



Informatica®
10.4.0

异常管理指南

Informatica 异常管理指南

10.4.0

2019 年 12 月

© 版权所有 Informatica LLC 1998, 2020

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。

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。

Informatica 和 Informatica 标志是 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。

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: 2020-02-07

目录

- 前言 6
 - Informatica 资源. 6
 - Informatica Network. 6
 - Informatica 知识库. 6
 - Informatica 文档. 6
 - Informatica 产品可用性矩阵. 7
 - Informatica Velocity. 7
 - Informatica Marketplace. 7
 - Informatica 全球客户支持部门. 7
- 第 1 章：异常管理简介..... 8
 - 异常管理概览. 8
 - 任务类型. 8
 - 任务和工作流. 9
 - 人工任务中的步骤. 9
 - 异常管理中的角色. 10
 - 异常管理工作区. 10
 - 启动工作区. 10
 - 异常项工作区. 12
 - 审计跟踪操作. 12
 - 异常管理流程. 13
 - 离散记录示例. 13
 - 重复记录示例. 14
 - 登录到 Analyst 工具. 14
- 第 2 章：异常任务..... 16
 - 异常任务概览. 16
 - 异常任务的类型. 16
 - 更正异常项步骤. 17
 - 审阅异常项步骤. 17
 - 异常项工作区中的异常项任务. 17
 - 在异常任务中修复错误. 19
 - 异常任务中的验证错误. 19
 - 异常记录状态指示器. 20
 - 设置异常记录的状态. 21
 - 异常任务的筛选器选项. 22
 - 记录数据筛选器. 22
 - 数据编辑选项卡上的元数据筛选器. 22
 - 数据审计选项卡上的元数据筛选器. 23
 - 在异常任务中筛选记录. 24

| | |
|-------------------------------|-----------|
| 查找并替换异常记录中的值. | 24 |
| 正在替换异常记录中的单个值. | 24 |
| 正在替换异常记录中的多个值. | 25 |
| 将说明添加到记录. | 26 |
| 审查中的异常任务. | 26 |
| 执行审阅异常项任务. | 26 |
| 异常任务的规则和准则. | 27 |
| 第 3 章：重复的记录任务. | 28 |
| 重复记录任务概览. | 28 |
| 首选记录. | 29 |
| 重复记录任务类型. | 29 |
| 更正重复项步骤. | 29 |
| 审阅重复项步骤. | 30 |
| 异常项工作区中的重复记录任务. | 30 |
| 编辑群集. | 32 |
| 重复记录状态指示器. | 32 |
| 设置群集的状态. | 33 |
| 在多个群集中查找记录. | 34 |
| 在群集之间移动记录. | 34 |
| 将记录移到空群集. | 35 |
| 将说明添加到群集. | 35 |
| 重复任务的筛选器选项. | 36 |
| 重复记录任务中的数据筛选器. | 36 |
| 重复记录任务中的审计跟踪筛选器. | 36 |
| 在重复记录任务中筛选群集. | 37 |
| 审阅重复项任务. | 37 |
| 执行审阅重复项任务. | 38 |
| 重复记录任务的规则和准则. | 39 |
| 第 4 章：任务管理. | 40 |
| 任务管理概览. | 40 |
| 任务执行者操作. | 40 |
| 查看任务. | 41 |
| 打开任务. | 41 |
| 释放任务. | 42 |
| 业务管理员操作. | 42 |
| 向用户分配任务. | 42 |
| 查看人工任务中的任务实例列表. | 43 |
| 任务摘要数据. | 43 |
| 多项任务完成. | 44 |
| 任务数据导出. | 45 |
| 异常任务元数据. | 45 |

| | |
|--------------|----|
| 重复任务元数据..... | 46 |
| 导出任务数据..... | 46 |
| 索引..... | 47 |

前言

要了解如何在 Informatica Analyst 中查看和更新异常，请阅读《*Informatica 异常管理指南*》。异常是指可能包含受损数据或重复数据从而需要人工评估进行确认的记录。要查找数据集中的异常，请对数据运行包含人工任务的工作流。使用 Informatica Analyst 确定记录是否为真异常、解决数据错误以及将重复的记录合并为一条记录。

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 上查找在线支持资源，请访问 <https://network.informatica.com>，然后选择 eSupport 选项。

第 1 章

异常管理简介

本章包括以下主题：

- [异常管理概览, 8](#)
- [任务类型, 8](#)
- [任务和工作流, 9](#)
- [异常管理中的角色, 10](#)
- [异常管理工作区, 10](#)
- [审计跟踪操作, 12](#)
- [异常管理流程, 13](#)
- [登录到 Analyst 工具, 14](#)

异常管理概览

异常是包含未解决的数据质量问题的记录。记录可能包含错误，也有可能包含冗余或重复信息。您可以在 Analyst 工具中审阅和修复异常记录。

要处理异常记录，请打开一个任务。任务用于标识数据库表中的一组异常记录。

Analyst 工具中的“启动”工作区显示您可以处理的任务。该工作区可能包含来自相同数据库表或不同数据库表的记录。从该工作区打开任务时，您会声明该任务的所有权，并且必须解决该任务指定的数据质量问题。

任务类型

工作流为您分配的任务类型取决于源数据库表所包含的数据质量问题的类型。表中的记录可能包含错误、空值，也可能包含当前数据项目中不准确的值。表可能包含冗余记录，因为这些记录包含相同信息的不同版本。

可以在 Analyst 工具中处理以下类型的任务实例：

更正异常项任务

具有可能包括错误或空值的记录。分析并修复找到的所有错误。记录不包含任何错误时，更新记录状态以指示该记录为有效记录。

更正重复项任务

具有可能包括重复信息的记录。该任务将记录分类到各个群集中。群集是表示源数据集中同一个业务实体的一组记录。分析每个群集，并定义该群集表示的记录的首选版本。更新群集状态以指示您已审阅首选记录。

如果记录不是群集中任何记录的重复项，则可以将该记录移至其他群集。可以创建包含单个唯一记录的群集。

审阅异常项任务

包含一组先前用户已在更正异常项任务中分析的异常记录。审阅先前用户所做的工作，并验证记录数据和记录状态正确无误。

审阅重复项任务

包含一组先前用户已在更正重复项任务中处理的群集。审阅先前用户所做的工作，并验证首选记录数据和群集状态正确无误。

注意：不能将记录从更正或审阅异常记录的任务传递给更正或审阅重复项的任务。包含异常记录的数据库表与包含重复记录的数据库表具有不同的结构。

任务和工作流

Informatica 进程（称为工作流）可生成您在 Analyst 工具中处理的任務。

工作流描述 Informatica 应用程序可以执行的任务序列。任务可能包含一些指令，以便运行映射，发送电子邮件或者将异常记录分发给 Analyst 工具用户。将异常记录分发给 Analyst 工具用户的任务称为人工任务。您在 Analyst 工具中处理的任务就是人工任务的实例。每个任务实例用于标识异常数据库表中的一组记录。

包含人工任务的工作流还必须包括映射任务。映射任务运行名为映射的数据进程。映射可验证源数据中的记录质量，并根据一组数据质量指标将记录写入不同的表中。人工任务读取包含异常记录的表，并将这些记录分配给 Analyst 工具用户进行分析和修复。

异常记录表包含许多记录时，可以用任务实例来标识一个记录子集。工作流将任务实例分发给工作流开发人员标识的 Analyst 工具用户。Analyst 工具用户可以并行处理多个任务实例。

人工任务中的步骤

工作流中的人工任务包含一个或多个步骤。步骤是通过人工任务定义数据流的元数据对象。这些步骤用于确定人工任务创建的任务实例类型并标识可以处理任务实例的用户。

工作流开发人员可以将以下步骤添加到人工任务：

异常步骤

异常步骤在 Analyst 工具中生成更正异常项任务实例。

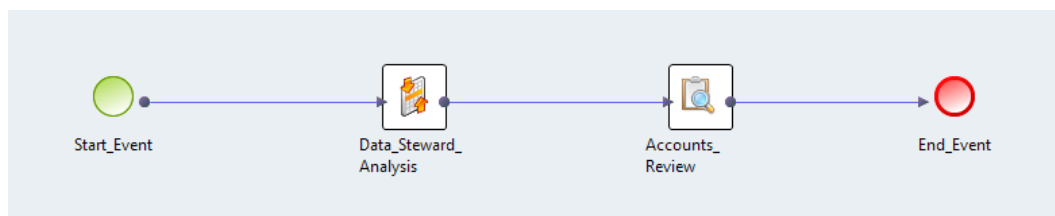
群集步骤

群集步骤在 Analyst 工具中生成更正重复项任务实例。

审阅步骤

审阅步骤在 Analyst 工具中生成审阅异常项任务实例或审阅重复项任务实例。

下图显示了 Developer tool 中人工任务的步骤：



人工任务包含多个步骤时，异常记录将从一个步骤传递到另一个步骤。配置工作流的开发人员会定义步骤的顺序。每个步骤都会标识一组用户。用户可以登录 Analyst 工具并处理该步骤定义的任务实例。

注意：配置工作流的开发人员可以为 Analyst 工具用户设置访问权限。这些权限可以确定用户能够查看的数据列以及用户能够在任务实例中执行的操作。开发人员可以设置每个步骤的访问权限。访问权限同样适用于处理该步骤定义的任务实例的所有用户。

默认情况下，所有用户对任务实例中的数据和选项都具有完全访问权限。

异常管理中的角色

人工任务将角色分配给处理任务实例的 Analyst 工具用户。角色用于确定该用户可以执行的操作范围。人工任务可以将角色分配给用户或组。如果您属于某个组，当您登录 Analyst 工具时，您将共享该组角色。

人工任务分配以下角色：

任务执行者

在 Analyst 工具中处理任务实例的用户或组。

业务管理员

负责管理人工任务生成的任务实例的用户或组。业务管理员也可以处理任务实例。

如果人工任务将任务实例分配给某个组，那么该组的任何成员都可以获得其中任务的所有权。当一个组成员获得某个任务的所有权后，该任务将从其他组成员的任务列表中消失。

工作流开发人员可以为人工任务中的每个步骤标识一个或多个任务执行者。您可以处理人工任务中单个步骤的任务实例，而不处理其他步骤的任务实例。

工作流开发人员可以为人工任务标识一个或多个业务管理员。人工任务的任何业务管理员也是该人工任务中每个步骤的业务管理员。

异常管理工作区

Analyst 工具使用工作区来组织您可以执行的不同类型的数据操作。

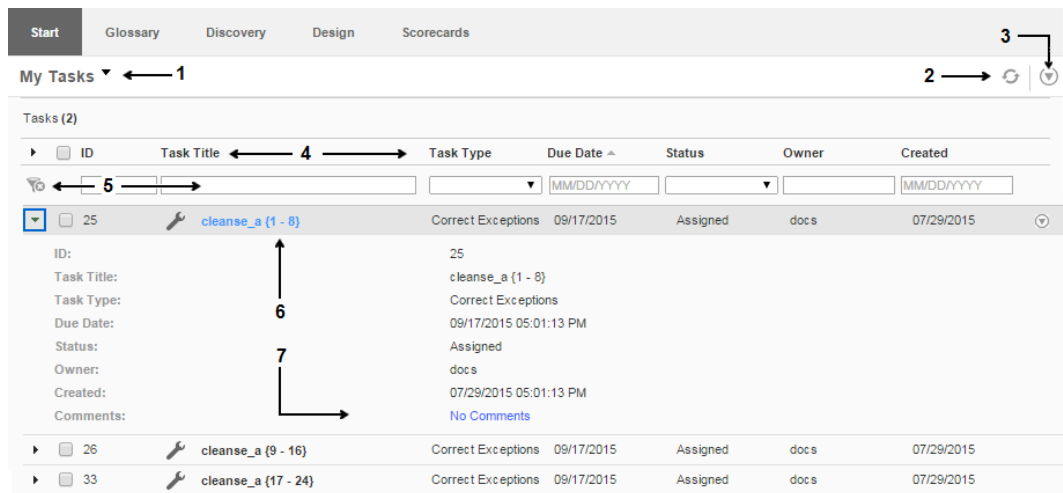
Analyst 工具在“启动”工作区的**我的任务**视图中列出您拥有的任务。打开该任务时，Analyst 工具将在**异常项**工作区中显示任务记录。

启动工作区

“启动”工作区显示您拥有的任务以及您管理的所有任务。使用“启动”工作区选项可打开任务、对任务执行操作以及审阅任务元数据。

用于登录 Analyst 工具的用户角色决定了您可以在“启动”工作区中查看的任务。选择**我的任务**视图可显示您拥有的任务。选择**任务管理**视图可显示您管理的任何任务。如果创建任务的工作流将您标识为业务管理员，则您可以管理任务。

下图显示了**我的任务**视图中的一个任务列表：



“我的任务”视图包含以下选项：

1. “我的任务”和“任务管理”选项。在您拥有的任务和您管理的任何任务之间切换工作区视图。您可以从“我的任务”视图和“任务管理”视图中打开任务。
2. 刷新。刷新工作区数据。
3. “操作”菜单。打开您可以对所选任务执行的操作列表。
4. 列标题。列出描述任务实例的列名称。
5. 筛选器。使用您输入的值筛选记录列表。
6. 任务名称。显示任务的名称。要打开任务，请选择任务名称。
7. 注释。打开一个编辑器，可以使用该编辑器输入与该任务有关的注释或读取注释。

启动工作区列标题

“启动”工作区显示表中的任务。每个任务是表中的一行。表列包含与任务有关的元数据。

使用以下可审阅您拥有或管理的任务的元数据：

ID

任务实例的唯一标识符。该标识符包含一个值，该值用于标识生成任务实例的人工任务。

任务标题

任务的名称。任务名称用于标识生成任务数据的人工任务中的步骤。

“任务标题”列包含一个筛选器字段，可用于筛选任务名称列表。可以在输入的筛选器字符串中包含通配符。

如果在筛选器字符串中添加了通配符，Analyst 工具将返回在任务名称的任何位置包含筛选器字符串的任务。

任务类型

您可以对任务数据执行的数据操作的类型。

任务可以是以下类型之一：

- 更正异常项。检查并修复记录中的错误。
- 更正重复项。检查重复记录群集，并通过记录中的值创建首选记录。
- 审阅异常项。在更正异常项的任务中审阅另一用户的工作。

- 审阅重复项。在根据记录群集创建首选记录的任务中审阅另一用户的工作。

注意: 任务类型可能包括投票任务。Business Glossary 用户处理投票任务。投票任务不是异常管理流程的一部分。

到期日期

任务的最终期限。人工任务定义更正异常项任务和更正重复项任务的到期日期。Analyst 工具计算更正异常项任务和更正重复项任务的到期日期。

状态

任务的状态

任务可具有以下状态之一：

- 已创建。任务没有所有者。
- 已分配。任务具有所有者。

所有者

当前任务所有者的名称。

创建时间

工作流创建任务的日期。

异常项工作区

“异常项”工作区是一个临时工作区，当您查看或打开某项任务时，将显示该工作区。“异常项”工作区包含**数据编辑**选项卡和**数据审计**选项卡。

数据编辑选项卡显示任务数据以及一些可用于在任务中更新记录或群集的选项。该选项卡还显示一些元数据列，Analyst 工具可使用这些列来跟踪您和其他用户对每个记录或群集所做的更新。

数据审计选项卡显示您和其他用户对任务数据所做更改的审计跟踪。您可以查看用户更改的字段、更改每个记录的用户的身份，以及更改日期。

在完成对任务的处理后，可以关闭“异常项”工作区。

审计跟踪操作

Analyst 工具会针对用户对任务中的记录所做的更改创建审计跟踪。要查看任务的审计跟踪，请在“异常”工作区中打开任务并选择**数据审计**选项卡。

审计跟踪显示用户在任务中更新的每个记录的记录数据和状态指示器值。此外，审计跟踪还标识执行更新操作的用户以及更新日期和时间。审计跟踪按时间顺序显示行。如果其他用户在之前的任务中更新了记录，审计跟踪将包括之前用户的更新。

查看审计跟踪时，请考虑以下规则和准则：

- 审计跟踪中的每一行都代表对记录的一次更新。如果您要对记录进行多次更新，则审计跟踪将为每次更新添加一行。
- 更改记录中的数据值时，审计跟踪将突出显示该值。将指针移动到该值上方可查看上一个值。审计跟踪还会标识更改记录的用户、更改日期以及更改的类型。

- 默认情况下，审计跟踪将显示每个记录的所有数据。如果 Analyst 工具没有向您授予查看任务中某个数据列的权限，审计跟踪将不会显示此列。如果用户在之前的任务中更新了记录值，并且您没有该值的查看权限，审计跟踪将不会显示此值。
- Informatica 域配置确定了审计跟踪数据的位置。Analyst 工具可以将所有审计跟踪数据写入单个数据库的表中。或者，Analyst 工具可以将每个任务的审计跟踪数据写入包含任务中记录的数据库。有关域中审计跟踪数据的信息，请联系 Informatica 域管理员。

异常管理流程

在 Analyst 工具中处理的任务表示数据质量周期中的一个阶段。该周期从组织决定验证数据集中数据的质量时开始。组织对数据质量感到满意时，该周期结束。组织可以连续运行数据质量周期。

异常管理阶段通常发生在数据质量周期即将结束时。早期阶段可以使用配置文件和映射来度量组织数据的质量以及提高数据质量。异常管理定义用户可以在当前周期内针对未达到数据质量目标的那些记录执行的手动操作。

您可能是负责实施数据质量流程的数据管理者团队中的一员，也可能是负责定义数据必须满足的数据质量标准的一个业务用户。无论您是上述哪种角色，您都可以拥有 Analyst 工具中的异常管理任务。

注意: 在许多组织中，数据管理者同时具有数据管理者角色和其他角色。您可能是组织内某个较大角色的一部分，负责某个数据质量目标。作为数据管理者，您可以对数据集执行任务，并将结果传递给承担数据管理职责的同事。

离散记录示例

您是一家零售组织的数据管理者。您担心产品库存记录可能包含错误。这些错误可能导致组织订购的产品量多于或少于可销售量。

您定义以下流程来调查和更正错误：

1. 您安排一名开发人员来配置一个或多个映射，以便查找和修复数据集中的错误。
这些映射还会计算数据集中每个记录的数值得分。得分表示记录的数据质量。一些记录具有边缘得分，这表示映射无法验证记录所包含的所有数据质量问题。
2. 开发人员配置另一个映射来读取数字得分。开发人员将该映射添加到一个包括映射任务和人工任务的工作流。
 - “映射”任务运行该映射。映射任务根据记录所包含的得分将记录写入不同的表中。
 - 人工任务将包含边缘得分的记录分发给您和其他用户可以在 Analyst 工具中打开的任务。
3. 您登录到 Analyst 工具，然后打开一个任务。任务将异常记录组织在一个或多个表中。每个表可以包含 100 个记录。
您对每个记录执行以下操作之一：
 - 更正记录中的错误，或者确定当前记录是否正确。
更新记录状态以指示记录有效。
 - 确定记录不包含任何有效数据。
更新记录状态以指示记录不是有效的。
 - 确定您无法验证记录的准确性。
更新记录状态以指示记录需要由另一个用户或另一个 Informatica 进程做进一步分析。

注意: 在更新记录之前，验证该任务在编辑模式下打开。要进入编辑模式，请在打开的任务中单击**编辑按钮**。

4. 完成对任务中所有记录的处理后，更新任务状态。任务状态指示记录已为数据质量流程中的下一阶段做好准备。

数据的下一阶段取决于人工任务的配置。例如，人工任务可能包含用于将记录分配给其他用户进行审阅的其他步骤。

在人工任务完成后，工作流的下一阶段随即开始。

重复记录示例

您是一家银行的数据管理者。您担心客户帐户表中的多个记录可能包含相同信息。重复记录可能表示数据输入错误，也可能表示具有欺诈性的客户活动。

您通过定义以下流程来查找重复记录并识别每个记录集的单个首选版本：

1. 您安排一名开发人员来配置一个或多个映射，以便识别重复记录。
映射将计算一组数字得分，这些得分用于表示记录中两个数据值之间的重复程度。高分表示重复记录，低分表示唯一记录。一些记录具有边缘得分，这表示记录的重复状态不确定。
2. 开发人员配置另一个映射来读取数字得分。开发人员将该映射添加到一个包括映射任务和人工任务的工作流。

- “映射”任务运行该映射。映射任务根据记录所包含的得分将记录写入不同的表中。
- 人工任务将包含边缘得分的记录分发给您和其他用户可以在 Analyst 工具中打开的任务。

3. 您登录到 Analyst 工具，然后打开一个任务。

Analyst 工具将记录组织为一系列群集。每个群集包含两个或更多个具有相似信息的记录。默认情况下，群集中的第一个记录是首选记录。

4. 打开一个群集，然后分析其包含的记录。

您可以在每个群集中执行以下操作：

- 检查每个记录数据列中的数据值。选择每个列中的最准确值并将该值升级为首选记录。
您可以编辑所选的值，也可以在其他群集中搜索包含通用值的记录。
- 如果某个记录不属于当前群集，您可以将其移动到另一个群集或者为该记录创建一个群集。
- 您更新群集状态以指示您已审阅该群集。在验证每个群集中的当前首选记录后，完成该任务。

注意：在更新记录之前，验证该任务在编辑模式下打开。要进入编辑模式，请在打开的任务中单击**编辑**按钮。

5. 完成对任务中所有群集的处理后，更新任务状态。任务状态指示记录已为数据质量流程中的下一阶段做好准备。

数据的下一阶段取决于人工任务的配置。例如，人工任务可能包含用于将群集分配给其他用户进行审阅的其他步骤。

在人工任务完成后，工作流的下一阶段随即开始。

登录到 Analyst 工具

使用 Analyst 工具 URL 可登录到 Analyst 工具。登录 Analyst 工具时，您需要指定 Informatica 登录名、密码及本地域或 LDAP 安全域。

1. 启动 Microsoft Internet Explorer 或 Google Chrome 浏览器。

2. 在“地址”字段中，输入 Analyst 工具的 URL：
`http[s]://<fully qualified host name>:<port number>/analyst/`
3. 如果域使用 LDAP 或本地身份验证，请在登录页面输入登录名和密码。
4. 选择**本地**或特定安全域的名称。
如果 Informatica 域使用 LDAP 或 Kerberos 身份验证，将显示“安全域”字段。如果您不知道用户帐户所属的安全域，请联系 Informatica 域管理员。
5. 单击**登录**。
Analyst 工具会在**启动**工作区上打开。

第 2 章

异常任务

本章包括以下主题：

- [异常任务概览, 16](#)
- [异常任务的类型, 16](#)
- [异常项工作区中的异常项任务, 17](#)
- [在异常任务中修复错误, 19](#)
- [异常记录状态指示器, 20](#)
- [异常任务的筛选器选项, 22](#)
- [查找并替换异常记录中的值, 24](#)
- [将说明添加到记录, 26](#)
- [审查中的异常任务, 26](#)
- [异常任务的规则和准则, 27](#)

异常任务概览

异常任务用于标识可能包含数据错误的记录。这些记录可能包含不准确的数据，也可能包含空字段。如果早期进程无法解决数据集中的所有数据质量问题，工作流会生成异常任务。打开异常任务时，Analyst 工具会识别每个记录中包含未解决问题的字段。

异常任务的类型

工作流可为您分配更正异常项任务或审阅异常项任务。

处理更正异常项任务时，需要检查任务中的每个记录并决定是否要更新记录数据。处理审阅异常项任务时，需要验证其他用户在早期任务中对每个记录做出的决定。

每个任务类型中的工作区布局均相同。在这两种任务类型中，记录数据和状态指示器的更新方式相同。执行审阅异常项任务时，可以使用其他选项来接受或拒绝先前用户的工作。

更正异常项步骤

从“启动”工作区中打开更正异常项任务。该任务将在“异常项”工作区的“数据编辑”选项卡中打开。

执行更正异常项任务时，请执行以下步骤：

1. 打开任务。
2. 检查每个记录中突出显示的数据字段。
可以使用箭头在任务的记录中移动。使用筛选器选项可显示或隐藏包含指定的数据质量问题或状态的记录。
3. 修复数据字段包含的错误。
4. 更新每个记录的状态以反映当前记录数据。
请选择以下选项之一：
 - 接受。确定当前数据对于业务是可接受的。
 - 拒绝。确定当前数据对于业务是不可接受的。
 - 重新处理。无法确定当前数据的状态。
5. 更新记录以指示您已审阅记录数据。您可以独立于其他状态值来设置“已审阅”状态值。
使用筛选器选项可显示或隐藏具有通用状态值的记录。
6. 可选择向记录添加说明。例如，可以添加一条说明来解释您拒绝某个记录的原因。
7. 完成对任务中所有记录的处理后，请更新任务状态。

审阅异常项步骤

从“启动”工作区中打开审阅异常项任务。该任务将在“异常项”工作区的“数据编辑”选项卡中打开。

执行审阅异常项任务时，请执行以下步骤：

1. 打开任务。
2. 审阅每个记录。检查记录数据，然后检查上一个用户为每个记录设置的状态指示器。
 - 如果您同意当前记录内容，则不用进行任何更改。如果您不同意任何记录的内容，可更新该记录。
 - 如果您同意当前记录状态，则不用进行任何更改。如果您不同意记录状态，可更新该状态。
使用筛选器选项可显示或隐藏具有通用状态值的记录。
3. 验证每个记录的审阅状态。审阅状态指示您批准还是拒绝记录。审阅状态将取代上一个用户应用于记录的状态。
4. 可选择向记录添加说明。如果用户已在早期任务中向记录添加一条说明，您添加的说明将替代早期的说明。
5. 完成对任务中所有记录的处理后，请更新任务状态。

异常项工作区中的异常项任务

打开某个异常项任务时，“异常项”工作区将列出该任务中的记录并显示您可以为该任务设置的选项。该任务将在“数据编辑”选项卡中打开。该选项卡显示一组适用于更正异常项任务和审阅异常项任务的通用选项。

Analyst 工具可更改包含未解决的数据质量问题的任何单元格的背景颜色。检查记录，并更新单元格。更新数据值时，Analyst 工具将清除单元格背景色，并在该单元格中添加一个绿色对勾。您可以使用光标显示单元格中的上一个值。

完成对记录的处理后，更新记录状态。完成对任务中所有记录的处理后，更新任务状态以指示记录已准备好进入工作流的下一阶段。

Start

Glossary

Discovery

Design

Scorecards

Exceptions

ExceptionStep

Customer_Cleanse_Analysis { 1 - 50 }

Save

Cancel

Undo

Redo

0 of 50

Records are reviewed

0

Accepted

0

Reprocessed

0

Rejected

Data Editing

Data Audit

Find & Replace

Filter

| In | All Strings | Find | Replace with | <input type="checkbox"/> Case sensitive | <input type="checkbox"/> Match whole field | < | > | Replace | Replace All | | |
|--------------------------|-------------|------------|------------------|---|--|-----------------|-----------------|----------------|----------------|-----|---------|
| <input type="checkbox"/> | | CUSTOMERID | NAME | COMPANY | ADDRESS1 | ADDRESS2 | ADDRESS3 | CITY | STATE | ZIP | COUNTRY |
| <input type="checkbox"/> | 1 | 12 | Sr. M. Adella | 74-76 HAUPTSTR | MEERBUSCH 4... | MEERBUS... | | 40668 | UNITED STAT... | | |
| <input type="checkbox"/> | 2 | 12 | Sr. M. Adella | 74-76 HAUPTSTR | MEERBUSCH 4... | MEERBUS... | | 40668 | USA | | |
| <input type="checkbox"/> | 3 | 496 | Sr. M. Adella | Kewill System... | ASH HOUSE | FARIFILED AVE | STAINES TW18... | STAINES T... | USA | | |
| <input type="checkbox"/> | 4 | 496 | Kewill System... | ASH HOUSE | FARIFILED AVE | STAINES TW18... | STAINES T... | USA | | | |
| <input type="checkbox"/> | 5 | 641 | Kewill System... | 8 M HLENSTR | FARIFILED AVE | L NINGEN | 49624 | UNITED STAT... | | | |
| <input type="checkbox"/> | 6 | 641 | 8 M HLENSTR | FARIFILED AVE | NONE | 96317 | USA | | | | |
| <input type="checkbox"/> | 7 | 641 | Tamara Am... | 4 OFENERDIEK... | FARIFILED AVE | OLDENBU... | 26125 | USA | | | |
| <input type="checkbox"/> | 8 | 10110088 | Konrad An... | 1 WALDSTR | BAD K NIG 64732 | BAD K NIG | 64732 | USA | | | |
| <input type="checkbox"/> | 9 | 10110088 | test | WESTWOOD | BAD K NIG 64732 | MARGATE... | USA | | | | |
| <input type="checkbox"/> | 10 | 10110094 | Phytopharm plc | WESTWOOD | MARGATE CT9 ... | GODMAN... | UNITED STAT... | | | | |

14

Go to page:

1

Show rows:

20

1-20 of 100

1. 撤消和恢复选项
撤消任务中数据或元数据的最新更新。恢复已选择撤消的任何更新。
无法撤消在任务中的所有页面上执行的查找和替换操作。
2. 任务操作菜单
对当前任务执行操作。导出任务数据，或将任务发送至工作流的下一步。
3. 查找和替换选项
查找并替换与您输入的字符串相匹配的数据值。
4. 突出显示工具
查看未使用状态指示器的记录。
5. 查找和替换
显示或隐藏用于在异常记录中查找和替换数据值的选项。
6. 元数据筛选器选项
根据记录元数据显示或隐藏任务中的记录，如数据质量问题的类型或记录状态。
7. 记录操作菜单
更新您选择的记录的状态。例如，接受记录、拒绝记录或清除记录状态。
8. 数据筛选器重置
清除您输入的作为列数据筛选器的任何值。
9. 说明列
阅读任务所有者向记录添加的说明。
10. 记录状态列
查看任务所有者对记录执行的最近操作的状态指示器。
11. 记录审阅状态列
查看任务所有者为记录定义的最近状态的状态指示器。
12. 列数据筛选器选项
根据一个或多个数据列中数据显示或隐藏任务中的记录。

13. 记录选择器
选择记录操作要应用到的记录。
14. 导航选项
移动到任务中的不同页面。

在异常任务中修复错误

可以在更正异常项任务或审阅异常项任务中更新记录。检查任务中的记录，然后更新包含数据质量错误的字段。

可以在**数据编辑**选项卡中编辑记录。

1. 打开异常任务。
验证该任务显示**数据编辑**选项卡。
2. 单击**编辑**。
3. 检查记录。选择一个包含可修复错误的字段。
将指针移动到该字段上方，以查看该字段包含的数据错误类型。
4. 在数据字段中输入正确的值。
字段外观发生变化。Analyst 工具清除该字段中的背景色，并在右上角添加一个绿色对勾。
5. 保存该任务。

对该任务中的其他记录重复这些步骤。

注意: 人工任务属性可能指示每个错误单元格必须包含一个值。在这种情况下，只有在每个错误单元格中输入一个值后，才能完成该任务。

异常任务中的验证错误

如果尝试输入一个对于所在列无效的数据值，Analyst 工具将显示验证错误。不能在数值列中输入字符串数据值。如果数据值超过列中的最大字符长度，则不能在列中输入该数据值。

Analyst 工具在查找和替换操作过程中也会显示验证错误。如果替换值在目标列中无效，则不能将数值替换为该替换值。

例如：印度的城市名称

您是跨国组织的数据管理员，并且正在处理标识印度古老城市名称的异常任务。您使用一个长期和古老城市名称的列表更新任务中的记录针对每个城市，使用“全部替换”选项更新名称。使用默认选项替换所有字符串数据中的值。

包含城市名称的某个列指定了无法满足当前名称的最大字符长度。例如，名称 *Thiruvananthapuram* 替换了印度喀拉拉邦中的 *Trivandrum* 名称。当 Analyst 工具尝试更新列中的名称时，操作将失败并且 Analyst 工具将显示错误。

异常记录状态指示器

完成对记录的处理后，设置记录状态。该状态指示另一用户或另一数据进程在任务完成时可对记录所执行的操作。设置表示记录处理完毕时记录数据状况的状态。

从“数据编辑”选项卡的**记录操作**菜单上选择状态指示器。状态指示器作为图标显示在任务中数据行旁边。如果用户在早期任务中设置记录状态，则可更新或删除状态。

完成对记录的处理后，选择下列状态指示器之一：

接受记录

指示可以接受将记录与组织数据一同存储。

可以在更正异常项任务或审阅异常项任务中接受记录。

拒绝记录

指示无法接受将记录与组织数据一同存储。

拒绝您确定无法正确标识数据库中的业务实体的记录。下游进程可能会丢弃您在数据库表中拒绝的记录。

可以在更正异常项任务或审阅异常项任务中拒绝记录。

重新处理记录

指示记录包含一个您无法验证的数据质量问题。

无法确定记录数据的准确性时，重新处理记录。下游任务中的另一个用户可能会验证或更新记录数据。或者，下游进程可能会将记录写入表，以便在映射中分析和更正。

可以在更正异常项任务或审阅异常项任务中重新处理记录。

标记为已查看

指示您已审阅记录数据。

将记录标记为已审阅可告知其他用户您已检查该记录。该状态不描述记录的数据质量，也不为记录数据指定任何其他操作。

将记录标记为已审阅时，可以设置其他状态指示器。例如，可以重新处理标记为已审阅的记录。

可以在任何任务中将记录标记为已审阅。

审批记录编辑

指示您已在审阅异常项任务中分析记录并且确定可以接受将记录与组织数据一同存储。

也可以在审阅重复项任务中批准记录编辑。

拒绝记录编辑

指示您已在审阅异常项任务中分析记录并且确定无法接受将记录与组织数据一同存储。

也可以在审阅重复项任务中拒绝记录编辑。

清除记录状态

从记录中清除以下任一状态指示器：

- 接受记录
- 拒绝记录
- 重新处理记录

如果确定当前记录状态不正确，则清除该状态。

可以在更正异常项任务或审阅异常项任务中清除状态指示器。

清除审阅者状态

从记录中清除以下任一状态指示器：

- 审批记录编辑
- 拒绝记录编辑
- 标记为已查看

如果确定当前审阅状态不正确，则清除该状态。例如，可以清除另一用户在早期任务中已审阅的记录状态。

可以在任何任务中清除状态指示器。

注意：状态指示器必须表示当前记录状态。例如，您可能因为在记录中发现错误而拒绝该记录。如果修复错误，请更新记录状态。

设置异常记录的状态

在完成对记录的处理后，可以为记录设置状态指示器。可以在单个操作中将一个状态指示器应用到多个记录。

可以在更正异常项任务和审阅异常项任务中设置记录状态。如果您不同意当前记录状态，可清除该状态。

1. 打开异常任务。

验证该任务显示**数据编辑**选项卡。

2. 单击**编辑**。

3. 选择一个或多个记录。

要选择某个记录，请选中位于该记录数据行开头的复选框。要选择任务中的所有记录，请选中用于标识列名称的行中的复选框。

选择所有记录时，请考虑以下因素：

- 用于选择所有记录的操作适用于当前工作区中的记录。
- 如果将筛选器应用到该任务，用于选择所有记录的操作将适用于工作区中满足筛选条件的记录。

4. 打开**记录操作**菜单，然后选择要应用到记录的状态指示器。

可以在更正异常项任务和审阅异常项任务中选择以下状态指示器：

- 要指示记录包含有效数据，请选择**接受记录**。
- 要指示记录不包含有效数据，请选择**拒绝记录**。
- 要指示记录需要进一步分析，请选择**重新处理记录**。
- 要指示您已检查记录，请选择**标记为已查看**。可以在选择该指示器时并行选择另一个指示器。

可以在审阅异常项任务中选择以下状态指示器：

- 要验证某个记录适合包含在业务数据中，请选择**标记为已接受**。
- 要验证某个记录不适合包含在业务数据中，请选择**标记为已拒绝**。

5. 如果您不同意当前记录状态，可选择以下两个选项或其中之一：

- 要清除用于接受、拒绝或重新处理记录的指示器，请选择**清除记录状态**。
- 要清除用于指定审阅类型的指示器，请选择**清除查看状态**。

对该任务中的其他记录重复这些步骤。

注意：“记录操作”菜单包括一个用于将说明添加到记录的选项。将说明添加到记录，用于说明所做的更改以及选择的状态指示器。

异常任务的筛选器选项

使用筛选器显示或隐藏 Analyst 工具为某个任务显示的记录。您可以在**数据编辑**选项卡和**数据审计**选项卡上定义筛选器。

您可以使用记录数据值或记录元数据定义筛选器。使用记录数据筛选器查找共享一个或多个通用数据值的记录。使用元数据筛选器查找共享通用数据质量问题或状态的记录。

例如，您可以定义一个筛选器以检索指定时间段内用户接受为有效记录的所有记录。Analyst 工具检索满足筛选条件的记录。可以审阅每个记录并验证用户做出的决定。

您可以对每个选项卡应用多个筛选器。应用多个筛选器时，Analyst 工具将返回满足所有筛选条件的记录。

记录数据筛选器

使用记录数据筛选器查找在指定分享列中共享某数据值的记录。

每一列都有一个筛选器选项。要应用筛选器，请在列顶部为该筛选器输入一个值。Analyst 工具将显示包含在所选字段中输入的值的记录

您可以在任务数据的多个列上定义筛选器。在多个列上定义筛选器时，Analyst 工具显示满足定义的所有筛选器的记录。

数据编辑选项卡上的元数据筛选器

使用数据编辑选项卡上的元数据筛选器标识一组具有通用元数据特性的记录。

“数据编辑”选项卡包含以下筛选器选项：

问题类型

指示工作流在记录数据中标识的数据质量问题的类型。选择某个问题时，Analyst 工具会显示包括该问题的所有记录。

将指针移动到字段上方可查看该字段表示的数据质量问题。

优先级

指示“映射”任务为记录中的数据质量问题分配的优先级。优先级是一个数值。选择某个优先级值时，Analyst 工具会显示包括该优先级的所有记录。

状态

指示您为任务中的记录分配的状态。选择某个状态时，Analyst 工具会显示使用该状态的所有记录。

可以选择以下状态选项：

- 任意。任务中的任意记录（不考虑状态）。
- 已接受。适合与组织数据永久存储在一起的记录。
- 已拒绝。不适合与组织数据永久存储在一起的记录。
- 已重新处理。需要在其他应用程序中进行进一步分析的记录。
- 空。记录不含当前状态。

查看

指示当前任务中记录的审阅状态。

可在以下审阅选项中进行选择：

- 所有记录。

- 已接受。用户在审阅任务中标记为已接受的记录。
- 已拒绝。用户在审阅任务中标记为已拒绝的记录。
- 已审阅。用户标记为已审阅的记录。
- 空。记录不含当前审阅状态。

数据审计选项卡上的元数据筛选器

使用“数据审计”选项卡上的元数据筛选器可标识用户在指定时间段内处理的记录。此外，还可以查找共享通用状态指示器的记录。

无法在“数据审计”选项卡上编辑记录数据。要编辑记录数据，请使用“数据编辑”选项卡。

“数据审计”选项卡包含以下筛选器选项：

更新日期

定义要应用到任务数据的时间。定义某个时间段时，Analyst 工具会返回用户在该时间段内更新的所有记录。

更新者

标识 Analyst 工具用户。标识某个用户时，Analyst 工具会返回该用户更新的所有记录。

状态

指示您为任务中的记录分配的状态。选择某个状态时，Analyst 工具会显示使用该状态的所有记录。

可以选择以下状态选项：

- 任意。任务中的任意记录（不考虑状态）。
- 已接受。适合与组织数据永久存储在一起的记录。
- 已拒绝。不适合与组织数据永久存储在一起的记录。
- 已重新处理。需要在其他应用程序中进行进一步分析的记录。
- 空。记录不含当前状态。
- 已清除。用户已清除状态的记录。

查看

指示当前任务中记录的审阅状态。

可在以下审阅选项中进行选择：

- 所有记录。
- 已接受。用户在审阅任务中标记为已接受的记录。
- 已拒绝。用户在审阅任务中标记为已拒绝的记录。
- 已审阅。用户标记为已审阅的记录。
- 空。记录不含当前审阅状态。
- 已清除。用户已清除审阅状态的记录。

在异常任务中筛选记录

使用元数据筛选器检索一组共享您指定的数据特性的记录。可以更新满足“数据编辑”选项卡上的筛选条件的记录。

- 1. 打开更正异常项任务或审阅异常项任务。
选择“数据编辑”选项卡或“数据审计”选项卡。
- 2. 单击**筛选器**。
“筛选器”面板将打开。
- 3. 选择要应用于任务数据的筛选器条件。
“数据编辑”选项卡和“数据审计”选项卡显示不同的筛选器选项组。
- 4. 单击**应用筛选器**。

Analyst 工具检索满足筛选条件的记录。Analyst 工具将您定义的筛选条件显示在任务群集列表之上。

注意: 如果应用筛选器后在“数据编辑”选项卡上选择所有记录，请在当前工作区视图中选择满足筛选条件的所有记录。不要选择任务中的所有记录。

查找并替换异常记录中的值

您可以在任务的记录中搜索指定的数据值。可以将该值替换为另一个值。可以在单个记录中查找并替换某个值，也可以在单个操作中查找并替换多个记录中的值。

在配置查找和替换选项时，可以按名称选择单个列。或者，也可以配置选项来搜索任何使用日期、字符串或数值数据类型的列。

当按数据类型搜索任务时，Analyst 工具会在使用该数据类型的任何列中搜索您输入的值。您可以查找并替换字符串数据列中包含的任何值，包括数值和日期。可以对字符串列数据执行区分大小写的搜索。可以搜索字符串数据列中某个字段内容和搜索值之间的部分匹配项或完整匹配项。查找部分匹配项时，Analyst 工具会替换您在记录字段中输入的字符的所有实例。

下表显示了对某个名称数据字符串执行部分匹配项查找和替换操作后的结果：

| 搜索值 | 替换值 | 要搜索的字符串 | 结果 |
|-------|-------|---------------------------|---------------------------|
| SMITH | SMYTH | JOHN SMITH AND MARY SMITH | JOHN SMYTH AND MARY SMYTH |

搜索日期时，请以日期列所使用的日期和时间戳格式输入日期值。可以使用您指定的日期值替换日期列中的空值。不能在日期数据列或数值数据列中搜索部分日期值或部分数值。无法对日期数据列或数值数据列的内容执行区分大小写的搜索。

筛选任务数据时，可以查找并替换值。查找和替换操作适用于满足筛选条件的记录。对筛选后的数据执行搜索时，搜索操作会打开包含符合筛选条件的记录的第一页任务数据。

单击**查找和替换**可显示或隐藏任务中的选项。

正在替换异常记录中的单个值

查找并替换指定列或者任意日期、字符串或数值列中的日期值。使用筛选器查找并替换任务中记录子集中的值。

- 1. 打开异常任务。

2. 在**数据编辑**选项卡上，单击**编辑**。
3. 单击**查找和替换**。

Analyst 工具将在记录列表上方显示查找和替换选项。
4. 在**查找**字段中输入一个数据值。
5. 指定一个要搜索的列。或者，指定一种数据类型。

如果选择一个数据类型，则 Analyst 工具将搜索使用该数据类型的列。默认情况下，Analyst 工具会在所有字符串数据列中搜索数据值。
6. 如果在字符串数据列中搜索数据值，请选择或清除以下选项：
 - 区分大小写。对字符串数据执行区分大小写的搜索。
 - 匹配整个字段。搜索输入值和记录中某个字段完整内容之间的匹配项。

无法对日期数据列或数值数据列执行区分大小写的搜索或部分数据搜索。
7. 单击**查找**。

Analyst 工具将突出显示第一个包含指定值的记录。可使用箭头在这些记录之间移动。
8. （可选）在**替换为**字段中输入一个值。
9. 要将突出显示的值替换为您输入的值，请单击**替换**。

重复该步骤可在任务中查找和替换该值的其他实例。

正在替换异常记录中的多个值

使用“全部替换”选项一次替换值的多个实例。您可以查找并替换当前页面、所有页面的记录中的值或所选记录中的值。使用筛选器查找并替换任务中记录子集中的值。

1. 打开异常任务。
2. 或者，选择多个记录。

选择多个记录将“全部替换”操作限制在任务的一组记录中。例如，您可以选择满足定义的筛选条件的所有记录。
3. 在**数据编辑**选项卡上，单击**编辑**。
4. 单击**查找和替换**。

Analyst 工具将在记录列表上方显示查找和替换选项。
5. 在**查找**字段中输入一个数据值。
6. 指定一个要搜索的列。或者，指定一种数据类型。
7. 在**替换为**字段中输入一个值。
8. 单击**全部替换**。

选择以下选项之一：

 - 在此页面上替换。替换您在当前页面的记录中输入的值的的所有实例。
 - 在所有页面上替换。替换您在任务的所有记录中输入的值的的所有实例。
 - 在所有选定对象上替换。替换您在所选记录中输入的值的的所有实例。

注意：此操作不会替换您没有查看和编辑权限的任何列中的数据值。

将说明添加到记录

添加说明可描述您对记录执行的操作。可以在“数据编辑”选项卡和“数据审计”选项卡上查看说明。

1. 打开包含该记录的任务。
2. 在“异常项”工作区中，选择**数据编辑**选项卡。
3. 单击**编辑**。
4. 可以使用复选框选项来选择某个记录。
可以选择多个记录。选择多个记录时，将相同的说明添加到选择的所有记录。
5. 打开“记录操作”菜单，然后选择**添加说明**。
Analyst 工具将打开一个文本输入对话框。
6. 将说明文本输入到该对话框。

审查中的异常任务

在审阅异常任务中审阅记录时，需要验证其他用户在早期任务中完成的工作。可以在审阅异常任务 and 更正异常任务中执行相同的操作。

检查任务中的记录，然后确认先前用户分配给每个记录的状态。如果在记录中发现错误，请更新该记录。如果您不同意当前记录状态，可更新或删除该状态。

使用**记录操作**菜单选项更新记录状态。完成对任务的处理后，使用**任务操作**菜单选项更新任务状态。

执行审阅异常项任务

验证该任务中的记录尽可能准确且完整。

1. 打开审阅异常项任务。
2. 验证记录数据，然后验证先前用户分配给每个记录的状态。
3. 如果您决定更新记录或者更改记录状态，请单击**编辑**。
要修复错误，请单击包含该错误的字段并输入正确的数据值。
要更新记录状态，请执行以下操作之一：
 - 要指示记录包含有效数据，请选择**接受记录**。
 - 要指示记录不包含有效数据，请选择**拒绝记录**。
 - 要指示记录需要进一步分析，请选择**重新处理记录**。
 - 要指示您已检查记录，请选择**标记为已查看**。可以在选择该指示器时并行选择另一个指示器。
 - 要验证某个记录适合包含在业务数据中，请选择**标记为已接受**。
 - 要验证某个记录不适合包含在业务数据中，请选择**标记为已拒绝**。
 - 要清除用于接受、拒绝或重新处理记录的指示器，请选择**清除记录状态**。
 - 要清除用于指定审阅类型的指示器，请选择**清除查看状态**。
4. 保存该任务。
完成对任务的处理后，更新任务状态。任务状态指示记录已为工作流的下一阶段做好准备。

异常任务的规则和准则

审阅和更新数据集中的异常记录的过程为协作式过程。您可能会处理其他用户在早期任务中分析的记录。在任务完成后，您处理的记录也可能会传递给其他用户。每个用户均可审阅和更新其他用户的工作。

执行异常记录任务时，请考虑以下规则和准则：

- 设置记录的状态时，可以指示您已审阅该记录。审阅状态并不描述该记录的准确性或数据质量。例如，可以接受、拒绝和重新处理一系列记录，也可以将每个记录标记为已审阅。

最佳做法是将检查过的每个记录都标记为已审阅。该状态将向下游任务中的用户确认其他用户已检查该记录。更新记录中的数据后，请将该记录标记为已审阅，而不考虑记录上是否存在或缺少其他状态指示器。

- 更新记录时，可以设置任何状态指示器。例如，可以更新重新处理的记录。您执行的更新可帮助下一个用户或下游数据进程分析和修复数据。
- 记录上的数据值和状态指示器可以在多个任务中彼此独立地进行更改。更新记录时，验证数据值和状态指示器最新且准确。您所做的更改可以否决另一个用户所做的决定。
- 审计跟踪将存储用户对记录的数据值和状态指示器所做的每次更改。审计跟踪不会存储对您在记录中添加的说明文本的更改。
- 您处理的数据可以传递给数据更正任务或数据审阅任务。例如，在工作流中配置人工任务的开发人员可能会依次指定多个更正异常项任务。该开发人员可能会在审阅重复项任务之后安排更正重复项任务。

第 3 章

重复的记录任务

本章包括以下主题：

- [重复记录任务概览, 28](#)
- [首选记录, 29](#)
- [重复记录任务类型, 29](#)
- [异常项工作区中的重复记录任务, 30](#)
- [编辑群集, 32](#)
- [重复记录状态指示器, 32](#)
- [在多个群集中查找记录, 34](#)
- [在群集之间移动记录, 34](#)
- [将记录移到空群集, 35](#)
- [将说明添加到群集, 35](#)
- [重复任务的筛选器选项, 36](#)
- [审阅重复项任务, 37](#)
- [重复记录任务的规则和准则, 39](#)

重复记录任务概览

重复记录任务标识数据集中可能包含重复信息或冗余信息的记录。该任务将记录显示为一系列群集。每个群集均标识一组包含相似或相同数据值的记录。

处理重复记录任务时，需要审阅每个群集并确定群集中的记录是否彼此重复。如果记录重复，则定义群集表示的记录的首选版本。如果群集包含唯一记录，则将该记录移至空群集。如果某一记录与其他群集中的记录匹配，则可以将该记录移至其他群集。

在审阅所有群集并为每个群集定义一个首选记录后，完成任务。

注意：当两个或更多记录表示源数据集中的同一实体时，即表明这些记录重复。记录可以包含相似数据，但不能代表相同的业务信息。组织可以定义用于标识重复记录的业务规则。

首选记录

处理群集时，需要创建或验证群集所代表的记录的最准确且最完整的版本。您创建或验证的记录就是首选记录。

群集中的第一行包含首选记录数据。要更新首选记录，需将来自群集中其他记录的数据值升级到首选记录。

默认情况下，Analyst 工具使用群集中第一个记录的数据来填充首选数据行。Analyst 工具将突出显示首选数据行。Analyst 工具还将突出显示包含默认数据的记录。将数据值升级到首选记录时，Analyst 工具将突出显示所升级的值。可以编辑添加到首选记录的值。无法编辑群集中其他记录的值。如果默认记录正确，则可接受默认记录。

注意：首选记录不是群集的成员。首选记录是工作流在重复记录数据库中创建的唯一记录。

重复记录任务类型

工作流可以为您分配更正重复项任务或审阅重复项任务。

处理更正重复项任务时，需要检查任务中每个群集的记录。打开群集并验证群集中包含可用的最准确数据的首选记录。处理审阅重复项任务时，需要验证其他用户在早期任务中对每个群集做出的决定。

每个任务类型中的工作区布局均相同。您在每个任务类型中使用的工作区选项均相似。在这两种任务类型中，群集数据和状态指示器的更新方式相同。执行审阅重复项任务时，可以使用其他选项来接受或拒绝先前用户的工作。

更正重复项步骤

从“启动”工作区中打开审阅重复项任务。该任务将在“异常项”工作区的“数据编辑”选项卡中打开。

执行更正重复项任务时，请执行以下步骤：

1. 打开任务并进入编辑模式。
Analyst 工具将显示任务中的第一个群集。可以在当前群集中工作，也可以打开其他群集。使用筛选器选项可显示或隐藏包含指定数据值或指定状态的群集。
2. 检查当前群集中的首选记录和其他记录。
如果其他记录中的字段所包含的值比首选记录中同一字段的值更准确，请将该值升级到首选记录。
3. 或者，执行下列任一任务来验证群集包含的数据是否最准确：
 - 编辑首选记录中的值。
 - 在其他群集中搜索包含相似值的记录。
 - 将记录从一个群集移至其他群集。
 - 创建一个群集并将记录移至该群集。
4. 更新群集状态以指示您已审阅该群集。
5. 完成对任务中所有群集的处理后，请更新任务状态。

审阅重复项步骤

从“启动”工作区中打开更正重复项任务。该任务将在“异常项”工作区的“数据编辑”选项卡中打开。

执行审阅异常项任务时，请执行以下步骤：

1. 打开任务并进入编辑模式。
Analyst 工具将显示任务中的第一个群集。可以在当前群集中工作，也可以打开其他群集。使用筛选器选项可显示或隐藏包含指定数据值或指定状态的群集。
2. 审阅每个群集的内容。验证或更新用户在其他任务中为该群集分配的状态。
使用筛选器查找用户已审阅或未审阅的群集。
 - 如果您同意群集的内容和状态，则不进行更改。
 - 如果您不同意任一群集的内容，请更新该群集。
 - 如果您不同意任一群集的状态，请更新状态。
3. 完成对任务中所有群集的处理后，请更新任务状态。

异常项工作区中的重复记录任务

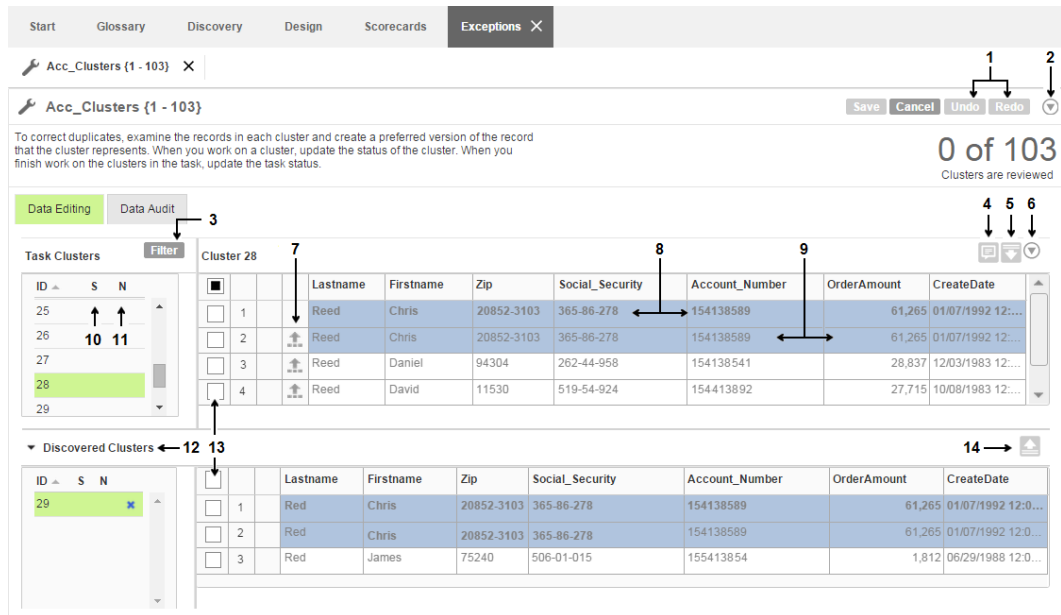
打开重复记录任务时，“异常项”工作区将列出任务中的群集并显示任务中第一个群集的记录。该任务将在“数据编辑”选项卡中打开。该选项卡显示一组适用于更正重复项任务和审阅重复项任务的通用选项。

Analyst 将在工作区左侧列出任务中的群集。选择一个群集以查看其包含的记录。群集中的第一个记录为当前首选记录。审阅群集中的其他记录，并确定其他记录是否包含属于首选记录的任何数据。要将值从某个记录添加到首选记录，请单击该值。要将首选记录替换为其他记录，请选择记录并使用列选项来升级记录。

将值添加到首选记录时，工作区将突出显示所添加的值。将首选记录替换为其他记录时，工作区将突出显示所升级的值。要查看对任务所做的更改的完整列表，请选择**数据审计**视图。

完成对群集的处理后，更新群集状态。完成对任务中群集的处理后，更新任务状态以指示群集已为工作流的下一阶段做好准备。

下图显示了在编辑模式下打开包含群集的任务时所出现的**数据编辑**选项卡：



处理任务时，请使用以下选项：

1. 撤消和恢复选项
撤消任务中数据或元数据的最新更新。 恢复已选择撤消的任何更新。
2. 任务操作菜单
对当前任务执行操作。 导出任务数据，或将任务发送至工作流的下一步。
3. 筛选器选项
查找任务中满足指定条件的群集。 应用筛选器时，请关闭当前群集并打开满足筛选条件的第一个群集。
筛选器还适用于您发现的群集。 应用筛选器时，“已发现的群集”列表将显示已发现的列表内满足筛选条件的群集。
4. 说明选项
读取用户在“任务群集”列表上为当前群集所添加的说明。
5. 将记录移至更低群集
将在更高群集中选择的一个或多个记录移至更低群集。 此选项在您打开“已发现的群集”列表中的群集时使用。
6. 群集操作菜单
更新“任务群集”列表上当前群集的状态。 可以向群集添加说明、将群集标记为已审阅，也可以清除群集的审阅状态。
7. 升级选项
使用选定行中的数据更新首选记录中的数据。
8. 首选记录数据
群集中首选记录的当前数据。
9. 首选记录数据的源记录
群集中为首选记录提供源数据的记录。 默认情况下，首选记录是群集中具有序列 ID 值的记录。
10. 群集审阅状态列
查看群集当前状态的状态指示器。
11. 群集说明列
指示用户添加说明的群集。
12. 已发现的群集选项
用于查找包含一个或多个指定数据值的群集以及用于创建群集。

13. 记录选择选项
用于选择一个或多个记录。
14. 将记录移至更高群集
将在更低群集中选择的一个或多个记录移至更高群集。此选项在您打开“已发现的群集”列表中的群集时使用。

编辑群集

检查群集中的首选记录 and 任何其他记录。验证首选记录是否包含群集所表示的业务实体的最准确数据。

如果当前首选记录正确，则不进行更改。

1. 打开某一更正重复项任务。
任务中的群集列表将以列的形式显示在“数据编辑”选项卡上。默认情况下，将打开第一个群集。
2. 也可以选择其他群集。
例如，如果当前群集的状态为“已审阅”，则选择其他群集。
3. 检查群集中的记录。确定首选记录是否包含群集所表示的业务实体的最准确数据。
首选记录是群集中的第一个记录。Analyst 工具突出显示首选数据的首选记录和源记录。
4. 单击**编辑**。
5. 要更新首选记录，请执行以下任一操作：
 - 将来自其他记录的数据值升级到首选记录。
要升级某个值，请单击源记录中的数据字段。您选择的值将替换首选记录中相同列的值。
 - 更新首选记录中的数据值。
如果群集中没有记录包含数据的准确版本，请更新首选记录。
 - 使用群集中的其他记录替换当前首选记录。
要替换首选记录，请单击包含该记录的行中的升级工具。
6. 将群集保存到任务。

在完成群集中的工作后，请将群集状态更新为“已审阅”。

相关主题：

- [“在多个群集中查找记录” 页面上 34](#)
- [“将记录移到空群集” 页面上 35](#)

重复记录状态指示器

完成对群集的处理后，设置群集状态。该状态指示另一用户或另一数据进程在任务完成时可对群集数据所执行的操作。设置表示群集处理完毕时首选记录状况的状态。

从“数据编辑”选项卡的**群集操作**菜单上选择状态指示器。状态指示器以图标形式显示在群集列表中。如果用户在早期任务中设置群集状态，则可更新或清除状态。

完成对群集的处理后，选择下列状态指示器之一：

标记为已查看

指示您已审阅群集数据。

将群集标记为已审阅可告知其他用户您已检查该群集。该状态不描述首选记录的状态，也不指定记录数据的任何进一步操作。

可以将群集标记为已审阅并为记录设置其他状态指示器。例如，您可以拒绝标记为已审阅的首选记录。

可以在任何任务中将群集标记为已审阅。

清除审阅者状态

清除群集的任一状态指示器。您如果确定当前状态不正确，则可以清除该状态。

可以在任何任务中清除状态指示器。

审批记录编辑

指示您已在审阅重复项任务中分析群集并同意当前首选记录。如果确定记录包含群集所表示的业务实体的最准确数据，则批准该记录。

也可以在审阅异常项任务中批准记录编辑。

拒绝记录编辑

指示您已在审阅重复项任务中分析群集并且不同意当前首选记录。如果确定记录不包含群集所表示的业务实体的最准确数据，则拒绝该记录。

也可以在审阅异常项任务中拒绝记录编辑。

注意: 状态指示器必须表示首选记录的当前状态。例如，您可能因为在首选记录中发现错误而拒绝该记录。如果修复了错误，请更新群集状态。

设置群集的状态

完成对群集的处理后，设置群集状态。请设置一个表示群集中记录的当前状况的状态。该状态将确定另一用户或另一数据进程在工作流的下一阶段对数据所执行的操作。

要设置群集状态，请使用**群集操作**菜单选项。

1. 打开更正重复项任务或审阅重复项任务。
任务中的群集列表将以列的形式显示在“数据编辑”选项卡上。默认情况下，将打开第一个群集。
2. 也可以选择其他群集。
3. 单击**编辑**。
4. 或者，更新首选记录或更新群集内容。
5. 要更新群集状态，请执行以下一个或多个操作：
 - 将群集标记为已审阅。
将群集标记为已审阅可告知其他用户您已检查该群集。可以在更正重复项任务和审阅重复项任务中将群集标记为已审阅。
 - 清除审阅者状态。
清除状态可删除群集的任一状态指示器。可以在更正重复项任务和审阅重复项任务中清除审阅者状态。
 - 批准群集中的首选记录。
批准记录以指示您同意群集中的当前首选记录。可以在审阅重复项任务中批准首选记录。
 - 拒绝群集中的首选记录。
拒绝记录以指示您不同意群集中的当前首选记录。可以在审阅重复项任务中拒绝首选记录。
6. 将群集保存到任务。

在任务中审阅完所有群集后，更新任务状态。任务状态指示群集记录已准备好进入工作流的下一阶段。

在多个群集中查找记录

可以在其他群集中搜索包含一个或多个指定数据值的记录。使用“已发现的群集”选项搜索记录。

在其他群集中搜索可能与当前群集中的记录匹配的记录。如果记录表示组织数据集中的相同业务实体，则将记录移至相同群集。

1. 打开某一更正重复项任务。

任务中的群集列表将以列的形式显示在“数据编辑”选项卡上。默认情况下，将打开第一个群集。

2. 也可以选择其他群集。

3. 展开**已发现的群集**选项。

默认情况下，这些选项不显示任何群集数据。

4. 单击**查找群集**。

5. 选择包含要搜索的数据值的列。

6. 输入要搜索的数据值。

按数据值在记录列中的显示输入该数据值，或者在文本模式下输入通配符。可以使用星号 (*) 或百分号 (%) 通配符。

7. 也可以为其他列输入额外的数据值。

8. 单击**查找**。

搜索操作将返回任务数据中包含您搜索的值的记录。如果搜索多个值，搜索操作将返回包含每个值的记录。

9. 在搜索结果中选择一个或多个记录，然后单击**打开**。

“数据编辑”选项卡将显示包含您选择的记录的群集。

10. 单击群集，在“数据编辑”选项卡上将其打开。

11. 比较群集中的记录。

要在“数据编辑”选项卡上更新群集，请单击**编辑**。

相关主题：

- [“将记录移到空群集” 页面上 35](#)
- [“编辑群集” 页面上 32](#)

在群集之间移动记录

如果两个或更多群集包含表示相同实体的记录，请将记录从一个群集移至另一群集。将可能提供数据的记录移至其他群集中的首选记录。

可以在“数据编辑”选项卡上打开群集并在群集之间移动记录。

1. 打开某一更正重复项任务。

任务中的群集列表将以列的形式显示在“数据编辑”选项卡上。默认情况下，将打开第一个群集。

2. 也可以选择其他群集。

例如，如果当前群集的状态为“已审阅”，则选择其他群集。

3. 单击**编辑**。

4. 在另一群集中查找与当前群集中的记录匹配的记录。

使用**已发现的群集**选项查找记录。搜索操作可能返回多个群集中的记录。

5. 打开包含您查找的记录的群集。
6. 比较群集中的记录。
7. 如果群集中的某个记录与另一群集中的多个记录更加匹配，请将该记录移至另一群集。

使用**移动记录**图标可移动记录。

8. 将群集保存到任务。

将记录移到空群集

标识不属于任务中的任何群集的有效记录时，请将记录移至空群集。将记录移至您创建的群集。

有效记录包含属于组织数据集的业务实体的相关信息。该记录可能唯一，也可能与其他群集中的记录匹配。可以将多个记录添加到您创建的群集。

1. 打开某一更正重复项任务。
任务中的群集列表将以列的形式显示在“数据编辑”选项卡上。默认情况下，将打开第一个群集。
2. 也可以选择其他群集。
3. 检查群集中的记录。
验证群集是否包含不与群集中的其他记录匹配的有效记录。
4. 单击**编辑**。
5. 展开**发现群集**选项。
6. 单击**创建群集**。
空群集显示在当前群集之下。
7. 选择要添加到该群集中的记录。
8. 将记录移至该群集。
该记录将成为您创建的群集的首选记录。
注意: 在当前群集中的**说明**图标旁边找到**移动记录**图标。
9. 移动与您创建的群集中的首选记录匹配的任何其他记录。

相关主题：

- [“在多个群集中查找记录” 页面上 34](#)
- [“编辑群集” 页面上 32](#)

将说明添加到群集

添加说明以描述对群集所执行的操作。可以在“数据编辑”选项卡和“数据审计”选项卡上查看说明。

1. 在“异常项”工作区中，选择**数据编辑**选项卡。
2. 打开群集。
3. 打开“群集操作”菜单，然后选择**添加说明**。

Analyst 工具将打开“说明”框。

4. 在“说明”框中输入注释。

重复任务的筛选器选项

可以使用筛选器检索任务中满足指定数据条件的群集。可以在“数据编辑”选项卡和“数据审计”选项卡上定义筛选器。

使用筛选器选项可显示或隐藏包含指定数据值或共享指定状态的群集。“数据审计”选项卡还包含其他选项，可根据群集的保存日期和审阅状态来显示或隐藏群集。

您可以对每个选项卡应用多个筛选器。应用多个筛选器时，Analyst 工具将返回满足所有筛选条件的记录。

重复记录任务中的数据筛选器

使用“数据编辑”选项卡上的筛选器可标识一组具有公共数据功能的群集。应用筛选器时，Analyst 工具将列出满足筛选条件的群集。

“数据编辑”选项卡包含以下筛选器选项：

值

返回包括一个或多个具有选定列中值的记录的群集。选择列名称并输入值。

状态

返回显示选定状态的群集。

可以选择以下状态指示器：

- 任意。任务中的任意群集（不考虑状态）。
- 已接受。包含适合与组织数据一起存储的首选记录的群集。
- 已拒绝。不包含适合与组织数据一起存储的首选记录的群集。
- 已审阅。已审阅的群集。状态不指示首选记录的状态。
- 空。不含当前状态的群集。

重复记录任务中的审计跟踪筛选器

使用“数据审计”选项卡上的筛选器可根据群集元数据显示或隐藏群集。可以按日期、用户和状态筛选群集。

“数据审计”选项卡包含以下筛选器选项：

更新日期

返回用户在指定时间段内更新的群集。

更新者

返回用户更新的群集。标识 Analyst 工具用户。要按用户筛选，请输入用户名。

状态

返回包含指定状态的群集。

可以选择以下状态选项：

- 任意。任务中的任意群集（不考虑状态）。
- 移至群集。包含用户移至另一群集的记录的任何群集。
- 从群集移出。包含用户从其他群集移出的记录的任何群集。
- 空。不含当前状态的群集。

查看

返回包含指定审阅状态的群集。

可以选择以下审阅选项：

- 所有群集。任务中的任意群集（不考虑审阅状态）。
- 已审阅。用户标记为已审阅的群集。
- 已批准。用户在审阅任务中标记为已接受的群集。
- 已拒绝。用户在审阅任务中标记为已拒绝的群集。
- 空。不含当前审阅状态的群集。
- 已清除。具有用户清除的早期审阅状态的群集。

在重复记录任务中筛选群集

使用筛选器可检索一组共享指定数据特征的记录。应用“数据编辑”选项卡上的筛选器时，Analyst 工具将显示包含一个或多个满足筛选条件的记录的群集。应用“数据审计”选项卡上的筛选器时，Analyst 工具将显示满足筛选条件的所有记录。

筛选器在“数据编辑”选项卡上返回的群集可能包含不满足筛选条件的记录。筛选器返回至少包含一个满足条件的记录的群集。

1. 打开更正重复项任务或审阅重复项任务。
2. 选择“数据编辑”选项卡或“数据审计”选项卡。
3. 单击**筛选器**。

“筛选器”面板将打开。

注意：“数据编辑”选项卡和“数据审计”选项卡显示不同的筛选器选项组。

4. 选择要应用于任务数据的筛选器条件。
5. 单击**应用筛选器**。

Analyst 工具检索满足筛选条件的记录。

Analyst 工具将您定义的筛选条件显示在任务群集列表之上。

审阅重复项任务

在审阅异常任务中审阅群集时，需要验证其他用户在早期任务中完成的工作。可以在审阅重复项任务和更正重复项任务中执行相同的操作。

检查群集中的记录，并确认首选记录代表该群集中的记录的最准确版本。如果您不同意先前用户的决定，可更新首选记录和群集状态。

使用**群集操作**菜单选项更新记录状态。完成对任务的处理后，使用**任务操作**菜单选项更新任务状态。

注意: 验证该群集中的其他记录不包括企业可能想要保留的任何记录。可以将记录移至其他群集，还可以创建用来存储记录的群集。

执行审阅重复项任务

验证每个群集中的首选记录是否代表该群集中的记录的最准确版本。

1. 打开审阅重复项任务。

任务中的群集列表将以列的形式显示在“数据编辑”选项卡上。默认情况下，将打开第一个群集。

2. 检查群集中的记录。

验证首选记录是否包含群集中的数据的最准确版本。

验证群集状态是否为“已审阅”。

3. 要更新群集内容或群集状态，请单击**编辑**。

要更新首选记录，请执行以下一个或多个操作：

- 将来自其他记录的数据值升级到首选记录。
要升级某个值，请单击源记录中的数据字段。您选择的值将替换首选记录中相同列的值。
- 更新首选记录中的数据值。
如果群集中没有记录包含数据的准确版本，请更新首选记录。
- 使用群集中的其他记录替换当前首选记录。
要替换首选记录，请单击包含该记录的行中的升级工具。

也可以将记录移至其他群集或从其他群集导入记录。要在两个群集之间移动记录，请使用**已发现的群集**选项。

要更新群集状态，请执行以下操作之一：

- 将群集标记为已审阅。
将群集标记为已审阅可告知其他用户您已检查该群集。
- 批准群集中的首选记录。
批准记录以指示您同意群集中的当前首选记录。
- 拒绝群集中的首选记录。
拒绝记录以指示您不同意群集中的当前首选记录。
- 清除审阅者状态。
清除状态可删除群集的任一状态指示器。

4. 完成对群集的处理后，将群集保存到任务。

在任务中审阅完所有群集后，更新任务状态。任务状态指示群集记录已准备好进入工作流的下一阶段。

重复记录任务的规则和准则

审阅和更新数据集中的异常记录的过程为协作式过程。您可能会处理其他用户在早期任务中分析的群集。在任务完成后，您处理的群集也可能会传递给其他用户。每个用户均可审阅和更新其他用户的工作。

执行重复记录任务时，请考虑以下规则和准则：

- 群集是数据库表中共享相似或相同数据值的一组记录。开发人员定义记录在群集中的排序条件。如果您认为记录不属于当前群集，请使用“已发现的群集”选项查找正确群集。如果您认为其他群集包含属于当前群集的记录，请使用“已发现的群集”选项查找正确群集。
- 处理群集时，请使用首选记录定义群集所代表的业务实体的最准确版本。首选记录不一定是记录的最终版本。其他用户或其他数据进程可能会在您完成任务后处理群集。
更新首选记录时，需要更新异常项数据库中表示群集记录首选形式的记录。不需要更新群集中的源数据。
- 设置群集的状态时，可以指示您已审阅该群集。审阅状态并不描述群集中首选记录的准确性或数据质量。
最佳做法是将检查过的每个群集都标记为已审阅。该状态将向下游任务中的用户确认其他用户已检查该群集。更新记录中的数据后，请将该记录标记为已审阅，而不考虑记录上是否存在或缺少其他状态指示器。
- 审计跟踪存储用户对首选记录所做的每次更改，但不存储群集数据的更改。
- 您处理的数据可以传递给数据更正任务或数据审阅任务。例如，在工作流中配置人工任务的开发人员可能会依次指定多个更正异常项任务。该开发人员可能会在审阅重复项任务之后安排更正重复项任务。

第 4 章

任务管理

本章包括以下主题：

- [任务管理概览, 40](#)
- [任务执行者操作, 40](#)
- [业务管理员操作, 42](#)
- [任务数据导出, 45](#)

任务管理概览

在工作流中配置人工任务的开发人员指定可以处理任务数据的用户。登录到 Analyst 工具后，“启动”工作区将列出工作流为您分配的任务实例。您可以任务执行者或业务管理员的身份来执行任务。

任务执行者可以更新任务中的记录和状态指示器。业务管理员除了可以执行与任务执行者相同的操作外，还可以将任务从一个用户重新分配给另一个用户并将任务移至工作流的下一阶段。

工作流可以根据不同的任务为您分配不同的角色。对于单一工作流中的不同任务，您可能会承担任务执行者和业务管理员的角色。

任务执行者操作

工作流开发人员选择可以处理人工任务实例的任务执行者。要查看可以处理的任务实例列表，请打开“我的任务”视图。

可以在**我的任务**视图中执行以下操作：

查看任务内容

查看任务中的记录或群集。任务内容显示在**异常项**工作区中。

打开任务

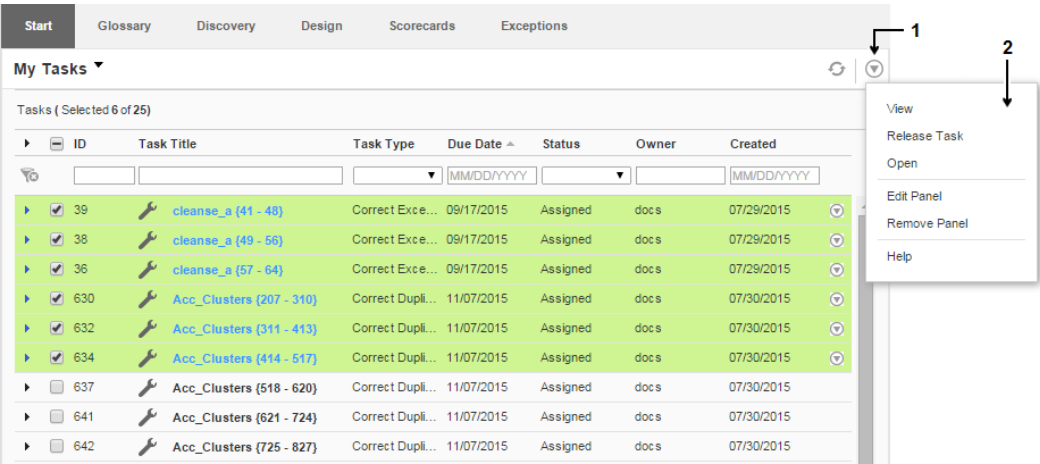
打开任务以审阅和编辑记录。当您打开任务时，您将成为任务所有者。其他任务执行者无法打开您拥有的任务。

释放任务

释放您拥有的任务。您释放的任务将没有所有者，直到业务管理员分配一个所有者或其他用户打开该任务为止。

您可以随时释放任务。无需完成对任务的处理便可释放任务。

下图在“启动”工作区中显示“我的任务”视图：



使用以下选项可处理任务：

1. “任务” 菜单图标。
2. “任务” 菜单选项。

可以一次查看、打开和释放多个任务。

查看任务

查看某项任务时，以只读模式打开该任务。任务执行者可以查看没有所有者的任务的内容。业务管理员可以在“任务管理”视图上查看任何任务的内容。

可以一次在只读模式下打开多个任务。每个任务均在“异常项”工作区的单独选项卡中打开。可以使用任务复选框来选择任务。

1. 在“启动”工作区上，打开“我的任务”视图。
2. 选择一个或多个任务。
注意：请勿单击任务名称。单击某个任务名称时，将打开该任务并声明任务的所有权。
3. 从任务菜单中，选择查看。

您选择的任务将在“异常项”工作区中打开。

打开任务

打开任务可处理任务数据。打开某项任务时，将声明该任务的所有权。其他任务执行者无法编辑您拥有的任务。

可以一次打开多个任务。每个任务均在“异常项”工作区的单独选项卡中打开。可以使用任务复选框来选择任务。

1. 在“启动”工作区上，打开“我的任务”视图。
2. 要打开单个任务，请单击任务名称。
或者，选择一个或多个任务。
3. 从任务菜单中，选择打开。

您选择的任务将在“异常项”工作区中打开。

释放任务

释放某项任务后，您将不再拥有该任务。其他用户可以声明任务的所有权，业务管理员可以将该任务分配给其他用户。任务会保存您对任务数据所执行的所有工作。

可以一次释放多个任务。可以使用任务复选框来选择任务。业务管理员可以在“任务管理”视图上释放任何任务。

1. 在“启动”工作区上，打开“我的任务”视图。
2. 选择一个或多个任务。
3. 从任务菜单中，选择**释放任务**。

Analyst 工具将更新任务列表以指示任务没有所有者。

业务管理员操作

workflows 开发人员可以选择您作为工作流生成的任务实例的业务管理员。要查看您管理的任务实例的列表，请打开“任务管理”视图。

业务管理员可以将任务分配给用户、查看用户执行的工作并将任务移至工作流的下一阶段。也可以声明任务的所有权并以与任务执行者相同的方式来更新任务数据。“任务管理”视图显示您管理和拥有的任务列表。

可以在“任务管理”视图中执行以下操作：

查看任务并打开您拥有的任务

查看任何任务中的数据并打开您拥有的所有任务。获得没有所有者的任务的所有权。

一次完成多个任务

完成源自与选定任务相同的人工任务的所有任务。完成任务后，任务数据将移至工作流的下一阶段。

重新分配任务

将任务从一个用户重新分配或委派给另一个用户。例如，如果当前用户无法按计划完成任务，则可以重新分配任务。

发布任务

发布您管理的任务。在您分配所有者或有其他用户打开任务之前，您发布的任务没有所有者。

审阅用户在任务中执行的工作

使用任务摘要数据可度量用户在任务中的进度。

查看具有公共父任务的任务列表

查看源自与选定任务相同的人工任务的任务列表。

可以一次查看和重新分配多个任务。

向用户分配任务

可在任务没有所有者时分配任务，或在当前用户无法完成某项任务时将该任务分配给其他用户或组。

可以一次分配多个任务。可以使用任务复选框来选择任务。

1. 打开**任务管理**视图。
2. 从任务列表中选择一项或多项任务。
3. 从任务菜单中，选择**重新分配任务**。

4. 选择用户或组名称。
使用筛选器可查找用户或组。
5. 或者，输入注释。
6. 单击**分配**。
“任务管理”视图将更新记录以显示您选择的任务所有者。

查看人工任务中的任务实例列表

可以查看工作流为人工任务生成的任务实例列表。

注意: 审阅任务实例列表时，不要单击**确定**。如果单击**确定**，则会将任务指定的数据移至工作流的下一阶段。

1. 打开“任务管理”视图。
2. 从任务列表中选择一项任务。
3. 从任务菜单中，选择**完成链接的任务**。
4. 审阅任务列表。

任务列表将显示每项任务的以下信息：

- 任务 ID。
人工任务内的任务实例的唯一标识符。
- 任务标题。
显示在“任务管理”视图中的任务的名称。
- 任务类型。
人工任务生成的任务类型。任务可能是更正异常项任务、审阅异常项任务、更正重复项任务或审阅重复项任务。
- 任务所有者
拥有任务的用户的名称。
- 到期日期
用户必须完成任务的到期日期。
- 状态
任务的所有权状态。用户声明任务之前，任务状态为“已创建”。用户声明任务或您将任务分配给用户时，任务状态为“已分配”。
- 已创建。
创建了任务的工作流的运行日期。

5. 单击**取消**。

任务摘要数据

可以读取在“异常项”工作区中查看或打开的任何任务的摘要数据。使用摘要数据可度量用户在任务中的进度。

重复记录任务显示任务中的群集数和已审阅的群集数。异常记录任务显示任务中的记录数和包括状态指示器的记录数。

下图显示了异常记录任务的摘要数据：

StartGlossaryDiscoveryDesignScorecardsExceptions X

*cleanse_a_rev {1 - 8} X

cleanse_a_rev {1 - 8}

SaveCancelUndoRedo

To review exceptions, verify the work that a user performed in an earlier task. Review the data in each record, and review the record status. You can accept or undo the work of the user. When you finish the review of the records, update the task status.

8 of 8620Records are reviewedAcceptedReprocessedRejected

1234

Find & ReplaceFilter

Data EditingData Audit

| | | | STRING_COL | INT_COL | BIGINT_COL | NUMBER_COL | NSTRING_COL | DOUBLE_COL | DATETIME_COL |
|-------------------------------------|---|--|------------|---------|------------------------|------------|-------------|------------|---------------------|
| <input checked="" type="checkbox"/> | 1 | | aaaaaaaa | 11,111 | 8,890,000,000,000,0... | 50 | bbbbbb | 7.527 | 01/01/2015 12:00:00 |
| <input checked="" type="checkbox"/> | 2 | | aaaaaaaa | 11,111 | 8,890,000,000,000,0... | 50 | bbbbbb | 7.527 | 01/02/2015 12:00:00 |

任务包含以下摘要数据值：

1. 用户审阅的记录数以及任务中的记录总数。
2. 用户已接受、适合与业务数据一起存储的记录数。
3. 用户为进一步分析所选择的记录数。
4. 用户已拒绝、不适合与业务数据一起存储的记录数。

审阅用户在任务中的进度

要度量用户在任务中的进度，请审阅摘要数据。

1. 在“启动”工作区上，打开“任务管理”视图。
2. 选择一个任务。
该任务将在“异常项”工作区中打开。
3. 审阅任务的摘要数据。

多项任务完成

可以标识源自单个人工任务的所有任务实例并一次性完成这些任务实例。

完成一项人工任务的所有任务实例后，该任务中的记录会移至工作流的下一阶段。Analyst 工具会将任务从处理任务的用户中删除。完成任务后，不要更新任何记录数据或状态数据。

在以下情况下，您可能会决定完成多个任务：

- 工作流已失败，并且您想再次运行该工作流。
- 用户无法按时完成多个任务实例。

完成多项任务

要完成人工任务的所有任务实例，请从人工任务中选择任一任务实例。

1. 打开“任务管理”视图。
2. 从任务列表中选择一项任务。
3. 从任务菜单中，选择**完成链接的任务**。
完成链接的任务对话框将显示源自与选定任务相同的人工任务的任务实例列表。
4. 或者，输入注释。

5. 单击**确定**。

任务完成。任务数据将移至工作流的下一阶段。

如果您在完成的任务后打开收件箱，则收件箱可能不会显示对任务列表所做的任何更改。要在收件箱中查看当前任务列表，请刷新收件箱。

任务数据导出

可将数据从任务实例导出到带分隔符的文件。导出数据以便与其他用户共享数据的当前状态。

导出任务时，需要导出记录数据、状态指示器数据和工作流标识符值。导出操作会为状态指示器数据和工作流标识符值添加列，但会排除审计跟踪数据。

异常任务元数据

从更正异常项任务或审阅异常项任务中导出数据时，将导出所有任务数据和元数据。导出操作会将元数据列添加到带分隔符文件的开头，并为更正异常项任务和审阅异常项任务创建相同的列。

下表介绍了您随任务数据一起导出的元数据列：

| 列名称 | 说明 |
|----------------|---|
| ROW_IDENTIFIER | 数据库表中的记录行编号。 |
| REVIEW_STATUS | 记录的当前审阅状态。 记录可以具有以下审阅状态值之一： <ul style="list-style-type: none">- 空。用户未更新审阅状态。- 已审阅。用户已审阅记录。- 已拒绝。用户已拒绝记录。- 已接受。用户已批准记录。 |
| WORKFLOW_ID | 生成任务实例的工作流的标识符。 |
| USER_COMMENT | 用户添加到 Analyst 工具中的记录的最新说明。 |
| UPDATED_STATUS | 记录数据的当前状态。 记录可以具有以下状态值之一： <ul style="list-style-type: none">- 已更新。用户已更新记录。- 已接受。用户已接受记录，记录有效。- 已拒绝。用户已拒绝记录，记录无效。- 重新处理。用户指示记录需要由其他用户或其他数据进程另外进行分析。- 空。用户未更新记录状态。 |
| RECORD_STATUS | 工作流设置的记录状态。工作流在将记录写入到异常数据表时会设置状态值。默认状态为“无效”。 |

重复任务元数据

从更正重复项任务或审阅重复项任务中导出数据时，将导出所有任务数据和元数据。 导出操作会将元数据列添加到带分隔符文件的开头， 并为更正重复项任务和审阅重复项任务创建相同的列。

下表介绍了您随任务数据一起导出的元数据列：

| 列名称 | 说明 |
|-----------------------|--|
| ROW_IDENTIFIER | 数据库表中的记录行编号。 |
| SEQUENTIAL_CLUSTER_ID | 数据库表中群集的标识符值。 工作流使用该值在数据库中为群集行排序。 |
| CLUSTER_ID | 标识记录所属的群集的值。 |
| MATCH_SCORE | 指示群集中两个记录之间的相似度的值。 得分为介于 0 与 1 之间的十进制值。 |
| IS_MASTER | 标识表中首选记录的值。 对于首选记录，值为 Y；对于其他记录，值为 N。 |
| UPDATED_STATUS | 群集中记录的更新状态。 记录可以具有以下状态值之一： <ul style="list-style-type: none">- 已更新。用户已更新记录。 用户可以更新群集中的首选记录。- 空。用户未更新记录。- 已提取。用户已将记录移出原始群集。 |
| USER_COMMENT | 用户添加到 Analyst 工具中的群集的最新说明。 |
| REVIEW_STATUS | 群集的当前审阅状态。 群集可以具有以下审阅状态值之一： <ul style="list-style-type: none">- 空。用户未更新审阅状态。- 已审阅。用户已审阅包含该记录的群集。- 已拒绝。用户已拒绝包含该记录的群集。- 已接受。用户已批准包含该记录的群集。 |
| WORKFLOW_ID | 生成任务实例的工作流的标识符。 |

导出任务数据

将任务数据导出到带分隔符的文件。

1. 打开任务。
2. 从任务操作菜单中，选择**导出数据**。
 此时将显示**导出**对话框。
3. 或者，更改导出文件名。
 默认情况下，该文件名为任务名称。
4. 选择或清除将列名称导出为导出数据第一行的选项。
5. 单击**导出**。

Analyst 工具便将数据文件导出到目录结构。

索引

C

查找和替换数据值
 异常记录 [24, 25](#)
重复记录
 编辑群集 [32](#)
 创建群集 [35](#)
 更新群集状态 [33](#)
 更正步骤 [29](#)
 更正重复项任务 [28, 29](#)
 审阅步骤 [29](#)
 审阅重复项任务 [30](#)
 已发现的群集选项 [34](#)
重复项任务
 导出文件结构 [46](#)

D

导出任务
 说明 [45](#)
导出文件
 重复项任务 [46](#)
 异常项任务 [45](#)

G

工作流
 说明 [9](#)

I

Informatica Analyst 界面
 登录 [14](#)

J

进程流
 重复记录示例 [14](#)
 离散记录示例 [13](#)
角色
 任务执行者 [10, 40](#)
 业务管理员 [10, 40](#)

Q

启动工作区
 列 [11](#)
清除群集状态
 说明 [33](#)
群集
 编辑 [32](#)

群集 (续)

 查找记录 [34](#)
 创建 [35](#)
 筛选群集数据 [36](#)
 添加说明 [35](#)
 状态更新 [33](#)

R

任务
 查看 [41](#)
 打开 [41](#)
 导出任务数据 [45, 46](#)
 更正重复项 [28, 29](#)
 更正异常项 [17](#)
 人工任务 [9](#)
 任务和工作流 [9](#)
 任务类型 [8](#)
 任务实例 [9](#)
 审阅重复项 [30](#)
 审阅异常项 [17](#)
 释放任务 [42](#)
 映射任务 [9](#)
 访问权限 [9](#)
 人工任务中的步骤 [9](#)
任务管理
 向用户分配任务 [42](#)
任务管理选项 [42](#)
任务实例
 定义 [9](#)
任务执行者
 我的任务视图 [40](#)

S

筛选器
 群集数据 [36](#)
筛选器选项
 异常数据 [22](#)
筛选群集
 步骤 [37](#)
审阅任务
 审阅重复项 [38](#)
 审阅群集的步骤 [37](#)
示例
 不良记录的进程流 [13](#)
 重复记录的进程流 [14](#)
 异常任务中的验证错误 [19](#)
首选记录
 更改 [32](#)
数据编辑面板
 筛选记录 [24](#)
 筛选群集 [37](#)
 异常数据筛选器 [22](#)

数据编辑选项卡
 重复记录任务 [30](#)
 筛选群集数据 [36](#)
说明
 添加到群集 [35](#)
审计跟踪
 重复记录筛选器 [36](#)
 异常记录筛选器 [23](#)
数据审计选项卡
 重复记录筛选器 [36](#)
 异常记录筛选器 [23](#)

W

我的任务视图 [10](#), [40](#)

Y

业务管理员 [10](#), [40](#)
异常管理
 概览 [8](#)
异常记录
 编辑异常记录 [19](#)

异常记录 (续)
 查找和替换数据值 [24](#), [25](#)
 更新记录状态 [21](#)
 更正步骤 [16](#)
 更正异常项任务 [17](#)
 筛选 [24](#)
 审阅步骤 [16](#)
 审阅异常项任务 [17](#)
异常任务
 验证错误 [19](#)
异常任务中的验证错误 [19](#)
异常项工作区
 更正异常项任务 [17](#)
 Informatica Analyst [12](#)
 数据编辑选项卡 [17](#)
异常项任务
 导出文件结构 [45](#)

Z

状态指示器
 重复记录任务 [32](#)
 异常项任务 [20](#)