将会话配置为多线程读取，而集成服务无法创建到单个文件源的多个线程，则会向会话日志写入一条消息并使用一个线程来读取源。
- 当集成服务使用多个线程来读取源文件时，可能无法按顺序读取文件中的行。如果排序顺序极为重要，请将会话配置为使用单个线程来读取文件。例如，如果映射包含已排序联接器转换并且文件源为排序来源，则排序顺序可能极为重要。
- 也可以组合使用直接文件和间接文件来实现负载平衡。
- 对于大型源文件，多线程读取的会话性能可达到最佳。如果输入数据量较少，可能无法实现负载平衡。
- 如果某一命令生成源数据，并且会话配置为在网格上运行或者是使用从上一检查点继续这一恢复策略配置的，则无法对文件源使用该命令。

使用一个线程读取文件源

集成服务使用一个线程来读取文件源时，会创建一个与源的连接。集成服务按顺序读取文件或文件列表中的行。可以为会话中的直接文件源或间接文件源配置单线程读取：

- **读取直接文件。**可以将集成服务配置为读取一个或多个直接文件。如果配置的会话具有多个直接文件，则集成服务会为每个文件创建一个并发连接。不会为一个文件创建多个连接。
- **读取间接文件。**集成服务读取间接文件时，会先读取文件列表，然后按顺序读取列表中的文件。如果会话中包含多个文件列表，则集成服务会并发读取文件列表，然后按顺序读取列表中的文件。

使用多个线程读取文件源

集成服务使用多个线程来读取源文件时，会创建与源的多个并发连接。集成服务可能会按顺序读取文件中的行，也可能不会。

可以为会话中的直接文件源或间接文件源配置多线程读取：

- **读取直接文件。**集成服务读取直接文件时，将创建多个读取器线程以并发读取该文件。可以将集成服务配置为读取一个或多个直接文件。例如，如果会话读取两个文件而您创建了五个分区，则集成服务可能会在两个分区之间分发一个文件，而在其他三个分区之间分发另一个文件。
- **读取间接文件。**集成服务读取间接文件时，将创建多个线程以并发读取文件列表。它还会创建多个线程以并发读取列表中的文件。集成服务可以使用多个线程来读取单个文件。

配置以实现文件分区

创建分区点并配置分区信息后，可以在“映射”选项卡的“转换”视图上配置源连接设置和文件属性。在“源”节点下单击要配置的源实例名称。单击文件源的源实例名称时，Workflow Manager 将在会话属性中显示连接和文件属性。

可以为每个源分区配置源文件名和目录。Workflow Manager 会为每个分区生成一个文件名和位置。

下表介绍了映射中文件源的文件属性设置：

属性	说明
输入类型	源输入的类型。您可以选择以下类型的源输入： <ul style="list-style-type: none"> - File。用于平面文件、COBOL 或 XML 源。 - 命令。用于由命令生成的源数据或文件列表。 不能使用命令生成 XML 源数据。
并发读取分区	多个分区从源文件中读取输入行的顺序。可以选择以下选项： <ul style="list-style-type: none"> - 优化吞吐量。集成服务不会保留输入行顺序。 - 保持相对输入行顺序。集成服务会为每个分区读取的行保留输入行顺序。 - 保持绝对输入行顺序。集成服务会为所有分区读取的所有行保留输入行顺序。
源文件目录	平面文件源的目录名称。默认情况下，集成服务在服务进程变量目录 \$PMSourceFileDir 中查找文件源。 如果您在“源文件名”字段中同时指定了目录和文件名，请清除此字段。集成服务在运行会话时会将此字段与“源文件名”字段相连。 您还可以使用 \$InputFileName 会话参数指定文件位置。
源文件名	平面文件源的文件名或文件名和路径。或者，也可以选择使用 \$InputFileName 会话参数作为文件名。 集成服务在运行会话时会将此字段与“源文件目录”字段相连。例如，如果您的“源文件目录”字段中的内容为“C:\data\”，然后在“源文件名”字段中输入“filename.dat”。当集成服务开始会话时，它会查找“C:\data\filename.dat”。 默认情况下，Workflow Manager 输入源定义中配置的文件名。
源文件类型	您可以选择以下源文件类型： <ul style="list-style-type: none"> - 直接。用于包含源数据的源文件。 - 间接。用于包含文件列表的源文件。当您选择“间接”时，集成服务将在运行会话时查找文件列表并读取其中所列的每个文件。
命令类型	命令生成的源数据的类型。您可以选择以下命令类型： <ul style="list-style-type: none"> - 生成数据的命令 - 适用于生成源数据输入行的命令。 - 生成文件列表的命令 - 适用于生成文件列表的命令。
命令	用于生成源文件数据的命令。
截断字符串空值	去除字符串值中的第一个空字符及其后面的所有字符。 请为字符串中包含空字符的带分隔符平面文件启用该选项。如果不启用此选项，PowerCenter 集成服务将为所有在字符串中包含空字符的行都生成一个行错误。 默认为已禁用。

将会话配置为使用单个线程

要将会话配置为使用单个线程读取文件，请在分区 2 至 n 中传递空数据。要传递空数据，请创建一个不包含任何数据的文件，如 empty.txt，并将其放置在源文件目录下。然后，将 empty.txt 用作源文件名。

注意：无法为使用命令生成源数据的已分区源配置单线程读取。

下表显示了当集成服务创建一个线程来读取 ProductsA.txt 时的源文件名和值。它将按顺序读取文件中的行。读取文件后，将数据传递到转换管道中的三个分区：

源文件名	值
分区 #1	ProductsA.txt
分区 #2	empty.txt
分区 #3	empty.txt

下表显示了当集成服务创建两个线程时的源文件名和值。它创建一个线程来读取 ProductsA.txt，同时创建另一个线程来读取 ProductsB.txt。它将并发读取文件，并按顺序读取文件中的行：

源文件名	值
分区 #1	ProductsA.txt
分区 #2	empty.txt
分区 #3	ProductsB.txt

如果使用 FTP 访问源文件，还可以为每个直接文件选择其他连接。

将会话配置为使用多个线程

要将会话配置为使用多个线程读取文件，请将分区 2 至 n 的源文件名留空。集成服务使用分区 2 至 n 来读取以前的分区文件或文件列表中的一部分。集成服务会忽略该分区的目录字段。

要将会话配置为使用多个线程读取命令，请为每个分区输入一个命令，或者将分区 2 至 n 的命令属性留空。如果为每个分区输入一个命令，则集成服务会创建一个线程来读取每个命令生成的数据。否则，集成服务将使用分区 2 至 n 来读取命令为第一个分区生成的部分数据。

下表显示了当集成服务创建三个线程来并发读取 ProductsA.txt 时的属性和值：

属性	值
分区 #1	ProductsA.txt
分区 #2	<blank>
分区 #3	<blank>

下表显示了当集成服务创建三个线程来并发读取 ProductsA.txt 和 ProductsB.txt 时的属性和值。其中两个线程读取 ProductsA.txt，一个线程读取 ProductsB.txt：

属性	值
分区 #1	ProductsA.txt
分区 #2	<blank>
分区 #3	ProductsB.txt

下表显示了当集成服务创建三个线程来并发读取命令中的管道数据时的属性和值：

属性	值
分区 #1	CommandA
分区 #2	<blank>
分区 #3	<blank>

下表显示了当集成服务创建三个线程来读取 CommandA 和 CommandB 中的管道数据时的属性和值。其中两个线程读取 CommandA 中的管道数据，一个线程读取 CommandB 中的管道数据：

属性	值
分区 #1	CommandA
分区 #2	<blank>
分区 #3	CommandB

配置并发读取分区

默认情况下，多个分区读取单个文件源时集成服务不保留行顺序。要在多个分区读取单个文件源时保留行顺序，请配置并发读取分区。可以配置以下选项：

- **优化吞吐量。**多个分区读取单个文件源时集成服务不保留行顺序。如果多个分区读取单个文件源的顺序不重要，则可以使用此选项。
- **保持相对输入行顺序。**保留每个分区读取的输入行的排序顺序。如果要保留每个分区读取的输入行的排序顺序，可以使用此选项。

下表显示了一个文件源的排序顺序示例（其中 10 行由两个分区读取）：

分区	读取的行数
分区 #1	1,3,5,8,9
分区 #2	2,4,6,7,10

- **保持绝对输入行顺序。**保留所有分区读取的全部输入行的排序顺序。如果要在每次会话运行时保留输入行的排序顺序，可以使用此选项。在具有被动转换的传递映射中，写入目标的行顺序将与输入行的顺序相同。

下表显示了一个文件源的排序顺序示例（其中 10 行由两个分区读取）：

分区	读取的行数
分区 #1	1,2,3,4,5
分区 #2	6,7,8,9,10

注意：默认情况下，集成服务在使用从上一检查点继续这一恢复策略配置的会话中使用“保持绝对输入行顺序”选项。

对关系目标进行分区

配置管道以将数据加载到关系目标时，集成服务将为目标实例上的每个分区创建与目标数据库的单独连接。它将每个分区的数据并发加载到目标数据库中。

在会话属性的“映射”选项卡上配置管道中目标的分区属性。对于关系目标，配置拒绝文件名和目录。集成服务为每个目标分区创建一个拒绝文件。

下表介绍了管道中关系目标的分区属性：

属性	说明
拒绝文件目录	目标拒绝文件的位置。默认为 \$PMBadFileDir。
拒绝文件名	拒绝文件的名称。默认为 <i>target name partition number.bad</i> 。也可以使用参数文件中定义的会话参数 \$BadFileName。

数据库兼容性

配置目标实例上具有多个分区的会话时，集成服务将为每个分区创建一个与目标的连接。如果配置某个会话中的多个目标分区，而该分区加载到的数据库或 ODBC 目标不支持与表的多个并发连接，则该会话将失败。

在将数据加载到 Informix 数据库的会话中创建多个目标分区时，必须使用行级锁定创建目标表。如果将具有多个分区的会话中的数据插入配置为页级锁定的 Informix 目标时，会话将失败并返回以下消息：

WRT_8206 Error: The target table has been created with page level locking. The session can only run with multi partitions when the target table is created with row level locking.

Sybase IQ 不允许与表的多个并发连接。如果在加载到 Sybase IQ 的会话中创建多个目标分区，则集成服务会将所有数据加载到一个分区中。

对文件目标进行分区

配置会话以写入文件目标时，可以将目标输出写入每个分区的单独文件，或写入包含目标输出的所有分区的合并文件。运行会话时，集成服务将并发写入单个输出文件或合并文件。还可以将单个分区或所有目标分区的数据发送到操作系统命令。

可以为每个目标分区配置连接设置和文件属性。在“映射”选项卡上的“转换”视图中配置这些设置。还可以将会话配置为使用已分区 FTP 文件目标。

配置连接设置

使用“映射”选项卡上“转换”视图中的“连接”设置可配置所有目标分区的连接类型。可以为每个分区选择不同的连接对象，但是这些对象必须属于同一类型。

对目标文件使用以下连接类型之一：

- **无。**将已分区目标文件写入本地计算机。
- **FTP。**将已分区目标文件传输到其他计算机。可以将这些文件传输到集成服务可连接到的任何计算机。
- **加载器。**使用可以从多个输出文件加载的外部加载器。如果管道将数据加载到关系目标，并且您在“映射”选项卡上的“写入器”设置中选择了文件写入器，则会出现此选项。如果选择的加载器无法从多个输出文件加载，集成服务将使会话失败。

- **消息队列。**将已分区目标文件传输到 WebSphere MQ 消息队列。

注意: 如果为所有目标分区选择本地或 FTP 连接类型，则可以合并目标文件。如果使用外部加载器或 WebSphere MQ 消息队列作为目标连接类型，将无法合并具有多个分区的会话的输出文件。

下表介绍了映射中文件目标的连接选项：

属性	说明
连接类型	选择 FTP、外部加载器或消息队列连接。为本地连接选择“无”。所有分区的连接类型均相同。
值	对于 FTP、外部加载器或消息队列连接，单击此字段中的“打开”按钮可选择连接对象。 可以为每个分区指定不同的连接对象。

配置文件属性

使用“映射”选项卡上“转换”视图中的“属性”设置可配置平面文件源的文件属性。

下表介绍了映射中文件目标的文件属性：

属性	说明
合并类型	集成服务对已分区目标的数据执行的合并类型。 合并目标文件后，集成服务将在会话运行时将所有分区的输出写入合并文件或命令。 如果会话使用外部加载器或消息队列，将无法合并文件。
合并文件目录	合并文件的位置。默认为 <code>\$PMTargetFileDir</code> 。
合并文件名	合并文件的名称。默认为 <code>target name.out</code> 。
若存在则追加	将输出数据追加到每个分区的目标文件和拒绝文件。如果您合并目标文件，将输出数据追加到合并文件。您不能为非磁盘文件的目标文件（如 FTP 目标文件）使用此选项。 如果您不选择此选项，集成服务将在向目标文件写入输出数据前截断每个目标文件。如果文件不存在，集成服务将创建该文件。
输出类型	会话目标的类型。选择“文件”将目标数据写入文件目标。选择“命令”可将目标数据发送到命令。无法为 FTP 或队列目标连接选择“命令”。
表头选项	在文件目标中创建表头行。
表头命令	用于在文件目标中生成表头行的命令。
表尾命令	用于在文件目标中生成表尾行的命令。
合并命令	用于处理已合并的目标数据的命令。
输出文件目录	目标文件的位置。默认为 <code>\$PMTargetFileDir</code> 。
输出文件名	目标文件的名称。默认为 <code>target name partition number.out</code> 。也可以使用参数文件中定义的会话参数 <code>\$OutputFileName</code> 。

属性	说明
拒绝文件目录	目标拒绝文件的位置。默认为 \$PMBadFileDir。
拒绝文件名	拒绝文件的名称。默认为 <i>target name partition number.bad</i> 。 也可以使用参数文件中定义的会话参数 \$BadFileName。
命令	用于处理单个分区的目标输出数据的命令。

配置已分区文件目标的命令

可使用命令来处理会话中单个分区的目标数据，也可使用命令来处理会话中所有目标分区的合并数据。在 UNIX 中，请使用任何有效的 UNIX 命令或 Shell 脚本。在 Windows 中，请使用任何有效的 DOS 或批处理文件。集成服务会将数据发送到命令，而不是发送到平面文件目标或合并文件。

可使用命令来处理下列类型的目标数据：

- 单个分区的目标数据。**可以为每个目标分区输入一个命令。集成服务在会话运行时将目标数据发送至命令。
要将单个分区的目标数据发送至命令，请选择“输出类型的命令”。在会话属性中为分区的“命令”属性输入一个命令。
- 合并所有目标分区的数据。**可以输入一个命令以处理所有分区的合并数据。集成服务在会话运行时将所有分区的目标数据并发送送至命令。该命令可能无法保持目标数据的顺序。
要将所有分区的合并数据发送至命令，请选择“命令作为输出类型”，并在会话属性中为“合并命令行”属性输入一个命令。

配置合并选项

您可以合并会话中分区的目标数据。合并目标数据时，集成服务将为所有目标分区创建一个合并文件。

可以配置以下合并文件选项：

- 连续合并。**集成服务会为所有分区创建一个输出文件，然后在会话结束时将这些文件合并为一个合并文件。集成服务按顺序将每个分区的输出数据添加到合并文件。集成服务使用分区的“输出文件名”和“输出文件目录”值创建单个目标文件。
- 文件列表。**集成服务会为所有分区创建一个目标文件，然后创建一个包含各个文件路径的文件列表。集成服务使用分区的“输出文件名”和“输出文件目录”值创建单个目标文件。如果将目标文件写入合并目录或合并目录的子目录，则文件列表将包含相对路径。否则，列表文件将包含绝对路径。如果将这些目标文件用作其他映射的源文件，则可将此文件用作源文件。
- 并发合并。**集成服务会将所有目标分区的数据并发写入合并文件。它不会为每个分区创建中间文件。因为集成服务会并发地为所有分区写入合并文件，所以合并文件中的数据顺序可能不是按序排列的。

对自定义转换进行分区

当映射包含自定义转换、Java 转换、SQL 转换或 HTTP 转换时，可以编辑以下分区信息：

- 添加多个分区。**当自定义转换允许多个分区时，可以创建多个分区。
- 创建分区点。**即使自定义转换不允许多个分区，也可以在该转换中创建分区点。

Java、SQL 和 HTTP 转换使用自定义转换构建，并具有相同的分区功能。不是所有使用自定义转换创建的转换都具有与自定义转换相同的分区功能。

将自定义转换配置为使用一个线程处理每个分区时，Workflow Manager 会根据映射配置添加分区点。

使用多个分区

可以对自定义转换进行配置以允许映射中有多个分区。如果为该转换设置了“是否可分区”属性，则可以向管道添加分区。可以为“是否可分区”选项选择以下值：

- **否**。不可对转换进行分区。转换和同一渠道中的其他转换只有一个分区。如果转换一起处理所有输入数据（例如数据清理），可选择“否”。
- **本地**。可以对转换进行分区，但集成服务必须在同一节点上运行管道中的所有分区。当转换的不同分区必须共享内存中的对象时，请选择“本地”。
- **在整个网格范围内**。可以对转换进行分区，并且集成服务可以将各个分区分发到不同的节点。

注意：向包含多个输入或输出组自定义转换的映射添加多个分区时，请为所有的组定义相同数量的分区。

创建分区点

即使自定义转换不允许多个分区，也可以在该转换中创建分区点。在自定义转换中创建分区点时，请遵循以下规则和准则：

- 可以为转换中的每个输入组定义分区类型。无法为输出组定义分区类型。
- 有效的分区类型包括传递、循环、键范围和哈希用户键。

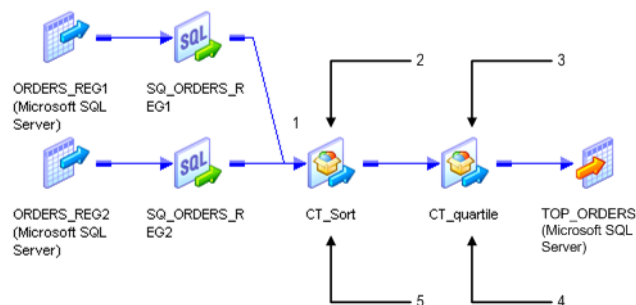
使用线程

要对自定义转换进行配置以便集成服务使用一个线程来为每个分区处理该转换，请启用“要求每个分区一个线程”自定义转换属性。Workflow Manager 根据输入组的数量和自定义转换在映射中的位置来创建传递分区点。

一个输入组

当单个输入组自定义转换是没有分区点的多个输入组自定义转换的下游时，Workflow Manager 会将传递分区点放置在最近的上游多个输入组转换。

例如，考虑以下映射：



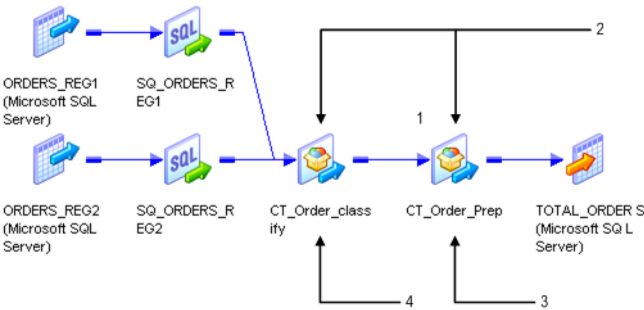
1. 分区点。
2. 多个输入组。
3. 单个输入组。
4. 每个分区需要一个线程。
5. 不是每个分区都需要一个线程。

CT_quartile 包含一个输入组，并且是多个输入组转换 CT_sort 的下游。CT_quartile 针对每个分区都需要一个线程，但上游自定义转换 CT_sort 不需要。Workflow Manager 会在最近的上游多个输入组转换 CT_Sort 创建分区点。

多个输入组

Workflow Manager 会在要求每个分区具有一个线程的多输入组自定义转换上放置一个分区点。

例如，考虑以下映射：



1. 分区点
2. 多个输入组。
3. 每个分区需要一个线程。
4. 不是每个分区都需要一个线程。

CT_Order_class 和 CT_Order_Prep 具有多个输入组，但只有 CT_Order_Prep 要求每个分区具有一个线程。Workflow Manager 会在 CT_Order_Prep 上创建一个分区点。

对联接器转换进行分区

在联接器转换中创建分区点时，如果转换范围为“全部输入”，则 Workflow Manager 会将分区类型设置为哈希自动键。如果转换范围为“事务”，则 Workflow Manager 会将分区类型设置为传递。

必须为主源和详细源创建相同数量的分区。如果为已排序输入配置联接器转换，可以将分区类型更改为传递。如果管道包含联接器转换的主源并且未在该联接器转换中添加分区点，则只能指定一个分区。

在联接器转换中创建了分区点时，集成服务将使用缓存分区。对联接器转换使用分区时，可以为联接器转换的主源和详细源创建多个分区。

如果没有在联接器转换中创建分区点，则可以为详细源创建 n 个分区，而为主源创建一个分区 ($1:n$)。

注意: 如果将联接器转换配置为使用行转换范围，将无法在联接器转换中添加分区点。

对已排序联接器转换进行分区

包含使用已排序输入的联接器转换时，必须验证该联接器转换是否接收已排序数据。如果源包含大量数量，您可能希望配置分区以提高性能。但是，重新分发的分区可以重新排列已排序数据的顺序，因此，将分区配置为维护已排序数据很重要。

例如，使用哈希自动键分区点时，集成服务将使用哈希函数来确定在分区中分发数据的最佳方式。但是，集成服务不会维护排序顺序，因此使用此类型的分区点时必须遵循特定的分区准则。

联接数据时，可以通过为主源和详细源配置相同数量的分区来对主管道和详细信息管道的数据进行分区。集成服务可并发处理多个分区。

您可能需要根据在联接器转换中使用的分区类型来对分区进行配置以维护排序顺序。如果联接器转换使用 1:n 分区，并且主管道和详细信息管道均在已排序端口上进行联接，则会话将意外终止。

请考虑以下分区准则：

- **使用已排序平面文件或已排序关系数据。** 当主管道和详细信息管道中存在一个大型平面文件时，请将分区配置为在第一个分区中传递所有已排序数据，而在其他分区中传递空文件数据。
- **使用排序器转换。** 如果在联接器转换中使用哈希自动键分区，请也将每个排序器转换配置为使用自动键分区点。

请仅在排序来源和联接器转换之间添加传递分区点。

使用已排序平面文件

当主管道中有一个平面文件，而详细信息管道中有多个平面文件时，请使用 1:n 分区。使用 1:n 分区时，集成服务会维护排序顺序，因为它不会在各分区中重新分发数据。当每个主管道和详细信息管道中有一个大型平面文件时，请使用 n:n 分区并在联接器转换中添加传递或哈希自动键分区。添加哈希自动键分区点时，必须将分区配置为在第一个分区中传递所有已排序数据以维护排序顺序。

使用 1:n 分区

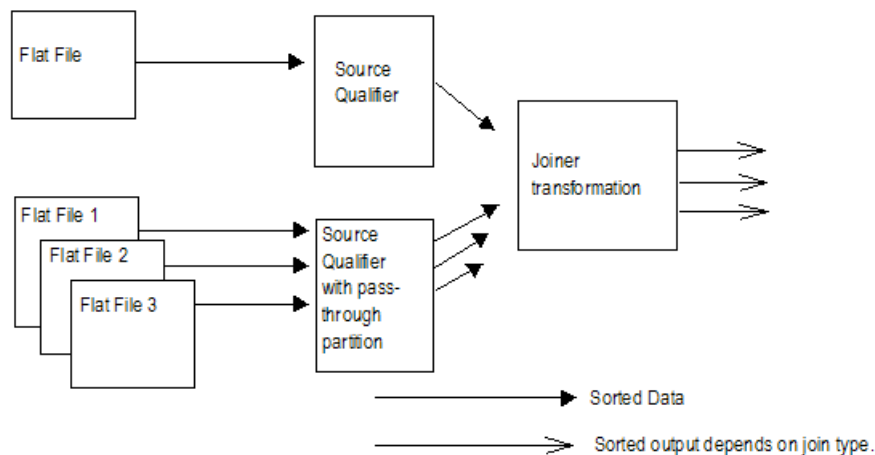
如果会话在主管道中使用一个平面文件，而在详细信息管道中使用多个平面文件，请为主源使用一个分区，而为详细文件源使用 n 个分区 (1:n)。在详细源限定符转换中添加传递分区点。请不要在联接器转换中添加分区点。为主源创建一个分区时，集成服务将维护排序顺序，因为它不会在各分区中重新分发已排序数据。

当详细信息管道中有多个结构相同的文件时，请按照以下准则将文件传递给联接器转换：

- 在每个管道中为该映射配置一个源和一个源限定符转换。
- 在会话属性的“映射”选项卡上“转换”视图的“属性”设置中指定各个平面文件的路径和文件名。
- 每个文件使用的文件属性必须与源定义中的配置相同。
- 平面文件中已排序数据的范围可以重叠。无需为每个文件使用唯一的数据范围。

在使用 1:n 分区对文件数据进行排序时，取决于联接类型，联接器转换可能会输出未排序的数据。如果使用完整外部联接或详细外部联接，则集成服务会在最后处理不匹配的主行，而这会导致数据未排序。

下图显示了使用 1:n 分区联接的已排序文件数据：



使用 n:n 分区

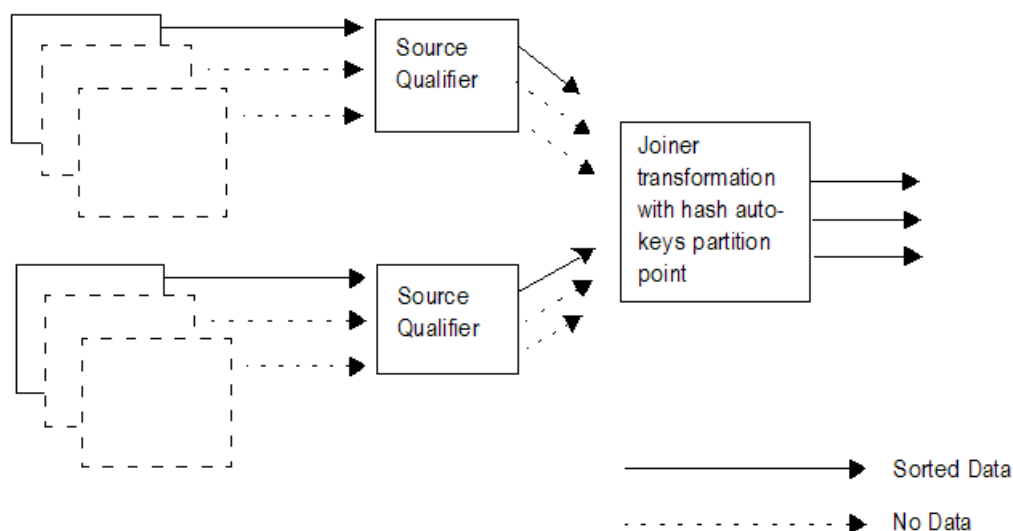
如果会话使用已排序平面文件数据，请为主管道和详细信息管道使用 $n:n$ 分区。您可以在联接器转换中添加传递分区或哈希自动键分区。

如果在联接器转换中添加传递分区，请保持映射中的排序顺序。如果在联接器转换中添加哈希自动键分区点，可以通过将所有已排序数据传递到一个分区中的联接器转换来保持排序顺序。当您在分区中传递已排序数据时，集成服务会在使用哈希函数重新分发数据时保持排序顺序。

要允许集成服务在一个分区中传递所有已排序数据，请将会话配置为为第一个分区使用已排序文件，而为其余分区使用空文件。

集成服务会在多个分区中重新分发并联接已排序数据。

下图显示了通过一个分区传递以保持排序顺序的已排序文件数据：



使用已排序关系数据

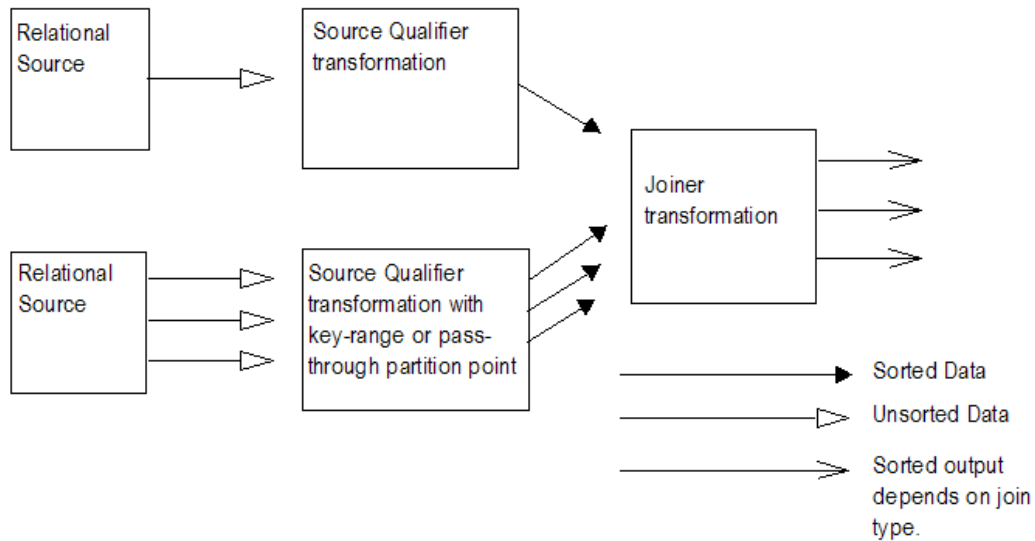
联接关系数据时，请为主管道和详细信息管道使用 $1:n$ 分区。使用 $1:n$ 分区时，将无法在联接器转换中添加分区点。如果使用 $n:n$ 分区，可以在联接器转换中添加传递或哈希自动键分区。如果使用哈希自动键分区点，则必须将分区配置为在第一个分区中传递所有已排序数据以维护排序顺序。

使用 1:n 分区

如果会话使用已排序关系数据，请为主源使用一个分区，而为详细源使用 n 个分区 ($1:n$)。在源限定符转换中添加键范围或传递分区点。请不要在联接器转换中添加分区点。为主源创建一个分区时，集成服务将维护排序顺序，因为它不会在各分区中重新分发数据。

在使用 $1:n$ 分区对关系数据进行排序时，取决于联接类型，联接器转换可能会输出未排序的数据。如果使用完整外部联接或详细外部联接，则集成服务会在最后处理不匹配的主行，而这会导致数据未排序。

下图显示了使用 1:n 分区的已排序关系数据：

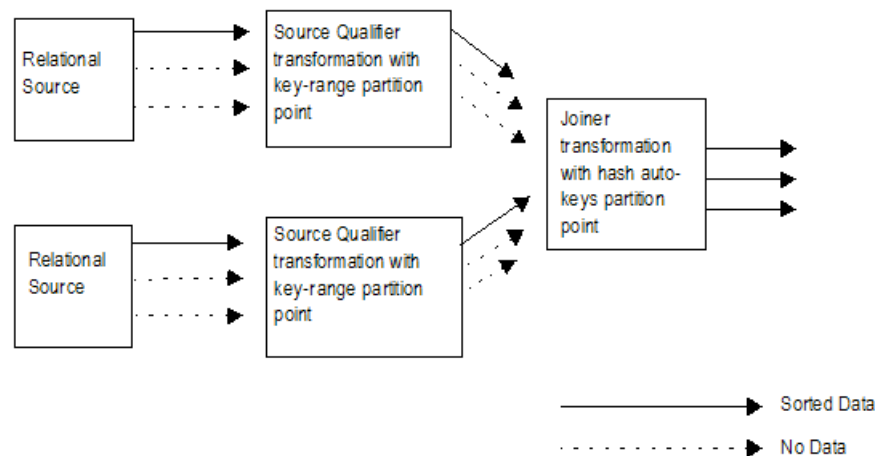


使用 n:n 分区

如果会话使用已排序关系数据，请为主管道和详细信息管道使用 $n:n$ 分区，并在联接器转换中添加传递或哈希自动键分区点。

在联接器转换中使用传递分区时，请保留映射中的已排序数据。使用哈希自动键分区点时，可通过将所有已排序数据传递到单个分区中的联接器转换来保持排序顺序。在第一个分区中包含所有源数据的源限定符转换中添加键范围分区点。在一个分区中传递已排序数据时，集成服务会使用哈希函数在多个分区间重新分发数据，并联接已排序数据。

下图显示了通过单个分区传递以保持排序顺序的已排序关系数据：

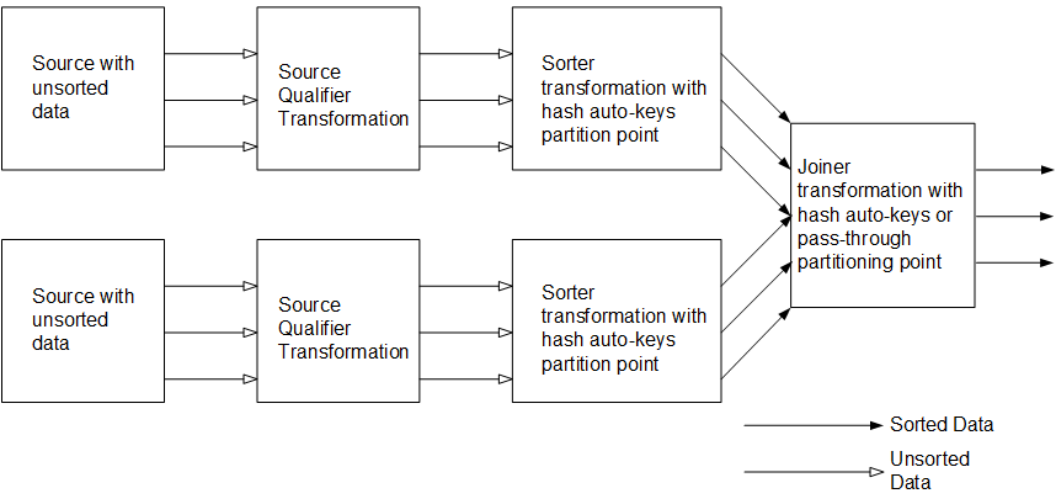


使用排序器转换

如果会话使用排序器转换对数据进行排序，请为主管道和详细信息管道使用 $n:n$ 分区。在排序器转换中使用哈希自动键分区点可对数据进行分组。可以在联接器转换中添加传递分区点或哈希自动键分区点。

集成服务可将数据分组到哈希值相同的分区中，并且排序器转换会先对数据进行排序，然后将其传递到联接器转换。集成服务在处理已配置了哈希自动键分区的联接器转换时，会使用与从每个排序器转换路由数据所用分区相同的分区来处理已排序数据，通过这种方法保持排序顺序。

下图显示了将排序器转换与哈希自动键分区一起使用以保持排序顺序：



注意: 为实现最佳性能，请使用已排序平面文件或已排序关系数据。您可能想要计算向映射添加排序器转换所需的处理开销。

使用分区优化已排序联接器转换

将分区与已排序联接器转换一起使用时，可以通过将数据分组并使用 $n:n$ 分区来优化性能。

添加排序来源的哈希自动键分区上游

要在对已排序联接器转换进行分区时获得预期的结果并实现最佳性能，必须对数据进行分组和排序。要对数据进行分组，确保将具有相同键值的行路由到同一分区。确保在分区之间均匀进行数据分组和分布的最好方法是在排序源之前添加一个哈希自动键或键范围分区点。在对数据进行排序之前放置分区点可确保维护分组并对各个组中的数据进行排序。

使用 $n:n$ 分区

使用 $n:n$ 分区能够提高已排序联接器转换的性能。使用 $n:n$ 分区时，联接器转换会并发读取主行和详细信息行，无需缓存所有主数据。这可减少内存使用量，提高处理速度。使用 $1:n$ 分区时，联接器转换将缓存主管道中的所有数据，并在内存缓存已满时将其写入磁盘。联接器转换从详细信息管道中接收数据后，必须从磁盘中读取数据以对主管道和详细信息管道进行比较。

对查找转换进行分区

可以为查找转换配置缓存分区。可以为静态查找缓存和动态查找缓存创建多个分区。

管道查找转换的缓存内置在包含该查找转换的管道中的独立管道中。可以在这两个管道中创建多个分区。

缓存分区查找转换

可为静态缓存和动态缓存，以及已命名缓存和未命名缓存使用缓存分区。在已连接的查找转换中创建分区点时，请在以下情况下使用缓存分区：

- 为查找转换使用哈希自动键分区类型。
- 查找条件必须仅包含等式运算符。
- 数据库配置为执行区分大小写的比较。

例如，如果查找条件包含字符串端口而数据库未配置为执行区分大小写的比较，则集成服务将不会执行缓存分区，并且向会话日志写入以下消息：

```
CMN_1799 Cache partitioning requires case sensitive string comparisons. Lookup will not use partitioned cache as the database is configured for case insensitive string comparisons.
```

在查找转换中创建哈希自动键分区点时，集成服务将使用缓存分区。

如果集成服务创建缓存分区，则当任意分区的第一行达到查找转换时，集成服务将开始为查找转换创建缓存。如果为并发缓存配置查找转换，则集成服务会并发构建分区的所有缓存。

共享已分区缓存

共享已分区查找缓存时，请遵循以下准则：

- 查找转换可在满足以下条件时共享已分区缓存：
 - 缓存结构相同。第一个共享转换的查找/输出端口必须与后续转换的查找/输出端口相匹配。
 - 转换的查找条件相同，并且查找条件列的顺序相同。
- 不能将已分区缓存与未分区缓存共享。
- 在目标加载顺序组中共享查找缓存时，必须配置具有相同数量的分区的目标加载顺序组。
- 如果集成服务检测到共享未命名缓存的查找转换之间存在不匹配，则会重建缓存文件。
- 如果集成服务检测到共享已命名缓存的查找转换之间存在不匹配，则会使会话失败。

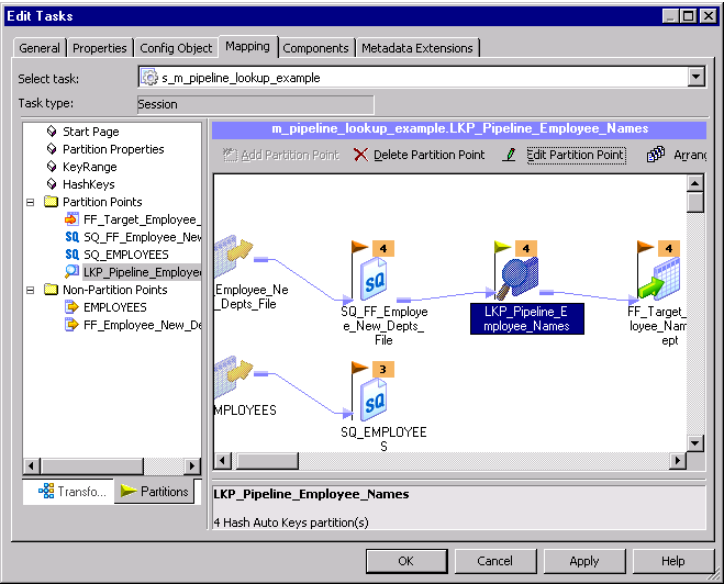
对管道查找转换缓存进行分区

默认情况下，将为管道查找转换启用缓存。在集成服务构建查找缓存时，可以对查找源进行分区以提高性能。缓存查找源时，查找转换将开始处理行。

配置管道查找转换时，查找源和源限定符位于查找转换的不同管道中。由于该管道不包含任何目标，因此为部分管道。集成服务可读取部分管道中的源数据。可以在该管道中创建多个分区以提高处理性能。

集成服务构建缓存时，将部分管道中的源数据传递到其他管道。当部分管道中的分区数量与查找转换的分区数量不同时，集成服务将创建一个分区点。如果查找转换包含哈希自动键分区点，则集成服务会在缓存中创建与查找转换中数量相同的分区。否则，缓存将仅有一个分区。

下图显示了包含一个管道查找转换和一个源限定符查找源的会话的分区：



集成服务在三个分区中处理“员工”行。包含查找转换的管道具有四个分区。由于查找转换包含一个哈希自动键分区点，因此缓存将分区为四个分区。

对序列生成器转换进行分区

如果在网格上配置使用未缓存序列生成器转换的会话中的多个分区，则集成服务为每个分区生成的序列号将不是连续的。

对排序器转换进行分区

如果配置使用排序器转换的会话中的多个分区，则集成服务将在每个分区中单独对数据进行排序。在排序器转换中添加分区点时，Workflow Manager 将支持您选择哈希自动键、键范围或传递分区。

在配置为使用已排序输入的汇总器转换之前放置排序器转换时，请使用哈希自动键分区。哈希自动键分区根据分区键将具有相同值的行分组到同一分区。将行分组后，集成服务将行传递给排序器转换。集成服务在每个分区中单独处理数据，但哈希键分区会对所有的源数据进行准确排序，因为具有匹配值的行将在同一分区中进行处理。您可以在汇总器转换删除默认分区点。

要将已分区会话中的所有行从多个分区发送到单个分区以进行排序时，可使用键范围分区。将所有行合并到单个分区以进行排序时，集成服务可以一起处理所有的数据。

如果已在管道中使用哈希分区，请使用传递分区。这可确保传递到排序器转换中的数据在各分区中正确分组。传递分区可在不增加管道中分区数量的情况下提高会话性能。

配置排序器转换工作目录

集成服务可为管道中的每个排序器转换创建临时文件。它在执行排序操作时向这些文件读取和写入数据。集成服务将这些文件存储在排序器转换工作目录中。

默认情况下，Workflow Manager 将排序器转换中所有分区的工作目录设置为 \$PMTempDir。您可以在会话属性中为每个分区指定不同的工作目录。

对 XML 生成器转换进行分区

在多个分区中生成 XML 时，将始终为各个分区生成单独的文档。无论“提交后标志”中的值为何，都会出现此种情况。如果使用 XML 生成器转换配置键范围分区，则会话可能会失败，并且转换中会出现孤行。出现这种情况是因为 XML 生成器转换在行之间创建了主键-外键关系。键范围分区可以将父行与子行分隔开来。

转换的限制

部分对分区数量的限制取决于管道中转换的类型。这些限制适用于所有转换，包括可重用转换、在映射和 Mapplet 中创建的转换，以及快捷方式所引用的转换、Mapplet 和映射。

下表介绍了对转换的分区数量的限制：

转换	限制
自定义转换	默认情况下，如果管道中包含自定义转换，将只能指定一个分区。 但是，此转换在“属性”选项卡上包含一个选项，可允许指定多个分区。如果启用此选项，则可在此转换中指定多个分区。如果自定义转换过程根据 <i>所有</i> 输入数据（如数据清理）来执行该过程，请不要选择“是否可分区”。
外部过程转换	默认情况下，如果管道中包含外部过程转换，将只能指定一个分区。 此转换在“属性”选项卡上包含一个选项，可允许指定多个分区。如果启用此选项，则可在此转换中指定多个分区。
联接器转换	如果管道包含联接器转换的主源并且未在该联接器转换中添加分区点，则只能指定一个分区。
XML 目标实例	如果管道中包含 XML 目标，则只能指定一个分区。

规范器和序列生成器转换所生成的序列号对于已分区源而言可能不是连续的，但却是唯一的。

数值函数的限制

数值函数 CUME、MOVINGSUM 和 MOVINGAVG 会逐行计算累加值和平均值。根据对管道进行分区的方式，数据行在包含这些函数之一的转换中传递的顺序可能会发生更改。因此，具有多个分区且使用 CUME、MOVINGSUM 或 MOVINGAVG 函数的会话可能不会始终返回相同的计算结果。

第 3 章

分区类型

本章包括以下主题：

- [分区类型概览, 50](#)
- [设置分区类型, 51](#)
- [数据库分区的分区类型, 53](#)
- [哈希自动键分区类型, 56](#)
- [哈希用户键分区类型, 56](#)
- [键范围分区类型, 57](#)
- [传递分区类型, 60](#)
- [循环分区类型, 61](#)

分区类型概览

PowerCenter 集成服务会在每个分区点上创建默认的分区类型。如果拥有“分区”选项，可以更改分区类型。分区类型可控制 PowerCenter 集成服务在分区点的分区之间分发数据的方式。

配置管道的分区信息时，必须在管道中的每个分区点定义分区类型。分区类型可确定 PowerCenter 集成服务在分区点之间重新分发数据的方式。

可以在 Workflow Manager 中定义以下分区类型：

- **数据库分区。**PowerCenter 集成服务会在 IBM DB2 或 Oracle 系统中查询表分区信息。它会从数据库中对应的节点读取已分区数据。对多节点表空间上的 Oracle 或 IBM DB2 源实例使用数据库分区。对 DB2 目标使用数据库分区。
- **哈希分区。**当您希望 PowerCenter 集成服务按组将行分发到分区中时，请使用哈希分区。例如，您需要按项目 ID 对项目进行排序，但是不知道多少项目具有特定的 ID 号。

可以使用以下类型的哈希分区：

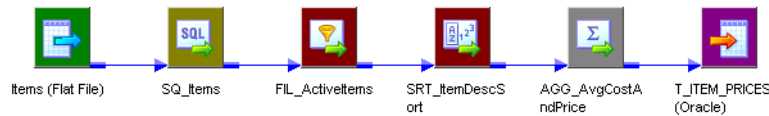
- **哈希自动键。**PowerCenter 集成服务使用所有已分组或已排序的端口作为复合分区键。在等级、排序器和未排序转换时可能需要使用哈希自动键分区。
- **哈希用户键。**PowerCenter 集成服务使用哈希函数来对各分区中的数据行进行分组。定义要生成分区键的端口数。
- **键范围。**可以指定一个或多个端口以形成复合分区键。PowerCenter 集成服务根据为每个端口指定的范围将数据传递到每个分区。当管道中的源或目标按键范围分区时，请使用键范围分区。
- **传递。**PowerCenter 集成服务会将某个分区点上的所有行传递给下一个分区点而不进行重新分发。如果您要创建一个额外的管道阶段以提高性能，但不希望更改数据在各分区之间的分发，请选择传递分区。

- **循环。**PowerCenter 集成服务会将数据块分发给一个或多个分区。使用循环分区以便每个分区均可基于块大小和数量处理行。

在管道中设置分区类型

可以在管道中的不同点创建不同分区类型。

下图显示了可以在其中创建分区类型以提高会话性能的映射：



该映射读取有关项目的数据并计算平均批发成本和价格。该映射必须从三个不同大小的平面文件中读取项目信息，然后筛选出已停止使用的项目。它按说明对活动项目进行排序，计算平均价格和批发成本，并将结果写入其中目标表按键范围进行分区的关系数据库中。

您可以删除汇总器转换上默认的分区分点，因为排序器转换上的哈希自动键分区分区会将包含具有相同说明的项目的所有行发送至同一分区。因此，汇总器转换会接收某个分区中具有相同说明的所有项目的数据，且可以正确计算此项目的平均成本和价格。

在会话中使用该映射时，可以通过在管道中的以下分区分点上定义不同的分区类型来提高会话性能：

- **源限定符。**要同时从三个平面文件中读取数据，必须在源限定符上指定三个分区。接受默认的分区分类型，传递。
- **筛选器转换。**由于源文件大小各不相同，每个分区处理不同的数据量。在筛选器转换设置分区分点，并选择循环分区以平衡进入过滤器转换的加载。
- **排序器转换。**要消除排序器和汇总器转换中的重叠组，请在排序器转换使用哈希自动键分区分区。这会导致在排序器和汇总器转换处理行之前，集成服务将具有同一说明的所有项分组到同一分区。您可以在汇总器转换删除默认分区分点。
- **目标。**由于目标表根据键范围分区分区，在目标指定键范围分区分区来优化将数据写入目标的性能。

设置分区分类型

Workflow Manager 可为管道中的每个分区分点设置一个默认的分区分类型。Workflow Manager 将传递指定为所有分区分点的默认分区分类型，除非某个转换的转换范围为“全部输入”。您可以更改默认类型。

例如，在源限定符和目标实例上，Workflow Manager 指定了传递分区。对于等级和未排序的汇总器转换，当转换范围为“全部输入”时，Workflow Manager 会指定哈希自动键分区分区。

您必须为位于生成提交的事务生成器或活动源下游以及位于具有“事务”转换范围的目标或转换上游的所有转换指定传递分区。此外，如果将会话配置为使用基于约束的加载，则必须为位于上一个活动源下游的所有转换指定传递分区。

如果已启用工作流恢复，Workflow Manager 会将分区分类型设置为传递，除非分区分点是汇总器转换或等级转换。

无法为以下转换创建分区分点：

- 源定义
- 序列生成器
- XML 解析器

- XML 目标
- 未连接的转换

下表列出了管道中不同分区点的有效分区类型和默认分区类型：

转换 (分区点)	循环	哈希自 动键	哈希用 户键	键范围	传递	数据库 分区
源限定符 (关系源)	否	否	否	是	是	是 (Oracle、 DB2)
源限定符 (平面文件源)	否	否	否	否	是	否
Web 服务源限定符	否	否	否	否	是	否
XML 源限定符	否	否	否	否	是	否
规范器 (COBOL 源)	否	否	否	否	是	否
规范器 (关系)	是	否	是	是	是	否
汇总器 (已排序)	否	否	否	否	是	否
汇总器 (未排序)	否	是	否	否	是	否
自定义	是	否	是	是	是	否
数据屏蔽	是	否	是	是	是	否
表达式	是	否	是	是	是	否
外部过程	是	否	是	是	是	否
筛选器	是	否	是	是	是	否
HTTP	否	否	否	否	是	否
Java	是	否	是	是	是	否
联接器	否	是	否	否	是	否
查找	是	是	是	是	是	否
等级	否	是	否	否	是	否
路由器	是	否	是	是	是	否
排序器	否	是	否	是	是	否
存储过程	是	否	是	是	是	否

转换 (分区点)	循环	哈希自动键	哈希用户键	键范围	传递	数据库分区
事务控制	是	否	是	是	是	否
联合	是	否	是	是	是	否
非结构化数据	是	否	是	是	是	否
更新策略	是	否	是	是	是	否
Web 服务使用者	否	否	否	否	是	否
XML 生成器	否	否	否	否	是	否
XML 解析器	否	否	否	否	是	否
关系目标定义	是	否	是	是	是	是 (DB2)
平面文件目标定义	是	否	是	是	是	否
Web 服务目标	否	否	否	否	是	否

对于以下转换，当转换范围为“事务”时，默认的分区类型为传递；当转换范围为“全部输入”时，默认的分区类型为哈希自动键：

- 汇总器（未排序）
- 联接器
- 等级
- 排序器

数据库分区的分区类型

可以通过对源和目标数据库使用数据库分区类型来优化会话性能。当使用源数据库分区时，集成服务查询数据库系统的表分区信息，并将数据提取到会话分区。当使用目标数据库分区时，集成服务将数据加载到相应的数据库分区节点。

对 Oracle 和 IBM DB2 源及 IBM DB2 目标使用数据库分区。使用任意数量的管道分区和任意数量的数据库分区。但是，当管道分区数等于数据库分区数时，可以提高性能。

数据库分区可以提高使用范围分区的 IBM DB2 源和目标的性能。

对于使用复合分区的 Oracle 源，当管道分区数等于数据库子分区数时，可以提高性能。例如，如果某 Oracle 源包含三个分区且每个分区有两个子分区，则将源上的管道分区数设置为 6。

对数据库源进行分区

使用源数据库分区时，集成服务会在数据库系统目录中查询分区信息。它会在会话分区之间分发数据库分区中的数据。

如果会话具有比数据库更多的分区，则集成服务会为每个数据库分区生成 SQL，并将数据重新分发给下一个分区点上的会话分区。

具有一个源的数据库分区

将数据库分区与具有一个源的源限定符结合使用时，集成服务会为每个数据库分区生成 SQL 查询，并在会话分区之间均匀分发数据库分区中的数据。

例如，当某会话具有三个分区且数据库具有五个分区时，集成服务会针对数据库分区在会话分区中执行 SQL 查询。第一个和第二个会话分区从两个数据库分区中接收数据。第三个会话分区从一个数据库分区中接收数据。

使用 Oracle 数据库时，集成服务会为分区 1 生成类似于下列语句的 SQL 语句：

```
SELECT <column list> FROM <table name> PARTITION <database_partition1 name> UNION ALL
SELECT <column list> FROM <table name> PARTITION <database_partition4 name> UNION ALL
```

使用 IBM DB2 数据库时，集成服务会为分区 1 创建类似于下列语句的 SQL 语句：

```
SELECT <column list> FROM <table name>
WHERE (nodenumber(<column 1>)=0 OR nodenumber(<column 1>) = 3)
```

如果 Oracle 源具有五个分区 (1-5)，每个分区中有两个子分区 (*a* 和 *b*)，且会话具有三个分区，则集成服务会针对数据库子分区在会话分区中执行 SQL 查询。第一个和第二个会话分区从四个数据库子分区中接收数据。第三个会话分区从两个数据库子分区中接收数据。

集成服务会为分区 1 生成类似于下列语句的 SQL 语句：

```
SELECT <column list> FROM <table name> SUBPARTITION <database_subpartition1_a name> UNION ALL
SELECT <column list> FROM <table name> SUBPARTITION <database_subpartition1_b name> UNION ALL
SELECT <column list> FROM <table name> SUBPARTITION <database_subpartition4_a name> UNION ALL
SELECT <column list> FROM <table name> SUBPARTITION <database_subpartition4_b name> UNION ALL
```

对具有多个源的源限定符进行分区

关系源限定符可以从多个源表接收数据。集成服务根据具有大多数分区的数据表中的分区数为数据库分区创建 SQL 查询。它会创建一个 SQL 联接条件，以从数据库分区中接收数据。

例如，源限定符从两个源表中接收数据。每个源表具有两个分区。如果会话具有三个分区且数据库表具有两个分区，则某个会话分区不会接收任何数据。

集成服务会为 Oracle 生成以下 SQL 语句：

```
Session Partition 1:
SELECT <column list> FROM t1 PARTITION (p1), t2 WHERE <join clause>
Session Partition 2:
SELECT <column list> FROM t1 PARTITION (p2), t2 WHERE <join clause>
Session Partition 3:
No SQL query.
```

集成服务会为 IBM DB2 生成以下 SQL 语句：

```
Session Partition 1:
SELECT <column list> FROM t1,t2 WHERE ((nodenumber(t1 column1)=0) AND <join clause>)
Session Partition 2:
SELECT <column list> FROM t1,t2 WHERE ((nodenumber(t1 column1)=1) AND <join clause>)
Session Partition 3:
No SQL query.
```

集成服务对源数据库分区的处理

集成服务对数据库分区使用以下规则：

- 如果为除 Oracle 或 IBM DB2 之外的数据库指定数据库分区，则集成服务会读取单个分区中的数据并向会话日志中写入消息。
- 如果会话分区数大于数据库中表的分区数，则超出的分区不会接收任何数据。会话日志描述了哪些分区不会接收数据。
- 如果会话分区数小于数据库中表的分区数，则集成服务会将数据均匀分发给会话分区。某些会话分区从多个数据库分区中接收数据。
- 将数据库分区与动态分区结合使用时，集成服务会在会话开始时确定会话分区的数量。
- 会话在分区方面的性能取决于数据库分区中的数据分发。集成服务会对数据库分区生成 SQL 查询。SQL 查询执行联合或联接命令，这会生成大量具有性能影响的查询语句。

源数据库分区的规则和准则

对关系源使用数据库分区类型时，请遵循以下规则和准则：

- 将会话配置为使用基于源或用户定义的提交、基于约束的加载或工作流恢复时，无法使用数据库分区。
- 将源限定符配置为执行数据库分区时，集成服务会在以下情况下还原为传递分区：
 - 数据库表存储在数据库分区上。
 - 在调试模式下运行会话。
 - 为具有一个分区的会话指定数据库分区。
 - 使用下推优化。下推优化使用其他分区类型。
- 创建 SQL 替代以读取数据库表以及设置数据库分区时，集成服务将还原为传递分区并向会话日志中写入消息。
- 如果创建用户定义的联接，则集成服务会将此联接添加到其为每个分区生成的 SQL 语句中。
- 如果创建源筛选器，则集成服务会将其添加到每个分区的 SQL 查询的 WHERE 子句中。

目标数据库分区

仅可对 IBM DB2 数据库使用目标数据库分区。将数据加载到多节点表空间上存储的 IBM DB2 表中时，可以通过使用数据库分区类型来优化会话性能。使用数据库分区时，集成服务会在 DB2 系统中查询表分区信息，并将已分区的数据加载到目标数据库中的对应节点。

默认情况下，当对非 DB2 目标使用数据库分区时，集成服务会使会话失败。但是，当对非 DB2 关系目标使用数据库分区时，可以将集成服务配置为默认为传递分区。在 Administrator 工具中将集成服务属性 TreatDBPartitionAsPassThrough 设置为“是”。

可以使用任意数量的管道分区和任意数量的数据库节点为目标分区类型指定数据库分区。但是，当管道分区数等于数据库节点数时，可以进一步提高加载性能。

目标数据库分区的规则和准则

对数据库目标使用数据库分区时，请遵循以下规则和准则：

- 将会话配置为使用基于源或用户定义的提交、基于约束的加载或会话恢复时，无法使用数据库分区。
- 当目标表按范围进行分区时，无法使用数据库分区。如果目标表按范围进行分区，请使用传递或键范围分区。
- 目标表必须包含分区键，且您必须将目标实例中的所有非空分区键列链接到映射中的转换。

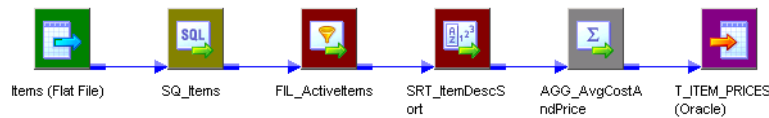
- IBM DB2 目标表分区键为小数列时，请为会话启用高精度。如果分区键为小数列而您没有为会话启用高精度，则集成服务可能会使会话失败。
- 如果为 DB2 批量加载会话创建多个分区，请对目标分区类型使用数据库分区。如果选择任何其他分区类型，集成服务将还原为正常加载并向会话日志中写入以下消息：

```
ODL_26097 Only database partitioning is support for DB2 bulk load. Changing target load type variable to Normal.
```
- 如果将会话配置为执行数据库分区，则集成服务会在以下情况下还原为传递分区：
 - DB2 目标表存储在一个节点上。
 - 使用调试器在调试模式下运行会话。
 - 将集成服务配置为将数据库分区分区类型视为传递分区，且对非 DB2 关系目标使用数据库分区。

哈希自动键分区类型

在等级、排序器、联接器和未排序的汇总器转换上或之前使用哈希自动键分区，以确保行在进入这些转换之前已正确分组。

下图显示了具有哈希自动键分区的映射。在行进入排序器和汇总器转换之前，集成服务会根据组将行分发给每个分区：

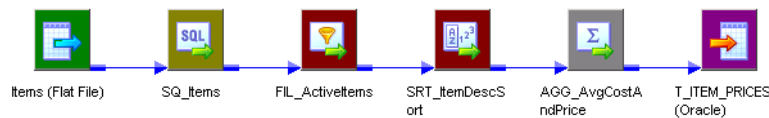


在该映射中，排序器转换按项目说明对项目进行排序。如果具有相同说明的项目存在于多个源文件中，则每个分区将包含具有相同说明的项目。如果没有哈希自动键分区，则汇总器转换可能会错误计算每个项目的平均成本和价格。

要防止成本和价格计算出现错误，请在排序器转换上设置分区点并将分区类型设置为哈希自动键。执行此操作时，集成服务会重新分发数据，以便所有具有相同说明的项目都到达单个分区中的排序器和汇总器转换。

哈希用户键分区类型

在哈希用户键分区中，集成服务使用哈希函数根据用户定义的分键对各分区中的数据行进行分组。您可以选择用于定义分区键的端口：



在先前映射中指定哈希自动键分区时，排序器转换会收到按排序键（如 ITEM_DESC）分组的数据行。如果项目说明较长，且您知道每个项目都具有唯一的 ID 号，则可以在排序器转换上指定哈希用户键分区并选择 ITEM_ID 作为哈希键。这可能会提高会话的性能，因为哈希函数处理数值数据的速度通常比处理字符串数据的速度快。

如果在任何分区点上选择哈希用户键分区，则必须指定哈希键。集成服务会使用哈希键根据组将行分发给相应的分区。

例如，如果在源限定符转换上指定键范围分区，当集成服务从源中选择数据时，会使用键和范围来创建 WHERE 子句。因此，您可以使集成服务将客户 ID 小于 135000 的所有行传递给一个分区，并将客户 ID 大于或等于 135000 的所有行传递给另一个分区。

如果在转换上指定哈希用户键分区，则集成服务将使用键根据选作键的端口对数据进行分组。例如，如果指定 ITEM_DESC 作为哈希键，则集成服务会分发数据，以便包含具有相同说明的项目的所有行都进入同一分区。

要指定哈希键，请在“映射”选项卡的“分区”视图上选择分区点，然后单击“编辑键”。这将显示“编辑分区键”对话框。“可用端口”列表显示了转换中的已连接输入及输入/输出端口。要指定哈希键，请从此列表中选择一个或多个端口，然后单击“添加”。

要重新排列用于定义键的端口的顺序，请在“选定端口”中选择端口，然后单击向上或向下箭头。

键范围分区类型

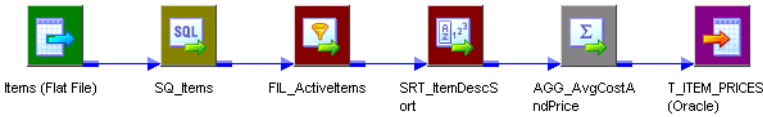
使用键范围分区，集成服务可根据定义为分区键的端口或端口集来分发数据行。对于每个端口，都可以定义值范围。集成服务使用键和范围将行发送至相应分区。

例如，如果在源限定符转换上指定键范围分区，当集成服务从源中选择数据时，会使用键和范围来创建 WHERE 子句。因此，您可以使集成服务将客户 ID 小于 135000 的所有行传递给一个分区，并将客户 ID 大于或等于 135000 的所有行传递给另一个分区。

如果在转换上指定哈希用户键分区，则集成服务将使用键根据选作键的端口对数据进行分组。例如，如果指定 ITEM_DESC 作为哈希键，则集成服务会分发数据，以便包含具有相同说明的项目的所有行都进入同一分区。

在其中源和目标表按键范围分区的映射中使用键范围分区。

下图显示了其中键范围分区可优化到目标表的写入的映射：



数据库中的目标表按 ITEM_ID 进行分区，如下所示：

- 分区 1： 0001–2999
- 分区 2： 3000–5999
- 分区 3： 6000–9999

要优化到目标表的写入，请完成以下任务：

1. 将目标实例上的分区类型设置为键范围。
2. 创建三个分区。
3. 选择 ITEM_ID 作为分区键。
集成服务使用此键将数据传递给相应分区。
4. 如下所示设置键范围：

ITEM_ID	Start Range	End Range
Partition #1	-	3000
Partition #2	3000	6000

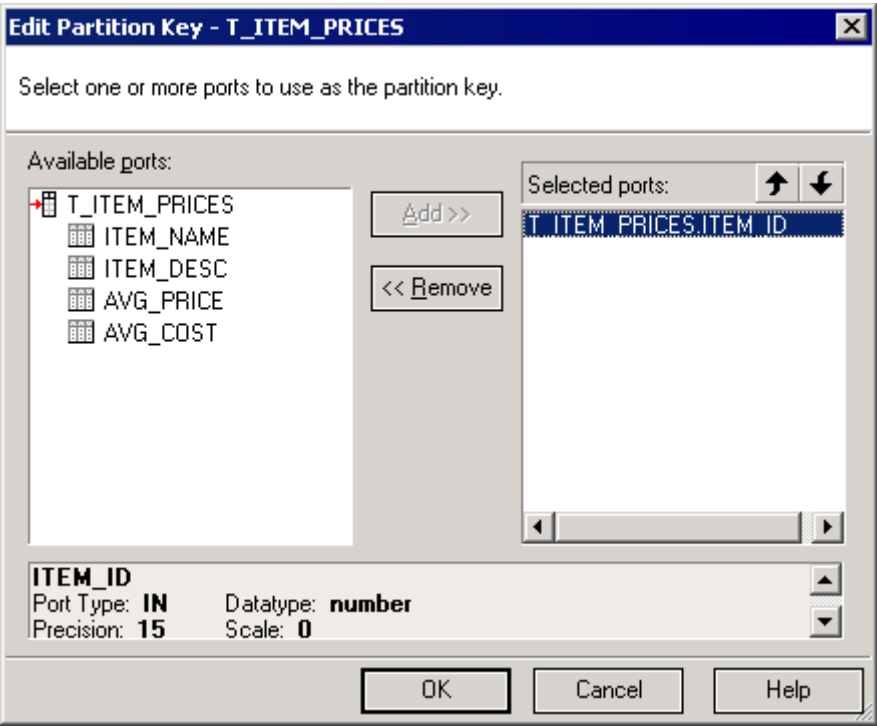
ITEM_ID	Start Range	End Range
Partition #3	6000	-

设置键范围时，集成服务会将 ID 小于 3000 的所有项目发送至第一个分区。它会将 ID 介于 3000 与 5999 之间的所有项目发送至第二个分区。ID 大于或等于 6000 的项目将进入第三个分区。

添加分区键

要为键范围分区指定哈希键，请在“映射”选项卡的“分区”视图上选择分区点，然后单击“编辑键”。这将显示“编辑分区键”对话框。“可用端口”列表显示了转换中的已连接输入及输入/输出端口。要指定分区键，请从此列表选择一个或多个端口，然后单击“添加”。

下图显示了“编辑分区键”对话框，其中已选择一个端口作为目标表 T_ITEM_PRICES 的分区键：



要重新排列用于定义分区键的端口的顺序，请在“选定端口”中选择端口，然后单击向上或向下箭头。

在键范围分区中，端口的顺序不会影响集成服务在分区之间重新分发行的方式，但是会影响会话性能。例如，您可以配置以下复合分区键：

Selected Ports

ITEMS.DESCRPTION

ITEMS.DISCONTINUED_FLAG

由于布尔比较的速度通常比字符串比较的速度快，如果按以下顺序排列端口，则会话运行的速度可能更快：

Selected Ports

ITEMS.DISCONTINUED_FLAG

Selected Ports

ITEMS.DESCRPTION

添加键范围

标识构成分区键的端口之后，必须在“映射”选项卡的“分区”视图上输入每个端口的范围。

可以将分区的开始或结束范围留空。 将开始范围留空时，集成服务会为开始范围使用最小数据值。 将结束范围留空时，集成服务会为结束范围使用最大数据值。

例如，可以根据包含两个分区的管道中的 CUSTOMER_ID 添加以下键范围：

CUSTOMER_ID	Start Range	End Range
Partition #1		135000
Partition #2	135000	

集成服务读取客户表时，会将客户 ID 小于 135000 的所有行发送至第一个分区，并将客户 ID 等于或大于 135000 的所有行发送至第二个分区。 集成服务会消除包含空值或键范围之外的值的行。

配置管道以将数据加载到关系目标时，如果某行在定义分区键的任何列中包含空值，或者如果某行包含位于所有键范围之外的值，则集成服务会将该行发送至第一个分区。

配置管道以从关系源中读取数据时，集成服务会读取位于键范围之内内的行。 它不会读取任何分区键列中包含空值的行。

如果要读取分区键中包含空值的行，请使用传递分区并创建 SQL 替代。

添加筛选器条件

如果为关系源指定键范围分区，则可以指定可选的筛选器条件或替代 SQL 查询。

创建键范围的规则和准则

- 创建键范围时，请遵循以下规则和准则：
- 分区键必须至少包含一个端口。
 - 如果在任何分区点上选择键范围分区，则必须在分区键中指定每个端口的范围。
 - 使用标准的 PowerCenter 日期格式在键范围中输入日期。
 - Workflow Manager 不会验证重叠字符串或数值范围。
 - Workflow Manager 不会验证存在间隔或丢失的范围。
 - 如果选择键范围分区且需要为任何端口输入日期范围，请使用标准的 PowerCenter 日期格式。
 - 在源限定符转换上定义键范围分区时，如果在源限定符转换中更改 SQL 语句，集成服务将默认为传递分区。
 - Workflow Manager 不会验证重叠字符串范围、重叠数值范围、存在间隔或丢失的范围。
 - 如果某行在定义分区键的任何列中包含空值，或者如果某行包含位于所有键范围之外的值，则集成服务会将该行发送至第一个分区。

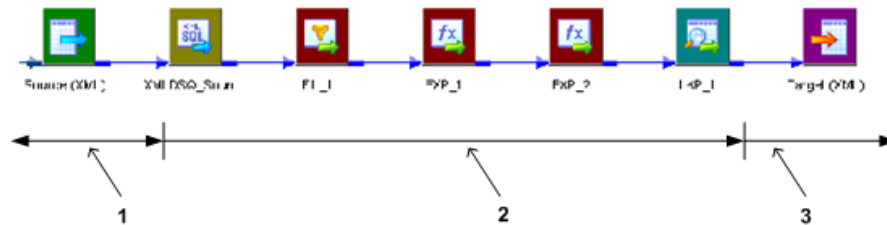
传递分区类型

在传递分区中，集成服务会在不将行重新分发给各分区的情况下处理数据。因此，单个分区中的所有行在错过传递分区点后将保留在该分区中。

将分区点添加到管道中时，主线程会创建一个额外的管道阶段。如果您要增加数据吞吐量，但不希望增加分区数，请使用传递分区。

可以在管道中的任何有效分区点上指定传递分区。

下图显示了其中传递分区可增加数据吞吐量的映射：



1. 读取器线程（第一阶段）。
2. 转换线程（第二阶段）。
3. 写入器线程（第三阶段）。

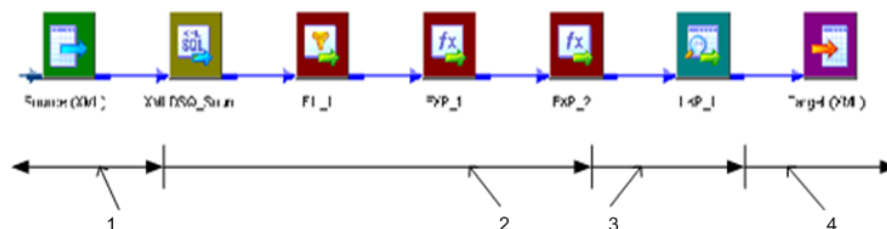
默认情况下，该映射在源限定符和目标实例上包含分区点。由于该映射包含 XML 目标，因此在任何分区点上只能配置一个分区。

在这种情况下，主线程会创建一个用于从源读取数据的读取器线程、一个用于处理数据的转换线程和一个用于将数据写入目标的写入器线程。每个管道阶段都按如下所示处理行：

源限定符 (第一阶段)	转换 (第二阶段)	目标实例 (第三阶段)
行集 1	-	-
行集 2	行集 1	-
行集 3	行集 2	行集 1
行集 4	行集 3	行集 2
...
行集 n	行集 (n-1)	行集 (n-2)

因为该管道包含三个阶段，所以集成服务能够同时处理三个行集。

如果表达式转换是非常复杂的，处理第二（转换）阶段可能会花费较长时间并导致数据吞吐量降低。为提高性能，请在表达式转换 EXP_2 上设置分区点并将分区类型设置为传递。这将创建一个额外的管道阶段。主线程会创建一个额外的转换线程：



1. 读取器线程（第一阶段）。
2. 转换线程（第二阶段）。
3. 转换线程（第三阶段）。
4. 写入器线程（第四阶段）。

集成服务现在能够同时处理四个行集，如下所示：

源限定符 (第一阶段)	FIL_1 & EXP_1 转换 (第二阶段)	EXP_2 & LKP_1 转换 (第三阶段)	目标实例 (第四阶段)
行集 1	-	-	-
行集 2	行集 1	-	-
行集 3	行集 2	行集 1	-
行集 4	行集 3	行集 2	行集 1
...
行集 n	行集 (n-1)	行集 (n-2)	行集 (n-3)

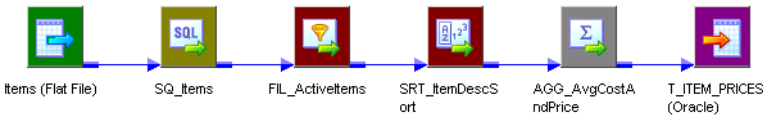
通过在表达式转换 EXP_2 上添加一个额外的分区点，可以使用两个短时间运行的转换阶段替换一个长时间运行的转换阶段。数据吞吐量取决于运行时间最长的阶段。因此在这种情况下，数据吞吐量会增加。

循环分区类型

在循环分区中，PowerCenter 集成服务会将数据块分发给一个或多个分区。每个分区根据块的数量和大小处理行。

如果不需要对各分区中的数据进行分组，请使用循环分区。在从不同大小的文件源读取数据的管道中，使用循环分区在分区之间分发行块。

下图显示了其中循环分区可帮助在行进入筛选器转换之前对其进行分发的映射：



基于该映射的会话从三个不同大小的平面文件中读取项目信息：

- 源文件 1：80,000 行
- 源文件 2：5,000 行
- 源文件 3：15,000 行

当 PowerCenter 集成服务读取源数据时，第一个分区将开始处理 80% 的数据，第二个分区处理 5% 的数据，第三个分区处理 15% 的数据。

要更均匀地分发工作负载，请在筛选器转换上设置分区点并将分区类型设置为循环。PowerCenter 集成服务将分发数据，以便每个分区处理大约三分之一的数据。

第 4 章

下推优化

本章包括以下主题：

- [下推优化概览, 62](#)
- [下推优化类型, 63](#)
- [活动和空闲数据库, 64](#)
- [使用数据库, 64](#)
- [下推兼容性, 67](#)
- [错误处理、日志记录和恢复, 83](#)
- [使用渐变维度, 84](#)
- [使用序列和视图, 84](#)
- [使用 \\$\\$PushdownConfig 映射参数, 87](#)
- [为会话配置下推优化, 89](#)

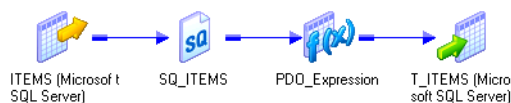
下推优化概览

可以使用下推优化将转换逻辑推送至源或目标数据库。运行已配置下推优化的会话时，集成服务会将转换逻辑转换为 SQL 查询，并将 SQL 查询发送至数据库。源或目标数据库将执行 SQL 查询以处理转换。

您可推送至数据库的转换逻辑量取决于数据库、转换逻辑、映射和会话配置。集成服务将处理所有无法推送至数据库的转换逻辑。

使用“下推优化查看器”预览集成服务可推送至源或目标数据库的 SQL 语句和映射逻辑。还可以使用“下推优化查看器”查看与下推优化相关的消息。

下图显示的映射中包含可推送至源数据库的转换逻辑：



此映射包含可基于店号 5419 和源中的项目 ID 创建项目 ID 的表达式转换。要将转换逻辑推送至数据库，集成服务会生成以下 SQL 语句：

```
INSERT INTO T_ITEMS(ITEM_ID, ITEM_NAME, ITEM_DESC) SELECT CAST((CASE WHEN 5419 IS NULL THEN '' ELSE 5419  
END) + '-' + (CASE WHEN ITEMS.ITEM_ID IS NULL THEN '' ELSE ITEMS.ITEM_ID END) AS INTEGER),  
ITEMS.ITEM_NAME, ITEMS.ITEM_DESC FROM ITEMS2 ITEMS
```

集成服务会生成 INSERT SELECT 语句以从源表中检索 ID、名称和说明值，创建新项目 ID，并将这些值插入目标表中的 ITEM_ID、ITEM_NAME 和 ITEM_DESC 列。它会连接店号 5419、下划线和原始 ITEM ID 以获取新项目 ID。

集成服务无法将表达式推送至数据库时，会在 workflow 日志和“下推优化查看器”中记录一条消息。使用此消息可确定无法将表达式推送至数据库的原因。

下推优化类型

配置以下下推优化类型：

- **源端下推优化。** 集成服务将尽可能多的转换逻辑推送到源数据库。
- **目标端下推优化。** 集成服务将尽可能多的转换逻辑推送到目标数据库。
- **完整下推优化。** 集成服务尝试将所有转换逻辑推送至目标数据库。如果集成服务无法将所有转换逻辑推送至数据库，则它会同时执行源端和目标端下推优化。

运行源端下推优化会话

运行已配置源端下推优化的会话时，集成服务将通过从源到目标或者直到到达无法推送至源数据库的下游转换来分析映射。

集成服务将基于可以推送至数据库的每个转换的转换逻辑来生成和执行 SELECT 语句。然后，会读取此 SQL 查询的结果并处理剩余转换。

运行目标端下推优化会话

运行已配置目标端下推优化的会话时，集成服务将通过从目标到源或者直到到达无法推送至目标数据库的上游转换来分析映射。它将基于可以推送至目标数据库的每个转换的转换逻辑来生成 INSERT、DELETE 或 UPDATE 语句。集成服务会在其将逻辑转换推送至数据库时处理转换逻辑。然后，在目标数据库上执行已生成的 SQL。

运行完整下推优化会话

要使用完整下推优化，源和目标数据库必须位于相同关系数据库管理系统中。当源和目标连接相同时，可以配置完整下推优化。运行已配置完整下推优化的会话时，集成服务将通过从源到目标或者直到到达无法推送至目标数据库的下游转换来分析映射。它将基于可以推送至数据库的转换逻辑针对源或目标生成和执行 SQL 语句。

运行具有大量数据和使用完整下推优化的会话时，数据库服务器必须运行长整型事务。生成长整型事务时，请考虑以下数据库性能问题：

- 长整型事务会使用更多数据库资源。
- 长整型事务会将数据库锁定较长时间。这会减少数据库并发并增加死锁可能性。
- 长整型事务会增加出现意外事件的可能性。

要最大程度地减少长整型事务的数据库性能问题，请考虑使用源端或目标端下推优化。

完整优化下的集成服务行为

为会话配置完整优化后，集成服务将通过从源到目标或者直到到达无法推送至目标数据库的下游转换来分析映射。如果集成服务无法将所有转换逻辑推送至目标数据库，则会尝试将所有转换逻辑推送至源数据库。如果无法将所有转换逻辑推送至源或目标，则集成服务会将尽可能多的转换逻辑推送至源数据库，处理其无法推送至任何

数据库的中间转换，然后将其余转换逻辑推送至目标数据库。集成服务将为每个要向其推送转换逻辑的数据库生成和执行 INSERT SELECT、DELETE 或 UPDATE 语句。

例如，映射可包含以下转换：



等级转换无法推送至源或目标数据库。如果已为会话配置完整下推优化，则集成服务会将源限定符转换和汇总器转换推送至源，处理等级转换，然后将表达式转换和目标推送至目标数据库。如果仅可将部分转换逻辑推送至数据库，则集成服务也不会使会话失败。

活动和空闲数据库

在下推优化期间，集成服务会将转换逻辑推送至一个数据库，也称为活动数据库。不处理转换逻辑的数据库称为空闲数据库。例如，映射包含两个由联接器转换联接的源。如果已将会话配置为源端下推优化，则集成服务会将联接器转换逻辑推送至详细管道中的源，即活动数据库。主管道中的源为空闲数据库，因为不处理转换逻辑。

集成服务使用以下条件来确定哪个数据库为活动或空闲数据库：

- 使用完整下推优化时，目标数据库处于活动状态，源数据库处于空闲状态。
- 在包含查找转换的会话中，源或目标数据库处于活动状态，查找数据库处于空闲状态。
- 在包含联接器转换的会话中，详细管道中的源处于活动状态，主管道中的源处于空闲状态。
- 在包含联合转换的会话中，第一个输入组中的源处于活动状态。其他输入组中的源处于空闲状态。

要将转换逻辑推送至活动数据库，该活动数据库的数据库用户帐户必须可以从空闲数据库中读取。

使用数据库

可以为以下数据库配置下推优化：

- Amazon Redshift
- Greenplum
- Google BigQuery
- IBM DB2
- Microsoft Azure SQL 数据仓库
- Microsoft SQL Server
- Netezza
- Oracle
- PostgreSQL
- SAP HANA

- Snowflake
- Sybase ASE
- Teradata
- Vertica

将转换逻辑推送至数据库时，数据库可能会产生与集成服务不同的输出。

使用 ODBC 连接下推优化到数据库

当您使用 ODBC 连接并配置下推优化时，PowerCenter 集成服务可以将转换逻辑推送到使用数据库特定 ODBC 驱动程序的数据库。

在 ODBC 连接中选择 ODBC 子类型，将转换逻辑推送到数据库。可以在 ODBC 连接对象定义中指定 ODBC 子类型。

可以为以下 ODBC 连接类型配置特定的 ODBC 子类型：

- AWS Redshift
- Azure DW
- Greenplum
- Google Big Query
- PostgreSQL
- Snowflake
- SAP HANA
- 无

默认值为“无”。当您将 ODBC 子类型选择为“无”时，PowerCenter 集成服务无法将转换逻辑推送到数据库。

比较集成服务的输出和数据库

处理相同的转换逻辑时，集成服务和数据库可能会产生不同的结果。读取数据时，集成服务有时会将数据转换为不同的格式。集成服务和数据库在如何处理空值和排序顺序以及是否区分大小写方面也可能会有所差别。

如果以下设置和转换不同，数据库和集成服务将生成不同的输出：

- **空值被视为最高值还是最低值。** 集成服务和数据库在如何处理空值方面可能有所差别。例如，您需要将排序器转换推送至 Oracle 数据库。在会话中，您将空值配置为排序顺序中的最低值。Oracle 将空值视为排序顺序中的最高值。
- **排序顺序。** 集成服务和数据库可以使用不同的排序顺序。例如，您需要将会话中的转换推送至 Microsoft SQL Server 数据库，而该数据库配置为使用不区分大小写的排序顺序。您将会话属性配置为使用区分大小写的二进制排序顺序。结果有所差别，具体取决于集成服务还是 Microsoft SQL Server 数据库负责处理转换逻辑。
- **区分大小写。** 集成服务和数据库在是否区分大小写方面的处理可能有所差别。例如，集成服务使用区分大小写的查询，但数据库不区分大小写。某个筛选器转换使用以下筛选条件：IIF(col_varchar2 = 'CA', TRUE, FALSE)。您需要数据库返回与“CA”匹配的行。但是，如果将此转换逻辑推送至不区分大小写的 Microsoft SQL Server 数据库，则将返回与值“Ca”、“ca”、“cA”和“CA”匹配的行。
- **转换为字符值的数值。** 集成服务和数据库可能会将相同的数值转换为不同格式的字符值。数据库可能会将数值转换为不可接受的字符格式。例如，一个表包含数字 1234567890。集成服务将该数字转换为字符值时，会插入字符“1234567890”。但是，数据库会将该数字转换为“1.2E9”。这两组字符表示相同的值。但是，如果您需要使用格式为“1234567890”的字符，则可以禁用下推优化。

- **精度。** 集成服务和数据库可能会对特定的数据类型使用不同的精度。转换数据类型使用默认的数值精度，该精度可能与本地数据类型的精度不同。例如，转换小数数据类型的精度为 1-28。对应的 Teradata 小数数据类型的精度为 1-18。如果数据库使用的精度与集成服务不同，结果可能有所差别。

IBM DB2 的规则和准则

向 IBM DB2 数据库应用下推优化时，需要类型转换的会话可能会失败。如果从浮点或双精度转换为字符串，或者如果会话需要 IBM DB2 数据库不允许的转换类型，则会话将失败。

Netezza 的规则和准则

对于下推优化到 Netezza 数据库，请遵循以下规则和准则：

- 如果 Netezza 数据库表包含日期、时间或时间戳列，则必须启用 Pre 85 时间戳兼容性会话属性才能在 Netezza 上执行目标端下推优化。如果禁用该选项，则集成服务会处理目标操作。
- 对于被动连接或未连接的查找转换，无法将转换逻辑推送到 Netezza。

Teradata 的规则和准则

对 Teradata 数据库的下推优化应遵循以下规则和准则：

- 如果下推优化会话将小数或双精度数据类型转换为字符串数据类型，则此会话将失败。
- 如果目标端下推优化会话将日期数据类型转换为字符串数据类型，则此会话将失败。

Vertica 的规则和准则

当您向使用 CHAR 列作为参数的函数应用下推优化时，Vertica 数据库脚本会删除 CHAR 列值中的填充空格。

Microsoft Azure SQL 数据仓库的规则和指南

使用以下规则和准则来使用完整下推优化，以在 Microsoft Azure SQL 数据仓库中读取数据或写入数据：

1. 安装适用于 Windows 和 Linux 操作系统的 Microsoft ODBC 驱动程序。要下载驱动程序，请参阅 <https://docs.microsoft.com/en-us/sql/connect/odbc/linux-mac/installing-the-microsoft-odbc-driver-for-sql-server#microsoft-odbc-driver-131-for-sql-server>。
2. 对于 Linux 操作系统，必须设置驱动程序的 ODBCINI 和 LD_LIBRARY_PATH 环境变量并创建 DSN 条目。将 ODBCINI 文件的值设置为 odbc.ini 文件的位置。例如，setenv ODBCINI "/data/home/adputf_9/cloud_td/ODBCINI/odbc.ini"。
3. 要设置 LD_LIBRARY_PATH 环境变量，请使用以下格式：
setenv LD_LIBRARY_PATH "/opt/microsoft/msodbcsql/lib64/libmsodbcsql-11.0.so.2270.0"
4. 在 odbc.ini 文件中添加 Microsoft Azure SQL 数据仓库数据源的条目。以下部分展示了 odbc.ini 文件中的条目示例：
[Sample Azure DW ODBC DSN][SD_Azure_DW]Driver=/opt/microsoft/msodbcsql/lib64/libmsodbcsql-11.0.so.2270.0Description=Microsoft ODBC Driver 11 for SQL
ServerServer=dghhg2ad3.database.windows.netDatabase=INFASQLDW_DEVLogonID=infadwadminPassword=QuotedId=YesAnsiNPW=YesEncryptionMethod=1SeedBeforeConnect=1EnableQuotedIdentifiers=1ValidateServerCertificate=0DriverUnicodeType=
5. 重新启动 PowerCenter 集成服务。

注意：在 ODBC 连接中，选择子类型作为 **AzureDW** 以使用完整下推优化。

下推兼容性

可以配置会话将日期转换推送到数据库。还可以将运算符、变量和函数推送到数据库。使用下推优化时，集成服务通过确定数据库中的等价运算符、变量和函数来对转换或工作流链接中的表达式进行转换。如果不存在等价运算符、变量和函数，则集成服务将处理转换逻辑。

要将具有多个连接的转换推送至数据库，则这些连接必须在下推操作上可兼容。

连接兼容性

如果它们连接到相同数据库管理系统中的数据库，并且集成服务可以确定这些连接访问的数据库表，则这些连接在下推操作上可兼容。

以下转换具有多个连接：

- **连接器。** 连接器转换可以联接来自多个源连接中的数据。
- **联合。** 联合转换可以合并来自多个源连接中的数据。
- **查找。** 查找转换的连接与源连接不同。
- **目标。** 目标连接与源连接不同。

每个连接对象均可与其自身在下推操作上兼容。如果将会话配置为使用源和目标连接的相同连接对象，则集成服务会将转换逻辑推送至源或目标数据库。

如果某些关系连接为相同数据库类型，具有相同的数据库用户名和密码，并且拥有某些相同属性，则这些连接在下推操作上可兼容。

下表列出了每个数据库类型都必须相同的连接属性：

数据库类型	必须相同的连接属性
IBM DB2	连接字符串 代码页 连接环境 SQL 事务环境 SQL
Greenplum	代码页 连接字符串 连接环境 SQL 事务环境 SQL
Microsoft SQL Server	代码页 服务器名称 域名 使用受信任连接 连接环境 SQL 事务环境 SQL
Oracle	连接字符串 代码页 连接环境 SQL 事务环境 SQL

数据库类型	必须相同的连接属性
Sybase ASE	代码页 服务器名称 连接环境 SQL 事务环境 SQL
Teradata	代码页 数据源名称 连接环境 SQL 事务环境 SQL
Vertica	代码页 连接字符串 连接环境 SQL 事务环境 SQL
Microsoft Azure SQL 数据仓库	代码页 服务器名称 域名 使用受信任连接 连接环境 SQL 事务环境 SQL

注意: 集成服务会执行区分大小写的字符串比较以验证连接属性是否相同。

相同关系数据库管理系统中的 Netezza 数据库在下推操作上不兼容。Netezza 数据库仅可与其自身在下推操作上兼容。

对于相同数据库类型的连接，如果 [“下推兼容性”页面上 67](#) 中的连接属性相同，但数据库用户名称和密码不同，将仍可使这些连接在下推操作上兼容。

数据库连接的不兼容用户

如果其他可兼容连接的数据库用户名和密码不匹配，必须提供其他信息以使连接兼容。

要使连接在下推操作上兼容，请执行以下操作：

1. 验证活动数据库的数据库用户是否具有所有空闲数据库上的读取权限。
2. 启用“对用户不兼容的连接允许下推”会话属性。
3. 对于与 Microsoft SQL Server 和 Sybase 的每个空闲连接，还必须指定连接属性中的数据库名称以及所有查找和源的表所有者。

空闲数据库中的限定名和限定表

集成服务生成 SQL 以将转换推送至活动数据库时，生成的 SQL 至少引用空闲数据库中的一个表。

对于以下情况，要确保集成服务可以确定所有表，必须限定空闲数据库中的表名称：

- 活动和空闲连接具有相同的连接属性并且为相同数据库类型，但是数据库用户名和密码不同。
- 源限定符转换包含源筛选器或用户定义的联接。

注意: 对于所有其他情况，集成服务都会限定空闲数据库中的表名称。

对于源限定符转换，在“所有者名称”会话属性中限定源表的名称。对于查找转换，在“查找表名称”会话属性中限定查找表的名称。

使用以下语法限定表名称：

数据库类型	语法
IBM DB2	<表所有者>.<表名称>
Microsoft SQL Server	<数据库名称>.<表所有者>.<表名称>
Netezza	不受支持
Oracle	<表所有者>.<表名称>
Sybase ASE	<数据库名称>.<表所有者>.<表名称>
Teradata	<数据库名称>.<表名称>
Vertica	<数据库名称>.<架构名称>.<表名称>
Microsoft Azure SQL 数据仓库	<数据库名称>.<表所有者>.<表名称>

日期兼容性

集成服务和数据库能够以不同的方式处理日期。将会话配置为将日期转换推送至数据库时，可能会产生意外结果或者会话可能会失败。

当以下日期设置和转换不同时，数据库可能会产生与集成服务不同的输出：

- **转换为字符值的日期值。**集成服务将转换日期/时间数据类型转换为在数据库中支持子秒精度的本地数据类型。如果将会话中的日期时间格式配置为数据库不支持的格式，则会话将失败。例如，当集成服务对日期执行 ROUND 函数时，将使用 MM/DD/YYYY HH:MI:SS.US 格式将该日期值存储在字符列中。当数据库执行此函数时，它会采用数据库的默认日期格式存储该日期。如果该数据库为 Oracle 数据库，会将日期存储为默认的 DD-MON-YY 格式。如果需要日期采用 MM/DD/YYYY HH:MI:SS.US 格式，可以禁用下推优化。
- **TO_CHAR 和 TO_DATE 函数的日期格式。**如果集成服务将函数推送至数据库，则使用 TO_CHAR 或 TO_DATE 函数中的日期格式。数据库将每个日期字符串转换为数据库所支持的日期时间值。

例如，集成服务将以下表达式推送至数据库：

```
TO_DATE( DATE_PROMISED, 'MM/DD/YY' )
```

数据库根据指定的日期格式字符串 MM/DD/YY 来解释 DATE_PROMISED 端口中的日期字符串。数据库将每个日期字符串（如 01/22/98）转换为支持的日期值（如 Jan 22 1998 00:00:00）。

如果集成服务将日期格式推送至数据库不支持的 IBM DB2、Microsoft SQL Server 或 Sybase 数据库，则集成服务将停止下推优化，然后再处理转换。

集成服务会先转换所有日期，然后将转换推送至 Oracle 或 Teradata 数据库。如果数据库不支持日期转换后的日期格式，会话将失败。

- **HH24 日期格式。**无法在 Teradata 的日期格式字符串中使用 HH24 格式。当集成服务为 Teradata 数据库生成 SQL 时，使用 HH 格式的字符串。
- **日期格式字符串中的空格。**不能在 Teradata 的日期格式字符串中使用空格。当集成服务为 Teradata 数据库生成 SQL 时，会将空格替换为 B。
- **处理查找转换的子秒精度。**如果为查找转换启用子秒精度，则数据库和集成服务将使用子秒精度执行查找比较，但返回不同的结果。与集成服务不同，数据库不会根据子秒精度截断查找结果。例如，将查找转换配置为

显示精确到毫秒的子秒精度。如果查找结果为 8:20:35.123456，则数据库将返回 8:20:35.123456，但集成服务将返回 8:20:35.123。

- **SYSDATE 内置变量。**使用 SYSDATE 内置变量时，集成服务将返回运行服务进程的节点的当前日期和时间。但是，将转换逻辑推送至数据库时，SYSDATE 变量将返回托管该数据库的计算机的当前日期和时间。如果托管该数据库的计算机与运行集成服务进程的计算机所在的时区不同，则结果可能有所不同。

运算符兼容性

使用下推优化时，集成服务通过确定数据库中的等价运算符来对转换或工作流链接中的表达式进行转换。如果不存在等价运算符，则集成服务将处理转换逻辑。

您可以使用完整下推优化将以下运算符推送到数据库：

+ - * / % || = > < >= <= <> != ^= not and or

请考虑运算符的以下例外情况：

- 不能将 * 运算符推送到 Amazon Redshift。
- 可以使用源下推优化将 = > < >= <= <> != ^= not and or 运算符下推到 Amazon Redshift。
- 不能将 % 运算符下推到 Greenplum。
- 可以使用源下推优化将 || 运算符下推到 IBM DB2、Microsoft SQL Server、Sybase ASE、Teradata 和 Azure DW。

变量兼容性

使用下推优化时，集成服务通过确定数据库中的等价变量来对转换或工作流链接中的表达式进行转换。如果不存在等价变量，则集成服务将处理转换逻辑。

您可以将以下变量推送到数据库：

- SESSSTARTTIME
- SYSDATE

适用于云数据仓库应用程序的函数

下表汇总了云数据仓库应用程序中 PowerCenter 函数的可用性：

Amazon Redshift*	Amazon Redshift	Google BigQuery	Snowflake
ABS()	源，完整	源，完整	源，完整
ADD_TO_DATE()	源，完整	源，完整 支持日期数据类型。	-
ASCII()	-	-	源，完整
AVG()	源	源，完整	源，完整
CEIL()	源，完整	源，完整	源，完整
CHR()	源，完整	源，完整	源，完整
CONCAT()	源，完整	源，完整	源，完整

Amazon Redshift*	Amazon Redshift	Google BigQuery	Snowflake
COS()	源, 完整	源, 完整	源, 完整
COSH()	-	-	源, 完整
COUNT()	源	源, 完整	源, 完整
DATE_COMPARE()	源, 完整	源, 完整	源, 完整
DATE_DIFF()	源, 完整	源, 完整	源, 完整
DECODE()	源, 完整	源, 完整	源, 完整
EXP()	源, 完整	源, 完整	源, 完整
FLOOR()	源, 完整	源, 完整	源, 完整
GET_DATE_PART()	源, 完整	源, 完整 支持日期数据类型。	源, 完整
IIF()	源, 完整	源, 完整	源, 完整
IN()	源	源, 完整	源, 完整
INITCAP()	源, 完整	-	源, 完整
INSTR()	源, 完整	源, 完整	源, 完整
IS_DATE()	-	源, 完整	-
ISNULL()	源	源, 完整	源, 完整
LAST_DAY()	源, 完整	源, 完整 支持日期数据类型。	源, 完整
LENGTH()	源, 完整	源, 完整	源, 完整
LN()	源, 完整	-	源, 完整
LOG()	-	-	源, 完整
查找	-	源, 完整	-
LOWER()	源, 完整	源, 完整	源, 完整
LPAD()	源, 完整	源, 完整	源, 完整
LTRIM()	源, 完整	源, 完整	源, 完整
MAX()	源	源, 完整 支持数字数据类型。	源, 完整
MEDIAN()	-	-	源, 完整

Amazon Redshift*	Amazon Redshift	Google BigQuery	Snowflake
MIN()	源	源，完整 支持日期、数字和字符串数据类型。	源，完整
MOD()	源	源，完整	源，完整
POWER()	源，完整	源，完整	源，完整
REPLACECHR()	-	源，完整	源，完整
REPLACESTR()	-	源，完整	源，完整
ROUND(DATE)	-	源，完整 支持时间戳数据类型。	-
ROUND(NUMBER)	源，完整	源，完整	源，完整
RPAD()	源，完整	源，完整	源，完整
RTRIM()	源，完整	源，完整	源，完整
SIGN()	源，完整	-	源，完整
SIN()	源，完整	源，完整	源，完整
SINH()	-	-	源，完整
SQRT()	源，完整	源，完整	源，完整
STDDEV()	源	-	源，完整
SUBSTR()	源，完整	源，完整	源，完整
SUM()	源	源，完整	源，完整
SYSDATE()	源	源，完整	源，完整
SYSTIMESTAMP()	源	源，完整 支持无格式。	源，完整
TAN()	源，完整	源，完整	源，完整
TANH()	-	-	源，完整
TO_BIGINT	源，完整	源，完整	源，完整
TO_CHAR(DATE)	源	源，完整	源，完整
TO_CHAR(NUMBER)	源，完整	-	源，完整
TO_DATE()	源，完整	源，完整	源，完整
TO_DECIMAL()	源，完整	-	源，完整

Amazon Redshift*	Amazon Redshift	Google BigQuery	Snowflake
TO_FLOAT()	源, 完整	源, 完整	源, 完整
TO_INTEGER()	源, 完整	源, 完整	源, 完整
TRUNC(DATE)	源	源, 完整 支持时间戳数据类型。	-
TRUNC(NUMBER)	源	-	源, 完整
UPPER()	源, 完整	源, 完整	源, 完整
VARIANCE()	源	-	源, 完整
*如果未列出某个函数, 则集成服务无法将该函数推送到任何数据库。			

Amazon Redshift 函数异常

将函数推送至 Amazon Redshift 时, 请遵循以下规则和准则:

- 要将 TRUNC(DATE) 推送到 Amazon Redshift, 必须定义日期参数和格式参数。否则, PowerCenter 集成服务无法将该函数推送到 Amazon Redshift。
- Amazon Redshift 的汇总器函数仅接受一个参数和一个字段集。不接受筛选条件参数。此外, 请确保在 GROUP BY 子句中列出映射到输出的所有端口。
- 对于 Amazon Redshift, 如果仅为 TO_DATE() 和 TO_CHAR() 定义 string 参数, 则 PowerCenter 集成服务会考虑采用会话属性中存在的默认日期格式。会话属性中的默认日期格式为: MM/DD/YYYY HH24:MI:SS.US
- 要将 SYSTIMESTAMP 推送到 Amazon Redshift, 请勿指定 SYSTIMESTAMP() 的格式。Amazon Redshift 数据库会返回完整时间戳。
- 要将 INSTR() 推送到 Amazon Redshift, 必须仅定义 string、search_value 和 start 参数。Amazon Redshift 不支持 occurrence 和 comparison_type 参数。
- 将 TO_BIGINT 和 TO_INTEGER 推送到 Amazon Redshift 时, flag 参数会被忽略。
- 将 IN() 推送到 Amazon Redshift 时, CaseFlag 参数会被忽略。
- 如果在 ADD_TO_DATE() 函数中使用 NS 格式, 则 PowerCenter 集成服务无法将函数推送到 Amazon Redshift。
- 如果在 TO_CHAR() 和 TO_DATE() 函数中使用以下任意格式, 则 PowerCenter 集成服务无法将 NS、SSSS、SSSSS 和 RR 函数推送到 Amazon Redshift:
- 要将 TRUNC(DATE) 和 DATE_DIFF() 推送到 Amazon Redshift, 必须使用以下格式, 例如 D、HH24、MI、MM、MS、SS、US 和 YYYY。
- 要将 GET_DATE_PART() 推送到 Amazon Redshift, 必须使用以下格式, 例如 D、DDD、HH24、MI、MM、MS、SS、US 和 YYYY。

Google BigQuery 函数异常

将函数推送至 Google BigQuery 数据库时, 请遵循以下规则和准则:

- 要将 ADD_TO_DATE() 函数推送到 Google BigQuery 数据库, 必须定义 Date 数据类型的参数。
- 要将 GET_DATE_PART() 函数推送到 Google BigQuery 数据库, 必须定义 Date、DateTime 或 Timestamp 数据类型的参数。

- 要将 INSTR() 函数推送到 Google BigQuery 数据库，必须使用以下格式：
INSTR(string, search_value)
- 要将 IS_DATE() 或 LAST_DAY() 函数推送到 Google BigQuery 数据库，必须定义 Date 数据类型的参数。
- 要将函数推送到 Google BigQuery 数据库，必须定义 Date 数据类型的参数。
- 要将 MAX() 函数推送到 Google BigQuery 数据库，必须定义 Number 数据类型的参数。
- 要将 MIN() 函数推送到 Google BigQuery 数据库，必须定义 Date、Number 或 String 数据类型的参数。
- 要将 ROUND(DATE) 函数推送到 Google BigQuery 数据库，必须使用以下格式定义 Timestamp 数据类型的参数：
 - D
 - DD
 - DDD
 - DY
 - HH
 - HH24
 - MI
 - SS
 - MS
- 要将 TRUNC(DATE) 函数推送到 Google BigQuery 数据库，必须使用以下格式定义 Timestamp 数据类型的参数：
 - Y
 - YY
 - YYYY
 - YYYY
 - MM
 - MON
 - D
 - DD
 - DDD
 - DY
 - HH
 - HH24
 - MI
 - SS
 - MS
 - US
- 要将 TO_CHAR(DATE) 函数推送到 Google BigQuery 数据库，必须使用以下格式定义 Timestamp 数据类型的参数：
 - YYYY
 - MM
 - DD

- HH24
- MI
- CN
- -
- /
- .
- ;
- :
- "text"
- 在 TO_DATE() 函数中定义 Timestamp 数据类型的参数时，必须使用以下格式：
 - YYYY
 - MM
 - DD
 - HH24
 - MI
 - SS
 - MS
 - CN
 - -
 - /
 - .
 - ;
 - :
- 在 GET_DATE_PART() 函数中定义 Timestamp 数据类型的参数时，必须使用以下格式：
 - Y
 - YY
 - YYYY
 - MM
 - MON
 - D
 - DD
 - DDD
 - DY
 - HH
 - HH24
 - MI
 - SS
 - MS

- CN

- 将 TO_DATE() 函数推到 Google BigQuery 数据库时，必须将输出映射到 Google BigQuery 表中的 Timestamp 列。
- 在将 SYSTIMESTAMP() 函数推送到 Google BigQuery 数据库时，不要指定任何格式。Google BigQuery 数据库会返回完整的时间戳。

Snowflake 函数异常

将函数推送至 Snowflake 时，请遵循以下规则和准则：

- 要将 TRUNC(DATE) 或 TO_CHAR() 函数推送到 Snowflake 数据库，必须定义日期和格式参数。
- Snowflake 汇总函数仅接受一个参数，即用于汇总函数的字段集。PowerCenter 集成服务会忽略参数中定义的所有筛选条件。确保在 GROUP BY 子句中列出映射到目标的所有字段。
- 在将 SYSTIMESTAMP() 或 SYSDATE() 函数推送到 Snowflake 数据库时，不要指定任何格式。Snowflake 数据库会返回完整时间戳。
- 您不能将包含多个参数的 TO_BIGINT() 或 TO_INTEGER() 函数推送到 Snowflake 数据库。
- 将 REPLACECHR() 或 REPLACESTR() 函数推送到 Snowflake 数据库时，PowerCenter 集成服务将忽略 caseFlag 参数。
例如，REPLACECHR(false, in_F_CHAR, 'a', 'b') 和 REPLACECHR(true, in_F_CHAR, 'a', 'b') 参数将返回相同的值。
- 在将函数推送到 Snowflake 数据库时，不能使用毫秒和微秒值。
- 在 ADD_TO_DATE() 和 TRUNC(DATE) 函数中可以使用纳秒值。
- 要将 TRUNC(DATE)、GET_DATE_PART() 或 DATE_DIFF() 函数推送到 Snowflake 数据库，必须使用 D、DDD、HH、MI、MM、SS 和 YYYY 时间格式作为参数。
例如，TRUNC(<datefieldname>, 'dd')。

有关日期和时间相关函数的信息，请参阅以下网站：
<https://docs.snowflake.net/manuals/sql-reference/functions-date-time.html#label-supported-date-time-parts>

适用于数据仓库应用程序的函数

下表汇总了数据仓库应用程序中 PowerCenter 函数的可用性：

函数*	Greenplum	Netezza	Teradata	Vertica	PostgreSQL
ABS()	源，完整	源，完整	源，完整	源，完整	源，完整
ADD_TO_DATE()	源，完整	源，完整	源，完整	源，完整	源，完整
ASCII()	源，完整	源，完整	-	源，完整	源，完整
AVG()	源，完整	源，完整	源，完整	源，完整	源，完整
CEIL()	源，完整	源，完整	源	源	源，完整
CHR()	源，完整	源，完整	-	源，完整	源，完整
CONCAT()	源，完整	-	源	源，完整	源，完整

函数*	Greenplum	Netezza	Teradata	Vertica	PostgreSQL
COS()	源, 完整	源, 完整	源, 完整	源, 完整	源, 完整
COSH()	源, 完整	源, 完整	源, 完整	源, 完整	-
COUNT()	源, 完整	源, 完整	源, 完整	源, 完整	源, 完整
DATE_COMPARE()	源, 完整	源, 完整	源	源, 完整	-
DATE_DIFF()	源, 完整	-	-	源, 完整	源, 完整
DECODE()	源, 完整	源, 完整	源, 完整	源, 完整	源, 完整
EXP()	源, 完整	源, 完整	源, 完整	源, 完整	源, 完整
FLOOR()	源, 完整	源, 完整	源	源, 完整	源, 完整
GET_DATE_PART()	源, 完整	源, 完整	源, 完整	源, 完整	-
IIF()	源, 完整	源, 完整	源, 完整	源, 完整	源, 完整
IN()	源, 完整	-	源, 完整	-	源, 完整
INITCAP()	源, 完整	-	-	源, 完整	源, 完整
INSTR()	-	-	源	源, 完整	-
ISNULL()	源, 完整	源, 完整	源, 完整	源, 完整	源, 完整
LAST_DAY()	-	-	-	源, 完整	源, 完整
LENGTH()	源, 完整	源, 完整	源, 完整	源, 完整	源, 完整
LN()	源, 完整	源, 完整	-	源, 完整	源, 完整
LOG()	源, 完整	源, 完整	源	源	源, 完整
查找	-	源, 完整	源, 完整	源, 完整	源, 完整
LOWER()	源, 完整	源, 完整	源, 完整	源, 完整	源, 完整
LPAD()	源, 完整	源, 完整	-	源, 完整	源, 完整
LTRIM()	源, 完整	源, 完整	源, 完整	源, 完整	源, 完整
MAX()	源, 完整	源, 完整	源, 完整	源, 完整	源, 完整
MIN()	源, 完整	源, 完整	源, 完整	源, 完整	源, 完整
MOD()	源, 完整	源, 完整	源, 完整	源, 完整	源, 完整
POWER()	源, 完整	源, 完整	源, 完整	源, 完整	源, 完整
ROUND(NUMBER)	源, 完整	源, 完整	源	源	源, 完整

函数*	Greenplum	Netezza	Teradata	Vertica	PostgreSQL
RPAD()	源, 完整	源, 完整	-	源, 完整	源, 完整
RTRIM()	源, 完整	源, 完整	源, 完整	源, 完整	源, 完整
SIGN()	源, 完整	源, 完整	源	源	源, 完整
SIN()	源, 完整	源, 完整	源, 完整	源, 完整	源, 完整
SINH()	源, 完整	源, 完整	源, 完整	源, 完整	-
SQRT()	源, 完整	源, 完整	源, 完整	源, 完整	源, 完整
STDDEV()	源, 完整	-	源, 完整	源, 完整	源, 完整
SUBSTR()	源, 完整	源, 完整	源	源, 完整	源, 完整
SUM()	源, 完整	源, 完整	源, 完整	源, 完整	源, 完整
SYSDATE()	源, 完整	源, 完整	源, 完整	源, 完整	源, 完整
SYSTIMESTAMP()	源, 完整	源, 完整	源, 完整	源, 完整	源, 完整
TAN()	源, 完整	源, 完整	源, 完整	源, 完整	源, 完整
TANH()	源, 完整	源, 完整	源, 完整	源, 完整	-
TO_BIGINT	源, 完整	源, 完整	源, 完整	源, 完整	源, 完整
TO_CHAR(DATE)	源, 完整	源, 完整	源	源, 完整	源, 完整
TO_CHAR(NUMBER)	源, 完整	源, 完整	源, 完整	源, 完整	源, 完整
TO_DATE()	源, 完整	源, 完整	源, 完整	源, 完整	源, 完整
TO_DECIMAL()	源, 完整	源, 完整	源, 完整	源, 完整	源, 完整
TO_FLOAT()	源, 完整	源, 完整	源, 完整	源, 完整	源, 完整
TO_INTEGER()	源, 完整	源, 完整	源, 完整	源, 完整	源, 完整
TRUNC(DATE)	源, 完整	源, 完整	-	源	-
TRUNC(NUMBER)	源, 完整	源, 完整	源	源	源, 完整
UPPER()	源, 完整	源, 完整	源, 完整	源, 完整	源, 完整
VARIANCE()	源, 完整	-	源, 完整	源, 完整	源, 完整
*如果未列出某个函数, 则集成服务无法将该函数推送到任何数据库。					

Greenplum 函数异常

将函数推送至 Greenplum 时，请遵循以下规则和准则：

- 要将 TRUNC(DATE) 推送至 Greenplum，必须使用以下格式，例如 YYYY、DD、DOY、HH、US、MS、MI、MM 和 SS。

Netezza 函数异常

将函数推送至 Netezza 时，请遵循以下规则和准则：

- 可以将 SYSTIMESTAMP('SS') 推送至 Netezza 数据库，但不能将 SYSTIMESTAMP('MS') 或 SYSTIMESTAMP('US') 推送至该数据库。
- 将 TO_CHAR(DATE) 或 TO_DATE() 推送至 Netezza 时，精确到子秒的日期必须使用 YYYY-MM-DD HH24:MI:SS.US 格式。如果格式不同，PowerCenter 集成服务不会将函数推送至 Netezza。

PostgreSQL 函数异常

将函数推送至 PostgreSQL 时，请遵循以下规则和准则：

- 要将 TRUNC(DATE) 函数推送至 PostgreSQL 数据库，必须定义日期和格式参数。
- 如果仅为 TO_DATE() 和 TO_CHAR() 函数定义字符串参数而忽略格式参数，PowerCenter 集成服务将根据在会话属性中所指定的默认日期格式 MM/DD/YYYY HH24:MI:SS 返回一个字符串。
- 将 SYSTIMESTAMP() 函数推送至 PostgreSQL 数据库时，请勿指定格式参数。如果为 SYSTIMESTAMP 指定格式，数据库将忽略该格式并返回完整的时间戳。
- 将 TO_BIGINT 或 TO_INTEGER 函数推送至 PostgreSQL 数据库时，PowerCenter 集成服务将忽略这些标志参数。
- 将 IN() 函数推送至 PostgreSQL 数据库时，PowerCenter 集成服务将忽略 CaseFlag 参数。
- 当使用 NS 格式字符串为 ADD_TO_DATE() 函数设置纳秒时，PowerCenter 集成服务不会将 ADD_TO_DATE() 函数推送至 PostgreSQL。
- 如果使用 NS、JQW、SSSSS 和 RR 格式，PowerCenter 集成服务无法将 TO_CHAR() 和 TO_DATE() 函数推送至 PostgreSQL。
- 将 TRUNC(DATE) 函数推送至 PostgreSQL 数据库时，可以使用 D、HH24、MI、MM、MS、SS、US 和 YYYY 等格式。

Teradata 函数异常

将函数推送至 Teradata 时，请遵循以下规则和准则：

- 如果在转换逻辑中使用 ADD_TO_DATE 更改天、小时、分钟或秒，则无法将函数推送至 Teradata 数据库。

Vertica 函数异常

将函数推送至 Vertica 时，请遵循以下规则和准则：

- 如果您将 YYY、MON、MONTH、HH12 和 HH24 格式用作 DATE_DIFF() 函数的一部分，则 PowerCenter 集成服务不会将该函数推送至 Vertica。
- 当您向 DATE_DIFF 函数推送至 Vertica 时，Vertica 会将日期差异值四舍五入为最近的整数。但是，PowerCenter 集成服务将返回浮点值。例如，如果第一个日期为 2000-08-15，第二个日期为 1997-08-16，Vertica 会将日期差异值四舍五入为 3，而 PowerCenter 集成服务则返回 2.99731182795699。如果您希望在 Vertica 数据库中将日期差异视为浮点值，可以禁用下推优化。

- 当您格式指定为 Y 并将 DATE_DIFF 函数推送至 Vertica 时，Vertica 将按天数计算日期差异。但是，PowerCenter 集成服务则按年数计算差异。如果要按年数处理差异值，可以禁用下推优化。

适用于企业版应用程序的函数

下表汇总了可以使用源端或完整下推优化推送到 SAP HANA 数据库的 PowerCenter 函数的可用性：

ABS()	EXP()	LTRIM()	SUM()
ADD_TO_DATE()	FLOOR()	MAX()	SYSDATE()
ASCII()	GET_DATE_PART()	MIN()	SYSTIMESTAMP()
AVG()	IIF()	MOD()	TAN()
CEIL()	IN()	POWER()	TANH()
CHR()	INITCAP()	ROUND(NUMBER)	TO_BIGINT
CONCAT()	ISNULL()	RPAD()	TO_CHAR(DATE)
COS()	LAST_DAY()	RTRIM()	TO_CHAR(NUMBER)
COSH()	LENGTH()	SIGN()	TO_DATE()
COUNT()	LN()	SIN()	TO_DECIMAL()
DATE_COMPARE()	LOG()	SINH()	TO_FLOAT()
DATE_DIFF()	LOWER()	SQRT()	TO_INTEGER()
DECODE()	LPAD()	SUBSTR()	UPPER()
*如果未列出某个函数，则集成服务无法将该函数推送到数据库。			

适用于关系数据库的函数

下表汇总了关系数据库中 PowerCenter 函数的可用性。带有“全部”标记的列表示可以使用源端、目标端或完整下推优化将函数推送至数据库。如果未列出某个函数，则集成服务无法将该函数推送到任何数据库。

函数*	DB2	Microsoft SQL 服务器	Oracle	Sybase ASE
ABS()	全部	全部	全部	全部
ADD_TO_DATE()	全部	源	全部	源
ASCII()	全部	全部	全部	全部
AVG()	全部	全部	全部	全部
CEIL()	全部	全部	全部	全部

函数*	DB2	Microsoft SQL 服务器	Oracle	Sybase ASE
CHR()	全部	全部	全部	全部
CONCAT()	源	源	全部	源
COS()	全部	全部	全部	全部
COSH()	全部	源	全部	源
COUNT()	全部	全部	全部	全部
DATE_COMPARE()	源	源	源	源
DECODE()	全部	全部	全部	全部
EXP()	全部	全部	全部	全部
FLOOR()	全部	全部	全部	全部
GET_DATE_PART()	全部	全部	全部	全部
GREATEST()	-	-	全部	-
IIF()	全部	全部	全部	全部
IN()	全部	全部	全部	全部
INITCAP()	-	-	全部	-
INSTR()	源	全部	全部	源
ISNULL()	全部	全部	全部	全部
LAST_DAY()	-	-	全部	-
LEAST()	-	-	全部	-
LENGTH()	全部	全部	全部	全部
LOG()	全部	源	全部	源
查找	全部	全部	全部	全部
LOWER()	全部	全部	全部	全部
LPAD()	-	-	全部	-
LTRIM()	全部	全部	全部	全部
MAX()	全部	全部	全部	全部
MIN()	全部	全部	全部	全部

函数*	DB2	Microsoft SQL 服务器	Oracle	Sybase ASE
MOD()	全部	全部	全部	全部
POWER()	全部	全部	全部	全部
ROUND(DATE)	-	-	全部	-
ROUND(NUMBER)	全部	全部	全部	全部
RPAD()	-	-	全部	-
RTRIM()	全部	全部	全部	全部
SIGN()	全部	全部	全部	全部
SIN()	全部	全部	全部	全部
SINH()	全部	源	全部	源
SOUNDEX()	全部	全部	全部	全部
SQRT()	全部	全部	全部	全部
STDDEV()	全部	全部	全部	-
SUBSTR()	源	源	全部	源
SUM()	全部	全部	全部	全部
SYSDATE()	全部	全部	全部	全部
SYSTIMESTAMP()	全部	全部	全部	全部
TAN()	全部	全部	全部	全部
TANH()	全部	源	全部	源
TO_BIGINT	全部	全部	全部	全部
TO_CHAR(DATE)	全部	全部	全部	全部
TO_CHAR(NUMBER)	全部	全部	全部	全部
TO_DATE()	全部	全部	全部	全部
TO_DECIMAL()	全部	全部	全部	全部
TO_FLOAT()	全部	全部	全部	全部
TO_INTEGER()	全部	源	全部	全部
TRUNC(DATE)	-	-	全部	-

函数*	DB2	Microsoft SQL 服务器	Oracle	Sybase ASE
TRUNC(NUMBER)	全部	全部	全部	源
UPPER()	全部	全部	全部	全部
VARIANCE()	全部	全部	全部	-

关系数据库函数异常

将函数推送至关系数据库时，请遵循以下规则和准则：

- 将 LAST_DAY() 推送至 Oracle 时，Oracle 将返回精确到秒的日期。如果输入日期包含子秒，则 Oracle 会将日期截减至秒。
- 当您 LTRIM、RTRIM 或 SOUNDEX 推送至数据库时，数据库会将参数 (' ') 视作 NULL，但 PowerCenter 集成服务会将参数 (' ') 视作空格。
- 当您 SYSDATE 或 SYSTIMESTAMP 推送至数据库时，数据库服务器会以它自己（而非 PowerCenter 集成服务）的时区返回时间戳。
- 如果将 SYSTIMESTAMP 推送至 IBM DB2 或 Sybase 数据库，并且指定了 SYSTIMESTAMP 的格式，则数据库将忽略该格式并返回完整的时间戳。
- IBM DB2 数据库和 PowerCenter 集成服务生成的 STDDEV 和 VARIANCE 结果不同。IBM DB2 使用不同于其他数据库的算法来计算 STDDEV 和 VARIANCE。
- 要将 TO_DATE() 函数推送到 IBM DB2 数据库，必须使用以下格式，例如 YYYYMMDD、YYYYMMDD HH24MISS、YYYY-MM-DD HH24MISS、YYYYMMDD HH24:MI:SS、YYYY/MM/DD HH24:MI:SS 和 YYYY/MM/DD HH24MISS。

错误处理、日志记录和恢复

集成服务和数据库以不同方式处理错误处理、日志记录和恢复。

错误处理

集成服务将转换逻辑推送至数据库时，无法跟踪数据库中出现的错误。因此，它处理错误的方式与在会话中处理转换的方式不同。集成服务运行已配置完整下推优化的会话并出现错误时，数据库会处理这些错误。数据库处理错误时，集成服务无法将拒绝行写入拒绝文件。

日志记录

集成服务将转换逻辑推送至数据库时，无法跟踪数据库服务器内部出现的所有事件。集成服务可跟踪的统计信息取决于下推优化的类型。将转换逻辑推送至数据库时，集成服务会生成存在以下差异的会话日志：

- 会话日志不包含由数据库处理的转换的详细信息。
- 如果会话已配置完整下推优化，则会话日志不包含线程繁忙百分比。
- 如果集成服务使用完整下推优化并将所有转换逻辑推送至数据库，则会话日志不包含从源中读取的行数。
- 如果集成服务使用源端下推优化，则会话日志包含从已优化源中读取的行数。

恢复

如果为会话配置了完整下推优化，并且会话失败，则集成服务无法执行增量恢复，因为数据库会处理转换。相反，数据库会回滚这些事务。如果数据库服务器失败，则其会在重新启动时回滚事务。如果集成服务失败，则数据库服务器会回滚事务。

如果集成服务在数据库中创建临时序列对象或视图时（即在处理任何行之前）出现此故障，则集成服务会在数据库上再次运行生成的 SQL。

如果在数据库处理所有行之前出现此故障，则集成服务会执行以下任务：

1. 如果适用，集成服务会在数据库中停止并重新创建临时视图或序列对象以确保不会产生重复值。
2. 集成服务会在数据库上再次运行生成的 SQL。

如果集成服务从数据库中删除临时视图或序列对象时（即在处理所有行之后）出现此故障，则集成服务会尝试再次删除临时对象。

使用渐变维度

可以将类型 1 和类型 3 渐变维度逻辑推送至数据库。映射中的渐变维度逻辑可以包含多个转换。每个转换的规则和准则决定了可以向数据库推送渐变维度逻辑的多少。

对集成服务进行配置以将渐变维度转换逻辑推送至数据库时，请遵循以下规则和准则：

- 可以将类型 1 和类型 3 渐变维度映射中包含的转换推送至 Oracle 或 IBM DB2 数据库。
- 源数据不得包含重复行。如果对同一行进行多次更新，数据库可能会死锁。
- 必须使用 Slowly Changing Dimensions Wizard 8.5 或更高版本来创建渐变维度映射。如果渐变维度逻辑是由以前版本的 Slowly Changing Dimensions Wizard 创建的，将无法将其推送至数据库。

使用序列和视图

要将转换逻辑推送至数据库，集成服务可以在数据库中创建临时序列或视图。数据库事务完成后，集成服务可删除创建用于下推优化的序列和视图对象。

序列

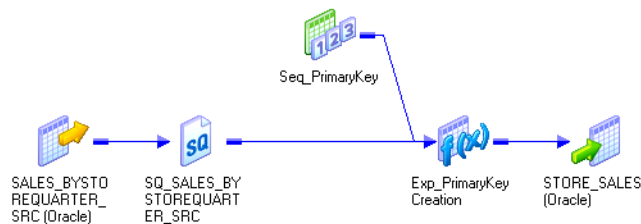
要将序列生成器转换逻辑推送至数据库，必须将会话配置为使用序列执行下推优化。

如果将会话配置为将序列生成器转换逻辑推送至数据库，则集成服务将完成以下任务：

1. **在数据库中创建序列对象。** 集成服务根据序列生成器转换逻辑在数据库中创建序列对象。集成服务将为每个序列对象创建一个唯一名称。要创建唯一的序列对象名称，它会将前缀 PM_S 添加到由哈希函数生成的值中。
2. **生成 SQL 查询，并针对数据库执行此查询。** 集成服务将生成并执行 SQL 查询，以将序列生成器转换逻辑推送至数据库。
3. **从数据库中删除序列对象。** 事务完成时，集成服务会删除在数据库中创建的序列对象。

序列创建示例

创建使用序列生成器转换为关系目标生成主键的以下映射：



当集成服务将转换逻辑推送至数据库时，会执行以下 SQL 语句以在源数据库中创建序列对象：

```
CREATE SEQUENCE PM_S6UHW420GXTY7NICHYIOSRMC5XQ START WITH 1 INCREMENT BY 1 MINVALUE 0 MAXVALUE 9223372036854775807 NOCYCLE CACHE 9223372036854775807
```

在集成服务创建序列对象后，集成服务会执行 SQL 查询以处理映射中包含的转换逻辑：

```
INSERT INTO STORE_SALES(PRIMARYKEY, QUARTER, SALES, STORE_ID) SELECT  
CAST(PM_S6UHW420GXTY7NICHYIOSRMC5XQ.NEXTVAL AS FLOAT), CAST(CAST(SALES_BYSTOREQUARTER_SRC.QUARTER AS  
FLOAT) AS VARCHAR2(10)), CAST(CAST(SALES_BYSTOREQUARTER_SRC.SALES AS NUMBER(10, 2)) AS NUMBER(25, 2)),  
CAST(SALES_BYSTOREQUARTER_SRC.STORE_ID AS NUMBER(0, 0)) FROM SALES_BYSTOREQUARTER_SRC
```

在会话完成后，集成服务会从数据库中删除序列对象。如果会话失败，集成服务会在执行恢复任务之前删除并重新创建序列对象。

视图

必须为会话配置视图下推优化才能启用集成服务以在数据库中创建视图对象。

在以下条件下，集成服务会创建视图对象：

- 为配置了 SQL 替代的源限定符或查找转换配置下推优化。
- 为配置了筛选器的查找转换配置下推优化。
- 为未连接的查找转换配置下推优化。

当集成服务将源限定符或查找转换推送到数据库时，集成服务会根据转换定义创建视图。例如，当集成服务根据具有筛选器的查找转换创建视图时，集成服务会创建仅包含未筛选行的视图。当集成服务将具有 SQL 替代的查找转换推送到数据库时，集成服务会根据所有查找端口而不仅仅是计划查找端口来创建视图。

集成服务不解析或验证 SQL 替代。如果配置会话以将具有 SQL 替代的源限定符或查找转换推送到数据库，请在运行该会话之前针对该数据库测试 SQL 替代。

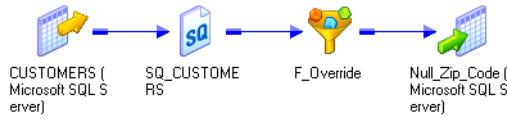
如果将源限定符转换逻辑推送到具有临时视图的 Teradata，则 Teradata 中的数据字典可能会导致 SQL 语句失败。SQL 语句由于使用多个下推优化会话的环境中视图的动态创建和删除而失败。如果源限定符转换包含源筛选器、用户定义的联接或 SQL 替代，则可以下推优化到 Teradata 禁用临时视图的创建。集成服务会创建派生表代替视图。

如果为会话配置视图下推优化，则集成服务会完成以下任务：

1. **在数据库中创建视图。** 集成服务会根据源限定符或查找转换中的查找筛选器、未连接的查找或 SQL 替代在数据库中创建视图。为创建唯一视图名称，集成服务会将前缀 PM_V 添加到由哈希函数生成的值。
2. **针对视图执行 SQL 查询。** 在集成服务创建视图对象后，集成服务会针对在数据库中创建的视图执行 SQL 查询以将转换逻辑推送到源。
3. **从数据库中删除视图。** 转换完成后，集成服务会删除其创建的视图。

查看创建示例

您可以创建以下映射以在客户数据库中搜索 94117 邮政编码：



您希望搜索返回其姓名与姓名 Johnson 的变体（包括 Johnsen、Jonssen 和 Jonson 等姓名）匹配的客户。要执行姓名匹配，需要为源限定符转换输入以下 SQL 替代：

```
SELECT CUSTOMERS.CUSTOMER_ID, CUSTOMERS.COMPANY, CUSTOMERS.FIRST_NAME, CUSTOMERS.LAST_NAME,
CUSTOMERS.ADDRESS1, CUSTOMERS.ADDRESS2, CUSTOMERS.CITY, CUSTOMERS.STATE, CUSTOMERS.POSTAL_CODE,
CUSTOMERS.PHONE, CUSTOMERS.EMAIL FROM CUSTOMERS WHERE CUSTOMERS.LAST_NAME LIKE 'John%' OR
CUSTOMERS.LAST_NAME LIKE 'Jon%'
```

集成服务将此会话的转换逻辑推送到数据库时，会执行以下 SQL 语句以在源数据库中创建视图：

```
CREATE VIEW PM_V4RZRW5GWCKUEWH35RKDMDPRNXI (CUSTOMER_ID, COMPANY, FIRST_NAME, LAST_NAME, ADDRESS1,
ADDRESS2, CITY, STATE, POSTAL_CODE, PHONE, EMAIL) AS SELECT CUSTOMERS.CUSTOMER_ID, CUSTOMERS.COMPANY,
CUSTOMERS.FIRST_NAME, CUSTOMERS.LAST_NAME, CUSTOMERS.ADDRESS1, CUSTOMERS.ADDRESS2, CUSTOMERS.CITY,
CUSTOMERS.STATE, CUSTOMERS.POSTAL_CODE, CUSTOMERS.PHONE, CUSTOMERS.EMAIL FROM CUSTOMERS WHERE
CUSTOMERS.LAST_NAME LIKE 'John%' OR CUSTOMERS.LAST_NAME LIKE 'Jon%'
```

集成服务创建视图后，会执行 SQL 查询以在映射中执行转换逻辑：

```
SELECT PM_V4RZRW5GWCKUEWH35RKDMDPRNXI.CUSTOMER_ID, PM_V4RZRW5GWCKUEWH35RKDMDPRNXI.COMPANY,
PM_V4RZRW5GWCKUEWH35RKDMDPRNXI.FIRST_NAME, PM_V4RZRW5GWCKUEWH35RKDMDPRNXI.LAST_NAME,
PM_V4RZRW5GWCKUEWH35RKDMDPRNXI.ADDRESS1, PM_V4RZRW5GWCKUEWH35RKDMDPRNXI.ADDRESS2,
PM_V4RZRW5GWCKUEWH35RKDMDPRNXI.CITY, PM_V4RZRW5GWCKUEWH35RKDMDPRNXI.STATE,
PM_V4RZRW5GWCKUEWH35RKDMDPRNXI.POSTAL_CODE, PM_V4RZRW5GWCKUEWH35RKDMDPRNXI.PHONE,
PM_V4RZRW5GWCKUEWH35RKDMDPRNXI.EMAIL FROM PM_V4RZRW5GWCKUEWH35RKDMDPRNXI WHERE
(PM_V4RZRW5GWCKUEWH35RKDMDPRNXI.POSTAL_CODE = 94117)
```

会话完成后，集成服务会从数据库中删除该视图。如果会话失败，集成服务会删除并重新创建该视图，然后执行恢复任务。

孤立序列和视图故障排除

如果集成服务、会话或连接失败，则集成服务可能无法从数据库中删除序列或视图对象。在这种情况下，请从数据库中手动删除这些对象。

注意：数据库中的孤立序列和视图对象不影响性能。

完成以下任务可从数据库中删除孤立序列或视图对象：

1. **标识数据库中的孤立对象。**您可以根据数据库上的会话日志或查询标识孤立对象。分析会话日志以确定来自会话运行的孤立对象。在给定时间运行数据库查询以确定数据库中的所有孤立对象。
2. **从数据库中删除孤立对象。**可以执行 SQL 语句以删除您标识的孤立对象。

使用会话日志标识孤立对象

在集成服务创建和删除视图或序列对象时，集成服务会写入事件日志。如果集成服务、会话或连接在会话正在运行时失败，可以检查会话日志以确定会话期间未删除的序列或视图对象。

例如，如果集成服务删除视图 PM_V4RZRW，则会话日志将显示以下消息：

映射 > TM_6356 开始源 [客户] 的下推清理 SQL。：(2006 年 2 月 14 日星期二 13:23:46)

映射 > TM_6358 执行源的下推清理 SQL：删除视图 PM_V4RZRW

映射 > TM_6360 已成功完成源 [客户] 的下推清理 SQL。：(2006 年 2 月 14 日星期二 13:23:46)]

使用 SQL 查询标识孤立对象

如果集成服务不删除序列对象或视图对象，则可以对数据库执行 SQL 查询，以标识由集成服务创建的所有孤立的序列对象或视图对象。如果集成服务对同一数据库帐户运行多个会话或多个集成服务写入，则 SQL 查询会从运行且未删除序列对象或视图对象的每个会话中返回所有孤立的对象。

集成服务在数据库中创建序列对象或视图对象时，会将前缀 PM_S 添加到序列对象的名称中，并将 PM_V 添加到视图对象的名称中。可以根据用于标识这些对象的前缀来搜索对象。

以下查询显示了搜索由集成服务创建的序列对象的语法：

IBM DB2：

```
SELECT SEQNAME FROM SYSCAT.SEQUENCES
WHERE SEQSHEMA = CURRENT SCHEMA
AND SEQNAME LIKE 'PM\_S%' ESCAPE '\'
```

Oracle：

```
SELECT SEQUENCE_NAME FROM USER_SEQUENCES
WHERE SEQUENCE_NAME LIKE 'PM\_S%' ESCAPE '\'
```

以下查询显示了搜索由集成服务创建的视图对象的语法：

IBM DB2：

```
SELECT VIEWNAME FROM SYSCAT.VIEWS
WHERE VIEWSHEMA = CURRENT SCHEMA
AND VIEW_NAME LIKE 'PM\_V%' ESCAPE '\'
```

Oracle：

```
SELECT VIEW_NAME FROM USER_VIEWS
WHERE VIEW_NAME LIKE 'PM\_V%' ESCAPE '\'
```

Microsoft SQL Server 或 Sybase ASE：

```
SELECT NAME FROM SYSOBJECTS
WHERE TYPE= 'V' AND NAME LIKE 'PM\_V%' ESCAPE '\'
```

Teradata：

```
SELECT TableName FROM DBC.Tables
WHERE CreatorName = USER
AND TableKind = 'V'
AND TableName LIKE 'PM\_V%' ESCAPE '\'
```

删除孤立对象

获取由集成服务创建的序列对象和视图对象的列表后，执行 SQL DROP 语句以从数据库中删除序列对象或视图对象。

以下查询显示了在任何数据库上删除由集成服务创建的序列对象的语法：

```
DROP SEQUENCE <sequence name>
```

以下查询显示了在任何数据库上删除由集成服务创建的视图对象的语法：

```
DROP VIEW <view name>
```

使用 \$\$PushdownConfig 映射参数

您可能需要在不同的时间使用源端、目标端或完整下推优化，具体取决于数据库工作负载。例如，在一天的高峰时段使用源端或目标端下推优化，但从午夜到凌晨 2 点数据库活动不频繁时使用完整下推优化。

要在不同时间使用不同下推优化配置，请使用 \$\$PushdownConfig 映射参数。该参数可允许您使用不同类型的下推优化运行会话。\$\$PushdownConfig 参数中的设置会替代会话属性中的下推优化设置。

完成以下步骤以配置映射参数：

1. 在 Mapping Designer 中创建 \$\$PushdownConfig。
2. 在 Mapping Designer 中添加 \$\$PushdownConfig 映射参数时，请使用以下值：

字段	值
名称	\$\$PushdownConfig
类型	参数
数据类型	字符串
精度或小数位数	20
汇总	不适用
初始值	无
描述	可选

3. 配置会话时，为下推优化属性选择 \$\$PushdownConfig。
4. 在参数文件中定义该参数。
5. 在参数文件中为 \$\$PushdownConfig 输入以下值之一：

值	说明
无	集成服务会处理会话的所有转换逻辑。
源 [序列视图连接]	集成服务会尽可能多地将转换逻辑推送至源数据库。
目标 [序列视图连接]	集成服务会尽可能多地将转换逻辑推送至目标数据库。
完整 [序列视图连接]	集成服务会尽可能多地将转换逻辑推送至源和目标数据库。集成服务将处理任何无法推送至数据库的转换逻辑。

或者，指定以下一个或多个选项：

- **序列。**允许集成服务在数据库中创建序列对象。
- **视图。**允许集成服务在数据库中创建视图对象。
- **连接。**指示活动数据库的数据库用户是否具有空闲数据库上的读取权限，需要该权限才能将转换逻辑推送至活动数据库。

例如，输入“Full View Conn”以使用完整下推优化，允许在活动数据库中创建视图对象，以及指示活动数据库是否具有空闲数据库上的读取权限。

为会话配置下推优化

在会话属性中为会话配置下推优化。但是，可能需要编辑转换、映射或会话配置才能将多个转换逻辑推送至数据库。使用“下推优化查看器”检查可推送至数据库的转换。

下推选项

可以在会话属性中配置以下下推优化选项：

- **下推优化。**下推优化的类型。如果使用 `$$PushdownConfig` 映射参数，请确保已配置该映射参数并且已在参数文件中为其定义值。
- **允许临时视图下推。**允许 PowerCenter 集成服务在其将会话推送至数据库时在数据库中创建临时视图对象。如果会话包含源限定符转换或查找转换中的 SQL 替代、已筛选的查找或未连接的查找，PowerCenter 集成服务会在数据库中创建视图。

如果使用包含源筛选器、用户定义的联接或 SQL 替代的 Teradata 源和源限定符转换，则无需允许临时下推视图。如果将源限定符转换逻辑推送到具有临时视图的 Teradata，则 Teradata 中的数据字典可能会导致 SQL 语句失败。SQL 语句由于使用多个下推优化会话的环境中视图的动态创建和删除而失败。

- **允许临时序列下推。**允许 PowerCenter 集成服务在数据库中创建临时序列对象。如果会话包含序列生成器转换，则 PowerCenter 集成服务必须在数据库中创建序列对象。
- **对用户不兼容的连接允许下推。**指示活动数据库的数据库用户是否具有空闲数据库上的读取权限。如果指示活动数据库的数据库用户具有空闲数据库上的读取权限，而其没有，则会话将失败。如果未指示活动数据库的数据库用户具有空闲数据库上的读取权限，则 PowerCenter 集成服务不会将转换逻辑推送至数据库。

使用“下推优化查看器”确定是否需要编辑映射、转换或会话配置以将多个转换逻辑推送至数据库。“下推优化查看器”可指示是否可以使用源端、目标端或完整下推优化将转换逻辑推送至数据库。如果可以将转换逻辑推送至数据库，“下推优化查看器”会列出可推送至数据库的所有转换。

还可以在“下推优化查看器”中选择下推选项或下推组以查看为指定选择生成的相应 SQL 语句。

注意：选择下推选项或下推组时，请勿更改下推配置。要更改配置，必须更新会话属性中的下推选项。

分区

如果分区类型为传递分区或键范围分区，则可以将具有多个分区的会话推送至数据库。

传递分区的下推优化

为使用传递分区的会话配置下推优化时，数据库会处理数据，而不会在分区之间重新分发。单个分区中的所有行在错过传递分区点后将保留在该分区。

必须为传递分区配置所有分区点才能将所有转换逻辑推送至数据库。例如，会话有四个分区点。将前三个分区点配置为传递分区，将最后一个分区点配置为哈希自动键分区。除了最后一个分区点及其之后的转换之外，集成服务会将所有转换逻辑推送至数据库。集成服务会在最后一个分区点及其之后处理转换。

键范围分区的下推优化

为在源限定符转换时使用键范围分区的会话配置下推优化时，集成服务会将所有行合并到第一个分区并为每个后续分区传递空数据。集成服务为每个分区创建一个 SQL 语句。如果集成服务仅将部分转换逻辑推送至数据库，则在其运行会话时不会在分区之间重新分发。

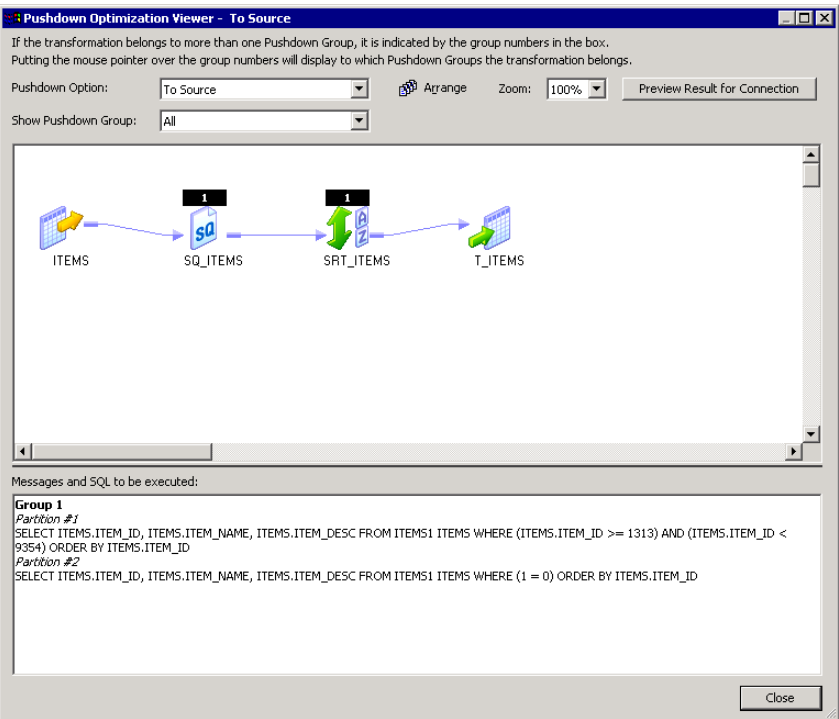
会话必须满足以下条件才能使集成服务将所有转换逻辑推送至数据库：

- 每个分区的结束键范围必须等于下一个分区的开始范围，才能将所有行合并到第一个分区。结束键范围不能与下一个分区重叠。例如，如果第一个分区的结束范围是 3386，则第二个分区的开始范围必须为 3386。

- 必须在源限定符转换配置分区点以使用键范围分区，并配置所有后续分区点以使用哈希自动键或传递分区。

具有多个分区的会话的下推优化示例

下图显示的映射中包含使用哈希自动键分区的排序器转换：



第一个键范围是 1313 - 3340，第二个键范围是 3340 - 9354。SQL 语句会将所有数据合并到第一个分区：

```
SELECT ITEMS.ITEM_ID, ITEMS.ITEM_NAME, ITEMS.ITEM_DESC FROM ITEMS1 ITEMS WHERE (ITEMS.ITEM_ID >= 1313)
AND (ITEMS.ITEM_ID < 9354) ORDER BY ITEMS.ITEM_ID
```

SQL 语句选择 1313 到 9354 的项目，包括键范围中的所有值，并将两个分区中的数据都合并到第一个分区。

第二个分区的 SQL 语句会传递空值：

```
SELECT ITEMS.ITEM_ID, ITEMS.ITEM_NAME, ITEMS.ITEM_DESC FROM ITEMS1 ITEMS WHERE (1 = 0) ORDER BY
ITEMS.ITEM_ID
```

具有多个分区的会话的规则和准则

配置集成服务以将具有多个分区的会话推送至数据库时，请遵循以下规则和准则。

在以下情况下，集成会将具有多个分区的会话推送至数据库：

- 如果会话在分区点（在源限定符转换时）以及所有后续分区点使用传递分区，则集成服务会使用源端、目标端或完整下推优化将转换逻辑推送至数据库。
- 如果会话在源限定符转换时使用键范围分区，并且在下游分区点包含哈希自动键或传递分区，则集成服务会使用源端或完整下推优化将转换逻辑推送至数据库。

如果下推优化将某个转换的多个分区中的数据合并到第一个分区，并且集成服务处理下游转换的转换逻辑，则集成服务无法在下游转换中在分区之间重新分发。它会继续将行传递到第一个分区并传递其他分区中的空数据。

目标加载规则

目标加载规则可影响您是否可以将会话推送至数据库。

下表显示了不同目标加载选项的下推优化：

目标选项	源	目标	完整
插入	X	X	X
删除	X	X	X
更新为更新	X	X	X
更新为插入	X	X	X
更新否则插入	X	X	是/否

配置集成服务以将目标加载逻辑推送至数据库时，请遵循以下规则和准则：

- 如果在使用完整下推优化时未实现性能提升并将源行视为删除或更新，请使用源端下推优化。
- 如果会话包含联合转换，并且集成服务将转换逻辑推送至 Sybase 数据库，则无法使用完整下推优化并将源行视为删除或更新。

查看下推组

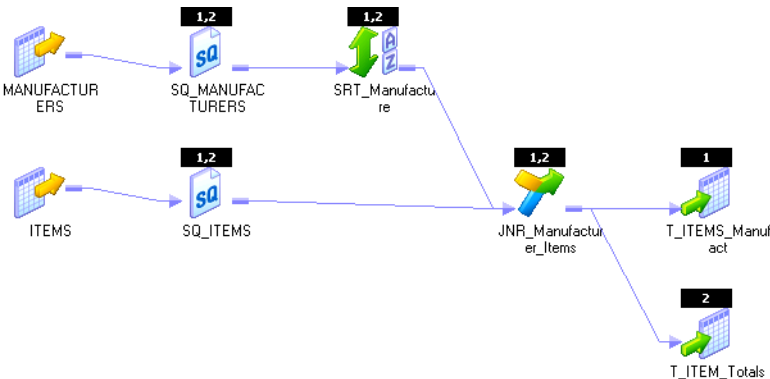
为会话配置下推优化时，集成服务会基于转换逻辑生成 SQL 语句。按照一个 SQL 语句处理的转换组也称为下推组。

将转换逻辑推送至数据库时，集成服务可能会根据管道数量、源、目标和使用的下推优化类型创建多个下推组。如果会话具有多个分区，则集成服务会为组中的每个分区执行 SQL 语句。如果联接管道，则会将每个管道中的转换合并到一个下推组中。如果同一转换属于已推送至两个或多个目标的转换逻辑，则该转换属于每个目标的下推组。

可以使用“下推优化查看器”查看下推组。在“下推优化查看器”中查看下推组时，可以确定可推送至数据库的转换以及集成服务可处理的转换。“下推优化查看器”还会显示可用于确定如何编辑转换或映射以将多个转换逻辑推送至数据库的消息。如果使用映射变量或将会话配置为在网格上运行，则“下推优化查看器”无法显示会话中运行的 SQL。

查看生成的 SQL 时，临时视图和序列对象的名称与会话期间的视图和序列对象的名称不同。集成服务使用哈希函数为其生成的每个序列和视图对象创建唯一名称。

下图显示了“下推优化查看器”中显示的映射。它包含可推送至源和目标数据库的两个下推组：



管道 1 和管道 2 源自不同源，并且包含对下推优化有效的转换。集成服务为每个目标创建下推组，并为每个下推组生成 SQL 语句。因为两个管道已联接，因此直至并包含联接器转换的转换属于这两个管道，并且会包含在两个下推组中。

要查看下推组，请打开“下推优化查看器”。“下推优化查看器”可预览下推组和集成服务在运行时生成的 SQL 语句。

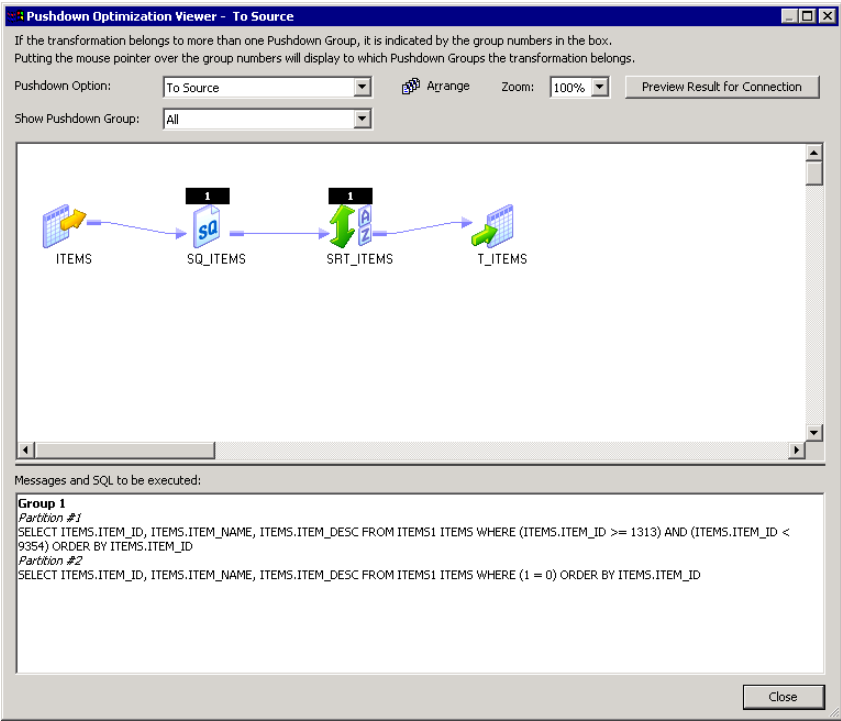
要查看下推组，请执行以下操作：

1. 在 Workflow Manager 中，打开配置了下推优化的会话。
2. 在“映射”选项上，选择左侧窗格中的“下推优化”或右侧窗格中的“查看下推优化”。

“下推优化查看器”会显示下推组和构成每个组的转换。如果在管道中配置了多个分区，则它会显示每个分区的 SQL 语句。可以查看为每个下推组生成的消息和 SQL 语句以及下推选项。下推选项包括“无”、“到源”、“到目标”、“完整”和 \$\$PushdownConfig。

下图显示的映射中包含一个可推送至源数据库的具有两个分区的管道：

图 1. 下推优化查看器



3. 选择“下推优化查看器”中的下推选项可预览 SQL 语句。
查看器中的下推选项不影响运行时进行的优化。要为会话更改下推优化，请编辑会话属性。
4. 如果将会话配置为使用连接变量，请单击“预览连接结果”以选择要预览的连接值。
如果会话使用连接变量，则必须在每次打开“下推优化查看器”时选择连接值。Workflow Manager 不会保存您选择的值，并且集成服务在运行时不使用此值。
如果 SQL 替代包含 \$\$\$SessStartTime 变量，则在您预览下推优化时，“下推优化查看器”不会展开此变量。

第 5 章

下推优化和转换

本章包括以下主题：

- [下推优化和转换概览, 93](#)
- [汇总器转换, 94](#)
- [表达式转换, 95](#)
- [筛选器转换, 96](#)
- [联接器转换, 97](#)
- [查找转换, 98](#)
- [路由器转换, 100](#)
- [序列生成器转换, 100](#)
- [排序器转换, 101](#)
- [源限定符转换, 102](#)
- [目标, 104](#)
- [联合转换, 105](#)
- [更新策略转换, 105](#)

下推优化和转换概览

配置下推优化时，集成服务会尝试将每个转换推送至数据库。下列标准会影响集成服务是否可以将转换推送至数据库：

- 转换类型
- 映射中转换的位置
- 转换的映射和会话配置
- 转换中包含的表达式

该标准可能还会影响集成服务可以执行的下推优化的类型以及可以向其中推送转换的数据库的类型。

集成服务可以将以下转换逻辑推送至数据库：

- 汇总器
- 表达式
- 筛选器
- 联接器

- 查找
- 路由器
- 序列生成器
- 排序器
- 源限定符
- 目标
- 联合
- 更新策略

常规下推限制

您可推送至数据库的转换逻辑量取决于数据库、转换逻辑、映射和会话配置。集成服务将处理所有无法推送至数据库的转换逻辑。

如果以下任一转换或映射条件为 true，则集成服务会处理该逻辑而不是将其推送至数据库：

- 转换逻辑更新映射变量并将其保存至存储库数据库。
- 转换包含变量端口。
- 转换满足以下所有条件：
 - 不是排序器转换、联合转换或目标。
 - 推送至 Microsoft SQL Server、Sybase 或 Teradata。
 - 是排序器转换中的下游，即是联合转换中的下游或包含相异排序。
- 将会话配置为替代输入或输出端口的默认值。
- 数据库没有在转换中的表达式中使用的等价运算符、变量或函数。
- 映射包含过多分支。对管道进行分支时，表示映射逻辑所需的 SQL 语句变得更为复杂。集成服务无法为以下映射生成 SQL 查询：包含的分支数超出 64 个双向分支、43 个三向分支或 32 个四向分支的映射。如果分支数量超出这些限制，则集成服务会处理下游转换。

如果以下任一会话属性为 true，则集成服务会处理该逻辑而不是将其推送至数据库：

- 会话为调试会话。
- 将会话配置为日志行错误。

如果上述所有条件均为 false，可以查看各个转换和数据库的下推规则。

汇总器转换

下表显示了您要向其推送汇总器转换的每个数据库的下推类型：

数据库	下推类型
Amazon Redshift	源端，完整
Greenplum	源端，完整

数据库	下推类型
IBM DB2	源端，完整
Microsoft SQL Server	源端，完整
Netezza	源端，完整
Oracle	源端，完整
PostgreSQL	源端，完整
SAP HANA	源端，目标端，完整
Snowflake	源端，完整
Sybase ASE	源端，完整
Teradata	源端，完整
Vertica	源端，完整
Microsoft Azure SQL 数据仓库	源端，完整

如果以下任一条件为 **true**，则集成服务会处理汇总器转换：

- 会话和映射配置为增量汇总。
- 转换包含嵌套的汇总函数。
- 转换包含汇总表达式中的条件子句。
- 转换在任何端口表达式中均使用 FIRST()、LAST()、MEDIAN() 或 PERCENTILE() 函数。
- 输出端口不是汇总端口或分组依据端口的一部分。
- 转换已推送至 Microsoft SQL Server、Sybase 或 Teradata，并且是排序器转换的下游。

表达式转换

下表显示了您要向其推送表达式转换的每个数据库的下推类型：

数据库	下推类型
Amazon Redshift	源端，目标端，完整
Greenplum	源端，目标端，完整
IBM DB2	源端，目标端，完整
Microsoft SQL Server	源端，目标端，完整
Netezza	源端，目标端，完整

数据库	下推类型
Oracle	源端，目标端，完整
PostgreSQL	源端，完整
Snowflake	源端，完整
SAP HANA	源端，目标端，完整
Sybase ASE	源端，目标端，完整
Teradata	源端，目标端，完整
Vertica	源端，目标端，完整
Microsoft Azure SQL 数据仓库	源端，目标端，完整

如果转换调用未连接的存储过程，则集成服务会处理表达式转换。

筛选器转换

将筛选器转换推送至数据库以减少 PowerCenter 集成服务处理的数据量。如果筛选器表达式无法推送至数据库，则 PowerCenter 集成服务会处理筛选器转换。例如，如果筛选器表达式包含一个无法推送至数据库的运算符，则集成服务不会将筛选器表达式推送至数据库。

下表显示了可将筛选器转换推送到的每个数据库的下推类型：

数据库	下推类型
Amazon Redshift	源端，完整
Greenplum	源端，完整
IBM DB2	源端，完整
Microsoft SQL Server	源端，完整
Netezza	源端，完整
Oracle	源端，完整
PostgreSQL	源端，完整
SAP HANA	源端，目标端，完整
Snowflake	源端，完整
Sybase ASE	源端，完整
Teradata	源端，完整

数据库	下推类型
Vertica	源端，完整
Microsoft Azure SQL 数据仓库	源端，完整

联接器转换

将联接器转换推送至数据库以优化数据库中的索引和统计信息的使用。

下表显示了您要向其推送联接器转换的每个数据库的下推类型：

数据库	下推类型
Amazon Redshift	源端，完整
Greenplum	源端，完整
IBM DB2	源端，完整
Microsoft SQL Server	源端，完整
Netezza	源端，完整
Oracle	源端，完整
PostgreSQL	源端，完整
SAP HANA	源端，目标端，完整
Snowflake	源端，完整
Sybase ASE	源端，完整
Teradata	源端，完整
Vertica	源端，完整
Microsoft Azure SQL 数据仓库	源端，完整

如果以下任一条件为 true，则集成服务会处理联接器转换：

- 集成服务无法将联接器转换的主管道和详细管道推送至数据库。
- 联接条件基于使用二进制数据类型的列。
- 联接器转换的传入组源自不同关系数据库管理系统上的数据库。
- 会话已配置为将所有源行标记为更新，并已配置下推优化到 Teradata。
- 转换已配置有外部联接，并且主源或详细源为多表联接。集成服务无法生成 SQL 以代表与多表联接合并的外部联接。
- 转换已配置有完整外部联接，并已配置下推优化到 Sybase。

- 集成服务基于主分支中的转换创建了视图或序列，而主分支和详细分支来自于不同数据库。
- 转换已推送至 Microsoft SQL Server、Sybase 或 Teradata，并且是排序器转换的下游，即汇总器转换的下游。
- 转换是排序器转换的下游，并且转换已推送至 Microsoft SQL Server、Sybase 或 Teradata，而主表和详细表源于相同源限定符转换实例。

查找转换

配置查找转换以进行下推优化时，数据库会对数据库查找表执行查找。如果将查找转换推送至数据库，数据库将为每行引入额外的子查询成本。在 PowerCenter 中启用查找缓存（而不是执行下推优化），以通过大量查找提高映射的性能。

下表显示了可将查找转换推送到的每个数据库的下推类型：

数据库	下推类型
Amazon Redshift	源端，完整
Greenplum	源端，完整
IBM DB2	源端，目标端，完整
Microsoft SQL Server	源端，完整
Netezza	源端，完整
Oracle	源端，目标端，完整
PostgreSQL	源端，完整
SAP HANA	源端，目标端，完整
Snowflake	源端，完整
Sybase ASE	源端，完整
Teradata	源端，完整
ODBC	源端，完整

配置集成服务以将查找转换逻辑推送至数据库时，请遵循以下规则和准则：

- 数据库在处理转换逻辑时不会使用 PowerCenter 缓存。
- 当多个查找转换存在于不同的管道分支中且这些分支在下游合并时，集成服务会在某个管道分支之后处理所有转换。
- 如果会话需要数据类型转换，为目标端下推优化配置的会话将失败。
- 与集成服务不同，Netezza 数据库可能会为单个查找返回多行。
- 如果查找转换包含 SQL 替代、包含筛选器或是未连接的查找转换，请使用视图配置下推优化。

- 当映射包含查找 Netezza、Redshift 或 Snowflake 表，并且设置的查找匹配策略选项不是**使用所有值**时，下推优化会停止于查找转换。设置的查找匹配策略选项不是**使用所有值**或**报告错误**时，将停止所有其他数据库的下推优化。

如果满足以下任一条件，则集成服务将处理查找转换：

- 转换是管道查找。
- 转换使用动态缓存。
- 转换配置为返回第一个、最后一个或任何匹配值。要使用下推优化，必须将查找转换配置为对多个匹配报告错误。
- 转换要求在数据库中创建视图，且提供查找输入的数据库与在其中创建视图的数据库不同。
- 转换已推送至 Microsoft SQL Server、Sybase 或 Teradata，并且是排序器转换的下游，即汇总器转换的下游。
- 会话已配置为将所有源行标记为更新，并已配置下推优化到 Teradata。
- 会话针对源端下推优化配置，且查找表和源表位于不同的关系数据库管理系统中。
- 会话针对目标端下推优化配置，且查找表和目标表位于不同的关系数据库管理系统中。
- 集成服务尝试将转换推送至 Netezza 数据库目标。

未连接的查找转换

配置集成服务以将未连接的查找转换推送至数据库时，请遵循以下规则和准则：

- 如果会话包含多个未连接的查找，则数据库执行速度可能比集成服务慢。因为集成服务会在其每次调用未连接的查找时创建外部联接，因此生成的 SQL 可能会很复杂。使用或不使用下推优化的情况下测试会话以确定哪种会话的性能更佳。
- 为会话配置视图下推优化。

如果以下任一条件为 true，则集成服务会处理未连接的查找转换：

- 查找连接与源连接在下推操作上不兼容。
- 可以配置目标端下推优化。
- 该转换是汇总器转换的下游。
- 该转换是主动转换，并且会在 Netezza 数据库中进行查找。

具有 SQL 替代的查找转换

配置集成服务以将包含 SQL 替代的查找转换推送至数据库时，请遵循以下规则和准则：

- 无法在查找替代中将 ORDER BY 子句附加到 SQL 语句。如果附加 ORDER BY 子句，则会话将失败。
- 验证 SQL 替代是否按端口在查找转换中的显示顺序选择查找转换中的所有端口。
- 如果 SQL 替代中的 SELECT 语句引用数据库序列，则会话将失败。

如果转换在 SQL 替代中包含 Informatica 外部联接语法，则集成服务将使用 SQL 替代处理查找转换。在 SQL 替代中使用 ANSI 外部联接语法，以将转换推送至数据库。

路由器转换

下表显示了您要向其推送路由器转换的每个数据库的下推类型：

数据库	下推类型
Amazon Redshift	源端，完整
Greenplum	源端，完整
IBM DB2	源端，完整
Microsoft SQL Server	源端，完整
Netezza	源端，完整
Oracle	源端，完整
PostgreSQL	源端，完整
SAP HANA	源端，目标端，完整
Snowflake	源端，完整
Sybase ASE	源端，完整
Teradata	源端，完整
ODBC	源端，完整
Microsoft Azure SQL 数据仓库	源端，完整

当所有输出组合并到可推送至源数据库的一个转换中时，可以使用源端下推。

如果路由器表达式无法推送至数据库，则集成服务会处理路由器转换。例如，如果表达式包含无法向数据库推送的运算符，则集成服务不会向数据库推送表达式。

序列生成器转换

下表显示了您要向其推送序列生成器转换的每个数据库的下推类型：

数据库	下推类型
Greenplum	不受支持
IBM DB2	源端，目标端，完整
Microsoft SQL Server	不受支持
Netezza	不受支持

数据库	下推类型
Oracle	源端，目标端，完整
PostgreSQL	不受支持
Sybase	不受支持
Teradata	不受支持
ODBC	不受支持
Microsoft Azure SQL 数据仓库	不受支持

如果以下任一条件为 true，则集成服务会处理序列生成器转换：

- 转换可重用。
- 转换已连接到多个目标。
- 转换连接 CURRVAL 端口。
- 转换向来自已配置为选择相异行的源限定符转换的转换下游提供序列值。
- 集成服务无法将序列生成器转换的所有逻辑推送到数据库。例如，序列生成器转换创建提供给两个管道分支的序列值。配置下推优化后，数据库仅可为一个管道分支创建序列值。集成服务无法将所有序列生成器逻辑推送到数据库，并显示以下消息：

Pushdown optimization stops at the transformation <transformation name> because the upstream Sequence Generator <Sequence Generator transformation name> cannot be pushed entirely to the database.

- 管道在序列生成器转换之前形成分支，然后在序列生成器转换之后重新联接在一起。
- 管道在序列生成器转换之后形成分支，并且无法重新联接在一起。
- 序列值传递到汇总器、筛选器、联接器、排序器或联合转换。
- 在其中创建了序列对象的数据库必须是活动数据库或与活动数据库相同的数据库类型。

如果集成服务在 CASE 表达式中使用序列生成器转换的 NEXTVAL 端口，并且已配置为下推优化到 IBM DB2，则集成服务会处理序列生成器转换。

排序器转换

下表显示了可将排序器转换推送到的每个数据库的下推类型：

数据库	下推类型
Amazon Redshift	源端，完整
Greenplum	源端，完整
IBM DB2	源端，完整
Microsoft SQL Server	源端，完整

数据库	下推类型
Netezza	源端，完整
Oracle	源端，完整
PostgreSQL	源端，完整
Snowflake	源端，完整
Sybase ASE	源端，完整
Teradata	源端，完整
Vertica	源端，完整
Microsoft Azure SQL 数据仓库	不受支持

配置集成服务以将排序器转换逻辑推送至数据库时，请遵循以下规则和准则：

- 当排序器转换配置为相异排序且推送至 Microsoft SQL Server、Sybase 或 Teradata 数据库时，集成服务会将排序器转换推送至数据库并处理下游转换。
- 如果映射包含多个连续的排序器转换且至少一个排序器转换配置为相异排序，则会出现以下结果：
 - 下推优化将相异排序应用于链中的最后一个排序器转换，除非某个排序器转换未计划所有输出端口。
 - 下推优化将相异排序应用于第一个未计划所有输出端口的排序器转换。

如果满足以下任一条件，则集成服务将处理排序器转换：

- 排序器转换是联合转换的下游，且排序器转换中用作排序键的端口未从联合转换计划到排序器转换。
- 排序器转换未计划所有输出端口，且它是映射中多个连续排序器转换中的一个。
- 排序器转换未计划所有输出端口，且以下某个语句为真：
 - 排序器转换配置为相异排序。
 - 该排序器转换的前面紧跟着一个或多个排序器转换，其中某个转换配置为相异排序。

源限定符转换

下表显示了您要向其推送源限定符转换的每个数据库的下推类型：

数据库	下推类型
IBM DB2	源端，完整
Microsoft SQL Server	源端，完整
Netezza	源端，完整
Oracle	源端，完整

数据库	下推类型
PostgreSQL	源端，完整
SAP HANA	源端，目标端，完整
Sybase ASE	源端，完整
Teradata	源端，完整
Vertica	源端，完整
Microsoft Azure SQL 数据仓库	源端，完整

配置 PowerCenter 集成服务以将源限定符转换逻辑推送至数据库时，请遵循以下规则和准则：

- 当序列生成器转换处于下游联合转换、连接器转换或目标中的空闲连接中，并且其他连接为不同数据库类型时，请限定您为源筛选器或用户定义的联接输入的表名。如果在此情况下不限定表名，则 PowerCenter 集成服务不会向数据库推送所有转换。
- 如果为快捷方式对象配置源限定符转换中用户定义的联接并启用下推优化，则会话将失败。

当以下任一条件为 true 时，PowerCenter 集成服务会处理源限定符转换逻辑：

- 转换包含 SQL 替代中的 Informatica 外部联接语法或用户定义的联接。使用 SQL 替代中的 ANSI 外部联接语法以使 PowerCenter 集成服务向数据库推送源限定符转换。
- 已为数据库分区配置源。
- 源为使用 XMLType 数据类型的 Oracle 源。

相关主题：

- [“空闲数据库中的限定名和限定表” 页面上 68](#)

具有 SQL 替代的源限定符转换

为包含具有 SQL 替代的源限定符转换的会话配置下推优化时，请遵循以下规则和准则：

- 自定义 SQL 查询中的 SELECT 语句必须按照端口名称在转换中显示的顺序列出这些端口名称。如果端口顺序不正确，则会话可能失败或输出意外结果。
- 为会话配置视图下推优化。
- 如果 SQL 替代中的 SELECT 语句引用数据库序列，则会话将失败。
- 如果 SQL 替代包含 ORDER BY 子句并且将源限定符转换逻辑推送到 IBM DB2、Microsoft SQL Server、Sybase ASE 或 Teradata 数据库，则会话将失败。
- 如果将源限定符转换配置为选择相异值并包含 SQL 替代，则集成服务会忽略相异配置。
- 如果会话包含多个分区，请为所有分区指定 SQL 替代。
- 如果源限定符转换包含 SQL 替代，则必须为下推优化到 Teradata 禁用临时视图的创建。PowerCenter 集成服务会创建派生表代替视图。
- 在将源数据库上的 SQL 替代推送到数据库之前请测试该 SQL 替代，因为 PowerCenter 不会验证替代 SQL 语法。如果 SQL 语法与源数据库不兼容，则会话将失败。

目标

下表显示了可以向每个数据库中推入目标逻辑的下推类型：

数据库	下推类型
IBM DB2	目标端，完整
Microsoft SQL Server	目标端，完整
Netezza	目标端，完整
Oracle	目标端，完整
SAP HANA	目标端，完整
Sybase ASE	目标端，完整
Teradata	目标端，完整
Vertica	目标端，完整
Microsoft Azure SQL 数据仓库	目标端，完整

配置会话以实现完整下推优化且以下其中一项条件为真时，集成服务会处理目标逻辑：

- 目标包括目标更新替代。
- 配置会话以进行基于约束的加载，且目标加载顺序组包含多个目标。
- 会话使用外部加载程序。
- 在空闲数据库中创建视图或序列生成器。

如果配置完整下推优化，但目标和源连接不兼容，则集成服务无法将所有转换逻辑推送至一个数据库。相反，如果彼此兼容，则集成服务会尽可能多地将转换逻辑推送至源数据库，并将所有剩余的转换逻辑推送至目标数据库。

配置会话以实现目标端下推优化且以下任何一项条件为真时，集成服务会处理目标逻辑：

- 目标包括目标更新替代。
- 配置目标以进行数据库分区。
- 配置会话以实现批量加载，且目标为 IBM DB2、Microsoft SQL Server、Oracle 或 Sybase ASE。
- 会话使用外部加载程序。将源端下推优化与外部加载程序结合使用可允许集成服务将转换逻辑推送至源数据库。

联合转换

下表显示了可将联合转换推送到的每个数据库的下推类型：

数据库	下推类型
Amazon Redshift	源端，完整
Greenplum	源端，完整
IBM DB2	源端，完整
Microsoft SQL Server	源端，完整
Netezza	源端，完整
Oracle	源端，完整
PostgreSQL	源端，完整
SAP HANA	源端，目标端，完整
Snowflake	源端，完整
Sybase ASE	源端，完整
Teradata	源端，完整
Vertica	源端，完整
Microsoft Azure SQL 数据仓库	源端，完整

如果满足以下任一条件，则集成服务将处理联合转换逻辑：

- 集成服务无法将所有输入组推送至源数据库。
- 输入组并不源自同一关系数据库管理系统。
- 联合转换的某个输入管道包含一个相异联合或排序器。
- 该转换是需要在数据库中创建视图或序列生成器的转换的下游，且这些连接位于不同数据库上。

更新策略转换

下表显示了您要向其推送更新策略转换的每个数据库的下推类型：

数据库	下推类型
Amazon Redshift	完整
Greenplum	目标端

数据库	下推类型
IBM DB2	完整
Microsoft SQL Server	完整
Netezza	完整
Oracle	完整
PostgreSQL	源端，完整
SAP HANA	源端，目标端，完整
Snowflake	源端，完整
Sybase ASE	完整
Teradata	完整
Vertica	完整
Microsoft Azure SQL 数据仓库	完整

配置集成服务以将更新策略转换逻辑推送至数据库时，请遵循以下规则和准则：

- 使用更新操作为更新策略转换生成的 SQL 可以很复杂。使用或不使用下推优化的情况下运行会话以确定哪种配置更快。
- 如果对同一行执行多个操作，则集成服务和数据库可以不同的方式处理这些操作。要确保向数据库推送新行时不会删除或更新这些新行，将按以下顺序处理源行：删除事务，更新事务，然后再插入事务。
- 如果转换包含多个插入、更新或删除操作，则集成服务会连续地生成和运行插入、更新和删除 SQL 语句。即使不需要这三种语句，集成服务也会运行它们。这可能会降低性能。
- 使用完整下推优化时，集成服务会忽略拒绝的行。不会将拒绝行写入拒绝文件。

如果以下任一条件为 true，则集成服务会处理更新策略转换：

- 如果集成服务无法向数据库推送更新策略表达式。例如，如果表达式包含无法向数据库推送的运算符，则集成服务不会向数据库推送表达式。
- 转换使用除了插入操作之外的操作，并且集成服务无法向数据库推送所有转换逻辑。
- 更新策略表达式返回不是数值且不是布尔值的值。

第 6 章

实时处理

本章包括以下主题：

- [实时处理概览, 107](#)
- [了解实时数据, 108](#)
- [配置实时会话, 110](#)
- [终止条件, 111](#)
- [刷新延迟, 112](#)
- [提交类型, 112](#)
- [消息恢复, 113](#)
- [恢复文件, 114](#)
- [恢复表, 116](#)
- [恢复队列和恢复主题, 117](#)
- [恢复忽略列表, 117](#)
- [停止实时会话, 118](#)
- [重新启动和恢复实时会话, 118](#)
- [实时会话的规则和准则, 119](#)
- [消息恢复的规则和准则, 119](#)
- [实时处理示例, 120](#)
- [PowerCenter 实时产品, 121](#)

实时处理概览

实时处理行为取决于实时源。异常项在本章或相应产品文档中进行介绍。

您可以使用 PowerCenter 实时处理数据。实时处理是指按需处理来自实时源的数据。实时会话连续读取和处理数据并将其写入目标。除非您为实时处理配置了会话，否则会话默认按计划的间隔读取和写入批量数据。

数据必须源自实时源，才能实时处理数据。实时源包括 JMS、WebSphere MQ、TIBCO、webMethods、MSMQ、SAP、Web 服务和 PowerExchange。您可能希望对需要直接访问动态数据（例如财务数据）的进程使用实时处理。

要了解使用 PowerCenter 完成的实时处理，您需要熟悉以下概念：

- **实时数据。**实时数据包括消息和消息队列、Web 服务消息以及来自 PowerExchange 更改数据捕获源的更改。实时数据源自实时源。

- **实时会话。**实时会话是指处理实时源数据的会话。如果集成服务根据刷新延迟配置生成实时刷新，并且所有转换都将刷新传播到目标，则会话为实时会话。延迟是指从源中的源数据更改时起到会话将数据写入目标时止的时间段。
- **实时属性。**实时属性决定集成服务处理数据并将数据提交至目标的时间。
 - **终止条件。**如果不希望会话持续运行，终止条件将决定集成服务何时停止从源读取数据并结束会话。
 - **刷新延迟。**刷新延迟可确定集成服务刷新来自源的实时数据的频率。
 - **提交类型。**提交类型决定集成服务何时将实时数据提交到目标。
- **消息恢复。**如果实时会话失败，您可以恢复消息。如果为实时会话启用了消息恢复，集成服务会将源消息或消息 ID 存储在恢复文件或表中。如果会话失败，您可以在恢复模式下运行该会话，以恢复集成服务无法处理的消息。

了解实时数据

您可以处理以下类型的实时数据：

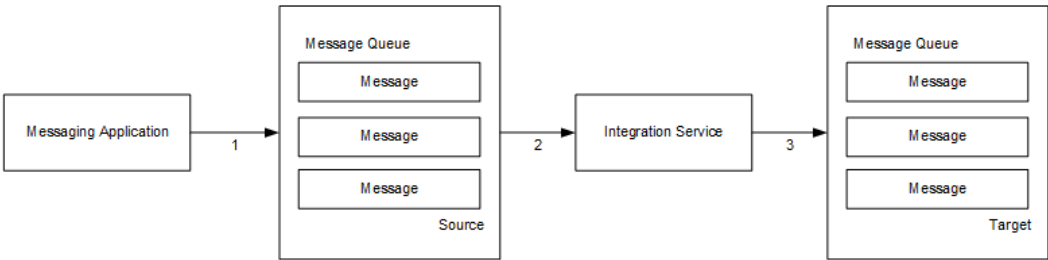
- **消息和消息队列。**处理来自 WebSphere MQ、JMS、MSMQ、SAP、TIBCO 和 webMethods 源的消息和消息队列。您可以从消息和消息队列读取。您可以写入消息、消息传送应用程序和消息队列。
- **Web 服务消息。**通过 Web 服务中心接收来自 Web 服务客户端的消息并转换数据。您可以将数据写入目标或将消息发送回 Web 服务客户端。
- **PowerExchange 从异构源捕获的更改数据。**提取 PowerExchange 从 i5/OS、Linux、UNIX、Windows 和 z/OS 系统上的各个关系源和非关系源捕获的更改数据。PowerExchange 更改数据捕获 (CDC) 功能与 PowerCenter 相集成，可实时地在整个企业内捕获、转换和交付更改数据。

消息和消息队列

集成服务使用消息传送和排队体系结构来处理实时数据。集成服务可以从消息队列中读取消息、处理消息数据以及将消息写入消息队列。

您还可以将消息写入其他消息传送应用程序。例如，集成服务可以从 JMS 源读取消息并将数据写入 TIBCO 目标。

下图显示了消息传送应用程序和集成服务如何处理来自消息队列的消息：



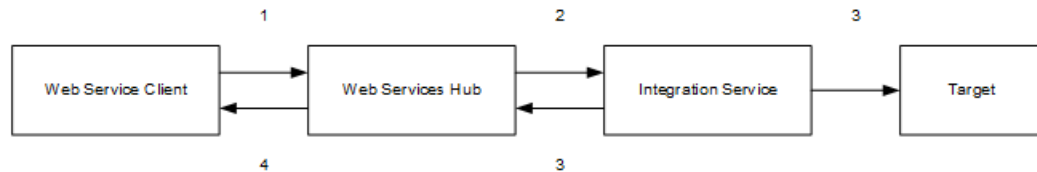
消息传送应用程序和集成服务完成以下任务以处理来自消息队列的消息：

1. 消息传送应用程序将消息添加到队列。
2. 集成服务从队列中读取消息并提取数据。
3. 集成服务处理数据并将回复写入消息队列。

Web 服务消息

Web 服务消息是指来自 Web 服务客户端的 SOAP 请求或来自 Web 服务中心的 SOAP 响应。集成服务处理来自 Web 服务客户端的实时数据，方法是通过 Web 服务中心接收消息请求并处理该请求。集成服务可以通过 Web 服务中心将回复发送回 Web 服务客户端，也可以将数据写入目标。

下图显示了 Web 服务客户端、Web 服务中心和集成服务如何处理 Web 服务消息：



Web 服务客户端、Web 服务中心和集成服务完成以下任务以处理 Web 服务消息：

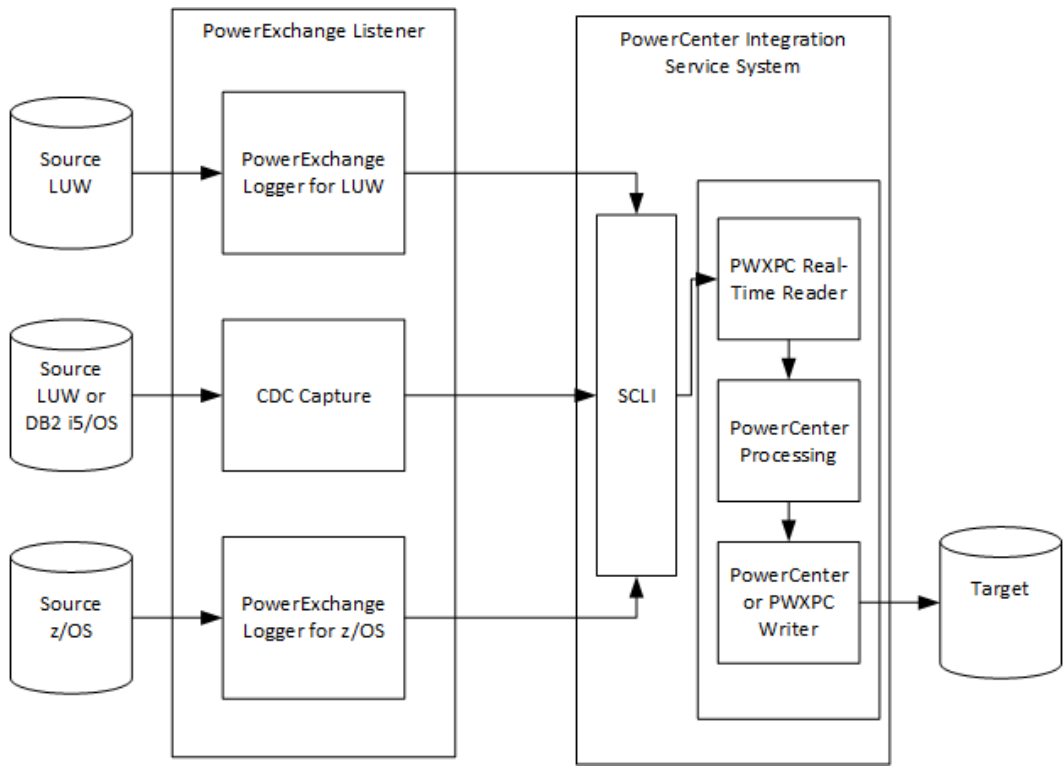
1. Web 服务客户端将 SOAP 请求发送到 Web 服务中心。
2. Web 服务中心处理 SOAP 请求并将该请求传递到集成服务。
3. 集成服务运行该服务请求。集成服务将响应发送到 Web 服务中心，或者将数据写入目标。
4. 如果集成服务将响应发送到 Web 服务中心，Web 服务中心将生成 SOAP 消息回复，并将该回复传递到 Web 服务客户端。

来自 PowerExchange CDC 源的更改数据

使用 PowerExchange 更改数据捕获 (Change Data Capture, CDC) 可将联机事务处理系统和数据库进行的最新更改发布到数据仓库、操作数据存储和应用程序，以帮助及时制定业务决策。

PowerExchange CDC 可从 i5/OS、Linux、UNIX、Windows 和 z/OS 系统上对各个关系源和非关系源提交的事务捕获更改数据。当选定的源对象上发生插入、更新和删除时，PowerExchange 可立即捕获这些操作并将更改存储在实时更改流或 PowerExchange 日志记录器日志文件中，直至 PowerExchange Client for PowerCenter (PWXP) 组件请求更改数据为止。

PWXPC 插件随 PowerCenter 一起安装，并要求在本地安装 PowerExchange。PWXPC 可与 PowerCenter 集成服务和客户端工具配合工作，以在特定时间段内运行实时会话或持续运行。PowerExchange 侦听器在源系统上运行或独立于源系统运行。下图显示了简单的 PowerExchange 和 PowerCenter 配置：



以下流程概括了 PowerExchange 与 PowerCenter 进行的交互：

1. 一旦发生更改，PowerExchange 立即从选定的源表和列捕获更改数据。
2. 启动一个包含 PowerExchange 源并使用 PWX CDC 实时应用程序连接的 PowerCenter 工作流。
3. PWXPC 通过 PowerExchange 调用级接口 (SCLI) 连接到 PowerExchange，以代表工作流通过 PowerExchange 侦听器检索更改数据。
4. PowerExchange 从映射中的所有源的更改流提取更改数据，并将数据传递到 PWXPC。
5. PWXPC 将更改数据传递到集成服务。
6. PowerCenter 工作流对更改数据进行处理和转换。
7. PowerCenter 或 PowerExchange 写入器将更改数据写入到一个或多个目标。

配置实时会话

配置会话以实时处理数据时，请配置用于控制会话何时停止从源中读取的会话属性。您可以将会话配置为在一段固定时间内停止接收消息后、会话达到消息计数上限时或会话已读取一段固定时间内的消息时停止从源中读取。您还可以配置集成服务如何将数据提交至目标并为失败的会话启用消息恢复。

您可以为实时会话配置以下属性：

- **终止条件。** 定义终止条件以确定集成服务何时停止从源中读取并结束会话。

- **刷新延迟。** 定义配置了刷新延迟的会话以读取和写入实时数据。刷新延迟决定会话将数据提交至目标的频率。
- **提交类型。** 定义基于源或目标的实时会话提交类型。如果是基于源的提交，集成服务将根据提交间隔和刷新延迟间隔提交消息。如果是基于目标的提交，集成服务将根据刷新延迟间隔提交消息。
- **消息恢复。** 为实时会话启用恢复以恢复来自失败会话的消息。

终止条件

终止条件决定集成服务停止从实时源读取消息并终止会话的时间。集成服务达到终止条件时将停止从实时源中读取。集成服务处理其读取的消息并将数据提交至目标。然后将结束会话。

您可以配置以下终止条件：

- 空闲时间
- 消息计数
- 读取器时间限制

如果您配置了多个终止条件，集成服务将在满足第一个条件时停止从源中读取。默认情况下，集成服务连续读取消息并使用刷新延迟来确定何时刷新来自源的数据。刷新后，集成服务将重置终止条件的计数器。

空闲时间

空闲时间是指集成服务在停止从源中读取之前等待接收消息的时间（秒）。-1 指示无限期等待。

例如，如果 JMS 会话的空闲时间为 30 秒，集成服务在从 JMS 读取后将等待 30 秒。如果在 30 秒内 JMS 中未收到新消息，集成服务将停止从 JMS 中读取。集成服务将处理消息并结束会话。

消息计数

消息计数是指集成服务在停止从源中读取之前从实时源中读取的消息数。-1 指示消息数不受限制。

例如，如果 JMS 会话中的消息计数为 100，集成服务在读取 100 条消息后将停止从源中读取。集成服务将处理消息并结束会话。

注意：消息计数终止条件的名称因 Informatica 产品而异。例如，PowerExchange for SAP NetWeaver 的消息称为“数据包计数”。PowerExchange Client for PowerCenter 的消息计数称为“UOW 计数”。

读取器时间限制

读取器时间限制是指集成服务在停止从源中读取之前从实时源中读取源消息的时间（秒）。使用读取器时间限制可在一段固定时间内从实时源读取消息。0 指示无限期读取。

例如，如果您使用的时间限制为 10 秒，集成服务将在 10 秒后停止从消息传送应用程序中读取。集成服务将处理消息并结束会话。

刷新延迟

使用刷新延迟可实时运行会话。刷新延迟决定集成服务刷新来自源的数据的频率。例如，如果您将刷新延迟设置为 10 秒，集成服务将每隔 10 秒刷新一次来自源的数据。

对于来自 PowerExchange 更改数据捕获源的更改数据，刷新延迟间隔由刷新延迟和工作单元 (UOW) 计数属性决定。有关详细信息，请参阅《PowerExchange Interfaces for PowerCenter》。

集成服务在从实时源中读取数据并且为会话配置了刷新延迟时执行以下过程：

1. 集成服务从源中读取数据。
刷新延迟间隔在集成服务从源中读取第一条消息时开始计时。
2. 在刷新延迟间隔结束时，集成服务将停止从源中读取数据。
3. 集成服务处理消息并将其写入目标。
4. 集成服务再次从源中读取，直至达到下一个刷新延迟间隔。

配置刷新延迟（秒）。默认值为零，指示刷新延迟处于禁用状态并且会话不实时运行。

请根据数据的活动频率和用户需要访问数据的速率来配置刷新延迟间隔。如果数据很快过期（例如金融交易信息），请配置一个较低的刷新延迟间隔，以便目标表尽可能在发生更改时更新。例如，用户需要每隔几分钟获取一次更新的财务数据。但是，他们每天只需获取一次更新的客户地址变更数据。请为财务数据配置一个较低的刷新延迟间隔，为地址变更数据配置一个较高的刷新延迟间隔。

配置刷新延迟时请遵循以下规则和准则：

- 集成服务不缓冲超过刷新延迟间隔的消息。
- 设置的刷新延迟时间间隔越低，集成服务将消息提交至目标的频率就越高。
- 如果您使用低刷新延迟间隔，会话会占用更多系统资源。

如果您配置了一个提交间隔，刷新延迟与提交间隔的组合决定将数据提交至目标的时间。

提交类型

集成服务根据刷新延迟和提交类型将数据提交至目标。您可以将会话配置为使用以下提交类型：

- **基于源的提交。**配置基于源的提交时，集成服务将根据提交间隔与刷新延迟间隔的组合将数据提交至目标。集成服务满足的第一个条件将触发刷新延迟期限的结束。刷新后，计数器将重置。
例如，您将刷新延迟设置为 5 秒，基于源的提交间隔为 1000 条消息。集成服务将在从源中读取 1000 条消息后或 5 秒后将消息提交至目标。
- **基于目标的提交。**配置基于目标的提交时，集成服务将忽略提交间隔，并根据刷新延迟间隔将数据提交至目标。

在实时会话中写入目标时，集成服务将连续处理提交并实时将数据提交至目标。集成服务不在 DTM 缓冲区内内存存储数据。

相关主题：

- [“提交点” 页面上 123](#)

消息恢复

为实时会话启用消息恢复时，集成服务可以恢复来自失败会话的未处理消息。集成服务将源消息或消息 ID 存储在恢复文件、恢复表、恢复队列或恢复主题中。如果会话失败，请在恢复模式下运行该会话，以恢复集成服务未处理的消息。

消息或消息 ID 存储在以下存储类型中，具体取决于实时源和目标类型：

- **恢复文件。** 消息或消息 ID 存储在指定的本地恢复文件中。与实时源和非关系目标或非队列目标的会话使用恢复文件。
- **恢复表。** 消息 ID 存储在目标数据库上的恢复表中。与 JMS 或 WebSphere MQ 源和关系目标的会话使用恢复表。
- **恢复队列和恢复主题。** 消息 ID 存储在恢复队列或恢复主题中。与 JMS 或 WebSphere MQ 源和 JMS 或 WebSphere MQ 目标的会话使用恢复队列。与 JMS 或 WebSphere MQ 源和主题目标的会话使用恢复主题。

会话可以使用几种存储类型的组合。例如，与 JMS 和 TIBCO 源的会话使用恢复文件和恢复表。

恢复实时会话时，集成服务将从中断点开始还原操作的状态。集成服务将读取和处理恢复文件、恢复表、恢复队列或恢复主题中的消息。然后将结束会话。

恢复过程中，终止条件不影响集成服务从恢复文件、恢复表、恢复队列或恢复主题中读取的消息。例如，如果您为会话指定了消息计数和空闲时间，终止条件将应用于集成服务从源中读取的消息，不应用于从恢复文件、恢复表、恢复队列或恢复主题中读取的消息。

如果会话在某些条件下失败，则除上述存储类型外，集成服务还使用恢复忽略列表。

与 MSMQ 源的会话、Web 服务消息或来自 PowerExchange 更改数据捕获源的更改数据使用不同的恢复策略。

前提条件

为与 JMS 或 WebSphere MQ 源和 JMS 或 WebSphere MQ 目标的会话启用消息恢复之前，请完成以下必备条件：

- 在 JMS 提供程序或 WebSphere MQ 中创建恢复队列。或者，在 JMS 提供程序中创建恢复主题。
- 在与消息队列相同的队列管理器下创建恢复队列，这样提交范围将相同。
- 将恢复队列配置为持久性队列。如果恢复队列不是持久性队列，则会出现数据重复问题。

启用消息恢复的步骤

要为会话启用消息恢复，请完成以下步骤：

1. 在会话属性中，选择“从上一检查点继续”作为恢复策略。
2. 在会话属性中指定每个分区点位置的恢复缓存目录。

集成服务在恢复缓存目录指示的位置存储消息。恢复缓存目录的默认值为 \$PMCacheDir。

恢复文件

对于为进行恢复而启用的包含以下源和目标类型的实时会话，集成服务将消息或消息 ID 存储在恢复文件中：

- JMS 源与非关系目标、非 JMS 目标或非 WebSphere MQ 目标
- WebSphere MQ 源与非关系目标、非 JMS 目标或非 WebSphere MQ 目标
- SAP ECC 源和所有目标
- webMethods 源和所有目标
- TIBCO 源和所有目标

集成服务将消息或消息 ID 暂时存储在您在会话属性中配置的本地恢复文件中。恢复过程中，集成服务将处理此恢复文件中的消息以确保数据不会丢失。

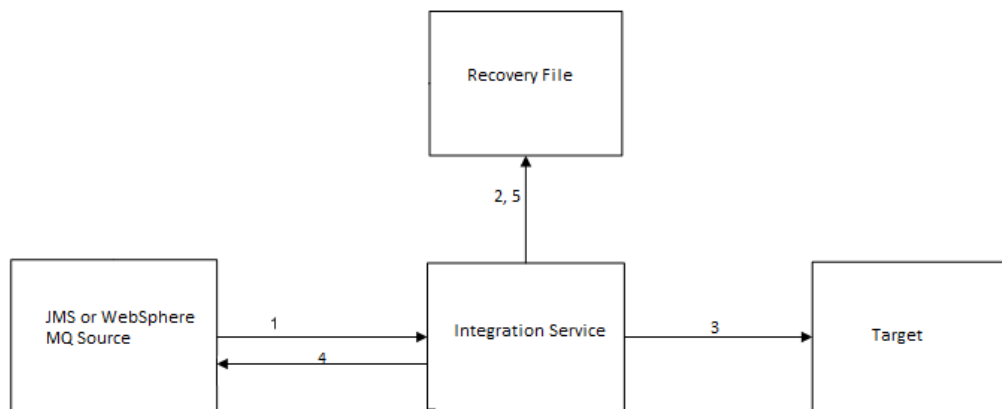
针对 JMS 和 WebSphere MQ 源的消息恢复

您可以针对与 JMS 和 WebSphere MQ 源的会话启用消息恢复，以恢复集成服务无法处理的消息。集成服务可以从中断点恢复运行状态。

集成服务完成以下任务以使用恢复文件处理消息：

1. 集成服务从源中读取消息。
2. 集成服务将消息 ID 写入恢复文件。集成服务重复步骤 1 到 2，直至达到刷新延迟。
3. 集成服务处理消息并将其写入目标。目标提交消息。
4. 集成服务将批处理确认发送给源，确认其读取消息。源将删除这些消息。
5. 集成服务清除恢复文件。

下图显示了集成服务如何使用恢复文件处理消息：



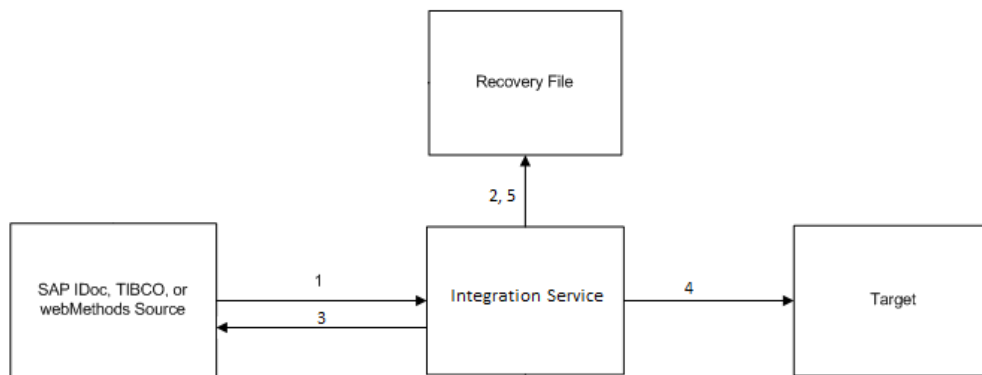
针对 SAP IDoc、TIBCO 和 webMethods 源的消息恢复

您可以针对与 SAP IDoc、TIBCO 和 webMethods 源的会话启用消息恢复，以恢复集成服务无法处理的消息。集成服务可以从中断点恢复运行状态。

集成服务完成以下任务以使用恢复文件处理消息：

1. 集成服务从源中读取消息。
2. 集成服务将消息写入恢复文件。
3. 集成服务将确认发送给源，以确认其读取消息。源将删除该消息。集成服务重复步骤 1 到 3，直至达到刷新延迟。
4. 集成服务处理消息并将其写入目标。目标提交消息。
5. 集成服务清除恢复文件。

下图显示了集成服务如何使用恢复文件处理消息：



消息恢复

恢复实时会话时，集成服务将读取并处理缓存的消息。集成服务读取所有缓存的消息后将结束会话。

对于与 JMS 和 WebSphere MQ 源的会话，集成服务将使用恢复文件中的消息 ID 来检索源中的消息。

集成服务在刷新延迟期限到期后以及成功的会话结束时清除恢复文件。如果会话在集成服务将消息提交至目标之后、从恢复文件中删除消息之前失败，目标在恢复过程中会收到重复的行。

会话恢复数据刷新

恢复数据刷新是指集成服务将会话恢复数据从操作系统缓存区刷新到恢复文件时执行的过程。如果集成服务无法将恢复数据写入恢复文件，您可以防止数据丢失。如果出现操作系统故障、硬件故障或文件系统故障，集成服务将无法写入恢复数据。恢复数据刷新适用于包括 JMS 或 WebSphere MQ 源和非关系目标、非-JMS 目标或非 WebSphere MQ 目标的会话。

您可以将集成服务配置为将恢复数据从操作系统缓冲区刷新到恢复文件，方法是在 Administrator 工具中将集成服务属性“刷新会话恢复数据”设置为“自动”或“是”。

恢复表

对于为进行恢复而启用的包含以下源和目标类型的实时会话，集成服务将消息 ID 存储在恢复表中：

- JMS 源与关系目标
- WebSphere MQ 源与关系目标

集成服务将消息 ID 和提交号暂时存储在每个目标数据库上的恢复表中。提交号指示集成服务提交至目标的提交数量。恢复过程中，集成服务使用提交号确定是否将相同数量的消息提交至所有目标。消息 ID 和提交号对照恢复表进行验证，以确保数据未丢失或不重复。

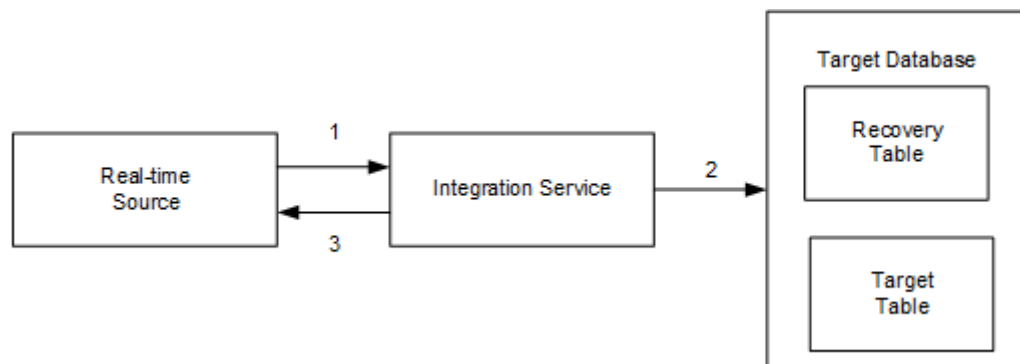
注意：源必须使用唯一的消息 ID 并提供通过消息 ID 访问消息的权限。

PM_REC_STATE 表

集成服务运行使用恢复表且启用了恢复的实时会话时，将在目标数据库中创建一个恢复表 PM_REC_STATE 以存储消息 ID 和提交号。集成服务恢复会话时，将使用恢复表中的信息来确定是否需要将消息写入目标表。

消息处理

下图显示了集成服务如何使用恢复表处理消息：



集成服务完成以下任务以使用恢复表处理消息：

1. 集成服务一次读取一条消息，直至达到刷新延迟。
2. 集成服务将消息 ID、提交号和转换状态写入目标数据库上的恢复表，同时将消息写入目标。
3. 目标提交消息时，集成服务将向实时源发送确认消息，以确认所有消息都已处理并写入目标。

集成服务继续从源中读取消息。

如果会话具有多个分区，上述任务将适用于每个分区。

消息恢复

恢复实时会话时，集成服务使用恢复表中的消息 ID 和提交号来确定是否已将消息提交至所有目标。

如果恢复表中存在消息 ID，并且所有目标的目标号相同，集成服务则将消息提交至所有目标。恢复过程中，集成服务将向其处理了消息的源发送确认消息。

如果目标的提交号不同，则集成服务不将消息提交至所有目标。恢复过程中，集成服务将从恢复表中读取消息 ID 和转换状态。集成服务处理消息并将其写入没有相应消息的目标。集成服务从恢复表中读取所有消息后将结束会话。

如果会话在集成服务将消息提交至所有目标之前失败，并且您在冷启动模式下重新启动该会话，目标会收到重复的行。

恢复队列和恢复主题

对于为进行恢复而启用的包含以下源和目标类型的实时会话，集成服务将消息 ID 存储在恢复队列或恢复主题中：

- JMS 源与 JMS 或 WebSphere MQ 目标
- WebSphere MQ 源与 JMS 或 WebSphere MQ 目标

集成服务将消息 ID 和提交号暂时存储在您在 JMS 提供程序或 WebSphere MQ 中创建的恢复队列或恢复主题中。提交号指示集成服务提交至目标的提交数量。恢复过程中，集成服务使用提交号确定是否将相同数量的消息提交至所有目标。消息 ID 和提交号对照恢复队列或恢复主题进行验证，以确保数据未丢失或不重复。

集成服务对每个会话中的所有队列目标使用相同的恢复队列或恢复主题。请为会话创建多个恢复队列或恢复主题以提高性能。

如果您未在会话属性或 JMS 连接对象中指定恢复队列或恢复主题名称，集成服务会将恢复信息存储在恢复文件中。为实现最优性能，请配置恢复队列或恢复主题名称而不是恢复文件。

消息处理

集成服务使用恢复队列或恢复主题处理消息，方法与使用恢复表处理消息类似。集成服务将恢复信息写入恢复队列或恢复主题而不是恢复表。

消息恢复

集成服务从恢复队列或恢复主题恢复消息，方法与从恢复表恢复消息类似。集成服务从恢复队列或恢复主题中检索恢复信息，而不是从恢复表中检索。

恢复忽略列表

与 JMS 或 WebSphere MQ 源的会话失败时，集成服务将恢复信息写入恢复忽略列表。如果源有可能未收到确认消息，集成服务会将恢复信息写入该列表。例如，会话在集成服务向源发送确认消息之前、将消息写入目标之后失败。在这种情况下，源可以回滚当前事务，但该事务中的消息可能无法立即获取。如果消息包含在恢复会话中，则会出现数据重复问题。为防止数据重复，集成服务创建了恢复忽略列表。

恢复忽略列表将存储集成服务写入目标的失败会话的消息 ID。集成服务在该会话使用的存储类型中创建恢复忽略列表，例如恢复文件、恢复表、恢复队列或恢复主题。恢复过程中，集成服务使用恢复忽略列表和存储类型来确定是否已将消息写入目标。集成服务对照恢复忽略列表和存储类型验证消息 ID，以确保数据不重复。

会话失败时，集成服务将消息写入恢复忽略列表并添加时间戳。默认情况下，集成服务在该时间戳后一小时从恢复忽略列表中删除消息。如果集成服务在默认时间段内在源中找到该消息，则会将其从恢复忽略列表中删除。

如果您在冷启动模式下重新启动了已停止或失败的会话，目标可能会收到重复的行。重新启动启用了恢复的会话以防止数据重复。或者，如果您可以确保从源中删除恢复忽略列表中的消息，请在冷启动模式下重新启动会话。使用会话日志可查看消息 ID。如果您配置了详细数据跟踪，集成服务会将恢复忽略列表中的消息 ID 写入会话日志。

停止实时会话

除非失败或者您手动停止，否则实时会话将连续运行。您可以通过在 *pmcmd* 中或从 Workflow Monitor 发出停止命令来停止会话。您可能需要停止会话以执行例行维护。

停止实时会话时，集成服务将根据以下实时源处理管道中的消息：

- **JMS 和 WebSphere MQ。** 集成服务处理其读取的消息，直至您发出停止命令。集成服务将消息写入目标。
- **MSMQ、SAP、TIBCO、webMethods 和 Web 服务消息。** 如果您在集成服务将所有消息写入目标之前停止会话，集成服务将不处理消息。

停止与 JMS 或 WebSphere MQ 源的实时会话时，集成服务将执行以下任务：

1. 集成服务停止从源中读取消息。
如果您停止了实时恢复会话，集成服务将在恢复所有消息后停止从源中读取。
2. 集成服务处理管道中的消息并将其写入目标。
3. 集成服务向源发送确认消息。
4. 集成服务清除恢复表或恢复文件，以避免重新启动会话时出现数据重复问题。

重新启动会话时，集成服务将开始从源中读取。集成服务将从中断点开始还原操作的会话和转换状态以继续进行会话。

注意：如果实时会话在您将其停止后挂起，会话可能保留在正在停止状态。如果实时会话保留在正在停止模式，您可以将其中止。集成服务处理在您发出停止命令之前读取的消息。

重新启动和恢复实时会话

您可以继续进行停止或失败的实时会话。要继续进行会话，必须重新启动或恢复该会话。如果您为会话启用了自动任务恢复，集成服务可以自动恢复会话。

以下部分介绍了实时会话特定的恢复信息。

重新启动实时会话

重新启动会话时，集成服务将根据实时源继续进行会话。集成服务将根据实时源重新启动启用或未启用恢复的会话。

您可以在冷启动模式下重新启动任务或工作流。在冷启动模式下重新启动任务或工作流时，集成服务将丢弃恢复信息并重新启动任务或工作流。

恢复实时会话

如果您启用了会话恢复，则可以恢复失败或中止的会话。恢复会话时，集成服务将从中断点开始继续处理消息。集成服务将根据实时源恢复消息。

集成服务使用以下会话恢复类型：

- **自动恢复。** 如果您将工作流配置为自动恢复终止的任务，集成服务将重新启动会话。集成服务将恢复所有未处理的数据并重新启动会话，而无论使用何种实时源。
- **手动恢复。** 使用 Workflow Monitor 或 Workflow Manager 菜单命令或 *pmcmd* 命令恢复会话。对于某些实时源，您必须在重新启动之前恢复会话，否则集成服务将不处理来自失败会话的消息。

重新启动和恢复命令

您可以在 Workflow Manager、Workflow Monitor 或 *pmcmd* 中重新启动或恢复会话。集成服务将根据实时源继续进行会话。

下表介绍了当您使用以下命令重新启动或恢复会话时的行为：

命令	说明
<ul style="list-style-type: none">- 重新启动任务- 重新启动工作流- 从任务中重新启动工作流	重新启动任务或工作流。对于 JMS 和 WebSphere MQ 会话，集成服务将在重新启动任务或工作流之前恢复。 注意: 如果会话包括 JMS、WebSphere MQ 源以及其他实时源，集成服务将在重新启动任务或工作流之前对所有实时源执行恢复。
<ul style="list-style-type: none">- 恢复任务- 恢复工作流- 通过恢复此任务来重新启动工作流	恢复任务或工作流。
<ul style="list-style-type: none">- 冷启动任务- 冷启动工作流- 从任务中冷启动工作流	放弃恢复信息并重新启动任务或工作流。

实时会话的规则和准则

运行实时会话时，请遵循以下规则和准则：

- 如果映射中包含事务控制转换，会话将失败。
- 如果映射中包含启用了生成事务功能的任何转换，会话将失败。
- 如果映射中包含转换范围设置为所有输入的任何转换，会话将失败。
- 如果映射中包含设置了行转换范围并接收来自多个事务控制点的输入的任何转换，会话将失败。
- 如果目标的加载范围设置为所有输入，会话将失败。
- 当您在调试模式下运行会话时，集成服务将忽略刷新延迟。
- 如果映射中包含关系目标，请将目标的加载类型配置为普通。
- 如果映射中包含 XML 目标定义，请在目标定义中为“提交后”选项选择“附加到文档”。
- 集成服务可以将连接故障复原到 WebSphere MQ 和 JMS。集成服务无法复原到任何其他消息传送系统。
- 当实时会话中包含请求和响应时（例如在 Web 服务中），会话日志中将包含请求和响应的开始和结束时间。当实时会话中包含发布/订阅或点到点体系结构时，会话日志中将包含用于说明集成服务何时将行提交至目标的统计信息。

消息恢复的规则和准则

PowerCenter 集成服务将使启用了消息恢复且包含以下任意条件的会话失败：

- 源定义为联接器转换的主源。
- 为相同的目标加载顺序组配置了多个并发运行的源定义。

- 映射中包含 XML 目标定义。
- 您在重新启动会话之前编辑了恢复文件或映射，并且运行的会话配置了重新启动或继续的恢复策略。
- 集成服务无法连接到恢复队列或恢复主题。
- 集成服务不将恢复消息写入恢复队列或恢复主题。

如果集成服务在消息队列中读取或写入的消息数超过消息大小限制，请增大消息大小限制或降低刷新延迟。

实时处理示例

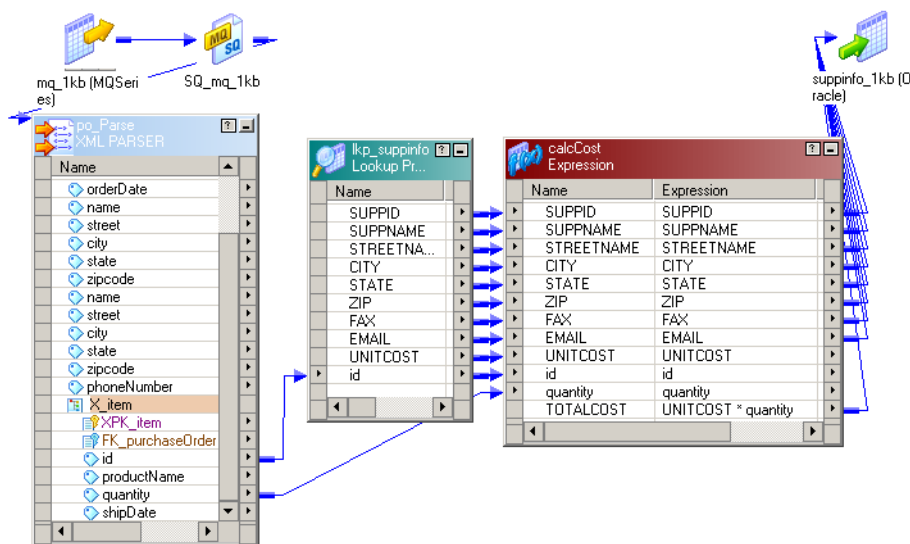
下列显示了您可以如何使用 PowerExchange for IBM WebSphere MQ 和 PowerCenter 处理实时数据。

您需要实时处理订单。订单中可以包含多个供应商提供的多个项目。但是，订单中不包含供应商或项目成本。当您收到订单时，必须计算每个供应商的总成本。您的主数据库中包含您的供应商及其各自的项目和项目成本。您使用 PowerCenter 根据项目 ID 查找供应商和项目成本。您还使用 PowerCenter 将总供应商成本写入关系数据库。

您的数据库管理员建议您将目标更新为在一次提交中最多包含 1000 条消息。您还需要每隔 2000 毫秒更新一次目标，以确保目标始终是最新的。

要实时处理订单，请创建并配置一个映射。

下图显示了实时处理订单的映射：



示例映射中包含以下组件：

- **源。** WebSphere MQ. 每条消息都是 XML 格式的消息，并且包含一个订单。
- **XML 解析器转换。** 接收来自 MQ 源限定符转换的订单信息。XML 解析器转换解析 XML 文件中的订单 ID 和数量。
- **查找转换。** 查找供应商详细信息中的订单 ID。查找转换将供应商信息、订单项目 ID 和项目成本传递到表达式转换。
- **表达式转换。** 计算供应商的订单成本。
- **目标。** Oracle 关系数据库。目标中包含供应商信息和总供应商成本。

您创建并配置了具有以下属性的会话和工作流：

属性	值
消息计数	1,000
刷新延迟间隔	2000 毫秒
提交类型	基于源的提交
工作流计划	连续运行

以下步骤介绍了集成服务如何实时处理会话：

1. 集成服务从 WebSphere MQ 队列中读取消息，直至读取 1000 条消息或 2000 毫秒以后。满足其中任一条件时，集成服务将停止从 WebSphere MQ 队列中读取。
2. 集成服务查找供应商信息并计算订单成本。
3. 集成服务将供应商信息和订单成本写入 Oracle 关系目标。
4. 集成服务重新开始从 WebSphere MQ 队列中读取消息。
5. 集成服务重复步骤 1 到 4，因为您将工作流配置为连续运行。

PowerCenter 实时产品

可以使用以下产品在 PowerCenter 中读取、转换和写入实时数据：

- **PowerExchange Client for PowerCenter**。使用 PowerExchange 可从 i5/OS、Linux、UNIX、Windows 和 z/OS 系统上的各种关系源和非关系源中捕获更改数据。PowerExchange 更改数据捕获 (CDC) 功能与 PowerCenter 相集成，可实时地在整个企业内捕获、转换和交付更改数据。

PowerExchange Client for PowerCenter (PWXPC) 软件随 PowerCenter 一起安装，可与 PowerCenter 集成服务和 PowerCenter 客户端工具配合使用，用于从 PowerExchange 侦听器检索更改数据。侦听器在源系统上运行或独立于源系统运行。您可以在工作流中包含 PowerExchange 目标，以便将更改数据写入 PowerCenter 不支持的目标类型，包括 i5/OS 和 z/OS 系统上的目标。另外，还可使用 PowerExchange 批量提取和加载大批量数据，以便具体化或刷新 CDC 目标。

- **PowerExchange for JMS for PowerCenter**。使用 PowerExchange for JMS 可从 JMS 源读取数据，并向 JMS 目标写入数据。可以根据消息主题从 JMS 消息、JMS 提供程序消息队列或 JMS 提供程序中读取。可以根据消息主题写入 JMS 提供程序消息队列或 JMS 提供程序。

JMS 提供程序是面向消息的中间件系统，可以发送和接收 JMS 消息。会话过程中，集成服务连接到 Java 命名和目录接口 (JNDI) 以确定连接信息。集成服务确定连接信息时，将连接到 JMS 提供程序以读取或写入 JMS 消息。

- **PowerExchange for WebSphere MQ for PowerCenter**。使用 PowerExchange for WebSphere MQ 可从 WebSphere MQ 消息队列读取消息，并向 WebSphere MQ 消息队列或数据库目标写入数据。数据提取和加载过程中，PowerExchange for WebSphere MQ 与 WebSphere MQ 队列管理器、消息队列和 WebSphere MQ 消息交互。

- **PowerExchange for TIBCO for PowerCenter**。使用 PowerExchange for TIBCO 可从 TIBCO 中读取消息，并向 TIBCO 写入 TIB/Rendezvous 或 AE 格式的消息。

集成服务接收来自 TIBCO 后台程序的 TIBCO 消息，并通过 TIBCO 后台程序写入消息。TIBCO 后台程序通过局域网或广域网传输目标消息。目标侦听器根据消息主题订阅 TIBCO 目标消息。

- **PowerExchange for webMethods for PowerCenter。**使用 PowerExchange for webMethods 可从 webMethods 源中读取文档，并向 webMethods 目标写入文档。

集成服务连接到用于发送和接收 webMethods 文档以及对其进行排队的 webMethods Broker。集成服务根据定义的文档类型或客户端 ID 读取和写入 webMethods 文档。集成服务还读取和写入 webMethods 请求/回复文档。

- **PowerExchange for MSMQ for PowerCenter。**使用 PowerExchange for MSMQ 可从 MSMQ 源读取数据，并向 MSMQ 目标写入数据。

集成服务连接到 Microsoft 消息传送队列以从消息中读取数据或将数据写入消息。队列可以是公共或专用的事务或非事务型队列。

- **PowerExchange for SAP NetWeaver for PowerCenter。**使用 PowerExchange for SAP NetWeaver 并借助应用程序链接启用 (ALE) 功能，可以使用出站 IDoc 从 SAP 读取数据，或者使用入站 IDoc 向 SAP 写入数据。

集成服务可以从外出 IDoc 中读取并写入关系目标。集成服务可以从关系源中读取数据并将数据写入传入 IDoc。集成服务可以实时捕获对 SAP 应用程序数据库中的主数据或事务数据所做的更改。

- **PowerCenter Web Services Provider。**使用 PowerCenter Web Services Provider 可通过 Web 服务中心将转换逻辑公开为服务，并编写客户端应用程序来运行实时 Web 服务。您可以创建一个服务映射以接收来自 Web 服务客户端的消息、转换该消息并将其写入 PowerCenter 支持的任何目标。您还可以创建带有 Web 服务源和目标定义的服务映射，以接收来自 Web 服务客户端的消息请求、转换数据并将响应发送回 Web 服务客户端。

Web 服务中心接收来自 Web 服务客户端的请求并将其传递至网关。集成服务或存储库服务通过 Web 服务中心处理请求并将响应发送至 Web 服务客户端。

第 7 章

提交点

本章包括以下主题：

- [提交点概览, 123](#)
- [基于目标的提交, 123](#)
- [基于源的提交, 124](#)
- [用户定义的提交, 127](#)
- [了解事务控制, 130](#)
- [设置提交属性, 133](#)

提交点概览

提交间隔是集成服务在会话期间将数据提交给目标的间隔。提交点可以是提交间隔的因子、提交间隔类型和缓冲区块的大小。提交间隔是要用作提交点基础的行数。提交间隔类型是要用作提交点基础的行类型。可以在以下提交类型之间进行选择：

- **基于目标的提交。**集成服务根据目标行数和目标表的键约束提交数据。提交点还依赖于缓冲区块大小、提交间隔和写入器超时的集成服务配置。
- **基于源的提交。**集成服务根据源行数提交数据。提交点是在会话属性中配置的提交间隔。
- **用户定义的提交。**集成服务根据在映射属性中定义的事务提交数据。还可以在会话属性中配置一些提交和回滚选项。

基于源的提交会话和用户定义的提交会话具有分区约束。如果将具有多个分区的会话配置为使用基于源或用户定义的提交，则可以在管道中的特定分区点上选择传递分区。

基于目标的提交

在基于目标的提交会话期间，集成服务根据目标行数和目标表的键约束提交行。提交点取决于以下因素：

- **提交间隔。**要用作提交基础的行数。在会话属性中配置目标提交间隔。
- **写入器等待超时。**写入器在发出提交命令之前等待的时间量。在集成服务设置中配置写入器等待超时。
- **缓冲区块。**在会话期间用于保存数据行的内存块。可以在会话属性中配置缓冲区块大小，但不能配置块保存的行数。

运行基于目标的提交会话时，集成服务可能会在配置的提交间隔之前、期间或之后发出提交命令。集成服务使用以下过程发出提交命令：

- 集成服务达到提交间隔时，会继续填充写入器缓冲区块。写入器缓冲区块填满时，集成服务将发出提交命令。
- 如果写入器缓冲区块在提交间隔之前填满，则集成服务将写入目标，但会等待发出提交命令。当满足以下任一条件时，它便会发出提交命令：
 - 写入器空闲集成服务写入器等待超时选项指定的时间量。
 - 集成服务达到提交间隔并填充另一个写入缓冲区块。

注意：为包含 XML 目标的会话选择基于目标的提交时，Workflow Manager 会在“映射”选项卡的“转换”视图上禁用“提交时”会话属性。

基于源的提交

在基于源的提交会话期间，集成服务根据目标加载顺序组内某些活动源中的行数将数据提交到目标。这些行称为源行。

集成服务运行基于源的提交会话时，会标识映射中每个管道的提交源。集成服务会在每个提交间隔内从这些活动源生成提交行。集成服务会将用于基于源的提交间隔的转换名称写入会话日志中：

```
Source-based commit interval based on... TRANSFORMATION_NAME
```

与活动源生成的行数相比，集成服务可能会向目标提交更少的行。例如，您具有一个基于源的提交会话，它通过活动源传递了 10,000 行，由于转换逻辑删除了 3,000 行。当剩余 7,000 行达到目标时，集成服务会向目标发出提交命令。

写入器缓冲区块中保存的行数不会影响基于源的提交会话的提交点。例如，您具有一个基于源的提交会话，它通过活动源传递了 10,000 行。当此 10,000 行达到目标时，集成服务会发出提交命令。如果会话成功完成，集成服务会在 10,000、20,000、30,000 和 40,000 源行之后发出提交命令。

如果目标位于同一事务控制单元中，集成服务会同时向目标提交数据。如果会话失败或中止，集成服务会将事务控制单元中的所有未提交数据回滚至同一源行。

如果目标位于不同的事务控制单元中，集成服务会在每个目标接收提交行时执行提交。如果会话失败或中止，集成服务会将每个目标回滚至最后一个提交点。它可能不会回滚至单独事务控制单元中目标的同一源行。

注意：如果会话使用一对一映射，则基于源的提交可能会降低会话性能。一对一映射是将数据从源限定符、XML 源限定符或应用程序源限定符转换直接移至目标的映射。

确定提交源

运行基于源的提交会话时，集成服务会在不传播事务边界的所有源限定符和转换上生成提交。这包括以下活动源：

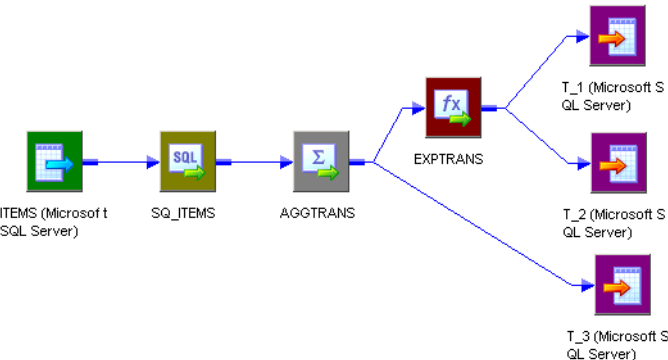
- 源限定符
- 应用程序源限定符
- MQ 源限定符
- XML 源限定符（仅连接一个输出组中的端口时）
- 规范器 (VSAM)
- 具有“全部输入”转换范围的汇总器
- 具有“全部输入”转换范围的联接器

- 具有“全部输入”转换范围的等级
- 具有“全部输入”转换范围的排序器
- 具有一个输出组和“全部输入”转换范围的自定义
- 具有一个连接到多个上游事务控制点的输出组的多输入组转换
- Mapplet（如果包含上述转换之一）

一个映射可以具有一个或多个目标加载顺序组，而一个目标加载顺序组可以具有一个或多个用于生成提交的来源。集成服务使用由最接近于目标定义的活动源生成的提交。这称为提交源。

下图显示了具有一个提交源的映射：

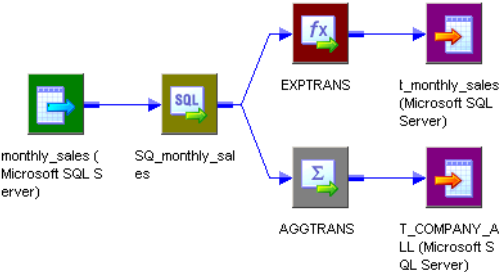
图 2. 具有一个提交源的映射



该映射包含一个源限定符转换和一个具有“全部输入”转换范围的汇总器转换。汇总器转换比源限定符转换更接近于目标，因此它用作基于源的提交会话的提交源。

下图显示了具有多个提交源的映射：

图 3. 具有多个提交源的映射



“转换范围”属性为“全部输入”。

该映射包含一个目标加载顺序组，该组具有一个从源限定符转换分支到两个目标的源管道。一个管道分支包含一个具有“全部输入”转换范围的汇总器转换，而另一个管道分支包含一个表达式转换。集成服务将源限定符转换标识为 t_monthly_sales 的提交源，将汇总器标识为 T_COMPANY_ALL 的提交源。它对这两个目标都执行基于源的提交，但对每个目标使用不同的提交源。

从基于源的提交切换到基于目标的提交

如果集成服务在目标加载顺序组中标识了不从生成提交的活动源接收提交的目标，则它仅针对此目标还原为基于目标的提交。

集成服务会将用于基于源的提交间隔的转换名称写入会话日志中。集成服务切换到基于目标的提交时，会在会话日志中写入一条消息。

在以下情况下，目标可能不会从提交源接收提交：

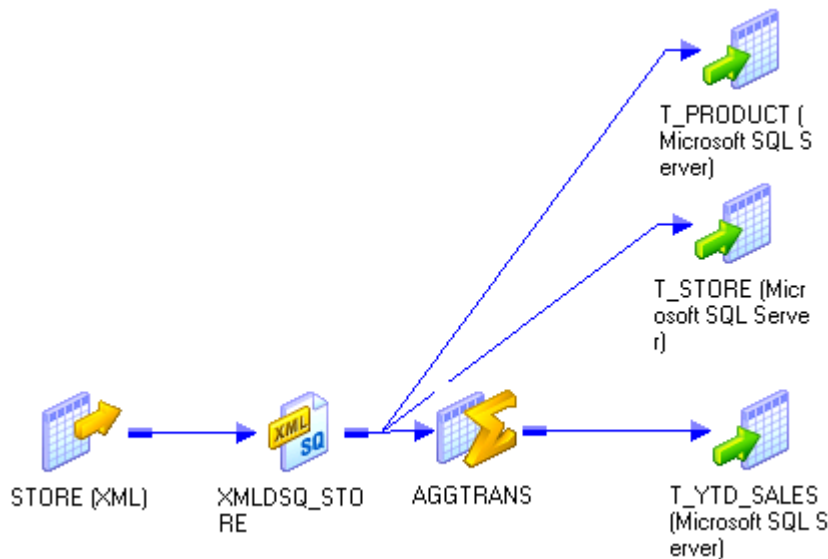
- **目标从 XML 源限定符转换接收数据，且您将多个输出组从 XML 源限定符转换连接到下游转换。**在下游连接多个输出组时，XML 源限定符转换不会生成提交。
- **目标从具有多个输出组的活动源（而不是 XML 源限定符转换）接收数据。**例如，目标从未配置为生成事务的自定义转换接收数据。多个输出组活动源既不会生成也不会传播提交。

连接映射中的 XML 源

在下游连接多个输出组时，XML 源限定符转换不会生成提交。连接映射中的 XML 源限定符转换时，集成服务可以根据映射中使用的转换对此会话中的目标使用不同的提交类型：

- **在 XML 源限定符转换和目标之间放置一个提交源。**集成服务对目标使用基于源的提交，因为它从提交源接收提交。活动源是目标的提交源。
- **不在 XML 源限定符转换和目标之间放置提交源。**集成服务对目标使用基于目标的提交，因为它不接收任何提交。

下图显示了具有 XML 源限定符转换的映射：



该映射包含一个在下游连接了多个输出组的 XML 源限定符转换。由于在下游连接了多个输出组，XML 源限定符转换不会生成提交。您将 XML 源限定符转换连接到两个关系目标（T_STORE 和 T_PRODUCT）。因此，这些目标不会接收由活动源生成的任何提交。集成服务在加载到这些目标时使用基于目标的提交。

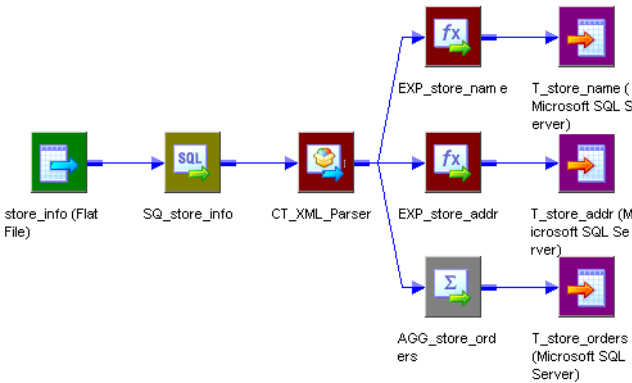
但是，该映射在 XML 源限定符转换和 T_YTD_SALES 之间包含一个可生成提交的活动源（即 AGG_Sales）。集成服务在加载到 T_YTD_SALES 时使用基于源的提交。

连接映射中的多个输出组自定义转换

未配置为生成事务的多个输出组自定义转换既不会生成也不会传播提交。因此，集成服务可以根据映射中使用的转换对此会话中的目标使用不同的提交类型：

- **在自定义转换和目标之间放置一个提交源。**集成服务对目标使用基于源的提交，因为它从活动源接收提交。活动源是目标的提交源。
- **不在自定义转换和目标之间放置提交源。**集成服务对目标使用基于目标的提交，因为它不接收任何提交。

下图显示了具有多个输出组自定义转换的映射：



该映射包含一个多输出组自定义转换（即 CT_XML_Parser），该转换可删除由源限定符转换生成的提交。因此，目标 T_store_name 和 T_store_addr 不会接收由活动源生成的任何提交。集成服务在加载到这些目标时使用基于目标的提交。

但是，该映射在自定义转换和 T_store_orders 之间包含一个可生成提交的活动源（即 AGG_store_orders）。AGG_store_orders 的转换范围是“全部输入”。集成服务在加载到 T_store_orders 时使用基于源的提交。

注意：可以将自定义转换配置为在自定义转换过程输出事务时生成事务。执行此操作时，请为会话配置用户定义的提交。

用户定义的提交

在用户定义的提交会话期间，集成服务根据通过事务控制转换传递的一行或一组行提交和回滚事务。集成服务针对输入转换的每行计算事务控制表达式。事务控制表达式的返回值定义了提交或回滚点。

还可以在映射包含配置为生成事务的自定义转换时创建用户定义的提交会话。执行此操作时，与自定义转换关联的过程定义了事务边界。

集成服务计算提交行时，会将事务中的所有行提交给一个或多个目标。它计算回滚行时，会从一个或多个目标回滚事务中的所有行。集成服务会在每个提交点和回滚点向会话日志中写入一条消息。会话详细信息是可以累积的。以下消息是会话日志中的示例提交消息：

```
WRITER_1_1_1> WRT_8317
USER-DEFINED COMMIT POINT Wed Oct 15 08:15:29 2003
=====
WRT_8036 Target: TCustOrders (Instance Name: [TCustOrders])
WRT_8038 Inserted rows - Requested: 1003      Applied: 1003      Rejected: 0      Affected: 1023
```

集成服务将事务中的所有行写入所有目标时，会按顺序对每个目标发出提交命令。

集成服务根据事务控制表达式或错误处理配置的返回值回滚数据。如果事务控制表达式返回回滚值，则集成服务会回滚事务。如果出现错误，可以选择在下一个提交点回滚或提交。

如果事务控制表达式的计算结果为除提交、回滚或继续之外的值，集成服务将导致会话失败。

会话完成时，集成服务可能会将数据写入未被提交行绑定的目标中。您可以选择在文件结尾提交或回滚此打开事务。

注意：如果对用户定义的提交会话使用批量加载，目标可能无法识别事务边界。如果目标连接组不支持事务，集成服务会向会话日志中写入以下消息：

```
WRT_8324 Warning: Target Connection Group's connection doesn't support transactions. Targets may not be loaded according to specified transaction boundaries rules.
```

回滚事务

在以下情况下，集成服务将回滚事务：

- **回滚计算。**事务控制表达式返回回滚值。
- **打开事务。**您选择在文件结尾回滚。
- **出错时回滚。**您选择在集成服务遇到非致命错误时回滚提交事务。
- **提交失败时回滚。**如果事务控制单元中的任何目标连接组都无法提交，集成服务会将所有未提交数据回滚至最后一个成功的提交点。

回滚计算

如果事务控制表达式返回回滚值，则集成服务会回滚事务，并向会话日志中写入一条指示事务已回滚的消息。它还指示了已回滚的行数。

以下消息是在事务控制表达式返回回滚值时集成服务向会话日志中写入的示例消息：

```
WRITER_1_1_1> WRT_8326 User-defined rollback processed
WRITER_1_1_1> WRT_8331 Rollback statistics
WRT_8162 =====
WRT_8330 Rolled back [333] inserted, [0] deleted, [0] updated rows for the target [TCustOrders]
```

回滚打开事务

如果事务控制表达式中的最后一行的计算结果为 TC_CONTINUE_TRANSACTION，会话将完成，但有打开事务。如果选择回滚此打开事务，则集成服务会回滚此事务，并向会话日志中写入一条指示事务已回滚的消息。

以下消息是指示已在会话属性中禁用“在文件结尾提交”的示例消息：

```
WRITER_1_1_1> WRT_8168 End loading table [TCustOrders] at: Wed Nov 05 10:21:56 2003
WRITER_1_1_1> WRT_8325 Final rollback executed for the target [TCustOrders] at end of load
```

以下消息是指示已在会话属性中启用“在文件结尾提交”的示例消息：

```
WRITER_1_1_1> WRT_8143
Commit at end of Load Order Group Wed Nov 05 08:15:29 2003
```

出错时回滚

您可以选择在集成服务遇到非致命错误时在下一个提交点回滚事务。集成服务在遇到非致命错误时，会处理错误行并继续处理事务。如果事务边界是提交行，集成服务会回滚整个事务并将其写入拒绝文件中。

下表介绍了已回滚事务的拒绝文件中的行指示器：

行指示器	说明
4	回滚插入
5	回滚更新
6	回滚删除

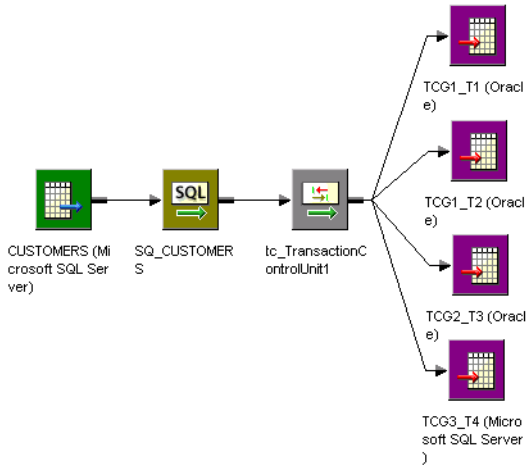
注意: 如果集成服务在通过事务控制转换处理任何行之前遇到错误，则它不会回滚事务。

提交失败时回滚

当集成服务达到事务控制单元中所有目标的提交点时，会按顺序对每个目标发出提交命令。如果事务控制单元中的任何目标连接组都无法提交，集成服务会将所有数据回滚至最后一个成功的提交点。集成服务无法回滚已提交的事务，但会将这些事务写入拒绝文件中。

例如，您创建了一个具有一个事务控制单元和三个目标连接组的映射。目标名称包含有关目标连接组的信息。TCG1_T1 表示第一个目标连接组和第一个目标。

下图显示了集成服务在提交失败时回滚的情况下的行为：



集成服务在处理该映射时使用以下逻辑：

1. 集成服务达到所有目标的第三个提交点。
2. 它开始按顺序对每个目标发出提交命令。
3. 集成服务成功提交到 TCG1_T1 和 TCG1_T2。
4. 针对 TCG2_T3 的提交失败。
5. 集成服务不会对 TCG3_T4 发出提交命令。
6. 集成服务将 TCG2_T3 和 TCG3_T4 回滚至第二个提交点，但无法将 TCG1_T1 和 TCG1_T2 回滚至第二个提交点，因为它已在第三个提交点成功提交。
7. 集成服务将 TCG2_T3 和 TCG3_T4 中的行写入拒绝文件。这些是与第三个提交点关联的回滚行。
8. 集成服务将 TCG_T1 和 TCG1_T2 中的行写入拒绝文件。这些是与第三个提交点关联的提交行。

下表介绍了失败的事务控制单元中已提交事务的拒绝文件中的行指示器：

行指示器	说明
7	已提交的插入
8	已提交的更新
9	已提交的删除

了解事务控制

PowerCenter 允许您定义集成服务在处理转换及提交并回滚目标数据时使用的事务。可以根据不同数量的输入行来定义事务。事务是一组由提交行或回滚行绑定的行，即事务边界。一些行可能未由事务边界绑定。这一组行是打开事务。您可以选择在文件结尾提交或在配置会话时回滚打开事务。

集成服务可以针对转换（每次一行）、事务中的所有行或所有源行处理输入行。针对事务中的所有行处理转换允许您在实时会话中包含汇总器等转换。

事务边界源自事务控制点。事务控制点是使用以下方式定义或重新定义事务边界的转换：

- **生成事务边界。**用于定义事务边界的转换因会话提交类型而异：
 - **基于目标的提交和用户定义的提交。**事务生成器会生成事务边界。事务生成器是可生成提交行和回滚行的转换。事务控制转换和自定义转换是事务生成器。
 - **基于源的提交。**某些活动源可生成提交。它们不会生成回滚行。此外，事务生成器也会生成提交行和回滚行。
- **删除传入事务边界。**转换删除传入事务边界且不生成提交时，集成服务会将所有行输出到打开事务中。所有生成提交的活动源和事务生成器会删除传入事务边界。

转换范围

可以使用“转换范围”转换属性配置集成服务如何将转换逻辑应用于传入数据。集成服务在处理转换时，会删除事务边界或保留事务边界，具体取决于转换范围和映射配置。

您可以为转换范围选择以下值之一：

- **行。**Applies the transformation logic to one row of data at a time. 当某个数据行不依赖于任何其他行时，请选择“行”。为连接到多个上游事务控制点的转换选择“行”时，集成服务会删除事务边界并将转换中的所有行作为打开事务输出。为连接到单个上游事务控制点的转换选择“行”时，集成服务会保留事务边界。
- **事务。**将转换逻辑应用至事务中的所有行。当一行数据取决于同一事务中的所有行，但不取决于其他事务中的行时，请选择“事务”。选择“事务”时，集成服务会保留传入事务边界。它会在接收到新事务时重置任何缓存（如汇总器或查找缓存）。
为多输入组转换选择“事务”时，必须将所有输入组连接至同一上游事务控制点。
- **全部输入。**将转换逻辑应用于所有传入数据。选择“全部输入”时，集成服务会删除传入事务边界并将转换中的所有行作为打开事务输出。当一行数据取决于源中的所有行时，请选择“全部输入”。

下表列出了每个转换可用的转换范围值：

转换	行	事务	全部输入
汇总器	-	可选。	默认值。 事务控制点。
应用程序源限定符	不适用 事务控制点。	-	-
自定义	可选。 配置为生成提交或连接到 多个上游事务控制点时的 事务控制点。	可选。 配置为生成提交时的 事务控制点。	默认值。 始终为事务控制点。 当它具有一个输出组或配置 为生成提交时生成提交。否 则，它会生成打开事务。

转换	行	事务	全部输入
数据屏蔽	默认值。只读。	-	-
表达式	默认值。不显示。	-	-
外部过程	默认值。不显示。	-	-
筛选器	默认值。不显示。	-	-
HTTP	默认值。只读。	-	-
Java	对于被动转换来说是默认的。	对于主动转换来说是可选的。	对于主动转换来说是默认的。
联接器	-	可选。	默认值。 事务控制点。
查找	默认值。不显示。	-	-
MQ 源限定符	不适用 事务控制点。	-	-
规范器 (VSAM)	不适用 事务控制点。	-	-
规范器 (关系)	默认值。不显示。	-	-
等级	-	可选。	默认值。 事务控制点。
路由器	默认值。不显示。	-	-
排序器	-	可选。	默认值。 事务控制点。
序列生成器	默认值。不显示。	-	-
源限定符	不适用 事务控制点。	-	-
SQL	对于脚本或查询模式 SQL 转换来说是默认的。	可选。 配置为生成提交时的事务控制点。	可选。
存储过程	默认值。不显示。	-	-
事务控制	默认值。不显示。 事务控制点。	-	-
联合	默认值。不显示。	-	-
非结构化数据	默认值。只读。	-	-

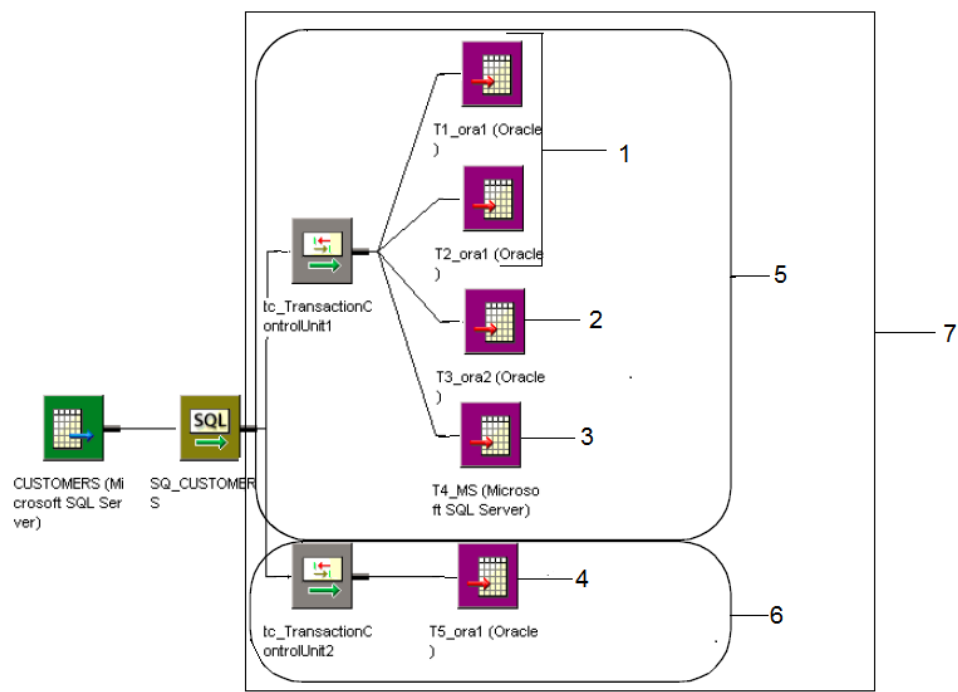
转换	行	事务	全部输入
更新策略	默认值。不显示。	-	-
XML 生成器	-	可选。 设置提交时刷新以创建新文档时的事务。	默认值。不显示。
XML 解析器	默认值。不显示。	-	-
XML 源限定符	不适用 事务控制点。	-	-

了解事务控制单元

事务控制单元是连接到可生成提交的活动源或有效事务生成器的目标组。事务控制单元是目标加载顺序组的子集，可能包含多个目标连接组。

当集成服务达到事务控制单元中所有目标的提交点时，会按顺序对每个目标发出提交命令。

下图显示了包含事务控制转换的事务控制单元：



1. 目标连接组 1
2. 目标连接组 2
3. 目标连接组 3
4. 目标连接组 4
5. 目标连接单元 1
6. 目标连接单元 2

7. 目标加载顺序组

在此示例中，T5_oral 使用与 T1_oral 和 T2_oral 相同的连接名称。由于 T5_oral 连接到单独的事务控制转换，因此它位于单独的事务控制单元和目标连接组中。如果将 T5_oral 连接到 tc_TransactionControlUnit1，它将与所有目标位于同一个事务控制单元中，并与 T1_oral 和 T2_oral 位于同一个目标连接组中。

使用事务控制的规则和准则

使用事务控制时，请考虑以下规则和准则：

- 具有“事务”转换范围的转换必须从单个事务控制点接收数据。
- 集成服务对具有“事务”转换范围的转换使用由第一个上游事务控制点定义的事务边界。
- 事务生成器可以对目标有效或无效。集成服务在将数据加载到目标时使用由有效事务生成器生成的事务。
- Workflow Manager 可防止您在包含汇总器转换（具有“事务”转换范围）的会话中使用增量汇总。
- 具有“全部输入”转换范围的转换会导致事务生成器变得对用户定义的提交会话中的目标无效。
- 集成服务会在具有“事务”转换范围的汇总器、联接器、等级和排序器转换的每个事务开始重置任何缓存。
- 您可以在使用已排序输入时为联接器转换选择“事务”转换范围。
- 在具有“事务”转换范围的转换上添加分区点时，Workflow Manager 会默认使用传递分区类型。您无法更改分区类型。

按事务创建目标文件

每次集成服务启动新事务时，可以生成单独的输出文件。可以动态命名每个目标平面文件。

要为每个事务生成单独的输出文件，请向平面文件目标定义中添加一个 FileName 端口。连接映射中的 FileName 端口时，PowerCenter 会在每次提交时写入单独的目标文件。集成服务使用每个事务中第一行的 FileName 端口值命名输出文件。

设置提交属性

创建会话时，可以配置提交属性。设置的属性取决于映射的类型以及您希望集成服务执行的提交类型。在“属性”选项卡的“常规选项”设置中配置提交属性。

下表介绍了在“属性”选项卡的“常规选项”设置中设置的会话提交属性：

属性	基于目标	基于源	用户定义
提交类型	如果映射中不存在任何事务生成器或仅存在无效的事务生成器，则默认情况下会选择此选项。	如果映射中不存在任何事务生成器或仅存在无效的事务生成器，则会为基于源的提交选择此选项。	如果映射中存在有效的事务生成器，则默认情况下会选择此选项。
提交间隔	默认值为 10,000。	默认值为 10,000。	不适用

属性	基于目标	基于源	用户定义
在文件结尾提交	在文件结尾提交数据。默认已启用。无法禁用此选项。	在文件结尾提交数据。如果您希望集成服务回滚打开事务，请清除此选项。	在文件结尾提交数据。如果您希望集成服务回滚打开事务，请清除此选项。
出错时回滚事务	如果集成服务遇到非致命错误，则可以选择在下一个提交点回滚事务。 集成服务遇到转换错误时，如果错误出现在目标的有效事务生成器之后，则集成服务会回滚事务。	如果集成服务遇到非致命错误，则可以选择在下一个提交点回滚事务。 集成服务遇到转换错误时，如果错误出现在目标的有效事务生成器之后，则集成服务会回滚事务。	如果集成服务遇到非致命错误，则可以选择在下一个提交点回滚事务。 集成服务遇到转换错误时，如果错误出现在目标的有效事务生成器之后，则集成服务会回滚事务。

提示: 批量加载到 Microsoft SQL Server 或 Oracle 目标时，请定义较大的提交间隔。Microsoft SQL Server 和 Oracle 每次提交后启动新的批量加载事务。增加提交间隔可减少批量加载事务的数目并提高性能。

第 8 章

行错误日志记录

本章包括以下主题：

- [行错误日志记录概览, 135](#)
- [了解错误日志表, 136](#)
- [了解错误日志文件, 140](#)
- [配置错误日志选项, 142](#)

行错误日志记录概览

配置会话时，可以在中心位置记录行错误。出现行错误时，集成服务会记录错误信息，通过此错误信息可以确定错误的原因和根源。集成服务会记录源名称、行 ID、当前行数据、转换、时间戳、错误代码、错误消息、存储库名称、文件夹名称、会话名称和映射信息等信息。

您可以将行错误记录到关系表或平面文件中。启用错误日志记录时，集成服务会在首次运行会话时创建错误表或错误日志文件。错误日志是可以累积的。如果存在错误日志，集成服务会将错误数据附加到现有错误日志中。

可以从平面文件或关系源中记录源行数据。源行数据包括来自出现错误的源限定符的行数据、源行 ID 和源行类型。您无法记录 XML 文件源的行错误。可以在会话日志中查看 XML 源错误。

如果错误出现在具有多个分区或以下某个活动源的非传递分区点之后，则集成服务无法标识包含错误的源限定符中的行：

- 汇总器
- 自定义，配置为一个活动转换
- 联接器
- 规范器（管道）
- 等级
- 排序器

默认情况下，集成服务在会话日志中记录转换错误，在拒绝文件中记录拒绝行。启用错误日志记录时，集成服务不会生成拒绝文件或将已删除的行写入会话日志。如果没有拒绝文件，则集成服务不会记录事务控制转换回滚或提交错误。如果除了将行写入到行错误日志之外还要写入到会话日志，则可以启用详细数据跟踪。

注意：记录行错误时，会话性能可能会降低，因为集成服务一次处理一行，而不是一次处理许多行。

错误日志代码页

集成服务根据集成服务进程操作系统以不同方式将数据写入错误日志文件：

- **UNIX.**集成服务使用集成服务进程代码页将数据写入错误日志文件。 但是，您可以通过启用 LogsInUTF8 集成服务属性将集成服务配置为使用 UTF-8 写入错误日志文件。
- **Windows.**集成服务使用 UTF-8 编码格式将所有字符写入错误日志文件。

存在错误表的关系数据库的代码页必须是目标代码页的子集。 如果错误日志表代码页不是目标代码页的子集，则集成服务可能会在错误日志表中写入不一致数据。

了解错误日志表

选择关系数据库错误日志记录时，集成服务会在首次运行会话时创建以下错误表：

- **PMERR_DATA.**存储有关转换行错误及其对应源行的数据和元数据。
- **PMERR_MSG.**存储有关错误和错误消息的元数据。
- **PMERR_SESS.**存储有关会话的元数据。
- **PMERR_TRANS.**在出现转换错误时，存储有关源和转换端口的元数据（如名称和数据类型）。

可以指定与集成服务在其上创建表的数据库的数据库连接。 如果会话存在错误表，则集成服务会将行错误附加到这些表中。

通过关系数据库错误日志记录，您可以在一组错误表中收集来自多个会话的行错误。 要执行该操作，需要为所有会话指定相同的错误日志表名称前缀。 可以对生成的错误表发出 select 语句，以便针对特定会话检索错误数据。

可以为错误表指定前缀。 错误表名称最多可具有十一个字符。 命名 Oracle、Sybase 或 Teradata 错误日志表时，不要指定超过 19 个字符的前缀，因为这些数据库中表名称的长度不能超过 30 个字符。 可以对表名称前缀使用参数或变量。 可使用您在参数文件中可定义的任何参数或变量类型。 例如，可以使用会话参数 \$ParamMyErrPrefix 作为错误日志表名称前缀，并在参数文件中将 \$ParamMyErrPrefix 设置为表前缀。

集成服务可以在不指定主键和外键的情况下创建错误表。 但是，可以指定键列。

PMERR_DATA

集成服务在遇到行错误时，会向 PMERR_DATA 表中插入一个条目。 此表存储有关转换行错误及其对应源行的数据和元数据。

下表介绍了 PMERR_DATA 表的结构：

列名称	数据类型	说明
REPOSITORY_GID	变长字符型	存储库的唯一标识符。
WORKFLOW_RUN_ID	整数	工作流的唯一标识符。
WORKLET_RUN_ID	整数	工作集的唯一标识符。 如果会话不属于工作集，则该值为“0”。
SESS_INST_ID	整数	会话的唯一标识符。
TRANS_MAPPLET_INST	变长字符型	出现错误的 Mapplet 的名称。

列名称	数据类型	说明
TRANS_NAME	变长字符型	出现错误的转换的名称。
TRANS_GROUP	变长字符型	出现错误的输入组或输出组的名称。如果转换不具有组，则默认为“输入”或“输出”。
TRANS_PART_INDEX	整数	指定出现错误的转换的分区编号。
TRANS_ROW_ID	整数	指定由上一个活动源生成的行 ID。
TRANS_ROW_DATA	大型变长字符串	<p>包含所有列数据（包括列指示器）的分隔字符串。列指示器如下：</p> <p>D - 有效</p> <p>N - 空</p> <p>T - 已截断</p> <p>B - 二进制</p> <p>U - 数据不可用</p> <p>列数据与列指示器之间的固定分隔符为冒号 (:)。各个列之间的分隔符为管道 ()。可以在错误处理设置中替代列分隔符。集成服务会将错误表中的所有列数据转换为文本字符串。对于二进制数据，集成服务仅使用列指示器。</p> <p>该值可以跨越多行。当数据超过 2000 字节时，集成服务会创建新行。每个行错误条目的行号存储在 LINE_NO 列中。</p>
SOURCE_ROW_ID	整数	源限定符分配给其读取的每行的值。如果集成服务无法标识行，则该值为 -1。
SOURCE_ROW_TYPE	整数	<p>告知行是被标记为插入、更新、删除还是拒绝的行指示器。</p> <p>0 - 插入</p> <p>1 - 更新</p> <p>2 - 删除</p> <p>3 - 拒绝</p>
SOURCE_ROW_DATA	大型变长字符串	<p>包含所有列数据（包括列指示器）的分隔字符串。列指示器如下：</p> <p>D - 有效</p> <p>O - 溢出</p> <p>N - 空</p> <p>T - 已截断</p> <p>B - 二进制</p> <p>U - 数据不可用</p> <p>列数据与列指示器之间的固定分隔符为冒号 (:)。各个列之间的分隔符为管道 ()。可以在错误处理设置中替代列分隔符。集成服务会将错误表或错误文件中的所有列数据转换为文本字符串。对于二进制数据，集成服务仅使用列指示器。</p> <p>该值可以跨越多行。当数据超过 2000 字节时，集成服务会创建新行。每个行错误条目的行号存储在 LINE_NO 列中。</p>
LINE_NO	整数	为跨越多行的 SOURCE_ROW_DATA 和 TRANS_ROW_DATA 中的每个行错误条目指定行号。
注意: 使用以粗体显示的列名称可联接表。		

PMERR_MSG

集成服务在遇到行错误时，会向 PMERR_MSG 表中插入一个条目。此表存储有关错误和错误消息的元数据。

下表介绍了 PMERR_MSG 表的结构：

列名称	数据类型	说明
REPOSITORY_GID	变长字符型	存储库的唯一标识符。
WORKFLOW_RUN_ID	整数	工作流的唯一标识符。
WORKLET_RUN_ID	整数	工作集的唯一标识符。如果会话不属于工作集，则该值为“0”。
SESS_INST_ID	整数	会话的唯一标识符。
MAPPLET_INST_NAME	变长字符型	转换所属的 Mapplet。如果转换不属于 Mapplet，则该值为 n/a。
TRANS_NAME	变长字符型	出现错误的转换的名称。
TRANS_GROUP	变长字符型	出现错误的输入组或输出组的名称。如果转换不具有组，则默认为“输入”或“输出”。
TRANS_PART_INDEX	整数	指定出现错误的转换的分区编号。
TRANS_ROW_ID	整数	指定由上一个活动源生成的行 ID。
ERROR_SEQ_NUM	整数	每个转换组中每行错误数的计数器。如果会话具有多个分区，则集成服务会为每个分区保持此计数器。 例如，如果转换在分区 1 中生成三个错误，在分区 2 中生成两个错误，则 ERROR_SEQ_NUM 会为分区 1 生成值 1、2 和 3，为分区 2 生成值 1 和 2。
ERROR_TIMESTAMP	日期/时间	出现错误时集成服务的时间戳。
ERROR_UTC_TIME	整数	出现错误时的协调世界时（称为格林尼治标准时间）。
ERROR_CODE	整数	错误生成的错误代码。
ERROR_MSG	大型变长字符串	可跨越多行的错误消息。当数据超过 2000 字节时，集成服务会创建新行。每个行错误条目的行号存储在 LINE_NO 列中。
ERROR_TYPE	整数	所出现错误的类型。集成服务使用以下值： 1 - 读取器错误 2 - 写入器错误 3 - 转换错误
LINE_NO	整数	为跨越多行的 ERROR_MSG 中的每个行错误条目指定行号。
注意: 使用以粗体显示的列名称可联接表。		

PMERR_SESS

选择关系数据库错误日志记录时，集成服务会向 PMERR_SESS 表中插入条目。此表存储出现错误时会话的相关元数据。

下表介绍了 PMERR_SESS 表的结构：

列名称	数据类型	说明
REPOSITORY_GID	变长字符型	存储库的唯一标识符。
WORKFLOW_RUN_ID	整数	工作流的唯一标识符。
WORKLET_RUN_ID	整数	工作集的唯一标识符。如果会话不属于工作集，则该值为“0”。
SESS_INST_ID	整数	会话的唯一标识符。
SESS_START_TIME	日期/时间	会话启动时集成服务的时间戳。
SESS_START_UTC_TIME	整数	会话启动时的协调世界时（称为格林尼治标准时间）。
REPOSITORY_NAME	变长字符型	存储会话的存储库的名称。
FOLDER_NAME	变长字符型	指定映射和会话所在的文件夹。
WORKFLOW_NAME	变长字符型	指定运行正记录的会话的工作流。
TASK_INST_PATH	变长字符型	可跨越多行的完全限定会话名称。集成服务会为会话名称创建一个新行。集成服务还会为限定的会话名称中的每个工作集创建一个新行。例如，您具有一个名为 WL1.WL2.S1 的会话。该名称的每个组成部分都会显示在新行上： WL1 WL2 S1 集成服务会在 LINE_NO 列中写入行号。
MAPPING_NAME	变长字符型	指定会话使用的映射。
LINE_NO	整数	为跨越多行的 TASK_INST_PATH 中的每个行错误条目指定行号。
注意： 使用以粗体显示的列名称可联接表。		

PMERR_TRANS

集成服务在遇到转换错误时，会向 PMERR_TRANS 表中插入一个条目。此表存储源和转换端口的名称和数据类型等元数据。

下表介绍了 PMERR_TRANS 表的结构：

列名称	数据类型	说明
REPOSITORY_GID	变长字符型	存储库的唯一标识符。
WORKFLOW_RUN_ID	整数	工作流的唯一标识符。
WORKLET_RUN_ID	整数	工作集的唯一标识符。 如果会话不属于工作集，则该值为“0”。
SESS_INST_ID	整数	会话的唯一标识符。
TRANS_MAPPLET_INST	变长字符型	指定 Mapplet 的实例。
TRANS_NAME	变长字符型	出现错误的转换的名称。
TRANS_GROUP	变长字符型	出现错误的输入组或输出组的名称。 如果转换不具有组，则默认为“输入”或“输出”。
TRANS_ATTR	变长字符型	列出出现错误的输入或输出组的端口名称和数据类型。 端口名称和数据类型对以逗号分隔，例如：portname1:datatype, portname2:datatype. 该值可以跨越多行。 数据超过 2000 字节时，集成服务会为转换属性创建一个新行并在 LINE_NO 列中写入行号。
SOURCE_MAPPLET_INST	变长字符型	源所在的 Mapplet 的名称。
SOURCE_NAME	变长字符型	源限定符的名称。 当行错误出现在并非源限定符的活动源或具有多个分区的非传递分区点的下游时，将显示 n/a。
SOURCE_ATTR	变长字符型	在出现错误的源限定符中列出连接的字段。 在多个字段中出现错误时，将在新行上输入每个字段名称。 在 LINE_NO 列中写入行号。
LINE_NO	整数	为跨越多行的 TRANS_ATTR 和 SOURCE_ATTR 中的每个行错误条目指定行号。
注意: 使用以粗体显示的列名称可联接表。		

了解错误日志文件

可以创建一个错误日志文件以收集会话中出现的所有错误。 此错误日志文件是一个列分割线顺序文件。 通过指定唯一的错误日志文件名，可以为工作流中的每个会话创建一个单独的日志文件。 如果要分析某个会话的行错误，请使用错误日志文件。

在错误日志文件中，双管道 “||” 用于分隔错误日志记录列。 默认情况下，管道 “|” 用于分隔行数据。 可以通过设置 “数据列分隔符” 错误日志选项来更改此行数据分隔符。

错误日志文件具有以下结构：

```
[Session Header]
[Column Header]
[Column Data]
```

会话标头包含类似 PMERR_SESS 表中存储信息的会话运行信息。列标头包含数据列名称。列数据包含行数据和错误消息信息。

下表介绍了错误日志文件中的列：

日志文件列标头	说明
转换	出现错误的映射使用的转换名称。
转换 Mapplet 名称	包含转换的 Mapplet 的名称。当此信息不可用时，将显示 n/a。
转换组	出现错误的输入组或输出组的名称。如果转换不具有组，则默认为“输入”或“输出”。
分区索引	指定出现错误的转换分区的分区编号。
转换行 ID	指定错误行的行 ID。
错误序列	每个转换组中每行错误数的计数器。如果会话具有多个分区，则集成服务会为每个分区保持此计数器。 例如，如果转换在分区 1 中生成三个错误，在分区 2 中生成两个错误，则 ERROR_SEQ_NUM 会为分区 1 生成值 1、2 和 3，为分区 2 生成值 1 和 2。
错误时间戳	出现错误时集成服务的时间戳。
错误 UTC 时间	出现错误时的协调世界时（称为格林尼治标准时间）。
错误代码	对应于错误消息的错误代码。
错误消息	错误消息。
错误类型	所出现错误的类型。集成服务使用以下值： 1 - 读取器错误 2 - 写入器错误 3 - 转换错误
转换数据	包含所有列数据（包括列指示器）的分隔字符串。列指示器如下： D - 有效 O - 溢出 N - 空 T - 已截断 B - 二进制 U - 数据不可用 列数据与列指示器之间的固定分隔符为冒号 (:)。各个列之间的分隔符为管道 ()。可以在错误处理设置中替代列分隔符。 集成服务会将错误文件中的所有列数据转换为文本字符串。对于二进制数据，集成服务仅使用列指示器。
源名称	源限定符的名称。当行错误出现在并非源限定符的活动源或具有多个分区的非传递分区点的下游时，将显示 N/A。
源行 ID	源限定符分配给其读取的每行的值。如果集成服务无法标识行，则该值为 -1。

日志文件列标头	说明
源行类型	告知行是被标记为插入、更新、删除还是拒绝的行指示器。 0 - 插入 1 - 更新 2 - 删除 3 - 拒绝
源数据	包含所有列数据（包括列指示器）的分隔字符串。列指示器如下： D - 有效 O - 溢出 N - 空 T - 已截断 B - 二进制 U - 数据不可用 列数据与列指示器之间的固定分隔符为冒号 (:)。各个列之间的分隔符为管道 ()。 可以在错误处理设置中替代列分隔符。 集成服务会将错误表或错误文件中的所有列数据转换为文本字符串。对于二进制数据，集成服务仅使用列指示器。

配置错误日志选项

可以在会话属性的“配置对象”选项卡上为每个会话配置错误日志记录。启用错误日志记录时，可以选择在关系数据库中或以平面文件形式创建错误日志。如果不启用错误日志记录，则集成服务不会创建错误日志。

提示: 使用 Workflow Manager 可为“配置对象”选项卡创建一组可重用的属性。

要配置错误日志记录选项，请执行以下操作：

1. 双击会话任务打开会话属性。
2. 选择“配置对象”选项卡。
3. 指定错误日志类型。

下表介绍了“配置对象”选项卡的错误日志记录设置：

错误日志选项	说明
错误日志类型	指定要创建的错误日志的类型。可以指定关系数据库、平面文件或无。默认值为“无”。 注意: 您无法记录 XML 文件源的行错误。可以在会话日志中查看 XML 源错误。
错误日志数据库连接	为关系日志指定数据库连接。启用关系数据库日志记录时，需要此选项。

错误日志选项	说明
错误日志表名称前缀	为关系日志指定表名称前缀。集成服务会将 11 个字符附加到前缀名称。Oracle 和 Sybase 限制表名称不超过 30 个字符。如果表名称超过 30 个字符，会话将失败。 可以对错误日志表名称前缀使用参数或变量。可使用您在参数文件中可定义的任何参数或变量类型。
错误日志文件目录	指定错误日志文件的目录。默认情况下，错误日志文件的目录为 \$PMBadFilesDir\。启用平面文件日志记录时，需要此选项。
错误日志文件名	指定错误日志文件名。错误日志文件名的字符限制为 255。默认情况下，错误日志文件名为 PMError.log。启用平面文件日志记录时，需要此选项。
记录行数据	指定是否记录转换行数据。当您启用错误日志记录时，集成服务默认会记录转换行数据。如果您禁用此属性，转换行数据字段中将显示 n/a 或 -1。
记录源行数据	如果选择不记录源行数据，或者如果源行数据不可用，则集成服务会写入 n/a 或 -1 等指示器，具体取决于列数据类型。 如果不需要捕获源行数据，请考虑禁用此选项以提高集成服务性能。
数据列分隔符	字符串类型的源行数据和转换组行数据的分隔符。默认情况下，集成服务使用竖线 () 分隔符。请确认没有为行数据使用与错误日志列相同的分隔符。如果您使用相同的分隔符，可能会发现错误日志文件难以阅读。

4. 单击“确定”。

第 9 章

workflow 恢复

本章包括以下主题：

- [workflow 恢复概览, 144](#)
- [运行状态, 145](#)
- [恢复选项, 148](#)
- [挂起 workflow, 149](#)
- [配置 workflow 恢复, 149](#)
- [配置任务恢复, 150](#)
- [继续进行会话, 152](#)
- [使用可重复数据, 153](#)
- [恢复 workflow 和任务的步骤, 157](#)
- [会话恢复的规则和准则, 158](#)

workflow 恢复概览

通过 workflow 恢复，您可以从中断点继续处理 workflow 和 workflow 任务。如果集成服务可以访问 workflow 运行状态，则可以恢复 workflow。 workflow 运行状态包括 workflow 中的任务状态和 workflow 变量值。集成服务将状态存储在内存或磁盘中，具体取决于 workflow 的配置方式：

- **启用恢复。**为 workflow 启用恢复后，集成服务会将 workflow 运行状态保存在一个共享位置。可以在 workflow 终止、停止或中止时将其恢复。 workflow 无须处于运行状态。
- **挂起。**如果将 workflow 配置为在出错时挂起，则集成服务会将 workflow 运行状态存储在内存中。可以在任务失败时恢复已挂起 workflow。可以修复该任务错误并恢复 workflow。

集成服务根据 workflow 中任务的恢复策略来恢复这些任务。默认情况下，会话和命令任务的恢复策略是使任务失败并继续运行 workflow。可以为会话和命令任务配置恢复策略。所有其他任务的策略是重新启动任务。

具有高可用性时，PowerCenter 会在正在运行某一 workflow 的服务进程故障转移到其他节点时自动恢复该 workflow。可以对正在运行的 workflow 进行配置以在某一任务终止时自动恢复该任务。PowerCenter 还可在数据库连接中断后恢复会话和 workflow。

当集成服务以安全模式运行时，将存储配置为恢复的 workflow 的运行状态。如果 workflow 无法使集成服务故障转移到备份节点，则集成服务不会自动恢复该 workflow。如果您对集成服务具有相应的特权，可以手动恢复该 workflow。

运行状态

恢复工作流或会话时，集成服务会还原运行的工作流或会话运行状态以确定开始进行恢复处理的位置。集成服务会根据您配置工作流的方式将运行的工作流状态存储在内存或磁盘中。集成服务根据您的配置会话的方式存储会话运行状态。

工作流运行状态

为工作流启用恢复或挂起时，集成服务将存储工作流运行状态。将工作流挂起时，运行状态存储在内存中。

为工作流启用恢复后，集成服务会将工作流运行状态存储在共享位置 `$PMStorageDir`。集成服务可以还原运行状态以恢复已停止、中止或终止的工作流。执行恢复时，将还原运行状态以从中断点恢复工作流。工作流完成后，集成服务将从共享文件夹中删除工作流运行状态。

工作流运行状态包含以下信息：

- 活动服务请求
- 已完成和正在运行的任务状态
- 工作流变量值

运行并发工作流时，集成服务将实例名称或工作流运行 ID 附加到 `$PMStorageDir` 中的工作流恢复存储文件。

为工作流启用恢复后，默认情况下，集成服务将不会存储会话运行状态。可以对会话恢复策略进行配置以保存会话运行状态。

会话运行状态

将会话恢复策略配置为从上一个检查点恢复时，集成服务会将会话运行状态存储在共享位置 `$PMStorageDir` 中。集成服务还会将关系目标恢复信息保存在目标数据库表中。集成服务在执行恢复时，会还原运行状态以从中断点恢复会话。它使用目标恢复数据来确定如何恢复目标表。

您可以将会话配置为保存会话运行状态，即使未保存工作流运行状态。可以恢复会话，也可以从会话恢复工作流。

会话运行状态包括以下信息：

- **源。**如果源的输出不确定且不可重复，则集成服务会将 SQL 查询的结果保存到 `$PMStorageDir` 中的共享存储文件。
- **转换。**集成服务会在 `$PMStorageDir` 中创建检查点，以确定在运行恢复会话时开始处理管道的位置。
运行增量汇总器转换的会话时，集成服务会在会话开始运行时在 `$PMCacheDir` 目录中创建汇总器缓存文件的备份。集成服务会在会话恢复开始运行时将备份缓存升级至初始缓存。
- **关系目标恢复数据。**集成服务会将恢复信息写入目标数据库中的恢复表，以确定在会话中断时提交到目标的最后一行。

目标恢复表

集成服务运行具有继续恢复策略的会话时，将写入目标数据库系统上的恢复表中。集成服务恢复会话时，会使用恢复表中的信息来确定开始将数据加载到目标表的位置。

如果希望集成服务创建恢复表，请向目标数据库连接中配置的数据库用户名授予表创建特权。如果不希望集成服务创建恢复表，请手动创建恢复表。

集成服务会在目标数据库中创建以下恢复表：

- **PM_RECOVERY.**包含供会话运行的目标加载信息。集成服务会在每次成功会话后从此表中删除该信息，并在后续会话开始时初始化该信息。
- **PM_TGT_RUN_ID.**包含集成服务用来标识数据库上的每个目标的信息。该信息在两次会话运行之间仍位于表中。如果手动创建此表，必须创建行并为 LAST_TGT_RUN_ID 输入非零值以确保会话可成功恢复。
- **PM_REC_STATE.**包含集成服务用来确定在实时会话恢复期间是否需要将消息写入目标表的信息。

如果在恢复会话前编辑或删除恢复表，则集成服务将无法恢复会话。如果禁用恢复，则集成服务无法从目标数据库中删除恢复表。必须手动删除恢复表。

下表介绍了 PM_RECOVERY 的格式：

列名称	数据类型
REP_GID	VARCHAR(240)
WFLOW_ID	INTEGER
WFLOW_RUN_ID	INTEGER
WFLOW_RUN_INS_NAME	VARCHAR(240)
SUBJ_ID	INTEGER
TASK_INST_ID	INTEGER
TGT_INST_ID	INTEGER
PARTITION_ID	INTEGER
TGT_RUN_ID	INTEGER
RECOVERY_VER	INTEGER
CHECK_POINT	INTEGER
ROW_COUNT	INTEGER

下表介绍了 PM_TGT_RUN_ID 的格式：

列名称	数据类型
LAST_TGT_RUN_ID	INTEGER

下表介绍了 PM_REC_STATE 的格式：

列名称	数据类型
OWNER_TYPE_ID	INTEGER
REP_GID	VARCHAR(240)
FOLDER_ID	INTEGER

列名称	数据类型
WFLOW_ID	INTEGER
WFLOW_RUN_INS_NAME	VARCHAR(240)
WLET_ID	INTEGER
TASK_INST_ID	INTEGER
WID_INST_ID	INTEGER
GROUP_ID	INTEGER
PART_ID	INTEGER
PLUGIN_ID	INTEGER
APPL_ID	VARCHAR (38)
SEQ_NUM	INTEGER
版本	INTEGER
CHKP_NUM	INTEGER
STATE_DATA	VARCHAR (1024)

Oracle 使用 NUMBER 数据类型，而不是 INTEGER 数据类型。

注意: 将并发恢复会话写入相同目标数据库时，集成服务可能会在 PM_RECOVERY 上遇到死锁。要在死锁时重试写入 PM_RECOVERY，可以配置“死锁时重试会话”选项以重试会话死锁。

相关主题：

- [“PM_REC_STATE 表” 页面上 116](#)

创建目标恢复表

可以手动创建目标恢复表。Informatica 在以下目录中提供 SQL 脚本：

<PowerCenter installation_dir>\server\bin\RecoverySQL

运行以下脚本之一可在目标数据库中创建恢复表：

脚本	数据库
create_schema_db2.sql	IBM DB2
create_schema_inf.sql	Informix
create_schema_ora.sql	Oracle
create_schema_sql.sql	Microsoft SQL Server

脚本	数据库
create_schema_syb.sql	Sybase
create_schema_ter.sql	Teradata

恢复选项

要执行恢复，必须配置映射、工作流任务以及要恢复的工作流。

下表介绍了可为恢复配置的选项：

选项	位置	说明
出错时挂起工作流	工作流	工作流中的任务失败时挂起工作流。您可以修复失败的任务以及恢复挂起的工作流。
挂起电子邮件	工作流	工作流挂起时发送电子邮件。
启用 HA 恢复	工作流	在共享位置保存操作的工作流状态。您无需有高可用性选项即可启用工作流恢复。
自动恢复终止的任务	工作流	在工作流运行过程中恢复终止的会话和命令任务。您必须具有高可用性选项。
最大自动恢复尝试次数	工作流	集成服务尝试恢复会话或命令任务的次数。
恢复策略	会话、命令	会话或命令任务的恢复策略。决定集成服务如何在工作流恢复过程中恢复会话或命令任务以及如何在会话恢复过程中恢复会话+。
任何命令失败时使任务失败	命令	如果任务中的任何命令失败，则允许命令任务失败。如果您未设置此选项，任意命令失败后任务将继续运行。如果任务中的任何命令失败，您可以将此选项与“出错时挂起工作流”结合使用以挂起工作流。
输出是确定性的	转换	指示转换始终基于相同的输入数据生成一组相同的数据。当输出是可重复的确定性输出时，集成服务可以从上次检查点恢复会话。为关系源限定符启用了此选项以及“输出是可重复的”选项时，集成服务不会将 SQL 结果保存到共享存储。
输出具有可重复性	转换	指示转换是否在会话运行之间以相同的顺序生成行。如果输出是可重复的且具有确定性，集成服务可以从上一检查点继续运行会话。为关系源限定符启用了此选项以及“输出是确定性的”选项时，集成服务不会将 SQL 结果保存到共享存储。

警告: 如果将转换配置为可重复和确定性的，则您有责任确保数据是可重复的。如果尝试使用不会生成可重复和确定性数据的转换恢复会话，则恢复过程可能会导致数据损坏。

挂起工作流

如果工作流中的任务失败，您可能会希望挂起工作流，修复错误，然后再恢复工作流。如果启用了工作流属性中的“出错时挂起”选项，则集成服务会挂起工作流。或者，可以设置挂起电子邮件，以使集成服务在其挂起工作流时发送电子邮件。

如果对工作流启用了“出错时挂起”，则集成服务会在以下任务之一失败时挂起工作流：

- 会话
- 命令
- 工作集
- 电子邮件

工作流中的任务失败时，集成服务会停止运行路径中的任务。集成服务不会评估失败任务的输出链接。如果工作流中没有其他任务正在运行，则 Workflow Monitor 会将工作流的状态显示为“已挂起”。

如果您具有高可用性选项，则集成服务会根据自动任务恢复的设置方式挂起工作流。如果将工作流配置为“出错时挂起”并且未启用自动任务恢复，则工作流会在任务失败时挂起。如果启用了自动任务恢复，则集成服务会首先尝试重新启动任务，直至达到指定的恢复限制，如若无法重新启动失败的任务，然后才会挂起工作流。

如果某个任务失败时工作流中的一个或多个任务仍在运行，则集成服务会停止运行失败的任务并继续运行其他路径中的任务。Workflow Monitor 将工作流的状态显示为“正在挂起”。

工作流的状态为“已挂起”或“正在挂起”时，可以修复错误，如目标数据库错误，然后恢复 Workflow Monitor 中的工作流。恢复工作流时，集成服务会重新启动失败的任务并继续评估工作流中的其余任务。集成服务不会运行已成功完成的任何任务。

注意：编辑已挂起工作流或已挂起工作流中的任务可能导致存储库不一致。

要挂起工作流，请执行以下操作：

1. 在 Workflow Designer 中，打开工作流。
2. 单击“工作流”>“编辑”。
3. 在“常规”选项卡中，启用“出错时挂起”。
4. 单击“确定”。

配置挂起电子邮件

您可以对工作流进行配置以便集成服务在挂起工作流时发送电子邮件。请为挂起电子邮件选择现有的可重用电子邮件任务。任务失败时，集成服务将开始挂起工作流并发送挂起电子邮件。如果另一项任务在集成服务挂起工作流过程中失败，您不会再次收到挂起电子邮件。

如果另一项任务在您继续运行工作流后失败，集成服务将发送挂起电子邮件。

配置工作流恢复

要为工作流配置恢复，必须为工作流启用恢复或将工作流配置为在任务出错时挂起。为工作流配置了恢复后，可以在它停止、中止、终止或挂起时将其恢复。

下表介绍了每个可恢复的工作流状态：

状态	说明
已中止	可以在 Workflow Monitor 中或通过 <i>pmcmd</i> 中止工作流。还可以选择在 Administrator 工具中禁用服务进程时中止所有正在运行的工作流。如果您为工作流启用了恢复功能，可以恢复已中止的工作流。可以在 Workflow Monitor 中或通过使用 <i>pmcmd</i> 恢复已中止的工作流。
已停止	可以在 Workflow Monitor 中或通过 <i>pmcmd</i> 停止工作流。还可以选择在 Administrator 工具中禁用服务或服务进程时停止所有正在运行的工作流。如果您为工作流启用了恢复功能，可以恢复已停止的工作流。可以在 Workflow Monitor 中或通过使用 <i>pmcmd</i> 恢复已停止的工作流。
已挂起	任务失败，且工作流配置为在任务出错时挂起。如果多个任务正在运行，则集成服务会在所有正在运行的任务成功或失败时挂起工作流。可以在运行恢复之前修复导致一个或多个任务失败的错误。 默认情况下，工作流会在任务失败后继续。要在任务失败时挂起工作流，请将工作流配置为在任务出错时挂起。
已终止	运行工作流的服务进程意外关闭。运行工作流的所有节点上的任务终止。当工作流中的任务终止且没有高可用性选项时，工作流会终止。如果您为工作流启用了恢复功能，可以恢复已终止的工作流。当具有高可用性时，服务进程会故障转移到其他节点且工作流恢复将启动。
注意： 失败的工作流是完成时出现错误的工作流。您无法恢复失败的工作流。	

恢复已停止、已中止和已终止的工作流

为工作流启用恢复时，集成服务会在工作流运行期间将工作流的运行状态保存到文件中。您可以恢复已停止、已终止或已中止的工作流。在工作流的“属性”选项卡上启用恢复。

恢复挂起的工作流

如果工作流中的任务失败，则可以将工作流配置为挂起。默认情况下，工作流会在任务失败时继续运行。您可以在任务失败时挂起工作流，修复失败的任务，并恢复工作流。挂起工作流时，工作流的运行状态会保留在内存中。您可以修复导致任务失败的错误，并从中断点恢复工作流。如果任务再次失败，则集成服务会再次挂起工作流。您可以恢复已挂起的工作流，但无法重新启动它。在工作流属性的“常规”选项卡上将工作流配置为挂起。

还可以将工作流配置为在任务挂起时发送电子邮件。

配置任务恢复

恢复工作流时，集成服务将根据每个任务的恢复策略来恢复任务。恢复策略可能会使任务失败并继续执行工作流、从上一检查点恢复或重新启动任务，具体取决于任务。

启用工作流恢复时，可以恢复中止或停止的任务。您可以恢复由于网络或服务进程故障而终止的任务。如果您将工作流配置为出错时挂起，则可以在恢复工作流时恢复失败的任务。

下表介绍了每个可恢复任务的状态：

状态	说明
已中止	在 Workflow Monitor 中或通过 <i>pmcmd</i> 中止工作流或任务。在 Administrator 工具中禁用了服务或服务进程时，还可以选择中止所有正在运行的工作流。您还可以将会话配置为根据映射条件中止。 可以在 Workflow Monitor 中恢复工作流以恢复任务，也可以使用 <i>pmcmd</i> 恢复工作流。
已停止	在 Workflow Monitor 中或通过 <i>pmcmd</i> 停止工作流或任务。还可以选择在 Administrator 工具中禁用服务或服务进程时停止所有正在运行的工作流。 可以在 Workflow Monitor 中恢复工作流以恢复任务，也可以使用 <i>pmcmd</i> 恢复工作流。
已失败	集成服务由于错误使任务失败。如果工作流配置为出现任务故障时挂起，则可以使用工作流恢复来恢复失败的任务。如果工作流未挂起，您可以通过仅恢复会话或恢复会话中的工作流来恢复失败的任务。 可以在 Workflow Monitor 中修复错误和恢复工作流，也可以使用 <i>pmcmd</i> 恢复工作流。
已终止	集成服务意外停止或断开与主服务进程的网络连接。可以在 Workflow Monitor 中恢复工作流，也可以在集成服务重新启动后使用 <i>pmcmd</i> 恢复工作流。

任务恢复策略

工作流中的每个任务都具有恢复策略。集成服务恢复工作流时，会根据恢复策略恢复任务：

- **重新启动任务。**集成服务恢复工作流时，会重新启动配置了重新启动策略的每个可恢复任务。可以为会话和命令任务配置重新启动恢复策略。默认情况下，所有其他任务都具有重新启动恢复策略。
- **使任务失败并继续工作流。**集成服务恢复工作流时，不恢复任务。任务状态变为已失败，并且集成服务将继续运行工作流。
如果希望完成工作流但不希望恢复任务，请配置失败恢复策略。可以为会话和命令任务配置使任务失败并继续工作流恢复策略。
- **从上一检查点恢复。**集成服务可从上一检查点恢复已停止、已中止或已终止的会话。可以为会话任务配置恢复策略。

下表介绍了每种任务类型的恢复策略：

任务类型	恢复策略	注释
分配	重新启动任务。	-
命令	重新启动任务。 使任务失败并继续工作流。	默认值为使任务失败并继续工作流。
控制	重新启动任务。	-
判定	重新启动任务。	-
电子邮件	重新启动任务。	集成服务可能会发送重复的电子邮件。
事件触发	重新启动任务。	-
事件等待	重新启动任务。	-

任务类型	恢复策略	注释
会话	从上一检查点恢复。 重新启动任务。 使任务失败并继续工作流。	默认值为使任务失败并继续工作流。
计时器	重新启动任务。	如果使用任务或工作流的开始时间的相对时间，请将计时器设置为小于已过去时间的原始值。
工作集	不适用	集成服务不会恢复工作集。通过在 Workflow Monitor 中展开工作集并选择“恢复任务”，可以恢复工作集中的会话。

命令任务策略

配置命令任务时，可以选择要重新启动或失败的恢复策略：

- **使任务失败并继续工作流。**如果要在命令任务出错时挂起工作流，必须使用失败策略配置任务。如果命令任务具有多个命令，且您配置了失败策略，则需要将任务配置为在任何命令失败时失败。
- **重新启动任务。**集成服务恢复工作流时，会重新启动使用重新启动策略配置的命令任务。

在命令任务的“属性”页面上配置恢复策略。

会话任务策略

配置会话进行恢复时，可以在恢复工作流时恢复会话，或者可以在未运行其余工作流的情况下恢复会话。

配置会话时，可以选择的恢复策略包括失败、重新启动或恢复：

- **从上一检查点恢复。**集成服务会保存会话运行状态并维护目标恢复表。如果会话中止、停止或终止，则集成服务会使用保存的恢复信息从中断点恢复会话。
如果使用映射变量，则无法使用恢复策略配置会话。
- **重新启动任务。**集成服务将在其恢复工作流时再次运行会话。通过重新启动任务进行恢复时，可能需要移除目标中部分加载的数据或设计映射以跳过重复行。
- **使任务失败并继续工作流。**集成服务恢复工作流时，不会恢复会话。会话将处于失败状态，且集成服务会继续运行工作流。

在会话任务的“属性”页面中配置恢复策略。

自动恢复终止的任务

如果具有高可用性选项，则可以配置自动恢复终止的任务。启用自动任务恢复后，如果工作流仍在运行，集成服务会恢复终止的会话和命令任务而无需用户干预。可以配置集成服务尝试恢复任务的次数。在工作流属性中启用自动任务恢复。

继续进行会话

将会话恢复配置为从上一检查点继续时，集成服务将在 \$PMStorageDir 中创建检查点以确定开始处理会话恢复的位置。集成服务继续进行会话时将还原会话运行状态，包括每个源、目标和转换的状态。集成服务决定需要处理的源数据量。

集成服务继续进行会话时，恢复会话必须生成与原始会话相同的数据。如果您将恢复配置为从上一检查点继续，但会话无法生成可重复数据，会话将无效。

集成服务可以恢复包括 FTP 源在内的平面文件源。集成服务可以截断或附加到平面文件和 FTP 目标。

从上一检查点恢复会话时，集成服务将还原会话运行状态以确定能够执行的恢复类型：

- **增量。**集成服务开始在中断点处理数据。集成服务不读取或转换其在中断前处理的行。默认情况下，集成服务尝试执行增量恢复。
- **完整。**如果无法执行增量恢复，集成服务将重新读取所有源行并执行所有转换逻辑。集成服务开始在上一提交点写入目标。如果任意会话组件需要执行完整恢复，集成服务将对该会话执行完整恢复。

下表介绍了集成服务何时执行增量恢复或完整恢复，具体取决于会话配置：

组件	增量恢复	完整恢复
提交类型	会话使用基于源的提交。映射中不包含生成提交的任何转换。	会话使用基于目标的提交或用户定义的提交。
转换范围	转换传播事务，并且转换范围必须为“事务”或“行”。	至少一个转换配置了“全部”转换范围。
文件源	文件源支持增量读取。	不适用
FTP 源	FTP 服务器必须支持搜索操作才能允许执行增量读取。	FTP 服务器不支持搜索操作。
关系源	当输出是确定性的且可重复时，关系源支持增量读取。如果输出不是确定性的且不可重复，则集成服务通过将 SQL 结果暂存到存储文件来支持增量关系源读取。	不适用
VSAM 源	不适用	集成服务执行完整恢复。
XML 源	不适用	集成服务执行完整恢复。
XML 生成器转换	必须为 XML 生成器转换配置“事务”转换范围。	不适用
XML 目标	必须将 XML 目标配置为在提交时生成新 XML 文档。	不适用

使用可重复数据

将恢复配置为从上一检查点继续时，恢复会话必须能够以与原始会话相同的顺序产生相同的数据。

验证会话时，Workflow Manager 会验证是否将转换配置为产生可重复的确定性数据。如果将恢复配置为从上一检查点继续，而未将转换配置为产生可重复的确定性数据，则会话无效。

当所有目标从以下映射对象接收可重复数据时，会话数据为可重复数据：

- **源。**来自源的输出数据在原始运行与恢复运行之间是可重复的。
- **转换。**从各个转换到目标的输出数据是可重复的。

源可重复性

如果每个源生成相同数据集并且输出的顺序可在两次运行之间重复，则可以从上一检查点继续会话。基于会话中的源类型，源数据为可重复数据。

关系源

在工作流运行之间，关系源可能会生成不同的数据或按相同顺序生成数据。将恢复策略配置为从上一个检查点恢复时，集成服务会将 SQL 结果存储在缓存文件中以保证恢复的输出顺序。

如果您知道工作流运行之间的 SQL 结果将是相同的，则可以将源限定符配置为指示数据可重复且具有确定性。当关系源输出具有确定性且输出始终可重复时，集成服务不会将 SQL 结果存储在缓存文件中。当关系输出不可重复时，如果映射中的转换始终生成有序数据，则集成服务可跳过缓存文件创建过程。

SDK 源

如果 SDK 源可以生成可重复数据，则可以允许输出具有确定性且输出在 SDK 源限定符转换中具有可重复性。

平面文件源

平面文件在会话和恢复运行期间不会发生更改。如果在恢复会话之前更改源文件，则恢复会话可能会产生意外的结果。

转换可重复性

您可以将会话配置为在会话中的转换在会话运行和恢复运行之间生成相同数据时从上一检查点恢复。所有转换都具有可确定转换是否可以生成可重复数据的属性。如果输出是确定性且可重复的，则转换可以在会话运行和恢复运行之间生成相同的数据。

警告: 如果将转换配置为可重复和确定性的，则您有责任确保数据是可重复的。如果尝试使用不会生成可重复和确定性数据的转换恢复会话，则恢复过程可能会导致数据损坏。

输出是确定性的

转换始终从相同的输入数据中创建相同的输出数据时，转换会生成确定性的输出。

输出具有可重复性

当转换在运行会话之间按相同的顺序生成行时，会生成可重复数据。转换会基于转换类型、转换配置或映射配置生成可重复数据。

在以下情况下，转换会生成可重复数据：

- **始终。**即使输入数据的顺序在会话运行之间不一致，输出数据的顺序也会在会话运行之间保持一致。
- **基于输入顺序。**所有输入组中的输入数据的顺序在会话运行之间保持一致时，转换会在会话运行之间生成可重复数据。如果所有输入组中的输入数据均未排序，输出数据也不会排序。

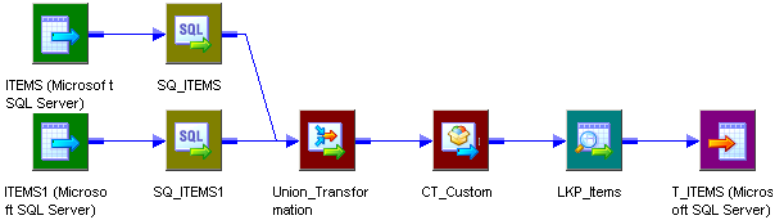
由于转换生成的可重复数据基于输入顺序，因此在会话验证期间，Workflow Manager 会验证映射以确定该转换是否可以生成可重复数据。例如，表达式转换只有在收到可重复数据时才会生成可重复数据。
- **从不。**输出数据的顺序在会话运行之间不一致。

配置映射以进行恢复

可以将映射配置为在会话中启用转换以在会话运行和恢复运行之间生成相同数据。映射包含从不生成可重复数据的转换时，可以在此转换之后紧接着添加一个始终生成可重复数据的转换。

例如，将从不生成可重复数据的转换直接连接到根据输入顺序生成可重复数据的转换。无法配置恢复以从上一个检查点恢复，除非数据是可重复的。要为会话启用恢复，可以在从不生成可重复数据的转换之后添加一个始终生成可重复数据的转换。

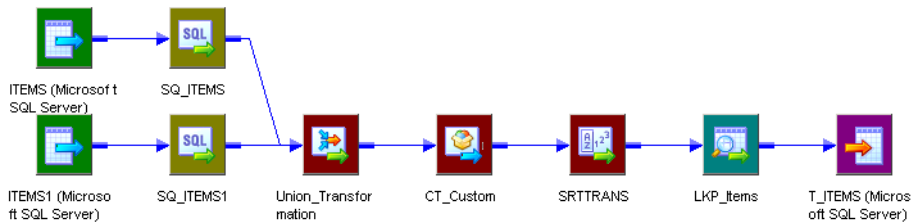
下图显示了无法通过从上一个检查点恢复而进行恢复的映射：



该映射包含两个可生成可重复数据的源限定符转换。还包含一个联合转换和一个自定义转换，它们从不生成可重复数据。查找转换会在接收可重复数据时生成可重复数据。因此，目标不会接收可重复数据，且您无法将会话配置为“恢复”恢复策略。

可以修改映射以启用“恢复”恢复策略。在从不输出可重复数据的转换之后紧接着添加一个配置为生成相异输出行的排序器转换。在自定义转换之后添加排序器转换。

下图显示了具有连接到自定义转换的排序器转换的映射：



查找转换会生成可重复数据，因为它从排序器转换接收可重复数据。

下表介绍了转换何时生成可重复数据：

转换	可重复数据
汇总器	始终。
应用程序源限定符	基于输入顺序。
自定义	基于输入顺序。根据转换过程行为配置属性。
数据屏蔽	基于输入顺序。根据转换过程行为配置属性。要生成可重复数据，请为每个端口配置可重复屏蔽或键屏蔽。
表达式	基于输入顺序。
外部过程	基于输入顺序。根据转换过程行为配置属性。
筛选器	基于输入顺序。
HTTP	基于输入顺序。根据转换过程行为配置属性。
联接器	基于输入顺序。
Java	基于输入顺序。根据转换过程行为配置属性。

转换	可重复数据
查找（动态）	始终。查找源必须与会话中的目标相同。
查找（静态）	基于输入顺序。
MQ 源限定符	始终。
规范器（管道）	基于输入顺序。
规范器 (VSAM)	始终。规范器以唯一主键形式生成源数据。会话在进行恢复时生成的键值可能与在成功完成时生成的键值不同。
等级	始终。
路由器	基于输入顺序。
序列生成器	始终。集成服务会将当前值存储到存储库中。
排序器（配置为生成相异输出行）	始终。
排序器（未配置为生成相异输出行）	基于输入顺序。
源限定符（平面文件）	始终。
源限定符（关系）	基于输入顺序。根据源数据配置转换。如果数据不可重复，则集成服务会暂存数据。
SQL 转换	基于输入顺序。根据源数据配置转换。
存储过程	基于输入顺序。根据转换过程行为配置属性。
事务控制	基于输入顺序。
联合	从不。
非结构化数据	基于输入顺序。根据转换过程行为配置属性。
更新策略	基于输入顺序。
XML 生成器	始终。
XML 解析器	基于输入顺序。根据源数据配置转换。
XML 源限定符	始终。

可以为以下转换配置“输出是可重复的”和“输出是确定性的”属性，也可以在这些转换之后紧接着添加一个可生成可重复数据的转换：

- 应用程序源限定符
- 自定义
- 外部过程
- 源限定符（关系）

- 存储过程

恢复工作流和任务的步骤

如果为工作流配置了恢复功能，则可以恢复工作流。配置会话恢复策略后，可以恢复会话。配置会话恢复策略后，无需启用工作流恢复即可恢复会话。

要恢复工作流或任务，可以使用下列方法之一：

- **恢复工作流。** 从中断点继续处理工作流。
- **恢复会话。** 恢复会话而不是其余工作流。
- **从会话中恢复工作流。** 恢复会话并继续处理工作流。

如果集成服务使用操作系统配置文件，请使用集成服务用来运行会话或工作流的同一操作系统配置文件来恢复会话或工作流。

如果要在不进行恢复的情况下重新启动工作流或任务，可以以冷启动模式重新启动工作流或任务。实时会话的恢复行为因实时源不同而异。

恢复工作流

恢复工作流时，集成服务会还原运行的工作流状态并继续从故障点进行处理。集成服务会使用任务恢复策略恢复出现故障的任务。

通过将工作流配置为在任务失败时挂起或在工作流属性中启用恢复的方式来配置工作流进行恢复。

可以使用 Workflow Manager、Workflow Monitor 或 *pmcmd* 恢复工作流。恢复工作流时，集成服务会将日志事件附加到现有会话日志中。

使用 Workflow Monitor 恢复工作流

要使用 Workflow Monitor 恢复工作流，请执行以下操作：

1. 在 Workflow Monitor 中选择工作流。
2. 右键单击工作流并选择“恢复”。

集成服务会恢复失败的任务并运行其余工作流。

还可以使用 *pmcmd recoverworkflow* 命令恢复工作流。

恢复会话

可以恢复已失败、已终止、已中止或已停止的会话而无需恢复工作流。如果工作流已完成，可以恢复会话而不运行其余工作流。必须配置从上一检查点重新启动或继续的恢复策略才能恢复会话。集成服务将根据任务恢复策略恢复会话。无需挂起工作流或启用工作流恢复即可恢复会话。当您恢复会话时，集成服务会另外创建会话日志。

要从 Workflow Monitor 中恢复会话，请执行以下操作：

1. 在 Workflow Monitor 中双击工作流以展开工作流并显示任务。
2. 右键单击会话，然后选择“恢复任务”。

集成服务将根据恢复策略恢复已失败的会话。

还可以结合使用 *pmcmd starttask* 与 *-recover* 选项来恢复会话。

从会话中恢复工作流

如果会话停止、中止或终止并且工作流尚未完成，若配置了会话恢复策略，则可以从会话中恢复工作流。恢复会话时，集成服务使用恢复策略来恢复会话并继续工作流。即使未挂起工作流或启用工作流恢复，也可以恢复会话。从会话中恢复工作流时，集成服务会创建另一个会话日志。

要在 Workflow Monitor 中从会话恢复工作流，请执行以下操作：

1. 在 Workflow Monitor 中双击工作流以展开工作流并显示会话。
2. 右键单击会话，然后通过恢复此任务选择“重新启动工作流”。

集成服务将根据恢复策略恢复已失败的会话。

可以结合使用 `pmcmd startworkflow` 与 `-recover` 选项从会话中恢复工作流。

注意：要恢复工作集中的会话，请展开该工作集，然后选择恢复任务。

会话恢复的规则和准则

恢复会话时，请遵循以下规则和准则：

- 集成服务在运行恢复会话时会创建一个新的会话日志。
- 会话报告上次成功运行的性能统计信息。
- 如果提供种子参数，则可以恢复包含使用随机数生成器 (RAND) 函数的转换的会话。
- 在会话恢复期间，PowerCenter 集成服务会将映射变量重置为起始值。

将恢复配置为从上一检查点继续

将恢复配置为从上一检查点继续时，请遵循以下规则和准则：

- 必须对每个转换使用传递分区。
- 不能将恢复配置为从在网格上运行的会话的上一检查点继续。
- 配置会话以实现完整下推优化时，集成服务将对数据库运行会话。因此，如果会话失败，集成服务将无法执行增量恢复。对包含 SQL 替代的会话执行恢复时，集成服务必须删除并重新创建视图。
- 修改中断的运行与恢复运行之间的工作流或会话时，可能会获得意外结果。集成服务不阻止恢复修改的工作流。如果工作流或任务自上次运行后被修改，恢复工作流或会话日志将显示一条消息。
- 从上一检查点继续进行会话时，会话前命令和前期 SQL 命令仅运行一次。如果前期或后期命令或 SQL 命令失败，集成服务将在恢复过程中重新运行该命令。设计命令以便您能够重新运行。
- 如果会话在批量模式下写入关系目标，则无法将其配置为继续进行。

不可恢复的工作流或任务

在某些情况下，集成服务无法恢复工作流或任务。在下列情况下，您无法恢复工作流或任务：

- **更改分区数量。**如果在会话失败后更改分区数量，则恢复会话将失败。
- **中断的任务具有失败恢复策略。**如果将命令或会话恢复配置为失败并继续工作流恢复，则该任务将不可恢复。
- **缺少恢复存储文件。**如果 `$PMStorageDir` 中缺少恢复存储文件或 `$PMStorageDir` 的定义在原始运行和恢复运行之间发生了更改，则集成服务会使恢复会话或工作流失败。

- **恢复表为空或目标数据库中缺少恢复表。**在下列情况下，集成服务会使恢复会话失败：

- 在集成服务创建该表后删除了该表。
- 在集成服务从该表中删除了恢复信息后为会话启用恢复立即失败。

如果在下列情况下执行恢复，您可能会获得不一致的数据：

- **源或目标在初始会话后发生更改。**如果在恢复会话之前删除或创建索引或者编辑源或目标表中的数据，则集成服务可能会返回缺少或重复的行。
- **源或目标代码页在初始会话失败后发生更改。**如果更改了源或目标代码页，集成服务可能会返回错误数据。如果代码页与原始代码页双向兼容，则可以执行恢复。

第 10 章

停止和中止

本章包括以下主题：

- [停止和中止概览, 160](#)
- [错误类型, 161](#)
- [会话失败的集成服务处理, 161](#)
- [停止或中止工作流, 162](#)
- [停止或中止的步骤, 163](#)

停止和中止概览

可随时停止或中止任务、工作流或工作集。

正如可以停止或中止任何任务一样，您可以停止或中止会话。还可以使用映射逻辑中的 ABORT() 函数中止会话。会话错误可能会导致集成服务过早停止会话。可以通过设置会话中的错误阈值、使用映射中的 ABORT 函数或请求集成服务停止会话来控制停止点。集成服务遇到致命错误时，则无法控制停止点，如丢失目标数据库连接。

如果会话由于失误而导致失败，则可以通过恢复工作流来恢复会话。

停止工作流时，集成服务尝试停止工作流中当前正在运行的所有任务。如果工作流包含工作集，集成服务还尝试停止工作集中当前正在运行的所有任务。如果无法停止工作流，则需要中止该工作流。

集成服务可以完全停止以下任务：

- 会话
- 命令
- 计时器
- 事件等待
- 工作集

停止包含多个命令的命令任务时，集成服务会执行完当前命令但不会运行其他命令。集成服务无法停止一些任务，如电子邮件任务。例如，如果在发出停止命令时集成服务已启动发送电子邮件，则集成服务会先完成发送电子邮件，然后再停止运行工作流。

如果存储库服务进程关闭，则集成服务会中止工作流。

错误类型

会话错误可以是致命错误或非致命错误。非致命错误是指不会在首次发生时就强制停止会话的错误。集成服务无法访问源、目标或存储库时，会出现致命错误。

阈值错误

可以根据指定的非致命错误数选择停止会话。非致命错误是指不会在首次发生时就强制停止会话的错误。在会话属性中使用“出错时停止”选项建立错误阈值。启用此选项时，集成服务会对读取器、写入器和转换线程中出现的非致命错误进行计数。

读取源、转换数据和写入目标时，集成服务会保留相关错误计数。在会话属性中设置“出错时停止”选项时，集成服务会对以下非致命错误进行计数：

- **读取器错误。**集成服务在读取源数据库或源文件期间遇到的错误。在 Unicode 模式下运行会话时，读取器阈值错误可以包括对齐错误。
- **写入器错误。**集成服务在写入到目标数据库或目标文件期间遇到的错误。写入器阈值错误可以包括违反键约束，从而导致向非空字段中加载空值，且数据库会触发响应。
- **转换错误。**集成服务在转换数据期间遇到的错误。转换阈值错误可以包括转换错误以及设置为错误的任何条件（例如空输入）。

在管道中创建多个分区时，集成服务会为每个分区保留单独的错误阈值。集成服务达到任何分区的错误阈值时，便会停止会话。写入器可能会继续写入一个或多个分区中的数据，但不会影响执行成功恢复的能力。

注意：如果在非换行 VSAM 文件中出现对齐错误，则集成服务会将该错误阈值设置为 1 并停止该会话。

致命错误

集成服务无法访问源、目标或存储库时，会出现致命错误。致命错误可能包括丢失连接或目标数据库错误，如缺少加载数据的数据库空间。如果会话使用规范器或序列生成器转换，则集成服务无法更新存储库中的序列值并出现致命错误。

如果会话未使用规范器或序列生成器转换，且集成服务失去与存储库的连接，则集成服务不会停止该会话。会话完成，但集成服务不会将会话统计信息记录到存储库中。

您可以从 Workflow Manager 或通过 *pmcmd* 停止会话。

可以从 Workflow Manager 中止会话。集成服务遇到指定的转换错误时，也可以使用映射逻辑中的 ABORT 函数中止会话。

会话失败的集成服务处理

集成服务会按不同的方式处理会话错误，具体取决于导致会话失败的错误或事件。

下表介绍了会话失败时的集成服务行为：

会话错误的原因	集成服务行为
<ul style="list-style-type: none">- 由于读取器错误而出现的错误阈值- 使用 Workflow Manager 或 <i>pmcmd</i> 停止命令	集成服务会执行以下任务： <ul style="list-style-type: none">- 停止读取。- 继续处理数据。- 继续向目标中写入并提交数据。 如果集成服务无法完成处理并提交数据，则需要发出中止命令以停止会话。
使用 Workflow Manager 中止命令	集成服务会执行以下任务： <ul style="list-style-type: none">- 停止读取。- 继续处理数据。- 继续向目标中写入并提交数据。 如果集成服务在 60 秒内无法完成处理并提交数据，则会终止 DTM 进程并终止该会话。
<ul style="list-style-type: none">- 数据库的致命错误- 由于写入器错误而出现的错误阈值	集成服务会执行以下任务： <ul style="list-style-type: none">- 停止读取和写入。- 回滚未提交到目标数据库的所有数据。 如果会话由于致命错误而停止，则提交或回滚操作可能会失败。
<ul style="list-style-type: none">- 由于转换错误而出现的错误阈值- <code>ABORT()</code>- 事务控制表达式的计算无效	集成服务会执行以下任务： <ul style="list-style-type: none">- 停止读取。- 将行标记为中止行并继续处理数据。- 继续写入到目标数据库，直到其遇到中止行。- 基于提交间隔来提交问题。- 回滚未提交到目标数据库的所有数据。

停止或中止工作流

可以使用工作流中的控制任务来指定希望集成服务停止或中止工作流的时间和方式。启动工作流后，可以通过 Workflow Monitor 或 *pmcmd* 停止或中止工作流。在执行工作流期间可以随时发出停止或中止命令。

可以通过执行以下操作之一来停止或中止工作流：

- 使用工作流中的控制任务。
- 在 Workflow Monitor 中发出停止或中止命令。
- 使用 *pmcmd* 发出停止或中止命令。

停止或中止任务

可以通过 Workflow Monitor 停止或中止工作流内的任务。停止或中止任务时，集成服务会停止处理该任务。集成服务不会在已停止或已中止任务的路径中处理其他任务。集成服务会继续处理工作流中的并发任务。如果集成服务无法停止该任务，您可以中止该任务。

中止任务时，集成服务会终止该任务的进程。中止任务时，集成服务会继续处理工作流中的并发任务。

还可以停止或中止工作集。与停止和中止任务类似，集成服务会停止和中止工作集。集成服务会在执行工作流中的并发任务期间停止工作集。也可以停止或中止工作集内的任务。

停止或中止会话任务

如果在您发出停止命令的同时集成服务正在执行会话任务，则集成服务会停止读取数据。但会继续处理和写入数据并将数据提交到目标。如果集成服务无法完成处理并提交数据，则可以发出中止命令。

集成服务处理会话任务的中止命令的方式与停止命令相似，但处理中止命令会有 60 秒的超时期间。如果集成服务无法在超时期间内完成处理并提交数据，则会终止 DTM 进程并终止会话。

停止或中止的步骤

您可以随时在 Workflow Monitor 中停止或中止任务、工作流或工作集。当您停止工作流中的某项任务时，集成服务停止处理该任务及其路径中的所有其他任务。集成服务会继续运行并发任务。如果集成服务无法停止处理任务，您需要中止任务。集成服务中止某个任务时，会结束 DTM 进程并终止该任务。

实时会话的行为取决于实时源。

要在 Workflow Monitor 中停止或中止工作流、任务或工作集，请完成以下步骤：

1. 在导航器中，选择要停止或中止的任务、工作流或工作集。
2. 单击“任务” > “停止”或单击“任务” > “中止”。

Workflow Monitor 将在“输出”窗口中显示停止或中止命令的状态。

第 11 章

并发工作流

本章包括以下主题：

- [并发工作流概览, 164](#)
- [配置唯一的工作流实例, 164](#)
- [配置同一名称的并发工作流, 165](#)
- [使用参数和变量, 166](#)
- [配置并发工作流的步骤, 167](#)
- [启动和停止并发工作流, 168](#)
- [监视并发工作流, 169](#)
- [查看会话和工作流日志, 169](#)
- [并发工作流的规则和准则, 170](#)

并发工作流概览

并发工作流是指可以作为多个实例并发运行的工作流。工作流实例是工作流的表示形式。配置并发工作流时，可以允许集成服务多次并发运行工作流的一个实例，或者定义并发运行的工作流的唯一实例。

使用以下工作流选项之一配置并发工作流：

- **允许具有相同实例名称的并发工作流。** 可以将一个工作流实例配置为多次并发运行。每个实例均具有相同的源、目标和变量参数。集成服务可以通过运行 ID 标识每个实例。运行 ID 是指用于标识已运行的工作流实例的编号。
- **将唯一工作流实例配置为并发运行。** 定义每个工作流实例名称，并为该实例配置工作流参数文件。可以在参数文件中定义不同的源、目标和变量。

运行并发工作流时，Workflow Monitor 会按工作流名称和实例名称显示各个工作流。如果工作流没有唯一的实例名称，则 Workflow Monitor 会显示每个并发工作流运行的同一工作流名称。

集成服务会将实例名称或运行 ID 和时间戳附加到工作流和会话日志名称中，以便为并发工作流创建唯一的日志文件。

配置唯一的工作流实例

可以配置多个工作流实例并同时运行每个实例。配置工作流实例时，会提供该实例的唯一的名称并为该实例配置工作流参数文件。

对工作流实例进行配置以运行具有不同源和目标的工作流。例如，组织可以接收三个地区的销售数据。创建一个可以读取销售数据并将其写入到数据库的工作流。配置三个有关该工作流的实例。每个实例具有不同的工作流参数文件，可用于定义要处理的销售文件。可以并发运行该工作流的所有实例。

启动工作流时，可以选择要运行的实例。将并发工作流配置为使用唯一的实例运行时，可以并发运行实例。要多次并发运行一个实例，请将工作流配置为使用同一实例名称运行。

按实例名称恢复工作流实例

可以从 Workflow Monitor 或 *pmcmd* 恢复工作流实例。为工作流启动恢复后，集成服务会将工作流运行 ID 附加到恢复存储文件名中。

恢复并发工作流时，请标识您要恢复的实例。在 Workflow Monitor 中右键单击要恢复的实例。使用 *pmcmd* 恢复时，请输入实例名称参数。

运行同一实例名称的并发实例的规则和准则

运行具有同一实例名称的并发实例时，请遵守以下规则和准则：

- 如果每次并发工作流运行时的变量均相同，则集成服务会覆盖并发工作流运行之间的变量。
- 可以在 *pmcmd* 中通过运行 ID 停止或中止工作流。
- 可以在 *pmcmd* 中通过运行 ID 停止或中止工作流任务。
- Workflow Monitor 不会显示每个实例的运行 ID。运行 ID 显示在工作流日志、会话日志以及 Workflow Monitor 的“运行属性”面板中。
- 将并发工作流配置为使用同一实例名称运行时，日志文件名称会始终包含时间戳。

配置同一名称的并发工作流

可以允许工作流在未定义唯一的实例名称的情况下并发运行。可以运行多个具有同一工作流名称的实例。PowerCenter 集成服务通过运行标识符编号或运行 ID 来区分各个工作流实例。每个工作流运行均有唯一的运行 ID。PowerCenter 集成服务会将运行 ID 附加到工作流和会话日志名称、恢复文件名称以及其他临时文件名称中，来为每个工作流创建单独的文件。

从实时源（消息队列或 Web 服务）读取工作流时，可以运行具有同一实例名称的并发工作流。例如，可以管理多个项目团队中的数据。创建一个可从用于确定源数据和目标的消息队列中读取数据的工作流。可以多次并发运行该实例，并将不同的连接参数从消息队列传递至工作流实例。

运行并发 Web 服务工作流

运行 Web 服务工作流时，集成服务可以运行工作流的多个实例以提高性能。将工作流配置为 Web 服务进行运行时，可以在 Hub 上配置要运行的工作流实例数目和启动新工作流实例的时间。

启用工作流作为 Web 服务时，Workflow Designer 会允许工作流使用同一工作流名称并发运行。Web 服务中心会根据您为 Web 服务配置的“每 Hub 最大运行计数”和“服务时间”属性来确定启动 Web 服务工作流的新实例的时间。

Web 服务中心启动 Web 服务工作流实例时，该实例的名称与其他工作流实例的名称相同。

注意：启动工作流作为 Web 服务时，默认情况下，Workflow Designer 允许工作流并发运行。

配置同一名称的工作流实例

允许工作流使用同一实例名称并发运行时，还可以为该工作流配置工作流实例和参数文件。可以多次并发启动每个实例。

例如，如果要定义一个工作流并创建两个实例，则可以启动该工作流并同时运行两个实例。可以再次启动该工作流以并发运行相同的实例。

此时，Workflow Monitor 任务视图会显示正在并发运行的四个实例：

```
wf_sales [Instance1]
wf_sales [Instance2]
wf_sales [Instance1]
wf_sales [Instance2]
```

恢复相同名称的工作流实例

当您为工作流启用了恢复时，PowerCenter 集成服务会将运行 ID 附加到工作流恢复存储文件。您可以通过 *pmcmd* 恢复同名工作流。不能通过 Workflow Monitor 进行恢复。恢复并发工作流时，必须输入运行 ID 参数。

恢复并发工作流时，必须标识要恢复的实例。在 Workflow Monitor 中右键单击要恢复的实例。使用 *pmcmd* 恢复时，请输入运行 ID 参数。

注意：如果工作流更新关系目标，则无法从上一检查点恢复会话。

运行同一实例名称的并发实例的规则和准则

运行具有同一实例名称的并发实例时，请遵守以下规则和准则：

- 如果每次并发工作流运行时的变量均相同，则集成服务会覆盖并发工作流运行之间的变量。
- 可以在 *pmcmd* 中通过运行 ID 停止或中止工作流。
- 可以在 *pmcmd* 中通过运行 ID 停止或中止工作流任务。
- Workflow Monitor 不会显示每个实例的运行 ID。运行 ID 显示在工作流日志、会话日志以及 Workflow Monitor 的“运行属性”面板中。
- 将并发工作流配置为使用同一实例名称运行时，日志文件名称会始终包含时间戳。

使用参数和变量

为防止出现冲突，请为每个工作流实例配置参数文件。

下表列出了可为并发工作流配置的参数：

参数类型	参数名
数据库连接	<i>\$DBConnectionName</i>
源文件	<i>\$InputFileName</i>
目标文件	<i>\$OutputFileName</i>

参数类型	参数名
拒绝文件	<code>\$BadFileName</code>
查找文件	<code>\$LookupFileName</code>

集成服务通过工作流运行实例名称保留工作流变量。

访问运行实例名称或运行 ID

启用工作流以使用唯一实例名称并发运行时，集成服务将按运行实例名称区分工作流运行实例。您可以为多个工作流配置相同的运行实例名称，因为每个工作流实例都是由工作流名称和运行实例名称的组合定义的。启用工作流以使用相同的实例名称并发运行时，集成服务将按运行 ID 区分工作流运行实例。

内置变量 `$PMWorkflowRunInstanceName` 和 `$PMWorkflowRunId` 会返回工作流运行实例名称和运行 ID 作为字符串值。这些变量为只读变量。可以在工作流或映射中访问这些变量以检索工作流实例的名称或运行 ID。可以将这些变量应用于表达式、文件监视事件或数据。还可以使用这些变量配置唯一的文件名称。

例如，创建一个预定义的事件等待任务以在显示指示器文件后将其删除。使用 `$PMWorkflowRunInstance` 名称定义文件名称。运行两个具有唯一的实例名称的并发工作流时，每个工作流事件等待任务均会等待并删除其他指示器文件。

注意：运行不允许并发运行的工作流时，`$PMWorkflowRunInstanceName` 不包含任何值。

配置并发工作流的步骤

创建或编辑工作流时，可以允许工作流并发执行。

要允许工作流并发执行，请执行以下操作：

1. 在 Workflow Manager 中，打开工作流。
2. 在工作流“常规”选项卡上，启用并发执行。
允许工作流使用同一实例名称并发运行。
3. 要配置不同的实例名称，请单击“配置并发执行”。
此时将显示“配置并发执行”对话框。
4. 请选择以下选项之一：
 - **仅允许使用唯一实例名称执行并发运行。**如果实例名称是唯一的，则集成服务可以运行并发工作流。
 - **允许使用同一实例名称执行并发运行。**集成服务可以运行具有同一名称的并发工作流。
5. 或者，单击“添加”按钮可添加工作流实例名称。
工作流实例名称不区分大小写。Workflow Designer 会对实例名称中的字符进行验证。不能在实例名称中使用以下特殊字符：
`$. + - = ~ ` ! % ^ & * () [] { } ' \ " ; : / ? , < > \ \ | \t \r \n`
6. 或者，输入实例的工作流参数文件的路径。要对每个工作流实例使用不同的源、目标或变量，请为每个实例配置参数文件。
7. 单击“确定”。

启动和停止并发工作流

可以在 Workflow Designer 或 Workflow Monitor 中启动并发工作流。还可以从 *pmcmd* 启动工作流。要运行唯一的工作流实例，请在启动工作流时选择要运行的实例。

从 Workflow Designer 启动工作流实例

可以选择从 Workflow Designer 中启动工作流时要运行的工作流实例。请按照以下步骤启动至少已定义一个实例的工作流。

要从 Workflow Designer 启动工作流实例，请执行以下操作：

1. 打开包含工作流的文件夹。
2. 从“导航器”中选择要启动的工作流。
3. 右键单击工作流，然后选择“启动高级工作流”。
4. 选择要启动的工作流运行实例。默认情况下，所有实例均处于选中状态。可以取消选中所有工作流实例并选择要启动的实例。
5. 单击“确定”启动工作流实例。

此时，Workflow Monitor 会显示每个并发工作流名称和实例名称。

启动一个并发工作流

并发工作流没有唯一的实例名称或您不需要运行配置的实例时，可以使用 Workflow Designer 启动一个工作流。如果使用“启动工作流”选项启动并发工作流，则集成服务运行的工作流中会包含您在工作流“属性”和“变量”选项卡上定义的属性和变量。集成服务不会运行任何已配置的工作流实例。

要启动一个并发工作流实例，请执行以下操作：

1. 打开包含工作流的文件夹。
2. 从“导航器”中选择要启动的工作流。
3. 在“导航器”中右键单击工作流并选择“启动工作流”。

集成服务运行的一个工作流实例中会包含工作流“属性”和“变量”选项卡中的属性。

从命令行启动并发工作流

可以从命令行一次启动一个工作流实例。*pmcmd* *startworkflow* 命令包含实例名称的参数。从命令行启动工作流并输入实例名称参数时，集成服务会运行该工作流的实例。要运行多个工作流实例，请多次运行 *pmcmd* *startworkflow* 命令。

如果未使用 *startworkflow* 输入实例名称参数，则集成服务运行的工作流中会包含您在工作流“属性”和“变量”选项卡上定义的属性和变量。集成服务不会运行任何已配置的工作流实例。

从命令行创建工作流实例

使用 *pmcmd* 启动工作流时，可以动态创建实例。输入实例名称和参数文件名称。如果未配置实例名称，则集成服务会生成一个实例。集成服务可以将实例的变量保留在存储库中，但实例不会显示在工作流的“并发执行配置”对话框中。

停止或中止并发工作流

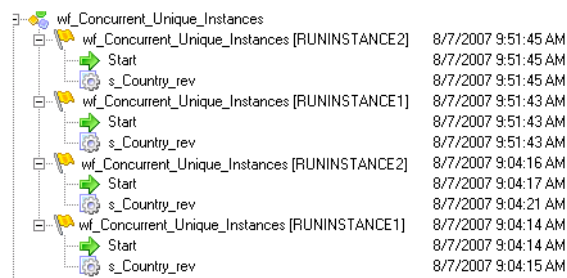
可以从 Workflow Monitor 或 *pmcmd* 停止或中止并发工作流。要从 Workflow Monitor 停止或中止工作流，请在导航器中右键单击工作流，然后选择“停止”或“中止”。Workflow Monitor 将在“输出”窗口中显示停止或中止命令的状态。

要从 *pmcmd* 停止或中止并发工作流，请通过在命令行输入实例名称或工作流运行 ID 参数来标识工作流实例。要停止或中止并发工作流中的任务，请输入包含要停止的任务的并发工作流的工作流实例名称或运行 ID。并发工作流没有唯一的实例名称时，可以在工作流日志或 Workflow Monitor 的工作流运行属性中查找工作流运行 ID。

监视并发工作流

运行并发工作流时，Workflow Monitor 会按工作流名称显示每个工作流运行情况。如果工作流具有唯一的实例名称，则 Workflow Monitor 会显示具有工作流名称的实例名称。

下图显示了 Workflow Monitor 任务视图中的并发工作流和实例名称：



在甘特图视图中查看并发工作流时，Workflow Monitor 会显示每个工作流名称或工作流实例名称的时间轴。如果工作流具有唯一的实例名称，则 Workflow Monitor 会显示每个工作流运行的实例名称，如 RunInstance1 和 RunInstance2。可以滚动“时间范围”查看有关特定工作流运行的信息。

查看会话和工作流日志

集成服务会根据您配置并发的方式对并发工作流会话和工作流日志文件进行命名：

- **唯一的实例名称。**集成服务会将实例名称附加到日志文件名称中。
- **同一名称的实例。**集成服务会将运行 ID 和时间戳附加到日志文件名称中。

集成服务会将运行 ID 和工作流类型写入到工作流日志。工作流类型会描述该工作流是否为并发工作流。

例如：

Workflow SALES_REV started with run id [108], run instance name [WF_CONCURRENT_SALES1], run type [Concurrent Run with Unique Instance Name].

每个会话日志还包括一个描述工作流运行 ID 和实例名称的条目：

Workflow: [SALES_REV] Run Instance Name: [WF_CONCURRENT_SALES1] Run Id: [108]

注意：如果您在错误严重级别为警告级别时无法查看所有工作流日志消息，可更改工作流日志的错误严重级别。在 PowerCenter 集成服务进程的高级属性中将日志级别从“警告”更改为“信息”。

唯一的工作流实例的日志文件

将工作流配置为使用唯一实例名称并发运行时，集成服务会为每个实例创建日志。每个日志文件名称均包括实例名称：

```
<workflow_name>.<workflow_instance_name>  
<session_name>.<workflow_instance_name>
```

例如，如果工作流日志文件名称为 wf_store_sales.log，且实例名称为 store1_workflow，则集成服务会为二进制工作流日志文件和文本工作流日志文件创建以下日志文件名称：

```
wf_store_sales.log.store1_workflow.bin  
wf_store_sales.log.store1_workflow
```

为避免覆盖日志文件，可以通过时间戳存档日志文件或保存日志文件。

同一名称的工作流实例的日志文件

将工作流配置为使用同一实例名称并发运行时，集成服务会为每个实例创建日志。默认情况下，每个日志文件名称均包括运行 ID 和时间戳：

```
<workflow_name>.<runID>.<timestamp>  
<session_name>.<run ID>.<timestamp>
```

例如，如果工作流日志文件名称为 wf_store_sales.log，且运行 ID 为 845，则如果工作流运行的时间为 2007 年 7 月 12 日的 11:20:45，集成服务会为二进制工作流日志文件和文本工作流日志文件创建以下日志文件名称：

```
wf_store_sales.log.845.20070712112045.bin  
wf_store_sales.log.845.20070712112045
```

将工作流配置为使用同一实例名称并发运行，且同时定义实例名称时，集成服务会将实例名称和时间戳附加到日志文件名称中。例如：

```
<workflow_name>.<instance_name>.<run ID>.20070712112045.bin  
<session_name>.<instance_name>.<run ID>.20070712112045.bin
```

集成服务会将实例名称和运行 ID 写入到工作流日志。例如：

```
Workflow wf_Stores started with run ID[86034], run instance name[Store1_workflow]
```

并发工作流的规则和准则

对于并发工作流，请遵守以下规则和准则：

- 无法引用参数文件中的工作流运行实例。要对每个实例使用单独的参数，必须配置不同的参数文件。
- 如果对多个并发工作流实例使用相同的缓存文件名称，则每个工作流实例均必须有效。但是，如果在写入到缓存时出现冲突，则会话失败。
- 可以使用 *pmcmd* 按运行 ID 或实例名称重新启动并发工作流。
- 如果配置多个工作流实例并调度该工作流，则集成服务会在调度时运行所有实例。无法在不同的调度中运行实例。
- 在工作集“常规”选项卡上将工作集配置为并发运行。
- 如果父工作流支持并发运行，则必须允许工作集并发运行。否则，工作流无效。
- 可以允许工作集并发运行，并将其放置在两个非并发工作流中。集成服务可以并发运行两个工作集。
- 允许并发运行的两个工作流可以运行相同的工作集。如果工作集不包含持久变量，则一个工作流可以运行同一工作集的两个实例。
- 会话不包含持久变量时，工作集中的会话可以与同一实例名称的另一个工作集中的会话并发运行。

以下转换对并发工作流存在一些限制：

- **汇总器转换。**在并发工作流中不能使用增量汇总。会话将失败。
- **查找转换。**对于并发工作流中的查找转换，请遵守以下规则和准则：
 - 可以将静态或动态查找缓存与并发工作流结合使用。
 - 如果缓存是非持久性缓存，则集成服务会将工作流运行 ID 作为前缀添加到缓存文件名称中。
 - 如果缓存是未命名持久缓存，则集成服务会将运行实例名称作为前缀添加到缓存文件名称中。
 - 如果缓存是动态未命名的持久缓存，且当前工作流配置为允许使用相同实例名称并发运行，则会话失败。
 - 如果查找缓存名称可进行参数化，则集成服务会使用参数值命名缓存文件。为每个运行实例传递不同的文件名称。
- **序列生成器转换。**为避免为并发工作流生成一组相同的序列编号，请在序列生成器转换中配置缓存值数。

第 12 章

网格处理

本章包括以下主题：

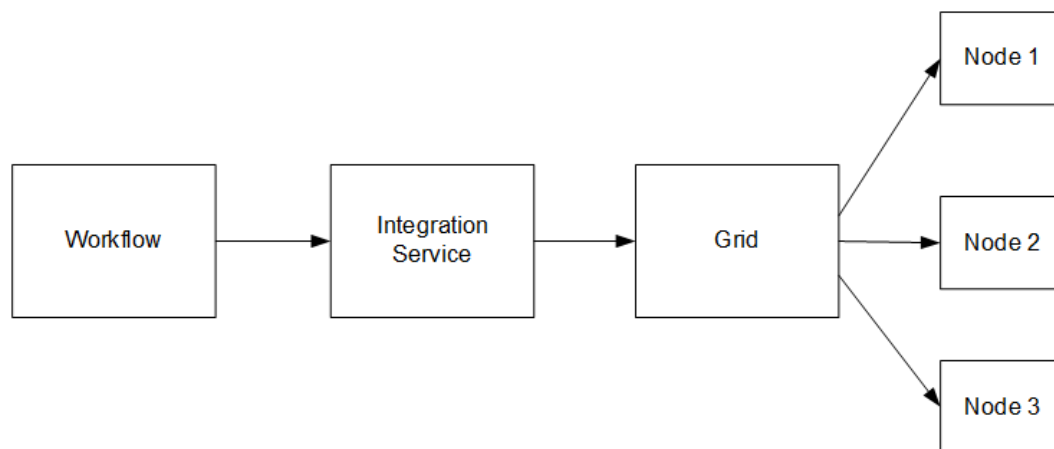
- [网格处理概览, 172](#)
- [在网格上运行工作流, 173](#)
- [在网格上运行会话, 173](#)
- [使用分区组, 174](#)
- [网格连接和恢复, 176](#)
- [将工作流或会话配置为在网格上运行, 176](#)

网格处理概览

PowerCenter 域包含多个节点，您可以将工作流和会话配置为在网格上运行。在网格上运行工作流时，集成服务会在网格的每个可用节点上运行服务进程以提高性能和扩展性。在网格上运行会话时，集成服务会将会话线程分发给网格中的节点上的多个 DTM 进程以提高性能和扩展性。

在 Administrator 工具中创建网格和配置集成服务。要在网格上运行工作流，可以将工作流配置为在与网格关联的集成服务上运行。要在网格上运行会话，可以将会话配置为在网格上运行。

下图显示了在网格上运行工作流时工作流和节点之间的关系：



集成服务根据您为工作流和会话配置的运行方式分发任务和会话线程：

- **在网格上运行工作流。** 集成服务在网格中的节点之间分发工作流。集成服务还会在网格中的节点之间分发工作流中的会话、命令和预定义的事件等待任务。

- **在网络上运行会话。** 集成服务在网格中的节点之间分发会话线程。

注意: 要在网络上运行工作流，必须拥有“服务器网格”选项。要在网络上运行会话，必须拥有“网络上的会话”选项。

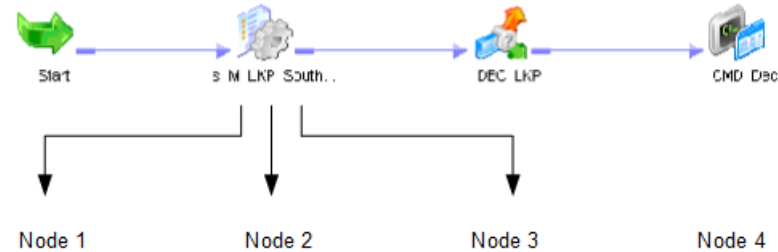
在网络上运行工作流

在网络上运行工作流时，主服务进程会运行工作流以及除了会话、命令和预定义的事件等待任务（这些服务会分发给其他节点）之外的所有任务。主服务进程是运行工作流、监视其他节点上运行的服务进程以及运行负载均衡器的集成服务进程。计划程序在主服务进程节点上运行，因此使用主服务进程节点的日期和时间启动计划的工作流。

负载均衡器是集成服务的一个组件，将会话、命令和预定义的事件等待任务分派到网格中的节点。负载均衡器根据节点可用性分发任务。如果集成服务配置为检查资源，则负载均衡器还会根据资源可用性分发任务。

例如，工作流包含会话任务、判定任务和命令任务。您可为会话任务指定资源要求。网格包含 4 个节点，节点 4 不可用。主服务进程会运行启动和判定任务。负载均衡器会根据资源可用性和节点可用性将会话和命令任务分发给网格上的节点。

下图显示了分发给网格上的节点的工作流：



1. 读取器线程在资源可用的节点 1 上运行。
2. 转换线程在资源可用的节点 2 上运行。
3. 写入器线程在资源可用的节点 3 上运行。
4. 节点 4 不可用，因此没有线程在其上运行。

在网络上运行会话

在网络上运行会话时，主服务进程会运行工作流以及除了会话、命令和预定义的事件等待任务之外的所有任务，与在网络上运行工作流一样。计划程序在主服务进程节点上运行，因此使用主服务进程节点的日期和时间启动计划的工作流。此外，负载均衡器会将会话线程分发给不同节点上运行的 DTM 进程。

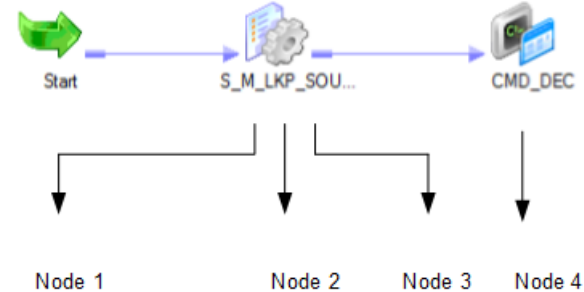
在网络上运行会话时，负载均衡器会根据以下因素分发会话线程：

- **节点可用性。** 负载均衡器会验证哪些节点当前正在运行、已启用以及可用于任务分派。
- **资源可用性。** 如果集成服务配置为检查资源，其会标识拥有在会话中映射对象所需资源的节点。
- **分区配置。** 负载均衡器会根据分区配置将多组会话线程分派到单独的节点。

当工作流包含需要花费较长时间运行的会话时，您可能希望将会话配置为在网络上运行。

例如，工作流包含使用一个分区的会话。要平衡负载，可以将会话配置为在网格上运行，并将集成服务配置为检查资源。负载平衡器会将读取器、写入器和转换线程分发给网格中的节点上运行的 DTM 进程。读取器线程需要资源，因此负载平衡器会将其分发给资源可用的节点上的 DTM 进程。

下图显示了分发给网格上的节点上运行的 DTM 进程的会话线程：



1. 读取器线程在资源可用的节点上运行。
2. 转换线程在可用节点上运行。
3. 写入器线程在可用节点上运行。
4. 命令任务在可用节点上运行。

使用分区组

在网格上运行会话时，Data Transformation Manager 进程 (DTM) 将形成会话线程组，称为分区组。分区组是在单个 DTM 进程中运行的一组读取器、写入器或转换线程。一个分区组可以包含一个或多个管道阶段。管道阶段是在两个分区点之间执行的管道的一部分。某些转换不可在网格范围内分区。当一个转换不可在网格范围内分区时，DTM 会为该转换的线程创建单个分区组，并在单个节点上运行这些线程。

不根据资源要求形成分区组

如果会话具有多个分区，DTM 会根据分区配置形成分区组。

例如，您配置了一个具有两个分区的会话。DTM 会为每个分区中的线程创建分区组，负载平衡器会将这些组分发给两个节点。分区组 1 在节点 1 上运行，分区组 2 在节点 2 上运行。

下图显示了一个包含两个分区的会话的两个分区组：



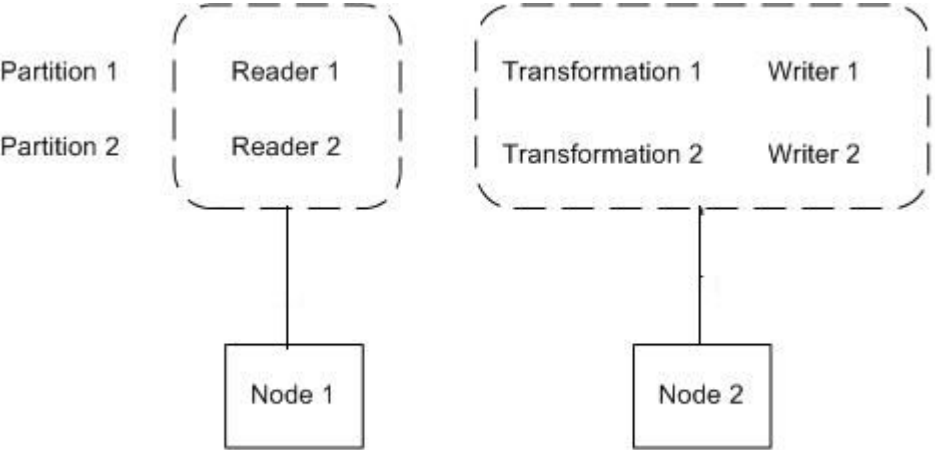
根据资源要求形成分区组

为映射对象指定资源要求时，DTM 进程会根据特定节点上可用的资源创建分区组。例如，如果特定节点上提供了会话的源文件，且您为源限定符转换指定了资源要求，则 DTM 进程会根据此要求形成分区组。

为了满足源限定符转换的资源要求，DTM 进程将从读取器线程创建分区组。负载均衡器将读取器线程分发给提供了资源的节点。

注意: 要使负载均衡器将线程分发给提供了所需资源的节点，您必须将集成服务配置为检查资源。

下图显示了根据分区配置和资源可用性分发的两个分区组：



创建分区组的规则和准则

集成服务使用以下规则和准则创建分区组：

- 集成服务会将分区组数限制在网格中的节点数内。
- 如果转换可在本地进行分区，则 DTM 进程会对转换线程形成一个分区组，并在一个 DTM 进程中运行该组。
下列转换可进行本地分区：
 - 配置为本地分区的自定义转换
 - 外部过程转换
 - 缓存查找转换
 - 未排序联接器转换
 - 配置为本地分区的 SDK 读取器或写入器转换

使用缓存

集成服务可为汇总器、等级、联接器、排序器和查找转换创建索引和数据缓存。当会话中包含多个分区时，可以将转换线程分发到网格中的多个节点。要为这些转换线程创建单个数据和索引缓存，请验证根目录和缓存目录是否对于网格中的所有节点均指向同一位置。

当集成服务在共享位置为查找转换创建缓存时，将为第一个分区组构建缓存，并且后续分区组也将使用此缓存。如果没有为查找转换缓存文件配置共享位置，则单独节点上的每个服务进程都将从数据库或源文件中提取数据以创建缓存。如果源数据经常发生更改，则在单独节点上创建的缓存可能会不一致。

相关主题：

- [“会话缓存” 页面上 248](#)

网络连接和恢复

在网络上运行工作流或会话时，服务进程和 DTM 进程在不同节点上运行。网络故障可能导致不同节点上运行的进程之间失去连接。在工作流或会话正在运行时，服务可能会意外关闭，或者您可以禁用集成服务或服务进程。在这些情况下，集成服务故障转移和恢复行为取决于已禁用、关闭或失去连接的服务进程。恢复行为还取决于以下因素：

- **高可用性选项。** 具有高可用性选项时，如果节点或服务关闭，则工作流会故障转移到其他节点。如果不具有高可用性选项，则可以手动在其他节点上重新启动工作流以恢复该工作流。
- **恢复策略。** 可以配置工作流以在出错时挂起。为工作流中的任务配置恢复策略。工作流挂起时，恢复行为取决于您为工作流中每个任务配置的恢复策略。
- **关闭模式。** 禁用集成服务或服务进程时，可以指定服务完成、中止或停止在服务上运行的进程。禁用集成服务或禁用服务进程时，行为有所不同。禁用主服务进程或执行工作的服务进程时，行为也有所不同。集成服务或服务进程也可能会意外关闭。在这种情况下，故障转移和恢复行为取决于哪些服务进程关闭以及已配置的恢复策略。
- **运行模式。** 如果工作流在网络上运行，则集成服务可以在其他节点上恢复工作流和任务。如果会话在网络上运行，则无法配置继续恢复策略。
- **操作模式。** 如果集成服务在安全模式下运行，则会为会话和工作流禁用恢复。

注意：如果集成服务在网络上运行，则无法将其配置为在安全模式下进行故障转移。

将工作流或会话配置为在网络上运行

在网络上运行会话或工作流之前，必须将网络分配给多个节点，并且必须将集成服务配置为在网络上运行。在 Administrator 工具中创建网络和分配集成服务。您可能需要以域管理员身份验证这些设置。

要在网络上运行工作流或会话，可以配置以下属性和设置：

- **工作流属性。** 在工作流属性的“常规”选项卡上，分配运行工作流的集成服务。验证是否已将集成服务配置为在网络上运行。
- **会话属性。** 要在网络上运行会话，可以在会话属性的“配置对象”选项卡上将会话启用为在网络上运行。
- **资源要求。** 可以在会话、命令和预定义的事件等待任务的“常规”选项卡上配置资源要求。

将工作流或会话配置为在网络上运行的规则和准则

将会话或工作流配置为在网络上运行时，请遵循以下规则和准则：

- 要在网络上运行会话，请验证操作系统和位模式是否与网络的每个节点相同。如果节点在不同操作系统或位模式下运行，则会话可能不会在网络上运行。
- 如果替代服务进程变量，请确保集成服务可以访问输入文件、缓存、日志、存储和临时目录以及源和目标文件目录。
- 要确保会话、命令或预定义的事件等待任务在特定节点上运行，请配置集成服务以检查资源并指定该任务的资源要求。

- 要确保映射对象的会话线程在特定节点上运行，请配置集成服务以检查资源并指定该对象的资源要求。
- 运行创建缓存文件的会话时，请将根目录和缓存目录配置为使用共享位置以确保缓存文件之间的一致性。
- 在联接器转换添加分区点，并且转换已配置为 1:n 分区时，确保集成服务在共享位置中构建缓存。必须共享详细管道的缓存。
- 在查找转换添加分区点，并且分区类型不是哈希自动键时，确保集成服务在共享位置中构建缓存。
- 运行使用动态分区的会话并且希望在网格中的所有节点之间分发会话线程时，请为会话配置动态分区以使用“基于网格中的节点数量”方法。
- 无法在网格上运行调试会话。
- 无法为网格上运行的会话配置继续恢复策略。
- 使用需要较长时间运行的会话时，请将该会话配置为在网格上运行。
- 当您拥有多个并发会话时，请将工作流配置为在网格上运行。
- 可以在网格上运行持久性配置文件会话，但无法在网格上运行临时配置文件会话。
- 使用序列生成器转换时，请增加缓存值数以减少主和执行工作的 DTM 进程以及存储库之间所需的通信。
- 要确保日志查看器在您在网格上运行工作流或会话时可以准确地对日志事件进行排序，请使用时间同步软件来确保网格的节点使用同步的日期/时间。
- 如果工作流使用 Windows 环境中的电子邮件任务，请在每个节点上配置相同的 Microsoft Outlook 配置文件以确保电子邮件任务可以运行。

第 13 章

负载均衡器

本章包括以下主题：

- [负载均衡器概览, 178](#)
- [向工作流分配服务级别, 178](#)
- [向任务分配资源, 179](#)

负载均衡器概览

负载均衡器向节点上运行的集成服务进程分派任务。运行工作流时，负载均衡器可在此工作流内分派会话、命令和预定义的事件等待任务。如果集成服务配置为检查资源，负载均衡器会将任务要求与资源可用性进行匹配，以标识运行任务的最佳节点。它可将任务分派给单个节点或在多个节点之间分派任务。

为标识可以运行任务的节点，负载均衡器会将任务所需的资源与每个节点上的可用资源进行匹配。负载均衡器按接收任务的顺序分派任务。当负载均衡器需要分派的会话和命令任务数超过集成服务当前能够运行的数量时，负载均衡器会将任务置于分派队列中。节点可用时，负载均衡器将按工作流服务级别确定的顺序分派队列中正在等待的任务。

请使用 Workflow Manager 分配资源和服务级别。您可以执行以下任务：

- **分配服务级别。** 请向工作流分配服务级别。服务级别在正在等待进行分派的工作流任务之间建立优先级。
- **分配资源。** 请向任务分配资源。会话、命令和预定义的事件等待任务需要 PowerCenter 资源才能成功运行。如果集成服务配置为检查资源，负载均衡器将向资源可用的节点分派这些任务。

向工作流分配服务级别

服务级别决定负载均衡器分派分派队列中的任务的顺序。如果有多个任务正在等待分派，负载均衡器将先分派高优先级任务，然后分派低优先级任务。请在 Administrator 工具中创建服务级别并配置分派优先级。

请在工作流属性的“常规”选项卡上向工作流分配服务级别。

向任务分配资源

PowerCenter 资源是指任务所需的数据库连接、文件、目录、节点名称和操作系统类型，任务需要这些资源才能成功运行。负载均衡器可能会使用资源来分派任务。如果集成服务未配置为在网格上运行或未配置为检查资源，负载均衡器将忽略资源要求。负载均衡器向节点上运行的主集成服务进程分派所有任务。

如果集成服务在网格上运行且配置为检查资源，负载均衡器将使用资源来分派任务。集成服务将工作流中的任务所需的资源与网格中的每个节点上可用的资源进行匹配，以确定能够运行任务的节点。负载均衡器将会话、命令和预定义的事件等待任务分发给具有可用资源的节点。例如，如果会话需要对预留文件使用文件资源，负载均衡器将向能够访问该文件的节点分派该会话。如果集成服务无法标识所需资源可用的节点，任务将失败。

在 Administrator 工具中，您定义了对每个节点可用的资源。资源为预定义的资源或用户定义的资源。预定义的资源包括对节点可用的连接、节点名称和操作系统类型。用户定义的资源包括文件/目录资源和自定义资源。

在任务属性中，您向需要 PowerCenter 资源的不可重用任务分配这些资源。不能向可重用任务分配资源。

下表列出了资源类型以及可以向其分配这些类型的存储库对象：

资源类型	预定义/ 用户定义	使用资源的存储库对象
自定义	用户定义	会话、命令和预定义的事件等待任务实例以及会话中的所有映射对象。
文件/目录	用户定义	会话、命令和预定义的事件等待任务实例以及会话中的以下映射对象： <ul style="list-style-type: none">- 源限定符- 汇总器转换- 自定义转换- 外部过程转换- 联接器转换- 查找转换- 排序器转换- 自定义转换- Java 转换- HTTP 转换- SQL 转换- 联合转换- 目标
节点名称	预定义	会话、命令和预定义的事件等待任务实例以及会话中的所有映射对象。
操作系统类型	预定义	会话、命令和预定义的事件等待任务实例以及会话中的所有映射对象。

如果您尝试分配不适用于存储库对象的资源类型，Workflow Manager 将显示以下错误消息：

The selected resource cannot be applied to this type of object. Please select a different resource.

Workflow Manager 分配连接资源。当您使用关系连接、FTP 连接或外部加载器连接时，Workflow Manager 将向会话实例中的源、目标和转换分配连接资源。在 Workflow Manager 中不能手动分配连接资源。

要向任务实例分配资源，请执行以下操作：

1. 在工作集或 Workflow Designer 中打开任务属性。
如果任务为事件等待任务，则仅当任务等待预定义的事件时才能分配资源。
2. 在“常规”选项卡上，单击“编辑”。

3. 在“编辑资源”对话框中，单击“添加”按钮添加资源。
4. 在“选择资源”对话框中，选择需要向其分配资源的对象。“资源列表”显示运行集成服务的节点可用的资源。
5. 选择要分配的资源并单击“选择”。
6. 在“编辑资源”对话框中，单击“确定”。

第 14 章

工作流变量

本章包括以下主题：

- [工作流变量概览, 181](#)
- [预定义工作流变量, 182](#)
- [用户定义的工作流变量, 186](#)
- [使用工作集变量, 189](#)
- [在工作集中分配变量值, 189](#)

工作流变量概览

可以在工作流中创建和使用变量以引用值和记录信息。例如，在判定任务中使用变量可确定上一个任务是否正确运行。如果正确运行，则可以运行下一个任务。如果未正确运行，则可以停止工作流。

使用以下类型的工作流变量：

- **预定义的工作流变量。** Workflow Manager 为工作流中的任务提供预定义的工作流变量。
- **用户定义的工作流变量。** 可在创建工作流时创建用户定义的工作流变量。

在配置以下类型的任务时使用工作流变量：

- **分配任务。** 使用分配任务将某个值分配给用户定义的工作流变量。例如，可以通过将变量设置为其当前值加 1 来增加用户定义的计数器变量。
- **判定任务。** 判定任务决定了集成服务运行工作流的方式。例如，只有在成功完成第一个会话后才能使用状态变量运行第二个会话。
- **链接。** 链接可连接各个工作流任务。在链接中使用工作流变量可在工作流中创建分支。例如，在判定任务之后，可以创建一个在判定条件计算结果为 true 时要打开的链接，以及另一个在判定条件计算结果为 false 时要打开的链接。
- **计时器任务。** 计时器任务可指定集成服务开始运行工作流中下一个任务的时间。使用用户定义的日期/时间变量可指定集成服务开始运行下一个任务的时间。

使用表达式编辑器可创建使用变量的表达式。构建表达式时，可以在“预定义”选项卡上选择预定义的变量。可以在“用户定义”选项卡上选择用户定义的变量。“函数”选项卡包含可与工作流变量结合使用的函数。使用“指向并单击”方法可输入使用变量的表达式。

使用以下关键字可为用户定义和预定义的工作流变量写入表达式：

- AND
- OR
- NOT

- TRUE
- FALSE
- NULL
- SYSDATE

预定义 workflow 变量

每个 workflow 都包含一组可用于评估 workflow 和任务条件的预定义变量。使用以下类型的预定义变量：

- **特定于任务的变量。**Workflow Manager 为 workflow 中的每个任务都提供了一组特定于任务的变量。使用链接条件中特定于任务的变量可控制集成服务在运行 workflow 时采用的路径。Workflow Manager 在表达式编辑器中的任务名称下列出了特定于任务的变量。
- **内置变量。**使用 workflow 中的内置变量可返回运行时或系统信息，如文件夹名称、集成服务名称、系统日期或 workflow 开始时间。Workflow Manager 在表达式编辑器中的内置节点下列出了内置变量。

提示：在集成服务中将日志文件的错误严重级别设置为“跟踪”时，workflow 日志将显示 workflow 变量的值。仅使用此日志记录级别以进行故障排除。

下表列出了 Workflow Manager 中可用的特定于任务的工作流变量：

特定于任务的变量	说明	任务类型	数据类型
Condition	判定条件表达式的计算结果。 如果任务失败，则 Workflow Manager 将条件设置保持为空。 示例语法： <code>\$Dec_TaskStatus.Condition = <TRUE FALSE NULL any integer></code>	判定	整数
EndTime	关联任务结束的日期和时间。精确到秒。 示例语法： <code>\$s_item_summary.EndTime > TO_DATE('11/10/2004 08:13:25')</code>	所有任务	日期/时间
ErrorCode	关联任务的最后一个错误代码。如果没有错误，则集成服务会在任务完成时将 ErrorCode 设置为 0。 示例语法： <code>\$s_item_summary.ErrorCode = 24013</code> 注意： 如果任务始终失败并显示此最终错误消息，则可以使用此变量。	所有任务	整数

特定于任务的变量	说明	任务类型	数据类型
ErrorMsg	<p>关联任务的最后一个错误消息。</p> <p>如果没有错误，则集成服务会在任务完成时将 ErrorMsg 设置为空字符串。</p> <p>示例语法：</p> <pre>\$s_item_summary.ErrorMsg = 'PETL_24013 Session run completed with failure'</pre> <p>Nstring 类型的变量最大长度为 600 个字符。</p> <p>注意：如果任务始终失败并显示此最终错误消息，则可以使用此变量。</p>	所有任务	Nstring
FirstErrorCode	<p>会话中第一个错误消息的错误代码。</p> <p>如果没有错误，则集成服务会在会话完成时将 FirstErrorCode 设置为 0。</p> <p>示例语法：</p> <pre>\$s_item_summary.FirstErrorCode = 7086</pre>	会话	整数
FirstErrorMsg	<p>会话中的第一个错误消息。</p> <p>如果没有错误，则集成服务会在任务完成时将 FirstErrorMsg 设置为空字符串。</p> <p>示例语法：</p> <pre>\$s_item_summary.FirstErrorMsg = 'TE_7086 Tscrubber: Debug info... Failed to evalWrapUp'</pre> <p>Nstring 类型的变量最大长度为 600 个字符。</p>	会话	Nstring
PrevTaskStatus	<p>集成服务运行的工作流中上一个任务的状态。状态包括：</p> <ul style="list-style-type: none"> - ABORTED - FAILED - STOPPED - SUCCEEDED <p>在编写表达式以评估上一个任务的状态时，请使用这些关键字。</p> <p>示例语法：</p> <pre>\$Dec_TaskStatus.PrevTaskStatus = FAILED</pre>	所有任务	整数
SrcFailedRows	<p>集成服务未能从源读取的总行数。</p> <p>示例语法：</p> <pre>\$s_dist_loc.SrcFailedRows = 0</pre>	会话	整数
SrcSuccessRows	<p>成功从源读取的总行数。</p> <p>示例语法：</p> <pre>\$s_dist_loc.SrcSuccessRows > 2500</pre>	会话	整数

特定于任务的变量	说明	任务类型	数据类型
StartTime	<p>关联任务启动的日期和时间。精确到秒。</p> <p>示例语法：</p> <pre>\$s_item_summary.StartTime > TO_DATE('11/10/2004 08:13:25')</pre> <p>注意: SESSSTARTTIME 会返回集成服务初始化会话后运行会话的节点上的当前日期和时间值。如果映射或 Mapplet 使用 SESSSTARTTIME，则会话的 StartTime 和 SESSSTARTTIME 值将不同。</p>	所有任务	日期/时间
状态	<p>工作流中上一个任务的状态。状态包括：</p> <ul style="list-style-type: none"> - ABORTED - DISABLED - FAILED - NOTSTARTED - STARTED - STOPPED - SUCCEEDED <p>在编写表达式以评估当前任务的状态时，请使用这些关键字。</p> <p>示例语法：</p> <pre>\$s_dist_loc.Status = SUCCEEDED</pre>	所有任务	整数
TgtFailedRows	<p>集成服务未能写入目标的总行数。</p> <p>示例语法：</p> <pre>\$s_dist_loc.TgtFailedRows = 0</pre>	会话	整数
TgtSuccessRows	<p>成功写入目标的总行数</p> <p>示例语法：</p> <pre>\$s_dist_loc.TgtSuccessRows > 0</pre>	会话	整数
TotalTransErrors	<p>转换错误总数。</p> <p>示例语法：</p> <pre>\$s_dist_loc.TotalTransErrors = 5</pre>	会话	整数

除状态之外的所有预定义工作流变量默认值均为空。集成服务遇到尚未在工作流中运行的任务中预定义的变量时，会使用该为空的默认值。因此，依赖于尚未运行的任务的表达式和链接条件有效。状态的默认值为 NOTSTARTED。

使用表达式中的预定义工作流变量

使用表达式中的工作流变量时，集成服务会计算该表达式并返回 True 或 False。如果条件的计算结果为 true，集成服务将运行下一个任务。集成服务会在工作流日志中写入类似于以下消息的条目：

```
INFO : LM_36506 : (1980|1040) Link [Session2 --> Session3]: condition is TRUE for the expression [Session2.PrevTaskStatus = SUCCEEDED].
```


表达式编辑器在“预定义”选项卡上显示预定义工作流变量。Workflow Manager 按任务对特定于任务的变量进行分组，并在内置节点下列出内置变量。要使用表达式中的变量，请双击该变量。表达式编辑器在“表达式”字段中以下列格式显示特定于任务的变量：

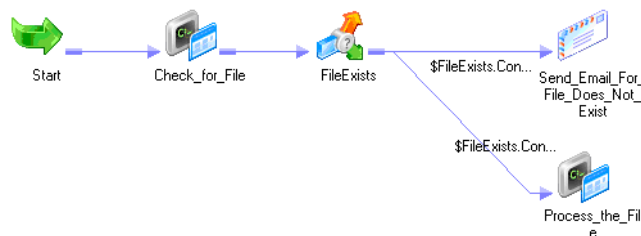
`$<TaskName>.<predefinedVariable>`

评估工作流中的条件

使用链接条件中的“条件”来评估判定条件表达式的结果。

下图显示了具有使用“条件”的链接条件的工作流：

图 4. 条件变量示例



FileExist 判定任务的判定条件表达式是 `$Check_for_file.Status = SUCCEEDED`。映射包含两个链接条件：
`$FileExists.Condition = False` 触发电子邮件任务，`$FileExists.Condition = True` 触发命令任务，
`Process_the_File`。

运行工作流时，集成服务会评估链接条件，然后根据 FileExists 判定任务的判定条件表达式返回值。集成服务会根据 Check_for_File 任务结果触发电子邮件任务或命令任务。

评估工作流中的任务状态

使用链接条件中的“状态”测试工作流中上一个任务的状态。

下图显示了具有使用“状态”的链接条件的工作流：

图 5. 状态变量示例



运行工作流时，集成服务会评估链接条件，`$Session2.Status = SUCCEEDED`，然后根据 Session2 的状态返回值。

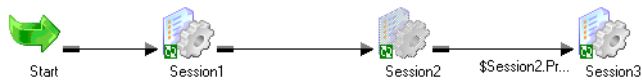
在工作流中评估上一个任务状态

在链接条件下使用 `PrevTaskStatus`，以便在集成服务运行的工作流中测试上一个任务的状态。

如果在工作流中禁用任务，请使用 `PrevTaskStatus`。状态和 `PrevTaskStatus` 将返回相同值，除非条件使用已禁用任务。

下图显示了包含使用 PrevTaskStatus 的连接条件的工作流：

图 6. PrevTaskStatus 变量示例



运行工作流时，由于已禁用会话，集成服务会跳过 Session2。集成服务评估连接条件 `$Session2.PrevTaskStatus = SUCCEEDED` 时，将根据 Session1 的状态返回值。

提示：如果未禁用 Session2，集成服务将根据 Session2 的状态返回值。启用和禁用 Session2 时，无需更改连接条件。

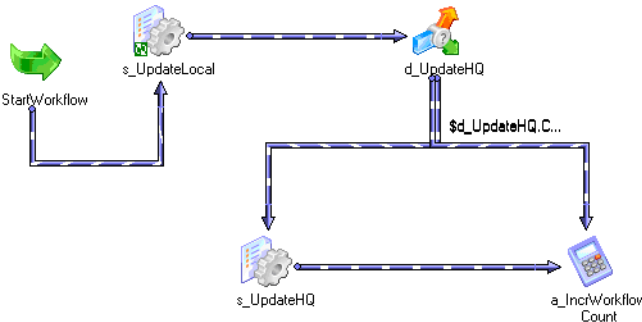
用户定义的工作流变量

可以在工作流中创建变量。在工作流中创建变量时，该变量仅在该工作流中有效。在该工作流中的任务中使用该变量。可以编辑和删除用户定义的工作流变量。

需要基于您指定的条件做出工作流决定时，请使用用户定义的变量。例如，创建工作流以在夜间将数据加载到订单数据库。还需要定期（本地订单数据库每更新十次时）将此数据的子集加载到总部。创建独立的会话以更新本地数据库以及总部的数据库。

下图显示了该工作流：

图 7. 使用工作流变量的工作流



使用用户定义的变量来决定何时运行在总部更新订单数据库的会话。

要配置用户定义的工作流变量，请完成以下步骤：

1. 创建持久工作流变量 `$$WorkflowCount` 以代表工作流已运行的次数。
2. 将启动任务和两个会话添加到工作流。
3. 将判定任务置于更新本地订单数据库的会话之后。
设置判定条件以检查是否工作流运行次数可由 10 整除。使用模数 (MOD) 函数来执行此操作。
4. 创建分配任务以将 `$$WorkflowCount` 变量的值增加 1。
5. 在判定条件计算结果为 true 时，将判定任务链接到在总部更新数据库的会话。在判定条件计算结果为 false 时，将判定任务链接到分配任务。

使用这些条件配置工作流变量时，更新本地数据库的会话便会在每次工作流运行时运行。工作流每运行十次便会在总部更新数据库的会话。

工作流变量起始值和当前值

从概念上讲，集成服务在工作流运行期间为工作流变量保存两个不同的值：

- 工作流变量的起始值
- 工作流变量的当前值

起始值为工作流开始时变量的值。起始值可以是在参数文件中为变量定义的值、存储库中保存的源自上一个工作流运行的值、用户为变量定义的起始值，或基于变量数据类型的默认值。

集成服务按以下顺序查找变量的起始值：

1. 参数文件中的值
2. 存储库中保存的值（如果变量为持久变量）
3. 用户指定的默认值
4. 数据类型默认值

例如，在工作流中创建一个工作流变量并输入默认值，但未在参数文件中为该变量定义值。集成服务首次运行工作流时，会将变量的起始值计算为用户定义的默认值。

如果声明该变量为持久变量，则集成服务会在工作流运行结束时将该变量的值保存到存储库中。工作流下次运行时，集成服务会将变量的起始值计算为存储库中保存的值。

如果变量为非持久变量，集成服务将不会保存该变量的值。工作流下次运行时，集成服务会将变量的起始值计算为用户定义的默认值。

如果要在运行工作流之前替代存储库中保存的值，需要在参数文件中为变量定义值。在参数文件中定义工作流变量时，集成服务将使用此值，而不是存储库中保存的值或为变量配置的初始值。

随着工作流的进行，当前值为变量的值。当工作流启动时，变量的当前值与起始值相同。如果创建了更新变量值的分配任务，则随着工作流的进行，变量的值可能会发生更改。

如果变量为持久变量，集成服务会在工作流运行成功结束时将变量的当前值保存到存储库中。如果工作流无法完成，集成服务将不会更新存储库中变量的值。

集成服务会在工作流日志中指出保存到存储库的每个工作流变量的值。

数据类型默认值

如果集成服务无法以任何其他方式确定变量的起始值，则会根据其数据类型对该变量使用默认值。

下表列出了用户定义的工作流变量的数据类型默认值：

数据类型	Workflow Manager 默认值
日期/时间	1/1/1753 00:00:00.000000000 A.D.
双精度	0
整数	0
Nstring	空白字符串

创建用户定义的工作流变量

您可以在工作流属性中为工作流创建工作流变量。

要创建工作流变量，请执行以下操作：

1. 在 Workflow Designer 中，创建新的工作流或编辑现有工作流。
2. 选择“变量”选项卡。
3. 单击“添加”。
4. 输入下表中的信息，然后单击“确定”：

字段	说明
名称	变量名称。正确格式为 <code>\$\$VariableName</code> 。工作流变量名称不区分大小写。不要对用户定义的工作流变量使用单个美元符号 (\$)。单个美元符号专为预定义的工作流变量预留。
数据类型	变量的数据类型。可以从以下数据类型中进行选择： <ul style="list-style-type: none">- 日期/时间- 双精度- 整数- Nstring
持久性	变量是否是持久性的。如果您希望变量值从某次工作流执行保留到下次工作流执行，请启用此选项。
默认值	<p>变量的默认值。如果您未在参数文件中设置变量值且存储库中未存储任何值，集成服务将在会话期间使用此变量值。</p> <p>日期/时间类型的变量可具有以下格式：</p> <ul style="list-style-type: none">- MM/DD/RR- MM/DD/YYYY- MM/DD/RR HH24:MI- MM/DD/YYYY HH24:MI- MM/DD/RR HH24:MI:SS- MM/DD/YYYY HH24:MI:SS- MM/DD/RR HH24:MI:SS.MS- MM/DD/YYYY HH24:MI:SS.MS- MM/DD/RR HH24:MI:SS.US- MM/DD/YYYY HH24:MI:SS.US- MM/DD/RR HH24:MI:SS.NS- MM/DD/YYYY HH24:MI:SS.NS <p>可以使用以下分隔符：短划线 (-)、斜杠 (/)、反斜杠 (\)、冒号 (:)、句点 (.) 和空格。集成服务会忽略多余空格。对于年份，不能使用一位数或三位数值；对于小时，不能使用“HH12”格式。</p> <p>Nstring 类型的变量最大长度为 600 个字符。</p>
为空	变量的默认值是否为空。如果默认值为空，请启用此选项。
说明	与变量相关的说明。

5. 要验证新工作流变量的默认值，请单击“验证”按钮。
6. 单击“应用”以保存新的工作流变量。
7. 单击“确定”。

使用工作集变量

工作集变量与工作流变量类似。工作集具有一组与任何任务相同的预定义变量。还可以创建用户定义的工作集变量。与用户定义的工作流变量类似，用户定义的工作集变量可以具有持久性或非持久性。

持久性工作集变量

用户定义的工作集变量可以具有持久性或非持久性。要创建持久性工作集变量，请在创建变量时选择持久性。如果创建持久性工作集变量，则下次集成服务运行父工作流中的工作集时，该工作集变量会保留其值。

例如，您有一个具有持久性变量的工作集。使用工作流中的两个工作集实例可以运行该工作集两次。将工作集的第一个实例和第二个实例分别命名为 Worklet1 和 Worklet2。

运行工作流时，持久性工作集变量会保留 Worklet1 中的值，并在 Worklet2 中变为初始值。集成服务运行 Worklet2 后，会将持久性变量的值保留在存储库中，并在下次运行工作流时使用该值。

运行同一个工作流时，仅会保留工作集变量。使用不同工作流中的工作集实例时，工作集变量不会保留其值。

替代初始值

对于每个工作集实例，可以通过向其分配工作流变量来替代工作集变量的初始值。

要替代工作集变量的初始值，请执行以下操作：

1. 在 Workflow Designer 工作区中双击工作集实例。
2. 在“变量”选项卡上，单击前期工作集变量分配中的“添加”按钮。
3. 在“用户定义的工作集变量”字段中单击“打开”按钮以选择工作集变量。
4. 单击“应用”。

此工作集实例中的工作集变量使用选定的工作流变量作为其初始值。

使用工作集变量的规则和准则

在使用工作集变量时，请遵循以下规则和准则：

- 无法使用工作集中的父工作流变量。
- 可以将工作流变量的值分配到工作集变量以进行初始化。
- 无法使用父工作流中用户定义的工作集变量。
- 可以与使用工作流中其他任务的预定义变量一样使用父工作流中预定义的工作集变量。

在工作集中分配变量值

可以在工作集运行之前或之后更新变量值。这允许您将信息从一个工作集传递到同一工作流或父工作集中的另一个工作集。例如，您拥有一个工作流，其包含两个需要增加相同计数器的工作集。您可以在第一个工作集中增加计数器，将更新后的计数器值传递给第二个工作集，然后在第二个工作集中增加计数器。

还可以将信息从工作集传递到不可重用的会话或从不可重用的会话传递到工作集，只要该工作集和该会话位于同一工作流或父工作集中。可以在可重用和不可重用的工作集中分配变量。

可以根据是在工作集运行之前还是之后分配变量来更新不同变量的值。可以在工作集运行前后更新以下类型的变量：

- **工作集前变量分配。**可以在工作集运行之前更新用户定义的工作集变量。可以通过工作流或父工作集中的其他任务为这些变量分配父工作流或工作集变量的值或映射变量的值。

可以使用工作集父项中的值更新工作集变量。因此，如果某个工作集位于工作流内的其他工作集中，则可以从父工作集变量（而非工作流变量）分配值。

- **工作集后变量分配。**可以在工作集完成之后更新父工作流或工作集变量。可以为这些变量分配用户定义的工作集变量的值。

编辑工作集时，可以在“变量”选项卡上分配变量。

在工作集之间传递变量值

您可以分配工作集中的变量值，以将值从一个工作集传递到同一工作流中的任意后续工作集或父工作集。例如，某个工作流中包含两个工作集 wklt_CreateCustList 和 wklt_UpdateCustOrders。工作集 wklt_UpdateCustOrders 需要使用在 wklt_CreateCustList 中更新的工作集变量的值。

下图显示了该工作流：



要将工作集变量值从 wklt_CreateCustList 传递到 wklt_UpdateCustOrders，请完成以下步骤：

1. 将工作集 wklt_CreateCustList 配置为使用工作集变量，例如 \$\$URLString1。
2. 将工作集 wklt_UpdateCustOrders 配置为使用工作集变量，例如 \$\$URLString2。
3. 将工作流配置为使用工作流变量，例如 \$\$PassURLString。
4. 配置工作集 wklt_CreateCustList，使其在工作集完成后将工作集变量 \$\$URLString1 的值分配给工作流变量 \$\$PassURLString。
5. 配置工作集 wklt_UpdateCustOrders，使其在工作集启动之前将工作流变量 \$\$PassURLString 的值分配给工作集变量 \$\$URLString2。

配置变量分配

编辑工作集时，可以在“变量”选项卡上分配变量。在工作集运行之前或之后为以下类型的变量分配值：

- **工作集前变量分配。**在此工作集之前运行的工作流或父工作集中使用父工作流或工作集变量的值或其他任务中的映射变量的值更新用户定义的工作集变量。
- **工作集后变量分配。**使用用户定义的工作集变量的值更新父工作流和工作集变量。

要在工作集中分配变量，请执行以下操作：

1. 编辑您要为其分配变量的工作集。
2. 单击“变量”选项卡。
3. 选择变量分配类型：
 - **工作集前变量分配。**在工作集运行之前向用户定义的工作集变量分配值。
 - **工作集后变量分配。**在工作集完成后向父工作流和工作集变量分配值。
4. 在“变量分配”字段中，单击“编辑”按钮。
5. 在工作集前或工作集后变量分配区域中，单击“添加”按钮添加变量分配语句。

6. 在用户定义的工作集变量和父工作流/工作集变量字段中单击“打开”按钮以选择要读取或分配其值的变量。对于工作集前变量分配，可以在这些字段中输入参数和变量名称。Workflow Manager 不验证参数和变量名称。

Workflow Manager 将分配语句右侧的值分配给语句左侧的变量。因此，如果变量分配语句为“`$$SiteURL_WFVar=$$SiteURL_WkltVar`”，则 Workflow Manager 会将 `$$SiteURL_WkltVar` 的值分配给 `$$SiteURL_WFVar`。

7. 重复步骤 5 到 6 以添加多个变量分配语句。
要删除变更分配语句，请单击分配语句中的其中一个字段，然后单击“剪切”按钮。
8. 单击“确定”。

第 15 章

会话中的参数和变量

本章包括以下主题：

- [使用会话参数, 192](#)
- [会话中的映射参数和变量, 196](#)
- [在会话中分配参数和变量值, 197](#)

使用会话参数

会话参数表示可以在会话运行之间更改的值，如数据库连接数或源和目标文件数。

会话参数可以是用户定义的，也可以是内置的。在会话或工作流属性中使用用户定义的会话参数并在参数文件中定义相应的值。运行会话时，集成服务会将参数文件中的参数与会话中的参数进行匹配。它使用参数文件中的值作为会话属性值。在参数文件中，文件夹和会话名称均区分大小写。

例如，可以将会话日志写入日志文件。在会话属性中，使用 `$PMSessionLogFile` 作为会话日志文件名，并在参数文件中将 `$PMSessionLogFile` 设置为 `TestRun.txt`。运行会话时，集成服务将创建一个名为 `TestRun.txt` 的会话日志。

用户定义的会话参数没有默认值，因此必须在参数文件中对其进行定义。如果集成服务无法为用户定义的会话参数找到值，则会话将失败，将空字符串作为默认值，或者无法在运行时展开参数。

使用 `pmcmd` 启动会话时，可以使用不同的参数文件运行会话。使用 `pmcmd` 设置的参数文件会替代会话或工作流属性中的参数文件。

使用内置会话参数可获取运行时信息，如文件夹名称、服务名称或会话运行统计信息。可以在会话后 `shell` 命令、`SQL` 命令以及电子邮件消息中使用内置会话参数。还可以在接受会话参数的 `Designer` 和 `Workflow Manager` 的输入字段中使用这些参数。集成服务可设置内置会话参数的值。无法在参数文件中定义内置会话参数值。集成服务在会话运行时展开这些参数。

下表介绍了用户定义的会话参数：

参数类型	命名约定	说明
会话日志文件	<code>\$PMSessionLogFile</code>	定义会话运行之间会话日志的名称。
分区数	<code>\$DynamicPartitionCount</code>	定义会话的分区数量。
源文件	<code>\$InputFileName</code>	定义源文件名。 使用适当的前缀定义参数名称。

参数类型	命名约定	说明
查找文件	<i>\$LookupFileName</i>	定义查找文件名。 使用适当的前缀定义参数名称。
目标文件	<i>\$OutputFileNames</i>	定义目标文件名。 使用适当的前缀定义参数名称。
拒绝文件	<i>\$BadFileName</i>	定义拒绝文件名。 使用适当的前缀定义参数名称。
数据库连接	<i>\$DBConnectionName</i>	定义源、目标、查找或存储过程的关系数据库连接。 使用适当的前缀为参数命名。
外部加载器连接	<i>\$LoaderConnectionName</i>	定义外部加载器连接。 使用适当的前缀定义参数名称。
FTP 连接	<i>\$FTPConnectionName</i>	定义 FTP 连接。 使用适当的前缀定义参数名称。
队列连接	<i>\$QueueConnectionName</i>	定义消息队列的数据库连接。 使用适当的前缀定义参数名称。
源或目标应用程序连接	<i>\$AppConnectionName</i>	定义与源和目标应用程序的连接。 使用适当的前缀定义参数名称。
常用会话参数	<i>\$ParamName</i>	定义任何其他会话属性。例如，可以使用此参数来定义表所有者名称、表名称前缀、FTP 文件或目录名称、查找缓存文件名前缀或电子邮件地址。可以使用此参数来定义源、查找、目标和拒绝文件名，但不能定义会话日志文件名或数据库连接。 使用适当的前缀定义参数名称。

下表介绍了内置会话参数：

参数类型	命名约定	说明
文件夹名称	<i>\$PMFolderName</i>	返回文件夹名称。
集成服务名称	<i>\$PMIntegrationServiceName</i>	返回集成服务名称。
映射名称	<i>\$PMMappingName</i>	返回映射名称。
存储库服务名称	<i>\$PMRepositoryServiceName</i>	返回存储库服务名称。
存储库用户名	<i>\$PMRepositoryUserName</i>	返回存储库用户名。
会话名称	<i>\$PMSessionName</i>	返回会话名称。
会话运行模式	<i>\$PMSessionRunMode</i>	返回会话运行模式（正常或恢复）。

参数类型	命名约定	说明
受影响的源行数	<code>\$PMSourceQualifierName@numAffectedRows</code>	返回集成服务从指定源限定符成功读取的行数。 使用适当的前缀和后缀定义参数名称。
已应用的源行数	<code>\$PMSourceQualifierName@numAppliedRows</code>	返回集成服务从指定源限定符成功读取的行数。 使用适当的前缀和后缀定义参数名称。
已拒绝的源行数	<code>\$PMSourceQualifierName@numRejectedRows</code>	返回集成服务从指定源限定符读取时删除的行数。 使用适当的前缀和后缀定义参数名称。
源表名称	<code>\$PMSourceName@TableName</code>	返回指定源实例的表名称。 使用适当的前缀和后缀定义参数名称。
受影响的目标行数	<code>\$PMTargetName@numAffectedRows</code>	返回指定目标实例的指定操作所影响的行数。 使用适当的前缀和后缀定义参数名称。
已应用的目标行数	<code>\$PMTargetName@numAppliedRows</code>	返回集成服务成功应用于指定目标实例的行数。 使用适当的前缀和后缀定义参数名称。
已拒绝的目标行数	<code>\$PMTargetName@numRejectedRows</code>	返回集成服务在写入指定目标实例时拒绝的行数。 使用适当的前缀和后缀定义参数名称。
目标表名称	<code>\$PMTargetName@TableName</code>	返回指定目标实例的表名称。 使用适当的前缀和后缀定义参数名称。
工作流名称	<code>\$PMWorkflowName</code>	返回工作流名称。
工作流运行 ID	<code>\$PMWorkflowRunId</code>	返回工作流运行 ID。
工作流运行实例名称	<code>\$PMWorkflowRunInstanceName</code>	返回工作流运行实例名称。

使用适当的前缀和后缀定义参数名称。例如，对于名称为“Customers”的源实例，源表名称的参数为 `$PMCustomers@TableName`。如果源限定符的名称为“SQ_Customers”，则受影响的源行数的参数为 `$PMSQ_Customers@numAffectedRows`。

更改会话日志名称

您可以对会话进行配置，使其将日志事件写入文件。在会话属性中，“会话日志文件目录”默认为服务进程变量 `$PMSessionLogDir`。“会话日志文件名”默认为 `$PMSessionLogFile`。

在参数文件中，您将 `$PMSessionLogFile` 设置为 `TestRun.txt`。在 Administrator 工具中，您将 `$PMSessionLogDir` 定义为 `\\server\infa_shared\SessLogs`。集成服务运行会话时，将在 `\\server\infa_shared\SessLogs` 目录中创建一个名为 `TestRun.txt` 的会话日志文件。

更改目标文件和目录

使用会话属性中的目标文件参数来更改会话的目标文件和目录。可以在“输出文件名”字段中输入包括目录和文件名的路径。如果“输出文件名”字段中包括目录，则必须清除“输出文件目录”。集成服务会连接输出文件目录和输出文件名以确定目标文件位置。

例如，会话使用文件参数来读取内部和外部 Web 日志。您希望将内部 Web 日志会话的结果写入一个位置，并将外部 Web 日志会话写入另一个位置。

在会话属性中，您将目标文件命名为 `$OutputFileName`，并清除“输出文件目录”字段。在参数文件中，将 `$OutputFileName` 设置为 `E:/internal_weblogs/November_int.txt` 以为内部 Web 日志会话创建目标文件。在会话完成后，为外部 Web 日志会话将 `$OutputFileName` 更改为 `F:/external_weblogs/November_ex.txt`。

可以为每个目标创建不同的参数文件，并使用 `pmcmd` 启动使用特定参数文件的会话。此参数文件可替代会话属性中的参数文件名。

更改文件中的源参数

您可以为参数文件中的会话属性定义多个参数并在会话中使用其中一个参数。可以在会话属性中更改参数名称并使用其他参数值重新运行会话。

例如，您在参数文件中创建了名为 `$InputFile_Products` 的会话参数。您将参数值设置为“products.txt”。在同一参数文件中，创建另一个名为 `$InputFile_Items` 的参数。您将参数值设置为“items.txt”。

当您在会话属性中将源文件名设置为 `$InputFile_Products` 时，集成服务会读取 `products.txt`。当您将源文件名更改为 `$InputFile_Items` 时，集成服务会读取 `items.txt`。

更改连接参数

使用连接参数可重新运行具有不同源、目标、查找表或存储过程的会话。可以在任何会话的会话属性中创建连接参数。可以引用参数中的任何连接。可以使用适当的前缀，后跟任何字母数字和下划线字符的方式对所有连接会话参数进行命名。

例如，运行一个从两个关系源进行读取的会话。可以访问一个数据库连接命名为“市场”的源以及另一个数据库连接命名为“销售”的源。在会话属性中，创建一个名为 `$DBConnection_Source` 的源数据库连接参数。在参数文件中，可以将 `$DBConnection_Source` 定义为市场并运行该会话。运行下一个会话时，请在参数文件中将 `$DBConnection_Source` 设置为销售。

如果使用连接参数替代源或目标的连接，则可以替代参数文件中的连接属性。将非关系连接参数用于源或目标实例时，可以替代连接属性。在参数文件中定义连接时，集成服务会搜索定义连接属性的用户设定的会话参数。例如，创建一个名为 `$FTPConnectionMyFTPConn` 的 FTP 连接参数，然后在参数文件中对其进行定义。集成服务会搜索以下参数的参数文件：

- `$Param_FTPConnectionMyFTPConn_Remote_Filename`
- `$Param_FTPConnectionMyFTPConn_Is_Staged`
- `$Param_FTPConnectionMyFTPConn_Is_Transfer_Mode_ASCII`

如果未对上述任一参数定义值，则集成服务将使用在连接对象中定义的值。

在以下模板文件中列出了可以替代的连接属性：

```
<PowerCenter Installation Directory>/server/bin/ConnectionParam.prm
```

获取运行时信息

使用内置会话参数可获取运行时信息，如文件夹名称、集成服务名称以及源和目标表名称。可以在会话后 shell 命令、SQL 命令以及电子邮件消息中使用内置会话参数。还可以在接受会话参数的 Designer 和 Workflow Manager 的输入字段中使用这些参数。

例如，希望在会话“s_UpdateCustInfo”完成后发送一封会话后电子邮件，其中包括源限定符“SQ_Customers”和目标“T_CustInfo”的会话运行统计信息。在电子邮件消息的正文中输入以下文本：

```
Statistics for session $PMSessionName
Integration service: $PMIntegrationServiceName
Source number of affected rows: $PMSQ_Customers@numAffectedRows
Source number of dropped rows: $PMSQ_Customers@numRejectedRows
Target number of affected rows: $PMT_CustInfo@numAffectedRows
Target number of applied rows: $PMT_CustInfo@numAppliedRows
Target number of rejected rows: $PMT_CustInfo@numRejectedRows
```

还可以使用电子邮件变量获取会话名称、集成服务名称、加载的行数以及拒绝的行数。

创建文件参数和数据库连接参数的规则和准则

通过会话文件参数和数据库连接参数，可以灵活地对不同文件和数据库运行会话。

创建文件参数时，请遵循以下规则和准则：

- 将参数文件定义为节点的资源时，验证集成服务是否在可访问该参数文件的节点上运行会话。定义节点的资源，将集成服务配置为检查资源，并将会话编辑为需要该资源。
- 创建文件参数时，使用字母数字和下划线字符。例如，要命名源文件参数，请使用 *\$InputFileName*（如 *\$InputFile_Data*）。
- 特定类型的所有会话文件参数必须具有不同名称。例如，如果创建两个源文件参数，可以将它们命名为 *\$SourceFileAccts* 和 *\$SourceFilePrices*。
- 在文件中定义参数时，可以引用集成服务的任何本地目录。
- 使用参数定义文件的位置。在会话属性中清除定义了该文件位置的条目。在参数文件中输入该文件的完整路径。
- 您可以在会话运行之间在参数文件中更改参数值，也可以创建多个参数文件。如果使用多个参数文件，请将 *pmcmd Startworkflow* 命令与 *-paramfile* 或 *-localparamfile* 选项结合使用以指定要使用哪个参数文件。

创建数据库连接参数时，请遵循以下规则和准则：

- 可以为关系源、目标、查找和存储过程更改连接。
- 定义参数时，可以引用存储库中的任何数据库连接。
- 对话话中的多个连接使用同一 *\$DBConnection* 参数。

会话中的映射参数和变量

在会话属性中使用映射参数更改特定的映射属性。例如，在转换替代中使用映射参数替代源限定符转换中的筛选器或用户定义的联接。

如果在会话中使用映射变量，则可以通过编辑会话清除存储库中保存的任何变量值。清除变量值时，集成服务会在下次运行会话时使用参数文件中的值。如果会话不使用参数文件，则集成服务会使用在会话前变量分配中分配的值。如果不存在任何分配的值，集成服务会使用在映射中定义的初始值。

要查看或删除存储库中保存的映射变量的值，请执行以下操作：

1. 在 Workflow Manager 的“导航器”窗口中，右键单击会话任务并选择“查看永久性值”。
您可以查看变量名称及其值。
2. 单击“删除值”以删除现有的变量值。
3. 单击“确定”。

在会话中分配参数和变量值

您可以在不可重用会话运行前后更新特定参数和变量的值。这允许您将信息从一个会话传递到同一工作流或工作集中的另一个会话。例如，您拥有一个工作流，其包含两个需要增加相同计数器的会话。您可以在第一个会话中增加计数器，将更新后的计数器值传递给第二个会话，然后再在第二个会话中增加计数器。或者，您拥有一个工作集，其包含可访问同一网站的会话。您可以将第一个会话配置为从网站中获取会话 ID，然后将会话 ID 值传递给后续会话。

还可以将信息从会话传递到工作集或从工作集传递到会话，只要该会话和该工作集位于同一工作流或父工作集中。

注意：无法在可重用会话中分配参数和变量。

可更新的参数和变量的类型取决于是在会话运行之前还是之后分配它们。可以在会话运行前后更新以下类型的参数和变量：

- **会话前变量分配。**可以在会话运行之前更新映射参数、映射变量和会话参数。可以为这些参数和变量分配父工作流或工作集中工作流或工作集变量的值。因此，如果某个会话位于工作流内的工作集中，则可以从工作集变量（而非工作流变量）分配值。

无法在会话前变量分配中更新 Mapplet 变量。

- **成功会话后的变量分配。**可以在会话成功完成之后更新父工作流或工作集中的工作流或工作集变量。可以为这些变量分配映射参数和变量的值。
- **失败会话后的变量分配。**可以在会话失败时更新父工作流或工作集中的工作流或工作集变量。可以为这些变量分配映射参数和变量的值。

可以在会话属性的“组件”选项卡上分配参数和变量。

在会话之间传递参数和变量值

可以在会话中分配参数和变量值，以便将这些值从一个会话传递到同一工作流或工作集中的任何后续会话。例如，某个工作流包含两个会话 s_NewCustomers 和 s_MergeCustomers。会话 s_MergeCustomers 需要使用在 s_NewCustomers 中更新的映射变量的值。

下图显示了该工作流：



要将映射变量值从 s_NewCustomers 传递到 s_MergeCustomers，请完成以下步骤：

1. 将与会话 s_NewCustomers 关联的映射配置为使用映射变量（如 \$\$Count1）。
2. 将与会话 s_MergeCustomers 关联的映射配置为使用映射变量（如 \$\$Count2）。
3. 将工作流配置为使用用户定义的工作流变量（如 \$\$PassCountValue）。

4. 将会话 s_NewCustomers 配置为在会话成功完成之后将映射变量 \$\$Count1 的值分配给工作流变量 \$PassCountValue。
5. 将会话 s_MergeCustomers 配置为在会话启动之前将工作流变量 \$\$PassCountValue 的值分配给映射变量 \$\$Count2。

配置变量分配

编辑工作集时，可以在“变量”选项卡上分配变量。在工作集运行之前或之后为以下类型的变量分配值：

- **工作集前变量分配。**在此工作集之前运行的工作流或父工作集中使用父工作流或工作集变量的值或其他任务中的映射变量的值更新用户定义的工作集变量。
- **工作集后变量分配。**使用用户定义的工作集变量的值更新父工作流和工作集变量。

要在工作集中分配变量，请执行以下操作：

1. 编辑您要为其分配变量的工作集。
2. 单击“变量”选项卡。
3. 选择变量分配类型：
 - **工作集前变量分配。**在工作集运行之前向用户定义的工作集变量分配值。
 - **工作集后变量分配。**在工作集完成后向父工作流和工作集变量分配值。
4. 在“变量分配”字段中，单击“编辑”按钮。
5. 在工作集前或工作集后变量分配区域中，单击“添加”按钮添加变量分配语句。
6. 在用户定义的工作集变量和父工作流/工作集变量字段中单击“打开”按钮以选择要读取或分配其值的变量。对于工作集前变量分配，可以在这些字段中输入参数和变量名称。Workflow Manager 不验证参数和变量名称。

Workflow Manager 将分配语句右侧的值分配给语句左侧的变量。因此，如果变量分配语句为“\$ \$\$SiteURL_WFVar=\$\$SiteURL_WkltVar”，则 Workflow Manager 会将 \$\$SiteURL_WkltVar 的值分配给 \$ \$\$SiteURL_WFVar。
7. 重复步骤 5 到 6 以添加多个变量分配语句。

要删除变更分配语句，请单击分配语句中的其中一个字段，然后单击“剪切”按钮。
8. 单击“确定”。

第 16 章

形参文件

本章包括以下主题：

- [参数文件概览, 199](#)
- [参数和变量类型, 200](#)
- [在何处使用参数和变量, 200](#)
- [替代参数文件中的连接属性, 207](#)
- [参数文件结构, 208](#)
- [配置参数文件名称和位置, 210](#)
- [参数文件示例, 213](#)
- [创建参数文件的准则, 213](#)
- [参数和参数文件故障排除, 214](#)
- [参数和参数文件的提示, 215](#)

参数文件概览

参数文件列出了参数和变量及其关联值。这些值可以定义服务、服务进程、工作流、工作集或会话的属性。运行使用该参数文件的工作流或会话时，集成服务会应用这些值。

参数文件使您能够在每次运行会话或工作流时灵活地更改参数和变量值。可以将多个服务、服务进程、工作流、工作集和会话的信息包括在一个参数文件中。还可以创建多个参数文件，并在每次运行会话或工作流时使用不同的文件。集成服务会在工作流或会话开始时读取参数文件以确定在该文件中定义的参数和变量的开始值。可以使用文本编辑器（如 WordPad 或 Notepad）创建参数文件。

使用参数文件时，请注意以下信息：

- **参数和变量的类型。**可以在参数文件中定义不同类型的参数和变量。其中包括服务变量、服务进程变量、工作流和工作集变量、会话参数以及映射参数和变量。
- **可以在参数文件中设置的属性。**使用参数和变量可在 Designer 和 Workflow Manager 中定义多个属性。例如，可以输入会话参数作为关系目标实例的更新替代值，然后在参数文件中将此参数设置为 UPDATE 语句。集成服务会在会话运行时展开该参数。
- **参数文件结构。**通过在单个行中按格式 *name=value* 输入参数或变量名称和值的方式为参数文件中的参数或变量分配值。参数和变量组的前面必须有一个标题，该标题用于标识要应用参数或变量的服务、服务进程、工作流、工作集或会话。
- **参数文件位置。**为工作流或会话指定要使用的参数文件。可以在工作流或会话属性或 *pmcmd* 命令行中输入参数文件名称和目录。

参数和变量类型

参数文件可以包含不同类型的参数和变量。运行使用参数文件的会话或工作流时，集成服务会读取参数文件并展开在该文件中定义的参数和变量。

可以在参数文件中定义以下类型的参数和变量：

- **服务变量。**定义集成服务的常规属性，如电子邮件地址、日志文件计数以及错误阈值。
\$PMSuccessEmailUser、\$PMSessionLogCount 和 \$PMSessionErrorThreshold 是服务变量的一些示例。
在参数文件中定义的服务变量值会替代在 Administrator 工具中设置的值。
- **服务进程变量。**为每个集成服务进程定义集成服务文件的目录。\$PMRootDir、\$PMSessionLogDir 和 \$PMBadFileDir 是服务进程变量的一些示例。在参数文件中定义的服务进程变量值会替代在 Administrator 工具中设置的值。如果集成服务使用操作系统配置文件，则在操作系统配置文件中指定的操作系统用户必须有权访问为服务进程变量定义的目录。
- **工作流变量。**评估工作流中的任务条件和记录信息。例如，可以在判定任务中使用工作流变量以确定上一个任务是否正常运行。在工作流中，\$TaskName.PrevTaskStatus 属于预定义的工作流变量，\$VariableName 属于用户定义的工作流变量。
- **工作集变量。**评估工作集中的任务条件和记录信息。可以在父工作流中使用预定义的工作集变量，但不可以在工作集中使用父工作流中的工作流变量。在工作集中，\$TaskName.PrevTaskStatus 属于预定义的工作集变量，\$\$VariableName 属于用户定义的工作集变量。
- **会话参数。**定义可以在会话之间进行更改的值，如数据库连接或文件名称。\$PMSessionLogFile 和 \$ParamName 均属于用户定义的会话参数。
- **映射参数。**定义在整个会话中保持不变的值，如州销售税率。在映射或 Mapplet 中声明时，\$ParameterName 属于用户定义的映射参数。
- **映射变量。**定义可以在会话期间更改的值。集成服务会在每次成功会话运行结束时将映射变量的值保存到存储库，并在下次运行会话时使用该值。在映射或 Mapplet 中声明时，\$\$VariableName 属于映射变量。

在参数文件中无法定义以下类型的变量：

- **\$Source 和 \$Target 连接变量。**为关系源、关系目标、查找表或存储过程定义数据库位置。
- **电子邮件变量。**定义电子邮件消息中的会话信息，如加载的行数、会话完成时间以及读取和写入统计信息。
- **局部变量。**将数据临时存储在汇总器、表达式以及等级转换的变量端口中。
- **内置变量。**用于返回运行时或系统信息的变量，如集成服务名称或系统日期。
- **事务控制变量。**定义在处理数据库行期间提交或回滚事务的条件。
- **ABAP 程序变量。**表示 SAP 结构、SAP 结构中的字段或 ABAP 程序中的值。

在何处使用参数和变量

可以使用参数和变量来向 Designer 和 Workflow Manager 中的属性分配值并替代某些服务和进程属性。例如，可以使用参数来指定查找缓存文件名前缀或 FTP 连接的默认远程目录。

如果属性为 SQL 语句或命令，可以在该语句或命令中使用参数和变量，也可以在该属性的输入字段中输入参数或变量，然后在参数文件中将该参数或变量设置为整个语句或命令。

例如，您要在关系目标替代中使用参数或变量。可以在关系目标替代的 UPDATE 语句中输入参数或变量，并在参数文件中的相应标题下定义参数或变量。或者，要在参数文件中定义 UPDATE 语句，请完成以下步骤：

1. 在 Designer 中，编辑目标实例，在“更新替代”字段中输入会话参数 \$ParamMyOverride，然后保存映射。

- 2. 在 Workflow Manager 中，将工作流或会话配置为使用参数文件。
- 3. 在参数文件的相应标题下将 \$ParamMyOverride 设置为 SQL UPDATE 语句。

还可以使用参数文件替代 Administrator 工具中定义的服务和服务进程属性。例如，可以替代会话日志目录 \$PMSessionLogDir。要执行此操作，请将工作流或会话配置为使用参数文件，并在参数文件中将 \$PMSessionLogDir 设置为新的文件路径。

可以为以下 PowerCenter 对象指定参数和变量：

- **源。**可以在与源相关的输入字段中使用参数和变量。
- **目标。**可以在与目标相关的输入字段中使用参数和变量。
- **转换。**可以在与转换相关的输入字段中使用参数和变量。
- **任务。**可以在与 Workflow Manager 中任务相关的输入字段中使用参数和变量。
- **会话。**可以在与会话任务相关的输入字段中使用参数和变量。
- **工作流。**可以在与工作流相关的输入字段中使用参数和变量。
- **连接。**可以在与连接对象相关的输入字段中使用参数和变量。

下表列出了可以在其中指定参数和变量的与源相关的输入字段：

源类型	字段	有效的参数和变量类型
关系	源表名称	工作流变量、工作集变量、会话参数、映射参数和映射变量。 在 Workflow Manager 的会话属性（“映射”选项卡）中替代此字段时，可以在此字段中指定参数和变量。
PeopleSoft	SetID、 生效日期、 树名称、 设置控制值、 提取日期	全部。
TIBCO	TIB/Adapter SDK 存储库 URL	服务和服务进程变量。
Web 服务	端点 URL	映射参数和变量 在 Workflow Manager 的会话属性（“映射”选项卡）中替代此字段时，可以在此字段中指定参数和变量。

下表列出了可以在其中指定参数和变量的与目标相关的输入字段：

目标类型	字段	有效的参数和变量类型
关系	更新替代 会话前和会话后 SQL 命令	全部。 在 Workflow Manager 的会话属性（“映射”选项卡）中替代这些字段时，可以在这些字段中指定参数和变量。
关系	目标表名称	工作流变量、工作集变量、会话参数、映射参数和映射变量。 在 Workflow Manager 的会话属性（“映射”选项卡）中替代此字段时，可以在此字段中指定参数和变量。
XML	缓存目录	服务和进程变量。 在 Workflow Manager 的会话属性（“映射”选项卡）中替代此字段时，可以在此字段中指定参数和变量。
TIBCO	TIB/Adapter SDK 存储库 URL	服务和进程变量。
Web 服务	端点 URL	映射参数和变量。 在 Workflow Manager 的会话属性（“映射”选项卡）中替代此字段时，可以在此字段中指定参数和变量。

下表列出了可以在其中指定参数和变量的与转换相关的输入字段：

转换类型	字段	有效的参数和变量类型
允许您使用表达式编辑器的转换	转换表达式	映射参数和变量。
汇总器、联接器、查找、等级、XML 生成器	缓存目录	服务和进程变量。 在 Workflow Manager 的会话属性（“映射”选项卡）中替代此字段时，可以在此字段中指定参数和变量。
汇总器、联接器、查找、等级、排序器	缓存大小	映射参数。 在 Workflow Manager 的会话属性（“映射”选项卡）中替代此字段时，可以在此字段中指定参数和变量。
自定义、外部过程、HTTP、XML 解析器	运行时位置	服务和进程变量。 在 Workflow Manager 的会话属性（“映射”选项卡）中替代此字段时，可以在此字段中指定参数和变量。
数据屏蔽	种子	映射参数和变量。
外部过程	初始化属性	服务和进程变量。
HTTP	基础 URL	映射参数和变量。

转换类型	字段	有效的参数和变量类型
查找	SQL 替代 缓存文件名前缀	全部。 在 Workflow Manager 的会话属性（“映射”选项卡）中替代此字段时，可以在此字段中指定参数和变量。
查找	连接信息	会话参数 <i>\$DBConnectionName</i> 和 <i>\$AppConnectionName</i> 、连接变量 <i>\$Source</i> 和 <i>\$Target</i> 、映射参数和变量。 在 Workflow Manager 的会话属性（“映射”选项卡）中替代此字段时，可以在此字段中指定参数和变量。
排序器	默认工作目录	服务和进程变量。 在 Workflow Manager 的会话属性（“映射”选项卡）中替代此字段时，可以在此字段中指定参数和变量。
源限定符（关系源）	SQL 查询 用户定义的联接 源筛选条件 会话前和会话后 SQL 命令	全部。 在 Workflow Manager 的会话属性（“映射”选项卡）中替代此字段时，可以在此字段中指定参数和变量。
SQL	脚本文件名	映射参数和变量。 在 Workflow Manager 的会话属性（“映射”选项卡）中替代此字段时，可以在此字段中指定参数和变量。
存储过程	调用文本（未连接存储过程）	全部。 在 Workflow Manager 的会话属性（“映射”选项卡）中替代此字段时，可以在此字段中指定参数和变量。
存储过程	连接信息	会话参数 <i>\$DBConnectionName</i> 、连接变量 <i>\$Source</i> 和 <i>\$Target</i> 。 在 Workflow Manager 的会话属性（“映射”选项卡）中替代此字段时，可以在此字段中指定参数和变量。
Web 服务使用者	端点 URL	映射参数和变量。 在 Workflow Manager 的会话属性（“映射”选项卡）中替代此字段时，可以在此字段中指定参数和变量。

下表列出了可以在其中指定参数和变量的与 Workflow Manager 任务相关的输入字段：

任务类型	字段	有效的参数和变量类型
分配任务	分配（用户定义的变量和表达式）	工作流和工作集变量
命令任务	命令（名称和命令）	服务、服务进程、工作流和工作集变量

任务类型	字段	有效的参数和变量类型
命令任务	会话前和会话后 Shell 命令	全部
判定任务	判定名称（要评估的条件）	工作流和工作集变量
电子邮件任务	电子邮件用户名、主题和文本	服务、服务进程、工作流和工作集变量
事件等待任务	文件监视名称（预定义事件）	服务、服务进程、工作流和工作集变量
链接	链接条件	服务、服务进程、工作流和工作集变量
会话	请参阅 the table 页面上 204 。	
计时器任务	绝对时间：用于计算等待时间的工作流日期时间变量	工作流和工作集变量

下表列出了可以在其中指定参数和变量的与会话相关的输入字段：

选项卡	字段	有效的参数和变量类型
“属性”选项卡	会话日志文件名	内置会话参数 <code>\$PMSessionLogFile</code> 。
“属性”选项卡	会话日志文件目录	服务和进程变量。
“属性”选项卡	参数文件名	工作流和工作集变量。
“属性”选项卡	<code>\$Source</code> 和 <code>\$Target</code> 连接值	会话参数 <code>\$DBConnectionName</code> 和 <code>\$AppConnectionName</code> 、连接变量 <code>\$Source</code> 和 <code>\$Target</code> 。
“属性”选项卡	下推优化会话属性	映射参数 <code>\$PushdownConfig</code> 。
“配置对象”选项卡	会话日志计数	服务变量 <code>\$PMSessionLogCount</code> 。
“配置对象”选项卡	会话错误阈值	服务变量 <code>\$PMSessionErrorThreshold</code> 。
“配置对象”选项卡	关系错误日志的表名称前缀	全部。
“配置对象”选项卡	错误日志文件名和目录	服务变量、服务进程变量、工作流变量、工作集变量、会话参数。
“配置对象”选项卡	动态分区的分区数量	内置会话参数 <code>\$DynamicPartitionCount</code> 。
“映射”选项卡	替代在映射中配置的属性的转换属性	因属性不同而异。有关详细信息，请参阅 the table 页面上 202 。
“映射”选项卡	关系连接值	会话参数 <code>\$DBConnectionName</code> 、连接变量 <code>\$Source</code> 和 <code>\$Target</code> 。
“映射”选项卡	队列连接值	会话参数 <code>\$QueueConnectionName</code> 。 可以在参数文件中替代此连接类型的连接属性。

选项卡	字段	有效的参数和变量类型
“映射”选项卡	FTP 连接值	会话参数 <i>\$FTPConnectionName</i> 。 可以在参数文件中替代此连接类型的连接属性。
“映射”选项卡	应用程序连接值	会话参数 <i>\$AppConnectionName</i> 。 可以在参数文件中替代此连接类型的连接属性。
“映射”选项卡	外部加载器连接值	会话参数 <i>\$LoaderConnectionName</i> 。 可以在参数文件中替代此连接类型的连接属性。
“映射”选项卡	FTP 远程文件名	全部。
“映射”选项卡	查找源文件名和目录	服务变量、服务进程变量、工作流变量、工作集变量、会话参数。
“映射”选项卡	会话前和会话后 SQL 命令（源和目标）	全部。
“映射”选项卡	文件源和目标的代码页	工作流变量、工作集变量、会话参数 <i>\$ParamName</i> 。
“映射”选项卡	源输入文件名和目录	服务变量、服务进程变量、工作流变量、工作集变量、会话参数。
“映射”选项卡	源输入文件命令	服务变量、服务进程变量、工作流变量、工作集变量、会话参数。
“映射”选项卡	关系源的表所有者名称	全部。
“映射”选项卡	目标合并文件名和目录	服务变量、服务进程变量、工作流变量、工作集变量、会话参数。
“映射”选项卡	目标合并命令	服务变量、服务进程变量、工作流变量、工作集变量、会话参数。
“映射”选项卡	目标表头和表尾命令	服务变量、服务进程变量、工作流变量、工作集变量、会话参数。
“映射”选项卡	目标输出文件名和目录	服务变量、服务进程变量、工作流变量、工作集变量、会话参数。
“映射”选项卡	目标拒绝文件名和目录	服务变量、服务进程变量、工作流变量、工作集变量、会话参数。
“映射”选项卡	目标表名称前缀	全部。
“映射”选项卡	Teradata FastExport 临时文件	服务和进程变量。
“映射”选项卡	Teradata 外部加载器的控制文件内容替代	全部。

选项卡	字段	有效的参数和变量类型
“映射”选项卡	WebSphere MQ、JMS、SAP ALE IDoc、TIBCO、webMethods、Web Service Provider 源的恢复缓存目录	服务和进程变量。
“映射”选项卡	持久订阅名称	会话参数 <i>\$ParamName</i> 。
“映射”选项卡	MQ 源限定符筛选条件	全部。
“映射”选项卡	SAP 暂存文件名和目录	服务变量、服务进程变量、工作流变量、工作集变量、会话参数。
“映射”选项卡	SAP 源文件目录	服务变量、服务进程变量、工作流变量、工作集变量、会话参数。
“组件”选项卡	会话后电子邮件（用户名、主题和文本）	全部。
“组件”选项卡	会话后电子邮件附件文件名	全部。

下表列出了可以在其中指定参数和变量的与工作流相关的输入字段：

选项卡	字段	有效的参数和变量类型
“属性”选项卡	工作流日志文件名和目录	服务、服务进程、工作流和工作集变量。
“属性”选项卡	工作流日志计数	服务变量 <i>\$PMWorkflowLogCount</i> 。
“常规”选项卡	挂起电子邮件（用户名、主题和文本）	服务、服务进程、工作流和工作集变量。

下表列出了可以在其中指定参数和变量的与 Workflow Manager 中连接对象相关的输入字段：

连接类型	字段	有效的参数和变量类型
关系	数据库用户名、密码	会话参数 <i>\$ParamName</i> 。 在参数文件中使用 <code>pmpasswd</code> 命令行程序通过 <code>CRYPT_DATA</code> 加密类型加密密码。
关系：源、目标、查找、存储过程	连接和事务环境 SQL	全部。
FTP	主机的用户名、密码	会话参数 <i>\$ParamName</i> 。 在参数文件中使用 <code>pmpasswd</code> 命令行程序通过 <code>CRYPT_DATA</code> 加密类型加密密码。
FTP	默认远程目录	全部。
应用程序	应用程序用户名、密码	会话参数 <i>\$ParamName</i> 。 在参数文件中使用 <code>pmpasswd</code> 命令行程序通过 <code>CRYPT_DATA</code> 加密类型加密密码。

连接类型	字段	有效的参数和变量类型
应用程序：Web 服务使用者	端点 URL	会话参数 <i>\$ParamName</i> 、映射参数和变量。
应用程序：HTTP	基础 URL	会话参数 <i>\$ParamName</i> 。
应用程序：JMS	JMS 目标	会话参数 <i>\$ParamName</i> 。
加载器	数据库用户名、密码	会话参数 <i>\$ParamName</i> 。 在参数文件中使用 <code>pmpasswd</code> 命令行程序通过 <code>CRYPT_DATA</code> 加密类型加密密码。

替代参数文件中的连接属性

如果使用会话参数定义源或目标的连接，则可以替代参数文件中的连接属性。使用 *\$FTPConnectionName*、*\$QueueConnectionName*、*\$LoaderConnectionName* 或 *\$AppConnectionName* 会话参数。

在参数文件中定义连接时，集成服务会搜索用户设定的会话参数来定义连接属性。例如，创建一个名为 *\$QueueConnectionMyMQ* 的消息队列连接参数，并在参数文件的 “[s_MySession]” 部分对其定义。集成服务会在参数文件的此部分搜索 “每条消息的行数” 参数 *\$Param_QueueConnectionMyMQ_Rows_Per_Message*。

安装 PowerCenter 时，安装程序会创建一个名为 *ConnectionParam.prm* 的模板文件，该文件列出了可针对 FTP、队列、加载程序以及应用程序连接替代的连接属性。 *ConnectionParam.prm* 文件位于以下目录中：

```
<PowerCenter Installation Directory>/server/bin
```

在参数文件中定义连接时，请复制相应连接类型的模板并将其粘贴到参数文件中。然后提供参数值。

例如，要替代参数文件中 FTP 连接的连接属性，请执行以下步骤：

- 1. 将会话或工作流配置为使用参数文件运行。
- 2. 在会话属性的 “映射” 选项卡上，请在 “连接” 节点上选择源或目标实例。
- 3. 单击值字段中的 “打开” 按钮，然后将该连接配置为使用会话参数。例如，将 *\$FTPConnectionMyFTPConn* 用于 FTP 连接。
- 4. 在文本编辑器中打开 *ConnectionParam.prm* 模板文件，然后滚动至您要替代的属性的连接类型部分。例如，对于 FTP 连接，请查找 “连接类型：FTP” 部分：

```
Connection Type : FTP
-----

...

Template
=====

$FTPConnection<VariableName>=
$Param_FTPConnection<VariableName>_Remote_Filename=
```

```
$Param_FTPConnection<VariableName>_Is_Staged=  
$Param_FTPConnection<VariableName>_Is_Transfer_Mode_ASCII=
```

5. 复制您要替代的连接属性的模板文本。例如，要替代“远程文件名称”和“是否暂存”属性，请复制以下行：

```
$FTPConnection<VariableName>=  
$Param_FTPConnection<VariableName>_Remote_Filename=  
$Param_FTPConnection<VariableName>_Is_Staged=
```

6. 将文本粘贴到参数文件中。将 <VariableName> 替换为连接名称，然后提供参数值。例如：

```
[MyFolder.WF:wf_MyWorkflow.ST:s_MySession]  
  
$FTPConnectionMyFTPConn=FTP_Conn1  
$Param_FTPConnectionMyFTPConn_Remote_Filename=ftp_src.txt  
$Param_FTPConnectionMyFTPConn_Is_Staged=YES
```

注意：集成服务会将等号前后的空格或引号解释为参数名称或值的一部分。

如果未定义属性值，则集成服务会使用为该连接对象定义的值。

参数文件结构

参数文件包含具有已分配值的参数和变量的列表。您可以对参数和变量分组放入参数文件的不同部分。各个部分的前面均有一个标题，该标题可标识要为其定义参数或变量的集成服务、集成服务进程、工作流、工作集或会话。您可以直接在标题下方定义参数和变量，在一个新行输入每个参数或变量。在该部分内可以按任意顺序列出参数和变量。

按 *name=value* 格式输入参数或变量定义。例如，以下行可向服务变量 \$PMSuccessEmailUser 和会话参数 \$ParamTgtOverride 分配值：

```
$PMSuccessEmailUser=rsmith@mail.com  
$ParamTgtOverride=UPDATE T_SALES SET DATE_SHIPPED = :TU.DATE_SHIPPED, TOTAL_SALES = :TU.TOTAL_SALES  
WHERE :TU.EMP_NAME = EMP_NAME and EMP_NAME = 'MIKE SMITH'
```

集成服务会将行开头与第一个等号之间的所有字符解释为参数名称，并将第一个等号与行结尾之间的所有字符解释为参数值。因此，如果在参数名称与等号之间输入空格，集成服务会将该空格解释为参数名称的一部分。如果行中包含多个等号，则集成服务会将第一个等号之后的所有等号均解释为参数值的一部分。

警告：使用参数文件运行工作流时，集成服务会使用句点字符 (.) 来验证文件夹、工作流和会话名称。如果文件夹名称包含句点 (.)，则集成服务无法正确验证名称，工作流将失败。

参数文件部分

可以在参数文件的任何部分定义参数和变量。如果在工作流、工作集或会话部分定义服务或服务进程变量，则变量将应用于运行任务的服务进程。同样，如果在会话部分定义工作流变量，则该工作流变量的值只有在运行会话时才适用。

下表介绍了在参数文件中定义各个部分的参数文件标题和在各个部分定义的参数和变量范围：

标题	范围
[全局]	所有集成服务、集成服务进程、工作流、工作集和会话。
[服务：服务名称]	命名的集成服务和运行此服务的工作流、工作集和会话。
[Service:service name.ND:node name]	命名的集成服务进程和运行此服务进程的工作流、工作集和会话。
[文件夹名称.WF:工作流名称]	命名的工作流和工作流中的所有会话。
[文件夹名称.WF:工作流名称.WT:工作集名称]	命名的工作集和工作集中的所有会话。
[文件夹名称.WF:工作流名称.WT:工作集名称.WT:工作集名称...]	嵌套工作集和嵌套工作集中的所有会话。
[文件夹名称.WF:工作流名称.ST:会话名称] -或- [文件夹名称.WF:工作流名称.WT:工作集名称.ST:会话名称] -或- [文件夹名称.会话名称] -或- [会话名称]	命名的会话。

在参数文件中一次只能创建一个标题。如果在参数文件中多次指定相同的标题，则集成服务将使用第一个标题下方的部分中的信息，并忽略后续相同标题下方的部分中的信息。例如，参数文件包含以下相同的标题：

```
[HET_TGTS.WF:wf_TCOMMIT1]
$$platform=windows
...
[HET_TGTS.WF:wf_TCOMMIT1]
$$platform=unix
$DBConnection_ora=Ora2
```

在工作流 wf_TCOMMIT1 中，映射参数 *\$\$platform* 的值为“windows”而不是“unix”，且未定义会话参数 *\$DBConnection_ora*。

如果你在参数文件中的多个部分中定义相同的参数或变量，最小范围的参数或变量的优先级高于更大范围的参数或变量。例如，参数文件包含以下部分：

```
[HET_TGTS.WF:wf_TGTS_ASC_ORDR]
$DBConnection_ora=Ora2
[HET_TGTS.WF:wf_TGTS_ASC_ORDR.ST:s_TGTS_ASC_ORDR]
$DBConnection_ora=Ora3
```

在会话 s_TGTS_ASC_ORDR 中，会话参数 *\$DBConnection_ora* 的值为“Ora3”。在工作流的所有其他部分中，它是“Ora2”。

注释

在参数文件中可以包括注释。集成服务会忽略无效标题且不包含等号字符 (=) 的行。以下行是一些参数文件注释的示例：

```
-----
Created 10/11/06 by JSmith.
```

```
*** Update the parameters below this line when you run this workflow on Integration Service Int_01. ***  
; This is a valid comment because this line contains no equals sign.
```

空值

可以向参数文件中的参数和变量分配空值。向参数和变量分配空值时，集成服务会根据参数或变量类型获取以下位置的值：

- **服务和进程变量。**集成服务将使用在 Administrator 工具中设置的值。
- **工作流和工作集变量。**集成服务将使用保存在存储库中的值（如果变量具有持久性）、用户定义的默认值或数据类型默认值。
- **会话参数。**会话参数没有默认值。如果集成服务找不到会话参数值，则会话可能会失败并将空字符串视为默认值，或无法在运行时展开该参数。例如，如果集成服务执行会话时，该会话未定义会话参数 `$DBConnectionName`，则会话将失败。
- **映射参数和变量。**集成服务将使用保存在存储库中的值（仅限映射变量）、配置的初始值或数据类型默认值。

要分配空值，请将参数或变量值设置为 “<null>” 或将该值保留为空白。例如，以下行可向服务进程变量 `$PMBadFileDir` 和 `$PMCacheDir` 分配空值：

```
$PMBadFileDir=<null>  
$PMCacheDir=
```

示例参数文件

以下文本摘自参数文件，其中包含一个集成服务的服务变量和四个工作流的参数：

```
-----  
File created by RSmith 11/12/2005  
-----  
[Service:IntSvs_01]  
$PMSuccessEmailUser=pcadmin@mail.com  
$PMFailureEmailUser=pcadmin@mail.com  
[HET_TGTS.WF:wf_TCOMMIT_INST_ALIAS]  
$$platform=unix  
[HET_TGTS.WF:wf_TGTS_ASC_ORDR.ST:s_TGTS_ASC_ORDR]  
$$platform=unix  
$DBConnection_ora=Ora2  
$ParamAscOrderOverride=UPDATE T_SALES SET CUST_NAME = :TU.CUST_NAME, DATE_SHIPPED = :TU.DATE_SHIPPED,  
TOTAL_SALES = :TU.TOTAL_SALES WHERE CUST_ID = :TU.CUST_ID  
[ORDERS.WF:wf_PARAM_FILE.WT:WL_PARAM_Lvl_1]  
$$DT_WL_lvl_1=02/01/2005 01:05:11  
$$Double_WL_lvl_1=2.2  
[ORDERS.WF:wf_PARAM_FILE.WT:WL_PARAM_Lvl_1.WT:NWL_PARAM_Lvl_2]  
$$DT_WL_lvl_2=03/01/2005 01:01:01  
$$Int_WL_lvl_2=3  
$$String_WL_lvl_2=cccc
```

配置参数文件名称和位置

启动工作流或会话时，请使用参数文件向集成服务传递参数和变量值。可以在工作流或会话属性中或通过 `pmcmd` 命令行指定参数文件名称和目录。如果集成服务使用操作系统配置文件，则在操作系统配置文件中指定的操作系统用户必须有权访问参数文件。

与 `pmcmd` 结合使用的参数文件会替代工作流或会话属性中的参数文件。如果在 `pmcmd` 命令行未输入参数文件名称，则集成服务会将在工作流属性中指定的参数文件用于工作流和工作流中的所有会话。如果在 `pmcmd` 命令行或工作流属性中均未输入参数文件名称，则集成服务将使用在会话属性中指定的参数文件。

将参数文件与工作流或会话结合使用

可以在工作流或会话属性中指定参数文件名称和目录。如果将工作流配置为并发运行，且需要将不同的参数文件用于各个工作流运行实例中的会话，请将工作流或工作集变量指定为会话参数文件名称。

如果为工作流或会话指定参数文件，但集成服务找不到该参数文件，则工作流或会话失败。

在工作流属性中输入参数文件

要在工作流属性中输入参数文件，请执行以下操作：

1. 在 Workflow Manager 中打开工作流。
2. 单击“工作流”>“编辑”。
3. 单击属性选项卡。
将显示“编辑工作流”对话框。
4. 在“参数文件名称”字段中输入参数文件位置和名称。
可以输入直接路径或服务进程变量。为集成服务操作系统使用恰当的分隔符。如果配置 PowerCenter 环境以实现高可用性，请将服务进程变量包括在该路径中
5. 单击“确定”。

在会话属性中输入参数文件

要在会话属性中输入参数文件，请执行以下操作：

1. 在 Workflow Manager 中打开会话。
将显示“编辑任务”对话框。
2. 单击“属性”选项卡，然后打开“常规选项”设置。
3. 在“参数文件名称”字段中输入参数文件位置和名称。
可以输入直接路径或服务进程变量。为集成服务操作系统使用恰当的分隔符。如果配置 PowerCenter 环境以实现高可用性，请将服务进程变量包括在该路径中。
还可以输入用户定义的工作流或工作集变量。输入工作流或工作集变量以在工作流参数文件中定义会话参数文件名称。
4. 单击“确定”。

使用变量指定会话参数文件

可以输入工作流或工作集变量作为会话参数文件名称。如果将工作流配置为并发运行，且需要为各个工作流运行实例中的会话定义不同的参数和变量值，请输入工作流或工作集变量作为会话参数文件名称。

为工作流内的会话定义工作流参数文件和会话参数文件时，集成服务将使用工作流参数文件并忽略会话参数文件。要使用变量定义会话参数文件名称，必须定义会话参数文件名称并在工作流参数文件中将 \$PMergeSessParamFile 设置为 TRUE。\$PMergeSessParamFile 属性会导致集成服务同时读取会话和工作流参数文件。

例如，将工作流配置为运行包含以下三个会话的两个并发实例：



对于第一个和第二个工作流实例，需要会话使用以下会话参数文件：

会话	会话参数文件名 (第一个工作流运行实例)	会话参数文件名 (第二个工作流运行实例)
s_1	s_1Inst1.txt	s_1Inst2.txt
s_2	s_2Inst1.txt	s_2Inst2.txt
s_3	s_3Inst1.txt	s_3Inst2.txt

创建工作流变量以存储会话参数文件名称。例如，您可以创建用户定义的工作流变量 `$$s_1ParamFileName`、`$$s_2ParamFileName` 和 `$$s_3ParamFileName`。在每个会话的会话属性中，将参数文件名称设置为工作流变量：

会话	会话参数文件名 在会话属性中
s_1	<code>\$\$s_1ParamFileName</code>
s_2	<code>\$\$s_2ParamFileName</code>
s_3	<code>\$\$s_3ParamFileName</code>

在每个工作流实例的工作流参数文件中，将每个工作流变量设置为正确的会话参数文件名称，然后将 `$PMergeSessParamFile` 设置为 `TRUE`。

如果将变量用作会话参数文件名称，且在会话和工作流参数文件中定义相同的参数或变量，则集成服务会根据以下规则设置参数和变量值：

- 在工作流和会话参数文件的相同部分定义参数或变量时，集成服务会使用工作流参数文件中的值。
- 在会话参数文件的会话部分和工作流参数文件的工作流部分定义参数或变量时，集成服务会使用会话参数文件中的值。

将参数文件与 pmcmd 结合使用

可以将参数文件与 `pmcmd startworkflow` 或 `starttask` 命令结合使用。通过这些命令，您可以指定在启动工作流或会话时要使用的参数文件。

`pmcmd -paramfile` 选项用于定义在运行会话或工作流时要使用的参数文件。`-localparamfile` 选项用于定义在无权访问集成服务计算机中的参数文件时可以在本地计算机上引用的参数文件。

以下命令使用参数文件 `myfile.txt` 启动 `workflowA`：

```
pmcmd startworkflow -uv USERNAME -pv PASSWORD -s SALES:6258 -f east -w wSalesAvg -paramfile '\$PMRootDir/myfile.txt' workflowA
```

以下命令使用参数文件 `myfile.txt` 启动 `taskA`：

```
pmcmd starttask -uv USERNAME -pv PASSWORD -s SALES:6258 -f east -w wSalesAvg -paramfile '\$PMRootDir/myfile.txt' taskA
```

参数文件示例

本节中的示例介绍了可能需要使用参数文件的一个会话。可以使用不同的状态和时间值重新运行该会话。该示例会显示可能需要设置的参数和变量、列出参数和变量值以及解释在重新运行会话时要更改的内容。

生产文件夹中有一个名为 s_MonthlyCalculations 的会话。会话使用会话参数连接到源文件和目标数据库并写入会话日志文件。如果会话失败，则集成服务会向 pccadmin@mail.com 发送一封电子邮件消息。会话会使用设置为“MA”的字符串映射参数 \$\$State 和日期时间映射变量 \$\$Time。\$\$Time 在存储库中的初始值为“9/30/2005 05:04:00”，但可以将此值替代为“10/1/2005 05:04:11”。

下表介绍了 s_MonthlyCalculations 会话的参数和变量：

参数或变量类型	参数或变量名称	定义
服务变量	\$PMFailureEmailUser	pccadmin@mail.com
字符串映射参数	\$\$State	MA
日期时间映射变量	\$\$Time	10/1/2005 05:04:11
源文件（会话参数）	\$InputFile1	Sales.txt
数据库连接（会话参数）	\$DBConnection_Target	销售
会话日志文件（会话参数）	\$PMSessionLogFile	d:/session logs/firstrun.txt

会话的参数文件包括文件夹和会话名称以及每个参数和变量：

```
[Production.s_MonthlyCalculations]
$PMFailureEmailUser=pccadmin@mail.com
$$State=MA
$$Time=10/1/2005 05:04:11
$InputFile1=sales.txt
$DBConnection_target=sales
$PMSessionLogFile=D:/session logs/firstrun.txt
```

下次运行会话时，可能需要编辑参数文件以将状态更改为 MD，并删除 \$\$Time 变量。这样，集成服务可以使用存储在存储库中的上一会话的变量的值。

创建参数文件的准则

创建参数文件时，请遵守以下规则和准则：

- **列出所有会话参数。**会话参数没有默认值。如果集成服务找不到会话参数值，则会话可能会失败并将空字符串视为默认值，或无法在运行时展开该参数。会话参数名称不区分大小写。
- **列出所有必需的映射参数和变量。**映射参数和变量值会变为映射中的参数和变量的起始值。映射参数和变量名称不区分大小写。
- **为非唯一会话名称输入文件夹名称。**如果存储库中多次存在一个会话名称，请输入文件夹名称以指示会话的位置。
- **在 Mapplet 中的参数和变量前面附加 Mapplet 名称。**使用以下格式：

```
mapplet_name.parameter_name=value
mapplet2_name.variable_name=value
```

- **使用多个参数文件。**可以单独向工作流、工作集和会话分配参数文件。可以为上述所有任务指定相同的参数文件或创建多个参数文件。
- **定义参数值时，请勿使用不必要的换行或空格。**集成服务会将其余空格解释为参数名称或值的一部分。
- **请为日期时间值使用正确的日期格式。**将以下日期格式用于日期时间值：
 - MM/DD/RR
 - MM/DD/YYYY
 - MM/DD/RR HH24:MI
 - MM/DD/YYYY HH24:MI
 - MM/DD/RR HH24:MI:SS
 - MM/DD/YYYY HH24:MI:SS
 - MM/DD/RR HH24:MI:SS.MS
 - MM/DD/YYYY HH24:MI:SS.MS
 - MM/DD/RR HH24:MI:SS.US
 - MM/DD/YYYY HH24:MI:SS.US
 - MM/DD/RR HH24:MI:SS.NS
 - MM/DD/YYYY HH24:MI:SS.NS

可以使用以下分隔符：短划线 (-)、斜杠 (/)、反斜杠 (\)、冒号 (:)、句点 (.) 和空格。集成服务会忽略多余空格。对于年份，不能使用一位数或三位数值；对于小时，不能使用“HH12”格式。
- **请勿将参数或变量值括在引号中。**集成服务会将第一个等号之后的所有字符均解释为值的一部分。
- **将具有正确长度的参数或变量值用于错误日志表名称前缀。**如果将参数或变量用于错误日志表名称前缀，则在命名 Oracle、Sybase 或 Teradata 错误日志表时，请勿指定长度超过 19 个字符的前缀。错误表名称最多可以有 11 个字符，且 Oracle、Sybase 和 Teradata 数据库中的表名称的最大长度为 30 个字符。参数或变量名称可以超过 19 个字符。

参数和参数文件故障排除

在参数文件中包含会话部分，但集成服务似乎无法读取该部分。

确保输入显示在 Workflow Manager 中的文件夹和会话名称。同时，对所有用户定义的会话参数使用相应的前缀。

我尝试使用源文件参数指定源文件和位置，但集成服务无法找到源文件。

确保清除会话属性中的源文件目录。集成服务会将源文件目录与源文件名称进行连接以查找源文件。

同时，确保输入集成服务的本地目录，并对操作系统使用相应的分隔符。

我尝试使用参数文件运行工作流，但其中一个会话一直失败。

会话可能包含参数文件中未列出的参数。集成服务会使用参数文件启动工作流中的所有会话。检查会话属性，然后验证在参数文件中是否正确定义所有会话参数。

我运行使用参数文件的工作流或会话，但执行失败。在恢复运行期间集成服务应使用哪些参数和变量值？

对于服务变量、服务进程变量、会话参数和映射参数，集成服务会使用参数文件中指定的值（如果存在）。如果参数文件中未指定值，集成服务会使用存储在恢复存储文件中的值。对于工作流、工作集和映射变量，集成服务会始终使用存储在恢复存储文件中的值。

参数和参数文件的提示

使用单个参数文件对相关会话的参数信息进行分组。

当会话可能要使用相同的数据库连接或目录时，可能需要将其包括在同一个参数文件中。连接或目录发生更改时，可以通过编辑一个参数文件更新所有会话的信息。

定期将 *pmcmd* 和多个参数文件用于会话。

有时可能会循环重用会话参数。例如，您可能会每天都运行销售数据库的会话，但对于销售和市场数据库使用的同一个会话，则一周运行一次。可以为各个会话运行创建单独的参数文件。每次按周运行会话时，除更改会话属性中的参数文件之外，还可以使用 *pmcmd* 指定启动该会话时要使用的参数文件。

将拒绝文件和会话日志参数与目标文件或目标数据库连接参数结合使用。

将目标文件或目标数据库连接参数与会话结合使用时，可以使用拒绝文件参数随时跟踪拒绝文件。还可以使用会话日志参数将会话日志写入到目标计算机。

使用资源以验证会话是否在有权访问参数文件的节点上运行。

在 Administrator 工具中，您可以为每个有权访问参数文件的节点定义文件资源，并配置集成服务以检查资源。然后，编辑使用参数文件的会话并分配资源。运行工作流时，集成服务会在具有可用资源的节点上运行具有所需资源的会话。

可以在会话部分进行定义来替代会话的工作流变量的初始值。

如果工作流包含的分配任务可以更改工作流变量的值，则工作流中的下一个会话将使用变量的最新值作为该会话的初始值。要替代会话的初始值，请在参数文件的会话部分为变量定义新值。

可以使用其他参数和变量定义参数和变量。

例如，在参数文件中，可以使用服务进程变量定义会话参数 *\$PMSessionLogFile*，如下所示：

```
$PMSessionLogFile=$PMSessionLogDir/TestRun.txt
```

第 17 章

FastExport

本章包括以下主题：

- [使用 FastExport 概览, 216](#)
- [步骤 1.创建 FastExport 连接, 217](#)
- [步骤 2.更改读取器, 218](#)
- [步骤 3.更改源连接, 219](#)
- [步骤 4.替代控制文件（可选）, 219](#)
- [使用 FastExport 的规则和准则, 220](#)

使用 FastExport 概览

FastExport 是一个实用工具，它使用多个 Teradata 会话来快速从 Teradata 数据库导出大量数据。可以创建使用 FastExport 读取 Teradata 源的 PowerCenter 会话。

要使用 FastExport，使用 Teradata 源数据库创建一个映射。映射可以包括来自联接在单个源限定符转换中的同一 Teradata 源数据库的多个源定义。在会话中，使用 FastExport 读取器而不是关系读取器。使用 FastExport 连接到要在会话中导出的 Teradata 表。

FastExport 使用定义导出内容的控制文件。会话启动后，集成服务会从 FastExport 连接属性中创建控制文件。如果为 Teradata 表创建了 SQL 替代，则集成服务会使用 SQL 生成控制文件。通过在会话属性中定义控制文件，可以替代会话的控制文件。

集成服务会在会话日志中写入 FastExport 消息，并在 FastExport 日志中写入有关 FastExport 性能的信息。PowerCenter 将 FastExport 日志储存在由“临时文件名”会话属性定义的文件夹中。FastExport 日志的默认扩展名为 .log。

要在会话中使用 FastExport，请完成以下步骤：

1. 在 Workflow Manager 中创建 FastExport 连接并配置连接属性。
2. 打开会话，并将读取器从关系更改为 Teradata FastExport。
3. 更改连接类型，并为会话选择 FastExport 连接。
4. 或者，在文本编辑器中创建 FastExport 控制文件，并将其保存到存储库中。

步骤 1.创建 FastExport 连接

在 Workflow Manager 中创建 FastExport 连接。如果编辑 FastExport 连接，则使用该连接的所有会话都会使用更新的连接。

要创建 FastExport 连接，请执行以下操作：

1. 在 Workflow Manager 中单击“连接”>“应用程序”。
将显示“连接浏览器”对话框。
2. 单击“新建”。
3. 选择 Teradata FastExport 连接，然后单击“确定”。
4. 输入 FastExport 连接的名称。
5. 输入数据库用户名。
6. 输入数据库用户名的密码，或单击“在密码中使用参数”以使用数据库密码的会话参数 `$ParamName`。
如果启用“在密码中使用参数”，请在工作流或会话参数文件中定义密码，并使用 `pmpasswd` CRYPT_DATA 选项对其进行加密。
7. 选择 FastExport 用来读取 Teradata 源的代码页。
FastExport 使用 `fexpcodemapfile.dat` 文件将代码页名称映射到 FastExport 支持的 Teradata 字符集。验证该文件是否包含代码页以及在 Teradata 数据库上是否已启用分配的字符集。
8. 输入 FastExport 属性，然后单击“确定”。

下表介绍了可为 Teradata FastExport 连接配置的属性：

属性	默认值	说明
TDPID	不适用	Teradata 数据库 ID。
韧度	4	FastExport 尝试登录到 Teradata 数据库的小时数。当 FastExport 尝试登录但已在运行最大数量的 Teradata 会话时，FastExport 会等待由 SLEEP 选项定义的时间。在 SLEEP 时间后，FastExport 会重新尝试登录到 Teradata 数据库。 FastExport 会重复此过程，直到其为所需数量的会话登录或超过 TENACITY 小时时间段。
最大会话数量	1	每个 FastExport 作业的最大 FastExport 会话数量。最大会话数量必须介于 1 和您系统上访问模块进程 (AMP) 的总数之间。
休眠	6	FastExport 在重新尝试之前暂停的分钟数。FastExport 会尝试登录，直到登录成功或 Tenacity 小时数过去。
块大小	64000	用于导出数据的最大块大小。
数据加密	禁用	为 FastExport 启用数据加密。可以对版本 8 的 Teradata 客户端使用数据加密。
日志表名称	FE_<source_table_name>	重新启动日志表名称。FastExport 实用程序使用重新启动日志表中的信息来重新启动由于 Teradata 数据库或客户端系统故障而停止的作业。每个 FastExport 作业都应使用单独的对数表。如果指定不存在的表，则 FastExport 实用程序会创建该表并将其用作重新启动日志。 PowerCenter 不支持重新启动 FastExport，但如果您暂存输出，则可以手动重新启动 FastExport。

属性	默认值	说明
可执行文件名称	fexp	用于读取源数据的 Teradata 命令。使用默认值。
数据库名称	不适用	要连接的 Teradata 数据库的名称。集成服务会生成 SQL 语句，使用数据库名称作为表名称的前缀。

验证代码页映射文件

创建 FastExport 连接时，可以选择 FastExport 用来读取 Teradata 源的 PowerCenter 代码页。FastExport 使用 fexpcodepagemapfile.dat 文件将 PowerCenter 代码页名称映射到 FastExport 支持的 Teradata 字符集。例如，如果选择“MS Windows Latin 1 (ANSI), Latin 1 的超集”作为连接代码页，则 PowerCenter 会使用名为“MS1252”的代码页，而 Teradata 使用名为“Latin1252_0A”的字符集。

fexpcodepagemapfile.dat 文件会将最适合的 Teradata 字符集映射到 PowerCenter 代码页，如 *Teradata 国际字符集支持* 中所指定。必须在数据库上启用 Teradata 字符集才能使用这些字符集。默认情况下，仅可在 Teradata 数据库上启用以下字符集：

- ASCII
- EBCDIC
- UTF8
- UTF16

如果您在 FastExport 连接中选择的 PowerCenter 代码页在 fexpcodepagemapfile.dat 中不存在，或者在数据库上未启用分配的 Teradata 字符集，则集成服务将使会话失败。

fexpcodepagemapfile.dat 文件位于 <PowerCenter 安装目录>\server\bin 中。验证该文件是否包含您在 FastExport 连接中选择的 PowerCenter 代码页以及在 Teradata 数据库上是否已启用分配的字符集。可以使用文本编辑器为 Teradata 字符集分配其他 PowerCenter 代码页或修改现有映射。采用以下格式在一行中输入名称，从而为 Teradata 字符集分配 PowerCenter 代码页：

```
<PowerCenter_code_page> = <Teradata_character_set>.
```

例如，MS1252 = Latin1252_0A

编辑文件时，请遵循以下规则和准则：

- 要指定注释，行应以感叹号 (!) 开头。
- 行不能仅包含空格或制表符字符。
- 如果文件将 PowerCenter 代码页映射到多个 Teradata 字符集，则 FastExport 会使用文件中最后分配的字符集。

注意: 对于 Unicode 的 UTF-16 编码，Teradata 不区分 Big Endian 和 Lower Endian。如果处理 UTF-16 字符，请在 Workflow Manager 中创建 FastExport 连接时选择“Unicode 的 UTF-16 编码 (平台字节序)”。

步骤 2.更改读取器

Teradata 的默认读取器是关系。要使用 FastExport，请将读取器更改为 Teradata FastExport。

步骤 3.更改源连接

要在会话中使用 FastExport，请将 Teradata 源连接更改为 Teradata FastExport 连接。可以替代某些会话属性。

下表介绍了可为 FastExport 更改的会话属性：

属性	默认值	精度
已暂存	禁用	如果启用，FastExport 会将数据写入暂存文件。否则，FastExport 会将数据写入命名管道。
小数秒数精度	0	时间戳中小数点后小数秒数的精度。可以输入 0 到 6。例如，精度为 6 的时间戳为“hh:mi:ss.ss.ss”。小数秒数精度必须匹配 Teradata 数据库中的设置。
临时文件	\$PMTempDir\	PowerCenter 使用临时文件名称为日志文件、控制文件和暂存输出文件生成名称。输入文件的完整路径。
控制文件替代	空	控制文件文本。使用此属性可替代集成服务为会话创建的控制文件。

步骤 4.替代控制文件（可选）

默认情况下，集成服务根据使用 FastExport 运行会话时的会话和连接属性生成 FastExport 控制文件。集成服务将其生成的控制文件保存在临时文件目录中，并在您下次运行会话时将其覆盖。

您可以替代集成服务生成的控制文件。替代控制文件时，Workflow Designer 会将控制文件保存到存储库。集成服务会在您运行会话时使用保存的控制文件。

每个 FastExport 语句都必须满足下列条件：

- 在新行开始。
- 以句点 (.) 开头。
- 以分号 (;) 结束。

下表介绍了可与 PowerCenter 一起使用的控制文件语句：

控制文件语句	说明
.LOGTABLE utillog;	重新启动日志表名称。
LOGON tdpz/user,pswd;	数据库登录字符串，包括数据库、用户名和密码。
BEGIN EXPORT	第一个导出命令。
.SESSIONS 20;	Teradata 会话的数量。
.EXPORT OUTFILE ddname2;	导出的数据的目标文件。
SELECT EmpNo, Hours FROM charges	用于选择数据的 SQL 语句。

控制文件语句	说明
WHERE Proj_ID = 20	-
ORDER BY EmpNo ;	-
.END EXPORT ;	指示导出任务结束和启动导出进程。
LOGOFF ;	从数据库断开连接。

要替代控制文件，请执行以下操作：

1. 在文本编辑器中创建控制文件。
2. 将控制文件文本复制到剪贴板。
3. 将控制文件文本粘贴到控制文件替代字段。

Workflow Manager 不验证控制文件语法。Teradata 在您运行会话时验证控制文件语法。如果控制文件无效，会话将失败。

提示： 可以将控制文件更改为只读以将控制文件用于每个会话。集成服务不会覆盖只读文件。

使用 FastExport 的规则和准则

结合使用 FastExport 和 PowerCenter 时，请遵循以下规则和准则：

- 为 Teradata 使用 SQL 替代时，PowerCenter 使用该替代来创建 FastExport 控制文件。如果不使用 SQL 替代，则 PowerCenter 会根据源限定中的已连接端口生成一个控制文件。
- FastExport 在 UNIX MP-RAS 操作系统上支持 2 GB 的最大导出文件大小。其他操作系统文件没有文件大小限制。
- 无法连接已导出的数据文件。
- 如果使用会话前 SQL 命令和 FastExport，则会话将失败。

第 18 章

外部加载

本章包括以下主题：

- [外部加载概览, 221](#)
- [外部加载器行为, 222](#)
- [加载到 IBM DB2, 223](#)
- [加载到 Oracle, 228](#)
- [加载到 Sybase IQ, 230](#)
- [加载到 Teradata, 231](#)
- [在会话中配置外部加载, 238](#)
- [对外部加载进行故障排除, 240](#)

外部加载概览

可以将会话配置为使用 IBM DB2、Oracle、Sybase IQ 和 Teradata 外部加载器将会话目标文件加载到其各自数据库中。外部加载器可以通过直接从文件或管道中加载信息（而不是运行 SQL 命令以将相同数据插入到数据库）来提高会话性能。

在一个会话中使用多个外部加载器。例如，如果映射包含两个目标，则可以创建使用 Oracle 外部加载器连接和 Sybase IQ 外部加载器连接的会话。

开始之前

运行外部加载器之前，请完成以下任务：

- **禁用约束。**可以在执行加载之前禁用构建到接收数据的表中的约束。有关禁用约束的信息，请参阅数据库文档。
- **关闭或禁用数据库日志记录。**要保持高性能，可以增加提交间隔并关闭数据库日志记录。但是，要对失败的会话执行数据库恢复，必须打开数据库日志记录。
- **配置代码页。**IBM DB2、Oracle、Sybase IQ 和 Teradata 数据库服务器必须使用与目标平面文件代码页相同的代码页。集成服务使用目标平面文件代码页创建控制文件和目标平面文件。如果对目标平面文件使用除 7 位 ASCII 以外的代码页，请在 Unicode 数据移动模式下运行集成服务。
- **将外部加载器连接配置为资源。**如果将集成服务配置为在网络上运行，请在提供了外部加载器的节点上将外部加载器连接配置为资源。

外部加载器行为

运行使用外部加载器的会话时，集成服务会创建控制文件和目标平面文件。控制文件包含外部加载器的数据格式和加载说明等信息。控制文件具有 .ctl 扩展名。可以在目标文件目录中查看控制文件和目标平面文件。

运行会话时，集成服务会删除并重新创建目标文件。外部加载器使用控制文件将会话输出加载到数据库。集成服务会在使用以下方式加载到数据库之前处理日期时间数据：

- 如果将会话配置为截减子秒，集成服务会根据精度 19 处理日期时间数据。
- 如果未将会话配置为截减子秒，集成服务会根据在目标平面文件中指定的精度处理日期时间数据。精度范围为 19 到 29。子秒根据指定的精度进行截减。
- 如果在目标文件中指定的精度大于为数据库指定的精度，则集成服务会将精度限制为为数据库指定的最大精度。

集成服务等待所有外部加载完成后，才会执行会话后命令、运行外部过程以及发送会话后电子邮件。

集成服务会将外部加载器初始化和完成消息写入会话日志中。有关外部加载器性能的详细信息，请检查外部加载器日志。加载器将日志保存在目标平面文件所在的相同目录中。外部加载器日志的默认扩展名为 .ldrlog。

外部加载器的行为取决于您选择如何加载数据。可以将数据加载到命名管道或平面文件中。

将数据加载到命名管道

一旦数据显示在管道中，外部加载器便开始将数据加载到数据库中。加载器会在完成加载后删除命名管道。

在 UNIX 上，集成服务会写入在配置的目标文件名之后命名的命名管道中。

在 Windows 上，集成服务会将数据写入使用指定格式命名的管道中：

```
\\.\pipe\<pipe name>
```

管道名称与配置的目标文件名相同。

将数据暂存到平面文件

在 Windows 或 UNIX 上将数据暂存到平面文件中时，集成服务会将数据写入在配置的目标文件名之后命名的平面文件中。集成服务将所有数据写入目标平面文件之后，外部加载器便开始将数据加载到目标数据库中。外部加载器不会在将数据加载到数据库之后删除目标平面文件。确保目标文件目录可以容纳目标平面文件的大小。

注意：将数据暂存到平面文件中时，集成服务会根据端口的小数位舍入数值。当您使用外部加载器将数据加载到命名管道中时，或者如果将目标配置为正常加载，它将不会舍入结果。

如果在集成服务将所有数据写入平面文件目标之前会话中止或失败，则外部加载器不会启动。如果在集成服务将所有数据写入平面文件目标之后会话中止或失败，则外部加载器会在向目标数据库加载完数据后退出。

使用外部加载器对会话进行分区

使用平面文件目标在会话中配置多个分区时，集成服务会为每个分区创建一个单独的平面文件。某些外部加载器无法从多个文件中加载数据。在具有多个分区的会话中使用外部加载器时，必须根据使用的外部加载器配置目标分区类型。

使用可从多个文件中加载数据的外部加载器时，可以选择可用于平面文件目标的任何分区类型。还可以为每个分区选择一个外部加载器连接。集成服务会为每个分区创建一个输出文件，且外部加载器会将每个目标文件的输出加载到数据库中。使用以下加载器时，可对目标使用任何分区类型：

- Oracle（已启用并行加载）
- Teradata T pump

如果使用无法从多个文件加载的加载器，请使用循环分区将数据路由到单个目标文件。可以为每个分区选择一个外部加载器连接。但是，集成服务会对第一个分区使用加载器连接。集成服务会创建单个输出文件，且外部加载器会将目标文件的输出加载到数据库中。如果您为目标选择任何其他分区类型，集成服务将会使会话失败。使用以下加载器时，可对目标使用循环分区类型：

- IBM DB2 EE
- IBM DB2 EEE 自动加载器
- Oracle（已禁用并行加载）
- Sybase IQ
- Teradata MultiLoad
- Teradata Fastload

加载到 IBM DB2

加载到 IBM DB2 目标时，可使用 IBM DB2 EE 或 IBM DB2 EEE 外部加载器。这两个外部加载器都可以对目标执行插入和替换操作。它们还可以重新启动或终止加载操作。它们可以对数据进行分区，并将已分区的数据同时加载到相应数据库分区中。

IBM DB2 EE 外部加载器

使用 IBM DB2 EE 外部加载器加载到以下数据库之一：

- IBM DB2 EE 版本 8.x
- IBM DB2 EEE 版本 8.x
- IBM DB2 版本 9.x

IBM DB2 EE 外部加载器可调用位于集成服务安装目录中的以下可执行文件之一：

- **db2load**。用于版本 9.5 之前的 IBM DB2 客户端。
- **db2load95**。用于 IBM DB2 客户端版本 9.5 及更高版本。

创建外部加载器连接时，根据运行集成服务进程的计算机上安装的 IBM DB2 客户端版本指定可执行文件的名称。

IBM DB2 EE 外部加载器可以将数据加载到远程（相对于集成服务来说）计算机上的 IBM DB2 服务器中。

处理 LOB 数据

IBM DB2 EE 外部加载器无法加载 LOB 数据，如 Blob、Clob 或 Dbclob 数据。运行使用 IBM DB2 EE 外部加载器的会话且源包含 LOB 数据时，外部加载器会根据以下映射配置将剩余数据成功加载到目标：

- **LOB 端口未连接**。外部加载器将所有剩余数据成功加载到目标。
- **LOB 端口已连接**。加载到数据库版本 8.x 时，外部加载器将 LOB 数据加载为 NULL 并正确加载剩余数据。加载到数据库版本 9.x 时，外部加载器不会加载任何数据。它会将拒绝的行记录到外部加载器日志中。

IBM DB2 EEE 外部加载器

使用 IBM DB2 EEE 外部加载器可加载到 IBM DB2 EEE 版本 8.x 数据库中。IBM DB2 EEE 外部加载器调用 IBM DB2 自动加载器程序加载数据。自动加载器程序使用 db2atld 可执行文件。IBM DB2 EEE 加载器要求 IBM DB2 服务器位于托管集成服务的相同计算机上。

注意: 如果 IBM DB2 EEE 服务器位于远程（相对于集成服务来说）计算机上，请使用 IBM DB2 EE 外部加载器，或使用关系数据库连接连接到 IBM DB2 EEE 数据库。请对 IBM DB2 目标使用数据库分区。使用数据库分区时，集成服务会在 IBM DB2 系统中查询表分区信息，并将已分区的数据加载到目标数据库中的对应节点。

IBM DB2 EEE 外部加载器的规则和准则

使用外部加载器加载到 IBM DB2 时，请遵循以下规则和准则：

- IBM DB2 外部加载器从带分隔符的平面文件中进行加载。验证目标表列是否足够宽可以存储所有数据。
- 对于使用 IBM DB2 客户端身份验证的连接，请在创建外部加载器连接时输入 PmNullUser 用户名和 PmNullPasswd。当连接用户名为 PmNullUser 并且所连接的是 IBM DB2 数据库时，PowerCenter 使用 IBM DB2 客户端身份验证。
- 对于具有多个分区的会话，请使用循环分区类型将数据路由到单个目标文件。
- 如果将同一管道中的多个目标配置为使用 IBM DB2 外部加载器，则每个加载器必须加载到目标数据库上的不同表空间中。
- 必须具有正确的授权级别和特权，才能将数据加载到数据库表。

设置操作模式

IBM DB2 操作模式指定了外部加载器运行的加载类型。可以将 IBM DB2 EE 或 IBM DB2 EEE 外部加载器配置为在以下某种操作模式下运行：

- **Insert.** 将已加载的数据添加到表中，而不会更改现有表数据。
- **替换.** 删除表中的所有现有数据，并插入已加载的数据。表和索引定义不会更改。
- **Restart.** 重新启动先前中断的加载操作。
- **终止.** 终止先前中断的加载操作，并将操作回滚到起点，即使已经过一致点。表空间返回到正常状态，且外部加载器使所有表对象一致。

配置授权、特权和权限

IBM DB2 特权允许您创建或访问数据库资源。授权级别允许您对特权进行分组并执行更高级别的数据库管理器维护和实用程序操作。它们一起作用来控制对数据库管理器及其数据库对象的访问权限。您可以访问对其具有所需特权或授权的对象。

要将数据加载到表中，必须具有以下授权之一：

- SYSADM 授权
- DBADM 授权
- 对数据库的 LOAD 授权及以下特权之一：
 - 在插入、终止或重新启动模式下调用加载实用程序时，对表的 INSERT 特权。
 - 在替换、终止或重新启动模式下调用加载实用程序时，对表的 INSERT 和 DELETE 特权。

此外，还必须具有适当的读取访问权限和读/写权限：

- 数据库实例所有者必须具有对外部加载器输入文件的读取访问权限。
- 如果在 Windows 上将 IBM DB2 作为服务运行，则必须使用具有读/写权限的用户帐户配置服务启动帐户，才能使用 LAN 资源（包括驱动器、目录和文件）。
- 如果加载到 IBM DB2 EEE，则数据库实例所有者必须具有对加载转储文件和加载临时文件的写入访问权限。

配置 IBM DB2 EE 外部加载器属性

IBM DB2 EE 外部加载器根据要加载的以下数据库创建单个或多个日志：

- **IBM DB2 EE 版本 8.x 或未分区的 IBM DB2 版本 9.x。**外部加载器在目标平面文件所在的相同目录中使用扩展名 `.ldrlog` 创建单个外部加载器日志。
- **IBM DB2 EEE 版本 8.x 或已分区的 IBM DB2 版本 9.x。**外部加载器在目标平面文件所在的相同目录中创建多个外部加载器日志。加载器日志具有以下扩展名：
 - `ldrlog.load.编号`。由 Load Agent 外部加载器进程创建。Load Agent 会创建两个日志文件。
 - `ldrlog.part.partition_number`。由 Partitioning Agent 外部加载器进程创建。Partitioning Agent 可根据目标表中分区数量创建多个日志文件。
 - `ldrlog.prep.partition_number`。由 Pre-partitioning Agent 外部加载器进程创建。Pre-partitioning Agent 可根据目标表中分区数量创建多个日志文件。
 - `ldrlog`。由 IBM DB2 EE 外部加载器创建。

下表介绍了 IBM DB2 EE 外部加载器连接的属性：

属性	默认值	说明
操作模式	插入	IBM DB2 外部加载器操作模式。选择以下操作模式之一： <ul style="list-style-type: none">- 插入- 替换- 重新启动- 终止
外部加载器可执行文件	db2load	IBM DB2 EE 外部加载器可执行文件的名称。根据运行集成服务进程的计算机上安装的 IBM DB2 客户端版本输入以下文件名之一： <ul style="list-style-type: none">- <code>db2load</code>。用于版本 9.5 之前的 IBM DB2 客户端。- <code>db2load95</code>。用于 IBM DB2 客户端版本 9.5 及更高版本。
DB2 服务器位置	远程	IBM DB2 数据库服务器相对于集成服务的位置。如果数据库服务器位于托管集成服务的计算机上，请选择“本地”。如果数据库服务器位于其他计算机上，请选择“远程”。
已暂存	禁用	加载数据的方法。选择“已暂存”可在将数据加载到数据库之前将其加载到平面文件暂存区域中。默认情况下，会使用命名管道将数据加载到数据库中。
可恢复	启用	如果已启用向前恢复，请在备份挂起状态下设置表空间。如果禁用向前恢复，IBM DB2 表空间将不会设置为备份挂起状态。如果 IBM DB2 表空间处于备份挂起状态，则必须在对表空间执行任何其他操作之前完全备份数据库。

使用 IBM DB2 EE 外部加载器加载空格

如果您需要通过 IBM DB2 EE 外部加载器加载空格，则必须配置会话。在暂存模式下，将平面文件配置为使用可选的双引号。在非暂存模式下，将以下行添加到控制文件：

MODIFIEDBY = keepblanks

将控制文件配置为只读。

IBM DB2 EE 外部加载器返回代码

IBM DB2 EE 外部加载器通过返回代码指示加载操作是成功还是失败。集成服务会将外部加载器返回代码写入会话日志中。返回代码 (0) 表示加载操作成功。如果外部加载器成功完成加载操作，集成服务会将以下消息写入会话日志中：

WRT_8029 External loader process <external loader name> exited successfully.

任何其他返回代码均指示加载操作失败。集成服务会将以下错误消息写入会话日志中：

WRT_8047 Error: External loader process <external loader name> exited with error <return code>.

下表介绍了 IBM DB2 EE 外部加载器的返回代码：

代码	说明
0	外部加载器操作已成功完成。
1	外部加载器找不到控制文件。
2	外部加载器无法打开外部加载器日志文件。
3	外部加载器无法访问控制文件，因为控制文件已被另一进程锁定。
4	IBM DB2 数据库返回错误。

配置 IBM DB2 EEE 外部加载器属性

可以将 IBM DB2 EEE 外部加载器配置为在加载到数据库时使用不同的加载模式。加载模式可确定 IBM DB2 EEE 外部加载器如何在数据库中的分区之间加载数据。可以将 IBM DB2 EEE 外部加载器配置为使用以下加载模式：

- **拆分并加载。**对数据进行分区，并使用相应的数据库分区同时加载数据。
- **仅拆分。**对数据进行分区，并将输出写入指定拆分文件目录中的文件内。
- **仅加载。**不对数据进行分区。它使用相应的数据库分区加载现有拆分文件中的数据。
- **分析。**通过所有数据库分区之间的平均分布生成最佳分区映射。如果在分析模式下运行外部加载器之后再拆分并加载模式下运行外部加载器，外部加载器将使用最佳分区映射对数据进行分区。

IBM DB2 EEE 外部加载器会根据其加载到的数据库分区数量创建多个日志。对于每个分区，外部加载器会将分区编号对应的数字附加到外部加载器日志文件名。IBM DB2 EEE 外部加载器日志文件格式为 *file_name.ldrlog.partition_number*。

集成服务不会存档或覆盖 IBM DB2 EEE 外部加载器日志。如果在外部加载器运行时存在相同名称的外部加载器日志，则外部加载器会将新的外部加载器日志消息附加到现有外部加载器日志文件的末尾。必须手动存档或删除外部加载器日志文件。

有关 IBM DB2 EEE 外部加载器返回代码的信息，请参阅 IBM DB2 文档。

下表介绍了 IBM DB2 EEE 外部加载器连接的属性：

属性	默认值	说明
操作模式	插入	IBM DB2 外部加载器操作模式。选择以下操作模式之一： <ul style="list-style-type: none"> - 插入 - 替换 - 重新启动 - 终止
外部加载器可执行文件	db2atld	IBM DB2 EEE 外部加载器可执行文件的名称。
拆分文件位置	不适用	拆分文件的位置。如果配置 SPLIT_ONLY 加载模式，则外部加载器会创建拆分文件。
输出节点	不适用	要在其上执行加载操作的数据库分区。
拆分节点	不适用	用于确定如何拆分数据的数据库分区。如果不指定此属性，则外部加载器会确定最佳拆分方法。
模式	拆分并加载	外部加载器加载数据所使用的加载模式。选择以下加载模式之一： <ul style="list-style-type: none"> - 拆分并加载 - 仅拆分 - 仅加载 - 分析
拆分器最大数量	25	拆分器进程的最大数量。
强制	否	强制外部加载器操作继续，即使外部加载器断定（在启动时）有些目标分区或表空间脱机。
状态时间间隔	100	在将进度消息写入外部加载器日志之前外部加载器加载的数据量 (MB)。指定介于 1 和 4,000 MB 之间的值。
端口	6000-6063	外部加载器为与 IBM DB2 服务器的内部通信创建套接字时使用的 TCP 端口范围。
检查级别	Nocheck	在输入或输出期间检查记录截断。
映射文件输入	不适用	指定分区映射的文件的名称。要使用自定义的分区映射，请指定此属性。在“分析”加载模式下运行外部加载器时生成自定义分区映射。
映射文件输出	不适用	在“分析”加载模式下运行外部加载器时分区映射的名称。如果要在“分析”加载模式下运行外部加载器，则必须指定此属性。
跟踪	0	在需要查看数据转换进程的转储和哈希值的输出时外部加载器跟踪的行数。

属性	默认值	说明
已暂存	禁用	加载数据的方法。选择“已暂存”可在将数据加载到数据库之前将其加载到平面文件暂存区域中。否则，会使用命名管道将数据加载到数据库中。
日期格式	mm/dd/yyyy	日期格式。必须与您在目标定义中定义的日期格式相匹配。IBM DB2 支持以下日期格式： <ul style="list-style-type: none"> - MM/DD/YYYY - YYYY-MM-DD - DD.MM.YYYY - YYYY-MM-DD

加载到 Oracle

加载到 Oracle 目标时，可使用 Oracle SQL 加载器对目标执行插入、更新和删除操作。

Oracle 外部加载器会为数据库拒绝的数据创建一个拒绝文件。拒绝文件具有 .ldrreject 扩展名。加载器将拒绝文件保存在目标文件目录中。

Oracle 外部加载器的规则和准则

使用外部加载器加载到 Oracle 时，请遵循以下规则和准则：

- 如果选择 Oracle 外部加载器，则默认的外部加载器可执行文件名为 sqlload。这对于大多数 UNIX 平台来说很准确，但如果使用 Windows，请查看 Oracle 文档以查找外部加载器可执行文件的名称。
- 对于使用 Oracle 操作系统身份验证的连接，请在创建外部加载器连接时输入 PmNullUser 用户名和 PmNullPasswd。当连接用户名为 PmNullUser 并且所连接的是 Oracle 数据库时，PowerCenter 使用 Oracle 操作系统身份验证。
- Oracle 外部加载器的目标平面文件可以是固定宽度，也可以带分隔符。
- 要在写入已分区目标时获得最佳性能，请选择“直接路径”。有关详细信息，请参阅 Oracle 文档。
- 如果配置会话以将子秒数据写入 Oracle 10.x 或 Oracle 11.x 目标中的“时间戳”列，集成服务会在默认情况下写入高达微秒的子秒数据。为确保获得更高的精度，请编辑控制文件并更改时间戳精度。例如，指定 TIMESTAMP(9) 以处理纳秒。
- 为获得最佳性能，请使用以下准则确定已分区和未分区目标的设置：

目标	加载方法	并行加载	加载模式
已分区	直接路径	启用	附加
已分区	常规路径	启用	不适用
未分区	不适用	禁用*	不适用

* 如果禁用并行加载，则必须选择循环分区以将数据路由到单个目标文件。

将多字节数据加载到 Oracle

将多字节数据加载到 Oracle 时，数据精度按字节数（对于固定宽度的文件）和字符数（对于带分隔符的文件）进行衡量。确保目标表列足够宽可以存储所有数据。

Oracle 支持面向字符的数据类型（如 Nchar），其精度按字符数进行衡量。如果使用 Nchar 数据类型，请将最大字符数乘以 K，其中 K 是选定目标代码页中字符包含的最大字节数。这可确保集成服务不会在加载目标文件之前截断数据。

配置 Oracle 外部加载器属性

下表介绍了 Oracle 外部加载器连接的属性：

属性	默认值	说明
错误限制	1	在外部加载器停止加载操作之前允许的错误的数。
加载模式	附加	外部加载器加载数据所使用的加载模式。选择以下加载模式之一： <ul style="list-style-type: none">- 附加- 插入- 替换- 截断
加载方法	使用常规路径	外部加载器加载数据所使用的方法。选择以下加载方法之一： <ul style="list-style-type: none">- 使用常规路径。- 使用直接路径（可恢复）。- 使用直接路径（不可恢复）。
启用并行加载	启用并行加载	确定 Oracle 外部加载器是否将数据并行加载到已分区的 Oracle 目标表。 <ul style="list-style-type: none">- 启用并行加载以加载到已分区的目标。- 不启用并行加载以加载到未分区的目标。
每次提交的行数	10000	对于“常规路径”加载方法，此属性可指定加载操作的绑定数组中的行数。对于“直接路径”加载方法，此属性可指定外部加载器在将数据保存到数据库之前从目标平面文件中读取的行数。
外部加载器可执行文件	sqlload	外部加载器可执行文件的名称。
日志文件名	不适用	外部加载器日志文件的路径和名称。
已暂存	禁用	加载数据的方法。选择“已暂存”可在将数据加载到数据库之前将其加载到平面文件暂存区域中。否则，会使用命名管道将数据加载到数据库中。

加载到 Sybase IQ

加载到 Sybase IQ 时，可使用 Sybase IQ 外部加载器执行插入操作。集成服务可将多字节数据加载到 Sybase IQ 目标。当 Sybase IQ 服务器与集成服务位于相同计算机或不同计算机上时，集成服务都可写入平面文件。如果集成服务是 Sybase IQ 数据库服务器的本地服务，则它可写入命名管道。

Sybase IQ 外部加载器的规则和准则

使用外部加载器加载到 Sybase IQ 时，请遵循以下规则和准则：

- 确保目标表未违反主键约束。
- 在使用 Sybase IQ 外部加载器之前，通过读/写访问权限配置 Sybase IQ 用户。
- Sybase IQ 外部加载器的目标平面文件可以是固定宽度，也可以带分隔符。
- Sybase IQ 外部加载器无法对目标执行更新或删除操作。
- 对于具有多个分区的会话，请使用循环分区类型将数据路由到单个目标文件。
- 如果集成服务和 Sybase IQ 服务器位于不同计算机上，请将托管集成服务的计算机中的驱动器映射或装载到托管 Sybase IQ 服务器的计算机中。

将多字节数据加载到 Sybase IQ

将多字节数据加载到 Sybase IQ 目标时，请使用以下准则。

带分隔符的平面文件目标

对于带分隔符的平面文件，数据精度按字符数进行衡量。在目标中插入多字节字符数据时，无需考虑多字节数据的其他精度。Sybase IQ 不允许使用可选引号。如果您具有带分隔符的目标平面文件，则必须为“可选引号”选择“无”。

将多字节数据加载到 Sybase IQ 时，空字符数和分隔符数最多分别为四个字节。为了避免将分隔符读取为正常字符，分隔符的每个字节的 ASCII 值必须小于 0x40。

固定宽度的平面文件目标

对于固定宽度的平面文件，数据精度按字节数（而非字符数）进行衡量。将多字节数据加载到固定宽度的平面文件目标时，配置精度以容纳多字节数据。如果精度的高低不足以容纳多字节数据，则集成服务会将行写入到拒绝文件中。

配置 Sybase IQ 外部加载器属性

可以在 PowerCenter 中对 Sybase IQ 使用外部加载器连接类型。使用连接字符串属性提供 Sybase IQ 数据库登录凭据。

Sybase IQ 15.x 的连接字符串必须包含以下属性：

```
uid=user ID; pwd=password; eng=Sybase IQ database server name
```

例如，您可以使用以下连接字符串：

```
uid=qasrvr65;pwd=qasrvr65;eng=SUNQA2SybaseIQ
```

注意：如果在连接字符串中使用引号，会话可能会失败。

下表介绍了 Sybase IQ 外部加载器连接的属性：

属性	默认值	说明
块因子	10000	目标 Sybase 表中每个块的记录数。外部加载器仅将“块因子”属性应用于最高至 Sybase IQ 15.x 版本的固定宽度平面文件目标的加载操作。
块大小	50000	Sybase 数据库操作中所用块的大小。外部加载器仅将“块大小”属性应用于最高至 Sybase IQ 15.x 版本的带分隔符平面文件目标的加载操作。
检查点	已启用	如果已启用，Sybase IQ 数据库会在成功加载表后发出检查点。如果已禁用，数据库不会发出任何检查点。
通知间隔	1000	Sybase IQ 外部加载器在将状态消息写入外部加载器日志之前加载的行数。
数据文件目录	不适用	可从运行集成服务的计算机访问 Sybase IQ 数据文件的目录。如果此目录在 Windows 系统上，请在目录路径中使用反斜杠 (\)：D:\mydirectory\inputfile.out 如果此目录在 UNIX 系统上，请在目录路径中使用正斜杠 (/)：/mydirectory/inputfile.out 输入运行集成服务的计算机的目录路径。
外部加载器可执行文件	对于 Sybase 15.x: dbisql -host<主机名> -port<端口号>	Sybase IQ 外部加载器可执行文件的名称。在创建 Sybase IQ 外部加载器连接时，Workflow Manager 会将外部加载器可执行文件的名称默认设置为 dbisql。如果使用具有不同名称的可执行文件，则必须更新“外部加载器可执行文件”字段。如果外部加载器可执行文件目录不位于系统路径中，则必须在此字段中输入文件路径和文件名。
已暂存	已启用	加载数据的方法。选择“已暂存”可在将数据加载到数据库之前将其加载到平面文件暂存区域中。清除此属性以从命名管道中加载数据。如果集成服务是 Sybase IQ 数据库的本地服务，则它可写入命名管道。

加载到 Teradata

加载到 Teradata 目标时，可使用以下某种外部加载器：

- **Multiload.**对大量增量加载执行插入、更新、删除和更新插入操作。运行具有单个分区的会话时，请使用此加载器。Multiload 可获取表级锁定，从而使其适用于脱机加载。
- **TPump.**对相对较少的更新执行插入、更新、删除和更新插入操作。运行具有多个分区的会话时，请使用此加载器。TPump 可获取对表的行哈希锁定，以便在 TPump 加载到表时允许其他用户访问此表。
- **FastLoad.**对大量初始加载或大量截断和重新加载操作执行插入操作。运行具有单个分区的会话时，请使用此加载器。请对不具有二级索引的空表使用此加载器。

如果使用 Teradata 外部加载器执行更新或更新插入操作，请在 Mapping Designer 中使用“目标更新替代”选项替代外部加载器控制文件中的 UPDATE 语句。对于更新插入，外部加载器控制文件中的 INSERT 语句保持不变。

Teradata 外部加载器的规则和准则

使用外部加载器加载到 Teradata 时，请遵循以下规则和准则：

- 集成服务可以使用 Teradata 外部加载器将固定宽度及带分隔符的平面文件加载到 Teradata 数据库。由于所有 Teradata 加载器都使用换行符 (\n) 分隔各个记录，因此不能使用换行符作为 Teradata 加载器的分隔符。
- 如果会话包含一个分区，则目标输出文件名（包括文件扩展名）不能超过 27 个字符。如果会话包含多个分区，则目标输出文件名（包括文件扩展名）不能超过 25 个字符。
- 不能使用 Teradata 外部加载器加载二进制数据。
- 使用命名管道加载到 Teradata 时，将检查点值设置为 0 以防止外部加载器执行检查点操作。
- 可以根据所用加载器指定错误、日志或工作表名称。还可以指定错误、日志或工作数据库名称。
- 可以在会话属性中替代控制文件。
- 使用 Teradata 时，可以输入 PmNullPasswd 作为数据库密码以防止密码显示在控制文件中。集成服务将在控制文件中写入空白字符串代替密码。

替代控制文件

在会话中编辑加载器连接时，可以替代控制文件。您可能希望替代控制文件，以更改一些无法在加载器连接中编辑的加载器属性。例如，可以在控制文件中指定跟踪选项。

替代控制文件时，Workflow Manager 会将控制文件保存到存储库中。当运行会话且在以后每次运行会话时，集成服务都会使用已保存的控制文件，直到您清除控制文件属性。如果在编辑控制文件后更改目标或加载器连接设置，则控制文件不包含这些更改。要包含这些更改，必须重新生成控制文件并对其进行编辑。

如果不替代控制文件，集成服务会在您每次运行会话时根据会话和加载器属性生成新的控制文件。集成服务会在输出文件目录中生成控制文件。每当您运行会话时，它都会进行覆盖。

注意：Workflow Manager 不验证控制文件语法。Teradata 在您运行会话时验证控制文件语法。如果控制文件无效，会话将失败。

可以通过打开控制文件编辑器来查看已编辑的控制文件。

要替代控制文件，请执行以下操作：

1. 在 Workflow Manager 中，打开会话属性。
2. 单击“映射”选项卡并打开“转换”视图。
3. 单击“目标”节点。
4. 在“连接”设置的“值”字段中，单击“更改”。
5. 在“控制文件内容替代”字段中，单击“打开”。
此时将显示“控制文件编辑器”对话框。
6. 单击“生成”。

Workflow Manager 会根据会话和加载器属性生成控制文件。

7. 编辑已生成的控制文件，然后单击“确定”以保存更改。

在控制文件中创建用户变量

配置 MultiLoad 或 TPump 外部加载器属性时，可以创建用户变量。用户变量是您在控制文件中使用的自定义置换变量。用户变量可捕获可能在连接对象属性中不可用的会话特定信息。用户变量通常用于加载前或加载后处理。

您可以在连接对象中定义用户变量名称和置换值。在控制文件中，可以将置换变量前缀和用户变量名称添加到相应命令。运行会话时，集成服务会使用置换值替换控制文件中的置换变量前缀和用户变量名称。如果在编辑控制文件后更改置换值，则控制文件将使用新值。

创建用户变量时，请遵循以下规则和准则：

- 创建用户变量时，请使用以下语法：
`<User_Variable_Name>=<Substitution_Value>`
- 如果在用户变量名称或置换值中包含空格，会话可能会失败。
- 将用户变量添加到控制文件中时，请使用以下语法：
`:CF.<User_Variable_Name>`

示例

集成服务将数据加载到目标后，您希望在输出文件中显示系统日期。在连接对象中，您配置了以下用户变量：

```
OutputFileName=output_file.txt
```

在控制文件中，您配置了以下内容：

```
DISPLAY '&SYSDATE' TO FILE ':CF.OutputFileName'
```

运行会话时，集成服务会将控制文件中的 `:CF.OutputFileName` 替换成 `output_file.txt`。

配置 Teradata MultiLoad 外部加载器属性

使用 MultiLoad 外部加载器时，请遵循以下规则和准则：

- 可以对目标执行插入、更新、删除和更新插入操作。还可以使用数据驱动模式根据更新策略转换或自定义转换执行插入、更新或删除操作。
- 对于具有多个分区的会话，请使用循环分区类型将数据路由到单个目标文件。
- 如果要调用的会话数比数据库允许的最大并发会话数多，会话可能会挂起。您可以设置“韧度”和“休眠”的最小值，以确保会话失败而不是挂起。

要配置 Teradata MultiLoad 外部加载器的属性，请单击“连接”>“加载器”，选择“类型”，然后单击“编辑”。

下表显示了为 Teradata MultiLoad 外部加载器配置的属性：

属性	默认值	说明
TDPID	不适用	Teradata 数据库 ID。
数据库名称	不适用	可选数据库名称。如果不指定数据库名称，则集成服务会使用在映射中定义的目标表数据库名称。
日期格式	不适用	日期格式。连接对象中的日期格式必须与您在目标定义中定义的日期格式相匹配。集成服务支持以下日期格式： <ul style="list-style-type: none">- DD/MM/YYYY- MM/DD/YYYY- YYYY/DD/MM- YYYY/MM/DD
错误限制	0	MultiLoad 可写入 MultiLoad 错误表的已拒绝记录总数。唯一性违反并不算作已拒绝记录。 错误限制为 0 表示对已拒绝记录数没有限制。

属性	默认值	说明
检查点	10,000	检查点之间的时间间隔。可以将此时间间隔设置为以下值： <ul style="list-style-type: none"> - 60 或更大。MultiLoad 会在处理此记录数的每个倍数之后执行检查点操作。 - 1-59. MultiLoad 以指定的时间间隔（以分钟为单位）指定检查点操作。 - 0. MultiLoad 不会在导入任务期间执行任何检查点操作。
韧度	10,000	MultiLoad 尝试登录到所需会话的时间量（以小时为单位）。如果登录失败，MultiLoad 会延迟在“休眠”属性中指定的分钟数，然后重试登录。MultiLoad 将一直尝试，直到登录成功或超过在“韧度”属性中指定的小时数。
加载模式	更新插入	用于生成 SQL 命令的模式：插入、删除、更新、更新插入或数据驱动。选择“数据驱动”加载时，集成服务会按照更新策略转换或自定义转换中的说明来确定如何标记要插入、删除或更新的行。集成服务会在目标文件或命名管道中写入一行，以指示更新策略。控制文件使用这些值来确定如何将数据加载到目标。集成服务使用以下值指示更新策略： <ul style="list-style-type: none"> 0 - 插入 1 - 更新 2 - 删除
删除错误表	已启用	在开始下次会话之前删除 MultiLoad 错误表。选择此选项可删除表或清除表以将其保留。
外部加载器可执行文件	mload	Teradata 外部加载器可执行文件的名称和可选文件路径。如果外部加载器可执行文件目录不位于系统路径中，则必须输入完整路径。
最大会话数量	1	每个 MultiLoad 作业的最大 MultiLoad 会话数。“最大会话数量”必须介于 1 和 32,767 之间。运行多个 MultiLoad 会话会导致客户端和数据库使用更多资源。因此，将此值设置为较小的数字可提高性能。
休眠	6	MultiLoad 在重试登录之前等待的分钟数。MultiLoad 将一直尝试，直到登录成功或超过在“韧度”属性中指定的小时数。 “休眠”必须大于 0。如果指定 0，则 MultiLoad 将发出错误消息并使用默认值（6 分钟）。
已暂存	禁用	加载数据的方法。选择“已暂存”可在将数据加载到数据库之前将其加载到平面文件暂存区域中。否则，会使用命名管道将数据加载到数据库中。
错误数据库	不适用	错误数据库名称。使用此属性可替代默认的错误数据库名称。如果不指定数据库名称，则集成服务会使用目标表数据库。
工作表数据库	不适用	工作表数据库名称。使用此属性可替代默认的工作表数据库名称。如果不指定数据库名称，则集成服务会使用目标表数据库。
日志表数据库	不适用	日志表数据库名称。使用此属性可替代默认的日志表数据库名称。如果不指定数据库名称，则集成服务会使用目标表数据库。
用户变量	不适用	默认的控制文件中使用的用户定义的变量。

下表显示了在会话属性中替代 Teradata MultiLoad 外部加载器连接对象时配置的属性：

属性	默认值	说明
错误表 1	不适用	第一个错误表的表名称。使用此属性可替代默认的错误表名称。如果不指定错误表名称，则集成服务会使用 ET_<target_table_name>。
错误表 2	不适用	第二个错误表的表名称。使用此属性可替代默认的错误表名称。如果不指定错误表名称，则集成服务会使用 UV_<target_table_name>。
工作表	不适用	工作表名称会替代默认的工作表名称。如果不指定工作表名称，则集成服务会使用 WT_<target_table_name>。
日志表	不适用	日志表名称会替代默认的日志表名称。如果不指定日志表名称，则集成服务会使用 ML_<target_table_name>。
控制文件内容替代	不适用	控制文件文本。使用此属性可替代集成服务在加载到 Teradata 时使用的控制文件。

配置 Teradata TPump 外部加载器属性

可以对目标执行插入、更新、删除和更新插入操作。还可以使用数据驱动模式根据更新策略转换或自定义转换执行插入、更新或删除操作。

如果运行具有多个分区的会话，请为每个分区选择一个 Teradata TPump 外部加载器。

要配置 Teradata TPump 外部加载器的属性，请单击“连接”>“加载器”，选择“类型”，然后单击“编辑”。

下表显示了为 Teradata TPump 外部加载器配置的属性：

属性	默认值	说明
TDPID	不适用	Teradata 数据库 ID。
数据库名称	不适用	可选数据库名称。如果不指定数据库名称，则集成服务会使用在映射中定义的目标表数据库名称。
错误限制	0	限制针对错误拒绝的行数。超出错误限制时，TPump 会回滚导致出现最后一个错误的事务。错误限制为 0 会导致 TPump 在出现任何错误后停止处理。
检查点	15	检查点之间的分钟数。必须将检查点设置为 0 到 60 之间的值。
韧度	4	TPump 尝试登录到所需会话的时间量（以小时为单位）。如果登录失败，TPump 会延迟在“休眠”属性中指定的分钟数，然后重试登录。TPump 将一直尝试，直到登录成功或超过在“韧度”属性中指定的小时数。 要禁用“韧度”，请将此值设置为 0。

属性	默认值	说明
加载模式	更新插入	<p>用于生成 SQL 命令的模式：插入、删除、更新、更新插入或数据驱动。</p> <p>选择“数据驱动”加载时，集成服务会按照更新策略转换或自定义转换中的说明来确定如何标记要插入、删除或更新的行。集成服务会在目标文件或命名管道中写入一行，以指示更新策略。控制文件使用这些值来确定如何将数据加载到数据库。集成服务使用以下值指示更新策略：</p> <p>0 - 插入 1 - 更新 2 - 删除</p>
删除错误表	已启用	在开始下次会话之前删除 TPump 错误表。选择此选项可删除表或删除表以将其保留。
外部加载器可执行文件	tpump	Teradata 外部加载器可执行文件的名称和可选文件路径。如果外部加载器可执行文件目录不位于系统路径中，则必须输入完整路径。
最大会话数量	1	每个 TPump 作业的最大 TPump 会话数。会话中的每个分区将启动其自己的 TPump 作业。运行多个 TPump 会话会导致客户端和数据库使用更多资源。因此，将此值设置为较小的数字可提高性能。
休眠	6	TPump 在重试登录之前等待的分钟数。TPump 将一直尝试，直到登录成功或超过在“韧度”属性中指定的小时数。
打包因子	20	每个会话缓冲区保存的行数。打包通过减少目标平面文件和 Teradata 数据库之间的发送和接收数目来提高网络/通道效率。
语句速率	0	TPump 可执行文件向 Teradata 数据库发送语句的每分钟最大初始速率。如果将此属性设置为 0，则未指定语句速率。
序列化	禁用	<p>确定针对给定键组合（行）的操作是否连续发生。</p> <p>如果 TPump 作业包含对某行的多次更改，则您可能希望启用此属性。如果会话包含多个具有相同键范围但具有不同筛选条件的分区，则可能会导致对单行作出多次更改。在这种情况下，您可能希望启用“序列化”以防止在 Teradata 数据库中出现锁定冲突，尤其是在将“包”属性设置为大于 1 的值时。</p> <p>如果启用“序列化”，集成服务会使用在目标表中指定的主键作为键列。如果目标表中不存在主键，则必须清除此选项或在控制文件的数据布局部分中指示键列。</p>
可靠	禁用	未选择“可靠”时，将会向 TPump 发送信号以使用简单的重新启动逻辑。在这种情况下，重新启动会导致 TPump 在上一个检查点处开始。TPump 将重新加载在此检查点之后加载的任何数据。此方法在可靠逻辑中没有其他数据库写入的额外开销。
无监视器	已启用	选中时，此属性将防止 TPump 通过 TPump 监视应用程序检查语句速率更改或为 TPump 监视应用程序更新状态信息。
已暂存	禁用	加载数据的方法。选择“已暂存”可在将数据加载到数据库之前将其加载到平面文件暂存区域中。否则，会使用命名管道将数据加载到数据库中。
错误数据库	不适用	错误数据库名称。使用此属性可替代默认的错误数据库名称。如果不指定数据库名称，则集成服务会使用目标表数据库。

属性	默认值	说明
日志表数据库	不适用	日志表数据库名称。使用此属性可替代默认的日志表数据库名称。如果不指定数据库名称，则集成服务会使用目标表数据库。
用户变量	不适用	默认的控制文件中使用的用户定义的变量。

下表显示了在会话属性中替代 Teradata TPump 外部加载器连接对象时配置的属性：

属性	默认值	说明
错误表	不适用	错误表名称。使用此属性可替代默认的错误表名称。如果不指定错误表名称，则集成服务会使用 ET_<target_table_name><partition_number>。
日志表	不适用	日志表名称。使用此属性可替代默认的日志表名称。如果不指定日志表名称，则集成服务会使用 TL_<target_table_name><partition_number>。
控制文件内容替代	不适用	控制文件文本。使用此属性可替代集成服务在加载到 Teradata 时使用的控制文件。

配置 Teradata FastLoad 外部加载器属性

使用 FastLoad 外部加载器时，请遵循以下准则：

- 每个 FastLoad 作业会将数据加载到一个 Teradata 数据库表中。如果要使用 FastLoad 将数据加载到多个表中，则必须创建多个 FastLoad 作业。
- 对于具有多个分区的会话，请使用循环分区类型将数据路由到单个目标文件。
- 目标表必须为空，且不包含任何定义的二级索引。
- 如果 Teradata 数据库中的目标表具有主键，则 FastLoad 不会将输出文件中的重复行加载到目标表。
- 如果将日期值加载到目标表，则必须使用 YYYY-MM-DD 格式为目标表中的列配置日期格式。
- 无法使用 FastLoad 加载二进制数据。
- 可以使用逗号 (,)、制表符 (\t) 和管道 (|) 作为分隔符。

要配置 Teradata FastLoad 外部加载器的属性，请单击“连接”>“加载器”，选择“类型”，然后单击“编辑”。

下表显示了为 Teradata FastLoad 外部加载器配置的属性：

属性	默认值	说明
TDPID	不适用	Teradata 数据库 ID。
数据库名称	不适用	数据库名称。
错误限制	1,000,000	FastLoad 在停止将数据加载到数据库表之前拒绝的最大行数。
检查点	0	在检查点之间传输到 Teradata 数据库的行数。如果在 FastLoad 作业运行时处理停止，则可以在最近检查点处重新启动作业。 如果输入 0，则 FastLoad 不会执行检查点操作。

属性	默认值	说明
韧度	4	当已在 Teradata 数据库上运行最大加载作业数时，FastLoad 尝试登录到所需 FastLoad 会话的小时数。当 FastLoad 尝试登录到新会话且 Teradata 数据库指示已在运行最大加载会话数时，FastLoad 会注销已登录的所有新会话，延迟在“休眠”属性中指定的分钟数，然后重试登录。FastLoad 将一直尝试，直到其登录所需会话数或超过在“韧度”属性中指定的小时数。
删除错误表	已启用	在开始下次会话之前删除 FastLoad 错误表。如果之前作业中存在非空错误表，则 FastLoad 将不会运行。 选择此选项可删除表或清除表以将其保留。
外部加载器可执行文件	fastload	Teradata 外部加载器可执行文件的名称和可选文件路径。如果外部加载器可执行文件目录不位于系统路径中，则必须输入完整路径。
最大会话数量	1	每个 FastLoad 作业的最大 FastLoad 会话数。“最大会话数量”必须介于 1 和系统上的访问模块进程 (AMP) 总数之间。
休眠	6	FastLoad 在重试登录之前暂停的分钟数。FastLoad 将一直尝试，直到登录成功或超过在“韧度”属性中指定的小时数。
截断目标表	禁用	在开始 FastLoad 作业之前截断目标数据库表。FastLoad 无法将数据加载到非空表。
已暂存	禁用	加载数据的方法。选择“已暂存”可在将数据加载到数据库之前将其加载到平面文件暂存区域中。否则，会使用命名管道将数据加载到数据库中。
错误数据库	不适用	错误数据库名称。使用此属性可替代默认的错误数据库名称。如果不指定数据库名称，则集成服务会使用目标表数据库。

下表显示了在会话属性中替代 Teradata FastLoad 外部加载器连接对象时配置的属性：

属性	默认值	说明
错误表 1	不适用	第一个错误表的表名称会替代默认的错误表名称。如果不指定错误表名称，则集成服务会使用 ET_<target_table_name>。
错误表 2	不适用	第二个错误表的表名称会替代默认的错误表名称。如果不指定错误表名称，则集成服务会使用 UV_<target_table_name>。
控制文件内容替代	不适用	控制文件文本。使用此属性可替代集成服务在加载到 Teradata 时使用的控制文件。

在会话中配置外部加载

在会话中配置外部加载之前，必须在 Workflow Manager 中创建外部加载器连接并配置外部加载器属性。

完成以下步骤以对会话使用外部加载器：

1. 将会话配置为写入平面文件，而不是关系数据库。
2. 配置文件属性。

3. 在会话属性中选择外部加载器连接。

将会话配置为写入文件

要使用外部加载器，请根据目标数据库类型在映射中创建目标定义。默认情况下，会话将配置关系目标类型。要选择外部加载器连接，必须将会话配置为写入文件，而不是关系目标。要将会话配置为写入文件，请将写入器类型从关系写入器更改为文件写入器。可以使用“映射”选项卡上的“写入器”设置更改写入器类型。

要更改目标的写入器类型，请选择目标实例并将写入器类型从“关系写入器”更改为“文件写入器”。

配置文件属性

将会话配置为写入文件后，可以设置文件属性。您需要指定输出文件名和目录以及拒绝文件名和目录。可以在“映射”选项卡上的“属性”设置中配置这些属性。要设置文件属性，请选择目标实例。

下表显示了“属性”设置中的属性：

属性	说明
输出文件目录	输出文件目录的名称和路径。在此字段中输入目录名称。默认情况下，集成服务会将输出文件写入 \$PMTargetFileDir 目录。 如果在“输出文件名”字段中输入完整目录和文件名，请清除此字段。如果在输出文件的路径中使用双空格，外部加载器会话可能会失败。
输出文件名	输出文件的名称。输入文件名或者文件名和路径。默认情况下，Workflow Manager 根据映射中使用的目标定义为目标文件命名： <i>target_name.out</i> 。如果在输出文件的路径中使用双空格，外部加载器会话可能会失败。
拒绝文件目录	拒绝文件目录的名称和路径。默认情况下，集成服务会将所有拒绝文件写入 \$PMBadFileDir 目录。 如果在“拒绝文件名”字段中输入完整目录和文件名，请清除此字段。
拒绝文件名	拒绝文件的名称。输入文件名，或输入文件名和目录。集成服务会将此字段中的信息附加到在“拒绝文件目录”字段中输入的目录中。例如，如果在“拒绝文件目录”字段中具有“C:/reject_file/”且在“拒绝文件名”字段中输入“filename.bad”，则集成服务会将拒绝的行写入 C:/reject_file/filename.bad 中。 默认情况下，集成服务以目标实例名称命名拒绝文件： <i>target_name.bad</i> 。 还可以输入拒绝文件会话参数，以表示拒绝文件或拒绝文件和目录。命名所有拒绝文件参数 \$BadFileName。
设置文件属性	平面文件属性的定义。使用外部加载器时，必须通过单击“设置文件属性”链接来定义平面文件属性。 对于 Oracle 外部加载器，目标平面文件可以是固定宽度，也可以带分隔符。 对于 Sybase IQ 外部加载器，目标平面文件可以是固定宽度，也可以带分隔符。 对于 Teradata 外部加载器，目标平面文件必须是固定宽度或带分隔符。 对于 IBM DB2 外部加载器，目标平面文件必须带分隔符。

注意：不要选择“合并已分区文件”或输入合并文件名。在使用外部加载器时，不能合并已分区的输出文件。

选择外部加载器连接

配置文件属性后，可以选择外部加载器连接。要选择外部加载器连接，请选择连接类型和连接对象。在“映射”选项卡的“连接”设置中配置连接选项。

如果会话包含多个分区，且选择可从多个输出文件加载的加载器，则可以为每个分区选择一个不同的连接，但每个连接必须属于相同类型。例如，可以为每个分区选择不同的 Teradata TPump 外部加载器连接，但不能为一个分区选择 Teradata TPump 连接，而为另一个分区选择 Oracle 连接。

如果会话包含多个分区，且选择只能从一个输出文件加载的加载器，请使用循环分区将数据路由到单个目标文件。可以为每个连接选择一个加载器，但集成服务会对第一个分区使用连接。

要选择外部加载器连接，请执行以下操作：

1. 在“映射”选项卡上的导航器中选择目标实例。
2. 选择加载器连接类型。
3. 在“值”字段中单击“打开”按钮。
4. 选择连接对象或变量：
 - **使用对象。**选择加载器连接对象。单击“替代”按钮替代连接属性。可替代的属性因加载器类型而异。
 - **使用连接变量。**使用 `$LoaderConnectionName` 会话参数，并在参数文件中定义此参数。替代参数文件中的连接属性。
5. 单击“确定”。

对外部加载进行故障排除

我正在尝试设置会话以将数据加载到外部加载器，但不能在会话属性中选择外部加载器连接。

验证映射是否包含关系目标。创建会话时，在会话属性中“映射”选项卡的“写入器”设置中选择一个文件写入器。然后，打开“连接”设置并选择外部加载器连接。

我正在尝试运行使用 TPump 的会话，但会话失败。会话日志显示一条错误，指出 Teradata 输出文件名太长。

集成服务使用 Teradata 输出文件名生成 TPump 错误和日志文件的名称以及日志表名称。为了生成这些名称，集成服务会将多个字符的前缀添加到输出文件名中。它将为具有一个分区的会话添加三个字符，为具有多个分区的会话添加五个字符。

Teradata 允许日志表名称最多具有 30 个字符。由于集成服务会添加前缀，如果您正在运行具有单个分区的会话，请指定最多具有 27 个字符的目标输出文件名，包括文件扩展名。如果您正在运行具有多个分区的会话，请指定最多具有 25 个字符的目标输出文件名，包括文件扩展名。

我尝试使用 TPump 将数据加载到 Teradata，但会话失败。我更正了错误，但会话仍然失败。

有时，Teradata 不会在您重新运行会话时删除日志表。检查 Teradata 数据库，并手动删除日志表（如果存在）。然后重新运行会话。

第 19 章

FTP

本章包括以下主题：

- [FTP 概览, 241](#)
- [SFTP, 242](#)
- [集成服务行为, 242](#)
- [在会话中配置 FTP, 243](#)

FTP 概览

可以将会话配置为使用文件传输协议 (FTP) 以从平面文件或 XML 源中读取或写入平面文件或 XML 目标。PowerCenter 集成服务可以使用 FTP 访问其可以连接到的任何计算机，包括大型机。对于源和目标文件，使用 FTP 可直接传输文件或将其暂存在本地目录中。直接访问源文件或使用文件列表访问会话中的间接源文件。

要在会话中使用 FTP 文件源和目标，请执行以下任务：

1. 在 Workflow Manager 中创建 FTP 连接对象并配置连接属性。
2. 在会话属性中将会话配置为使用 FTP 连接对象。

使用 FTP 的规则和准则

将 FTP 与平面文件或 XML 源和目标结合使用时，请遵循以下准则：

- 在会话属性中指定源或目录输出目录。如果未指定目录，则集成服务会将文件暂存在集成服务在 UNIX 或 Windows 系统目录中运行的目录中。
- 如果会话使用位于大型机上的相同 FTP 源文件或目标文件，则无法并发运行会话。
- 如果中止包含暂存大型机中的 FTP 源或目标的工作流，则可能需要先等待连接超时，然后才能再次运行该工作流。
- 要使用需要公钥身份验证的 SFTP 服务器的 FTP 连接运行会话，必须可以在将运行会话的节点上访问公钥和私钥文件。

SFTP

如果通过网络发送敏感数据，则可以使用安全文件传输协议 (SFTP) 保护数据。要连接到 SFTP 服务器，请配置 FTP 连接以使用 SFTP。SFTP 使文件通过安全数据流传输。PowerCenter 集成服务会创建 SSH2 传输层，进而启用安全连接以及 SFTP 服务器上文件的访问权限。

SFTP 在两个计算机系统之间创建一个加密通道，可防御以下攻击：

- IP 欺骗，在远程主机发出伪装成来自其他受信任主机的数据包时发生。
- IP 源路由，在主机可以伪称 IP 数据包来自其他受信任主机时发生。
- DNS 欺骗，当攻击者伪造名称服务器记录时发生。
- 通过中间主机截取明文密码和其他数据。
- 攻击者控制中间主机的数据的操作。

SFTP 将不对称密码和对称密码结合使用，可提供强大的加密能力和最佳性能。大多数商用服务器和许多开源服务器都支持 SFTP。SFTP 也是一个用于传输大型文件的有效协议，因为它在加密前压缩数据流。

要在会话中使用 SFTP 文件源和目标，请执行以下任务：

1. 创建 FTP 工作流连接，并为 SFTP 配置 FTP 连接对象。
2. 在会话属性中选择并配置 SFTP 连接对象。
3. 配置源文件属性。
4. 配置目标文件属性。

集成服务行为

使用 FTP 或 SFTP 的集成服务的行为取决于您配置 FTP 或 SFTP 连接和会话的方式。通过以下方法，Integration Service 集成服务可以使用 FTP 或 SFTP 访问源和目标文件：

- **源文件。** 在托管集成服务的计算机上暂存源文件或直接从 FTP 或 SFTP 主机访问源文件。使用单个源文件或包含单个源实例的间接源文件的文件列表。
- **目标文件。** 在托管集成服务的计算机上暂存目标文件或在 FTP 或 SFTP 主机上写入目标文件。

可以暂存 FTP 或 SFTP 文件以消除由于网络故障造成的部分传输风险。在托管集成服务的主机上创建暂存文件。在 FTP 或 SFTP 进程创建暂存文件后，集成服务会启动读取操作。在目标上使用 FTP 或 SFTP 时，FTP 或 SFTP 进程会在集成服务写入暂存文件后启动。如果在暂存文件完成前出现网络故障，可以删除暂存文件并重新运行会话。

可以在 FTP 或 SFTP 连接对象中或通过会话前或会话后 shell 命令配置暂存。

将 FTP 与源文件结合使用

在读取平面文件或 XML 文件源的会话中使用 FTP。可以在托管集成服务的计算机上暂存会话的源文件。使用单个源文件或每个源实例的文件列表。

暂存源数据时，集成服务会使用 FTP 创建本地文件。它将本地文件用作会话的源。在暂存文件完成前，集成服务不会将数据移至管道中。

如果不暂存源数据，则集成服务会使用 FTP 直接访问源文件。如果出现网络故障，必须重新运行会话。

下表介绍了结合使用 FTP 与源文件的集成服务的行为：

源类型	已暂存	集成服务行为
直接	是	集成服务会在会话开始后将文件从 FTP 主机复制到托管集成服务的计算机。
直接	否	集成服务使用 FTP 直接访问源文件。
间接	是	集成服务会在会话开始后读取文件列表并将文件列表和源文件复制到托管集成服务的计算机。
间接	否	集成服务会在会话开始后将文件列表复制到托管集成服务的计算机。集成服务使用 FTP 直接访问源文件。

将 FTP 与目标文件结合使用

在写入平面文件或 XML 文件目标的会话中使用 FTP。在将目标文件复制到 FTP 主机前，可以在托管集成服务的计算机上暂存目标文件。

暂存目标数据时，集成服务会在本地创建目标文件，然后在会话完成时将该文件传输到 FTP 主机。如果不暂存目标文件，则集成服务会直接在 FTP 主机上写入目标文件。如果出现网络故障，必须重新运行会话。

如果拥有“分区”选项，请将 FTP 用于多个目标分区实例。可以在集成服务或 FTP 主机上写入多个目标文件或合并文件。

在会话中配置 FTP

在将会话配置为使用 FTP 前，必须在 Workflow Manager 中创建 FTP 连接对象。集成服务使用 FTP 连接属性连接到 FTP 服务器。

在 Workflow Manager 中创建 FTP 连接后，可以将会话配置为使用 FTP。要使用安全连接，请选择为 SFTP 配置的 FTP 连接对象。将任何会话与平面文件或 XML 源或目标结合使用。

要配置会话，请为每个需要 FTP 连接的源和目标完成以下任务：

- 选择 FTP 连接。
- 配置源文件属性。
- 配置目标文件属性。

要在集成服务计算机上暂存源或目标文件，请在会话属性中编辑 FTP 连接以为暂存文件配置目录和文件名。

在会话中配置 SFTP

要使用需要公钥身份验证的 SFTP 连接对象运行会话，必须可以在将运行会话的节点上访问公钥和私钥文件。

如果集成服务已配置为在主节点和备份节点上运行，应使已配置用于运行集成服务进程的每个节点上均可访问密钥文件。

如果集成服务已配置为在网络上运行，应使已配置用于在网络上运行的每个节点上均可访问密钥文件。如果无法将文件放置于网络中的每个节点上，请在域中创建资源并将其分配给要放置文件的每个节点。创建会话时，请将其配置为使用该资源。

例如，创建一个名为 SFTP 的自定义资源。创建会话时，可能需要会话使用该 SFTP 资源。负载均衡器仅将会话分派给可访问密钥文件的节点。

选择 FTP 连接

要将会话配置为使用 FTP，请选择连接类型和连接对象。为将使用该 FTP 连接的每个源和目标选择 FTP 连接对象。要使用 SFTP，请选择已为 SFTP 配置的 FTP 连接对象。在“映射”选项卡的“连接”设置中配置连接选项。

要为每个源或目标实例选择 FTP 连接，请执行以下操作：

1. 在“映射”选项卡上，在“转换”视图中选择源或目标实例。
2. 选择 FTP 连接类型。
3. 在“值”字段中单击“打开”按钮。
4. 选择连接对象或变量：
 - **使用对象。**选择 FTP 连接对象。单击“替代”按钮替代连接属性。
 - **使用连接变量。**使用 `$FTPConnectionName` 会话参数，并在参数文件中定义该参数。替代参数文件中的连接属性。

可以替代以下属性：

属性	说明
远程文件名	源或目标的远程文件名。如果使用间接源文件，请输入间接源文件名。 文件名必须使用 7 位 ASCII 字符。如果使用具有 Unicode 字符的远程文件名，会话将失败。 如果为源文件名输入完全限定名称，集成服务会忽略在“默认远程目录”字段中输入的路径。如果用单引号或双引号引起完全限定名称，则会话将失败。 可以为远程文件名使用参数或变量。可使用您在参数文件中可定义的任何参数或变量类型。例如，可以将会话参数 <code>\$ParamMyRemoteFile</code> 用作源或目标远程文件名，并在参数文件中将 <code>\$ParamMyRemoteFile</code> 设置为文件名。
已暂存	在集成服务上暂存源或目标文件。默认值为“不暂存”。
传输模式 ASCII	更改传输模式。如果启用，则集成服务使用 ASCII 传输模式。在 Windows 计算机上传输文件时使用 ASCII 模式可确保行字符的末尾在文本文件中可以正确转换。如果禁用，则集成服务使用二进制传输模式。在 UNIX 计算机上传输文件时，使用二进制传输模式。默认为已禁用。

5. 单击“确定”。

配置源文件属性

如果使用 FTP 或 SFTP 访问源文件，请在为源实例选择 FTP 或 SFTP 连接对象后配置源文件属性。源文件属性可确定源文件类型和暂存位置。在“映射”选项卡的“属性”设置中配置源文件属性。

如果要暂存源文件，请选择源文件名、目录和文件类型。

如果不暂存源文件，请指定源文件类型。PowerCenter 集成服务使用 FTP 连接对象中的远程文件名和目录，并忽略源文件名和目录。

1. 在 Workflow Manager 中，打开任务开发程序，然后单击任务 > 创建。
2. 选择“会话”作为任务类型。

3. 输入会话任务的名称。不要在任务名称中使用 句号字符 (.)。Workflow Manager 不允许任务名称包含句号字符。
4. 单击**创建**。
任务开发程序将创建会话任务。
5. 在**映射**对话框中，选择要在“会话”任务中使用的映射，然后单击**确定**。
6. 单击**完成**。
7. 双击工作区中的图标以打开会话属性。
此时将显示**编辑任务**对话框。
8. 在**映射**选项卡“源”节点的“属性”设置中配置以下源文件属性：

属性	说明
源文件类型	指示源文件是否包含源数据或具有相同文件属性的文件列表。 如果源文件包含源数据，选择“直接”。 如果源文件包含文件列表，选择“间接”。
源文件目录	用于暂存源数据的本地源文件目录的名称和路径。 默认情况下，PowerCenter 集成服务对文件源使用服务进程变量目录 \$PMSourceFileDir。PowerCenter 集成服务会在其运行会话时连接此字段与“源文件名”字段。 如果不暂存源文件，则 PowerCenter 集成服务会使用为 SFTP 配置的 FTP 连接对象中的文件名和目录。 如果在“源文件名”字段中输入完全限定文件名，则 PowerCenter 集成服务会忽略此字段。
源文件名	用于暂存源数据的本地源文件的名称。 可以输入文件名或文件名和路径。 如果输入完全限定文件名，则 PowerCenter 集成服务会忽略“源文件目录”字段。 如果不暂存源文件，则 PowerCenter 集成服务会使用为 SFTP 配置的 FTP 连接对象中的远程文件名和默认目录。

9. 单击**应用**。
10. 要关闭**编辑任务**对话框，请单击**确定**。

配置目标文件属性

如果使用 FTP 或 SFTP 写入目标文件，请在为目标实例指定 FTP 或 SFTP 连接对象后指定目标文件属性。目标文件属性可确定拒绝文件和目录以及暂存位置。在“映射”选项卡的“属性”设置中指定目标文件属性。

如果要暂存目标文件，请配置目标文件名和目录以及拒绝文件名和目录。如果不暂存目标文件，请配置拒绝文件和目录。PowerCenter 集成服务使用 FTP 连接对象中的远程文件名和目录。

如果拥有“分区”选项，还可以选择合并文件属性。

1. 在 Workflow Manager 中，打开任务开发程序，然后单击**任务 > 创建**。
2. 选择“会话”作为任务类型。
3. 输入会话任务的名称。不要在任务名称中使用 句号字符 (.)。Workflow Manager 不允许任务名称包含句号字符。
4. 单击**创建**。
任务开发程序将创建会话任务。
5. 在**映射**对话框中，选择要在“会话”任务中使用的映射，然后单击**确定**。
6. 单击**完成**。

7. 双击工作区中的图标以打开会话属性。
此时将显示**编辑任务**对话框。
8. 在**映射**选项卡“目标”节点的“属性”设置中配置以下目标文件属性：

属性	说明
输出文件目录	用于暂存目标数据的本地目标文件目录的名称和路径。默认情况下，PowerCenter 集成服务使用服务进程变量目录 \$PMTargetFileDir。PowerCenter 集成服务会在其运行会话时连接此字段与“输出文件名”字段。 如果不暂存目标文件，则 PowerCenter 集成服务会使用 FTP 连接对象中的文件名和目录。 如果在“输出文件名”字段中输入完全限定文件名，则 PowerCenter 集成服务会忽略此字段。
输出文件名	用于暂存目标数据的本地目标文件的名称。可以输入文件名，或文件名和路径。如果输入完全限定文件名，则 PowerCenter 集成服务会忽略“输出文件目录”字段。 如果不暂存源文件，则 PowerCenter 集成服务会使用 FTP 连接对象中的远程文件名和默认目录。

9. 单击**应用**。
10. 要关闭**编辑任务**对话框，请单击**确定**。
11. 要保存目标文件属性中的更改，请单击**存储库 > 保存**。

分区 FTP 文件目标

为会话中的已分区目标选择 FTP 连接类型时，为目标分区配置 FTP 设置。

可以合并目标文件或每个的各个目标文件。

为目标分区配置 FTP 设置时，请遵循以下规则和准则：

- 必须为每个目标分区使用 FTP 连接。
- 为目标分区选择连接对象时，可以选择暂存文件。必须暂存文件才能使用连续合并。
- 如果目标分区的 FTP 连接拥有除远程文件名之外的任何设置，则集成服务不会创建合并文件。

下表介绍了集成服务为已分区的 FTP 文件目标完成的操作：

合并类型	集成服务行为
不合并	集成服务为每个分区生成一个目标文件。如果暂存文件，则集成服务会在会话结束时将目标文件传输到远程位置。如果不暂存文件，则集成服务会在远程位置生成目标文件。
连续合并	在连接对象中启用 已暂存 选项。集成服务为每个分区创建一个输出文件。会话结束时，集成服务会完成以下操作： 1. 将各个输出文件合并到一个合并文件中。 2. 删除各个输出文件。 3. 将合并文件传输到远程位置。

合并类型	集成服务行为
文件列表	<p>如果暂存文件，则集成服务会创建以下文件：</p> <ul style="list-style-type: none"> - 每个分区的输出文件 - 包含本地文件的名称和路径的文件列表 - 包含远程文件的名称和路径的文件列表 <p>会话结束时，集成服务会将文件传输到远程位置。如果各个目标文件位于合并文件目录中，则文件列表包含相对路径。否则，文件列表包含绝对路径。</p> <p>如果不暂存文件，则集成服务会在远程位置写入每个分区的数据，并创建包含各个目标文件列表的远程文件列表。</p> <p>将文件列表用作其他映射中的源文件。</p>
并发合并	<p>如果暂存文件，则集成服务会并发地将所有目标分区的数据写入本地合并文件。会话结束时，集成服务会将合并文件传输到远程位置。集成服务无法写入任何中间输出文件。</p> <p>如果不暂存文件，则集成服务会并发地将所有分区的目标数据写入远程位置的合并文件。</p>

第 20 章

会话缓存

本章包括以下主题：

- [会话缓存概览, 248](#)
- [高速缓存, 249](#)
- [缓存文件, 250](#)
- [配置缓存大小, 252](#)
- [缓存分区, 254](#)
- [汇总器缓存, 255](#)
- [联接器缓存, 256](#)
- [查找缓存, 259](#)
- [等级缓存, 260](#)
- [排序器缓存, 261](#)
- [XML 目标缓存, 262](#)
- [优化缓存大小, 262](#)

会话缓存概览

集成服务会为映射中的 XML 目标和汇总器、联接器、查找、等级以及排序器转换分配高速缓存。集成服务会为 XML 目标和汇总器、联接器、查找和等级转换创建索引和数据缓存。集成服务将键值存储在索引缓存中，将输出值存储在数据缓存中。集成服务会为排序器转换创建一个缓存以存储排序键和要排序的数据。

可以在会话属性中为缓存配置内存参数。首次配置缓存大小时，可以计算处理转换所需的内存量或配置集成服务以便在运行时自动配置内存要求。

运行会话后，可以在会话中调整转换的缓存大小。可以分析转换统计信息以确定为实现最优会话性能所需的缓存大小，然后更新已配置的缓存大小。

如果集成服务所需的内存大小超过您配置的内存大小，则会将溢出值存储在缓存文件中。会话完成后，集成服务会释放高速缓存，且在多数情况下，会删除缓存文件。

如果会话包含多个分区，则集成服务会为每个分区创建一个内存缓存。在特定情况下，集成服务会使用缓存分区为每个分区创建单独的缓存。

下表介绍了集成服务在每个缓存中存储的信息类型：

映射对象	缓存类型和说明
汇总器	- 索引。根据分组端口中的配置存储组值。 - 数据。基于分组端口存储计算。
联接器	- 索引。将具有唯一键的所有主行存储在联接条件中。 - 数据。存储主源行。
查找	- 索引。存储查找条件信息。 - 数据。存储未存储在索引缓存中的查找数据。
等级	- 索引。根据分组端口中的配置存储组值。 - 数据。根据分组端口存储等级信息。
排序器	- 排序器。存储排序键和数据。
XML 目标	- 索引。将主键和外键信息存储在单独的缓存中。 - 数据。生成 XML 目标时存储 XML 行数据。

高速缓存

集成服务会根据配置的缓存大小创建各个内存缓存。创建会话时，可以在会话属性中为每个转换实例配置缓存大小。

出于以下某个原因，集成服务可能会增加配置的缓存大小：

- **配置的缓存大小低于处理操作所需的最低缓存大小。**集成服务要求提供最小内存量来初始化每个会话。如果配置的缓存大小低于所需的最低缓存大小，则集成服务会增加配置的缓存大小以满足最低要求。如果集成服务无法分配所需的最低内存，则会话失败。
- **配置的缓存大小不是缓存页面大小的倍数。**集成服务将缓存数据存储在缓存页面中。缓存页面必须均匀放入缓存。因此，如果将缓存大小配置为 10 MB（1,048,576 个字节），并将缓存页面大小配置为 10,000 个字节，则集成服务会将配置的缓存大小增加到 1,050,000 个字节，从而使其成为页面大小为 10,000 个字节的倍数。

集成服务增加配置的缓存大小时，会继续运行会话并在会话日志中写入一条类似于以下内容的消息：

```
MAPPING> TE_7212 Increasing [Index Cache] size for transformation <transformation name> from <configured index cache size> to <new index cache size>.
```

查看会话日志以验证是否满足最低要求分配了足够内存。

为实现最优性能，请将缓存大小设置为处理转换所需的总内存。如果处理转换的高速缓存不足，则集成服务会先处理内存中的一些转换，然后对磁盘的信息进行分页以处理剩余的转换。

通过以下信息了解集成服务如何在 32 位和 64 位计算机上按不同的方式处理内存缓存：

- 如果所有已配置的会话缓存的总大小超过 2 GB，则在 32 位计算机上运行的集成服务进程将无法运行会话。如果在网格上运行会话，则在单个节点上运行的所有会话线程的总缓存大小必须不能超过 2 GB。
- 如果网格具有 32 位和 64 位集成服务进程，且会话的内存大小超过 2 GB，则必须将会话配置为在 64 位计算机的集成服务上运行。

缓存文件

运行会话时，集成服务应为每个转换至少创建一个缓存文件。如果集成服务无法处理内存中的转换，则会将溢出值写入到缓存文件。

下表介绍了集成服务为不同的映射对象创建的缓存文件类型：

映射对象	缓存文件
汇总器、联接器、查找和等级转换	集成服务会创建以下类型的缓存文件： <ul style="list-style-type: none">- 为每个索引缓存和数据缓存创建一个表头文件- 为每个索引缓存和数据缓存创建一个数据文件
排序器转换	集成服务会创建一个排序器缓存文件。
XML 目标	集成服务会创建以下类型的缓存文件： <ul style="list-style-type: none">- 为每个 XML 目标组创建一个数据缓存文件- 为每个 XML 目标组创建一个主键索引缓存文件- 为每个 XML 目标组创建一个外键索引缓存文件

集成服务会根据集成服务代码页创建缓存文件。

运行会话时，集成服务会在会话日志中写入一条指示缓存文件名称和转换名称的消息。会话完成后，集成服务会释放高速缓存，且通常删除缓存文件。在下列情况下，可以在缓存目录中查找索引和数据缓存文件：

- 会话执行增量汇总。
- 配置查找转换以使用持久性缓存。
- 会话完成失败。下次运行会话时，集成服务会删除现有的缓存文件并创建新的缓存文件。

注意：由于写入到缓存文件可能会降低会话性能，因此请配置缓存大小以处理内存中的转换。

缓存文件的命名约定

对于索引、数据和排序器缓存文件，集成服务使用不同的命名约定。

下表介绍了每个缓存文件类型的命名约定：

缓存文件	命名约定
数据和排序器	[<名称前缀> <前缀> <会话 ID>_<转换 ID>]_[分区索引]_[OS][BIT]。<后缀> [溢出索引]
索引	<前缀> <会话 ID>_<转换 ID>_<组 ID>_<键类型>。<后缀> <溢出>

下表介绍了缓存文件名称的各个组件：

文件名组件	说明
名称前缀	在查找转换中配置的缓存文件名称前缀。用于查找转换缓存文件。
前缀	描述转换的类型： <ul style="list-style-type: none"> - 汇总器转换为 PMAGG。 - 联接器转换为 PMJNR。 - 查找转换为 PMLKUP。 - 等级转换为 PMAGG。 - 排序器转换为 PMSORT。 - XML 目标为 PMXML。
会话 ID	会话实例 ID 号。
转换 ID	转换实例 ID 号。
组 ID	层次结构 XML 目标中的各个组的 ID。集成服务会为每个组创建一个索引缓存。用于 XML 目标缓存文件。
键类型	键的类型。可以是外键或主键。用于 XML 目标缓存文件。
分区索引	如果会话包含多个分区，则通过分区号进行标识。分区索引从零开始，因此第一个分区没有分区索引。分区索引 2 表示在第三个分区中创建的缓存文件。
操作系统	标识正在运行集成服务进程的计算机的操作系统： <ul style="list-style-type: none"> - W 表示 Windows。 - S 表示 Solaris。 - A 表示 AIX。 - L 表示 Linux。 - M 表示大型机。 用于查找转换缓存文件。
位	标识正在运行集成服务进程的计算机的位平台：32 位或 64 位。用于查找转换缓存文件。
后缀	标识缓存文件的类型： <ul style="list-style-type: none"> - 对于表头文件，索引缓存文件为 .idx0；对于数据文件，索引缓存文件为 .idx<i>n</i>。 - 对于表头文件，数据缓存文件为 .dat0；对于数据文件，数据缓存文件为 .dat<i>n</i>。 - 排序器缓存文件为 .PMSORT()。
溢出索引	如果缓存文件处理数据的大小超过 2 GB，则集成服务会创建多个缓存文件。创建这些文件时，集成服务会向文件名称附加溢出索引，如 PMAGG*.idx2 和 PMAGG*.idx3。缓存文件的数量受缓存目录中的可用磁盘空间量的限制。 注意： 排序器转换缓存文件处理数据的大小超过 2 GB 时，PowerCenter 集成服务不会创建多个缓存文件。

例如，索引缓存的数据文件的名称为 PMLKUP748_2_5S32.idx1。PMLKUP 将转换类型标识为查找，748 表示会话 ID，2 表示转换 ID，5 表示分区索引，S (Solaris) 表示操作系统以及 32 表示位平台。

缓存文件目录

默认情况下，集成服务会在 \$PMCacheDir 目录中创建缓存文件。如果集成服务进程找不到该目录，则会话失败并向会话日志中写入一条消息，指示无法创建或打开缓存文件。

集成服务可能会创建多个缓存文件。缓存文件的数量受缓存目录中的可用磁盘空间量的限制。

如果在网格上运行集成服务，并且只有部分集成服务节点能够快速访问共享缓存文件目录，请将每个具有大型缓存的会话配置为在能够快速访问目录的节点上运行。要将会话配置为在能够快速访问目录的节点上运行，请完成以下步骤：

1. 创建 PowerCenter 资源。
2. 使资源可用于能够快速访问目录的节点。
3. 将资源分配给会话。

如果网格中的所有集成服务进程对缓存文件的访问都十分缓慢，请为每个集成服务进程设置单独的本地缓存文件目录。如果集成服务进程在包含缓存目录的同一台计算机上运行，它可以更快地访问缓存文件。

配置缓存大小

可以在会话属性中为缓存配置内存量。在会话属性中指定的缓存大小会替代在转换属性中设置的值。

配置的内存量取决于要使用的内存缓存和磁盘缓存。如果您配置的缓存大小不足以在内存中处理转换，集成服务将在内存中处理部分转换，并将信息分页到缓存文件，以处理其余的转换。为实现最优会话性能，请配置缓存大小以便集成服务可以处理内存中的所有数据。

如果会话可重用，则该会话的所有实例会使用在可重用会话属性中配置的缓存大小。但无法替代会话实例中的缓存大小。

使用以下方法之一配置缓存大小：

- **缓存计算器。**使用计算器可估算处理转换所需的总内存量。
- **自动高速缓存。**使用自动内存可指定为处理转换所分配的缓存大小的最大限制。如果运行集成服务进程的计算机对高速缓存存在限制，则使用此方法。
- **数值。**为缓存大小配置特定值。如果要调整缓存大小，请配置特定值。

集成服务使用缓存分区时，可以按不同的方式配置内存需求。如果集成服务使用缓存分区，请为每个分区分配配置的缓存大小。要使用缓存分区配置转换的内存需求，请计算转换的总需求并除以分区数。

转换的输入发生更改时，转换的缓存大小需求可能也会发生更改。定期在会话日志中监视缓存大小可帮助您调整缓存大小。

计算缓存大小

使用缓存计算器可估算处理转换所需的总内存量。必须提供输入才能计算缓存大小。输入取决于转换的类型。例如，要计算汇总器转换的缓存大小，请提供组数。

可以在缓存计算器中选择以下模式之一：

- **自动。**如果希望集成服务在运行时根据在“配置对象”选项卡上配置的最大内存来确定缓存大小，请选择自动模式。
- **计算。**选择基于输入计算转换的总需求。缓存计算器要求每个转换提供不同的输入。必须选择适用的缓存类型以应用计算的缓存大小。例如，要将计算的缓存大小应用于数据缓存而不是索引缓存，仅需选择“数据缓存大小”选项。

缓存计算器会根据您的输入估算为实现最优会话性能所需的缓存大小。配置缓存大小并运行会话后，可以在会话日志中查看转换统计信息以调整配置的缓存大小。

注意：无法使用缓存计算器估算 XML 目标的缓存大小。

自动缓存大小

默认情况下，转换的内存缓存设置为自动模式。集成服务会将高速缓存自动分配给其缓存已设置为自动模式的所有转换中。可以设置集成服务能为转换分配的最大高速缓存量。

要在自动缓存模式下为转换设置最大高速缓存，请配置以下会话属性：

自动内存属性允许使用的内存上限

用于分配给会话缓存的最大内存量。集成服务会将会话缓存中的内存分配到高速缓存设置为自动的所有转换中。默认单位为字节。为值附加 KB、MB 或 GB 可指定其他单位。例如，1048576 或 1,024 KB 或 1 MB。

自动内存属性允许使用的总内存百分比上限

用于分配给会话缓存的计算机内存的百分比。集成服务会将会话缓存中的内存分配到高速缓存设置为自动的所有转换中。

为会话设置最大缓存大小时，集成服务会先计算内存的最大百分比并根据您指定的最大内存量进行比较。然后在自动缓存模式下，为转换分配较低的内存量。如果在自动缓存模式下具有多个转换，则集成服务会为自动缓存模式下的所有转换分配内存。

例如，托管集成服务的计算机具有 1 GB 的内存。将“自动内存属性允许使用的内存上限”属性设置为 800 MB。同时将“自动内存属性允许使用的总内存百分比上限”属性设置为 10%。集成服务会将 102.4 MB 的内存分配至会话缓存，并在自动缓存模式下的所有转换之间分配高速缓存。

您设置的最大会话缓存大小仅会影响缓存模式设置为自动的转换。集成服务会为配置特定缓存大小的转换单独分配内存。

如果会话具有多个需要缓存的转换，则可以将其中一些转换的缓存模式设置为自动，并指定其他转换的缓存大小。集成服务除了向转换分配使用数值缓存大小配置的内存外，还会在自动模式下分配为转换指定的内存。

例如，会话具有三个需要缓存的转换。可以先将两个转换设置为自动缓存模式，并指定会话的最大内存缓存大小为 800 MB。同时指定第三个转换的缓存大小为 500 MB。此时集成服务便会分配 1,300 MB 的总内存量。

如果集成服务使用缓存分区，则集成服务会在会话的所有转换之间分发给自动高速缓存指定的最大缓存大小，并在其所有分区中分配每个转换的高速缓存。

配置数值缓存大小

可以为缓存大小配置特定值。可以在调整缓存大小时配置特定值。首次配置缓存大小时，可以使用缓存计算器或自动高速缓存。配置缓存大小并运行会话后，可以分析会话日志中的转换统计信息以调整缓存大小。会话日志会显示在不对磁盘进行分页的情况下处理内存中的转换所需的缓存大小。使用会话日志中指定的缓存大小以实现最优会话性能。

配置缓存大小的步骤

可以在会话属性中为转换配置缓存大小。配置缓存大小时，请指定转换的总需求，除非集成服务使用缓存分区。

如果集成服务使用缓存分区，可以按不同的方式配置缓存大小。要在集成服务使用缓存分区时计算缓存大小，请先计算转换的总需求，然后除以分区数。

要在会话中配置缓存大小，请执行以下操作：

1. 在 Workflow Manager 中打开会话。
2. 单击“映射”选项卡。

3. 在左侧窗格中选择映射对象。
- “映射”选项卡的右侧窗格用于显示可以配置缓存大小的对象属性。
4. 使用以下其中一个方法设置缓存大小：
- 为缓存大小输入相应值，单击“确定”，然后跳至步骤 8。如果输入值，默认情况下所有值均以字节为单位。但是，也可以输入值并指定以下单位之一：KB、MB 或 GB。如果输入单位，请勿在值与单位之间输入空格。例如，输入 350000KB、200MB 或 1GB。
- 或-
- 将缓存大小输入为“自动”，单击“确定”，然后跳至步骤 8。
- 或-
- 单击“打开”按钮打开缓存计算器。
5. 选择模式。
- 选择“自动”模式可限制分配给转换的缓存量。请跳至步骤 8。
- 或-
- 选择“计算”模式可计算转换的总内存需求。
6. 根据转换类型提供输入，然后单击“计算”。
- 注意:** 如果输入值过大且您无法在缓存计算器中输入该值，请使用自动内存缓存。
- 缓存计算器会计算缓存大小，以千字节表示。
7. 如果转换包含数据缓存和索引缓存，请选择“数据缓存大小”、“索引缓存大小”或两者同时选中。
8. 单击“确定”可将计算的值应用于在步骤 7 中选择的缓存大小。

缓存分区

创建具有多个分区的会话时，集成服务会将缓存分区用于汇总器、联接器、查找、等级或排序器转换。集成服务分区缓存时，会为每个分区创建单独的缓存并将已配置的缓存大小分配给每个分区。集成服务在每个缓存中存储不同数据，每个缓存仅包含该分区所需的行。因此，集成服务需要总缓存内存的一部分用于每个分区。

集成服务使用缓存分区时，会并行访问每个分区的缓存。如果集成服务未使用缓存分区，则串行访问每个分区的缓存。

下表介绍了集成服务为每个适用转换使用缓存分区时的情况：

转换	说明
汇总器转换	在使用汇总器转换的会话中创建多个分区。无需在汇总器转换时设置分区点。
联接器转换	在联接器转换时创建分区点。
查找转换	在查找转换时创建哈希自动键分区点。
等级转换	在使用等级转换的会话中创建多个分区。无需在等级转换时设置分区点。
排序器转换	在使用排序器转换的会话中创建多个分区。无需在排序器转换时设置分区点。

为缓存分区配置缓存大小

集成服务使用缓存分区时，可以按不同的方式配置内存需求。如果集成服务使用缓存分区，请为每个分区分配配置的缓存大小。要使用缓存分区配置转换的内存需求，请计算转换的总需求并除以分区数。

例如，在具有汇总器转换的会话中创建四个分区。先确定汇总器转换所需的数据缓存的内存大小为 400 MB。为汇总器转换的数据缓存大小配置 100 MB。运行会话时，集成服务会使用汇总器转换的总 400 MB 的内存大小为每个分区各自分配 100 MB。

使用缓存计算器计算转换的总需求。如果使用动态分区，则可以根据动态分区方法确定分区数。如果根据网格中的节点使用动态分区，则集成服务会为每个节点创建一个分区。如果根据源分区使用动态分区，则使用源数据库中的分区数。

汇总器缓存

集成服务会使用高速缓存来处理未排序输入的汇总器转换。运行会话时，集成服务会在完成汇总计算之前先将数据存储在内​​存中。

集成服务会为汇总器转换创建以下缓存：

- **索引缓存。**根据分组端口中的配置存储组值。
- **数据缓存。**基于分组端口存储计算。

默认情况下，集成服务会分别为转换中的数据和索引创建一个内存缓存和一个磁盘缓存。

在具有汇总器转换的会话中创建多个分区时，集成服务会使用缓存分区。集成服务会为所有分区创建一个磁盘缓存并为每个分区创建单独的内存缓存。

增量汇总

首次运行增量汇总会话时，集成服务会处理源。会话结束后，集成服务会将汇总的数据分别存储在两个缓存文件中（索引和数据缓存文件）。集成服务将缓存文件保存在缓存文件目录中。下次运行会话时，集成服务会将新行与缓存文件中的缓存汇总值进行汇总。

运行增量汇总器转换的会话时，集成服务会在会话开始运行时在 \$PMCacheDir 目录中创建汇总器缓存文件的备份。集成服务会在会话恢复开始运行时将备份缓存升级至初始缓存。如果会话中止，则集成服务无法还原备份缓存文件。

在使用增量汇总的会话中创建多个分区时，集成服务会为每个分区创建一组缓存文件。

为汇总器转换配置缓存大小

可以为无排序端口的汇总器转换配置缓存大小。

您无需配置使用已排序端口的汇总器转换的高速缓存。集成服务使用系统内存来处理具有已排序端口的汇总器转换。

下表介绍了为计算汇总器缓存大小所提供的输入：

选项名称	说明
组数	组数。汇总器转换按组汇总数据。使用分组端口计算组数。例如，如果按存储 ID 和项目 ID 进行分组，具有 5 个存储和 25 个项目，且每个存储包含所有 25 个项目，则将组数计算为： $5 * 25 = 125 \text{ groups}$
数据移动模式	集成服务的数据移动模式。缓存需求因数据移动模式而异。每个 ASCII 字符使用一个字节。每个 Unicode 字符使用两个字节。

输入输入内容，然后单击“计算”可计算数据和索引缓存大小。计算后的值显示在“数据缓存大小”和“索引缓存大小”字段中。

汇总器缓存故障排除

使用本节中的信息可帮助对汇总器转换的缓存进行故障排除。

使用缓存计算器计算汇总器转换的缓存大小时，显示以下警告：

CMN_2019 Warning: The estimated data cache size assumes the number of aggregate functions equals the number of connected output-only ports. If there are more aggregate functions, increase the cache size to cache all data in memory.

可以使用汇总器转换中的一个或多个汇总函数。如果输出基于其中一个汇总函数，则缓存计算器会估算缓存大小。如果使用多个汇总函数以确定一个输出端口值，则必须增加缓存大小。

查看会话日志中的转换统计信息，并在会话中调整汇总器转换的缓存大小。

联接器缓存

集成服务使用高速缓存来处理联接器转换。运行会话时，集成服务会并发读取主源和详细信息源中的行，并根据主行构建索引和数据缓存。集成服务会基于详细信息源数据和缓存主数据执行联接。

集成服务会基于联接器转换的类型在缓存中存储不同的行数。

下表介绍了针对不同的联接器转换类型，集成服务存储在缓存中的信息：

联接器转换类型	索引缓存	数据缓存
未排序输入	将具有唯一索引键的所有主行存储在联接条件中。	存储所有主行。
具有不同源的已排序输入	将具有唯一索引键的 100 条主行存储在联接条件中。	存储与存储在索引缓存中的行对应的主行。如果主数据包含多个具有相同键的行，则集成服务在数据缓存中存储的行数将超过 100。
具有相同源的已排序输入	将具有唯一键的所有主行或详细信息行存储在联接条件中。如果集成服务处理详细信息管道的速度比处理主管道的速度快，则存储详细信息行。否则，存储主行。集成服务存储的行数取决于主管道和详细信息管道的处理速率。如果一个管道处理其行的速度比其他管道处理的速度快，则集成服务会缓存所有已处理的行并将其保持缓存状态，直至其他管道处理完其行。	存储已存储在索引缓存中的行的数据。如果索引缓存存储主管道的键值，则数据缓存会存储主管道的数据。如果索引缓存存储详细信息管道的键值，则数据缓存会存储详细信息管道的数据。

如果数据已排序，则集成服务会为所有分区创建一个磁盘缓存，并为每个分区创建单独的内存缓存。集成服务联接行中的数据后，会从缓存中释放每一行。

如果数据未排序且在联接器转换时没有分区，则集成服务会为每个分区创建一个磁盘缓存和一个单独的内存缓存。如果数据未排序但在联接器转换时存在一个分区，则集成服务会为每个分区创建一个单独的磁盘缓存和内存缓存。如果数据未排序，则集成服务会将所有主数据均保留在缓存中，直到其联接完所有数据。

在会话中创建多个分区时，可以使用 1:*n* 分区或 *n*:*n* 分区。使用 1:*n* 分区和使用 *n*:*n* 分区时，集成服务会按不同的方式处理联接器转换。

1:*n* 分区

可以将 1:*n* 分区与具有已排序输入的联接器转换结合使用。使用 1:*n* 分区时，会为主管道创建一个分区，并在详细信息管道中创建多个分区。集成服务处理联接时，会将详细信息分区中的行与主源中的行进行比较。处理外部联接的主数据和详细信息数据时，集成服务会在处理完所有详细信息分区后输出不匹配的主行。

n:*n* 分区

可以将 *n*:*n* 分区与具有排序输入或未排序输入的联接器转换结合使用。将 *n*:*n* 分区用于联接器转换时，会在主管道和详细信息管道中创建 *n* 个分区。集成服务处理联接时，会将详细信息分区中的行与对应的主分区中的行进行比较，但会忽略其他主分区中的行。处理外部联接的主数据和详细信息数据时，集成服务会在处理完每个详细信息缓存的分区后输出不匹配的主行。

提示：如果主源包含大量行，则使用 *n*:*n* 分区可实现更优的会话性能。

要使用 *n*:*n* 分区，必须在会话中创建多个分区并在联接器转换时创建一个分区点。在联接器转换时创建分区点可同时为联接器转换的主源和详细信息源创建多个分区。

如果在联接器转换时创建分区点，则集成服务使用缓存分区。集成服务会为每个分区创建一个内存缓存。每个分区的内存缓存仅包含该分区所需的行数。因此，集成服务需要总缓存内存的一部分用于每个分区。

为联接器转换配置缓存大小

可以为联接器转换会话属性配置索引和数据缓存大小。

使用 1:*n* 分区时，集成服务会复制每个分区的内存缓存。每个分区所需的内存量与转换的总需求一样。使用 1:*n* 分区配置联接器转换的缓存大小时，请将缓存大小设置为该转换的总需求。

使用 *n*:*n* 分区时，每个分区仅需要处理转换所需的总内存中的一小部分内存。使用 *n*:*n* 分区配置联接器转换的缓存大小时，请先计算该转换的总需求，然后除以分区数。

可以使用缓存计算器来确定处理转换所需的缓存大小。例如，使用缓存计算器以确定联接器转换需要 2,000,000 个字节的内存量用于索引缓存和 4,000,000 个字节的内存量用于数据缓存。可以为该管道创建四个分区。如果使用 1:*n* 分区，则为索引缓存配置 2,000,000 个字节并为数据缓存配置 4,000,000 个字节。如果使用 *n*:*n* 分区，则为索引缓存配置 500,000 个字节并为数据缓存配置 1,000,000 个字节。

下表介绍了为计算联接器缓存大小所提供的输入：

输入	说明
主行数	主源中的行数。适用于未排序输入的联接器转换。主行数不会影响已排序的联接器转换的缓存大小。 注意: 如果主源中的行共享唯一键，则缓存计算器会过高估计索引缓存大小。
数据移动模式	集成服务的数据移动模式。缓存需求因数据移动模式而异。ASCII 字符使用一个字节。Unicode 字符使用两个字节。

输入输入内容，然后单击“计算”可计算数据和索引缓存大小。计算后的值显示在“数据缓存大小”和“索引缓存大小”字段中。

联接器缓存故障排除

使用本节中的信息可帮助对联接器转换的缓存进行故障排除。

使用缓存计算器计算具有已排序输入的联接器转换的缓存大小时，显示以下警告：

CMN_2020 Warning: If the master and detail pipelines of a sorted Joiner transformation are from the same source, the Integration Service cannot determine how fast it will process the rows in each pipeline. As a result, the cache size estimate may be inaccurate.

主管道和详细信息管道并发处理行。如果联接同一源中的数据，则管道处理行的速率可能会有所不同。如果一个管道处理其行的速度比其他管道处理的速度快，则集成服务会缓存所有已处理的行并将其保持缓存状态，直至其他管道处理完其行。缓存的行数取决于两个管道之间的处理速率差异。

缓存大小必须足够大才能存储所有已缓存行以实现最优会话性能。如果缓存大小不够大，请增加缓存大小。

注意: 不仅联接同一源中的数据时会显示此消息，联接不同源中的数据时也会显示此消息。

使用缓存计算器计算具有已排序输入的联接器转换的缓存大小时，显示以下警告：

CMN_2021 Warning: Increase the data cache size if the sorted Joiner transformation processes master rows that share the same key. To determine the new cache size, divide the existing cache size by 2.5 and multiply the result by the average number of master rows per key.

计算具有已排序输入的联接器转换的缓存大小时，缓存计算器会按估算的缓存需求为基准，即平均每个唯一键 2.5 个主行。如果每个唯一键的主行的平均数量大于 2.5，则相应地增加缓存大小。例如，如果每个唯一键的主行的平均数量为 5（2.5 的两倍），则需要将缓存计算器计算的缓存大小翻倍。

查找缓存

如果在查找转换中启用缓存，则集成服务会在内存中构建缓存以存储查找数据。集成服务在内存中构建查找缓存时，会先处理转换中的第一行的数据，然后对输入该转换的每一行的缓存进行查询。如果未启用缓存，则集成服务会对每个输入行的查找源进行查询。

无论您是否缓存查找源，查询查询和处理的结果均相同。但是，使用查找缓存可以提高会话性能。源较大时，可以通过缓存查找源来优化性能。

如果查找在会话之间未更改，则可以将转换配置为使用持久性查找缓存。运行会话时，如果任何缓存文件丢失或无效，集成服务均会重建持久性缓存。

集成服务会为查找转换创建以下缓存：

- **数据缓存。**对于已连接的查找转换，会存储已连接的输出口（不包括在查找条件中使用的端口）的数据。对于未连接的查找转换，则存储返回端口中的数据。
- **索引缓存。**存储在查找条件中使用的各列的数据。

集成服务会根据查找缓存和分区信息创建磁盘和内存缓存。

下表介绍了集成服务根据缓存和分区信息所创建的缓存：

查找条件	磁盘缓存	内存缓存
- 静态缓存 - 无哈希自动键分区点	为所有分区创建一个磁盘缓存。	为每个分区创建一个内存缓存。
- 动态缓存 - 无哈希自动键分区点	为所有分区创建一个磁盘缓存。	为所有分区创建一个内存缓存。
- 静态或动态缓存 - 哈希自动键分区点	为每个分区创建一个磁盘缓存。	为每个分区创建一个内存缓存。

在具有查找转换的会话中创建多个分区并在查找转换上创建哈希自动键分区点时，集成服务会使用缓存分区。

集成服务使用缓存分区时，如果任意分区的第一行达到查找转换的条件，集成服务便会为查找转换创建缓存。如果为并发缓存配置查找转换，则集成服务会并发构建分区的所有缓存。

共享缓存

集成服务会按不同的方式处理共享查找缓存，具体取决于该缓存是静态还是动态：

- **静态缓存。**如果两个查找转换共享一个静态缓存，则集成服务不会为同一管道阶段的共享转换分配额外内存。对于不同管道阶段的共享转换，集成服务不会分配额外内存。

使用相同数据或数据子集创建磁盘缓存的静态查找转换可以共享磁盘缓存。但是，查找键可能会有所不同，因此转换必须具有单独的内存缓存。

- **动态缓存。**查找转换共享动态缓存时，集成服务会更新内存缓存和磁盘缓存。为使缓存保持同步，集成服务必须在转换之间共享磁盘缓存和对应的内存缓存。

为查找转换配置缓存大小

可以在会话属性中为查找转换配置缓存大小。

下表介绍了为计算查找缓存大小所提供的输入：

输入	说明
具有唯一查找键的行数	在查找源中具有唯一查找键的行数。
数据移动模式	集成服务的数据移动模式。缓存需求因数据移动模式而异。ASCII 字符使用一个字节。Unicode 字符使用两个字节。

输入输入内容，然后单击“计算”可计算数据和索引缓存大小。计算后的值显示在“数据缓存大小”和“索引缓存大小”字段中。

等级缓存

集成服务使用高速缓存来处理等级转换。集成服务在完成等级操作之前会先将数据存储在等级内存中。

集成服务运行等级转换会话时，会将输入行与数据缓存中的行进行比较。如果输入行的等级高于存储行，则集成服务会将存储行替换为输入行。

例如，配置一个等级转换以查找前三位销售额。集成服务会读取以下输入数据：

```
SALES
10,000
12,210
5,000
2,455
6,324
```

集成服务会缓存前三行（10,000、12,210 和 5,000）。集成服务读取下一行（2,455）时，会将其与缓存值进行比较。由于该行等级低于缓存行，则会丢弃 2,455 这一行。但是，读取下一行（6,324）时，其等级高于任意一个缓存行。因此，集成服务会将缓存行替换为较高等级的输入行。

如果将等级转换配置为在多个组之间进行评级，则集成服务会对其查找的每个组增加等级。

集成服务会为等级转换创建以下缓存：

- **数据缓存。**根据分组端口存储等级信息。
- **索引缓存。**根据分组端口中的配置存储组值。

默认情况下，集成服务会为所有分区创建一个内存缓存和一个磁盘缓存。

如果为会话创建多个分区，则集成服务使用缓存分区。集成服务会为等级转换创建一个磁盘缓存并为每个分区创建一个内存缓存，然后根据转换的组键值将数据从一个分区路由至另一个分区。

为等级转换配置缓存大小

可以在会话属性中为等级转换配置缓存大小。

下表介绍了为计算等级缓存大小所提供的输入：

输入	说明
组数	组数。等级转换按组对数据进行评级。使用分组端口确定组数。例如，如果按存储 ID 和项目 ID 进行分组，具有 5 个存储和 25 个项目，且每个存储包含 25 个项目，则将组数计算为： $5 * 25 = 125 \text{ groups}$
等级数	评级中的项目数。例如，如果要对前 10 位销售额进行评级，应具有 10 个等级。缓存计算器会根据在等级转换中设置的值来填充此值。
数据移动模式	集成服务的数据移动模式。缓存需求因数据移动模式而异。ASCII 字符使用一个字节。Unicode 字符使用两个字节。

输入输入内容，然后单击“计算”可计算数据和索引缓存大小。计算后的值显示在“数据缓存大小”和“索引缓存大小”字段中。

排序器缓存

集成服务使用高速缓存来处理排序器转换。集成服务会在执行排序操作之前，将所有传入数据传递到排序器转换中。

集成服务对数据进行排序时，会创建一个排序器缓存以存储排序键和数据。默认情况下，集成服务会为所有分区创建一个内存缓存和一个磁盘缓存。

如果在会话中创建多个分区，则集成服务使用缓存分区。集成服务会为排序器转换创建一个磁盘缓存，并为每个分区创建一个内存缓存。集成服务会为每个分区创建一个单独的缓存，并单独对每个分区进行排序。

如果未配置缓存大小以对内存中的所有数据进行排序，则在会话日志中会显示一条警告，指示集成服务已对源数据执行多次传递。如果集成服务必须对磁盘的信息进行分页才能完成排序，则该服务会对数据进行多次传递。该消息用于指定单个传递所需的字节数，这也就意味着集成服务会一次性读取数据，并在未对磁盘进行分页的情况下执行内存中的排序操作。要增加会话性能，请配置缓存大小以便集成服务对数据执行一次传递。

为排序器转换配置缓存大小

可以在会话属性中为排序器转换配置排序器缓存。

下表介绍了为计算排序器缓存大小所提供的输入：

输入	说明
行数	行数。
数据移动模式	集成服务的数据移动模式。缓存需求因数据移动模式而异。ASCII 字符使用一个字节。Unicode 字符使用两个字节。

输入输入，然后单击“计算”可计算排序器缓存大小。计算后的值显示在“排序器缓存大小”字段中。

XML 目标缓存

集成服务使用高速缓存来创建 XML 目标。集成服务在生成 XML 目标时将数据和 XML 层次结构保存在高速缓存中。

集成服务为 XML 目标创建以下类型的缓存：

- **数据缓存。**在生成 XML 目标文档时存储 XML 行数据。为所有组存储一个数据缓存。
- **索引缓存。**存储主键或外键。为每个组创建一个主键索引缓存和一个外键索引缓存。

为 XML 目标配置缓存大小

您可以在目标或会话属性中为 XML 目标配置缓存大小。默认情况下，缓存大小设置为“自动”。集成服务可确定在运行时所需的缓存内存量。

还可以配置缓存大小并指定缓存内存量（以字节为单位）。完成以下步骤以计算缓存大小：

1. 估算每组中的行数。
2. 使用以下公式计算每组的缓存大小：

$$\text{Group cache size} = \text{Data cache size} + \text{Primary key index cache size} + \text{Foreign key index cache size}$$

3. 使用以下公式计算总缓存大小：

$$\text{总缓存大小} = \text{Sum}(\text{所有组的缓存大小})$$

以下方程式显示了如何计算组的数据缓存大小：

$$(\text{组中的行数}) \times (\text{组的行大小})$$

以下方程式显示了如何计算组的主键树大小：

$$(\text{组中的行数}) \times (\text{主键索引缓存大小})$$

以下方程式显示了如何计算组的外键树大小：

$$\text{Sum} ((\text{父组中的行数}) \times (\text{外键索引缓存大小}))$$

注意：无法使用缓存计算器为 XML 目标配置缓存大小。

优化缓存大小

为实现最优会话性能，请配置缓存大小，以便集成服务可以在未对磁盘进行分页的情况下处理内存中的转换。集成服务对磁盘进行分页时，会话性能会降低。

使用缓存计算器计算缓存大小时，缓存计算器会根据您的输入估算为实现最优会话性能所需的缓存大小。可以使用会话日志中指定的缓存大小调整估算值。运行会话后，请查看会话日志中的转换统计信息以获取缓存大小。

例如，运行一个名为 AGGTRANS 的汇总器转换。会话日志包含以下文本：

```
MAPPING> TT_11031 Transformation [AGGTRANS]:
MAPPING> TT_11114 [AGGTRANS]: Input Group Index = [0], Input Row Count [110264]
MAPPING> TT_11034 [SQ_V_PETL]: Input - 110264
MAPPING> TT_11115 [AGGTRANS]: Output Group Index = [0]
MAPPING> TT_11037 [FILTRANS]: Output - 1098,Dropped - 0
MAPPING> CMN_1791 The index cache size that would hold [1098] aggregate groups of input rows for
[AGGTRANS], in memory, is [286720] bytes
MAPPING> CMN_1790 The data cache size that would hold [1098] aggregate groups of input rows for
[AGGTRANS], in memory, is [1774368] bytes
```

该日志显示了索引缓存需要 286,720 个字节以及数据缓存需要 1,774,368 个字节才能在未对磁盘进行分页的情况下处理内存中的转换。

缓存大小根据会话或源数据的更改可能会有所不同。后续会话运行后查看会话日志可监视对缓存大小所做的更改。

在会话属性中必须将跟踪级别设置为“详细初始化”以允许集成服务将转换统计信息写入到会话日志。

注意: 会话日志不包含排序器、具有已排序输入的联接器转换、具有已排序输入的汇总器转换或 XML 目标的转换统计信息。

第 21 章

增量汇总

本章包括以下主题：

- [增量汇总概览, 264](#)
- [增量汇总的集成服务处理, 265](#)
- [重新初始化汇总文件, 265](#)
- [移动或删除汇总文件, 265](#)
- [增量汇总的分区准则, 266](#)
- [为增量汇总做准备, 267](#)

增量汇总概览

使用增量汇总时，在源中应用捕获的更改以汇总会话中的计算。如果源发生增量更改并且您可以捕获这些更改，则可以配置会话以处理这些更改。这将允许集成服务增量更新目标，而不是强制其在您每次运行会话时处理整个源并重新计算相同数据。

例如，可能有会话使用每天接收新数据的源。由于您已将筛选条件添加到从数据流中删除已有数据的映射，因此您可以捕获这些增量更改。随后启用增量汇总。

如果在 3 月 1 日第一次启用增量汇总的情况下运行会话，则可以使用整个源。这将允许集成服务读取和存储必需的汇总数据。在 3 月 2 日，再次运行会话时，可以筛选出除了时间戳为 3 月 2 日之外的所有记录。集成服务随后将处理新数据，并相应地更新目标。

在以下情况下，请考虑使用增量汇总：

- **可以捕获新源数据。**如果在您每次运行会话时都可以捕获新源数据，请使用增量汇总。使用存储过程或筛选器转换来处理新数据。
- **增量更改不会显著更改目标。**如果更改未显著更改目标，请使用增量汇总。如果处理增量更改的源更改了超过一半的现有目标，则会话可能不会从使用增量汇总中受益。在这种情况下，请删除该表并重新创建具有完整源数据的目标。

注意：如果映射包含 percentile 或 median 函数，请勿使用增量汇总。除了您在会话属性中配置的缓存内存之外，集成服务还使用系统内存处理这些函数。因此，集成服务不会在磁盘缓存中存储 percentile 或 median 函数的增量汇总值。

增量汇总的集成服务处理

第一次运行增量汇总会话时，集成服务会处理整个源。在会话结束时，集成服务会在索引文件和数据文件这两个文件中存储该会话运行中的汇总数据。集成服务会在汇总器转换属性中指定的缓存目录中创建这些文件。

以后每次使用增量汇总运行会话时，都会在会话中使用增量源更改。对于每个输入记录，集成服务都会检查对应组的索引文件中的历史信息。如果找到对应组，则集成服务会使用该组的汇总数据执行增量汇总操作，并保存增量更改。如果未找到对应组，则集成服务会创建新组并保存记录数据。

写入目标时，集成服务会将这些更改应用到现有目标。它会将修改的汇总数据保存在要在下次运行会话时用作历史数据的索引和数据文件中。

如果源发生显著更改，并且您希望集成服务继续保存将来增量更改的汇总数据，请将集成服务配置为使用新汇总数据覆盖现有汇总数据。

以后每次使用增量汇总运行会话时，集成服务都会创建增量汇总文件的备份。汇总器转换的缓存目录必须包含可用于两组文件的足够磁盘空间。

对使用增量汇总的会话进行分区时，集成服务会为每个分区都创建一组缓存文件。

执行以下任务之一时，集成服务会创建新汇总数据，而不是使用历史数据：

- 保存新版本的映射。
- 配置会话以重新初始化汇总缓存。
- 移动汇总文件，而不是在会话属性中更正文件的已配置路径或目录。
- 更改汇总文件的已配置路径或目录，而不是将文件移至新位置。
- 删除缓存文件。
- 减少分区数量。

集成服务重建增量汇总文件时，之前文件中的数据将丢失。

注意：要保护增量汇总文件免受文件损坏或磁盘故障影响，请定期备份这些文件。

重新初始化汇总文件

如果源表发生显著更改，您可能希望集成服务创建新汇总数据，而不是使用历史数据。要使集成服务创建新汇总数据，请配置会话以重新初始化汇总缓存。

例如，如果会话源每天增量更改并且每月完全更改，则可以重新初始化汇总缓存。接收该月的新源数据时，可以配置会话以重新初始化汇总缓存、截断现有目标和在会话期间使用新源表。

在运行重新初始化汇总缓存的会话后，编辑会话属性以禁用“重新初始化汇总缓存”选项。如果未清除重新初始化汇总缓存，则集成服务会在您每次运行会话时覆盖汇总缓存。

注意：从 Windows 移至 UNIX 时，必须重新初始化缓存。因此，无法从 Latin1 代码页更改为 MSLatin1 代码页，即使这些代码可兼容。

移动或删除汇总文件

在运行增量汇总会话后，请避免移动或修改存储历史汇总信息的索引和数据文件。

如果将这些文件移至其他目录中并且希望集成服务使用汇总文件，还必须在会话属性中更改这些文件的路径。此外，如果更改文件的路径但不移动文件，则集成服务会在您下次运行会话时重建文件。

如果更改某些会话或集成服务属性，集成服务将无法使用增量汇总文件，并且会使会话失败。要避免会话失败，请在执行以下任一任务时删除现有增量汇总文件：

- 将集成服务数据移动模式从 ASCII 更改为 Unicode，或从 Unicode 更改为 ASCII。
- 将集成服务代码页更改为不兼容的代码页。
- 当集成服务在 Unicode 模式下运行时，更改会话排序顺序。
- 更改“启用高精度”会话选项。

查找索引和数据文件

默认情况下，集成服务将索引和数据文件存储在 Workflow Manager 中的进程变量 \$PMCacheDir 中输入的目录中。集成服务将索引文件命名为 PMAGG*.idx*。集成服务将数据文件命名为 PMAGG*.dat*。

运行会话时，集成服务在会话日志中写入文件名。要找到这些文件，请在之前的会话日志中查找指示缓存文件名和位置的 SM_7034 和 SM_7035 消息。以下消息显示了会话日志中的示例条目：

```
MAPPING> SM_7034 Aggregate Information: Index file is [C:\Informatica\PowerCenter8.0\server\infa_shared
\Cache\PMAGG8_4_2.idx2]
MAPPING> SM_7035 Aggregate Information: Data file is [C:\Informatica\PowerCenter8.0\server\infa_shared
\Cache\PMAGG8_4_2.dat2]
```

增量汇总的分区准则

在具有多个分区的会话中使用增量汇总时，集成服务会为每个分区都创建一组缓存文件。

更改分区数量或缓存目录时，请遵循以下准则：

- **更改分区的缓存目录。**如果更改分区的目录并且希望集成服务重用缓存文件，必须移动与已更改目录关联的分区的缓存文件。
 - 如果更改第一个分区的目录，并且不移动缓存文件，则集成服务会为所有分区重建缓存文件。
 - 如果更改第 2-*n* 个分区的目录，并且不移动缓存文件，则集成服务会重建其无法找到的缓存文件。
- **减少分区数量。**如果删除分区并且希望集成服务重用缓存文件，必须将已删除分区的缓存文件移至为第一个分区配置的目录中。如果未将文件移至第一个分区的目录中，则集成服务会重建其无法找到的缓存文件。

注意：如果增加分区的数量，则集成服务会在您下次运行会话时重新调整索引和数据缓存文件。无需重建这些文件。
- **移动缓存文件。**如果移动分区的缓存文件并且希望集成服务重用这些文件，还必须更改分区目录。如果不更改目录，则集成服务会在您下次运行会话时重建这些文件。
- **删除缓存文件。**如果删除缓存文件，则集成服务会在您下次运行会话时重建这些文件。

如果更改分区的数量和缓存目录，可能需要为其移动缓存文件。例如，如果更改第一个分区的缓存目录并减少分区的数量，则需要移动已删除分区的缓存文件以及与已更改目录关联的分区的缓存文件。

为增量汇总做准备

使用增量汇总时，需要配置映射和会话属性：

- 实施映射逻辑或筛选器以删除已有的数据。
- 为会话配置增量汇总并验证文件目录是否具有用于汇总文件的足够磁盘空间。

配置映射

在启用增量汇总之前，必须捕获源数据中的更改。可以使用映射中的筛选器或存储过程转换在会话期间删除已有的源数据。

配置会话

为会话配置增量汇总时，请遵循以下准则：

- **确认要存储汇总文件的位置。**索引和数据文件增长与源数据成比例。确保缓存目录具有足够磁盘空间来存储会话的历史数据。

使用增量汇总运行多个会话时，决定要存储文件的位置。然后，为 Workflow Manager 中的进程变量 `$PMCacheDir` 输入相应目录。可以为索引和数据文件输入特定于会话的目录。但是，通过为使用增量汇总的所有会话使用进程变量，可在必要时通过更改 `$PMCacheDir` 轻松更改缓存目录。

更改缓存目录而不移动文件会导致集成服务重新初始化汇总缓存并收集新汇总数据。

在网格中，集成服务会重建其无法找到的增量汇总文件。当集成服务重建增量汇总文件时，会丢失汇总历史记录。

- **验证会话属性中的增量汇总设置。**可在“属性”选项卡上的“性能”设置中为会话配置增量汇总。
也可以配置会话以重新初始化汇总缓存。如果选择重新初始化缓存，则 Workflow Manager 会显示一条指示集成服务将覆盖现有缓存的警告以及一条在运行会话后清除此选项的提醒。

注意：如果映射包含具有事务转换范围的汇总器转换，则无法使用增量汇总。Workflow Manager 将会话标记为无效。

第 22 章

会话日志界面

本章包括以下主题：

- [会话日志界面概览, 268](#)
- [实现会话日志界面, 268](#)
- [会话日志界面中的函数, 269](#)
- [会话日志界面示例, 272](#)

会话日志界面概览

默认情况下，集成服务会将会话事件写入到运行服务进程所在的节点上的二进制日志文件中。此外，集成服务还可以将会话事件信息传递到外部库。在外部共享库中，可以提供集成服务如何写入日志事件的过程。

PowerCenter 可通过会话日志界面访问会话事件信息。创建共享库时，可以实现会话日志界面中提供的函数。

集成服务写入会话事件时，会调用会话日志界面中指定的函数。您创建的共享库中的函数必须与会话日志界面中定义的函数签名相匹配。

实现会话日志界面

要将集成服务配置为使用自定义过程来处理会话事件信息，请完成以下步骤：

1. 创建一个可以实现会话日志界面的共享库。
2. 在 Administrator 工具中配置集成服务属性时，请将 ExportSessionLogLibName 属性设置为您创建的共享库的名称。

集成服务和会话日志界面

将集成服务的 ExportSessionLogLibName 属性设置为共享库的名称时，集成服务除创建事件日志文件之外，还会执行在共享库中定义的过程。

集成服务将按如下方式使用共享库：

1. 集成服务在记录会话中的第一个事件之前，会先加载共享库并调用 INFA_InitSessionLog() 函数。
2. 每次集成服务将事件记录到会话日志文件时，均会调用 INFA_OutputSessionLog() 函数以便将消息、代码和会话信息传递到共享库。
3. 会话完成并记录完最后一个事件后，集成服务会调用 INFA_EndSessionLog()，然后卸载共享库。

为确保集成服务能够正确调用共享库，请遵守写入共享库的准则。

实现会话日志界面的规则和准则

写入代码以实现会话日志界面时，请遵守以下规则和准则：

- 必须实现会话日志界面中的所有函数。
- 除异常终止之外，从集成服务到会话日志界面中的函数的所有调用均应序列化。集成服务将事件记录到会话日志中时，会对这些函数执行调用。因此，实现会话日志界面中的函数时，并不需要使用 mutex 对象，从而确保一次仅有一个线程执行代码段。
- 在 UNIX 中实现会话日志界面时，请勿在函数内执行任何信号处理。这样可确保函数不会与 PowerCenter 处理信号的方式互相干扰。请勿注册或取消注册任何信号处理程序。
- 由于集成服务属于多线程进程，因此必须将共享库编译为多线程库才能进行正常加载。

会话日志界面中的函数

会话日志界面中的函数不会返回值。因此，会话不会因为集成服务调用会话日志界面中的函数而失败。

下表介绍了会话日志界面中的函数：

函数	说明
INFA_InitSessionLog	提供有关集成服务将向其中写入事件日志的会话的信息。
INFA_OutputSessionLogMsg	每次记录事件时均调用。传递事件的相关信息。
INFA_OutputSessionLogFatalMsg	在异常终止前记录上一个事件时调用。
INFA_EndSessionLog	上一条消息发送至会话日志且会话正常终止后调用。
INFA_AbnormalSessionTermination	上一条消息发送至会话日志且会话异常终止后调用。

本节介绍的函数使用在标准 C 表头文件 *time.h* 中声明的时间结构。函数同时会假定以下声明：

```
typedef int          INFA_INT32;
typedef unsigned int  INFA_UINT32;
typedef unsigned short INFA_UNICHAR;
typedef char          INFA_MBCSCHAR;
typedef int           INFA_MBCS_CODEPAGE_ID;
```

INFA_InitSessionLog

```
void INFA_InitSessionLog(void ** dllContext,
    const INFA_UNICHAR * sServerName,
    const INFA_UNICHAR * sFolderName,
    const INFA_UNICHAR * sWorkflowName,
```

```
const INFA_UNICHAR * sessionHierName[]);
```

集成服务在写入任何会话日志事件之前，会先调用 INFA_InitSessionLog 函数。传递至此函数的参数会提供有关集成服务向其中写入事件日志的会话的信息。

INFA_InitSessionLog 具有以下参数：

参数	数据类型	说明
dllContext	未指定	用户定义的特定于共享库的信息。在后续函数调用中，会将此参数传递给所有函数。可以使用此参数存储与网络连接相关的信息或分配在处理会话日志输出期间所需的内存。共享库必须分配和解除分配与此参数关联的任何内存。
sServerName	无符号短整型	运行会话的集成服务的名称。
sFolderName	无符号短整型	包含会话的文件夹的名称。
sWorkflowName	无符号短整型	与会话关联的工作流的名称
sessionHierName[]	无符号短数组	包含会话层次结构的数组。该数组包括会话所属的存储库、工作流和工作集（如果有）。数组的大小除以指示器的大小所得到的值等于数组元素数。

INFA_OutputSessionLogMsg

```
void INFA_OutputSessionLogMsg(  
    void * dllContext,  
    time_t curTime,  
    INFA_UINT32 severity,  
    const INFA_UNICHAR * msgCategoryName,  
    INFA_UINT32 msgCode,  
    const INFA_UNICHAR * msg,  
    const INFA_UNICHAR * threadDescription);
```

每次集成服务记录事件时均会调用此函数。传递至该函数的参数包括日志事件消息的不同元素。可以使用这些参数自定义日志输出的格式或筛选消息。

INFA_OutputSessionLogMsg 具有以下参数：

参数	数据类型	说明
dllContext	未指定	用户定义的特定于共享库的信息。可以使用此参数存储与网络连接相关的信息或分配在处理会话日志输出期间所需的内存。共享库必须分配和解除分配与此参数关联的任何内存。
curTime	time_t	集成服务记录事件的时间。

参数	数据类型	说明
severity	无符号整形	用于指示日志事件消息类型的代码。事件日志使用以下严重性代码： 32: 调试消息 8: 信息性消息 2: 错误消息 4: 警告消息
msgCategoryName	常量 无符号短整型	用于指示日志事件消息类别的代码前缀。 在以下示例消息中，字符串 <i>BLKR</i> 是在 msgCategoryName 参数中传递的值。 READER_1_1_1> BLKR_16003 Initialization completed successfully.
msgCode	无符号整形	用于标识日志事件消息的数字。 在以下示例消息中，字符串 <i>16003</i> 是在 msgCode 参数中传递的值。 READER_1_1_1> BLKR_16003 Initialization completed successfully.
msg	常量 无符号短整型	日志事件消息的文本。 在以下示例消息中，字符串 <i>Initialization completed successfully</i> 是在 msg 参数中传递的值。 READER_1_1_1> BLKR_16003 Initialization completed successfully.
threadDescription	常量 无符号短整型	用于指示线程生成事件日志的代码。 在以下示例消息中，字符串 <i>READER_1_1_1</i> 是在 threadDescription 参数中传递的值。 READER_1_1_1> BLKR_16003 Initialization completed successfully.

INFA_OutputSessionLogFatalMsg

```
void INFA_OutputSessionLogFatalMsg(void * dllContext, const char * msg);
```

在异常终止之前，集成服务会调用此函数记录上一个事件。在集成服务代码页中，参数消息是指 MBCS 字符。

在 UNIX 中实现此函数时，请确保仅从此函数内部调用异步信号安全函数。

INFA_OutputSessionLogFatalMsg 具有以下参数：

参数	数据类型	说明
dllContext	未指定	用户定义的特定于共享库的信息。可以使用此参数存储与网络连接相关的信息或分配在处理会话日志输出期间所需的内存。共享库必须分配和解除分配与此参数关联的任何内存。
msg	常量字符	错误消息的文本。通常，这些消息属于声明错误消息或操作系统错误消息。

INFA_EndSessionLog

```
void INFA_EndSessionLog(void * dllContext);
```

在上一条消息发送至会话日志且会话正常终止后，集成服务会调用此函数。可以使用此函数执行清理操作并释放内存和资源。

INFA_EndSessionLog 具有以下参数：

参数	数据类型	说明
dllContext	未指定	用户定义的特定于共享库的信息。可以使用此参数存储与网络连接相关的信息或分配在处理会话日志输出期间所需的内存。共享库必须分配和解除分配与此参数关联的任何内存。

INFA_AbnormalSessionTermination

```
void INFA_AbnormalSessionTermination(void * dllContext);
```

在上一条消息发送至会话日志且会话异常终止后，集成服务会调用此函数。集成服务会在调用 INFA_OutputSessionLogFatalMsg 函数之后调用此函数。如果集成服务调用此函数，则不会调用 INFA_EndSessionLog。

例如，DTM 中止或超时，集成服务会调用此函数。在 UNIX 中，当发生信号异常时，集成服务会调用此函数。实现此函数时，仅能实现最低功能。在 UNIX 中，请确保仅从此函数内部调用异步信号安全函数。

INFA_AbnormalSessionTermination 具有以下参数：

参数	数据类型	说明
dllContext	未指定	用户定义的特定于共享库的信息。可以使用此参数存储与网络连接相关的信息或分配在处理会话日志输出期间所需的内存。共享库必须分配和解除分配与此参数关联的任何内存。

会话日志界面示例

Informatica 提供了一个使用会话日志界面的示例程序。该示例程序会将会话日志事件发送到名为 *sesslog.log* 的文本文件中。可以查看示例程序获得有关如何根据您的需求使用会话日志界面处理会话日志事件的详细信息。还可以编译示例程序并构建外部库以将会话日志事件发送至文本文件中。

通过 Informatica Development Platform 安装程序安装 PowerCenter SDK 文件时，可以使用会话日志示例程序。默认情况下，会话日志示例程序安装在以下目录中：

```
<SDKInstallationDir>/SessionLog_API/samples
```

构建外部会话日志库

使用随示例程序 demo_sesslog.cpp 提供的 make 文件来构建外部库。编译库的命令取决于构建该库的平台。

在 UNIX 中构建库

下表显示了在不同的平台上构建库的命令：

平台	编译程序	命令
Solaris	CC	make -f makefile_sol
HP-UX	aCC	make -f makefile_hpx
HP-UX 64 位	aCC	make -f makefile_hpx64
AIX	xCl_r	make -f makefile_aix
AIX 64 位	xCl_r	make -f makefile_aix64
Linux	g++	make -f makefile_linux

在 Windows 中构建库

在 Windows 中可使用 Microsoft Visual C++ 6.0 构建示例会话日志库。在 Visual C++ 6.0 中打开示例程序 demo_sesslog.dsw，然后构建项目。

使用外部会话日志库

构建库之后，可以使用该库将会话日志的输出写入到文件中。

要使用示例外部会话日志库，请完成以下步骤：

1. 登录到 Administrator 工具，然后选择您要设置会话日志文件的集成服务。
2. 在集成服务的“属性”选项卡上，编辑配置属性。
3. 将 ExportSessionLogLibName 属性设置为从会话日志示例程序创建的会话日志库的路径和文件名。

第 23 章

了解缓冲区内内存

本章包括以下主题：

- [了解缓冲区内内存概览, 274](#)
- [自动缓冲区内内存设置, 275](#)
- [配置缓冲区内内存, 275](#)
- [配置会话缓存内存, 275](#)

了解缓冲区内内存概览

运行会话时，集成服务进程会启动 Data Transformation Manager (DTM)。DTM 根据会话属性中的“DTM 缓冲区大小”设置在运行时为会话分配缓冲区内内存。

DTM 根据会话属性的“默认缓冲区块大小”设置中的配置将内存划分为多个缓冲区块。读取器、转换和写入器线程使用缓冲区块将数据从源移至目标。缓冲区块大小应大于源或目标中最大数据行的精度。

集成服务为分区中的每个源和目标至少分配两个缓冲区块。对于 XML 源和目标，缓冲区块数必须至少是源和目标中组数的两倍。具有未规范化列的 XML 读取器和具有循环引用的 XML 架构可能需要其他缓冲区块。

可以通过调整以下会话属性配置缓冲区内内存设置：

DTM 缓冲区大小

DTM 缓冲区大小指定了集成服务在 DTM 处理会话时使用的缓冲区内内存量。在会话属性中的“属性”选项卡上配置 DTM 缓冲区大小。

默认缓冲区块大小

缓冲区块大小指定了用于将数据块从源移至目标的缓冲区内内存量。在会话属性中的“配置对象”选项卡上配置缓冲区块大小。

集成服务会为缓冲区内内存和缓冲区块计算最小内存分配。默认情况下，集成服务为每个块分配 64,000 字节，或分配映射中任何源或目标的最大行大小（取较大者）。

如果 DTM 无法为会话分配配置的缓冲区内内存量，则会话无法初始化。通常，无需为缓冲区内内存分配 1 GB 以上的内存量。

您可以手动设置缓冲区大小的值，也可以将会话配置为允许集成服务确定会话所需的缓冲区内内存大小。

自动缓冲区内内存设置

在初始化时，DTM 会分配会话将在运行时使用的缓冲区内内存。您可以将集成服务配置为自动分配缓冲区内内存大小，也可以设置内存缓冲区和块大小。

默认情况下，PowerCenter 集成服务根据转换要求和映射中的源和目标自动计算会话所需的缓冲区内内存。此计算不基于主机上的内存量或该内存的可用量。在某些情况下，PowerCenter 集成服务可能仅将一小部分可用内存分配给会话。

如果由 PowerCenter 集成服务自动计算的缓冲区内内存似乎并未生成会话所需的内存占用量，则可以指定 PowerCenter 集成服务分配给会话的缓冲区内内存大小和块大小。

相关主题：

- [“会话缓存概览” 页面上 248](#)

将会话配置对象用于内存配置

可以使用会话配置对象为多个会话配置内存设置。可以为每个会话配置对象设置不同的内存设置。

存储库中的每个文件夹有一个默认的会话配置对象，其中包含提交和加载设置、日志选项和错误处理设置等会话属性。创建会话时，Workflow Manager 会将默认的配置对象设置应用于会话。您也可以选择配置对象以用于会话。

如果您希望为多个会话应用不同的配置设置，可以创建多个配置对象。例如，当从测试环境迁移至生产环境时，或您拥有多个具有不同自动内存要求的会话时，可以在会话配置对象中配置内存设置。

配置缓冲区内内存

集成服务可以确定会话的内存要求，或者您可以手动设置 DTM 缓冲区大小和默认的缓冲区块大小。

可以在会话属性中配置缓冲区内内存设置。

1. 打开会话，然后单击“配置对象”选项卡。
2. 为“默认缓冲区块大小”输入值。

您可以指定“自动”或一个数字值。

默认单位为字节。为值附加 KB、MB 或 GB 可指定其他单位。例如，1048576 或 1024KB 或 1MB。

3. 单击属性选项卡。
4. 为 DTM 缓冲区大小输入值。

您可以指定“自动”或一个数字值。如果会话所需的内存大于为 DTM 缓冲区大小设置的内存，则会话性能会降低且会话可能会失败。

默认单位为字节。为值附加 KB、MB 或 GB 可指定其他单位。例如，1048576 或 1024KB 或 1MB。

配置会话缓存内存

集成服务可以确定以下会话缓存的内存要求：

- 查找转换索引和数据缓存

- 汇总器转换索引和数据缓存
- 等级转换索引和数据缓存
- 联接器转换索引和数据缓存
- 排序器转换缓存
- XML 目标缓存

您可以在转换属性中或在会话属性的映射选项卡上将索引和数据缓存大小配置为自动。

会话缓存限制

集成服务使用会话缓存作为内存缓存。您可以为会话缓存设置自动模式的转换分配内存。您可以在会话缓存中限制内存量。限制会话缓存时，可以限制集成服务可用于会话的内存量，以便为其他进程保留一些内存。

必须将缓存内存限制同时指定为数值和总内存百分比。集成服务将百分比值基于运行集成服务的计算机的总物理内存。

集成服务将数值与百分比值相比较，以确定哪个值更小。它使用较小值作为总内存分配给会话缓存。

会话配置对象的以下属性针对为会话分配的内存缓存设置了限制：

自动内存属性允许使用的内存上限

要为会话缓存分配的内存量。自动缓存内存总量不能超过此属性的值，即使所用计算机内存的百分比小于“自动内存属性允许使用的总内存百分比上限”属性中的值。当会话在具有大量物理内存的计算机上运行时，可能会发生这种情况。

自动内存属性允许使用的总内存百分比上限

用于分配给会话缓存的计算机内存的百分比。自动缓存内存总量不能超过此百分比，即使“自动内存属性允许使用的内存上限”属性中的值更大。当会话在具有极少物理内存的计算机上运行时，可能会发生这种情况。

集成服务会将会话缓存中的内存分配到高速缓存设置为自动的所有转换中。它在所有转换缓存之间划分内存。

例如，您在会话中为三个查找转换配置自动缓存。然后，为会话配置内存缓存限制 500 MB。运行会话时，集成服务会在所有三个查找转换的索引和数据缓存之间划分 500 MB 的已分配内存。会话的内存缓存限制不适用于未配置为自动缓存的转换。

如果将会话缓存设置为自动分配，则对于设置为自动缓存分配的每个转换，集成服务将至少为其索引缓存分配 1 MB，而至少为其数据缓存分配 2 MB。如果会话缓存限制没有为最小索引和数据缓存分配提供足够内存，则集成服务将替代缓存限制并为索引和数据缓存分配最小内存量。

例如，将会话缓存限制为 4 MB 内存，将两个转换设置为自动缓存分配。集成服务将替代会话缓存限制，并针对设置为自动缓存分配的每个转换至少为其索引缓存分配 1 MB，而至少为其数据缓存分配 2 MB。分配给转换缓存的总内存量为 6 MB。

在网格上运行会话并配置“自动内存属性允许使用的内存上限”时，集成服务会在网格中的所有节点之间划分已分配的内存缓存。配置“自动内存属性允许使用的总内存百分比上限”时，集成服务会将指定百分比的内存缓存分配给网格中的每个节点。

配置会话缓存的自动内存设置

要配置会话缓存的自动内存设置，请执行以下操作：

1. 在 Transformation Developer 中或在会话属性的“映射”选项卡中打开转换。

2. 在转换属性中，为以下缓存大小设置选择或输入“自动”：
 - 索引和数据缓存
 - 排序器缓存
 - XML 缓存
3. 在任务开发程序或 Workflow Designer 中打开会话，然后单击“配置对象”选项卡。
4. 为“自动内存属性允许使用的内存上限”输入值。

该值指定了要用于会话缓存的最大内存量。

默认单位为字节。为值附加 KB、MB 或 GB 可指定其他单位。例如，1048576 或 1024KB 或 1MB。
5. 为“自动内存属性允许使用的总内存百分比上限”输入值。

该值指定了会话缓存可能使用的总内存百分比上限。

第 24 章

高精度数据

本章包括以下主题：

- [高精度数据概览, 278](#)
- [长整型, 278](#)
- [小数, 279](#)

高精度数据概览

高精度数据确定了多少数字用更高的准确性表示。归属于数字的精度包括该数字的小数位数。例如，值 11.47 的精度为 4，小数位数为 2。大量数字都会因为在用于产生溢出的计算时进行舍入而失去准确性。由于无法截断高精度数据，可能会出现不正确的结果。

高精度数据值更准确。如果您需要准确的值，则启用高精度。

您可以在会话的属性选项卡上启用高精度。集成服务会以不同的方式针对长整型值和小数值处理高精度数据。

长整型

在可以生成小数值的计算中，集成服务会将长整型值处理为双精度或小数。会话包含可以生成小数值的计算且在未启用高精度的情况下运行时，集成服务会在执行计算之前将长整型值转换为双精度。转换双精度数据类型支持最多 15 位精度，而长整型数据类型支持最多 19 位精度。因此，在生成精度高于 15 位的长整型值的计算中，精度可能会有损失。

例如，表达式转换包含以下计算：

```
POWER( BIGINTVAL, EXPVAL )
```

在执行计算前，集成服务将 POWER 函数的输入转换为双精度值。如果 BIGINTVAL 端口包含长整型值 9223372036854775807，则集成服务将该值转换为 9.22337203685478e+18，失去了最后四位的精度。如果 EXPVAL 端口包含值 1.0，结果端口为长整型，该计算将产生一个行错误，因为结果 9223372036854780000 超出了最大长整型值。

如果在可以生成小数值的计算中使用长整型值，并且在启用高精度的情况下运行会话，集成服务会将长整型值转换为小数。转换小数数据类型支持最多 28 位精度。因此，除非结果生成精度大于 28 位的值，否则计算中不会出现精度损失的情况。在本例中，集成服务将结果存储为双精度。

小数

在未启用高精度的情况下运行会话时，集成服务会将小数值转换为双精度。转换小数数据类型支持最多 28 位精度，而双精度数据类型支持最多 15 位精度。因此，如果小数值的精度超出 15 位，精度将有损失。

例如，假设有一个具有小数 (20,0) 数据类型的映射传递数字 40012030304957666903。如果在启用高精度的情况下会话未运行，集成服务会将小数值转换为双精度，并传递 $4.00120303049577 \times 10^{19}$ 。

为保证最多 28 位精度，请使用小数数据类型，并在会话属性中启用高精度。在启用高精度的情况下运行会话时，集成服务会将小数值处理为小数。除非结果生成精度大于 28 位的值，否则计算中不会出现精度损失的情况。在本例中，集成服务将结果存储为双精度。

索引

符號

{Teradata 中的下推优化
派生表 [85](#)
\$PMStorageDirPMStorageDir
会话运行状态[pm storage dir] [145](#)

A

ABORT 函数
会话失败 [161](#)
\$AppConnection
使用 [192](#)

B

\$BadFile
命名规则 [192](#)
使用 [192](#)
保留绝对输入行顺序
会话属性 [37](#)
保留相对输入行顺序
会话属性 [37](#)
变量值
跨分区计算 [22](#)
表名称
符合下推兼容性要求 [68](#)
空闲数据库的语法 [68](#)
标头
在文件目标中创建 [39](#)
参数和变量类型 [200](#)
表头命令
平面文件目标 [39](#)
表头选项
平面文件目标 [39](#)
表尾
在文件目标中创建 [39](#)
参数和变量类型 [200](#)
表尾命令
平面文件目标 [39](#)
并发读取分区
会话属性 [34](#)
并发工作集
说明 [170](#)
并发工作流
查看日志 [169](#)
从命令行启动 [168](#)
从命令行停止 [169](#)
规则和准则 [170](#)
计划 [170](#)
配置步骤 [167](#)
配置为使用相同的名称运行 [165](#)
配置唯一实例 [164](#)
“启动高级工作流”选项 [168](#)

并发工作流 (续)
“启动工作流”选项 [168](#)
启动和停止 [168](#)
使用不同的会话参数文件 [211](#)
使用参数 [166](#)
说明 [164](#)
添加实例名称 [167](#)
通过 pmcmd 创建工作流实例 [168](#)
运行 Web 服务工作流 [165](#)
在 Workflow Monitor 中查看 [169](#)
转换限制 [170](#)
并发合并
文件目标 [40](#)
并发连接
在已分区的管道中 [38](#)
不可重用会话
缓存 [252](#)
变量
\$PMWorkflowRunId [167](#)
\$PMWorkflowRunInstanceName [167](#)
工作流 [181](#)
类型概览 [200](#)
接受变量的输入字段 [200](#)
在参数文件中定义 [200](#)
表达式
参数和变量类型 [200](#)
表达式转换
下推优化 [95](#)
表名称前缀
关系错误日志, 长度限制 [213](#)
关系错误日志, 参数和变量类型 [200](#)
目标, 参数和变量类型 [200](#)
表所有者名称
参数和变量类型 [200](#)

C

参数文件
标题 [208](#)
部分 [208](#)
参数和变量的范围 [208](#)
参数和变量类型 [200](#)
创建时遵循的准则 [208, 213](#)
创建提示 [215](#)
对工作流使用 [211](#)
对会话使用 [211](#)
概览 [199](#)
故障排除 [214](#)
会话参数文件名, 变量类型 [200, 211](#)
结构 [208](#)
空值, 输入 [210](#)
名称, 配置 [210](#)
模板文件 [207](#)
配置并发工作流实例 [166](#)
日期时间格式 [213](#)

- 参数文件 (续)
 - 示例参数文件 [210](#)
 - 使用变量指定 [211](#)
 - 说明 [199](#)
 - 替代连接属性 [207](#)
 - 通过 pmcmd 使用 [212](#)
 - 位置, 配置 [210](#)
 - 用例 [213](#)
 - 优先级 [212](#)
 - 指定要使用的 [199](#)
 - 注释, 添加 [209](#)
 - 定义属性 [200](#)
 - 接受参数和变量的输入字段 [200](#)
- 长整型
 - 高精度处理 [278](#)
- 查找
 - 持久性缓存 [259](#)
- 查找转换
 - 缓存 [259](#)
 - 缓存分区 [47, 254, 259](#)
 - 缓存计算器的输入 [259](#)
 - 配置缓存 [259](#)
 - 添加到并发工作流 [170](#)
 - 连接信息, 参数和变量类型 [200](#)
 - 下推优化 [98](#)
 - 源文件, 参数和变量类型 [200](#)
- 持久性变量
 - 定义 [187](#)
 - 在工作集中 [189](#)
- 重新初始化
 - 汇总缓存 [265](#)
- 重新启动任务恢复策略
 - 说明 [151, 152](#)
- 传递分区类型
 - 处理 [60](#)
 - 概览 [50](#)
 - 说明 [20](#)
 - 下推优化 [90](#)
 - 性能 [60](#)
- 创建
 - 错误日志表 [136](#)
 - FTP 会话 [243](#)
 - 工作流变量 [187](#)
 - 数据文件目录 [267](#)
 - 索引目录 [267](#)
 - 已分区的源的文件列表 [34](#)
- 从上一检查点继续
 - 恢复策略 [151, 152](#)
- ConnectionParam.prm 文件
 - 使用 [207](#)
- CPU 数
 - 为动态分区设置 [21](#)
- CUME 函数
 - 分区约束 [49](#)
- 错误
 - 阈值 [161](#)
 - 致命 [161](#)
- 错误处理
 - 错误日志文件 [140](#)
 - 概览 [161](#)
 - PMEError_MSG 表架构 [138](#)
 - PMEError_ROWDATA 表架构 [136](#)
 - PMEError_Session 表架构 [139](#)
 - 事务控制 [128](#)
 - 下推优化 [83](#)
 - 选项 [142](#)
- 错误日志
 - 概览 [135](#)
- 错误日志 (续)
 - 会话错误 [161](#)
 - 选项 [142](#)
- 错误日志表
 - 创建 [136](#)
 - 概览 [136](#)
- 错误日志文件
 - 表名称前缀长度限制 [213](#)
 - 概览 [140](#)
 - 名称, 参数和变量类型 [200](#)
 - 目录, 参数和变量类型 [200](#)
- 错误消息
 - 外部加载器 [222](#)
- 错误阈值
 - 出错时停止 [161](#)
 - 管道分区 [161](#)
 - 变量类型 [200](#)
- 参数
 - 类型概览 [200](#)
 - 数据库连接 [195](#)
 - 会话 [192](#)
 - 接受参数的输入字段 [200](#)
 - 在参数文件中定义 [200](#)
- 操作状态
 - 会话恢复 [145](#)
 - 检查点 [145, 152](#)
 - 工作流恢复 [145](#)
- 查找 SQL 替代选项
 - 参数和变量类型 [200](#)
- 查找缓存
 - 说明 [259](#)
 - 文件名前缀, 参数和变量类型 [200](#)
- 查找数据库
 - 数据库连接会话参数 [192](#)
- 查找文件
 - 查找文件会话参数 [192](#)
- 查找源文件
 - 使用参数 [192](#)
- 持久订阅名称
 - JMS 的变量类型 [200](#)
- 存储过程转换
 - 调用文本, 参数和变量类型 [200](#)
 - 连接信息, 参数和变量类型 [200](#)

D

- 代码页
 - 外部加载器文件 [221](#)
- 弹性
 - 实时会话 [119](#)
- \$DBConnection
 - 命名规则 [192](#)
 - 使用 [192](#)
- 等级缓存
 - 说明 [260](#)
- 等级转换
 - 缓存 [260](#)
 - 缓存分区 [254, 260](#)
 - 缓存计算器的输入 [260](#)
 - 配置缓存 [260](#)
 - 使用分区点 [30](#)
- 电子邮件任务
 - 挂起电子邮件 [149](#)
 - 变量类型 [200](#)
- 动态分区
 - 根据 CPU 数 [21](#)
 - 根据分区数 [21](#)

- 动态分区 (续)
 - 根据网格中的节点数 [21](#)
 - 规则和准则 [21](#)
 - 禁止 [21](#)
 - 使用源分区 [21](#)
 - 说明 [20](#)
 - 性能 [20](#)
 - 与分区类型结合使用 [21](#)
 - 分区数量, 参数类型 [200](#)
- DTM (Data Transformation Manager)
 - 缓冲区大小 [275](#)
- DTM 缓冲区大小要求
 - 配置 [275](#)
- 多种输入组转换
 - 创建分区点 [30](#)
- 多字节数据
 - Oracle 外部加载器 [229](#)
 - Sybase IQ 外部加载器 [230](#)
 - Teradata FastExport [217](#)
- 多组转换
 - 分区 [19](#)
- 读取器
 - 为 Teradata FastExport 选择 [218](#)
- 读取器时间限制
 - 配置 [111](#)
- \$DynamicPartitionCount
 - 说明 [192](#)
- 电子邮件
 - 挂起, 变量类型 [200](#)
 - 会话后, 参数和变量类型 [200](#)
- 端点 URL
 - Web 服务, 参数和变量类型 [200](#)
 - 参数和变量类型 [200](#)
- 队列连接
 - 参数类型 [200](#)
 - 会话参数 [192](#)

F

- 非持久变量
 - 定义 [187](#)
- 分配任务
 - 变量 [181](#), [200](#)
- 分区点
 - 编辑 [24](#)
 - 查找转换 [46](#)
 - 概览 [18](#)
 - HTTP 转换 [40](#), [41](#)
 - Java 转换 [40](#), [41](#)
 - 联接器转换 [42](#)
 - 添加和删除 [29](#)
 - 自定义转换 [40](#), [41](#)
 - 添加, 步骤 [25](#)
- 分区键
 - 规则和准则 [59](#)
 - 具有空值的行 [59](#)
 - 添加 [56](#), [58](#)
 - 添加键范围 [59](#)
- 分区级别属性
 - 说明 [22](#)
- 分区类型
 - 传递 [60](#)
 - 概览 [20](#)
 - 键范围 [57](#)
 - 默认值 [51](#)
 - 设置 [51](#)
 - 说明 [50](#)

- 分区类型 (续)
 - 性能 [51](#)
 - 循环 [61](#)
 - 与分区点一起使用 [51](#)
 - 更改 [25](#)
- 分区约束
 - 分区数 [23](#)
 - 关系目标 [38](#)
 - Informix [38](#)
 - 数值函数 [49](#)
 - Sybase IQ [38](#)
 - 未连接的转换 [30](#)
 - XML 目标 [49](#)
 - XML 生成器 [49](#)
 - 转换 [49](#)
- 分区组
 - 阶段 [174](#)
 - 说明 [174](#)
- FTP
 - 创建会话 [243](#)
 - 对目标进行分区 [246](#)
 - 访问目标文件 [244](#)
 - 访问源文件 [244](#)
 - 连接到文件目标 [38](#)
 - SFTP [242](#)
 - 概览 [241](#), [242](#)
 - 远程目录, 参数和变量类型 [200](#)
 - 远程文件名, 参数和变量类型 [200](#)
- FTP 连接
 - 参数类型 [200](#)
 - 会话参数 [192](#)
 - 密码, 参数类型 [200](#)
 - 用户名参数类型 [200](#)
- \$FTPConnection
 - 使用 [192](#)
- 服务变量
 - 在参数文件中 [200](#)
- 服务级别
 - 分配给任务 [178](#)
- 服务进程变量
 - 在参数文件中 [200](#)
- 负载均衡器
 - 工作流设置 [178](#)
 - 为任务分配优先级 [178](#)
 - 向任务分配资源 [179](#)
- 分区
 - 对多个目标使用 FTP [243](#)
 - 管道查找源表 [47](#)
 - 合并目标数据 [40](#)
 - 会话属性 [39](#)
 - 联接器转换 [256](#)
 - 缩放 [20](#)
 - 通过 XML 生成器 [49](#)
 - 为下推优化合并 [90](#)
 - 性能 [61](#)
 - 增量汇总 [266](#)
 - 删除 [25](#)
 - 输入说明 [25](#)
 - 说明 [18](#)
 - 添加 [25](#)
- 分区计数
 - 会话参数 [192](#)
- 分区名称
 - 设置 [25](#)
- 分区数
 - 为动态分区设置 [21](#)
 - 概览 [18](#)
 - 会话参数 [192](#)

分区数 (续)
性能 [18](#)

G

高精度

Bigint 数据类型 [278](#)
处理 [278](#)
小数数据类型 [278](#)

更新

增量 [267](#)

根据 CPU 数

设置 [21](#)

根据分区数

设置 [21](#)

共享库

实现会话日志界面 [269](#)

工作集

变量 [189](#)
参数选项卡 [189](#)
持久性变量 [189](#)
持久性变量示例 [189](#)
分配前期工作集和后期工作集变量 [189](#)
分配前期工作集和后期工作集变量, 过程 [190](#), [198](#)
添加到并发工作流 [170](#)
替代变量值 [189](#)
在对象之间传递信息 [189](#)
在对象之间传递信息, 示例 [190](#)

工作集变量

将值传入和传出会话 [197](#)
在参数文件中 [200](#)
在工作集之间传递值 [189](#)

工作流变量

持久性变量 [187](#)
创建 [187](#)
非持久变量 [187](#)
关键字 [181](#)
将值传入和传出工作集 [189](#)
将值传入和传出会话 [197](#)
命名约定 [187](#)
默认值 [182](#), [187](#)
内置变量 [182](#)
起始值和当前值 [187](#)
日期时间格式 [187](#)
使用 [181](#)
数据类型 [182](#), [187](#)
用户定义 [186](#)
预定义 [182](#)
在表达式中使用 [184](#)
在参数文件中 [200](#)

工作流管理器

在网络上运行工作流 [172](#)
在网络上运行会话 [172](#)

工作流日志文件

查看并发工作流 [169](#)

工作流实例

从命令行启动 [168](#)
动态创建 [168](#)
启动和停止 [168](#)
说明 [164](#)
添加工作流实例 [167](#)
在 Workflow Monitor 中查看 [169](#)
使用 \$PMWorkflowRunInstanceName 变量 [167](#)

工作流属性

服务级别 [178](#)

工作流运行 ID

说明 [165](#)

工作流运行 ID (续)

在工作流日志中查看 [170](#)

管道

说明 [17](#), [29](#), [50](#)

管道查找

对源表进行分区 [47](#)

管道分区

编辑分区点 [24](#)
并发连接 [38](#)
传递分区类型 [60](#)
错误阈值 [161](#)
动态分区 [20](#)
对间接文件进行分区 [34](#)
对象验证 [23](#)
多组转换 [19](#)
分区键 [56](#), [58](#)
FTP 文件目标 [246](#)
管道阶段 [17](#)
关系目标 [38](#)
规则 [23](#)
哈希用户键分区 [56](#)
哈希自动键分区 [56](#)
合并目标文件 [38](#), [39](#)
HTTP 转换 [40](#)
缓存 [22](#)
恢复 [161](#)
Java 转换 [40](#)
键范围 [57](#)
加载到 Informix [38](#)
联接器转换 [42](#)
排序器转换 [45](#), [48](#)
配置会话 [24](#)
配置下推优化 [89](#)
配置以优化联接性能 [42](#)
筛选条件 [32](#)
事务控制转换 [51](#)
数据库兼容性 [38](#)
说明 [17](#), [29](#), [50](#)
数值函数限制 [49](#)
SQL 查询 [32](#)
添加哈希键 [56](#)
添加键范围 [59](#)
外部加载器 [38](#), [222](#)
为已排序的数据配置 [42](#)
文件列表 [34](#)
文件目标 [38](#)
文件源 [33](#)
消息队列 [38](#)
性能 [56](#), [58](#), [61](#)
序列生成器转换 [48](#)
循环分区 [61](#)
映射变量 [22](#)
已排序关系数据 [44](#)
已排序平面文件 [43](#)
用例 [51](#)
有效分区类型 [51](#)
在网络上 [174](#)
指导原则 [33](#)
自定义转换 [40](#)
线程和分区 [18](#)

管道阶段

说明 [17](#)

关系目标

分区 [38](#)

分区约束 [38](#)

关系数据库日志记录

错误日志类型, 配置 [142](#)

- 规范器转换
 - 使用分区点 [30](#)
- 更改数据
 - PowerExchange 实时更改数据捕获 [108](#)
- 更新策略转换
 - 下推优化 [105](#)
- 工作流
 - 变量 [181](#)
 - 并发实例 [164](#)
 - 参数文件 [187](#)
 - 分派任务 [178](#)
 - 服务级别 [178](#)
 - 计划并发工作流 [170](#)
 - 配置实例名称 [167](#)
 - 配置同名并发 [165](#)
 - 配置唯一实例 [164](#)
 - 通过 pmcmd 启动并发工作流 [168](#)
 - 通过网格分发 [173](#), [176](#)
 - 在网格上恢复 [176](#)
 - 在网格上运行 [173](#)
 - 正在挂起 [149](#)
 - 正在停止 [162](#)
 - 中止 [162](#)
 - 状态 [149](#)
 - 操作状态 [145](#)
- 工作流日志
 - 工作流日志计数, 变量类型 [200](#)
 - 文件名和目录, 变量类型 [200](#)
- 挂起电子邮件
 - 变量类型 [200](#)

H

- 哈希分区
 - 说明 [50](#)
 - 添加哈希键 [56](#)
- 哈希用户键
 - 说明 [20](#)
- 哈希用户键分区
 - 概览 [56](#)
 - 性能 [56](#)
- 哈希自动键分区
 - 概览 [56](#)
 - 说明 [20](#)
- 合并类型
 - 说明 [39](#)
- 合并命令
 - 说明 [39](#)
 - 参数和变量类型 [200](#)
- 合并目标文件
 - 本地连接 [38](#), [39](#)
 - 并发合并 [40](#)
 - FTP [38](#)
 - FTP 文件目标 [246](#)
 - 会话属性 [39](#)
 - 连续合并 [40](#)
 - 命令 [40](#)
 - 文件列表 [40](#)
- 合并文件名
 - 说明 [39](#)
 - 参数和变量类型 [200](#)
- 合并文件目录
 - 说明 [39](#)
 - 参数和变量类型 [200](#)
- 后期工作集变量分配
 - 执行 [189](#)

- HTTP 转换
 - 管道分区 [40](#)
 - 线程 [41](#)
- 缓冲区块大小
 - 配置 [274](#)
- 缓冲区内存在
 - 分配 [274](#)
 - 缓冲区块 [274](#)
 - 配置 [274](#)
- 缓存
 - 不可重用会话的 [252](#)
 - 查找转换 [259](#)
 - 持久性查找 [259](#)
 - 等级转换 [260](#)
 - 分区 [22](#)
 - 概览 [248](#)
 - 缓存计算器 [252](#), [253](#)
 - 会话缓存文件 [248](#)
 - 汇总器转换 [255](#)
 - 可重用会话的 [252](#)
 - 联接器转换 [256](#)
 - 内存 [249](#)
 - 排序器转换 [261](#)
 - 配置 [253](#)
 - 配置方法 [252](#)
 - 配置内存上限 [276](#)
 - 数值 [253](#)
 - 替代 [252](#)
 - 通过实时会话重置 [130](#)
 - 网格上的数据缓存 [175](#)
 - 网格上的索引缓存 [175](#)
 - 为 XML 目标配置 [262](#)
 - 为查找转换配置 [259](#)
 - 为等级转换配置 [260](#)
 - 为汇总器转换配置 [255](#)
 - 为联接器转换配置 [257](#), [261](#)
 - XML 目标 [262](#)
 - 已排序输入汇总器转换 [255](#)
 - 优化 [262](#)
 - 转换的 [248](#)
 - 自动内存 [253](#)
- 缓存大小
 - 会话内存要求, 配置 [275](#)
 - 配置 [252](#)
 - 优化 [262](#)
- 缓存分区
 - 查找转换 [47](#), [254](#), [259](#)
 - 等级转换 [254](#), [260](#)
 - 汇总器转换 [254](#), [255](#)
 - 联接器转换 [254](#), [257](#)
 - 排序器转换 [254](#), [261](#)
 - 配置缓存大小 [254](#)
 - 性能 [22](#)
 - 已介绍 [22](#)
 - 增量汇总 [255](#)
 - 转换 [254](#)
- 缓存计算器
 - 查找转换输入 [259](#)
 - 等级转换输入 [260](#)
 - 汇总器转换输入 [255](#)
 - 联接器转换输入 [257](#)
 - 排序器转换输入 [261](#)
 - 使用 [253](#)
 - 说明 [252](#)
- 缓存文件
 - 查找 [266](#)
 - 命名约定 [250](#)

恢复

- 按实例名称恢复 [165](#)
 - 按运行 ID 恢复工作流 [166](#)
 - 包含增量汇总器的会话 [145](#)
 - 策略 [151](#)
 - 从任务中恢复工作流 [158](#)
 - 从上一检查点继续 [151](#), [152](#)
 - 概览 [144](#)
 - 管道分区 [161](#)
 - 规则和准则 [158](#)
 - 恢复任务 [157](#)
 - 会话运行状态 [145](#)
 - 来自检查点的会话 [152](#)
 - 目标恢复表 [145](#)
 - 平面文件 [154](#)
 - PM_RECOVERY 表格式 [145](#)
 - PM_TGT_RUN_ID 表格式 [145](#)
 - SDK 源 [154](#)
 - 删除数据库视图 [86](#)
 - 删除数据库序列 [86](#)
 - 实时会话 [113](#)
 - 使用会话中的可重复数据 [153](#)
 - 完成不可恢复的会话 [158](#)
 - 网络上的工作流 [176](#)
 - 网络上的会话 [176](#)
 - 完整恢复 [152](#)
 - 下推优化 [83](#)
 - 验证会话 [153](#)
 - 增量 [152](#)
 - 工作流运行状态 [145](#)
- ## 恢复表
- 从脚本手动创建 [147](#)
 - 说明 [145](#)
 - 消息处理 [116](#)
 - 消息恢复 [113](#), [116](#)
- ## 恢复策略
- 重新启动任务 [151](#), [152](#)
 - 从上一检查点继续 [151](#), [152](#)
 - 使任务失败并继续运行工作流 [151](#), [152](#)
- ## 恢复队列
- 消息处理 [117](#)
 - 消息恢复 [113](#), [117](#)
- ## 恢复工作流
- 按实例名称恢复工作流 [165](#)
 - 按运行 ID 恢复实例 [166](#)
- ## 恢复文件
- 消息恢复 [113](#), [115](#)
- ## 恢复主题
- 消息处理 [117](#)
 - 消息恢复 [113](#), [117](#)
- ## 回滚数据
- 事务控制 [127](#)
- ## 会话
- 操作状态 [145](#)
 - 分配资源 [179](#)
 - 故障 [23](#), [161](#)
 - 会话前和会话后变量 [197](#)
 - 目标端下推优化 [63](#)
 - 配置以优化联接性能 [42](#)
 - 使用 FTP [243](#)
 - 使用 SFTP [243](#)
 - 通过网格分发 [173](#), [176](#)
 - 外部加载 [221](#), [238](#)
 - 完整下推优化 [63](#)
 - 为下推优化配置 [89](#)
 - 下推优化 [63](#)
 - 源端下推优化 [63](#)
 - 在对象之间传递信息 [197](#)

会话 (续)

- 在对象之间传递信息, 示例 [197](#)
 - 在网络上恢复 [176](#)
 - 在网络上运行 [173](#)
 - 正在停止 [160](#), [163](#)
 - 中止 [160](#), [163](#)
 - 参数 [192](#)
- ## 会话错误
- 处理 [161](#)
- ## 会话后变量分配
- 成功时执行 [197](#)
 - 失败后执行 [197](#)
- ## 会话恢复数据刷新
- 消息恢复 [115](#)
- ## 会话前变量分配
- 执行 [197](#)
- ## 会话日志
- 传递到外部库 [268](#)
 - 工作流恢复 [158](#)
 - 外部加载器错误消息 [222](#)
 - 会话参数 [192](#)
 - 目录, 变量类型 [200](#)
 - 文件名, 参数类型 [200](#)
- ## 会话日志界面
- 函数 [269](#)
 - INFA_AbnormalSessionTermination [272](#)
 - INFA_EndSessionLog [271](#)
 - INFA_InitSessionLog [269](#)
 - INFA_OutputSessionLogFatalMsg [271](#)
 - INFA_OutputSessionLogMsg [270](#)
 - 集成服务调用 [268](#)
 - 实现 [268](#)
 - 说明 [268](#)
 - 指导原则 [269](#)
- ## 会话属性
- FastExport 源 [219](#)
 - 基于目标的提交 [133](#)
 - 排序顺序 [265](#)
- ## 汇总缓存
- 重新初始化 [265](#)
- ## 汇总器缓存
- 概览 [255](#)
 - 说明 [255](#)
- ## 汇总器转换
- 缓存 [255](#)
 - 缓存分区 [254](#), [255](#)
 - 缓存计算器的输入 [255](#)
 - 配置缓存 [255](#)
 - 使用分区点 [30](#)
 - 添加到并发工作流 [170](#)
 - 已排序端口 [255](#)
 - 下推优化 [94](#)
- ## 汇总文件
- 重新初始化 [265](#)
 - 删除 [265](#)
 - 移动 [265](#)
- ## 活动数据库
- 说明 [64](#)
- ## 活动源
- 基于源的提交 [124](#)
 - 生成提交 [124](#)
- ## 函数
- 会话日志界面 [269](#)
 - 下推优化 [70](#), [76](#), [80](#)
 - 在关系数据库中可用 [80](#)
 - 在企业版应用程序中可用 [80](#)
 - 在数据库仓库应用程序中可用 [76](#)
 - 在云数据仓库应用程序中可用 [70](#)

- 环境 SQL
 - 参数和变量类型 [200](#)
- 缓存目录
 - 共享 [252](#)
 - 最优, 选择 [252](#)
 - 变量类型 [200](#)
- 恢复缓存文件夹
 - JMS 的变量类型 [200](#)
 - TIBCO 的变量类型 [200](#)
 - webMethods 的变量类型 [200](#)
 - WebSphere MQ 的变量类型 [200](#)
- 会话参数
 - FTP 连接参数 [192](#)
 - 设为资源 [196](#)
 - 文件名, 变量类型 [200](#), [211](#)
 - 在参数文件中 [200](#)
 - 在会话之间传递值 [197](#)
 - 队列连接参数 [192](#)
 - 分区数 [192](#)
 - 概览 [192](#)
 - 会话日志参数 [192](#)
 - 拒绝文件参数 [192](#)
 - 命名约定 [192](#)
 - 目标文件参数 [192](#)
 - 内置 [192](#)
 - 数据库连接参数 [192](#)
 - 外部加载器连接参数 [192](#)
 - 应用程序连接参数 [192](#)
 - 用户定义 [192](#)
 - 源文件参数 [192](#)
- 会话参数文件名
 - 变量类型 [200](#), [211](#)
- 会话后 shell 命令
 - 参数和变量类型 [200](#)
- 会话后电子邮件
 - 参数和变量类型 [200](#)
- 会话前和会话后 SQL
 - 命令, 参数和变量类型 [200](#)
- 会话日志计数
 - 变量类型 [200](#)

I

- IBM DB2
 - 数据库分区 [50](#), [53](#), [55](#)
- IBM DB2 EE
 - 加载空格 [225](#)
 - 外部加载 [223](#)
 - 属性 [225](#)
- IBM DB2 EEE
 - 属性 [226](#)
 - 外部加载 [223](#)
- INFA_AbnormalSessionTermination
 - 会话日志界面 [272](#)
- INFA_EndSessionLog
 - 会话日志界面 [271](#)
- INFA_InitSessionLog
 - 会话日志界面 [269](#)
- INFA_OutputSessionLogFatalMsg
 - 会话日志界面 [271](#)
- INFA_OutputSessionLogMsg
 - 会话日志界面 [270](#)
- Informix
 - 行级锁定 [38](#)
- \$InputFile
 - 命名规则 [192](#)
 - 使用 [192](#)

J

- Java 转换
 - 管道分区 [40](#)
 - 线程 [41](#)
- 检查点
 - 会话恢复 [152](#)
 - 会话运行状态 [145](#), [152](#)
- 键范围分区
 - 分区视图 [24](#)
 - 说明 [20](#), [50](#)
 - 添加 [57](#)
 - 添加键范围 [59](#)
 - 下推优化 [90](#)
 - 添加分区键 [58](#)
 - 性能 [58](#)
- 集成服务
 - 分配网络 [176](#)
 - 提交间隔概览 [123](#)
 - 外部加载器支持 [221](#)
 - 网络概览 [172](#)
 - 网络上的行为 [176](#)
 - 在会话日志界面中调用函数 [268](#)
 - 在网络上运行会话 [173](#)
- 集成服务代码页
 - 影响增量汇总 [265](#)
- 计划工作流
 - 并发工作流 [170](#)
- 计时器任务
 - 变量 [181](#), [200](#)
- 继续执行恢复策略
 - 使用恢复目标表 [145](#)
 - 使用可重复数据 [153](#)
- 基于目标的提交
 - 配置 [112](#)
 - 实时会话 [112](#)
 - WriterWaitTimeOut [123](#)
- 基于目标的提交间隔
 - 说明 [123](#)
- 基于源的提交
 - 活动源 [124](#)
 - 配置 [112](#)
 - 实时会话 [112](#)
 - 说明 [124](#)
- JMS 目标
 - 参数和变量类型 [200](#)
- 拒绝文件
 - Oracle 外部加载器 [228](#)
 - 事务控制 [128](#)
 - 参数和变量类型 [200](#)
 - 会话参数 [192](#)
- 拒绝文件名
 - 说明 [39](#)
- 拒绝文件目录
 - 目标文件属性 [39](#)
 - 参数和变量类型 [200](#)
- 基础 URL
 - 参数和变量类型 [200](#)
- 脚本文件
 - 参数和变量类型 [200](#)

K

- 开放式事务
 - 定义 [130](#)
- 可重复任务
 - 说明 [150](#)

- 可重复数据
 - 恢复工作流 [153](#)
 - 通过源 [154](#)
 - 通过转换 [154](#)
- 可重用会话
 - 缓存 [252](#)
- 可执行文件名称
 - FastExport 属性 [217](#)
- 空闲时间
 - 配置 [111](#)
- 空闲数据库
 - 说明 [64](#)
- 控制任务
 - 停止或中止工作流 [162](#)
- 控制文件替代
 - 加载 Teradata [232](#)
 - 设置 Teradata FastExport 语句 [219](#)
 - 说明 [219](#)
 - 替代 Teradata FastExport 的步骤 [219](#)
- 块大小
 - FastExport 属性 [217](#)

L

- 冷启动
 - 实时会话 [118](#)
- 链接
 - 变量 [181](#)
 - 变量类型 [200](#)
- 联接器缓存
 - 说明 [256](#)
- 联接器转换
 - 分区 [256](#)
 - 分区准则 [49](#)
 - 缓存 [256](#)
 - 缓存分区 [254](#), [257](#)
 - 缓存计算器的输入 [257](#)
 - 联接排序关系数据 [44](#)
 - 联接已排序平面文件 [43](#)
 - 配置缓存 [257](#), [261](#)
 - 下推优化 [97](#)
- 连续合并
 - 文件目标 [40](#)
- 临时文件
 - Teradata FastExport 属性 [219](#)
- \$LoaderConnection
 - 使用 [192](#)
- \$LookupFile
 - 命名规则 [192](#)
 - 使用 [192](#)
- 连接环境 SQL
 - 参数和变量类型 [200](#)
- 联合转换
 - 下推优化 [105](#)
- 路由器转换
 - 下推优化 [100](#)

M

- Microsoft Access
 - 管道分区 [38](#)
- Microsoft Azure SQL 数据库连接
 - 下推优化, 规则和准则 [66](#)
- 命令
 - 已分区的目标 [40](#)
 - 已分区的源 [33](#)

- 命令类型
 - 对文件源进行分区 [34](#)
- 命令任务
 - 分配资源 [179](#)
 - 变量类型 [200](#)
 - “命令” 属性
 - 对文件源进行分区 [34](#)
 - 配置已分区的目标 [39](#)
- MOVINGAVG 函数
 - 分区约束 [49](#)
- MOVINGSUM 函数
 - 分区约束 [49](#)
- 目标
 - 分区 [38](#)
 - 合并输出文件 [38](#), [39](#)
 - 删除分区点 [30](#)
 - 使用下推优化 [91](#)
 - 通过 FTP 访问 [241](#), [242](#), [244](#)
 - 下推优化 [104](#)
- 目标端下推优化
 - 说明 [63](#)
- 目标恢复表
 - 手动创建 [147](#)
 - 说明 [145](#)
- 目标连接组
 - 提交数据 [124](#)
 - 事务控制转换 [132](#)
- 目标命令
 - 目标 [40](#)
 - 在分区中使用 [40](#)
- 目标文件
 - 附加 [39](#)
 - 会话参数 [192](#)
- 目录
 - 共享缓存 [252](#)
 - 历史汇总数据的 [267](#)
- 命名约定
 - 会话参数 [192](#)
- 目标表
 - 参数和变量类型 [200](#)
- 目标更新
 - 参数和变量类型 [200](#)
- \$Target connection value
 - 参数和变量类型 [200](#)
- 目标数据库
 - 数据库连接会话参数 [192](#)

N

- 内存
 - 缓存 [249](#)
 - 配置自动设置 [275](#)
 - 为多个会话配置设置 [275](#)
- 内存设置
 - 为多个会话配置 [275](#)
- 内存要求
 - DTM 缓冲区大小 [275](#)
 - 会话缓存大小 [275](#)
- Netezza 连接
 - 下推优化, 规则和准则 [66](#)

O

- Oracle
 - 数据库分区 [50](#), [53](#)

Oracle 外部加载器
带分隔符的平面文件目标 [229](#)
多字节数据 [229](#)
固定宽度平面文件目标 [229](#)
拒绝文件 [228](#)
数据精度 [229](#)
属性 [229](#)
外部加载器支持 [221](#), [228](#)
已分区的目标文件 [229](#)
\$OutputFile
命名规则 [192](#)
使用 [192](#)

P

排序器缓存
命名约定 [250](#)
说明 [261](#)
排序器转换
分区 [48](#)
缓存 [261](#)
缓存分区 [254](#), [261](#)
缓存计算器的输入 [261](#)
已优化的联接性能的分区分区 [45](#)
工作目录, 变量类型 [200](#)
下推优化 [101](#)
排序顺序
为输入行保留 [37](#)
影响增量汇总 [265](#)
判定任务
变量 [181](#)
变量类型 [200](#)
配置并发执行
配置工作流实例 [167](#)
批量加载
使用用户定义的提交 [127](#)
平面文件
保留输入行顺序 [37](#)
“表头命令”属性 [39](#)
“表头选项”属性 [39](#)
“表尾命令”属性 [39](#)
配置恢复 [154](#)
输出文件会话参数 [192](#)
源文件会话参数 [192](#)
平面文件日志记录
错误日志类型, 配置 [142](#)
错误日志文件名, 配置 [142](#)
错误日志文件目录, 配置 [142](#)
PM_REC_STATE 表
实时会话 [116](#)
手动创建 [147](#)
说明 [145](#)
PM_RECOVERY 表
格式 [145](#)
手动创建 [147](#)
说明 [145](#)
死锁重试 [145](#)
PM_TGT_RUN_ID
格式 [145](#)
手动创建 [147](#)
说明 [145](#)
PMErrors_MSG 表
架构 [138](#)
PMErrors_ROWDATA 表
架构 [136](#)
PMErrors_Session 表
架构 [139](#)

\$PMSessionLogFile
使用 [192](#)
\$PMStorageDir
工作流运行状态 [145](#)
\$PMWorkflowRunId
并发工作流 [167](#)
\$PMWorkflowRunInstanceName
并发工作流 [167](#)
PowerCenter 实时产品
概览 [121](#)
PowerExchange Client for PowerCenter
实时更改数据 [108](#)
“Pre 85 时间戳兼容性”选项, 适用于 Netezza 上的下推优化 [66](#)
\$\$PushdownConfig
说明 [87](#)

Q

前期工作集变量分配
执行 [189](#)
启动高级工作流
启动并发工作流 [168](#)
\$QueueConnection
使用 [192](#)

R

韧度
FastExport 属性 [217](#)
任务
分配资源 [179](#)
负载均衡器设置 [178](#)
恢复策略 [151](#)
正在停止 [162](#)
中止 [162](#)
自动恢复 [152](#)
日志表名称
FastExport 属性 [217](#)
日志记录
下推优化 [83](#)
若存在则追加
平面文件目标属性 [39](#)

S

SDK 源
恢复 [154](#)
SetID
PeopleSoft, 参数和变量类型 [200](#)
SFTP
创建会话 [243](#)
在网络上运行会话 [243](#)
密钥文件位置 [243](#)
说明 [242](#)
筛选条件
添加 [59](#)
WebSphere MQ, 参数和变量类型 [200](#)
在已分区的管道中 [32](#)
参数和变量类型 [200](#)
shell 命令
参数和变量类型 [200](#)
生成
包含基于源的提交的提交 [124](#)
失败任务恢复策略
说明 [151](#), [152](#)

- 实例
 - 工作流实例说明 [164](#)
- 实时刷新延迟
 - 配置 [112](#)
- 事务
 - 已定义 [130](#)
- 事务边界
 - 丢弃 [130](#)
 - 事务控制 [130](#)
- 事务控制
 - 点 [130](#)
 - 概览 [130](#)
 - 规则和准则 [133](#)
 - 集成服务处理 [127](#)
 - 拒绝文件 [128](#)
 - 开放式事务 [130](#)
 - 批量加载 [127](#)
 - 实时会话 [130](#)
 - 文件末尾 [128](#)
 - 用户定义的提交 [127](#)
 - 转换错误 [128](#)
 - 转换范围 [130](#)
- 事务控制转换
 - 分区准则 [51](#)
 - 目标连接组 [132](#)
- 事务生成器
 - 事务控制点 [130](#)
- 刷新会话恢复数据（属性）
 - 集成服务 [115](#)
- 刷新延迟
 - 配置 [112](#)
 - 说明 [112](#)
- 输出具有确定性（属性）
 - 关于 [154](#)
 - “输出类型”属性
 - 对文件目标进行分区 [39](#)
- 输出是可重复的（属性）
 - 关于 [154](#)
 - “输出文件名”属性
 - 对目标文件进行分区 [39](#)
 - 参数和变量类型 [200](#)
- 输出文件目录属性
 - 对目标文件进行分区 [39](#)
 - 参数和变量类型 [200](#)
- 数据
 - 捕获增量源更改 [264](#), [267](#)
- 数据缓存
 - 命名约定 [250](#)
 - 增量汇总的 [266](#)
- 数据加密
 - FastExport 属性 [217](#)
- 数据库分区
 - 多个源 [54](#)
 - 集成服务的规则和准则 [55](#)
 - 目标 [55](#)
 - 目标的规则和准则 [55](#)
 - 说明 [20](#), [50](#)
 - 性能 [53](#), [55](#)
 - 一个源 [53](#)
 - 源的规则和准则 [55](#)
 - 源的集成服务处理 [55](#)
- 数据库视图
 - 故障排除 [86](#)
 - 恢复过程中删除 [86](#)
 - 删除孤立的视图 [86](#)
 - 使用下推优化创建 [85](#)
 - 下推优化 [86](#)
- 数据库序列
 - 故障排除 [86](#)
 - 恢复过程中删除 [86](#)
 - 删除孤立的序列 [86](#)
 - 下推优化 [86](#)
- 数据文件
 - 查找 [266](#)
 - 创建目录 [267](#)
- 数据移动模式
 - 影响增量汇总 [265](#)
- 输入类型
 - 文件源分区属性 [34](#)
- 属性
 - 分区级别 [22](#)
- SQL
 - 为下推优化生成的 [63](#)
 - 已分区的管道中的队列 [32](#)
- SQL 替代
 - 下推优化 [85](#)
- SQL 查询
 - 参数和变量类型 [200](#)
- 索引
 - 查找 [266](#)
 - 创建目录 [267](#)
- 索引缓存
 - 命名约定 [250](#)
 - 增量汇总的 [266](#)
- Sybase IQ
 - 分区约束 [38](#)
- Sybase IQ 外部加载器
 - 带分隔符的平面文件目标 [230](#)
 - 多字节数据 [230](#)
 - 概览 [230](#)
 - 固定宽度的平面文件目标 [230](#)
 - 数据精度 [230](#)
 - 支持 [221](#)
 - 属性 [230](#)
- 筛选器转换
 - 下推优化 [96](#)
- 设置控制值
 - PeopleSoft, 参数和变量类型 [200](#)
- 生效日期
 - PeopleSoft, 参数和变量类型 [200](#)
- 实时处理
 - 示例映射 [120](#)
 - 说明 [107](#)
- 实时会话
 - 重新启动 [118](#)
 - 弹性 [119](#)
 - 读取器时间限制, 配置 [111](#)
 - 规则和准则 [119](#)
 - 恢复 [118](#)
 - 继续 [118](#)
 - 空闲时间, 配置 [111](#)
 - 冷启动 [118](#)
 - 配置 [110](#)
 - PM_REC_STATE 表 [116](#)
 - 示例映射 [120](#)
 - 刷新延迟, 配置 [112](#)
 - 提交类型, 配置 [112](#)
 - 消息处理 [116](#), [117](#)
 - 消息恢复 [115-117](#)
 - 消息计数, 配置 [111](#)
 - 正在停止 [118](#)
 - 支持的产品 [121](#)
 - 中止 [118](#)
 - 终止条件, 配置 [111](#)
 - 转换 [119](#)

- 实时会话 (续)
 - 转换范围 [130](#)
 - 概览 [107](#)
 - 说明 [107](#)
- 实时数据
 - Web 服务消息 [108](#)
 - 支持的产品 [121](#)
 - 概览 [108](#)
 - 来自 PowerExchange 源的更改数据 [108](#)
 - 消息, 消息队列, 和更改数据捕获 [108](#)
- 事件等待任务
 - 文件监视名称, 变量类型 [200](#)
- 事务环境 SQL
 - 参数和变量类型 [200](#)
- 事务控制单元
 - 说明 [132](#)
- 树
 - PeopleSoft, 参数和变量类型 [200](#)
- 数据库连接
 - 参数 [195](#)
 - 参数类型 [200](#)
 - 会话参数 [192](#)
 - 密码, 参数类型 [200](#)
 - 下推兼容 [67](#)
 - 用户名参数类型 [200](#)

T

- TDPID
 - 说明 [217](#)
- Teradata FastExport
 - 创建连接 [217](#)
 - 读取多字节字符 [217](#)
 - fexp 命令 [217](#)
 - 更改源连接 [219](#)
 - 规则和准则 [220](#)
 - 会话属性说明 [219](#)
 - 连接属性 [217](#)
 - 使用步骤 [216](#)
 - 说明 [216](#)
 - TDPID 属性 [217](#)
 - 替代控制文件 [219](#)
 - 选择读取器 [218](#)
 - 暂存数据 [219](#)
 - 临时文件, 变量类型 [200](#)
- Teradata 外部加载器
 - 代码页 [231](#)
 - FastLoad 属性 [237](#)
 - MultiLoad 属性 [233](#)
 - 日期格式 [231](#)
 - 替代控制文件 [232](#)
 - TPump 属性 [235](#)
 - 支持 [221](#)
 - 控制文件内容替代, 参数和变量类型 [200](#)
- TIB/存储库
 - TIB/Adapter SDK 存储库 URL, 变量类型 [200](#)
- 替代
 - Teradata 加载器控制文件 [232](#)
- 提交间隔
 - 基于源和基于目标 [123](#)
 - 配置 [133](#)
 - 说明 [123](#)
- 提交类型
 - 配置 [112](#)
 - 实时会话 [112](#)
- 提交数据
 - 目标连接组 [124](#)

- 提交数据 (续)
 - 事务控制 [127](#)
- 提交源
 - 基于源的提交 [124](#)
- 停止
 - 错误阈值 [161](#)
- 提取日期
 - PeopleSoft, 参数和变量类型 [200](#)

U

- UNIX 系统
 - 外部加载器行为 [222](#)

V

- Vertica 连接
 - 下推优化, 规则和准则 [66](#)

W

- 外部过程转换
 - 分区准则 [49](#)
 - 初始化属性, 变量类型 [200](#)
- 外部加载器
 - 处理子秒 [222](#)
 - 错误消息 [222](#)
 - 代码页 [221](#)
 - DB2 [223](#)
 - 对已分区的管道使用 [38](#)
 - 概览 [221](#)
 - 加载多字节数据 [229](#), [230](#)
 - 集成服务支持 [221](#)
 - Oracle [228](#)
 - 配置为资源 [221](#)
 - 设置 Workflow Manager [238](#)
 - Sybase IQ [230](#)
 - Teradata [231](#)
 - 行为 [222](#)
 - 在 Windows 系统中 [222](#)
- 网格
 - 分发工作流 [173](#), [176](#)
 - 分发会话 [173](#), [176](#)
 - 概览 [172](#)
 - 管道分区 [174](#)
 - 缓存要求 [175](#)
 - 恢复工作流 [176](#)
 - 恢复会话 [176](#)
 - 集成服务属性设置 [176](#)
 - 集成服务行为 [176](#)
 - 配置工作流属性 [176](#)
 - 配置会话属性 [176](#)
 - 配置资源 [176](#)
 - 要求 [176](#)
 - 运行会话 [173](#)
 - 指定内存上限 [276](#)
- 网格上的会话
 - 说明 [173](#)
 - 序列生成器转换的分区 [48](#)
- 网格中的节点数
 - 通过动态分区设置 [21](#)
- 完整恢复
 - 说明 [152](#)
- 完整下推优化
 - 说明 [63](#)

Web Services Hub
 运行并发工作流 [165](#)
Web 服务消息
 实时数据 [108](#)
未连接的转换
 分区约束 [30](#)
文件列表
 合并目标文件 [40](#)
 为已分区的源创建 [34](#)
文件末尾
 事务控制 [128](#)
文件目标
 分区 [38](#)
 代码页, 参数和变量类型 [200](#)
文件源
 分区 [33](#)
 代码页, 参数和变量类型 [200](#)
 名称, 参数和变量类型 [200](#)
 目录, 参数和变量类型 [200](#)
 输入文件命令, 参数和变量类型 [200](#)
Windows 系统
 外部加载器行为 [222](#)
Workflow Monitor
 查看并发工作流 [169](#)
WriterWaitTimeOut
 基于目标的提交 [123](#)
外部加载器连接
 参数类型 [200](#)
 会话参数 [192](#)
 密码, 参数类型 [200](#)
 用户名, 参数类型 [200](#)

X

小数
 高精度处理 [278](#)
小数秒数精度
 Teradata FastExport 属性 [219](#)
消息处理
 规则和准则 [119](#)
 恢复表 [116](#)
 恢复队列 [117](#)
 恢复主题 [117](#)
 实时会话 [116](#), [117](#)
消息队列
 对已分区的管道使用 [38](#)
 处理实时数据 [108](#)
消息恢复
 规则和准则 [119](#)
 恢复表 [113](#), [116](#)
 恢复队列 [113](#), [117](#)
 恢复文件 [113](#), [115](#)
 恢复主题 [113](#), [117](#)
 会话恢复数据刷新 [115](#)
 启用 [113](#)
 实时会话 [113](#), [115](#)–[117](#)
 说明 [113](#)
 前提条件 [113](#)
消息计数
 配置 [111](#)
下推优化
 \$\$PushdownConfig 参数[下推优化
 下推配置] [87](#)
 AWS Redshift [65](#)
 Azure DW [65](#)
 传递分区类型 [90](#)
 创建数据库视图 [85](#)

下推优化 (续)
 错误处理 [83](#)
 Google Big Query [65](#)
 Greenplum [65](#)
 规则和准则 [91](#)
 合并分区 [90](#)
 恢复 [83](#)
 会话 [63](#)
 键范围分区, 使用 [90](#)
 加载到目标 [91](#)
 临时视图 [85](#)
 临时序列 [84](#)
 目标端优化 [63](#)
 ODBC 连接 [65](#)
 配置分区 [89](#)
 配置会话 [89](#)
 PostgreSQL [65](#)
 日志记录 [83](#)
 数据库视图 [86](#)
 数据库序列 [86](#)
 Snowflake [65](#)
 SQL 生成的 [63](#)
 SQL 与 ANSI SQL [64](#)
 完整下推优化 [63](#)
 向映射中添加转换 [91](#)
 性能问题 [63](#)
 源端优化 [63](#)
 源数据库分区 [55](#)
 转换 [93](#)
 本地数据库驱动程序 [64](#)
 表达式转换 [95](#)
 参数类型 [200](#)
 概览 [62](#)
 更新策略转换 [105](#)
 函数 [70](#), [76](#), [80](#)
 汇总器转换 [94](#)
 联合转换 [105](#)
 联接器转换 [97](#)
 路由器转换 [100](#)
 目标 [104](#)
 排序器转换 [101](#)
 筛选器转换 [96](#)
 序列生成器转换 [100](#)
 映射变量 [70](#)
 源限定符转换 [102](#)
 运算符 [70](#)
下推优化查看器
 查看下推组 [91](#)
下推组
 查看 [91](#)
 说明 [91](#)
 下推优化查看器, 使用 [91](#)
性能
 缓存设置 [252](#)
 提交间隔 [124](#)
休眠
 FastExport 属性 [217](#)
XML 目标
 分区约束 [49](#)
 缓存 [262](#)
 基于目标的提交 [123](#)
 配置缓存 [262](#)
XML 生成器转换
 分区约束 [49](#)
XML 目标缓存
 说明 [262](#)
 变量类型 [200](#)

- 序列生成器转换
 - 分区 [48](#)
 - 分区准则 [30, 49](#)
 - 添加到并发工作流 [170](#)
 - 下推优化 [100](#)
- 循环分区
 - 说明 [20, 50, 61](#)
- 下推兼容性
 - 不兼容的数据库用户 [68](#)
 - 说明 [67](#)
 - 要求 [67](#)
- 线程
 - HTTP 转换 [41](#)
 - Java 转换 [41](#)
 - 自定义转换 [41](#)
 - 分区 [18](#)
- 消息和消息队列
 - 实时数据 [108](#)

Y

- 验证
 - 会话是否可恢复 [153](#)
- 已挂起
 - 状态 [149](#)
- 映射
 - 源自分区的会话故障 [23](#)
- 映射变量
 - 在参数文件中 [200](#)
 - 在会话之间传递值 [197](#)
 - 在已分区的管道中 [22](#)
 - 下推优化 [70](#)
 - 在数据库中提供 [70](#)
- 已排序端口
 - 缓存要求 [255](#)
- 已排序关系数据
 - 已优化的联接性能的分区分区 [44](#)
- 已排序平面文件
 - 已优化的联接性能的分区分区 [43](#)
- 已暂存
 - FastExport 会话属性 [219](#)
- 用户定义的提交
 - 批量加载 [127](#)
- 优化吞吐量
 - 会话属性 [37](#)
- 优先级
 - 分配给任务 [178](#)
- 源
 - 保留输入行排序顺序 [37](#)
 - 并发读取 [34](#)
 - 分区 [33](#)
 - 会话属性 [34](#)
 - 命令 [33](#)
- 源端下推优化
 - 说明 [63](#)
- 源管道
 - 说明 [17, 29, 50](#)
- 源数据
 - 捕获汇总更改 [264](#)
- 源位置
 - 会话属性 [34](#)
- 源文件
 - 会话属性 [34](#)
 - 通过 FTP 访问 [241, 242, 244](#)
 - 会话参数 [192](#)
 - 使用参数 [192](#)

- 源文件类型
 - 说明 [34](#)
- 源文件名
 - 说明 [34](#)
- 源限定符转换
 - 使用分区点 [30](#)
 - 下推优化, SQL 替代 [85](#)
 - 下推优化 [102](#)
- 运行时分区
 - 在会话属性中设置 [21](#)
- 延迟
 - 说明 [107](#)
- 应用程序连接
 - 参数类型 [200](#)
 - 会话参数 [192](#)
 - 密码, 参数类型 [200](#)
 - 用户名, 参数类型 [200](#)
- 映射参数
 - \$\$PushdownConfig[映射参数下推配置] [87](#)
 - 替代 [196](#)
 - 在参数文件中 [200](#)
 - 在会话属性中 [196](#)
 - 在会话之间传递值 [197](#)
- 用户定义的联接
 - 参数和变量类型 [200](#)
- 源表
 - 参数和变量类型 [200](#)
- \$Source connection value
 - 参数和变量类型 [200](#)
- 源数据库
 - 数据库连接会话参数 [192](#)
- 运算符
 - 下推优化 [70](#)
 - 在数据库中提供 [70](#)
- 运行时位置
 - 变量类型 [200](#)

Z

- 增量更改
 - 捕获 [267](#)
- 增量恢复
 - 说明 [152](#)
- 增量汇总
 - 重新初始化缓存 [265](#)
 - 处理 [265](#)
 - 分区数据 [266](#)
 - 概览 [264](#)
 - 更改会话排序顺序 [265](#)
 - 缓存分区 [255](#)
 - 集成服务数据移动模式 [265](#)
 - 配置会话 [267](#)
 - 删除文件 [265](#)
 - 移动文件 [265](#)
 - 准备启用 [267](#)
- 正在挂起
 - 电子邮件 [149](#)
 - 工作流 [149](#)
 - 行为 [149](#)
 - 状态 [149](#)
- 正在停止
 - 工作流 [162](#)
 - 会话 [163](#)
 - 集成服务处理 [160](#)
 - 任务 [162](#)

- 致命错误
 - 会话失败 [161](#)
- 中止
 - 工作流 [162](#)
 - 会话 [163](#)
 - 集成服务处理 [160](#)
 - 任务 [162](#)
- 终止条件
 - 配置 [111](#)
- 状态
 - 已挂起 [149](#)
 - 正在挂起 [149](#)
- 转换
 - 分区约束 [49](#)
 - 缓存 [248](#)
 - 恢复包含增量汇总器的会话 [145](#)
 - 配置下推优化 [93](#)
 - 生成可重复数据 [154](#)
 - 实时会话 [119](#)
- 转换范围
 - 实时处理 [130](#)
 - 说明 [130](#)
- 转换范围 (续)
 - 转换 [130](#)
- 自定义转换
 - 分区准则 [49](#)
 - 管道分区 [40](#)
 - 线程 [41](#)
- 自动内存设置
 - 配置 [275](#)
- 自动任务恢复
 - 配置 [152](#)
- 子秒
 - 外部加载 [222](#)
- 资源
 - 分配给任务 [179](#)
 - 分配外部加载器 [221](#)
- 最大会话数量
 - FastExport 属性 [217](#)
- 最大内存限制
 - 网格上的会话 [276](#)
 - 为缓存配置 [276](#)
- 暂存文件
 - SAP 文件名和目录, 变量类型 [200](#)
- 转换表达式
 - 参数和变量类型 [200](#)