



Informatica® Cloud Data Integration
Spring 2020 April

Tasks

Informatica Cloud Data Integration Tasks
Spring 2020 April
April 2020

© Copyright Informatica LLC 2006, 2020

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

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

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

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

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

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

Publication Date: 2020-08-24

Table of Contents

Preface	10
Informatica Resources.	10
Informatica Documentation.	10
Informatica Intelligent Cloud Services web site.	10
Informatica Intelligent Cloud Services Communities.	10
Informatica Intelligent Cloud Services Marketplace.	10
Data Integration connector documentation.	11
Informatica Knowledge Base.	11
Informatica Intelligent Cloud Services Trust Center.	11
Informatica Global Customer Support.	11
 Chapter 1: Data Integration tasks.....	 12
Data filters.	12
Simple data filters.	13
Advanced data filters.	13
Data filter operators.	14
Data filter variables.	15
Rules and guidelines for data filters.	15
Field expressions.	16
Creating a field expression.	16
Transformation language components for expressions.	17
Expression syntax.	17
String and numeric literals.	17
Rules and guidelines for expressions.	18
Adding comments to expressions.	18
Reserved words.	19
Parameter files.	20
Schedules.	20
Repeat frequency.	21
Time zones and schedules.	22
Daylight Savings Time changes and schedules.	22
Creating a schedule.	23
Running a task on a schedule.	23
Email notification.	24
Monitoring a job.	24
Data catalog discovery for sources.	25
Catalog search.	25
Discovering and selecting a catalog object.	27
Stopping a job.	28

Chapter 2: Mapping tasks.....	29
Mapping task templates.	29
Advanced connection properties for Visio templates.	30
Related objects.	30
Advanced relationships.	31
Advanced session properties.	32
Advanced session properties for elastic mappings.	37
Pushdown optimization.	40
Pushdown optimization types.	40
Cross-schema pushdown optimization.	41
Pushdown optimization user-defined parameters.	41
Field metadata.	41
Sequence Generator values.	41
Mapping task configuration.	42
Defining a mapping task.	42
Configuring sources.	43
Configuring targets.	45
Configuring parameters.	47
Configuring a schedule and advanced options.	50
CLAIRE Tuning.	51
Guidelines to get an accurate recommendation.	51
Configuring tuning.	52
Viewing tuning results.	53
Viewing and editing mapping task details.	54
Running a mapping task.	54
 Chapter 3: Synchronization tasks.....	 55
Task operations.	55
Synchronization task sources.	56
Rules and guidelines for multiple-object databases.	56
Synchronization task targets.	56
Flat file target creation.	56
Database target truncation	57
Salesforce targets and IDs for related objects.	57
Update columns.	57
Column names in flat files.	58
Rules and guidelines for synchronization task sources and targets.	58
Rules and guidelines for flat file sources and targets.	58
Rules and guidelines for database sources and targets.	59
Field mappings.	59
Field data types.	59
Mapplets in field mappings.	60

Lookup conditions.	60
Lookup return values.	60
Rules and guidelines for lookups.	61
Preprocessing and postprocessing commands.	62
Preprocessing and postprocessing SQL commands.	62
Preprocessing and postprocessing operating system commands.	62
Configuring a synchronization task.	62
Synchronization prerequisite tasks.	63
Defining a synchronization task.	63
Configuring the source.	63
Configuring the target.	66
Configuring the data filters.	67
Configuring the field mapping.	68
Configuring a schedule and advanced options.	70
Viewing synchronization task details.	71
Running a synchronization task.	71
Rules and guidelines for running a synchronization task.	72
Chapter 4: Replication tasks.	73
Load types.	73
Full load.	73
Replication task sources.	74
Replication task targets.	74
Replicate data to a database target.	74
Replicate data to a flat file target.	75
Reset a database target.	75
Resetting a target table.	75
Rules and guidelines for resetting a target table.	75
Table and column names in a database target.	76
Table name truncation.	76
Duplicate tables names from same replication task.	76
Duplicate table names from different replication tasks.	76
Column name truncation.	77
Target prefixes.	77
Creating target tables.	77
Replication task schedules.	78
Preprocessing and postprocessing commands.	78
Preprocessing and postprocessing SQL commands.	78
Preprocessing and postprocessing operating system commands.	78
Configuring a replication task.	79
Rules and guidelines for configuring replication tasks.	79
Replication prerequisite tasks.	79
Defining a replication task.	80

Configuring the source.	80
Configuring the target.	81
Configuring the field exclusions.	82
Configuring the data filters.	82
Configuring a schedule and advanced options.	83
Viewing replication task details.	83
Running a replication task.	84
Rules and guidelines for running a replication task.	84

Chapter 5: Mass ingestion tasks. 85

Mass ingestion task sources.	85
Mass ingestion task targets.	86
Mass ingestion task schedules.	86
File processing actions.	87
File security.	87
Mass ingestion task configuration.	88
Before you begin.	88
Defining a mass ingestion task.	88
Configuring the source.	89
Configuring the target.	93
Defining the schedule and adding file actions.	94
Viewing mass ingestion task details.	95
Running a mass ingestion task.	95
Key ring command reference.	95
createKeyRing.	96
createKeyPair.	96
listKeys.	97
importKeys.	97
exportKeyPairs.	98
exportPublicKeys.	99
deleteKeys.	99
changePassphrase.	100

Chapter 6: Masking tasks. 101

Rules and guidelines for masking tasks.	101
Masking task options.	102
Source objects.	102
Schema graph.	102
Target task operations.	103
Inplace masking.	104
Update partial sandbox.	104
Refresh fields.	104
Validation reports.	104

Staging database.	105
Start the staging connection.	105
H2 database configuration requirements.	106
Installing and configuring H2 database manually on Windows.	106
Installing H2 database manually on Linux.	107
Data subset.	108
Data subset options.	109
Automatic task recovery.	109
Parameter files in data filters.	110
Configure relationship behavior.	111
Data subset use cases for two objects.	112
Case 1. Select the default path with filter on Account.	112
Case 2. Select the configured path with filter on Account.	113
Case 3. Select the default path with filter on Contact.	115
Case 4. Select the configured path with filter on Contact.	116
Data subset use cases for three objects.	117
Case 1. Default path.	118
Case 2. Configured path.	119
Data subset rows.	119
Data subset rows example.	119
Refresh metadata.	121
Reset task.	122
Apply masking rules.	122
Masking rule assignments.	123
Add mapplets.	123
Target fields.	124
Default masking rules package.	124
Schedule options.	127
Email notification options.	127
Advanced options.	127
Configuring a masking task.	128
Prerequisites.	128
Step 1. Define the masking task.	128
Step 2. Configure the source	129
Step 3. Configure the target.	130
Step 4. Configure the data subset.	130
Step 5. Define data masking rules.	131
Step 6. Schedule the masking task.	131
Masking task maintenance.	132
Editing a masking task.	132
Running a masking task manually.	133
Refreshing the metadata in a masking task.	133

Stopping a masking task.	134
Resetting a masking task.	135
Configuring masking task permissions.	135
Copying a masking task.	135
Renaming a masking task.	135
Deleting a masking task.	136
Exporting a masking task.	136
Downloading mapping XML.	136
Downloading validation reports.	136
Dictionary files for data masking.	136
Consistent masked output.	138
Rules and guidelines.	138
Example.	139
Chapter 7: Masking rules.	140
Masking rules.	140
Repeatable output.	141
Seed value.	141
Preprocessing and postprocessing expressions.	142
Credit card masking.	142
Credit card parameters.	143
Email masking.	143
Advanced email masking.	143
IP address masking.	144
Key masking.	145
Key string masking.	145
Key numeric masking.	147
Key date masking.	147
Nullification masking.	147
Phone number masking.	147
Random masking.	148
Random string masking.	148
Random numeric masking.	149
Random date masking.	150
SIN masking.	151
SSN masking.	151
Substitution masking.	152
Substitution masking with custom dictionaries.	152
Substitution masking parameters.	153
Custom substitution lookup example.	153
Custom substitution dictionary lookup use cases.	154
URL masking.	154
Custom masking.	155

Mapplet masking.	155
--------------------------	-----

Chapter 8: PowerCenter tasks..... 158

PowerCenter workflows.	158
Supported transformations and mapping objects.	159
Exception handling in stored procedures.	160
Pre-session and post-session commands.	160
Sources and targets.	160
FTP/SFTP connections for PowerCenter tasks.	161
Web Service connections for PowerCenter tasks.	161
PowerCenter task configuration.	161
Configuring a PowerCenter task.	161
Running a PowerCenter task.	163

Index..... 164

Preface

Use *Tasks* to learn how to set up and run Data Integration tasks manually or on a schedule.

Informatica Resources

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

Informatica Documentation

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

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

Informatica 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 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:

<https://network.informatica.com/community/informatica-network/products/cloud-integration>

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

<https://network.informatica.com/community/informatica-network/products/cloud-integration/cloud-developers>

Informatica Intelligent Cloud Services Marketplace

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

<https://marketplace.informatica.com/>

Data Integration connector documentation

You can access documentation for Data Integration Connectors at the Documentation Portal. To explore the Documentation Portal, visit <https://docs.informatica.com>.

Informatica Knowledge Base

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

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

Informatica Intelligent Cloud Services Trust Center

The Informatica Intelligent Cloud Services Trust Center provides information about Informatica security policies and real-time system availability.

You can access the trust center at <https://www.informatica.com/trust-center.html>.

Subscribe to the Informatica Intelligent Cloud Services Trust Center to receive upgrade, maintenance, and incident notifications. The [Informatica Intelligent Cloud Services Status](#) page displays the production status of all the Informatica cloud products. All maintenance updates are posted to this 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 receive updates for a single component or all Informatica Intelligent Cloud Services components. Subscribing to all components is the best way to be certain you never miss an update.

To subscribe, go to <https://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 <https://network.informatica.com/welcome>.

The telephone numbers for Informatica Global Customer Support are available from the Informatica web site at <https://www.informatica.com/services-and-training/support-services/contact-us.html>.

CHAPTER 1

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:

Field	Operator	Field Value
AnnualRevenue	>=	100000
NumberOfEmployees	>	500

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:

Field type	Operators
Boolean	=, !=, Is Null, Is Not Null
Currency	=, !=, <, <=, >, >=, Is Null, Is Not Null
Date	=, !=, <, <=, >, >=, Is Null, Is Not Null
Datetime	=, !=, <, <=, >, >=, Is Null, Is Not Null
Double	=, !=, <, <=, >, >=, Is Null, Is Not Null
ID	=, !=, Is Null, Is Not Null
Int	=, !=, <, <=, >, >=, Is Null, Is Not Null
Reference	=, !=, Is Null, Is Not Null
String	=, !=, LIKE'_%', LIKE'%_', LIKE'%_%', Is Null, Is Not Null, <, <=, >, >=
Textarea	=, !=, LIKE'_%', LIKE'%_', LIKE'%_%', Is Null, Is Not Null, <, <=, >, >=
All other field types	=, !=, Is Null, Is Not Null

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 data filter variables:

Variable	Description
\$LastRunDate	The start date in GMT time zone of the last task run that was successful or ended with a warning. Does not include time. For example, 2018-09-24. Can be used as a value for filter where the field type is DATE.
\$LastRunTime	The start date and time in GMT time zone of the last task run that was successful or ended with a warning. For example, 2018-09-24 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.

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 Informatica Intelligent Cloud Services. 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 data types of the fields change. If a data type changes, edit the task.
- You can select Equals, Not Equals, Is Null, or Is Not Null operators on fields of the Other data type.
- 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 data types.
- 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 data type 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.

- When you use a parameter in a data filter, start the data filter with the parameter. For example, use `$$Sales=100000` instead of `100000=$$Sales`.

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.

For detailed information about the Data Integration transformation language, see *Function Reference*.

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.
To include source fields and system variables in the expression, you can select them from the **Source Fields** and **System Variables** tabs to insert them into the expression or you can add them to the expression manually.
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 data types 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 data type of a field in an expression matches the data type expected by the containing function. However, the expression validator does not check for incompatible data types 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 data types 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 data types. 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 data types 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.
- When you use a parameter in an expression, use the appropriate function to convert the value to the necessary data type. For example, you might use the following expression to define a quarterly bonus for employees:

```
IIF((EMP_SALES < TO_INTEGER($$SalesQuota), 200, 0)
```

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
- :INFA
- :LKP
- :MCR
- :SD
- :SEQ
- :SP
- :TD
- AND
- DD_DELETE
- DD_INSERT
- DD_REJECT
- DD_UPDATE
- FALSE
- NOT
- NULL
- OR
- PROC_RESULT
- SESSSTARTTIME
- SPOUTPUT
- SYSDATE
- TRUE
- WORKFLOWSTARTTIME

The following words are reserved for Informatica Intelligent Cloud Services:

- ABORTED
- DISABLED
- FAILED

- NOTSTARTED
- STARTED
- STOPPED
- SUCCEEDED

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

Parameter files

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

Use a parameter file to define values that you want to update without having to edit the task. You update the values in the parameter file instead of updating values in the task. The parameter values are applied when the task runs.

You can use a parameter file to define values for parameters in data filters, expressions, and lookup expressions in the following tasks:

- Synchronization tasks
- Mapping tasks

You can use a parameter file to define parameter values for source and target connections and source and target data objects in mapping tasks.

Note: Not all connectors support parameter files. To see if a connector supports runtime override of source and target connections and objects, see the help for the appropriate connector.

You cannot use a parameter file if the mapping task is based on an elastic mapping.

You enter the parameter file name and location on the **Schedule** tab when you configure the task.

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:

Option	Description
Does not repeat	Tasks run as scheduled and do not repeat.
Every N minutes	Tasks run on an interval based on a specified number of minutes. You can configure the following options: <ul style="list-style-type: none">- Repeat frequency. Select a frequency in minutes. Options are 5, 10, 15, 20, 30, 45.- Days. Days of the week when you want tasks to run. You can select one or more days of the week.- Time range. Hours of the day when you want tasks to start. Select All Day or configure a time range. You can configure a time range between 00:00-23:55.- Repeat option. The range of days when you want tasks to run. You can select Repeat Indefinitely or configure an end date and time.
Hourly	Tasks run on an hourly interval based on the start time of the schedule. You can configure the following options: <ul style="list-style-type: none">- Repeat frequency. Select a frequency in hours. Options are 1, 2, 3, 4, 6, 8, 12.- Days. Days of the week when you want tasks to run. You can select one or more days of the week.- Time range. Hours of the day when you want tasks to start. Select All Day or configure a time range. You can configure a time range between 00:00-23:55.- Repeat option. The range of days when you want tasks to run. You can select Repeat Indefinitely or configure an end date and time.
Daily	Tasks run daily at the start time configured for the schedule. You can configure the following options: <ul style="list-style-type: none">- Repeat frequency. The frequency at which you want tasks to run. Select Every Day or Every Weekday.- Repeat option. The range of days when you want tasks to run. You can select Repeat Indefinitely or configure an end date and time.

Option	Description
Weekly	<p>Tasks run on a weekly interval based on the start time of the schedule.</p> <p>You can configure the following options:</p> <ul style="list-style-type: none"> - 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. <p>If you do not specify a day, the schedule runs regularly on the same day of the week as the start date.</p>
Monthly	<p>Tasks run on a monthly interval based on the start time of the schedule.</p> <p>You can configure the following options:</p> <ul style="list-style-type: none"> - Day. Day of the month when you want tasks to run. You can configure one of the following options: <ul style="list-style-type: none"> - 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. <p>Tip: With the Day option, you can configure tasks to run on the First Day or the Last Day of the month.</p> - 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:

Property	Description
Schedule Name	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.
Description	Description of the schedule. Maximum length is 255 characters.
Starts	Date and time when the schedule takes effect. 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.
Time Zone	Select the time zone for the schedule to use. The time zone can differ from the organization time zone or user time zone.
Repeats	Repeat frequency for the schedule. Select one of the following options: <ul style="list-style-type: none">- Does Not Repeat- Every N Minutes- Hourly- Daily- Weekly- Monthly Default is Does Not Repeat.

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:

Field	Description
Use Default Email Notification Options for my Organization	Use the email notification options configured for the organization.
Use Custom Email Notification Options for this Task	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.

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*.

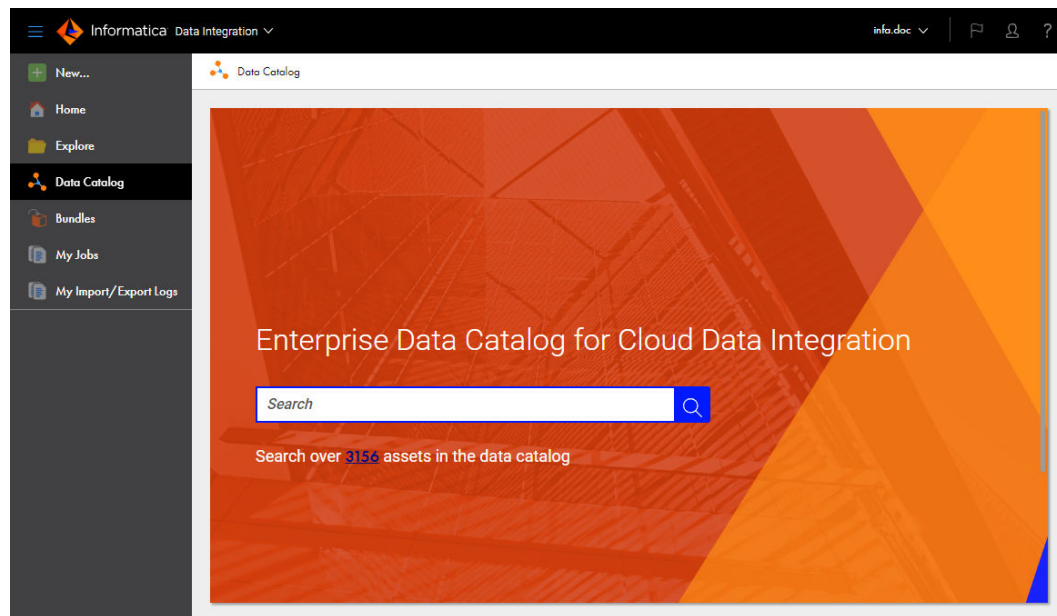
Data catalog discovery for sources

If your organization uses Enterprise Data Catalog and you have the appropriate license, you can perform a search against the catalog and discover catalog assets. You can use the assets that you discover as sources, targets, and lookup objects in mappings and as sources in synchronization and mass ingestion tasks.

Note: Before you can use data catalog discovery, your organization administrator must configure the Enterprise Data Catalog integration properties on the **Organization** page in Administrator. For more information about configuring Enterprise Data Catalog integration properties, see the Administrator help.

Perform data catalog discovery on the **Data Catalog** page.

The following image shows the **Data Catalog** page:



The page displays a **Search** field and the total number of table, view, and flat file assets in the catalog.

In the **Search** field, enter a search phrase that might occur in the object name, description, or other metadata such as the data domain or associated business glossary term. When you select an object from the search results, Data Integration asks you where you want to use the object.

To use the object as a source in a synchronization or mass ingestion task, select **Create a new asset** and choose the task. Data Integration imports the connection if it does not exist in your organization. Data Integration then creates the task and adds the object to the task as the source object. You cannot add the object as a source in an existing task.

Catalog search

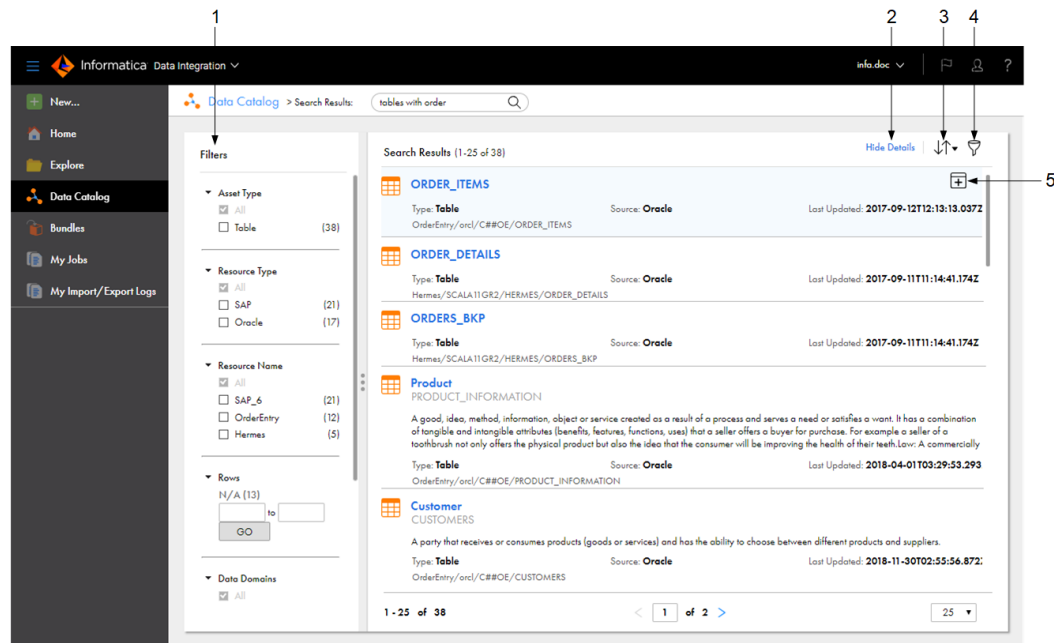
Use the search on the **Data Catalog** page to find an Enterprise Data Catalog object. Enter the object name, part of the name, or keywords associated with the object in the **Search** field, and then click the search icon. Data Integration returns all tables, views, and flat files in the catalog that match the search criteria.

You can use the * and ? wildcard characters in the search phrase. For example, to find objects that start with the string "Cust", enter Cust* in the **Search** field.

You can also enter keyword searches. For example, if you enter `tables with order` in the **Search** field, Data Integration returns tables with "order" in the name or description, tables that have the associated business term "order," and tables that contain columns for which the "order" data domain is inferred or assigned.

For more information about Enterprise Data Catalog searches and search results, see the Enterprise Data Catalog documentation.

The following image shows an example of search results when you enter "tables with order" as the search phrase:



1. Filter search results.
2. Show or hide object details.
3. Sort search results.
4. Apply or remove all filters.
5. Use the selected object in a synchronization task, a mapping, or an elastic mapping.

You can perform the following actions on the search results page:

Filter search results.

Use the filters to filter search results by asset type, resource type, resource name, number of rows, data domains, and date last updated.

Show details.

To display details about the object, click **Show Details**.

Sort results.

Use the **Sort** icon to sort results by relevance or name.

Open an object in Enterprise Data Catalog.

To open an object in Enterprise Data Catalog, click the object name. To view the object, you must log in to Enterprise Data Catalog with your Enterprise Data Catalog user name and password.

Use the object in a synchronization task, mass ingestion task, or mapping.

To use the object in a synchronization task, mass ingestion task or mapping, click **Use Object**. You can select an object if the object is a valid source, target, or lookup type for a mapping or a valid source type

for the task. For example, you can select an Oracle table to use as the source in a new synchronization task, but you cannot select a Hive table.

When you select the object, Data Integration prompts you to select the task where you want to use the object and imports the connection if it does not exist.

Connection properties vary based on the object type. Data Integration imports most connection properties from the resource configuration in Enterprise Data Catalog, but you must enter other required properties, such as the connection name and password.

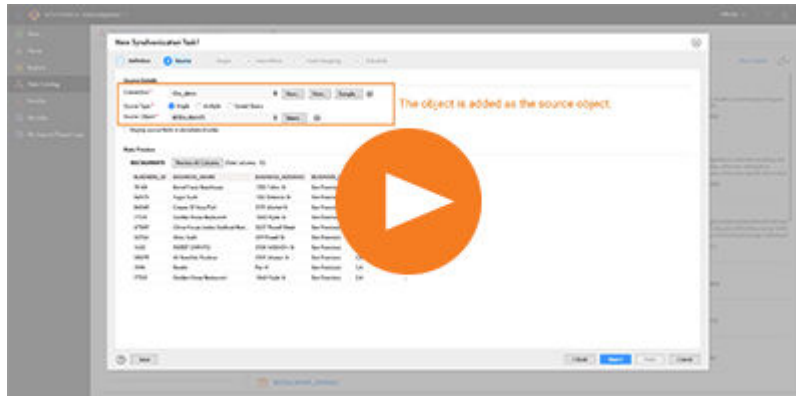
After you configure the connection or if the connection already exists, Data Integration adds the object to a new synchronization task, mass ingestion task, or to the inventory of a new or open mapping.

Discovering and selecting a catalog object

Discover and select a catalog object so that you can use the object as a source in a new synchronization or mass ingestion task.

Before you can use data catalog discovery, your organization administrator must configure the Enterprise Data Catalog integration properties on the **Organization** page in Administrator.

The following video shows you how to discover and select a catalog object as the source in a new synchronization task:



1. Open the **Data Catalog** page.
2. Enter the search phrase in the search field.
For example, to find customer tables, you might enter "Customer," "Cust*," or "tables with customer."
3. On the search results page, click **Use Object** in the row that contains the object.
You can select one object at a time.
Data Integration prompts you to select where to use the object.
4. Select one of the following options:
 - To add the object to a new synchronization task, click **New Synchronization Task**.
 - To add the object to a new Mass ingestion task, click **New Mass Ingestion Task**.
 - To add the object to a new mapping, click **New Mapping**.
 - To add the object to an open mapping or elastic mapping, click **Add to an open asset**, and then select the mapping.

5. Click **OK**.

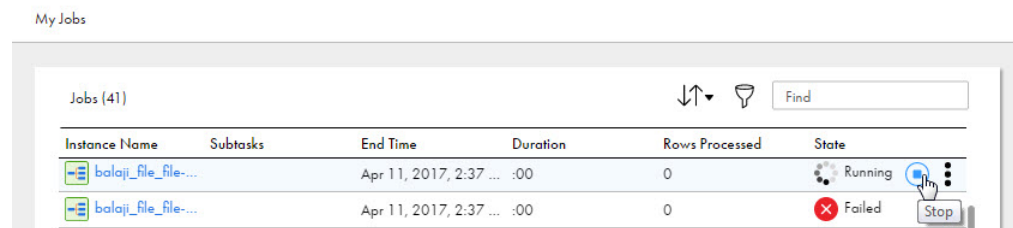
If the connection does not exist in your organization, Data Integration prompts you to import the connection. Enter the missing connection properties such as the connection name and password.

If you use the object in a synchronization or mass ingestion task, Data Integration creates the task with the object as the source. Configure other task properties such as the target, data filters, field mapping, and scheduling information.

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.

CHAPTER 2

Mapping tasks

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.

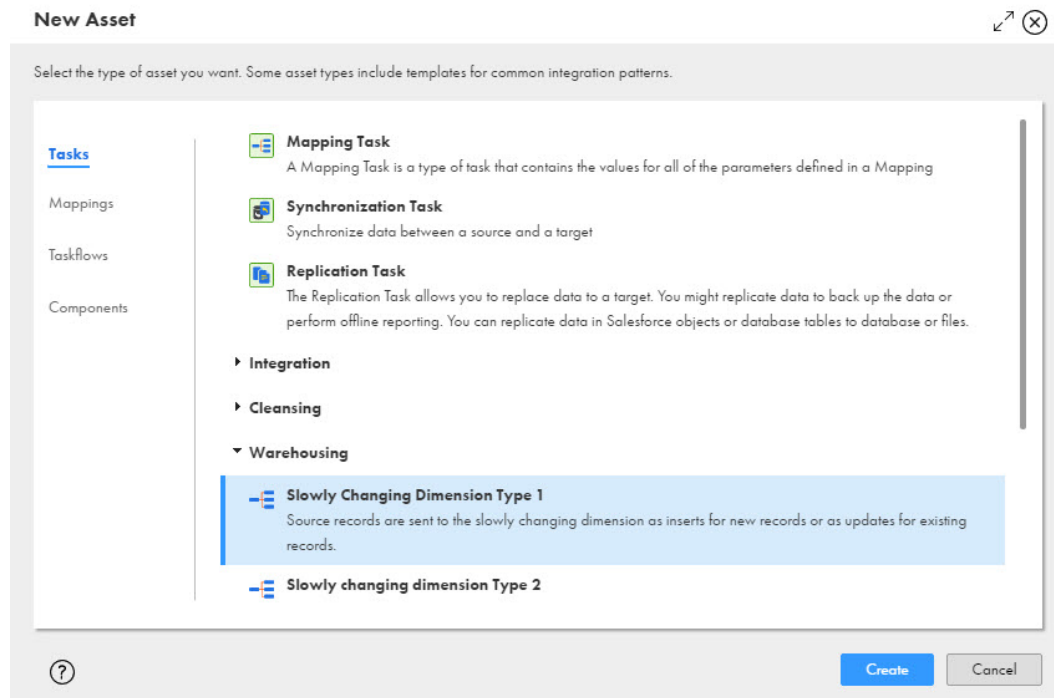
At run time, a mapping task processes task data based on the data flow logic from the mapping or Visio template, the parameters defined in the task, and the user-defined parameters defined in a parameter file, when available.

Mapping task templates

Use a mapping task template 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, left outer, or right outer 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, specify keys for the primary and related objects, and configure the join type and operator. For more information about related source objects, see the Source Transformation section in *Transformations*.

The screenshot shows the 'New MappingTask7' dialog box with the 'Sources' tab selected. The 'Source Parameter Details' section shows 'NewSource Connection' as 'mysql' and 'Source Type' as 'Multiple'. The 'Objects and Relationships' table lists three objects: 'a_r104_mar24_al_df_account' (highlighted), 'a_r10_2_account', and 'a_r10_c_lookup_c'. The 'Join Condition' for 'a_r10_2_account' is 'a_r10_2_account.ID = a_r104_mar24_al_df_account.ID', and for 'a_r10_c_lookup_c' it is 'a_r10_c_lookup_c.ID = a_r104_mar24_al_df_account.ID'. The 'Query Options' section includes 'Filter' and 'Sort' buttons, and a checkbox for 'Display source fields in alphabetical order'. The 'Data Preview' section shows 'a_r104_mar24_al_df_account' and 'Preview All Columns' (Total columns: 45). The bottom of the dialog has 'Save', '< Back', 'Next >', 'Finish', and 'Cancel' buttons.

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

If you configure advanced session properties for a mapping task and the mapping task is based on an elastic mapping, the advanced session properties are different.

General options

The following table describes the general options:

General options	Description
Write Backward Compatible Session Log File	Writes the session log to a file.
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: <ul style="list-style-type: none">- \$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.
Source File Directory	Source file directory path. Use for flat file connections only.
Target File Directory	Target file directory path. Use for flat file connections only.
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: <ul style="list-style-type: none">- 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.

General options	Description
Commit Type	<p>Commit type to use. Use one of the following options.</p> <ul style="list-style-type: none"> - 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. <p>When you do not configure a commit type, the task performs a target commit.</p>
Commit Interval	<p>Interval in rows between commits.</p> <p>When you do not configure a commit interval, the task commits every 10,000 rows.</p>
Commit on End of File	Commits data at the end of the file.
Rollback Transactions on Errors	<p>Rolls back the transaction at the next commit point when the task encounters a non-fatal error.</p> <p>When the task encounters a transformation error, it rolls back the transaction if the error occurs after the effective transaction generator for the target.</p>
Java Classpath	<p>Java classpath to use.</p> <p>The Java classpath is added to the beginning of the system classpath when the task runs.</p> <p>Use this option when you use third-party Java packages, built-in Java packages, or custom Java packages in a Java transformation.</p>

Performance settings

The following table describes the performance settings:

Performance settings	Description
DTM Buffer Size	<p>Amount of memory allocated to the task from the DTM process.</p> <p>By default, a minimum of 12 MB is allocated to the buffer at run time.</p> <p>Use one of the following options:</p> <ul style="list-style-type: none"> - 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. <p>You might increase the DTM buffer size in the following circumstances:</p> <ul style="list-style-type: none"> - 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.
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.

Performance settings	Description
Enable High Precision	Processes the Decimal data type to a precision of 28.
Session Retry on Deadlock	The task retries a write on the target when a deadlock occurs.
Pushdown Optimization	<p>Type of pushdown optimization. Use one of the following options:</p> <ul style="list-style-type: none"> - None. The task processes all transformation logic for the task. - To Source. The task pushes as much of the transformation logic to the source database as possible. - To Target. The task pushes as much of the transformation logic to the target database as possible. - 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. - \$\$PushdownConfig. The task uses the pushdown optimization type specified in the user-defined parameter file for the task. <p>When you use \$\$PushdownConfig, ensure that the user-defined parameter is configured in the parameter file.</p> <p>When you use pushdown optimization, do not use the Error Log Type property.</p> <p>For more information, see “Pushdown optimization” on page 40.</p>
Enable cross-schema pushdown optimization	<p>Enables pushdown optimization for tasks that use source or target objects associated with different schemas within the same database.</p> <p>To see if the connector you use supports cross-schema pushdown optimization, see the help for the relevant connector.</p> <p>This property is enabled by default.</p>
Allow Temporary View for Pushdown	<p>Allows the task to create temporary view objects in the database when it pushes the task to the database.</p> <p>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.</p>
Allow Temporary Sequence for Pushdown	<p>Allows the task to create temporary sequence objects in the database.</p> <p>Use when the task is based on a Visio template that includes a Sequence Generator transformation.</p>
Allow Pushdown for User Incompatible Connections	<p>Indicates that the database user of the active database has read permission on idle databases.</p> <p>If you indicate that the database user of the active database has read permission on idle databases, and it does not, the task fails.</p> <p>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.</p>
Session Sort Order	Order to use to sort character data for the task.

Advanced options

The following table describes the advanced options:

Advanced options	Description
Constraint Based Load Ordering	Currently not supported in Informatica Intelligent Cloud Services.
Cache Lookup() Function	<p>Caches lookup functions in Visio templates with unconnected lookups. Overrides lookup configuration in the template.</p> <p>By default, the task performs lookups on a row-by-row basis, unless otherwise specified in the template.</p>
Default Buffer Block Size	<p>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.</p> <p>Use one of the following options:</p> <ul style="list-style-type: none"> - 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. <p>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.</p> <p>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.</p>
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	<p>Maximum memory allocated for automatic cache when you configure the task to determine the cache size at run time.</p> <p>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.</p> <p>If the value is set to zero, the task uses default values for memory attributes that you set to auto.</p>
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	<p>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.</p> <p>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.</p> <p>To configure the number of pipelines that the task can create concurrently, select one of the following options:</p> <ul style="list-style-type: none"> - 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.

Advanced options	Description
Custom Properties	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.
Pre-build Lookup Cache	<p>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.</p> <p>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.</p> <p>Configure one of the following options:</p> <ul style="list-style-type: none"> - 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. <p>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.</p>
DateTime Format String	<p>Date time format for the task. You can specify seconds, milliseconds, or nanoseconds.</p> <p>To specify seconds, enter <code>MM/DD/YYYY HH24:MI:SS</code>.</p> <p>To specify milliseconds, enter <code>MM/DD/YYYY HH24:MI:SS.MS</code>.</p> <p>To specify microseconds, enter <code>MM/DD/YYYY HH24:MI:SS.US</code>.</p> <p>To specify nanoseconds, enter <code>MM/DD/YYYY HH24:MI:SS.NS</code>.</p> <p>By default, the format specifies microseconds, as follows: <code>MM/DD/YYYY HH24:MI:SS.US</code>.</p>
Pre 85 Timestamp Compatibility	Do not use with Data Integration.

Error handling

The following table describes the error handling options:

Error handling options	Description
Stop on Errors	<p>Indicates how many non-fatal errors the task can encounter before it stops the session. Non-fatal errors include reader, writer, and DTM errors.</p> <p>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.</p>
Override Tracing	Overrides tracing levels set on an object level.
On Stored Procedure Error	<p>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:</p> <ul style="list-style-type: none"> - 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. <p>By default, the task stops.</p>

Error handling options	Description
On Pre-Session Command Task Error	<p>Determines the behavior when a task that includes pre-session shell commands encounters errors. Use one of the following options:</p> <ul style="list-style-type: none"> - Stop Session. The task stops when errors occur while executing pre-session shell commands. - Continue Session. The task continues regardless of errors. <p>By default, the task stops.</p>
On Pre-Post SQL Error	<p>Determines the behavior when a task that includes pre-session or post-session SQL encounters errors:</p> <ul style="list-style-type: none"> - Stop Session. The task stops when errors occur while executing pre-session or post-session SQL. - Continue. The task continues regardless of errors. <p>By default, the task stops.</p>
Error Log Type	<p>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.</p> <p>Do not use this property when you use the Pushdown Optimization property.</p>
Error Log File Directory	<p>Specifies the directory where errors are logged. By default, the error log file directory is \$PMBadFilesDir\.</p>
Error Log File Name	<p>Specifies error log file name. By default, the error log file name is PLError.log.</p>
Log Row Data	<p>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.</p>
Log Source Row Data	<p>Specifies whether or not to log source row data. By default, the check box is clear and source row data is not logged.</p>
Data Column Delimiter	<p>Delimiter for string type source row data and transformation group row data. By default, the task uses a pipe () delimiter.</p> <p>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.</p>

Advanced session properties for elastic mappings

For a mapping task that is based on an elastic mapping, configure optional advanced session properties to run the task on the Serverless Spark engine.

You can configure the following types of advanced session properties for an elastic mapping:

- General
- Spark
- Custom

General properties

The following table describes the general properties:

Advanced session properties	Description
Date/Time Format String	Date time format for the task. 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.
Override Mapping Task Timeout	Overrides the mapping task timeout set in the elastic configuration that is associated with the runtime environment.
Override Tracing	Overrides tracing levels set on an object level.

Spark properties

The Secure Agent configures the properties on the Serverless Spark engine using best practices and the average computational requirements of in-house mapping tasks. If the default values do not fit the requirements of a specific mapping task, reconfigure the properties to override the default values.

To get an optimal set of Spark properties for the mapping task, see [“CLAIRE Tuning” on page 51](#).

The following table describes the Spark properties:

Spark advanced properties	Description
spark.driver.extraJavaOptions	Additional JVM options for the Spark driver process. Default is -Djava.security.egd=file:/dev/./urandom -XX:MaxMetaspaceSize=256M -XX:+UseG1GC -XX:MaxGCPauseMillis=500.
spark.driver.maxResultSize	Maximum total size of serialized results of all partitions for each Spark action. Default is 4G.
spark.driver.memory	Amount of memory for the Spark driver process. Default is 4G.
spark.dynamicAllocation.maxExecutors	Maximum number of Spark executors if dynamic allocation is enabled. Default is 1000.
spark.executor.cores	Number of cores that run each Spark executor. Default is 2.
spark.executor.extraJavaOptions	Additional JVM options for Spark executors. Default is -Djava.security.egd=file:/dev/./urandom -XX:MaxMetaspaceSize=256M -XX:+UseG1GC -XX:MaxGCPauseMillis=500.

Spark advanced properties	Description
spark.executor.memory	Amount of memory for each Spark executor. Default is 6G.
spark.memory.fraction	Fraction of the heap that is allocated to the Spark engine. When set to 1, the Spark engine uses the full heap space except for 300 MB that is reserved memory. Default is 0.6.
spark.memory.storageFraction	Fraction of memory that the Spark engine uses for storage compared to processing data. Default is 0.5.
spark.rdd.compress	Indicates whether to compress serialized RDD partitions. Default is false.
spark.reducer.maxSizeInFlight	Maximum size of the data that each reduce task can receive from a map task while shuffling data. The size represents a network buffer to make sure that the reduce task has enough memory for the shuffled data. Default is 48M.
spark.shuffle.file.buffer	Size of the in-memory buffer that each map task uses to write the intermediate shuffle output. Default is 32K.
spark.sql.autoBroadcastJoinThreshold	Indicates whether broadcast join is enabled or disabled. When broadcast join is enabled, the Spark driver sends data to Spark executors that are running on the elastic cluster. Default is 256000000.
spark.sql.shuffle.partitions	Number of partitions that Spark uses to shuffle data to process joins or aggregations in an elastic mapping. Default is 100.
spark.custom.property	Configure custom properties for the Spark engine. Use & : to separate custom properties.

Custom properties

The following table describes the custom property:

Advanced custom properties	Description
advanced.custom.property	Configure custom properties to run the elastic mapping. 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. Use & : to separate custom properties.

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.

You cannot configure pushdown optimization for a mapping task that is based on an elastic mapping.

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 help for the relevant connector.

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 to 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.

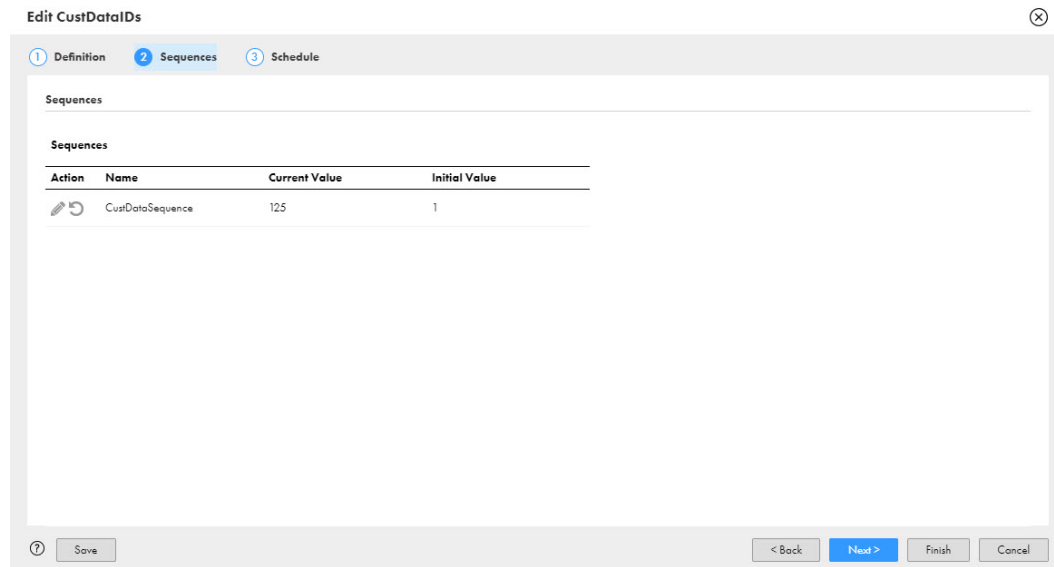
To view and edit field metadata, use the **Edit Types** option on the **Sources** page of the mapping task wizard.

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:

Field	Description
Task Name	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.
Location	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.
Description	Description of the task. Maximum length is 4000 characters.
Runtime Environment	Runtime environment that contains the Secure Agent to run the task.
Task Based On	The basis for the task. Select Mapping or Visio Template.
Mapping	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 .
Visio Template	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.

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:

Source parameter detail	Description
Connection	<p>Select a connection.</p> <p>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.</p> <p>To configure advanced properties for the connection, click Advanced. Not available for all connection types.</p> <p>For information about a particular connector's properties, see the help for the appropriate connector.</p>
Object	<p>Select a source object.</p> <p>If a list of objects does not appear, click Select.</p> <p>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.</p>
Display Technical Field Names Instead of Labels	<p>Displays technical names instead of business names.</p> <p>Not available for all connection types.</p>
Display Source Fields in Alphabetical Order	<p>Displays source fields in alphabetical order.</p> <p>By default, fields appear in the order returned by the source system.</p>

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**:

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

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.

Tip: All of the properties display in one list, even when the task includes multiple objects. Place your the cursor over the properties to determine which objects the properties apply to. For example, in the following image, the **Enable target bulk load** parameter applies to the target named `TargetSQLAllCust`.

New mt_CustAll

1 Definition 2 Targets 3 Schedule

Validate

Target Parameter Details

TargetSQLAllCust Connection: SQL_Server2012_READONLY_LT View... Advanced...

TgtParameter Object: AUTO_ALL_DT_2012_TRG Select... Create Target...

Operation: Insert

☒ Truncate Target

☒ Enable target bulk load \$TargetSQLAllCust\$

☐ Display target fields in alphabetical order

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** page, configure the following details as required:

Target parameter detail	Description
Connection	<p>Select a connection.</p> <p>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.</p> <p>To create a connection, click New. To edit a connection, click View, and in the View Connection dialog box, click Edit.</p> <p>To configure advanced properties for the connection, click Advanced. Not available for all connection types.</p> <p>For information about a particular connector's properties, see the help for the appropriate connector.</p>
Object	<p>Select a target object.</p> <p>If a list of objects does not appear, click Select.</p> <p>The Select Target 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.</p>
Display Technical Field Names Instead of Labels	<p>Displays technical names instead of business names.</p> <p>Not available for all connection types.</p>
Display Target Fields in Alphabetical Order	<p>Displays target fields in alphabetical order.</p> <p>By default, fields appear in the order returned by the target system.</p>

Target parameter detail	Description
Formatting Options	<p>For Flat File and FTP/SFTP connections only. Select a delimiter and text qualifier. Optionally, select an escape character.</p> <p>If you choose Other for the delimiter, the delimiter cannot be an alphanumeric character or a double quotation mark.</p> <p>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.</p>
Create Target	<p>For flat file and relational database connections only. Creates a target file.</p> <p>Enter a name for the target file.</p> <p>If you want the file name to include a time stamp, select Handle Special Characters and add special characters to the file name, for example, <code>Accounts_%d%m%y%.csv</code>.</p>

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*.

The **In-Out Parameters** page does not appear if the mapping task is based on an elastic mapping.

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 the 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 maplet details.

This displays when a maplet requires a connection and the maplet 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 maplet 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:

Input control type	Description
Text box	Enter any valid value. Note: You cannot use blank spaces. Also, leading and trailing spaces are removed at run time.
Data filter dialog box	To define the template parameter, click New . To create a simple data filter, in the Data Filters dialog box, select a column and operator and enter the value you want to use. To create an advanced data filter, click Advanced . Enter the field expression that you want to use, and click OK . If the template parameter is already defined with a data filter, delete the existing data filter before creating a new data filter. Note: For a template parameter included in a source filter, use an advanced data filter.
Field expression dialog box	To define the template parameter, click New . In the Field Expressions dialog box, enter the expression you want to use and click OK . For more information about configuring field expressions, see "Field expressions" on page 16 .
Field list	To define the template parameter, select a field from the list.
Field mapping dialog box	To define the template parameter, configure the field mappings you want to use: <ul style="list-style-type: none"> - 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. - To match fields with the same name, click Automatch > Exact Field Name. Or, to match fields with similar names, click Automatch > Smart Match. - You can also select and drag the source fields to the applicable target fields. - To clear all field mappings, click Clear Mapping. - To clear mapped field, click the Clear Mapped Field icon for the target field. <p>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 data type, precision, and scale.</p> <p>If the target fields have the same name but different data types, precision, or scale, you can map one of the target fields.</p> <p>If fields from a lookup do not display, configure the lookup connection and object, save the task, then edit the task again.</p>
Custom Dropdown	To define the template parameter, select an option from the list.

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:

Field	Description
Pre-Processing Commands	Commands to run before the task.
Post-Processing Commands	Commands to run after the task completes.
Maximum Number of Log Files	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.

These options do not appear if the mapping task is based on an elastic mapping.

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.

This option does not appear if the mapping task is based on an elastic mapping.
5. Optionally, configure advanced session properties.
 - a. Click **Add**.
 - b. Select an advanced session property.
 - c. Configure the advanced session property.
6. Optionally, if you want to create a parameter file based on the parameters and default values specified in the mapping on which the task is based, click **Download Parameter File Template**.

For more information about parameter file templates, see *Mappings*.
7. Optionally, if the mapping task contains parameters, you can use parameter values from a parameter file. Choose one of the following options:

- To use a parameter file on a local machine, select **Local**. Enter the following information:

Field	Description
Parameter File Directory	Absolute path for the directory that contains the parameter file, excluding the parameter file name. The directory must be accessible by the Secure Agent. If you do not enter a location, the following directory is used: <Secure Agent installation directory>/apps/Data_Integration_Server/data/userparameters
Parameter File Name	Name of the file that contains the definitions and values of user-defined parameters used in the task. You can provide the file name or the relative path and file name in this field.

- To use a cloud-hosted file, select **Cloud Hosted**. Enter the following information about the file:

Field	Description
Connection	Connection where the parameter file is stored. You can use the following connection types: <ul style="list-style-type: none"> - Amazon S3 - Google Storage V2 - Azure Data Lake Store Gen2
Object	Name of the file that contains the definitions and values of user-defined parameters used in the task.

8. Click **Finish**.

CLAIRE Tuning

You can use CLAIRE Tuning to tune a mapping task that is based on an elastic mapping.

CLAIRE, Informatica's AI engine, runs the mapping task several times and uses machine learning to assess the performance of each run. It uses the information to create a tuning recommendation for the set of Spark properties that optimizes task performance.

You can view the tuning recommendation to see a list of recommended Spark properties and their values. You can apply the recommendation to use the values in the mapping task.

The recommendation considers parameters such as the complexity of the elastic mapping, the size of the data, and the processing capacity on the elastic cluster.

Guidelines to get an accurate recommendation

Use the following guidelines to get an accurate recommendation during the tuning job:

- Use sample data that closely matches the actual volume of the data that the mapping task will process.
- Make sure that the mapping logic handles duplicate data in the target. The tuning job will write data to the target multiple times.

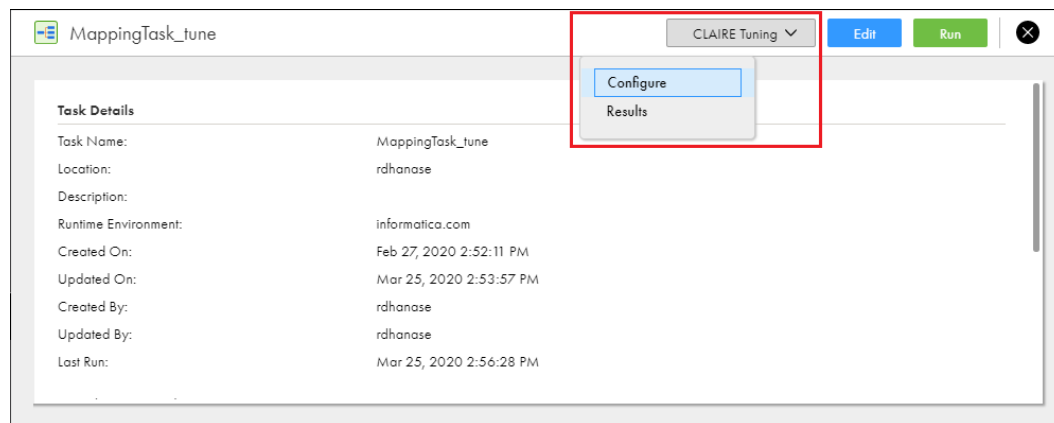
- Set resource limits on your cloud environment by configuring the appropriate Spark properties before you tune the mapping task. Your cloud service provider charges you for the resources that each run uses.

For example, if you know that you can allocate only 4 GB to the Spark driver, you can configure `spark.driver.memory=4G` in the mapping task. CLAIRE will honor the pre-defined Spark property to create a tuning recommendation for other Spark properties.

Configuring tuning

Configure CLAIRE Tuning in the mapping task details.

The following image shows where you can configure tuning in the mapping task details:



Set the number of times that CLAIRE runs the mapping task and begin tuning. When tuning begins, Data Integration creates a tuning job with multiple subtasks to represent each run of the mapping task. You must wait for all subtasks to complete before you can view the tuning results.

Each time that CLAIRE runs the mapping task, CLAIRE gathers task performance data to improve its recommendation for an optimal set of Spark properties.

Viewing tuning results

When tuning is complete, you can view the tuning recommendation and the performance improvement. The improvement is measured in the amount of time that it takes for the mapping task to run using the recommended set of Spark properties.

The following image shows the tuning results for a particular mapping task:

The screenshot shows a 'Tuning Result' dialog box. At the top, it says 'CLAIRE Tuning found a recommendation to optimize performance.' Below this, there's a section titled 'Performance Improvements' which lists: 'Difference in task duration: 37.34%', 'Estimated task duration without tuning: 158', and 'Estimated task duration with tuning: 99'. The next section is 'Tuning Recommendation', which contains a table with three columns: 'Property Name', 'Property Value', and 'Tuning Recommendation'. The table lists four properties: 'spark.executor.memory' (6G), 'spark.rdd.compress' (true), 'spark.shuffle.file.buffer' (8k), and 'spark.executor.cores' (1). At the bottom right, it says 'POWERED BY CLAIRE™'. At the bottom left, there's a question mark icon. At the bottom right, there are two buttons: 'Apply Tuning Recommendation' and 'Cancel'.

Property Name	Property Value	Tuning Recommendation
spark.executor.memory		6G
spark.rdd.compress		true
spark.shuffle.file.buffer		8k
spark.executor.cores		1

Apply the tuning recommendation to use the Spark property values in the mapping task.

Guidelines to apply a tuning recommendation

Use the following guidelines when you apply a tuning recommendation to make sure that job performance is optimal:

- Use the full set of Spark properties to achieve the performance improvement. Using a partial set of the recommended Spark properties might not be optimal.
- Do not edit the Spark properties in the mapping task in between the time that you begin tuning and the time that you apply the tuning recommendation. If you make significant changes to the Spark properties, tune the mapping task again.

Viewing and editing 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.
- Pre-processing and post-processing commands.

To view details for a mapping task, perform the following steps:

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.

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 in to 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 add sources to a synchronization task when you configure the task or when you perform data catalog discovery. If you add a source when you configure the task and the source connection is a database connection, you can add a single object or multiple related objects.

You can add a source to a synchronization task in the following ways:

When you configure the task

Select the source connection and source object on the **Source** tab when you configure the synchronization task.

If the source connection is a database connection, you can use a single object as a source or multiple related objects as sources. Define relationships based on key columns or create a user-defined join condition.

Through data catalog discovery

If your organization administrator has configured Enterprise Data Catalog integration properties, you can perform data catalog discovery to find the source object in the catalog.

Search for the source object on the **Data Catalog** page, select the object in the search results, and then add it to a new synchronization task.

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.
```

- You can use a tab or any printable special character as a flat file delimiter. The delimiter must be different from the escape character and text qualifier.
- 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 data types or create field expressions to convert data types 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:

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

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

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

Field data types

When you create a synchronization task, Data Integration assigns a data type 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 data types 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:

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

The synchronization task performs the following lookup:

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

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:

```
$OutputField+100
```

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

```
'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:

Lookup Return Value Property	Description
Output Field	The field from the lookup table that you want to use.
Multiplicity	How Data Integration handles multiple return values. <ul style="list-style-type: none">- 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 \$OutputField to represent the selected output field. By default, Data Integration passes the lookup return value without alteration with the following expression: \$OutputField.

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 data types. If the data types are not compatible, the following error appears:

```
Source field [<source field name> (<source field data type>)] and lookup field  
[<lookup field name> (<lookup field data type>)] have incompatible data types.
```

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 data type. The synchronization task uses the larger precision and scale of the source field data types as the precision and scale for the target field. If the source fields do not have the same data type, the following error appears:

```
Lookup field <field name> in <file name> has conflict data types inferenced: <data  
type 1> and <data type 2>.
```

- You cannot include lookup fields of particular data types in a lookup condition. When the lookup field in a flat file has the Text or Ntext data type or the target field of a lookup has the Text or Ntext data type, 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:

Field	Description
Task Name	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.
Location	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.
Description	Description of the synchronization task. Maximum length is 4000 characters.
Task Operation	Select one of the following task operation types: <ul style="list-style-type: none">- Insert- Update- Upsert- Delete The list of available targets in a subsequent step depend on the operation you select.

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:

Field	Description
Connection	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 .
Target Object	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 .
Display Technical Field Names Instead of Labels	Displays technical names instead of business names. Not available for all connection types.
Display Target Fields in Alphabetical Order	Displays target fields in alphabetical order instead of the order returned by the target system.
Formatting Options	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.
Create Target	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.
Truncate Target	Database targets with the Insert task operation only. Truncates a database target table before inserting new rows. <ul style="list-style-type: none">- True. Truncates the target table before inserting all rows.- False. Inserts new rows without truncating the target table. Default is False.
Enable Target Bulk Load	Select this option to write data in bulk mode. The default value is false.

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 `$$Sales < 100000` instead of `100000 > $Sales`.

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

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

Data sorting option	Description
Object	Source object.
Sort By	Source field to use to sort data.
Sort Direction	Sort direction: <ul style="list-style-type: none">- ASC. Ascending order.- DESC. Descending order.

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.
 - data type of a field.
3. Some source types allow you to configure field data types. To configure field data types 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 data type attributes and click **OK**:

data type attribute	Description
Datatype	The data type of data in the column.
Precision	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.
Scale	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 data type is 65535.

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 the same name, click **Automatch > Exact Field Name**. Or, to match fields with similar names, click **Automatch > Smart Match**.

You can also 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 data types 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 data type attributes and click **OK**:

data type attribute	Description
Datatype	The data type of data in the column.
Precision	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.
Scale	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 data type is 65535.

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**:

Option	Description
Lookup Connection	Connection for the lookup object.
Lookup Object	Object on which you want to lookup a value.
Display Technical Field Names Instead of Labels	Displays technical names instead of business names. Not available for all connection types.
Display Fields in Alphabetical Order	Displays lookup fields in alphabetical order. By default, fields appear in the order returned by the lookup system.
Source Fields	Source column to use in the lookup condition.

Option	Description
Lookup Fields	The column in the lookup table to use in the lookup condition.
Output Field	The column in the lookup table that contains the output value.
Multiplicity	Determines how to handle cases when a lookup returns multiple values: <ul style="list-style-type: none"> - 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:

Advanced options	Description
Preprocessing Commands	Command to run before the task.
Postprocessing Commands	Command to run after the task completes.
Parameter File Name	Name of the file that contains the definitions and values of user-defined parameters used in the task.
Maximum Number of Log Files	Number of session log files, error 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.
Update Columns	Database targets only. Temporary primary key columns to update target data. If the database target does not include a primary key column, and the task performs an update, upsert, or delete task operation, click Add to add a temporary key.
Upsert Field Name	The target field to use to perform upserts.

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 help for Salesforce Connector.

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 nonreplicable within Salesforce.
If you run an incremental load on a replication task that contains nonreplicable objects, the replication task runs a full load on the object. Contact the Salesforce administrator to get a list of replicable Salesforce objects.
- The data type of a Salesforce field changed.
If the replication task detects a data type 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 data types are not compatible. For more information about the `AutoAlterColumnType` property, see the help for Salesforce Connector.

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:

```
TenLetters1234567890TenLettersXXXXXXXXXX
```

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:

Property	Description
Task Name	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.
Location	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.
Description	Description of the task. Maximum length is 255 characters.

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:

Property	Description
Connection	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 .
Target Prefix	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_.
Load Type	Type of load. Select one of the following options: <ul style="list-style-type: none"> - Incremental loads after initial full load. Loads all data the first time the task runs. In subsequent runs, loads changed data only. - Incremental loads after initial partial load. 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. - Full Load each run. Loads all data every time the task runs. <p>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.</p>

Property	Description
Delete Options	Select one of the following options: <ul style="list-style-type: none"> - Remove Deleted Columns and Rows. Deletes columns and rows from the target if they no longer exist in the source. - Retain Deleted Columns and Rows. Retains columns and rows in the target that were removed from the source.
Commit Size	Number of rows to commit. Default for full load replication is 5,000 rows. Default for incremental load replication is 999,999,999.

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:

Property	Description
Row Limit	Select one of the following options: <ul style="list-style-type: none"> - Process all Rows. Replicates all rows of the source. - 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. <p>You cannot specify a row limit on replication tasks with non-Salesforce sources. If you select a non-Salesforce source, the option is disabled.</p>
Data Filters	Click New to create a data filter on a Salesforce or database source. You can create simple or advanced data filters.

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.

Advanced option	Description
High Precision Calculations	In calculated fields, allows for precisions of up to 28. Recommended for Salesforce calculation fields.
Use Float Semantic	When enabled, the task uses a target-specific floating point data type.
Preprocessing Commands	Commands to run before the task.
Postprocessing Commands	Commands to run after the task completes.
Maximum Number of Log Files	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.

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.

CHAPTER 5

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 and the target endpoints for file transfer. To enhance the performance of a mass ingestion task, you can define the number of files a mass ingestion task can transfer in a batch.

You can define a schedule by which the task runs. You can also configure the task to perform actions, such as compression, decompression, encryption, or decryption of files.

Mass ingestion task sources

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

You can add a source to a mass ingestion task in the following ways:

Through data catalog discovery

If your organization administrator has configured Enterprise Data Catalog integration properties, you can perform data catalog discovery to find the source object in the catalog. You can discover Amazon S3 V2, Microsoft Azure Blob Storage V3, Hadoop Files V2 objects to use as sources in new mass ingestion tasks.

Search for the source object on the **Data Catalog** page, select the object in the search results, and then add it to a new mass ingestion task.

When you configure the task

Select the source connection and source object on the **Source** tab when you configure the mass ingestion task.

Mass Ingestion supports the following source types:

- Source Connection— Supports the following sources:
 - Local folder
 - Advanced FTP V2
 - Advanced FTPS V2
 - Advanced SFTP V2
 - Amazon S3 V2

- Google Cloud Storage V2
- Hadoop Files V2
- Microsoft Azure Blob Storage V3
- Microsoft Azure Data Lake Store Gen2
- Microsoft Azure Data Lake Store V3
- Databricks Delta
- File Listener— Use a file listener component as a source.

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 V2
- Advanced FTPS V2
- Advanced SFTP V2
- Amazon S3 V2
- Amazon Redshift V2
- Google BigQuery V2
- Google Cloud Storage V2
- Hadoop Files V2
- Microsoft Azure Blob Storage V3
- Microsoft Azure Data Lake Store Gen2
- Microsoft Azure Data Lake Store V3
- Microsoft Azure SQL Data Warehouse V3
- Snowflake Cloud Data Warehouse V2
- Databricks Delta

Mass ingestion task schedules

You can manually run a mass ingestion task. You can also schedule the task to run at a specific time or when a file is ready. For more information about scheduling a mass ingestion task, refer to [“Defining the schedule and adding file actions” on page 94](#).

File processing actions

You can apply actions on the files that the mass ingestion task transfers.

Use one or more of the following file processing actions:

- **Encryption.** Uses the Pretty Good Program (PGP) method to encrypt files. The mass ingestion task encrypts files and flattens the file structure in the target directory.
For more information about securing the files that the mass ingestion task transfers, refer to [“File security” on page 87](#).
- **Decryption.** Uses the PGP method to decrypt files. The mass ingestion task decrypts files and flattens the file structure in the target directory.
For more information about securing the files that the mass ingestion task transfers, refer to [“File security” on page 87](#).
- **Compression.** Use one of the following file compression methods: Zip, Tar, or Gzip to compress files. The mass ingestion task compresses files and flattens the file structure in the target directory.
- **Decompression.** Use one of the following file decompression methods: Unzip, Untar, or Gunzip to decompress files. The mass ingestion task decompresses files and flattens the file structure in the target directory.
- **Flatten file structure.** Moves the files from multiple folders to a single folder in the target directory. Flattening of files might result in the loss of files if the same file name exists in different folders. The session log displays the overridden files.
- **Virus scan.** Reviews and identifies viruses in the files that the mass ingestion task transfers. Mass ingestion uses the Internet Content Adaptation Protocol (ICAP) to scan files and detect malwares. The ICAP server scans the files and sends a response code as 200 when the scan does not identify any virus in the files. The mass ingestion task fails when the scan detects virus.

Note: Use the ICAP server of the organization.

Note: The mass ingestion task retains the file structure if you do not configure any action.

The mass ingestion task performs the file processing actions that you configure in a sequential order. For more information about adding file processing actions to a mass ingestion task, see [“Defining the schedule and adding file actions” on page 94](#).

For example, consider that you want to transfer files from a local repository to an FTP server. You want to compress the files and encrypt the compressed file before you transfer.

In this scenario, add the following file processing actions:

1. Compression with an action type of Zip.
2. Encryption with an action type of PGP.

File security

Use the encryption and decryption methods to secure files that a mass ingestion task transfers.

You can apply the encryption and decryption methods as actions to perform on the files. For more information about actions that you can perform on the files that the mass ingestion task transfers, see [“File processing actions” on page 87](#).

Mass ingestion tasks uses the PGP method to encrypt and decrypt files. You use the key ID to encrypt files and key passphrase to decrypt files. If you have the privilege to update files in the agent location, you can use

the key ring command line interface (CLI) to manage key IDs. For more information about using key ring CLIs, see [“Key ring command reference” on page 95](#). If you do not have permissions to access the agent location, ask your administrator for the key ID and key passphrase.

Note: GPG method is compatible with the PGP method to encrypt and decrypt files.

Encryption

When you configure the encryption action for a mass ingestion task, you provide a key ID. The key ID is the public key ID of the receiver who decrypts the file. You can also add your private key ID and key passphrase to sign the files.

Decryption

When you configure the decryption action for a mass ingestion task, you provide a key passphrase. The key passphrase is the private key passphrase of the receiver who decrypts the file.

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 88](#) 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:

Property	Description
Task Name	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.
Location	Project folder in which the task resides.
Description	Optional description of the task. Maximum length is 1024 characters.
Runtime Environment	Runtime environment that contains the Secure Agent used to run the task. The Mass Ingestion application must run on the Secure Agent.

3. Click **Next**.

Configuring the source

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

Options that appear on the **Source** page of the task wizard vary based on the type of source that you select. Refer to the following sections for configuration information:

- For information about local folder sources, see [“Configuring a local folder source” on page 89](#).
- For information about configuring a file listener as a source, see the file listener configuration in the *Components* help module.
- For information about other sources, see the applicable connector help modules.

Configuring a local folder source

Perform the following steps to configure a local folder as a source connection:

1. Select a task and click **Source**. In the **Source Type** list select **Source Connection**.
2. In the **Connection Type** list select **Local Folder**.
3. Configure the following **Source Options**:

Option	Description
Source Directory	Directory from where files are transferred. The Secure Agent must be able to access the directory. The use of slashes around the source folder path differs between connectors. Using slashes incorrectly will result in connection failures. For more information, see the Knowledge Base article 625869 .
Include files from sub folders	Transfer files from all sub folders under the defined source directory.

Option	Description
Skip Duplicate Files	Indicates whether to skip duplicate files. If you select this option, the mass ingestion task does not transfer the files that have the same name and creation date. The mass ingestion task marks files as duplicate in the job log. If you do not select this option, the task transfers all files.
Transfer Mode	<p>The data type to use when the files are transferred. This option is applicable if you transfer files from the FTP or FTPS sources.</p> <p>The following options are available:</p> <ul style="list-style-type: none"> - Auto. Determines the mode of transfer based on file extensions, type of content in the file, or both. - ASCII. Transfers files with the following extensions as text: .txt, .asp, .html, and .php - Binary. Transfers files with the following extensions as raw data: .wav, .jpg, .gif, and mp3 <p>Default is Auto.</p>
File Pattern	<p>File name pattern used to select the files to transfer. Based on the file pattern that you have selected, enter the file name patterns.</p> <p>The following file patterns are available:</p> <ul style="list-style-type: none"> - Wildcard. Use the following wildcard character filters: <ul style="list-style-type: none"> - An asterisk (*) matches any number of characters. - A question mark (?) matches a single character. - Regex. Use regular expression to match the pattern type. Consider the following samples: <ul style="list-style-type: none"> - <code>^(?!.*(?:out baz foo)).*\$</code> all except Identifies all files except for files whose name contains out, foo, and baz. - <code>([a-zA-Z0-9\s_\.\-\\(\):])+(\.doc \.docx \.pdf)\$</code> Identifies all files that have an extension of doc, docx, or pdf. - <code>^(?!out).*\.txt\$</code> Identifies all text files except for files whose name contains out.txt.
File Date	<p>Filters files to transfer based on the file modification date and time.</p> <p>The following options are available:</p> <ul style="list-style-type: none"> - Greater than or Equal. Filters files that are modified after or on the specified date and time. Click the calendar to select the date and the clock to select the time. - Less than or Equal. Filters files that are modified before or on the specified date and time. Click the calendar to select the date and the clock to select the time. - Equal. Filters files that are modified on the specified date and time. Click the calendar to select the date and the clock to select the time. - Days before today. Filters files that are modified within the specified number of days before the current date. The current date calculation starts from 00:00 hours. Enter the number of days. For example, you configure the mass ingestion task to run weekly. To filter the files that were modified in the previous week set the value to seven.

Option	Description
Time Zone	Timezone for the specified time. Select the time zone based on the location of the source file.
File Size	<p>Filters files to transfer based on the size of files. Enter the file size, select the file size unit, and filter options.</p> <p>The following filter options are available:</p> <ul style="list-style-type: none"> - Greater than or Equal. Filters files that are greater than or equal to the specified size. - Less than or Equal. Filters files that are that are lesser than or equal to the specified size. - Equal. Filters files that are of the specified size.
Batch Size	<p>The maximum number of files a mass ingestion task transfers in a batch. Default is 5.</p> <p>The maximum batch size varies, based on the following conditions:</p> <ul style="list-style-type: none"> - If the task transfers files from source to target with no intermediate staging, the maximum number of files the task can transfer in a batch is 8000. - If the task transfers files from source to target with intermediate staging, the maximum number of files the task can transfer in a batch is 20. - If the task transfers files from any source to a Snowflake target, the maximum number of files the task can transfer in a batch is 1000. <p>Consider the following guidelines when you define the batch size:</p> <ul style="list-style-type: none"> - The task transfers files with no intermediate staging in the following scenarios: <ul style="list-style-type: none"> - File transfers from Amazon S3 to Amazon Redshift when Amazon Redshift Connector is configured to upload files with no intermediate staging - File transfers from Google Cloud Storage to Google BigQuery - File transfers from Azure Blob to Microsoft Azure Data Warehouse - File transfers from Amazon S3 and from Azure Blob to Snowflake - When you use a command line to transfer files, the task transfers files with intermediate staging.
After File Pickup	<p>Determines what to do with source files after the files transfer.</p> <p>The following options are available:</p> <ul style="list-style-type: none"> - Keep files in the source directory. - Delete files from the source directory. - Rename 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.

4. Click **Next**.

The **Target** page appears.

Configuring a file listener source

Configure a file listener as a source type when you use the file listener to trigger the mass ingestion task.

To configure a file listener as a source, you must create a file listener by using the Components service. For more information about creating a file listener, see the *Components* help.

Note: You cannot run the mass ingestion task with a file listener as a source from the mass ingestion user interface. A mass ingestion task with a file listener as a source runs automatically when the file listener starts.

1. Select a task and click **Source**.

2. Click **File Listener** in the Source Type.
3. Select a file listener from the **File Listener Details** list.
The file listener configuration appears.
4. Configure the following options:

Option	Description
Skip Duplicate Files	Indicates whether to skip duplicate files. If you select this option, the mass ingestion task does not transfer the files that have the same name and creation date. The mass ingestion task marks files as duplicate in the job log. If you do not select this option, the task transfers all files.
File Pattern	<p>File name pattern used to select the files to transfer. Based on the file pattern that you have selected, enter the file name patterns.</p> <p>The following file patterns are available:</p> <ul style="list-style-type: none"> - Wildcard. Use the following wildcard character filters: <ul style="list-style-type: none"> - An asterisk (*) matches any number of characters. - A question mark (?) matches a single character. - Regex. Use regular expression to match the pattern type. Consider the following samples: <ul style="list-style-type: none"> - <code>^(?!.*(?:out baz foo)).*\$</code> all except Identifies all files except for files whose name contains out, foo, and baz. - <code>([a-zA-Z0-9\s_\.\-\\(\):])+(\.doc \.docx \.pdf)\$</code> Identifies all files that have an extension of doc, docx, or pdf. - <code>^(?!out).*\.txt\$</code> Identifies all text files except for files whose name contains out.txt.
File Date	<p>Filters files to transfer based on the file modification date and time.</p> <p>The following options are available:</p> <ul style="list-style-type: none"> - Greater than or Equal. Filters files that are modified after or on the specified date and time. Click the calendar to select the date and the clock to select the time. - Less than or Equal. Filters files that are modified before or on the specified date and time. Click the calendar to select the date and the clock to select the time. - Equal. Filters files that are modified on the specified date and time. Click the calendar to select the date and the clock to select the time. - Days before today. Filters files that are modified within the specified number of days before the current date. The current date calculation starts from 00:00 hours. Enter the number of days. For example, you configure the mass ingestion task to run weekly. To filter the files that were modified in the previous week set the value to seven.
Time Zone	Timezone for the specified time. Select the time zone based on the location of the source file.

Option	Description
File Size	Filters files to transfer based on the size of files. Enter the file size, select the file size unit, and filter options. The following filter options are available: <ul style="list-style-type: none"> - Greater than or Equal. Filters files that are greater than or equal to the specified size. - Less than or Equal. Filters files that are that are lesser than or equal to the specified size. - Equal. Filters files that are of the specified size.
Batch Size	The maximum number of files a mass ingestion task transfers in a batch. Default is 5.

5. 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 93](#). For information about other targets, see the help for the applicable connector.

Configuring a local folder target

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

Option	Description
Target Directory	Directory to where files are transferred. The Secure Agent must be able to access the directory.
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: <ul style="list-style-type: none"> - Overwrite - Append Timestamp
Transfer Mode	The data type to use when the files are transferred. This option is applicable if you transfer files from the FTP or FTPS sources. <p>The following options are available:</p> <ul style="list-style-type: none"> - Auto. Determines the mode of transfer based on file extensions, type of content in the file, or both. - ASCII. Transfers files with the following extensions as text: .txt, .asp, .html, and .php - Binary. Transfers files with the following extensions as raw data: .wav, .jpg, .gif, and mp3 Default is Auto.

3. Click **Next**.

The **Schedule** page appears.

Defining the schedule and adding file actions

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. You can also define actions, such as encryption, decryption to process files.

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 notifies the task of a file event. A file event occurs when files arrive to the monitored folder or the files in the monitored folder are updated or deleted.
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.
2. Click **Add Action**.
The **Action Details** window appears.
3. Perform one or more of the following file processing actions:

Action	Description
Compress	Select Compress , and select one of the following file compression methods: Zip, Tar, or Gzip.
Decompress	Select Decompress , and select one of the following file decompression methods: Unzip, Untar, or Gunzip. Note: To decompress a file, ensure that you use the method that corresponds to the compression method that you used to compress the file. For example, if you use the Zip method for compression, use the Unzip method for decompression.
Encrypt	Select Encrypt , select PGP , and enter the key ID of the user who decrypts the file. <ul style="list-style-type: none">- To add your sign key, select Sign. The key ID and the key passphrase enables.- Enter your private key ID and key passphrase. Note: For more information about securing files that mass ingestion transfers, see "File security" on page 87 .
Decrypt	Select Decrypt , select PGP , and enter the key passphrase of the user of the target directory.
Flatten file structure	Select File Structure , and then select Flatten to move the files from multiple folders to a single folder at the target directory.
Virus scan	Select Virus Scan , and then select ICAP . Enter the ICAP Server URL that scans the files and sends a response code that indicates whether the malwares are detected in the files.

4. Click **OK**.
5. To add another action, click **Add Action**.
6. To delete an action, click **Delete**.
7. To change the order in which the mass ingestion task processes files, use arrows to change the sequence of actions.
8. 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.

Key ring command reference

A mass ingestion task encrypts and decrypts files using the Pretty Good Program (PGP) method. An Informatica Intelligent Cloud Services administrator uses the command line interface (CLI) to create key IDs and key passphrases. The administrator can then share them with the Informatica Intelligent Cloud Services user to encrypt and decrypt files.

You can run the key ring commands if you have the privileges to update files in the agent location. A PGP configuration file is created when you install the agent. The PGP configuration file consists of the properties that lists the location of the public key ring and the secret key ring. You must update the properties to change the location of the existing key ring. For more information about updating the properties, see the *Administrator* help.

The default location of the PGP configuration file is `<agent location>/apps/MassIngestionRuntime/<latest version no>/conf/pgp-configuration.properties`.

Use the `createKeyRing` command to create a key ring in the key ring location that is defined in the PGP configuration file.

To create key IDs and add them to the key ring, use the `createKeyPair` command. A key ID consists of a public key and a private key. To import public keys from different partners or use an existing key pair and import it to the current agent key ring location, use the `importKeys` command.

createKeyRing

Creates a key ring. A key ring consists of a public key ring and a secret key ring.

If the key ring exists, the command displays an error indicating that a key ring already exists.

The createKeyRing command uses the following syntax:

```
<--command|-c> createKeyRing
```

The following sample command creates a key ring and saves the key ring in the location that is defined in the PGP configuration file:

```
./pgp_cli.sh -c createKeyRing
```

The command displays the following output:

```
KeyRing created successfully
```

createKeyPair

Creates a key pair. The key pair or the key ID consists of a public key and a private keys.

The createKeyPair command uses the following syntax:

```
<--command|-c> createKeyPair
<--name|-n> key_name
<--passphrase|-p> passphrase
[<--type|-t> type]
[<--size|-s> size]
<--expiration|-e> expiration_date>
<--email|-m> email
```

The following table describes createKeyPair options and arguments:

Option	Argument	Description
--name -n	key_name	Required. The name of the key pair.
--passphrase -p	passphrase	Required. The passphrase of the PGP key.
--type -t	type	Optional. The type of PGP key. The following types of PGP keys are supported: - RSA - Diffie Hellman/DSS Default is RSA.
--size -s	size	Optional. The size of the PGP key in bits. Enter one of the following values: - 512 - 1024 - 2048 - 4096 Default is 512.

Option	Argument	Description
--expiration -e	expiration_date	Required. The date when the PGP key pair expires. Use the following date format: dd-mm-yyyy
--email -m	email	Required. The email ID of the user.

The following sample command creates a key pair and adds the key pair to the key ring.

```
./pgp_cli.sh -c createKeyPair -n Mykeypair -p Mykeypassphrase -t RSA -s 1024 -e
10-12-2019 -m abc@informatica.com
```

The command displays the following output:

```
12:09:2017 INFO Key pair was successfully created and added to your key ring. The key
ID is '0x23149FC8C38658EA'.
12:09:2017 INFO Key Pair created successfully.
```

listKeys

Lists all keys in key ring.

The listKeys command uses the following syntax:

```
<--command|-c> listKeys
```

The following sample command lists keys that are in the key ring:

```
./pgp_cli.sh -c listKeys
```

The command displays the following output:

```
12:10:38 INFO Default system locale: English (United States)
12:10:38 INFO Listing Keys.
12:10:38 INFO Total keys : 2
Key ID : 0x23149FC8C38658EA User : Mykeypair <abc@informatica.com./pgp_cli.sh>
Description : Key Pair Key Type : RSA Key Size : 1024 Expiration Date : Tue Dec 10
23:59:59 IST 2019

Key ID : 0x7B1E52AFB29030A6 User : new <a@b.com> Description : Key Pair Key Type : RSA
Key Size : 1024 Expiration Date : Sat Sep 28 23:59:59 IST 2019
```

importKeys

Imports keys from an external file to the key ring.

To import public keys from an external file or to use an existing key pair and import it to the current agent key ring location, use the importKeys command.

The importKeys command uses the following syntax:

```
<--command|-c> importKeys <--location|-l> location
```

The following table describes importKeys options and arguments:

Option	Argument	Description
-location -l	location	Required. The file name and location of the file that contains key pairs or public keys to import.

The following sample command imports keys from the key pair to the key ring:

```
./pgp_cli.sh -c importKeys -l /root/RSFiles/SubFolder1/SubFolder2/file1.asc
```

The command displays the following output:

```
12:37:09 INFO Default system locale: English (United States)
12:37:10 INFO Importing Keys.
12:37:10 INFO Public key '0x23149FC8C38658EA' with user ID 'doctest
<abc@informatica.com./pgp_cli.sh>' was imported successfully.
12:37:10 INFO 1 public keys and 0 secret keys were successfully imported into your key
ring.
12:37:10 INFO Import Finished.
```

exportKeyPairs

Exports key pairs from the key ring to a file.

The exportKeyPairs command uses the following syntax:

```
<--command|-c> exportKeyPairs
<--ids|-i> list_of_key_ids
<--location|-l> location
```

The following table describes exportKeyPairs options and arguments:

Option	Argument	Description
--ids -i	list_of_key_ids	Required. Comma-separated list of key IDs in the key ring.
--location -l	location	Required. The file name and location of the file to export key pairs from the key ring.

The following sample command exports key pairs from the key ring to a local repository:

```
./pgp_cli.sh -c exportKeyPairs -i 0x23149FC8C38658EA -l /root/RSFiles/SubFolder1/file.asc
```

The command displays the following output:

```
12:28:18 INFO Default system locale: English (United States)
12:28:18 INFO Exporting Key Pairs.
12:28:18 INFO Export Finished.
```

exportPublicKeys

Exports public keys from the key ring to a file.

The exportPublicKeys command uses the following syntax:

```
<--command|-c> exportPublicKeys  
<--ids|-i> list_of_key_ids  
<--location|-l> location
```

The following table describes exportPublicKeys options and arguments:

Option	Argument	Description
--ids -i	list_of_key_ids	Required. Comma-separated list of PGP key IDs in the key ring.
--location -l	location	The file name and location file to export public key from the key ring.

The following sample command exports public keys to a local repository:

```
./pgp_cli.sh -c exportPublicKeys -i 0x23149FC8C38658EA -l /root/RSFiles/SubFolder1/  
SubFolder2/file1.asc
```

The command displays the following output:

```
12:32:10 INFO Default system locale: English (United States)  
12:32:10 INFO Exporting Public Keys.  
12:32:10 INFO Export Finished.
```

deleteKeys

Deletes keys from the key ring.

The deleteKeys command uses the following syntax:

```
<--command|-c> deleteKeys <--ids|-i> list_of_key_ids
```

The following table describes deleteKeys options and arguments:

Option	Argument	Description
--ids -i	list_of_key_ids	Required. Comma-separated list of key IDs in the key ring.

The following sample command deletes keys:

```
./pgp_cli.sh -c deleteKeys -i 0x23149FC8C38658EA
```

The command displays the following output:

```
12:36:46 INFO Default system locale: English (United States)  
12:36:46 INFO Deleting Key.  
12:36:47 INFO Key '0x23149FC8C38658EA' was deleted  
12:36:47 INFO Delete Finished.
```

changePassphrase

Changes the passphrase of the key.

The changePassphrase command uses the following syntax:

```
<--command|-c> changePassphrase  
<--ids|-i> key_id  
<--old-passphrase|-o> old_passphrase  
<--passphrase|-p> new_passphrase
```

The following table describes changePassphrase options and arguments:

Option	Argument	Description
--ids -i	key_id	Required. Comma-separated list of PGP key IDs in the key ring.
--old-passphrase -o	old_passphrase	Required. The old passphrase of the PGP key ring.
--passphrase -p	new_passphrase	Required. The new passphrase of the PGP key ring.

The following sample command replaces the old key passphrase to the new key passphrase:

```
./pgp_cli.sh -c changePassphrase -i 0xDA70CEEDF703DCBE -o Mykeypassphrase -p  
Mynewkeypassphrase
```

The command displays the following output:

```
12:46:36 INFO Default system locale: English (United States)  
12:46:36 WARN Unable to load pgp configuration file : ./conf/pgp-  
configuration.properties (No such file or directory)  
12:46:36 INFO Changing Key Pair.  
12:46:36 INFO Passphrase for the key '0xDA70CEEDF703DCBE' was changed successfully.  
Please make sure to save this passphrase in a secure place.  
12:46:36 INFO Key Passphrase changed successfully.
```

CHAPTER 6

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:

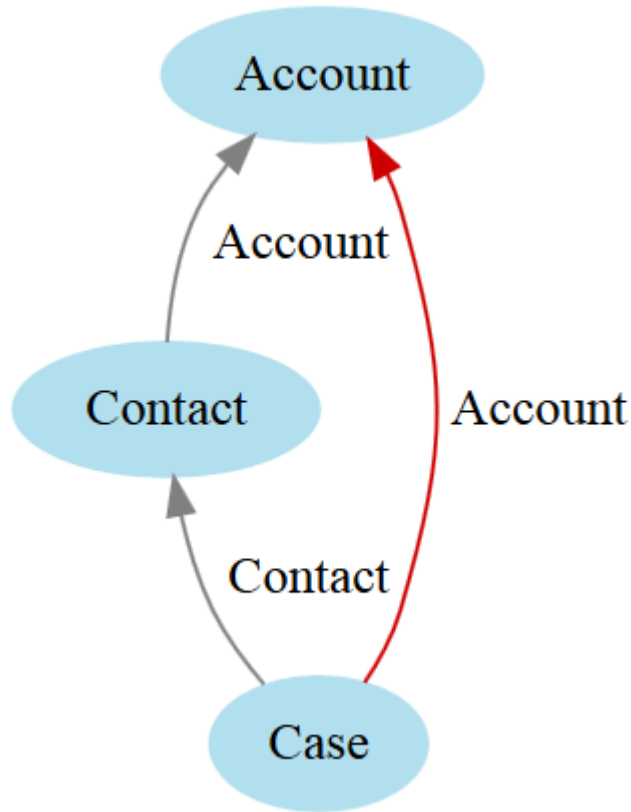
```
<filename>.dot
```

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.

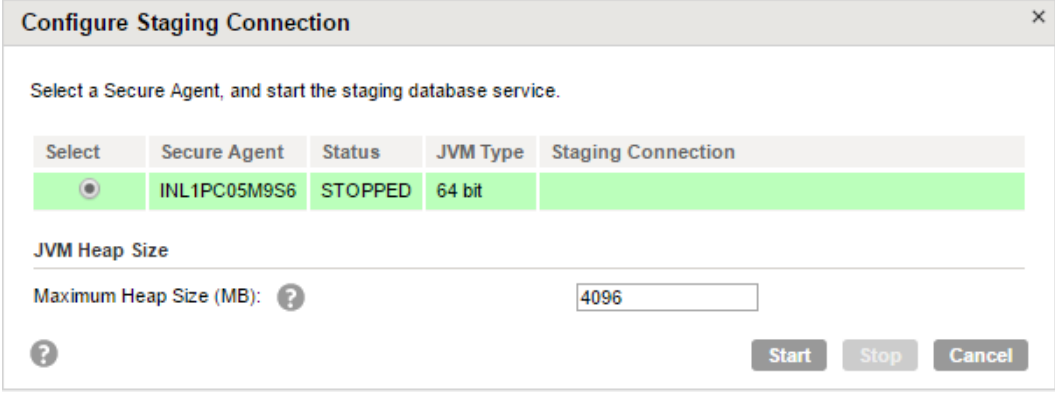
You can view the subset record count in the logs. The task reads the record IDs that are selected for subset from the staging database and uses standard API to read those records from the source. The task uses standard or bulk API to load data into the target.

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:



The dialog box titled "Configure Staging Connection" contains the following elements:

- Instruction: "Select a Secure Agent, and start the staging database service."
- Table with 5 columns: Select, Secure Agent, Status, JVM Type, and Staging Connection.
- Row 1: A radio button in the "Select" column, the text "INL1PC05M9S6" in the "Secure Agent" column, the text "STOPPED" in the "Status" column, the text "64 bit" in the "JVM Type" column, and an empty field in the "Staging Connection" column.
- Section: "JVM Heap Size"
- Text: "Maximum Heap Size (MB):" followed by a question mark icon and a text input field containing "4096".
- Buttons: "Start", "Stop", and "Cancel".

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:

```
@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. If you estimate the subset and then run the task, 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 before you run the task. 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.

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 updated data, you must first reset the task. When you reset the task, the task status returns to the start stage. You can then estimate the subset and then run the task or directly run the task. In both cases, the task runs through staging and estimation stages and then loads the tables into the target.

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: \$\$<parameter>

Save the parameter file local to the following directory:

```
<Secure Agent installation directory>/apps/Data_Integration_Server/data/userparameters
```

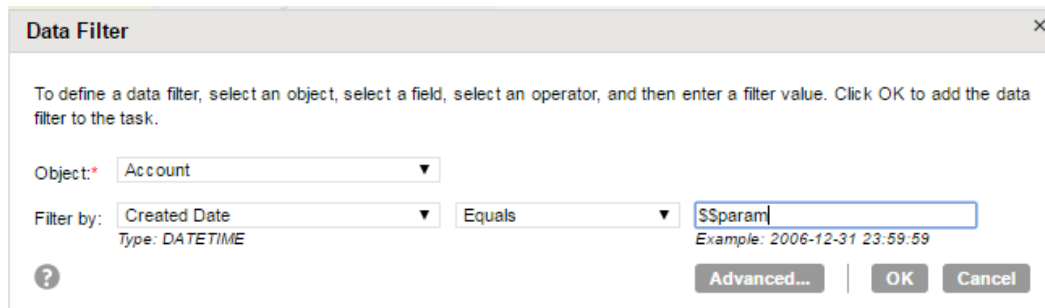
You can specify the parameter file name on the Schedule page of the task wizard. The parameter values are applied when the task runs.

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 \$\$param. Then create a parameter file with the following content:

```
$$param=('1991-10-03')
```

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



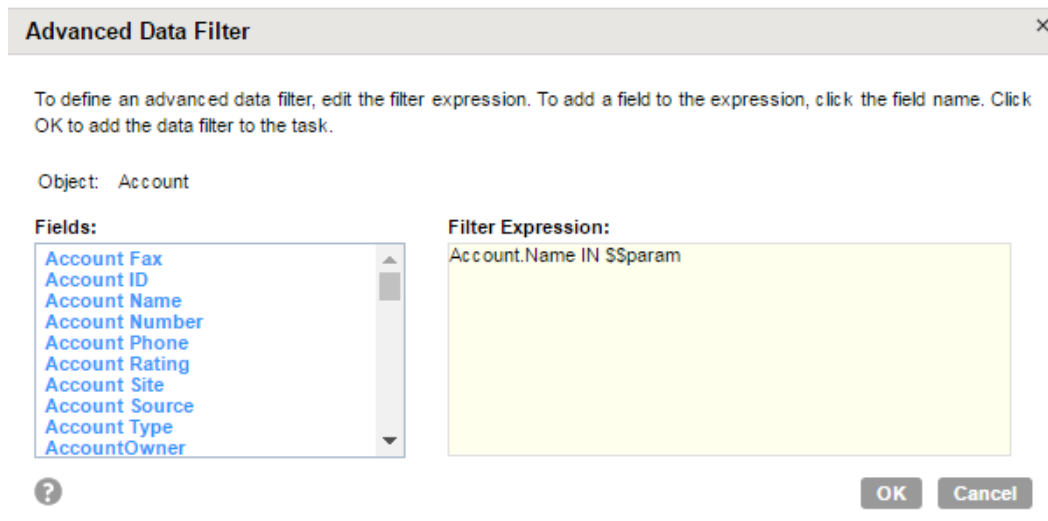
The 'Data Filter' dialog box contains instructions: 'To define a data filter, select an object, select a field, select an operator, and then enter a filter value. Click OK to add the data filter to the task.' It features three dropdown menus: 'Object:' set to 'Account', 'Filter by:' set to 'Created Date' (with 'Type: DATETIME' below it), and an operator set to 'Equals'. A text input field contains '\$\$param' with an example '2006-12-31 23:59:59' below it. At the bottom are buttons for '?', 'Advanced...', 'OK', and 'Cancel'.

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 \$\$param. Then create a parameter file with the following content:

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

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

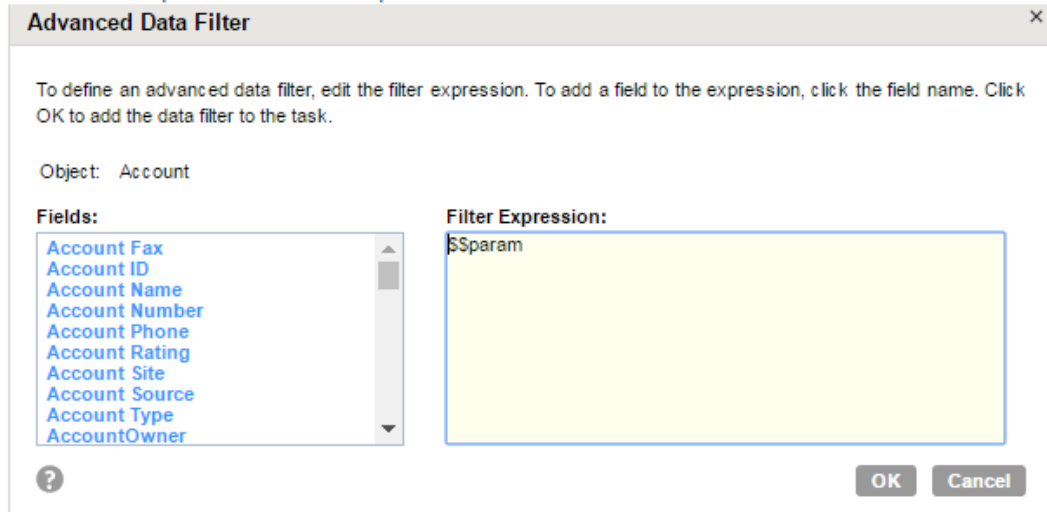


The 'Advanced Data Filter' dialog box contains instructions: 'To define an advanced data filter, edit the filter expression. To add a field to the expression, click the field name. Click OK to add the data filter to the task.' It shows 'Object: Account'. On the left, a 'Fields:' list includes 'Account Fax', 'Account ID', 'Account Name', 'Account Number', 'Account Phone', 'Account Rating', 'Account Site', 'Account Source', 'Account Type', and 'AccountOwner'. On the right, the 'Filter Expression:' text area contains 'Account.Name IN \$\$param'. At the bottom are buttons for '?', 'OK', and 'Cancel'.

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:



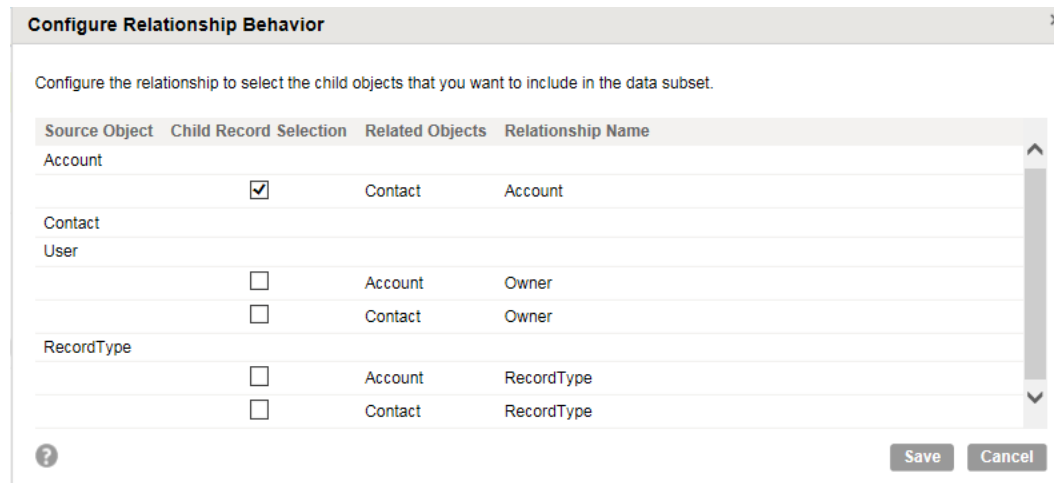
Configure relationship behavior

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:



The dialog box titled "Configure Relationship Behavior" contains a table with the following data:

Source Object	Child Record Selection	Related Objects	Relationship Name
Account	<input checked="" type="checkbox"/>	Contact	Account
Contact	<input type="checkbox"/>	Account	Owner
User	<input type="checkbox"/>	Contact	Owner
RecordType	<input type="checkbox"/>	Account	RecordType
	<input type="checkbox"/>	Contact	RecordType

At the bottom right are "Save" and "Cancel" buttons. At the bottom left is a help icon.

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.

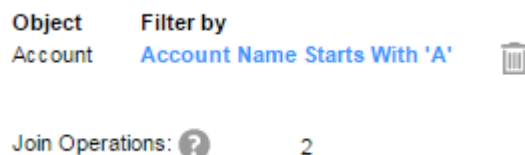
Consider that there are two objects Account and Contact. Account is the parent object, and Contact is the child object of Account. You can apply filter on the Account object or the Contact object. You can either use the default path that the task selects or configure the path. The number of paths traversed 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 configuration shows a table with two columns: "Object" and "Filter by".

Object	Filter by
Account	Account Name Starts With 'A'

Below the table, there is a "Join Operations" section with a question mark icon and the number 2.

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.


```

classDiagram
    class Account
    class Contact
    Account --> Contact : Account
    Contact --> Account : ParentAccount__r
  
```

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

Source	Target	Relationship	Rows
	Account		4
Account	Contact	Account	5
Contact	Account	ParentAccount__r	3

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.

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

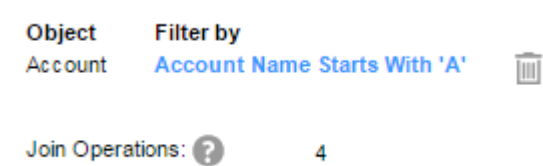
Source Object	Total Rows	Subset Rows	Subset Size (in KBs)
Contact	103	5	10
Account	100	4	8
Total Size	203	9	18

Use 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.

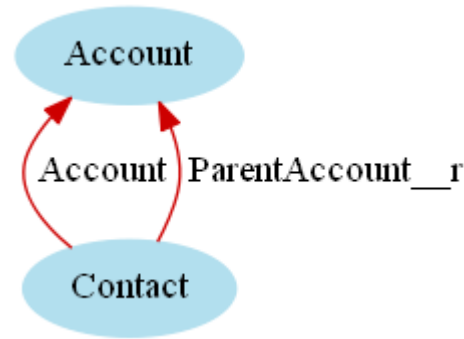
Consider that you apply a filter on the Account object with the condition that the account name starts with the letter A.

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



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:



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:

Source	Target	Relationship	Rows
	Account		5
Account	Contact	Account	5
Account	Contact	ParentAccount__r	5
Contact	Account	Account	4
Contact	Account	ParentAccount__r	3

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.

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

Subset Statistics			
Source Object	Total Rows	Subset Rows	Subset Size (in KBs)
Contact	103	7	14
Account	100	5	10
Total Size	203	12	24

Case 3. Select the default path with filter on Contact

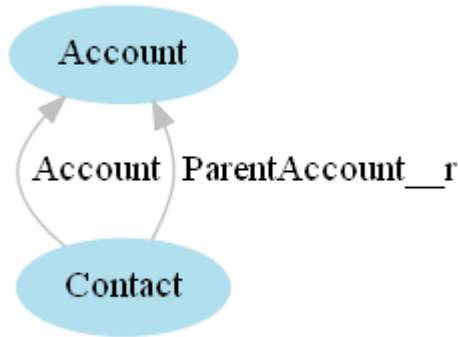
Consider that you apply a filter on the Contact object with the condition that the first names start with the letter A.

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



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:

Source	Target	Relationship	Rows
	Contact		12
Contact	Account	Account	12
Contact	Account	ParentAccount__r	1

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:

Subset Statistics			
Source Object	Total Rows	Subset Rows	Subset Size (in KBs)
Contact	103	12	24
Account	100	12	24
Total Size	203	24	48

Case 4. Select the configured path with filter on Contact

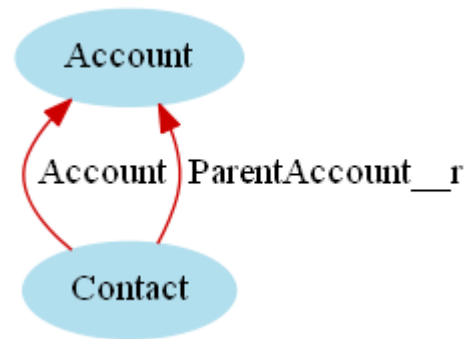
Consider that you apply a filter on the Contact object with the condition that the first names start with the letter A.

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



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.

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

Source	Target	Relationship	Rows
	Contact		15
Contact	Account	Account	12
Contact	Account	ParentAccount__r	1
Account	Contact	Account	13
Account	Contact	ParentAccount__r	3
Contact	Account	Account	14
Contact	Account	ParentAccount__r	3

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:

Subset Statistics			
Source Object	Total Rows	Subset Rows	Subset Size (in KBs)
Contact	103	15	30
Account	100	14	28
Total Size	203	29	58

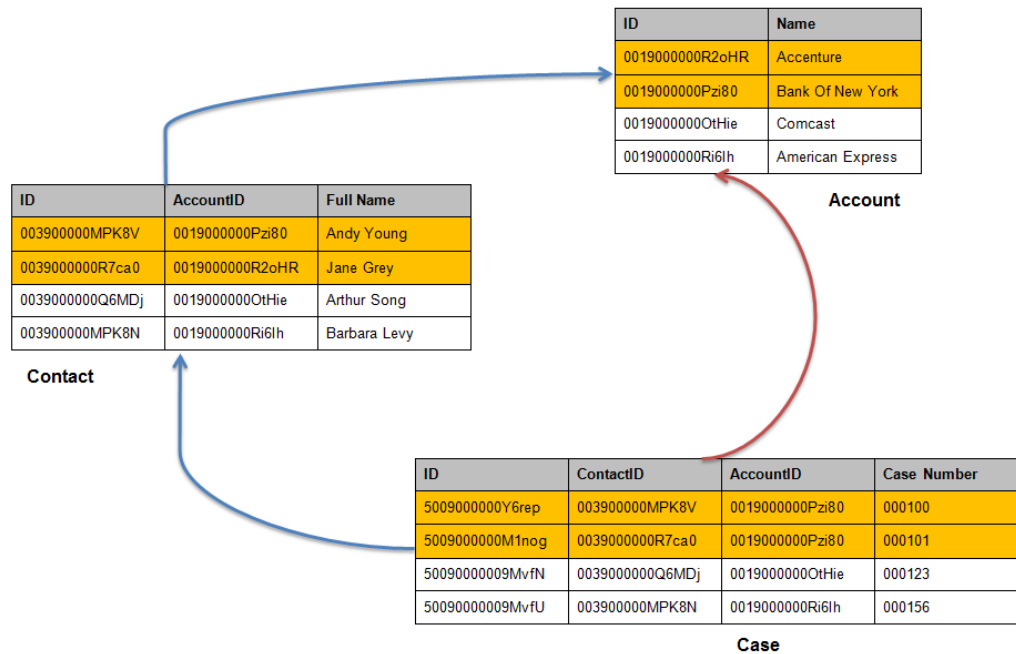
Data subset use cases for three objects

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:

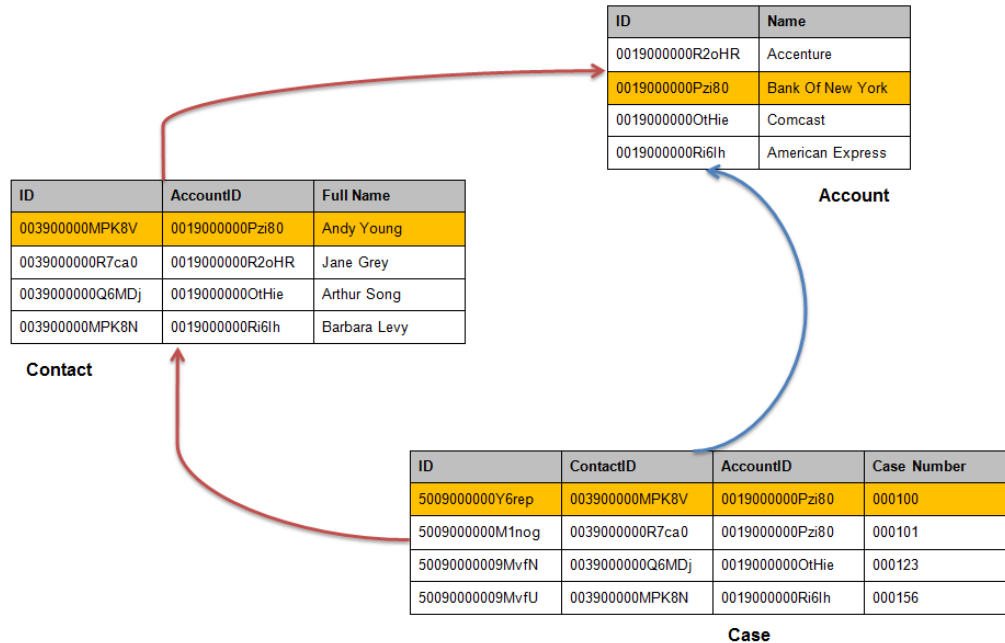


Based on the default path selection, the task traverses to the Case object directly, and then traverses to the Contact object.

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 shows 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

Connection:* cmaskSource View... New...

Source Type:* ☐ Single ☒ Multiple

Add... View Schema...

Actions	Source Object	Related Objects	Relationship Name
<input type="checkbox"/>	Case		
<input type="checkbox"/>		Account	Account
<input type="checkbox"/>		Contact	Contact
<input type="checkbox"/>	Contact		
<input type="checkbox"/>		Account	Account
<input type="checkbox"/>	Account		

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

Case	Contact
00002541	Victor
00002542	Jack

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:

The screenshot shows the Informatica Data Integration interface. The left sidebar contains navigation options: Home, Explore, Bundles, My Jobs, My Import/Export Logs, Account Contact Sta..., and Account Contact Sta... The main panel displays the 'Account Contact Subset Masking-6' task details. The 'Job Properties' section shows: Task Name: Account Contact Subset Masking, Instance ID: 6, Task Type: Data Masking, Started By: vdi@informatica.com, Start Time: Nov 15, 2017 2:38:27 AM, End Time: Nov 15, 2017 2:40:10 AM, Duration: 1 minute, 43 seconds, Runtime Environment: iku10dag619, Secure Agent: iku10dag619. The 'Results' section shows: Status: Success, Success Rows: 216, Error Rows: 0, Subset Rows: 10. The 'Activity Log Entries' table is as follows:

Name	End Time	Status	Success Rows	Error Rows	Error Message	Recommendation
Creating Staging Tables	Nov 15, 2017 2:39:00 AM	Success	0	0		
Staging Contact	Nov 15, 2017 2:39:21 AM	Success	100	0		
Staging Account	Nov 15, 2017 2:39:32 AM	Success	101	0		
Creating Indices	Nov 15, 2017 2:39:32 AM	Success	0	0		
Applying Filters	Nov 15, 2017 2:39:44 AM	Success	5	0		
Computing Subset	Nov 15, 2017 2:39:44 AM	Success	0	0		
Fetching Subset Results	Nov 15, 2017 2:39:45 AM	Success	0	0		
Loading Account	Nov 15, 2017 2:39:58 AM	Success	5	0		
Loading Contact	Nov 15, 2017 2:40:10 AM	Success	5	0		
Dropping Staging Tables	Nov 15, 2017 2:40:10 AM	Success	0	0		

Refresh metadata

When you create a masking task, the task imports the source and target metadata. Over time, you might update the Salesforce objects and add or delete objects. You might also add or delete objects in the masking task.

Because of changes to Salesforce objects or objects in the task, the metadata imported when you created the task can get outdated.

If you run the same masking task at regular intervals, the metadata imported in the task might not be the latest. The masking task requires the latest metadata to define relationships between objects and to determine fields that you can mask.

A masking task might fail if it does not use the updated metadata in the Salesforce source and target.

You can refresh the metadata before you run a masking task to ensure that the source and target metadata in the task is up to date. The time required for a metadata refresh might differ based on the number of objects, latency, and the Salesforce API response time.

You can refresh the metadata in a masking task in one of the following ways:

Refresh the metadata without editing a task

When you refresh the metadata without editing the task, the refresh runs as a separate job. You cannot run an instance of a masking task and a metadata refresh of the task at the same time. If the refresh job fails at any point, the metadata does not update. So the source metadata and target metadata remain consistent.

Refresh the metadata without editing the task from the list of tasks on the **Explore** page or from the task view page. View the progress and status of the refresh job from the **My Jobs** page.

Refresh the metadata from within a masking task

You can refresh the source and target fields from within a masking task when you create or update a masking task. The **Target** page in the create task and edit task workflows includes an option to refresh

fields. You cannot view the progress of the refresh or perform other tasks during the refresh. You can continue to create or update and save the masking task after the refresh finishes. The refresh process might take some time, based on the number of objects and the size of the metadata.

Choose how you want to refresh the metadata based on the number of objects to refresh. As a best practice, if you need to update many objects, it is recommended that you refresh the metadata without editing the task. To update fewer objects or less metadata, you can edit the metadata from within the task.

Reset task

You can reset a masking task that has a different source and target and contains data filters if the task fails at any point and you want to restart the task from the first step.

A masking task with data filters performs different steps including staging data, subset computation, load to target, and drop staging tables. If a task fails at any of the steps, it continues from the point of failure when you restart the task.

For example, you configure a masking task with data filters and run the task. The task fails at the load to target step. If you restart the task at a later date, it skips the steps of data staging and subset computation. You want to rerun the task, but you want to ensure that the subset computation is accurate and includes any changes to the data. If the task skips the staging and subset computation steps, it uses the estimation and staged tables from the previous failed task run.

You might want to restart a failed task from the first step if you reinstall the staging database or the database file is corrupted and you use a different staging database. The staged files are not available on the new installed or different database. If you restart a task that failed at the load to target step, the task restarts from the same step but cannot access the previous staged tables. The task fails again. In such cases, reset the task before you restart the task. The status returns to start and when you restart the task, it stages and estimates the data again.

Choose to reset the task before you restart the task. The reset returns the task status to Start. When you restart the task, the task starts from the first step. It performs all steps of staging, subset computation, load to target, and drop staging tables, based on how you configure the task.

You can reset masking tasks that have different source and target connections and include data filters. Tasks that use the same source and target or do not include data filters do not require subset computation or staging tables.

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.

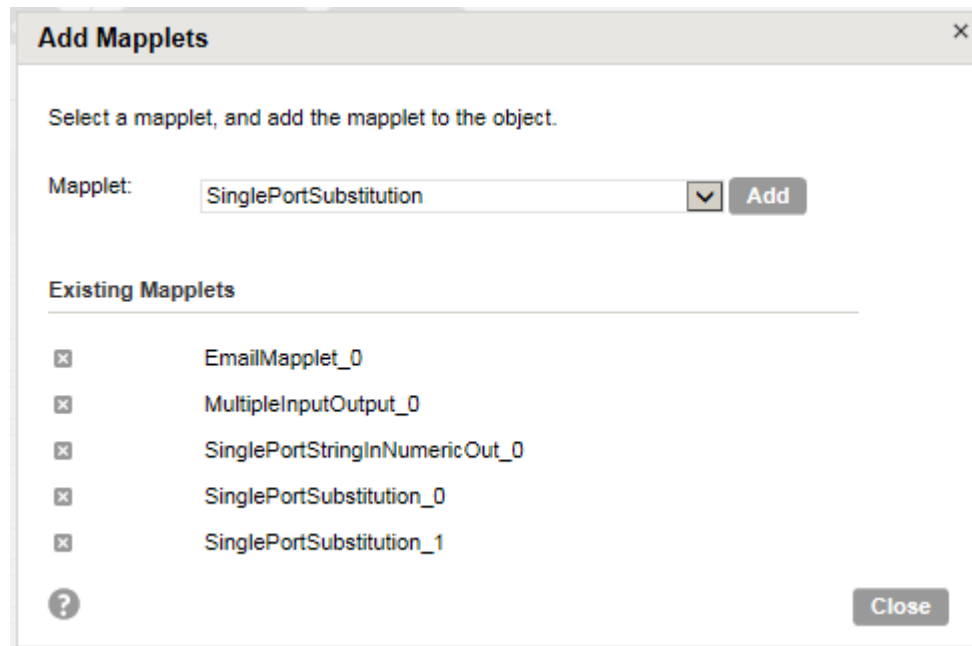
You can use a mapplet that requires an extra connection to a relational database or a flat file. Before you add the mapplet, you must add the connection.

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:



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\main\dmask
```

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:

Parameter	Description
isSeeded	To configure repeatable output. Enter True or False. If you enter True, specify a seed value.
seedValue	A starting number to create repeatable output. Enter a number from 1 through 999.
keepCardIssuer	Masks a credit card number with the same credit card type. Enter True or False. If you enter False, specify the targetIssuer parameter.
targetIssuer	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.
firstNameColumn	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.
firstNameLength	The maximum number of characters of the first name to include in the masked email addresses.
delimiter	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.
lastNameColumn	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.
lastNameLength	The maximum number of the characters of the last name to include in the masked email address.
domainConstantValue	A domain string name to include in the masked email addresses.
useMaskFormat	Specifies if you want to use a mask format. Enter True or False.
maskFormat	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.
useSrcFilter	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.
srcFilterOption	Defines a filter that determines which characters to mask in the source. Enter one of the following options: <ul style="list-style-type: none">- 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.

Parameter	Description
srcFilterStr	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.
usetargetFilter	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.
targetFilterOption	Defines a filter that determines which characters to use in target mask. Enter one of the following options: <ul style="list-style-type: none"> - 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.
targetFilterStr	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.
useRange	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.
minWidth	The minimum value of the range. You can specify the minimum width for date, string, and numeric data types.
maxWidth	The maximum value of the range. You can specify the minimum width for date, string, and numeric data types.
startDigit	Defines the first digit of the masked SIN.
startDigitValue	The value for the first digit of the masked SIN.
DicConn	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.
DicName	The dictionary that you want to select. The dictionary file must be present in the rdtmDir directory of the Secure Agent.
outputPort	The output port column from the dictionary.
useBlurring	Masks data with a variance of the source data if specify that you want to blur the target data.
blurringUnit	Unit of the date to apply the variance to. You can enter the following values: Year, Month, Day, Hour, Minute, or Second.
blurringOption	The unit of numeric blurring. Enter Fixed or Percent.
blurLow	The low boundary of the variance from the source. Enter the value for numeric and date data types.

Parameter	Description
blurHigh	The high boundary of the variance from the source. Enter the value for numeric and date data types.
expText	An expression that you can configure to mask the target data.
Preprocessing expression	An expression to define changes to make to the data before masking.
Preprocessing expression	An expression to define changes to make to the masked data before saving the data to the target.

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**.

If you create simple data filters for the same object, the simple data filters for the object merge with the advanced filter. For more information about configuring advanced data filters, see [“Configuring advanced data filters” on page 14](#)

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 maplet rule, you must configure input and output fields of the maplet.
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. On the **Explore** page, navigate to the masking task.

2. In the row that contains the masking task, click **Actions** and select **Edit**. You can also open the task and click the **Edit** button on the task view page.

You can edit the task information in the **Edit Task** window.

3. Click **Finish** to save the changes.

Running a masking task manually

You can start a masking task manually to run it immediately.

1. On the **Explore** page, navigate to the masking task.
2. You can run a task manually in one of the following ways:
 - In the row that contains the task, click **Actions** and select **Run**.
 - 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.

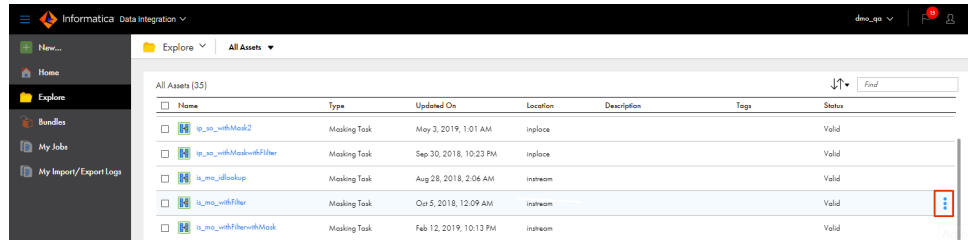
Refreshing the metadata in a masking task

You can refresh the metadata in a masking task if the metadata of the objects changes in Salesforce or you make changes to the list of objects in the masking task.

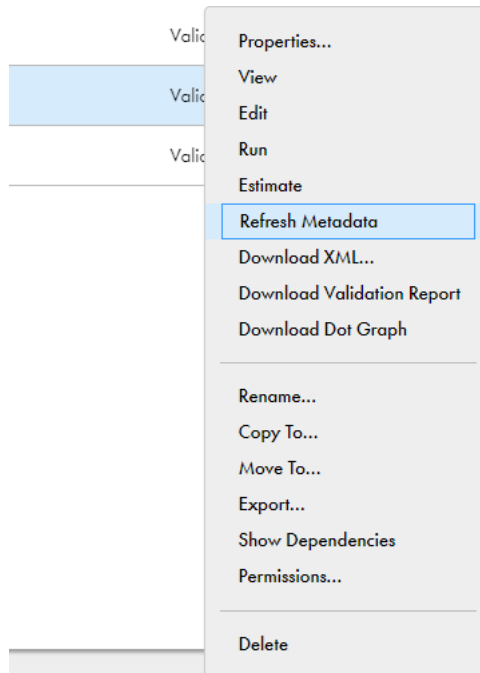
You can choose to refresh the metadata without editing a task. You can also refresh the metadata from the create or update workflow of a masking task. Choose how you want to refresh the metadata based on the number of objects to refresh.

1. On the **Explore** page, navigate to the masking task.
2. You can refresh the metadata in one of the following ways:
 - Refresh the metadata without editing the task. You can refresh the metadata without editing the task in one of the following ways:
 - In the row that contains the masking task, click **Actions** and select **Refresh Metadata**.

- Open the masking task. In the task view page, click **Actions** and select **Refresh Metadata**. The following image shows the **Actions** icon on the **Explore** page:



The following image shows the Refresh Metadata option available in the **Actions** menu:



You cannot refresh the metadata if an instance of the task is in progress. Wait for the task to finish and then refresh the metadata.

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.

- Edit or update the task to refresh the metadata.

Note: Use this option if you need to update a small number of objects.

1. Open the masking task.
2. Click **Edit** to open the task edit page.
3. Click **Target** to open the **Target** page.
4. Click **Refresh Fields**.
5. Click **Save** and then click **Finish**.

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.

Resetting a masking task

Reset a failed masking task to ensure that the task starts from the first step when you restart the task.

Reset a failed task that includes a data filter and uses a different source and target. Tasks that do not include a data filter or use the same connection as the source and target do not require staging or subset computation.

1. On the **Explore** page, navigate to the masking task.
2. Open the masking task. In the task view page, click **Actions** and select **Reset Task**.

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. On the **Explore** page, navigate to the masking task.
2. In the row that contains the masking task, click **Actions** and select **Permissions**.
The **Task Permissions** page opens with a list of users and groups that have permissions on the task.
3. To remove users or groups, select the users or groups and click **Remove** and then click **Save**.
4. To edit the level of permission assigned, select the user or group, change the permission level as required and click **Save**.
5. To add users or groups, click **Add**.
6. 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.
7. 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. On the **Explore** page, navigate to the masking task.
2. In the row that contains the masking task, click **Actions** and select **Copy To**.
3. 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. On the **Explore** page, navigate to the masking task.
2. In the row that contains the masking task, click **Actions** and select **Rename**.
3. 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. On the **Explore** page, navigate to the masking task.
2. In the row that contains the masking task, click **Actions** and select **Delete**.
3. 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. On the **Explore** page, navigate to the masking task.
2. In the row that contains the masking task, click **Actions** and select **Export**.
3. Optional. Edit the job name if required.
4. 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. On the **Explore** page, navigate to the masking task.
2. In the row that contains the masking task, click **Actions** and select **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. On the **Explore** page, navigate to the masking task.
2. In the row that contains the masking task, click **Actions** and select **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.

Themasking task downloads the following dictionary files:

```
informatica_mask_address.dic
informatica_mask_cc_american_express.dic
informatica_mask_cc_diners_club.dic
informatica_mask_cc_american_express.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:

```
SNO, 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.

CHAPTER 7

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:

Masking rule	Description
Credit Card Masking	Masks a credit card number with a number that has a valid checksum. You can mask the string data type.
Email Masking	Masks an email address with a random email address. You can mask the string data type.
Email Advanced Masking	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.
IP Address Masking	Applies a built-in mask format to disguise an IP address. You can mask the string data type.
Key Masking	Creates repeatable results for the same source data. You can mask date, numeric, and string data types.
Nullification Masking	Changes input values in a field to null. You can mask date, numeric, and string data types.
Phone Masking	Masks a phone number with random numbers in the same format as the original number. You can mask the string data type.
Random Masking	Produces random, nonrepeatable results. You can mask date, numeric, and string data types.
SIN Masking	Masks a Social Insurance number. You can mask a string data type.

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:

Parameter	Description
Repeatable	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.
Seed Value	A starting number to create repeatable output. Enter a number from 1 through 999. Default seed value is 190.
Keep Issuer	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.
Mask Issuer	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.

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 KtrlupQAPyk@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:

Parameter	Description
Repeatable	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.
Seed Value	A starting number to create repeatable output. Enter a number from 1 through 999. Default seed value is 190.
First Name	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.
First Name Length	The number of the characters of the first name to include in the email address.
Delimiter	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.
Last Name	Name of the masked column to use in the email name. The email name contains the masked value of the column you choose.
Last Name Length	The number of the characters of the last name to include in the email address.
Domain Name	A string value that identifies represents an Internet Protocol (IP) resource such as gmail.com.

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:

Masking parameter	Description
Repeatable	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.
Seed Value	A starting number to create repeatable output. Enter a number from 1 through 999. Default seed value is 190.
Mask Format	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.
Filter Source	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.
Source Filter Type	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: <ul style="list-style-type: none">- 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.
Target Filter Type	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: <ul style="list-style-type: none">- 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.

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 mask 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:

Character	Description
A	Alphabetical characters. For example, ASCII characters a to z and A to Z.
D	Digits. From 0 through 9.
N	Alphanumeric characters. For example, ASCII characters a to z, A to Z, and 0-9.
X	Any character. For example, alphanumeric or symbol.
+	No masking.
R	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.

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:

Masking parameters	Descriptions
Mask Format	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.
Filter Source	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.
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: <ul style="list-style-type: none">- 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: <ul style="list-style-type: none"> - 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:

Masking parameter	Description
Range	A range that you want to set for the numeric data. Select the check box to enter a minimum and maximum range.
Minimum Range	The minimum value of the range.
Maximum Range	The maximum value of the range. The maximum value must be greater than the minimum value.
Blurring	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.
Blurring Option	The unit of blurring. Select Fixed or Percent. Default is Fixed.
Low Bound	The low boundary of the variance from the source number.
High Bound	The high boundary of the variance from the source number.

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:

Blurring type	Low	High	Result
Fixed	0	10	Between 66 and 76
Fixed	10	0	Between 56 and 66
Fixed	10	10	Between 56 and 76
Percent	0	50	Between 66 and 99
Percent	50	0	Between 33 and 66
Percent	50	50	Between 33 and 99

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:

Masking parameter	Description
Range	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.
Minimum Range	The minimum value to return for the selected datetime value. The default datetime format is MM/DD/YYYY HH24:MI:SS.
Maximum Range	The maximum value to return for the selected datetime value. The maximum datetime must be later than the minimum datetime.
Blurring	Mask a date as a variance of the source date.
Blurring Unit	Unit of the date to apply the variance to. Select the year, month, day, or hour. Default is year.
Low Bound	The low boundary of the variance from the source date.
High Bound	The high boundary of the variance from the source date.

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:

Masking parameter	Description
Repeatable	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.
Seed Value	A starting number to create repeatable output. Enter a number from 1 through 999. Default seed value is 190.
Start Digit	When enabled, you can define the first digit of the masked SIN.
Start Digit Value	The value for the first digit of the masked SIN.

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:

123456789

123 45 6789

123-45-6789

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:

Parameter	Description
Dictionary File	<p>Appears if you select Custom Substitution from the list of masking rules.</p> <p>Click Select and specify the following parameters:</p> <p>Flat File Connection</p> <p>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.</p> <p>Dictionary File</p> <p>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:</p> <pre><Secure Agent installation directory>\apps\Data_Integration_Server\data</pre>
Dictionary Column	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.
Lookup Input Port	Optional. The source input column based on which you perform a lookup operation on the dictionary.
Lookup Output Port	Optional. The dictionary column that you can look up based on the input port.
Lookup Error Constant	Optional. A constant value that you can configure when there are no matching values for the lookup condition from the dictionary.
Repeatable	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.
Seed Value	A starting number to create repeatable output. Enter a number from 1 through 999. Default seed value is 190.

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://MgLaHjCa.VsD/`. 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 16](#)

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.

The following image shows the mapplet parameters that you can configure:

Configure Mapplets

Select Mapplet: EmailMapplet_0

Configure Source to Mapplet Mapping

FirstName3	FirstName
LastName3	LastName

Configure Mapplet to Target Mapping

Email	Email3
-------	--------

?

Save Cancel

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:

Configure Mapplets

Select Mapplet: AccountNameMapplet_0

Connection Reference	Connection
Lookup Connection	AccName_Lookup
Dictionary Connection	AccName_Dict_Con

Configure Source to Mapplet Mapping

Source	Mapplet
Account Name	input_port

Configure Mapplet to Target Mapping

Mapplet	Target
output_port	Account Name

Save Cancel

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.

CHAPTER 8

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 data types, such as the following IIF expressions:

```
[IIF(ANNUALREVENUE >0,NAME)]  
[IIF(empid_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 help for the appropriate Data Integration connector.

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:

```
Exception
when NO_DATA_FOUND
then NULL;
END;
```

For example, you have the following stored procedure in a PowerCenter workflow:

```
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;
```

Before you export the workflow, add exception handling as follows:

```
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:

Field	Description
Task Name	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.
Location	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.
Description	Description of the PowerCenter task. Maximum length is 255 characters.
Runtime Environment	Runtime environment that contains the Secure Agent to run the task.
Workflow XML File	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.

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 [41](#)

A

add sources
multiple objects [102](#)
advanced data filters
configuring [13](#), [14](#)
deleting [13](#), [14](#)
advanced email masking
description [143](#)
advanced options
configuring for synchronization tasks [70](#)
advanced relationships
in mapping tasks [31](#)
aggregate functions
in mapping tasks [47](#)
AND
reserved word [19](#)
automatic task recovery
estimate subset [109](#)

C

CHR function
inserting single quotation mark [17](#)
CLAIRE tuning
guidelines [51](#), [53](#)
Cloud Application Integration community
URL [10](#)
Cloud Developer community
URL [10](#)
column names
duplicates [77](#)
columns
names, truncating [77](#)
target database, replication [76](#)
commands
in synchronization and replication tasks [62](#), [78](#)
comments
adding to field expressions [18](#)
configure
staging connection [105](#)
constants
definition [17](#)
credit card issuer
description [142](#)
credit card masking
overview [142](#)
cross-schema pushdown optimization [41](#)
custom masking
expression [155](#)

D

data catalog discovery
discovering and selecting objects [27](#)
searching for objects [25](#)
synchronization task sources [25](#)
data filters
configuring [12](#)
configuring advanced filters [13](#), [14](#)
configuring for replication tasks [82](#)
configuring simple filters [13](#)
deleting [13](#)
operators [14](#)
rules for configuring [15](#)
synchronization task [67](#)
variables [15](#)
Data Integration community
URL [10](#)
data masking
common fields [124](#)
email notification [127](#)
inplace masking [104](#)
missing mandatory fields [124](#)
overview [101](#)
refresh fields [104](#)
Data Masking
parameter files [110](#)
data subset
automatic task recovery [109](#)
configure relationships [108](#)
data filters [108](#)
options [109](#)
relationship behavior [108](#)
schema graph [108](#)
subset rows [119](#)
data types
configuring [59](#)
database sources and targets
rules in synchronization tasks [59](#)
database user accounts
requirements for synchronization [63](#)
user privileges for replication tasks [79](#)
databases
configuring multiple source tables for synchronization tasks [64](#)
Daylight Savings Time
schedules [22](#)
DD_DELETE constant
reserved word [19](#)
DD_INSERT constant
reserved word [19](#)
DD_REJECT constant
reserved word [19](#)
DD_UPDATE constant
reserved word [19](#)
default assignment
masking rules [124](#)

- delete criteria
 - description [81](#)
- delete task operation [55](#)
- deleting
 - advanced data filters [13](#), [14](#)
 - data filters [13](#)
- dictionary files
 - data masking [136](#)
 - substitution masking [152](#)
- DOT format
 - download [102](#)
- download
 - mapping XML [136](#)
 - validation reports [136](#)

E

- elastic mappings
 - CLAIRE tuning [53](#)
 - CLAIRE Tuning [51](#), [52](#)
- email masking
 - description [143](#)
- email notification
 - configuring for tasks [24](#)
- Enterprise Data Catalog
 - discovering and selecting objects [27](#)
 - finding synchronization task sources [25](#)
 - searching for objects [25](#)
- exception handling
 - in stored procedures for PowerCenter tasks [160](#)
- expressions
 - configuring for a lookup [60](#)
 - rules [18](#)
- :EXT reference qualifier
 - reserved word [19](#)
- external IDs
 - for related Salesforce objects [57](#)

F

- FALSE constant
 - reserved word [19](#)
- field exclusions
 - configuring for replication tasks [82](#)
- field expressions
 - comments, adding [18](#)
 - components [17](#)
 - creating [16](#)
 - literals [17](#)
 - reserved words [19](#)
 - rules [18](#)
 - rules for validating [16](#)
 - syntax [17](#)
- field mappings
 - configuring for synchronization tasks [68](#)
 - required fields [59](#)
 - understanding for synchronization tasks [59](#)
- field metadata [41](#)
- fields
 - required in field mappings [59](#)
- flat files
 - column names in synchronization tasks [58](#)
 - creating target files in synchronization tasks [56](#)
 - directory for replication tasks [79](#)
 - directory for synchronization tasks [63](#)
 - editing metadata in mapping tasks [41](#)

- flat files (*continued*)
 - rules for sources and targets in synchronization tasks [58](#)
- FTP/SFTP connections
 - for PowerCenter tasks [161](#)
- full load
 - replication task load type [73](#)
- functions
 - definition [17](#)

H

- H2 database
 - configure [106](#)
 - data subset [105](#)
 - install [106](#)
 - Linux [107](#)
 - requirements [106](#)
 - staging [105](#)
 - Windows [106](#)

I

- IDs
 - types for objects [57](#)
- in-out parameters
 - in mapping tasks [47](#)
- :INFA reference qualifier
 - reserved word [19](#)
- Informatica Global Customer Support
 - contact information [11](#)
- Informatica Intelligent Cloud Services
 - web site [10](#)
- inplace masking
 - update [104](#)
- input parameters
 - in mapping tasks [47](#)
- insert task operation [55](#)
- IP address masking
 - description [144](#)

J

- jobs
 - monitoring [24](#)
 - scheduling [20](#)

K

- key masking
 - date [147](#)
 - description [145](#)
 - numeric [147](#)
 - string [145](#)

L

- \$LastRunDate
 - data filter variable [15](#)
- \$LastRunTime
 - data filter variable [15](#)
- linear taskflows
 - creating schedules [23](#)

- literals
 - definition [17](#)
 - single quotation mark requirement [17](#)
 - string and numeric [17](#)
- :LKP reference qualifier
 - reserved word [19](#)
- load types
 - for replication [73](#)
 - full load [73](#)
- lookups
 - configuring return values and expressions [60](#)
 - configuring the condition [60](#)
 - rules for creating [61](#)

M

- maintenance outages [11](#)
- mapping tasks
 - advanced connection properties [30](#)
 - advanced session properties for elastic mappings [37](#)
 - advanced session properties for mappings [32](#)
 - advanced session properties for Visio templates [32](#)
 - CLAIRE tuning [53](#)
 - CLAIRE Tuning [51](#), [52](#)
 - configuring [42](#)
 - configuring schedules, email, and advanced options [50](#)
 - configuring sources [43](#)
 - creating [42](#)
 - editing [42](#)
 - input and in-out parameters [47](#)
 - joining related sources [30](#)
 - overview [29](#)
 - running [54](#)
 - Sequence Generator configuration [41](#)
 - templates [29](#)
 - using aggregate functions [47](#)
 - using parameter files [20](#)
 - viewing details [54](#)
- mapplets
 - in synchronization tasks [60](#)
- mask all except
 - key masking [145](#)
- mask only
 - key masking [145](#)
- masking
 - advanced email [143](#)
 - advanced options [127](#)
 - configuring rules [131](#)
 - configuring the source [129](#)
 - configuring the target [130](#)
 - credit cards [142](#)
 - defining the task [128](#)
 - email addresses [143](#)
 - IP addresses [144](#)
 - overview [101](#)
 - phone numbers [147](#)
 - random [148](#)
 - rules descriptions [140](#)
 - scheduling a task [131](#)
 - SSN [151](#)
 - URL [154](#)
 - validation reports [104](#)
- masking parameters
 - random date [150](#)
 - random numeric [149](#)
 - random string [148](#)
- masking rule
 - mapplet [155](#)
 - postprocessing expression [142](#)
 - preprocessing expression [142](#)
- masking rules
 - description [140](#)
 - source field [122](#)
- masking task
 - data subset [108](#)
 - metadata refresh [121](#)
 - refresh metadata [133](#)
 - reset [122](#)
 - resetting [135](#)
- Masking task
 - copying [135](#)
 - deleting [136](#)
 - exporting [136](#)
 - rename [135](#)
 - running immediately [133](#)
 - steps to create [128](#)
- Masking tasks
 - using parameter files [20](#)
- mass ingestion
 - schedule configuration [94](#)
- mass ingestion task
 - configuration [88](#)
 - defining [88](#)
 - description [85](#), [87](#)
 - file listener [91](#)
 - local folder source [89](#)
 - local folder target [93](#)
 - running [95](#)
 - source configuration [89](#)
 - target configuration [93](#)
 - viewing details [95](#)
- mass ingestion tasks
 - prerequisites [88](#)
 - sources [85](#)
 - targets [86](#)
- :MCR reference qualifier
 - reserved word [19](#)
- monitoring jobs [24](#)
- multiplicity
 - configuring for a lookup [60](#)

N

- NOT
 - reserved word [19](#)
- NULL constant
 - reserved word [19](#)

O

- operating system commands
 - rules for replication tasks [62](#), [78](#)
 - rules for synchronization tasks [62](#), [78](#)
- operations
 - for synchronization tasks [55](#)
- operators
 - definition [17](#)
- OR
 - reserved word [19](#)
- \$OutputField
 - lookup expression
 - description [60](#)

\$OutputField (*continued*)
lookup return value variable [60](#)

P

parameter file
 configuring for mapping tasks [50](#)
parameter files
 overview [20](#)
parameters
 configuring [48](#)
 user defined [20](#)
partial sandbox
 update [104](#)
phone numbers
 masking [147](#)
postprocessing commands
 SQL commands [62, 78](#)
PowerCenter tasks
 configuring [161](#)
 description [158](#)
 editing [161](#)
 rules for PowerCenter workflows [158](#)
 rules for PowerCenter XML files [158](#)
 rules for sources and targets [160](#)
 running [163](#)
 supported transformations and mapping objects [159](#)
 Web Service connections [161](#)
PowerCenter XML files
 rules for PowerCenter tasks [158](#)
 supported transformations and mapping objects [159](#)
pre- and post-session commands
 PowerCenter tasks
 using pre-session and post-session commands [160](#)
preprocessing and postprocessing commands
 configuring for mapping tasks [50](#)
 configuring for synchronization tasks [70](#)
 rules for operating system commands [62, 78](#)
 rules for SQL commands [62, 78](#)
preprocessing commands
 SQL commands [62, 78](#)
prerequisites
 for synchronization tasks [63](#)
 replication tasks [79](#)
PROC_RESULT variable
 reserved word [19](#)
pushdown optimization
 cross-schema [41](#)
 mapping tasks [40](#)
 types [40](#)
 user-defined parameter [41](#)

Q

quotation marks
 inserting single using CHR function [17](#)

R

random masking
 date [150](#)
 numeric [149](#)
 string [148](#)
related objects
 configuring external IDs [57](#)

repeat frequency
 schedules [21](#)
replication
 source types [74](#)
 target types [74](#)
replication tasks
 configuring [79](#)
 configuring data filters [82](#)
 configuring schedules and email [83](#)
 configuring sources [80](#)
 configuring targets [81](#)
 creating [80](#)
 creating target tables [77](#)
 database targets [74](#)
 editing [80](#)
 field exclusions [82](#)
 flat file targets [75](#)
 load types [73](#)
 preprocessing and postprocessing commands [62, 78](#)
 prerequisites [79](#)
 replication process [73](#)
 rules for creating [79](#)
 rules for running [84](#)
 running [84](#)
 viewing details [83](#)
reserved words
 list [19](#)
resetting
 target tables [75](#)
rules
 configuring [131](#)

S

Salesforce IDs
 for objects [57](#)
schedule
 configure [127](#)
 configuring for a PowerCenter task [161](#)
schedules
 configuring [23](#)
 creating [23](#)
 Daylight Savings Time [22](#)
 repeat frequency [21](#)
 time zones [22](#)
scheduling
 mapping tasks [50](#)
 Masking task [131](#)
 replication tasks [83](#)
 synchronization tasks [70](#)
 taskflows [20](#)
 tasks [20](#)
Schema graph
 DOT format [102](#)
 multiple source objects [102](#)
:SD reference qualifier
 reserved word [19](#)
:SEQ reference qualifier
 reserved word [19](#)
SESSSTARTIME variable
 reserved word [19](#)
simple data filters
 configuring [13](#)
Social Security masking
 description [151](#)
source filter characters
 description [146](#)

- source filter type
 - description [145](#)
- source objects
 - multiple [102](#)
 - single [102](#)
- sources
 - configuring for mapping tasks [43](#)
 - configuring for replication tasks [80](#)
 - configuring for synchronization task [63](#)
 - configuring multiple database tables for synchronization tasks [64](#)
 - configuring saved query sources [64](#)
 - configuring single-object sources [64](#)
 - rules for multiple-object database in synchronization tasks [56](#)
- :SP reference qualifier
 - reserved word [19](#)
- SPOUTPUT
 - reserved word [19](#)
- SQL commands
 - rules for replication tasks [62, 78](#)
 - rules for synchronization tasks [62, 78](#)
- SSN masking
 - description [151](#)
- status
 - Informatica Intelligent Cloud Services [11](#)
- stopping a running job [28](#)
- stored procedures
 - exception handling for PowerCenter tasks [160](#)
- string literals
 - single quotation mark requirement [17](#)
- substitution masking
 - custom dictionaries [152](#)
 - description [152](#)
 - parameters [153](#)
- synchronization tasks
 - adding mapplets [60](#)
 - configuring [62](#)
 - configuring schedules, email, and advanced options [70](#)
 - creating [63](#)
 - creating flat file targets [56](#)
 - data filters [67](#)
 - description [55](#)
 - editing [63](#)
 - preprocessing and postprocessing commands [62, 78](#)
 - prerequisites [63](#)
 - rules for database sources and targets [59](#)
 - rules for flat file sources and targets [58](#)
 - rules for multiple-object database [56](#)
 - rules for running [72](#)
 - rules for sources and targets [58](#)
 - running [71](#)
 - sources overview [56](#)
 - sources, configuring [63](#)
 - targets overview [56](#)
 - targets, configuring [66](#)
 - task operations [55](#)
 - truncating database targets [57](#)
 - understanding field mappings [59](#)
 - update columns [57](#)
 - using parameter files [20](#)
 - viewing details [71](#)
- syntax
 - for field expressions [17](#)
- SYSDATE variable
 - reserved word [19](#)
- system status [11](#)

T

- table names
 - duplicates [76](#)
- tables
 - resetting for replication [75](#)
 - target database, replication [76](#)
 - user-defined indexes [75](#)
- target
 - refresh fields [104](#)
- target databases
 - column names, truncating [77](#)
 - for data replication [74](#)
 - requirements for a synchronization task [63](#)
 - requirements for replication tasks [79](#)
 - resetting tables for replication [75](#)
 - table and column names, replication [76](#)
- target filter characters
 - description [146](#)
- target flat files
 - creating in synchronization tasks [56](#)
- target prefix
 - description [81](#)
 - guidelines [77](#)
- target tables
 - creating for replication tasks [77](#)
 - rules for resetting [75](#)
- target task operations
 - insert [103](#)
 - update [103](#)
 - upsert [103](#)
- targets
 - configuring for mapping tasks [45](#)
 - configuring for replication tasks [81](#)
 - configuring for synchronization tasks [66](#)
 - truncating in synchronization tasks [57](#)
- taskflows
 - monitoring [24](#)
 - scheduling [20](#)
- tasks
 - configuring email notification [24](#)
 - configuring field expressions [16](#)
 - creating schedules [23](#)
 - monitoring [24](#)
 - scheduling [20](#)
 - stopping a running job [28](#)
- :TD reference qualifier
 - reserved word [19](#)
- time zones
 - description [22](#)
- transformations
 - in PowerCenter tasks [159](#)
- TRUE constant
 - reserved word [19](#)
- trust site
 - description [11](#)

U

- update columns
 - adding [57](#)
- update task operation [55](#)
- upgrade notifications [11](#)
- upsert task operation [55](#)
- URL masking
 - description [154](#)
- user parameters [20](#)

using parameter files
 advanced filter [110](#)
 simple filter [110](#)

V

validation
 rules [16](#)
validation report
 target [104](#)
variables
 for data filters [15](#)

W

Web Service connections
 and PowerCenter tasks [161](#)
web site [10](#)
workflow XML file
 downloading from a PowerCenter task [161](#)
WORKFLOWSTARTTIME variable
 reserved word [19](#)