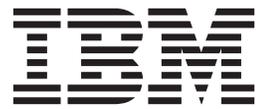


IBM InfoSphere DataStage
Version 11 Release 3

Data Masking Guide



IBM InfoSphere DataStage
Version 11 Release 3

Data Masking Guide



Note

Before using this information and the product that it supports, read the information in “Notices and trademarks” on page 45.

Contents

Data masking	1	Random replacement data masking policy	27
Overview	1	Hash data masking policy	28
Installing and configuring	2	Hash lookup data masking policy	30
Designing a data masking job	2	Appendix A. Product accessibility	33
Creating a data masking job	3	Appendix B. Reading command-line	
Setting up column definitions.	4	syntax	35
Configuring stage and link properties	5	Appendix C. How to read syntax	
Assigning a data masking policy to a column	6	diagrams	37
Compiling and running data masking jobs	7	Appendix D. Contacting IBM	39
Setting up sample reference tables	7	Appendix E. Accessing the product	
Data masking policies	8	documentation	41
Credit card number data masking policy	8	Appendix F. Providing feedback on the	
Email address data masking policy	9	product documentation	43
US national ID data masking policy - National ID		Notices and trademarks	45
(US)	10	Index	51
Canada national ID data masking policy -			
National ID (CA)	13		
French national ID data masking policy -			
National ID (FR).	15		
Italy national ID data masking policy - National			
ID (IT)	18		
Spain national ID data masking policy - National			
ID (ES)	20		
UK national ID data masking policy - National			
ID (UK)	22		
Date age data masking policy	25		
Repeatable replacement data masking policy	25		

Data masking

Use the Data Masking stage to mask sensitive data that must be included for analysis, in research, or for the development of new software. By using this pack, you can comply with company and government standards for data privacy, including the Sarbanes-Oxley (SOX) Act (and its equivalents around the world).

Overview

The Data Masking stage has a variety of predefined masking policies to mask different types of data.

These predefined data masking policies can be used to mask information in one of the following data types:

Context-aware data types

Context-aware business data types such as email addresses, national identification numbers, or credit card numbers.

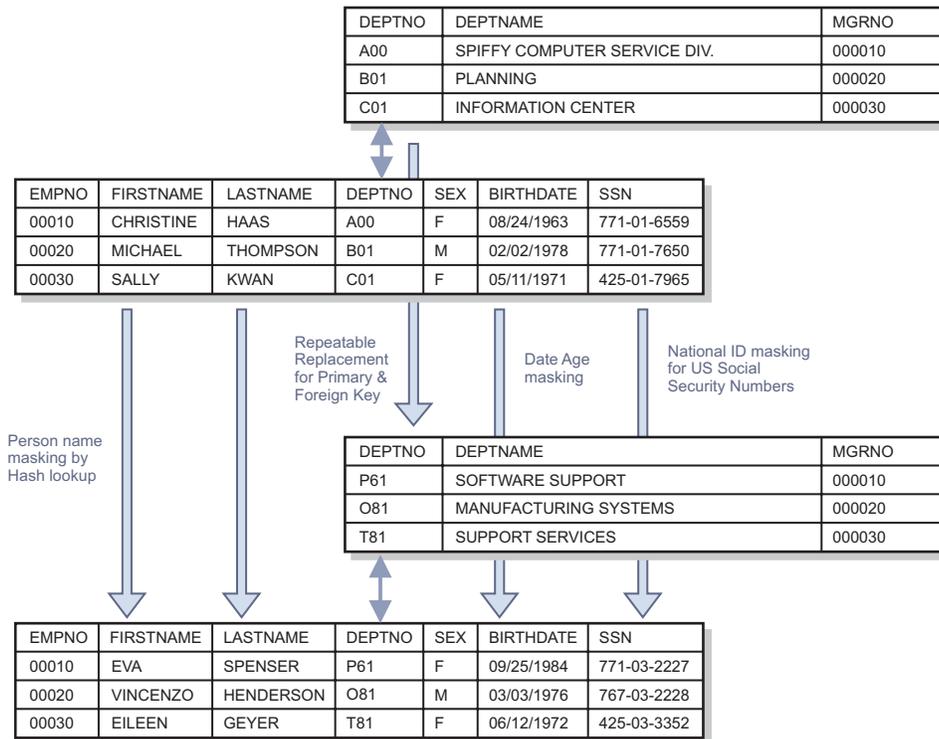
Generic data types

Generic data types such as dates or text strings are supported.

Some of the key features of the Data Masking stage are:

- Consistently mask an identifier in all data sources across the enterprise.
- Mask individual records, while maintaining analytical integrity.
- Mask data values with fictional but valid values for data types or business element types, while maintaining application integrity.
- Mask data repeatedly, while maintaining the referential integrity.
- Create masked test databases.

Source data before masking



Destination data after masking

Installing and configuring

The Data Masking stage requires Optim Data Privacy Providers, which is a library to mask privacy data. To use the Data Masking stage, you must install Optim Data Privacy Providers on the InfoSphere Information Server engine tier.

For more information about installing and configuring, see Installation instructions.

Designing a data masking job

You must create Data Masking stage jobs in order to assign data masking policies to relevant columns. You must also set up column definitions for stage operations.

A Data Masking stage job contains:

Input link

The input source can be a file, database or any other supported stage.

Output link

The output source can be a file, database or any other supported stage.

Reject link

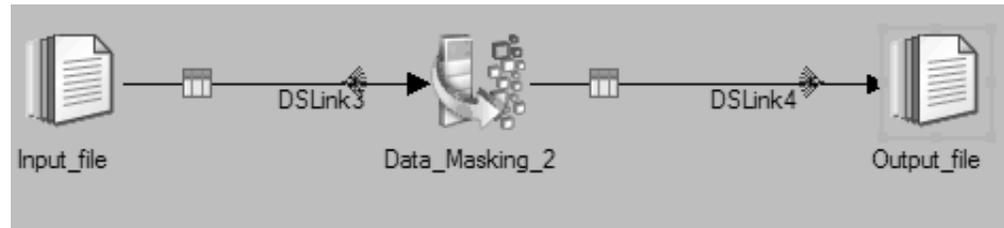
When a reject link is configured, invalid and rejected records are copied to this file or any other supported stage.

A Data Masking stage job can be created in one of the following ways:

One input link and one output link

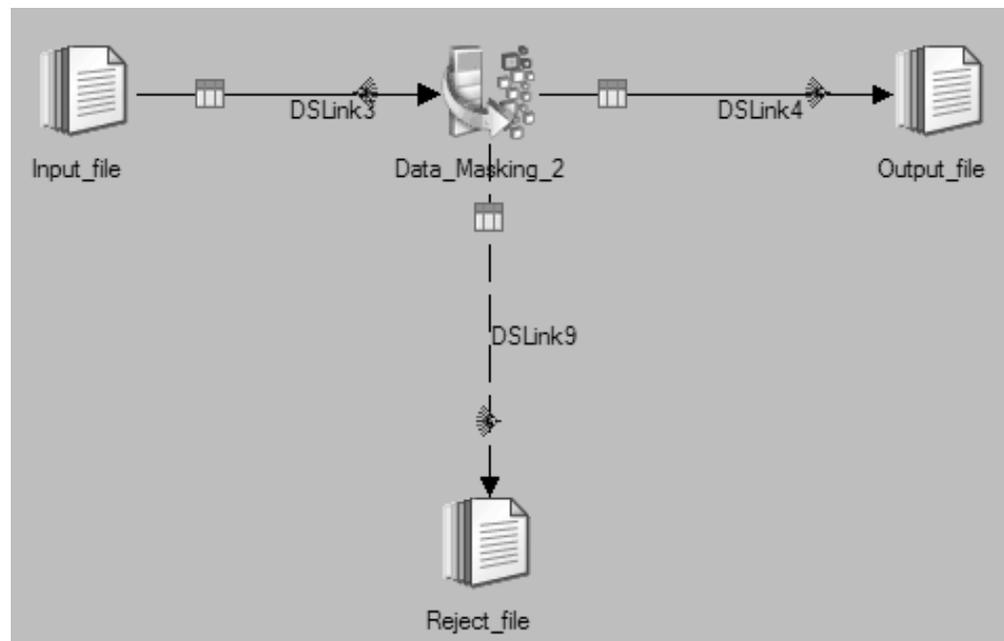
The Data Masking stage job represented in the following image is a simple job with one input link and one output link.

When a column is associated with a masking policy, data in that column is masked in the Data Masking stage.



One input link, one output link, and one reject link

The Data Masking stage job represented in the following image contains one input link, one output link, and one reject link:



If the input source data fails to validate, an error occurs. When a reject link is configured in the job, the record with the invalid data is copied to the configured destination. You can configure the error handling behavior in the stage property.

Creating a data masking job

Use the Masking Policy Editor to create the Data Masking stage job.

Procedure

1. From the Designer client, select **File > New**.
2. Select the **Parallel Job** icon, and click **OK**.

3. In the Parallel Job canvas, create input, output, and optionally reject stages.
4. In the Designer client palette area, click **Processing**.
5. In the processing section of the palette, select the **Data Masking** stage icon and drag the stage to your open job. Position the stage in between the input, output, and reject stages.
6. Link the different stages.
7. Rename the links and stages.
8. Select **File > Save** to save the job.

What to do next

“Setting up column definitions”

Setting up column definitions

You can create a set of columns and save the column definitions for later use, or load predefined column definitions. When the column definitions of output columns of Data Masking stage are saved or loaded, the data masking policy is saved or loaded along with other metadata.

Before you begin

“Creating a data masking job” on page 3

Procedure

1. On the parallel canvas, double-click the Data Masking stage icon.
2. Select the input link.
3. On the **Columns** tab, modify the columns grid to specify the metadata that you want to define.
 - a. Right-click within the grid, and select **Properties** from the menu.
 - b. In the Grid properties window, select the properties and the order in which you want the selected properties to be displayed. Then, click **OK**.
4. To save the column definitions as a table definition in the repository:
 - a. Click **Save**.
 - b. In the Save Table Definition window, enter the appropriate information, and then click **OK**.
 - c. In the Save Table Definition As window, select the folder where you want to save the table definition, and then click **Save**.
5. To load column definitions from the repository:
 - a. Click **Load**.
 - b. In the Table Definitions window, select the table definition to load, and click **OK**.
 - c. In the Select Columns window, use the arrow buttons to move columns from the **Available columns** list to the **Selected columns** list. Click **OK**.

What to do next

“Configuring stage properties” on page 5

Configuring stage and link properties

Every Data Masking stage job contains stages and links representing the flow of data. The links join the various stages in a job together and are used to specify how data flows when the job is run. The Data Masking stage job has an input link, output link and a reject link.

Configuring stage properties

When you create a Data Masking stage job, you can configure actions that you want to be performed when validation errors occur. Validation errors can include errors caused by invalid source data formats.

Before you begin

“Creating a data masking job” on page 3

Procedure

1. On the parallel canvas, double-click the Data Masking stage icon.
2. On the **Properties** tab, use the **Fail on Validation Error** field to specify how you want to handle validation errors. Selecting **Fail** aborts the job if validation errors occur, and **Continue** copies records to the reject link, when a reject link exists, or to the output link, when a reject link does not exist.
3. Optional: If you selected **Continue** in the previous step, then in the **Warning** field, select the options to log warning messages.
4. Click **OK** to save the changes.

Configuring the output link

The properties on an output link define the data to be read from a data source. When the data masking policy is applied to a column, the output link displays the applied policy. If you want to use the hash lookup data masking policy, you must configure data source connection properties and the usage properties in the output link.

Before you begin

“Creating a data masking job” on page 3

Procedure

1. On the parallel canvas, double-click the Data Masking stage icon.
2. Select the output link.
3. To configure a connection to the database:
 - a. On the **Properties** tab in the **Connectors** section, select the database.
 - b. Select **Variant**.
 - c. Specify details of the database that you want to connect to.
4. To configure reference table properties:
 - a. On the **Properties** tab in the **Usage Properties** section, select the **Source column for Hash Key generation** field, then select the source column from **Available columns**. The value from the specified column is used to generate a hash key.
 - b. In the **Table name** field, specify the table that you want to use for hash lookup.

- c. Optional: In the **Seed Value** field, specify a value. A seed value is used to generate a hash key value. The seed value must be an integer from 0 to 2,000,000,000. The default value is -1, which means that no seed is used.
 - d. In the **Hash key column name** field, specify the name of the hash key column in the reference table.
 - e. To add additional reference tables, right click one of the numbered tables, and select **Add Property Value**.
 - f. To delete a reference table, right click the numbered table you want to delete, and select **Remove Property Value**.
5. Click **OK** to save the connection information.

Configuring the reject link

If you create a job with the reject link, the records rejected due to validation errors are copied to the reject link.

Before you begin

“Creating a data masking job” on page 3

Procedure

1. On the parallel canvas, double-click the Data Masking stage icon.
2. Select the reject link.
3. On the **Reject** tab, select **ERRORCODE** or **ERRORTTEXT**, or both, in the **Add to reject row** section to specify the error code and the corresponding error message to describe the reason for rejection in the error message.
4. In the **Reject From Link** field, select the input link.
5. In the **Abort when** field, specify when you want to stop a job because of too many rejected rows.
6. Click **OK** to save.

Assigning a data masking policy to a column

Use the Masking Policy Editor to assign the data masking policies to the relevant columns.

Before you begin

“Creating a data masking job” on page 3

Procedure

1. On the parallel canvas, double-click the Data Masking stage icon.
2. Select the output link.
3. Select the **Columns** tab.
4. Click the **Masking Policy Editor** button. The Masking Policy Editor is displayed.
5. In the **Output Column** field, select the column whose data you want to mask.
6. In the **Masking Policy** option, select the required data masking policy.
7. In the **Masking Policy Options** section, configure the parameters for the data masking policy.
8. Click **OK** to save the changes.

What to do next

“Compiling and running data masking jobs”

Compiling and running data masking jobs

You must compile the Data Masking stage jobs into executable scripts that you can schedule and run.

Procedure

1. In the InfoSphere® DataStage® and QualityStage® Designer Client, open the Data Masking stage job that you want to compile.
2. Click the Compile icon.
3. If the **Compilation Status** area shows errors, edit the job to resolve the errors. After resolving the errors, click **Re-compile**.
4. When the job compiles successfully, click the Run icon, and specify the job run options:
 - a. Specify the job parameters as required.
 - b. Optional: Click **Validate** to verify if the job can run successfully.
 - c. Click **Run** to extract, convert, or write data.
5. To view the results of validating or running a job:
 - a. In the InfoSphere DataStage and QualityStage Designer Client, select **Tools > Run Director** to open the Director client
 - b. In the **Status** column, verify that the job was validated or completed successfully.
 - c. If the job or validation fails, select **View > Log** to identify any runtime problems.
6. If the job has runtime problems, fix the problems, recompile, validate (optional), and run the job until it completes successfully.

Setting up sample reference tables

The Data Masking stage includes sample reference data for hash lookup in a CSV file that you can import into the IBM® InfoSphere DataStage and QualityStage Designer Client.

About this task

You can use your own reference data for the hash lookup masking policy or set up the sample reference tables. The sample reference tables include the following data:

Address

Sample address data for Australia (AU), Canada (CA), Germany (DE), Spain (ES), France (FR), Italy (IT), Japan (JP), United Kingdom (UK), and United States of America (USA).

Name – First name, Last name

Sample name data includes first name, male first name, female first name, and last name for the supported countries.

Company name

Company name in English.

Personal Information

A set of data associated to a person in a record. For example, the personal

information for the USA contains information such as first name, last name, company name, national identification number, gender, phone number, birth date, and email address.

Procedure

1. Set up a database to store the reference data, and create an ODBC DSN for this database.
2. Uncompress the sample_reference_data.zip file on the engine tier machine.
3. Uncompress the setup_dsjobs.zip file on the client tier machine.
4. Import the setup_dsjobs.dsx file in the IBM InfoSphere DataStage and QualityStage Designer Client.
5. Compile and run all the imported jobs in the IBM InfoSphere DataStage and QualityStage Designer Client. The sample jobs create tables and store the reference data.

Data masking policies

The Data Masking stage provides a variety of predefined data masking policies.

Credit card number data masking policy

The credit card number data masking policy generates an appropriate mask for credit card numbers based on the source data. The Data Masking stage supports data masking for American Express, MasterCard, Visa, and Discover credit cards.

Supported data types

The credit card number masking policy can be applied to one of the following data types:

Table 1. Supported data types for credit card number data masking policy

SQLType	Extended	Length	Scale	Nullable	Note
Char		13 or longer	N/A	Yes or No	Cannot contain null characters.
Char	Unicode	13 or longer	N/A	Yes or No	Cannot contain null characters.
NChar		13 or longer	N/A	Yes or No	Cannot contain null characters.
VarChar		13 or longer	N/A	Yes or No	Cannot contain null characters.
VarChar	Unicode	13 or longer	N/A	Yes or No	Cannot contain null characters.
NVarChar		13 or longer	N/A	Yes or No	Cannot contain null characters.
BigInt	Unsigned	N/A	N/A	Yes or No	

Masking policy options

Mask Mode

Use one of the following options to specify modes of masking data:

Repeatable Masking

The first four digits of the credit card number are copied from the source to the output and the rest of the digits are masked. This type of masking is repeatable for data from the same source, regardless of the order.

Use 4 issuer digits

The first four digits of the credit card number are copied from the source to the output. The remaining part of the credit card number is appended with the masked account number and a check digit. A check digit is a digit added to a number that validates the authenticity of the number. When this option is used, different runs for the same input can result in different numbers. The uniqueness of the number is guaranteed only when the Data Masking stage job runs in the sequential mode or runs on one node.

Use 6 issuer digits

The first six digits of the credit card number are copied from the source to the output. The remaining part of the credit card number is appended with the masked account number and a check digit. When this option is used, different runs for the same input can result in different numbers. The uniqueness of the number is guaranteed only when the Data Masking stage job runs in the sequential mode or runs on one node.

Examples

The following examples show what the masked data might look like after the masking policy is applied. In these examples, the original value is 3400 1100 0000 063.

Table 2. Data masking examples for credit card number

Parameter	Example of masked data
Repeatable masking	3400 1065 4300 068
Use 4 issuer digits	3400 4100 0000 011
Use 6 issuer digits	3400 1165 4300 066

Email address data masking policy

The email address data masking policy generates an appropriate mask for source email addresses. You can mask the entire email address, only the user name, or only the domain name.

Supported Data Types

The email address data masking policy can be applied to one of the following data types:

Table 3. Supported data types for the email address data masking policy

SQLType	Extended	Length	Scale	Nullable
Char		3 or longer	N/A	Yes or No

Table 3. Supported data types for the email address data masking policy (continued)

SQLType	Extended	Length	Scale	Nullable
Char	Unicode	3 or longer	N/A	Yes or No
NChar		3 or longer	N/A	Yes or No
VarChar		3 or longer	N/A	Yes or No
VarChar	Unicode	3 or longer	N/A	Yes or No
NVarChar		3 or longer	N/A	Yes or No

Masking policy options

Mask Mode

Use one of the following options to specify modes of masking data:

All Masks the entire email address. This is the default option.

User name only

Masks only the user name. The domain name is copied from the source data.

Domain name only

Masks only the domain name. The user name is copied from the source data.

Domain Name

Specify the following information if you selected the **All** or **Domain name only** option for the **Mask Mode** option.

Domain Mask Mode

Select **Auto-generated domain name** to automatically generate the domain name. This is the default option.

Select **Selected from a list of domain names** to select the domain name from a list of large email service providers.

Seed Seed value in integer up to 31 digits.

Examples

The following examples show what the masked data might look like after the masking policy is applied. In these examples, the original value is lewis.maria@university.edu.

Table 4. Examples for masked email addresses

Parameter	Example of masked data
All	bkgfgohrnrtseqq85@bkgiplpsrhsl16.org
User name only	hhdighponbprc100@university.edu
Domain name only	Lewis.Maria@bkgiplpsrhsl16.org

US national ID data masking policy - National ID (US)

The national identification number for USA is the Social Security number. The US national ID data masking policy generates an appropriate mask for the Social Security numbers based on the source data.

The Social Security number is represented in AAA-GG-SSSS format, where, AAA indicates the three digit area code, GG indicates the two digit group code, and SSSS indicates the four digit serial number.

Supported data types

The National ID (US) data masking policy can be applied to columns of one of the following data types:

Table 5. Supported data types for National ID (US) data masking policy

SQLType	Extended	Length	Scale	Nullable	Note
Char		9 or longer	N/A	Yes or No	Cannot contain null characters. The length must be enough to contain the national ID with the selected separator.
Char	Unicode	9 or longer	N/A	Yes or No	Cannot contain null characters. The length must be enough to contain the national ID with the selected separator.
NChar		9 or longer	N/A	Yes or No	Cannot contain null characters. The length must be enough to contain the national ID with the selected separator.
VarChar		9 or longer	N/A	Yes or No	Cannot contain null characters. The length must be enough to contain the national ID with the selected separator.

Table 5. Supported data types for National ID (US) data masking policy (continued)

SQLType	Extended	Length	Scale	Nullable	Note
VarChar	Unicode	9 or longer	N/A	Yes or No	Cannot contain null characters. The length must be enough to contain the national ID with the selected separator.
NVarChar		9 or longer	N/A	Yes or No	Cannot contain null characters. The length must be enough to contain the national ID with the selected separator.
BigInt	Unsigned	N/A	N/A	Yes or No	The separator must be None.
Integer	Unsigned	N/A	N/A	Yes or No	The separator must be None.
Decimal		9	0	Yes or No	The separator must be None.

Masking policy options

Mask Mode

Use one of the following options to specify modes of masking data:

Repeatable masking

The result is always the same for different runs of the same source data. The source area number is copied without altering, while the group and serial number are masked.

Randomize area number

The result might be different each time the source data is processed. It generates a random source area number, and an appropriate group number. The uniqueness of the generated number is guaranteed only when the Data Masking stage is run in the sequential mode or is run on one node.

Separator

Use one of the following options to specify the output format of masked data:

Keep source format

To use the input format as the output format. This is the default option.

No separator

No separators are used in the output format.

DASH

To use the dash as a separator.

SPACE

To use the space as a separator.

DOT

To use the dot as a separator.

Examples

The following examples show what the masked data might look like after the masking policy is applied with specific formatting options selected. In these examples, the original value is 987654321.

Table 6. Data masking examples for National ID (US) masking policy

Separator	Example of masked data
No separators	867923415
Dash	867-92-3415
Space	867 92 3415
Dot	867.92. 4321

Canada national ID data masking policy - National ID (CA)

The national identification number for Canada is the Social Insurance Number. The National ID (CA) data masking policy generates a valid Canada Social Insurance Number based on the source data.

When this policy is used to mask data, the first three digits are copied from the source, and the remaining parts are masked. The result is always the same for different runs of the same data.

Supported data types

The National ID (CA) data masking policy can be applied to columns of one of the following data types:

Table 7. Supported data types for National ID (CA) data masking policy

SQLType	Extended	Length	Scale	Nullable	Note
Char		9 or longer	N/A	Yes or No	Cannot contain null characters. The length must be enough to contain the national ID with the selected separator.

Table 7. Supported data types for National ID (CA) data masking policy (continued)

SQLType	Extended	Length	Scale	Nullable	Note
Char	Unicode	9 or longer	N/A	Yes or No	Cannot contain null characters. The length must be enough to contain the national ID with the selected separator.
NChar		9 or longer	N/A	Yes or No	Cannot contain null characters. The length must be enough to contain the national ID with the selected separator.
VarChar		9 or longer	N/A	Yes or No	Cannot contain null characters. The length must be enough to contain the national ID with the selected separator.
VarChar	Unicode	9 or longer	N/A	Yes or No	Cannot contain null characters. The length must be enough to contain the national ID with the selected separator.
NVarChar		9 or longer	N/A	Yes or No	Cannot contain null characters. The length must be enough to contain the national ID with the selected separator.

Table 7. Supported data types for National ID (CA) data masking policy (continued)

SQLType	Extended	Length	Scale	Nullable	Note
BigInt	Unsigned	N/A	N/A	Yes or No	The separator must be None.
Integer	Unsigned	N/A	N/A	Yes or No	The separator must be None.
Decimal		9	0	Yes or No	The separator must be None.

Masking policy options

Separator

Use one of the following options to specify the output format of masked data:

Keep source format

To use the input format as the output format. This is the default option.

No separator

No separators are used in the output format.

DASH

To use the dash as a separator.

SPACE

To use the space as a separator.

DOT To use the dot as a separator.

Examples

The following examples show what the masked data might look like after the masking policy is applied with specific formatting options selected. In these examples, the original value is 987654321.

Table 8. Data masking examples for National ID (CA) masking policy

Separator	Example of masked data
No separators	987923415
Dash	987-923-415
Space	987 923 415
Dot	987.923.415

French national ID data masking policy - National ID (FR)

The national identification number for France is the French National Institute for Statistics and Economic Studies number. The National ID (FR) data masking policy generates a valid French National Institute for Statistics and Economic Studies number based on the source data.

The general format of French National Institute for Statistics and Economic Studies number is SYMMDDCCCOOK, where:

- S is the gender and citizenship information
- YY is the last two digits of the year of birth
- MM is the month of birth
- DD is the department of origin
- CCC is the commune of origin
- OOO is the order number
- KK is the control key or the check digit.

When the identification number is masked, the part containing the department of origin DD is copied from the source data, while the other parts are masked. The result is always the same for different runs of the same data.

Supported data types

The French National Institute for Statistics and Economic Studies number masking policy can be applied to columns of one of the following data types:

Table 9. Source Column

SQLType	Extended	Length	Scale	Nullable	Note
Char		15 or longer	N/A	Yes or No	Cannot contain null characters. The length must be enough to contain the national ID with the selected separator.
Char	Unicode	15 or longer	N/A	Yes or No	Cannot contain null characters. The length must be enough to contain the national ID with the selected separator.
NChar		15 or longer	N/A	Yes or No	Cannot contain null characters. The length must be enough to contain the national ID with the selected separator.

Table 9. Source Column (continued)

SQLType	Extended	Length	Scale	Nullable	Note
VarChar		15 or longer	N/A	Yes or No	Cannot contain null characters. The length must be enough to contain the national ID with the selected separator.
VarChar	Unicode	15 or longer	N/A	Yes or No	Cannot contain null characters. The length must be enough to contain the national ID with the selected separator.
NVarChar		15 or longer	N/A	Yes or No	Cannot contain null characters. The length must be enough to contain the national ID with the selected separator.

Masking policy options

Separator

Use one of the following options to specify the output format of masked data:

Keep source format

To use the input format as the output format. This is the default option.

No separators

No separators are used in the output format.

DASH

To use the dash as a separator.

SPACE

To use the space as a separator.

Examples

The following examples show what the masked data might look like after the masking policy is applied with specific formatting options selected. In these examples, the original value is 287091821012345.

Table 10. Data masking examples for National ID (FR) data masking policy

Separator	Example of masked data
No separators	150318378987654
Dash	1503183789876-54
Space	1503183789876 54

Italy national ID data masking policy - National ID (IT)

The national identification number for Italy is the Fiscal Code. The National ID (IT) data masking policy generates a valid Fiscal Code number based on the source data. When the Fiscal Code number is masked, the part containing the name is copied from the source and the other parts are masked. The result is always the same for different runs of the same source data.

The general format of the Italy Fiscal Code number is FFF-NNN-YYMDD-RRRRC, where:

- FFF is the encoded family name string
- NNN is the encoded first name string, YY is the year of birth
- M is an alphabet representing the month of birth
- DD is the day of birth
- RRRR is the region code
- C is the control character calculated from the first 15 characters

Supported data types

The Italian Fiscal Code data masking policy can be applied to columns of one of the following data types:

Table 11. Supported data types for National ID (IT) data masking policy

SQLType	Extended	Length	Scale	Nullable	Note
Char		16 or longer	N/A	Yes or No	Cannot contain null characters. The length must be enough to contain the national ID with the selected separator.

Table 11. Supported data types for National ID (IT) data masking policy (continued)

SQLType	Extended	Length	Scale	Nullable	Note
Char	Unicode	16 or longer	N/A	Yes or No	Cannot contain null characters. The length must be enough to contain the national ID with the selected separator.
NChar		16 or longer	N/A	Yes or No	Cannot contain null characters. The length must be enough to contain the national ID with the selected separator.
VarChar		16 or longer	N/A	Yes or No	Cannot contain null characters. The length must be enough to contain the national ID with the selected separator.
VarChar	Unicode	16 or longer	N/A	Yes or No	Cannot contain null characters. The length must be enough to contain the national ID with the selected separator.
NVarChar		16 or longer	N/A	Yes or No	Cannot contain null characters. The length must be enough to contain the national ID with the selected separator.

Masking policy options

Separator

Use one of the following options to specify the output format of masked data:

Keep source format

To use the input format as the output format. This is the default option.

No separator

No separators are used in the output format.

DASH

To use the dash as a separator.

SPACE

To use the space as a separator.

Examples

The following examples show what the masked data might look like after the masking policy is applied with specific formatting options selected. In these examples, the original value is ABCDEF12E34F567G.

Table 12. Data masking examples for National ID (IT) data masking policy

Separator	Example of masked data
No separators	EFGHAB34D12H789I
Dash	EFG-HAB-34D12-H789I
Space	EFG HAB 34D12 H789I

Spain national ID data masking policy - National ID (ES)

The national identification number for Spain is the Fiscal Identification number (NIF) or the Foreigner's Identification number (NIE). The Fiscal Identification number is given to citizens of Spain and the Foreign Identification number is given to foreign residents.

The general format of Fiscal Identification Number is SSSSSSS-A, where SSSSSSS is the seven digit serial number and A is the literal which is computed based on serial number. The general format of the Foreigner's Identification Number is X-SSSSSS-A, where X is the literal, SSSSSSS is the seven digit serial number, and A is the literal which is computed based on the serial number. When the Foreigner's Identification number is masked, the first literal is copied from the source data and the other parts are masked. When the Fiscal Identification number is masked, all parts are masked. The result is always the same for different runs of the same source data.

Supported data types

The Spain national identification number masking policy can be applied to columns of one of the following data types:

Table 13. Supported data types for National ID (ES) data masking policy

SQLType	Extended	Length	Scale	Nullable	Note
Char		8 or longer	N/A	Yes or No	Cannot contain null characters. The length must be enough to contain the national ID with the selected separator.
Char	Unicode	8 or longer	N/A	Yes or No	Cannot contain null characters. The length must be enough to contain the national ID with the selected separator.
NChar		8 or longer	N/A	Yes or No	Cannot contain null characters. The length must be enough to contain the national ID with the selected separator.
VarChar		8 or longer	N/A	Yes or No	Cannot contain null characters. The length must be enough to contain the national ID with the selected separator.
VarChar	Unicode	8 or longer	N/A	Yes or No	Cannot contain null characters. The length must be enough to contain the national ID with the selected separator.

Table 13. Supported data types for National ID (ES) data masking policy (continued)

SQLType	Extended	Length	Scale	Nullable	Note
NVarChar		8 or longer	N/A	Yes or No	Cannot contain null characters. The length must be enough to contain the national ID with the selected separator.

Masking policy options

Separator

Use one of the following options to specify the output format of masked data:

Keep source format

To use the input format as the output format. This is the default option.

No separator

No separators are used in the output format.

DASH

To use the dash as a separator.

SPACE

To use the space as a separator.

Examples

The following examples show what the masked data might look like after the masking policy is applied with specific formatting options selected. In these examples, the original value is 9876543L.

Table 14. Data masking examples for National ID (ES) data masking policy

Separator	Example of masked data
No separators	8679234L
Dash	8679234-L
Space	8679234 L

UK national ID data masking policy - National ID (UK)

The national identification number for UK is the National Insurance Number (NINO). The National ID (UK) data masking policy generates a valid National Insurance Number based on the source data.

The general format of UK National Insurance number is PP-NNNNNN-S, where:

- PP is the prefix pattern
- NNNNNN is a number between 000001 to 999999
- S is the suffix is limited to A, B, C, or D.

When the national identification number is masked, the prefix and the suffix are not be masked, and the other parts are masked.

Supported data types

The National ID (UK) data masking policy can be applied to columns of one of the following data types:

Table 15. Supported data types for National ID (UK) data masking policy

SQLType	Extended	Length	Scale	Nullable	Note
Char		8 or longer	N/A	Yes or No	Cannot contain null characters. The length must be enough to contain the national ID with the selected separator.
Char	Unicode	8 or longer	N/A	Yes or No	Cannot contain null characters. The length must be enough to contain the national ID with the selected separator.
NChar		8 or longer	N/A	Yes or No	Cannot contain null characters. The length must be enough to contain the national ID with the selected separator.
VarChar		8 or longer	N/A	Yes or No	Cannot contain null characters. The length must be enough to contain the national ID with the selected separator.

Table 15. Supported data types for National ID (UK) data masking policy (continued)

SQLType	Extended	Length	Scale	Nullable	Note
VarChar	Unicode	8 or longer	N/A	Yes or No	Cannot contain null characters. The length must be enough to contain the national ID with the selected separator.
NVarChar		8 or longer	N/A	Yes or No	Cannot contain null characters. The length must be enough to contain the national ID with the selected separator.

Masking policy options

Separator

Use one of the following options to specify the output format of masked data:

Keep source format

To use the input format as the output format. This is the default option.

No separators

No separators are used in the output format.

DASH

To use the dash as a separator.

SPACE

To use the space as a separator.

Separation Format

Select one of the following separation formats:

XX-123456-Y

Separates the output format into 3 parts.

XX-12-34-56-Y

Separates the output format into 5 parts.

Examples

The following examples show what the masked data might look like after the data masking policy is applied with specific formatting options selected. In these examples, the original value is AB987654C.

Table 16. Data masking examples for National ID (UK) data masking policy

Separator	Example of masked data
No separators	AB123456C
Dash	AB-123456-C, AB-12-34-56-C
Space	AB 123456 C, AB 12 34 56 C

Date age data masking policy

The date age data masking policy generates a new date based on the source data value. The date age data masking policy does not generate random dates.

Supported data types

The date age data masking policy can be applied to columns of one of the following data types:

Table 17. Supported data types for date age data masking policy

SQLType	Extended	Length	Scale	Nullable
Date		N/A	N/A	Yes or No
Timestamp		N/A	N/A	Yes or No
Timestamp	Microseconds	N/A	N/A	Yes or No

Masking policy options

Aging Amount

Use this option to specify integers to increment or decrement the year, month, week, or day. A positive number increments the age, and a negative number decrements the age.

Specific Year

Use this option to specify a year to replace the year in the source data. When this option is specified, the values specified in the **Aging Amount** option are disabled and ignored at runtime.

Repeatable replacement data masking policy

The repeatable replacement data masking policy masks source data in the format of the source data. You must always provide input data to the repeatable replacement data masking policy. You can use this data masking policy to convert keys such as the primary key or the foreign key.

In repeatable replacement data masking, capital letters are masked to random capital letters, lowercase letters are masked to random lowercase letters, and numbers are masked to random numbers. Any other character is copied to the output unchanged. For example, the string AB-123\$xyz might be masked to OW-159\$bgo. Other characters that appear in the input are copied to the output unchanged.

Supported data types

The repeatable replacement data masking policy can be applied to columns of one of the following data types:

Table 18. Supported data types for the repeatable replacement data masking policy

SQLType	Extended	Length	Scale	Nullable
Char		Any	N/A	Yes or No
Char	Unicode	Any	N/A	Yes or No
NChar		Any	N/A	Yes or No
VarChar		Any	N/A	Yes or No
VarChar	Unicode	Any	N/A	Yes or No
NVarChar		Any	N/A	Yes or No
BigInt		N/A	N/A	Yes or No
BigInt	Unsigned	N/A	N/A	Yes or No
Integer		N/A	N/A	Yes or No
Integer	Unsigned	N/A	N/A	Yes or No
SmallInt		N/A	N/A	Yes or No
SmallInt	Unsigned	N/A	N/A	Yes or No
TinyInt		N/A	N/A	Yes or No
TinyInt	Unsigned	N/A	N/A	Yes or No
Decimal		Any	Any	Yes or No
Float		N/A	N/A	Yes or No
Double		N/A	N/A	Yes or No

Masking policy options

Copy Specify the position and length numbers in the string. It replaces the string from the source data with the specified options. Multiple specifications must be made from left to right with no overlap. For example, "(1,2)(3,5)".

Number Mode

Select **Yes** or **No** to use the masking logic for numbers. If you select **Yes**, the masking logic for numbers is used even for strings.

Seed Specify seed literal in integer. This value is optional. If no value is specified, the default seed value is used.

Result of number mode option for different data types

Table 19. Result of number mode option for data types

Data Type	Number Mode	Description
TinyInt, SmallInt, Integer, BigInt, Float, Double	Yes	Result is the same as the result of character data types when Number Mode is set to Yes . Result might exceed the storage size of the data type when the most significant digit of the data type in the decimal expression is not zero. For example, the maximum value of Unsigned SmallInt is 65535. When input is greater than or equal to 10000, the result might exceed 65535. Use a data type that is capable of storing the result. The result is unique unless it exceeds the storage size of the data type.
TinyInt, SmallInt, Integer, BigInt, Float, Double	No	Result is unique and within the storage size of the data type.
Char, NChar, VarChar, NVarChar	Yes	Result is the same as the result of numeric data types when Number Mode is set to Yes . The input value must be a string expression of a numeric value without which, the Number Mode option is ignored. The result is unique.
Char, NChar, VarChar, NVarChar	No	Input can be any string. The result is unique.
Decimal	Yes or No	Result is the same as the result of character data types when Number Mode is set to Yes . The result is unique.

Random replacement data masking policy

The random replacement data masking policy masks source data in different formats for different runs of the source data.

In random replacement data masking, capital letters are masked to random capital letters, lowercase letters are masked to random lowercase letters, and numbers are masked to random numbers. Any other character is copied to the output unchanged. For example, the string AB-123\$xyz might be masked to OW-159\$bgo. However, the output generated for the same input will be different each time the data is generated.

Supported data types

The random replacement data masking policy can be applied to columns of one of the following data types:

Table 20. Supported data types for the random replacement data masking policy

SQL Type	Extended	Length	Scale	Nullable
Char		Any	N/A	Yes or No
Char	Unicode	Any	N/A	Yes or No
NChar		Any	N/A	Yes or No
VarChar		Any	N/A	Yes or No
VarChar	Unicode	Any	N/A	Yes or No
NVarChar		Any	N/A	Yes or No
BigInt		N/A	N/A	Yes or No
BigInt	Unsigned	N/A	N/A	Yes or No
Integer		N/A	N/A	Yes or No
Integer	Unsigned	N/A	N/A	Yes or No
SmallInt		N/A	N/A	Yes or No
SmallInt	Unsigned	N/A	N/A	Yes or No
TinyInt		N/A	N/A	Yes or No
TinyInt	Unsigned	N/A	N/A	Yes or No
Decimal		Any	Any	Yes or No
Float		N/A	N/A	Yes or No
Double		N/A	N/A	Yes or No

Masking policy options

Copy Specify the position and length numbers in the string. It replaces the string from the source data with the specified options. Multiple specifications must be made from left to right with no overlap. For example, "(1,2)(3,5)".

Hash data masking policy

The hash data masking policy generates an integer hash value that is based on the value of the source column that you specified. The output column can be different from the source column, but the input link must contain a column with the same name as the output column.

You can use the hash data masking policy instead of the hash lookup data masking policy to perform the hash lookup operations with a Lookup stage that is downstream from the data masking stage. Unlike the hash lookup data masking policy, the hash data masking policy does not access the reference table to get the maximum value for a hash key or to perform the lookup operation to replace data in the columns. To use the hash data masking policy to perform normal or sparse lookup operations, the Lookup stage must be downstream from the data masking stage.

Supported data types

The hash data masking policy can be applied to output columns of type Integer, SmallInt, or TinyInt. Data of type BigInt, decimal, numeric, real, double, float data types are not supported.

Table 21. Supported data types for output column for hash data masking policy

SQL Type	Extended	Length	Scale	Nullable	Note
TinyInt		N/A	N/A	Yes or No	An unsigned integer is not supported.
SmallInt		N/A	N/A	Yes or No	An unsigned integer is not supported.
Integer		N/A	N/A	Yes or No	An unsigned integer is not supported.

Table 22. Supported data types for source column for hash data masking policy

SQL Type	Extended	Length	Scale	Nullable	Note
Char		Any	N/A	Yes or No	
Char	Unicode	Any	N/A	Yes or No	
NChar		Any	N/A	Yes or No	
VarChar		Any	N/A	Yes or No	
VarChar	Unicode	Any	N/A	Yes or No	
NVarChar		Any	N/A	Yes or No	
BigInt		N/A	N/A	Yes or No	
BitInt	Unsigned	N/A	N/A	Yes or No	
Integer		N/A	N/A	Yes or No	
Integer	Unsigned	N/A	N/A	Yes or No	
SmallInt		N/A	N/A	Yes or No	
SmallInt	Unsigned	N/A	N/A	Yes or No	
TinyInt		N/A	N/A	Yes or No	
TinyInt	Unsigned	N/A	N/A	Yes or No	
Decimal		Any	Any	Yes or No	
Float		N/A	N/A	Yes or No	
Double		N/A	N/A	Yes or No	
Date		N/A	N/A	Yes or No	
Time		N/A	N/A	Yes or No	
Time	Microseconds	N/A	N/A	Yes or No	
Timestamp		N/A	N/A	Yes or No	
Timestamp	Microseconds	N/A	N/A	Yes or No	

Masking policy options

Source Column Name

Name of the source column on the input link from which the hash key value is calculated.

Maximum value

Specify any 32-bit signed integer value. The value of the generated hash key must be equal to or less than the specified value. If the output column is SmallInt (16-bit integer) or TinyInt (8-bit integer) and the specified maximum value is greater than the maximum value that is allowed for the data type, the value that you specify is replaced by the default maximum value when the job is run. The maximum value for the data type is 32767 for SmallInt and 127 for TinyInt.

Seed

The value specified here is used as a seed initialize hash key generation. Specify an integer value in the range 0 - 2000000000 as the seed value. The default value is -1, which means that no seed is used.

Hash lookup data masking policy

The hash lookup data masking policy masks input columns by using a reference table on a database. It calculates hash value based on the value of a source column and retrieves a record whose hash key column matches the hash value.

The Data Masking stage includes sample reference data for hash lookup. You can use your own reference data for the hash lookup masking policy or set up the sample reference tables.

When the hash lookup data masking policy is assigned to a column:

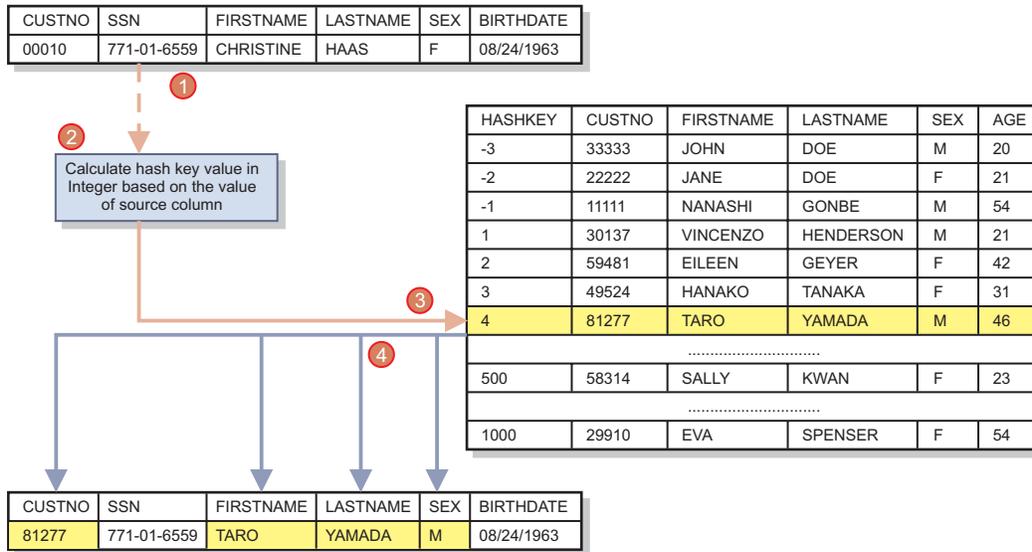
1. The value specified in the source column is read.
2. A hash key value is calculated for the selected value.
3. This hash key value is internally used to lookup on the reference tables, and to locate the matching record.
4. The value specified in the relevant column of the matched record is retrieved and copied to the output link.

If the source value contains either a zero-length variable character, a space, or a NULL, the following negative value is used as the hash key:

- -3 for NULL
- -2 for all spaces
- -1 for zero-length variable character

To use the hash lookup data masking policy, you must configure the database connection information in the output link. This version of Data Masking stage supports DB2, Oracle, and ODBC databases. The hash lookup masking policy also requires the reference table name to be specified in the output link property. In order to establish association of reference table options specified in the output link and the column, you must ensure that the table name specified for the **Reference Table name** option in the hash lookup data masking policy and the table name specified in the output link match.

The following figure illustrates how the hash lookup data masking policy works:



Supported data types

The hash lookup data masking policy can be applied to any data types, but it should match with the data type of associated column in the reference table. The source column for hash key generation can be one of the following data types:

Table 23. Supported data types for source column for hash key generation

SQL Type	Extended	Length	Scale	Nullable
Char		Any	N/A	Yes or No
Char	Unicode	Any	N/A	Yes or No
NChar		Any	N/A	Yes or No
VarChar		Any	N/A	Yes or No
VarChar	Unicode	Any	N/A	Yes or No
NVarChar		Any	N/A	Yes or No
BigInt		N/A	N/A	Yes or No
BigInt	Unsigned	N/A	N/A	Yes or No
Integer		N/A	N/A	Yes or No
Integer	Unsigned	N/A	N/A	Yes or No
SmallInt		N/A	N/A	Yes or No
SmallInt	Unsigned	N/A	N/A	Yes or No
TinyInt		N/A	N/A	Yes or No
TinyInt	Unsigned	N/A	N/A	Yes or No
Decimal		Any	Any	Yes or No
Float		N/A	N/A	Yes or No
Double		N/A	N/A	Yes or No
Date		N/A	N/A	Yes or No
Time		N/A	N/A	Yes or No
Time	Microseconds	N/A	N/A	Yes or No
Timestamp		N/A	N/A	Yes or No
Timestamp	Microseconds	N/A	N/A	Yes or No

Masking policy options

Reference Table Name

The name of the reference table that is specified in the output link.

Column Name in Reference Table

The name of the column that is specified in the reference table.

Appendix A. Product accessibility

You can get information about the accessibility status of IBM products.

The IBM InfoSphere Information Server product modules and user interfaces are not fully accessible.

For information about the accessibility status of IBM products, see the IBM product accessibility information at http://www.ibm.com/able/product_accessibility/index.html.

Accessible documentation

Accessible documentation for InfoSphere Information Server products is provided in IBM Knowledge Center. IBM Knowledge Center presents the documentation in XHTML 1.0 format, which is viewable in most web browsers. Because IBM Knowledge Center uses XHTML, you can set display preferences in your browser. This also allows you to use screen readers and other assistive technologies to access the documentation.

The documentation that is in IBM Knowledge Center is also provided in PDF files, which are not fully accessible.

IBM and accessibility

See the IBM Human Ability and Accessibility Center for more information about the commitment that IBM has to accessibility.

Appendix B. Reading command-line syntax

This documentation uses special characters to define the command-line syntax.

The following special characters define the command-line syntax:

- [] Identifies an optional argument. Arguments that are not enclosed in brackets are required.
- ... Indicates that you can specify multiple values for the previous argument.
- | Indicates mutually exclusive information. You can use the argument to the left of the separator or the argument to the right of the separator. You cannot use both arguments in a single use of the command.
- { } Delimits a set of mutually exclusive arguments when one of the arguments is required. If the arguments are optional, they are enclosed in brackets ([]).

Note:

- The maximum number of characters in an argument is 256.
- Enclose argument values that have embedded spaces with either single or double quotation marks.

For example:

```
wsetsrc[-S server] [-l label] [-n name] source
```

The *source* argument is the only required argument for the **wsetsrc** command. The brackets around the other arguments indicate that these arguments are optional.

```
wlsac [-l | -f format] [key... ] profile
```

In this example, the **-l** and **-f** format arguments are mutually exclusive and optional. The *profile* argument is required. The *key* argument is optional. The ellipsis (...) that follows the *key* argument indicates that you can specify multiple key names.

```
wrb -import {rule_pack | rule_set}...
```

In this example, the *rule_pack* and *rule_set* arguments are mutually exclusive, but one of the arguments must be specified. Also, the ellipsis marks (...) indicate that you can specify multiple rule packs or rule sets.

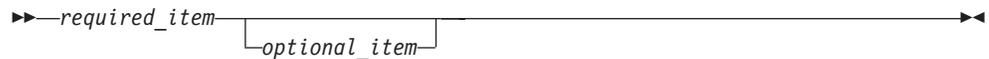
Appendix C. How to read syntax diagrams

The following rules apply to the syntax diagrams that are used in this information:

- Read the syntax diagrams from left to right, from top to bottom, following the path of the line. The following conventions are used:
 - The >>--- symbol indicates the beginning of a syntax diagram.
 - The ---> symbol indicates that the syntax diagram is continued on the next line.
 - The >--- symbol indicates that a syntax diagram is continued from the previous line.
 - The --->< symbol indicates the end of a syntax diagram.
- Required items appear on the horizontal line (the main path).



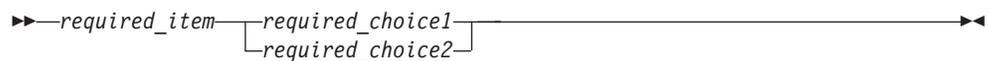
- Optional items appear below the main path.



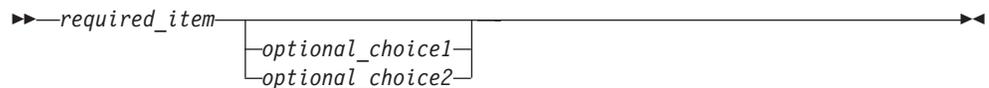
If an optional item appears above the main path, that item has no effect on the execution of the syntax element and is used only for readability.



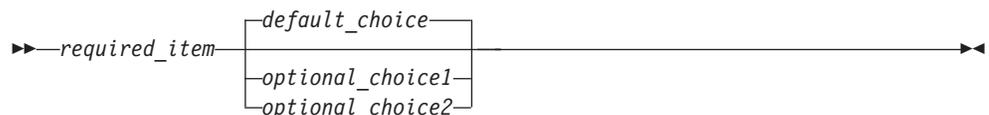
- If you can choose from two or more items, they appear vertically, in a stack. If you must choose one of the items, one item of the stack appears on the main path.



If choosing one of the items is optional, the entire stack appears below the main path.



If one of the items is the default, it appears above the main path, and the remaining choices are shown below.



- An arrow returning to the left, above the main line, indicates an item that can be repeated.



If the repeat arrow contains a comma, you must separate repeated items with a comma.



A repeat arrow above a stack indicates that you can repeat the items in the stack.

- Sometimes a diagram must be split into fragments. The syntax fragment is shown separately from the main syntax diagram, but the contents of the fragment should be read as if they are on the main path of the diagram.



Fragment-name:



- Keywords, and their minimum abbreviations if applicable, appear in uppercase. They must be spelled exactly as shown.
- Variables appear in all lowercase italic letters (for example, *column-name*). They represent user-supplied names or values.
- Separate keywords and parameters by at least one space if no intervening punctuation is shown in the diagram.
- Enter punctuation marks, parentheses, arithmetic operators, and other symbols, exactly as shown in the diagram.
- Footnotes are shown by a number in parentheses, for example (1).

Appendix D. Contacting IBM

You can contact IBM for customer support, software services, product information, and general information. You also can provide feedback to IBM about products and documentation.

The following table lists resources for customer support, software services, training, and product and solutions information.

Table 24. IBM resources

Resource	Description and location
IBM Support Portal	You can customize support information by choosing the products and the topics that interest you at www.ibm.com/support/entry/portal/Software/Information_Management/InfoSphere_Information_Server
Software services	You can find information about software, IT, and business consulting services, on the solutions site at www.ibm.com/businesssolutions/
My IBM	You can manage links to IBM Web sites and information that meet your specific technical support needs by creating an account on the My IBM site at www.ibm.com/account/
Training and certification	You can learn about technical training and education services designed for individuals, companies, and public organizations to acquire, maintain, and optimize their IT skills at http://www.ibm.com/training
IBM representatives	You can contact an IBM representative to learn about solutions at www.ibm.com/connect/ibm/us/en/

Appendix E. Accessing the product documentation

Documentation is provided in a variety of formats: in the online IBM Knowledge Center, in an optional locally installed information center, and as PDF books. You can access the online or locally installed help directly from the product client interfaces.

IBM Knowledge Center is the best place to find the most up-to-date information for InfoSphere Information Server. IBM Knowledge Center contains help for most of the product interfaces, as well as complete documentation for all the product modules in the suite. You can open IBM Knowledge Center from the installed product or from a web browser.

Accessing IBM Knowledge Center

There are various ways to access the online documentation:

- Click the **Help** link in the upper right of the client interface.
- Press the F1 key. The F1 key typically opens the topic that describes the current context of the client interface.

Note: The F1 key does not work in web clients.

- Type the address in a web browser, for example, when you are not logged in to the product.

Enter the following address to access all versions of InfoSphere Information Server documentation:

```
http://www.ibm.com/support/knowledgecenter/SSZJPZ/
```

If you want to access a particular topic, specify the version number with the product identifier, the documentation plug-in name, and the topic path in the URL. For example, the URL for the 11.3 version of this topic is as follows. (The ⇒ symbol indicates a line continuation):

```
http://www.ibm.com/support/knowledgecenter/SSZJPZ_11.3.0/⇒  
com.ibm.swg.im.iis.common.doc/common/accessingiidoc.html
```

Tip:

The knowledge center has a short URL as well:

```
http://ibm.biz/knowctr
```

To specify a short URL to a specific product page, version, or topic, use a hash character (#) between the short URL and the product identifier. For example, the short URL to all the InfoSphere Information Server documentation is the following URL:

```
http://ibm.biz/knowctr#SSZJPZ/
```

And, the short URL to the topic above to create a slightly shorter URL is the following URL (The ⇒ symbol indicates a line continuation):

```
http://ibm.biz/knowctr#SSZJPZ_11.3.0/com.ibm.swg.im.iis.common.doc/⇒  
common/accessingiidoc.html
```

Changing help links to refer to locally installed documentation

IBM Knowledge Center contains the most up-to-date version of the documentation. However, you can install a local version of the documentation as an information center and configure your help links to point to it. A local information center is useful if your enterprise does not provide access to the internet.

Use the installation instructions that come with the information center installation package to install it on the computer of your choice. After you install and start the information center, you can use the **iisAdmin** command on the services tier computer to change the documentation location that the product F1 and help links refer to. (The `⇒` symbol indicates a line continuation):

Windows

```
IS_install_path\ASBServer\bin\iisAdmin.bat -set -key ⇒  
com.ibm.iis.infocenter.url -value http://<host>:<port>/help/topic/
```

AIX® Linux

```
IS_install_path/ASBServer/bin/iisAdmin.sh -set -key ⇒  
com.ibm.iis.infocenter.url -value http://<host>:<port>/help/topic/
```

Where `<host>` is the name of the computer where the information center is installed and `<port>` is the port number for the information center. The default port number is 8888. For example, on a computer named `server1.example.com` that uses the default port, the URL value would be `http://server1.example.com:8888/help/topic/`.

Obtaining PDF and hardcopy documentation

- The PDF file books are available online and can be accessed from this support document: <https://www.ibm.com/support/docview.wss?uid=swg27008803&wv=1>.
- You can also order IBM publications in hardcopy format online or through your local IBM representative. To order publications online, go to the IBM Publications Center at <http://www.ibm.com/e-business/linkweb/publications/servlet/pbi.wss>.

Appendix F. Providing feedback on the product documentation

You can provide helpful feedback regarding IBM documentation.

Your feedback helps IBM to provide quality information. You can use any of the following methods to provide comments:

- To provide a comment about a topic in IBM Knowledge Center that is hosted on the IBM website, sign in and add a comment by clicking **Add Comment** button at the bottom of the topic. Comments submitted this way are viewable by the public.
- To send a comment about the topic in IBM Knowledge Center to IBM that is not viewable by anyone else, sign in and click the **Feedback** link at the bottom of IBM Knowledge Center.
- Send your comments by using the online readers' comment form at www.ibm.com/software/awdtools/rcf/.
- Send your comments by e-mail to comments@us.ibm.com. Include the name of the product, the version number of the product, and the name and part number of the information (if applicable). If you are commenting on specific text, include the location of the text (for example, a title, a table number, or a page number).

Notices and trademarks

This information was developed for products and services offered in the U.S.A. This material may be available from IBM in other languages. However, you may be required to own a copy of the product or product version in that language in order to access it.

Notices

IBM may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not grant you any license to these patents. You can send license inquiries, in writing, to:

IBM Director of Licensing
IBM Corporation
North Castle Drive
Armonk, NY 10504-1785 U.S.A.

For license inquiries regarding double-byte character set (DBCS) information, contact the IBM Intellectual Property Department in your country or send inquiries, in writing, to:

Intellectual Property Licensing
Legal and Intellectual Property Law
IBM Japan Ltd.
19-21, Nihonbashi-Hakozakicho, Chuo-ku
Tokyo 103-8510, Japan

The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law:

INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this IBM product and use of those Web sites is at your own risk.

IBM may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you.

Licensees of this program who wish to have information about it for the purpose of enabling: (i) the exchange of information between independently created programs and other programs (including this one) and (ii) the mutual use of the information which has been exchanged, should contact:

IBM Corporation
J46A/G4
555 Bailey Avenue
San Jose, CA 95141-1003 U.S.A.

Such information may be available, subject to appropriate terms and conditions, including in some cases, payment of a fee.

The licensed program described in this document and all licensed material available for it are provided by IBM under terms of the IBM Customer Agreement, IBM International Program License Agreement or any equivalent agreement between us.

Any performance data contained herein was determined in a controlled environment. Therefore, the results obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurements may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

All statements regarding IBM's future direction or intent are subject to change or withdrawal without notice, and represent goals and objectives only.

This information is for planning purposes only. The information herein is subject to change before the products described become available.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

COPYRIGHT LICENSE:

This information contains sample application programs in source language, which illustrate programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to IBM, for the purposes of developing, using, marketing or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. IBM, therefore, cannot guarantee or imply reliability, serviceability, or function of these programs. The sample programs are provided "AS IS", without warranty of any kind. IBM shall not be liable for any damages arising out of your use of the sample programs.

Each copy or any portion of these sample programs or any derivative work, must include a copyright notice as follows:

© (your company name) (year). Portions of this code are derived from IBM Corp. Sample Programs. © Copyright IBM Corp. _enter the year or years_. All rights reserved.

If you are viewing this information softcopy, the photographs and color illustrations may not appear.

Privacy policy considerations

IBM Software products, including software as a service solutions, ("Software Offerings") may use cookies or other technologies to collect product usage information, to help improve the end user experience, to tailor interactions with the end user or for other purposes. In many cases no personally identifiable information is collected by the Software Offerings. Some of our Software Offerings can help enable you to collect personally identifiable information. If this Software Offering uses cookies to collect personally identifiable information, specific information about this offering's use of cookies is set forth below.

Depending upon the configurations deployed, this Software Offering may use session or persistent cookies. If a product or component is not listed, that product or component does not use cookies.

Table 25. Use of cookies by InfoSphere Information Server products and components

Product module	Component or feature	Type of cookie that is used	Collect this data	Purpose of data	Disabling the cookies
Any (part of InfoSphere Information Server installation)	InfoSphere Information Server web console	<ul style="list-style-type: none"> • Session • Persistent 	User name	<ul style="list-style-type: none"> • Session management • Authentication 	Cannot be disabled
Any (part of InfoSphere Information Server installation)	InfoSphere Metadata Asset Manager	<ul style="list-style-type: none"> • Session • Persistent 	No personally identifiable information	<ul style="list-style-type: none"> • Session management • Authentication • Enhanced user usability • Single sign-on configuration 	Cannot be disabled

Table 25. Use of cookies by InfoSphere Information Server products and components (continued)

Product module	Component or feature	Type of cookie that is used	Collect this data	Purpose of data	Disabling the cookies
InfoSphere DataStage	Big Data File stage	<ul style="list-style-type: none"> • Session • Persistent 	<ul style="list-style-type: none"> • User name • Digital signature • Session ID 	<ul style="list-style-type: none"> • Session management • Authentication • Single sign-on configuration 	Cannot be disabled
InfoSphere DataStage	XML stage	Session	Internal identifiers	<ul style="list-style-type: none"> • Session management • Authentication 	Cannot be disabled
InfoSphere DataStage	IBM InfoSphere DataStage and QualityStage Operations Console	Session	No personally identifiable information	<ul style="list-style-type: none"> • Session management • Authentication 	Cannot be disabled
InfoSphere Data Click	InfoSphere Information Server web console	<ul style="list-style-type: none"> • Session • Persistent 	User name	<ul style="list-style-type: none"> • Session management • Authentication 	Cannot be disabled
InfoSphere Data Quality Console		Session	No personally identifiable information	<ul style="list-style-type: none"> • Session management • Authentication • Single sign-on configuration 	Cannot be disabled
InfoSphere QualityStage Standardization Rules Designer	InfoSphere Information Server web console	<ul style="list-style-type: none"> • Session • Persistent 	User name	<ul style="list-style-type: none"> • Session management • Authentication 	Cannot be disabled
InfoSphere Information Governance Catalog		<ul style="list-style-type: none"> • Session • Persistent 	<ul style="list-style-type: none"> • User name • Internal identifiers • State of the tree 	<ul style="list-style-type: none"> • Session management • Authentication • Single sign-on configuration 	Cannot be disabled
InfoSphere Information Analyzer	Data Rules stage in the InfoSphere DataStage and QualityStage Designer client	Session	Session ID	Session management	Cannot be disabled

If the configurations deployed for this Software Offering provide you as customer the ability to collect personally identifiable information from end users via cookies and other technologies, you should seek your own legal advice about any laws applicable to such data collection, including any requirements for notice and consent.

For more information about the use of various technologies, including cookies, for these purposes, see IBM's Privacy Policy at <http://www.ibm.com/privacy> and IBM's Online Privacy Statement at <http://www.ibm.com/privacy/details> the section entitled "Cookies, Web Beacons and Other Technologies" and the "IBM Software Products and Software-as-a-Service Privacy Statement" at <http://www.ibm.com/software/info/product-privacy>.

Trademarks

IBM, the IBM logo, and [ibm.com](http://www.ibm.com)[®] are trademarks or registered trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the Web at www.ibm.com/legal/copytrade.shtml.

The following terms are trademarks or registered trademarks of other companies:

Adobe is a registered trademark of Adobe Systems Incorporated in the United States, and/or other countries.

Intel and Itanium are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both.

Microsoft, Windows and Windows NT are trademarks of Microsoft Corporation in the United States, other countries, or both.

UNIX is a registered trademark of The Open Group in the United States and other countries.

Java[™] and all Java-based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates.

The United States Postal Service owns the following trademarks: CASS, CASS Certified, DPV, LACS^{Link}, ZIP, ZIP + 4, ZIP Code, Post Office, Postal Service, USPS and United States Postal Service. IBM Corporation is a non-exclusive DPV and LACS^{Link} licensee of the United States Postal Service.

Other company, product or service names may be trademarks or service marks of others.

Index

C

- Canada Social Insurance Number
 - masking 13
- command-line syntax
 - conventions 35
- commands
 - syntax 35
- compiling and running data stage jobs 7
- configuring stage properties for data masking 5
- configuring the reject link 6
- creating a Data Masking stage job 3
- credit card number masking 8
- customer support
 - contacting 39

D

- data masking 1, 6, 8, 9, 11, 13, 15, 18, 20, 22, 25, 27, 30
- data masking for hash lookup reference table 30
- data masking job 7
- Data Masking pack 1
- data masking policy 6, 8, 9, 11, 13, 15, 18, 20, 22, 25, 27
- Data Masking stage job 3
- Data Masking stage reject link 6
- data privacy 6, 8, 9, 11, 13, 15, 18, 20, 22, 25, 27, 30
- date age masking 6, 25, 27
- designing data stage jobs 2

E

- email address masking 9

H

- hash lookup
 - masking policy 30
 - reference table 7

I

- InfoSphere DataStage Pack for Data Masking 1
- Institute for Statistics and Economic Studies (INSEE) number masking 15
- insurance number masking 22
- invalid records 6
- Italy Fiscal code number masking 18

L

- legal notices 45
- link properties for data masking 5

M

- masking credit card number 8, 11, 13
- masking date age 6, 25, 27
- masking email address 9
- masking Institute for Statistics and Economic Studies (INSEE) 15
- masking Italy Fiscal code number 18
- masking Spain national ID 20
- masking UK National Insurance number 22

P

- Pack for Data Masking 1
- product accessibility
 - accessibility 33
- product documentation
 - accessing 41

R

- repeatable replacement masking 25

S

- sample reference table 7
- setting up column definitions 4
- software services
 - contacting 39
- Spain national ID masking 20
- special characters
 - in command-line syntax 35
- stage properties for data masking 5
- support
 - customer 39
- syntax
 - command-line 35

T

- trademarks
 - list of 45

U

- UK National 22
- US Social Security Number masking 11

W

- web sites
 - non-IBM 37



Printed in USA

SC19-4281-00

