Informatica®
10.5

Exception Management Guide

Informatica Exception Management Guide
10.5
April 2021

© Copyright Informatica LLC 1998, 2021

This software and documentation are provided only under a separate license agreement containing restrictions on use and disclosure. No part of this document may be reproduced or transmitted in any form, by any means (electronic, photocopying, recording or otherwise) without prior consent of Informatica LLC.

U.S. GOVERNMENT RIGHTS Programs, software, databases, and related documentation and technical data delivered to U.S. Government customers are "commercial computer software" or "commercial technical data" pursuant to the applicable Federal Acquisition Regulation and agency-specific supplemental regulations. As such, the use, duplication, disclosure, modification, and adaptation is subject to the restrictions and license terms set forth in the applicable Government contract, and, to the extent applicable by the terms of the Government contract, the additional rights set forth in FAR 52.227-19, Commercial Computer Software License.

Informatica and the Informatica logo are trademarks or registered trademarks of Informatica LLC in the United States and many jurisdictions throughout the world. A current list of Informatica trademarks is available on the web at https://www.informatica.com/trademarks.html. Other company and product names may be trade names or trademarks of their respective owners.

Portions of this software and/or documentation are subject to copyright held by third parties. Required third party notices are included with the product.

The information in this documentation is subject to change without notice. If you find any problems in this documentation, report them to us at infa_documentation@informatica.com.

Informatica products are warranted according to the terms and conditions of the agreements under which they are provided. INFORMATICA PROVIDES THE INFORMATION IN THIS DOCUMENT "AS IS" WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING WITHOUT ANY WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND ANY WARRANTY OR CONDITION OF NON-INFRINGEMENT.

Publication Date: 2021-04-02

Table of Contents

Preface 6

Informatica Resources. 6

Informatica Network. 6

Informatica Knowledge Base. 6

Informatica Documentation. 6

Informatica Product Availability Matrices. 7

Informatica Velocity. 7

Informatica Marketplace. 7

Informatica Global Customer Support. 7

Chapter 1: Introduction to Exception Management. 8

Exception Management Overview. 8

Task Types. 8

Tasks and Workflows. 9

Steps in a Human Task. 10

Roles in Exception Management. 10

Exception Management Workspaces. 11

Start Workspace. 11

Exceptions Workspace. 13

Audit Trail Operations. 13

Exception Management Process Flow. 13

Bad Records Example. 14

Duplicate Records Example. 15

Logging In to the Analyst Tool. 16

Chapter 2: Exception Tasks. 17

Exception Tasks Overview. 17

Types of Exception Task. 17

Steps to Correct Exceptions. 18

Steps to Review Exceptions. 18

Exception Tasks in the Exceptions Workspace. 19

Fixing Errors in an Exception Task. 20

Validation Errors in Exception Tasks. 21

Exception Record Status Indicators. 21

Setting the Status of an Exception Record. 22

Filter Options for Exception Tasks. 23

Record Data Filters. 23

Metadata Filters on the Data Editing Tab. 24

Metadata Filters on the Data Audit Tab. 24

Filtering Records in an Exception Task. 25

Find and Replace Values in Exception Records. . . . 26
Replacing a Single Value in Exception Records. . . . 26
Replacing Multiple Values in Exception Records. . . . 27
Adding a Note to a Record. . . . 28
Exception Tasks in Review. . . . 28
Working in a Review Exceptions Task. . . . 28
Rules and Guidelines for Exception Tasks. . . . 29

Chapter 3: Duplicate Record Tasks. . . . 30

Duplicate Record Tasks Overview. . . . 30
Preferred Records. . . . 31
Types of Duplicate Record Task. . . . 31
Steps to Correct Duplicates. . . . 31
Steps to Review Duplicates. . . . 32
Duplicate Record Tasks in the Exceptions Workspace. . . . 32
Editing a Cluster . . . 34
Duplicate Record Status Indicators. . . . 35
Setting the Status of a Cluster. . . . 35
Finding Records in Multiple Clusters. . . . 36
Moving Records Between Clusters. . . . 37
Moving a Record to an Empty Cluster. . . . 38
Adding a Note to a Cluster. . . . 38
Filter Options for Duplicate Tasks. . . . 39
Data Filters in Duplicate Record Tasks. . . . 39
Audit Trail Filters in Duplicate Record Tasks. . . . 39
Filtering Clusters in a Duplicate Record Task. . . . 40
Review Duplicates Tasks. . . . 40
Working in a Review Duplicates Task. . . . 41
Rules and Guidelines for Duplicate Record Tasks. . . . 42

Chapter 4: Task Management. . . . 43

Task Management Overview. . . . 43
Task Performer Operations. . . . 43
Viewing Tasks. . . . 44
Opening a Task. . . . 44
Releasing a Task. . . . 45
Business Administrator Operations. . . . 45
Assigning a Task to a User. . . . 46
Viewing the List of Task Instances in a Human Task. . . . 46
Task Summary Data. . . . 47
Multiple Task Completion. . . . 47
Task Data Export. . . . 48
Exception Task Metadata. . . . 48

Duplicate Task Metadata. 49
Exporting Task Data. 50

Index. 51

Preface

Read the *Informatica Exception Management Guide* to understand how you can review and update exceptions in Informatica Analyst. An exception is a record that might contain bad data or duplicate data and that needs human evaluation to verify. To find exceptions in a data set, run a workflow that contains a Human task on the data. Use Informatica Analyst to determine whether the records are genuine exceptions, to resolve data errors, and to consolidate duplicate records into a single record.

Informatica Resources

Informatica provides you with a range of product resources through the Informatica Network and other online portals. Use the resources to get the most from your Informatica products and solutions and to learn from other Informatica users and subject matter experts.

Informatica Network

The Informatica Network is the gateway to many resources, including the Informatica Knowledge Base and Informatica Global Customer Support. To enter the Informatica Network, visit https://network.informatica.com.

As an Informatica Network member, you have the following options:

- Search the Knowledge Base for product resources.
- View product availability information.
- Create and review your support cases.
- Find your local Informatica User Group Network and collaborate with your peers.

Informatica Knowledge Base

Use the Informatica Knowledge Base to find product resources such as how-to articles, best practices, video tutorials, and answers to frequently asked questions.

To search the Knowledge Base, visit https://search.informatica.com. If you have questions, comments, or ideas about the Knowledge Base, contact the Informatica Knowledge Base team at KB_Feedback@informatica.com.

Informatica Documentation

Use the Informatica Documentation Portal to explore an extensive library of documentation for current and recent product releases. To explore the Documentation Portal, visit https://docs.informatica.com.

If you have questions, comments, or ideas about the product documentation, contact the Informatica Documentation team at infa_documentation@informatica.com.

Informatica Product Availability Matrices

Product Availability Matrices (PAMs) indicate the versions of the operating systems, databases, and types of data sources and targets that a product release supports. You can browse the Informatica PAMs at https://network.informatica.com/community/informatica-network/product-availability-matrices.

Informatica Velocity

Informatica Velocity is a collection of tips and best practices developed by Informatica Professional Services and based on real-world experiences from hundreds of data management projects. Informatica Velocity represents the collective knowledge of Informatica consultants who work with organizations around the world to plan, develop, deploy, and maintain successful data management solutions.

You can find Informatica Velocity resources at http://velocity.informatica.com. If you have questions, comments, or ideas about Informatica Velocity, contact Informatica Professional Services at ips@informatica.com.

Informatica Marketplace

The Informatica Marketplace is a forum where you can find solutions that extend and enhance your Informatica implementations. Leverage any of the hundreds of solutions from Informatica developers and partners on the Marketplace to improve your productivity and speed up time to implementation on your projects. You can find the Informatica Marketplace at https://marketplace.informatica.com.

Informatica Global Customer Support

You can contact a Global Support Center by telephone or through the Informatica Network.

To find your local Informatica Global Customer Support telephone number, visit the Informatica website at the following link:
https://www.informatica.com/services-and-training/customer-success-services/contact-us.html.

To find online support resources on the Informatica Network, visit https://network.informatica.com and select the eSupport option.

CHAPTER 1

Introduction to Exception Management

This chapter includes the following topics:

- Exception Management Overview, 8
- Task Types, 8
- Tasks and Workflows, 9
- Roles in Exception Management, 10
- Exception Management Workspaces, 11
- Audit Trail Operations, 13
- Exception Management Process Flow, 13
- Logging In to the Analyst Tool, 16

Exception Management Overview

An exception is a record that contains unresolved data quality issues. The record might contain errors, or the record might contain redundant or duplicate information. You can review and fix exception records in the Analyst tool.

To work on exception records, you open a task. A task identifies a set of exception records in a database table.

The Start workspace in the Analyst tool displays the tasks that you can work on. The workspace might contain tasks from the same database table or from different database tables. When you open a task from the workspace, you claim ownership of the task, and you must resolve the data quality issues that the task specifies.

Task Types

The type of task that a workflow assigns to you depends on the type of data quality issues that the source database tables contain. The records in the tables might contain errors, null values, or values that are

inaccurate in the current data project. The tables might contain records that are redundant because they contain different versions of the same information.

You can work on the following types of task instance in the Analyst tool:

Correct exceptions task

Contains records that might include errors or null values. Analyze and fix any error that you find. When a record is free of errors, update the record status to indicate that the record is valid.

Correct duplicates task

Contains records that might contain duplicate information. The task sorts the records into clusters. A cluster is a group of records that represent the same business entity in the source data set. Analyze each cluster, and define a preferred version of the record that the cluster represents. Update the cluster status to indicate that you reviewed the preferred record.

If a record is not a duplicate of any record in the cluster, move the record to another cluster. You can create a cluster that contains a single unique record.

Review exceptions task

Contains a set of exception records that an earlier user analyzed in a correct exceptions task. Review the work done by the earlier user, and verify that the record data and the record status are correct.

Review duplicates task

Contains a set of clusters that an earlier user worked on in a correct duplicates task. Review the work done by the earlier user, and verify that the preferred record data and the cluster status are correct.

Note: Records cannot pass from a task that corrects or reviews exception records to a task that corrects or reviews duplicate records. The database tables that contain exception records and duplicate records have different structures.

Tasks and Workflows

An Informatica process called a workflow generates the tasks that you work on in the Analyst tool.

A workflow describes a sequence of tasks that Informatica applications can perform. A task might contain instructions to run a mapping, to send an email, or to distribute exception records to Analyst tool users. A task that distributes exception records to Analyst tool users is called a Human task. The tasks that you work on in the Analyst tool are instances of a Human task. Each task instance identifies a set of records in an exception database table.

A workflow that contains a Human task must also contain a Mapping task. A Mapping task runs a data process called a mapping. The mapping verifies the quality of the records in the source data and writes the records to different tables based on a set of data quality indicators. The Human task reads the table that contains the exception records and assigns the records to Analyst tool users for analysis and repair.

When the exception record table contains many records, a task instance identifies a subset of the records. The workflow distributes the task instances to the Analyst tool users that the workflow developer identifies. Analyst tool users can work on the task instances concurrently.

Steps in a Human Task

A Human task in a workflow contains one or more steps. A step is a metadata object that defines the flow of data through the Human task. The steps determine the types of task instance that the Human task creates and identifies the users who can work on the task instances.

The workflow developer can add the following steps to a Human task:

Exception step

An exception step generates a correct exceptions task instance in the Analyst tool.

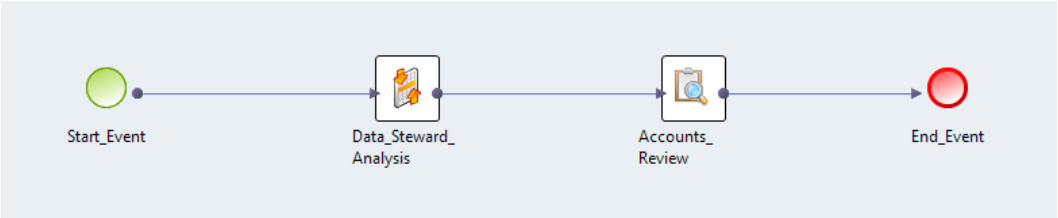
Cluster step

A cluster step generates a correct duplicates task instance in the Analyst tool.

Review step

A review step generates a review exceptions task instance or a review duplicates task instance in the Analyst tool.

The following image shows the steps in a Human task in the Developer tool:



When a Human task contains multiple steps, the exception records pass from one step to another. The developer who configures the workflow defines the sequence of the steps. Each step identifies a set of users. The users can log in to the Analyst tool and work on the task instances that the step defines.

Note: The developer who configures the workflow can set access permissions for the Analyst tool users. The permissions can determine the columns of data that the users can view and the actions that the users can perform in the task instances. The developer can set access permissions on each step. The access permissions apply equally to all users who work on the task instances that the step defines.

By default, all users have full access to the data and the options in a task instance.

Roles in Exception Management

A Human task assigns roles to the Analyst tool users who work on the task instances. The roles determine the range of actions that the user can perform. The Human task can assign roles to a user or to a group. If you belong to a group, you share the group role when you log in to the Analyst tool.

The Human task assigns the following roles:

Task performer

A user or a group that works on task instances in the Analyst tool.

Business administrator

A user or a group that manages the tasks instances that a Human task generates. A business administrator can also work on task instances.

If a Human task assigns task instances to a group, any member of the group can take ownership of a task. When a group member takes ownership of a task, the task disappears from the task lists of the other group members.

The workflow developer identifies one or more task performers for each step in the Human task. You might work on a task instance for a single step in the Human task and not work on a task instance for other steps.

The workflow developer identifies one or more business administrators for a Human task. Any business administrator for a Human task is also a business administrator for each step in the Human task.

Exception Management Workspaces

The Analyst tool uses workspaces to organize the different types of data operations that you can perform.

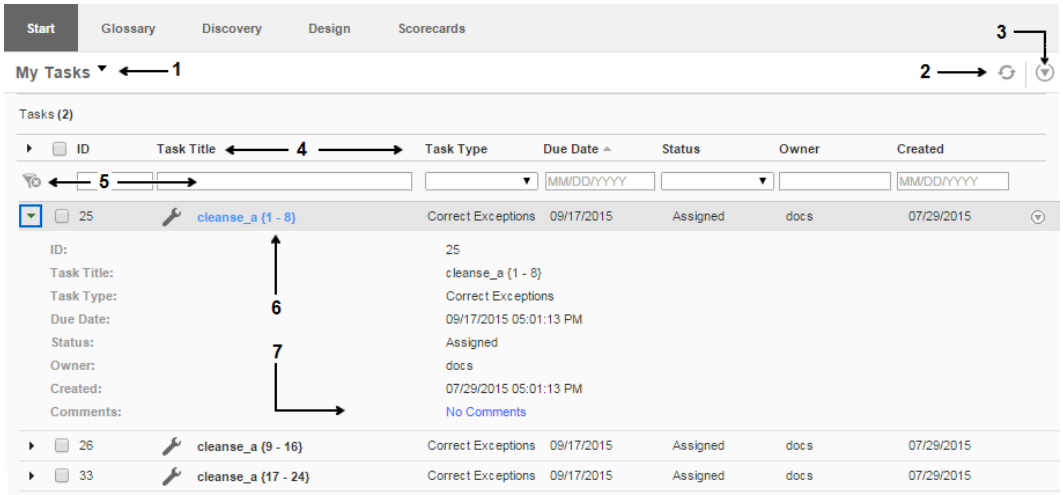
The Analyst tool lists the tasks that you own in the **My Tasks** view of the Start workspace. When you open the task, the Analyst tool displays the task records in the **Exceptions** workspace.

Start Workspace

The Start workspace displays the tasks that own and any task that you administer. Use the Start workspace options to open tasks, to perform actions on tasks, and to review the task metadata.

The user role that you use to log in to the Analyst tool determines the tasks that you can view in the Start workspace. Select the **My Tasks** view to display the tasks that you own. Select the **Task Administration** view to display any task that you administer. You can administer tasks if the workflow that creates the tasks identifies you as a business administrator.

The following image shows a list of tasks in the **My Tasks** view:



The My Tasks view includes the following options:

1. My Tasks and Task Administration options. Toggle the workspace view between the tasks that you own and any task that you administer. You can open a task from the My Tasks view and from the Task Administration view.
2. Refresh. Refresh the workspace data.
3. Actions menu. Open a list of actions that you can perform on the tasks that you select.

4. Column headings. List the names of the columns that describe the task instances.
5. Filter. Use the values that you enter to filter the list of records.
6. Task name. Show the name of the task. To open a task, select the task name.
7. Comments. Open an editor that you can use to enter a comment or to read a comment about the task.

Start Workspace Column Headings

The Start workspace displays the tasks in a table. Each task is a row in the table. The table columns contain metadata about the tasks.

Use the following columns to review the metadata for the tasks that you own or administer:

ID

The unique identifier for the task instance. The identifier includes a value that identifies the Human task that generated the task instance.

Task Title

The name of the task. The task name identifies the step in the Human task that the generated the task data.

The Task Title column includes a filter field that you can use to filter the list of task names. You can include wildcard characters with the filter string that you enter.

If you add a wildcard character to a filter string, the Analyst tool returns the tasks that include the filter string in any position in the task name.

Task Type

The type of data operation that you can perform on the task data.

The task can be one of the following types:

- Correct exceptions. Examine and fix errors in the records.
- Correct duplicates. Examine a cluster of duplicate records and create a preferred record from the values in the records.
- Review exceptions. Review the work of another user in a task that corrects exceptions.
- Review duplicates. Review the work of another user in a task that creates a preferred record from a cluster of records.

Note: The task types might include Voting tasks. Business Glossary users work on Voting tasks. Voting tasks are not a part of the exception management process.

Due Date

The deadline for the task. The Human task defines the due date for a correct exceptions task and a correct duplicates task. The Analyst tool calculates the due date for a correct exceptions task and a correct duplicates task.

Status

The status of the task

The task can have one of the following statuses:

- Created. The task has no owner.
- Assigned. The task has an owner.

Owner

The name of the current task owner.

Created

The date on which the workflow created the task.

Exceptions Workspace

The Exceptions workspace is a temporary workspace that appears when you view or open a task. The Exceptions workspace contains a **Data Editing** tab and a **Data Audit** tab.

The **Data Editing** tab displays the task data and the options that you can use to update the records or clusters in the task. The tab also displays the metadata columns that the Analyst tool uses to track the updates that you and other users make to each record or cluster.

The **Data Audit** tab displays an audit trail of the changes that you and other users made to the task data. You can view the fields that users changed, the identity of the user who changed each record, and the date of the change.

When you finish work on a task, you can close the Exceptions workspace.

Audit Trail Operations

The Analyst tool creates an audit trail of the changes that users make to the records in a task. To view the audit trail for a task, open the task and select the **Data Audit** tab in the Exceptions workspace.

The audit trail displays the record data and the status indicator values for each record that a user updates in a task. In addition, the audit trail identifies the user who made the update and the date and time of the update. The audit trail displays the rows in chronological order. If another user updated a record in an earlier task, the audit trail includes the update from the earlier user.

Consider the following rules and guidelines when you view the audit trail:

- Each row in the audit trail represents a single update to a record. If you make multiple updates to a record, the audit trail adds a row for each update.
- When you change a data value in a record, the audit trail highlights the value. Move the pointer over the value to see the previous value. The audit trail also identifies the user who changed the record, the date of the change, and the type of change.
- By default, the audit trail displays all of the data in each record. If the Analyst tool does not grant you permission to view a column of data in the task, the audit trail does not display the column. If a user updated a record value in an earlier task and you do not have viewing permissions for the value, the audit trail does not display the value.
- The Informatica domain configuration determines the location of the audit trail data. The Analyst tool can write all audit trail data to tables in a single database. Or, the Analyst tool can write audit trail data for each task to the database that contains the records in the task. For information about the audit trail data in your domain, contact the Informatica domain administrator.

Exception Management Process Flow

The tasks that you work on in the Analyst tool represent a stage in a data quality cycle. The cycle begins when the organization decides to verify the quality of the data in a data set. The cycle ends when the

organization is satisfied with the quality of the data. An organization might run a data quality cycle on a continuous basis.

The exception management stage often occurs toward the end of the data quality cycle. Earlier stages might use profiles and mappings to measure the quality of the organization data and to enhance the quality of the data. Exception management defines the manual operations that users can perform on the records that fall short of the data quality targets in the current cycle.

You might work in a team of data stewards that implements the data quality process. Or, you might be a business user who defines the data quality standards that the data must meet. In both cases, you can own exception management tasks in the Analyst tool.

Note: In many organizations, data stewards combine the data stewardship role with other roles. You might be responsible for a data quality objective as part of a larger role in the organization. As a data steward, you might perform a task on a data set and pass the results to a colleague who assumes the data stewardship duties.

Bad Records Example

You are a data steward at a retail organization. You are concerned that the product inventory records might contain errors. The errors might cause the organization to order more products or fewer products than it can sell.

You define the following process to investigate and correct the errors:

1. You ask a developer to configure one or more mappings to find and fix the errors in the data set.

   The mappings also calculate a numeric score for each record in the data set. The scores represent the data quality of the records. Some records have marginal scores that indicate that the mappings cannot verify all of the data quality issues that the records contain.

2. The developer configures an additional mapping that reads the numeric scores. The developer adds the mapping to a workflow that includes a Mapping task and a Human task.

   - The Mapping task runs the mapping. The Mapping task writes the records to different tables based on the scores that they contain.
   - The Human task distributes the records with marginal scores to tasks that you and other users can open in the Analyst tool.

3. You log in to the Analyst tool, and you open a task. The task organizes the exception records in one or more tables. Each table can contain 100 records.

   You perform one the following actions on each record:

   - You correct the errors in the record, or you decide that the current record is correct.

     You update the record status to indicate that the record is valid.

   - You determine that the record does not contain any valid data.

     You update the record status to indicate that the record is not valid.

   - You decide that you cannot verify the accuracy of the record.

     You update the record status to indicate that the record needs further analysis by another user or by another Informatica process.

   Note: Before you update a record, verify that the task is open in edit mode. To enter edit mode, click the **Edit** button in the open task.

4. When you finish work on all of the records in the task, you update the task status. The task status indicates that the records are ready for the next stage in the data quality process.

   The next stage for the data depends on the configuration of the Human task. For example, the Human task might include additional steps that assign the records to other users for review.

When the Human task completes, the next stage of the workflow begins.

Duplicate Records Example

You are a data steward at a bank. You are concerned that multiple records in the customer account tables might contain the same information. The duplicate records might represent data entry errors, or they might represent fraudulent customer activity.

You define the following process to find the duplicate records and to identify a single preferred version of each set of records:

1. You ask a developer to configure one or more mappings to identify the duplicate records.

   The mappings calculate a set of numeric scores that represent the levels of duplication between the data values in the records. High scores indicate duplicate records, and low scores indicate unique records. Some records have marginal scores that indicate that the duplicate status of the records is uncertain.

2. The developer configures an additional mapping that reads the numeric scores. The developer adds the mapping to a workflow that includes a Mapping task and a Human task.

   - The Mapping task runs the mapping. The Mapping task writes the records to different tables based on the scores that they contain.
   - The Human task distributes the records with marginal scores to tasks that you and other users can open in the Analyst tool.

3. You log in to the Analyst tool, and you open a task.

   The Analyst tool organizes the records in a series of clusters. Each cluster contains two or more records that contain similar information. By default, the first record in a cluster is the preferred record.

4. Open a cluster, and analyze the records that it contains.

   You perform the following actions in each cluster:

   - You examine the data values in each column of record data. You select the most accurate value in each column and promote the value to the preferred record.

     You can edit the values that you select, and you can search for records that contain common values in other clusters.

   - If a record does not belong in the current cluster, you move it to another cluster or you create a cluster for the record.
   - You update the cluster status to indicate that you reviewed the cluster. You complete the task when you verify the current preferred record in every cluster.

   Note: Before you update a record, verify that the task is open in edit mode. To enter edit mode, click the **Edit** button in the open task.

5. When you finish work on all of the clusters in the task, you update the task status. The task status indicates that the records are ready for the next stage in the data quality process.

   The next stage for the data depends on the configuration of the Human task. For example, the Human task might include additional steps that assign the clusters to other users for review.

When the Human task completes, the next stage of the workflow begins.

Logging In to the Analyst Tool

Use the Analyst tool URL to log in to the Analyst tool. When you log in to the Analyst tool, you specify an Informatica login name, a password, and the native domain or the LDAP security domain.

1. Start a Microsoft Internet Explorer or Google Chrome browser.
2. In the Address field, enter the URL for the Analyst tool:

   ```
   http[s]://<fully qualified host name>:<port number>/analyst/
   ```

3. If the domain uses LDAP or native authentication, enter a login name and a password on the login page.
4. Select **Native** or the name of a specific security domain.

   The Security Domain field appears when the Informatica domain uses LDAP or Kerberos authentication. If you do not know the security domain that your user account belongs to, contact the Informatica domain administrator.
5. Click **Log In**.

   The Analyst tool opens on the **Start** workspace.

CHAPTER 2

Exception Tasks

This chapter includes the following topics:

- Exception Tasks Overview, 17
- Types of Exception Task, 17
- Exception Tasks in the Exceptions Workspace, 19
- Fixing Errors in an Exception Task, 20
- Exception Record Status Indicators, 21
- Filter Options for Exception Tasks, 23
- Find and Replace Values in Exception Records, 26
- Adding a Note to a Record, 28
- Exception Tasks in Review, 28
- Rules and Guidelines for Exception Tasks, 29

Exception Tasks Overview

An exception task identifies records that might contain data errors. The records might contain inaccurate data, or they might contain null fields. A workflow generates exception tasks when earlier processes cannot resolve all of the data quality issues in a data set. When you open an exception task, the Analyst tool identifies the fields in each record that contain the unresolved issues.

Types of Exception Task

A workflow can assign you a correct exceptions task or a review exceptions task.

When you work on a correct exceptions task, you examine each record in the task and you decide whether to update the record data. When you work on a review exceptions task, you verify the decisions that another user made for each record in an earlier task.

The workspace layout in each task type is identical. You update the record data and the status indicators in the same manner in both types of task. When you work in a review exceptions task, you can use additional options to accept or reject the work of an earlier user.

Steps to Correct Exceptions

Open a correct exceptions task from the Start workspace. The task opens on the Data Editing tab of the Exceptions workspace.

When you work in a correct exceptions task, perform the following steps:

1. Open the task.
2. Examine the highlighted data fields in each record.
   You can use the arrows to move through records in the task. Use the filter options to show or hide records that contain a data quality issue or a status that you specify.
3. Fix the errors that the data fields contain.
4. Update the status of each record to reflect the current record data.
   Choose one of the following options:
   - Accept. You determine that the current data is acceptable to the business.
   - Reject. You determine that the current data is unacceptable to the business.
   - Reprocess. You cannot determine the status of the current data.
5. Update the record to indicate that you reviewed the record data. You can set a status value of Reviewed independently from the other status values.
   Use the filter options to show or hide records with a common status value.
6. Optionally, add a note to the record. For example, you might add a note to explain why you rejected a record.
7. When you finish work on all records in the task, update the task status.

Steps to Review Exceptions

Open a review exceptions task from the Start workspace. The task opens on the Data Editing tab of the Exceptions workspace.

When you work in a review exceptions task, perform the following steps:

1. Open the task.
2. Review each record. Examine the record data, and examine the status indicators that the previous user set for each record.
   - If you agree with the current content of the records, make no change. If you disagree with the content of any record, update the record.
   - If you agree with the current record status, make no change. If you disagree with the record status, update the status.

   Use the filter options to show or hide records with a common status value.
3. Verify the review status of each record. The review status indicates that you approve or reject the record. The review status supersedes the status that a previous user applies to the record.
4. Optionally, add a note to the record. If a user added a note to a record in an earlier task, the note that you add replaces the older note.
5. When you finish work on all records in the task, update the task status.

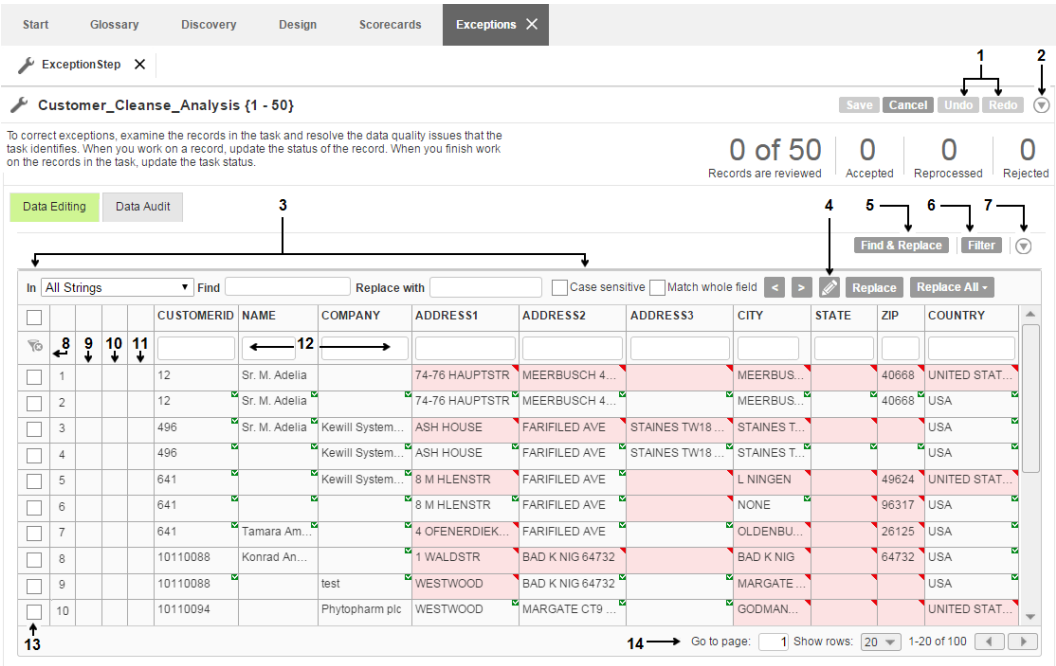
Exception Tasks in the Exceptions Workspace

When you open an exception task, the Exceptions workspace lists the records in the task and displays the options that you can set for the task. The task opens on the Data Editing tab. The tab displays a common set of options for correct exceptions tasks and review exceptions tasks.

The Analyst tool changes the background color of any cell that contains an unresolved data quality issue. Examine the records, and update the cells. When you update a data value, the Analyst tool clears the cell background color and adds a green check mark to the cell. You can use the cursor to reveal any previous value in the cell.

When you finish work on a record, update the record status. When you finish work on all of the records in the task, update the task status to indicate that the records are ready for the next stage in the workflow.

The following image shows the **Data Editing** tab when you open a task that contains exception records in edit mode:



Use the following options when you work on the task:

1. Undo and redo options
   Undo the most recent update to the data or the metadata in the task. Redo any update that you chose to undo.

   You cannot undo a find and replace operation that you perform on all of the pages in the task.
2. Task actions menu
   Perform actions on the current task. Export the task data, or send the task to the next step in the workflow.
3. Find and replace options
   Find and replace the data values that match the strings that you enter.
4. Highlighter
   View the records that do not use a status indicator.
5. Find and Replace
   Show or hide the options to find and replace data values in the exception records.

6. Metadata filter options
   Show or hide the records in the task based on record metadata, such as the type of data quality issue or the record status.
7. Record actions menu
   Update the status of the records that you select. For example, accept the record, reject the record, or clear the record status.
8. Data filter reset
   Clear any value that you entered as a column data filter.
9. Note column
   Read the note that the task owner added to the record.
10. Record status column
    View a status indicator for the most recent action that the task owner performed on the record.
11. Record review status column
    View a status indicator for the most recent status that the task owner defined for the record.
12. Column data filter options
    Show or hide the records in the task based on the data in one or more data columns.
13. Record selector
    Select the records that a record action applies to.
14. Navigation options
    Move to different pages in the task.

Fixing Errors in an Exception Task

You can update the records in a correct exceptions task and in a review exceptions task. Examine the records in the task, and update the fields that contain the data quality errors.

You edit the records in the **Data Editing** tab.

1. Open the exception task.

   Verify that the task displays the **Data Editing** tab.
2. Click **Edit**.
3. Examine the records. Select a data field that contains an error that you can fix.

   Move the pointer over the field to see the type of data error that the field contains.
4. Enter the correct value in the data field.

   The field appearance changes. The Analyst tool clears the background color from the field and adds a green check mark in the top right corner.
5. Save the task.

Repeat the steps for the other records in the task.

Note: The Human task properties might indicate that every error cell must contain a value. In this case, you cannot complete the task until you enter a value in every error cell.

Validation Errors in Exception Tasks

The Analyst tool displays a validation error if you try to enter a data value that is not valid for the column that stores it. You cannot enter a string data value to a numeric column. You cannot enter a data value to a column if the value is longer than the maximum character length in the column.

The Analyst tool can also display validation errors during find and replace operations. You cannot replace a value with another value if the replacement value is not valid in the destination column.

Example: City Names in India

You are a data steward in a multinational organization, and you work on an exception task that identifies older city names in India. You use a list of current and older city names to update the records in the task. For each city, you use the Replace All option to update the name. You use the default option to replace the values in all string data.

One of the columns that contains a city name specifies a maximum character length that cannot accommodate the current name. For example, the name *Thiruvananthapuram* replaces the name *Trivandrum* in the Indian state of Kerala. When the Analyst tool tries to update the name in the column, the operation fails and the Analyst tool displays an error.

Exception Record Status Indicators

When you finish work on a record, set the record status. The status indicates the action that another user or another data process can perform on the record when you complete the task. Set the status that represents the condition of the record data when you finish work on the record.

Select a status indicator from the **Record Actions** menu on the Data Editing tab. The status indicators appear as icons beside the data rows in the task. If a user set a record status in an earlier task, you can update or clear the status.

Select one of the following status indicators when you finish work on a record:

Accept Record

Indicates that the record is acceptable for storage with the organization data.

You can accept a record in a correct exceptions task and in a review exceptions task.

Reject Record

Indicates that the record is not acceptable for storage with the organization data.

Reject a record that you determine cannot correctly identify a business entity in the database. A downstream process might drop the records that you reject from the database tables.

You can reject a record in a correct exceptions task and in a review exceptions task.

Reprocess Record

Indicates that the record contains a data quality issue that you cannot verify.

Reprocess a record when you cannot determine the accuracy of the record data. Another user in a downstream task might verify or update the record data. Or, a downstream process might write the record to a table for analysis and correction in a mapping.

You can reprocess a record in a correct exceptions task and in a review exceptions task.

Mark as Reviewed

Indicates that you reviewed the record data.

Mark a record as reviewed to indicate to other users that you examined the record. The status does not describe the data quality of the record or specify any further action for the record data.

You can set other status indicators when you mark a record as reviewed. For example, you can reprocess a record that you mark as reviewed.

You can mark a record as reviewed in any task.

Approve Record Edit

Indicates that you analyzed the record in a review exceptions task and you determined that the record is acceptable for storage with the organization data.

You can also approve a record edit in a review duplicates task.

Reject Record Edit

Indicates that you analyzed the record in a review exceptions task and you determined that the record is not acceptable for storage with the organization data.

You can also reject a record edit in a review duplicates task.

Clear Record Status

Clears any of the following status indicators from the record:

- Accept Record
- Reject Record
- Reprocess Record

Clear the status if you determine that the current status is incorrect for the record.

You can clear the status indicator in a correct exceptions task and in a review exceptions task.

Clear Reviewer Status

Clears any of the following status indicators from the record:

- Approve Record Edit
- Reject Record Edit
- Mark as Reviewed

Clear the status if you determine that the current review status is incorrect. For example, you might clear the status of a record that another user reviewed in an earlier task.

You can clear the status indicator in any task.

Note: The status indicator must represent the current state of the record. For example, you might reject a record because you identify an error in the record. If you fix the error, update the record status.

Setting the Status of an Exception Record

When you complete work on a record, set a status indicator for the record. You can apply a status indicator to multiple records in a single operation.

You can set the record status in a correct exceptions task and a review exceptions task. If you disagree with the current record status, you can clear the status.

1. Open the exception task.

   Verify that the task shows the **Data Editing** tab.
2. Click **Edit**.
3. Select one or more records.

To select a record, select the check box that begins the data row for the record. To select all records in the task, select the check box in the row that identifies the column names.

Consider the following factors when you select all records:

- The operation to select all records applies to the records in the current workspace.
- If you apply a filter to the task, the operation to select all records applies to the records in the workspace that meet the filter criteria.

4. Open the **Record Actions** menu, and select the status indicator to apply to the records.

   You can select the following status indicators in a correct exceptions task and a review exceptions task:

   - To indicate that a record contains valid data, select **Accept Record**.
   - To indicate that a record does not contain valid data, select **Reject Record**.
   - To indicate that the record needs further analysis, select **Reprocess Record**.
   - To indicate that you examined the record, select **Mark as Reviewed**. You can select the indicator in parallel to another indicator.

   You can select the following status indicators in a review exceptions task:

   - To verify that a record is valid for inclusion in the business data, select **Mark as Accepted**.
   - To verify that a record is not valid for inclusion in the business data, select **Mark as Rejected**.

5. If you disagree with the current status of the record, select one or both of the following options:

   - To clear an indicator that accepts, rejects, or reprocesses a record, select **Clear Record Status**.
   - To clear an indicator that specifies a type of review, select **Clear Review Status**.

Repeat the steps for the other records in the task.

Note: The Record Actions menu includes an option to add a note to a record. Add a note to a record to describe the changes that you made and the status indicators that you selected.

Filter Options for Exception Tasks

Use filters to show or hide the records that the Analyst tool displays for a task. You can define filters on the **Data Editing** tab and on the **Data Audit** tab.

You can use the record data values or the record metadata to define filters. Use record data filters to find records that share one or more data values in common. Use metadata filters to find records that share a common data quality issue or status.

For example, you might define a filter to retrieve every record that a user accepted as valid within a time period that you specify. The Analyst tool retrieves the records that meet the filter criteria. You can review each record and verify the decisions that the user made.

You can apply multiple filters on each tab. When you apply multiple filters, the Analyst tool returns the records that meet all of the filter criteria.

Record Data Filters

Use the record data filters to find records that share a data value in a column that you specify.

Each column has a filter option. To apply a filter, enter a value to the filter at the top of the column. The Analyst tool shows the records that contain the value that you enter in the field that you select.

You can define a filter on multiple columns in the task data. When you define filters on multiple columns, the Analyst tool shows the records that meet all of the filters that you define.

Metadata Filters on the Data Editing Tab

Use the metadata filters on the Data Editing tab to identify a set of records with common metadata characteristics.

The Data Editing tab has the following filter options:

Type of issue

Indicates the types of data quality issue that the workflow identified in the record data. When you select an issue, the Analyst tool displays all records that include the issue.

Move the pointer over a field to view the data quality issue that it represents.

Priority

Indicates the priority that the Mapping task assigned to the data quality issue in the record. The priority is a numeric value. When you select a priority value, the Analyst tool displays all records that include the priority.

Status

Indicates the status that you assigned to the records in the task. When you select a status, the Analyst tool displays all records that use the status.

You can choose the following status options:

- Any. Any record in the task, regardless of status.
- Accepted. Records that are suitable for permanent storage with the organization data.
- Rejected. Records that are not suitable for permanent storage with the organization data.
- Reprocessed. Records that need further analysis in another application.
- Empty. Records with no current status.

Review

Indicates the review status of the record in the current task.

You can choose from the following review options:

- All Records.
- Accepted. Records that a user marked as accepted in a review task.
- Rejected. Records that a user marked as rejected in a review task.
- Reviewed. Records that a user marked as reviewed.
- Empty. Records with no current review status.

Metadata Filters on the Data Audit Tab

Use the metadata filters on the Data Audit tab to identify the records that a user worked on in a time period that you specify. You can also find records that share common status indicators.

You cannot edit the record data on the Data Audit tab. To edit the record data, use the Data Editing tab.

The Data Audit tab has the following filter options:

Date of update

Defines a time to apply to the task data. When you define a time period, the Analyst tool returns all of the records that a user updated during the period.

Updated by

Identifies an Analyst tool user. When you identify a user, the Analyst tool returns all of the records that the user updated.

Status

Indicates the status that you assigned to the records in the task. When you select a status, the Analyst tool displays all records that use the status.

You can choose the following status options:

- Any. Any record in the task, regardless of status.
- Accepted. Records that are suitable for permanent storage with the organization data.
- Rejected. Records that are not suitable for permanent storage with the organization data.
- Reprocessed. Records that need further analysis in another application.
- Empty. Records with no current status.
- Cleared. Records for which a user cleared the status.

Review

Indicates the review status of the record in the current task.

You can choose from the following review options:

- All Records.
- Accepted. Records that a user marked as accepted in a review task.
- Rejected. Records that a user marked as rejected in a review task.
- Reviewed. Records that a user marked as reviewed.
- Empty. Records with no current review status.
- Cleared. Records for which a user cleared the review status.

Filtering Records in an Exception Task

Use the metadata filters to retrieve the set of records that share the data characteristics that you specify. You can update the records that meet the filter criteria on the Data Editing tab.

1. Open a correct exceptions task or a review exceptions task.

   Select the Data Editing tab or the Data Audit tab.
2. Click **Filter**.

   The Filter panel opens.
3. Select the filter criteria to apply to the task data.

   The Data Editing tab and the Data Audit tab display different sets of filter options.

4. Click **Apply Filter**.

   The Analyst tool retrieves the records that meet the filter criteria. The Analyst tool displays the filter criteria that you define above the task cluster list.

   Note: If you select all of the records on the Data Editing tab after you apply a filter, you select every record that meets the filter criteria in the current workspace view. You do not select every record in the task.

Find and Replace Values in Exception Records

You can search the records in a task for a data value that you specify. You can replace the value with another value. You can find and replace a value in a single record, and you can find and replace values in multiple records in a single operation.

When you configure the find and replace options, you can select a single column by name. Or, you can configure the options to search any column that uses a date, string, or numeric data type.

When you search the task by data type, the Analyst tool searches for the value that you enter in any column that uses the data type. You can find and replace any value that a string data column contains, including numbers and dates. You can perform case-sensitive searches on string column data. You can search for a partial match or a complete match between the search value and the contents of a field in a string data column. When you find a partial match, the Analyst tool replaces all instances of the characters that you enter in the record field.

The following table shows the result of a partial find and replace operation on a string of name data:

Search Value	Replacement Value	String to Search	Result
SMITH	SMYTH	JOHN SMITH AND MARY SMITH	JOHN SMYTH AND MARY SMYTH

When you search for a date, enter the date value in the date and time stamp format that the date columns use. You can replace a null value in a date column with a date value that you specify. You cannot search for part of a date value or part of a number in a date data column or a numeric data column. You cannot perform case-sensitive searches on the contents of a date data column or a numeric data column.

You can find and replace values when you filter the task data. The find and replace operations apply to the records that meet the filter criteria. When you search on filtered data, the search operation opens the first page of task data that contains a record that meets the filter criteria.

Click **Find & Replace** to show or hide the options in the task.

Replacing a Single Value in Exception Records

Find and replace a data value in a column that you specify or in any date, string, or numeric column. Use the filters to find and replace the value in a subset of the records in the task.

1. Open the exception task.
2. On the **Data Editing** tab, click **Edit**.
3. Click **Find & Replace**.

   The Analyst tool shows the find and replace options above the list of records.

4. Enter a data value in the **Find** field.
5. Specify a column to search. Or, specify a data type.

   If you select a data type, the Analyst tool searches the columns that use the data type. By default, the Analyst tool searches for the data value in all string data columns.
6. If you search for a data value in a string data column, select or clear the following options:
   - Case sensitive. Perform a case-sensitive search on the string data.
   - Match whole field. Search for a match between the value that you enter and the complete contents of a field in a record.

   You cannot perform case-sensitive searches or partial data searches on columns of date data or numeric data.
7. Click **Find**.

   The Analyst tool highlights the first record that contains the value that you specify. Use the arrows to move through the records.
8. Optionally, enter a value in the **Replace with** field.
9. To replace the highlighted value with the value that you entered, click **Replace**.

   Repeat the step to find and replace additional instances of the value in the task.

Replacing Multiple Values in Exception Records

Use the Replace All option to replace multiple instances of a value in a single operation. You can find and replace the value in the records on the current page, on all pages, or in the records that you select. Use the filters to find and replace the value in a subset of the records in the task.

1. Open the exception task.
2. Optionally, select multiple records.

   Select multiple records to restrict the Replace All operation to a set of records within the task. For example, you might select all of the records that meet the filter criteria that you defined.
3. On the **Data Editing** tab, click **Edit**.
4. Click **Find & Replace**.

   The Analyst tool shows the find and replace options above the list of records.
5. Enter a data value in the **Find** field.
6. Specify a column to search. Or, specify a data type.
7. Enter a value in the **Replace with** field.
8. Click **Replace All**.

   Select one of the following options:
   - Replace on this page. Replaces all instances of the value that you entered in the records on the current page.
   - Replace on all pages. Replaces all instances of the value that you entered in all of the records in the task.
   - Replace on all selected. Replaces all instances of the value that you entered in the records that you selected.

   Note: The operation does not replace data values in any column that you do not have permission to view and edit.

Adding a Note to a Record

Add a note to describe the action that you perform on a record. You can view the note on the Data Editing tab and on the Data Audit tab.

1. Open the task that contains the record.
2. In the Exceptions workspace, select the **Data Editing** tab.
3. Click **Edit**.
4. Use the check box options to select a record.

   You can select multiple records. When you select multiple records, you add the same note to all of the records that you select.
5. Open the Record Actions menu, and select **Add Note**.

   The Analyst tool opens a text entry dialog box.
6. Enter the note text to the dialog box.

Exception Tasks in Review

When you review the records in a review exceptions task, you validate the work done by another user in an earlier task. You can perform the same operations in a review exceptions task and in a correct exceptions task.

Examine the records in the task and confirm the status that the earlier user assigned to each record. If you find an error in a record, update the record. If you disagree with the current status of a record, update or clear the status.

Use the **Record Actions** menu options to update the record status. When you finish work on the task, use the **Task Actions** menu options to update the task status.

Working in a Review Exceptions Task

Verify that the records in the task are as accurate and complete as possible.

1. Open a review exceptions task.
2. Verify the record data, and verify the status that earlier users assigned to each record.
3. If you decide to update a record or to change the record status, click **Edit**.

   To fix an error, click the field that contains the error and enter the correct data value.

   To update the record status, perform one of the following actions:

   - To indicate that a record contains valid data, select **Accept Record**.
   - To indicate that a record does not contain valid data, select **Reject Record**.
   - To indicate that the record needs further analysis, select **Reprocess Record**.
   - To indicate that you examined the record, select **Mark as Reviewed**. You can select the indicator in parallel to another indicator.
   - To verify that a record is valid for inclusion in the business data, select **Mark as Accepted**.
   - To verify that a record is not valid for inclusion in the business data, select **Mark as Rejected**.

- To clear an indicator that accepts, rejects, or reprocesses a record, select **Clear Record Status**.
- To clear an indicator that specifies a type of review, select **Clear Review Status**.

4. Save the task.

   When you finish work on the task, update the task status. The task status indicates that the records are ready for the next stage in the workflow.

Rules and Guidelines for Exception Tasks

The process to review and update the exception records in a data set is collaborative. You might work on records that another user analyzed in an earlier task. Or, the records that you work on might pass to another user when you complete a task. Each user can review and update the work of the other user.

Consider the following rules and guidelines when you work in an exception record task:

- When you set the status of a record, you can indicate that you reviewed the record. The review status does not describe the accuracy or the data quality of the record. For example, you can accept, reject, and reprocess a series of records, and you can mark each record as reviewed.

  As a best practice, mark every record that you examine as reviewed. The status confirms to a user in a downstream task that another user examined the record. When you update the data in a record, mark the record as reviewed regardless of the presence or absence of another status indicator on the record.
- You can set any status indicator when you update a record. For example, you can update a record that you reprocess. The update that you make can help the next user or a downstream data process to analyze and repair the data.
- The data values and the status indicators on a record can change independently of one another across multiple tasks. When you update a record, verify that the data values and the status indicators are current and accurate. The changes that you make can overrule the decisions that another user made.
- The audit trail stores every change that a user makes to the data values and the status indicators on a record. The audit trail does not store changes to the text in a note that you add to the record.
- The data that you work on can pass to a task that corrects data or a task that reviews data. For example, a developer who configures a Human task in a workflow might specify multiple correct exceptions tasks in sequence. The developer might follow a review duplicates task with a correct duplicates task.

CHAPTER 3

Duplicate Record Tasks

This chapter includes the following topics:

- Duplicate Record Tasks Overview, 30
- Preferred Records, 31
- Types of Duplicate Record Task, 31
- Duplicate Record Tasks in the Exceptions Workspace, 32
- Editing a Cluster , 34
- Duplicate Record Status Indicators, 35
- Finding Records in Multiple Clusters, 36
- Moving Records Between Clusters, 37
- Moving a Record to an Empty Cluster, 38
- Adding a Note to a Cluster, 38
- Filter Options for Duplicate Tasks, 39
- Review Duplicates Tasks, 40
- Rules and Guidelines for Duplicate Record Tasks, 42

Duplicate Record Tasks Overview

A duplicate record task identifies the records in a data set that might contain duplicate or redundant information. The task displays the records as a series of clusters. Each cluster identifies a set of records that contain similar or identical data values.

When you work on a duplicate record task, you review each cluster and you determine whether the records in the cluster are duplicates of each other. If the records are duplicates, you define a preferred version of the record that the cluster represents. If the cluster contains a unique record, you move the record to an empty cluster. If a record matches the records in another cluster, you can move the record to the other cluster.

Complete the task after you review all of the clusters and define a preferred record for each cluster.

Note: Two or more records are duplicates when they represent the same entity in the source data set. Records can contain similar data and not represent the same information to the business. Your organization can defines the business rules that identify duplicate records.

Preferred Records

When you work on a cluster, you create or verify the most accurate and complete version of the record that the cluster represents. The record that you create or verify is the preferred record.

The first row in the cluster contains the preferred record data. To update the preferred record, promote data values from the other records in the cluster to the preferred record.

By default, the Analyst tool populates the preferred data row with data from the first record in the cluster. The Analyst tool highlights the preferred data row. The Analyst tool also highlights the record that contains the default data. When you promote a data value to the preferred record, the Analyst tool highlights the value that you promote. You can edit a value that you add to the preferred record. You cannot edit a value in another record in the cluster. If the default record is correct, you can accept the default record.

Note: The preferred record is not a member of the cluster. The preferred record is a unique record that the workflow creates in the duplicate record database.

Types of Duplicate Record Task

A workflow can assign a correct duplicates task or a review duplicates task to you.

When you work on a correct duplicates task, you examine the records in each cluster in the task. You open a cluster, and you verify that the preferred record in the cluster contains the most accurate data available. When you work on a review duplicates task, you verify the decisions that another user made for each cluster in an earlier task.

The workspace layout in each task type is identical. The workspace options that you use in each task type are similar. You update the cluster data and the status indicators in the same manner in both types of task. When you work in a review duplicates task, you can use additional options to accept or reject the work of an earlier user.

Steps to Correct Duplicates

Open a review duplicates task from the Start workspace. The task opens on the Data Editing tab of the Exceptions workspace.

When you work in a correct duplicates task, perform the following steps:

1. Open the task, and enter edit mode.
   The Analyst tool displays the first cluster in the task. You can work in the current cluster or you can open another cluster. Use the filter options to show or hide clusters that contain a data value or a status that you specify.
2. Examine the preferred record and the other records in the current cluster.
   If a field in another record contains a more accurate value than the same field in the preferred record, promote the value to the preferred record.
3. Optionally, perform any of the following tasks to verify that the cluster contain the most accurate data:
   - Edit the values in the preferred record.
   - Search for records that contain similar values in other clusters.
   - Move a record from one cluster to another.
   - Create a cluster and move a record to the cluster that you created.

4. Update the cluster status to indicate that you reviewed the cluster.
5. When you finish work on all of the clusters in the task, update the task status.

Steps to Review Duplicates

Open a correct duplicates task from the Start workspace. The task opens on the Data Editing tab of the Exceptions workspace.

When you work in a review exceptions task, perform the following steps:

1. Open the task, and enter edit mode.
   The Analyst tool displays the first cluster in the task. You can work in the current cluster or open another cluster. Use the filter options to show or hide clusters that contain a data value or a status that you specify.
2. Review the contents of each cluster. Verify or update the status that a user assigned to the cluster in another task.
   Use a filter to find the clusters that the user reviewed or did not review.
   - If you agree with the content and the status of the cluster, make no change.
   - If you disagree with the content of any cluster, update the cluster.
   - If you disagree with the status of any cluster, update the status.
3. When you finish work on all of the clusters in the task, update the task status.

Duplicate Record Tasks in the Exceptions Workspace

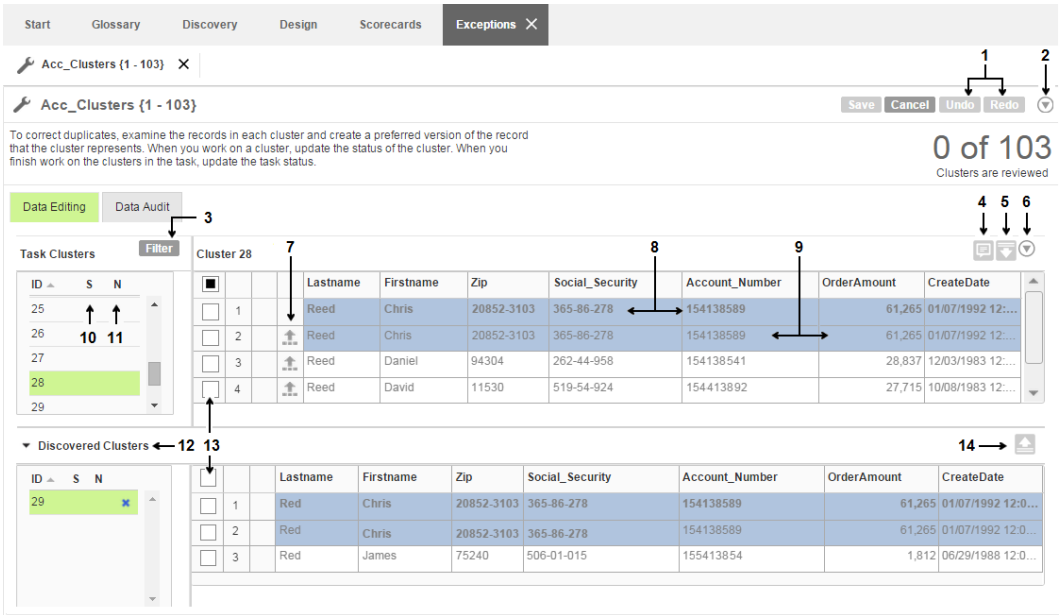
When you open a duplicate record task, the Exceptions workspace lists the clusters in the task and displays the records in the first cluster in the task. The task opens on the Data Editing tab. The tab displays a common set of options for correct duplicates tasks and review duplicates tasks.

The Analyst lists the clusters in the task on the left side of the workspace. Select a cluster to view the records that it contains. The first record in the cluster is the current preferred record. Review the other records in the cluster, and determine if the other records contain any data that belongs in the preferred record. To add a value from a record to the preferred record, click the value. To replace the preferred record with another record, select the record and use the column options to promote the record.

When you add a value to the preferred record, the workspace highlights the value that you added. When you replace the preferred record with another record, the workspace highlights the value that you promoted. To view a complete list of the changes that you make to the task, select the **Data Audit** view.

When you finish work on a cluster, update the cluster status. When you finish work on the clusters in the task, update the task status to indicate that the clusters are ready for the next stage in the workflow.

The following image shows the **Data Editing** tab when you open a task that contains clusters in edit mode:



Use the following options when you work on the task:

1. Undo and redo options
   Undo the most recent update to the data or the metadata in the task. Redo any update that you chose to undo.
2. Task actions menu
   Perform actions on the current task. Export the task data, or send the task to the next step in the workflow.
3. Filter options
   Find clusters in the task that meet the criteria that you specify. When you apply a filter, you close the current cluster and you open the first cluster that meets the filter criteria.

   The filter also applies to the clusters that you discover. When you apply a filter, the Discovered Clusters list shows the clusters within the discovered list that meet the filter criteria.
4. Note option
   Read a note that a user added to the current cluster on the Task Clusters list.
5. Move a record to the lower cluster
   Moves one or more records that you select from the upper cluster to the lower cluster. Use the option when you open a cluster from the Discovered Clusters list.
6. Cluster actions menu
   Update the status of the current cluster on the Task Clusters list. You can add a note to the cluster, mark the cluster as reviewed, or clear the review status of the cluster.
7. Promotion option
   Updates the data in the preferred record with the data from the row that you select.
8. Preferred record data
   The current data for the preferred record in the cluster.
9. Source record for the preferred record data
   The record in the cluster that provided the source data for the preferred record. By default, the preferred record is the record in the cluster with the sequence ID value.
10. Cluster review status column
    View a status indicator for the current status of the cluster.

11. Cluster note column
    Indicates the clusters to which a user added a note.
12. Discovered Clusters options
    Use to find clusters that contain one or more data values that you specify and to create clusters.
13. Record selection options
    Use to select one or more records.
14. Move a record to the upper cluster
    Moves one or more records that you select from the lower cluster to the upper cluster. Use the option when you open a cluster from the Discovered Clusters list.

Editing a Cluster

Examine the preferred record and any other record in the cluster. Verify that the preferred record contains the most accurate data for the business entity that the cluster represents.

If the current preferred record is correct, make no change.

1. Open a correct duplicates task.

   The list of the clusters in the task appears in a column on the Data Editing tab. The first cluster opens by default.
2. Optionally, select a different cluster.

   For example, select a different cluster if the current cluster has a status of Reviewed.
3. Examine the records in the cluster. Determine if the preferred record contains the most accurate data for the business entity that the cluster represents.

   The preferred record is the first record in the cluster. The Analyst tool highlights the preferred record and the source record for the preferred data.
4. Click **Edit**.
5. To update the preferred record, perform any of the following actions:
   - Promote a data value from another record to the preferred record.

     To promote a value, click the data field in the source record. The value that you select replaces the value in the same column in the preferred record.
   - Update a data value in the preferred record.

     Update the preferred record if no record in the cluster contains an accurate version of the data.
   - Replace the current preferred record with another record in the cluster.

     To replace the preferred record, click the promotion tool in the row that contains the record.
6. Save the cluster to the task.

When you complete work in a cluster, update the cluster status to Reviewed.

Related Topics:

- "Finding Records in Multiple Clusters" on page 36
- "Moving a Record to an Empty Cluster" on page 38

Duplicate Record Status Indicators

When you finish work on a cluster, set the cluster status. The status indicates the action that another user or another data process can perform on the cluster data when you complete the task. Set the status that represents the condition of the preferred record when you finish work on the cluster.

Select a status indicator from the **Cluster Actions** menu on the Data Editing tab. The status indicators appear as icons in the list of clusters. If a user set a cluster status in an earlier task, you can update or clear the status.

Select one of the following status indicators when you finish work on a cluster:

Mark as Reviewed

Indicates that you reviewed the cluster data.

Mark a cluster as reviewed to indicate to other users that you examined the cluster. The status does not describe the status of the preferred record or specify any further action for the record data.

You can mark a cluster as reviewed and set another status indicator for the record. For example, you can reject a preferred record that you mark as reviewed.

You can mark a cluster as reviewed in any task.

Clear Reviewer Status

Clears any status indicator from the cluster. Clear the status if you determine that the current status is incorrect.

You can clear the status indicator in any task.

Approve Record Edit

Indicates that you analyzed the cluster in a review duplicates task and you agree with the current preferred record. Approve the record if you determine that the record contains the most accurate data for the business entity that the cluster represents.

You can also approve a record edit in a review exceptions task.

Reject Record Edit

Indicates that you analyzed the cluster in a review duplicates task and you disagree with the current preferred record. Reject the record if you determine that the record does not contain the most accurate data for the business entity that the cluster represents.

You can also reject a record edit in a review exceptions task.

Note: The status indicator must represent the current state of the preferred record. For example, you might reject a preferred record because you identify an error in the record. If you fix the error, update the cluster status.

Setting the Status of a Cluster

When you complete work on a cluster, set the cluster status. Set a status that represents the current condition of the records in the cluster. The status determines the action that another user or another data process performs on the data in the next stage of the workflow.

To set the cluster status, use the **Cluster Actions** menu options.

1. Open a correct duplicates task or a review duplicates task.

   The list of the clusters in the task appears in a column on the Data Editing tab. The first cluster opens by default.

2. Optionally, select a different cluster.
3. Click **Edit**.
4. Optionally, update the preferred record or update the contents of the cluster.
5. To update the cluster status, perform one or more of the following actions:
   - Mark the cluster as reviewed.

     Mark a cluster as reviewed to indicate to other users that you examined the cluster. You can mark the cluster as reviewed in a correct duplicates task and in a review duplicates task.
   - Clear the reviewer status.

     Clear the status to remove any status indicator from the cluster. You can clear the reviewer status in a correct duplicates task and in a review duplicates task.
   - Approve the preferred record in the cluster.

     Approve the record to indicate that you agree with the current preferred record in the cluster. You can approve the preferred record in a review duplicates task.
   - Reject the preferred record in the cluster.

     Approve the record to indicate that you disagree with the current preferred record in the cluster. You can reject the preferred record in a review duplicates task.
6. Save the cluster to the task.

After you review all of the clusters in the task, update the task status. The task status indicates that the cluster records are ready for the next stage in the workflow.

Finding Records in Multiple Clusters

You can search for records in other clusters that contain one or more data values that you specify. Use the Discovered Clusters options to search for the records.

Search for records in other clusters that might match the records in the current cluster. If the records represent the same business entity in the organization data set, move the records to the same cluster.

1. Open a correct duplicates task.

   The list of the clusters in the task appears in a column on the Data Editing tab. The first cluster opens by default.
2. Optionally, select a different cluster.
3. Expand the **Discovered Clusters** options.

   By default, the options do not display any cluster data.
4. Click **Find Cluster(s)**.
5. Select the column that contains the data value to search for.
6. Enter the data value to search for.

   Enter the data value as it appears in the record column, or enter a wildcard character in a text pattern. You can use the asterisk (*) or percent sign (%) wildcard characters.
7. Optionally, enter additional data values for other columns.
8. Click **Find**.

   The search operation returns the records in the task data that contain the value that you searched for. If you searched for multiple values, the search operation returns the records that contain every value.

9. Select one or more records in the search results, and click **Open**.

   The Data Editing tab displays the clusters that contain the records that you selected.

10. Click a cluster to open it on the Data Editing tab.
11. Compare the records in the clusters.

To update a cluster in the Data Editing tab, click **Edit**.

Related Topics:

- "Moving a Record to an Empty Cluster" on page 38
- "Editing a Cluster " on page 34

Moving Records Between Clusters

If two or more clusters contain records that represent the same entity, move the records from one cluster to the other. Move a record that might contribute data to the preferred record in the other cluster.

You can open the clusters on the Data Editing tab and move a record between the clusters.

1. Open a correct duplicates task.

   The list of the clusters in the task appears in a column on the Data Editing tab. The first cluster opens by default.

2. Optionally, select a different cluster.

   For example, select a different cluster if the current cluster has a status of Reviewed.

3. Click **Edit**.
4. Find a record in another cluster that matches a record in the current cluster.

   Use the **Discovered Clusters** options to find the record. The search operation might return records in multiple clusters.

5. Open a cluster that contains the record that you find.
6. Compare the records in the clusters.
7. If a record in a cluster is a better match with the records in another cluster, move the record to the cluster.

   Use the **Move Records** icons to move the records.

8. Save the clusters to the task.

Moving a Record to an Empty Cluster

Move a record to an empty cluster when you identify a valid record that does not belong in any cluster in the task. Move the record to the cluster that you create.

A valid record contains information about a business entity that belongs in the organization data set. The record might be unique, or it might match a record in another cluster. You can add multiple records to the cluster that you create.

1. Open a correct duplicates task.

   The list of the clusters in the task appears in a column on the Data Editing tab. The first cluster opens by default.
2. Optionally, select a different cluster.
3. Examine the records in the cluster.

   Verify that the cluster contains a valid record that does not match the other records in the cluster.
4. Click **Edit**.
5. Expand the **Discover Clusters** options.
6. Click **Create Cluster**.

   An empty cluster appears below the current cluster.
7. Select the record to add to the cluster.
8. Move the record to the cluster.

   The record becomes the preferred record of the cluster you created.

   Note: Find the **Move Records** icon beside the **Note** icon in the current cluster.
9. Move any other record that matches the preferred record in the cluster that you created.

Related Topics:

- "Finding Records in Multiple Clusters" on page 36
- "Editing a Cluster " on page 34

Adding a Note to a Cluster

Add a note to describe the action that you performed on a cluster. You can view the note on the Data Editing tab and on the Data Audit tab.

1. In the Exceptions workspace, select the **Data Editing** tab.
2. Open a cluster.
3. Open the Cluster Actions menu, and select **Add Note**.

   The Analyst tool opens the Note box.
4. Enter a comment in the Note box.

Filter Options for Duplicate Tasks

You can use filters to retrieve the clusters from a task that meet the data criteria that you specify. You can define a filter on the Data Editing tab and on the Data Audit tab.

Use the filter options to show or hide the clusters that contain a data value or that share a status that you specify. The Data Audit tab contain additional options to show or hide clusters based on the saved date and the review status of the clusters.

You can apply multiple filters on each tab. When you apply multiple filters, the Analyst tool returns the records that meet all of the filter criteria.

Data Filters in Duplicate Record Tasks

Use the filters on the Data Editing tab to identify a set of clusters with common data features. When you apply a filter, the Analyst tool lists the clusters that meet the filter criteria.

The Data Editing tab has the following filter options:

Value

Returns the clusters that include one or more records with a value in a column that you select. Select the column name and enter the value.

Status

Returns the clusters that display the status that you select.

You can choose the following status indicators:

- Any. Any cluster in the task, regardless of status.
- Accepted. Clusters that contain a preferred record that is suitable for storage with the organization data.
- Rejected. Clusters that do not contain a preferred record that is suitable for storage with the organization data.
- Reviewed. Clusters that you reviewed. The status does not indicate the status of the preferred record.
- Empty. Clusters with no current status.

Audit Trail Filters in Duplicate Record Tasks

Use the filters on the Data Audit tab to show or hide clusters based on the cluster metadata. You can filter the clusters by date, by user, and by status.

The Data Audit tab has the following filter options:

Date of update

Returns the clusters that a user updated during the time period that you specify.

Updated by

Returns the clusters that a user updated. Identifies an Analyst tool user. To filter by user, enter a user name.

Status

Returns the clusters that contain the status that you specify.

You can choose the following status options:

- Any. Any cluster in the task, regardless of status.
- Moved to cluster. Any cluster that contained a record that a user moved to another cluster.
- Moved from cluster. Any cluster that contains a record that a user moved from another cluster.
- Empty. Clusters with no current status.

Review

Returns the clusters that contain the review status that you specify.

You can choose the following review options:

- All Clusters. Any cluster in the task, regardless of review status.
- Reviewed. Clusters that a user marked as reviewed.
- Approved. Clusters that a user marked as accepted in a review task.
- Rejected. Clusters that a user marked as rejected in a review task.
- Empty. Clusters with no current review status.
- Cleared. Clusters with an earlier review status that a user cleared.

Filtering Clusters in a Duplicate Record Task

Use filters to retrieve the set of records that share the data characteristics that you specify. When you apply a filter on the Data Editing tab, the Analyst tool displays the clusters that contain one or more records that meet the filter criteria. When you apply a filter on the Data Audit tab, the Analyst tool displays all of the records that meet the filter criteria.

The clusters that the filter returns on the Data Editing tab might contain records that do not meet the filter criteria. The filter returns the clusters that contain at least one record that meets the criteria.

1. Open a correct duplicates task or a review duplicates task.
2. Select the Data Editing tab or the Data Audit tab.
3. Click **Filter**.

   The Filter panel opens.

   Note: The Data Editing tab and the Data Audit tab display different sets of filter options.
4. Select the filter criteria to apply to the task data.
5. Click **Apply Filter**.

   The Analyst tool retrieves the records that meet the filter criteria.

   The Analyst tool displays the filter criteria that you define above the task cluster list.

Review Duplicates Tasks

When you review the clusters in a review exceptions task, you validate the work done by another user in an earlier task. You can perform the same operations in a review duplicates task and in a correct duplicates task.

Examine the records in the cluster and confirm that the preferred record represents the most accurate version of the records in the cluster. You can update the preferred record and the cluster status if you disagree with the decisions of the earlier user.

Use the **Cluster Actions** menu options to update the record status. When you finish work on the task, use the **Task Actions** menu options to update the task status.

Note: Verify that the other records in the cluster do not include any record that the business might want to keep. You can move the records to another cluster, and you can create a cluster to store the record.

Working in a Review Duplicates Task

Verify that the preferred record in each cluster represents the most accurate version of the records in the cluster.

1. Open a review duplicates task.

   The list of the clusters in the task appears in a column on the Data Editing tab. The first cluster opens by default.

2. Examine the records in the cluster.

   Verify that the preferred record contains the most accurate version of the data in the cluster.

   Verify that the cluster status is Reviewed.

3. To update the contents of the cluster or the cluster status, click **Edit**.

   To update the preferred record, perform one or more of the following actions:

   - Promote a data value from another record to the preferred record.

     To promote a value, click the data field in the source record. The value that you select replaces the value in the same column in the preferred record.

   - Update a data value in the preferred record.

     Update the preferred record if no record in the cluster contains an accurate version of the data.

   - Replace the current preferred record with another record in the cluster.

     To replace the preferred record, click the promotion tool in the row that contains the record.

   You can also move a record to another cluster or import a record from another cluster. To move records between clusters, use the **Discovered Clusters** options.

   To update the cluster status, perform one of the following actions:

   - Mark the cluster as reviewed.

     Mark a cluster as reviewed to indicate to other users that you examined the cluster.

   - Approve the preferred record in the cluster.

     Approve the record to indicate that you agree with the current preferred record in the cluster.

   - Reject the preferred record in the cluster.

     Approve the record to indicate that you disagree with the current preferred record in the cluster.

   - Clear the reviewer status.

     Clear the status to remove any status indicator from the cluster.

4. When you finish work on the cluster, save the cluster to the task.

After you review all of the clusters in the task, update the task status. The task status indicates that the cluster records are ready for the next stage in the workflow.

Rules and Guidelines for Duplicate Record Tasks

The process to review and update the exception records in a data set is a collaborative one. You might work on clusters that another user analyzed in an earlier task. Or, the clusters that you work on might pass to another user when you complete a task. Each user can review and update the work of the other user.

Consider the following rules and guidelines when you work in a duplicate record task:

- A cluster is a set of records in a database table that share similar or identical data values. A developer defines the criteria that sort the records into clusters. If you believe that a record does not belong in the current cluster, use the Discovered Clusters options to find the correct cluster. If you believe that another cluster contains a record that belongs in the current cluster, use the Discovered Clusters options to find the correct cluster.
- When you work on a cluster, use the preferred record to define the most version of the business entity that the cluster represents. The preferred record is not necessarily the final version of the record. Another user or another data process might work on the cluster after you complete the task.

  When you update the preferred record, you update a record in the exception database that represents the preferred form of the records in the cluster. You do not update the source data in the cluster.
- When you set the status of a cluster, you can indicate that you reviewed the cluster. The review status does not describe the accuracy or the data quality of the preferred record in the cluster.

  As a best practice, mark every cluster that you examine as reviewed. The status confirms to a user in a downstream task that another user examined the cluster. When you update the data in a record, mark the record as reviewed regardless of the presence or absence of another status indicator on the record.
- The audit trail stores every change that a user makes to the preferred record. The audit trail does not store changes to the cluster data.
- The data that you work on can pass to a task that corrects data or a task that reviews data. For example, a developer who configures a Human task in a workflow might specify multiple correct exceptions tasks in sequence. The developer might follow a review duplicates task with a correct duplicates task.

CHAPTER 4

Task Management

This chapter includes the following topics:

- Task Management Overview, 43
- Task Performer Operations, 43
- Business Administrator Operations, 45
- Task Data Export, 48

Task Management Overview

A developer who configures a Human task in a workflow nominates the users who can work on the task data. When you log in to the Analyst tool, the Start workspace lists the task instances that the workflow assigns to you. You can work on a task as a task performer or as a business administrator.

As a task performer, you can update records in the task and the status indicators in the task. As a business administrator, you can perform the same operations as a task performer. You can also reassign tasks from one user to another and move tasks to the next stage in the workflow.

A workflow can assign you different roles on different tasks. You might be a task performer and a business administrator for different tasks in a single workflow.

Task Performer Operations

The workflow developer selects the task performers who can work on the instances of a Human task. To view the list of task instances that you can work on, open the My Tasks view.

You can perform the following operations in the **My Tasks** view:

View the contents of a task

View the records or the clusters in a task. The task contents appear in the **Exceptions** workspace.

Open a task

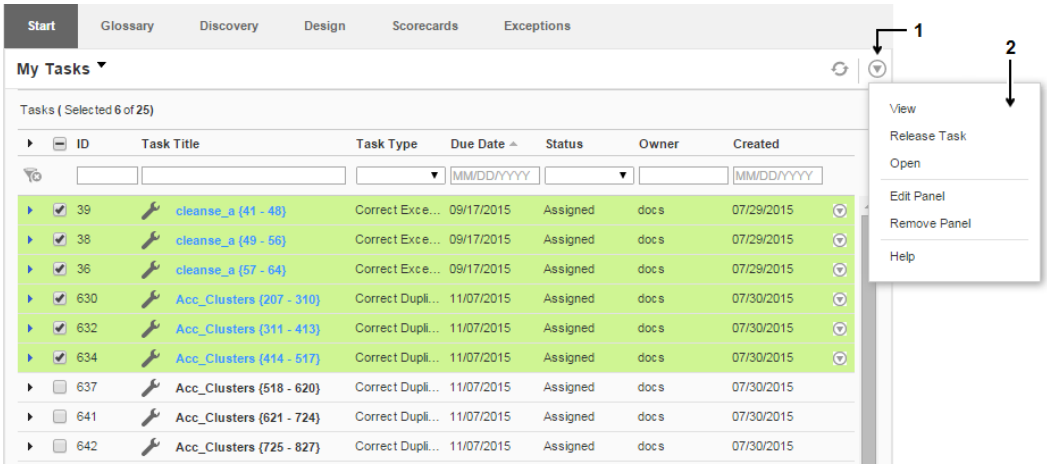
Open a task to review and edit the records. When you open the task, you become the task owner. Other task performers cannot open a task that you own.

Release a task

Release a task that you own. The task that you release has no owner until the business administrator assigns an owner or another user opens the task.

You can release a task at any time. You do not need to complete work on a task to release the task.

The following image shows the My Tasks view in the Start workspace:



Use the following options to work on tasks:

1. Task menu icon.
2. Task menu options.

You can view, open, and release multiple tasks in a single operation.

Viewing Tasks

When you view a task, you open the task in read-only mode. As a task performer, you can view the contents of a task that has no owner. A business administrator can view the contents of any task on the Task Administration view.

You can open multiple tasks in read-only mode in a single operation. Each task opens on a separate tab in the Exceptions workspace. Use the task check boxes to select the tasks.

1. On the Start workspace, open the My Tasks view.
2. Select one or more tasks.

   Note: Do not click the task name. When you click a task name, you open the task and claim ownership of the task.
3. From the task menu, select **View**.

The tasks that you select open in the Exceptions workspace.

Opening a Task

Open a task to work on the task data. When you open a task, you claim ownership of the task. Other task performers cannot edit a task that you own.

You can open multiple tasks in a single operation. Each task opens on a separate tab in the Exceptions workspace. Use the task check boxes to select the tasks.

1. On the Start workspace, open the My Tasks view.

2. To open a single task, click the task name.

   Or, select one or more tasks.

3. From the task menu, select **Open**.

   The tasks that you select open in the Exceptions workspace.

Releasing a Task

When you release a task, you no longer own the task. Another user can claim ownership of the task, or a business administrator can assign the task to another user. The task saves any work that you performed on the task data.

You can open release multiple tasks in a single operation. Use the task check boxes to select the tasks. A business administrator can release any task on the Task Administration view.

1. On the Start workspace, open the My Tasks view.
2. Select one or more tasks.
3. From the task menu, select **Release Task**.

   The Analyst tool updates the task list to indicate that the tasks have no owner.

Business Administrator Operations

The workflow developer might select you as a business administrator for the task instances that a workflow generates. To view the list of task instances that you administer, open the Task Administration view.

As a business administrator, you can assign tasks to users, view the work that the users perform, and move tasks to the next stage in the workflow. You can also claim ownership of tasks and update task data in the same manner as a task performer. The Task Administration view displays the list of tasks that you manage and own.

You can perform the following operations in the Task Administration view:

View tasks and open tasks that you own

View the data in any task, and open any task that you own. Take ownership of a task that does not have an owner.

Complete multiple tasks in a single operation

Complete all of the tasks that originate from the same Human task as the task that you select. When you complete the tasks, the task data moves to the next stage in the workflow.

Reassign tasks

Reassign or delegate a task from one user to another user. For example, reassign a task if the current user cannot complete the task on schedule.

Release tasks

Release a task that you administer. The task that you release has no owner until you assign an owner or another user opens the task.

Review the work that a user performed in a task

Use the task summary data to measure the rate of user progress in a task.

View a list of the tasks that have a common parent task

View a list of the tasks that originate from the same Human task as the task that you select.

You can view and reassign multiple tasks in a single operation.

Assigning a Task to a User

Assign a task when the task has no owner, or assign a task to another user or group when the current user cannot complete the task.

You can assign multiple tasks in a single operation. Use the task check boxes to select the tasks.

1. Open the **Task Administration** view.
2. Select one or more tasks from the task list.
3. From the task menu, select **Reassign Tasks**.
4. Select a user or group name.

   Use the filter to find a user or group.
5. Optionally, enter a comment.
6. Click **Assign**.

   The Task Administration view updates the records to show the task owner that you selected.

Viewing the List of Task Instances in a Human Task

You can view a list of the task instances that a workflow generates for a Human task.

Note: When you review the list of task instances, do not click **OK**. If you click **OK**, you move the data that the tasks specify to the next stage in the workflow.

1. Open the Task Administration view.
2. Select a task from the task list.
3. From the task menu, select **Complete Linked Tasks**.
4. Review the list of tasks.

   The task list displays the following information for each task:

   - Task ID.
     The unique identifier of the task instance within the Human task.
   - Task title.
     The name of the task that appears in the Task Administration view.
   - Task type.
     The type of task that the Human task generated. A task might be a correct exceptions task, a review exceptions task, a correct duplicates task, or a review duplicates task.
   - Task owner
     The name of the user who owns the task.
   - Due date
     The date by which the user must complete the task.
   - Status
     The ownership status of the task. Before a user claims a task, the task status is Created. When a user claims a task, or when you assign a task to a user, the task status is Assigned.

- Created.
  The date on which the workflow that created the tasks ran.

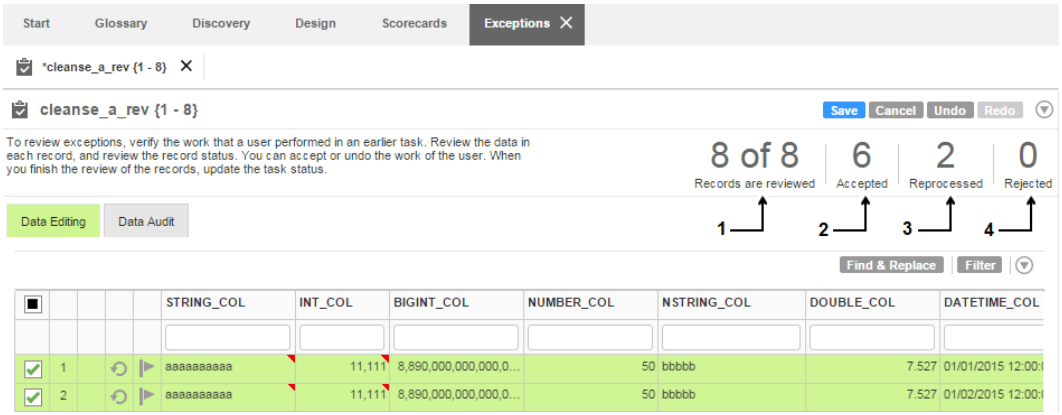
5. Click **Cancel**.

Task Summary Data

You can read summary data for any task that you view or open in the Exceptions workspace. Use the summary data to measure the progress that a user makes in a task.

A duplicate record task displays the number of clusters in the task and the number of clusters that are reviewed. An exception record task displays the number of records in the task and the number of records that include a status indicator.

The following image shows the summary data for an exception record task:



The task contains the following summary data values:

1. The number of records that the user reviewed and the total number of records in the task.
2. The number of records that the user accepted as valid for storage with the business data.
3. The number of records that the user selected for further analysis.
4. The number of records that the user rejected as not valid for storage with the business data.

Reviewing User Progress in a Task

To measure the progress that a user makes in a task, review the summary data.

1. On the Start workspace, open the Task Administration view.
2. Select a task.

   The task opens in the Exceptions workspace.
3. Review the summary data for the task.

Multiple Task Completion

You can identify all of the task instances that originate from a single Human task and complete the task instances in a single operation.

When all of the task instances for a Human task are complete, the records in the tasks move to the next stage in the workflow. The Analyst tool removes the tasks from the users who worked on the tasks. You do not update any record data or status data when you complete the tasks.

You might decide to complete the tasks in the following cases:

- The workflow failed, and you want to run the workflow again.
- The users cannot finish the task instances on time.

Completing Multiple Tasks

To complete all task instances for a Human task, select any task instance from the Human task.

1. Open the Task Administration view.
2. Select a task from the task list.
3. From the task menu, select **Complete Linked Tasks**.

   The **Complete Linked Tasks** dialog box displays the list of the task instances that originate from the same Human task as the task that you select.
4. Optionally, enter a comment.
5. Click **OK**.

   The tasks are complete. The task data moves to the next stage in the workflow.

If you open the Inbox after you complete the tasks, the Inbox might not display any change to the task list. To view the current list of tasks in the Inbox, refresh the Inbox.

Task Data Export

You can export the data from a task instance to a delimited file. Export the data to share the current state of the data with other users.

When you export a task, you export the record data, the status indicator data, and the workflow identifier values. The export operation adds columns for the status indicator data and the workflow identifier values. The export operation excludes audit trail data.

Exception Task Metadata

When you export data from a correct exceptions task or a review exceptions task, you export all of the task data and metadata. The export operation adds the metadata columns to the start of the delimited file. The export operation creates the same columns for a correct exceptions task and for a review exceptions task.

The following table describes the metadata columns that you export with the task data:

Column Name	Description
ROW_IDENTIFIER	The number of the record row in the database table.
REVIEW_STATUS	The current review status of the record. A record can have one of the following review status values: - NULL. The user did not update the review status. - REVIEWED. The user reviewed the record. - REJECTED. The user rejected the record. - ACCEPTED. The user approved the record.
WORKFLOW_ID	An identifier for the workflow that generated the task instance.

Column Name	Description
USER_COMMENT	The most recent note that a user added to the record in the Analyst tool.
UPDATED_STATUS	The current status of the record data. A record can have one of the following status values: - UPDATED. The user updated the record. - ACCEPTED. The user accepted the record as valid. - REJECTED. The user rejected the record as not valid. - REPROCESS. The user indicated that the record needs additional analysis by another user or by another data process. - NULL. The user did not update the record status.
RECORD_STATUS	The record status that the workflow sets. The workflow sets the status value when it writes the record to an exception data table. The default status is INVALID.

Duplicate Task Metadata

When you export data from a correct duplicates task or a review duplicates task, you export all of the task data and metadata. The export operation adds the metadata columns to the start of the delimited file. The export operation creates the same columns for a correct duplicates task and for a review duplicates task.

The following table describes the metadata columns that you export with the task data:

Column Name	Description
ROW_IDENTIFIER	The number of the record row in the database table.
SEQUENTIAL_CLUSTER_ID	An identifier value for the cluster in the database table. The workflow uses the value to sort the cluster rows in the database.
CLUSTER_ID	The value that the identifies the cluster to which the record belongs.
MATCH_SCORE	A value that indicates the degree of similarity between two records in the cluster. The score is a decimal value between 0 and 1.
IS_MASTER	A value that identifies preferred records in the table. The values are Y for preferred records and N for other records.
UPDATED_STATUS	The update status of the record in the cluster. A record can have one of the following status values: - UPDATED. The user updated the record. Users can update the preferred record in a cluster. - NULL. The user did not update the record. - EXTRACTED. The user moved the record from the original cluster.
USER_COMMENT	The most recent note that a user added to the cluster in the Analyst tool.

Column Name	Description
REVIEW_STATUS	The current review status of the cluster. A cluster can have one of the following review status values: - NULL. The user did not update the review status. - REVIEWED. The user reviewed the cluster that contains the record. - REJECTED. The user rejected the cluster that contains the record. - ACCEPTED. The user approved the cluster that contains the record.
WORKFLOW_ID	An identifier for the workflow that generated the task instance.

Exporting Task Data

Export task data to a delimited file.

1. Open the task.
2. From the task actions menu, select **Export Data**.

   The **Export** dialog box appears.
3. Optionally, change the export file name.

   By default, the file name is the task name.
4. Select or clear the option to export the column names as the first row of the export data.
5. Click **Export**.

The Analyst tool exports the data file to the directory structure.

INDEX

A

audit trail
  duplicate record filters 39
  exception record filters 24

B

business administrator 10, 43

C

Clear Cluster Status
  description 35
cluster
  adding a note 38
  creating 38
  editing 34
  finding records 36
  status updating 35
clusters
  filtering cluster data 39

D

Data Audit tab
  duplicate record filters 39
  exception record filters 24
Data Editing panel
  exception data filters 24
  filtering clusters 40
  filtering records 25
Data Editing tab
  duplicate record tasks 32
  filtering cluster data 39
duplicate records
  correct duplicates task 30, 31
  creating a cluster 38
  Discovered Clusters options 36
  editing clusters 34
  review duplicates task 32
  steps to correct 31
  steps to review 31
  updating cluster status 35
duplicate tasks
  export file structure 49

E

example
  process flow for bad records 14
  process flow for duplicate records 15
example *(continued)*
  validation errors in exception tasks 21
Exception Management
  overview 8
exception records
  correct exceptions task 18
  editing exception records 20
  filtering 25
  find and replace data values 26, 27
  review exceptions task 18
  steps to correct 17
  steps to review 17
  updating record status 22
exception tasks
  export file structure 48
  validation errors 21
Exceptions workspace
  correct exceptions task 19
  Data Editing tab 19
  Informatica Analyst 13
export file
  duplicate tasks 49
  exception tasks 48
exporting a task
  description 48

F

filter options
  exception data 23
filtering clusters
  steps 40
filters
  cluster data 39
find and replace data values
  exception records 26, 27

I

Informatica Analyst interface
  log in 16

M

My Tasks view 11, 43

N

notes
  adding to clusters 38

P

preferred record
  changing 34
process flow
  bad records example 14
  duplicate record example 15

R

review task
  review duplicates 41
  steps to review clusters 40
roles
  business administrator 10, 43
  task performer 10, 43

S

Start workspace
  columns 12
status indicators
  duplicate record tasks 35
  exception tasks 21

T

task
  access permissions 10
  correct duplicates 30, 31
task *(continued)*
  correct exceptions 18
  exporting task data 48, 50
  Human task 9
  Mapping task 9
  opening 44
  review duplicates 32
  review exceptions 18
  steps in a Human task 10
  task instances 9
  tasks and workflows 9
  types of task 9
task administration
  assigning a task to a user 46
task administration options 45
task instance
  definition 9
task performer
  My Tasks view 43
tasks
  release a task 45
  viewing 44

V

validation errors in exception tasks 21

W

workflows
  description 9