Informatica® Cloud Data Integration
Summer 2018 October

Tasks
# Table of Contents

**Preface** ....................................................................... 9
- Informatica Resources .................................................. 9
- Informatica Documentation ........................................... 9
- Informatica Intelligent Cloud Services web site. ................. 9
- Informatica Intelligent Cloud Services Communities. ........... 9
- Informatica Intelligent Cloud Services Marketplace ............. 10
- Data Integration connector documentation ....................... 10
- Informatica Knowledge Base ......................................... 10
- Informatica Intelligent Cloud Services trust site. ................. 10
- Informatica Global Customer Support ............................ 10

**Chapter 1: Data Integration tasks.** ................................. 11
- Data filters ................................................................. 11
  - Simple data filters .................................................... 12
  - Advanced data filters .............................................. 12
  - Data filter operators ................................................. 13
  - Data filter variables ................................................. 14
  - Rules and guidelines for data filters ............................ 14
- Field expressions ......................................................... 15
  - Creating a field expression ....................................... 15
  - Transformation language components for expressions ...... 16
  - Expression syntax .................................................... 16
  - String and numeric literals ...................................... 16
  - Rules and guidelines for expressions .......................... 16
  - Adding comments to expressions ............................... 17
  - Reserved words ...................................................... 17
- Parameter files ........................................................... 18
- Schedules ..................................................................... 19
  - Repeat frequency .................................................... 20
  - Time zones and schedules ...................................... 21
  - Daylight Savings Time changes and schedules ............ 21
  - Creating a schedule ................................................ 22
  - Running a task on a schedule ................................... 22
- Email notification .......................................................... 23
- Monitoring a job ........................................................... 23
- Stopping a job ............................................................. 24

**Chapter 2: Mapping tasks** ............................................. 25
- Mapping task templates .............................................. 25
- Advanced connection properties for Visio templates ........... 26
Chapter 3: Synchronization tasks

Task operations. .................................................. 46
Synchronization task sources. .................................. 47
  Rules and guidelines for multiple-object databases. .......... 47
Synchronization task targets. .................................. 47
  Flat file target creation. .................................... 47
  Database target truncation .................................. 47
  Salesforce targets and IDs for related objects. .............. 48
  Update columns. ............................................. 48
Column names in flat files. .................................... 48
Rules and guidelines for synchronization task sources and targets. .......... 49
  Rules and guidelines for flat file sources and targets. .......... 49
  Rules and guidelines for database sources and targets. .......... 49
Field mappings. .................................................. 49
  Field datatypes. ............................................. 50
  Mapplets in field mappings. ................................ 50
  Lookup conditions. ......................................... 50
  Lookup return values. ..................................... 51
  Rules and guidelines for lookups. ......................... 52
Preprocessing and postprocessing commands. ................. 52
  Preprocessing and postprocessing SQL commands. ............ 53
  Preprocessing and postprocessing operating system commands. ...... 53
Configuring a synchronization task. .......................... 53
  Synchronization prerequisite tasks. ......................... 53
  Defining a synchronization task. .......................... 54
# Table of Contents

**Chapter 5: Mass ingestion tasks**                        76
- Mass ingestion task sources.                           76
- Mass ingestion task targets.                          76
- Mass ingestion task configuration.                    77
  - Before you begin.                                    77
  - Defining a mass ingestion task.                      77
  - Configuring the source.                             78
  - Configuring the target.                             79
  - Defining the schedule.                               80
- Viewing mass ingestion task details.                   80
- Running a mass ingestion task.                         80

**Chapter 6: Masking tasks**                             81
- Rules and guidelines for masking tasks.               81
- Masking task options.                                 82
- Source objects.                                       82
  - Schema graph.                                        82
- Target task operations.                               83
  - Inplace masking.                                     84
  - Update partial sandbox.                              84
  - Refresh fields.                                      84
  - Validation reports.                                  84
- Staging database.                                      85
  - Start the staging connection.                        85
  - H2 database configuration requirements.              86
  - Installing and configuring H2 database manually on Windows.  86
  - Installing H2 database manually on Linux.            87
- Data subset.                                          88
- Data subset options.                                   89
- Automatic task recovery.                               89
- Parameter files in data filters.                       90
- Configure relationship behavior.                      91
- Data subset use cases for two objects.                92
  - Case 1. Select the default path with filter on Account. 92
  - Case 2. Select the configured path with filter on Account. 93
  - Case 3. Select the default path with filter on Contact. 95
  - Case 4. Select the configured path with filter on Contact. 96
- Data subset use cases for three objects.              97
  - Case 1. Default path.                                98
  - Case 2. Configured path.                             99
- Data subset rows.                                      99
- Data subset rows example.                              99
Chapter 7: Masking rules

Masking rules.

Repeatable output.

Seed value.

Preprocessing and postprocessing expressions.

Credit card masking.

Credit card parameters.

Email masking.

Advanced email masking.

IP address masking.

Key masking.
Preface

Tasks explains how to set up and run data integration tasks manually or on a schedule with Informatica Cloud Data Integration.

Informatica Resources

Informatica Documentation

To get the latest documentation for your product, browse the Informatica Knowledge Base at https://kb.informatica.com/_layouts/ProductDocumentation/Page/ProductDocumentSearch.aspx.

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

Informatica Intelligent Cloud Services web site

You can access the Informatica Intelligent Cloud Services web site at http://www.informatica.com/cloud. This site contains information about Data Integration editions and applications as well as information about other Informatica Cloud integration services.

Informatica Intelligent Cloud Services Communities

Use the Informatica Intelligent Cloud Services Community to discuss and resolve technical issues. You can also find technical tips, documentation updates, and answers to frequently asked questions.

Access the Informatica Intelligent Cloud Services Community at:


To find resources on using Application Integration (the Informatica Cloud Real Time service), access the community at:

https://network.informatica.com/community/informatica-network/products/cloud-integration/cloud-application-integration/content

Developers can learn more and share tips at the Cloud Developer community:

Informatica Intelligent Cloud Services Marketplace

Visit the Informatica Marketplace to try and buy Data Integration Connectors, templates, and mapplets:

Data Integration connector documentation

You can access documentation for Data Integration Connectors at the Informatica Cloud Community:
https://network.informatica.com/cloud/index.htm

Informatica Knowledge Base

Use the Informatica Knowledge Base to search Informatica Network for product resources such as
documentation, how-to articles, best practices, and PAMs.

To access the Knowledge Base, visit https://kb.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 Intelligent Cloud Services trust site

Subscribe to the Informatica trust site for upgrade, maintenance, and incident notifications.

Status.Informatica.com displays the production status of all the Informatica cloud products. All maintenance
updates are posted to this status page, and during an outage, it will have the most current information. To
ensure you are notified of updates and outages, you can subscribe to a single component, a single incident,
or the site as a whole. Subscribing to the site as a whole is the best way to be certain you never miss an
update. To subscribe, go to http://status.informatica.com and click SUBSCRIBE TO UPDATES. You can then
choose to receive notifications sent as emails, SMS text messages, webhooks, RSS feeds, or any
combination of the four.

Informatica Global Customer Support

You can contact a Customer Support Center by telephone or online.

For online support, click Submit Support Request in Informatica Intelligent Cloud Services. You can also use
Online Support to log a case. Online Support requires a login. You can request a login at

The telephone numbers for Informatica Global Customer Support are available from the Informatica web site
**Data Integration tasks**

A Data Integration task is a process that you configure to analyze, extract, transform, and load data. You can run individual tasks manually or set tasks to run on a schedule.

You can use the following Data Integration tasks to integrate data:

- **Mapping.** Use to process data based on the data flow logic defined in a mapping or Visio template.
- **Synchronization.** Use to load data and integrate applications, databases, and files. Includes add-on functionality such as mapplets.
- **Replication.** Use to replicate data from Salesforce or database sources to database or file targets. You might replicate data to archive the data, perform offline reporting, or consolidate and manage data.
- **Mass ingestion.** Use to transfer a large number of files of any file type between on-premises and cloud repositories and to track and monitor file transfers.
- **Masking.** Use to replace source data in sensitive columns with realistic test data for non-production environments. Masking rules define the logic to replace the sensitive data. Assign masking rules to the columns you need to mask.
- **PowerCenter.** Use to import a PowerCenter workflow and run it as a Data Integration PowerCenter task.

When you create a task, Data Integration walks you through the required steps. The options and properties that display depend on the task type, the options that you select, and the licenses enabled for the organization. For example, for a synchronization task, advanced Salesforce target options display on the **Schedule** page of the task wizard if you select a Salesforce target connection for the task on the **Target** page and your organization has the DSS Advanced Options license.

You can create a workflow of multiple tasks by linking the tasks in taskflows. For more information, see **Taskflows.**

**Data filters**

You can create the following type of data filters for any type of task, except for mass ingestion tasks:

- **Simple**
- **Advanced**

You can create a set of data filters for each object included in a replication task or synchronization task. Each set of data filters act independently of the other sets.
Simple data filters

You can create one or more simple data filters.

When you create multiple simple data filters, the associated task creates an AND operator between the filters and loads rows that apply to all simple data filters.

For example, you load rows from the Account Salesforce object to a database table. However, you want to load only accounts that have greater than or equal to $100,000 in annual revenue and that have more than 500 employees. You configure the following simple data filters:

<table>
<thead>
<tr>
<th>Field</th>
<th>Operator</th>
<th>Field Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>AnnualRevenue</td>
<td>&gt;=</td>
<td>100000</td>
</tr>
<tr>
<td>NumberOfEmployees</td>
<td>&gt;</td>
<td>500</td>
</tr>
</tbody>
</table>

Configuring simple data filters

You configure simple data filters in the task wizard.

1. On the Data Filters page, click Simple, and then click New to create a data filter.
   The Data Filter dialog box appears.
2. Specify the object on which to create the data filter.
   You create separate data filters for each source object included in the task.
3. Enter the filter condition based on the field, operator, and field value.
4. Click OK.
5. Create additional simple data filters as needed.
   To delete a data filter, click the Delete icon next to the data filter.
6. Click Next.

Advanced data filters

Create an advanced data filter to create complex expressions that use AND, OR, or nested conditions.

When you create an advanced data filter, you enter one expression that contains all filters. The expression that you enter becomes the WHERE clause in the query used to retrieve records from the source.

For example, you load rows from the Account Salesforce object to a database table. However, you want to load records where the billing state is California or New York and the annual revenue is greater than or equal to $100,000. You configure the following advanced filter expression:

```
BillingState = 'CA' OR BillingState = 'NY') AND (AnnualRevenue >= 100000)
```

When you create a data filter on a Salesforce object, the corresponding task generates a SOQL query with a WHERE clause. The WHERE clause represents the data filter. The SOQL query must be less than 20,000 characters. If the query exceeds the character limit, the following error appears:

```
Salesforce SOQL limit of 5000 characters has been exceeded for the object: <Salesforce object>. Please exclude more fields or decrease the filters.
```

Note: Filter conditions are not validated until runtime.
Configuring advanced data filters

Configure advanced data filters in the task wizard.

1. To create an advanced data filter, on the Data Filters page, click New > Advanced.
   To convert all simple data filters to one advanced data filter, on the Data Filters page, select a simple data filter and then click Advanced. You cannot convert an advanced data filter back to simple data filters.

2. When you configure a data filter, specify the object on which to create the data filter.
   You create separate data filters for each source object included in the task.

3. Enter the filter expression.
   Click the field name to add the field to the expression.

4. Click OK.
   To delete a data filter, click the Delete icon next to the data filter.

5. Click Next.

Data filter operators

You can use specific operators with each field type.

The following table shows the operators you can use for each field type:

<table>
<thead>
<tr>
<th>Field type</th>
<th>Supported operators</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boolean</td>
<td>=, !=, Is Null, Is Not Null</td>
</tr>
<tr>
<td>Currency</td>
<td>=, !=, &lt;, &lt;=, &gt;, &gt;=, Is Null, Is Not Null</td>
</tr>
<tr>
<td>Date</td>
<td>=, !=, &lt;, &lt;=, &gt;, &gt;=, Is Null, Is Not Null</td>
</tr>
<tr>
<td>Datetime</td>
<td>=, !=, &lt;, &lt;=, &gt;, &gt;=, Is Null, Is Not Null</td>
</tr>
<tr>
<td>Double</td>
<td>=, !=, &lt;, &lt;=, &gt;, &gt;=, Is Null, Is Not Null</td>
</tr>
<tr>
<td>ID</td>
<td>=, !=, Is Null, Is Not Null</td>
</tr>
<tr>
<td>Int</td>
<td>=, !=, &lt;, &lt;=, &gt;, &gt;=, Is Null, Is Not Null</td>
</tr>
<tr>
<td>Reference</td>
<td>=, !=, Is Null, Is Not Null</td>
</tr>
<tr>
<td>String</td>
<td>=, !=, LIKE'%', LIKE'%', LIKE'%<em>]', LIKE'</em>%', Is Null, Is Not Null, &lt;, &lt;=, &gt;, &gt;=</td>
</tr>
<tr>
<td>Textarea</td>
<td>=, !=, LIKE'%', LIKE'%', LIKE'%<em>]', LIKE'</em>%', Is Null, Is Not Null, &lt;, &lt;=, &gt;, &gt;=</td>
</tr>
<tr>
<td>All other field types</td>
<td>=, !=, Is Null, Is Not Null</td>
</tr>
</tbody>
</table>

Data filters 13
Data filter variables

Data filter variables represent the date or time that a task previously ran. Use data filter variables to help capture the source data that changed since the last task run. You can use data filter variables in simple and advanced data filter conditions.

You can use the following table data filter variables:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$LastRunDate</td>
<td>The last date in GMT time zone when the task ran successfully. Does not include time. For example, 2008-09-24. Can be used as a value for filter where the field type is ‘DATE.’</td>
</tr>
<tr>
<td>$LastRunTime</td>
<td>The last date and time in GMT time zone when the task ran successfully. For example, 2008-09-24 15:23:23. Can be used as a value for filter where the field type is ‘DATETIME.’ You cannot use the $LastRunTime variable with DATE fields.</td>
</tr>
</tbody>
</table>

For example, you can include the following simple filter condition:

```
LastModifiedDate > $LastRunTime
```

**Note:** Consider time zone differences when comparing dates across time zones. The date and time of the $LastRunDate and $LastRunTime variables are based on the time zone set in the Data Integration application. The date and time of the actual job is based on the GMT time zone for Salesforce sources and the database server for database sources. The difference in the time zones may yield unexpected results.

Rules and guidelines for data filters

Use the following rules and guidelines for data filters:

- Data filters must contain valid SQL or SOQL operators.
- You cannot include simple and advanced data filters in the same task.
- When you convert a simple data filter to an advanced data filter, you cannot convert the advanced data filter back to a simple data filter.
- A task fails if the fields included in the data filter no longer exist or if the datatypes of the fields change. If a datatype changes, edit the task.
- You can select Equals, Not Equals, Is Null, or Is Not Null operators on fields of the Other datatype.
- Applications do not apply filters with Equals, Starts With, or Ends With operators and string fields that contain data that starts or ends with a single quotation mark. To filter these records, use the Contains operator.
- You can only use IS NULL and LIKE operators in data filters for fields of the Text, Ntext, and Image datatypes.
- If you specify a date and no time for a date/time filter, Data Integration uses 00:00:00 (12:00:00 a.m.) as the time.
- You cannot create a simple data filter in a synchronization task that includes a flat file source. You can create an advanced data filter.
- The list of available operators in a simple data filter depends on the datatype of the field included in the data filter. Some operators do not apply to all fields included in data filters.
- When you enter more than one simple data filter, applications filter rows that meet the requirements of all data filters.
Field expressions

You can transform the source data before loading it into the target. When you configure field mappings, you can specify an expression for each field mapping. You can map multiple source fields to the same target field. For example, you map SourceFieldA and SourceFieldB to TargetFieldC.

Data Integration might suggest operations when you map multiple source fields to a single target field. For example, if you map multiple text fields to a target text field, Data Integration concatenates the source text fields by default. You can change the default expression.

Data Integration provides a transformation language that includes SQL-like functions to transform source data. Use these functions to write expressions, which modify data or test whether data matches the conditions that you specify.

Creating a field expression

Create a field expression in a task wizard.

1. In the Field Mappings page, select the target field for which you want to add an expression.
2. Click Add or Edit Expression.
   - By default, the Field Expression dialog box shows the source field as the expression, which indicates that the target contains the same value as the source.
3. Enter the new field expression.
4. Click Validate Mapping to validate the field mappings.
5. Click Save.

Validating expressions in field mappings

Use the following rules and guidelines when you validate an expression in a field mapping:

- When you validate mappings, Data Integration performs the following validations:
  - Verifies that the source and target fields in the task exist in the source or target. If the field does not exist, an error appears.
  - Verifies that all column datatypes are string and all field expressions contain string operations when the source and target are flat files.
  - Verifies that the correct parameters are used for each function and that the function is valid.
- The expression validator does not perform case-sensitive checks on field names.
- The expression validator verifies that the datatype of a field in an expression matches the datatype expected by the containing function. However, the expression validator does not check for incompatible datatypes between the following sets of objects:
  - Source and target fields of tasks.
  - Source field in a lookup condition and the lookup field.
  - Output of an expression or lookup and the target field.
    - The expression or lookup with these incompatible datatypes may validate successfully, but, at runtime, the task fails and an error appears.
- If you map a string source field to a number target field, the validation succeeds. Data Integration tries to convert the string to a number using the atoi (ASCII to Integer) C function.
- The expression validator does not validate lookups.
Transformation language components for expressions

The transformation language includes the following components to create simple or complex expressions:

- **Fields.** Use the name of a source field to refer to the value of the field.
- **Literals.** Use numeric or string literals to refer to specific values.
- **Functions.** Use these SQL-like functions to change data in a task.
- **Operators.** Use transformation operators to create expressions to perform mathematical computations, combine data, or compare data.
- **Constants.** Use the predefined constants to reference values that remain constant, such as TRUE.

Expression syntax

You can create a simple expression that only contains a field, such as ORDERS, or a numeric literal, such as 10. You can also write complex expressions that include functions nested within functions, or combine different fields using the transformation language operators.

**Note:** Although the transformation language is based on standard SQL, there are differences between the two languages.

String and numeric literals

You can include numeric or string literals.

Enclose string literals within single quotation marks. For example:

'Alice Davis'

String literals are case sensitive and can contain any character except a single quotation mark. For example, the following string is not allowed:

'Joan's car'

To return a string containing a single quotation mark, use the CHR function:

'Joan' || CHR(39) || 's car'

Do not use single quotation marks with numeric literals. Just enter the number you want to include. For example:

.05

or

$$Sales_Tax$

Rules and guidelines for expressions

Use the following rules and guidelines when you write expressions:

- For each source field, you can perform a lookup or create an expression. You cannot do both.
- You cannot use strings in numeric expressions.
  
  For example, the expression 1 + '1' is not valid because you can only perform addition on numeric datatypes. You cannot add an integer and a string.
- You cannot use strings as numeric parameters.
  
  For example, the expression SUBSTR(TEXT_VAL, '1', 10) is not valid because the SUBSTR function requires an integer value, not a string, as the start position.
• You cannot mix datatypes when using comparison operators. For example, the expression 123.4 = '123.4' is not valid because it compares a decimal value with a string.

• You can pass a value from a field, literal string or number, or the results of another expression.

• Separate each argument in a function with a comma.

• Except for literals, the transformation language is not case sensitive.

• The colon (:) , comma (,) , and period (.) have special meaning and should be used only to specify syntax.

• Data Integration tasks treat a dash (-) as a minus operator.

• If you pass a literal value to a function, enclose literal strings within single quotation marks. Do not use quotation marks for literal numbers. Data Integration tasks treat any string value enclosed in single quotation marks as a character string.

• Do not use quotation marks to designate fields.

• You can nest multiple functions within an expression. Data Integration tasks evaluate the expression starting with the innermost function.

Adding comments to expressions

You can use the following comment specifiers to insert comments in expressions:

• Two dashes:
  -- These are comments

• Two forward slashes:
  // These are comments

Data Integration tasks ignore all text on a line preceded by comment specifiers. For example, to concatenate two strings, enter the following expression with comments in the middle of the expression:

-- This expression concatenates first and last names for customers:
FIRST_NAME -- First names from the CUST table
|| // Concat symbol
LAST_NAME // Last names from the CUST table
// Joe Smith Aug 18 1998

Data Integration tasks ignore the comments and evaluates the expression as follows:

FIRST_NAME || LAST_NAME

You cannot continue a comment to a new line:

-- This expression concatenates first and last names for customers:
FIRST_NAME -- First names from the CUST table
|| // Concat symbol
LAST_NAME // Last names from the CUST table
Joe Smith Aug 18 1998

In this case, Data Integration tasks do not validate the expression because the last line is not a valid expression.

Reserved words

Some keywords, such as constants, operators, and system variables, are reserved for specific functions. These include:

• :EXT
• :NFA
Parameter files

A parameter file is a list of user-defined parameters and their associated values.

You can use user-defined parameters in data filters, expressions, and lookup expressions in the following tasks:

- Synchronization tasks

Note: You cannot use a reserved word to name a field. Reserved words have predefined meanings in expressions.
• Mapping tasks

Use a parameter file to define values that you want to update without having to edit the task. For example, you might use a parameter for a sales quota that changes quarterly.

Save a parameter file local to the Secure Agent to run the task. The parameter values are applied when the task runs.

Parameter values are treated as String values. When you use a parameter in an expression, use the appropriate function to convert the value to the necessary datatype. For example, you might use the following expression to define a quarterly bonus for employees:

\[
\text{IF\left(\text{EMP\_SALES} < \text{TO\_INTEGER}\left(\text{\$SalesQuota}\right), 200, 0\right)}
\]

When you use a parameter in a data filter, start the data filter with the parameter. For example, use \(\text{\$Sales < 100000} \text{ instead of 100000 > \$Sales}\).

To use a parameter file, perform the following steps:

1. Use the parameter in the task. You can use parameters in expressions and filters.
   - Use two dollar signs to name the parameter, as follows: \(\text{\$\langle\text{parameter}\rangle}\).
   - When you use the parameter in a filter, start the filter with the parameter.
   - When you use the parameter in an expression, convert the String parameter value to the appropriate datatype as necessary.

2. In the parameter file, define parameters as follows:

   \[
   \text{\$<\text{parameter}>=}\text{\text{value}} \\
   \text{\$<\text{parameter2}>=}\text{\text{value2}}
   \]

   For example:

   \[
   \text{\$SalesQuota=}1000 \\
   \text{\$Region=}\text{NW}
   \]

   **Note**: The parameter value includes any characters after the equals sign (=), including leading or trailing spaces. Parameter names are case-sensitive.

3. Save the parameter file to a directory local to the Secure Agent to run the task.
   Use the following directory:

   \(<\text{Secure Agent installation directory}>/\text{apps/Data\_Integration\_Server/data/userparameters}\)

4. Enter the parameter file name on the Schedule and Advanced Options page of the task wizard.

**Schedules**

You can run tasks manually or you can use schedules to run them at a specific time or interval such as hourly, daily, or weekly.

To use a schedule, you associate the task with a schedule when you configure the task. You can use an existing schedule or create a new schedule. If you want to create a schedule, you can create the schedule from the task's **Schedule** page during task configuration. You cannot create the schedule during task configuration in mass ingestion tasks.

When you create a schedule, you specify the date and time. You can configure a schedule to run associated assets throughout the day between 12:00 a.m. and 11:55 p.m. Informatica Intelligent Cloud Services might add a small schedule offset to the start time, end time, and all other time configurations. As a result, scheduled tasks and taskflows might start later than expected. For example, you configure a schedule to run hourly until noon, and the schedule offset for your organization is 10 seconds. Informatica Intelligent Cloud
Services extends the end time for the schedule to 12:00:10 p.m., and the last hourly task or taskflow starts at 12:00:10 p.m. To see the schedule offset for your organization, check the Schedule Offset organization property.

You can monitor scheduled tasks from the All Jobs page in Monitor. Scheduled tasks do not appear on the My Jobs page.

When you copy a task that includes a schedule, the schedule is not associated with the new task. To associate a schedule with the new task, edit the task.

If you remove a task from a schedule as the task runs, the job completes. Data Integration cancels any additional runs associated with the schedule.

Repeat frequency

The repeat frequency determines how often tasks run. The following table describes the repeat frequency options:

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Does not</td>
<td>Tasks run as scheduled and do not repeat.</td>
</tr>
<tr>
<td>repeat</td>
<td></td>
</tr>
<tr>
<td>Every N</td>
<td>Tasks run on an interval based on a specified number of minutes. You can</td>
</tr>
<tr>
<td>minutes</td>
<td>configure the following options:</td>
</tr>
<tr>
<td></td>
<td>- Repeat frequency. Select a frequency in minutes. Options are 5, 10, 15,</td>
</tr>
<tr>
<td></td>
<td>20, 30, 45.</td>
</tr>
<tr>
<td></td>
<td>- Days. Days of the week when you want tasks to run. You can select one or</td>
</tr>
<tr>
<td></td>
<td>more days of the week.</td>
</tr>
<tr>
<td></td>
<td>- Time range. Hours of the day when you want tasks to start. Select All</td>
</tr>
<tr>
<td></td>
<td>Day or configure a time range.</td>
</tr>
<tr>
<td></td>
<td>You can configure a time range between 00:00-23:55.</td>
</tr>
<tr>
<td></td>
<td>- Repeat option. The range of days when you want tasks to run. You can</td>
</tr>
<tr>
<td></td>
<td>select Repeat Indefinitely or configure an end date and time.</td>
</tr>
<tr>
<td>Hourly</td>
<td>Tasks run on an hourly interval based on the start time of the schedule.</td>
</tr>
<tr>
<td></td>
<td>You can configure the following options:</td>
</tr>
<tr>
<td></td>
<td>- Repeat frequency. Select a frequency in hours. Options are 1, 2, 3, 4,</td>
</tr>
<tr>
<td></td>
<td>6, 8, 12.</td>
</tr>
<tr>
<td></td>
<td>- Days. Days of the week when you want tasks to run. You can select one or</td>
</tr>
<tr>
<td></td>
<td>more days of the week.</td>
</tr>
<tr>
<td></td>
<td>- Time range. Hours of the day when you want tasks to start. Select All</td>
</tr>
<tr>
<td></td>
<td>Day or configure a time range.</td>
</tr>
<tr>
<td></td>
<td>You can configure a time range between 00:00-23:55.</td>
</tr>
<tr>
<td></td>
<td>- Repeat option. The range of days when you want tasks to run. You can</td>
</tr>
<tr>
<td></td>
<td>select Repeat Indefinitely or configure an end date and time.</td>
</tr>
<tr>
<td>Daily</td>
<td>Tasks run daily at the start time configured for the schedule.</td>
</tr>
<tr>
<td></td>
<td>You can configure the following options:</td>
</tr>
<tr>
<td></td>
<td>- Repeat frequency. The frequency at which you want tasks to run. Select</td>
</tr>
<tr>
<td></td>
<td>Every Day or Every Weekday.</td>
</tr>
<tr>
<td></td>
<td>- Repeat option. The range of days when you want tasks to run. You can</td>
</tr>
<tr>
<td></td>
<td>select Repeat Indefinitely or configure an end date and time.</td>
</tr>
</tbody>
</table>
Option | Description
--- | ---
Weekly | Tasks run on a weekly interval based on the start time of the schedule. You can configure the following options:
- Days. Days of the week when you want tasks to run. You can select one or more days of the week.
- Repeat option. The range of days when you want tasks to run. You can select Repeat Indefinitely or configure an end date and time.
If you do not specify a day, the schedule runs regularly on the same day of the week as the start date.

Monthly | Tasks run on a monthly interval based on the start time of the schedule. You can configure the following options:
- Day. Day of the month when you want tasks to run. You can configure one of the following options:
  - Select the exact date of the month, between 1-28. If you want the task to run on days later in the month, use the <n> <day of the week> option.
  - Select the <n> <day of the week>. Options for <n> include First, Second, Third, Fourth, and Last. Options for <day of the week> includes Day, and Sunday-Saturday.
  Tip: With the Day option, you can configure tasks to run on the First Day or the Last Day of the month.
- Repeat option. The range of days when you want tasks to run. You can select Repeat Indefinitely or configure an end date and time.

---

Time zones and schedules

Informatica Intelligent Cloud Services stores time in Coordinated Universal Time (UTC). When you log in, Informatica Intelligent Cloud Services converts the time and displays it in the time zone associated with your user profile.

When you create a schedule, you select the time zone for the scheduler to use. You can select a time zone that is different from your time zone or your organization time zone.

Daylight Savings Time changes and schedules

Informatica Intelligent Cloud Services applies Daylight Savings Time changes to tasks that run daily or weekly. It does not support Daylight Savings Time changes for schedules that run at other frequencies, such as hourly or monthly.

Daylight Savings Time does not trigger additional runs for tasks that are scheduled to run between 1:00 a.m. - 1:59 a.m. and run everyday or every week. For example, a task is scheduled to run everyday at 1:30 a.m. When the clock changes from 2 a.m. to 1 a.m. during Daylight Savings Time, the task does not run again at 1:30 a.m.

On the other hand, if a task is scheduled to run everyday or every week between 2:00 a.m. and 2:59 a.m., it does not run the day that the clock changes forward from 2:00 a.m. to 3:00 a.m.

**Tip:** To ensure that Informatica Intelligent Cloud Services does not skip any scheduled runs near the 2 a.m. time change, do not schedule jobs to run between 12:59 a.m. and 3:01 a.m.
Creating a schedule

You can create a schedule in Data Integration when you configure a task or linear taskflow. You can also create a schedule in Administrator if you have the appropriate permissions.

The following procedure describes how to create a schedule when you access the Schedule page from Data Integration during task or linear taskflow configuration.

1. Select Run this task on a schedule, and then click New.
2. Configure the following properties:

<table>
<thead>
<tr>
<th>Property</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schedule Name</td>
<td>Name of the schedule. Each schedule name must be unique within the organization. Schedule names can contain alphanumeric characters, spaces, and the following special characters: _ . + - Maximum length is 100 characters. Schedule names are not case sensitive.</td>
</tr>
<tr>
<td>Description</td>
<td>Description of the schedule. Maximum length is 255 characters.</td>
</tr>
<tr>
<td>Starts</td>
<td>Date and time when the schedule starts. The date format is MM/DD/YYYY. Time appears in the 24-hour format. Click the calendar button to select the start date. The start date and time can affect the repeat frequency for tasks and taskflow jobs that repeat at regular intervals. For example, if the start date is November 10 and the repeat frequency is monthly, the schedule runs associated assets on the tenth day of each month. If the start time is 3:10 and the repeat frequency is hourly, the assets run every hour at 10 minutes past the hour. Default is the current date, current time, and time zone of the user that creates the schedule.</td>
</tr>
<tr>
<td>Time Zone</td>
<td>Select the time zone for the schedule to use. The time zone can differ from the organization time zone or user time zone.</td>
</tr>
<tr>
<td>Repeats</td>
<td>Repeat frequency for the schedule. Select one of the following options: Does Not Repeat, Every N Minutes, Hourly, Daily, Weekly, Monthly. Default is Does Not Repeat.</td>
</tr>
</tbody>
</table>

3. Click Save to save the schedule and return to the task configuration page.

Running a task on a schedule

Associate a task with a schedule on the Schedule page when you configure the task. You can use an existing schedule or create a schedule.

1. On the Schedule page for the task, select Run this task on a schedule.
2. To specify whether to use an existing schedule or a new schedule, perform one of the following tasks:
   - To use an existing schedule, select the schedule that you want to use.
To create a schedule to use for the task, click **New**, and then configure the schedule properties. For more information on creating a schedule, see the Administrator help.

3. Click **Save**.

### Email notification

You can configure email notification for a task. When you configure custom email notification, Data Integration uses the custom email notification instead of the email notification options configured for the organization. You cannot configure email notifications for mass ingestion tasks.

To configure email notification options, perform the following steps in the task wizard:

1. Specify whether to use the default email notification options that have been set for your organization or create custom email notification for the task. Configure email notification using the following options:

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Use Default Email Notification Options for my Organization</td>
<td>Use the email notification options configured for the organization.</td>
</tr>
<tr>
<td>Use Custom Email Notification Options for this Task</td>
<td>Use the email notification options configured for the task. You can send email to different addresses based on whether the task failed, completed with errors, or completed successfully. Use commas to separate a list of email addresses. When you select this option, email notification options configured for the organization are not used.</td>
</tr>
</tbody>
</table>

2. Click **Save**.

### Monitoring a job

You can monitor tasks or taskflows that are currently running, have completed, or have stopped.

Monitor jobs on the following pages:

- Monitor the jobs that you initiated on the **My Jobs** page in Data Integration.
- Monitor running jobs in your organization on the **Running Jobs** page in Monitor.
- Monitor all jobs in your organization on the **All Jobs** page in Monitor.

For more information about monitoring jobs, see **Monitor**.
Stopping a job

A job is an instance of a mapping, task, or taskflow. You can stop a running job on the All Jobs, Running Jobs, or My Jobs page.

1. Open Monitor and select All Jobs or Running Jobs, or open Data Integration and select My Jobs.
2. In the row that contains the job that you want to stop, click the Stop icon.

To view details about the stopped job, click the job name.

Note: You cannot stop a subtask of a linear taskflow that is running.
Use the mapping task to process data based on the data flow logic defined in a mapping or Visio template.

When you create a mapping task, you select the mapping or Visio template for the task to use. The mapping or Visio template must already exist before you can create a mapping task for it. Alternatively, you can create a mapping task by using a template.

A Visio template includes template parameters for the source and target connections. A Visio template can also include other template parameters, such as filter conditions or lookup connections.

If the mapping includes parameters, you can define the parameters when you configure the task or define the parameters when you run the task. You can use user-defined parameters for data filters, expressions, and lookup expressions in a mapping task. You define user-defined parameters in a parameter file associated with the task.

At run time, a mapping task processes task data based on the data flow logic from the mapping or Visio template, the parameters defined in the task, and the user-defined parameters defined in a parameter file, when available.

Mapping task templates

Mapping task templates provide the ability to run a mapping task without creating a mapping beforehand. Each mapping task template is based upon a mapping template. Use a mapping task template when the mapping on which the mapping task template is based suits your needs. When you select a mapping task template, Data Integration creates a copy of the template for you to use. When you define the mapping task in the task wizard, you save a copy of the mapping template on which the mapping task template is based.
Templates are divided into three categories: Integration, Cleansing, and Warehousing, as shown in the following image:

The templates range from simple templates that you can use to copy data from one source to another, to complex templates that you can use for data warehousing-related tasks.

**Advanced connection properties for Visio templates**

For tasks based on Visio templates, you can configure advanced properties for Informatica Intelligent Cloud Services Connector connections. For tasks based on mappings, you define advanced connection properties in the mapping.

Connections for Informatica Intelligent Cloud Services Connectors can display advanced properties, such as page size, flush interval, or row limit. The advanced properties display based on the connection type and how the connection is used. Some Informatica Intelligent Cloud Services Connectors might not be configured to display advanced properties.

**Related objects**

When a mapping or Visio template includes a source that is a parameter and is configured for multiple objects, you can join related objects in the task.

You can join related objects based on existing relationships or custom relationships. Data Integration restricts the type of relationships that you can create based on the connection type.

Use the following relationships to join related objects:
Existing relationships

You can use relationships defined in the source system to join related objects. You can join objects with existing relationships for Salesforce, database, and some Data Integration Connectors connection types. After you select a primary object, you select a related object from a list of related objects.

Custom relationships

You can use custom relationships to join multiple source objects. You can create custom relationships for the database connection type.

When you create a custom relationship for database objects, you create an inner join on the source fields that you select.

To join source objects, you add the primary source object in the Objects and Relationships table. Then you add related objects and specify keys for the primary object and related objects. For more information about related source objects, see the Source Transformation section in Transformations.

Advanced relationships

You can create an advanced relationship for database sources when the source object in the mapping is a parameter and configured for multiple sources. You cannot create an advanced relationship between source objects that have been joined using a custom relationship.

When you create an advanced relationship, the wizard converts any relationships that you defined to an SQL statement that you can edit.

To create an advanced relationship, you add the primary source object in the Objects and Relationships table. Then you select fields and write the SQL statement that you want to use. Use an SQL statement that is valid for the source database. You can also add additional objects from the source.
Advanced session properties

Advanced session properties are optional properties that you can configure in mapping tasks and Visio templates. Use caution when you configure advanced session properties. The properties are based on PowerCenter advanced session properties and might not be appropriate for use with all tasks.

You can configure the following types of advanced session properties:

- General
- Performance
- Advanced
- Error handling

General options

The following table describes the general options:

<table>
<thead>
<tr>
<th>General options</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Write Backward Compatibl...</td>
<td>Writes the session log to a file.</td>
</tr>
</tbody>
</table>
| Session Log File Name                         | Name for the session log. Use any valid file name. You can use the following variables as part of the session log name:  
|                                              | - $CurrentTaskName. Replaced with the task name.  
|                                              | - $CurrentTime. Replaced with the current time.                                                                                                                                                           |
| Session Log File Directory                   | Directory where the session log is saved. Use a directory local to the Secure Agent to run the task. By default, the session log is saved to the following directory:  
|                                              | <Secure Agent installation directory>/apps/Data_Integration_Server/logs                                                                                                                                 |
| $Source Connection Value                     | Source connection name for Visio templates.                                                                                                                                                                 |
| $Target Connection Value                     | Target connection name for Visio templates.                                                                                                                                                                 |
| Treat Source Rows as                         | When the task reads source data, it marks each row with an indicator that specifies the target operation to perform when the row reaches the target. Use one of the following options:  
|                                              | - Insert. All rows are marked for insert into the target.  
|                                              | - Update. All rows are marked for update in the target.  
|                                              | - Delete. All rows are marked for delete from the target.  
|                                              | - Data Driven. The task uses the Update Strategy object in the data flow to mark the operation for each source row.                                                                                                                                                      |
| Commit Type                                  | Commit type to use. Use one of the following options:  
|                                              | - Source. The task performs commits based on the number of source rows.  
|                                              | - Target. The task performs commits based on the number of target rows.  
|                                              | - User Defined. The task performs commits based on the commit logic defined in the Visio template.  
|                                              | When you do not configure a commit type, the task performs a target commit.                                                                                                                                 |
| Commit Interval                               | Interval in rows between commits. When you do not configure a commit interval, the task commits every 10,000 rows.                                                                                           |
### General options

<table>
<thead>
<tr>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Commits data at the end of the file.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rolls back the transaction at the next commit point when the task encounters a non-fatal error. When the task encounters a transformation error, it rolls back the transaction if the error occurs after the effective transaction generator for the target.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Java classpath to use. The Java classpath is added to the beginning of the system classpath when the task runs. Use this option when you use third-party Java packages, built-in Java packages, or custom Java packages in a Java transformation.</td>
</tr>
</tbody>
</table>

### Performance settings

The following table describes the performance settings:

<table>
<thead>
<tr>
<th>Performance settings</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>DTM Buffer Size</td>
<td>Amount of memory allocated to the task from the DTM process. By default, a minimum of 12 MB is allocated to the buffer at run time. Use one of the following options: - Auto. Enter Auto to use automatic memory settings. When you use Auto, configure <strong>Maximum Memory Allowed for Auto Memory Attributes</strong>. - A numeric value. Enter the numeric value that you want to use. The default unit of measure is bytes. Append KB, MB, or GB to the value to specify a different unit of measure. For example, 512MB. You might increase the DTM buffer size in the following circumstances: - When a task contains large amounts of character data, increase the DTM buffer size to 24 MB. - When a task contains n partitions, increase the DTM buffer size to at least n times the value for the task with one partition. - When a source contains a large binary object with a precision larger than the allocated DTM buffer size, increase the DTM buffer size so that the task does not fail.</td>
</tr>
</tbody>
</table>

<p>| Incremental Aggregation | Performs incremental aggregation for tasks based on Visio templates. |
| Reinitialize Aggregate Cache | Overwrites existing aggregate files for a task that performs incremental aggregation. |
| Enable High Precision | Processes the Decimal datatype to a precision of 28. |
| Session Retry on Deadlock | The task retries a write on the target when a deadlock occurs. |</p>
<table>
<thead>
<tr>
<th>Performance settings</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pushdown Optimization</td>
<td>Type of pushdown optimization. Use one of the following options:</td>
</tr>
<tr>
<td></td>
<td>- None. The task processes all transformation logic for the task.</td>
</tr>
<tr>
<td></td>
<td>- To Source. The task pushes as much of the transformation logic to the source database as possible.</td>
</tr>
<tr>
<td></td>
<td>- To Target. The task pushes as much of the transformation logic to the target database as possible.</td>
</tr>
<tr>
<td></td>
<td>- Full. The task pushes as much of the transformation logic to the source and target databases as possible. The task processes any transformation logic that it cannot push to a database.</td>
</tr>
<tr>
<td></td>
<td>- $$PushdownConfig. The task uses the pushdown optimization type specified in the user-defined parameter file for the task.</td>
</tr>
<tr>
<td></td>
<td>When you use $$PushdownConfig, ensure that the user-defined parameter is configured in the parameter file.</td>
</tr>
<tr>
<td></td>
<td>When you use pushdown optimization, do not use the Error Log Type property.</td>
</tr>
<tr>
<td></td>
<td>For more information, see “Pushdown optimization” on page 33.</td>
</tr>
<tr>
<td>Enable cross-schema pushdown optimization</td>
<td>Enables pushdown optimization for tasks that use source or target objects associated with different schemas within the same database.</td>
</tr>
<tr>
<td></td>
<td>To see if the connector you use supports cross-schema pushdown optimization, see the relevant connector guide.</td>
</tr>
<tr>
<td></td>
<td>This property is enabled by default.</td>
</tr>
<tr>
<td>Allow Temporary View for Pushdown</td>
<td>Allows the task to create temporary view objects in the database when it pushes the task to the database.</td>
</tr>
<tr>
<td></td>
<td>Use when the task includes an SQL override in the Source Qualifier transformation or Lookup transformation. You can also use for a task based on a Visio template that includes a lookup with a lookup source filter.</td>
</tr>
<tr>
<td>Allow Temporary Sequence for Pushdown</td>
<td>Allows the task to create temporary sequence objects in the database.</td>
</tr>
<tr>
<td></td>
<td>Use when the task is based on a Visio template that includes a Sequence Generator transformation.</td>
</tr>
<tr>
<td>Allow Pushdown for User Incompatible</td>
<td>Indicates that the database user of the active database has read permission on idle databases.</td>
</tr>
<tr>
<td>Connections</td>
<td>If you indicate that the database user of the active database has read permission on idle databases, and it does not, the task fails.</td>
</tr>
<tr>
<td></td>
<td>If you do not indicate that the database user of the active database has read permission on idle databases, the task does not push transformation logic to the idle databases.</td>
</tr>
<tr>
<td>Session Sort Order</td>
<td>Order to use to sort character data for the task.</td>
</tr>
</tbody>
</table>
### Advanced options

The following table describes the advanced options:

<table>
<thead>
<tr>
<th>Advanced options</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Constraint Based Load Ordering</strong></td>
<td>Loads targets based on primary key-foreign key constraints when possible.</td>
</tr>
<tr>
<td><strong>Cache Lookup() Function</strong></td>
<td>Caches lookup functions in Visio templates with unconnected lookups. Overrides lookup configuration in the template. By default, the task performs lookups on a row-by-row basis, unless otherwise specified in the template.</td>
</tr>
</tbody>
</table>
| **Default Buffer Block Size**                          | Size of buffer blocks used to move data and index caches from sources to targets. By default, the task determines this value at run time. Use one of the following options:  
  - Auto. Enter Auto to use automatic memory settings. When you use Auto, configure Maximum Memory Allowed for Auto Memory Attributes.  
  - A numeric value. Enter the numeric value that you want to use. The default unit of measure is bytes. Append KB, MB, or GB to the value to specify a different unit of measure. For example, 512MB.  

  The task must have enough buffer blocks to initialize. The minimum number of buffer blocks must be greater than the total number of Source Qualifiers, Normalizers for COBOL sources, and targets. The number of buffer blocks in a task = DTM Buffer Size / Buffer Block Size. Default settings create enough buffer blocks for 83 sources and targets. If the task contains more than 83, you might need to increase DTM Buffer Size or decrease Default Buffer Block Size. |
| **Line Sequential Buffer Length**                      | Number of bytes that the task reads for each line. Increase this setting from the default of 1024 bytes if source flat file records are larger than 1024 bytes.                                                                                                                                                                                                                                                                                                                                 |
| **Maximum Memory Allowed for Auto Memory Attributes**  | Maximum memory allocated for automatic cache when you configure the task to determine the cache size at run time. You enable automatic memory settings by configuring a value for this attribute. Enter a numeric value. The default unit is bytes. Append KB, MB, or GB to the value to specify a different unit of measure. For example, 512MB. If the value is set to zero, the task uses default values for memory attributes that you set to auto. |
| **Maximum Percentage of Total Memory Allowed for Auto Memory Attributes** | Maximum percentage of memory allocated for automatic cache when you configure the task to determine the cache size at run time. If the value is set to zero, the task uses default values for memory attributes that you set to auto.                                                                                                                                                                                                                                                                                     |
| **Additional Concurrent Pipelines for Lookup Cache Creation** | Restricts the number of pipelines that the task can create concurrently to pre-build lookup caches. You can configure this property when the Pre-build Lookup Cache property is enabled for a task or transformation.  
  When the Pre-build Lookup Cache property is enabled, the task creates a lookup cache before the Lookup receives the data. If the task has multiple Lookups, the task creates an additional pipeline for each lookup cache that it builds. To configure the number of pipelines that the task can create concurrently, select one of the following options:  
  - Auto. The task determines the number of pipelines it can create at run time.  
  - Numeric value. The task can create the specified number of pipelines to create lookup caches. |
<table>
<thead>
<tr>
<th>Advanced options</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Custom Properties</td>
<td>Configure custom properties for the task. You can override the custom properties that the task uses after the job has started. The task also writes the override value of the property to the session log.</td>
</tr>
</tbody>
</table>
| Pre-build Lookup Cache           | Allows the task to build the lookup cache before the Lookup receives the data. The task can build multiple lookup cache files at the same time to improve performance. You can configure this option in a Visio template or in a task. The task uses the task-level setting if you configure the Lookup option as Auto for a Visio template. Configure one of the following options:  
  - Always allowed. The task can build the lookup cache before the Lookup receives the first source row. The task creates an additional pipeline to build the cache.  
  - Always disallowed. The task cannot build the lookup cache before the Lookup receives the first row.  
  When you use this option, configure the **Configure the Additional Concurrent Pipelines for Lookup Cache Creation** property. The task can pre-build the lookup cache if this property is greater than zero. |
| DateTime Format String           | Date time format for the task. You can specify seconds, milliseconds, or nanoseconds.  
  To specify seconds, enter **MM/DD/YYYY HH24:MI:SS**.  
  To specify milliseconds, enter **MM/DD/YYYY HH24:MI:SS.MS**.  
  To specify microseconds, enter **MM/DD/YYYY HH24:MI:SS.US**.  
  To specify nanoseconds, enter **MM/DD/YYYY HH24:MI:SS.NS**.  
  By default, the format specifies microseconds, as follows: **MM/DD/YYYY HH24:MI:SS.US**. |
| Pre 85 Timestamp Compatibility   | Do not use with Data Integration. |

**Error handling**

The following table describes the error handling options:

<table>
<thead>
<tr>
<th>Error handling options</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stop on Errors</td>
<td>Indicates how many non-fatal errors the task can encounter before it stops the session. Non-fatal errors include reader, writer, and DTM errors. Enter the number of non-fatal errors you want to allow before stopping the session. The task maintains an independent error count for each source, target, and transformation. If you specify 0, non-fatal errors do not cause the session to stop.</td>
</tr>
<tr>
<td>Override Tracing</td>
<td>Overrides tracing levels set on an object level.</td>
</tr>
</tbody>
</table>
| On Stored Procedure Error| Determines the behavior when a task based on a Visio template encounters pre-session or post-session stored procedure errors. Use one of the following options:  
  - Stop Session. The task stops when errors occur while executing a pre-session or post-session stored procedure.  
  - Continue Session. The task continues regardless of errors.  
  By default, the task stops. |
<table>
<thead>
<tr>
<th>Error handling options</th>
<th>Description</th>
</tr>
</thead>
</table>
| On Pre-Session Command Task Error                           | Determines the behavior when a task that includes pre-session shell commands encounters errors. Use one of the following options:  
- Stop Session. The task stops when errors occur while executing pre-session shell commands. 
- Continue Session. The task continues regardless of errors. 
By default, the task stops. |
| On Pre-Post SQL Error                                       | Determines the behavior when a task that includes pre-session or post-session SQL encounters errors:  
- Stop Session. The task stops when errors occur while executing pre-session or post-session SQL. 
- Continue. The task continues regardless of errors. 
By default, the task stops. |
| Error Log Type                                              | Specifies the type of error log to create. You can specify flat file or no log. Default is none. You cannot log row errors from XML file sources. You can view the XML source errors in the session log. 
Do not use this property when you use the Pushdown Optimization property. |
| Error Log File Directory                                   | Specifies the directory where errors are logged. By default, the error log file directory is $PMBadFilesDir. |
| Error Log File Name                                         | Specifies error log file name. By default, the error log file name is PMError.log. |
| Log Row Data                                                | Specifies whether or not to log transformation row data. When you enable error logging, the task logs transformation row data by default. If you disable this property, n/a or -1 appears in transformation row data fields. |
| Log Source Row Data                                         | Specifies whether or not to log source row data. By default, the check box is clear and source row data is not logged. |
| Data Column Delimiter                                       | Delimiter for string type source row data and transformation group row data. By default, the task uses a pipe ( | ) delimiter. 
**Tip:** Verify that you do not use the same delimiter for the row data as the error logging columns. If you use the same delimiter, you may find it difficult to read the error log file. |

**Pushdown optimization**

You can use pushdown optimization to push transformation logic to source databases or target databases for execution. Use pushdown optimization when using database resources can improve task performance.

When you run a task configured for pushdown optimization, the task converts the transformation logic to an SQL query. The task sends the query to the database, and the database executes the query.

The amount of transformation logic that you can push to the database depends on the database, transformation logic, and task configuration. The task processes all transformation logic that it cannot push to a database.

Use the **Pushdown Optimization** advanced session property to configure pushdown optimization for a task.
Pushdown optimization types

You can use the following pushdown optimization types:

**Source pushdown optimization**

The task analyzes the mapping from source to target until it reaches transformation logic that it cannot push to the source database.

The task generates and executes a Select statement based on the transformation logic for each transformation that it can push to the database. Then, the task reads the results of the SQL query and processes the remaining transformations.

**Target pushdown optimization**

The task analyzes the mapping from target to source or until it reaches transformation logic that it cannot push to the target database.

The task generates an Insert, Delete, or Update statement based on the transformation logic for each transformation that it can push to the target database. The task processes the transformation logic up to the point where it can push the transformation logic to the database. Then, the task executes the generated SQL on the target database.

**Full pushdown optimization**

The task analyzes the mapping from source to target or until it reaches transformation logic that it cannot push to the target database.

The task generates and executes SQL statements against the source or target based on the transformation logic that it can push to the database.

You can use full pushdown optimization when the source and target databases are in the same relational database management system.

When you run a task with large quantities of data and full pushdown optimization, the database server must run a long transaction. Consider the following database performance issues when you generate a long transaction:

- A long transaction uses more database resources.
- A long transaction locks the database for longer periods of time, which can reduce database concurrency and increase the likelihood of deadlock.
- A long transaction increases the likelihood of an unexpected event.

To minimize database performance issues for long transactions, consider using source or target pushdown optimization.

**Cross-schema pushdown optimization**

You can enable cross-schema pushdown optimization for tasks that use source or target objects associated with different schemas within the same database.

To use cross-schema pushdown optimization, you create a connection for each schema. The database and the database user name and password must be the same for both connections.

Cross-schema pushdown optimization is enabled by default.

To see if the connector you use supports cross-schema pushdown optimization and find instructions that are specific to your connector, see the relevant connector guide.
Pushdown optimization user-defined parameters

You can use a pushdown optimization user-defined parameter to perform pushdown optimization based on the parameter value defined in a parameter file. Use a pushdown optimization user-defined parameter when you want to perform different pushdown optimization options at different times.

For example, you might use source or target pushdown optimization during the peak hours of the day, but use full pushdown optimization from midnight until 2 a.m. when database activity is low.

To use the pushdown optimization user-defined parameter, perform the following steps:

1. Configure a parameter file to use the $$PushdownConfig user-defined parameter. Save the file to a directory local the Secure Agent to run the task.
   
   Use the following format to define the parameter:
   
   $$PushdownConfig=<pushdown optimization type>
   
   For example: $$PushdownConfig=Source.
   
   Configure a different parameter file of the same name for each pushdown type that you want to use.

2. In the task, add the Pushdown Optimization property and select the $$PushdownConfig option.

3. Configure the task to use the parameter file.

4. Replace the parameter file version as needed.

Field metadata

You can view and edit field metadata such as the type, precision, and scale for parameterized flat file source objects.

To view and edit field metadata, use the Edit Types option on the Sources page of the mapping task wizard.

Sequence Generator values

When you run a mapping task that includes a Sequence Generator transformation in the mapping, you can change the beginning value for the sequence.

To change the beginning value, you change the Current Value field in the Sequences page in the mapping task wizard. The Current Value field shows the first value the task will generate in the sequence, based on the last value generated in the last task execution.

For example, the last time you ran the CustDataIDs task, the last value generated was 124. The next time the task runs, the first number in the sequence is 125 because the Sequence Generator transformation is
configured to increment by 1. If you want the sequence to begin with 200, you change the Current Value to 200.

Mapping task configuration

Use the mapping task wizard to create a mapping task.

Complete the following steps in the wizard to create a mapping task:
1. Create the mapping task.
2. Configure the source.
3. Configure the target.
4. Define other parameters.
5. Optionally, configure a schedule and advanced options.

As you work through the task wizard, you can click Save to save your work at any time. When you have completed the wizard, you can click Finish to save and close the task wizard.

Defining a mapping task

1. To create a mapping task, click New > Tasks and then complete one of the following steps:
   - To create a mapping task based on a mapping, select Mapping Task and click Create.
   - To create a mapping task using a template, expand the appropriate template category and select the template you want to use, and then click Create.

To edit a mapping task, on the Explore page, navigate to the mapping task. In the row that contains the task, click Actions and select Edit.
2. Configure the following fields:

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task Name</td>
<td>Name of the task. Task names can contain alphanumeric characters, spaces, and the following special characters: _ . + -. Maximum length is 100 characters. Task names are not case sensitive.</td>
</tr>
<tr>
<td>Location</td>
<td>Project folder in which the task resides. If the Explore page is currently active and a project or folder is selected, the default location for the asset is the selected project or folder. Otherwise, the default location is the location of the most recently saved asset.</td>
</tr>
<tr>
<td>Description</td>
<td>Description of the task. Maximum length is 4000 characters.</td>
</tr>
<tr>
<td>Runtime Environment</td>
<td>Runtime environment that contains the Secure Agent to run the task.</td>
</tr>
<tr>
<td>Task Based On</td>
<td>The basis for the task. Select Mapping or Visio Template.</td>
</tr>
<tr>
<td>Mapping</td>
<td>Mapping associated with the task. If you are creating a mapping task based on a mapping, perform the following steps: 1. Click Select and navigate to the mapping that you want to use. 2. Select the mapping and click Select. If you are using a mapping task template, the name of the mapping template on which the mapping task template is based displays in the Mapping field. By default, the mapping is saved in the same folder as the mapping task. Optionally, perform the following steps: 1. In the Mapping field, change the name of the mapping. 2. To save the mapping in a different folder, click Browse.</td>
</tr>
<tr>
<td>Visio Template</td>
<td>The Visio template associated with the task. You can select any Visio template imported to the organization. To select a Visio template, click Select. The Select a Visio Template dialog box displays up to 200 templates. If the object you want to use does not display, enter a search string to reduce the number of objects that display. Select a Visio template and click OK. If you select a Visio template that includes a template image file, the image file displays below the Visio template name.</td>
</tr>
</tbody>
</table>

3. Click Next.

Configuring sources

The Sources page displays differently based on the basis for the task. If the mapping does not include source parameters, the Sources page might not appear.

You can add a single source object or multiple source objects based on the connection type and the mapping configuration. You can also configure a source filter.

After you configure the source, validate the configuration.
If the mapping specifies a connection parameter and after you create the mapping task, you edit the mapping to change the source object, you might need to edit the task. Check the specific connection in the mapping task to determine if you need to reset it and then validate the task again.

1. On the Sources page, configure the following details as required:

<table>
<thead>
<tr>
<th>Source parameter detail</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Connection</td>
<td>Select a connection. For a Visio template, the list of available connections depends on the connections associated with the runtime environment and the connection types allowed by the Visio template. To configure advanced properties for the connection, click Advanced. Not available for all connection types. For information about a particular connector’s properties, see the appropriate connector guide.</td>
</tr>
<tr>
<td>Object</td>
<td>Select a source object. If a list of objects does not appear, click Select. The Select Source Object dialog box displays up to 200 objects. If the object you want to use does not appear, enter a search string to reduce the number of objects that appear.</td>
</tr>
<tr>
<td>Display Technical Field Names Instead of Labels</td>
<td>Displays technical names instead of business names. Not available for all connection types.</td>
</tr>
<tr>
<td>Display Source Fields in Alphabetical Order</td>
<td>Displays source fields in alphabetical order. By default, fields appear in the order returned by the source system.</td>
</tr>
</tbody>
</table>

2. For a parameterized source object, configure field metadata if required. You can configure field metadata when the source connection is for a flat file or Informatica Intelligent Cloud Services Connector.

To configure field metadata, click Edit Types. In the Edit Field Datatypes dialog box, configure the following attributes and click OK:

<table>
<thead>
<tr>
<th>Data type attribute</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Datatype</td>
<td>The data type of the field.</td>
</tr>
<tr>
<td>Precision</td>
<td>Total number of digits in a number. For example, the number 123.45 has a precision of 5. The precision must be greater than or equal to 1.</td>
</tr>
<tr>
<td>Scale</td>
<td>Number of digits to the right of the decimal point of a number. For example, the number 123.45 has a scale of 2. Scale must be greater than or equal to 0. The scale of a number must be less than its precision. The maximum scale for a Numeric datatype is 65535.</td>
</tr>
</tbody>
</table>

3. For Visio templates, configure lookup details if necessary.
This appears when a lookup requires a connection or object, and the lookup is configured to display on this page. If you need to select a lookup object, select the object from the list. If a list does not appear, click **Select**.

For some connection types, you can click **Advanced** to configure advanced properties for the connection.

For some connection types, you can select **Display technical field names instead of labels** to display technical names instead of business names.

To display fields in alphabetical order, click **Display lookup fields in alphabetical order**.

4. If necessary, configure mapplet details.

This appears when a mapplet requires a connection, and the mapplet is configured to display on this page.

For some connection types, you can select **Display technical field names instead of labels** to display technical names instead of business names.

To display fields in alphabetical order, click **Display mapplet fields in alphabetical order**.

5. If necessary, configure stored procedure details.

This displays when a stored procedure requires a connection, and the stored procedure is configured to appear on this page.

6. To test the source connection, click **Validate**.

7. Click **Next**.

## Configuring targets

The **Targets** page displays differently depending on the basis for the task.

For mappings, the **Targets** page displays when the mapping includes parameters for target connections or target objects. The properties that you need to specify are based on the type of parameter. For example, if the target is parameterized but the connection is not, you must specify a target and you can optionally change the connection.

For Visio templates, the **Targets** page can display connection and template parameters for targets, lookups, and stored procedures, as well as mapplets that contain lookups.

Note the following additional information about template parameters:

- The connection and object names that display are based on the template parameter names in the Visio template.
- For a mapplet, you select a connection. You do not select objects for mapplets.
- When a connection name displays without surrounding dollar signs, it is a logical connection. If the logical connection is associated with multiple objects on the **Targets** page, you select the logical connection once, and then select each object.
- If the logical connection is associated with objects on other pages of the task wizard, be sure to use the same connection for logical connections with the same name.

When you select an object, the **Data Preview** area displays a portion of the data in the object. For a flat file connection, data preview displays all of the columns and the first ten rows of the object. For other connection types, data preview displays the first ten rows of the first five columns in the object. It also displays the total number of columns in the object.

If the page has more than one object, you can select the object in the **Data Preview** area to display its data.
Data preview does not display the following types of data:

- Mapplet data.
- Certain Unicode characters.
- Binary data. If the object contains binary data, data preview shows the following text:
  BINARY DATA

1. On the **Targets** tab, configure the following details as required:

<table>
<thead>
<tr>
<th>Target parameter detail</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Connection</td>
<td>Select a connection. For a Visio template, the list of available connections depends on the connections associated with the selected runtime environment and the connection types allowed by the Visio template. To create a connection, click <strong>New</strong>. To edit a connection, click <strong>View</strong>, and in the <strong>View Connection</strong> dialog box, click <strong>Edit</strong>. To configure advanced properties for the connection, click <strong>Advanced</strong>. Not available for all connection types. For information about a particular connector's properties, see the appropriate connector guide.</td>
</tr>
<tr>
<td>Object</td>
<td>Select a target object. If a list of objects does not appear, click <strong>Select</strong>. The <strong>Select Target Object</strong> dialog box displays up to 200 objects. If the object you want to use does not display, enter a search string to reduce the number of objects that display.</td>
</tr>
<tr>
<td>Display Technical Field Names Instead of Labels</td>
<td>Displays technical names instead of business names. Not available for all connection types.</td>
</tr>
<tr>
<td>Display Target Fields in Alphabetical Order</td>
<td>Displays target fields in alphabetical order. By default, fields appear in the order returned by the target system.</td>
</tr>
<tr>
<td>Formatting Options</td>
<td>For Flat File and FTP/SFTP connections only. Select a delimiter and text qualifier. Optionally, select an escape character. If you choose Other for the delimiter, the delimiter cannot be an alphanumeric character or a double quotation mark. If you choose a delimiter for an FTP/SFTP flat file, Data Integration applies the delimiter to the local file, not the remote file, when previewing and reading data. If the remote and local files are not synchronized, you might see unexpected results.</td>
</tr>
<tr>
<td>Create Target</td>
<td>For flat file and relational database connections only. Creates a target file. Enter a name for the target file. If you want the file name to include a time stamp, select <strong>Handle Special Characters</strong> and add special characters to the file name, for example, <code>Accounts_4d49y77T.csv</code>.</td>
</tr>
</tbody>
</table>

2. For a task based on a Visio template, configure the following details if necessary:

- Lookup details. This displays when a lookup requires a connection or object, and the lookup is configured to display on this page. If you need to select a lookup object, select the object from the list. If a list does not display, click **Select**.
- Mapplet details. This displays when a mapplet requires a connection and the mapplet is configured to display on this page.
- Stored procedure details. This displays when a stored procedure requires a connection and the stored procedure is configured to display on this page.

For a task based on a mapping, you define these properties in the mapping.

3. Click Next.

## Configuring parameters

The **Input Parameters** page and **In-Out Parameters** page display differently depending on the basis for the task.

For mappings, the **Input Parameters** page displays parameters that are not in the Source transformation or Target transformation. Depending on the mapping data flow, you might need to configure some parameters before the task wizard allows you to configure other parameters. For more information, see Mappings.

For Visio templates, the **Input Parameters** page displays lookups, mapplets, or stored procedures that require connections, and string template parameters. String template parameters display based on the template parameter properties in the imported Visio template.

## Aggregate functions

You can use aggregate functions for template parameters associated with an Aggregator object.

Aggregate functions display in the **Field Expression** dialog box when you configure the template parameter display options in the Visio template to allow aggregate functions. You can use the following aggregate functions:

- **AVG**
- **COUNT**
- **FIRST**
- **LAST**
- **MAX (Date)**
- **MAX (Number)**
- **MAX (String)**
- **MEDIAN**
- **MIN (Date)**
- **MIN (Number)**
- **MIN (String)**
- **PERCENTILE**
- **STDDEV**
- **SUM**
- **VARIANCE**
Configuring input or in-out parameters

1. For a task based on a mapping, on the Input Parameters or In-Out Parameters page, configure the parameters that display.
   Depending on the data flow of the mapping, you might need to configure certain parameters before the task wizard allows you to configure other parameters.
   For more information on using in-out parameters and input parameters in mappings, see Mappings.

2. For a task based on a Visio template, on the Input Parameters page, configure shared connection details if necessary.
   Logical connections display in the Shared Connection Details area.
   If the logical connection is associated with multiple objects, select the logical connection, and then select each object.
   If the logical connection is associated with objects on other pages of the wizard, be sure to use the same connection for logical connections with the same name.

3. If necessary, configure lookup details.
   This displays when a lookup requires a connection or object and the lookup is configured to display on this page. If you need to select a lookup object, select the object from the list. If a list of objects does not appear, click Select.
   For some connection types, click Advanced to configure advanced properties for the connection. For more information, see Connections.
   For some connection types, you can select Display technical field names instead of labels to display technical names instead of business names.
   To display fields in alphabetical order, click Display lookup fields in alphabetical order.

4. If necessary, configure mapplet details.
   This displays when a mapplet requires a connection and the mapplet is configured to display on this page.
   For some connection types, you can select Display technical field names instead of labels to display technical names instead of business names.
   To display fields in alphabetical order, click Display mapplet fields in alphabetical order.

5. If necessary, configure stored procedure details.
   This displays when a stored procedure requires a connection and the stored procedure is configured to display on this page.

6. Define the remaining template parameters, as needed.
String template parameters display individually based on the Visio template. The following table describes how to define a string template parameter based on the input control type:

<table>
<thead>
<tr>
<th>Input control type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text box</td>
<td>Enter any valid value. <strong>Note</strong>: You cannot use blank spaces. Also, leading and trailing spaces are removed at runtime.</td>
</tr>
<tr>
<td>Data filter dialog box</td>
<td>To define the template parameter, click <strong>New</strong>. To create a simple data filter, in the <strong>Data Filters</strong> dialog box, select a column and operator and enter the value you want to use. To create an advanced data filter, click <strong>Advanced</strong>. Enter the field expression that you want to use, and click <strong>OK</strong>. If the template parameter is already defined with a data filter, delete the existing data filter before creating a new data filter. <strong>Note</strong>: For a template parameter included in a source filter, use an advanced data filter.</td>
</tr>
<tr>
<td>Field expression dialog box</td>
<td>To define the template parameter, click <strong>New</strong>. In the <strong>Field Expressions</strong> dialog box, enter the expression you want to use and click <strong>OK</strong>. For more information about configuring field expressions, see &quot;Field expressions&quot; on page 15.</td>
</tr>
<tr>
<td>Field list</td>
<td>To define the template parameter, select a field from the list.</td>
</tr>
<tr>
<td>Field mapping dialog box</td>
<td>To define the template parameter, configure the field mappings you want to use:</td>
</tr>
<tr>
<td></td>
<td>- The left table can display fields from sources, mapplets, and lookups. The right table can display fields from multiple targets, as well as mapplets and lookups. Use the Object list to display fields from different objects. By default, all available fields display.</td>
</tr>
<tr>
<td></td>
<td>- To match fields with similar names, click <strong>Automatch</strong>.</td>
</tr>
<tr>
<td></td>
<td>- You can select and drag the source fields to the applicable target fields.</td>
</tr>
<tr>
<td></td>
<td>- To clear all field mappings, click <strong>Clear Mapping</strong>.</td>
</tr>
<tr>
<td></td>
<td>- To clear mapped field, click the <strong>Clear Mapped Field</strong> icon for the target field.</td>
</tr>
<tr>
<td></td>
<td>If you map a target field in a task with multiple targets, you also map any matching fields in other targets. A matching field is one with the same name and datatype, precision, and scale. If the target fields have the same name but different datatypes, precision, or scale, you can map one of the target fields. If fields from a lookup do not display, configure the lookup connection and object, save the task, then edit the task again.</td>
</tr>
<tr>
<td>Custom Dropdown</td>
<td>To define the template parameter, select an option from the list.</td>
</tr>
</tbody>
</table>

7. Click **Next**.
Configuring a schedule and advanced options

On the **Schedule** page, you specify whether to run a mapping task manually or schedule it to run at a specific time or interval. You can create a schedule or use an existing schedule.

You can also configure email notification and advanced options for the task on the **Schedule** page.

1. To specify whether to run the task on a schedule or without a schedule, choose one of the following options:
   - If you want to run the task on a schedule, click **Run this task on schedule**. Select the schedule you want to use or click **New** to create a schedule.
   - If you want to run the task without a schedule, click **Do not run this task on a schedule**.

2. Configure email notification options for the task.

3. Optionally, enter the following advanced options:

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preprocessing Commands</td>
<td>Command to run before the task.</td>
</tr>
<tr>
<td>Postprocessing Commands</td>
<td>Command to run after the task completes.</td>
</tr>
<tr>
<td>Parameter File Name</td>
<td>Name of the file that contains the definitions and values of user-defined parameters used in the task.</td>
</tr>
<tr>
<td>Maximum Number of Log Files</td>
<td>Number of session log files and import log files to retain. By default, Data Integration stores each type of log file for 10 runs before it overwrites the log files for new runs.</td>
</tr>
</tbody>
</table>

4. Choose whether to run the task in standard or verbose execution mode.
   - If you select Verbose mode, the mapping generates additional data in the logs that you can use for troubleshooting. It is recommended that you select verbose execution mode only for troubleshooting purposes. Verbose execution mode impacts performance because of the amount of data it generates.

5. Optionally, configure advanced session properties.
   a. Click **Add**.
   b. Select an advanced session property.
   c. Configure the advanced session property.

6. Click **Finish**.

Viewing mapping task details

You can view details about a mapping task, such as the mapping or Visio template used by the task.

The **Task Details** page includes the following information:

- The runtime environment used to run the mapping task.
- The date the task was created and the user who created the task.
- The last time the task was updated and the user who updated the task.
- The date of the last run.
The name and image of the mapping on which the task is based and the date the mapping was last updated.

1. On the Explore page, navigate to the task.
2. In the row that contains the task, click Actions and select View.
   On the Task Details page, you can click Edit to modify the mapping task or Run to run the task.

Running a mapping task

You can run a mapping task in the following ways:

- Manually. To run a mapping task manually, on the Explore page, navigate to the task. In the row that contains the task, click Actions and select Run.
  You can also run a mapping task manually from the Task Details page. To access the Task Details page, click Actions and select View.
- On a schedule. To run a mapping task on a schedule, edit the task in the mapping task wizard to associate the task with a schedule.
CHAPTER 3

Synchronization tasks

Use the synchronization task to synchronize data between a source and a target. For example, you can read sales leads from your sales database and write them into Salesforce. You can also use expressions to transform the data according to your business logic or use data filters to filter data before writing it to targets.

You can use the following source and target types in synchronization tasks:

- Database
- Flat file
- Salesforce

Task operations

When you configure a synchronization task, you specify the task operation and the type of target. The available target types depend on the task operation that you select.

You can use one of the following task operations:

Insert

When you run a task with the Insert task operation, Data Integration inserts all source rows into the target. If Data Integration finds a source row that exists in the target, the row fails.

If you write data to a flat file target, Data Integration truncates the flat file before it inserts the source rows into the file.

Update

When you run a task with the Update task operation, Data Integration updates rows in the target that exist in the source. If Data Integration finds a row in the source that does not exist in the target, the row fails.

Upsert

When you run a task with the Upsert task operation, Data Integration updates all rows in the target that also exist in the source and inserts all new source rows into the target.

If a source field contains a NULL value and the corresponding target field contains a value, Data Integration retains the existing value in the target field.

Delete

When you run a task with the Delete task operation, Data Integration deletes all rows from the target that exist in the source.
Synchronization task sources

You can use a single object as a source or multiple related objects as sources in a synchronization task. You can use multiple objects for database connection types. Define relationships based on key columns or create a user-defined join condition.

Rules and guidelines for multiple-object databases

Use the following rules and guidelines when you configure a multiple-object database:

- All objects must be available through the same source connection. All database tables in a multiple-object source must have valid relationships defined by key columns or user-defined join conditions.
- When you add multiple database tables as sources for a task, you can either create relationships or user-defined joins, but not both.
- The synchronization task wizard removes a user-defined join under the following conditions:
  - You remove one of two remaining database tables from the list of sources for the task.
  - You change the source connection from database to flat file or Salesforce.

Synchronization task targets

You can use a single object as a target for a synchronization task.

The target connections that you can use depend on the task operation you select for the task. For example, if you select the upsert task operation, you cannot use a flat file target connection because you cannot upsert records into a flat file target.

Flat file target creation

If a task has a flat file target, create the flat file before you save the task. Or, you can create the flat file target with the synchronization task wizard when all of the following are true:

- The source connection type is Salesforce, database, or ODBC.
- The source object is Single or Custom.
- The target connection type is Flat File.

The synchronization task wizard uses the source object name as the default name of the flat file target. It truncates the name of the flat file to the first 100 characters if the source name is too long. If the target name conflicts with the name of another target object, the following error appears:

Object named <object name> already exists in the target connection.

Database target truncation

You can configure a synchronization task to truncate a database target table before writing new data to the table when you configure the task to use an Insert task operation. By default, Data Integration inserts new rows without truncating the target table.
Salesforce targets and IDs for related objects

Data Integration identifies records of a Salesforce object based on one of the following types of IDs:

- Salesforce ID
  Salesforce generates an ID for each new record in a Salesforce object.

- External ID
  You can create a custom external ID field in the Salesforce object to identify records in the object. You might create an external ID to use the ID generated from a third-party application to identify records in the Salesforce object. You can use one or more external IDs to uniquely identify records in each Salesforce object.

If you create a synchronization task that writes to a Salesforce target, the source must provide either the Salesforce IDs or the external IDs for the records in the Salesforce target object and applicable related objects. A related object is an object that is related to another object based on a relationship defined in Salesforce. The synchronization task uses the Salesforce ID or external ID to update changes to related objects.

If the source in a task contains external IDs for Salesforce objects, you must specify the external IDs for all related objects when you create the Salesforce target for the task. If you do not specify the external ID, Data Integration requires the Salesforce ID to identify records in each related object.

For more information about creating and using Salesforce external IDs, see the Data Integration Community article, "Using External IDs and Related Objects in Informatica Cloud".

Update columns

Update columns are columns that uniquely identify rows in the target table. Add update columns when the database target table does not contain a primary key and the synchronization task uses an update, upsert, or delete task operation.

When you run the synchronization task, the synchronization task uses the field mapping to match rows in the source to the database table. If the synchronization task matches a source row to multiple target rows, it performs the specified task operation on all matched target rows.

Column names in flat files

If the column name in a flat file source contains nonalphanumeric characters, starts with a number, or contains more than 75 characters, the synchronization task modifies the column name in the flat file target.

The synchronization task truncates column names to 75 characters. For a flat file source, the Data Preview area and the Field Expression dialog box show modified column names. For a flat file target, the synchronization task changes the column name in the flat file when it generates the file at run time.
Rules and guidelines for synchronization task sources and targets

Use the following rules and guidelines for synchronization sources and targets:

- Field names must contain 65 characters or less.
- Field names must contain only alphanumeric or underscore characters. Spaces are not allowed.
- Field names cannot start with a number.
- Each field name must be unique within each source and target object.
- Data Integration truncates data if the scale or precision of a numeric target column is less than the scale or precision of the corresponding source column.

Rules and guidelines for flat file sources and targets

Use the following rules and guidelines for flat file sources and targets:

- All date columns in a flat file source must have the same date format. Rows that have dates in a different format than the one specified in the synchronization task definition are written to the error rows file.
- Each flat file target must contain all fields that will be populated by the synchronization task.
- The synchronization task truncates a flat file target before writing target data to the file.
  To avoid overwriting target data, you might use a post-session command to merge target data with a master target file in a different location.
- The flat file cannot contain empty column names. If a file contains an empty column name, the following error appears:
  
  Invalid header line: Empty column name found.

- Do not map binary fields when you use a flat file source or target in a synchronization task.
- Column names in a flat file must contain printable tab or ASCII characters (ASCII code 32-126). If the file contains a character that is not valid, the following error appears:
  
  Invalid header line: Non-printable character found. The file might be binary or might have invalid characters in the header line.

- For flat file sources and targets with multibyte data on Linux, the default locale must be UTF-8.

Rules and guidelines for database sources and targets

Use the following rules and guidelines for database sources and targets:

- You can use database tables as targets. You can use database tables, aliases, and views as sources.
- Relational targets must meet the minimum system requirements.
- The database user account for each database target connection must have DELETE, INSERT, SELECT, and UPDATE privileges.

Field mappings

Configure field mappings in a synchronization task to map source columns to target columns.

Configure field mapping on the Field Mapping page of the synchronization task wizard.
You must map at least one source column to a target column. Map columns with compatible datatypes or create field expressions to convert datatypes appropriately.

Depending on the task operation, the synchronization task requires certain fields to be included in the field mapping. By default, the synchronization task maps the required fields. If you configure the field mapping, ensure that the required fields remain mapped. If you do not map the required fields, the synchronization task fails.

The following table shows the required fields for each applicable task operation for a database target:

<table>
<thead>
<tr>
<th>Required field</th>
<th>Task operations</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary Keys</td>
<td>Delete</td>
<td>Map primary key columns to enable the synchronization task to identify records to delete, update, or upsert in a database target.</td>
</tr>
<tr>
<td></td>
<td>Update</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Upsert</td>
<td></td>
</tr>
<tr>
<td>Non-null fields</td>
<td>Insert</td>
<td>Map all fields that cannot be null in database.</td>
</tr>
<tr>
<td></td>
<td>Update</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Upsert</td>
<td></td>
</tr>
</tbody>
</table>

When you configure field mappings, you can also perform the following tasks:

- Edit field datatypes.
- Add a mapplet to the field mapping.
- Create lookups.

### Field datatypes

When you create a synchronization task, Data Integration assigns a datatype to each field in the source and target. You can edit the field datatypes on the Field Mapping page of the synchronization task wizard. You can edit field datatypes for any source or target type except for Data Integration Connector sources and targets, and mapplets.

### Mapplets in field mappings

You can add a mapplet to a field mapping. After you add a mapplet to a field mapping, you must map the source fields to the input fields of the mapplet and map the output fields of the mapplet to the target fields.

When a source field is mapped directly to a target field and you map an output field of a mapplet to the same target field, Data Integration concatenates the values of the source and output fields in the target field. Verify that the expression is correct for the target field.

**Note:** The names of the output fields of a mapplet do not match the source field names. Data Integration appends a number to the end of the source field name to determine the output field name. In addition, Data Integration may not display the output fields in the same order as the source fields.

### Lookup conditions

A lookup returns values based on a lookup condition. You can create a lookup condition based on information in the source on the Field Mapping page of the synchronization task wizard. For example, for a SALES source database table, you might set the ITEM_ID column equal to ITEM_ID column in a ITEMS flat file, and have the lookup return the item name for each matching item ID.
When you create a lookup condition, you define the following components:

- **Lookup connection and object.** The connection and object to use to perform the lookup. When possible, use a native connection. For example, to perform a lookup on an Oracle table, use an Oracle connection instead of an ODBC connection.
- **Source and lookup fields.** The fields used to define the lookup condition. The synchronization task compares the value of the source field against the lookup field and then returns a value based on the match. You can define multiple conditions in a lookup. If you define more than one lookup condition, all lookup conditions must be true to find the match.

For example, you define the following conditions for a lookup:

```plaintext
SourceTable.Name = LookupTable.Name
SourceTable.ID = LookupTable.ID
```

The synchronization task performs the following lookup:

```plaintext
Lookup (SourceTable.Name = LookupTable.Name, SourceTable.ID = LookupTable.ID)
```

### Lookup return values

When you configure a lookup, you configure a lookup return value. The lookup return value depends on the return value properties that you define, such as multiplicity or a lookup expression.

A lookup return value is the value that Data Integration returns when it finds a match based on the lookup condition. If the lookup returns an error, Data Integration writes the row to the error rows file.

You can configure a lookup expression as part of the lookup return value. Configure a simple expression that uses the $OutputField variable to represent the lookup return value.

For example, the following expression adds 100 to each lookup return value:

```plaintext
$OutputField+100
```

As another example, you can use the concatenate operator (||) to append a string to a string lookup return value as follows:

```plaintext
'Mighty' || '$OutputField'
```

You can use parameters defined in a parameter file in a lookup expression.

The following table describes the lookup return value properties that you can configure:

<table>
<thead>
<tr>
<th>Lookup Return Value Property</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Output Field</strong></td>
<td>The field from the lookup table that you want to use.</td>
</tr>
<tr>
<td><strong>Multiplicity</strong></td>
<td>How Data Integration handles multiple return values.</td>
</tr>
<tr>
<td></td>
<td>- Error If More Than 1 Output Value. Select if the synchronization task should display an error when the lookup condition returns multiple values. Data Integration rejects rows when multiple matches are found, writing them to the error rows file. This is the default.</td>
</tr>
<tr>
<td></td>
<td>- Randomly Pick 1 Output Value. Select if the synchronization task should choose the first returned value when a lookup condition returns multiple values. Different systems might use different orders to return lookup values.</td>
</tr>
<tr>
<td><strong>Expression</strong></td>
<td>A simple expression that uses $OutputField to represent the selected output field. By default, Data Integration passes the lookup return value without alteration with the following expression: $OutputField.</td>
</tr>
</tbody>
</table>
Rules and guidelines for lookups

Use the following rules and guidelines when creating a lookup:

- If the lookup is on a flat file, the file must use a comma delimiter. You cannot use any other type of delimiter.
- When you configure a lookup, you can configure a simple lookup expression as part of the lookup return value. Use the $OutputField variable to represent the expression. If you use a lookup expression that does not include $OutputField, you negate the action of the lookup.
- Tasks with a flat file lookup that run by a Secure Agent on Windows 7 (64 bit) might not complete. To resolve the issue, configure a network login for the Secure Agent service.
- On the Field Mapping page, you can perform a lookup or create an expression for each source field. You cannot do both.
- Each task can contain one or more lookups. To avoid impacting performance, include less than six lookups in a task.
- When performing the lookup, the task performs an outerjoin and does not sort the input rows. The lookup performs a string comparison that is not case-sensitive to determine matching rows.
- The source field and lookup field in the lookup condition must have compatible datatypes. If the datatypes are not compatible, the following error appears:
  
  Source field [<source field name> (<source field datatype>)] and lookup field [<lookup field name> (<lookup field datatype>)] have incompatible datatypes.

  If you create multiple lookup conditions on a lookup field and the lookup source is a flat file, all source fields must have the same datatype. The synchronization task uses the larger precision and scale of the source field datatypes as the precision and scale for the target field. If the source fields do not have the same datatype, the following error appears:

  Lookup field <field name> in <file name> has conflict data types inferred: <datatype 1> and <datatype 2>.

- You cannot include lookup fields of particular datatypes in a lookup condition. When the lookup field in a flat file has the Text or Ntext datatype or the target field of a lookup has the Text or Ntext datatype, the task fails.
- If you run a task with a lookup and the source field, lookup field, or output field of the lookup no longer exist in the lookup object, an error appears.

Preprocessing and postprocessing commands

You can run preprocessing and postprocessing commands to perform additional jobs. The task runs preprocessing commands before it reads the source. It runs postprocessing commands after it writes to the target.

You can use the following types of commands:

- SQL commands. Use SQL commands to perform database tasks.
- Operating system commands. Use shell and DOS commands to perform operating system tasks.

If any command in the preprocessing or postprocessing scripts fail, the task fails.
Preprocessing and postprocessing SQL commands

You can run SQL commands before or after a task. For example, you can use SQL commands to drop indexes on the target before the task runs, and then recreate them when the task completes. Data Integration does not validate the SQL.

Use the following rules and guidelines when creating the SQL commands:

- Use any command that is valid for the database type. However, Data Integration does not allow nested comments, even if the database allows them.
- Use a semicolon (;) to separate multiple statements. Data Integration issues a commit after each statement.
- Data Integration ignores semicolons within comments. If you need to use a semicolon outside of comments, you can escape it with a backslash (\).

Preprocessing and postprocessing operating system commands

Data Integration can perform operating system commands before or after the task runs. For example, use a preprocessing shell command to archive a copy of the target flat file before the task runs on a UNIX machine.

You can use the following types of operating system commands:

- UNIX. Any valid UNIX command or shell script.
- Windows. Any valid DOS or batch file.

Configuring a synchronization task

Configure a synchronization task using the synchronization task wizard.

To configure a synchronization task, complete the following steps:

1. Complete the prerequisite tasks.
2. Create the synchronization task.
3. Configure the source.
4. Configure the target.
5. Optionally, configure data filters.
6. Configure field mappings.
7. Optionally, configure a schedule and advanced options.

As you work through the task wizard, you can click **Save** to save your work at any time. When you have completed the wizard, click **Finish** to save and close the task wizard.

Synchronization prerequisite tasks

Before you create a synchronization task, complete the following prerequisite tasks:

- Create database users.
  To write source data to a database target, the database administrator must create a database user account in the target database. Each database user account must have the DELETE, INSERT, SELECT, and UPDATE privileges.
- Verify that the sources and targets meet your requirements.
Defining a synchronization task

1. To create a synchronization task, click New > Tasks. Select Synchronization Task and click Create. To edit a synchronization task, on the Explore page, navigate to the synchronization task. In the row that contains the task, click Actions and select Edit.

2. In Synchronization Task Details, configure the following fields:

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task Name</td>
<td>Name of the synchronization task. Task names can contain alphanumeric characters, spaces, and the following special characters: _ . + - Maximum length is 100 characters. Task names are not case sensitive.</td>
</tr>
<tr>
<td>Location</td>
<td>Project folder in which the task resides. If the Explore page is currently active and a project or folder is selected, the default location for the asset is the selected project or folder. Otherwise, the default location is the location of the most recently saved asset.</td>
</tr>
<tr>
<td>Description</td>
<td>Description of the synchronization task. Maximum length is 4000 characters.</td>
</tr>
<tr>
<td>Task Operation</td>
<td>Select one of the following task operation types: Insert, Update, Upsert, Delete. The list of available targets in a subsequent step depend on the operation you select.</td>
</tr>
</tbody>
</table>

3. Click Next.

Configuring the source

Select the source for the synchronization task. The steps to configure a source vary based on whether you use a single object or saved query as the source, or you use multiple database tables as the source.

Configuring a single object or saved query as the source

You can configure a single object or saved query as the source of a synchronization task.

1. On the Source page, select a connection.
   To create a connection, click New. To edit a connection, click View, and in the View Connection dialog box, click Edit.
2. To use a single source, select Single.
   To use a saved query, select Saved Query.
   You can use a saved query when you use a database connection.
3. If the connection includes less than 200 objects, select a source object or click Select.
   If the connection includes more than 200 objects, click Select.
The Select Source Object dialog box displays up to 200 objects. If the object you want to use does not display, enter a search string to reduce the number of objects that display.

Select an object and click Select.

4. To display technical names instead of business names, select Display technical field names instead of labels.
   This option is not available for all connection types.

5. To display source fields in alphabetical order, click Display source fields in alphabetical order.
   By default, fields appear in the order returned by the source system.

6. For a flat file or FTP/SFTP single source, click Formatting Options. Select a delimiter and text qualifier. Optionally, select an escape character.
   If you choose Other for the delimiter, the delimiter cannot be an alphanumeric character or a double quotation mark.
   If you choose a delimiter for an FTP/SFTP flat file, Data Integration applies the delimiter to the local file, not the remote file, when previewing and reading data. If the remote and local files are not synchronized, you might see unexpected results.

7. If preview data does not display automatically, click Show Data Preview to preview data.
   The Data Preview area shows the first ten rows of the first five columns in the object. It also displays the total number of columns in the object.
   The Data Preview area does not display certain Unicode characters correctly. If the data contains binary data, the Data Preview area shows the following text:
   BINARY DATA

8. To preview all source columns in a file, click Preview All Columns.
   The file shows the first ten rows of the source.

9. Click Next.

Configuring multiple database tables as the source

You can configure multiple database tables as the source of a synchronization task.

1. On the Source page, select a database connection.
   To create a connection, click New. To edit a connection, click View and then click Edit.

2. Select Multiple.
   The Source Objects table displays.

3. Click Add.

4. In the Select Source Objects dialog box, select the objects you want to use.
   The dialog box displays up to 200 objects. If the objects that you want to use do not display, enter a search string to reduce the number of objects that display.
   When you select an object, it appears in the Selected Objects list. To remove an object from the Selected Objects list, press Delete.

5. Click Select.
   The selected sources display in the Source Objects table. To remove a source, in the Actions column, click Remove.

6. To display source fields in alphabetical order, select Display source fields in alphabetical order.
   By default, source fields appear in the order returned by the source system.
7. Create source relationships or create a user-defined join, and click OK.
   To create a relationship, perform the following steps:
   a. Select a database table and click Create Relationship.
   b. Select the source key for the table and then select the related source object and matching object key.
   c. Click OK.
   d. Match the primary key of the source table to the corresponding foreign key of the related database table.
   e. Create relationships as necessary to include all sources in the task.
   To create a user-defined join to join all database tables, perform the following steps:
   a. Select User Defined Join and define the join.
   b. Any existing relationships are added to the join condition. To ensure that you enter field names correctly, use the Object list and Fields list to add field names to the join statement.
   c. To save the user-defined join, click OK.

8. To preview source data, select the source in the Source Objects table. If preview data does not appear automatically, click Show Data Preview.
   The Data Preview area shows the first ten rows of the first five columns in the source. It also displays the total number of columns in the source.
   The Data Preview area does not display certain Unicode characters as expected. If the data contains binary data, the Data Preview area shows the following text:
   BINARY DATA

9. To preview all source columns in a file, select the source in the Source Objects table and click Preview All Columns.
   The file shows the first ten rows of the source.

10. Click Next.

Configuring the target

You can configure a single target for a synchronization task. The options that appear on the page depend on the task type and target type that you select for the task.

1. On the Target page, enter the following information:

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Connection</td>
<td>Select a connection. The list of available connections depends on the task operation defined for the task. To create a connection, click New. To edit a connection, click View, and in the View Connection dialog box, click Edit.</td>
</tr>
<tr>
<td>Target Object</td>
<td>If the connection includes less than 200 objects, select a target object or click Select. If the connection includes more than 200 objects, click Select. The Select Target Object dialog box displays up to 200 objects. If the object you want to use does not appear, enter a search string to reduce the number of objects that display. Select a target object and click OK.</td>
</tr>
<tr>
<td>Field</td>
<td>Description</td>
</tr>
<tr>
<td>------------------------------------------------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Display Technical Field Names Instead of Labels</td>
<td>Displays technical names instead of business names. Not available for all connection types.</td>
</tr>
<tr>
<td>Display Target Fields in Alphabetical Order</td>
<td>Displays target fields in alphabetical order instead of the order returned by the target system.</td>
</tr>
<tr>
<td>Formatting Options</td>
<td>For Flat File and FTP/SFTP connections only. Select a delimiter and text qualifier. Optionally, select an escape character. If you choose Other for the delimiter, the delimiter cannot be an alphanumeric character or a double quotation mark. If you choose a delimiter for an FTP/SFTP flat file, Data Integration applies the delimiter to the local file, not the remote file, when previewing and reading data. If the remote and local files are not synchronized, you might see unexpected results.</td>
</tr>
<tr>
<td>Create Target</td>
<td>Flat File and relational database connections only. Creates a target file. You can create a target file when the source connection is Salesforce, database, or ODBC, and the source object is Single or Custom. Enter a file name and select the source fields that you want to use. By default, all source fields are used.</td>
</tr>
<tr>
<td>Truncate Target</td>
<td>Database targets with the Insert task operation only. Truncates a database target table before inserting new rows. - True. Truncates the target table before inserting all rows. - False. Inserts new rows without truncating the target table. Default is False.</td>
</tr>
<tr>
<td>Enable Target Bulk Load</td>
<td>Select this option to write data in bulk mode. The default value is false.</td>
</tr>
</tbody>
</table>

2. If preview data does not appear automatically, click **Show Data Preview** to preview data.

The **Data Preview** area shows the first ten rows of the first five columns in the target. It also shows the total number of columns in the target.

The **Data Preview** area does not display certain Unicode characters correctly. If the data contains binary data, the Data Preview area shows the following text:

```
BINARY DATA
```

3. To preview all target columns in a file, click **Preview All Columns**.

The file shows the first ten rows of the target.

4. Click **Next**.

### Configuring the data filters

Use a data filter to reduce the number of source rows that the synchronization task reads for the task. By default, the synchronization task reads all source rows.

You can also configure the sort order for the task.

1. On the **Data Filters** page, choose whether to read all rows in sources or to read the first set of rows in sources.
   - To read all rows, select **Process all rows**.
• To read the first set of rows, select **Process only the first** and enter a number.

2. To create a data filter, click **New**.
   • To create a simple data filter, select a source column and operator. Enter the value you want to use, and click **OK**.
   • To create an advanced data filter, click **Advanced**. Enter the field expression you want to use and click **OK**.

You can use parameters defined in a parameter file in data filters. When you use a parameter in a data filter, start the data filter with the parameter. For example, use $$\text{Sales} < 100000$$ instead of $$100000 > $\text{Sales}$$.

To delete a data filter, click **Delete**.

3. To configure sort criteria, configure the following sort options:

<table>
<thead>
<tr>
<th>Data sorting option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Object</td>
<td>Source object.</td>
</tr>
<tr>
<td>Sort By</td>
<td>Source field to use to sort data.</td>
</tr>
<tr>
<td>Sort Direction</td>
<td>Sort direction:</td>
</tr>
<tr>
<td></td>
<td>- ASC. Ascending order.</td>
</tr>
<tr>
<td></td>
<td>- DESC. Descending order.</td>
</tr>
</tbody>
</table>

To add additional sort criteria, click **Add**. Use the Move Up and Move Down arrows to define the order of the sort criteria.

To remove a sort criteria, click **Delete**.

4. Click **Next**.

## Configuring the field mapping

Configure field mappings to define the data that the synchronization task writes to the target.

1. On the **Field Mapping** page, configure field mappings.

2. If you included multiple source objects in the task, you can select each source object in the **Source** field to display the fields for the selected object. Or, you can view all source object fields.

When displaying all source object fields, the **Sources** table displays field names grouped by source object. You can place the cursor over the Status icon for a source field to determine the following information:

• Database table or Salesforce object to which the field belongs.
• Datatype of a field.

3. Some source types allow you to configure field datatypes. To configure field datatypes for a source, click **Edit Types**.

   If the task includes more than one source, first select the source you want to edit.
In the **Edit Field Datatypes** dialog box, configure the following datatype attributes and click **OK**:

<table>
<thead>
<tr>
<th>Datatype attribute</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Datatype</td>
<td>The datatype of data in the column.</td>
</tr>
<tr>
<td>Precision</td>
<td>Total number of digits in a number. For example, the number 123.45 has a precision of 5. The precision must be greater than or equal to 1.</td>
</tr>
<tr>
<td>Scale</td>
<td>Number of digits to the right of the decimal point of a number. For example, the number 123.45 has a scale of 2. Scale must be greater than or equal to 0. The scale of a number must be less than its precision. The maximum scale for a Numeric datatype is 65535.</td>
</tr>
</tbody>
</table>

4. To add a mapplet, complete the following steps:
   a. Click **Add Mapplet**.
   b. In the **Add Mapplet** dialog box, select the mapplet.
   c. To display technical names instead of business names, select **Display technical field names instead of labels**.
   d. To display fields in alphabetical order, click **Display mapplet fields in alphabetical order**. By default, fields appear in the order specified by the mapplet.
   e. If necessary, select a connection for the mapplet.
   f. Click **OK**.

5. To configure field mappings, for **Mapping Selection**, select one of the following options:
   - **Source to Target**. Displays the source and target. Map source fields to the applicable target fields.
   - **Source to Mapplet**. Displays the source and the input fields of the mapplet. Map the source fields to the applicable input fields of the mapplet.
   - **Mapplet to Target**. Displays the output fields of the mapplet and the target fields. Map the output fields of the mapplet to the applicable target fields.
   The **Clear Mapping**, **Automatch**, and **Validate Mapping** buttons apply to the selected area of the field mapping.

6. To match fields with similar names, click **Automatch**. Or, you can select and drag the source fields to the applicable target fields.

   Data Integration caches field metadata. If the fields do not appear correctly, click **Refresh Fields** to update the cache and view the latest field attributes.

7. To configure field datatypes for a target, click **Edit Types**. This option is not available for all target types. If the task includes more than one target, first select the target you want to edit.
In the **Edit Field Datatypes** dialog box, configure the following datatype attributes and click **OK**:

<table>
<thead>
<tr>
<th>Datatype attribute</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Datatype</td>
<td>The datatype of data in the column.</td>
</tr>
<tr>
<td>Precision</td>
<td>Total number of digits in a number. For example, the number 123.45 has a precision of 5. The precision must be greater than or equal to 1.</td>
</tr>
<tr>
<td>Scale</td>
<td>Number of digits to the right of the decimal point of a number. For example, the number 123.45 has a scale of 2. Scale must be greater than or equal to 0. The scale of a number must be less than its precision. The maximum scale for a Numeric datatype is 65535.</td>
</tr>
</tbody>
</table>

8. To create an expression to transform data, click the **Add or Edit Expression** icon in the **Actions** column. In the **Field Expressions** dialog box, enter the expression you want to use and click **OK**. You can use parameters defined in a parameter file in expressions.

9. To create a lookup, click the **Add or Edit Lookup** icon. In the **Field Lookup** dialog box, configure the following properties and click **OK**:

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lookup Connection</td>
<td>Connection for the lookup object.</td>
</tr>
<tr>
<td>Lookup Object</td>
<td>Object on which you want to lookup a value.</td>
</tr>
<tr>
<td>Display Technical Field Names Instead of Labels</td>
<td>Displays technical names instead of business names. Not available for all connection types.</td>
</tr>
<tr>
<td>Display Fields in Alphabetical Order</td>
<td>Displays lookup fields in alphabetical order. By default, fields appear in the order returned by the lookup system.</td>
</tr>
<tr>
<td>Source Fields</td>
<td>Source column to use in the lookup condition.</td>
</tr>
<tr>
<td>Lookup Fields</td>
<td>The column in the lookup table to use in the lookup condition.</td>
</tr>
<tr>
<td>Output Field</td>
<td>The column in the lookup table that contains the output value.</td>
</tr>
</tbody>
</table>
Option | Description
--- | ---
Multiplicity | Determines how to handle cases when a lookup returns multiple values:
- Error If More Than 1 Output Value. Select if the synchronization task should display an error when the lookup condition returns multiple values. Data integration rejects rows when multiple matches are found, writing them to the error rows file. This is the default.
- Randomly Pick 1 Output Value. Select if the synchronization task should choose the first returned value when a lookup condition returns multiple values. Different systems might use different orders to return lookup values.

Expression | A simple expression that uses the $OutputField variable to represent the lookup return value. Enter a simple expression, such as $OutputField*100. You can use parameters defined in a parameter file in lookup expressions. To return the lookup return value without an additional expression, use $OutputField.

10. To clear an expression or lookup and delete the field mapping, click the **Clear Expression/Lookup** icon next to the target field.

11. To clear all field mappings, click **Clear Mapping**.

12. To validate a mapping, click **Validate Mapping**.

13. Click **Next**.

### Configuring a schedule and advanced options

On the **Schedule** page of the synchronization task wizard, you can specify to run a synchronization task manually or schedule it to run at a specific time or interval. You can create a schedule or use an existing schedule.

You can also configure email notifications and advanced options for the task on the **Schedule** page.

1. On the **Schedule** page, choose whether to run the task on a schedule or without a schedule.
2. To run a task on a schedule, click **Run this task on schedule** and select the schedule you want to use. To create a new schedule, click **New**. Enter schedule details and click **OK**. To remove the task from a schedule, click **Do not run this task on a schedule**.
3. If necessary, select a runtime environment to run the task.
4. Configure email notification options for the task.
5. Optionally, configure advanced options.

You can configure the following advanced options:

<table>
<thead>
<tr>
<th>Advanced options</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preprocessing Commands</td>
<td>Command to run before the task.</td>
</tr>
<tr>
<td>Postprocessing Commands</td>
<td>Command to run after the task completes.</td>
</tr>
</tbody>
</table>
### Advanced options

<table>
<thead>
<tr>
<th><strong>Advanced options</strong></th>
<th><strong>Description</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameter File Name</td>
<td>Name of the file that contains the definitions and values of user-defined</td>
</tr>
<tr>
<td></td>
<td>parameters used in the task.</td>
</tr>
<tr>
<td>Maximum Number of Log Files</td>
<td>Number of session log files, error log files, and import log files to retain. By</td>
</tr>
<tr>
<td></td>
<td>default, Data Integration stores each type of log file for 10 runs before it</td>
</tr>
<tr>
<td></td>
<td>overwrites the log files for new runs.</td>
</tr>
<tr>
<td>Update Columns</td>
<td>Database targets only. Temporary primary key columns to update target data.</td>
</tr>
<tr>
<td></td>
<td>If the database target does not include a primary key column, and the task</td>
</tr>
<tr>
<td></td>
<td>performs an update, upsert, or delete task operation, click Add to add a</td>
</tr>
<tr>
<td></td>
<td>temporary key.</td>
</tr>
<tr>
<td>Upsert Field Name</td>
<td>The target field to use to perform upserts.</td>
</tr>
</tbody>
</table>

6. Choose whether to run the task in standard or verbose execution mode.
   
   If you select verbose execution mode, the mapping generates additional data in the logs that you can use for troubleshooting. It is recommended that you select verbose execution mode only for troubleshooting purposes. Verbose execution mode impacts performance because of the amount of data it generates.

7. Click **Finish**.

## Viewing synchronization task details

You can view details about a synchronization task, including the source and target connections, the field mapping, and the associated schedule.

1. On the **Explore** page, navigate to the task.
2. In the row that contains the task, click **Actions** and select **View**.

   On the **Task Details** page, you can click **Edit** to modify the synchronization task or **Run** to run the task.

## Running a synchronization task

You can run a synchronization task manually or on a schedule:

- To run a synchronization task manually, on the **Explore** page, navigate to the task. In the row that contains the task, click **Actions** and select **Run**.
  
  You can also run a synchronization task manually from the **Task Details** page. To access the **Task Details** page, click **Actions** and select **View**.

- To run a synchronization task on a schedule, edit the task in the synchronization task wizard to associate the task with a schedule.
Rules and guidelines for running a synchronization task

Use the following rules and guidelines when you run a synchronization task:

- Verify that the source and target definitions are current. If the source or target no longer contains fields that are mapped in the field mapping, the synchronization task fails.

- You cannot run multiple instances of a synchronization task simultaneously. If you run a synchronization task that is already running, the synchronization task fails with the following error:

  Data synchronization task <Data Synchronization task name> failed to run. Another instance of the task is currently executing.

If you configured the synchronization task to run on a schedule, increase the time interval between the scheduled tasks to prevent multiple instances of the synchronization task from running simultaneously. If you run the synchronization task manually, wait for the currently running instance of the synchronization task to complete before starting it again.

You can view currently running synchronization tasks on the All Jobs or Running Jobs page in Monitor or on the My Jobs page in Data Integration.

- The synchronization task does not load any data into an IBM DB2 target if one or more records fails to load.

- When you use an active mapplet with a synchronization task that includes a saved query, the synchronization task ignores the configured target option for the task and inserts data to the target.
Chapter 4

Replication tasks

Use the replication task to replicate data to a target. You might replicate data to back up the data or perform offline reporting.

You can replicate data in Salesforce objects or database tables to databases or flat files. You can configure a task to replicate all rows of a source object each time the task runs or to only replicate the rows that changed since the last time the task was run. You can use a replication task to reset target tables and create target tables.

Load types

The load type determines the type of operation to use when the replication task replicates data from the source to the target.

Use one of the following load types when you replicate data:

- **Incremental loads after initial full load**
  The first time the replication task runs, it performs a full load, replicating all rows of the source. For each subsequent run, the replication task performs an incremental load. In an incremental load, the replication task uses an upsert operation to replicate rows that changed since the last time the task ran. You can specify this load type when the task uses a Salesforce source and a database target.

- **Incremental loads after initial partial load**
  The replication task always performs an incremental load with this load type. The first time the replication task runs, the replication task processes rows created or modified after a specified point in time. For each subsequent run, the replication task replicates rows that changed since the last time the task ran. You can specify this load type when the task uses a Salesforce source and a database target.

- **Full load each run**
  The replication task replicates all rows of the source objects in the task during each run. You can specify this load type when the task uses a Salesforce or database source and a database or flat file target.

For information about incremental load, see the Data Integration Salesforce Connector Guide.

**Full load**

For a full load, the replication task replicates the data for all rows of the source objects in the task. Each time the task runs, the replication task truncates the target database tables or flat file and performs a full data refresh from the source.
Run a full load in the following situations:

- The replication task uses a database source.
- A Salesforce object in the replication task is configured to be nonreplicateable within Salesforce. If you run an incremental load on a replication task that contains nonreplicateable objects, the replication task runs a full load on the object. Contact the Salesforce administrator to get a list of replicateable Salesforce objects.
- The datatype of a Salesforce field changed. If the replication task detects a datatype change, you might need to reset the target table to create a table that matches the updated Salesforce object. Then run the replication task with full load to reload the data for all Salesforce objects included in the replication task. Alternatively, you can set the AutoAlterColumnType custom configuration property so that the target table column updates to match the Salesforce object. The AutoAlterColumnType property does not apply in certain situations, such as when the source and target datatypes are not compatible. For more information about the AutoAlterColumnType property, see the IData Integration Salesforce Connector Guide.

Replication task sources

You can use Salesforce and database sources in replication tasks. You can use database tables, aliases, and views as sources.

Replication task targets

You can replicate source data to database and flat file targets. The type of target affects how the replication task replicates data.

Replicate data to a database target

For a replication task configured with the full load type, the first time you run the task, the task creates the database tables. The replication task then writes the replicated data to the tables. During subsequent runs, the task truncates the database tables and then writes the source data to the tables.

For a replication task configured with the incremental load after initial full load type, the first time you run the task, the task creates the database tables. The replication task then writes the replicated data to the tables. During subsequent runs, the task performs an upsert operation.

For a replication task configured with the incremental load after initial partial load, the first time you run the task and for subsequent runs, the task performs an upsert to replicate source data from the specified period in time.
Replicate data to a flat file target

When you run a replication task for a flat file target for the first time, the replication task creates the flat files. The replication task then stores the files in the specified directory and writes the replicated data to the files. During subsequent runs, the replication task truncates the file and loads the data.

When the replication task includes flat file targets with multibyte data on Linux, the default locale must be UTF-8.

Reset a database target

To drop all of the target tables in the task, reset a relational target table in a replication task.

You might want to drop a target table when you want to re-create it based on the latest source object definition. When you reset target tables for a replication task, the task performs the following actions:

1. Drops all of the target tables included in the replication task from the database.
2. Sets the load type for the replication task to full load.

After you run the replication task to reset the target, you must run the replication task again to reload the data from all source objects included in the task. When you run the replication task after you reset the target, the replication task recreates each target table. The replication task then loads all of the data into the new table.

If the target table is damaged, the replication task might consistently fail to write to the target table. You might need to reset the relational target.

Resetting a target table

You can reset the relational target tables included in a replication task.

1. On the Explore page, navigate to the replication task.
2. In the row that contains the task, click Actions and select Reset Target.

Rules and guidelines for resetting a target table

Use the following rules and guidelines when you reset a target table:

- If you previously created indexes on a target table and reset the target table, the replication task drops the indexes and the target table. You must create the indexes again.
- If you try to reset a target table that does not exist in the target database, an error appears.
- The replication task drops the target table that was updated the last time the task ran. For example, if you change the prefix of the target table and do not run the replication task, the replication task resets the old target table.
Table and column names in a database target

The replication task replicates source objects and fields to target database tables and columns, respectively. In certain cases, the replication task does not give the target table and column names the same names as the source objects and fields.

The replication task might not give the same names in the following circumstances:

- You write replicated data to a database target and use a table name prefix.
  A table name prefix prevents you from overwriting database tables when you share a database account.

- You replicate case-sensitive data.
  When the replication task replicates data to a target database, it creates all table and column names in uppercase. If the target database is case sensitive, use uppercase table and column names when you query the database.

- You replicate objects with long object and field names.
  When a source object or field name contains more characters than the maximum allowed for the name in the target, the replication task truncates the table or column name in the target database.

Table name truncation

When you replicate a source object to a database, the replication task replicates the data to a database table with the same name as the source object.

If you replicate data to a database target and the length of the source object name exceeds the maximum number of characters allowed for the target table name, the replication task truncates the table name in the target database. It truncates the table name to the first X characters, where X is the maximum number of characters allowed for a table name in the target database.

Duplicate tables names from same replication task

If you replicate multiple source objects from the same replication task to the same database user account and truncation causes duplicate table names, the replication task replaces the last character of the duplicate table names with sequential numbers.

For example, the replication task contains the following Salesforce source objects:

```
TenLetters1234567890TenLettersXXX
TenLetters1234567890TenLettersYYY
TenLetters1234567890TenLettersZZZ
```

When you replicate the objects, the replication task creates the following truncated table names in the target database:

```
TenLetters1234567890TenLetters
TenLetters1234567890TenLetter1
TenLetters1234567890TenLetter2
```

Duplicate table names from different replication tasks

If you replicate multiple source objects with the same names from different replication tasks to the same database user account, the replication task creates one target table, and overwrites the table data each time you replicate one of the objects. If you run a full load, it overwrites the entire table. If you run an incremental load, it overwrites the changed rows.

To avoid overwriting tables, use a different target table name prefix for each replication task.
Column name truncation

If the length of the source field name exceeds the maximum number of characters allowed for a column name in a relational target, the replication task truncates the column name in the target database. It truncates the column name to the first X characters, where X is the maximum number of characters allowed for a column name in the target database.

For example, the replication task creates a column name in an Oracle database based on the following 40-character field name of a Salesforce object:

```
TenLetters1234567890TenLettersXXXXXXXXX
```

The replication task truncates the column name to the first 30 characters:

```
TenLetters1234567890TenLetters
```

If the truncation causes duplicate column names for the target table, the replication task replaces the last character of the duplicate column names with sequential numbers. The replication task also replaces the last character of duplicate table names from the same task.

Target prefixes

When you replicate data to a database table or flat file, the replication task names each database table or flat file based on the corresponding source object name.

By default, the replication task includes the target prefix SF_. For example, the default flat file name for the Account Salesforce object is `SF_ACCOUNT.CSV`. If you remove the default target prefix and do not specify another prefix, the replication task creates a flat file or database table with the same name as the corresponding source object.

You can use target prefixes to prevent overwriting data. For example, you and another user share a database user account. The other user ran a replication task on the Contact object from her Salesforce account. Her replication task created a database table named Contact in the shared database. You use no target prefix and run a replication task on the Contact object from your Salesforce account. The replication task overwrites the data in the existing Contact table with your data. If you use the SF_ prefix, the replication task creates a table named SF_CONTACT and does not overwrite the existing table named Contact.

Creating target tables

You can use Data Integration to create the database table for a target before you run the replication task. You might want to create the target table and then modify the table properties before the replication task loads the data into the table.

1. On the Explore page, navigate to the replication task.
2. In the row that contains the task, click Actions and select Create Target.
Replication task schedules

You can run a replication task manually or schedule it to run at a specific time or interval.

You might want to run a replication task manually for the following reasons:

- To verify that a replication task is configured properly.
- To replicate the data occasionally. You might not want to replicate data at regular intervals.

If you specify a schedule in the replication task wizard, you can select an existing schedule or create a schedule. You can include a repeat frequency to replicate the data at regular intervals.

If you remove a task from a schedule while the task is running, the task that is in progress completes and future tasks are cancelled.

Preprocessing and postprocessing commands

You can run preprocessing and postprocessing commands to perform additional jobs. The task runs preprocessing commands before it reads the source. It runs postprocessing commands after it writes to the target.

You can use the following types of commands:

- SQL commands. Use SQL commands to perform database tasks.
- Operating system commands. Use shell and DOS commands to perform operating system tasks.

If any command in the preprocessing or postprocessing scripts fail, the task fails.

Preprocessing and postprocessing SQL commands

You can run SQL commands before or after a task. For example, you can use SQL commands to drop indexes on the target before the task runs, and then recreate them when the task completes. Data Integration does not validate the SQL.

Use the following rules and guidelines when creating the SQL commands:

- Use any command that is valid for the database type. However, Data Integration does not allow nested comments, even if the database allows them.
- Use a semicolon (;) to separate multiple statements. Data Integration issues a commit after each statement.
- Data Integration ignores semicolons within comments. If you need to use a semicolon outside of comments, you can escape it with a backslash (\).

Preprocessing and postprocessing operating system commands

Data Integration can perform operating system commands before or after the task runs. For example, use a preprocessing shell command to archive a copy of the target flat file before the task runs on a UNIX machine.

You can use the following types of operating system commands:

- UNIX. Any valid UNIX command or shell script.
- Windows. Any valid DOS or batch file.
Configuring a replication task

Configure a replication task to replicate data from a source to a target. When you configure a replication task, you specify the source connection, target connection, and the objects to replicate.

A replication task can replicate data from one or more Salesforce objects or database tables. When you configure the task, you can replicate all available objects through the selected connection, or you can select objects for replication by including or excluding a set of objects. You can also exclude rows and columns from the replication task.

Configure a replication task to run full or incremental loads. Perform a full load to replicate all rows for each object. Perform an incremental load to replicate rows that are new or changed since the last time you ran the task.

Associate a schedule with a replication task to specify when and how often the task runs. If you remove the replication task from a schedule as the task runs, the task completes. The replication task cancels any additional task runs associated with the schedule.

To configure a replication task, use the replication task wizard to perform the following steps:

1. Complete the prerequisite tasks.
2. Create the replication task.
3. Configure the source.
4. Configure the target.
5. Optionally, exclude fields.
6. Optionally, configure data filters.
7. Optionally, configure a schedule and advanced options.

As you work through the task wizard, you can click Save to save your work at any time. When you have completed the wizard, you can click Finish to save and close the task wizard.

Rules and guidelines for configuring replication tasks

Use the following rules and guidelines for configuring replication tasks:

• The names of source tables and fields can contain at most 79 characters.
• Multiple replication tasks cannot write to the same database table or flat file.
• You cannot configure a replication task when the source and target object are the same. If the source and target connections are the same, you must enter a target prefix to distinguish the source and target objects.

Replication prerequisite tasks

Complete the following prerequisite tasks before you create a replication task:

1. Verify that a database target exists.
   
   To replicate data to a database target, the database target must exist before you create the replication task. If a database does not exist, the database administrator must create a target database. The database must meet the minimum system requirements.

2. Create database users.
   
   To replicate data to a database target, the database administrator must create a database user account in the target database. Each database user account must have the CREATE, DELETE, DROP, INSERT,
SELECT, and UPDATE privileges. You must have a database user account for each replication task that writes to that database. You can use the same database user account for multiple replication tasks. If you use the same database user account for multiple groups, ensure that the replication tasks do not overwrite data in the same target tables.

3. Create a directory for the flat files.
   - To replicate data to a flat file, create a directory to store the flat files.

4. Optionally, create a schedule.
   - To run replication tasks at specified times or on regular intervals, create a schedule.

### Defining a replication task

Define a replication task using the replication task wizard.

1. To create a replication task, click **New > Tasks**. Select **Replication Task** and then click **Create**.
   - To edit a replication task, on the **Explore** page, navigate to the task. In the row that contains the task, click **Actions** and select **Edit**.

2. In the Replication Task Details, configure the following properties:

<table>
<thead>
<tr>
<th>Property</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task Name</td>
<td>Name of the replication task. Task names can contain alphanumeric characters, spaces, and the following special characters: _, +. Maximum length is 100 characters. Task names are not case sensitive.</td>
</tr>
<tr>
<td>Location</td>
<td>Project folder in which the task resides. If the <strong>Explore</strong> page is currently active and a project or folder is selected, the default location for the asset is the selected project or folder. Otherwise, the default location is the location of the most recently saved asset.</td>
</tr>
<tr>
<td>Description</td>
<td>Description of the task. Maximum length is 255 characters.</td>
</tr>
</tbody>
</table>

3. Click **Next**.

### Configuring the source

Configure the source on the **Source** page of the replication task wizard.

**Note:** Column names of a database source must not contain spaces or hyphens.

If you replicate a source with a name that includes a dollar sign ($), the replication task replaces the dollar sign with an underscore (_) in the target name.

1. In the **Source Details** area, select a connection.
   - To create a connection, click **New**. To edit a connection, click **View**, and in the **View Connection** dialog box, click **Edit**.

2. To select the objects to replicate, select one of the following options:
   - **All Objects**. To replicate all objects in the database or Salesforce account, select **All Objects**.
• **Include Objects.** To select the objects you want to include, click Select. In the Include Source Objects dialog box, select the objects to replicate and click Select.

• **Exclude Objects.** To select the objects you want to exclude, click Select. In the Exclude Source Objects dialog box, select the objects to exclude and click Select. The task replicates all available objects except for the selected objects.

The Available Objects area displays up to 200 objects. If the objects that you want to use do not display, enter a search string to reduce the number of objects that display.

When you select an object, it displays in a list. To remove a selected object, select the object and press Delete.

3. If you want the replication task to stop processing when it encounters an error, click Cancel processing the remaining objects.

If you want the replication task to continue to process a task after it encounters an error, click Continue processing of the remaining objects.

By default, the replication task stops processing the task when it encounters an error.

4. To display technical names instead of business names for some source types, click Display technical names instead of labels.

5. Click Next.

### Configuring the target

1. On the **Target** page, enter the following information:

<table>
<thead>
<tr>
<th>Property</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Connection</strong></td>
<td>Connection to the target object. To create a connection, click New. To edit a connection, click View, and in the View Connection dialog box, click Edit.</td>
</tr>
<tr>
<td><strong>Target Prefix</strong></td>
<td>Prefix that is added to Salesforce object names to create the flat file names or table names in a target database. By default, the prefix is SF_.</td>
</tr>
<tr>
<td><strong>Load Type</strong></td>
<td>Type of load. Select one of the following options: - <strong>Incremental loads after initial full load.</strong> Loads all data the first time the task runs. In subsequent runs, loads changed data only. - <strong>Incremental loads after initial partial load.</strong> Loads data created or modified after a specified period in time. If you select this option, enter the date and time, for example, August 29, 2015 at 2:00. The replication task uses the time zone that is set for the user. If the server on which the data resides is located in a different time zone, adjust the date and time accordingly. For example, the time zone for the user is Pacific Time and the time zone for the server is Eastern Time, which is three hours ahead of Pacific Time. The user wants the initial load to replicate data modified on the server after August 29, 2015 at 2:00 AM. Because the user’s time zone is Pacific Time, the user specifies August 28, 2015 and 11:00 PM. - <strong>Full Load each run.</strong> Loads all data every time the task runs. This option is enabled for tasks with a Salesforce source and a relational target. For all other tasks, the replication task performs a full load.</td>
</tr>
</tbody>
</table>

72 Chapter 4: Replication tasks
<table>
<thead>
<tr>
<th>Property</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Delete Options</td>
<td>Select one of the following options:</td>
</tr>
<tr>
<td></td>
<td>- Remove Deleted Columns and Rows. Deletes columns and rows from the target if they no longer exist in the source.</td>
</tr>
<tr>
<td></td>
<td>- Retain Deleted Columns and Rows. Retains columns and rows in the target that were removed from the source.</td>
</tr>
<tr>
<td>Commit Size</td>
<td>Number of rows to commit.</td>
</tr>
<tr>
<td></td>
<td>Default for full load replication is 5,000 rows. Default for incremental load replication is 999,999,999.</td>
</tr>
</tbody>
</table>

2. Click Next.

**Configuring the field exclusions**

To limit the fields loaded in to a target, configure field exclusions for each source object. By default, the replication task loads all fields in to the target.

1. On the Field Exclusion page, click Exclude Fields.
2. In the Field Exclusion dialog box, select the source object that you want to use.
3. In the Included Fields list, select and move the fields that you want to exclude to the Excluded Fields list.
4. Click OK.
   The excluded fields display in the Field Exclusion page. To remove an excluded field, click Delete next to the field.
5. Click Next.

**Configuring the data filters**

By default, the replication task replicates all source rows to the target. To filter source rows that are replicated, configure data filters. If you replicate multiple source objects, create a different set of data filters for each object.

1. On the Data Filters page, enter the following details:

<table>
<thead>
<tr>
<th>Property</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Row Limit</td>
<td>Select one of the following options:</td>
</tr>
<tr>
<td></td>
<td>- Process all Rows. Replicates all rows of the source.</td>
</tr>
<tr>
<td></td>
<td>- Process Only the First... Rows. Replicates the first X rows, where X is the number of rows. You might choose to process the first set of rows to test the task.</td>
</tr>
<tr>
<td></td>
<td>You cannot specify a row limit on replication tasks with non-Salesforce sources. If you select a non-Salesforce source, the option is disabled.</td>
</tr>
<tr>
<td>Data Filters</td>
<td>Click New to create a data filter on a Salesforce or database source. You can create simple or advanced data filters.</td>
</tr>
</tbody>
</table>

2. Click the Delete icon next to the data filter to delete the filter.
3. Click Next.
Configuring a schedule and advanced options

Configure a schedule and advanced options for a replication task on the Schedule page of the task wizard.

1. On the Schedule page, choose whether to run the task on a schedule or without a schedule.
2. To run a task on a schedule, click Run this task on schedule and select the schedule you want to use.
   To create a new schedule, click New. Enter schedule details and click OK.
   To remove the task from a schedule, click Do not run this task on a schedule.
3. Configure email notification options for the task.
4. Optionally, enter the advanced options as required.

<table>
<thead>
<tr>
<th>Advanced option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preprocessing Commands</td>
<td>Commands to run before the task.</td>
</tr>
<tr>
<td>Postprocessing Commands</td>
<td>Commands to run after the task completes.</td>
</tr>
<tr>
<td>Maximum Number of Log Files</td>
<td>Number of session log files and import log files to retain. By default, Data Integration stores each type of log file for 10 runs before it overwrites the log files for new runs.</td>
</tr>
</tbody>
</table>

5. Choose whether to run the task in standard or verbose execution mode.
   If you select Verbose mode, the mapping generates additional data in the logs that you can use for troubleshooting. It is recommended that you select verbose execution mode only for troubleshooting purposes. Verbose execution mode impacts performance because of the amount of data it generates.
6. Click Finish.

Viewing replication task details

You can view details about a replication task, including the load criteria, source and target connections, field exclusions, data filters, and the associated schedule.

1. On the Explore page, navigate to the task.
2. In the row that contains the task, click Actions and select View.
   On the Task Details page, you can click Edit to modify the replication task or Run to run the task.

Running a replication task

You can run a replication task in the following ways:

- Manually. To run a replication task manually, on the Explore page, navigate to the task. In the row that contains the task, click Actions and select Run.
  You can also run a replication task manually from the Task Details page. To access the Task Details page, click Actions and select View.
• On a schedule. To run a replication task on a schedule, edit the task in the replication task wizard to associate the task with a schedule.

Rules and guidelines for running a replication task

Use the following guidelines when you run a replication task:

• You cannot run multiple instances of a replication task simultaneously. If you run a replication task that is already running, the replication task fails with the following error:

Data replication task failed to run. Another instance of task <Data Replication task name> is currently replicating the same objects.

If you configured the replication task to run on a schedule, increase the time interval between the scheduled jobs to prevent multiple instances of the replication task from running simultaneously. If you run the replication task manually, wait for the currently running instance of the replication task to complete before starting it again.

• If you run a replication task that excludes fields and that writes to a database target, the replication task drops any indexes defined on the excluded fields in the target table.

• If you replicate timestamp data, the replication task truncates the millisecond portion of the timestamp data.
Mass ingestion tasks

Use mass ingestion tasks to transfer a large number of files of any file type between on-premises and cloud repositories and to track and monitor file transfers.

When you create a mass ingestion task, you define the source from which to transfer files and the target to which to transfer the files. You can define a schedule by which the task runs.

Mass ingestion task sources

You can transfer files from on-premises and cloud sources that Mass Ingestion supports to any Mass Ingestion-supported targets.

Mass Ingestion supports the following sources:

- Local folder
- Advanced FTP
- Advanced FTPS
- Advanced SFTP
- Amazon S3 V2
- Microsoft Azure Blob Storage V3

Mass ingestion task targets

You can transfer files from any Mass Ingestion-supported sources to on-premises and cloud targets that Mass Ingestion supports.

Mass Ingestion supports the following targets:

- Local folder
- Advanced FTP
- Advanced FTPS
- Advanced SFTP
- Amazon S3 V2
- Amazon Redshift V2
Mass ingestion task configuration

Use the task wizard to configure mass ingestion tasks.

To configure a mass ingestion task, you perform the following actions:

1. Define the task.
2. Configure the source.
3. Configure the target.
4. Optionally, configure a schedule.

As you work through the task wizard, you can click **Save** to save your work at any time. When you have completed the wizard, click **Finish** to save the task and close the task wizard.

Before you begin, verify that the conditions that are described in “Before you begin” on page 77 exist.

**Before you begin**

Before you create mass ingestion tasks, verify that the following conditions exist:

- The organization has the following licenses:
  - Mass Ingestion
  - Mass Ingestion Runtime
- The Mass Ingestion application is running on the Secure Agent.
- Source and target connections exist, based on the sources from where you want to transfer files and the targets to where you want to transfer files.

**Defining a mass ingestion task**

1. To define a mass ingestion task, click **New > Tasks**. Select **Mass Ingestion** and then click **Create**.

To edit a mass ingestion task, on the **Explore** page, navigate to the task. In the row that contains the task, from the **Actions** menu, select **Edit**.

The task wizard appears.
2. In the **Definition** page of the task wizard, configure the following properties:

<table>
<thead>
<tr>
<th>Property</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task Name</td>
<td>Name of the mass ingestion task. The names of mass ingestion tasks must be unique within the organization. Task names can contain alphanumeric characters, spaces, and underscores. Names must begin with an alphabetic character or underscore. Task names are not case sensitive.</td>
</tr>
<tr>
<td>Location</td>
<td>Project folder in which the task resides.</td>
</tr>
<tr>
<td>Description</td>
<td>Optional description of the task. Maximum length is 1024 characters.</td>
</tr>
<tr>
<td>Runtime Environment</td>
<td>Runtime environment that contains the Secure Agent used to run the task. The Mass Ingestion application must run on the Secure Agent.</td>
</tr>
</tbody>
</table>

3. Click **Next**.

### Configuring the source

To configure the source, you select a connection or local directory from which to transfer files and then configure source options.

The options that appear on the **Source** page of the task wizard vary based on the type of source that you select.

For information about local folder sources, see "Configuring a local folder source" on page 78. For information about other sources, see the applicable connector guide.

### Configuring a local folder source

1. In the **Connection Type** list select **Local Folder**.
2. Configure the following **Source Options**:

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source Directory</td>
<td>Directory from where files are transferred. The Secure Agent must be able to access the directory.</td>
</tr>
<tr>
<td>Include files from sub-folders</td>
<td>Transfer files from all sub-folders under the defined source directory.</td>
</tr>
<tr>
<td>Skip Duplicate Files</td>
<td>Do not transfer duplicate files. If files with the same name and creation date were transferred by the same mass ingestion task, the task does not transfer them again, and the files are marked as duplicate in the job log. If this option is not selected the task transfers all files.</td>
</tr>
<tr>
<td>Option</td>
<td>Description</td>
</tr>
<tr>
<td>-------------------------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
</tbody>
</table>
| File Pattern            | File name pattern used to select the files to transfer. You can use the following wildcard character filters:  
                          - An asterisk (*) matches any number of characters.  
                          - A question mark (?) matches a single character.                                                                                     |
| After File Pickup       | Determines what to do with source files after the files transfer. The following options are available:  
                          - Keep the files in the source directory.  
                          - Delete the files from the source directory.  
                          - Rename the files in the source directory. You must specify a file name suffix that Mass Ingestion adds to the file name when renaming the files.  
                          - Archive the files to a different location. You must specify an archive directory.                                                  |

3. Click **Next**.
   
The Target page appears.

### Configuring the target

To configure the target, you select a connection or local directory to which to transfer files and then configure target options.

The options that appear on the **Target** page of the task wizard vary based on the type of target that you select.

For information about local folder targets, see "Configuring a local folder target" on page 79. For information about other targets, see the applicable connector guide.

### Configuring a local folder target

1. In the **Connection Type** list select **Local Folder**.
2. Configure the following **Target Options**:

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Target Directory</td>
<td>Directory to where files are transferred. The Secure Agent must be able to access the directory.</td>
</tr>
</tbody>
</table>
   | File Exists Action | Determines what to do with a file if a file with the same name exists in the target directory. The following options are available:  
                          - Overwrite  
                          - Append Timestamp                                                                                                                     |

3. Click **Next**.
   
The Schedule page appears.
Defining the schedule

You can run a mass ingestion task manually, or you can schedule the task to run at a specific time or when a file is ready.

1. On the Schedule page, select one of the following options:
   - Do not run this task on a schedule. The task does not run according to a defined schedule. You can run the task manually.
   - Run this task on schedule. Select a schedule by which the task runs.
   - Run this task by file listener. The task runs when the file listener that you select here notifies the task that files arrived at the source folder or when the files in the source folder are updated or deleted, based on the configuration of the listener. You must create a file listener that listens to the folder where files arrive. For more information about creating a file listener, see the section “File listeners” in the Mappings help.
     Applicable for local folder, Advanced FTP, Advanced SFTP, Advanced FTPS, Amazon S3 V2, and Microsoft Azure Blob Storage V3 sources.
2. Click Finish to save the task and close the task wizard.

Viewing mass ingestion task details

You can view details about a mass ingestion task, including the source and target connections and the associated schedule.

1. On the Explore page, navigate to the task.
2. In the row that contains the task, click Actions and select View.
   The Task Details page appears with task, source, target, and schedule details.
3. You can edit or run the task that you selected to view. On the Task Details page, click Edit to modify the task or click Run to run the task.

Running a mass ingestion task

You can run a mass ingestion task in the following ways:

1. To run a mass ingestion task manually, on the Explore page, navigate to the task. In the row that contains the task, click Actions and select Run.
   Alternatively, you can run the task manually from the Task Details page. To access the Task Details page, click Actions and select View. In the Task Details page, select Run.
2. To run a mass ingestion task on a schedule, edit the task in the mass ingestion task wizard to associate the task with a schedule.
Masking tasks

Use masking tasks to mask the sensitive fields in source data with realistic test data for nonproduction environments. You can choose to create a subset of the sensitive source data that reconciles object relationships.

When you configure a masking task, choose the source and target and then select a masking rule for each field in the source you want to mask. You can also use inplace masking to mask the data in the same system from which the masking task reads the data.

A data masking rule is a type of masking that you can apply to a selected field. The type of masking rule that you apply depends on the type of the field that you need to mask. You can select built-in rules for masking fields such as Social Security numbers, credit card numbers, phone numbers, and dates. You can apply substitution values for fields such as names, cities, countries, or positions. You can mask fields with random values or with repeatable values.

For example, you might need to test a Human Resources application. You need realistic employee data to test with. You can mask the fields in an Employee table to create the test data.

You can apply masking parameters to some data masking rules. Masking parameters are options that you can apply to customize the rules.

If the source and target locations are different, you can create a subset of the source data. Define data subset criteria to selectively process source data. For example, you can use a data subset to create a small environment for testing and development. You can define the type of data that you want to include in the subset database. The data subset retains foreign key relationships from the source data.

Rules and guidelines for masking tasks

Consider the following rules and guidelines when you run masking tasks:

- The Secure Agent must have access to Salesforce servers.

- Tasks that include data subset properties require a staging connection. You can create a staging connection on an H2 database.

- To improve batch processing, configure the EnableSalesForceStagingResponse flag in the Custom Configuration Details for the Secure Agent and set it to TRUE.

Bulk operations that contain large amounts of data to read in a single query might encounter a connection reset at regular intervals. The task might fail because of the connection reset. Improved batch processing reduces the chances of a connection reset during a task run.
Masking task options

You can configure different options when you create a masking task.

You can choose a single source object or multiple source objects. You can configure the task operation that you want to perform in the target. You can choose whether you want to perform inplace masking. You can set filters to create a subset of data. Apply data masking rules to source objects and configure the rule parameters. When you schedule to run a masking task, you can configure email notifications and advanced options.

Source objects

You can add a single object or multiple related objects in a masking task.

You can add a single object that does not contain any related objects. You can add multiple objects that have an explicit relationship defined in Salesforce. For example, if you use the Opportunity object as a source, you can add the related Campaign as well. You can also add the RecordType object because it is related to the Campaign object. All Salesforce objects in a multiple-object source must have a predefined relationship in Salesforce.

If you select multiple source objects, you can choose an object and add the related parent, child, and self-reference objects manually. When a source object references to itself within a task, it is called self-reference relationship.

When you select multiple source objects, you can download the subset graph from the Source page to view the relationships between the related objects.

Schema graph

A schema graph contains the graphical representation of the relationships between the multiple source objects in a masking task.

You can see the visual images of the multiple source objects, the relationships, and the assignments.

You can view the graph from the Source page or from the Data Filters page.

To view the object relationships on the local system, download the schema graph. The graph downloads in the following format:

```<filename>.dot```

The DOT language defines a graph and you can directly view the graph in the browser. If you download the DOT format file, you must use a graph visualizer tool to render, view, and manipulate graphs in the DOT language. You can download and install any open source graph visualizer tool, such as Graphviz, Canviz, Grappa, and Beluging.

Schema Graph Example

Consider that you select the following source objects in a task: Account, Case, and Contact. The parent object is Account, and the child objects are Case and Contact.
The following image shows a sample schema graph:

The path from Account to Case is the default path that the task takes after applying the filter. Then the task loads the Contact and Account parent records that are related to the Case object.

If you want to load all the contacts related to the filtered accounts and then the cases, you need to select the path from Account to Contact objects and then to the Case object. You must ensure that the subset is selected for all the objects that are included in the task. Edges marked in red indicate that the child record is selected from a parent object.

**Target task operations**

You can select a task operation that you want to perform in the target.

When the target is same as the source, you can only perform an update operation.

When the target is different from the source, you can select the following task operations:

**Insert**
- Ignores existing target data and inserts source data.

**Update**
- Updates data in the target location based on source data.

**Upsert**
- Updates existing target data. If data does not exist in the target, the masking task inserts data.
You can perform update, upsert, and insert operations on a partial sandbox. When you run an inplace masking task, you can perform an update operation. If the target is different from the source, you can perform insert and upsert operations.

**Inplace masking**

In inplace masking, you select the target same as the source.

When you configure a masking task, you can use inplace masking to mask the data in the same system from which the masking task reads the data. When you choose inplace masking, you can perform an Update operation in the target, but you cannot perform an Insert or an Upsert operation. You can apply data filters to a single source object when you select a Salesforce connection.

When you choose inplace masking in a masking task, the task does not create a custom field or an external ID for relationship reconciliation.

**Update partial sandbox**

You can update a partial sandbox if the record ID of the target sandbox matches the production or source record ID.

Salesforce upserts data based on the external ID values. A masking task creates fields for the external IDs in the target when you configure the task. If the external ID field is empty, Salesforce creates duplicate records. To avoid duplicate records in target for the upsert operation, you need the external IDs that the masking task creates. You can update the partial sandbox to load the external IDs to perform an upsert operation.

You can first update the partial sandbox to add the external IDs, and update the external ID values for the existing records in the target. Run the task with an upsert operation to update or insert records.

**Refresh fields**

You can refresh the fields whenever you change metadata in a masking task.

To refresh metadata, click **Refresh Fields** on the **Target** page.

You can refresh metadata when you make changes to the following fields in a masking task:

- Connection properties
- Source objects
- Metadata in Salesforce source
- Metadata in Salesforce target

**Validation reports**

You can view validation reports before you run a task.

When you click **Validation Report** on the **Target** page, the masking task validates whether the fields are visible in source and targets, the schema is proper, and then creates a validation report. The validation report is in plain ASCII text format.

The validation report lists mandatory relationships, fields, and field and relationship mismatch between source and target. The report lists the mandatory fields that are not available in the source that you need to fix. The report lists warning messages if the non-mandatory fields or relationships are not present in the source or target.

The validation report contains the following fields:
Object Name
The name of the source or target object.

Field Name
The name of the field in the source or the target object.

Data Type
The data type of the field.

Relationship
The relationship that is not present in the source or the target.

Staging database

You can perform data subset operations in a masking task. H2 is a database that you use to stage subsets of data.

The masking task uses the H2 database to stage record IDs, data subset for simple entities, entities with junction objects, multipath relationships, and lookup based relationships, and masking fields.

The H2 staging database installer is packaged with the Secure Agent. You can run the H2 package installer to install the staging connection. You can either start the staging database connection from the Schedule page of the masking task wizard, or manually run the H2 startup script.

You can view the subset record count in the logs. The task reads the record IDs that are selected for subset from the staging database and uses standard API to read those records from the source. The task uses standard or bulk API to load data into the target.

Start the staging connection

You can start the staging database connection from the Schedule page before you run a masking task.

Select a Secure Agent and start the staging database service. The task uses the staging database to store the data during the subset operations.

The following image shows the staging connection details:

![Configure Staging Connection](image)

You can view the JVM type and configure the JVM heap size for a staging connection. The JVM type specifies the type of JVM that the staging database uses. The JVM heap size specifies the heap memory for the staging database. Configure the JVM heap size to process large amount of data in the staging database.
in a short time. JVM type can be of 32-bit or 64-bit. When the JVM type is 32-bit, the default maximum heap size is 512 MB. When the JVM type is 64-bit, the default maximum heap size is 4096 MB. You can increase or decrease the heap size based on the amount of memory that the system can support.

You can select a Secure Agent and stop the staging database service.

**H2 database configuration requirements**

You must consider the maximum file size of the file system, cache size, and heap size when you configure the H2 database.

The H2 staging database has the following requirements:

- When you use a FAT or FAT32 file system, the data limit for H2 database is 4 GB.
- Configure the cache and heap size in the H2 startup script according to the source size and Secure Agent hardware configuration. An increase in the cache size improves staging, target load, and subset computation performance. Select the heap size based on the amount of physical memory that the system supports. The default cache size in the script is 2048 MB and the heap size is 4096 MB.
- The heap size must be at least twice the size of the cache memory.
- You can also configure the heap size for the staging connection on the Schedule page under Data Subset options.

The following example contains the section of code from the startup script file with the properties that you can configure:

```bash
@echo off
:: Script location
set H2_JAR_DIR=%~dp0
:: H2 Cache size in KBs
set H2_CACHE_SIZE=2097152
:: H2 Jar Name
set H2_JAR_PATH=%H2_JAR_DIR%\h2-1.3.176.jar
:: H2 DB Name
set H2_DB=dmask
:: JVM path
set JVM_PATH=%H2_JAR_DIR%\..\..\..\..\..\..\..\..\jre\bin
:: JVM Options. Initial and maximum heap size
set JVM_OPTS=-Xms128m -Xmx4g
```

You can change the H2_CACHE_SIZE value and the JVM_OPTS value to increase or decrease memory requirement for the H2 database. Higher memory allocation ensures better staging and subset computation performance.

**Note:** Do not make other changes to the script file. Changes to other properties can damage the file.

The H2 startup script is available in the following location:

```
<Agent installation directory>\apps\Data_Integration_Server\$Version\ICS\main\tomcat\cmask\h2_start.bat
```

**Installing and configuring H2 database manually on Windows**

After you install the Secure Agent, you must run the H2 staging database installer that is packaged with the Secure Agent. In Informatica Intelligent Cloud Services, you must create a connection to connect to the H2 database.

1. Browse to the following location:
   ```
   <Secure Agent installation directory>\apps\Data_Integration_Server\$version\ICS\main\tomcat\cmask
   ```
2. Run the h2_start.bat startup script.
   The database starts up and the Command Prompt displays the parameters that you need to configure
   the connection in Informatica Cloud. Keep the Command Prompt open.

3. Log in to Informatica Intelligent Cloud Services.

4. Perform one of the following steps:
   - In Administrator, select Connections.
   - In Data Integration, open a source or target object in a task.

5. Click New Connection.
   The New Connection page appears.

6. Enter the connection name for the H2 database.
   Connection names are not case sensitive. Connection names can contain alphanumeric characters,
   spaces, and the following special characters: _ . + -

7. Optionally, enter a description for the connection. Maximum length is 255 characters.

8. Select the JDBC_IC (Informatica Cloud) connection type.
   The JDBC_IC connection properties appears.

9. Enter the Secure Agent group that runs the masking task.

10. Enter the JDBC Connection URL that appears in the Command Prompt.

11. Enter the path to the JDBC jar directory that appears in the Command Prompt.

12. Enter the database schema.

13. Enter the user name and password that appears in the Command Prompt to connect to the H2 database.

14. To test the connection, click Test Connection.

15. To create the connection, click Save.

### Installing H2 database manually on Linux

You can install H2 database on Linux and configure the connection in Informatica Intelligent Cloud Services.

Provide read, write, and execute access to the following directory:

```
<Secure Agent installation directory>\apps\Data_Integration_Server\$version\ICS\main\tomcat\cmask
```

1. Browse to the following location:

```
<Secure Agent installation directory>\apps\Data_Integration_Server\$version\ICS\main\tomcat\cmask
```

2. To run the H2 startup script, enter the following command:

   ```
   nohup sh h2_start.sh &
   ```
   The H2 database starts up and lists all the parameter values that you need to configure the connection in Informatica Cloud.

   **Note:** To stop the database, identify the process ID and enter the following command:

   ```
   ps -ef | grep "h2", kill -9 processid
   ```

3. To view the parameter values, open nohup.out file in a text editor or run the following command:

   ```
   vi nohup.out
   ```

4. Log in to Informatica Intelligent Cloud Services.

5. Perform one of the following steps:
• In Administrator, select Connections.
• In Data Integration, open a source or target object in a task.

6. Click New Connection.
The New Connection page appears.

7. Enter the connection name for the H2 database.
   Connection names are not case sensitive. Connection names can contain alphanumeric characters, spaces, and the following special characters:
   `_.+`.

8. Optionally, enter a description for the connection. Maximum length is 255 characters.

9. Select the JDBC_IC (Informatica Cloud) connection type.
The JDBC_IC connection properties appears.

10. Enter the Secure Agent group that runs the masking task.

11. Enter the JDBC Connection URL that appears in the `nohup.out` file.

12. Enter the path to the JDBC jar directory that appears in the `nohup.out` file.

13. Enter the database schema.

14. Enter the user name and password that appears in the `nohup.out` file to connect to the H2 database.

15. To test the connection, click Test Connection.

16. To create the connection, click Save.

Data subset

You can extract a subset of data from the source and move to the target in a masking task.

The masking task maintains primary and foreign key relationships in the subset data and reconciles the object relationships in the subset data.

Configure the following data subset options from the Data Filters page:

Data Filters

The data filter that you want to apply on the source. You can create a simple or an advanced data filter for an object. You can apply a filter to a single object in a task. You can apply multiple filters on the same object. You can add one advanced filter in a task. You can also use the filter values from a parameter file and mention the file name in the task.

Relationship Behavior

You can configure relationships when you select multiple source objects. When you perform a data subset operation, the masking task selects all the parent records of an object to maintain referential integrity. The task selects the child record if you configure to include the child objects. You can configure relationships of the child objects after you apply a data filter. You can select the child objects that you want to include in the data subset. You can view and download the schema graph that shows the graphical representation of the relationships between the source objects.

You can view the number of join operations that are required to compute a subset operation. View the sequence in which the task selects the records to create a data subset.
**Subset Statistics**

You can view the subset statistics such as total number of rows, the number of subset rows, and the subset size for each source object. The source contains large amount of data, but the target in which you want to create a subset might not contain enough space. To evaluate the target size, you can estimate the data subset. After you estimate the subset, you can view the estimated target size on all the masking task pages. If the estimated target size is large, you can update the task and estimate the subset again.

You can estimate the data subset size for multiple source objects.

**Data subset options**

Configure the data subset options on the **Schedule** page. You can view the data subset options if you select multiple source objects.

Configure the following data subset options:

**Staging Connection**

The connection that the task uses to run the data subset operation.

**Source Lookup Batch Size**

The number of records to retrieve from the Salesforce source in one SOQL query when the task writes to the target. Uses Salesforce standard API because the standard API limit is higher than the bulk API. Enter a number between 10 and 200 based on the SOQL character limit restriction for Salesforce.

**Drop Staging Tables**

Drops the staging tables even if there are error rows in the task.

When you configure data subset filters and run the masking task, the task runs through the staging, subset computation, target load, and staging drop phases. By default, if there are error rows, the task does not drop the staging tables. You can correct the errors and restart the task. The task resumes from the phase at which it failed. The staged data that is saved in the Secure Agent machine consumes some storage space, and you can choose to drop the staging tables even if there are errors rows present.

**Automatic task recovery**

In a masking task, you can estimate a subset. When you run the task after subset estimation, the task recovers from the previous stage and continues to the next stage.

When you configure data subset filters and run a masking task, the task runs through the staging, subset estimation, target load, and staging drop stages.

You can estimate the subset to evaluate the target subset size. When you click **Estimate**, the task stages the records and estimates the subset. After you estimate the subset, if you click **Run**, the task resumes to load the target and then drops the staging tables.

If you save and run a task without estimating the subset, the task runs through all the stages and drops the staging tables at the end.

**Data Subset Task Reset**

You create a task, estimate a subset, and run the task. Every task has an associated staging schema. After the task runs through all the stages, the task drops staging tables if there are no errors. If you did not choose to drop the staging tables and if there are error rows present, the task does not drop the staging tables. If you run the same task after a few days, the task performs the data subset operation on the old data. To run the task with the updated data, you must first copy the task. After you copy the task, the task is associated to
another staging schema and the staging data is built again. You can estimate the subset or run the task that contains the updated data.

Parameter files in data filters

In a masking task, you can use user-defined parameters in simple and advanced data filters.

When you use a parameter in a filter, start the filter with the parameter. Use two dollar signs to name the parameter in the following format: $$<\text{parameter}>$

Save the parameter file local to the following directory:

$$<\text{Secure Agent installation directory}>/\text{apps/Data\_Integration\_Server/data/userparameters}$$

You can specify the parameter file name on the Schedule page of the task wizard. The parameter values are applied when the task runs.

Simple Filter Example

Consider that you apply a filter on the Account object. Configure the filter condition that the Created Date must be equal to $$\text{param}$$.

Then create a parameter file with the following content:

$$\text{param}= ('1991-10-03')$$

The following image shows a simple filter with the use of a parameter:

Advanced Filter Examples

Consider that you apply a filter on the Account object. In the Advanced Data Filter dialog box, you can specify the filter expression that the Account Name field must pick all the values from $$\text{param}$$.

Then create a parameter file with the following content:

$$\text{param}= ('Apple', 'Microsoft')$$
The following image shows an advanced filter with the use of parameter as a value:

![Advanced Data Filter](image1)

You can also specify a filter expression as a parameter. Enter $$param as the filter expression. Then create a parameter file with the following content:

```
$$param=Name IN ('Apple', 'Microsoft')
```

The following image shows an advanced filter with the use of a parameter as an expression:

![Advanced Data Filter](image2)

**Configure relationship behavior**

You can select the child objects that you want to include in the data subset.

When you define a filter criteria without configuring the child records selection, the task traverses through the object in the default path where the task includes required child records with minimal graph traversal. When you configure child records selection, the task follows the order in which you select the relationships. To configure child records selection, you can click Configure in the Relationship Behavior section on the Data Filters page.
You can view the number of join operations that are required to compute a subset operation. View the sequence in which the task selects the records to create a data subset.

The following image shows the child records that you can select for an object:

Data subset use cases for two objects

When you define a filter for an object in a task, the task selects a default path so that it can traverse through the entire graph at least once.

For every selected record, the task loads all the parent records to maintain referential integrity. You can configure relationship behavior to select child records from an object.

Consider that there are two objects Account and Contact. Account is the parent object, and Contact is the child object of Account. You can apply filter on the Account object or the Contact object. You can either use the default path that the task selects or configure the path. The number of paths traversed is minimal with the default path selection.

The number of subset rows is not the sum of all target rows that the task loads. The subset estimates show the number of unique records that are selected from all the relationships and loaded to target.

Case 1. Select the default path with filter on Account

Consider that you apply a filter on the Account object with the condition that the account name starts with the letter A.

The following image shows the data subset filter criteria that you can configure:

The task traverses the default path from Account to Contact through the relationship Account. The number of join operations to compute the subset with the default path selection is two.
The following image shows the graphical representation of the relationship between Account and Contact objects:

![Graphical representation of the relationship between Account and Contact objects]

The task first loads the records from the Account object on which the filter is applied. Then the task traverses through the default path from Account to Contact through the relationship Account and loads the records. To maintain referential integrity, the task traverses from Contact to Account through the relationship ParentAccount__r.

The following image shows the sequence in which the task selects the records:

<table>
<thead>
<tr>
<th>Source</th>
<th>Target</th>
<th>Relationship</th>
<th>Rows</th>
</tr>
</thead>
<tbody>
<tr>
<td>Account</td>
<td>Contact</td>
<td>Account</td>
<td>5</td>
</tr>
<tr>
<td>Contact</td>
<td>Account</td>
<td>ParentAccount__r</td>
<td>3</td>
</tr>
</tbody>
</table>

Based on the filter applied, the Account object has four rows. From Account to Contact, the Contact object has five rows through the relationship Account. From Contact to Account, the Account object has three rows through the relationship ParentAccount__r.

To view the number of subset rows, you estimate the subset. If there are common records from multiple join operations, the task updates the records. If there are new records, the task adds the subset rows. In this use case, though the Account object shows total seven rows, the task loads four subset rows that are unique for the Account object from all the relationships. The task loads five subset rows for the Contact object.

The following image shows the subset statistics that you can estimate in a task:

![Subset statistics]

**Case 2. Select the configured path with filter on Account**

Consider that you apply a filter on the Account object with the condition that the account name starts with the letter A.
The following image shows the data subset filter criteria that you can configure:

<table>
<thead>
<tr>
<th>Object</th>
<th>Filter by</th>
</tr>
</thead>
<tbody>
<tr>
<td>Account</td>
<td>Account Name Starts With 'A'</td>
</tr>
</tbody>
</table>

Join Operations: 4

Consider that you choose both the relationships Account and ParentAccount__r between Account and Contact objects. The number of join operations to compute the subset with the configured path is four.

The following image shows the graphical representation of the relationship between Account and Contact objects:

![Graphical representation of relationships](image)

The task first loads the records from the Account object on which the filter is applied. Then the task traverses through the selected paths from Account to Contact through both the relationships Account and ParentAccount__r. To maintain referential integrity, the task traverses from Contact to Account through both the relationships Account and ParentAccount__r.

The following image shows the selection sequence of the objects:

<table>
<thead>
<tr>
<th>Source</th>
<th>Target</th>
<th>Relationship</th>
<th>Rows</th>
</tr>
</thead>
<tbody>
<tr>
<td>Account</td>
<td>Contact</td>
<td>Account</td>
<td>5</td>
</tr>
<tr>
<td>Account</td>
<td>Contact</td>
<td>ParentAccount__r</td>
<td>5</td>
</tr>
<tr>
<td>Contact</td>
<td>Account</td>
<td>Account</td>
<td>4</td>
</tr>
<tr>
<td>Contact</td>
<td>Account</td>
<td>ParentAccount__r</td>
<td>3</td>
</tr>
</tbody>
</table>

Based on the filter applied, the Account object has five rows. From Account to Contact, the Contact object has ten rows with both the relationships Account and ParentAccount__r. From Contact to Account, the Account object has seven rows with the relationships Account and ParentAccount__r.

To view the number of subset rows, you estimate the subset. If there are common records from multiple join operations, the task updates the records. If there are new records, the task adds the subset rows. In this use case, though the Account object shows total 12 rows, the task loads five subset rows that are unique for the Account object from both the relationships. Though the Contact object shows total 10 rows, the task loads seven subset rows that are unique for the Contact object from both the relationships.
Case 3. Select the default path with filter on Contact

Consider that you apply a filter on the Contact object with the condition that the first names start with the letter A.

The following image shows the data subset filter criteria that you can configure:

<table>
<thead>
<tr>
<th>Object</th>
<th>Filter by</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contact</td>
<td>First Name Starts With 'A'</td>
</tr>
</tbody>
</table>

Join Operations: 2

Since the Account object is parent to the Contact object and the filter is applied on the Contact object, the task does not select a relationship for child record selection. To maintain referential integrity, the task traverses from Contact to Account through both the relationships Account and ParentAccount__r. The number of join operations to compute the subset with the default path selection is two.

The following image shows the graphical representation of the relationship between Account and Contact objects:

The task first loads the records from the Contact object on which the filter is applied. Then the task traverses through the paths from Contact to Account through the relationships Account and ParentAccount__r, and loads the records.

The following image shows the sequence in which the task selects the records:

<table>
<thead>
<tr>
<th>Source</th>
<th>Target</th>
<th>Relationship</th>
<th>Rows</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contact</td>
<td>Account</td>
<td>Account</td>
<td>12</td>
</tr>
<tr>
<td>Contact</td>
<td>Account</td>
<td>ParentAccount__r</td>
<td>1</td>
</tr>
</tbody>
</table>

Based on the filter applied, the Contact object has 12 rows. From Contact to Account, the Account object has 13 rows through the relationships Account and ParentAccount__r.
To view the number of subset rows, you estimate the subset. If there are common records from multiple join operations, the task updates the records. In this use case, though the Account object shows total 13 rows, the task loads 12 subset rows that are unique for the Account object from all the relationships. The task loads 12 subset rows for the Contact object.

The following image shows the subset statistics that you can estimate in a task:

<table>
<thead>
<tr>
<th>Source Object</th>
<th>Total Rows</th>
<th>Subset Rows</th>
<th>Subset Size (in K)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contact</td>
<td>103</td>
<td>12</td>
<td>24</td>
</tr>
<tr>
<td>Account</td>
<td>100</td>
<td>12</td>
<td>24</td>
</tr>
<tr>
<td>Total Size</td>
<td>203</td>
<td>24</td>
<td>40</td>
</tr>
</tbody>
</table>

Case 4. Select the configured path with filter on Contact

Consider that you apply a filter on the Contact object with the condition that the first names start with the letter A.

The following image shows the data subset filter criteria that you can configure:

<table>
<thead>
<tr>
<th>Object</th>
<th>Filter by</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contact</td>
<td>First Name Starts With 'A'</td>
</tr>
</tbody>
</table>

Join Operations: 6

Consider that you choose both the relationships Account and ParentAccount__r between Account and Contact objects. The number of join operations to compute the subset with the configured path selection is six.

The following image shows the graphical representation of the relationship between Account and Contact objects:

The task first loads the records from the Contact object on which the filter is applied. To maintain the referential integrity, the task traverses from Contact to Account through both the relationships Account and ParentAccount__r. Then the task traverses from Account to Contact through the configured paths for child records selection with the relationships Account and ParentAccount__r and loads additional records. To maintain referential integrity for the additional records, the task traverses from Contact to Account through both the relationships Account and ParentAccount__r.
Based on the filter applied, the Contact object has 15 rows. From Contact to Account, the Account object has 13 rows through both the relationships Account and ParentAccount__r. From Account to Contact, the Contact object has 16 rows through both the relationships Account and ParentAccount__r. From Contact to Account, the Account object has 17 rows through the relationships Account and ParentAccount__r.

To view the number of subset rows, you estimate the subset. If there are common records from multiple join operations, the task updates the records. In this use case, though the Account object shows total 30 rows, the task loads 14 subset rows that are unique for the Account object from both the relationships. Though the Contact object shows total 31 rows, the task loads 15 subset rows that are unique for the Contact object from both the relationships.

The following image shows the subset statistics that you can estimate in a task:

<table>
<thead>
<tr>
<th>Source Object</th>
<th>Total Rows</th>
<th>Subset Rows</th>
<th>Subset Size (in KBs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contact</td>
<td>103</td>
<td>15</td>
<td>50</td>
</tr>
<tr>
<td>Account</td>
<td>100</td>
<td>14</td>
<td>28</td>
</tr>
<tr>
<td>Total Size</td>
<td>203</td>
<td>29</td>
<td>58</td>
</tr>
</tbody>
</table>

**Data subset use cases for three objects**

Consider that you selected the Account, Case, and Contact objects in a task. Account is the parent object, and Case and Contact are the child objects of Account. The Contact object is also parent to the Case object. Multipath relationship exists between Account and Case objects.

The task uses the default path to select the records. You can also configure the path to select the records.
Case 1. Default path

The following image shows the default path that the task chooses:

Based on the default path selection, the task traverses to the Case object directly, and then traverses to the Contact object.

If you apply filter with Account Name equals to Bank of New York, the task marks the default minimal path Account -> Case to traverse through the complete graph. The task selects the Bank of New York account from the Account object and then marks the corresponding two records in the Case object based on the AccountID. The task traverses to the Contact object and selects the corresponding records based on ContactID. To maintain referential integrity, the task selects the additional account Accenture from the Account object.
Case 2. Configured path

The following images show the path that you configure:

If you configure to select the Contact -> Account and Case -> Account paths, the task first traverses to the Contact object and then traverses to the Case object.

If you apply filter with Account Name equals to Bank of New York, the task marks the corresponding ContactID in the Contact object based on the AccountID. For the selected ContactID in the Contact object, the task marks the corresponding Case ID in the Case object.

Data subset rows

When you perform a data subset operation, the task can return the number of target subset rows that is more than the applied filter criteria.

When you run a data subset operation, the task selects the records in an object based on the filter criteria. The filtered records of the object can contain child records. The child records might have references to other records from the same parent object. In such cases, the task loads additional records along with filtered record to maintain referential integrity.

Data subset rows example

Consider that there are Account, Case, and Contact objects. Account is the parent object, and Case and Contact are the child objects of Account. The Contact object is also parent to the Case object.
The following image shows the source objects and related objects:

![Source Details](image)

The following table shows the two sample cases and the respective contacts for the account ABC1:

<table>
<thead>
<tr>
<th>Case</th>
<th>Contact</th>
</tr>
</thead>
<tbody>
<tr>
<td>00002541</td>
<td>Victor</td>
</tr>
<tr>
<td>00002542</td>
<td>Jack</td>
</tr>
</tbody>
</table>

Apply a subset filter on the Account object to load the target with ABC1 accounts. When you run the masking task, the task first applies filter on the account to load ABC1. The account ABC1 contains case 00002541 and case 00002542. If you enable the child record selection from the Account object to the Case object, the task loads case 00002542 and case 00002541.

The case 00002541 refers to the contact Victor, and the case 00002542 refers to the contact Jack. The parent account of Jack is XYZ2. To maintain the referential integrity for the contact Jack with the case 00002542, the task loads the additional XYZ2 account. Though you applied filter to load the account ABC1, the task loads both ABC1 and XYZ2 accounts.

The task loads six subset rows, two rows from each Account, Contact, and Case object.
The following image shows the masking task activity log entries for a data subset operation:

![Image showing the masking task activity log entries](image)

**Apply masking rules**

You can apply a masking rule to a field from the Masking page.

You can select a rule from a list based on the data type of the source field. If some of the fields are different in source and target, the common fields are listed. The fields pick up the attributes, such as length, field type, and label, from the target connection.

When you select multiple source objects, the task lists fields from a single object at a time on the Masking page. Select the source objects individually if you want to apply masking rule to fields in different objects.

After you apply masking rules, you can configure masking rule properties. For each masking rule, you can configure preprocessing and postprocessing expressions.

You cannot apply masking rules to read-only objects.

**Masking rule assignments**

You can apply masking rules to the objects from the Masking page to mask the fields.

You can apply the masking rules to the objects based on the field data type. After you apply a masking rule to a field, you can configure the masking rule properties. You can either manually select the available data masking rules from the list for each field or assign the default masking rules to a set of fields at once. The masking task package contains default masking rules. To assign the default masking rules to the source objects, click **Default Assignment**.

You can clear default masking rule assignments and assign the rules manually. To delete a masking rule assignment, click **Clear Assignment**.
Add mapplets

Add mapplets in a masking task to mask the target fields.

Use passive mapplets to perform a masking task. Assign a mapplet rule to a source object. Map the source fields to the input fields of the mapplet, and map the output fields of the mapplet to the target fields.

You can add multiple mapplets to an object. You can also add multiple instances of a mapplet to multiple objects.

You can add multiple instances of a mapplet to a single object. Informatica Cloud appends a unique number to identify each instance of the mapplet. You must configure each instance to the object before you run the task.

You can use a mapplet that requires an extra connection to a relational database or a flat file. Before you add the mapplet, you must add the connection.

If the dictionary information for the mapplet is in a flat file, the flat file must be present in the following location:

`<Secure Agent installation directory>\apps\Data_Integration_Server\data`

If the lookup connection for the mapplet is a flat file connection, the connection name must be the name of the flat file.

The following image shows the mapplets that you can add to a masking task:

![Add Mapplets](image)

You cannot use active mapplets.

Target fields

You can view the common and missing mandatory fields from the **Masking** page.

Common fields list the fields that are common in both source and target. You can assign masking rules to all the common fields.

The missing mandatory fields list the target mandatory fields that are missing in the source. To mask the missing mandatory fields, you can configure an expression or specify a value in the expression builder.
For example, you need to populate the target with data for testing purposes. You create a mandatory field called AlternatePhone_c in the target Account object that is not present in the source. When you run a masking task, you need to populate the missing mandatory target field with a value. The task fails because the mandatory field is missing in the source. You can enter a specific value or configure an expression for the missing mandatory field to populate the target.

**Default masking rules package**

You can assign default masking rules to the target fields.

The masking task package contains files with the default masking rules. After you install the Secure Agent, you can view the default_rules.xml, fields.properties, and salesforce_default_values.properties files in the following location:

```
<Secure Agent installation directory>/apps/Data_Integration_Server$/version/ICS\mask
```

The default_rules.xml file contains the configured rule properties for each masking rule. The fields.properties file contains the default masking rules for all the fields in the objects. When you apply default masking rules to the common fields, the task picks the default rules from the default_rules.xml and fields.properties files.

The salesforce_default_values.properties file contains the default values for the target mandatory fields that are missing in the source.

You can edit these files to change default values or create rules for default assignment.

**Configure default rules parameters**

You can edit the default rules files and configure the parameters for default assignments.

The following table describes the parameters that you can configure in the default_rules.xml file:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>isSeeded</td>
<td>To configure repeatable output. Enter True or False. If you enter True, specify a seed value.</td>
</tr>
<tr>
<td>seedValue</td>
<td>A starting number to create repeatable output. Enter a number from 1 through 999.</td>
</tr>
<tr>
<td>keepCardIssuer</td>
<td>Masks a credit card number with the same credit card type. Enter True or False. If you enter False, specify the targetIssuer parameter.</td>
</tr>
<tr>
<td>targetIssuer</td>
<td>Masks the credit card numbers with the selected credit card type. You can enter the following credit card types: ANY, JCB, VISA, AMEX, DISCOVER, and MASTERCARD.</td>
</tr>
<tr>
<td>firstNameColumn</td>
<td>Name of the column to use as the first part of the email name. The email name contains the masked value of the column you choose.</td>
</tr>
<tr>
<td>firstNameLength</td>
<td>The maximum number of characters of the first name to include in the masked email addresses.</td>
</tr>
<tr>
<td>delimiter</td>
<td>Delimiter to separate the first name and the last name in masked email addresses. You can enter the following characters: . / - _ . If you do not want to separate the first name and last name in the email address, leave the delimiter blank.</td>
</tr>
<tr>
<td>lastNameColumn</td>
<td>Name of the column to use as the last part of the email name. The email name contains the masked value of the column you choose.</td>
</tr>
<tr>
<td>Parameter</td>
<td>Description</td>
</tr>
<tr>
<td>--------------------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>lastNameLength</td>
<td>The maximum number of the characters of the last name to include in the masked email address.</td>
</tr>
<tr>
<td>domainConstantValue</td>
<td>A domain string name to include in the masked email addresses.</td>
</tr>
<tr>
<td>useMaskFormat</td>
<td>Specifies if you want to use a mask format. Enter True or False.</td>
</tr>
<tr>
<td>maskFormat</td>
<td>Defines the type of character to substitute for each character in the source data. You can limit each character to an alphabetic numeric or alphanumeric character type. Use the following characters to define a mask format: A for alphabets, D for digits from 0 to 9, N for alphanumeric characters, X for any character, + for no masking, and R for the remaining characters in the string of any character type. R must appear at the end of the mask format.</td>
</tr>
<tr>
<td>useSrcFilter</td>
<td>Specifies if you want to skip masking some of the source characters. Enter True or False. If you enter True, you must specify the srcFilterOption and srcFilterStr parameters.</td>
</tr>
<tr>
<td>srcFilterOption</td>
<td>Defines a filter that determines which characters to mask in the source. Enter one of the following options: - Mask Only. Mask only the characters that you configure as source filter characters. - Mask All Except. Mask all characters except the characters you configure as source filter characters.</td>
</tr>
<tr>
<td>srcFilterStr</td>
<td>The source characters that you want to mask or the source characters that you want to skip masking. Each character is case-sensitive. Enter the source filter characters with no delimiters. For example, AaBbC.</td>
</tr>
<tr>
<td>useTargetFilter</td>
<td>Specifies if you want to limit the characters that can appear in the target. Enter True or False. If you enter True, you must specify the targetFilterOption and targetFilterStr parameters.</td>
</tr>
<tr>
<td>targetFilterOption</td>
<td>Defines a filter that determines which characters to use in target mask. Enter one of the following options: - Use Only. Limit the target to the characters that you configure as target filter characters. - Use All Except. Limits the target to all characters except the characters you configure as target filter characters.</td>
</tr>
<tr>
<td>targetFilterStr</td>
<td>The characters that you want to use in a mask or the characters that do not want to use in a mask, based on the values of target filter type. Each character is case-sensitive. Enter the target filter characters with no delimiters. For example, AaBbC.</td>
</tr>
<tr>
<td>useRange</td>
<td>Specifies whether you want to set a range for the masked data. Returns a value between the minimum and maximum values of the range depending on field precision. To define the range, configure the minimum and maximum ranges or configure a blurring range based on a variance from the original source value. You can configure ranges for string, date, and numeric data types.</td>
</tr>
<tr>
<td>minWidth</td>
<td>The minimum value of the range. You can specify the minimum width for date, string, and numeric data types.</td>
</tr>
<tr>
<td>maxWidth</td>
<td>The maximum value of the range. You can specify the minimum width for date, string, and numeric data types.</td>
</tr>
<tr>
<td>startDigit</td>
<td>Defines the first digit of the masked SIN.</td>
</tr>
<tr>
<td>Parameter</td>
<td>Description</td>
</tr>
<tr>
<td>-------------------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>startDigitValue</td>
<td>The value for the first digit of the masked SIN.</td>
</tr>
<tr>
<td>DicConn</td>
<td>The connection to the directory where the dictionary files are present. You must create a flat file connection with the directory that points to the dictionary files.</td>
</tr>
<tr>
<td>DicName</td>
<td>The dictionary that you want to select. The dictionary file must be present in the rdtmDir directory of the Secure Agent.</td>
</tr>
<tr>
<td>outputPort</td>
<td>The output port column from the dictionary.</td>
</tr>
<tr>
<td>useBlurring</td>
<td>Masks data with a variance of the source data if specify that you want to blur the target data.</td>
</tr>
<tr>
<td>blurringUnit</td>
<td>Unit of the date to apply the variance to. You can enter the following values: Year, Month, Day, Hour, Minute, or Second.</td>
</tr>
<tr>
<td>blurringOption</td>
<td>The unit of numeric blurring. Enter Fixed or Percent.</td>
</tr>
<tr>
<td>blurLow</td>
<td>The low boundary of the variance from the source. Enter the value for numeric and date data types.</td>
</tr>
<tr>
<td>blurHigh</td>
<td>The high boundary of the variance from the source. Enter the value for numeric and date data types.</td>
</tr>
<tr>
<td>expText</td>
<td>An expression that you can configure to mask the target data.</td>
</tr>
<tr>
<td>Preprocessing expression</td>
<td>An expression to define changes to make to the data before masking.</td>
</tr>
<tr>
<td>Preprocessing expression</td>
<td>An expression to define changes to make to the masked data before saving the data to the target.</td>
</tr>
</tbody>
</table>

**Schedule options**

You can run a masking task manually, or you can schedule the task to run at a specific time or at specified time intervals.

You must select the runtime environment that contains the Secure Agent to run the task. You can configure staging connections for a data subset operation. You can configure email notification options and advanced options before you run the task.

**Email notification options**

When you configure a masking task, you can set email notification options to receive the status of the task.

You can select the following email notification options:

*Use the default email notification options for my organization*

Use the email notification options for the organization.
Use custom email notification options for this task

Use the email notification options that you configure for the task. Use commas to separate the email addresses in a list. When you select this option, the masking task ignores email notification options for the organization.

You can send email to different addresses based on the status of the task:

- Failure Email Notification. Sends failure email notification when the task fails to complete.
- Warning Email Notification. Sends warning email notification when the task completes with errors.
- Success Email Notification. Sends success email notification when the task completes without errors.

Advanced options

You can configure the advanced options to optimize a masking task.

Configure the following advanced options:

Parameter File

Name of the file that contains the definitions and values of user-defined parameters used in the task.

Preprocessing Commands

Command to run before the task.

Postprocessing Commands

Command to run after the task completes.

Configuring a masking task

Configure a masking task to apply data masking to a source and to create a data subset.

To configure a masking task, perform the following steps:

1. Define the masking task.
2. Configure the source.
3. Configure the target.
4. Configure the data subset.
5. Define data masking rules.
6. Schedule the masking task.

When you configure a masking task, you can save your work after you enter all required properties. You can choose one of the following options:

- Save. Saves the masking task and keeps it open.
- Finish. Saves and closes the masking task.
- Cancel. Closes the masking task and discards changes made after the last save.
Prerequisites

Before you configure a masking task, perform the following tasks:

• Use API version 32.0 and above to perform the masking task.
• Set the EnableSalesForceStagingResponse flag to True in the runtime environment to improve batch processing.
• Disable triggers, validations, lookup filters, and workflow rules on the target object.
• Ensure that the source is synchronized with the target.
• Ensure that all the mandatory fields in the target are present in the source.
• Align the user profiles, and check the user permissions and the visibility of the objects in the target.

Step 1. Define the masking task

Create the masking task.

1. Open Data Integration and click New to open the New Asset window.
2. Select Tasks > Masking Task.
3. Click Create.
   The new task window opens on the Definition page.
4. Enter a masking task name and an optional description.
5. Click Browse and select a project or folder location to store the task.
6. Click Next.
   The Source page opens.

Step 2. Configure the source

To configure a source, select a connection on the Source page. You can edit a connection that you select. Alternatively, you can create another connection.

1. On the Source page, select a connection from the list of connections.
   The source must have a primary key.
2. Optional. Choose to edit the connection or create another connection.
   • To edit a connection, select the connection and click View. In the View Connection dialog box, click Edit and edit the connection details in the Edit Connection dialog box. Test the connection to verify that the connection is valid.
   • To create another connection, click New. In the New Connection dialog box, enter the connection information. Test the connection to verify that the connection is valid.
3. Choose a single source object or multiple source objects.
   • To select a single object from the list, click Single. Select a single source object from the list. You can preview the source object details in the Data Preview section.
   • To select many objects, click Multiple.
   A list of source objects appears.
   Note: Objects that you cannot update do not appear in the list of source objects. For example, you cannot update objects of type isSfidLookup(), isCreateable(), isupdateable(), or isreferenced(). You
cannot update objects with cyclic relationships. For example, you cannot update the User, Profile, Community, and Idea objects.

4. If you choose multiple source objects, perform the following steps:
   a. Click Add.
      The Select Source Object dialog box appears with a list of objects.
   b. Select a source object and click Select.
   c. Select the added object and click Add.
      The Select Related Objects dialog box appears.
   d. Select the related child, parent, or self-reference objects that you want to include in the source objects. The objects move to the list of selected objects.
   e. To add the related objects, click Select.
      You can view and download the schema graph in the DOT format to view the relationships between the related objects.

5. Click Next.
   The Target page opens.

Step 3. Configure the target

To configure a target, select a connection on the Target page. The target connection type must be the same as the source connection type.

1. On the Target page, you can choose to perform inplace masking or save the edited data to a different location.
   • To perform inplace masking, select Same as source. If you select the target same as the source, the Connection list is disabled. The Same as source check box is selected by default.
   • To save masked data to a different location, clear Same as source, and select a connection from the Connection list.

   The target object or objects are the same as the source.

2. Optional. You can choose to edit the connection or create another connection.
   • To edit a connection, select the connection and click View. In the View Connection dialog box, click Edit and edit the connection details in the Edit Connection dialog box. Test the connection to verify that the connection is valid.
   • To create another connection, click New. In the New Connection dialog box, enter the connection information. Test the connection to verify that the connection is valid.

3. From the Task Operation list, select the operation that you want to perform.
   On a partial sandbox, you can perform update, upsert, and insert operations. When you run an inplace masking task, you can perform an update operation. If the target is different from the source, you can perform insert and upsert operations.

4. Select a target field that can link to target records. Select an existing external ID, custom field, or unique field from the list, or create another target field. To create another target field, click Create. To add the target field, click Create.
   You can view the target field details and any errors or warnings. You can save the external IDs to perform another upsert operation.

5. If you change the source object or connection properties, click Refresh Fields.

6. If you want to validate the source and target fields, click Validation Reports.
7. Click Next.
The Data Filters page opens.

Step 4. Configure the data subset

To configure a subset operation, use the Data Filters and Relationship Behavior options on the Data Filters page. Skip this step if you do not want to create a data subset.

1. On the Data Filters page, click New to create a data filter.
   - To create a simple data filter, select an object, a field to filter by, and an operator. Enter the value you want to use, and click OK.
   - To create an advanced data filter, click Advanced. Select an object. Select the field and create the filter expression. Click OK.

You can apply filters to a single object within a task.

Note: Fields that you cannot update do not appear in the filter. You cannot apply filters on the following field types: TEXTAREA (RICH), TEXTAREA (RICH), and TEXT ENCRYPTED.

2. To configure child relationships, click Configure.
   The Configure Relationship Behavior dialog box appears.

3. Enable the child records that you want to include in the subset.

4. Click Save.
   Join Operation displays the number of join operations that are required to create the subset. You can view and download the schema graph in the DOT format to view the relationships between the objects.

5. Optional. Click View to view the sequence in which a task with multiple objects selects objects to create the data subset.

6. Click Next.
The Masking page opens.

Step 5. Define data masking rules

On the Masking page, choose the object and select masking rules to assign to each field in the target.

1. On the Masking page, select a source object to view the fields.
   The task lists the common fields and the missing mandatory fields.

2. To view information about a field in the source object, click Status.
   The field data type determines the masking rules that you can apply to it. Fields that you cannot mask do not display a masking rule list.

3. To assign a rule to a common field, select the rule in the Masking Rule list.
   If the rule you select requires additional parameters, a Configure button appears next to the rule.

4. To configure the masking rule properties, click Configure.
   Each masking rule can have different properties.

5. Configure the masking rule properties and click Save.
   When you select a mapplet rule, you must configure input and output fields of the mapplet.
6. To assign the default masking rules to the fields, click Default Assignment. To clear the masking rules assignment, click Clear Assignment.

7. To view and configure an expression for the mandatory fields that are missing in the source, click Missing Mandatory Fields.

8. In the Actions column, click Configure Expression and enter an expression in the expression builder. Click OK.

9. After you configure masking rules for all fields, click Next. The Schedule page opens.

**Step 6. Schedule the masking task**

Configure when to run the masking task from the Schedule page.

You can run the masking task manually or schedule it. You can schedule the masking task to run at a specific time or at specified time intervals.

**Running the masking task immediately**

You can run a masking task without scheduling it.

1. Click Explore to open the Explore page.

2. Choose to browse by projects or assets from the Explore list.
   Select Assets to view a list of all assets. Select Projects to view a list of all projects. You can then select a project to view the assets in the project.

3. You can run a task manually in one of the following ways:
   • Select the masking task that you want to run. Click the Actions icon and click Run.
   • Click to open the masking task that you want to run. In the task page, click Run.

You can view the progress and status of the job on the My Jobs page. You can also view and manage jobs from the All Jobs or Running Jobs page in Monitor.

**Scheduling the masking task**

Configure when to run the masking task from the Schedule page.

1. On the Schedule page, choose whether to run the masking task on a schedule or run manually without any schedule.

2. If you choose to run the task on a schedule, choose a schedule from the list or click New to create a new schedule.

3. To run the masking task, select the runtime environment that contains the Secure Agent to run the task.
   **Note:** You cannot use a cloud runtime environment to run a masking task.

4. If you selected multiple source objects, configure the data subset options. Select a staging connection that runs the data subset operation. Configure the staging connection and start the staging database service.

5. Optional. Select Drop Staging Tables to drop the staging tables even if there are error rows in the task.

6. Select an email notification option.

7. Optional. Configure the advanced options.

8. Optional. Configure the advanced Salesforce options if the API version is 32.0 and higher.

9. Click Save to save and keep the task open, or click Finish to save and close the task.
Masking task maintenance

You can view and maintain all masking tasks from the **Assets** list on the **Explore** page.

You can edit tasks, copy tasks, delete tasks, run tasks, estimate a data subset, and download XML mapping and validation reports. You can also edit permissions on a masking task.

**Editing a masking task**

You can edit a masking task if you want to make changes to the metadata.

1. Click **Explore** to open the **Explore** page.
2. Choose to browse by projects or assets from the **Explore** list.
   - Select **Assets** to view a list of all assets. Select **Projects** to view a list of all projects. You can then select a project to view the assets in the project.
3. Select the masking task that you want to edit. Click the **Actions** icon and click **Edit**.
   - You can edit the task information in the **Edit Task** window.

**Running a masking task manually**

You can start a masking task manually to run it immediately.

1. Click **Explore** to open the **Explore** page.
2. Choose to browse by projects or assets from the **Explore** list.
   - Select **Assets** to view a list of all assets. Select **Projects** to view a list of all projects. You can then select a project to view the assets in the project.
3. You can run a task manually in one of the following ways:
   - Select the masking task that you want to run. Click the **Actions** icon and click **Run**.
   - Click to open the masking task that you want to run. In the task page, click **Run**.
   - You can view the progress and status of the job on the **My Jobs** page. You can also view and manage jobs from the **All Jobs** or **Running Jobs** page in **Monitor**.

**Stopping a masking task**

You can view the progress and status of a job on the **My Jobs** page. You can also view and manage jobs from the **All Jobs** or **Running Jobs** page in **Monitor**.

To stop a masking task that is currently running, click the **Stop** icon in the row that contains the job.

**Configuring masking task permissions**

You can configure permissions for a masking task. Set permissions for users or groups to read, create, update, delete, and run the masking task. You can also allow other users or groups to change permissions for the masking task.

By default, the user group defines the objects that a user can access.

1. Click **Explore** to open the **Explore** page.
2. Choose to browse by projects or assets from the **Explore** list.
Select **Assets** to view a list of all assets. Select **Projects** to view a list of all projects. You can then select a project to view the assets in the project.

3. Select the masking task that you want to edit. Click the **Actions** icon and click **Permissions**.

   The **Task Permissions** page opens with a list of users and groups that have permissions on the task.

4. To remove users or groups, select the users or groups and click **Remove** and then click **Save**.

5. To edit the level of permission assigned, select the user or group, change the permission level as required and click **Save**.

6. To add users or groups, click **Add**.

7. Choose the users or groups from the list of users or groups in the organization and click **Add**.

   The users or groups are added to the list.

8. Configure the required level of permission for each user or group and click **Save**.

### Copying a masking task

You can copy a masking task to create another task with similar behavior.

When you copy a task, the masking task appends a number to the task name and creates a task name.

**Note:** When you copy a task with a schedule, the schedule is not copied.

1. Click **Explore** to open the **Explore** page.

2. Choose to browse by projects or assets from the **Explore** list.

   Select **Assets** to view a list of all assets. Select **Projects** to view a list of all projects. You can then select a project to view the assets in the project.

3. Select the masking task that you want to copy. Click the **Actions** icon and click **Copy To**.

4. Choose the project or folder where you want to create a copy of the masking task and click **Select**.

   You cannot create a copy of a task in the same location.

### Renaming a masking task

You can rename a masking task.

1. Click **Explore** to open the **Explore** page.

2. Choose to browse by projects or assets from the **Explore** list.

   Select **Assets** to view a list of all assets. Select **Projects** to view a list of all projects. You can then select a project to view the assets in the project.

3. Select the masking task that you want to rename. Click the **Actions** icon and click **Rename**.

4. In the **Rename Asset** window, enter a name for the masking task and click **Save**.

   The name of the masking task is updated.

### Deleting a masking task

You can delete a masking task at any time.

When you delete a task at run time, the masking task completes the task and then deletes it. You cannot retrieve a task after you delete it.

1. Click **Explore** to open the **Explore** page.
2. Choose to browse by projects or assets from the Explore list.
   Select Assets to view a list of all assets. Select Projects to view a list of all projects. You can then select a project to view the assets in the project.
3. Select the masking task that you want to delete. Click the Actions icon and click Delete.
4. Click Delete to confirm that you want to delete the task.

Exporting a masking task
You can export a masking task. You can use information in an exported masking task to analyze errors.

1. Click Explore to open the Explore page.
2. Choose to browse by projects or assets from the Explore list.
   Select Assets to view a list of all assets. Select Projects to view a list of all projects. You can then select a project to view the assets in the project.
3. Select the masking task that you want to export. Click the Actions icon and click Export.
4. Optional. Edit the job name if required.
5. Click Export.
   You can view and download the exported file from the My Import/Export Logs page. Click Export to view the export logs.

Downloading mapping XML
You can download a mapping XML file to view scope, control, and the workflows.
After you download the mapping XML, you can import the XML file into PowerCenter for debugging.

1. Click Explore to open the Explore page.
2. Choose to browse by projects or assets from the Explore list.
   Select Assets to view a list of all assets. Select Projects to view a list of all projects. You can then select a project to view the assets in the project.
3. Select the required masking task. Click the Actions icon and click Download XML.

Downloading validation reports
You can view validation reports after you run a masking task.
A validation report contains mandatory relationships, fields, and field and relationship mismatch between source and target.

1. Click Explore to open the Explore page.
2. Choose to browse by projects or assets from the Explore list.
   Select Assets to view a list of all assets. Select Projects to view a list of all projects. You can then select a project to view the assets in the project.
3. Select the required masking task. Click the Actions icon and click Download Validation Report.
Dictionary files for data masking

The masking task uses a set of built-in dictionary files or the custom dictionary files that you create. When you configure a substitution masking operation, you select a dictionary that contains substitute values. The masking task performs a lookup on the dictionary selected and replaces the source data with data from the dictionary.

When you install or upgrade the Secure Agent in a runtime environment, the masking task downloads and saves the dictionary files to the following location:

<Secure Agent installation directory>\apps\Data_Integration_Server\data

You cannot edit or rename these files, but you can change the content within the specified file structure.

**Note:** The data in the dictionary files is test data.

The masking task downloads the following dictionary files:

- informatica_mask_address.dic
- informatica_mask_cc_american_expression.dic
- informatica_mask_cc_diners_club.dic
- informatica_mask_cc_american_expression.dic
- informatica_mask_cc_diners_club.dic
- informatica_mask_cc_discover.dic
- informatica_mask_cc_jcb.dic
- informatica_mask_cc_master_card.dic
- informatica_mask_cc_visa.dic
- informatica_mask_countries.dic
- informatica_mask_email.dic
- informatica_mask_female_first_names.dic
- informatica_mask_first_names.dic
- informatica_mask_job_position.dic
- informatica_mask_last_names.dic
- informatica_mask_male_first_names.dic
- informatica_mask_states.dic
- informatica_mask_streets.dic
- informatica_mask_uk_ni.dic
- informatica_mask_us_telephone.dic
- informatica_mask_us_towns.dic
- informatica_mask_us_ssn.dic
- informatica_mask_us_zipcode.dic
- dict.csv
- defaultValue.xml

If you want to use a custom dictionary, you must add a connection to the custom flat file dictionary. If there are multiple Secure Agents in a runtime environment, you must also copy the custom dictionary file to the following location:

<Secure Agent installation directory>\apps\Data_Integration_Server\data
You can use any of the flat file formats, such as .txt, .dic, and .csv. There is no limit on the maximum number of fields that can be present in a flat file dictionary. All the fields must have a column header, and the fields must be separated by a comma. The structure within the file must contain a sequential number and a value separated by a comma. A file can contain more than two columns. When you configure a substitution rule with custom dictionaries, you can select the dictionary column. To support non-English characters, you can use different code pages from a flat file connection when you configure a substitution rule with custom dictionaries.

The following text is a sample from a flat file dictionary that you can use to mask the credit card numbers:

```
SN0, CC_JCB
1, 3500-0003-7382-2377
2, 3500-0092-0490-3652
3, 3500-0077-6261-9918
4, 3500-0039-3695-5973
5, 3500-0089-8551-0603
6, 3500-0064-5387-7207
7, 3500-0030-0361-1582
8, 3500-0042-8477-2366
```

**Consistent masked output**

You might want to use different tools to mask the source data.

You can use the following tools to generate the same masked output from the same source data:

- **Informatica Intelligent Cloud Services**
  - You can use the following tools in Informatica Intelligent Cloud Services:
    - Masking tasks on Informatica Intelligent Cloud Services
    - Mapping tasks that contain mappings with the data masking transformation on Informatica Intelligent Cloud Services
    - Mappings that contain the data masking transformation on Informatica Intelligent Cloud Services
- **Test Data Management (on-premise)**
- **PowerCenter mappings that contain the data masking transformation**

**Rules and guidelines**

Consider the following rules and guidelines before you run mappings, tasks, or workflows to generate consistent masked output:

- Use substitution masking rules to generate consistent masked output.
- The masking rules must use the same dictionaries.
- The Repeatable option must be set to ON.
- Use the same seed value.

Substitution masking rules use values from dictionaries to create masked output. The default dictionaries on Informatica Intelligent Cloud Services and on-premise Test Data Management are the same. When you use the same substitution rule, the workflow uses the same dictionary to substitute source data. The same seed value therefore ensures that the same substitute value is used for all rows provided the dictionaries are the same.

On Informatica Intelligent Cloud Services, the dictionary files are available at: `<Secure Agent installation directory>\apps\Data_Integration_Server\data`
In on-premise Test Data Management, the dictionary files are available at: `<Informatica installation directory>\server\infa_shared\LkpFiles`

The Repeatable option must be set to ON to ensure that the task or workflow repeats dictionary values for the same source value.

**Example**

Consider the following example:

The source data contains *First Name* and *Last Name* columns that you need to mask to ensure that you mask the full name in the target data.

You can use the following methods to generate the masked output:

- Run a masking task on Informatica Intelligent Cloud Services.
- Run a mapping task that contains a data masking transformation on Informatica Intelligent Cloud Services.
- Run a mapping that contains a data masking transformation on Informatica Intelligent Cloud Services.
- Run a data masking plan in Test Data Management.
- Run a PowerCenter mapping that contains a data masking transformation.

Perform the following high-level tasks to generate the masked output:

1. Use the Substitution Name masking rule to mask the *First Name* column. Set the Repeatable option to ON. Enter a seed value.
2. Use the Substitution Last Name masking rule to mask the *Last Name* column. Set the Repeatable option to ON. Enter a seed value.
3. Use the default dictionaries available with the setup. Do not make changes to the dictionaries.

When you run the masking task, mapping, or mapping task on Informatica Intelligent Cloud Services, Test Data Management workflow, or PowerCenter mapping to generate output, you generate the same masked output for the same source data.
Masking rules

A masking rule defines the logic that masks the data.

The type of masking rule that you can apply depends on the data type of the field that you need to mask. For example, if the field data type is numeric, you can define a masked value that is within a fixed or percent variance from the original value.

You can restrict the characters in a string to replace and the characters to apply in the mask. You can provide a range of numbers to mask numbers and dates.

### Masking rules

You can select a masking rule for each source field on the **Masking** page that you want to mask. The rules that you can select from depend on the data type of the field that you want to mask.

The following table describes the masking rules that you can apply to source fields:

<table>
<thead>
<tr>
<th>Masking rule</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Credit Card Masking</td>
<td>Masks a credit card number with a number that has a valid checksum. You can mask the string data type.</td>
</tr>
<tr>
<td>Email Masking</td>
<td>Masks an email address with a random email address. You can mask the string data type.</td>
</tr>
<tr>
<td>Email Advanced Masking</td>
<td>Masks an email address with a realistic email address from a first name, last name, and a domain name. You can mask the string data type.</td>
</tr>
<tr>
<td>IP Address Masking</td>
<td>Applies a built-in mask format to disguise an IP address. You can mask the string data type.</td>
</tr>
<tr>
<td>Key Masking</td>
<td>Creates repeatable results for the same source data. You can mask date, numeric, and string data types.</td>
</tr>
<tr>
<td>Nullification Masking</td>
<td>Changes input values in a field to null. You can mask date, numeric, and string data types.</td>
</tr>
<tr>
<td>Phone Masking</td>
<td>Masks a phone number with random numbers in the same format as the original number. You can mask the string data type.</td>
</tr>
<tr>
<td>Random Masking</td>
<td>Produces random, nonrepeatable results. You can mask date, numeric, and string data types.</td>
</tr>
<tr>
<td>SIN Masking</td>
<td>Masks a Social Insurance number. You can mask a string data type.</td>
</tr>
</tbody>
</table>
**Masking rule** | **Description**
---|---
SSN Masking | Masks a Social Security number with a valid number. You can mask the string data type.
Substitution Masking | Replaces a field with similar but unrelated data from a dictionary file. You can mask the string data type.
URL Masking | Masks a URL by searching for the ‘://’ string and masking the substring to the right of it. You can mask a string data type.
Custom Masking | Applies an expression to mask the target field. You can mask a string, numeric, and date data types.
Mapplet Masking | Applies masking rules from a PowerCenter mapplet. The mapplet contains the logic to mask the input fields and return data to the target.

## Repeatable output

Repeatable masking output returns deterministic values. Use repeatable masking when you run a masking task more than once and you need to return the same masked values each time it runs.

Configure repeatable output if you have the same value in multiple source tables and you want to return the masked value in all of the target tables. The tables in the target database receive consistent masked values.

For example, customer John Smith has two account numbers, 1234 and 5678, and the account numbers are in multiple tables. The masking task always masks the name John Smith as Frank Martinez, the account number 1234 as 6549, and the account number 5678 as 3214 in all the tables.

You can enter a seed value when you configure repeatable output. You can configure a dictionary file with data values that you can replace when you use substitution masking. When you configure repeatable output, the masking task returns the same value from the dictionary whenever a specific value appears in the source data.

### Seed value

Apply a seed value to create repeatable output for masking output. The seed value is a starting point for generating masked values.

You can define a seed value from 1 through 999. The default seed value is 190. Apply the same seed value to a field to return the same masked data values in different source data. For example, if you have the same Cust_ID field in four tables and you want all of them to generate the output with the same masked values, apply the same seed value when you mask each field.

You cannot use a seed value for random and custom masking rules.
Preprocessing and postprocessing expressions

When you configure masking rule properties, you can configure preprocessing and postprocessing expressions.

For each masking rule that you create, you can specify preprocessing and postprocessing expression parameters to apply changes before and after masking the data. A preprocessing expression defines the changes that you want to make to the data before masking. A postprocessing expression defines the changes that you want to make to the masked data before saving the data to the target. Use the expression builder to create an expression.

For example, the AccountName field can contain the name JOHN in uppercase and John in lowercase alphabetic characters. If you apply a masking technique and run the task, the task masks the same name with different values. If you want the same masked value in the target for both the names, you can apply preprocessing expression to convert all the uppercase alphabetic characters to lowercase. Use repeatable masking technique to mask with same values. To convert the masked values to uppercase alphabetic characters, apply postprocessing expression.

You cannot configure preprocessing and postprocessing expressions for custom masking.

Credit card masking

Credit card masking applies a built-in mask format to mask credit card numbers. You can create a masked number in the format for a specific credit card issuer.

The masking task generates a logically valid credit card number when it masks a valid credit card number. The length of the source credit card number must be from 13 through 19 digits. The input credit card number must have a valid checksum based on credit card industry rules.

The first six digits of a credit card number identify the credit card issuer. You can keep the original credit card issuer or you can select another credit card issuer to appear in the masking results.

The following examples of credit card numbers are in valid formats:

- 4567893453452
- 4567 8934 5345 2
- 4567-8934-5345-2
- 4567-8934-5345-2657
- 4567 8934 5345 2657

The following examples of credit card numbers are not in valid formats:

- 4563 x567 5674 7432
- 4563#7587 4666 9876

If the source does not contain a valid format, the task replaces with default values from the defaultValue.xml file.
Credit card parameters

You can mask a credit card with the same type of credit card number or you can replace the credit card number with a different type of credit card number.

The following table describes the parameters that you can configure for credit card masking:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Repeatable</td>
<td>Returns the same masked value when you run a task multiple times or when you generate masked values for a field that is in multiple tables.</td>
</tr>
<tr>
<td>Seed Value</td>
<td>A starting number to create repeatable output. Enter a number from 1 through 999. Default seed value is 190.</td>
</tr>
<tr>
<td>Keep Issuer</td>
<td>Returns the same credit card type for the masked credit card. For example, if the source credit card is a Visa card, generate a masked credit card number that is the Visa format.</td>
</tr>
<tr>
<td>Mask Issuer</td>
<td>Replaces the source credit card type with another credit card type. When you disable Keep Issuer, select which type of credit card to replace it with. You can choose credit cards such as AMEX, VISA, and MASTERCARD. Default is ANY.</td>
</tr>
</tbody>
</table>

Email masking

When you mask an email address, you can create a random email address or a realistic email address. Create a random email address by default.

When you configure email masking, you return random ASCII characters in the email address by default. For example, when you mask Georgesmith@yahoo.com, the results might be KtrIupQAPyk@vdSKh.BIC.

The source input data can contain alphanumeric characters, and the following special characters: _, ., and @.

The following examples of email addresses are valid formats:

- David_john.Waugh@gmail.com
- DavidJohn1988@yahoo.com

The following examples of email addresses are not in valid formats:

- david@john@gmail.com
- david$%^&john@gmail.com

If the source does not contain a valid format, the task replaces with default values from the defaultValue.xml file.

You can configure email masking as repeatable between tasks, or you can configure email masking to be repeatable in one task. To return realistic email addresses, configure advanced email masking.

Advanced email masking

You can create a realistic email address from a first name, last name, and domain name.

When you configure advanced email masking, you can configure parameters to mask the user name and the domain name in the email address. For example, a source table might contain columns called First_Name and Last_Name. You can configure the email address to contain the first character of First_Name, and seven
characters of the last name. Define a domain name for the email address. Masking creates an address with the following syntax:

VSingh@mycompany.com

The following table describes the parameters you can configure for advanced email masking:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Repeatable</td>
<td>Returns the same masked value when you run a task multiple times or when you generate masked values for a field that is in multiple tables.</td>
</tr>
<tr>
<td>Seed Value</td>
<td>A starting number to create repeatable output. Enter a number from 1 through 999. Default seed value is 190.</td>
</tr>
<tr>
<td>First Name</td>
<td>Name of the column to use as the first part of the email name. The email name contains the masked value of the column you choose.</td>
</tr>
<tr>
<td>First Name Length</td>
<td>The number of the characters of the first name to include in the email address.</td>
</tr>
<tr>
<td>Delimiter</td>
<td>Delimiter, such as a dot, hyphen, or underscore, to separate the first name and last name in the email address. If you do not want to separate the first name and last name in the email address, leave the delimiter blank.</td>
</tr>
<tr>
<td>Last Name</td>
<td>Name of the masked column to use in the email name. The email name contains the masked value of the column you choose.</td>
</tr>
<tr>
<td>Last Name Length</td>
<td>The number of the characters of the last name to include in the email address.</td>
</tr>
<tr>
<td>Domain Name</td>
<td>A string value that identifies represents an Internet Protocol (IP) resource such as gmail.com.</td>
</tr>
</tbody>
</table>

**IP address masking**

You can mask an IP address as another IP address.

The masking task masks an IP address as another IP address by splitting it into four numbers, separated by a period. The first number is the network. The masking task masks the network number within the network range.

The masking task masks a Class A IP address as a Class A IP Address and a 10.x.x.x address as a 10.x.x.x address. The masking task does not mask the class and private network address. For example, the masking task can mask 11.12.23.34 as 75.32.42.52. and 10.23.24.32 as 10.61.74.84.

You can configure repeatable output when you mask IP addresses. You must select **Repeatable** and enter a seed value.
Key masking

Key masking produces repeatable results for the same source data.

When you configure a field for key masking, the masking task creates a seed value for the field. The masking task uses the seed to create repeatable masking for the same source field values. Mask date, numeric, and string data types with key masking.

Key string masking

Configure key string masking to mask all or part of a string. To limit the masking output to certain characters, specify a mask format and result string replacement characters. If you need repeatable output, specify a seed value.

The following table describes the parameters that you can use with key masking:

<table>
<thead>
<tr>
<th>Masking parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Repeatable</td>
<td>Returns the same masked value when you run a task multiple times or when you generate masked values for a field that is in multiple tables.</td>
</tr>
<tr>
<td>Seed Value</td>
<td>A starting number to create repeatable output. Enter a number from 1 through 999. Default seed value is 190.</td>
</tr>
<tr>
<td>Mask Format</td>
<td>The type of character to substitute for each character in the source data. You can limit each character to an alphanumeric, numeric, or character type.</td>
</tr>
<tr>
<td>Filter Source</td>
<td>Determines whether to skip masking some of the source characters. Configure the Source Filter Type and the Source Filter Chars parameters when you enable this option. Default is disabled.</td>
</tr>
<tr>
<td>Source Filter Type</td>
<td>A filter that determines which characters to mask in the source. Use with the Source Filter Chars parameter. You must enable the Filter Source parameter to configure this parameter. Choose one of the following options:</td>
</tr>
<tr>
<td></td>
<td>- Mask Only. Mask only the characters that you configure as source filter characters.</td>
</tr>
<tr>
<td></td>
<td>- Mask All Except. Mask all characters except the characters you configure as source filter characters.</td>
</tr>
<tr>
<td>Source Filter Chars</td>
<td>The source characters that you want to mask or the source characters that you want to skip masking. Each character is case-sensitive. Enter the source filter characters with no delimiters. For example, AaBbC.</td>
</tr>
<tr>
<td>Filter Target</td>
<td>Determines whether to limit the characters that can appear in the target. Configure the Target Filter Type and the Target Filter Chars parameters when you enable this option. Default is disabled.</td>
</tr>
<tr>
<td>Target Filter Type</td>
<td>A filter that determines which characters to use in the target mask. Use with the Target Filter Chars parameter. You must enable the Filter Target parameter to configure this parameter. Choose one of the following options:</td>
</tr>
<tr>
<td></td>
<td>- Use Only. Limit the target to the characters that you configure as target filter characters.</td>
</tr>
<tr>
<td></td>
<td>- Use All Except. Limits the target to all characters except the characters you configure as target filter characters.</td>
</tr>
<tr>
<td>Target Filter Chars</td>
<td>The characters that you want to use in a mask or the characters that do not want to use in a mask, based on the values of target filter type. Each character is case-sensitive. Enter the target filter characters with no delimiters. For example, AaBbC.</td>
</tr>
</tbody>
</table>
Mask format

Configure a mask format to limit each character in the output field to an alphabetic, numeric, or alphanumeric character.

If you do not define a mask format, the masking task replaces each source character with any character. If the mask format is longer than the input string, the masking task ignores the extra characters in the mask format. If the mask format is shorter than the source string, the masking task does not mask the characters at the end of the source string.

When you configure a mask format, you must configure the source filter characters or target filter characters that you want to use the mask format with.

The following table describes mask format characters:

<table>
<thead>
<tr>
<th>Character</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Alphabetical characters. For example, ASCII characters a to z and A to Z.</td>
</tr>
<tr>
<td>D</td>
<td>Digits. From 0 through 9.</td>
</tr>
<tr>
<td>N</td>
<td>Alphanumeric characters. For example, ASCII characters a to z, A to Z, and 0-9.</td>
</tr>
<tr>
<td>X</td>
<td>Any character. For example, alphanumeric or symbol.</td>
</tr>
<tr>
<td>+</td>
<td>No masking.</td>
</tr>
<tr>
<td>R</td>
<td>Remaining characters. R specifies that the remaining characters in the string can be any character type. R must appear as the last character of the mask.</td>
</tr>
</tbody>
</table>

Source filter characters

Configure source filter characters to choose the characters that you want to mask.

When you set the a character as a source filter character, the character is masked every time it occurs in the source data. The position of the characters in the source string does not matter, and you can configure any number of characters. If you do not configure source filter characters, the masking replaces all the source characters in the field.

The source filter are case sensitive. The masking task does not always return unique data if the number of source string characters is fewer than the number of result string characters.

Target filter characters

Configure target filter characters to limit the characters that can appear in a target column.

Masking replaces characters in the target with the target filter characters. For example, enter the following characters to configure each mask to contain the uppercase alphabetic characters A through F:

```
ABCDEF
```

To avoid generating the same output for different input values, configure a wide range of substitute characters, or mask only a few source characters. The position of each character in the string does not matter.
Key numeric masking

You can configure key masking for numeric values and generate deterministic output.

When you configure a field for key numeric masking, you can select a seed value for the field. When the masking task masks the source data, it applies a masking algorithm that requires the seed. You can change the seed value for a field to produce repeatable results if the same source value occurs in a different field.

Key date masking

Key date masking produces repeatable results for the same source date. Date masking always generates valid dates.

You can change the seed value for a field to return repeatable datetime values between the fields.

Key date masking can mask dates between the years 1753 and 9999. If the source year is in a leap year, the masking task returns a year that is also a leap year. If the source month contains 31 days, the masking task returns a month that has 31 days. If the source month is February, the masking task returns February.

When you perform key date masking on Salesforce data, the masking task can insert dates up to 4000. If the masked date value is more than 4000, the masking task fails and throws exception.

Nullification masking

Nullification masking changes field input values to null.

You can mask String, Date, and Numeric data types. The appropriate masking type appears based on the field type. For example, if the field type is Date, the Date Nullification masking type appears.

You can perform insert operations in nullification masking. You cannot perform upsert and update operations in nullification masking. Use custom masking with upsert and update operations to mask the field input values with null values.

Phone number masking

You can mask phone numbers with random numbers.

The masking task masks a phone number without changing the format of the original phone number. For example, the masking task can mask the phone number (408) 382-0658 as (607) 256-3106.

The source data can contain numbers, spaces, hyphens, and parentheses. The following examples are valid formats of input phone numbers:

- 08040208950
- 080-4020-8950
- (080)-4020-3797
- (080)-(4020)-(3797)
- (080)-4020-3797
- (080) 4020 (3797)
Alphabetic and special characters are not masked. The following examples of input phone numbers are not in valid formats:

- x80-4020-8950
- x80-4020-x768
- x80-4020/789

To support additional formats, you need to make changes to the Data Masking transformation. If the input data contains alphabetic and special characters, the task replaces with default values from the defaultValue.xml file.

You can configure repeatable output when you mask phone numbers. You must select **Repeatable** and enter a seed value.

## Random masking

Random masking produces random, non-repeatable results for the same source data and masking rules.

Random masking does not require a seed value. The results of random masking are non-deterministic. Use random masking to mask string, numeric, and date data types.

## Random string masking

Configure random masking to generate random output for string data types.

To configure limitations for each character in the output string, configure a mask format.

The following table describes the parameters that you can use with random string masking:

<table>
<thead>
<tr>
<th>Masking parameters</th>
<th>Descriptions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mask Format</strong></td>
<td>The type of character to substitute for each character in the source data. You can limit each character to an alphanumeric, numeric, or character type.</td>
</tr>
<tr>
<td><strong>Filter Source</strong></td>
<td>Determines whether to skip masking some of the source characters. Configure the Source Filter Type and the Source Filter Chars parameters when you enable this option. Default is disabled.</td>
</tr>
</tbody>
</table>
| **Source Filter Type** | Defines a filter that determines which characters to mask in the source. Use with the Source Filter Chars parameter. You must enable the Filter Source parameter to configure this parameter. Choose one of the following options: 
  - Mask Only. Mask only the characters that you configure as source filter characters. 
  - Mask All Except. Mask all characters except the characters you configure as source filter characters. |
| **Source Filter Chars** | The source characters that you want to mask or the source characters that you want to skip masking. Each character is case-sensitive. Enter the source filter characters with no delimiters. For example, AaBbC. |
| **Filter Target**  | Determines whether to limit the characters that can appear in the target. Configure the Target Filter Type and the Target Filter Chars parameters when you enable this option. Default is disabled. |
Masking parameters | Descriptions
--- | ---
Target Filter Type | Defines a filter that determines which characters to use in the target mask. Use with the Target Filter Chars parameter. You must enable the Filter Target parameter to configure this parameter. Choose one of the following options:
- Use Only. Limit the target to the characters that you configure as target filter characters.
- Use All Except. Limits the target to all characters except the characters you configure as target filter characters.
Target Filter Chars | The characters that you want to use in a mask or the characters that do not want to use in a mask, based on the values of target filter type. Each character is case-sensitive. Enter the target filter characters with no delimiters. For example, AaBbC.

Random numeric masking

To mask numeric data, you configure a range of output values for a field.

The masking task returns a value between the minimum and maximum values of the range depending on field precision. To define the range, configure the minimum and maximum ranges or a blurring range based on a variance from the original source value.

The following table describes the parameters that you can configure for random masking of numeric data:

<table>
<thead>
<tr>
<th>Masking parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Range</td>
<td>A range that you want to set for the numeric data. Select the check box to enter a minimum and maximum range.</td>
</tr>
<tr>
<td>Minimum Range</td>
<td>The minimum value of the range.</td>
</tr>
<tr>
<td>Maximum Range</td>
<td>The maximum value of the range. The maximum value must be greater than the minimum value.</td>
</tr>
<tr>
<td>Blurring</td>
<td>A range of output values that are within a fixed variance or percent variance of the source data. Select the check box to enter blurring details.</td>
</tr>
<tr>
<td>Blurring Option</td>
<td>The unit of blurring. Select Fixed or Percent. Default is Fixed.</td>
</tr>
<tr>
<td>Low Bound</td>
<td>The low boundary of the variance from the source number.</td>
</tr>
<tr>
<td>High Bound</td>
<td>The high boundary of the variance from the source number.</td>
</tr>
</tbody>
</table>

Numeric blurring

To blur a numeric source value, select a fixed or percent variance, a high bound, and a low bound. The high and low bounds must be greater than or equal to zero.
The following table lists the masking results for blurring range values when the input source value is 66:

<table>
<thead>
<tr>
<th>Blurring type</th>
<th>Low</th>
<th>High</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed</td>
<td>0</td>
<td>10</td>
<td>Between 66 and 76</td>
</tr>
<tr>
<td>Fixed</td>
<td>10</td>
<td>0</td>
<td>Between 56 and 66</td>
</tr>
<tr>
<td>Fixed</td>
<td>10</td>
<td>10</td>
<td>Between 56 and 76</td>
</tr>
<tr>
<td>Percent</td>
<td>0</td>
<td>50</td>
<td>Between 66 and 99</td>
</tr>
<tr>
<td>Percent</td>
<td>50</td>
<td>0</td>
<td>Between 33 and 66</td>
</tr>
<tr>
<td>Percent</td>
<td>50</td>
<td>50</td>
<td>Between 33 and 99</td>
</tr>
</tbody>
</table>

Random date masking

Random date masking produces random, non-repeatable results for the same source date. You can assign a minimum and a maximum date for the results. You can also configure blurring to define a variance limit for the date results.

The following table describes the parameters that you can configure for random date masking:

<table>
<thead>
<tr>
<th>Masking parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Range</td>
<td>A range that you want to specify for the date data. Select the check box to set minimum and maximum values for a datetime value.</td>
</tr>
<tr>
<td>Minimum Range</td>
<td>The minimum value to return for the selected datetime value. The default datetime format is MM/DD/YYYY HH24:MI:SS.</td>
</tr>
<tr>
<td>Maximum Range</td>
<td>The maximum value to return for the selected datetime value. The maximum datetime must be later than the minimum datetime.</td>
</tr>
<tr>
<td>Blurring</td>
<td>Mask a date as a variance of the source date.</td>
</tr>
<tr>
<td>Blurring Unit</td>
<td>Unit of the date to apply the variance to. Select the year, month, day, or hour. Default is year.</td>
</tr>
<tr>
<td>Low Bound</td>
<td>The low boundary of the variance from the source date.</td>
</tr>
<tr>
<td>High Bound</td>
<td>The high boundary of the variance from the source date.</td>
</tr>
</tbody>
</table>

Date blurring

To blur a datetime source value, select a unit of time to blur, a high bound, and a low bound. You can select year, month, day, or hour as the unit of time. By default, the blur unit is year.

For example, to restrict the masked date to a date within two years of the source date, select year as the unit. Enter two as the low and high bound. If a source date is 02 February, 2006, the masking task returns a date between 02 February, 2004 and 02 February, 2008.
**SIN masking**

SIN masking applies a built-in mask format to change Social Insurance numbers.

You can mask a Social Insurance number that is nine digits. The digits can be delimited by any set of characters. The following delimiters are valid:

- space, no space, #, +, -, *, =, ~, !, @, $, %, ^, &, *, :, ;, ”, ,, /, and ,

If the number contains no delimiters, the masked number contains no delimiters. Otherwise the masked number has the following format:

```
xxx-xxx-xxx
```

The following examples of Social Insurance numbers are valid formats:

- 123456789
- 123 456 789
- 123-456-789

If the source does not contain a valid format, the task replaces with default values from the `defaultValue.xml` file.

You can define the first digit of the masked SIN. Enable **Start Digit** and enter the digit. The masking task creates masked SIN numbers that start with the number that you enter. You can configure repeatable masking for Social Insurance numbers. To configure repeatable masking for SIN numbers, select **Repeatable** and enter a seed value.

The following table describes the parameters you can configure for SIN masking:

<table>
<thead>
<tr>
<th>Masking parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Repeatable</td>
<td>Returns the same masked value when you run a task multiple times or when you generate masked values for a field that is in multiple tables.</td>
</tr>
<tr>
<td>Seed Value</td>
<td>A starting number to create repeatable output. Enter a number from 1 through 999. Default seed value is 190.</td>
</tr>
<tr>
<td>Start Digit</td>
<td>When enabled, you can define the first digit of the masked SIN.</td>
</tr>
<tr>
<td>Start Digit Value</td>
<td>The value for the first digit of the masked SIN.</td>
</tr>
</tbody>
</table>

**SSN masking**

SSN masking applies a built-in mask format to change Social Security numbers.

The SSN masking accepts any SSN format that contains nine digits. You can delimit the digits with any characters. The following delimiters are valid:

- space, no space, #, +, -, *, =, ~, !, @, $, %, ^, &, *, :, ;, ”, ,, /, and ,

For example, the SSN masking rule accepts the following format:

```
+=54-*9944$#789-,*(")
```

The following examples of Social Security numbers are valid formats:
If the source does not contain a valid format, the task replaces with default values from the `defaultValue.xml` file.

You can configure repeatable masking for Social Security numbers. You must select **Repeatable** and enter a seed value.

The masking task cannot return all unique Social Security numbers because it does not return valid Social Security numbers that the Social Security Administration has issued.

**Substitution masking**

Substitution masking replaces a column of data with similar but unrelated data from a dictionary file.

The masking task provides dictionaries that contain sample data for substitution masking. You can also use custom dictionaries. When you configure substitution masking, select the type of dictionary that contains the substitute values. The masking task performs a lookup on the dictionary that you choose and replaces source data with data from the dictionary.

You can substitute data with repeatable or non-repeatable values. When you choose repeatable values, you must configure a seed value to substitute data with deterministic results.

**Substitution masking with custom dictionaries**

Create and use custom dictionaries when you perform substitution masking.

You create and add a flat file dictionary to the masking task. Add a connection to the flat file dictionary from the **Configure | Connections** view.

When you configure a masking task, you can use the flat file dictionary connection to perform substitution masking. You cannot use a relational dictionary.

To support non-English characters, you can use different code pages from a flat file connection when you configure a substitution rule with custom dictionaries.
Substitution masking parameters

To perform substitution masking, you can select the custom dictionaries that you create or the dictionaries that the masking application provides.

The following table describes the parameters that you can configure for substitution masking:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dictionary File</td>
<td>Appears if you select Custom Substitution from the list of masking rules. Click Select and specify the following parameters:</td>
</tr>
<tr>
<td></td>
<td>Flat File Connection</td>
</tr>
<tr>
<td></td>
<td>The connection to the directory where the custom dictionaries are present. You must create a flat file connection with the directory that points to the dictionary files.</td>
</tr>
<tr>
<td></td>
<td>Dictionary File</td>
</tr>
<tr>
<td></td>
<td>The custom dictionary that you want to select. The dictionary file must be present for all the Secure Agents in a runtime environment in the following location:</td>
</tr>
<tr>
<td></td>
<td>&lt;Secure Agent installation directory&gt;\apps\Data_Integration_Server\data</td>
</tr>
<tr>
<td>Dictionary Column</td>
<td>The output column from the custom dictionary. Appears if you select Custom Substitution from the list of masking rules. You can select a dictionary column if the flat file contains column headers.</td>
</tr>
<tr>
<td>Lookup Input Port</td>
<td>Optional. The source input column based on which you perform a lookup operation on the dictionary.</td>
</tr>
<tr>
<td>Lookup Output Port</td>
<td>Optional. The dictionary column that you can look up based on the input port.</td>
</tr>
<tr>
<td>Lookup Error Constant</td>
<td>Optional. A constant value that you can configure when there are no matching values for the lookup condition from the dictionary.</td>
</tr>
<tr>
<td>Repeatable</td>
<td>Returns the same masked value when you run a task multiple times or when you generate masked values for a field that is in multiple tables.</td>
</tr>
<tr>
<td>Seed Value</td>
<td>A starting number to create repeatable output. Enter a number from 1 through 999. Default seed value is 190.</td>
</tr>
</tbody>
</table>

Custom substitution lookup example

Consider that you apply substitution masking on the S_City column and you select a dictionary file with city names, identification numbers, and serial numbers. Select CITY as the dictionary column. The lookup input port is Id and the lookup dictionary port is SNO. If there are no matching values between the Id and SNO columns, the task uses the error constant BANGALORE as the lookup value.
The following image shows the substitution parameters for masking with custom dictionaries:

Custom substitution dictionary lookup use cases

The task performs dictionary lookup in custom substitution masking in the following cases:

- Case 1. If there are valid target lookup records in a dictionary for all the corresponding source records, the task picks all the values from the dictionary and replaces in the target.
- Case 2. If there are some records in the source for which there are multiple lookup values in a dictionary, the task picks one of the lookup values from the dictionary and substitutes with the source value.
- Case 3. If some of the source values are same as the lookup values in a dictionary, the target contains the same data as the source.
- Case 4. If the source records do not have a lookup value in a dictionary and if you specify a valid error constant, the task uses the error constant for all the failed lookup conditions.
- Case 5. If the source records do not have a lookup value in a dictionary and if you do not specify a valid error constant, the task fails and generates an exception.

URL masking

You can configure URL masking to mask a source URL address.

The masking task parses a URL by searching for the ‘://’ string and parsing the substring to the right of it. The source URL must contain the ‘://’ string. The source URL can contain numbers and alphabetic characters.

The task cannot mask a URL without protocols, such as http://, https://, and ftp://. If the source does not contain a protocol, the task replaces with default values from the defaultValue.xml file.

The masking task does not mask the protocol of the URL. For example, if the URL is http://www.yahoo.com, the masking task can return http://MgL.aHjCa.VaD/. The masking task can generate a URL that is not valid.
You can configure repeatable output when you mask a URL address. You must select Repeatable and enter a seed value.

**Note:** The masking task always returns ASCII characters for a URL.

### Custom masking

Custom masking applies an expression to mask the target data. Use custom masking to mask string, numeric, and date data types.

When you apply custom masking to a field, click **Configure** and enter the expression. You can select the source fields, operators, and functions to build an expression. When you select a function, you can view the function description and the syntax.

You can concatenate data from multiple source fields to create a masked value for the target field. For example, you need to create a login name. The source has FirstName and LastName fields. Use substitution masking to mask the first and last names. In the Login field, configure an expression to concatenate the first letter of the first name with the last name:

```
CONCAT(SUBSTR(FirstName,1,1),LastName)
```

To mask field input values with null values, use custom masking. In the expression builder, enter single quotes separated by a space in the following format: ' '.

For more information about configuring expressions, see "Field expressions" on page 15

### Mapplet masking

You can assign a mapplet rule to the source fields to mask the output target fields.

A mapplet can contain multiple input and multiple output ports. A task fails if you do not configure any of the mapplet input or output ports that you add to a source object.

After you add a mapplet rule and assign the rule to a field, you must configure the mapplet parameters. Map the source fields to the input fields of the mapplet, and map the output fields of the mapplet to the target fields.

For example, an email mapplet contains the logic to concatenate the first name and last name of the source object to generate an email ID. Apply the email mapplet masking rule to the source fields. Map the FirstName3 input field of the source to the FirstName field of the mapplet. Map the LastName3 input field of the source to the LastName field of the mapplet. Map the Email output field of the mapplet to the Email3 field of the target.
You can use a passive mapplet that requires an extra connection to a relational database or a flat file. For example, mapplets that contain an SQL transformation, lookup transformation, or a data masking transformation that uses a dictionary connection. Before you add the mapplet, you must create the connection. When you configure a mapplet that requires an extra connection, you must configure the dictionary, SQL, or lookup connections. You select the connection reference based on the type of connection that the mapplet contains.

For example, you want to mask an account name with an AccountNameMapplet mapplet and the mapplet has connections to a dictionary and a relational database. After you add the mapplet and the connections, configure and assign the mapplet to the target. After you select the AccountNameMapplet mapplet, select the AccName_Lookup connection to perform the lookup operation. Select the AccName_Dict_Con connection to read the values from the dictionary connection. Map the Account Name input source field to the input mapplet field. Map the mapplet output port to the Account Name target field.
The following image shows the mapplet that contains the dictionary and lookup connections:

If the dictionary information for the mapplet is in a flat file, the flat file must be present in the following location:

   \<Secure Agent installation directory>\apps\Data_Integration_Server\data

If the lookup connection for the mapplet is a flat file connection, the connection name must be the name of the flat file.
PowerCenter tasks

Use the PowerCenter task to run a PowerCenter session in Data Integration.

To run a PowerCenter session in Data Integration, you create a workflow for the session in the PowerCenter Workflow Manager. You create a PowerCenter task in Data Integration. When you configure the task, you upload the PowerCenter XML file that contains the workflow.

If you want to make any changes to a session that is used in a PowerCenter task, you need to make the changes in PowerCenter. You can export the revised PowerCenter XML file and then edit the PowerCenter task to upload the updated XML file.

PowerCenter workflows

To use a PowerCenter workflow for a PowerCenter task, the workflow objects must be objects that the Data Integration PowerCenter task supports.

Consider the following rules when you use a PowerCenter workflow for a PowerCenter task:

- The PowerCenter XML file must only contain one workflow.
- The workflow must contain one Session task with one mapping.
- The workflow cannot include task types other than Session tasks.
- Do not edit the XML file after you've exported the workflow from PowerCenter. Instead, change the workflow in PowerCenter and then export it again.
- The session can contain up to 64 partitions for sources and targets.
- The session can use pre-session and post-session commands.
- The session must have the Enable High Precision session property enabled.
- The mapping must contain a source definition and target definition.
- The mapping cannot use parameters or parameter files.
- The mapping cannot contain an IIF expression with values of different datatypes, such as the following IIF expressions:
  
  ```
  [IIF(ANNUALREVENUE >0,NAME)]
  [IIF(emplid_offset = 'Y',LINE_NO + 1,LINE_NO]
  ```
- The mapping cannot include reusable objects such as reusable transformations or shortcuts because Data Integration doesn't use a repository like PowerCenter does, so reusable objects cannot be stored.
Supported transformations and mapping objects

The mapping objects in a workflow must be supported by Data Integration.

A mapping can include the following source and target types:

- Flat file
- FTP/SFTP
- Database
- Salesforce
- SAP
- Web service
- Most add-on connectors

To find out if the add-on connector you use supports PowerCenter tasks, see the appropriate Data Integration connector guide.

A mapping can include the following transformations:

- Aggregator transformation
- Data Masking transformation
- Expression transformation
- Filter transformation
- HTTP transformation
- Java transformation
- Joiner transformation
- Lookup transformation
- Normalizer transformation
- Router transformation
- Salesforce Lookup transformation
- Salesforce Picklist transformation
- Salesforce Merge transformation
- SAP IDOC Interpreter transformation
- SAP IDOC Prepare transformation
- Sequence Generator transformation
- Sorter transformation
- Stored Procedure transformation
- Transaction Control transformation
- Union transformation
- Update Strategy transformation
- Web Services Consumer transformation
- XML Parser transformation with file or database sources
- XML Generator transformation with file or database sources

If the workflow contains transformations or mapping objects other than the objects listed above, the workflow upload to Data Integration might fail.
Exception handling in stored procedures

When a mapping that you want to use in a PowerCenter task contains a Stored Procedure transformation, the stored procedure must include exception handling. Exception handling can be as complex as necessary. Or, you can use the following simple example:

```sql
Exception
when NO_DATA_FOUND
then NULL;
END;
```

For example, you have the following stored procedure in a PowerCenter workflow:

```sql
CREATE OR REPLACE PROCEDURE SP_GETSAL_WITH_EXCEPTION (EMP_ID NUMBER, EMP_NAME OUT VARCHAR, SAL OUT NUMBER) AS
BEGIN
  SELECT EMPNAME INTO EMP_NAME FROM EMPLOYEE WHERE EMPID=EMP_ID;
  SELECT SALARY INTO SAL FROM EMPLOYEE WHERE EMPID=EMP_ID;
END;
```

Before you export the workflow, add exception handling as follows:

```sql
CREATE OR REPLACE PROCEDURE SP_GETSAL_WITH_EXCEPTION (EMP_ID NUMBER, EMP_NAME OUT VARCHAR, SAL OUT NUMBER) AS
BEGIN
  SELECT EMPNAME INTO EMP_NAME FROM EMPLOYEE WHERE EMPID=EMP_ID;
  SELECT SALARY INTO SAL FROM EMPLOYEE WHERE EMPID=EMP_ID;
  Exception
  when NO_DATA_FOUND
  then NULL;
END;
```

Pre-session and post-session commands

You can use pre-session and post-session SQL or shell commands in a workflow that you want to use in a PowerCenter task.

You might use a pre-session or post-session command to start FTP/SFTP scripts or stored procedures, rename or archive files, or run post-processing commands. Configure pre-session and post-session commands in the PowerCenter session.

When you configure a pre-session or post-session command, you can enter a single command or you can call a batch file that contains a set of commands. If you use a batch file, be sure to use complete paths or directories. When you configure the pre-session or post-session command in PowerCenter, enter the complete path or directory along with the file name, such as `c:/IC PowerCenter Task Commands/pre-session1.bat`.

Sources and targets

Use the following rules and guidelines for sources and targets that are used in a PowerCenter task:

- Field names must contain only alphanumeric or underscore characters. Do not use spaces in field names.
- Field names cannot start with a number.
- Each field name must be unique within each source and target object.
- The scale or precision of a numeric target column should be the same or greater than the scale or precision of the corresponding source column. Otherwise, the PowerCenter task truncates the data.
• Do not include Nvarchar2 columns in Oracle targets. Due to an ODBC driver limitation, the PowerCenter task truncates the last half of Nvarchar2 data before writing it to Oracle targets.

• Do not write Decimal data of 2147483648 or larger to Microsoft SQL Server or ODBC Integer(10) columns. Doing so can cause unexpected results.

FTP/SFTP connections for PowerCenter tasks

If you create a PowerCenter task with an FTP/SFTP target connection and the IS_STAGED option is enabled for the underlying PowerCenter session, Data Integration writes the flat file to the remote machine and the following local directory:

<Secure Agent installation directory>/apps/Data_Integration_Server/data

For PowerCenter tasks, Data Integration ignores the Local Directory property specified in the FTP/SFTP connection. Instead, it uses properties specified in the PowerCenter session. To change the local directory or default local filename, change the Output File Directory and Output Filename session properties in PowerCenter. Then export the workflow from PowerCenter to an XML file and re-import the XML file into Data Integration.

Web Service connections for PowerCenter tasks

When a PowerCenter XML file contains Web Service connection information, you can configure a Web Service connection in the PowerCenter task. If you configure a different connection type, the PowerCenter task uses Web Service connection information that is saved in the workflow.

PowerCenter task configuration

To create and configure a PowerCenter task, export the workflow from the PowerCenter Repository Manager to an XML file and then upload the XML file in to Data Integration. After you upload the file, map the uploaded connections to Data Integration connections.

You can update an existing PowerCenter task to use a different PowerCenter XML file. When you upload a new PowerCenter XML file to an existing PowerCenter task, the PowerCenter task deletes the old XML file and updates the PowerCenter task definition based on new XML file content.

Configuring a PowerCenter task

Perform the following tasks to create a PowerCenter task in Data Integration.

1. To create a PowerCenter task, click New > Tasks. Select PowerCenter Task and then click Create. To edit a PowerCenter task, on the Explore page, navigate to the task. In the row that contains the task, click Actions and select Edit.
2. In the Task Details area, configure the following fields:

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task Name</td>
<td>Name of the PowerCenter task. Task names can contain alphanumeric characters, spaces, and the following special characters: _ . + -. Maximum length is 100 characters. Task names are not case sensitive.</td>
</tr>
<tr>
<td>Location</td>
<td>Project folder in which the task resides. If the Explore page is currently active and a project or folder is selected, the default location for the asset is the selected project or folder. Otherwise, the default location is the location of the most recently saved asset.</td>
</tr>
<tr>
<td>Description</td>
<td>Description of the PowerCenter task. Maximum length is 255 characters.</td>
</tr>
<tr>
<td>Runtime Environment</td>
<td>Runtime environment that contains the Secure Agent to run the task.</td>
</tr>
<tr>
<td>Workflow XML File</td>
<td>PowerCenter workflow XML file associated with the task. Only the first 30 characters of the XML file name appear. To upload a file, click Upload XML File. After you upload the workflow XML file, the connections and transformations appear in the Workflow XML File Details area. To download the workflow XML file from Data Integration, click Download XML File. You might download a file to import the workflow to the PowerCenter Workflow Manager for review.</td>
</tr>
</tbody>
</table>

3. In the Schedule Details area, choose whether to run the task on a schedule or without a schedule. Choose one of the following options:
   - To run a task on a schedule, select Run this task on schedule and select the schedule you want to use.
   - To create a new schedule, select New. Enter schedule details and click OK.
   - To run the task manually, select Do not run this task on a schedule.

4. Configure email notification options for the task.

5. In the Connections area, select a Connection for each connection reference. A connection reference is a source, target, or lookup connection defined in the workflow XML file. Alternatively, to create a connection, click New. To edit a connection, click View and then click Edit. The Transformations area displays all transformations defined in the workflow XML file.

6. Click Save.

7. To run the PowerCenter task, click Run.
   You can also run the task from the Explore page.
Running a PowerCenter task

Perform the following tasks before you run a PowerCenter task:

- Ensure that the source and target definitions are current. If the source or target no longer contains fields that are mapped in the field mapping, the PowerCenter task fails.

- If the PowerCenter workflow uses the $PMSourceFileDir\ or $PMTargetFileDir variables to specify the source or target file directory location, you must copy the source or target files to the following directory:

  `<Secure Agent installation directory>/apps/Data_Integration_Server/data`

  If you do not move the source or target files, the task fails.

**Note:** You cannot run multiple instances of a PowerCenter task simultaneously. If you run a PowerCenter task that is already running, the PowerCenter task fails.

You can run a PowerCenter task manually or on a schedule:

- To run a PowerCenter task manually, on the Explore page, navigate to the task. In the row that contains the task, click **Actions** and select **Run**.
  You can also run a PowerCenter task manually from the Task Details page. To access the Task Details page, click **Actions** and select **View**.

- To run a PowerCenter task on a schedule, edit the task in the PowerCenter task wizard to associate the task with a schedule.
INDEX

$$PushdownConfig
   user-defined parameter 35

A
   add sources
      multiple objects 82
   advanced data filters
      configuring 12, 13
      deleting 12, 13
   advanced email masking
      description 120
   advanced options
      configuring for synchronization tasks 61
   advanced relationships
      in mapping tasks 27
   aggregate functions
      in mapping tasks 41
   AND
      reserved word 17
   automatic task recovery
      estimate subset 89

C
   CHR function
      inserting single quotation mark 16
   Cloud Application Integration community
      URL 9
   Cloud Developer community
      URL 9
   column names
      duplicates 68
   columns
      names, truncating 68
      target database, replication 67
   commands
      in synchronization and replication tasks 52, 69
   comments
      adding to field expressions 17
   configure
      staging connection 85
   constants
      definition 16
   credit card issuer
      description 119
   credit card masking
      overview 119
   cross-schema pushdown optimization 34
   custom masking
      expression 132

D
   data filters
      configuring 11
      configuring advanced filters 12, 13
      configuring for replication tasks 73
   data subset
      automatic task recovery 89
      configure relationships 88
   data types
      configures for synchronization 53
   credit card issuer
      description 119
   Credit card masking
      overview 119
   cross-schema pushdown optimization 34
   custom masking
      expression 132
   Daylight Savings Time
      schedules 21
   DD_DELETE constant
      reserved word 17
   DD_INSERT constant
      reserved word 17
   DD_REJECT constant
      reserved word 17
   DD_UPDATE constant
      reserved word 17
   default assignment
      masking rules 103
   delete criteria
      description 72
   delete task operation 46
deleting
  advanced data filters 12, 13
data filters 12
dictionary files
data masking 114
  substitution masking 129
DOT format
download 82
download
  mapping XML 113
  validation reports 113

E
e-mail masking
description 120
e-mail notification
configuring for tasks 23
exception handling
  in stored procedures for PowerCenter tasks 137
expressions
configuring for a lookup 51
rules 16
:EXT reference qualifier
reserved word 17
external IDs
  for related Salesforce objects 48

F
FALSE constant
reserved word 17
field exclusions
configuring for replication tasks 73
field expressions
  comments, adding 17
  components 16
  creating 15
  literals 16
  reserved words 17
rules 16
rules for validating 15
syntax 16
field mappings
  configuring for synchronization tasks 58
  required fields 49
  understanding for synchronization tasks 49
field metadata 35
  fields
  required in field mappings 49
flat files
  column names in synchronization tasks 48
  creating target files in synchronization tasks 47
directory for replication tasks 70
directory for synchronization tasks 53
  editing metadata in mapping tasks 35
rules for sources and targets in synchronization tasks 49
FTP/SFTP connections
  for PowerCenter tasks 138
full load
  replication task load type 64
functions
  definition 16

H
H2 database
configure 86
data subset 85
install 86
Linux 87
requirements 86
staging 85
Windows 86

I
IDs
types for objects 48
in-out parameters
  in mapping tasks 41
:INFA reference qualifier
reserved word 17
Informatica Global Customer Support
  contact information 10
Informatica Intelligent Cloud Services
  web site 9
in-place masking
update 84
input parameters
  in mapping tasks 41
insert task operation 46
IP address masking
description 121

J
jobs
  monitoring 23
  scheduling 19

K
key masking
date 124
description 122
numeric 124
string 122

L
$LastRunDate
data filter variable 14
$LastRunTime
data filter variable 14
linear taskflows
  creating schedules 22
literals
  definition 16
  single quotation mark requirement 16
  string and numeric 16
:LKP reference qualifier
reserved word 17
load types
  for replication 64
  full load 64
lookups
  configuring return values and expressions 51
lookups (continued)
  configuring the condition 50
  rules for creating 52

M
  maintenance outages 10
  mapping tasks
    advanced connection properties 26
    advanced session properties 28
    configuring 36
    configuring schedules, email, and advanced options 44
    configuring sources 37
    creating 36
    editing 36
    input and in-out parameters 41
    joining related sources 26
    overview 25
    running 45
    Sequence Generator configuration 35
    templates 25
    using aggregate functions 41
    using parameter files 18
    viewing details 44
  mapplets
    in synchronization tasks 50
  mask all except
    key masking 122
  mask only
    key masking 122
  masking
    advanced email 120
    advanced options 106
    configuring rules 109
    configuring the source 107
    configuring the target 108
    credit cards 119
    defining the task 107
    email addresses 120
    IP addresses 121
    overview 81
    phone numbers 124
    random 125
    rules descriptions 117
    scheduling a task 110
    SSN 128
    URL 131
    validation reports 84
  masking parameters
    random date 127
    random numeric 126
    random string 125
  masking rule
    mapplet 132
    postprocessing expression 119
    preprocessing expression 119
  masking rules
    description 117
    source field 101
  masking task
    data subset 88
  Masking task
    copying 112
    deleting 112
    exporting 113
    rename 112
    running immediately 111

Masking task (continued)
  steps to create 106
  Masking tasks
    using parameter files 18
  mass ingestion
    schedule configuration 80
    mass ingestion task
      configuration 77
      defining 77
      description 76
      local folder source 78
      local folder target 79
      running 80
      source configuration 78
      target configuration 79
      viewing details 80
  mass ingestion tasks
    prerequisites 77
    sources 76
    targets 76
  MCR reference qualifier
    reserved word 17
  monitoring jobs 23
  multiplicity
    configuring for a lookup 51

N
  NOT
    reserved word 17
  NULL constant
    reserved word 17

O
  operating system commands
    rules for replication tasks 53, 69
    rules for synchronization tasks 53, 69
  operations
    for synchronization tasks 46
    operators
      definition 16
  OR
    reserved word 17
  $OutputField
    lookup expression
      description 51
      lookup return value variable 51

P
  parameter file
    configuring for mapping tasks 44
  parameter files
    overview 18
  parameters
    configuring 42
    partial sandbox
      update 84
  phone numbers
    masking 124
  postprocessing commands
    SQL commands 53, 69
  PowerCenter tasks
    configuring 138
PowerCenter tasks (continued)
description 135
editing 138
rules for PowerCenter workflows 135
rules for PowerCenter XML files 135
rules for sources and targets 137
running 140
supported transformations and mapping objects 136
Web Service connections 138
PowerCenter XML files
rules for PowerCenter tasks 135
supported transformations and mapping objects 136
pre- and post-session commands
PowerCenter tasks
using pre-session and post-session commands 137
preprocessing and postprocessing commands
configuring for mapping tasks 44
configuring for synchronization tasks 61
rules for operating system commands 53, 69
rules for SQL commands 53, 69
preprocessing commands
SQL commands 53, 69
prerequisites
for synchronization tasks 53
replication tasks 70
PROC_RESULT variable
reserved word 17
pushdown optimization
cross-schema 34
mapping tasks 33
types 34
user-defined parameter 35

Q
quotation marks
inserting single using CHR function 16

R
random masking
date 127
numeric 126
string 125
related objects
configuring external IDs 48
repeat frequency
schedules 20
replication
source types 65
target types 65
replication tasks
configuring 70
configuring data filters 73
configuring schedules and email 74
configuring sources 71
configuring targets 72
creating 71
creating target tables 68
database targets 65
ingesting 71
field exclusions 73
flat file targets 66
load types 64
preprocessing and postprocessing commands 52, 69
prerequisites 70
replication tasks (continued)
replication process 64
rules for creating 70
rules for running 75
running 74
viewing details 74
reserved words
list 17
resetting
target tables 66
rules
configuring 109

S
Salesforce IDs
for objects 48
schedule
configure 105
configuring for a PowerCenter task 138
schedules
configuring 22
creating 22
Daylight Savings Time 21
repeat frequency 20
time zones 21
scheduling
mapping tasks 44
Masking task 110
replication tasks 74
synchronization tasks 61
taskflows 19
tasks 19
Schema graph
DOT format 82
multiple source objects 82
:SD reference qualifier
reserved word 17
:SEQ reference qualifier
reserved word 17
SESSSTARTTIME variable
reserved word 17
simple data filters
configuring 12
Social Security masking
description 128
source filter characters
description 123
source filter type
description 122
source objects
multiple 82
single 82
sources
configuring for mapping tasks 37
configuring for replication tasks 71
configuring for synchronization task 54
configuring multiple database tables for synchronization tasks 55
configuring saved query sources 54
configuring single-object sources 54
rules for multiple-object database in synchronization tasks 47
:SP reference qualifier
reserved word 17
SPOUTPUT
reserved word 17
SQL commands
rules for replication tasks 53, 69
SQL commands (continued)
rules for synchronization tasks 53, 69
SSN masking
description 128
status
Informatica cloud services 10
stopping a running job 24
stored procedures
exception handling for PowerCenter tasks 137
string literals
single quotation mark requirement 16
substitution masking
custom dictionaries 129
description 129
parameters 130
synchronization tasks
adding mapplets 50
configuring 53
configuring schedules, email, and advanced options 61
creating 54
creating flat file targets 47
data filters 57
description 46
editing 54
preprocessing and postprocessing commands 52, 69
prerequisites 53
rules for database sources and targets 49
rules for flat file sources and targets 49
rules for multiple-object database 47
rules for running 63
rules for sources and targets 49
running 62
sources overview 47
sources, configuring 54
targets overview 47
targets, configuring 56
task operations 46
truncating database targets 47
understanding field mappings 49
update columns 48
using parameter files 18
viewing details 62
syntax
for field expressions 16
SYSDATE variable
reserved word 17
system status 10

T

table names
duplicates 67
tables
resetting for replication 66
target database, replication 67
user-defined indexes 66
target
refresh fields 84
target databases
column names, truncating 68
data replication 65
requirements for a synchronization task 53
requirements for replication tasks 70
resetting tables for replication 66
table and column names, replication 67
target filter characters
description 123
target flat files
creating in synchronization tasks 47
target prefix
description 72
guidelines 68
target tables
creating for replication tasks 68
rules for resetting 56
target task operations
insert 83
update 83
upsert 83
targets
configuring for mapping tasks 39
configuring for replication tasks 72
configuring for synchronization tasks 56
truncating in synchronization tasks 47
taskflows
monitoring 23
scheduling 19
tasks
configuring email notification 23
configuring field expressions 15
creating schedules 22
monitoring 23
scheduling 19
stopping a running job 24
:TD reference qualifier
reserved word 17
time zones
description 21
transformations
in PowerCenter tasks 136
TRUE constant
reserved word 17
trust site
description 10

U

U

update columns
adding 48
update task operation 46
upgrade notifications 10
upsert task operation 46
URL masking
description 131
using parameter files
advanced filter 90
simple filter 90
validation
rules 15
validation report
target 84
variables
for data filters 14
We

Web Service connections
and PowerCenter tasks 138
web site 9

Index 145
workflow XML file downloading from a PowerCenter task 138

WORKFLOWSTARTTIME variable reserved word 17